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)

Population bottlenecks, inbreeding, and artificial selection can all, in principle, influence levels of deleterious genetic variation. However, the relative importance of each of these effects on genome-wide patterns of deleterious variation remains controversial. Domestic and wild canids offer a powerful system to address the role of these factors in influencing deleterious variation because their history is dominated by known bottlenecks and intense artificial selection. Here, we assess genome-wide patterns of deleterious variation in 90 whole-genome sequences from breed dogs, village dogs, and gray wolves. We find that the ratio of amino acid changing heterozygosity to silent heterozygosity is higher in dogs than in wolves and, on average, dogs have 2–3% higher genetic load than gray wolves. Multiple lines of evidence indicate this pattern is driven by less efficient natural selection due to bottlenecks associated with domestication and breed formation, rather than recent inbreeding. Further, we find regions of the genome implicated in selective sweeps are enriched for amino acid changing variants and Mendelian disease genes. To our knowledge, these results provide the first quantitative estimates of the increased burden of deleterious variants directly associated with domestication and have important implications for selective breeding programs and the conservation of rare and endangered species. Specifically, they highlight the costs associated with selective breeding and question the practice favoring the breeding of individuals that best fit breed standards. Our results also suggest that maintaining a large population size, rather than just avoiding inbreeding, is a critical factor for preventing the accumulation of deleterious variants.

deleterious mutations | domestication | bottleneck | selective sweep

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.

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.

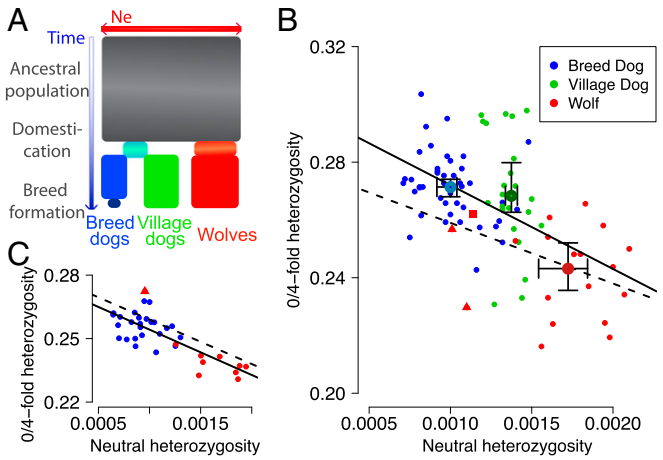

Fig. 1. Population history and deleterious genetic variation. (*A*) Conceptual model of dog domestication used in population genetic simulations. Box widths are proportional to estimated population sizes (*SI Appendix*, Table S4). (*B*) The ratio of zerofold to fourfold heterozygosity vs. neutral genetic 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 (lintercept = 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.

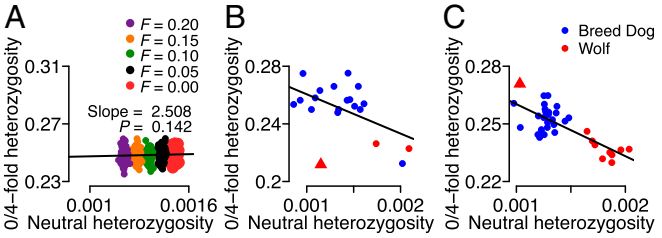

Fig. 2. Recent inbreeding does not drive the relationship between neutral heterozygosity and the zerofold/fourfold heterozygosity ratio. (*A*) Forward simulations using a demographic model that includes inbreeding over the last 100 generations, but not bottlenecks associated with domestication or breed 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).

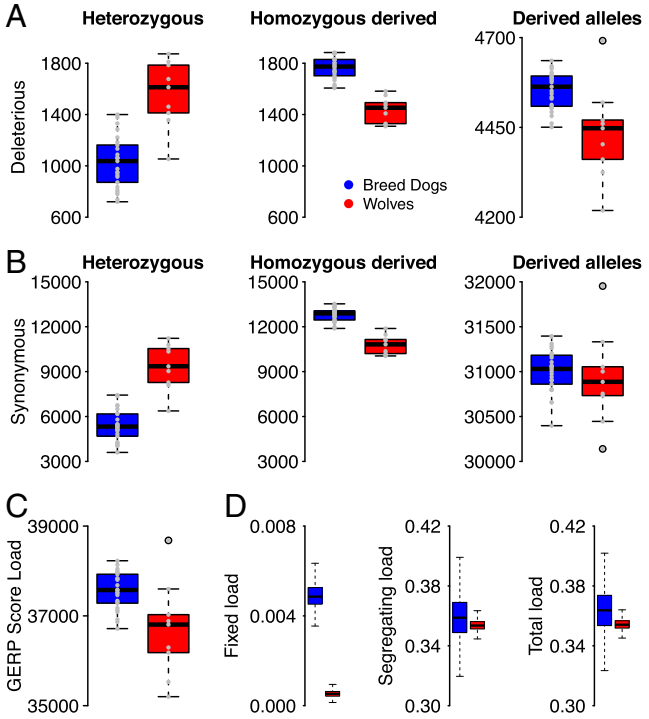

Fig. 3. Comparison of the burden of deleterious genetic variation between breed dogs (blue) and wolves (red) based on high-quality genomes. "Homozygous derived" refers to the number of genotypes per individual that are homozygous for the derived allele. The total number of derived alleles is based on 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

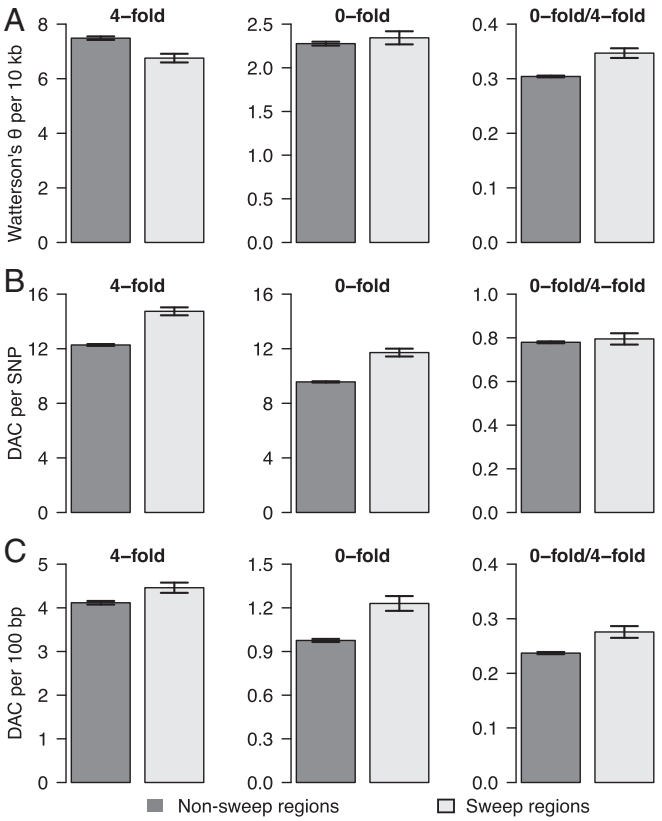

Fig. 4. Genetic variation surrounding nonsweep (dark gray) and sweep (light gray) regions in breed dogs. (*A*) Watterson's θ, an estimate of genetic diversity based on the number of SNPs. (*B*) The average derived allele count (DAC) per SNP. (*C*) Average DAC per 100 bp (considering invariant positions). 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.