RESEARCH ARTICLE SUMMARY

ZOONOMIA

Comparative genomics of Balto, a famous historic dog, captures lost diversity of 1920s sled dogs

Katherine L. Moon*†, Heather J. Huson*†, Kathleen Morrill*†, Ming-Shan Wang, Xue Li, Krishnamoorthy Srikanth, Zoonomia Consortium, Kerstin Lindblad-Toh, Gavin J. Svenson, Elinor K. Karlsson*‡, Beth Shapiro*‡

INTRODUCTION: It has been almost 100 years since the sled dog Balto helped save the community of Nome, Alaska, from a diphtheria outbreak. Today, Balto symbolizes the indomitable spirit of the sled dog. He is immortalized in statue and film, and is physically preserved and on display at the Cleveland Museum of Natural History. Balto represents a dog population that was reputed to tolerate harsh conditions at a time when northern communities were reliant on sled dogs. Investigating Balto's genome sequence using technologies for sequencing degraded DNA offers a new perspective on this historic population.

RATIONALE: Analyzing high-coverage (40.4-fold) DNA sequencing data from Balto through comparison with large genomic data resources offers an opportunity to investigate genetic diversity and genome function. We leveraged the genome sequence data from 682 dogs, including both working sled dogs and dog breeds, as well as evolutionary constraint scores from the Zoonomia alignment of 240 mammals, to reconstruct Balto's phenotype and investigate his ancestry and what might distinguish him from modern dogs.

RESULTS: Balto shares just part of his diverse ancestry with the eponymous Siberian husky breed and was more genetically diverse than both modern breeds and working sled dogs. Both Balto and working sled dogs had a lower burden of rare, potentially damaging variation than modern breeds and fewer potentially damaging variants, suggesting that they represent genetically healthier populations. We inferred Balto's appearance on the basis of genomic variants known to shape physical characteristics in dogs today. We found that Balto had a combination of coat features atypical for modern sled dog breeds and a slightly smaller stature, inferences that are confirmed by comparison to historical photographs. Ba… ability to digest starch was enhanced compa… to wolves and Greenland sled dogs but reduced compared to modern breeds. He carried a compendium of derived homozygous coding variants at constrained positions in genes connected to bone and skin development, which may have conferred a functional advantage.

CONCLUSION: Balto belonged to a population of small, fast, and fit sled dogs imported from Siberia. By sequencing his genome from his taxidermied remains and analyzing these data in the context of large comparative and canine datasets, we show that Balto and his working sled dog contemporaries were more genetically diverse than modern breeds and may have carried variants that helped them survive the harsh conditions of 1920s Alaska. Although the era of Balto and his contemporaries has passed, comparative genomics, supported by a growing collection of modern and past genomes, can provide insights into the selective pressures that shaped them. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: katielouisemoon@gmail.com (K.L.M.); hjh3@cornell.edu (H.J.H.); kathleen.morrill@umassmed.edu (K.M.); beth.shapiro@gmail.com (B.S.); elinor.karlsson@umassmed.edu (E.K.K.)
†These authors contributed equally to this work.
‡These authors contributed equally to this work.
Cite this article as K. L. Moon *et al.*, *Science* **380**, eabn5887 (2023). DOI: 10.1126/science.abn5887

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abn5887

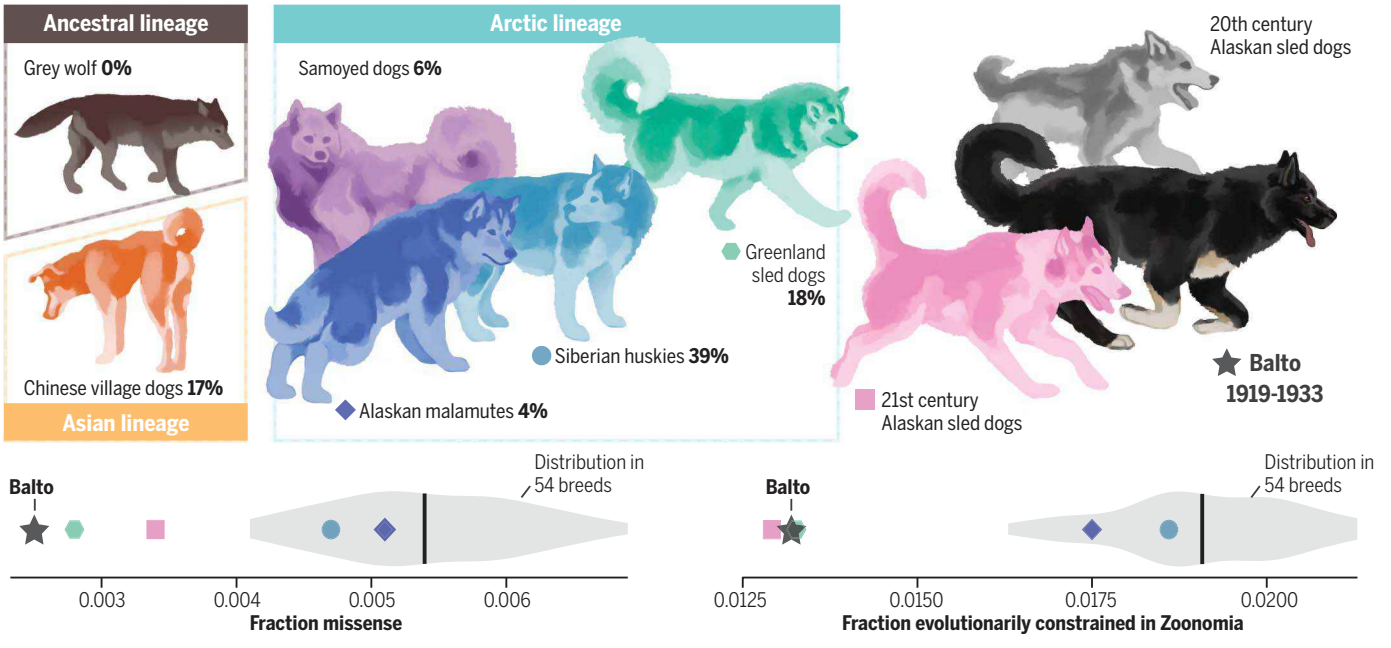

IMAGE CREDIT: K. MORRILL

Balto, famed 20th-century Alaskan sled dog, shares common ancestry with modern Asian and Arctic canine lineages. In an unsupervised admixture analysis, Balto's ancestry, representing 20th-century Alaskan sled dogs, is assigned predominantly to four Arctic lineage dog populations. He had no discernable wolf ancestry. The Alaskan sled dogs (a working population) did not fall into a distinct ancestry cluster but shared about a third of their ancestry with Balto in the supervised admixture analysis. Balto and working sled dogs carried fewer constrained and missense rare variants than modern dog breeds.

RESEARCH ARTICLE

ZOONOMIA

Comparative genomics of Balto, a famous historic dog, captures lost diversity of 1920s sled dogs

Katherine L. Moon[1,2]*†, Heather J. Huson[3]*†, Kathleen Morrill[4,5,6]*†, Ming-Shan Wang[1,2], Xue Li[4,5,6], Krishnamoorthy Srikanth[3], Zoonomia Consortium‡, Kerstin Lindblad-Toh[6,7], Gavin J. Svenson[8], Elinor K. Karlsson[4,5]*§, Beth Shapiro[1,2]*§

We reconstruct the phenotype of Balto, the heroic sled dog renowned for transporting diphtheria antitoxin to Nome, Alaska, in 1925, using evolutionary constraint estimates from the Zoonomia alignment of 240 mammals and 682 genomes from dogs and wolves of the 21st century. Balto shares just part of his diverse ancestry with the eponymous Siberian husky breed. Balto's genotype predicts a combination of coat features atypical for modern sled dog breeds, and a slightly smaller stature. He had enhanced starch digestion compared with Greenland sled dogs and a compendium of derived homozygous coding variants at constrained positions in genes connected to bone and skin development. We propose that Balto's population of origin, which was less inbred and genetically healthier than that of modern breeds, was adapted to the extreme environment of 1920s Alaska.

Technological advances in the recovery of ancient DNA make it possible to generate high-coverage nuclear genomes from historic and fossil specimens, but interpreting genetic data from past individuals is difficult without data from their contemporaries. Comparative genomic analysis offers a solution: By combining population-level genomic data and catalogs of trait associations in modern populations, we can infer the genetic and phenotypic features of long-dead individuals and the populations from which they were born. Zoonomia is a new comparative resource that addresses limitations of previous datasets (*1*) to support interpretation of paleogenomics data. With 240 placental mammal species, Zoonomia has sufficient power to distinguish individual bases under evolutionary constraint—a useful predictor of functional importance (*2*)—in coding and regulatory elements (*3*). Zoonomia's reference-free genome alignment (*4*, *5*) allows evolutionary constraint to be scored in any of its 240 species, including dogs.

Here, we generate a genome for Balto, the famous sled dog who delivered diphtheria serum to the children of Nome, Alaska, during a 1925 outbreak. Following his death, Balto was taxidermied, and his remains are held by the Cleveland Museum of Natural History. We generated a 40.4-fold coverage nuclear genome from Balto's underbelly skin using protocols for degraded samples. His DNA was well preserved, with an average endogenous content of 87.7% in sequencing libraries, low (<1%) damage rates (fig. S1), and short [68 base pairs (bp)] average fragment sizes, consistent with the age of the sample.

Balto was born in the kennel of sled dog breeder Leonard Seppala in 1919. Although Seppala's small fast dogs were known as Siberian huskies (*6*), they were a working population that differed from the dog breed recognized by the American Kennel Club (AKC) today. Modern dog breeds are genetically closed populations that conform to a tightly delineated physical standard (*7*). Balto's relationship to AKC-recognized sled dog breeds such as the Siberian husky (established in 1930) and Alaskan malamute (1935) (*8*) is unclear. Balto himself was neutered at 6 months of age and had no offspring.

Working populations of sled dogs survive. Alaskan sled dogs are bred solely for physical performance, including outcrossing with various breeds (*9*). Greenland sled dogs are an indigenous land-race breed that have been used for hunting and sledging by Inuit in Greenland for 850 years, where they have been isolated from contact with other dogs (*10*). Here, we use the term "breed" exclusively to refer to modern breeds recognized by the AKC or other kennel clubs (e.g., sled dog breeds), as distinct from the less rigidly defined populations of Greenland sled dogs and Alaskan sled dogs (working sled dogs). This is a genetic distinction; AKC-registered dogs can be successful working sled dogs.

We compared Balto to working sled dogs, sled dog breeds, other breeds, village dogs (free-breeding dogs without known breed ancestry), and other canids. Our whole-genome dataset comprised 688 dogs (table S1) representing 135 breeds or populations, including three Alaskan sled dogs and five Greenland sled dogs (*10*). We identified evolutionarily constrained bases using phyloP evolutionary constraint scores from the dog-referenced version of the 240-species Zoonomia alignment (*3*).

Ancestry analysis places Balto in a clade of sled dog breeds and working sled dogs and closest to the Alaskan sled dogs (Fig. 1, A and B). Most of his ancestry is assigned to clades of Arctic-origin dogs (68%) and, to a lesser extent, Asian-origin dogs (24%) in an unsupervised admixture analysis with 2166 dogs and 116 clusters (Fig. 1C and tables S2 and S3). He carried no discernible wolf ancestry. The more recently established Alaskan sled dog population (*9*) did not fall into a distinct ancestry cluster in the unsupervised analysis but comprised 34% of Balto's ancestry in a supervised analysis defining them as a cluster (fig. S2).

Balto was more genetically diverse than breed dogs today and similar to working sled dogs (Fig. 1D). Balto had shorter runs of homozygosity than any breed dog, and fewer runs of homozygosity than all but one Tibetan mastiff (table S4). When inbreeding is calculated from runs of homozygosity, Balto and dogs from the two working sled dog populations have lower inbreeding than almost any breed dog (fig. S3). When inbreeding is calculated using an allele frequency approach (method-of-moment), Greenland sled dogs have high inbreeding coefficients, reflecting their long genetic isolation in Greenland (fig. S3).

To evaluate the genetic health of Balto's population of origin, we developed an analytical approach that leveraged the Zoonomia 240-species constraint scores and required only a single dog from each population (necessary because Balto is the only available representative of his population). Briefly, we selected one individual at random from each breed or population (57 dogs in total) and scored variant positions as either evolutionarily constrained [and more likely to be damaging (*2*)] or not using the Zoonomia phyloP scores (*3*). We also identified variants likely to be "rare" (low frequency) in each dog's breed or population. Because we could not directly measure population allele frequencies with only a single representative dog, we defined "rare" variants as heterozygous or homozygous variants

[1]Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, CA, USA. [2]Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, CA, USA. [3]Department of Animal Sciences, Cornell University College of Agriculture and Life Sciences, Ithaca, NY 14853, USA. [4]Bioinformatics and Integrative Biology, UMass Chan Medical School, Worcester, MA 01655, USA. [5]Morningside Graduate School of Biomedical Sciences, UMass Chan Medical School, Worcester, MA 01655, USA. [6]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [7]Department of Medical Biochemistry and Microbiology, Science for Life Laboratory, Uppsala University; 751 32 Uppsala, Sweden. [8]Cleveland Museum of Natural History, Cleveland, OH 44106, USA.
***Corresponding author. Email: katielouisemoon@gmail.com (K.L.M.); hjh3@cornell.edu (H.J.H.); kathleen.morrill@umassmed.edu (K.M.); beth.shapiro@gmail.com (B.S.); elinor.karlsson@umassmed.edu (E.K.K.)**
†These authors contributed equally to this work.
‡Zoonomia Consortium collaborators and affiliations are listed at the end of this paper.
§These authors contributed equally to this work.

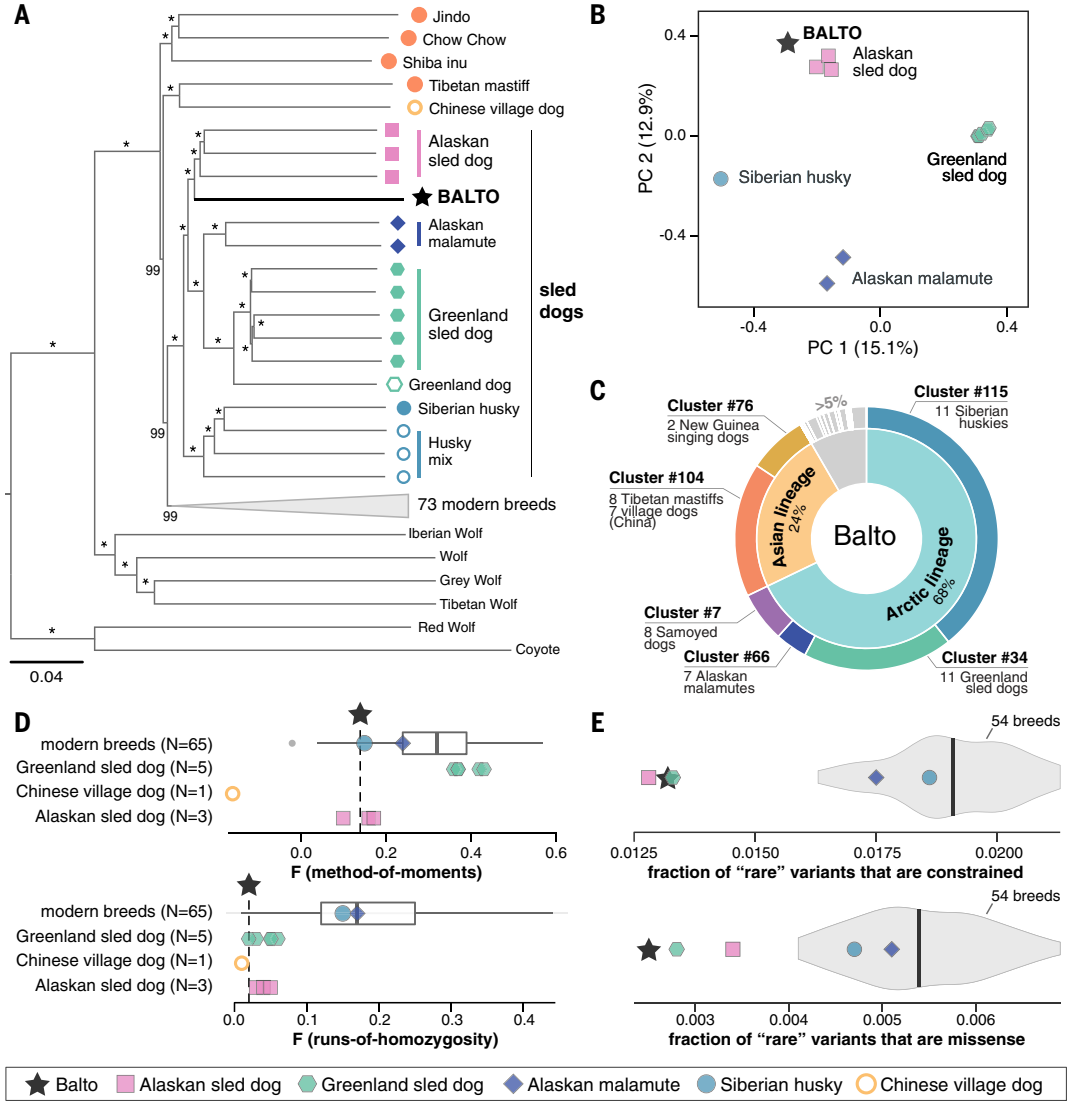

Fig. 1. Balto clusters most closely with Alaskan sled dogs, but had high genetic diversity and a lower burden of potentially damaging variants. (**A**) Neighbor-joining tree clusters Balto (★) most closely with the outbred, working population of Alaskan sled dogs, and a part of a clade of sled dog populations. (**B**) Similarly, principal component analysis puts Balto near, but not in, a cluster of Alaskan sled dogs. (**C**) Unsupervised admixture analysis of Balto alongside the Alaskan sled dogs and other dogs and canids (K = 116 putative populations and N = 2166 individuals) infers substantial ancestral similarity to Siberian huskies, Greenland sled dogs, and outbred dogs from Asia (table S2). The remainder of his ancestry (8%) matches poorly (<5%) to any other clusters. (**D** and **E**) Balto and working sled dogs (D) had lower levels of inbreeding and (E) carried fewer constrained (p_{wilcox} = 0.0019) and missense (p_{wilcox} = 0.0023) rare variants than modern dog breeds (table S10).

specific to that dog among all 57 representative dogs. This metric effectively identifies variants occurring at unusually low frequencies (fig. S4).

Balto and modern working sled dogs had a lower burden of rare, potentially damaging variation, indicating that they represent genetically healthier populations (*11*) than breed dogs. Balto and the working sled dogs had significantly fewer potentially damaging variants (missense or constrained) than any breed dog, including the sled dog breeds (Fig. 1E). The pattern persists even in the less genetically diverse Greenland sled dog. Selection for fitness in working sled dog populations appears more effective in removing damaging genetic variation than selection to meet a breed standard.

Balto's physical appearance predicted from his genome sequence (Fig. 2A and table S5) matches historical photos (Fig. 2B) and his taxidermied remains, indicating that the same variants that shaped modern breed phenotypes also explained natural variation in his pre-breed working population. We predict that he stood 55 cm tall at his shoulders (*12*) (Fig. 2C), within the acceptable range for today's Siberian husky breed [53 to 60 cm (*8*)], and had a double-layered coat (*13*) that was mostly black with only a small amount of white (*14*). He was homozygous for an allele conferring tan points (*15*) and one for blue eyes (*16*), but both were masked by his melanistic facial mask (*17*), and his predicted light-tan pigmentation (*18*) may have been indistinguishable from white. He carried neither the "wolf agouti" nor "Northern domino" patterns that are common in the Siberian husky and other sled dog breeds today (*19*).

Both Balto and Alaskan sled dogs had unexpected evidence of adaptation to starch-rich diets. They carry the dog version of *MGAM*, a gene involved in starch processing that is differentiated between dogs and wolves (*20*) and 1 of 14 regions analyzed for evidence of selective

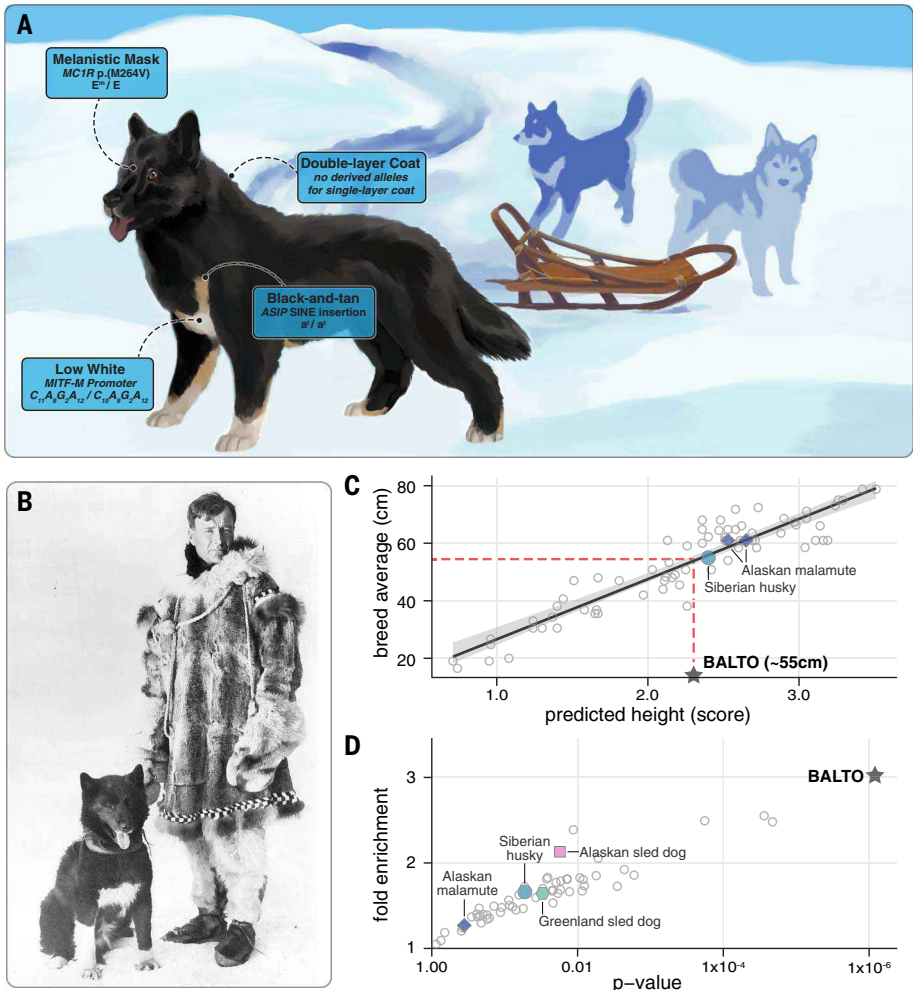

Fig. 2. Genomic recreation of Balto's physical appearance. (**A**) Prediction of Balto's coat features based on his genome sequence with details on each trait and genotype in blue boxes. (**B**) A photo of Balto with musher Gunnar Kaasen. From the photo and his taxidermied remains, Balto was a black dog with dark eyes and some white patches on his chest and legs. He had a double-layered coat and stood just under knee-high relative to Kaasen. [Photo credit: Cleveland Museum of Natural History] (**C**) Using a random forest model based on 1730 dogs and 2797 height-associated genetic variants (*12*), we predicted that Balto would stand around 55 cm tall (value: 2.3) at his withers, close to the average height for the Siberian husky breed. Circles show dogs from other breeds. (**D**) Gene set enrichment testing of genes with common and constrained missense variants in 57 different dog populations shows a significant enrichment (p_{FDR} = 0.013) in the GO Tissue Development pathway only for Balto's population.

pressure in Balto's lineage using a gene tree analysis (table S6). In earlier work, the high frequency of the wolf version of *MGAM* in Greenland sled dogs prompted speculation that reduced starch digestion might be a working sled dog trait (*10*). Our findings suggest that this phenomenon is specific to Greenland sled dogs. Gene tree analysis places one of Balto's chromosomes in the ancestral wolf cluster and one in the derived dog cluster (fig. S5). Most Alaskan sled dogs carry the dog version (frequency = 0.83). However, read coverage of the gene *AMY2B* suggests that Balto had fewer copies of this gene than many modern dogs and thus comparatively lower production of the starch-digesting enzyme amylase (*21*, *22*). Taken together, we suggest that Balto's ability to digest starch was enhanced compared to wolves and Greenland sled dogs but reduced compared to modern breeds.

Of the other 14 regions tested, most (10 out of 14) lacked sufficient diversity in dogs to resolve phylogenetic relationships. Bootstrap support was weak for two other genes selected in Greenland sled dogs (*CACNA1A* and *MAGI2*). As expected, Balto did not carry versions of *EPAS1* associated with high-altitude adaptation (*23*).

We found an enrichment for unusual function variation in Balto's population consistent with adaptation to the extreme environments in which early–20th century sled dogs worked. We identified variants in Balto's genome that were new (not seen in wolves) and likely to be common in his population (homozygous in Balto; fig. S4). We further filtered for variants that were both protein-altering (missense) and evolutionarily constrained [false discovery rate (FDR) <0.01], and thus likely to be functional. Balto was no more likely to carry such variants than dogs from 54 other populations (fig. S6), but in Balto these variants tended to disrupt tissue development genes [Gene Ontology (GO): 0009888; 24 genes; 3.02-fold enrichment; p_{FDR} = 0.013] (table S7). This enrichment was specific to Balto (Fig. 2D and fig. S7), and most of the variants were rare or missing in other dog populations (fig. S8). Even when all GO biological process gene sets are tested in all 57 dogs, Balto's enrichment in tissue development genes is highly unusual. It ranks fourth out of 888,573 dog/gene set pairs tested (fig. S7 and table S8). Phenotype associations from human disease studies suggest that these variants could have influenced skeletal and epithelial development including joint formation, body weight, coordination, and skin thickness (table S9) (*24*). Modern sled dog breeds and working sled dogs are only slightly more similar to Balto than other dogs at these variants (fig. S9).

Balto was part of a famed population of small, fast, and fit sled dogs imported from Siberia. After his famous run, the Siberian husky breed was recognized by the AKC. By sequencing his genome from his taxidermied remains and analyzing it in the context of large comparative and canine datasets, we show that Balto shared only part of his ancestry with today's Siberian huskies. Balto's working sled dog contemporaries were healthier and more genetically diverse than modern breeds and may have carried variants that helped them survive the harsh conditions of 1920s Alaska (*6*). Further work is still needed to assess the impact of the evolutionarily constrained missense variants that Balto carried. Although the era of Balto and his fellow huskies has passed, comparative genomics, supported by a growing collection of modern and past genomes, can provide a snapshot of individuals and populations from the past, as well as insights into the selective pressures that shaped them.

Materials and methods

Assembly of comparative canid genetic variants

We collated a reference set of comparative canid genetic variants starting from the curated Broad-UMass Canid Variant set (https://data.broadinstitute.org/DogData/) and comprising whole-genome sequencing data for 531 dogs of known breed ancestry distributed among 132 breeds, 28 dogs of mixed breed ancestry, 12 dogs of unknown ancestry, 69 worldwide indigenous or village dogs, 33 wolves, and 1 coyote (table S1).

Ancient DNA extraction, library preparation, and genome assembly

We extracted DNA from a ~5 mm by 5 mm piece of Balto's underbelly skin tissue, in two replicates (HM246 and HM247) with an extraction negative, using the ancient DNA–specific protocol in Dabney *et al.* 2013 (*25*). We prepared 32 ~1-pmol input Illumina libraries from these extracts following the Santa Cruz library preparation method (*26*), including positive and negative controls. All 32 libraries passed quality control (QC), and so we sequenced them to a depth of ~2.3 billion on a NovaSeq 6000 platform 150 bp paired end (see table S11 for the number of reads produced per library).

We used SeqPrep v.1.1 (*27*) to trim adapters, remove reads shorter than 28 bp, and merge remaining paired-end reads with a minimum overlap of 15 bp. We then used the Burrows-Wheeler Aligner (BWA) v.0.7.12 (*28*) with a minimum quality cut off of 20 to align reads to the *Canis lupus familiaris* (dog) reference genome (CanFam3.1) (NCBI: GCA_000002285.2). All 32 bam files (one for each library) were merged into one with PCR duplicates removed. We used both Qualimap (v2.2.1) and samtools (v1.7) to calculate metrics and assess the quality of the alignment (table S12).

Variant calling

We used GATK HaplotypeCaller to call variants in Balto as well as 10 previously published Greenland sled dogs (*10*) and 3 Alaskan sled dogs sequenced for this study (see materials and methods for details on sampling, DNA extraction, and sequencing) against the UMass-Broad Canid Variant set using parameter *–genotyping-mode* GENOTYPE_ GIVEN_ALLELES *–alleles* (known alleles). Then, we merged variant call records from these 14 dogs with records from the UMass-Broad Candid Variants set, for variant calls in a full set of 688 individuals: Balto (this study), 3 modern Alaskan sled dogs (this study), 10 modern Greenland sled dogs (*10*), 531 dogs from modern breeds, 40 dogs of unknown or admixed ancestry, 69 village or indigenous dogs, 33 wolves, and 1 coyote.

Phylogenetic analysis and neighbor-joining trees

Using a dataset of 100 representative canids (table S1 for samples selected in the "Phylogenetic Analysis") we confirmed Balto's phylogenetic position by generating a neighbor-joining (NJ) phylogenetic tree and conducting a principal component analysis (PCA). We converted the variant calls into a FASTA file and used MEGA-CC (*29*) with 1000 bootstraps to assess tree topology. We also ran a PCA on this set using *PLINK* (v1.9) and then visualized the first two principal components in R (v. 3.6.3) using the "ggplot2" package.

Global ancestry inference

We inferred Balto's ancestral similarity to that of modern dog breeds, sled dog type breeds, and working sled dogs using a custom built reference panel of modern dogs and canids of the 21st century (table S3). In *PLINK* (v2.00a3LM) (*30*), we identified 4,267,732 biallelic single nucleotide polymorphisms with <10% missing genotypes, and calculated Wright's F-statistics using Hudson method (*31*, *32*) for (i) each dog breed and sled dog population versus all other dogs; (ii) all village dogs versus all other dogs; (iii) each regional village dog population; (iv) all wolves versus all other dogs; (v) all coyotes versus all other canids; and (vi) North American wolves versus Eurasian wolves. We selected 1,858,634 single-nucleotide polymorphisms (SNPs) with F_{ST} > 0.5 across all comparisons, and performed LD-based pruning in 250-kb windows for r^2 > 0.2 to extract 136,779 markers for global ancestry inference. We merged Balto's genotypes for these SNPs with genotypes from the reference samples. For reference samples also represented in the whole-genome dataset, population labels used in the admixture analysis are given in the "Representative in Global Ancestry Inference" column of table S1. We performed global ancestry inference using *ADMIXTURE* (*33*) in both supervised mode (random seed: 43) with 20 bootstrap replicates to estimate parameter standard errors, and in unsupervised mode for the same number of populations (K = 116), which showed low levels of error (0.3) in 10-fold cross-validation analysis of chromosome 1 for K clusters between 50 and 150 (table S13).

Homozygosity and inbreeding metrics

We removed samples with any missing data from the dataset of 100 representative individuals used in the phylogenetic analyses, leaving 86 individuals (see table S1 for samples selected in the "Homozygosity Analysis"). Using this pruned dataset, we detected runs of homozygosity (RoH) using a window-based approach implemented in *PLINK* (v1.9) (*30*). We calculated two measures of inbreeding: the method-of-moments coefficient in *PLINK* (F_{MoM}) and the metric based on runs-of-homozygosity (F_{RoH}), as recommended by Zhao *et al.* 2020 (*34*) (table S4). Using the *R* (v. 3.6.3) function "cor.test," we confirmed that F_{RoH} and F_{MoM} are significantly correlated ($R_{Pearson}$= 0.6752819, p = 9.958e-13, t = 8.3913, df = 84).

Population representative sampling

As Balto is the sole representative of his population, we randomly selected one representative sample from each of 57 populations for the discovery of individually represented, population-relevant genetic variants (see table S1 for samples selected in the "Population Variants Analysis") among 67,085,518 biallelic SNPs. These populations included Balto, 1 Alaskan sled dog, 1 Greenland sled dog, and 54 modern breed dogs, including 1 Siberian husky and 1 Alaskan malamute. Likewise, we selected, where available, another 5 to 11 random samples from 10 modern breeds, and all remaining Greenland sled dog samples, to assess the population-wide allele frequency of these variants (see table S1, "Population Frequency Analysis").

Dog-referenced mammalian evolutionary constraint

We selected biallelic SNPs under evolutionary constraint by examining sites overlapping phyloP evolutionary constraint scores from the dog-referenced version of the 240 species Cactus alignment (*3*). We calculated the constraint score cutoffs at various FDRs.

Unique, rare, and potentially deleterious variants

We first identified all "population-unique" variants, defined as those observed in the representative dog from a population (either once or twice) and not observed in representatives from any of the other populations. With this method, we identified 206,164 population-unique variants for Balto, 120,279 for the Alaskan sled dog, 119,482 variants for the Greenland sled dog, 120,780 unique to the Alaskan malamute, and 133,200 unique to the Siberian husky. We confirmed that population-unique variants tend to be uncommon by calculating the allele frequencies in its population. We used Zoonomia phyloP scores and SnpEff (*35*) annotations to identify which population-unique variants were either "evolutionarily constrained" (phyloP score above the FDR 0.05 cutoff of 2.56) or a missense mutation and thus more likely to have functional consequences (table S15). We grouped the dogs into working dog groups including Balto, Alaskan sled dog, and Greenland sled dog, and modern breeds including all the other 54 dogs. We then applied Student's t test on the percentage of "evolutionarily constrained" or missense mutation for the two groups.

Derived, common, and potentially beneficial variants

We identified "homozygous derived" variants, defined as those observed twice in the representative dog from a population and not observed in wolves, for each of the populations. With this method, we identified 176,135 homozygous derived variants for Balto, 148,036 variants for Alaskan sled dog, 260,457 variants for Greenland sled dog, 225,270 variants for Alaskan Malamute, and 189,188 variants for Siberian husky. We confirmed that homozygous variants in each representative dog tend to be "common" in their population by calculating the allele frequency of the homozygous derived variants in its own breed. We also used a Wilcox test against randomly selected SNPs to show that population-unique SNPs are rare, whereas homozygous derived SNPs are rather common, among their population.

We further defined variants likely to be functional as those that were both "highly

evolutionarily constrained" (defined by phyloP score above the FDR >0.01 cutoff of 3.52) and a missense mutation. We annotated the variant by genes, and performed gene set enrichment against all Gene Ontology Biological Process gene sets (http://geneontology.org/) using the R package rbioapi v. 0.7.4 (*36*, *37*) (tables S7 and S8). We also tested for overlap between Balto's variant genes and genes implicated in particular phenotypes in human studies using the Human Phenotype Ontology (*24*) and the "Investigate gene sets" feature provided by GSEA (http://www.gsea-msigdb.org/) (table S9).

Prediction of Balto's aesthetic phenotypes

We extracted Balto's genotypes for a panel of 27 genetic variants associated with physical appearance in domestic dogs (table S5) to infer his coat coloration, patterning, and type. We also phased haplotypes from Balto's genotypes using *EAGLE* (v.2.4.1) (*38*) with reference haplotypes from the phased UMass-Broad Canid Variants and constructed the haplotype consensus sequences of the *MITF*-M promoter length polymorphism locus (chr 20: 21,839,331 to 21,839,366) and upstream SINE (short interspersed nuclear element) insertion locus (chr 20: 21,836,232 to 21,836,429) using *BCFtools* to investigate the *MITF* variants that putatively affect white spotting. We also ran a body-size prediction for Balto using a random forest model (*R* packages "caret" and "randomForest") built on the relative heights (defined as where a dog's shoulders fall relative to an "average person," and surveyed on a Likert scale from ankle-high and shorter, or survey option 0, to hip-high and taller, or survey option 4) of 1730 modern pet dogs surveyed and 2797 size-associated SNPs genotyped by the Darwin's Ark project described previously (*12*) (see supporting files for model and scripts used to run prediction).

Balto's physiological adaptations

We examined the genotypes underlying 14 regions (table S6), which included 1 region under selection in high altitude individuals (*39*) [endothelial PAS domain–containing protein 1 (*EPAS1*)], 2 regions previously identified as under selection in sled dogs (*10*) [calcium voltage-gated channel subunit alpha1 A (*CACNA1A*) and maltase-glucoamylase (*MGAM*)], 8 regions identified by population branch statistics as potentially under selection in sled dog breeds (*12*), and 3 regions responsible for aesthetic phenotypes described previously in domestic dogs [melanocortin 1 receptor (*MC1R*) (*40*), agouti signaling protein (*ASIP*) (*41*), and a chr 28 cis-regulatory region associated with single-layered coats (*13*)]. Following the method outlined in Bergström *et al.* (*21*), we also investigated the number of amylase alpha 2B (*AMY2B*) copies Balto had by quantifying the ratio of reads (reads/total length of region) mapping to the *AMY2B* regions in CanFam3.1 (ratio: 0.20) to the number of reads mapping to 75 randomly chosen 1-kb windows of the genome (ratio: 0.59), given that higher copy numbers are suggested for dog adaptation to starch-rich diets (*22*).

REFERENCES AND NOTES

1. K. Lindblad-Toh *et al*., A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482 (2011). doi: 10.1038/nature10530; pmid: 21993624
2. P. F. Sullivan *et al*., Leveraging base pair mammalian constraint to understand genetic variation and human disease. *Science* **380**, eabn2937 (2023). doi: 10.1123/science.abn2937
3. M. J. Christmas *et al*., Evolutionary constraint and innovation across hundreds of placental mammals. *Science* **380**, eabn3943 (2023). doi: 10.1123/science.abn3943
4. J. Armstrong *et al*., Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**, 246–251 (2020). doi: 10.1038/s41586-020-2871-y; pmid: 33177663
5. Zoonomia Consortium, A comparative genomics multitool for scientific discovery and conservation. *Nature* **587**, 240–245 (2020). doi: 10.1038/s41586-020-2876-6; pmid: 33177664
6. G. Salisbury, L. Salisbury, *The Cruelest Miles: The Heroic Story of Dogs and Men in a Race Against an Epidemic* (Norton, 2003).
7. N. B. Sutter, D. S. Mosher, M. M. Gray, E. A. Ostrander, Morphometrics within dog breeds are highly reproducible and dispute Rensch's rule. *Mamm. Genome* **19**, 713–723 (2008). doi: 10.1007/s00335-008-9153-6; pmid: 19020935
8. American Kennel Club, *The Complete Dog Book: 20th Edition* (Random House Publishing Group, 2007).
9. H. J. Huson, H. G. Parker, J. Runstadler, E. A. Ostrander, A genetic dissection of breed composition and performance enhancement in the Alaskan sled dog. *BMC Genet.* **11**, 71 (2010). doi: 10.1186/1471-2156-11-71; pmid: 20649949
10. M. S. Sinding *et al*., Arctic-adapted dogs emerged at the Pleistocene–Holocene transition. *Science* **368**, 1495–1499 (2020). doi: 10.1126/science.aaz8599; pmid: 32587022
11. A. V. Shindyapina *et al*., Germline burden of rare damaging variants negatively affects human healthspan and lifespan. *eLife* **9**, e53449 (2020). doi: 10.7554/eLife.53449; pmid: 32254024
12. K. Morrill *et al*., Ancestry-inclusive dog genomics challenges popular breed stereotypes. *Science* **376**, eabk0639 (2022). doi: 10.1126/science.abk0639; pmid: 35482869
13. D. T. Whitaker, E. A. Ostrander, Hair of the Dog: Identification of a *Cis*-Regulatory Module Predicted to Influence Canine Coat Composition. *Genes* **10**, 323 (2019). doi: 10.3390/genes10050323; pmid: 31035530
14. E. K. Karlsson *et al*., Efficient mapping of mendelian traits in dogs through genome-wide association. *Nat. Genet.* **39**, 1321–1328 (2007). doi: 10.1038/ng.2007.10; pmid: 17906626
15. D. L. Dreger, H. G. Parker, E. A. Ostrander, S. M. Schmutz, Identification of a mutation that is associated with the saddle tan and black-and-tan phenotypes in Basset Hounds and Pembroke Welsh Corgis. *J. Hered.* **104**, 399–406 (2013). doi: 10.1093/jhered/est012; pmid: 23519866
16. P. E. Deane-Coe, E. T. Chu, A. Slavney, A. R. Boyko, A. J. Sams, Direct-to-consumer DNA testing of 6,000 dogs reveals 98.6-kb duplication associated with blue eyes and heterochromia in Siberian Huskies. *PLOS Genet.* **14**, e1007648 (2018). doi: 10.1371/journal.pgen.1007648; pmid: 30286082
17. S. M. Schmutz, T. G. Berryere, N. M. Ellinwood, J. A. Kerns, G. S. Barsh, MC1R studies in dogs with melanistic mask or brindle patterns. *J. Hered.* **94**, 69–73 (2003). doi: 10.1093/jhered/esg014; pmid: 12692165
18. A. J. Slavney *et al*., Five genetic variants explain over 70% of hair coat pheomelanin intensity variation in purebred and mixed breed domestic dogs. *PLOS ONE* **16**, e0250579 (2021). doi: 10.1371/journal.pone.0250579; pmid: 34043658
19. H. Anderson, L. Honkanen, P. Ruotanen, J. Mathlin, J. Donner, Comprehensive genetic testing combined with citizen science reveals a recently characterized ancient MC1R mutation associated with partial recessive red phenotypes in dog. *Canine Med. Genet.* **7**, 16 (2020). doi: 10.1186/s40575-020-00095-7; pmid: 33292722
20. E. Axelsson *et al*., The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* **495**, 360–364 (2013). doi: 10.1038/nature11837; pmid: 23354050
21. A. Bergström *et al*., Origins and genetic legacy of prehistoric dogs. *Science* **370**, 557–564 (2020). doi: 10.1126/science.aba9572; pmid: 33122379
22. M. Arendt, T. Fall, K. Lindblad-Toh, E. Axelsson, Amylase activity is associated with *AMY2B* copy numbers in dog: Implications for dog domestication, diet and diabetes. *Anim. Genet.* **45**, 716–722 (2014). doi: 10.1111/age.12179; pmid: 24975239
23. X. Gou *et al*., Whole-genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-altitude hypoxia. *Genome Res.* **24**, 1308–1315 (2014). doi: 10.1101/gr.171876.113; pmid: 24721644
24. S. Köhler *et al*., The Human Phenotype Ontology in 2021. *Nucleic Acids Res.* **49** (D1), D1207–D1217 (2021). doi: 10.1093/nar/gkaa1043; pmid: 33264411
25. J. Dabney *et al*., Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15758–15763 (2013). doi: 10.1073/pnas.1314445110; pmid: 24019490
26. J. D. Kapp, R. E. Green, B. Shapiro, A Fast and Efficient Single-stranded Genomic Library Preparation Method Optimized for Ancient DNA. *J. Hered.* **112**, 241–249 (2021). doi: 10.1093/jhered/esab012; pmid: 33768239
27. J. S. John, SeqPrep: tool for stripping adaptors and/or merging paired reads with overlap into single reads. (2011); https://github.com/jstjohn/SeqPrep.
28. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010). doi: 10.1093/bioinformatics/btp698; pmid: 20080505
29. S. Kumar, G. Stecher, D. Peterson, K. Tamura, MEGA-CC: Computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. *Bioinformatics* **28**, 2685–2686 (2012). doi: 10.1093/bioinformatics/bts507; pmid: 22923298
30. S. Purcell *et al*., PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007). doi: 10.1086/519795; pmid: 17701901
31. B. S. Weir, C. C. Cockerham, Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984). pmid: 28563791
32. G. Bhatia, N. Patterson, S. Sankararaman, A. L. Price, Estimating and interpreting FST: The impact of rare variants. *Genome Res.* **23**, 1514–1521 (2013). doi: 10.1101/gr.154831.113; pmid: 23861382
33. D. H. Alexander, K. Lange, Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* **12**, 246 (2011). doi: 10.1186/1471-2105-12-246; pmid: 21682921
34. G. Zhao *et al*., Genome-Wide Assessment of Runs of Homozygosity in Chinese Wagyu Beef Cattle. *Animals (Basel)* **10**, 1425 (2020). doi: 10.3390/ani10081425; pmid: 32824035
35. P. Cingolani, snpEff: Variant effect prediction (2012).
36. M. Rezwani, A. A. Pourfathollah, F. Noorbakhsh, rbioapi: User-friendly R interface to biologic web services' API. *Bioinformatics* **38**, 2952–2953 (2022). doi: 10.1093/bioinformatics/btac172; pmid: 35561170
37. H. Mi *et al*., PANTHER version 16: A revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res.* **49** (D1), D394–D403 (2021). doi: 10.1093/nar/gkaa1106; pmid: 33290554
38. P.-R. Loh *et al*., Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016). doi: 10.1038/ng.3679; pmid: 27694958
39. B. vonHoldt, Z. Fan, D. Ortega-Del Vecchyo, R. K. Wayne, *EPAS1* variants in high altitude Tibetan wolves were selectively introgressed into highland dogs. *PeerJ* **5**, e3522 (2017). doi: 10.7717/peerj.3522; pmid: 28717592
40. S. M. Schmutz, T. G. Berryere, A. D. Goldfinch, TYRP1 and MC1R genotypes and their effects on coat color in dogs. *Mamm. Genome* **13**, 380–387 (2002). doi: 10.1007/s00335-001-2147-2; pmid: 12140685
41. T. G. Berryere, J. A. Kerns, G. S. Barsh, S. M. Schmutz, Association of an Agouti allele with fawn or sable coat color in domestic dogs. *Mamm. Genome* **16**, 262–272 (2005). doi: 10.1007/s00335-004-2445-6; pmid: 15965787

ACKNOWLEDGMENTS

We thank the Cleveland Museum of Natural History for their contributions to Balto's preservation and history and the owners of the three working Alaskan sled dogs sequenced for this work (IACUC 2014-0121). **Funding:** NIH grant R01 HG008742 (E.K.K.), NIH grant U19 AG057377 (E.K.K.), The Siberian Husky Club of America. **Author contributions:** Conceptualization: H.J.H., G.S., E.K.K., B.S. Data Acquisition: K.L.M., H.J.H., B.S., G.S. Analysis: K.L.M., H.J.H., K.M., M.S.W., X.L., K.S., E.K.K. Writing: K.L.M., H.J.H.,

K.M., X.L., E.K.K., B.S. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Raw sequencing reads for Balto and Alaskan sled dogs have been deposited to the NCBI Sequence Read Archive under BioProject accession PRJNA786530.

Zoonomia Consortium

Gregory Andrews[1], Joel C. Armstrong[2], Matteo Bianchi[3], Bruce W. Birren[4], Kevin R. Bredemeyer[5], Ana M. Breit[6], Matthew J. Christmas[3], Hiram Clawson[2], Joana Damas[7], Federica Di Palma[8,9], Mark Diekhans[2], Michael X. Dong[3], Eduardo Eizirik[10], Kaili Fan[1], Cornelia Fanter[11], Nicole M. Foley[5], Karin Forsberg-Nilsson[12,13], Carlos J. Garcia[14], John Gatesy[15], Steven Gazal[16], Diane P. Genereux[4], Linda Goodman[17], Jenna Grimshaw[14], Michaela K. Halsey[14], Andrew J. Harris[5], Glenn Hickey[18], Michael Hiller[19,20,21], Allyson G. Hindle[11], Robert M. Hubley[22], Graham M. Hughes[23], Jeremy Johnson[4], David Juan[24], Irene M. Kaplow[25,26], Elinor K. Karlsson[1,4,27], Kathleen C. Keough[17,28,29], Bogdan Kirilenko[19,20,21], Klaus-Peter Koepfli[30,31,32], Jennifer M. Korstian[14], Amanda Kowalczyk[25,26], Sergey V. Kozyrev[3], Alyssa J. Lawler[4,26,33], Colleen Lawless[23], Thomas Lehmann[34], Danielle L. Levesque[6], Harris A. Lewin[7,35,36], Xue Li[1,4,37], Abigail Lind[28,29], Kerstin Lindblad-Toh[3,4], Ava Mackay-Smith[38], Voichita D. Marinescu[3], Tomas Marques-Bonet[39,40,41,42], Victor C. Mason[43], Jennifer R. S. Meadows[3], Wynn K. Meyer[44], Jill E. Moore[1], Lucas R. Moreira[1,4], Diana D. Moreno-Santillan[14], Kathleen M. Morrill[1,4,37], Gerard Muntané[24], William J. Murphy[5], Arcadi Navarro[39,41,45,46], Martin Nweeia[47,48,49,50], Sylvia Ortmann[51], Austin Osmanski[14], Benedict Paten[2], Nicole S. Paulat[14], Andreas R. Pfenning[25,26], BaDoi N. Phan[25,26,52], Katherine S. Pollard[28,29,53], Henry E. Pratt[1], David A. Ray[14], Steven K. Reilly[38], Jeb R. Rosen[22], Irina Ruf[54], Louise Ryan[23], Oliver A. Ryder[55,56], Pardis C. Sabeti[4,57,58], Daniel E. Schäffer[25], Aitor Serres[24], Beth Shapiro[59,60], Arian F. A. Smit[22], Mark Springer[61], Chaitanya Srinivasan[25], Cynthia Steiner[55], Jessica M. Storer[22], Kevin A. M. Sullivan[14], Patrick F. Sullivan[62,63], Elisabeth Sundström[3], Megan A. Supple[59], Ross Swofford[4], Joy-El Talbot[64], Emma Teeling[23], Jason Turner-Maier[4], Alejandro Valenzuela[24], Franziska Wagner[65], Ola Wallerman[3], Chao Wang[3], Juehan Wang[16], Zhiping Weng[1], Aryn P. Wilder[55], Morgan E. Wirthlin[25,26,66], James R. Xue[4,57], Xiaomeng Zhang[4,25,26]

[1]Program in Bioinformatics and Integrative Biology, UMass Chan Medical School, Worcester, MA 01605, USA. [2]Genomics Institute, University of California Santa Cruz, Santa Cruz, CA 95064, USA. [3]Department of Medical Biochemistry and Microbiology, Science for Life Laboratory, Uppsala University, Uppsala, 751 32, Sweden. [4]Broad Institute of MIT and Harvard, Cambridge, MA 02139, USA. [5]Veterinary Integrative Biosciences, Texas A&M University, College Station, TX 77843, USA. [6]School of Biology and Ecology, University of Maine, Orono, ME 04469, USA. [7]The Genome Center, University of California Davis, Davis, CA 95616, USA. [8]Genome British Columbia, Vancouver, BC, Canada. [9]School of Biological Sciences, University of East Anglia, Norwich, UK. [10]School of Health and Life Sciences, Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, 90619-900, Brazil. [11]School of Life Sciences, University of Nevada Las Vegas, Las Vegas, NV 89154, USA. [12]Biodiscovery Institute, University of Nottingham, Nottingham, UK. [13]Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Uppsala, 751 85, Sweden. [14]Department of Biological Sciences, Texas Tech University, Lubbock, TX 79409, USA. [15]Division of Vertebrate Zoology, American Museum of Natural History, New York, NY 10024, USA. [16]Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA. [17]Fauna Bio Incorporated, Emeryville, CA 94608, USA. [18]Baskin School of Engineering, University of California Santa Cruz, Santa Cruz, CA 95064, USA. [19]Faculty of Biosciences, Goethe-University, 60438 Frankfurt, Germany. [20]LOEWE Centre for Translational Biodiversity Genomics, 60325 Frankfurt, Germany. [21]Senckenberg Research Institute, 60325 Frankfurt, Germany. [22]Institute for Systems Biology, Seattle, WA 98109, USA. [23]School of Biology and Environmental Science, University College Dublin, Belfield, Dublin 4, Ireland. [24]Department of Experimental and Health Sciences, Institute of Evolutionary Biology (UPF-CSIC), Universitat Pompeu Fabra, Barcelona, 08003, Spain. [25]Department of Computational Biology, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA. [26]Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA. [27]Program in Molecular Medicine, UMass Chan Medical School, Worcester, MA 01605, USA. [28]Department of Epidemiology & Biostatistics, University of California San Francisco, San Francisco, CA 94158, USA. [29]Gladstone Institutes, San Francisco, CA 94158, USA. [30]Center for Species Survival, Smithsonian's National Zoo and Conservation Biology Institute, Washington, DC 20008, USA. [31]Computer Technologies Laboratory, ITMO University, St. Petersburg 197101, Russia. [32]Smithsonian-Mason School of Conservation, George Mason University, Front Royal, VA 22630, USA. [33]Department of Biological Sciences, Mellon College of Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA. [34]Senckenberg Research Institute and Natural History Museum Frankfurt, 60325 Frankfurt am Main, Germany. [35]Department of Evolution and Ecology, University of California Davis, Davis, CA 95616, USA. [36]John Muir Institute for the Environment, University of California Davis, Davis, CA 95616, USA. [37]Morningside Graduate School of Biomedical Sciences, UMass Chan Medical School, Worcester, MA 01605, USA. [38]Department of Genetics, Yale School of Medicine, New Haven, CT 06510, USA. [39]Catalan Institution of Research and Advanced Studies (ICREA), Barcelona, 08010, Spain. [40]CNAG-CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology (BIST), Barcelona, 08036, Spain. [41]Department of Medicine and LIfe Sciences, Institute of Evolutionary Biology (UPF-CSIC), Universitat Pompeu Fabra, Barcelona, 08003, Spain. [42]Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, 08193, Cerdanyola del Vallès, Barcelona, Spain. [43]Institute of Cell Biology, University of Bern, 3012, Bern, Switzerland. [44]Department of Biological Sciences, Lehigh University, Bethlehem, PA 18015, USA. [45]BarcelonaBeta Brain Research Center, Pasqual Maragall Foundation, Barcelona, 08005, Spain. [46]CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology (BIST), Barcelona, 08003, Spain. [47]Department of Comprehensive Care, School of Dental Medicine, Case Western Reserve University, Cleveland, OH 44106, USA. [48]Department of Vertebrate Zoology, Canadian Museum of Nature, Ottawa, Ontario K2P 2R1, Canada. [49]Department of Vertebrate Zoology, Smithsonian Institution, Washington, DC 20002, USA. [50]Narwhal Genome Initiative, Department of Restorative Dentistry and Biomaterials Sciences, Harvard School of Dental Medicine, Boston, MA 02115, USA. [51]Department of Evolutionary Ecology, Leibniz Institute for Zoo and Wildlife Research, 10315 Berlin, Germany. [52]Medical Scientist Training Program, University of Pittsburgh School of Medicine, Pittsburgh, PA 15261, USA. [53]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA. [54]Division of Messel Research and Mammalogy, Senckenberg Research Institute and Natural History Museum Frankfurt, 60325 Frankfurt am Main, Germany. [55]Conservation Genetics, San Diego Zoo Wildlife Alliance, Escondido, CA 92027, USA. [56]Department of Evolution, Behavior and Ecology, School of Biological Sciences, University of California San Diego, La Jolla, CA 92039, USA. [57]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA. [58]Howard Hughes Medical Institute, Chevy Chase, MD, USA. [59]Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, CA 95064, USA. [60]Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, CA 95064, USA. [61]Department of Evolution, Ecology and Organismal Biology, University of California Riverside, Riverside, CA 92521, USA. [62]Department of Genetics, University of North Carolina Medical School, Chapel Hill, NC 27599, USA. [63]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [64]Iris Data Solutions, LLC, Orono, ME 04473, USA. [65]Museum of Zoology, Senckenberg Natural History Collections Dresden, 01109 Dresden, Germany. [66]Allen Institute for Brain Science, Seattle, WA 98109, USA.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn5887
Supplementary Text
Materials and Methods
Figs. S1 to S10
Tables S1 to S15
References (*42*–*53*)
MDAR Reproducibility Checklist
Data S1 and S2
Model and Scripts

View/request a protocol for this paper from *Bio-protocol*.

Submitted 7 December 2021; accepted 23 November 2022
10.1126/science.abn5887

Comparative genomics of Balto, a famous historic dog, captures lost diversity of 1920s sled dogs

Katherine L. Moon, Heather J. Huson, Kathleen Morrill, Ming-Shan Wang, Xue Li, Krishnamoorthy Srikanth, Zoonomia Consortium, Kerstin Lindblad-Toh, Gavin J. Svenson, Elinor K. Karlsson, Beth Shapiro, Gregory Andrews, Joel C. Armstrong, Matteo Bianchi, Bruce W. Birren, Kevin R. Bredemeyer, Ana M. Breit, Matthew J. Christmas, Hiram Clawson, Joana Damas, Federica Di Palma, Mark Diekhans, Michael X. Dong, Eduardo Eizirik, Kaili Fan, Cornelia Fanter, Nicole M. Foley, Karin Forsberg-Nilsson, Carlos J. Garcia, John Gatesy, Steven Gazal, Diane P. Genereux, Linda Goodman, Jenna Grimshaw, Michaela K. Halsey, Andrew J. Harris, Glenn Hickey, Michael Hiller, Allyson G. Hindle, Robert M. Hubley, Graham M. Hughes, Jeremy Johnson, David Juan, Irene M. Kaplow, Elinor K. Karlsson, Kathleen C. Keough, Bogdan Kirilenko, Klaus-Peter Koepfli, Jennifer M. Korstian, Amanda Kowalczyk, Sergey V. Kozyrev, Alyssa J. Lawler, Colleen Lawless, Thomas Lehmann, Danielle L. Levesque, Harris A. Lewin, Xue Li, Abigail Lind, Kerstin Lindblad-Toh, Ava Mackay-Smith, Voichita D. Marinescu, Tomas Marques-Bonet, Victor C. Mason, Jennifer R. S. Meadows, Wynn K. Meyer, Jill E. Moore, Lucas R. Moreira, Diana D. Moreno-Santillan, Kathleen M. Morrill, Gerard Muntan, William J. Murphy, Arcadi Navarro, Martin Nweeia, Sylvia Ortmann, Austin Osmanski, Benedict Paten, Nicole S. Paulat, Andreas R. Pfenning, BaDoi N. Phan, Katherine S. Pollard, Henry E. Pratt, David A. Ray, Steven K. Reilly, Jeb R. Rosen, Irina Ruf, Louise Ryan, Oliver A. Ryder, Pardis C. Sabeti, Daniel E. Schffer, Aitor Serres, Beth Shapiro, Arian F. A. Smit, Mark Springer, Chaitanya Srinivasan, Cynthia Steiner, Jessica M. Storer, Kevin A. M. Sullivan, Patrick F. Sullivan, Elisabeth Sundstrm, Megan A. Supple, Ross Swofford, Joy-El Talbot, Emma Teeling, Jason Turner-Maier, Alejandro Valenzuela, Franziska Wagner, Ola Wallerman, Chao Wang, Juehan Wang, Zhiping Weng, Aryn P. Wilder, Morgan E. Wirthlin, James R. Xue, and Xiaomeng Zhang

Science, **380** (6643), eabn5887.
DOI: 10.1126/science.abn5887