Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs

Clare D. Marsden[a,1], Diego Ortega-Del Vecchyo[b,1], Dennis P. O'Brien[c], Jeremy F. Taylor[d], Oscar Ramirez[e], Carles Vilà[f], Tomas Marques-Bonet[e,g], Robert D. Schnabel[d,h], Robert K. Wayne[a], and Kirk E. Lohmueller[a,b,i,2]

[a]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095; [b]Interdepartmental Program in Bioinformatics, University of California, Los Angeles, CA 90095; [c]Department of Veterinary Medicine and Surgery, University of Missouri, Columbia, MO 65211; [d]Division of Animal Sciences, University of Missouri, Columbia, MO 65211; [e]Institut Catala de Recerca i Estudis Avançats, Institut de Biologia Evolutiva (Centro Superior de Investigaciones Cientificas-Universitat Pompeu Fabra), 08003 Barcelona, Spain; [f]Conservation and Evolutionary Genetics Group, Estación Biológica de Doñana-Consejo Superior de Investigaciones Cientificas, 41092, Seville, Spain; [g]Centro Nacional Analasis Genomico, 08023, Barcelona, Spain; [h]Informatics Institute, University of Missouri, Columbia, MO 65211; and [i]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA 90095

Edited by Montgomery Slatkin, University of California, Berkeley, CA, and approved November 16, 2015 (received for review June 25, 2015)

Many of the mutations that arise in genomes are weakly deleterious and reduce fitness but are not always eliminated from the population by purifying natural selection. Consequently, understanding the reasons why deleterious mutations persist in populations and the role of demographic history in this process is of considerable interest (1–9). The radiation of domestic dogs offers a unique opportunity to address these questions. Dogs were originally domesticated from ancestral gray wolf populations >15,000 y ago in a process involving one or more severe population bottlenecks (10–12). The more recent isolation of modern dog breeds, which occurred over the last 300 y, involved additional population bottlenecks, intense artificial selection, and inbreeding (refs. 11 and 13–15; Fig. 1*A*). Although this history is predicted to have resulted in the accumulation of deleterious variants, its specific effect on genome-wide patterns of deleterious variation remains unclear.

Here, we use complete genome sequencing data from 46 dogs representing 34 breeds, 25 village dogs, and 19 wolves to directly examine patterns of deleterious genetic variation across the dog genome (Dataset S1). Because more than half of these data derive from our own sequencing efforts, this project represents the largest survey of dog genetic diversity based on genome sequences to date. Overall, we find that population bottlenecks associated with domestication have resulted in a proportional increase of amino acid changing variants in dogs relative to wolves and also have led to an increase in the additive genetic load in dogs relative to wolves. We also find an enrichment of amino acid changing variants surrounding regions of the genome that have been targeted by selective sweeps, suggesting that deleterious variants have increased in frequency because of hitchhiking with nearby positively selected variants. Finally, Mendelian disease genes are enriched in sweep regions, suggesting a link between disease and traits under strong artificial selection. Taken together, our results indicate that the domestication process has dramatically reshaped patterns of deleterious variation across the dog genome.

deleterious mutations | domestication | bottleneck | selective sweep

Results and Discussion

Description of the Data. Using a combination of in-house generated data ($n = 50$) and published sequences ($n = 40$; refs. 16–18), we collated a dataset of 90 canid whole genomes representing 46 breed dogs, 25 village dogs, 19 gray wolves, and a single genome from a golden jackal to polarize ancestral and derived states (Dataset S1). Our analyses focused on patterns of genetic diversity at putatively neutral sites far from genes (*SI Appendix, SI Materials and Methods*), fourfold degenerate sites (nonamino acid changing coding variants), and zerofold degenerate sites (amino acid changing coding variants).

Significance

Dogs have an integral role in human society, and recent evidence suggests they have a unique bond that elicits a beneficial hormonal response in both dogs and human handlers. Here, we show this relationship has a dark side. Small population size during domestication and strong artificial selection for breed-defining traits has unintentionally increased the numbers of deleterious genetic variants. Our findings question the overly typological practice of breeding individuals that best fit breed standards, a Victorian legacy. This practice does not allow selection to remove potentially deleterious variation associated with genes responsible for breed-specific traits.

Author contributions: C.D.M., D.O.-D.V., D.P.O., J.F.T., O.R., C.V., T.M.-B., R.D.S., R.K.W., and K.E.L. designed research; C.D.M., D.O.-D.V., D.P.O., J.F.T., O.R., C.V., T.M.-B., R.D.S., and K.E.L. performed research; D.P.O., J.F.T., O.R., C.V., T.M.-B., R.D.S., R.K.W., and K.E.L. contributed new reagents/analytic tools; C.D.M., D.O.-D.V., and K.E.L. analyzed data; and C.D.M., D.O.-D.V., R.K.W., and K.E.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Database deposition: NCBI Sequence Read Archive accessions for previously unpublished genomes are in Dataset S1. The data reported in this paper have been deposited in the Dryad Digital Repository, datadryad.org (doi:10.5061/dryad.012s5).

[1]C.D.M. and D.O.-D.V. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: klohmueller@ucla.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1512501113/-/DCSupplemental.

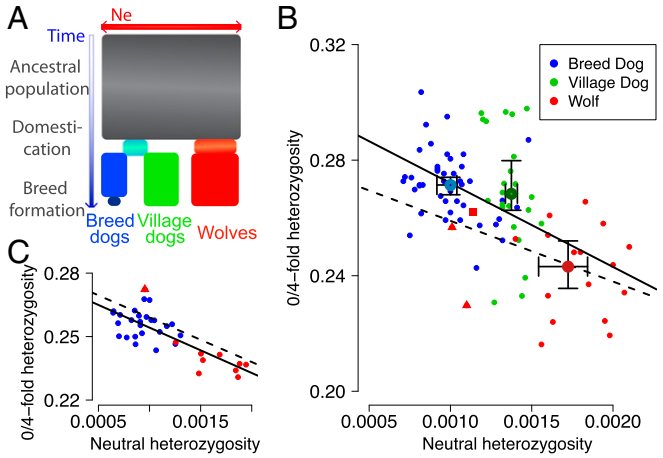

Fig. 1. Population history and deleterious genetic variation. (*A*) Conceptual model of dog domestication used in population genetic simulations. Box widths are proportional to estimated population sizes (*SI Appendix*, Table S4). (*B*) The ratio of zerofold to fourfold heterozygosity vs. neutral genetic diversity. Observed heterozygosity is based on four reads per individual. The larger circles represent the trimmed median values for each population group, and the error bars denote 95% confidence intervals on the trimmed median for each population group. Triangles denote the Tibetan wolves. A square denotes the Isle Royale wolf. The solid black line denotes the best-fit linear regression line (Iintercept = 0.301, slope = –29.00, $r = -0.534$, $P < 6 \times 10^{-8}$). The dashed line denotes the best-fit linear regression line from forward simulations of demography and negative selection (*SI Appendix*, Tables S4 and S7). (*C*) The ratio of zerofold to fourfold heterozygosity vs. neutral genetic diversity in the 35 high-coverage genomes where genotypes were called using GATK. The solid black line denotes the best-fit linear regression line (intercept = 0.276, slope = –21.43, $r = -0.777$, $P < 5 \times 10^{-8}$), and the dashed line is as described in *B*.

We divided our dataset into two groups based on sequencing coverage. The first group contains the subset of genomes with high sequencing coverage (>15×) comprising 25 breed dogs and 10 wolves. For this dataset, we called individual genotypes using the Genome Analysis Toolkit (GATK, ref. 19). The second group consists of all 90 canid genomes. Many of these genomes have low sequence depth where genotype calls are less reliable. For these data, we estimated per individual heterozygosity (i.e., average pairwise differences between sequences) using a maximum likelihood approach based directly on sequence reads (*Materials and Methods*). To assess the performance of this method, we compared our read-based estimates of heterozygosity to those from genotypes called using GATK (19) on a subset of high-coverage genomes. We found the two estimates of heterozygosity to be highly concordant, suggesting that our estimator performs well (*SI Appendix, SI Text* and Fig. S1). Importantly, because our read-based estimator was applied to subsamples of only four reads per individual, it is appropriate even for the lower-coverage genomes. Estimates of heterozygosity are not affected by sequencing coverage (*SI Appendix*, Fig. S2). Comparison of the high-coverage data to genotype array data shows negligible batch effects, a low-false discovery rate (~1%), and a false-negative rate of <8% (*SI Appendix, SI Text* and Fig. S3).

Genome-Wide Patterns of Deleterious Variation. Because we typically have only 1–2 genomes per breed or population, we first focus on patterns of heterozygosity. To evaluate the role of population size in affecting deleterious variation, we calculate the ratio of zerofold to fourfold heterozygosity (20–22). This ratio is an estimate of the proportion of amino acid changing mutations that are not removed by selection. Assuming constant selection coefficients across populations, changes in this ratio indicate that demographic effects modulate the efficacy of selection. We chose this metric because it quantifies how demography affects selection without estimating parameters in complex demographic models for all populations (21, 22).

In our data, the ratio of zerofold heterozygosity to fourfold heterozygosity shows a strong negative correlation with levels of neutral heterozygosity (Pearson's $r = -0.534$, $P < 6 \times 10^{-8}$, Fig. 1 *B* and *C* and *SI Appendix*, Fig. S4*A* and Table S1). Breed dogs have lower levels of neutral heterozygosity than wolves, consistent with their bottlenecked demographic history. However, they show disproportionately higher levels of amino acid (zerofold) heterozygosity (Fig. 1*B*). This result is concordant with previous estimates based on more limited data (a single boxer genome and mtDNA data; refs. 23 and 24) and suggests that the proportional elevation in deleterious amino acid variation in dogs relative to wolves is seen across a range of breeds. Much of this pattern is driven by the difference between breed dogs and wolves. It diminishes when analyzing them separately (*SI Appendix*, Fig. S4*B*), although statistical power also is reduced. Patterns of neutral heterozygosity in the village dogs fall between those of breed dogs and wolves, consistent with their intermediate effective population size and variable levels of admixture between modern and ancient breeds (25). However, the ratio of zerofold to fourfold heterozygosity in village dogs depends to some degree on the filters used and is either similar to that in breed dogs or intermediate to that of dogs and wolves (*SI Appendix*, Fig. S4*C*). Interestingly, several wolf populations appear to show lower levels of neutral heterozygosity and higher ratios of zerofold to fourfold heterozygosity than breed dogs. These include the Tibetan wolves, which were previously shown to have low genetic diversity (18), and the Isle Royale wolf, which is a highly inbred island population derived from two founders in the 1950s (26). The negative correlation in Fig. 1 is unlikely to be driven by hypermutable CpG sites (*SI Appendix*, Fig. S4*C*) or regions affected by selective sweeps (*SI Appendix*, Fig. S4*D*), because it persists after removing these genomic features. Further, dogs still show an elevated zerofold/fourfold ratio compared with wolves when accounting for the shared genealogical history of different individuals (*SI Appendix*, Fig. S5).

Although no individual gene showed a significantly higher zerofold/fourfold ratio in dogs relative to wolves after correction for multiple testing (*SI Appendix*, Table S2), the von Willebrand factor (*VWF*) gene had the highest ratio (*SI Appendix*, Table S2). *VWF* has been implicated in bleeding disorders in breed dogs (27), suggesting the increased level of amino acid changes in this gene may be of relevance to health.

To test whether our observed patterns could be explained solely by differences in demographic history between breed dogs and wolves, we conducted forward in time population genetic simulations. We examined different models of population history that had previously been fit to genetic variation data of dogs and wolves (Fig. 1*A* and *SI Appendix*, Tables S3–S5). We assumed the same distribution of selection coefficients across populations in all simulations (*SI Appendix, SI Text* and Table S6). Whereas simulations assuming additive effects predict a negative relationship between neutral heterozygosity and the ratio of zerofold heterozygosity to fourfold heterozygosity, previously inferred distributions of selective effects for humans and mice did not match the quantitative patterns seen in our data (*SI Appendix, SI Text* and Figs. S6 and S7). However, distributions including more weakly deleterious mutations provided a better fit (Fig. 1*B* and *SI Appendix*, Figs. S6 and S7). These results can be interpreted in the context of the nearly neutral theory (22, 28). Neutral heterozygosity is proportional to the effective population size and because selection is less effective at eliminating weakly deleterious variation in small populations relative to larger ones, we observe a negative correlation between neutral heterozygosity and the ratio of zerofold to fourfold heterozygosity. Thus, provided there are enough amino acid changing mutations that are weakly to moderately deleterious ($|s| < 0.001$), the population bottlenecks associated with dog domestication have reduced the ability of negative selection to remove deleterious variants. When assuming fully recessive effects for all deleterious mutations,

we observed a positive relationship between neutral heterozygosity and the ratio of zerofold heterozygosity to fourfold heterozygosity (*SI Appendix*, Fig. S8). This result is consistent with theoretical work showing the number of recessive deleterious alleles can decrease after a bottleneck (29). Taken together, our results argue that most segregating deleterious mutations in dogs and wolves are not fully recessive and more consistent with an additive model.

The Role of Recent Inbreeding. Dogs from some breeds are homozygous for large (>1 Mb) regions of the genome, suggesting recent mating among close relatives (i.e., inbreeding; ref. 30 and *SI Appendix*, Fig. S9*A*). This inbreeding can reduce the effective population size, allowing deleterious alleles to drift higher in frequency and is a mechanism commonly assumed to account for the accumulation of deleterious mutations in dog genomes (31) but has not been formally assessed. Based on three distinct analyses, we find that recent inbreeding is not driving the patterns shown in Fig. 1.

First, we conducted additional forward simulations including negative selection and recent inbreeding within breed dogs. Even strong inbreeding ($F = 0.2$) over the last 300 y, without the bottlenecks associated with domestication and breed formation, is insufficient to generate the observed negative relationship between the zerofold/fourfold heterozygosity ratio and neutral heterozygosity (Fig. 2*A*). Second, we attempted to remove the effects of recent inbreeding on our analysis of heterozygosity. Because recent inbreeding increases the probability that two chromosomes within a given individual share a common ancestor with each other rather than with a chromosome from another individual (*SI Appendix*, Fig. S9*A*), it will reduce within-individual heterozygosity relative to between-individual heterozygosity (32). Thus, we can obtain an estimate of heterozygosity removing the effects of inbreeding by sampling a single read from each individual at each site and determining whether the reads have different nucleotides (*SI Appendix, SI Text*). Forward simulations indicate that this approach removes the effects of recent inbreeding on heterozygosity (*SI Appendix*, Fig. S9). However, in contrast, in the actual data, neutral heterozygosity computed from two canids remains negatively correlated with the ratio of zerofold to fourfold heterozygosity (Fig. 2*B*), suggesting recent inbreeding is not the cause of the association. Finally, when removing large runs of homozygosity (>2 Mb) from our analyses (*SI Appendix, SI Text*), the negative relationship between neutral heterozygosity and the ratio of zerofold heterozygosity to fourfold heterozygosity remained strong (Fig. 2*C*), indicating that it was not driven by patterns of variation within regions of the genome most affected by inbreeding. These unexpected findings imply that population bottlenecks, rather than recent inbreeding, are responsible for the proportional increase in amino acid changing heterozygosity in breed dogs relative to wolves.

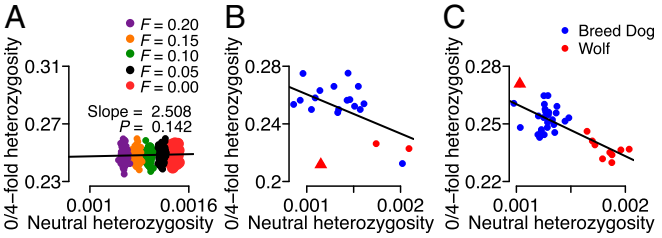

Fig. 2. Recent inbreeding does not drive the relationship between neutral heterozygosity and the zerofold/fourfold heterozygosity ratio. (*A*) Forward simulations using a demographic model that includes inbreeding over the last 100 generations, but not bottlenecks associated with domestication or breed formation ("wolf" demographic model in *SI Appendix*, Table S4). (*B*) Empirical results from computing heterozygosity using one read from each of two individuals per population. The solid line denotes the best-fit linear regression line (intercept = 0.288, slope = −27.25, $r = -0.502$, $P = 0.024$). (*C*) The relationship between neutral polymorphism and the ratio of zerofold to fourfold heterozygosity persists when removing runs of homozygosity. The solid black line denotes the best-fit linear regression line (intercept = 0.287, slope = −27.07, $r = -0.757$, $P < 5 \times 10^{-7}$). This plot uses the same data as in Fig. 1*C*, but removing ROHs. Red triangles denote the Tibetan wolves.

Genetic Load in Dogs Vs. Wolves. Our results indicate demography has affected the ability of purifying selection to remove weakly deleterious variants. However, these analyses do not directly assess the burden of deleterious variants per genome. To quantify this burden, we focused on a subset of the dog and gray wolf genomes with high coverage (Dataset S1), and tabulated the number of neutral and deleterious variants per genome. The Tibetan wolf that appears as an outlier in Fig. 1*C* was excluded from this analysis (results were similar with the Tibetan wolf; see *SI Appendix, SI Text* and Fig. S10). We defined deleterious variants as those amino acid changes that occurred at phylogenetically conserved sites as measured by the Genomic Evolutionary Rate Profiling (GERP) scores (33). Wolves carry significantly more deleterious amino acid changing variants in the heterozygous state than do breed dogs ($P < 2 \times 10^{-5}$, Mann–Whitney *U* test, Fig. 3*A*; *SI Appendix*, Table S7). However, breed dogs carry approximately 320 (22%) more derived deleterious amino acid changing genotypes in the homozygous state relative to wolves ($P < 4 \times 10^{-8}$; Fig. 3*A*). We then assessed the number of derived deleterious alleles per genome by counting heterozygous genotypes once and homozygous-derived genotypes twice. After correcting for the increased false negative rate for heterozygous genotypes compared with homozygous derived genotypes (*SI Appendix, SI Text* and Fig. S10*D* for counts before correction), breed dogs carry ∼115 more derived deleterious alleles than do wolves, corresponding to a 2.6% increase relative to wolves ($P < 0.002$). There are significantly more heterozygous genotypes in wolves than in dogs and significantly more homozygous-derived genotypes in dogs than in wolves at putatively neutral synonymous SNPs as well (Fig. 3*B*). However, the number of synonymous-derived alleles per individual does not differ between dogs and wolves (Fig. 3*B* and *SI Appendix*, Table S7), suggesting that neutral processes alone cannot explain these patterns. We also defined deleterious amino acid changes to be those that differ in polarity and volume, as measured by the Miyata distance (34), and observed qualitatively similar patterns (*SI Appendix*, Fig. S10*B*).

The counts of deleterious variants per individual imply that the genetic load is higher in dogs than in wolves. This conclusion holds if mutations act in an additive manner, because the average dog carries 2–3% more derived deleterious alleles than the average wolf. As a more direct measure of the genetic load, we calculated the GERP score load for each individual. The GERP score load is the sum of the GERP scores over all of the deleterious nonsynonymous variants carried by each individual. Dogs have a 2.1% higher GERP score load compared with wolves ($P < 0.008$, Mann–Whitney *U* test, Fig. 3*C*; *SI Appendix*, Fig. S10*C*). Further, simulations under our demographic and selective models predict that the genetic load will be 2–3% higher in dogs than wolves (Fig. 3*D* and *SI Appendix*, *SI Text*). The increase in load in dogs would be even more pronounced if deleterious mutations were partially recessive, because dogs carried more homozygous derived deleterious variants per individual. We caution, however, that statements about genetic load depend on the underlying demographic and selective models. Further, they assume positive selection that may increase the frequencies of many variants (i.e., polygenic selection) does not account for these patterns (6). However, polygenic selection does not appear to be the dominant force underlying phenotypic change in dogs, because association studies suggest that a small number of large-effect alleles that have been subjected to artificial selection can account for much of the variance in traits (30, 35, 36). Finally, after filtering previously identified selective sweep regions, both the number of derived deleterious alleles and GERP score load remains significantly higher in dogs than wolves ($P < 0.008$), arguing that the genome-wide patterns are not driven by the artificially selected regions (*SI Appendix*, Fig. S11).

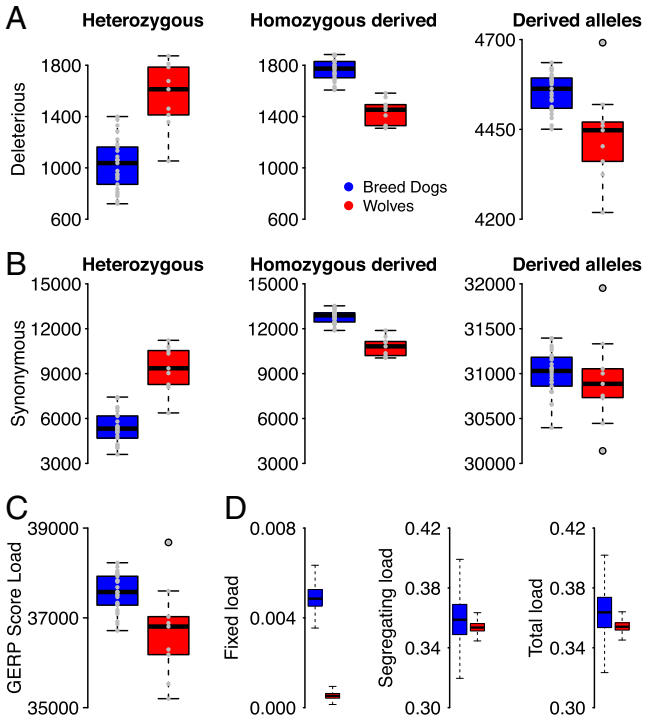

Fig. 3. Comparison of the burden of deleterious genetic variation between breed dogs (blue) and wolves (red) based on high-quality genomes. "Homozygous derived" refers to the number of genotypes per individual that are homozygous for the derived allele. The total number of derived alleles is based on counting each heterozygous genotype once and each homozygous derived genotype twice. Small points denote the genomes used for each species ($n = 25$ for breed dogs, $n = 9$ for wolves). (*A*) Nonsynonymous variants that are predicted to be deleterious (GERP score >4). (*B*) Synonymous variants. (*C*) GERP score load for each individual. (*D*) Genetic load computed from our forward simulations. Outlier points are not shown for clarity. *Left* shows the load due to mutations that became fixed within the most recent 2,480 generations. *Middle* shows the load contributed by segregating mutations only. *Right* shows the total load, combining fixed and segregating variants. $P < 0.008$ for all comparisons between dogs and wolves using a Mann–Whitney *U* test except the comparison of the total number of synonymous derived alleles (*SI Appendix*, Table S7).

Recently, questions have been raised concerning whether recent demographic history can affect the genetic load. Some studies showed similar numbers of putatively deleterious alleles per individual across human populations (3, 4). In contrast, other recent studies reported a significant increase in the number of deleterious alleles (5, 37) and a higher additive genetic load in non-African populations compared with African populations (5, 6, 8, 37, 38). Our present findings of a higher genetic load in dogs compared with wolves supports the view that recent demographic history can affect genetic load. The magnitude of the increase in additive genetic load in non-African human populations has been estimated to be slight, ∼1–3% (5, 6, 8, 37, 38), which is similar in magnitude to the increase we observed in dogs relative to wolves (Fig. 3). Given the differences in the timing and severity of the bottlenecks experienced by humans and dogs, it is surprising that both species show qualitatively similar trends. This similarity suggests that the genome-wide human-mediated demographic processes associated with domestication, although increasing the per individual counts of deleterious variants and the additive genetic load, have not enhanced the genome-wide burden beyond that caused by natural demographic processes in other species. More generally, these findings argue that even extreme recent population bottlenecks may only result in a subtle, but often statistically detectable, increase in the per-individual count of deleterious alleles and the additive genetic load.

Enrichment of Amino Acid Changing Variants Surrounding Selective Sweeps. Although the selective sweep regions were not driving the genome-wide patterns of deleterious variation, the extreme artificial selection during domestication could result in the hitchhiking of deleterious variants surrounding the sweeps (39–41). To evaluate this effect, we focused on a set of 421 selective sweep regions identified through a comparison of domestic dogs to wolves (12, 42). The sweep regions show the expected signatures of classic selective sweeps in dogs (43), such as decreased genetic diversity at putatively neutral fourfold sites using two different measures (Fig. 4*A* for Watterson's θ; see *SI Appendix, SI Text* for average pairwise differences) and an increase in neutral derived allele frequency (Fig. 4*B*). These sweep regions do not show a decrease in neutral diversity in wolves (*SI Appendix*, Fig. S12), supporting the idea that they are genuine targets of selection in dogs.

We next examined patterns of variation at amino acid changing (zerofold) sites. Watterson's θ is similar between sweep and nonsweep regions (Fig. 4*A*), suggesting that the number of deleterious variants in the sample per 10 kb in the sweep regions is similar to that of nonsweep regions. Because the number of neutral variants has been reduced in the sweep regions, there is an enrichment of zerofold variants within the sweep regions. The average derived allele count of zerofold SNPs is significantly elevated within the sweep regions (Fig. 4*B*), suggesting that zerofold variants experienced the same increase in frequency due to hitchhiking within the sweep regions as fourfold variants. Finally, we examined the number of derived zerofold alleles per 100 bp. This metric is influenced both by the number of SNPs and by their frequency in the population (*SI Appendix, SI Text*). The total number of derived alleles per 100 bp at zerofold sites is 1.26 fold higher in the sweep regions compared with the nonsweep regions ($Z = 9.6$, $P < 4 \times 10^{-22}$, two-sample Z test; Fig. 4*C*), indicating that, when normalized to have the same sequence length, sweep regions contribute more to the genetic load of amino acid changing variants than the nonsweep regions. However, because most of the genome lies outside of the sweep regions, the sweeps do not affect the overall genome-wide patterns of variation (*SI Appendix*, Fig. S4*D* and S11).

The enrichment of amino acid changing variants near selective sweeps in dogs is significantly greater than that in wolves, suggesting it is not driven by other factors like differences in mutation or recombination rates (*SI Appendix, SI Text* and Figs. S13 and S14). Conceivably, the excess of amino acid variation surrounding the selective sweeps could be the direct target of positive selection. However, we believe hitchhiking of deleterious mutations is a better explanation (*SI Appendix, SI Text*).

Enrichment of Mendelian Disease Genes Near Selective Sweeps. We assessed whether artificial selection may partially be responsible for the numerous Mendelian genetic diseases observed in breed dogs. Specifically, we determined whether the previously reported targets of selective sweeps (12, 42, 44, 45) were enriched for genes implicated in disease. We find slightly more overlap among 145 genes implicated in Mendelian disease in dogs and genes near recent selective sweeps than expected by chance ($P = 0.087$ and $P = 0.155$; *SI Appendix*, Table S8). To increase statistical power, we repeated our analyses by using 2,535 genes causing Mendelian diseases in humans based on the shared disease etiology between humans and dogs (46, 47). We find more Mendelian disease genes overlap with genes near the selective sweeps reported by Vaysse et al. (44) and Akey et al. (45) (i.e., sweeps related to breed formation) than expected by chance ($P = 0.005$ and $P = 0.057$, respectively; *SI Appendix*, Table S9). This enrichment could be explained by two different mechanisms. First, genes controlling artificially selected traits in dogs could be the same set of genes that confer Mendelian disease in humans. Alternatively, the human disease genes could also cause disease in dogs but be located in regions linked to those under selection for breed traits. Disease alleles would increase in frequency because of hitchhiking with the variants controlling the trait under intense

EVOLUTION

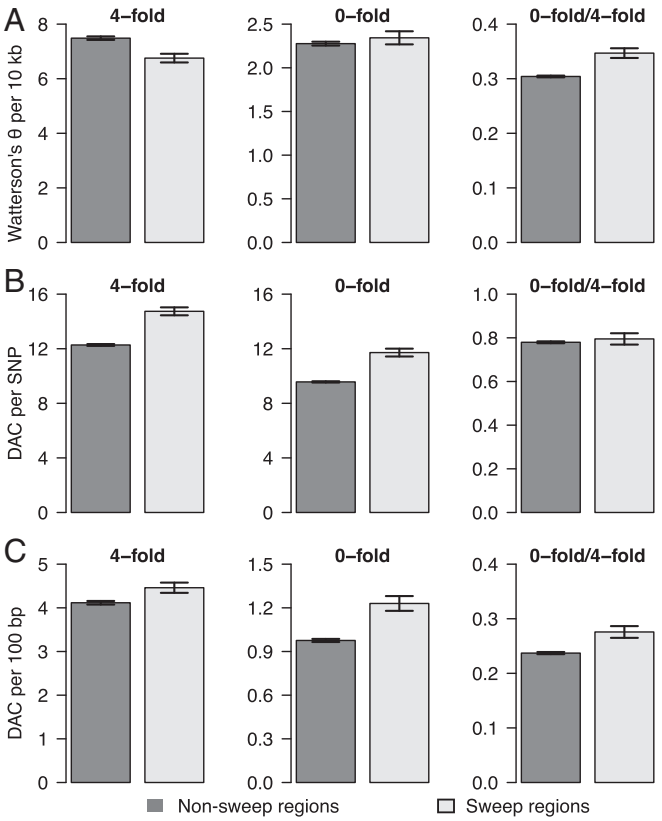

Fig. 4. Genetic variation surrounding nonsweep (dark gray) and sweep (light gray) regions in breed dogs. (*A*) Watterson's θ, an estimate of genetic diversity based on the number of SNPs. (*B*) The average derived allele count (DAC) per SNP. (*C*) Average DAC per 100 bp (considering invariant positions). Each variant site is counted the number of times its derived allele appears in the sample. Error bars are 95% confidence intervals. Note the decrease in diversity in *A* and the increase in derived allele frequency (*B* and *C*) at fourfold sites, the expected patterns surrounding a selective sweep. However, the total number of zerofold variants is not reduced near sweeps (*A*), and the average frequency of derived zerofold alleles is increased near the sweeps (*B* and *C*).

artificial selection. Under either mechanism, our results suggest that an associated cost of selection for specific traits in breed dogs is an enhanced likelihood for Mendelian disease. Considering that many modern breeds have been selected for unusual appearance and size, which reflects fashion more than function, our results raise ethical concerns about the creation of fancy breeds. For example, positive selection for black coat color in poodles may have caused a high frequency of copy number variants of the *KITLG* gene, resulting in an increased frequency of squamous cell carcinoma of the nail bed (48). Interestingly, we find no enrichment of Mendelian disease genes in selective sweeps that occurred early during dog domestication (i.e., sweeps identified through comparison of dogs and wolves), perhaps suggesting that early and breed-specific sweeps involve fundamentally different types of genes (*SI Appendix, SI Text* and Tables S8 and S9).

Conclusions

Our results show that the domestication process has dramatically affected patterns of deleterious variation across the dog genome. First, population history has had a genome-wide effect that increases the burden of deleterious variation in breed dogs as indicated by an elevated level of amino acid changing variation relative to wolves where selection is more efficacious. Comparison of the additive genetic load between dogs and wolves reveals qualitatively similar trends to those seen in comparisons of bottlenecked and nonbottlenecked human populations. This similarity indicates that, although detectable, the effect of recent demography on additive genetic load is likely to be subtle, even for extreme bottlenecks. Although dramatic fitness consequences in dogs are often thought to be caused by recessive mutations of large effect, we find that as in humans, most of the additive genetic load is accounted for by numerous weakly deleterious mutations (5, 6), which are particularly hard to remove from bottlenecked populations. Second, intense artificial selection for desirable traits results in a concomitant accumulation of deleterious variation in genes trapped in sweep regions. This finding is especially disconcerting because sweep regions are enriched for disease-related genes, a finding that highlights anew the controversy over intense selection for fancy traits in dog breeds and other domestic species. Importantly, selectively breeding a limited number of individuals during domestication or breed formation can reduce effective population size across the genome. Thus, selective breeding practices can increase deleterious variation genome-wide, not just at the loci controlling selected traits. Third, our demographic models suggest that repeated population bottlenecks and small effective population size have had a more profound effect on the accumulation of weakly deleterious variation than does recent inbreeding (i.e., mating between close relatives). Consequently, to minimize the accumulation of deleterious variation in the increasing number of species suffering from habitat loss and fragmentation, conservation efforts should focus on maintaining sufficient population sizes in the wild and captivity, rather than focusing exclusively on inbreeding avoidance. Finally, our approach provides a comprehensive method for evaluating deleterious variation from genome data in the small isolated and threatened populations worldwide that can help prioritize their genetic management.

Materials and Methods

Genomic Data. Breed dogs were sequenced at the University of Missouri on an Illumina GAIIx, 2000 or 2500. These studies were approved by the University of Missouri, Animal Care and Use Committee and performed with informed consent of the dogs' owners. Wolves were sequenced at BGI and the University of California, Berkeley sequencing core. Genomes generated here have been deposited into the Short Read Archive (Dataset S1). Data were processed by using standard bioinformatics pipelines (*SI Appendix, SI Text*), including alignment to CanFam 3.1 by using BWA (49), indel realignment, base quality score recalibration, and filtering of reads with quality <30. Neutral and coding regions were taken from ref. 10.

Estimation of Heterozygosity Without Calling Genotypes. Our approach to estimating heterozygosity from the low-coverage data, called FourSite (https://github.com/LohmuellerLab/FourSite), is similar to that described by Lynch (50) (*SI Appendix, SI Text*). For each site within a given genome, we sample four sequencing reads and tabulate whether: (*i*) all four reads are the same base, (*ii*) two reads are one base and two reads are a different base, or (*iii*) one read is one base, and three reads are a different base. We then computed the likelihood of the heterozygosity and sequencing error rate as function of these counts across a particular functional category (*SI Appendix, SI Text*).

Analysis of the High-Coverage Genomes. We selected a high coverage sample set consisting of the 36 samples (10 gray wolves, 25 breed dogs, and a golden jackal) with an average genomic coverage > 15× for SNP genotype calling (Dataset S1). Genotypes were called by using GATK (19) (*SI Appendix, SI Text*). Heterozygosity was calculated as the number of heterozygous genotypes for each individual divided by the number of called genotypes. Runs of homozygosity were identified by using PLINK (51).

Accumulation of Deleterious Derived Alleles. To assess the accumulation of deleterious derived alleles in dogs and wolves, we counted the number of variants in each of 25 dog genomes and 9 or 10 gray wolf genomes (*SI Appendix, SI Text*). We used the golden jackal as an outgroup to classify the ancestral state and considered only those sites where the jackal was homozygous as the ancestral allele. Because the jackal has evolved since the common ancestor with dogs and wolves, it may not perfectly represent the true ancestral state. However, this error is not expected to bias the relative comparison of variants between dogs and wolves because both show similar levels of divergence with jackal (ref. 10, *SI Appendix, SI Text*). We normalized for differences in missing data across individuals and corrected the number

of derived alleles per animal for the fact that the false-negative rate for calling heterozygous genotypes is higher than for calling homozygous genotypes (*SI Appendix, SI Text*).

Forward Simulations. To determine whether we could recapitulate the negative correlation between the zerofold/fourfold ratio and neutral heterozygosity using realistic models of demography and purifying selection, we performed forward in time simulations under the Wright Fisher model in the Poisson Random Field framework (2, 52, 53). We explored a variety of different distributions of selective effects, including those fit to mouse (54) and human (55) data, as well as several custom distributions (*SI Appendix, SI Text*).

Analysis of Coding Genetic Diversity near Vs. far from Sweeps. We used sweep regions that have been identified in the ancestral population of breed dogs, presumably related to domestication (12, 42). To assess whether there were differences in patterns of variation between sweep and nonsweep regions, we performed a jackknife over chromosomes. The SE on our point estimates of diversity were computed from the SD of these jackknife estimates. Given these SEs, 95% confidence intervals were determined under the standard normality assumptions.

Testing for Overlap Between Mendelian Disease Genes and Genes Located in Selective Sweeps. We tested whether genomic regions implicated in selective sweeps are enriched for genes that cause Mendelian diseases. We used genes that were reported in the Online Mendelian Inheritance in Animals database to cause Mendelian disease in dogs as well as genes in the Online Mendelian Inheritance in Man "morbidmap" implicated in Mendelian diseases in humans. We then examined three different sets of selective sweep regions identified in dogs, including the set of sweeps associated with domestication that are shared across breeds and were described above for the deleterious mutation analysis as well as two sets of breed-specific sweeps (44, 45) (*SI Appendix, SI Text*). We then computed the probability of observing as many or more overlapping genes by chance alone using a hypergeometric distribution.

ACKNOWLEDGMENTS. We thank Bogdan Pasaniuc, Rena Schweizer, Pedro Silva, Emilia Huerta Sanchez, Bridgett vonHoldt, Evan Koch, Tom Smith, Brian Davis, Elaine Ostrander, and members of the K.E.L. laboratory for discussions and comments on the manuscript. Part of the sequencing costs were supported by a grant from the "Programa de Captación del Conocimiento para Andalucía" (Spain) (to C.V.). C.D.M. is supported by a University of California, Los Angeles Quantitative Computational Biosciences Postdoctoral Fellowship. D.O.-D.V. is supported by a University of California Institute for Mexico and the United States and El Consejo Nacional de Ciencia y Tecnología Doctoral Fellowship 213627. O.R. was supported by a Fundació Barcelona Zoo and Ajuntament de Barcelona Grant. T.M.-B. was supported by the Ministry of Science and Innovation, Spain, BFU2014-55090-P, BFU2015-7116-ERC, and BFU2015-6215-ERC. K.E.L. is supported by a Searle Scholars Fellowship and an Alfred P. Sloan Research Fellowship in Computational and Molecular Biology. We appreciate grant support from the Missouri Advantage program and National Science Foundation Grants DEB-1021397 and DEB-1257716 (to R.K.W.).

1. Lohmueller KE (2014) The distribution of deleterious genetic variation in human populations. *Curr Opin Genet Dev* 29:139–146.
2. Lohmueller KE, et al. (2008) Proportionally more deleterious genetic variation in European than in African populations. *Nature* 451(7181):994–997.
3. Simons YB, Turchin MC, Pritchard JK, Sella G (2014) The deleterious mutation load is insensitive to recent population history. *Nat Genet* 46(3):220–224.
4. Do R, et al. (2015) No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. *Nat Genet* 47(2):126–131.
5. Fu W, Gittelman RM, Bamshad MJ, Akey JM (2014) Characteristics of neutral and deleterious protein-coding variation among individuals and populations. *Am J Hum Genet* 95(4):421–436.
6. Henn BM, Botigué LR, Bustamante CD, Clark AG, Gravel S (2015) Estimating the mutation load in human genomes. *Nat Rev Genet* 16(6):333–343.
7. Gazave E, Chang D, Clark AG, Keinan A (2013) Population growth inflates the per-individual number of deleterious mutations and reduces their mean effect. *Genetics* 195(3):969–978.
8. Peischl S, Dupanloup I, Kirkpatrick M, Excoffier L (2013) On the accumulation of deleterious mutations during range expansions. *Mol Ecol* 22(24):5972–5982.
9. Schubert M, et al. (2014) Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc Natl Acad Sci USA* 111(52):E5661–E5669.
10. Freedman AH, et al. (2014) Genome sequencing highlights the dynamic early history of dogs. *PLoS Genet* 10(1):e1004016.
11. Boyko AR (2011) The domestic dog: Man's best friend in the genomic era. *Genome Biol* 12(2):216.
12. vonHoldt BM, et al. (2010) Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* 464(7290):898–902.
13. Lindblad-Toh K, et al. (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438(7069):803–819.
14. Ash EC (1927) *Dogs: Their History and Development* (E. Benn Limited, London).
15. American Kennel Club (1997) *The Complete Dog Book* (Howell Book House, New York).
16. Auton A, et al. (2013) Genetic recombination is targeted towards gene promoter regions in dogs. *PLoS Genet* 9(12):e1003984.
17. Wang GD, et al. (2013) The genomics of selection in dogs and the parallel evolution between dogs and humans. *Nat Commun* 4:1860.
18. Zhang W, et al. (2014) Hypoxia adaptations in the grey wolf (Canis lupus chanco) from Qinghai-Tibet Plateau. *PLoS Genet* 10(7):e1004466.
19. DePristo MA, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43(5):491–498.
20. Fay JC, Wyckoff GJ, Wu CI (2001) Positive and negative selection on the human genome. *Genetics* 158(3):1227–1234.
21. Elyashiv E, et al. (2010) Shifts in the intensity of purifying selection: An analysis of genome-wide polymorphism data from two closely related yeast species. *Genome Res* 20(11):1558–1573.
22. Akashi H, Osada N, Ohta T (2012) Weak selection and protein evolution. *Genetics* 192(1):15–31.
23. Cruz F, Vilà C, Webster MT (2008) The legacy of domestication: Accumulation of deleterious mutations in the dog genome. *Mol Biol Evol* 25(11):2331–2336.
24. Björnerfeldt S, Webster MT, Vilà C (2006) Relaxation of selective constraint on dog mitochondrial DNA following domestication. *Genome Res* 16(8):990–994.
25. Shannon LM, et al. (2015) Genetic structure in village dogs reveals a Central Asian domestication origin. *Proc Natl Acad Sci USA* 112(44):13639–13644.
26. Wayne RK, et al. (1991) Conservation genetics of the endangered Isle Royale gray wolf. *Conserv Biol* 5(1):44–51.
27. Boudreaux MK (2012) Inherited platelet disorders. *J Vet Emerg Crit Care (San Antonio)* 22(1):30–41.
28. Ohta T (1972) Population size and rate of evolution. *J Mol Evol* 1(3):305–314.
29. Balick DJ, Do R, Cassa CA, Reich D, Sunyaev SR (2015) Dominance of deleterious alleles controls the response to a population bottleneck. *PLoS Genet* 11(8):e1005436.
30. Boyko AR, et al. (2010) A simple genetic architecture underlies morphological variation in dogs. *PLoS Biol* 8(8):e1000451.
31. McGreevy PD, Nicholas FW (1999) Some practical solutions to welfare problems in dog breeding. *Anim Welf* 8(4):329–341.
32. Wright S (1951) The genetical structure of populations. *Ann Eugen* 15(4):323–354.
33. Davydov EV, et al. (2010) Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLOS Comput Biol* 6(12):e1001025.
34. Miyata T, Miyazawa S, Yasunaga T (1979) Two types of amino acid substitutions in protein evolution. *J Mol Evol* 12(3):219–236.
35. Rimbault M, Ostrander EA (2012) So many doggone traits: Mapping genetics of multiple phenotypes in the domestic dog. *Hum Mol Genet* 21(R1):R52–R57.
36. Wayne RK, vonHoldt BM (2012) Evolutionary genomics of dog domestication. *Mamm Genome* 23(1-2):3–18.
37. Henn BM, et al. (2015) Distance from Sub-Saharan Africa predicts mutational load in diverse human genomes. *bioRxiv*, 10.1101/019711.
38. Gravel S (2014) When is selection effective? *bioRxiv*, 10.1101/010934.
39. Chun S, Fay JC (2011) Evidence for hitchhiking of deleterious mutations within the human genome. *PLoS Genet* 7(8):e1002240.
40. Hartfield M, Otto SP (2011) Recombination and hitchhiking of deleterious alleles. *Evolution* 65(9):2421–2434.
41. Lu J, et al. (2006) The accumulation of deleterious mutations in rice genomes: A hypothesis on the cost of domestication. *Trends Genet* 22(3):126–131.
42. Axelsson E, et al. (2013) The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495(7441):360–364.
43. Nielsen R (2005) Molecular signatures of natural selection. *Annu Rev Genet* 39(1):197–218.
44. Vaysse A, et al.; LUPA Consortium (2011) Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLoS Genet* 7(10):e1002316.
45. Akey JM, et al. (2010) Tracking footprints of artificial selection in the dog genome. *Proc Natl Acad Sci USA* 107(3):1160–1165.
46. Karlsson EK, Lindblad-Toh K (2008) Leader of the pack: Gene mapping in dogs and other model organisms. *Nat Rev Genet* 9(9):713–725.
47. Ostrander EA (2012) Franklin H. Epstein Lecture. Both ends of the leash–the human links to good dogs with bad genes. *N Engl J Med* 367(7):636–646.
48. Karyadi DM, et al. (2013) A copy number variant at the KITLG locus likely confers risk for canine squamous cell carcinoma of the digit. *PLoS Genet* 9(3):e1003409.
49. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26(5):589–595.
50. Lynch M (2008) Estimation of nucleotide diversity, disequilibrium coefficients, and mutation rates from high-coverage genome-sequencing projects. *Mol Biol Evol* 25(11):2409–2419.
51. Purcell S, et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81(3):559–575.
52. Sawyer SA, Hartl DL (1992) Population genetics of polymorphism and divergence. *Genetics* 132(4):1161–1176.
53. Lohmueller KE (2014) The impact of population demography and selection on the genetic architecture of complex traits. *PLoS Genet* 10(5):e1004379.
54. Halligan DL, et al. (2013) Contributions of protein-coding and regulatory change to adaptive molecular evolution in murid rodents. *PLoS Genet* 9(12):e1003995.
55. Boyko AR, et al. (2008) Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet* 4(5):e1000083.

Supplementary Appendix

Table of Contents

SI Text

Sequence processing, read alignment and filtering

We converted all fastq files to Sanger quality score encoding and then trimmed adaptor sequences and low quality sequences (quality score < 20) from reads using SCYTHE (https://github.com/vsbuffalo/scythe) and SICKLE (https://github.com/najoshi/sickle). The trimmed reads were then aligned to the dog genome (CanFam 3.1) using BWA 0.6.2 (ref.1) with default parameters. We removed reads not mapped in a proper pair or with low mapping quality scores (< 30) using SAMTOOLS 0.1.19 (ref. 2), and then we marked and removed any duplicate reads using PICARD TOOLS 1.77 (http://picard.sourceforge.net/).

Indel realignment and base quality score recalibration

We performed indel realignment and base quality score recalibration (BQSR) on the extracted coding and neutral region bam files. The presence of indels in reads that are not present in the reference genome often result in mismatches around the indel that appear as SNPs. Consequently, we conducted local realignment around indels using the RealignerTargetCreator and IndelRealigner tools in the Genome Analysis Toolkit v 3.2.2 (GATK, refs. 3, 4). These tools were run with default settings without a database of known indels.

Sequence base quality scores reflect the probability that the called base is an error. However, sequencers often produce inaccurate and biased base quality scores. We therefore applied BQSR with GATK. In order to run BQSR, it is necessary to provide a database of known variant sites which is used to distinguish between true variants and errors. Therefore, we first conducted a round of SNP calling with Unified Genotyper in GATK with a minimum base quality score threshold of 20. SNP calling was conducted on individual samples (*vs* joint calling)

to prevent biases in SNP calling accuracy between groups with different numbers of samples. Multiple rounds of BQSR were conducted until convergence was observed.

Additional quality filters

Standard variant calling pipelines routinely apply filters to reads prior to calling, and *post hoc* filters are applied to the called variants. These filters reduce the potential contribution of errors. As the read based estimates of heterozygosity are directly applied to the reads, it was important to apply additional filters to reduce the potential for errors to contribute to heterozygosity estimates. Therefore, we took the BQSR bam files and used a custom python script to change any bases within a read with a quality score < 30 to 'N' so that they would be ignored in downstream read-based analyses (https://github.com/cdmarsden/replace_lowqualitybases). Then, we removed reads with low overall quality. Specifically, reads where the length was shorter than 40 bp after trimming, or where more than 20% of the bases in the read had quality scores less than 30 were removed (https://github.com/cdmarsden/remove_lowqualityreads).

Extracting Neutral and Coding regions

From the above processed files, we extracted a set of neutral and coding regions using BEDTOOLS v2.17.0 (ref. 5). The neutral regions were a set of 5,139 1 kb neutral regions previously identified by (6). In brief these regions were selected based on an absence of coding DNA, having a distance far from known and predicted genes (>100 kb) and >50 kb from another selected neutral region, with PhastCons scores >0.5, no evidence of elevated GC content or consecutive 50 bp windows with mappability score >2 and an absence of N's in the reference genome. The set of extracted coding regions also was based on those regions previously identified by (6). These regions were derived from NCBI and Ensembl annotation databases,

with the longest transcripts for each gene selected to prevent the same sites from being included more than once after all transcripts without proper start and stop codons, and with premature stop codons were removed. Finally, for both the coding and neutral regions, we masked CpG islands and repeat regions. CpG sites were removed for certain analyses as described throughout the text. This was done by filtering any position that came after a C nucleotide or before a G nucleotide in the CanFam3.1 reference sequence (7).

Identifying zero and four fold sites

Within the coding regions, we identified the zero and four-fold degenerate sites by iterating across all four possible bases at each site along a transcript and recording the changes in the resulting amino acid. Sites were classed as zero-fold degenerate when the four different bases resulted in four different amino acids, and four-fold degenerate when no changes in amino acids were observed.

Estimation of heterozygosity without calling genotypes

Here we fully describe our approach to estimating heterozygosity from sequencing reads without calling genotypes. First, a pileup is made, and an iterator used to move along a bam file site by site. The base calls at each site are extracted and assessed to determine whether to include the site. Specifically, we excluded sites where more than two bases are observed, or where the number of reads observed at the site is less than four or greater than four times the average genome coverage of the individual (which may be indicative of copy number variants / misalignment). Otherwise, we select a random sample of 4 reads and count whether 1) all four reads are the same base, 2) two reads are one base and two reads are a different base, or 3) one read is one base, and three reads are a different base. We repeated these counts 10 times for each site, and averaged the results across the ten iterations. Down sampling to 4 reads per genome per

site is meant to eliminate systematic biases that may be present due to uneven depths of coverage across different regions of the genome or between different genomes. We note that the distributional properties of this estimator based on averaging over 10 samples of reads per site will be influenced by the genome-wide coverage as the same reads have a higher probability of being chosen more often in low-coverage samples. However, because our goal in performing the resampling process is to simply try to obtain a better point estimate of the heterozygosity, rather than assess the uncertainty in our estimate, we do not think that this should cause any systematic biases. In support of this, we found that our estimates of heterozygosity were not correlated with the coverage (SI Appendix, Fig. S2), suggesting that our approach is robust to variable degrees of coverage across samples. However, the ratio of 0-fold to 4-fold heterozygosity showed a weak negative correlation with coverage (r = -0.24, P = 0.024). Nevertheless, neutral heterozygosity was still a significant predictor of the 0-fold/4-fold ratio in a multiple regression including coverage as a predictor (intercept = 0.316, slope = -30.4, $P = 3.96 \times 10^{-9}$), suggesting that our results are not driven by differences in coverage. The python script (CountReadFoursite.py) used for these calculations is available from the Lohmueller lab Github page (https://github.com/LohmuellerLab).

Let X_i refer to the average number of reads (averaged across the 10 resamplings as described above) showing the alternate allele at site i in the genome of a particular individual. Thus, X_i is our observed data. Because we are considering only 4 reads per site, $X_i \in \{0,1,2,3,4\}$. Let G_i refer to the number of copies of the minor allele in the genotype at site i in the same individual. G_i is unknown. $G_i \in \{0,1,2\}$ for homozygous reference, heterozygous, and homozygous non-reference genotypes, respectively. Further, let us also assume that sequencing errors occur at a rate of ε per site per read. In practice, we estimate ε from the data.

We can then write down a genotype likelihood or $P(X_i \mid G_i)$. For $G_i = 0$, the only way that $X_i > 0$ is if there are sequencing errors. Then,

$$P(X_i \mid G_i = 0) = \begin{cases} (1-\varepsilon)^4 & ,X_i = 0 \\ 4\varepsilon(1-\varepsilon)^3 & ,X_i = 1 \\ 6\varepsilon^2(1-\varepsilon)^2 & ,X_i = 2 \\ 4\varepsilon^3(1-\varepsilon) & ,X_i = 3 \\ \varepsilon^4 & ,X_i = 4 \end{cases}.$$

For $G_i = 1$ (heterozygous genotypes), we assume that sequencing errors have a symmetric effect, as previously presumed (8). Thus, we do not consider them in the likelihood. Second, we assume that X_i is binomially distributed and that the probability of drawing a read from each of the two alleles is 0.5. Then, we can write down a trivial binomial likelihood:

$$P(X_i \mid G_i = 1) = \begin{cases} 0.5^4 & ,X_i = 0 \\ 4(0.5)(0.5)^3 & ,X_i = 1 \\ 6(0.5)^2(0.5)^2 & ,X_i = 2 \\ 4(0.5)(0.5)^3 & ,X_i = 3 \\ 0.5^4 & ,X_i = 4 \end{cases}.$$

The genotype likelihood for $G_i = 2$ is essentially the same as for $G_i = 0$, except switching $X_i = 0$ and $X_i = 4$ and switching $X_i = 1$ and $X_i = 3$.

Following Lynch (8), we can now write the likelihood of the parameters, π and ε based on the observed data at a particular site. Specifically, for observed data configuration $X_i = k$:

$$L(\pi, \varepsilon \mid X_i = k) = \left[P(X_i = k \mid G_i = 0) + P(X_i = k \mid G_i = 2)\right](1-\pi) + P(X_i = k \mid G_i = 1)\pi .$$

Let N_k denote the number of sites with the functional annotation of interest where there are k reads for the non-reference allele (and 4-k reads for the reference allele). In other words, N_k is the number of sites where $X_i = k$. Assuming that sites are independent, we can write down the log-likelihood of the full data for that annotation as:

$$l(\pi,\varepsilon \mid N) = \sum_{k=0}^{4} N_k \ln\left[L(\pi,\varepsilon \mid X_i = k)\right].$$

To find the maximum likelihood estimates (MLEs) of π and ε, we maximized the above likelihood function using a grid-search. For our data, we searched a range of 0.00005 to 0.003 for π. This is likely to be a reasonable range as previous studies of canid genetic diversity have provided estimates well within this range (6, 9). For ε, we searched a range from 0.0001 to 0.001. This is also reasonable as we filtered bases with quality scores less than 30 (corresponding to 0.001). Our error model does not explicitly account for read mapping errors. Thus, while in principle the error rates per base could be much greater than our upper bound of 0.001 in the grid search, we found that they were always less than this upper bound (Dataset S1), suggesting that our upper bound in the grid search is reasonable. We used step sizes of 10^{-5} for both parameters.

Intuitively, this estimator allows us to jointly estimate the error rate and heterozygosity by considering four reads at a site. Because error rates are typically low (<0.001), it is unlikely that more than one random error will occur at a particular site. Thus, those sites where there are 2 reads for each allele ($X_i = 2$) are unlikely to be due to sequencing errors and instead are likely to be true heterozygotes. The $X_i = 1$ and $X_i = 3$ sites are a mixture of truly heterozygous sites and sequencing errors. By using the information from the $X_i = 2$ sites, we correctly account for both possibilities. Importantly, while the approach assumes that errors are randomly distributed, there are likely to be systematic errors that can lead to $X_i = 2$. However, our data indicate this has little effect in practice as we find that the estimates of heterozygosity from our low-coverage approach agree reasonably well with those obtained on a subset of individuals using GATK (see below and SI Appendix, Fig. S1). Note that the GATK-based estimates of heterozygosity were from only high-coverage individuals using all the reads. However, for the low-coverage estimation, we followed the approach described above and used only 4 reads per site.

We have implemented these calculations in an R script called FourSite.R which is available on the Lohmueller Github page (https://github.com/LohmuellerLab/FourSite).

Evaluation of the quality of the low-coverage data

To evaluate the quality of the estimates of heterozygosity using FourSite on the low-coverage sequencing data, we compared them to estimates from the higher-coverage genomes. Specifically, we focused on the 25 breed dogs and 10 wolves that were within our high-coverage genome set. We treated these individuals as though they had lower coverage and applied our FourSite approach to estimate heterozygosity using four reads per site as described above. We then compared these estimates using the low-coverage data to estimates made from these same individuals when calling genotypes from the deeper sequencing data using GATK. Overall, estimates of heterozygosity from the low-coverage data show excellent concordance with the estimates from GATK (SI Appendix, Fig. S1). There is a slight degree of overestimation of neutral heterozygosity in the low-coverage analysis, but some of this may be due to underestimating heterozygosity in the high-coverage genomes (see "***Evaluation of the quality of the high-coverage data***" for further discussion). There is a very slight underestimation of 4-fold heterozygosity in the low-coverage data. However, this effect is consistent across populations, so it should not bias our estimates of differences in patterns of genetic variation between dogs and wolves.

Next, we tested for a relationship between the ratio of 0-fold heterozygosity to 4-fold heterozygosity and neutral variation using the FourSite estimates (based only on the low-coverage data) of heterozygosity from the 35 high coverage genomes. The 0-fold to 4-fold ratio shows a strong negative correlation with neutral heterozygosity when using the estimates of heterozygosity derived from the low-coverage data (SI Appendix, Fig. S4A). This pattern is

similar to that obtained when using the genotypes called using GATK (Fig. 1C). Quantitatively, the regression parameters are similar to each other from these two different datasets (SI Appendix, Table S1). The 95% CIs on the slope and intercept parameters obtained from the GATK called genotypes overlap with those from the FourSite approach. However, we observe a slight upward bias in the estimate of the intercept parameter from the FourSite data relative to the GATK genotypes. This bias may be driven by the slight under-calling of 4-fold heterozygosity in the low-coverage analysis relative to the called genotypes. Importantly, this bias only affects the overall 0-fold to 4-fold ratio and not the relationship between that ratio and neutral heterozygosity, which is captured in the slope parameter. The estimates of the slope show greater agreement between the two methods of processing the data. Finally, because we compare both the low-coverage and high-coverage estimates to the predictions from the forward simulations, this slight upward bias does not affect any of our conclusions.

Estimation of heterozygosity from two dogs

Estimates of heterozygosity computed from a single individual may be affected by recent inbreeding. Recent inbreeding will result in the two alleles at a given locus within a given dog sharing a common ancestor more recently than two alleles sampled from different animals (SI Appendix, Fig S9A). This idea has been formalized in Wright's F_{IS} statistic, which in its simplest form, can be written as:

$$F_{IS} = \frac{H_S - H_I}{H_S}.$$

where H_S is the expected heterozygosity within a sub-population, and H_I is the observed heterozygosity within an individual sampled from the sub-population.

This statistic suggests that heterozygosity computed between different individuals can provide an alternate estimator of heterozygosity that should not be affected by recent inbreeding.

Simulations show that this is in fact the case (SI Appendix, Fig S9B and S9C). Specifically, we simulated data under a model of population history that included bottlenecks associated with domestication and breed formation as well as inbreeding over the most recent 100 generations. This inbreeding results in a reduction in heterozygosity and a weak negative relationship between neutral heterozygosity and the 0-fold to 4-fold heterozygosity ratio (SI Appendix, Fig 9B). Our estimator, which samples one allele from each of two individuals, is not affected by this recent inbreeding (SI Appendix, Fig. 9C).

Thus, we applied this estimator to our data. For the 17 populations where we had at least 2 different animals, we re-computed heterozygosity, using one sequence read at a given site from each of the two individuals (see code at https://github.com/LohmuellerLab/FourSite). Thus, as long as the two individuals are not closely related, this approach should remove the effects of recent inbreeding reducing heterozygosity within an individual.

We estimated heterozygosity from a pair of sequence reads from individuals *j* and *k* using the following approach. Let $X_i = 0$ if the two reads are the same and $X_i = 1$ if the two reads are different at site *i*. First, consider sites that are true differences between a single allele sampled from animals *j* and *k* ($G_i = 1$). Under the assumption that sequencing error rates are low and that the number of sites that differ between animals are low, then $P(X_i = 1 \mid G_i = 1) = 1$.

Second, consider the case where the two alleles are the same. Because this is likely to be the case for most of the sites across the genome, we need to account for sequencing errors, even when the error rate is low. Here,

$$P(X_i = 0 \mid G_i = 0) = (1-\varepsilon_j)(1-\varepsilon_k) + \varepsilon_j\varepsilon_k$$
$$P(X_i = 1 \mid G_i = 0) = (1-\varepsilon_k)\varepsilon_j + (1-\varepsilon_j)\varepsilon_k \text{ ,}$$

where ε_j and ε_k are the sequencing error rates for dogs *j* and *k*, respectively.

In practice, we use the MLEs for ε_j and ε_k that were obtained for each animal from each class of sites using the FourSite approach described above. Consequently, the likelihood can be written as:

$$L(\pi,\varepsilon \mid X_i = 0) = \left[P(X_i = 0 \mid G_i = 0)\right](1-\pi) + 0\pi$$
$$L(\pi,\varepsilon \mid X_i = 1) = \left[P(X_i = 1 \mid G_i = 0)\right](1-\pi) + P(X_i = 1 \mid G_i = 1)\pi \,.$$

Let N_1 denote the number of sites with the functional annotation of interest where there is a difference between the reads for dogs *j* and *k* and N_0 denote the number of sites where the reads are the same. Assuming that sites are independent, we can write the log-likelihood of the full data for that annotation as:

$$l(\pi,\varepsilon \mid N) = N_0 \ln\left[L(\pi,\varepsilon \mid X_i = 0)\right] + N_1 \ln\left[L(\pi,\varepsilon \mid X_i = 1)\right].$$

We also used a grid-search to find the MLE of π.

SNP and genotype calling on high-coverage genomes

We selected a high coverage sample set consisting of the 36 samples (10 grey wolves, 25 breed dogs, and a golden jackal) with an average genomic coverage > 15X for SNP genotype calling (Dataset S1). Using the coding and neutral BQSR bam files, we conducted single sample SNP calling with GATK (refs. 3, 4) using a minimum base quality score of 30, and emitting both variant and invariant sites. We then merged all of the individual sample vcfs, and applied GATK best practices on *post hoc* quality filters (QD < 2.0, FS > 60.0, MQ < 40.0, MQRankSum < -12.5, ReadPosRankSum < -8.0), as well as setting a per sample minimum and maximum depth of coverage of 2 and 80 respectively, and minimum genotype quality of 20. Finally, we removed any non-biallelic variant sites and clustered SNPs (> 3 SNPs within 10 bp). Genotype calls, functional annotations, Miyata scores, and GERP scores for coding SNPs can be found in the

"Dogwolf_vcf_Filtered_coding_region_with_Miyata_and_GerpScore_annotated.vcf" file on Dryad (doi:10.5061/dryad.012s5).

Evaluation of the quality of the high-coverage data

Because we employed a consistent bioinformatics pipeline to all the samples used in our study, we expect bioinformatics batch effects between samples and populations to be minimized. However, the genome sequences were obtained from different labs, sequencing machines, and different time points. Thus, it is important to assess the overall quality of the data to ensure it is not confounded by these batch effects. We assessed the overall quality of our high-coverage genotype calls described above by comparing them to Affymetrix genotype data from VonHoldt et al. (10). One challenge to comparing these data is that they only contain one individual that was among our 35 high-coverage genomes was also included in vonHoldt et al. Thus, we utilized two different approaches to assess concordance between the genotype and the sequencing data.

First, as done in Freedman et al. (6), we assessed whether our sequenced genomes clustered with other individuals from the same population in the genotype data. If systematic batch effects plague our data, then we would expect all our sequenced genomes to form clusters that are separate from those seen in the SNP genotype data. On the other hand, if systematic batch effects are small relative to differences in genetic variation due to population structure, then we expect our sequenced individuals to cluster with SNP genotyped individuals from the same population. We focused on the 12 dog breeds where we had high-coverage genomes that were represented by at least 12 individuals in the vonHoldt study. In addition, we classified the sequenced wolves into 3 main groups: Chinese (n=5), North American (n=2), and Iranian (n=1). These groupings had 9, 54 and 1 individuals, respectively in the vonHoldt data. To implement this approach, we first selected the SNPs that were among the 47,844 high-quality SNPs in

vonHoldt et al. with those SNPs where we called genotypes from our high converge sequencing data within the neutral and coding regions. This left a set of 589 SNPs.

We then performed principal components analysis (PCA) with the program "smartpca" from the EIGENSOFT package (11) using these 589 SNPs. Our sequenced genomes cluster based upon patterns of population structure, rather than by type of data (SI Appendix, Fig. S3A). Specifically, we find that PC1 primarily separates dogs from wolves and, less dramatically, Basenji (an ancient breed) from the other dogs. This pattern is seen both within the SNP genotype data as well as our sequenced genomes. PC2 separates the boxer from the other genomes. This is due to the known ascertainment effect of the SNPs (the dog reference genome is a boxer), but because we focused on using the same set of SNPs from the SNP genotype data in our sequencing data, we see this same effect in both types of data. Overall, for the first 5 PCs, we find that patterns of genetic variation are concordant between the genotype data and the sequencing data. In other words, the sequenced individuals have PC values that are in line with what was seen in the genotyped individuals. These findings suggest that batch effects due to different sequencing centers or technology are small relative to the true signals of population structure in our data.

Nonetheless, we did observe a few instances where the sequenced individuals showed slight differences in PC values compared to the genotyped individuals. For example, on PC1, some of the sequenced Chinese wolves appear to be slightly higher than the genotyped Chinese wolves. This pattern may be due to differences in population structure between the two sets of Chinese wolves, as the sampling locations of wolves our study differed to those in vonHoldt et al. Nevertheless, these differences are slight relative to the main axes of variation present in the data.

The Iranian wolf also shows differences between the genotype and sequence data. Because the sequencing data and the genotype data were from the same individual, we examined concordance between these two technologies in greater detail. Specifically, considering the intersection of the neutral and coding regions in the sequencing data with the SNPs on the Affymetrix chip, 131 sites were called heterozygous in the sequencing data. Of these, 130 were also called heterozygous in our genotype data, indicating that we have a false-discovery rate of <1% for heterozygotes identified in our high-coverage sequencing data. Of 141 sites were called heterozygous in the SNP genotype data, 130 were called heterozygous in our sequencing data. This suggests that we have a false-negative rate of 7.8% for heterozygous genotypes in our data. Interestingly, we estimate a similar false-negative rate for heterozygous SNPs within our data using a completely orthogonal approach and pattern in the data (see "***Accumulation of deleterious derived alleles***" below). The majority of heterozygotes (9/11) called in the genotype data that were not detected in our sequencing data had lower than average sequencing depth (≤10x). In principle, we could impose a more strict depth filter on the final genotype calls, which would remove many of these sites. While this would improve our estimates of the apparent false-negative rate (because we would no long be counting these sites) and may improve the genotypes used in the PCA analysis, it will not reduce the overall false-negative rate in our data. This false-negative rate is a function of coverage that can only be improved by drastically increasing the depth beyond 15x.

Nevertheless, our false-negative rate is comparable to or lower than that in recent whole-genome sequencing studies of non-human mammals. For example, when comparing to Sanger sequencing, Wang et al. (9) reported false-negative rates of at least 15-20% for SNPs in six dog and wolf genomes (their Table S3). Additionally, Auton et al. (12) reported a false-negative rate

of 10-15% in their genome resequencing data (their Fig. S2). Finally, Fig. S2 of Nevado et al. (13), suggests that in simulated data, the false-negative rate for heterozygous genotypes individually called from > 15X data using GATK was close to 20% with the false-negative rate for homozygous non-reference genotypes was approximately 8%. We estimate our false-negative rate to be lower than these numbers.

The estimates of heterozygosity seen for the neutral regions in our high coverage genomes are in agreement with or exceed estimates presented in previous work. For example, Wang et al. (9) report a mean genome-wide estimate of heterozygosity of 1.41×10^{-3} per bp in wolves, which falls within the range seen for wolves in our data (0.95×10^{-3} to 1.9×10^{-3}). They report a mean genome-wide estimate of heterozygosity of 0.73×10^{-3} in breed dogs, which again falls within the range seen in our data (0.64×10^{-3} to 1.3×10^{-3}). It is not surprising that our estimates are generally higher as we have a lower false-negative rate and we are considering data from putatively neutral regions of the genome which generally harbor greater diversity than the genome-wide average (7, 14, 15).

Importantly, our analyses are either robust to this false-negative rate, or we explicitly correct for it as described below. First, our analysis of the ratio of 0-fold to 4-fold heterozygosity should be robust to under-calling heterozygotes. If the rate of under-calling does not systematically differ across the genome or across populations, then heterozygotes at both 0-fold and 4-fold sites will be under-called to a similar extent and will directly cancel when taking the ratio. Similarly, even if one functional category of sites has a greater degree of under-calling than the other, this still should not lead to differences in the 0-fold to 4-fold ratio between dogs and wolves. For example, if we under-call heterozygosity at 4-fold sites more than at 0-fold sites, this would lead to an increase in the 0-fold/4-fold ratio in both dogs and wolves. It should not lead to

a difference in the ratio between the populations. The only way that under calling could lead to the patterns seen in Fig. 1 is if the rate of under calling differs between dogs and wolves as well as across functional categories in a specific way. There is no evidence to support this, and the fact that we see a negative correlation between the 0-fold/4-fold heterozygosity ratio and neutral heterozygosity in the much larger low-coverage dataset argues against under calling driving these patterns.

Naïve analysis of the number of derived alleles per individual, however, is confounded by under-calling heterozygous sites. The reason for this is that the total number of derived alleles is a composite count of both heterozygous and homozygous derived genotypes. Because wolves have more heterozygous sites than dogs, under-calling heterozygous genotypes (even by the same amount in both dogs and wolves) will lead to an apparent decrease in the number of derived alleles in wolves relative to dogs. We have developed a method to correct for this bias (see "***Accumulation of deleterious derived alleles***"). Unless otherwise specified, all analyses of counts of variants per individual include this correction.

Lastly, our analysis of deleterious variation surrounding selective sweeps should be robust to under-calling heterozygotes genotypes. The reason for this is that we compare the sweep regions to the rest of the genome in the same individuals. This comparison serves as a built-in control.

Classifying variant sites as non-synonymous and synonymous with SIFT

We classified variants as non-synonymous or synonymous with SIFT (ref. 16) implemented through Ensemble's Variant Effect Predictor (VEP) v77 (ref. 17). Following VEP recommendations, sites were classified as nonsynonymous if annotated as missense variant, initiation codon variant, inframe insertion or inframe deletion, and synonymous if annotated as

stop retained variant or synonymous variant. All other annotations were ignored. Variants annotated as both synonymous and nonsynonymous (due to differing predictions from different transcripts) were classified as nonsynonymous.

Analysis of runs of homozygosity (ROHs)

To assess whether the negative correlation between neutral heterozygosity and the 0-fold/4-fold heterozygosity ratio was driven by recent inbreeding, we repeated our analysis removing large ROHs. Longer ROHs are thought to have arisen due to recent common ancestry or inbreeding, while shorter ROHs could be due to long-term demographic effects, such as population bottlenecks (18). By analyzing only patterns of variation found outside of the long ROHs, we should remove the effects on genetic variation that are most driven by recent inbreeding.

To identify ROHs, we first applied our GATK genotype calling pipeline described above to the entire autosomal genomes for the 35 high-coverage individuals. It was important to use the entire genome so that ROH partially overlapping with our coding and neutral regions could be most accurately identified. We then used PLINK (19) to identify ROHs that were >2 MB, with a separate analysis conducted for each animal. We chose a cutoff of 2 MB as it was slightly higher than that suggested to identify ROHs due to recent inbreeding in humans (18). Specifically, following Freedman et al. (6) the command used was:

./plink --tfile INFILENAME --homozyg --homozyg-snp 200 --homozyg-kb 2000 --homozyg-window-missing 100 --homozyg-window-het 10 --allow-no-sex --dog --out OUTFILENAME

Then, for each animal separately, we removed any of the coding and neutral sites located within the ROH identified in that animal, and re-ran our analysis of patterns of variation.

Fig. 2C shows that the negative relationship between neutral heterozygosity and the 0-fold/4-fold heterozygosity ratio still persists after removing patterns of variation within the ROHs. This analysis indicates that this pattern is not driven solely by patterns of genetic variation most affected by recent inbreeding. It is further evidence of population size driving the accumulation of amino acid changing mutations.

Accumulation of deleterious derived alleles

We used the following approach to normalize for differences in missing data across individuals. Let L_i refer to the total number of successfully called genotypes in individual i at a particular category of sites (e.g., nonsynonymous). Let H_i refer to the number of called genotypes where individual i is heterozygous. We then computed the normalized number of heterozygous genotypes in individual i ($\tilde{H}_i$) as:

$$\tilde{H}_i = \left(\frac{H_i}{L_i} \right) \bar{L},$$

where $\bar{L}$ is the average number of sites successfully genotyped across all individuals. A similar procedure was used for the number of homozygous derived genotypes and total number of derived alleles.

Using this procedure, we find a significantly higher number of derived (i.e., non-jackal) alleles in the dogs than the wolves for all categories of sites (SI Appendix, Table S7 and Fig. S10D for synonymous SNPs). For neutral variants, this finding is unexpected, as standard coalescent theory predicts that the bottlenecks in dogs relative to wolves should not affect the number of non-jackal alleles per genome, but whether they are partitioned into heterozygous or

homozygous genotypes. The increase in non-jackal alleles in dogs relative to wolves could also be driven by other demographic processes such as admixture between jackal and wolves, though previous studies detected evidence of only limited admixture (6). Further divergence between jackal and dog was similar to the divergence between jackal and wolf (0.2%) indicating that admixture probably does not account for this pattern.

Instead, we believe that this increase in non-jackal (i.e., derived alleles) in dogs is due to differential missing data rates for heterozygous and homozygous genotypes. Note that the false-negative rate (i.e., the probability that we do not call a genotype) will be higher for heterozygous than homozygous genotypes. The reason for this is that it is generally harder to call heterozygous genotypes than homozygous genotypes. As long as this effect is roughly constant across populations and individuals, it should not bias the findings of different numbers of heterozygous and homozygous genotypes between dogs and wolves. However, because the number of derived alleles per genome is a function of both the number of heterozygous and homozygous derived genotypes and the relative contribution of the two genotypes to the number of derived alleles significantly differs between dogs and wolves, the apparent overall number of derived alleles per genome may be biased. Because wolves carry more of their derived (non-jackal) alleles in the heterozygous state than do dogs, the difficulty in calling heterozygous genotypes will cause fewer derived alleles to be detected in wolves compared to dogs. This effect could then lead to an apparent increase in the number of derived alleles shared between dog and jackal. We stress that this effect does not require the false-negative rate for heterozygous genotypes to differ between dogs and wolves.

We developed a simple strategy to correct for the different false-negative rates across genotypes. Let β be the false-negative rate or the probability of not calling a heterozygous

genotype. Further Let $He_{O,D}$ refer to the observed number of heterozygous genotypes in dogs, $He_{O,W}$ be the observed number of heterozygous genotypes in wolves, Ho_D be the observed number of homozygous derived genotypes in dogs and Ho_W be the observed number of homozygous derived genotypes in wolves. We can now write the true number of heterozygous genotypes in wolves as:

$$He_{T,W} = \frac{He_{O,W}}{1-\beta}.$$

The total number of derived alleles in dogs, D_D, is equal to the true number of heterozygotes plus twice the number of homozygous genotypes, and is written as:

$$D_D = He_{T,D} + 2Ho_D.$$

Under the assumption that the total number of derived alleles per animal is equal between dogs and wolves, we can write:

$$He_{T,W} + 2Ho_W = He_{T,D} + 2Ho_D.$$

We can then re-write the above expression as a function of the observed heterozygous counts and the false-negative rate:

$$\frac{He_{O,W}}{1-\beta} + 2Ho_W = \frac{He_{O,D}}{1-\beta} + 2Ho_D.$$

Rearranging,

$$\beta = 1 - \frac{He_{O,W} - He_{O,D}}{2(Ho_D - Ho_W)}.$$

We applied the above estimator of β to the observed median counts of heterozygous genotypes in dogs and wolves at synonymous variants. We found that $\beta = 0.0964$, suggesting that approximately 10% of true heterozygous genotypes are called as missing data in both dogs and wolves. We then used this value of β to estimate $He_{T,D}$ and $He_{T,W}$ and the numbers of derived

alleles for all categories of sites, defining the total number of derived alleles per genome as $He_T + 2(Ho)$. After applying this correction, dogs no longer show an increase in the number of derived alleles per genome at synonymous sites, suggesting that this correction is appropriate (SI Appendix, Table S7).

Note that our estimate of β assumes that the false-negative rate for homozygous genotypes is 0. This is likely to be an underestimate. However, the relative number of derived alleles between populations is sensitive to the difference in false-negative rates between heterozygous and homozygous derived genotypes, rather than the actual values. Furthermore, using genotype concordance between our sequencing data and SNP genotype data (see "***Evaluation of the quality of the high-coverage data***"), we had estimated a comparable false-negative rate of 9%. Thus, our approach should be a reasonable correction for comparing the number of derived alleles per genome between dogs and wolves.

We compared counts of the per individual number of variants both with (SI Appendix, Fig. S10A) and without (Fig. 3) the Tibetan wolf genome. We excluded this genome because as it was clearly an outlier in terms of having low neutral heterozygosity and an elevated 0-fold/4-fold heterozygosity ratio (Fig. 1C; Fig. 2B and 2C), consistent with its recent history of being from a small isolated population (20). As our goal is to assess the extent to which deleterious variation may have accumulated in dogs since domestication, we believe it is most appropriate to not include this genome as it has been extensively shaped by population history not typical for wolf populations and does not reflect levels of deleterious variation in extant wolf populations. Nevertheless, we observe similar patterns of deleterious variants in dogs and wolves even if we include the Tibetan wolf (SI Appendix, Fig. S10, Table S7).

GERP scores

In order to predict which variants are more likely to be deleterious, we utilized GERP scores (21). This approach computes the number of rejected substitutions at each site by comparing the observed number of differences across the phylogeny to the number expected based on neutral models of evolution. Because it has been previously shown that methods that predict functional consequences based on phylogenetic information, like GERP, are biased against calling variants that are present in the reference genome as deleterious (22–24), we omitted the dog reference genome when computing the GERP scores. As such, we could not use the publicly available previously computed GERP scores for humans. We chose GERP over SIFT because it was unclear how the SIFT scores were affected by this bias and recomputing SIFT scores without the dog would have been more challenging.

Instead, we recomputed GERP scores using the GERP++ code (http://mendel.stanford.edu/SidowLab/downloads/gerp/). We downloaded FASTA alignments of 45 vertebrate genomes aligned to hg19 for the UCSC "KnownGenes" list (knownGene.exon*.fa.gz) from the UCSC genome browser (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/multiz46way/alignments/). We then ran each exon through GERP++, omitting the dog sequence. We used the phylogenetic tree "46way.nh" from the UCSC genome browser that was made using the phyloFit program from 4-fold degenerate sites. We then converted the CanMap3.1 coordinates of our dog SNPs to hg19 coordinates using LiftOver to annotate them with a GERP score. Sites with >4 rejected substitutions per site were considered deleterious (25).

Additionally, we also computed the sum of the GERP score across all nonsynonymous SNPs carried by an individual. As the GERP score is a quantitative metric regarding the number

of substitutions that were rejected by evolution, larger scores indicate that a mutation at that site is likely to be more deleterious than a mutation at a site with a lower score. The sum of the GERP scores per individual provides an estimate of the genetic load as it is influenced both by the number of deleterious variants carried by an individual as well as the degree to which variants are deleterious (26, 27). Specifically, let G_i refer to the GERP score load for individual i:

$$G_i = \left(\frac{\sum_{\text{all het}} R_i}{L_i} \right) \frac{\bar{L}}{(1-\beta)} + \left(\frac{\sum_{\text{all hom}} 2R_i}{L_i} \right) \bar{L},$$

where R_i is the sum of the GERP scores, L_i is the total number of sites with successfully called genotypes in individual i, and $\bar{L}$ is the number of successfully called sites per individual, averaged across all individuals. β is the false-negative rate for calling heterozygous sites. This expression accounts for differences in missing data across individuals as well as under calling heterozygous sites. We count homozygous scores twice to reflect an additive model of load. Importantly, we only considered those sites with GERP scores >2, as scores less than this indicate neutral evolution, accelerated evolution, or poor quality alignments. Regardless, it is not clear that mutations at these sites would lead to an increase in fitness.

Miyata prediction of which amino acid changes were most likely to be deleterious

In addition to GERP scores, we also utilized the Miyata amino acid distance (28). Briefly, amino acid mutations that result in changes to the volume or polarity are considered to be more deleterious than those that do not result in such a physiochemical change. We chose to use the Miyata amino acid distance because previous population genetic analyses have shown that it has a reasonable ability to distinguish putatively neutral human SNPs from those that are deleterious (29). Specifically, Williamson et al. (29) used the Poisson Random Field approach to estimate the strength of purifying selection acting on those variants predicted to be deleterious using the

Miyata distance and variants predicted to be neutral. They found that the variants predicted to be deleterious using the Miyata distance were under stronger purifying selection than those that were predicted to be neutral. The difference in the estimated strength of selection between the predicted neutral and predicted deleterious variants was greater for the Miyata distance than other popular metrics, such as the Grantham distance or BLOSUM matrices. While other popular methods like PolyPhen (30) perform well, the application of PolyPhen to non-human species remains challenging. Thus, we chose to use the Miyata distance.

We applied the Miyata distance to each nonsynonymous variant in our high-coverage genomes where individual genotypes were called using GATK. This was done by assigning a score to each nonsynonymous variant. Scores were taken from Table 1 of Miyata et al. (28). Larger scores indicate a greater chance of an amino acid change disrupting the polarity or volume. As done in Williamson et al. (29), amino acid changes with scores >1.85 were considered to be deleterious.

Forward simulations

Here we describe additional details and parameters used for the forward in time simulations. First we discuss the overall simulation framework, followed by the mutation rates, then demographic parameters, and lastly, the distribution of selective effects.

We conducted simulations under the Poisson Random Field framework. The number of mutations that occur in generation i follows a Poisson distribution with a mean equal to $2N_i ul$, where N_i is the effective population size in generation i, u is the mutation rate and l is the number of independent sites being simulated. Each simulation replicate was performed assuming a sequence length of 10 million independent sites, where each site can contain one or two alleles

and only one mutation can take place in each site. The frequency of a mutation, p_{i+1}, follows a binomial distribution $Bin(N_{i+1},p')$, where:

$$p' = \frac{(1-s)(p^2+Fpq)+(1-hs)pq(1-F)}{q^2+pqF+(1-hs)2pq(1-F)+(1-s)(p^2+Fpq)},$$

p is the frequency of the derived allele in generation i (the previous generation), q is the frequency of the ancestral allele, s is the selection coefficient, F is the inbreeding coefficient, and h is the dominance coefficient. For neutral sites, the value of s is equal to 0. Synonymous sites were also assumed to be neutral and had a value of s equal to 0. Values of s for nonsynonymous mutations were drawn from a gamma distribution of selective effects assuming that, unless otherwise noted, all the mutations were additive (h = 0.5). Simulations were performed under a variety of models of population history. See below for further details on the mutations rates, demographic models, and distributions of selective effects used in the simulations.

At the end of the simulation, we sampled individual animals and computed their heterozygosity. Because some of our simulations included recent inbreeding, and inbreeding results in an increase in homozygosity relative to a randomly mating population, we included its effects in computing heterozygosity. Specifically, the genotype for animal i at site j was drawn from a multinomial distribution with probabilities:

$$P(Genotype) = \begin{cases} p_j^2 + p_j q_j F & \text{Homozygous for the derived allele} \\ 2p_j q_j (1-F) & \text{Heterozygous} \\ q_j^2 + p_j q_j F & \text{Homozygous for the ancestral allele} \end{cases}$$

where p_j is the frequency of the derived allele in the population at site j, q_j is the frequency of the ancestral allele at site j, and F is the inbreeding coefficient used in the simulation.

Heterozygosity was computed by dividing the number of heterozygous sites over the number of sites simulated per simulation replicate (10 million). We also determined the proportion of heterozygous sites when sampling one allele from two different animals. To do this, we sampled two genotypes following the previous equation. Then, we sampled one allele from each of the two genotypes and determined that the site was heterozygous if the two sampled alleles were different. The total proportion of heterozygous sites in two dogs was obtained by dividing the number of heterozygous sites by the number of sites simulated.

We accounted for differences in mutation rates (μ) across 0-fold, 4-fold, and neutral sites using the following approach. First, we assume that CpG sites have a 10-fold higher mutation rate than do non-CpG sites. The mutation rates employed for the neutral, 0-fold and 4-fold sites used for the simulations were dependent on the proportion of CpG sites found within those three categories of sites in humans and our data. We then obtained estimates for the proportion of CpG sites in different functional categories from the literature. Veeramah et al. (31) found that in humans 5.75% of the autosomal 0-fold sites were CpG sites while 9.1% of the autosomal 4-fold sites were CpG sites. To estimate the proportion of CpG sites for neutral sites, we examined the proportions of all sites occurring within a 'CG' motif (i.e., any site after a C and any site before a G) from our data. We observed that 'CG' motifs were 9.1% more frequent in 0-fold sites than in neutral sites. While not all of the 'CG' motif sites are necessarily CpG sites, we predict there should be around 9.1% more CpG sites at 0-fold sites compared to neutral sites. Therefore, we reasoned that the proportion of CpG sites in neutral sites was equal to 5.27%. Using these numbers, we obtained the mutation rates of different categories of sites as:

$$\mu = P(CpG)(10)(B) + (1 - P(CpG))(B),$$

where P(CpG) is the proportion of sites that are CpGs in that category of sites and B is the background mutation rate for non-CpG sites. The factor of 10 indicates the assumed 10-fold increase in the mutation rate at CpG cites. We began with a neutral mutation rate of $\mu = 2 \times 10^{-8}$. Then, for neutral sites with P(CpG) = 5.27%, we obtain $B = 1.356 \times 10^{-8}$. That equation can also be used to obtain the mutation rate for 0-fold and 4-fold sites by using that same value of B and replacing P(CpG) by 5.75% and 9.1%, respectively. Using this procedure we obtain mutation rates of $\mu = 2.468 \times 10^{-8}$ for 4-fold sites and $\mu = 2.059 \times 10^{-8}$ for 0-fold sites. These mutation rates were used for the forward simulations.

We examined three different models of population history for canids. First, we used the demographic model for wolves and dogs inferred in Freedman et al. (6) to simulate genetic variation that mimics the demographic history of wolves, village dogs, and breed dogs (SI Appendix, Table S3). Because we assumed a per-base pair per-generation neutral mutation rate of 2×10^{-8}, and Freedman et al. assumed a mutation rate of 1×10^{-8}, we rescaled the N_e and divergence times from the Freedman et al. study. Breed dogs were assumed to have been formed 100 generations ago to be consistent with the historical records and previous work (32, 33). This breed formation was modeled as a decrease in population size. We explored different realistic effective population sizes for the most recent (around 2,500) generations in all populations to assess their effect on neutral heterozygosity and the ratio of nonsynonymous to neutral site heterozygosity. The values shown in SI Appendix, Table S3 show the final parameter values used in the simulations. Our second scenario also used the Freedman et al. (6) demographic model as a backbone. But, here we increased the effective population size in the second epoch from 44,993 to 60,000 individuals. As expected, this model showed higher values of neutral heterozygosity. The parameters of this model are given in SI Appendix, Table S4. Finally, the

third model that we considered was that fit to village dogs and wolves by Wang et al. (9) (SI Appendix, Table S5). We made several simplifying assumptions and replaced exponential growth with piece-wise constant population sizes (SI Appendix, Table S5).

Because the distribution of selective effects has not been estimated for new nonsynonymous mutations in dogs, the optimal parameters to use are not immediately clear. Thus, we examined different distributions of selective effects for new nonsynonymous mutations (SI Appendix, Table S6). First we used estimates from other species. We fully acknowledge the distribution of selective effects may vary across species and these values may not be appropriate for dogs. However, our goal here is to determine whether plausible distributions of selective effects combined with demography can generate the qualitative patterns seen in our data, rather than perform a rigorous assessment of model fit. First, we used the gamma distribution that had been fit to human nonsynonymous SNP data by Boyko et al. (34). Second, we used a gamma distribution that had been fit to nonsynonymous SNPs in 10 *M. m. castaneus* individuals (35). Importantly, because the ß (or scale) parameters of the gamma distribution are typically estimated as the population scaled selection coefficients ($2Ns$), we converted values of $2Ns$ drawn from the distribution into values of s by dividing by twice the relevant population size (SI Appendix, Table S6).

However, we found that both of these distributions of selective effects did not match the regression parameters relating the 0-fold/4-fold ratio and neutral heterozygosity for the observed data (SI Appendix, Fig. S7). In particular the Boyko et al. (34) model from humans predicted a 0-fold/4-fold ratio that was too low compared to our data. This suggests that our data contains more nearly neutral ($|s| < 0.0001$) mutations than had been estimated from humans. Models including a few percent more mutations with $|s| < 0.0001$ better fit the observed data. The

Halligan et al. mouse (35) model (SI Appendix, Table S6), which includes more mutations with $s < 0.0001$, better matches the observed 0-fold/4-fold ratio in dogs, but does not have a steep enough slope. This model contains too few moderately deleterious mutations ($0.0001 < |s| < 0.01$) that could be effectively removed by selection from the wolf population, but persist due to drift in dogs.

There are several possible reasons for this lack of fit of previous models to our data. First, the distribution of selective effects could be different in canids than in humans and mice. The human and mouse distributions appear to differ from each other (SI Appendix, Table S6), supporting the notion that this distribution may not be constant across species. Second, our simulations used to generate the relationship between the 0-fold/4-fold heterozygosity ratio and neutral heterozygosity assume that all variants are independent of each other. If the real data includes substantial Hill-Robertson effects, then the data could differ from our simulations, even when the correct distribution of selective effects was used. Third, the distribution of selective effects may differ between dogs and wolves, perhaps because of domestication. If more new genetic variants in dogs became neutral after domestication, they may be able to drift to higher frequency, increasing the 0-fold/4-fold ratio. More detailed work on the distribution of selective effects in dogs and wolves is needed to distinguish among these possibilities.

Because previously published distributions of selective effects appeared to not match the observed data, we explored several additional custom gamma distributions (SI Appendix, Table S6). We found that models including a greater proportion of weakly deleterious ($|s| < 0.001$) and fewer strongly deleterious ($|s| > 0.01$) mutations provided a better fit to the data. In particular, a gamma distribution with a shape parameter of 0.3 and scale parameter of 0.05 (in terms of s) predicted regression coefficients intermediate between those seen in the low and high coverage

datasets (Fig. 1B) under the Freedman et al. demographic model shown in SI Appendix, Table S3. Under the Wang et al. demographic model (SI Appendix, Table S5), a gamma distribution with a shape parameter of 0.25 and scale parameter of 0.125 reasonably predicted the observed regression parameters (SI Appendix, Fig. S7). While these distributions mimic the empirical patterns, other more complex distributions may be more biologically reasonable. As discussed above, further work on the distribution of selective effects is necessary to distinguish among these possibilities.

We also performed a set of simulations where all mutations were recessive. Here we used the demographic models shown in SI Appendix, Table S3. We used two different distributions of selective effects, the gamma distribution inferred in Boyko et al. as well as our Gamma Test 2 distribution. Overall, we found that the intercept of the regression of the 0-fold/4-fold heterozygosity ratio on neutral heterozygosity was higher with recessive effects than additive effects (compare SI Appendix, Fig S6 with SI Appendix, Fig S8). This finding is not surprising because, for the same distribution of s, recessive mutations are only selected against in the homozygous state and can thus drift up in frequency and persist in the population more easily than variants with additive effects. In contrast to the additive case, the slope of the regression was weakly positive when assuming fully recessive mutations. This result is in agreement with the recent theoretical findings of Balick et al. (36). Essentially, recessive alleles that survive during a bottleneck will have drifted to higher frequency and have a higher probability of being in the homozygous state compared to the same alleles in non-bottlenecked populations. When in the homozygous state, the recessive deleterious mutations can be removed by selection, leading to the decrease in the 0-fold/4-fold ratio in the bottlenecked population relative to the non-bottlenecked population. Because these simulations do not match the patterns seen in our data,

and simulations including additive effects provide a better fit, we conclude that most segregating amino acid changing variants in dogs and wolves are probably not fully recessive. They may be fully additive, however.

We used our forward simulations to explore whether genetic load is predicted to be higher in dogs than wolves. Specifically, we used the demographic parameters from the Freedman et al. demographic model (SI Appendix, Table S3) and the Custom Test 1 gamma distribution of selective effects (SI Appendix, Table S6). At the end of the simulation, we computed the additive genetic load from the segregating variants. The load for variant *i* was calculated as $L_i = 1 - \bar{w}_i = 1 - (1 - 2q_i(1-q_i)s_i h - s_i q_i^2) = s_i q_i (1-q_i) + s_i q_i^2$, assuming *h*=0.5. Assuming fitness effects are independent and additive across sites, the total genetic load due to *m* segregating deleterious mutations was calculated as $L_m = \sum_{i=1}^{m} L_i$. An additional component of the genetic load comes from variants which recently have become fixed in dogs or wolves. To assess the contribution of the recently fixed mutations, we tabulated the selection coefficients of mutations that fixed within the last 2480 generations. This time represents the recent history after dogs and wolves had split from each other. The total genetic load was computed summing the components from the segregating and fixed deleterious mutations. Overall, the genetic load is approximately 2-3% higher in dogs than wolves (Fig. 3D).

Analysis of coding genetic diversity near vs. far from sweeps

To examine patterns of deleterious variation surrounding selective sweeps, we focused on the 25 dog and 10 wolf high-coverage genomes where we had called genotypes using GATK (3, 4). We assumed that the ancestral state was the allele present in the golden jackal. We used the same sweep regions as those determined in the original studies (10, 37). These regions

represented putative selective sweeps in the ancestral population of breed dogs, presumably related to domestication. We used LiftOver to convert the sweep coordinates to CanFam 3.1.

We then examined four different summaries of genetic variation, Watterson's θ, π, the average derived allele count per SNP, and the average number of derived alleles per 100 bp. We computed these four summaries of genetic variation for four categories of sites: 4-fold sites near sweeps, 4-fold sites not near sweeps, 0-fold sites near sweeps, and 0-fold sites not near sweeps.

Watterson's θ per-site was calculated as:

$$\theta_W = \frac{S}{\sum_{i=1}^{2n-1} \frac{1}{i}},$$

where n is the number of diploids and S is the number of variant sites normalized by the size of the regions. Specifically, S for 0-fold sites near sweeps would be computed as:

$$S = \frac{\sum_{j=1}^{L} I_j}{L},$$

where L refers to the total number of 0-fold sites (both variant and invariant) in the sweep region and I_j represents an indicator function that equals 1 if the site is variable in at least one individual. Fig. 4 shows Watterson's θ per 10 kb.

The average number of pairwise differences per site, π, was calculated for a specific category of sites (e.g., 0-fold sites in sweep regions) as:

$$\pi = \frac{\sum_{j=1}^{L} 2p_j(1-p_j)}{L},$$

where p_j is the frequency of the derived allele in the sample at site j. We found that π was reduced within the sweep regions in dogs consistent with the expected signature of a selective

sweep. Specifically, π per site was 5.0 x 10^{-4} (95% CI: 4.8 x 10^{-4} – 5.2 x 10^{-4}) within the sweep regions, compared to 6.5 x 10^{-4} (95% CI: 6.4 x 10^{-4} – 6.6 x 10^{-4}) outside of them. We did not observe this reduction in wolves. In wolves π was 1.35 x 10^{-3} (95% CI: 1.32 x 10^{-3} – 1.37 x 10^{-3}) at 4-fold sites within the sweep regions and was 1.25 x 10^{-3} (95% CI: 1.24 x 10^{-3} – 1.25 x 10^{-3}) outside of the sweep regions.

The average derived allele count per SNP (DAC_SNP), was calculated for a specific category of sites (e.g., 0-fold sites in sweep regions) as:

$$DAC\_SNP = \left(\frac{2n\sum_{j=1}^{L} p_j}{L} \right) \left(\frac{L}{\sum_{j=1}^{L} I_j} \right) = \frac{2n\sum_{j=1}^{L} p_j}{\sum_{j=1}^{L} I_j},$$

where *n* represents the total number of individuals in the sample (*n*=25 for breed dogs, *n*=10 for wolves. In other words, this metric represents the average count of the derived allele, divided by the number of variable sites. Fig. 4B shows the DAC_SNP.

The average number of derived alleles, or the derived allele count (DAC), was calculated for a specific category of sites (e.g., 0-fold sites in sweep regions) as:

$$DAC = \frac{2n\sum_{j=1}^{L} p_j}{L}.$$

Fig. 4C shows the DAC per 100 bp.

We found significantly more derived alleles (i.e. DAC was higher) per bp at 4-fold sites surrounding selective sweeps than in 4-fold sites across the rest of the genome (Fig. 4C). Models of selective sweeps, however, do not predict an increase in the total number of derived alleles. Rather, hitchhiking models predict that the derived allele count per SNP (DAC_SNP) will be elevated by hitchhiking. In other words, sweep regions are predicted to have fewer SNPs (lower

S), but those that are variable are predicted to have higher derived allele frequency, giving higher DAC_SNP. Consistent with this prediction, we found that DAC_SNP at 4-fold sites is significantly higher in surrounding the sweeps compared to the rest of the genome in dogs (Fig. 4B).

Why is the DAC per 100 bp of 4-fold sites elevated surrounding the sweep? Two hypotheses are that the mutation rate is higher or that the level of selective constraint is lower in regions surrounding the sweeps than the rest of the genome (see next section for further discussion). Under the first hypothesis, regions having a higher mutation rate would tend to show more derived alleles than regions with a lower mutation rate. Note, the fact that Watterson's θ and π are both reduced in the sweep regions could still be compatible with a higher mutation rate near the sweeps; the reduction in these estimators of diversity is simply due to the effect of the sweeps. Put another way, if the mutation rate in the sweep regions were to be the same as in the rest of the genome, then we would expect to see an even greater reduction in neutral diversity surrounding the sweeps. Importantly, whatever the cause, even if the sweep regions have higher mutation rates, they still show the expected signatures of sweeps in dogs but not in wolves. Specifically, finding that DAC_SNP is higher in sweep regions than non-regions cannot be explained by differences in mutation rate. Further, the enrichment of 0-fold variants in the sweep regions compared to the rest of the genome in dogs also cannot be explained by differences in mutation rates as this enrichment is significantly greater in dogs than wolves (SI Appendix, Fig. S14).

We chose to present analyses of the DAC per 100 bp because this statistic is a more meaningful statistic when considering potentially deleterious 0-fold variants. DAC provides an approximate summary of the burden of deleterious variants, or genetic load, contained in

sequence surrounding the sweeps versus the rest of the genome. We found that sweep regions contained 1.2 derived 0-fold alleles per 100 bp, compared to 0.97 derived 0-fold alleles per 100 bp (Fig. 4C). This suggests that the genetic load contributed by regions surrounding the sweep is 1.26 fold greater per 100 bp than the genome wide background (Fig. 4C). Summary statistics of genetic variation within and outside of the sweep regions can be found in "DOG_jacknife_on_sweep_nonsweep_data.txt" (for dogs) and "WOLF_jacknife_on_sweep_nonsweep_data.txt" (for wolves) on Dryad (doi:10.5061/dryad.012s5).

Consideration of additional factors leading to an enrichment of 0-fold variants in the sweep regions

Curiously, levels of neutral and deleterious variation in wolves are higher in the sweep regions compared to the rest of the genome (SI Appendix, Fig. S12). This finding suggests that sweeps may preferentially occur in regions of the genome that have higher mutations rates, although the mean GC content of the sweep regions is similar to the rest of the genome (SI Appendix, Fig. S13). Alternatively, the sweep regions could have lower selective constraint and tolerate more amino acid changing mutations. Another possibility is that deleterious variants may be removed less effectively in wolves due to their linkage to other nearby selected variants (i.e., Hill-Robertson interference (38)). If these effects are greater in the sweep regions than across the rest of the genome, they may explain the increase in deleterious variation in the sweep regions in wolves. Hill-Robertson effects are seen most often in regions of the genome with low recombination rates (39, 40). Because the average recombination rate in the sweep regions (1.09 cM/Mb; SI Appendix, Fig. S13) is similar to the genome-wide average (0.97 cM/Mb from (41)), we conclude that this is explanation is less likely. Lastly, the higher diversity in the sweep

regions in wolves could be due to an ascertainment bias. The sweep regions were initially ascertained through a comparison of dogs to wolves. Power to detect differences in genetic diversity between dogs and wolves may be greater for those regions of the genome showing higher levels of diversity in wolves. Whatever the mechanism for the increased variation in the sweep regions in the wolves, it is unlikely to explain the enrichment of deleterious variants in the sweep regions in dogs because the increase in the 0-fold to 4-fold ratio seen in the sweep vs. non-sweep regions is significantly stronger in the dogs than in the wolves (SI Appendix, Fig. S14).

Another possibility is that the excess of amino acid variation surrounding the selective sweeps could be the direct targets of positive selection. However, hitchhiking of deleterious mutations is a better explanation. The amino acid changes considered in our analyses were not adaptive mutations fixed by hard selective sweeps because we only consider polymorphic amino acid changes and not fixed differences. Thus, in order for positive selection to explain the excess amino acid variation, it would have occurred via partial sweeps or polygenic adaptation acting on coding SNPs. Partial sweeps and polygenic adaptation, however, have a lower probability than hard sweeps of showing elevated XP-EHH values, which were used to ascertain sweeps, and patterns of diversity seen in Fig. 4 (42, 43)). Additionally, there is little evidence from previous studies of domestication loci to suggest that polygenic adaptation on coding SNPs in the main mechanism of adaptation during domestication. Large-effect mutations tend to be structural variants rather than coding SNPs (44, 45). Further, studies in humans and domesticated rabbits have reported greater effects of polygenic adaptation on noncoding rather than coding SNPs (46–48). However, because coding SNPs may also play a role in polygenic adaptation (46, 49), we

cannot formally rule out the possibility that the enrichment of amino acid changing variants in the sweep loci in dogs has been driven by polygenic selection.

Testing for overlap between Mendelian disease genes and genes located in selective sweeps

We tested whether regions of the genome that have been implicated in selective sweeps are enriched for genes that cause Mendelian diseases. To do this, we used two different datasets of Mendelian disease genes. First, we used a set of 170 genes that were reported in the Online Mendelian Inheritance in Animals database (http://omia.angis.org.au/results/?search_type=advanced&gb_species_id=9615&characterised=yes&result_type=gene) to cause Mendelian disease in dogs. We searched the IDs of these genes against a list of 18,514 autosomal genes on the CanFam3.1 dog genome assembly. 145 of the dog Mendelian disease genes were present in our list of 18,514 dog genes and were used for subsequent analyses.

Second, to obtain a larger set of disease genes, we used genes implicated in Mendelian diseases in humans. Specifically, we examined genes reported on the Online Mendelian Inheritance in Man "morbidmap" (http://omim.org/). We restricted our analysis to those autosomal genes confidently identified as causing a Mendelian disease. Consequently, we excluded those entries starting with a "[]", "{}", or "?". A total of 3,545 genes met these criteria. We then intersected these genes with our list of genes found in the dogs and found that 2,535 of the Mendelian disease genes were also found in dogs. This is the final set of genes used for subsequent analyses.

We then intersected these data with three different sets of selective sweeps. First, we used the set of sweeps associated with domestication that are shared across breeds and were described above for the deleterious mutation analysis. Second, Akey et al. (50) identified 155 1-Mb

windows of the dog genome that showed unusually high levels of population differentiation across dog breeds. We used LiftOver to convert the coordinates of the sweep regions to CanFam3.1. 149 of the sweep regions were successfully converted and were used for subsequent analyses. We then examined which of our 18,514 dog genes overlapped with these 1-Mb sweep regions. If a gene had any degree of overlap with the sweep window, it was considered overlapping. Third, we used a set of 523 sweep regions identified by Vaysse et al. (51). Like those from the Akey et al. (50) study, these are putatively recent sweeps which occurred in a sub-set of dog breeds. Specifically, we used the sweeps identified using their S_i statistic, which is sensitive to regions of the genome with lower levels of heterozygosity. We took the regions of the genome from Supplementary Table 5 of their paper showing a FDR P-value < 0.05. We used LiftOver to convert the coordinates of the sweep regions from CanFam2 to CanFam3.1, resulting in 516 regions being successfully converted. Again, any of our 18,514 dog genes that overlapped with any portion of the sweep region were considered to be overlapping the sweep.

We then intersected our list of Mendelian disease genes with the list of genes found in selective sweeps. We computed the expected number of overlapping genes as the expected number of successes for a hypergeometric distribution. Specifically,

$$E[overlap] = n\frac{K}{N},$$

where n is the number of genes in the sweep regions, K is the number of Mendelian disease genes, and N is the total number of genes in the dataset (18,514). We then computed the probability of observing as many or more overlapping genes by chance alone. Lists of genes overlapping the sweep regions can be found in the "Gene_IDs_overlapping_sweep_regions.final.xlsx" file in Dryad (doi:10.5061/dryad.012s5).

Interestingly, we did not detect an enrichment of Mendelian disease genes in selective sweeps that occurred early during dog domestication (i.e., sweeps identified through comparison of dogs and wolves), perhaps suggesting that early and breed selective sweeps involve fundamentally different types of genes (SI Appendix, Table S8, Table S9). Genes selected for breed specific phenotypes tend to have large phenotypic effects, mirroring those responsible for Mendelian disease (52, 53). In contrast, selection of traits in early domestic dogs, before the science of selective breeding, likely operated on a more heterogeneous genetic base over a greater time period allowing for selection of genetic variants with smaller phenotypic effects. Alternatively, if the enrichment of Mendelian disease genes within dog sweeps was due to linkage of Mendelian disease genes to those controlling traits under artificial selection, our results indicate genes selected during domestication would be less likely to be linked to Mendelian disease genes.

Another possible explanation for not detecting an enrichment of Mendelian disease genes in selective sweeps that occurred early during dog domestication could be lower statistical power for the ancient sweep regions (which had 711 genes within the sweep regions compared to around 1600 for the other sweeps). To test this, we performed simulations. Specifically, of the 1663 genes located within the sweeps in the Vaysse et al study, 263 (15.81%) were also disease genes. In the smaller dataset of ancient sweeps, we found only 92 out of 711 genes in the sweep regions also were disease genes (12.94%). Assuming the same level of overlap between sweep genes and disease genes as seen in the Vaysse study, (i.e. p=15.81%), we wished to determine the probability of finding 92 or fewer overlapping genes out of a set of 711 genes. To do this, we, drew the number of genes within the sweep regions that overlapped with a disease gene from a binomial distribution with p=15.81% and n=711. We then tabulated the proportion of replicates

with <93 overlapping genes. We found about 2% of the simulation replicates had fewer than 93 overlapping genes. We then repeated this experiment using the data from the Akey et al study. Here, of the 1632 genes in the sweep regions, 245 overlapped with disease genes (p=15.01%). When assuming p=15.01% and n=711, we found only 6-7% of the simulation replicates had <93 genes in the sweep region also being disease genes. Thus, it is unusual to see a proportion of overlapping genes as small as or smaller than that observed in the ancient sweeps if we assume they have the same level of enrichment as the more recent sweeps.

These results taken together strongly argue that there is a difference between the genes within the newer sweeps and the older sweeps. The lack of concordance across the different sweep datasets is unlikely to be due to limited power due to having a smaller set of ancient sweeps. Assuming that the ancient sweeps have the same degree of overlap with disease genes as the modern sweeps, it is very unlikely (<10%) to see as little overlap as observed in our ancient sweeps.

Testing which genes have the greatest enrichment of 0-fold variants in dogs vs. wolves

We tested whether individual genes showed differential 0-fold to 4-fold ratios in dogs and wolves. To do this, we tabulate the number of 0-fold and 4-fold SNPs segregating in each gene in the 25 high-coverage dogs and the 10 high-coverage wolves. Those 614 genes containing at least 10 coding SNPs in either dogs or wolves were analyzed further. We focused on genes containing at least 10 variants because of lower statistical power for genes with fewer SNPs. We then applied Fisher's exact test to each of these 614 genes. The 10 genes with the smallest *P*-value are shown in SI Appendix, Table S2. None of these genes pass a Bonferroni correction for the number of genes analyzed, suggesting that the genome-wide enrichment of 0-fold to 4-fold variants in dogs relative to wolves is not confined to specific regions. However, we note that, due

to limited numbers of coding SNPs per gene, the gene-based analysis likely has low statistical power.

We applied a similar strategy to test which selective sweep regions showed the greatest difference in the 0-fold to 4-fold ratio between dogs and wolves. Here we tabulated the number of 0-fold and 4-fold SNPs within each sweep region in dogs and wolves. We then applied Fisher's exact test to the counts from those sweep regions containing at least 10 coding SNPs in dogs or wolves.

The single sweep with the most extreme enrichment of amino acid changes is on chromosome 30, the second highest outlier region in a genome-wide selection scan (10), which contains a gene implicated in memory and behavior (*RYR3*). In dogs, this region contains 12 amino acid changing SNPs, but only 4 silent variants. In wolves there are 24 and 25 amino acid changing and silent variants, respectively.

Additional References

1. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26(5):589–595.

2. Li H, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.

3. McKenna A, et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20(9):1297–303.

4. DePristo MA, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43(5):491–8.

5. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842.

6. Freedman AH, et al. (2014) Genome sequencing highlights the dynamic early history of dogs. *PLoS Genet* 10(1):e1004016.

7. McVicker G, Gordon D, Davis C, Green P (2009) Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet* 5(5):e1000471.

8. Lynch M (2008) Estimation of nucleotide diversity, disequilibrium coefficients, and mutation rates from high-coverage genome-sequencing projects. *Mol Biol Evol* 25(11):2409–2419.

9. Wang GD, et al. (2013) The genomics of selection in dogs and the parallel evolution between dogs and humans. *Nat Commun* 4:1860.

10. vonHoldt BM, et al. (2010) Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* 464(7290):898–902.

11. Price AL, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38(8):904–909.

12. Auton A, et al. (2012) A fine-scale chimpanzee genetic map from population sequencing. *Science* 336(6078):193–198.

13. Nevado B, Ramos-Onsins SE, Perez-Enciso M (2014) Resequencing studies of nonmodel organisms using closely related reference genomes: optimal experimental designs and bioinformatics approaches for population genomics. *Mol Ecol* 23(7):1764–1779.

14. Hammer MF, et al. (2010) The ratio of human X chromosome to autosome diversity is positively correlated with genetic distance from genes. *Nat Genet* 42(10):830–831.

15. Lohmueller KE, et al. (2011) Natural selection affects multiple aspects of genetic variation at putatively neutral sites across the human genome. *PLoS Genet* 7(10):e1002326.

16. Ng PC, Henikoff S (2001) Predicting deleterious amino acid substitutions. *Genome Res* 11(5):863–874.

17. McLaren W, et al. (2010) Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* 26(16):2069–2070.

18. Pemberton TJ, et al. (2012) Genomic patterns of homozygosity in worldwide human populations. *Am J Hum Genet* 91(2):275–292.

19. Purcell S, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81(3):559–575.

20. Zhang W, et al. (2014) Hypoxia adaptations in the grey wolf (*Canis lupus chanco*) from Qinghai-Tibet plateau. *PLoS Genet* 10(7):e1004466.

21. Davydov EV, et al. (2010) Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol* 6(12):e1001025.

22. Simons YB, Turchin MC, Pritchard JK, Sella G (2014) The deleterious mutation load is insensitive to recent population history. *Nat Genet* 46(3):220–4.

23. Do R, et al. (2015) No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. *Nat Genet* 47(2):126–131.

24. Fu W, Gittelman RM, Bamshad MJ, Akey JM (2014) Characteristics of neutral and deleterious protein-coding variation among individuals and populations. *Am J Hum Genet* 95(4):421–436.

25. Henn BM, et al. (2015) Distance from Sub-Saharan Africa predicts mutational load in diverse human genomes. *bioRxiv*. doi:10.1101/019711.

26. Casals F, et al. (2013) Whole-exome sequencing reveals a rapid change in the frequency of rare functional variants in a founding population of humans. *PLoS Genet* 9(9):e1003815.

27. Schubert M, et al. (2014) Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc Natl Acad Sci* 111(52):E5661–E5669.

28. Miyata T, Miyazawa S, Yasunaga T (1979) Two types of amino acid substitutions in protein evolution. *J Mol Evol* 12(3):219–236.

29. Williamson SH, et al. (2005) Simultaneous inference of selection and population growth from patterns of variation in the human genome. *Proc Natl Acad Sci U S A* 102(22):7882–7887.

30. Adzhubei IA, et al. (2010) A method and server for predicting damaging missense mutations. *Nat Methods* 7(4):248–249.

31. Veeramah KR, Gutenkunst RN, Woerner AE, Watkins JC, Hammer MF (2014) Evidence for increased levels of positive and negative selection on the X chromosome versus autosomes in humans. *Mol Biol Evol* 31(9):2267-2282.

32. Gray MM, et al. (2009) Linkage disequilibrium and demographic history of wild and domestic canids. *Genetics* 181(4):1493–505.

33. Boyko AR (2011) The domestic dog: man's best friend in the genomic era. *Genome Biol* 12(2):216.

34. Boyko AR, et al. (2008) Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet* 4(5):e1000083.

35. Halligan DL, et al. (2013) Contributions of protein-coding and regulatory change to adaptive molecular evolution in murid rodents. *PLoS Genet* 9(12):e1003995.

36. Balick DJ, Do R, Cassa CA, Reich D, Sunyaev SR (2015) Dominance of deleterious alleles controls the response to a population bottleneck. *PLoS Genet* 11(8):e1005436.

37. Axelsson E, et al. (2013) The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495(7441):360–364.

38. Hill WG, Robertson A (1966) The effect of linkage on limits to artificial selection. *Genet Res* 8(3):269–294.

39. Comeron JM, Williford A, Kliman RM (2008) The Hill-Robertson effect: evolutionary consequences of weak selection and linkage in finite populations. *Heredity* 100(1):19–31.

40. Kaiser VB, Charlesworth B (2009) The effects of deleterious mutations on evolution in non-recombining genomes. *Trends Genet* 25(1):9–12.

41. Wong AK, et al. (2010) A comprehensive linkage map of the dog genome. *Genetics* 184(2):595–605.

42. Pennings PS, Hermisson J (2006) Soft wweeps III: The signature of positive selection from recurrent mutation. *PLoS Genet* 2(12):e186.

43. Liu X, et al. (2013) Detecting and characterizing genomic signatures of positive selection in global populations. *Am J Hum Genet* 92(6):866–881.

44. Andersson L (2013) Molecular consequences of animal breeding. *Curr Opin Genet Dev* 23(3):295–301.

45. Wayne RK, vonHoldt BM (2012) Evolutionary genomics of dog domestication. *Mamm Genome* 23(1-2):3–18.

46. Carneiro M, et al. (2014) Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science* 345(6200):1074–1079.

47. Vernot B, et al. (2012) Personal and population genomics of human regulatory variation. *Genome Res* 22(9):1689–1697.

48. Enard D, Messer PW, Petrov DA (2014) Genome-wide signals of positive selection in human evolution. *Genome Res* 24(6):885–895.

49. Hancock AM, et al. (2010) Human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. *Proc Natl Acad Sci* 107(Supplement 2):8924–8930.

50. Akey JM, et al. (2010) Tracking footprints of artificial selection in the dog genome. *Proc Natl Acad Sci U A* 107(3):1160–5.

51. Vaysse A, et al. (2011) Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLoS Genet* 7(10):e1002316.

52. Karlsson EK, Lindblad-Toh K (2008) Leader of the pack: gene mapping in dogs and other model organisms. *Nat Rev Genet* 9(9):713–725.

53. Ostrander EA (2012) Both ends of the leash — the human links to good dogs with bad genes. *N Engl J Med* 367(7):636–646.

54. Auton A, et al. (2013) Genetic recombination is targeted towards gene promoter regions in dogs. *PLoS Genet* 9(12):e1003984.

55. Lindblad-Toh K, et al. (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438(7069):803–19.

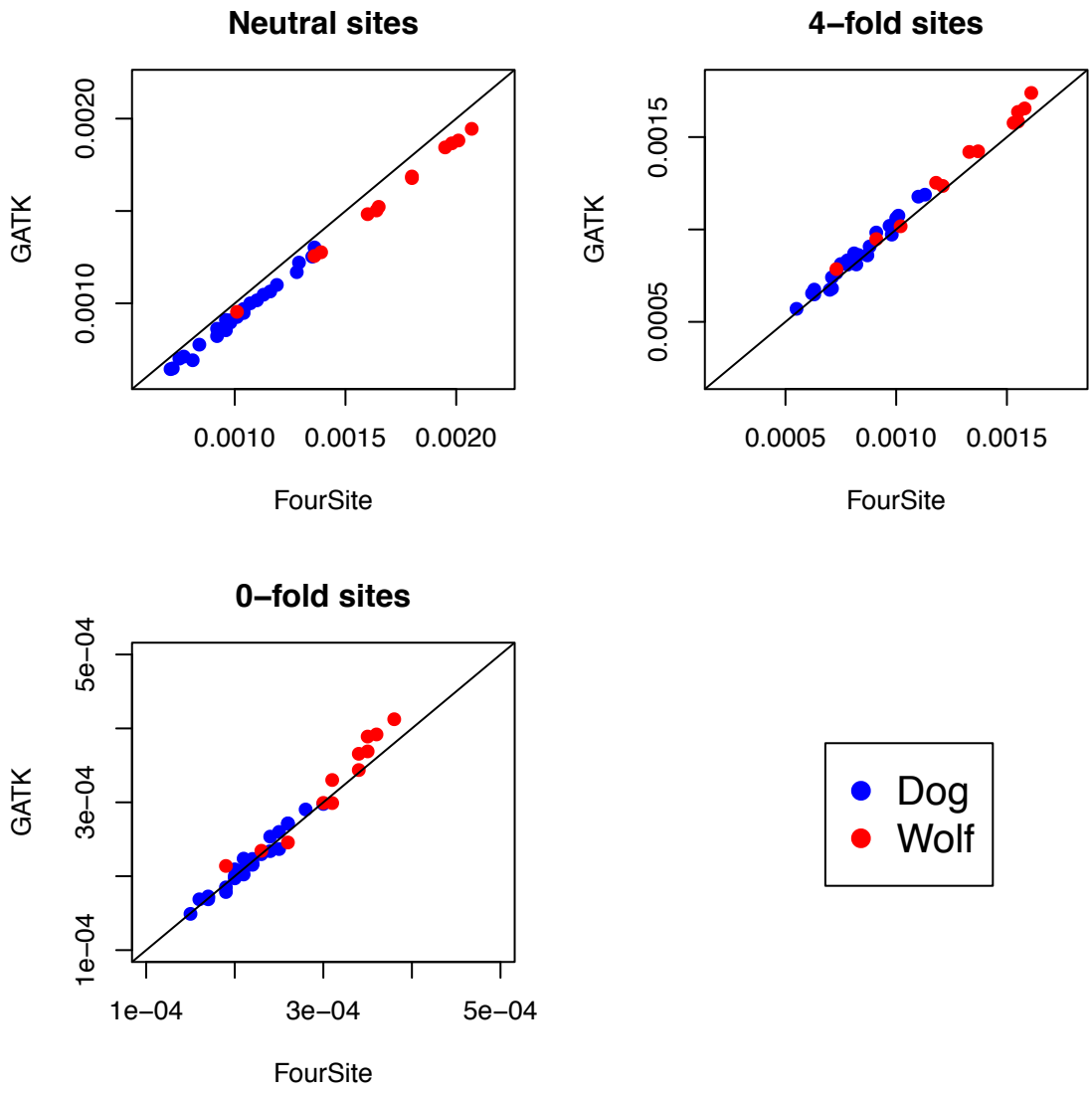

Fig. S1: Comparison of the read-based estimator of heterozygosity (FourSite) to the estimates based on GATK for high coverage individuals.
Lines denote the diagonal. Each blue point represents a breed dog. Each red point represents a wolf. Note the close correspondence between the estimates of heterozygosity obtained using FourSite to those from GATK. Importantly, only 4 reads per individual per site were used with FourSite while all the reads that passed our quality filters were used for calling genotypes with GATK.

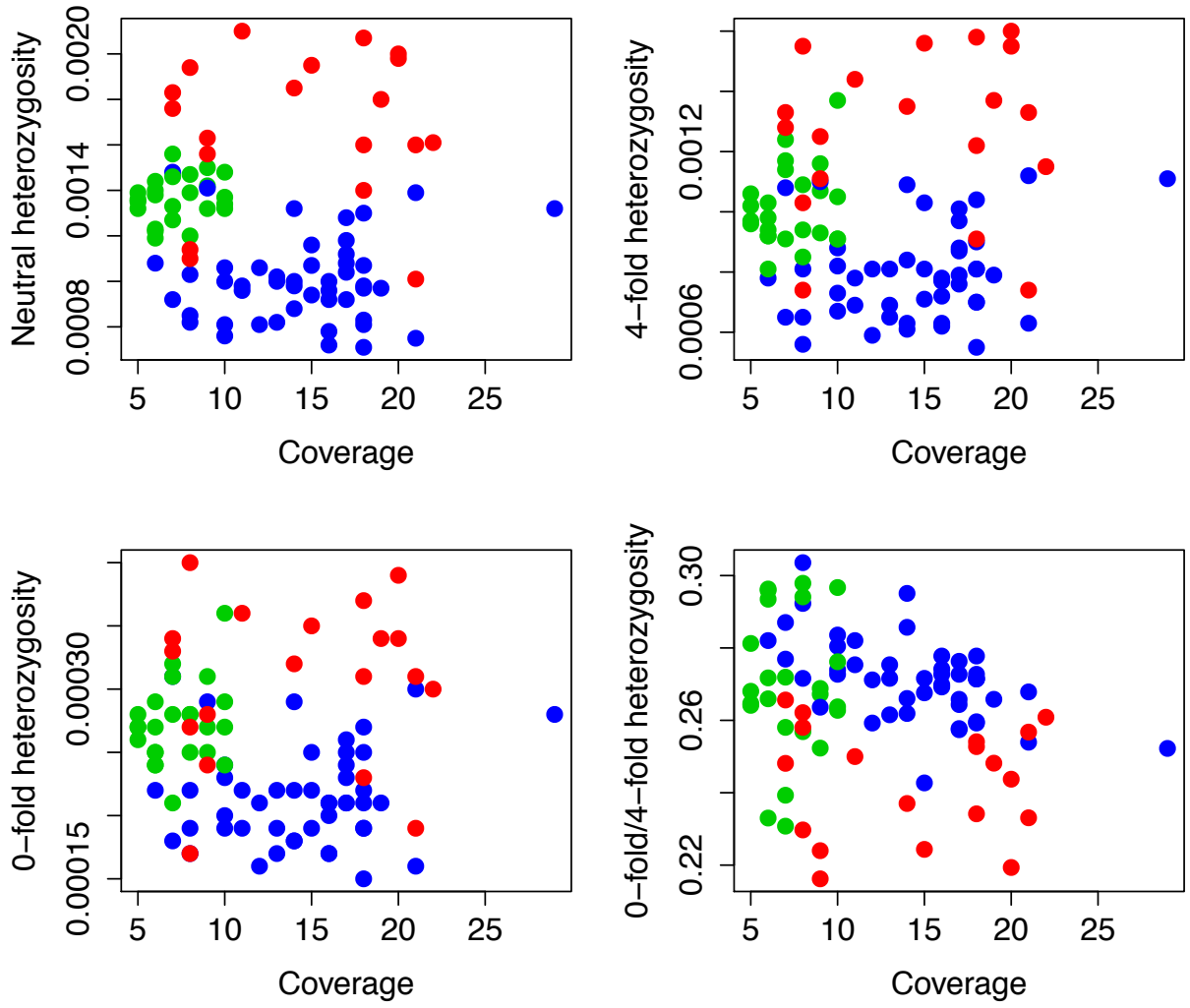

Fig. S2: Estimated heterozygosity vs. average read depth.
Each point represents one of the 90 genomes analyzed in our study. Breed dogs are in blue, village dogs in green, and wolves in red. Note that the average coverage of village dogs is lower than that of breed dogs and wolves, but this does not affect the estimates of heterozygosity. Estimates of heterozygosity for all genomes are made using a subset of 4 reads per site analyzed using the FourSite approach. Neutral heterozygosity (r = -0.09, P = 0.40), 4-fold heterozygosity (r = 0.01, P = 0.95), and 0-fold heterozygosity (r = -0.08, P = 0.45) show no correlation with coverage. There is a weak negative correlation between the ratio of 0-fold to 4-fold heterozygosity (r = -0.24; P = 0.024; see SI Text).

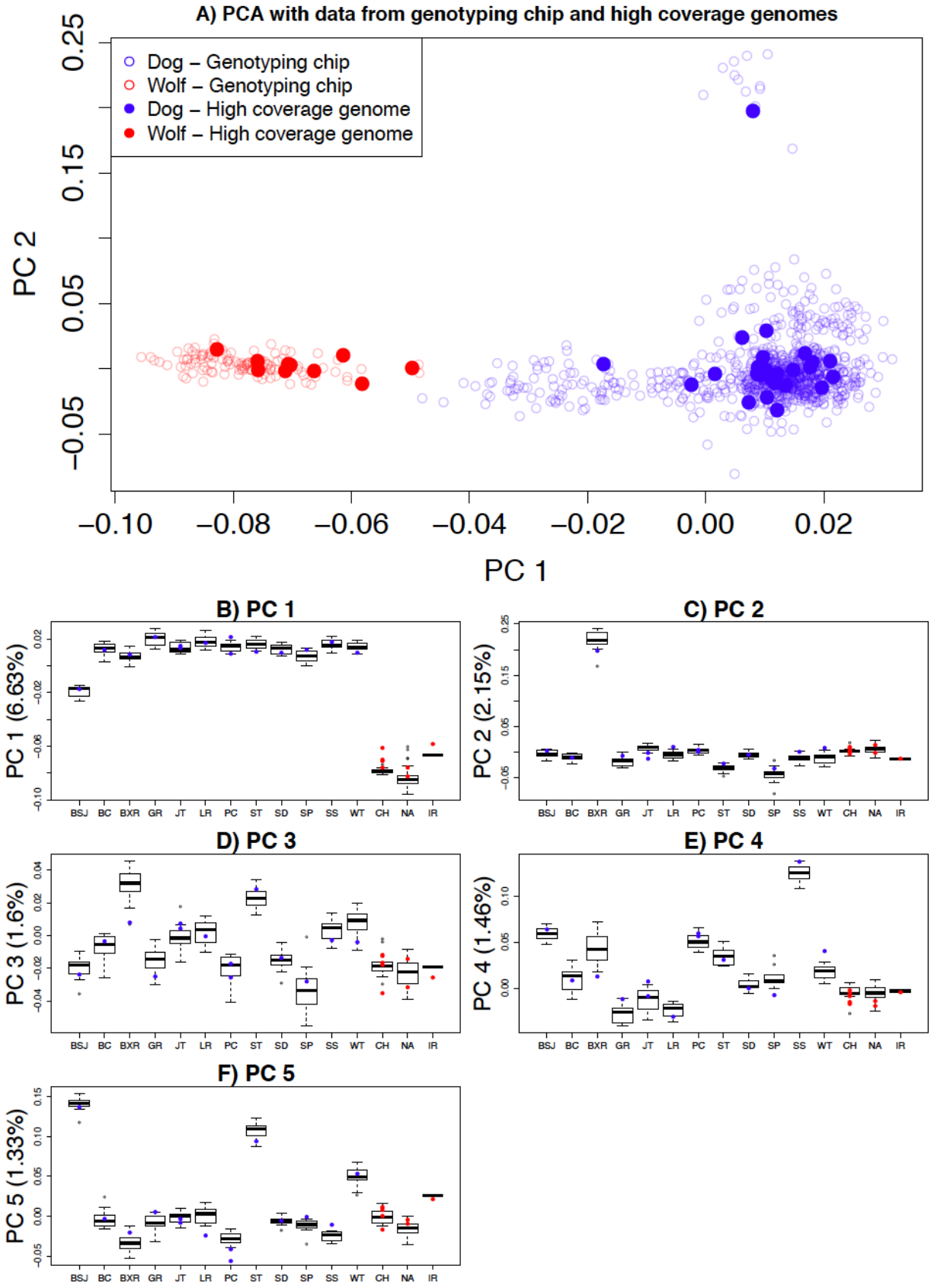
A) PCA with data from genotyping chip and high coverage genomes
Dog – Genotyping chip
Wolf – Genotyping chip
Dog – High coverage genome
Wolf – High coverage genome
PC 2
PC 1
B) PC 1
PC 1 (6.63%)
C) PC 2
PC 2 (2.15%)
D) PC 3
PC 3 (1.6%)
E) PC 4
PC 4 (1.46%)
F) PC 5
PC 5 (1.33%)
BSJ BC BXR GR JT LR PC ST SD SP SS WT CH NA IR

Fig. S3: Principal components analysis (PCA) of the 25 high-coverage genomes along with Affymetrix SNP genotype data from 145 dogs and 64 wolves.
(A) First two PCs. Open circles denote data from SNP genotypes. Filled circles denote the high-coverage genomes. Note that the sequenced dogs cluster with the genotyped dogs and the sequenced wolves cluster with the genotyped wolves. The cluster of points with elevated values on PC 2 are the boxers. **(B-F)** Boxplots of the top 5 PCs for dogs and wolves. Boxes denote the ranges of the PCs from the genotyped dogs. Colored points denote the sequenced genomes. Across all 5 PCs, the sequenced dogs are close to, or within the range of the genotyped dogs. The fact that individuals cluster based on population structure and not by assay suggests that batch effects, sequencing errors, or differences between genotyping and sequencing make a negligible contribution to overall patterns of genetic variation in our data. Population abbreviations for dogs are as follows: BSJ: Basenji, BC: Border Collie, BXR: Boxer, GR: Golden Retriever, JT: Jack Russell Terrier, LR: Labrador Retriever, PC: Pembroke Welsh Corgi, ST: Scottish Terrier, SD: Scottish Deerhound, SP: Standard Poodle, SS: Shetland Sheepdog, WT: West Highland White Terrier. Abbreviations for wolf populations are CH: China, NA: North America, IR: Iran.

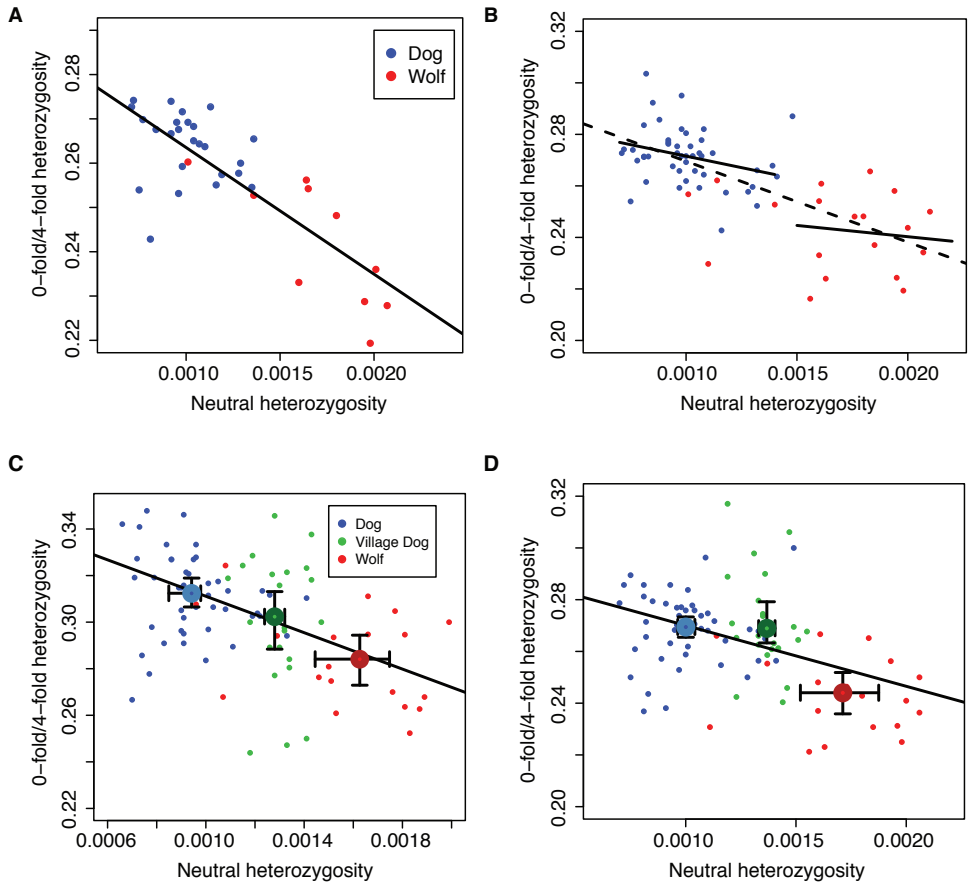

Fig. S4: The ratio of 0-fold to 4-fold heterozygosity is negatively correlated with neutral genetic diversity.
(A) This analysis uses 35 high coverage genomes treated as though they have lower-coverage. We sampled 4 reads per site and estimated heterozygosity using FourSite. The solid line denotes the best-fit linear regression line (Intercept = 0.292, slope = -28.59, $r = 0.795$, $P < 2 \times 10^{-8}$). **(B)** This analysis considers breed dogs (blue) and wolves (red) separately. Heterozygosity was computed using four reads per individual. The dashed line denotes the best-fit linear regression line (Intercept = 0.30, slope = -31.4, $r = -0.674$, $P < 6 \times 10^{-10}$) for both breed dogs and wolves together. Breed dogs show a slight negative relationship (solid line over blue points; Intercept = 0.29, slope = -17.82, $r = -0.30$ $P = 0.043$) while wolves do not (solid line over red points; Intercept = 0.26, slope = -8.75, $r = -0.185$, $P = 0.45$). However, due to the limited sample size, statistical power is diminished within each group. **(C)** This analysis filters CpG sites (see SI Text). Heterozygosity was computed using four reads per individual. Error bars denote 95% confidence intervals on the trimmed median for each population group. The solid line denotes the best-fit linear regression line (Intercept = 0.350, slope = -39.09, $r = -0.468$, $P < 4 \times 10^{-6}$). **(D)** This analysis filters sites near a selective sweep (see SI Text). Heterozygosity was computed using four reads per individual. Error bars denote 95% confidence intervals on the trimmed median for each population group. The solid line denotes the best-fit linear regression line (Intercept = 0.294, slope = -23.50, $r = -0.430$, $P < 3 \times 10^{-5}$).

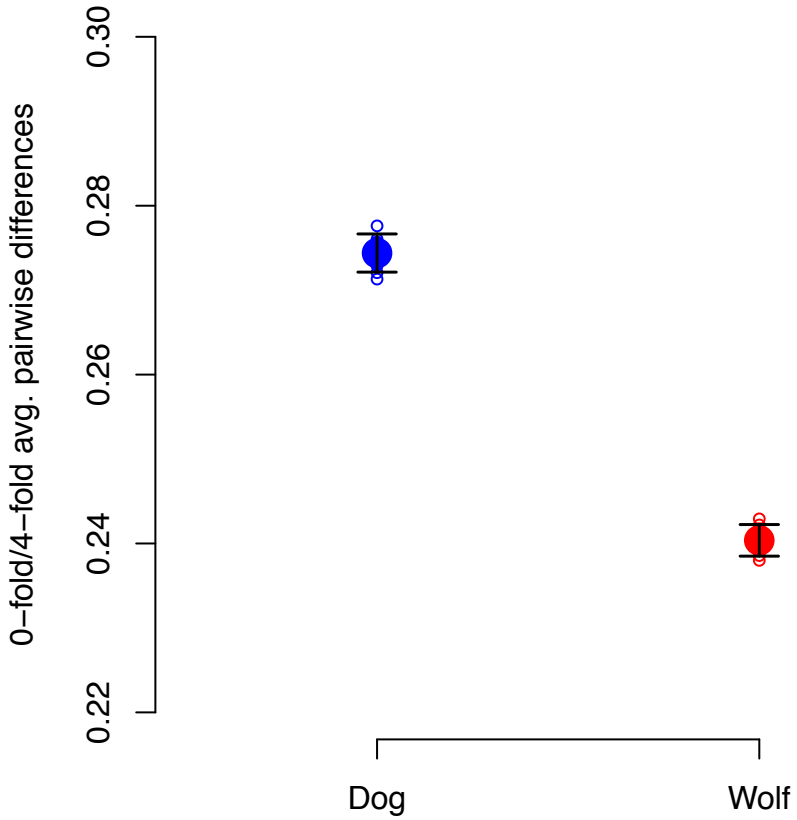

Fig. S5: 0-fold/4-fold ratio of the average pairwise differences is significantly higher in dogs than wolves.
This analysis considers the 25 high-coverage breed dogs and the 10 high-coverage wolves. We computed the average number of pairwise differences at 0-fold sites and 4-fold sites. Regions implicated in selective sweeps (see text) were excluded from this analysis. Error bars represent 95% confidences intervals calculated from estimates of the standard error obtained from jackknifing over chromosomes. Small dots show the estimates for each individual jackknife replicate. Note that this analysis does not assume that individuals are independent of one another and appropriately accounts for shared genealogical history between individuals.

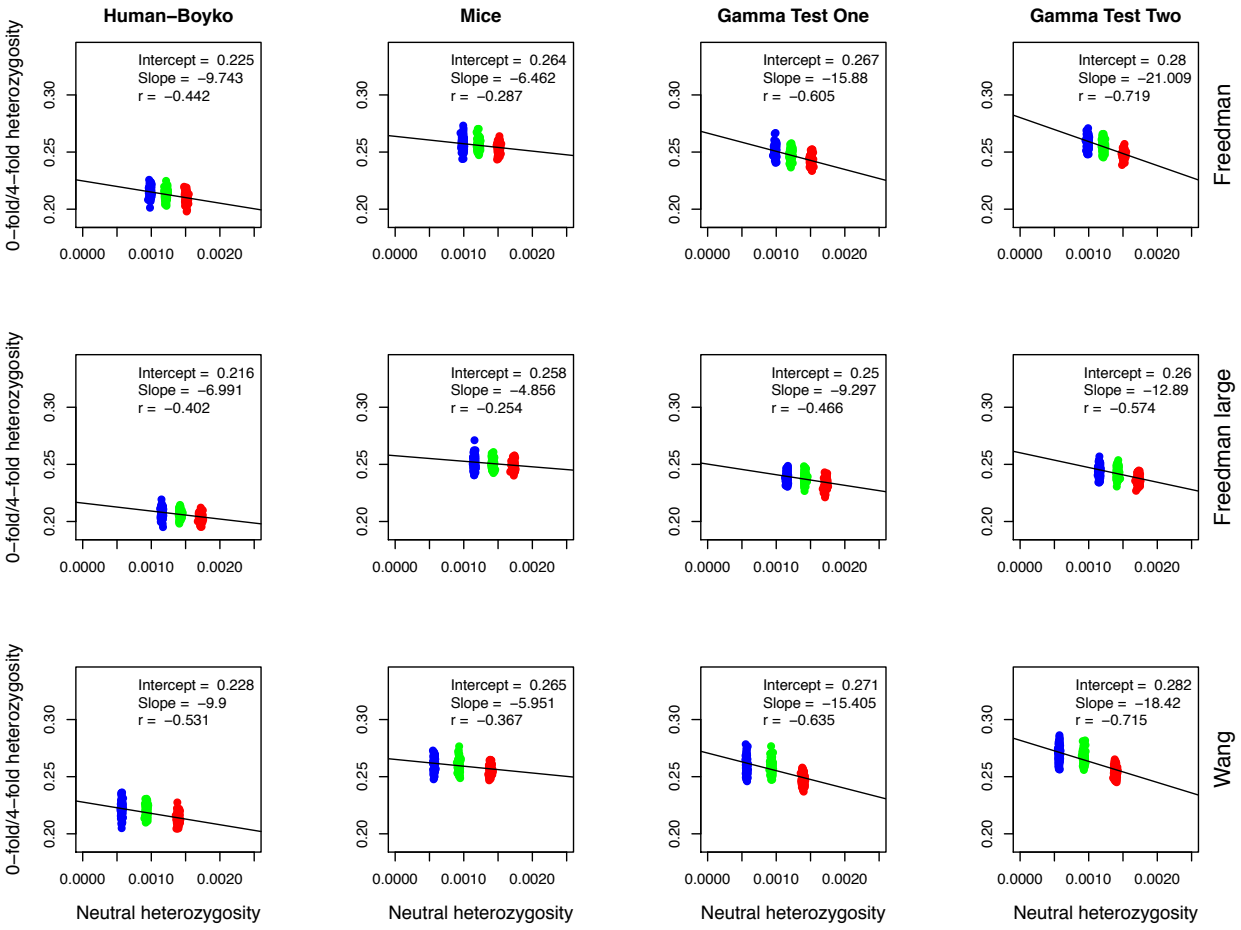

Fig. S6: Ratio of 0-fold to 4-fold heterozygosity vs. neutral heterozygosity from the forward simulations under different models of demography and selection.
Rows denote the different demographic models. “Freedman” refers to the Freedman et al. model (SI Appendix, Table S3). “Freedman large” refers to the Freedman et al. model, but increasing the size of the ancient population size (SI Appendix, Table S4). “Wang” denotes our implementation of the model fit in Wang et al. (SI Appendix, Table S5). Columns denote different distributions of selective effects (SI Appendix, Table S6). Lines are from the best-fit linear regression. Blue points denote breed dogs, green points denote village dogs, and red points represent wolves. In all cases, models of demography and selection predict a negative relationship between the ratio of 0-fold to 4-fold heterozygosity vs. neutral heterozygosity.

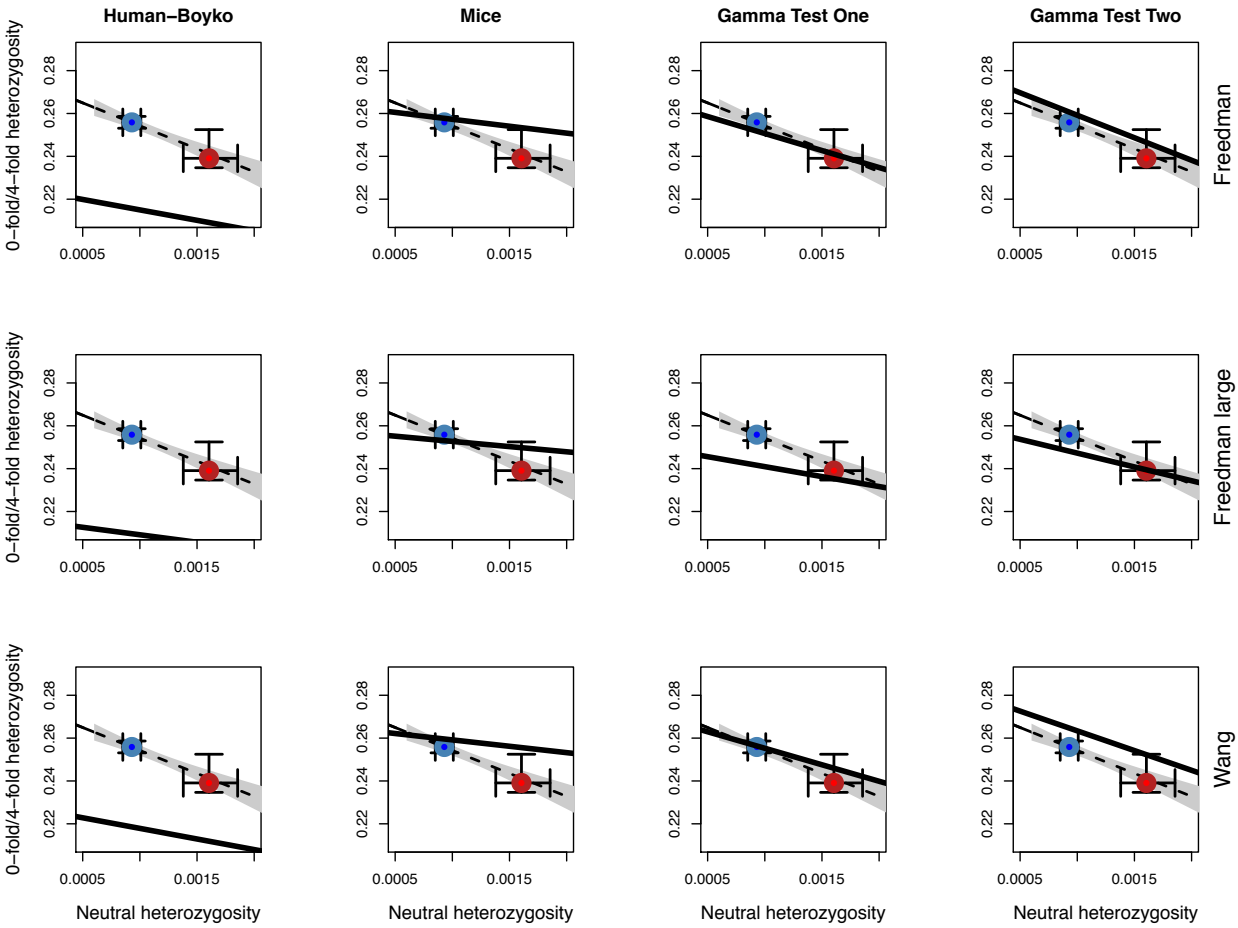

Fig. S7: Models of purifying selection and demography predict a similar negative relationship between 0-fold/4-fold heterozygosity and neutral heterozygosity as seen in the high quality genomes.

Rows denote the different demographic models (SI Appendix, Tables S3-5). Columns denote different distributions of selective effects (SI Appendix, Table S6). Dark solid black lines are from the best-fit linear regression of the simulations under the particular model (e.g., best fit lines in SI Appendix, Fig. S6). The gray shaded region denotes the 95% CI on the linear regression line calculated from the 35 high quality genomes (e.g., the data shown in Fig. 1B). The dark blue and red points represent the trimmed medians from the observed data from the breed dogs and wolves, respectively. The whiskers denote 95% CIs on the trimmed medians. Note that the Gamma Test 2 distribution of selective effects best fits the observed relationship between 0-fold/4-fold heterozygosity and neutral heterozygosity under the Freedman demographic model. The Gamma Test 1 distribution also provides a good fit under the Wang demographic model.

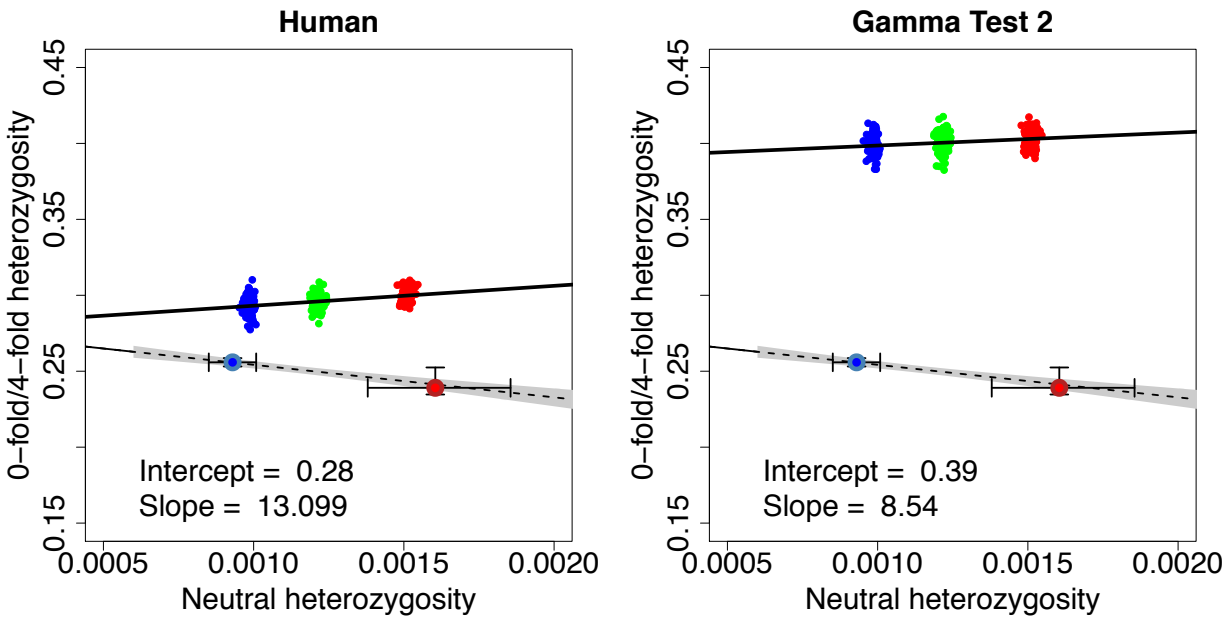

Fig. S8: Models with recessive effects predict a positive relationship between 0-fold/4-fold heterozygosity and neutral heterozygosity.
Breeds dogs are in blue, village dogs in green, and wolves in red. All simulations assumed h=0 and the demographic parameters shown in SI Appendix, Table S3. Columns denote different distributions of selective effects (SI Appendix, Table S6). The shaded gray lines denote the regression parameters from the simulations including additive effects. The clouds of blue, green, and red points denote the results of the simulations assuming recessive effects.

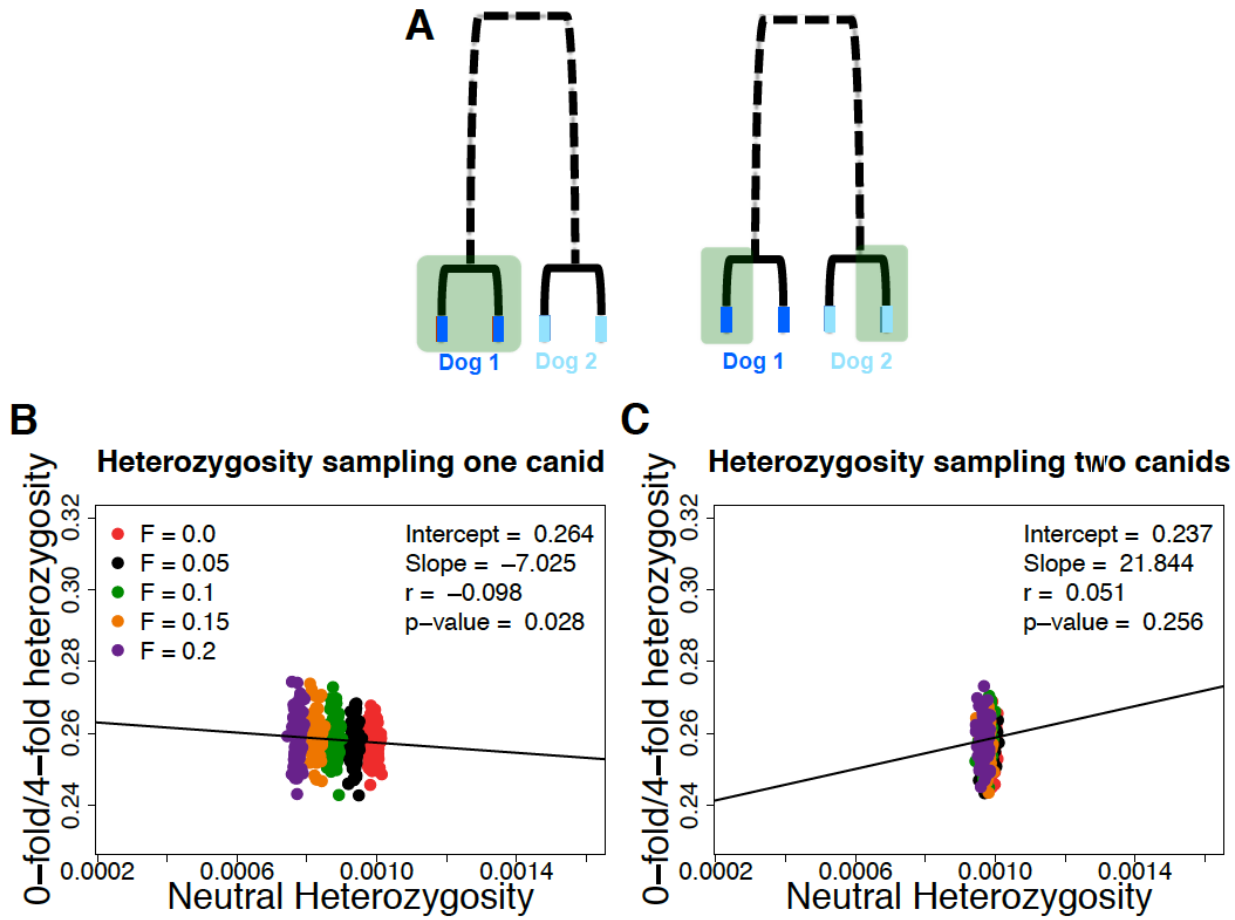

Fig. S9: Estimating heterozygosity using one chromosome from each of two individuals removes the effects of recent inbreeding.
(A) Inbreeding results in an increase in the probability that two chromosomes within an individual share a recent common ancestor with each other than with a chromosome in a different individual (i.e. chromosomes of the same color have a higher probability of coalescing with each other than with chromosomes of a different color). This will lead to a reduction in heterozygosity (left panel). By computing heterozygosity from one read from each individual (i.e. from different colored chromosomes, right panel), we will remove this effect of inbreeding. **(B)** Forward simulations using the breed dog demographic model including population bottleneck along with 100 generations of inbreeding (model from Freedman et al., SI Appendix, Table S3) show a slight negative correlation between the ratio of 0–fold to 4-fold heterozygosity and neutral heterozygosity. This suggests recent inbreeding in certain dog breeds may slightly increase the 0-fold to 4-fold ratio. Lines denote the regression between the ratio of 0–fold to 4-fold heterozygosity vs. neutral heterozygosity. **(C)** Same simulations as in **(B)**, except here heterozygosity is computed using one chromosome from each of two dogs. Sampling from two dogs eliminates the reduction in heterozygosity due to recent inbreeding as well as the weak negative correlation seen in **(B)**.

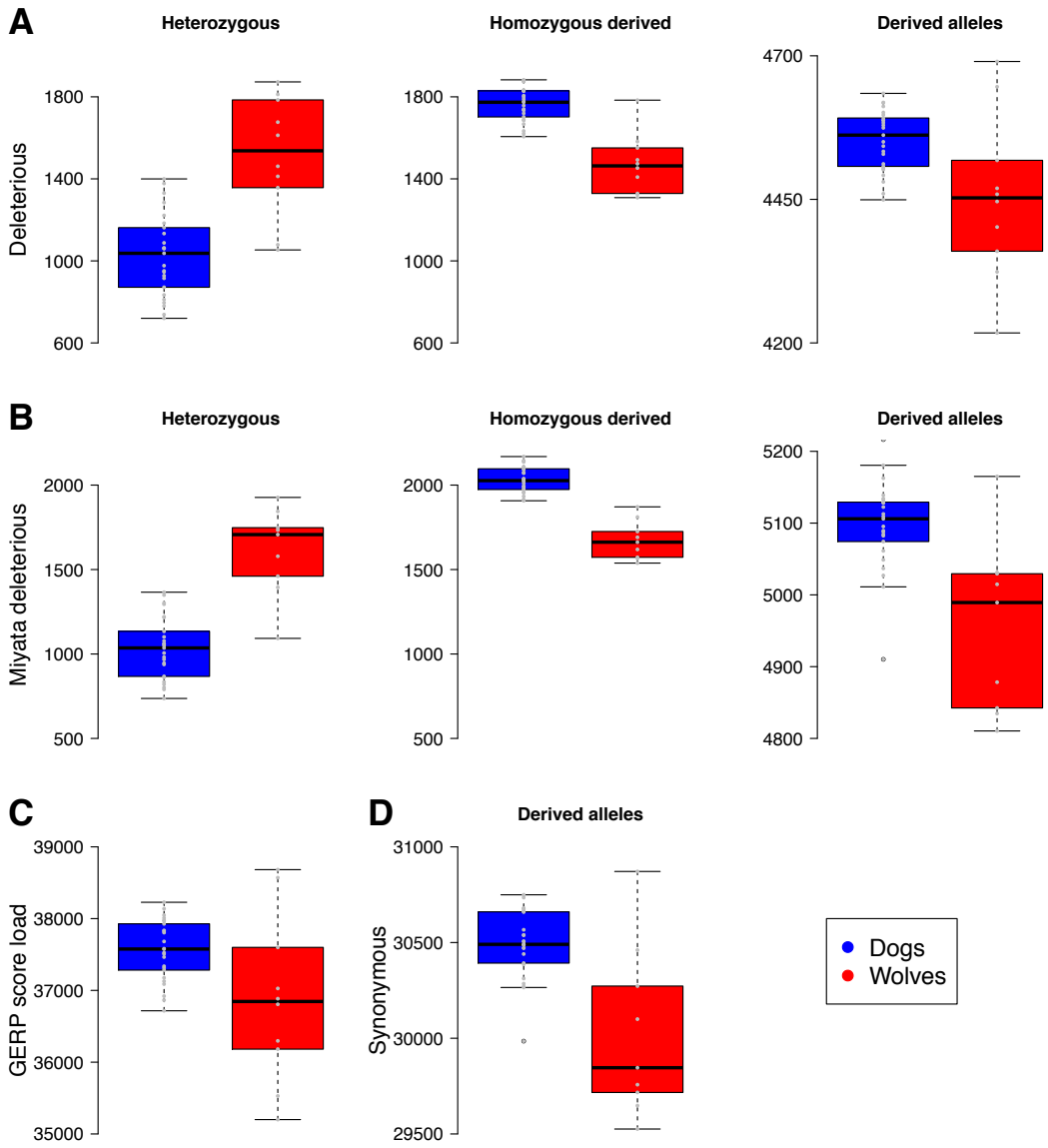

Fig. S10: Additional comparisons of the burden of deleterious genotypes per individual using the high-quality genomes.
Homozygous derived refers to the number of genotypes per individual that are homozygous for the derived allele when using the golden jackal to assign ancestral state. The total number of derived alleles counts each heterozygous genotype once and each homozygous derived genotype twice. **(A)** Nonsynonymous variants with GERP score >4. Here, the Tibetan wolf is included. There is a 2.5% increase in the number of derived deleterious alleles in dogs compared to wolves (P<0.019). **(B)** Nonsynonymous variants that are predicted to be deleterious based on the Miyata classification. Note the 2.3% increase of derived deleterious alleles in dogs compared to wolves (P<0.001). **(C)** Distribution of GERP score load per individual when including the Tibetan wolf. Dogs have a significantly higher GERP score load than do wolves (P<0.046). **(D)** Number of derived synonymous alleles per individual before correction for under calling heterozygotes (P<0.003). Unless otherwise noted above, $P<6 \times 10^{-5}$ for all comparisons using a Mann-Whitney U test (SI Appendix, Table S7).

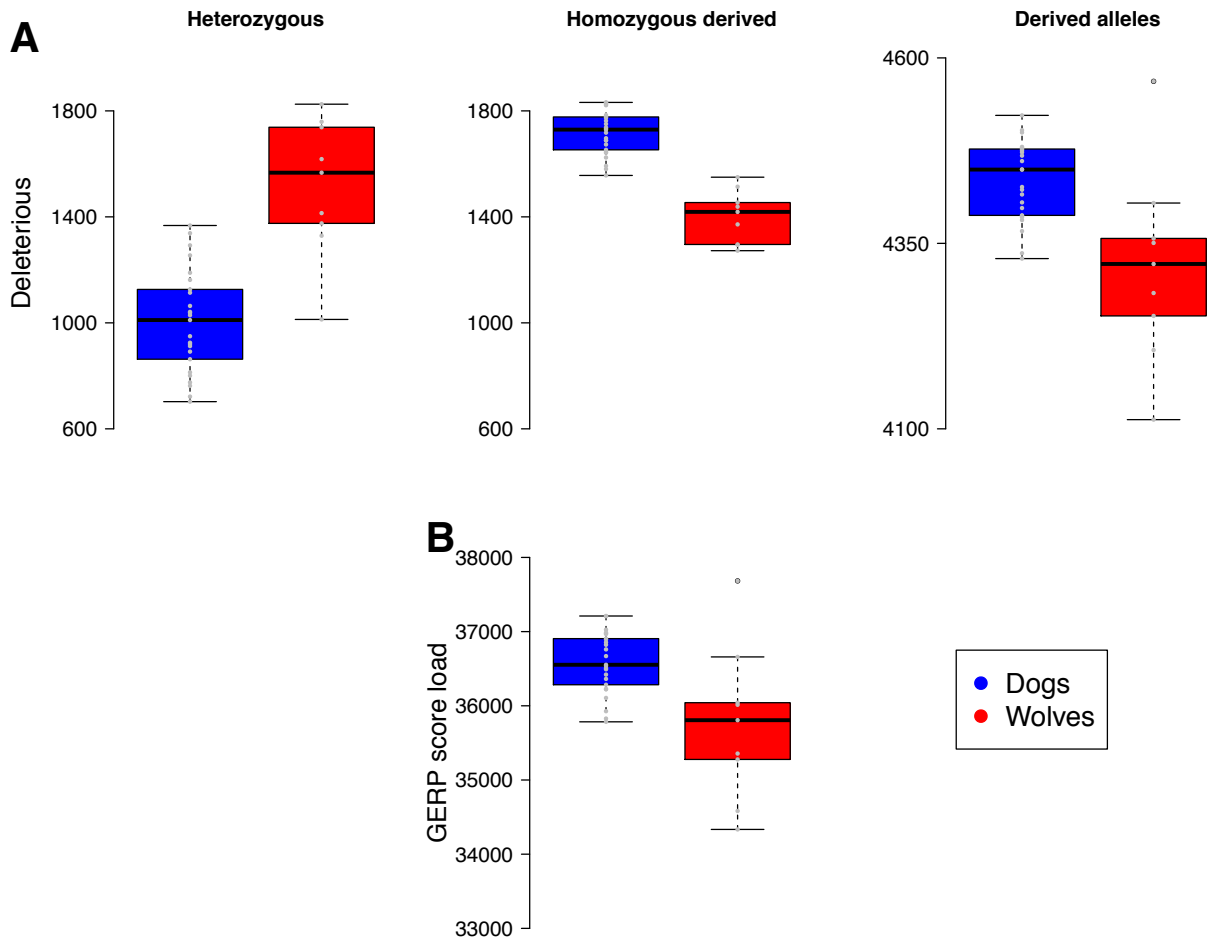

Fig. S11: Comparison of the burden of deleterious genetic variation between breed dogs and wolves based on high-quality genomes when removing selective sweep regions. "Homozygous derived" refers to the number of genotypes per individual that are homozygous for the derived allele when using the golden jackal to assign ancestral state. The total number of derived alleles is based on counting each heterozygous genotype once and each homozygous derived genotype twice after correction for the higher false-negative rate for heterozygous genotypes. Small points denote the number of high-coverage genomes used for each species (n = 25 for breed dogs, n = 9 for wolves). **(A)** Nonsynonymous variants that are predicted to be deleterious based on having a GERP score >4. Breed dogs have approximately 2.9% more derived deleterious alleles than wolves. **(B)** GERP score load for each individual. $P < 0.008$ for all comparisons between dogs and wolves using a Mann-Whitney U test unless otherwise noted (SI Appendix, Table S7).

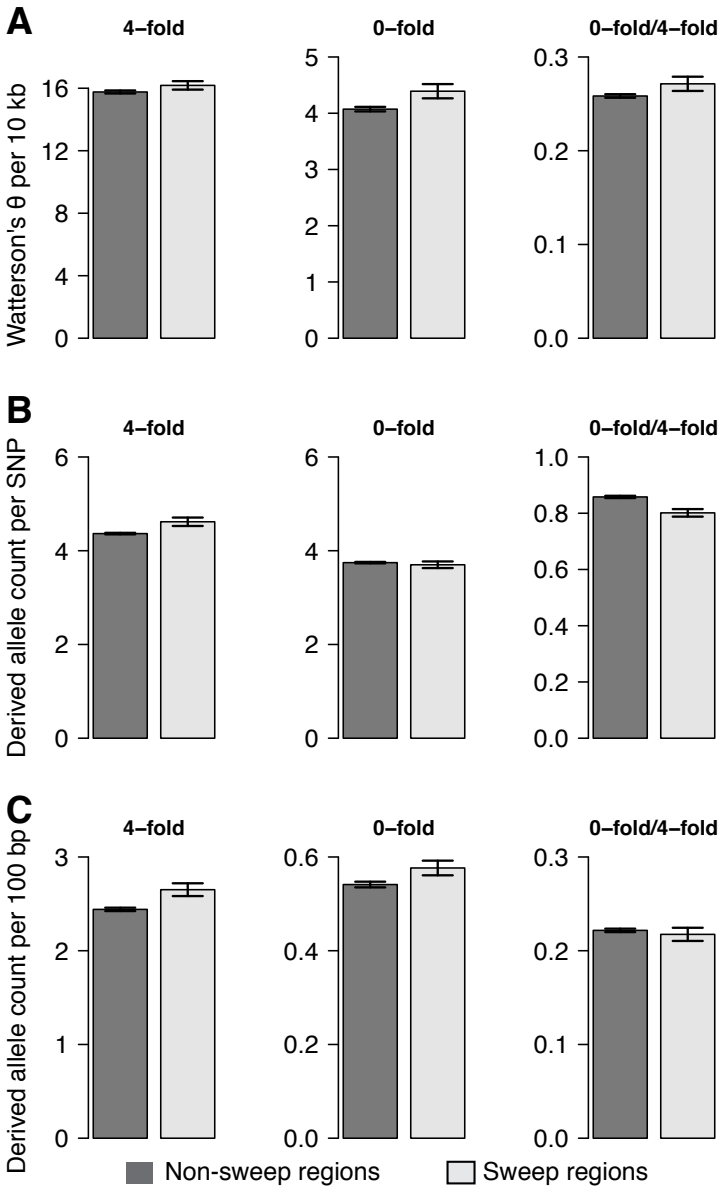

Fig. S12: Genetic variation in wolves surrounding selective sweeps identified in dogs. **(A)** Watterson's θ, an estimate of genetic diversity based on the number of variants, regardless of their frequency. **(B)** The average derived allele count per SNP. **(C)** Average number of derived alleles per 100 bp (considering invariant positions). Each variant site is counted the number of times its derived allele appears in the sample. Error bars are 95% confidence intervals obtained by jackknifing over chromosomes. 4-fold denotes those changes in the coding sequence that do not change the amino acid and are putatively neutral. Note that both 4-fold (putatively neutral) and 0-fold variants (possibility deleterious) are elevated near the breed dog sweeps compared to the rest of the genome. This finding may be explained by the sweep regions having a higher mutation rate or by having a reduction in selective constraint. In both scenarios, more genetic variation may be available on which positive natural and artificial selection could act. Further, under this scenario, there would be greater neutral and deleterious variation expected near sweeps.

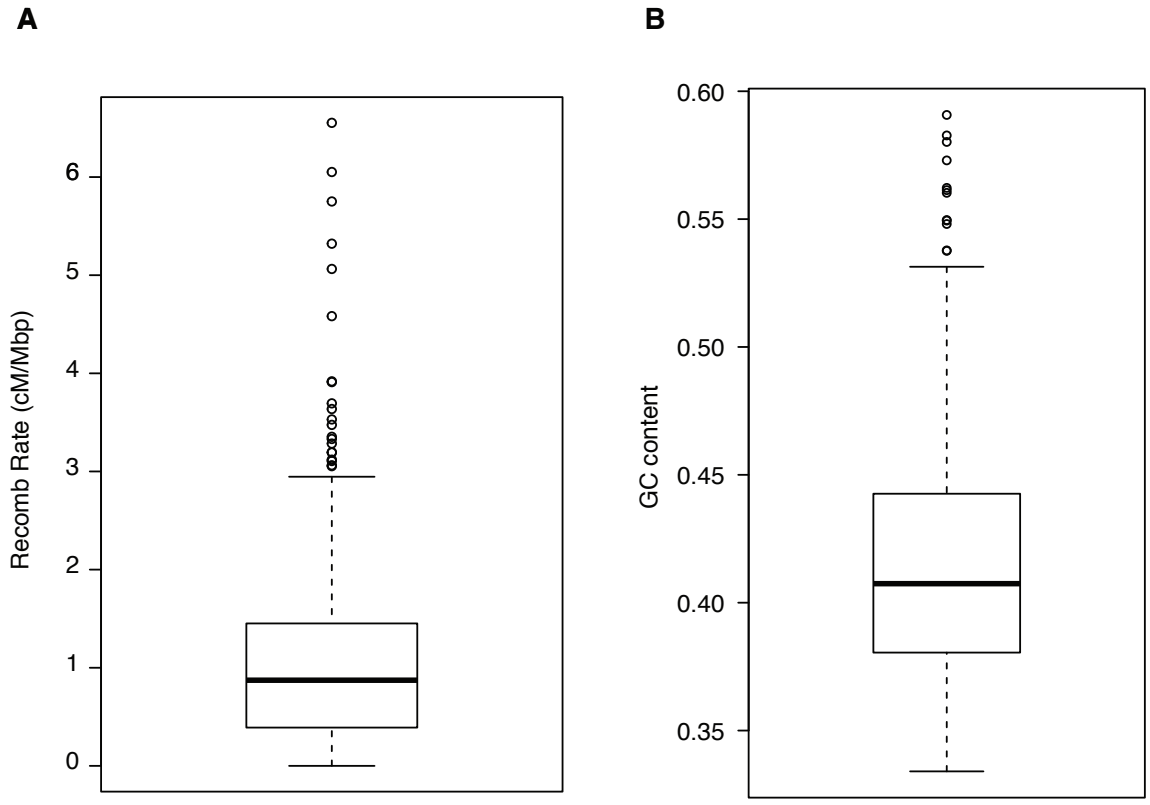

Fig. S13: Properties of 421 selective sweep regions.
(A) Distribution of recombination rates. Recombination rates were taken from those estimated in (54). The average recombination rate in the selective sweep regions is 1.09 cM/Mb, similar to the genome wide-average of 0.97 cM/Mb reported in ref. (41). **(B)** Distribution of GC content. Estimates of GC content were taken from the "gc5BaseBw" table on the CanFam 3.1 build on the UCSC Genome Browser. The average GC content in the selective sweep regions is 41.7%, similar to the genome wide-average of 41% reported in (55), and seen for random regions of chromosome 14.

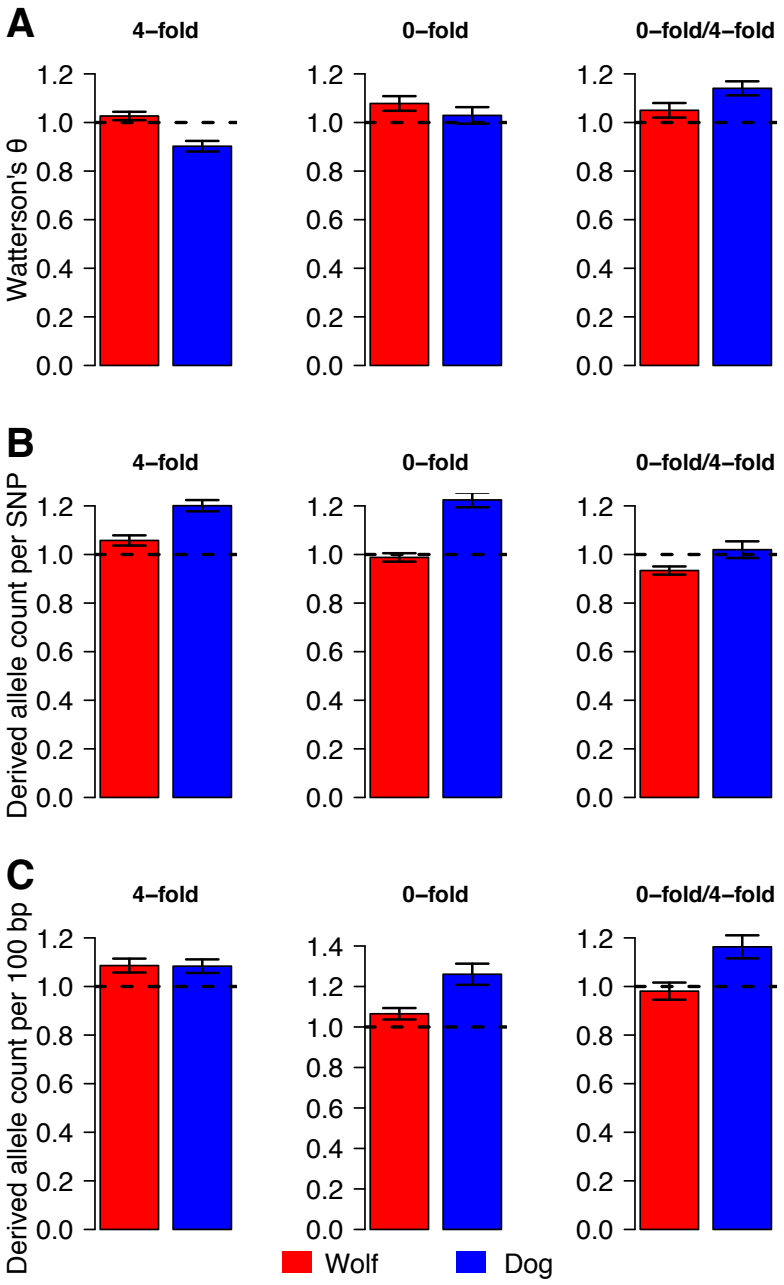

Fig. S14: Ratio of genetic diversity near sweeps vs. the rest of the genome in breed dogs and wolves.

(A) Watterson's θ, an estimate of genetic diversity based on the number of variants, regardless of their frequency. **(B)** The average derived allele count per SNP. **(C)** Average number of derived alleles per 100 bp (considering invariant positions). Each variant site is counted the number of times its derived allele appears in the sample. Error bars are 95% confidence intervals obtained by jackknifing over chromosomes. Note that the ratio is significantly > 1 for 4-fold diversity in wolves indicating that the sweeps have more putatively neutral variation than the rest of the genome. In breed dogs, however, diversity (**A**) is reduced near sweeps. Wolves show increased deleterious diversity (**A**) near the sweeps while dogs show no change or a reduction. However, the increase in the total number of derived 0-fold alleles near the sweeps is significantly higher in

the breed dogs than the wolves. Finally, the enrichment of 0-fold to 4-fold variants is significantly higher in the dogs than the wolves for all summaries of genetic variation. This finding suggests that a higher mutation rate near the sweeps or reduction in selective constraint cannot by itself explain the enrichment of putatively deleterious variations near the sweeps in breed dogs compared to wolves. Instead, this pattern can be explained by hitchhiking of deleterious variations with the sweep.

Table S1: Comparison of the regression parameter estimates across different data sets

Dataset	Intercept	SE	Lower CI	Upper CI	Slope	SE	Lower CI	Upper CI
High coverage*	0.2759	0.0035	0.2691	0.2827	-21.4348	3.0244	-15.5069	-27.3627
Foursite on high coverage#	0.2921	0.0048	0.2827	0.3016	-28.5896	3.7965	-21.1484	-36.0308
FourSite low-coverage	0.3010	0.0063	0.2887	0.3134	-29.0043	4.8938	-19.4126	-38.5961

*Denotes the estimates from the 35 high coverage genomes where genotypes were called using GATK.
#Denotes the estimates from the 35 high coverage genomes that were treated analyzed as low-coverage genomes. We sampled four reads per site and estimated heterozygosity using FourSite.

Table S2: Top 10 genes showing the greatest difference in the number of 0-fold and 4-fold variants between dogs and wolves

Ensemble Transcript ID	Gene	# 0-fold SNPs in dog	# 0-fold SNPs wolf	# 4-fold SNPs in dog	# 4-fold SNPs wolf	0-dog)/0-wolf	4-dog)/4-wolf	*P*-value
ENSCAFG00000015228	*vWF*	10	7	0	8	1.43	0.00	0.0077
ENSCAFG00000003887	*APOB*	18	16	3	14	1.13	0.21	0.0188
ENSCAFG00000015314	*UBR4*	4	3	1	11	1.33	0.09	0.0379
n/a		11	4	0	3	2.75	0.00	0.0429
ENSCAFG00000009561	*DNMBP*	0	6	5	4	0.00	1.25	0.0440
ENSCAFG00000013255	*MKI67*	9	24	8	5	0.38	1.60	0.0441
ENSCAFG00000023780	*MUC5B*	19	18	8	23	1.06	0.35	0.0466
ENSCAFG00000008686	*CSMD1*	7	2	7	13	3.50	0.54	0.0502
ENSCAFG00000013453	Uncharacterized	0	9	3	4	0.00	0.75	0.0625
ENSCAFG00000017941	*CYP1A2*	11	6	1	5	1.83	0.20	0.0686

"0-dog/0-wolf" denotes the ratio of the number of 0-fold variants in dogs to the number of 0-fold variants in wolves. "4-dog/4-wolf" denotes the ratio of the number of 4-fold variants in dogs to the number of 4-fold variants in wolves. "*P*-value" denotes the *P*-value from Fisher's exact test for homogeneous 0-fold to 4-fold ratios between dogs and wolves.

Table S3: Forward simulation parameters based on the Freedman et al. demographic model

	Wolves		Village Dogs		Breed Dogs	
	Number of chromosomes ($2N_e$)	**Number of generations**	**Number of chromosomes ($2N_e$)**	**Number of generations**	**Number of chromosomes ($2N_e$)**	**Number of generations**
Epoch 1	18,169	145,352	18,169	145,352	18,169	145,352
Epoch 2	44,993	63,898	44,993	63,898	44,993	63,898
Epoch 3	24,000	237	1999	347	1999	347
Epoch 4	30,000	2243	15,000	2133	8000	2033
Epoch 5	-	-	-	-	1000	100

Epoch 1 denotes the ancestral population size. Epoch 4 denotes the current effective population size for wolves and village dogs while Epoch 5 represents the current effective population size for breed dogs. This demographic model was used for the regression line presented in Fig. 1B.

Table S4: Forward simulation parameters based on the Freedman et al. model with larger ancestral population sizes

	Wolves		Village Dogs		Breed Dogs	
	Number of chromosomes ($2N_e$)	**Number of generations**	**Number of chromosomes ($2N_e$)**	**Number of generations**	**Number of chromosomes ($2N_e$)**	**Number of generations**
Epoch 1	18,169	145,352	18,169	145,352	18,169	145,352
Epoch 2	60,000	63,898	60,000	63,898	60,000	63,898
Epoch 3	2400	237	1999	347	1999	347
Epoch 4	30,000	2243	15,000	2133	8000	2033
Epoch 5	-	-	-	-	1000	100

Epoch 1 denotes the ancestral population size. Epoch 4 denotes the current effective population size for wolves and village dogs while Epoch 5 represents the current effective population size for breed dogs.

Table S5: Forward simulation parameters based on the Wang et al. model

	Wolves		Village Dogs		Breed Dogs	
	Number of chromosomes ($2N_e$)	**Number of generations**	**Number of chromosomes ($2N_e$)**	**Number of generations**	**Number of chromosomes ($2N_e$)**	**Number of generations**
Epoch 1	35,000	280,000	35,000	280,000	35,000	280,000
Epoch 2	33,020	3556	5666	2556	5666	2,556
Epoch 3	-	-	11,332	1000	11,332	900
Epoch 4	-	-	-	-	200	100

Epoch 1 denotes the ancestral population size. Epoch 2 denotes the current effective population size for wolves, Epoch 3 is the current effective population size for village dogs and Epoch 4 represents the current effective population size for breed dogs.

Table S6: Parameters for the gamma distributions of selective effects on new mutations used in forward simulations of demography and selection

Model	α	β	*N*	% mutations *s*<0.0001	% mutations 0.0001<*s* < 0.001	% mutations 0.001<*s*<0.01	% mutations *s*>0.01
Boyko (34)	0.184	319.8626	1000	27.89	14.68	21.90	35.54
Mice (35)	0.11	8,636,364	10^6	32.6	9.40	12.10	45.86
Gamma Test 1	0.25	250	10000	33.00	24.86	32.91	9.23
Gamma Test 2	0.3	100	10000	34.30	31.44	32.06	2.21

α denotes the shape parameter of the distribution of selective effects while β denotes the scale parameter. *N* refers to the population size used to scale the β parameter. Remaining columns provide the proportions of new mutations having different selection coefficients. The regression line from the Gamma Test 2 distribution is shown in Fig. 1B.

Table S7: *P*-values for Mann-Whitney *U*-tests comparing numbers of genotypes, derived alleles and corrected derived alleles (see 'Accumulation of deleterious derived alleles') per individual between dogs and wolves

	Observed number of heterozygous genotypes	Observed number of homozygous genotypes	Observed number of derived alleles	Corrected number of derived alleles
Omitting Tibetan wolf				
Synonymous	7.245×10^{-7}	3.813×10^{-8}	0.0023	0.3354
Nonsynonymous	1.716 x 10-^{6}	3.813×10^{-8}	4.75×10^{-4}	0.0092
Miyata damaging	2.555×10^{-6}	3.813×10^{-8}	3.699×10^{-6}	9.851×10^{-4}
GERP deleterious	1.838×10^{-5}	3.813×10^{-8}	2.129×10^{-4}	0.0019
GERP deleterious, omitting sweeps	1.838×10^{-5}	3.813×10^{-8}	2.129×10^{-4}	0.0023
4-fold	1.716×10^{-6}	3.813×10^{-8}	0.0026	0.0451
0-fold	1.716×10^{-6}	3.813×10^{-8}	0.0030	0.0451
Including Tibetan wolf				
Synonymous	3.359×10^{-5}	9.478×10^{-6}	0.009457	0.4604
Nonsynonymous	1.609×10^{-5}	3.987×10^{-6}	0.006567	0.0496
Miyata damaging	1.239×10^{-5}	2.114×10^{-6}	1.239×10^{-5}	0.001944
GERP deleterious	5.310×10^{-5}	5.382×10^{-6}	0.0045	0.0186
4-fold	6.622×10^{-5}	2.931×10^{-6}	0.001944	0.041
0-fold	2.647×10^{-5}	3.987×10^{-6}	0.006567	0.070

Note, *P*-values for differences in the number of heterozygous genotypes are the same both before and after correction for false-negative genotype calls.

Table S8: Overlap between dog Mendelian disease genes and selective sweeps

Sweep dataset	Number of genes in sweeps	Observed number of disease genes in sweeps	Expected number of disease genes in sweeps	*P*-value for enrichment
Ancient (10, 37)	711	3	5.6	0.921
Recent (50)	1632	18	12.8	0.087
Recent (51)	1663	17	13.0	0.155

This analysis compares the overlap between genes near selective sweeps and 145 genes implicated in dog Mendelian disease. This analysis includes a total of 18,514 genes. The expected number of genes in sweeps was computed using a hypergeometric distribution (See Methods).

Table S9: Overlap between human Mendelian disease genes and dog selective sweeps

Sweep dataset	Number of genes in sweeps	Observed number of disease genes in sweeps	Expected number of disease genes in sweeps	*P*-value for enrichment
Ancient (10, 37)	711	92	97.4	0.740
Recent (50)	1632	245	223.5	0.057
Recent (51)	1663	263	227.7	0.005

This analysis compares the overlap between genes near selective sweeps and 2,535 genes implicated in human Mendelian disease. This analysis includes a total of 18,514 genes. The expected number of genes in sweeps was computed using a hypergeometric distribution (See Methods).