GENETICS

The Australian dingo is an early offshoot of modern breed dogs

Matt A. Field[1,2], Sonu Yadav[3], Olga Dudchenko[4,5], Meera Esvaran[6], Benjamin D. Rosen[7], Ksenia Skvortsova[2], Richard J. Edwards[3], Jens Keilwagen[8], Blake J. Cochran[9], Bikash Manandhar[9], Sonia Bustamante[10], Jacob Agerbo Rasmussen[11,12], Richard G. Melvin[13], Barry Chernoff[14], Arina Omer[4], Zane Colaric[4], Eva K. F. Chan[2,15], Andre E. Minoche[2], Timothy P. L. Smith[16], M. Thomas P. Gilbert[11,17], Ozren Bogdanovic[2,3], Robert A. Zammit[18], Torsten Thomas[6], Erez L. Aiden[4,5,19,20,21], J. William O. Ballard[22,23]*

Dogs are uniquely associated with human dispersal and bring transformational insight into the domestication process. Dingoes represent an intriguing case within canine evolution being geographically isolated for thousands of years. Here, we present a high-quality de novo assembly of a pure dingo (CanFam_DDS). We identified large chromosomal differences relative to the current dog reference (CanFam3.1) and confirmed no expanded pancreatic amylase gene as found in breed dogs. Phylogenetic analyses using variant pairwise matrices show that the dingo is distinct from five breed dogs with 100% bootstrap support when using Greenland wolf as the outgroup. Functionally, we observe differences in methylation patterns between the dingo and German shepherd dog genomes and differences in serum biochemistry and microbiome makeup. Our results suggest that distinct demographic and environmental conditions have shaped the dingo genome. In contrast, artificial human selection has likely shaped the genomes of domestic breed dogs after divergence from the dingo.

INTRODUCTION

Dogs are a highly successful model for informing the prehistoric movement of humans, the development of human culture, and the processes of domestication. Dingoes represent a unique lineage within canine history as they have been geographically isolated from both wolves and domestic dogs for thousands of years. It is thought that they arrived in Australia 5000 to 8500 years ago, possibly as a single introduction, and have been the continent's apex predator since the extinction of thylacines (*1*–*3*). Since their introduction, dingo populations have been naturally selected to thrive on a diet of marsupials and reptiles (*4*, *5*). The first domestic dogs were brought to Australia in 1788, and with the subsequent expansion of settlers, domestic dog DNA has introgressed into the dingo gene pool (*6*).

There is controversy concerning the evolutionary affinities of the dingo (*3*, *7*, *8*), but a recent paper concluded that the most appropriate taxonomic name is *Canis familiaris* (*9*). Within dingoes, at least two ecotypes exist, desert and alpine (*2*). Inconsistent conclusions about the evolutionary relationships of dingoes may arise from the use of the boxer genome (CanFam v3) as a reference genome for mapping reads, which by default assumes low chromosomal divergence. The boxer genome also contains 23,876 gaps, which can cloud syntenic relationships (*10*, *11*) and hinder the detection of structural variants (SVs) and copy number variants that might provide additional clarity. Phylogenetic studies using short-read assemblies for a golden jackal, three gray wolves, a dingo, a basenji, and the boxer reference suggested that the dingo is a sister lineage to domestic breed dogs (*7*). A second read mapping study (*3*) of 10 dingo genomes proposed that the dingo is closely related to Indonesian dogs, but relationships with breed dogs are clouded due to low support for many clades. A third study (*8*) used a dog single-nucleotide polymorphism (SNP) genotyping platform to conclude that the dingo is within a primitive domestic dog clade, including the Greenland sled dog and Chinese chow chow.

Here, we report the construction of the first high-quality desert dingo de novo genome (CanFam_DDS). We compared this assembly to five existing high-quality de novo dog assemblies that span the diversity of breed dogs to more accurately determine the evolutionary relationship between these canids. In addition to the standard boxer reference [CanFam3.1 (*12*)], we included highly contiguous de novo assemblies of the German shepherd dog (GSD), basenji,

[1]Centre for Tropical Bioinformatics and Molecular Biology, College of Public Health, Medical and Veterinary Sciences, James Cook University, Cairns, QLD 4878, Australia. [2]Garvan Institute of Medical Research, Victoria Street, Darlinghurst, NSW 2010, Australia. [3]School of Biotechnology and Biomolecular Sciences, UNSW Sydney, High St, Kensington, NSW 2052, Australia. [4]The Center for Genome Architecture, Baylor College of Medicine, Houston, TX 77030, USA. [5]Center for Theoretical Biological Physics, Rice University, Houston, TX 77005, USA. [6]School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia. [7]Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705, USA. [8]Julius Kühn-Institut, Erwin-Baur-Str. 27, 06484 Quedlinburg, Germany. [9]School of Medical Sciences, University of New South Wales, Sydney, NSW 2052, Australia. [10]Bioanalytical Mass Spectrometry Facility, Mark Wainwright Analytical Centre, University of New South Wales, Sydney, NSW 2052, Australia. [11]Laboratory of Genomics and Molecular Biomedicine, Department of Biology, University of Copenhagen, Copenhagen 2100, Denmark. [12]Center for Evolutionary Hologenomics, Faculty of Health and Medical Sciences, The GLOBE Institute University of Copenhagen, Copenhagen, Denmark. [13]Department of Biomedical Sciences, University of Minnesota Medical School, 1035 University Drive, Duluth, MN 55812, USA. [14]College of the Environment, Departments of Biology, and Earth and Environmental Sciences, Wesleyan University, Middletown, CT 06459, USA. [15]Statewide Genomics, New South Wales Health Pathology, 45 Watt St, Newcastle, NSW 2300, Australia. [16]U.S. Meat Animal Research Center, Agricultural Research Service, USDA, Rd 313, Clay Center, NE 68933, USA. [17]University Museum, NTNU, Trondheim, Norway. [18]Vineyard Veterinary Hospital, 703 Windsor Rd, Vineyard, NSW 2765, Australia. [19]UWA School of Agriculture and Environment, The University of Western Australia, Perth, WA 6009, Australia. [20]Shanghai Institute for Advanced Immunochemical Studies, ShanghaiTech University, Pudong 201210, China. [21]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [22]Department of Environment and Genetics, SABE, La Trobe University, Melbourne, VIC 3086, Australia. [23]School of Biosciences, University of Melbourne, Royal Parade, Parkville, VIC 3052, Australia.
*Corresponding author. Email: b.ballard@latrobe.edu.au

Great Dane, and Labrador retriever (*13*–*16*). The GSD is intermediate in the domestic dog genealogy (*17*), and CanFam_GSD was included as it has the most contiguous assembly (table S1) (*13*). The basenji is considered the most primitive breed (*7*, *17*), and we included the near-complete CanFam_BAS (*14*). The Great Dane and Labrador retriever both have contiguous long-read de novo assemblies (*15*, *16*), with a contig N50 of >1 Mb. As an outgroup, we include a recently released assembly of the Greenland wolf (also known as the Polar wolf, *Canis lupus orion*) (*18*). At least three grounds justify this as a suitable choice for describing chromosomal rearrangements in dog genomes and for rooting the phylogenomic analyses. First, it is the highest-quality wolf genome released to date. Second, Greenland wolves fall within the North American wolf clade; thus, they are basal to both dogs and Eurasian wolves (*19*). Third, Greenland wolf genomes exhibit extremely low levels of admixture with other canids such as dogs and coyotes (*19*).

We hypothesized that genetic variation between dingoes and domestic dogs would cause functional differences to arise. Sundman *et al.* (*20*) report substantial differences in wolf and dog methylation profiles and record breed-specific patterns. We, therefore, assayed the DNA methylation status of transcription start sites (TSSs), as it may serve as a proxy for gene activity (*13*). Dingoes have been shown to have only a single *AMY2B* gene copy in contrast to the copy number expansion observed in most dogs (*21*). Therefore, dingoes are expected to have reduced serum amylase, resulting in decreased ability to digest starch (*22*). We further considered that genomic differences between dingoes and dogs might affect their gut microbiomes. For example, bacteria have been shown to modulate nutrient-specific appetites in *Drosophila* (*23*). Similarly within canids, GSDs with the *AMY2B* expansion (*13*) are expected to have higher amylase levels and thus harbor a microbial community rich in species able to ferment and degrade starch products. We, therefore, surveyed the components of the microbiomes of dingoes and dogs to identify differences in content and diversity.

We selected the GSD breed for the experimental comparison because these dogs are morphologically like the dingo and are common feral canines (*17*, *24*, *25*). GSDs have been used in two previous comparative studies with dingoes (*1*, *26*) and have been used for behavioral comparisons with wolves (*27*) and wolfdogs (*28*). Yadav *et al.* (*26*) reported 62 significant plasma metabolite differences between dingoes and two domestic dog breeds (GSD and basenji). Ballard *et al.* (*1*) compared dingo with GSD and basenji. They concluded that the dingo is behaviorally intermediate between captive wolves and basenji dogs, with GSDs being most social.

RESULTS

The dingo genome is structurally distinct from five domestic breed dogs

Differences in the genome organization between carnivores have been used as examples for studying the role of chromosomal rearrangements in speciation and diet preference (*29*, *30*). Potential chromosomal differences between the dingo and domestic breed dogs have not previously been described. The "Sandy" dingo (Fig. 1A) genome (CanFam_DDS) was assembled, yielding a size of 2.35 Gb consisting of 228 contigs and 159 scaffolds (1834 contigs) with 69 gaps (Fig. 1B, figs. S1 to S3, and table S1), and was estimated to have very low error (table S2). It had a contig N50 length of 40.7 Mb and a scaffold N50 of 64.2 Mb (table S1). The chromosome-assigned scaffolds in the assembly accounted for 99.46% of the genome. Final assembly quality assessed by BUSCO analysis (*31*) identified that 5815 of 6253 conserved genes (93.0%) are present and complete in the assembly, with only 213 genes (3.40%) not found (table S1). Modified analysis using the longest isoform per annotated gene using BUSCOMP (which considers all assembly versions) increased this number to 6036 (96.5%) complete, with only 142 (2.3%) missing (Fig. 1B). Assembly quality assessed using a KAT *k*-mer analysis (*32*) showed no sign of missing data or large duplications (fig. S3).

We compared our de novo assembly with five domestic breeds that span the domestic dog genealogy, boxer (*12*), GSD (*13*), basenji (*14*), Great Dane (*15*), and Labrador retriever (*16*), as well as Greenland wolf (table S1) (*18*). As expected, the dingo assembly is highly concordant with the five domestic dog assemblies. On average, it covers 99.36% of the breed assemblies, while 99.43% of the dingo assembly aligns with the breeds. These levels amount to between 7 and 24 Mb of unique dingo sequence relative to the other five domestic dog genome assemblies. The most similar genome to CanFam_DDS was CanFam_GSD, and the most unique was CanFam3.1.

Synteny plots were generated to detect large structural genomic rearrangements [typically >1000 base pairs (bp)] (fig. S4). Rearrangements were then mapped onto a phylogenetic tree including the Greenland wolf as an outgroup (Fig. 2). We found that the Greenland wolf has one unique inversion on chromosome 26 and the dingo on chromosome 16 (~3.45 Mb located in dingo reference chr16:10.55-14.0Mb). A chromosomal hotspot of genomic rearrangements occurs on chromosome 26 in the Greenland wolf and all the breed dogs (~1.5 Mb located in dingo reference chr26:25.5-27.0Mb), suggesting that the complex rearrangements are dingo specific. We further corroborate previous reports that the dingo, like most wolves and some arctic dog breeds, has a single copy of *AMY2B* (*21*) and

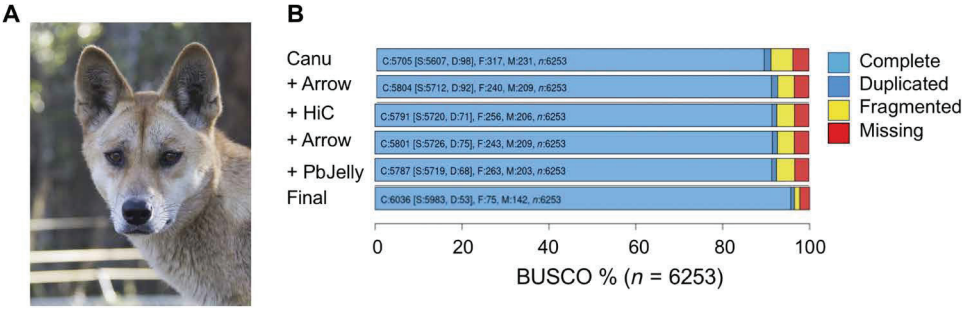

Fig. 1. The dingo Sandy and assembly statistics. (**A**) Sandy as a 3-year-old. She was found as a 4-week-old puppy in a remote region of South Australia in 2014. Subsequent genetic testing showed that she was a pure desert dingo. (**B**) BUSCOMP completeness scores for different stages of the genome assembly (C, complete; S, single; D, duplicated; F, fragmented; M, missing). BUSCOMP uses BUSCO v3 to calculate values for the longest isoform per annotated gene across all assemblies.

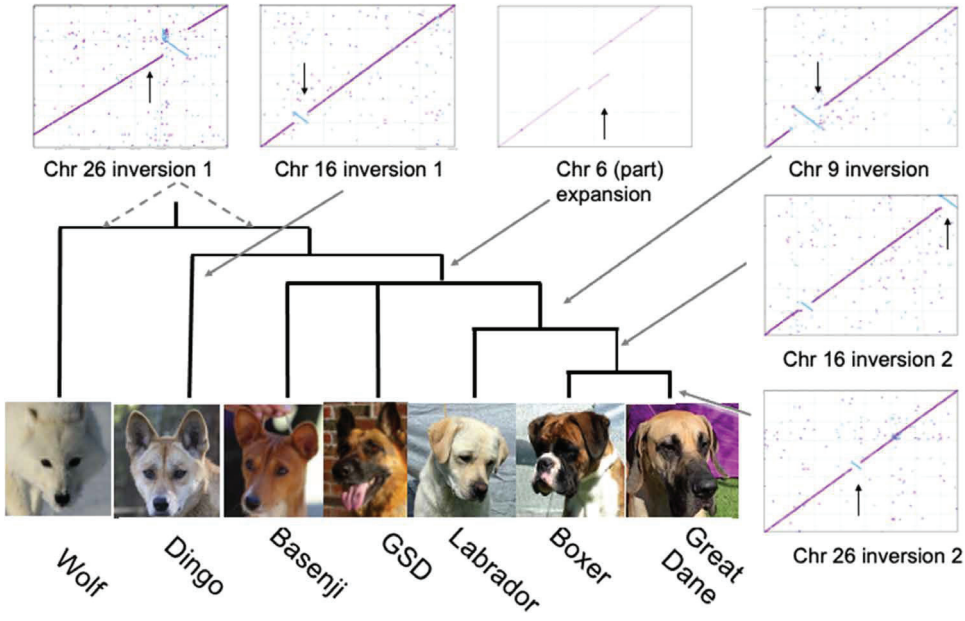

Fig. 2. Comparative mapping analyses. Mapping analysis showing the dingo is distinct from five domestic breed dogs. Phenogram plotting the major structural changes in the de novo genome assemblies of seven canids, with Greenland wolf as the outgroup. Within each box is the comparison of the breed dog (*y* axis) versus dingo (*x* axis). Except for chromosome (Chr) 6, all other boxes are whole chromosome alignments. Chromosome 6 (part) has a single copy of the gene coding for pancreatic amylase, *AMY2B*, in the dingo and an expansion in breed dogs. Chromosome 9 has a common inversion in Labrador retriever, boxer, and Great Dane. Chromosome 16 has two inversions. Inversion 1 appears only in the desert dingo. Inversion 2 occurs in boxer and Great Dane. Chromosome 26 has an inversion in the wolf that could be either lineage specific or rearranged in dingoes and breed dogs (uncertainty denoted by dashed line) and a unique inversion in the Great Dane. There is also a "hotspot" in all breed dogs but not the dingo (cluster of dots on the diagonal line). Not shown is a polymorphic inversion on chromosome 11 (see fig. S4C). Photo credits: Dingo photographer: Barry Eggleton, Pure Dingo Sanctuary; basenji photographer: Jenifer Power, Zanzipow Kennels; GSD photographer: Alan Brooks, Outdoor Action Photography; Greenland wolf photographer: Morten Petersen, Morten Petersen photography; Labrador, boxer, and Great Dane photographer: J.W.O.B.

there is no evidence of duplication loss (Fig. 2). However, in this region, we observed a 6.4-kb long interspersed nuclear element (LINE) element in the wolf genome compared to the dingo (fig. S4A), with such transposable elements well known in canines (*33*). A heterozygous 203-bp deletion was also detected in the Pacific Biosciences (PacBio) dingo reads relative to CanFam_GSD and CanFAM_BAS (*13*, *14*). Overall, there are at least three large chromosomal differences between CanFam_DDS and CanFam3.1: Two occur on chromosome 16 (3.45 and 4.99 Mb) and one on chromosome 9 (9.55 Mb) (Fig. 2). The 3.45-Mb dingo-specific chromosome 16 rearrangement overlaps 60 unique ENSEMBL transcripts (table S3) and was enriched for gene ontology terms of cellular metabolic processes according to the PANTHER (protein annotation through evolutionary relationship) classification system (www.pantherdb.org/) (fig. S5). Five pathways were overrepresented, including glycolysis and glucose metabolism (table S4). Limitations of this approach mean that we cannot rule out the possibility that some of these rearrangements are due to assembly errors in individual breeds other than dingo. Also, the data do not establish whether these rearrangements occur in all wolves and dingoes, respectively, or just these animals.

A complete list of SVs (>50 bp) were identified by mapping Nanopore and PacBio dingo long reads to the domestic breed assemblies. A conservative list of SVs was generated, consisting of the intersection of Nanopore and PacBio calls that account for potential false positives specific to either technology (*34*). This consensus approach showed a high degree of overlap in the technologies, resulting in a mean of 65,000 SVs called across the five breed dogs. Of the total SV calls, more than 99.5% are either insertions or deletions, with the remainder consisting of inversions, duplications, and translocations. SVs were prioritized for further investigation if they were homozygous and overlapped existing protein-coding gene annotations in the respective assemblies for CanFam3.1 (24,489 SVs representing 8434 unique genes), CanFam_GSD (21,921 SVs/7627 unique genes), CanFam_BAS (20,743 SVs/7002 unique genes), Great Dane (18,621 SVs/7370 unique genes), and Labrador retriever (21,575 SVs/7077 unique genes). These prioritized SVs represented a mean of 24.2 Mb of deleted and 2.0 Mb of inserted sequence in dingo compared with the five breed dogs.

Next, small indels (<50 bp) and single-nucleotide variations (SNVs) were called between dingo and each of the five domestic breeds, which was compared to variation among the five domestic breeds. This highlighted the difference between the dingo and domestic breeds, with an average of 6,598,389 small indels between dingoes and domestic dogs compared to an average of 6,005,034 small indels between the breed dogs (Table 1). Similarly for SNVs, variant analysis detected an average of 4,227,702 SNVs between the dingo and the dog breeds and 3,601,013 SNVs between the dog breeds (Table 1).

Phylogenetic analyses from these distance matrices and additional Greenland wolf information (Table 1) show that the dingo is highly differentiated and an outgroup to the five domestic dog breeds with 100% bootstrap support (Fig. 3, A and C). This result is supported by nonmetric multidimensional scaling (NMDS) (Fig. 3, B and D). Both small indels and SNV datasets strongly suggest that the basenji is the basal breed dog, but there is conflict over the evolutionary position of the Labrador, GSD, Great Dane, and boxer. Further genomic studies including other breed dogs are required.

Table 1. Distance matrix tables. SNVs above diagonal and indels below. All possible pairwise alignments were generated using MUMmer4 (*72*) (v4.0.0 beta 2), and SNVs/indels numbers were calculated using MUMmer4 "show-snp" script.

	Dingo	Basenji	German shepherd dog	Labrador	Boxer	Great Dane	Greenland wolf
Dingo	–	4,379,273	4,157,347	4,266,975	4,476,850	3,858,069	7,273,469
Basenji	6,813,866	–	3,893,739	3,922,731	4,223,170	3,605,316	7,606,227
German shepherd dog	6,290,364	6,237,235	–	3,477,794	3,645,169	3,007,546	7,339,762
Labrador	7,072,684	6,758,598	6,029,553	–	3,745,847	3,162,230	8,130,863
Boxer	6,213,950	6,063,455	5,144,715	5,900,366	–	3,326,597	7,215,476
Great Dane	6,601,081	6,455,860	5,582,788	6,440,463	5,437,308	–	7642,878
Greenland wolf	7,273,469	7,606,227	7,339,762	8,130,863	7,215,476	7,642,878	–

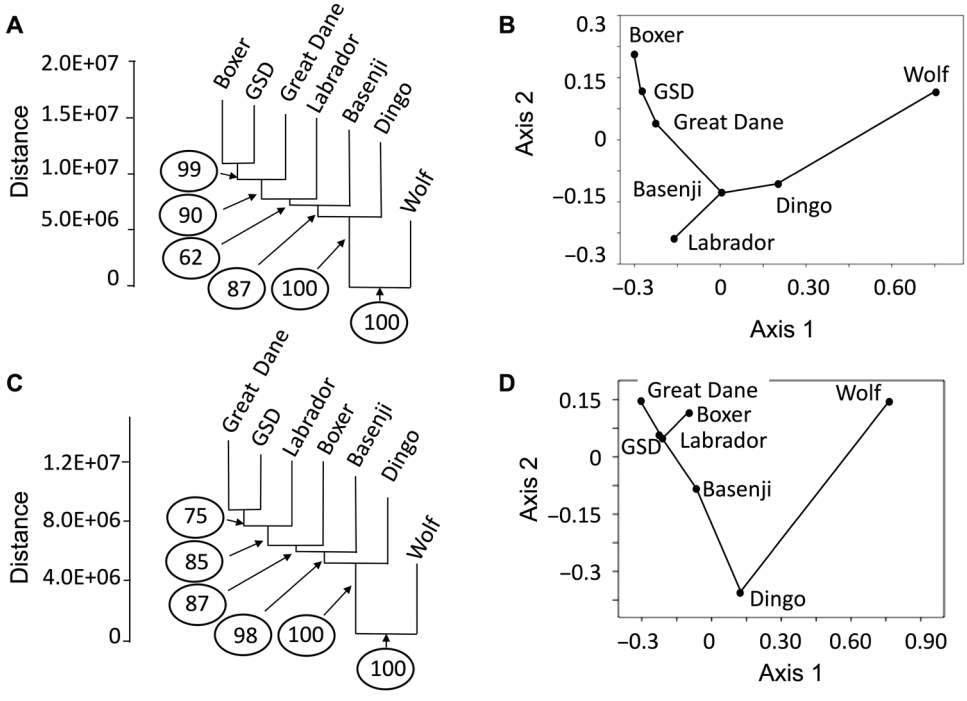

Fig. 3. Phylogenetic and ordination analyses from indels and SNVs of seven canines. (**A**) Phylogenetic tree from indels with bootstrapping (percentages of times that node appeared in 500,000 bootstraps). (**B**) Ordination analyses of first two axes from nonmetric multidimensional scaling from indels. (**C**) Phylogenetic tree from SNVs with bootstrapping (500,000 bootstraps). (**D**) Ordination analyses of first two axes from nonmetric multidimensional scaling from SNVs. GSD, German shepherd dog; Wolf, Greenland wolf.

The dingo genome's assembly, annotation, and comparative analyses show that it has diverged from domestic dog breeds. This divergence has previously been shown to influence the plasma metabolome (*26*), gastric capacity, and the digestion of proteins (*35*, *36*), behaviors (*1*), and cranial morphology (*37*). In the next section, we test whether these differences influence the functions of epigenetic signaling, the circulatory metabolome, and the gut microbiome of dingoes and GSDs.

The dingo is functionally distinct from GSDs

We hypothesized that differences between the dingo and domestic dog genomes could functionally influence the DNA methylation patterns and nutrient bioavailability. Highly methylated gene promoters often indicate a transcriptionally repressed state, while unmethylated gene promoters specify a permissive state (*38*). Genome-wide analysis showed reduced DNA methylation in dingo relative to GSD for *GAL3ST1*, *NAP1L5*, *FAM83F*, *MAB21L1*, and *UPK3A* gene promoters. In contrast, *LIME1* and *GGT5* gene promoters had hypermethylation in the dingo (Fig. 4A and fig. S6A). Consistent with the observed decreased methylation, we observed elevated *GAL3ST1* and *MAB21L1* transcript abundance in pure dingoes (Fig. 4B and fig. S6B). *GAL3ST1* is of particular interest as it may differentially influence nutrient metabolism in dingoes and dogs (*39*, *40*). We also identify a 192-bp homozygous insertion in *GAL3ST1*

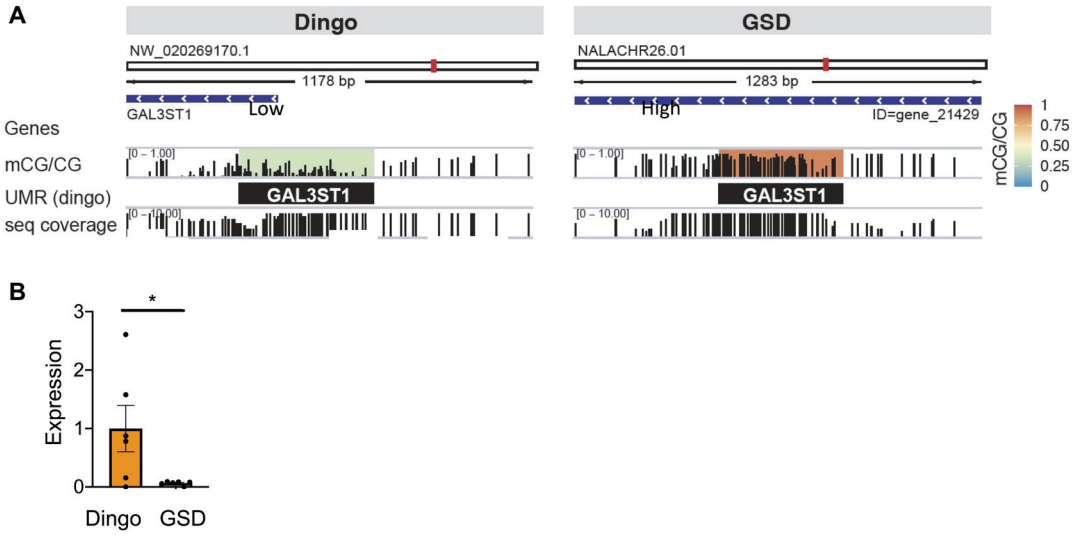

Fig. 4. Methylation differences between dingo and GSD. (**A**) Integrative Genomics Viewer (IGV) browser tracks depicting DNA methylation differences at GAL3ST1 gene promoter. GAL3ST1 gene promoter overlaps with the unmethylated region (UMR) in the dingo while showing increased DNA methylation in the GSD. The green box shows the low activity in the dingo, and the brown box shows higher activity in the GSD. The color scale depicting average promoter DNA methylation is shown on the right. It goes from blue (unmethylated) to red (fully methylated). (**B**) Significant difference in expression of GAL3ST1 between dingo and GSD ($t_{10} = 2.361$, $P = 0.03$, dingo $n = 6$, GSD $n = 6$). Means ± SE are shown on plots across all assemblies; *$P < 0.05$.

in the Great Dane assembly and a 553-bp deletion in *GGT5* in the CanFam 3.1 assembly. The functional impact of these changes in the two domestic breeds is unknown.

To observe dietary differences, food and water were standardized in a dietary study of 17 dingoes and 15 GSDs (table S5), with microbial priority effects minimized by treating animals with an antibiotic and feeding them a probiotic at the commencement of the study. Blood was drawn, and fresh scat was collected at the study's beginning and end (15 days). The number of animals analyzed for each phenotype varied and is indicated in the legend of Fig. 5. Consistent with the single copy of *AMY2B* in the genome, serum amylase levels were lower in the dingoes than in the GSDs (fig. S7, A and B) (*41*). Total cholesterol and low-density lipoprotein (LDL) were significantly higher on day 15 in the dingoes than in the GSDs (Fig. 5, A and B), while there were no apparent differences in high-density lipoprotein, lipase, or triglyceride levels (Fig. 5B and fig. S7, C and D). The difference in cholesterol levels predicts that bile acid levels would differ between canids, as primary bile acids are synthesized from cholesterol in the liver (*42*). We did not observe any significant difference in the concentration of primary bile acids, but the levels of two secondary bile acids differed (Fig. 5C and table S6). Levels of ursodeoxycholic acid and lithocholic acid were higher in GSDs than in dingoes. Ursodeoxycholic acid is a naturally occurring secondary bile acid produced by the bacterial metabolism of the primary bile acid, chenodeoxycholic acid. It is known to be metabolized to lithocholic acid in the colon (*43*).

The gut microbiomes of dingo and GSD were compared by 16*S* ribosomal RNA (rRNA) gene sequencing of the scat microbiome (table S7). Scat was collected on days 1 and 15 (Fig. 6A). For dingoes, the statistical effect size is moderate-large (Hedges $g = 0.99$), with mean alpha diversity declining by 6% (Fig. 6A). For GSDs, the effect size is large (Hedges $g = 1.34$), with mean alpha diversity increasing by 10% (Fig. 6A). At day 15, dingoes have lower alpha diversity and microbial richness than GSDs (Fig. 6A). Analysis of the microbiome composition showed that one microbial phylum, 17 families, and 51 genera differed between the canids (fig. S8C and table S8). The family Clostridiaceae and the genus *Clostridium sensu stricto 1*, which can use complex resistant starch (*44*), were enriched in dingoes (Fig. 6, B and C, and fig. S8C). Unexpectedly, 2 of the 17 dingoes included in the study scavenged upon a brushtail possum that fell into the Pure Dingo Sanctuary during the experiment. Those two animals had high numbers of Prevotellaceae (Fig. 6C). Bacteria of the families Lactobacillaceae, Ruminococcaceae, and Prevotellaceae, which are involved in fermentation and degradation of starch products (*45*, *46*), were elevated in GSDs in comparison to dingoes. The genera *Lactobacillus* and *Eubacterium*, which have a demonstrated capacity for reducing cholesterol levels (*45*, *47*), were higher in the domestic breed, consistent with previous results. Functional prediction based on 16*S* rRNA gene data showed higher metabolic potential for cholesterol and protein metabolism and lower metabolic potential for secondary acid bile secretion in the microbial communities of dingoes (fig. S8D), which is in alignment with the higher observed serum cholesterol (Fig. 5A) and lower secondary bile acid levels (Fig. 5C).

DISCUSSION

In Australia, dingoes have been isolated from both wolves and domesticated canines for thousands of years. This geographic isolation has prevented ongoing introgression and thereby provides unique insight into lineage-specific effects in dingoes and the evolutionary history of dogs. Our data show that the dingo genome has diverged substantially from the five high-quality domestic dog assemblies tested but forms a monophyletic group with these breeds relative to the Greenland wolf. Likely, this divergence is due to the ancient separation, recovery of genetic variation since the bottleneck of colonization, and natural selection for feeding on marsupials (*1*–*5*). In comparison, the evolution of domestic dogs has likely been shaped

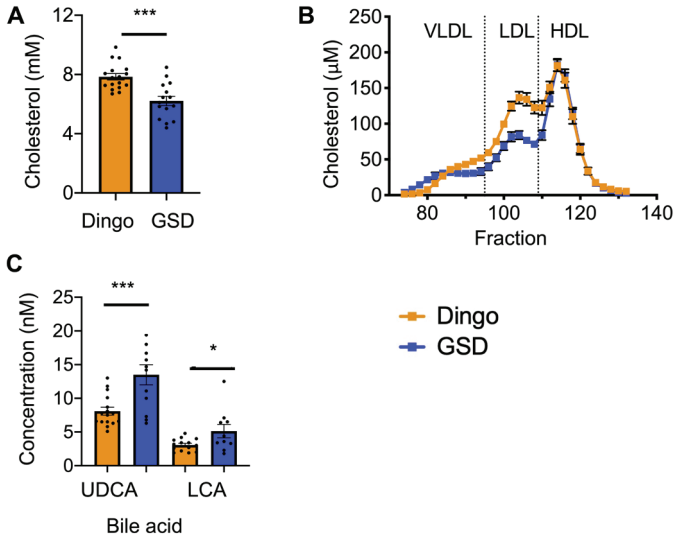

Fig. 5. Biochemical and physiological differences between dingoes and GSD. (**A**) Total cholesterol is significantly higher in the dingoes as compared to GSDs ($t_{30} = 4.36$, $P = 0.0001$; dingo $n = 17$, GSD $n = 15$). (**B**) Low-density lipoprotein cholesterol (LDL-C) is elevated 2.2-fold in dingoes ($t_{10} = 4.64$, $P < 0.001$; dingo $n = 6$, GSD $n = 6$) but no obvious difference in high-density lipoprotein cholesterol (HDL-C) levels. Individual points are within symbol size. VLDL, very-low-density lipoprotein. (**C**) Two secondary bile acids ursodeoxycholic acid (UDCA) ($t_{26} = 3.732$, $P < 0.001$; dingo $n = 16$, GSD $n = 12$) and lithocholic acid (LCA) ($t_{22} = 2.314$, $P = 0.030$; dingo $n = 14$, GSD $n = 10$) are significantly lower in dingoes; *$P < 0.05$ and ***$P < 0.001$.

by feeding on starch-rich diets in the Neolithic, high-fat diets during the agricultural revolution and artificial selection for breed-specific traits over the past 200 years (*13*, *14*, *21*).

We present multiple independent lines of evidence that distinguish the dingo genome from that of the domestic dog breeds. There are at least three large chromosomal differences between CanFam_DDS and CanFam3.1 in addition to the previously described chromosome 6 *AMY2B* copy number expansion in most breed dogs. These three inversions and the *AMY2B* duplication occur in different subsets of breeds, so it is unlikely that they are assembly errors. Furthermore, chromosome 26 has a unique inversion in the Great Dane, showing that chromosomal rearrangements are found in domestic dogs. On average, we identified 21.78 Mb of large SVs in the dingo compared to breed dog reference genomes. In comparison, there are estimated to be 18.7 Mb of large variations in the average human genome (*48*).

Phylogenetic analyses derived from indels and SNVs strongly support the hypothesis that dingoes are distinct from the five breed dogs tested with 100% bootstrap support (*7*). This result is supported by NMDS calculated from both the indel and SNV distance matrices. The basenji genome is also highly differentiated but in ways that are distinct from dingoes. The genetic differentiation of the basenji may be due to a past hybridization event that is worth further exploration (*7*). It remains possible that one or more domestic dog breeds not included may be basal to the ones studied. Two fascinating breeds to include in detailed future studies are the Greenland sled dog and the Australian cattle dog. The Greenland sled dog has a low *AMY2B* copy number without introgression from wolves (*21*). It is proposed that the Australian cattle dog is derived from a cross between a merle dog imported from England and the dingo. A third dog worthy of inclusion in future studies is the New Guinea singing dog, as it is proposed to be the sister to the dingo (*3*, *8*). De novo sequencing of these dogs, and the inclusion of multiple outgroups, will test the conclusions reached here.

We probed the genomic differences between the dingo and a GSD and detected variation in the methylation status of seven gene promoters. Among these, *GAL3ST1* promoters showed reduced DNA methylation and higher mRNA expression levels in the dingo. *GAL3ST1* is associated with galactose metabolism through its substrate galactosphingolipids for which uridine 5′-diphosphogalactose (UDPgal) is the source of galactose (*39*, *40*). *GAL3ST1* is a sulfotransferase involved in sulfoglycolipid synthetic pathways that lead to myelination of nervous tissue and spermatogenesis (*49*). Yadav *et al.* (*26*) showed that UDPgal levels were significantly lower in GSD than in dingoes. This is consistent with the observations that the GSD *GAL3ST1* promoter has increased methylation and lower transcript abundance. We note that blood was necessarily used as a proxy tissue due to animal welfare concerns when studying captive dingoes and kennel GSDs (*50*), and that determination of methylation patterns in the brain could be valuable if it becomes possible to collect appropriate samples in the future. Sundman *et al.* (*20*) compared the DNA methylation differences in three wolf brains and 38 dogs of eight breeds and concluded that epigenetic factors may have been necessary for canid speciation and the divergence of different dog breeds.

The dietary study detected differences in the serum metabolome that appear to be driven by genomic variations and potentially feedback from the microbiome. Dingoes had lower serum amylase than GSDs due to the decreased copy number of *AMY2B*. Dingoes also had higher cholesterol, higher LDL levels, and lower levels of two secondary bile acids. Elevated cholesterol and LDL levels are protective against infection in humans (*51*), suggesting that captive dingoes may have immune response that differs to GSDs. This hypothesis is supported by reports that the secondary bile acids ursodeoxycholic acid and lithocholic acid are involved in immune responses in human cell lines (*52*, *53*) and exert anti-inflammatory actions in mouse colons (*43*). Ursodeoxycholic acid has long been recognized to have broad-ranging protective actions. For centuries, it has been used in traditional Chinese medicine as a component of bear bile to treat hepatic disorders (*54*). More recently, in Western medicine, it has been used to treat liver inflammation (*43*). A constraint of our dietary study is that it was conducted under controlled conditions and limits extrapolation to natural differences. Future studies testing the immune response in wild dingoes may offer insights into the evolution of the dingo and mammalian immune systems.

We argue that the genomic disparities and the resulting physiological differences between the dingo and GSD alter their respective microbiomes. When gut microbiomes of the canids were provided identical diets for 15 days, 17 bacterial families differed between dingoes and dogs. Only the genus *Clostridium sensu stricto 1* was enriched in dingoes compared to 16 genera with higher relative abundances in GSDs. This taxonomic shift and their associated functional changes are consistent with the prediction that decreased amylase levels in dingoes under the same rice intake will result in greater microbial accessibility to carbohydrates (*55*). We suggest that the distinct microbiome patterns in dingoes result from genome-driven variation and are not simply related to their prior history for three main reasons. First, microbial diversity of wild-born and captive-born dingoes differed by <3% at completion of the study (Shannon diversity 4.39 ± 0.09, $n = 5$ and 4.27 ± 0.08, $n = 12$, respectively).

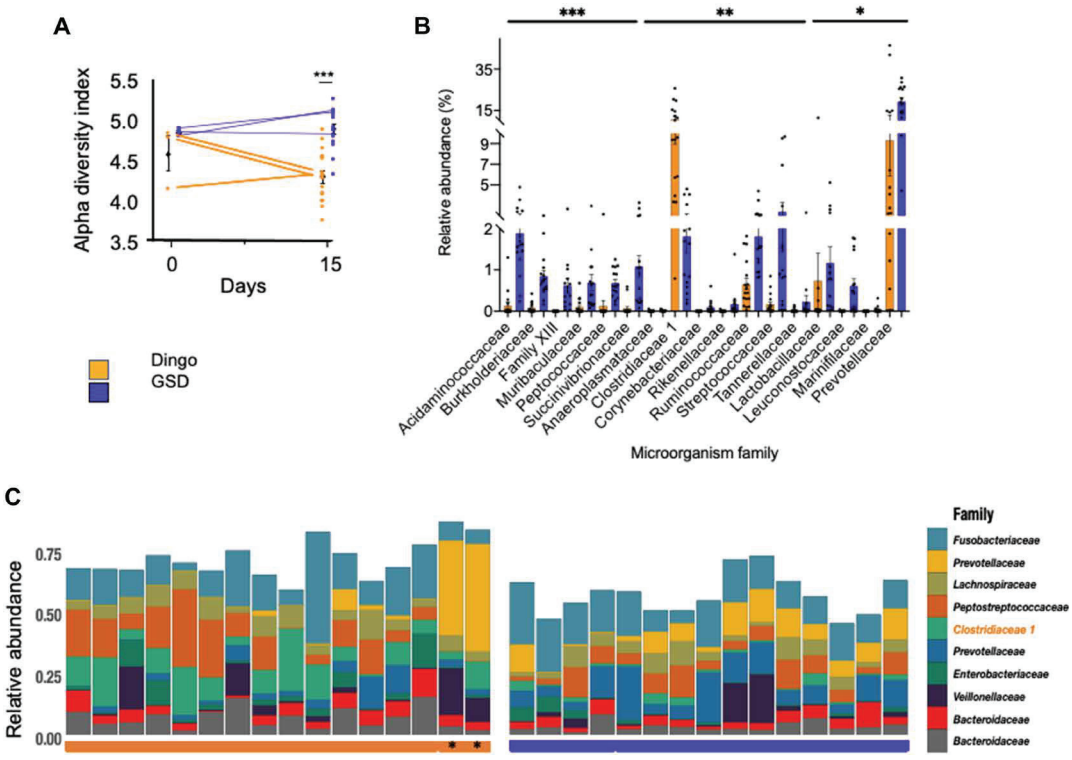

Fig. 6. Microbial diversity. (**A**) Microbial diversity is lower in the dingo than GSD on day 15. Lines connect the same individual on the two dates assayed. Wilcoxon rank sum test $P_{adj} = 0.00003$; dingo $n = 17$, GSD $n = 15$. (**B**) Relative abundance of significantly different families between the dingo and GSD as shown by Analysis of the Composition of the Microbiome (ANCOM) (Benjamini-Hochberg corrected) (dingo $n = 17$, GSD $n = 15$). (**C**) Relative abundance of the top 10 most abundant zOTUs at the end of the diet study on the *y* axis for the dingoes ($n = 16$) and GSDs ($n = 15$) along the *x* axis. *Clostridiaceae 1* is highlighted in the legend as it is elevated in dingoes. Dingoes 15 and 16 (shown with *) fed upon a brushtail possum during the experiment. For (B) and (C), dingo is orange and GSD blue; $^*P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 0.001$.

Second, mean microbial alpha diversity of animals from Bargo and Pure Dingo sanctuaries differed by <6% at completion of the study (Shannon diversity 4.35 ± 0.09, $n = 14$ and 4.10 ± 0.1, $n = 3$, respectively). Third, the dingo microbiome is similar in broad taxonomic composition to that of wolves (*56*), which have a single copy of *AMY2B* (*21*). Testing the microbial communities of breed dogs with a single copy of amylase would test whether the observed differences are driven by the *AMY2B* copy number or there is a broader relationship. More detailed studies examining amylase levels and the microbial communities of hybrids may provide a roadmap for field testing of wild dogs in Australia.

Our inclusive study reinforces the view that the dingo genome is structurally and evolutionarily distinct from domestic breed dogs, which may translate into functional differences in the ecosystem. Dingoes often consume the most abundant species in native ecosystems, including marsupials and reptiles with high protein (P):low fat (F):low carbohydrate (C) content (*2*). The preferred P:F:C profile of dingoes is currently unknown, but Bosch *et al.* (*57*) reported that the selected ratio of wolves is 54:45:1 P:F:C. In contrast, it seems likely that domestic dog evolution is shaped by feeding on starch-rich diets in the Neolithic, high-fat diets during the agricultural revolution and by artificial selection for breed-specific traits. Genomic signatures of adaptation to high-fat diets during the agricultural revolution have been documented in humans (*58*). Most domestic dog breeds have been created in the past 200 years (*13*, *14*). The selected P:F:C profile of breed dogs is 30:63:7 (*59*), but see (*60*), with similar dietary profiles between the five dog breeds indicating that they predate their recent phenotypic divergences (*59*).

We have shown that the dingo genome is distinct from the five breed dogs tested, but have not established whether the dingo was ever domesticated or determined the consequences of ongoing introgression with domestic breeds. It is unlikely that dingoes were domesticated in Australia, but it is possible that it occurred before their arrival (*2*). Incorporation of ancient DNA from dogs in known archeological contexts may help to resolve this dilemma. Introgression of domestic dog DNA into dingo populations is now prevalent (*6*), but the impacts of introgression on dingo behavior and physiology are unknown, as most ecological studies have unknowingly combined genetically pure dingoes and dingo-dog hybrids. The inability to distinguish the physiology and behavior of pure dingoes has led to a scientific debate on their role in the ecosystem (*61*) and underpins politicians questioning the value of conservation efforts of this ancient dog. Focused studies examining the roles of pure dingoes in the ecosystem and the consequences of hybridization are urgently required.

MATERIALS AND METHODS

Experimental design

The desert dingo genome data generation pipeline is described in detail in fig. S1.

Sampling/ethics

The desert dingo named Sandy was found in a remote region of South Australia in 2014. She was rescued with her two siblings and transported to eastern Australia. Subsequent genetic testing (*62*) showed that she was a pure dingo. All samples were collected under University of New South Wales Ethics Approval IDs 16/77B and 18/148B.

Sequencing

The genome was assembled using PacBio single-molecule real-time (SMRT) sequencing, Oxford Nanopore Technologies (ONT) PromethION sequencing, 10X Genomics Chromium genome sequencing, and Hi-C scaffolding (fig. S1). Contigs were assembled using SMRT and ONT sequencing with the Canu assembler (Canu, RRID: SCR_015880; v1.8.0) and then polished with Arrow to minimize error propagation (fig. S1). After scaffolding, gaps were filled using the SMRT and ONT reads, followed by a final round of polishing, including aligning the 10X Chromium reads to the assembly and Pilon polishing. The resulting chromosome-length genome assembly has been deposited to the National Center for Biotechnology Information (NCBI) (GCA_003254725.2). The mitochondrial genome has been submitted (ID 2385777) and is linked with the BioProject and BioSample.

Genomic DNA was prepared from a skin biopsy. Extraction was performed with supplemental ribonuclease (Astral Scientific, Taren Point, Australia) and proteinase K (New England Biolabs, Ipswich, MA, USA) treatment, as per the manufacturer's instructions. Isolated genomic DNA was further purified using AMPure XP beads (Beckman Coulter, Brea, CA, USA). DNA purity was calculated using NanoDrop (Thermo Fisher Scientific), and molecular integrity was assessed using pulse-field gel electrophoresis. Sage Science Pippin Pulse assessed DNA integrity. A 0.75% KBB gel was run on the 9-hour 10- to 48-kb (80 V) program. DNA ladder used was the Invitrogen 1 Kb Extension DNA Ladder (catalog no. 10511-012). One hundred fifty nanograms of DNA was loaded on the gel.

We generated two libraries that were size-selected on Sage BluePippin gels (Sage Science, Beverly, MA, USA). Libraries were sequenced on Sequel machines with 2.0 chemistry recording 10-hour movies. Sequencing was conducted at the Arizona Genomics Institute, University of Arizona.

ONT PromethION sequencing

DNA (1 μg) was prepared for ONT sequencing using the one-dimensional (1D) genomic DNA by ligation kit (SQK-LSK109, ONT) according to the standard protocol. Long fragment buffer was used for the final elution to exclude fragments shorter than 1000 bp. In total, 119 ng of adapted DNA was loaded onto an FLO-PRO002 PromethION flow cell and run on an ONT PromethION sequencing device (PromethION, RRID: SCR_017987) using MinKNOW (18.08.2) with MinKNOW core (v1. 14.2).

Base calling was performed after sequencing with the graphics processing unit (GPU)–enabled guppy base caller (v3.0.3) using the PromethION high-accuracy flip-flop model with conFig. "dna_r9.4.1_450bps_hac.cfg." Sequencing was conducted at Kinghorn Centre for Clinical Genomics at the Garvan Institute of Medical Research, Sydney, Australia.

10X Genomics Chromium sequencing

DNA was prepared following the protocol described above for SMRT sequencing. A 10X GEM library was barcoded from high–molecular weight DNA according to the manufacturer's recommended protocols. The protocol used was the Chromium Genome Reagent Kits v2 User Guide, manual part number CG00043 Rev B. Quality Control was performed using LabChip GX (PerkinElmer, MA, USA) and Qubit 2.0 Fluorometer (Life Technologies, CA, USA). The library was run on a single lane of a v2 patterned flowcell. Paired-end sequencing with 150-bp read length was performed using the Illumina HiSeq X (Illumina HiSeq X Ten, RRID: SCR_016385) within the Kinghorn Centre for Clinical Genomics at the Garvan Institute of Medical Research, Sydney, Australia.

Long-read genome assembly

The SMRT and ONT reads were corrected and assembled. With a total length of 2,427,850,753 bp, the assembled genome consisted of 1834 contigs with an N50 length of 24.1 Mb (including 152 repeats of total length 16,671,837 bp) with no bubbles. There were 2,000,973 unassembled sequences of total length 13,107,822,345 bp. The resulting contigs were polished by aligning the raw reads to the assembly and correcting the sequencing errors using Arrow polishing. There were 2,934,153 fixes implemented. Following Arrow polishing, there were 1834 sequences with a total length of 243,110,9461 bp.

Chromosome-length assembly using Hi-C data

An in situ Hi-C library was prepared from a blood sample from the same individual (fig. S2). The Hi-C data were aligned to the polished contig set using Juicer (*63*) and input into the 3D-DNA pipeline (*64*) to produce a candidate chromosome-length genome assembly. We performed additional finishing on the resulting scaffolds using Juicebox Assembly Tools (*65*). Figure S2 shows the contact matrices generated by aligning the Hi-C dataset to the genome assembly before the Hi-C upgrade (on the left) and after Hi-C scaffolding (on the right). The matrices are visualized in Juicebox.js, a cloud-based visualization system for Hi-C data (*66*), and are available for browsing at multiple resolutions at DNA Zoo. This process reduced the number of scaffolds to 210 (N50 64.2 Mb), introducing 197 gaps. Subsequent polishing with Arrow closed 29 of these gaps, increasing contig N50 to 26.2 Mb (fig. S2).

Gap filling, Pilon polishing, and final cleanup

After scaffolding and correction, all raw SMRT and ONT reads were aligned to the assembly with Minimap2 (v2.16) (-ax map-PB/map-ont) and used by PBJelly (pbsuite v.15.8.24) to fill gaps. It was able to close 74 gaps, increasing contig N50 to 36.2 Mb. The third round of Arrow polishing closed a further 25 gaps, increasing contig N50 to 40.7 Mb.

To further improve the assembly, another round of polishing was performed by aligning the 10X Chromium reads to the assembly using the linked-read analysis software provided by 10X Genomics, Long Ranger, v2.2.2 (fig. S1). Small indels were then corrected using Pilon v1.23 (diploid mode) (fig. S1).

The Pilon-polished genome was mapped onto CanFam v3.1 chromosomes with PAFScaff v0.3.0. It then underwent a final scaffold cleanup with Diploidocus v0.9.6 ("Nala" purge mode) to generate a high-quality core assembly, remove low-coverage artifacts and haplotig sequences, and annotate remaining scaffolds with potential issues. PacBio subreads (15.8 M subreads; 149.9Gb) and ONT "pass" reads (6.12 M reads; 49.8Gb) were mapped onto the assembly using Minimap2 v2.17 (-ax map-PB or -ax map-ont --secondary = no) (*67*), and read depth summaries were calculated with BBMap

v38.51 pileup.sh (*68*). Any scaffolds with a median coverage of less than three (e.g., less than 50% of the scaffold covered by at least three reads) were filtered out as low-coverage scaffolds. Single-copy read depth was estimated using the modal read depth of 75X across the 5736 single-copy complete genes identified by BUSCO v3.0.2b (*31*). This was used to set low-, mid-, and high-depth thresholds for PurgeHaplotigs v20190612 (*69*) (implementing Perl v5.28.0, BEDTools v2.27.1, R v3.5.3, and SAMTools v1.9) at 18X, 56X, and 150X, respectively, to remove allelic contigs. PurgeHaplotig coverage parameter was adjusted to exclude gap regions. Any scaffolds with ≥80% bases in the low/haploid coverage bins and ≥95% of their length mapped by PurgeHaplotigs onto another scaffold were filtered as haplotigs or assembly artifacts. Any other scaffolds with ≥80% low coverage bases were filtered as Low Coverage. In total, 11 sequences (93.2 kb) were removed as low-coverage artifacts, and a further 31 (438.7 kb) were removed as probably haplotigs. Evaluation of the completion of the conserved single-copy genes was performed by BUSCO v3.0.2b, short mode, implementing BLAST+ v2.2.31, HMMer v3.2.1, AUGUSTUS v3.3.2, and EMBOSS v6.6.0) against Laurasiatheria_ob9 dataset (n = 6253).

Following the second round of read mapping and depth filtering, no other scaffolds were identified for removal. The remaining 159 of the 201 Pilon-polished scaffolds were further classified on the basis of read depth profiles, and 51 scaffolds with <20% diploid coverage and ≥50% high coverage were marked as likely collapsed repeats. A single scaffold with "Diploid" depth as the dominant PurgeHaplotigs coverage bin and >50% match to another scaffold was marked as a possible repeat sequence (*13*).

Additional *k*-mer analysis of the final assembly was performed using KAT v2.4.2. KAT comp was used to compare *k*-mer frequencies from the 10X reads (16-bp barcode trimmed from read 1) with their copy number in the assembly. This comparison revealed no sign of missing data or large duplications, including retention of haplotigs (fig. S3).

The genome was annotated using the homology-based gene prediction program GeMoMa (version 1.6.2beta) (*70*) and nine reference organisms: *Canis lupus familiaris*, *Vulpes vulpes*, *Felis catus*, *Sus scrof*, *Bos taurus*, *Ailuropoda melanoleuca*, *Ursus maritimus*, *Mus musculus*, and *Homo sapiens*. The assembled contigs were then aligned to CanFam3.1 for chromosome assignments (*13*).

Last, Diploidocus v0.9.6 "vecpurge" mode (implementing BLAST+/2.9.0 tblastn) was used to screen the assembly for contaminants from the NCBI UniVec database (downloaded 05/08/2019) and the PacBio control sequence (MG551957.1). No additional scaffolds were masked, trimmed, or purged. The estimated base error rate of the assembly is 0.00014 (table S2) and similar to other long-read genome assemblies (*71*).

Mitochondrial genome assembly

The mitochondrion for Sandy was filtered out of the assembly at the initial haplotig purging step due to a high read depth. This 68.8-kb contig (tig00007654) was used as the basis for the mitochondrial chromosome. GABLAM v2.30.5 (implementing BLAST+ v2.9.0 blastn) mapped the 16,727-bp CanFam 3.1 mitochondrion onto tig00007654. The sequence was circularized by extracting the best complete match (positions 15,268 to 31,991) as the basis for the mitochondrial genome. Final Pilon polishing was performed by adding the mitochondrial DNA to the main Sandy assembly and mapping 10X Genomics linked reads using Long Ranger v2.2.2 before running Pilon v1.23 with the same settings as the main assembly. The 16,726-bp polished mitochondrial genome was then extracted and added back to the main nuclear genome assembly.

Genetic variation

Several approaches were used to detect large-scale and smaller variations (fig. S4). SVs from both Oxford Nanopore and PacBio sequence data were called relative to other assemblies using a combination of minimap2 v2.17-r943-dirty, SAMTools v1.9, and sniffles v1.0.11 (fig. S1). A conservative list of SVs detected by both Nanopore and PacBio was taken forward for annotation and analysis. Small-scale variation generally smaller than 50 bases was detected in the dingo assembly and five domestic dog assemblies using pairwise MUMmer4 (*72*) (v4.0.0 beta 2) alignment databases.

The 60 ENSEMBL gene IDs were used to identify gene ontology using PANTHER classification system (www.pantherdb.org/). PANTHER was searched for biological processes and overrepresented pathways. Statistical overrepresentation test for pathways was performed using Fisher test on Reactome pathway database. The highest number of gene hits was further investigated for functional annotations (fig. S5).

Phylogenetic and ordination analyses

All possible pairwise alignments were generated using MUMmer4 (*72*) (v4.0.0 beta 2), and indel/SNV numbers were calculated using MUMmer4 "show-snp" script.

Indels and SNVs were analyzed separately due to the different evolutionary processes that produce differences. Distance matrices were generated from the intercanid differences in indels and SNVs and then transformed to WA distance. Glazko *et al.* (*73*) show the derivation and that WA has better phylogenetic properties against normalization of genome sizes.

Phylogenetic analyses using maximum parsimony were generated from the R package "phangorn" version 2.8.1, 15 December 2021 (https://github.com/KlausVigo/phangorn), described in (*74*). The analyses were run as unrooted networks to test the hypothesis that the wolf was the outgroup. To test the stability of the nodes, a Bayesian bootstrap was applied to the original distance matrix using the program bayesian_bootstrap on GitHub and the phylogenetic analysis was recalculated. This process was iterated 500,000 times on the Wesleyan computing cluster. The consensus phylogenetic trees were rooted on the branch leading to wolf (fig. S6, A and C), and the values indicate the percentage of times that a node occurred. The y axis and branch lengths were rescaled to the original number of differences in indels and SNVs among the taxa. The retention index that measures the fit of the network to the distance matrix exceeded 0.96 for all 500,000 trees of indels and SNVs.

NMDS was calculated from the distance matrices and scores for the taxa calculated from the largest two axes. These axes describe 75% of the variance in indel and 73% of the variance in SNVs (fig. S6, B and D). Minimum spanning trees were calculated among the scores in NMDS space. NMDS and minimum spanning trees were calculated in Past 4.04 (*75*).

DNA methylome

We profiled DNA methylation of the dingo and GSD genomes using MethylC-seq (Fig. 4A and fig. S6A) (*76*). DNA methylation data of GSD blood were downloaded from GSE136348. Dingo's blood DNA methylation library was sequenced on the Illumina HiSeq X

platform (150 bp, PE), generating 281 million read pairs and yielding 14.5x sequencing coverage. Sequenced reads were trimmed using Trimmomatic and mapped to the ASM325472v1 genome reference using WALT with the following parameters: -m 10 -t 24 -N 10000000 -L 2000. The mappability of the MethylC-seq library was 85.36%. Duplicate reads were removed using Picard Tools v2.3.0. Genotype and methylation bias correction were performed using MethylDackel with additional parameters: --minOppositeDepth 5 --maxVariantFrac 0.5 --OT 20,148,20,120 --OB 25,145,25,145. The numbers of methylated and unmethylated calls at each CpG site were determined using MethylDackel (https://github.com/dpryan79/MethylDackel). Bisulfite conversion efficiency was 99.7%, estimated using unmethylated lambda phage spike-in control.

Segmentation of dingo and GSD blood DNA methylomes into CpG-rich unmethylated regions (UMRs) was performed using MethylSeekR (*76*) [segmentUMRsLMRs (m = meth, meth.cutoff = 0.5, nCpG.cutoff = 5, PMDs = NA, nCpG.smoothing = 3, minCover = 5)]. To compare DNA methylation levels between proximal gene regulatory regions, we lifted over dingo TSS-associated UMRs to the GSD genome and GSD UMRs to the dingo genome. Next, we calculated average CpG methylation at UMRs and their corresponding lifted-over regions. UMRs showing more than 30% CpG methylation difference between dingo and GSD were selected for the subsequent analysis. The TSS-associated UMRs correspond to transcriptionally permissive gene promoters in each genome (Fig. 4A and fig. S6A).

To validate the difference in expression in *GAL3ST1* and *MAB21L1*, we performed quantitative reverse transcription polymerase chain reaction (RT-qPCR) on six dingoes and six GSDs. RNA was extracted from blood using TRI Reagent protocol, and extracted total RNA was treated with DNase I Amplification Grade (Sigma-Aldrich). Complementary DNA (cDNA) was prepared from an RNA template in a 20-μl reaction mixture using a ProtoScript cDNA synthesis kit (New England Biolabs, MA, USA). The comparative cycle threshold (Ct) method was used to analyze the RT-qPCR results. The expression of *GAL3ST1* was quantified using the following primer: GAL_F2 forward 5′-CTTGGCCCCGTTGTCCTCG-3′ and GAL_F2 reverse 5′- TGACCGCAGAGGCAGCCT-3′ (Fig. 4B). The expression of *MAB21L1* was quantified using the following primers: MAB_F1 forward 5′-AGTGCATCTGGGCTCTTAGAC-3′ and MAB_R1 reverse 5′-AACAAAAGTTGCGCTGAGACC-3′ (fig. S6B).

The RT-qPCR program included an initial step of 95°C for 10 min, followed by 40 cycles of 95°C for 10 s and 60°C for 45 s. To confirm that a single product was produced, amplification followed a melting curve from 60° to 95°C, rising by steps of 1°C. The gene expression was normalized using two housekeeping genes *HNRNPH* (forward 5′-CTCACTATGATCCACCACG-3′ and reverse 5′-TAGCCTCCATAACCTCCAC-3′) and *GAPDH* (forward 5′-TGTCCCCACCCCCAATGTATC-3′ and reverse 5′- CTCCGATGCCTGCTTCACTACCTT-3′). The unpaired *t* test was performed to detect significance (fig. S6B).

Sample collection for dietary study

Biochemical studies were performed on 17 dingoes from two different dingo sanctuaries and 15 GSDs from two different kennels in December 2018. Of these, 14 dingoes were from Dingo Sanctuary Bargo (seven males and seven females), and three were from Pure Dingo Sanctuary (one male and two females) in southeast New South Wales (NSW), Australia. Dingo Sanctuary Bargo feeds kibble and chicken, while Pure Dingo feeds kangaroo meat. All dingoes were pure as determined by microsatellite testing (*62*). The age was 3.8 ± 0.44 (SE) years. Five dingoes were born in the wild but humanized before 6 weeks of age. The remaining 12 dingoes were sanctuary-born. All had daily interactions with humans. No consistent differences were observed between these groups that would suggest that results were influenced by whether dingoes were wild versus sanctuary born. The dingoes were housed in mated pairs and were not kept as pets. Volunteers fed and walked the dingoes daily. The dingoes were socialized but rarely traveled from the sanctuary.

Fifteen GSDs included in the study were tested in December 2018. Eleven were from Kingvale Kennels (three males and eight females), and four were Allendell Kennels (one male and three females). Both Kennels in southeast NSW Australia fed kibble and chicken. The mean age was 3.66 ± 0.44 years. Two female GSDs were subsequently excluded from the study as they came into estrus within 10 days of the study. All GSDs were registered with the Australian Kennel Club and showed no evidence of genetic disease. The GSDs were kept in large runs with fewer males than females in each kennel. All GSDs were socialized and used to traveling distances in cars and trailers.

Experimental diets and treatments

To avoid a possible bias in the results due to diet differences between kennels and sanctuaries, we standardized the diet of dingoes and GSDs for 15 days. Canids were fed throughout the evening on standard "Blackhawk" commercially available dog kibble for the first 10 days. From day 11, canids were transitioned to rice and Blackhawk, so they were fed rice only on day 14 (25% rice + 75% Blackhawk on day 11, 50% rice + 50% Blackhawk on day 12, 75% rice + 25% Blackhawk on day 13, and 100% rice on day 14). Fresh untreated rainwater was transported to all sanctuaries and kennels.

On the evening of day 1, all canids were treated for fleas, ticks, and worms with Advocate for dogs (Bayer) and given the antibiotic Neo-Sulcin (Jurox Animal Health) in kilogram per dependent doses. On days 2 and 3, they were then given the probiotic Protexin (Protexin Veterinary) to recolonize the gut microbiota in kilogram per dependent doses. Warm scat was collected, and blood was drawn on days 1 and 15. Scat was stored at −80°C and blood at 4°C.

Amylase copy number

We used droplet digital PCR (ddPCR) to quantify the amylase copy number. ddPCR was performed using a QX100 ddPCR system (Bio-Rad). Each reaction was performed in a 20-μl reaction volume containing 10 μl of 2× ddPCR Supermix (Bio-Rad), 1 μl of each 20× primer/probe, 1 μl of Dra I restriction enzyme (New England Biolabs #R0129S), 5 μl of DNA template (4 ng/μl), and 2 μl of ddH_2O. Copy number data were rounded to the nearest whole number and presented as copies per individual chromosome. Primer sequence for *Amy2B*: forward 5′-CCAAACCTGGACGGACATCT-3′ and reverse 5′-TATCGTTCGCATTCAAGAGCAA-3′ with FAM probe: 6FAM-TTTGAGTGGCGCTGGG-MGBNFQ. Primer sequence for *C7orf28b-3*: forward 5′-GGGAAACTCCACAAGCAATCA-3′ and reverse 5′-GAGCCCATGGAGGAAATCATC-3′ with HEX probe HEX-CACCTGCTAAACAGC-MGBNFQ. Statistical significance in amylase copy number difference and biochemical studies was analyzed using simple *t* tests using GraphPad Prism software version 8.0 (www.graphpad.com) (fig. S7A).

Serum metabolites

We tested for amylase, cholesterol, triglycerides, and lipase differences associated with starch digestion and fat metabolism (*21*). Amylase, cholesterol, triglycerides, and lipase were assayed using the Thermo Scientific Konelab Prime 30i at the Veterinary Pathology Diagnostic Services Laboratory, University of Sydney. Statistical significance was determined as described above.

Serum lipoprotein profile analysis

Serum was fractionated on an AKTA FPLC system (GE Healthcare Life Sciences) using two Superdex 200 columns (GE Healthcare Life Sciences) connected in series. Plasma (200 μl) was loaded onto the columns, which had been pre-equilibrated with phosphate-buffered saline [10 mM NaH_2PO_4, 137 mM NaCl, and 2.7 mM KCl (pH 7.4)]. Lipoproteins were separated at a flow rate of 0.25 ml/min. Fractions were collected at 1-min intervals and immediately analyzed on AU480 Auto-Analyzer (Beckman Coulter) for total cholesterol levels using the Wako Cholesterol E reagent (Wako Diagnostics).

Bile acid quantification analysis

To examine whether differences in cholesterol levels influence bile acid production in dingoes and GSDs, we quantified free bile acids in canine plasma using a liquid chromatography–tandem mass spectrometry (LC-MS/MS) assay (*77*). We measured the concentration of the primary bile acids cholic acid and chenodeoxycholic acid. We also measured the secondary bile acids ursodeoxycholic acid, deoxycholic acid, and lithocholic acid. Standards for all bile acids were prepared at 0, 10, 20, 40, 60, 100, and 200 nM concentrations from a 1 μM combined stock solution. Deuterium-labeled standards, d_4CA, d_4DCA, d_4CDCA, and d_4LCA, were combined to a final concentration of 4 μM and used as internal standards (ISs) to correct for variability and losses during processing, and a 10-μl aliquot of IS was added to each calibrator (final volume, 200 μl). Each canine plasma sample (30 to 100 μl) was mixed with 10 μl of combined deuterated IS and four volumes of acetonitrile. The mixture was vortexed and centrifuged at 10,000 rpm for 10 min to remove proteins. The supernatant was transferred to a clean tube and vacuum-dried before reconstitution in a 50:50 solution of methanol and water (200 μl). The sample was filtered into reduced volume vials and ready for LC-MS/MS analysis.

The ultraperformance LC-MS detector assembly consisted of an Accela AS injector, Accela UPLC pump, and a TSQ Vantage benchtop mass spectrometer (Thermo Fisher Scientific, Waltham, MA) fitted with a heated electrospray probe. Solutions of the five bile acids, including labeled analogs (200 μM in 50% methanol), were infused at 10 μl/min using a syringe pump into the detector. Collision-induced dissociation experiments in negative ion mode were carried out to determine the parameters at which optimum sensitivity was achieved for these metabolites. The selected reaction monitoring (SRM) transitions were then set in the MS detector parameters. Mass spectra were accumulated during 0.2 s per SRM. Capillary voltage, capillary temperature, and collision gas pressure (Argon) were set to 3000 V, 300°C, and 1.0 torr, respectively. Sheath and auxiliary gas valves (nitrogen) were set at 20 and 10 arbitrary units.

Standards and samples (20 μl) were injected into a Waters Acquity BEH18 column (100 mm by 2.1 mm by 1.7 μm) heated at 40°C. The binary solvent gradient consisted of 5 mM ammonium formate (mobile phase A) and acetonitrile (mobile phase B) at a constant flow rate of 200 μl/min. Initial solvent composition at injection was 40% B, followed by a 5-min gradient to 50% B and a fast gradient ramp to 80% B (1 min), and B was increased again to 95% (2 min), held for 4 min, and then reverted to initial conditions (0.1 min) for equilibration, with a total run time of 18 min. The column flow was directed into the compact mass spectrometer (CMS) detector. Retention times and mass transitions are shown in table S6. The differences in retention times were observed between plasma samples, likely due to sample matrix components.

Calibration curves for each bile acid were plotted using the peak area ratios of the bile acid divided by the peak area of its corresponding deuterated counterpart (*y* axis) versus nanomolar standard concentration (*x* axis). All spectra were processed, and peak areas were integrated using Xcalibur software (version 2.2, 2011, Thermo Fisher Scientific, Waltham, MA). Automated data processing was performed using the LCQuan feature of the software. The concentrations of the endogenous metabolites in the sample extracts were obtained from these calibration curves and calculated using dilution factors. *T* tests were performed on individual bile acids using GraphPad Prism software program version 8.0 (www.graphpad.com) with two outliers removed for dingo and two for GSD using the ROUT method of Prism and a false discovery rate (FDR) (*Q*) of 1%.

Microbiome analysis

Scat from the same set of dingoes and GSDs were sampled on days 1 and 15. The samples were placed into sterile tubes and placed into liquid nitrogen. In the laboratory, they were transferred to −80°C until assayed.

According to the manufacturer's instruction, DNA was extracted from thawed stool samples (0.3 g) using the Qiagen PowerSoil kit (catalog no. 1288-100; Hilden, Germany). However, instead of vortexing, samples were subjected to physical lysis in a bead-beater (TissueLyser II, Qiagen) for 3 min at 30 Hz. DNA was eluted in molecular-grade water and stored at −80°C. The V3-V4 region of the 16*S* rRNA gene was amplified and sequenced. Library preparation and pair-end sequencing were performed (2 × 300 cycles) on the Illumina MiSeq platform at the Ramaciotti Centre for Genomics (University of New South Wales, Sydney, Australia).

16*S* rRNA gene sequence data were quality-filtered and trimmed using trimmomatic version 0.36 truncating reads if the quality was below 12 in a sliding window of 4 bp (Fig. 6 and fig. S8). USEARCH version 10.0.240 was used to merge and quality filter the sequencing reads between 350 and 500 nucleotides (Fig. 6 and fig. S8). Sequences that appeared less than eight times were removed. Processed reads were then concatenated into a single file and dereplicated to form unique sequences. Unique sequences were clustered into zero-radius operational taxonomic units (zOTUs) using the UNOISE3 algorithm implemented in USEARCH (Fig. 6 and fig. S8) (*78*). Chimeras were removed in reference mode using UCHIME and the SILVA SSURef NR99 database version 132.

The zOTU sequences were taxonomically classified using BLASTn alignments against the SILVA database. zOTUs without any taxonomic assignment were removed from the dataset. No zOTUs were found to be assigned to the chloroplast in the dataset. The number of final zOTUs was 9951. Data were visualized using the ggpubr package. For alpha diversity measures, each sample was subsampled 100 times to a count of 164,700 counts per sample, and the average was taken. zOTU richness and Shannon diversity index were calculated in R (version 3.6.0) using the vegan package and statistically compared dingoes and GSDs using the Wilcoxon rank sum test (Fig. 6 and fig. S8).

For beta diversity, the rarefied data were square root–transformed. Bray-Curtis distances were calculated and visualized on an NMDS plot. The zOTU sequences were aligned using MAFFT (*79*), and a phylogenetic tree was calculated using FastTree (*80*) to calculate weighted UniFrac distances, which were visualized on a principal coordinate analysis plot. Differences in the beta diversity of dingoes and GSD communities were analyzed using a pairwise adonis test (https://github.com/bwemheu/pairwise.adonis).

The bacterial composition of the dingo and GSD was visualized with bar plots for each location of sampling, using phyloseq R package version 1.32.0. We visualized the relative abundance of the top 10 most abundant OTUs. We analyzed the differential abundance of the OTUs using metacoder R package version 0.3.4. Before differential abundance analysis, taxa were taxonomy concatenated according to species, using phyloseq. The $\log_2$ ratio of median proportions of relative abundance between the dingo and GSD and significance was determined using a Wilcoxon rank sum test followed by a Benjamini-Hochberg (FDR) correction for multiple comparisons. The $\log_2$ ratio of median proportions of the insignificant different abundant taxon was removed for clarity.

The metabolic potential of the microbial community was evaluated using the predictive metagenomic analysis tool, Tax4Fun2 (fig. S8). Relative abundances of taxa or predicted functions were examined using the phyloseq package. Significant differences of microbial taxa between the canids were analyzed at phylum, family, and genus levels using the Analysis of the Composition of the Microbiome (ANCOM; v2.0) (table S8). ANCOM evaluates the statistical significance of the taxa or predicted functions using log ratio–transformed data. Benjamini-Hochberg correction was applied to correct for multiple comparison testing, and $P < 0.05$ was considered statistically significant.

Statistical analysis

Statistical analyses are described throughout in the relevant sections and always followed recommended practices and cutoffs within the individual algorithms.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.abm5944

REFERENCES AND NOTES

1. J. W. O. Ballard, C. Gardner, L. Ellem, S. Yadav, R. I. Kemp, Eye-contact and sociability data suggests that Australian dingoes were never domesticated. *Curr. Zool.* **2021**, zoab024 (2021).
2. J. W. O. Ballard, L. A. B. Wilson, The Australian dingo: Untamed or feral? *Front. Zool.* **16**, 2 (2019).
3. S.-j. Zhang, G.-D. Wang, P. Ma, L.-l. Zhang, T.-T. Yin, Y.-h. Liu, N. O. Otecko, M. Wang, Y.-p. Ma, L. Wang, B. Mao, P. Savolainen, Y.-p. Zhang, Genomic regions under selection in the feralization of the dingoes. *Nat. Commun.* **11**, 671 (2020).
4. T. S. Doherty, N. E. Davis, C. R. Dickman, D. M. Forsyth, M. Letnic, D. G. Nimmo, R. Palmer, E. G. Ritchie, J. Benshemesh, G. Edwards, J. Lawrence, L. Lumsden, C. Pascoe, A. Sharp, D. Stokeld, C. Myers, G. Story, P. Story, B. Triggs, M. Venosta, M. Wysong, T. M. Newsome, Continental patterns in the diet of a top predator: Australia's dingo. *Mamm. Rev.* **49**, 31–44 (2019).
5. L. Corbett, *The Dingo in Australia and Asia* (J.B. Books Australia, 2001), 200 p.
6. D. Stephens, A. N. Wilton, P. J. S. Fleming, O. Berry, Death by sex in an Australian icon: A continent-wide survey reveals extensive hybridization between dingoes and domestic dogs. *Mol. Ecol.* **24**, 5643–5656 (2015).
7. A. H. Freedman, I. Gronau, R. M. Schweizer, D. O.-D. Vecchyo, E. Han, P. M. Silva, M. Galaverni, Z. Fan, P. Marx, B. Lorente-Galdos, H. Beale, O. Ramirez, F. Hormozdiari, C. Alkan, C. Vilà, K. Squire, E. Geffen, J. Kusak, A. R. Boyko, H. G. Parker, C. Lee, V. Tadigotla, A. Wilton, A. Siepel, C. D. Bustamante, T. T. Harkins, S. F. Nelson, E. A. Ostrander, T. Marques-Bonet, R. K. Wayne, J. Novembre, Genome sequencing highlights the dynamic early history of dogs. *PLOS Genet.* **10**, e1004016 (2014).
8. S. Surbakti, H. G. Parker, J. K. Mc Intyre, H. K. Maury, K. M. Cairns, M. Selvig, M. P. Adam, A. Safonpo, L. Numberi, D. Y. P. Runtuboi, B. W. Davis, E. A. Ostrander, New Guinea highland wild dogs are the original New Guinea singing dogs. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 24369–24376 (2020).
9. S. M. Jackson, P. J. S. Fleming, M. D. B. Eldridge, M. Archer, S. Ingleby, R. N. Johnson, K. M. Helgen, Taxonomy of the dingo: It's an ancient dog. *Aust. Zool.* **41**, 347–357 (2021).
10. V. Jagannathan, C. Drögemüller, T. Leeb; Dog Biomedical Variant Database Consortium (DBVDC), A comprehensive biomedical variant catalogue based on whole genome sequences of 582 dogs and eight wolves. *Anim. Genet.* **50**, 695–704 (2019).
11. D. Liu, M. Hunt, I. J. Tsai, Inferring synteny between genome assemblies: A systematic evaluation. *BMC Bioinf.* **19**, 26 (2018).
12. K. L. Toh, C. M. Wade, T. S. Mikkelsen, E. K. Karlsson, D. B. Jaffe, M. Kamal, M. Clamp, J. L. Chang, E. J. Kulbokas III, M. C. Zody, E. Mauceli, X. Xie, M. Breen, R. K. Wayne, E. A. Ostrander, C. P. Ponting, F. Galibert, D. R. Smith, P. J. deJong, E. Kirkness, P. Alvarez, T. Biagi, W. Brockman, J. Butler, C.-W. Chin, A. Cook, J. Cuff, M. J. Daly, D. D. Caprio, S. Gnerre, M. Grabherr, M. Kellis, M. Kleber, C. Bardeleben, L. Goodstadt, A. Heger, C. Hitte, L. Kim, K.-P. Koepfli, H. G. Parker, J. P. Pollinger, S. M. J. Searle, N. B. Sutter, R. Thomas, C. Webber; Broad Sequencing Platform members, E. S. Lander, Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803–819 (2005).
13. M. A. Field, B. D. Rosen, O. Dudchenko, E. K. F. Chan, A. E. Minoche, R. J. Edwards, K. Barton, R. J. Lyons, D. E. Tuipulotu, V. M. Hayes, A. D. Omer, Z. Colaric, J. Keilwagen, K. Skvortsova, O. Bogdanovic, M. A. Smith, E. L. Aiden, T. P. L. Smith, R. A. Zammit, J. W. O. Ballard, CanFam_GSD: *De novo* chromosome-length genome assembly of the German shepherd dog (*Canis lupus familiaris*) using a combination of long reads, optical mapping and Hi-C. *GiGaScience* **9**, giaa027 (2020).
14. R. J. Edwards, M. A. Field, J. M. Ferguson, O. Dudchenko, J. Keilwagen, B. D. Rosen, G. S. Johnson, E. S. Rice, L. D. Hillier, J. M. Hammond, S. G. Towarnicki, A. Omer, R. Khan, K. Skvortsova, O. Bogdanovic, R. A. Zammit, E. L. Aiden, W. C. Warren, J. W. O. Ballard, Chromosome-length genome assembly and structural variations of the primal Basenji dog (*Canis lupus familiaris*) genome. *BMC Genomics* **22**, 188 (2021).
15. J. V. Halo, A. L. Pendleton, F. Shen, A. J. Doucet, T. Derrien, C. Hitte, L. E. Kirby, B. Myers, E. Sliwerska, S. Emery, J. V. Moran, A. R. Boyko, J. M. Kidd, Long-read assembly of a Great Dane genome highlights the contribution of GC-rich sequence and mobile elements to canine genomes. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2016274118 (2021).
16. R. A. Player, E. R. Forsyth, K. J. Verratti, D. W. Mohr, A. F. Scott, C. E. Bradburne, A novel *Canis lupus familiaris* reference genome improves variant resolution for use in breed-specific GWAS. *Life Sci. Alliance* **4**, e202000902 (2021).
17. H. G. Parker, D. L. Dreger, M. Rimbault, B. W. Davis, A. B. Mullen, G. Carpintero-Ramirez, E. A. Ostrander, Genomic analyses reveal the influence of geographic origin, migration, and hybridization on modern dog breed development. *Cell Rep.* **19**, 697–708 (2017).
18. M.-H. S. Sinding, S. Gopalakrishnan, K. Raundrup, L. Dalén, J. Threlfall; Darwin Tree of Life Barcoding collective; Wellcome Sanger Institute Tree of Life programme; Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective; Tree of Life Core Informatics collective; Darwin Tree of Life Consortium, T. Gilbert, The genome sequence of the grey wolf, *Canis lupus* Linnaeus 1758. *Wellcome Open Res.* **6**, 310 (2021).
19. M.-H. S. Sinding, S. Gopalakrishan, F. G. Vieira, J. A. S. Castruita, K. Raundrup, M. P. H. Jørgensen, M. Meldgaard, B. Petersen, T. Sicheritz-Ponten, J. B. Mikkelsen, U. M. Petersen, R. Dietz, C. Sonne, L. Dalén, L. Bachmann, Ø. Wiig, A. J. Hansen, M. T. P. Gilbert, Population genomics of grey wolves and wolf-like canids in North America. *PLOS Genet.* **14**, e1007745 (2018).
20. A.-S. Sundman, F. Pértille, L. L. Coutinho, E. Jazin, C. G. Bosagna, P. Jensen, DNA methylation in canine brains is related to domestication and dog-breed formation. *PLOS ONE* **15**, e0240787 (2020).
21. M. Arendt, K. M. Cairns, J. W. O. Ballard, P. Savolainen, E. Axelsson, Diet adaptation in dog reflects spread of prehistoric agriculture. *Heredity* **117**, 301–306 (2016).
22. E. Axelsson, A. Ratnakumar, M.-L. Arendt, K. Maqbool, M. T. Webster, M. Perloski, O. Liberg, J. M. Arnemo, Å. Hedhammar, K. L. Toh, The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* **495**, 360–364 (2013).
23. R. L. Gonçalves, Z. Carvalho-Santos, A. P. Francisco, G. T. Fioreze, M. Anjos, C. Baltazar, A. P. Elias, P. M. Itskov, M. D. W. Piper, C. Ribeiro, Commensal bacteria and essential amino acids control food choice behavior and reproduction. *PLOS Biol.* **15**, e2000862 (2017).
24. T. McKnight, *Feral Livestock in Ango-America* (University of California Press, 1964).
25. S. Westfall, Posts tagged "German Shepherds banned in Australia"; https://retrieverman.net/tag/german-shepherds-banned-in-australia/2019 [Posted in German Shepherd Dog, tagged German Shepherd Dog, German Shepherds banned in Australia on November 17, 2019].
26. S. Yadav, R. Pickford, R. A. Zammit, J. W. O. Ballard, Metabolomics shows the Australian dingo has a unique plasma profile. *Sci. Rep.* **11**, 5245 (2021).

27. E. Kubinyi, M. Bence, D. Koller, M. Wan, E. Pergel, Z. Ronai, M. S. Szekely, Á. Miklósi, Oxytocin and opioid receptor gene polymorphisms associated with greeting behavior in dogs. *Front. Psychol.* **8**, 1520 (2017).
28. A. Sommese, K. Nováková, N. F. Šebková, L. Bartoš, A wolfdog point of view on the impossible task paradigm. *Anim. Cogn.* **22**, 1073–1083 (2019).
29. W. Nie, J. Wang, W. Su, D. Wang, A. Tanomtong, P. L. Perelman, A. S. Graphodatsky, F. Yang, Chromosomal rearrangements and karyotype evolution in carnivores revealed by chromosome painting. *Heredity* **108**, 17–27 (2012).
30. M. Vozdova, S. Kubickova, H. Cernohorska, J. Fröhlich, R. Vodicka, J. Rubes, Comparative study of the Bush Dog (*Speothos venaticus*) karyotype and analysis of satellite DNA sequences and their chromosome distribution in six species of Canidae. *Cytogenet. Genome Res.* **159**, 88–96 (2019).
31. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
32. D. Mapleson, G. G. Accinelli, G. Kettleborough, J. Wright, B. J. Clavijo, KAT: A K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics* **33**, 574–576 (2017).
33. S. Kim, S. Mun, T. Kim, K.-H. Lee, K. Kang, J.-Y. Cho, K. Han, Transposable element-mediated structural variation analysis in dog breeds using whole-genome sequencing. *Mamm. Genome* **30**, 289–300 (2019).
34. A. J. Waardenberg, M. A. Field, consensusDE: An R package for assessing consensus of multiple RNA-seq algorithms with RUV correction. *PeerJ* **7**, e8206 (2019).
35. H. Hoffmann, J. Gill, R. Oiekarz, Studies on the physiology of digestion in wolf (*Canis lupus* L.) and dingo (*Canis dingo* L.) and jackal (*Canis aureus* L.). *Acta Physiol. Pol.* **XV**, 105–115 (1964).
36. J. Gill, H. Hoffmannowa, R. Piekarz, Studies on the digestive physiology of the wolf (*Canis lupus* L.), dingo (*Canis dingo* L.) and jackal (*Canis aureus* L.). II. Digestive capacity of the pancreas, duodenum and salivary glands; size of the digestive system; weight of internal organs. *Acta Physiol. Pol.* **15**, 137–148 (1964).
37. A. E. Newsome, L. K. Corbett, S. M. Carpenter, The identity of the dingo I. Morphological discriminants of dingo and dog skulls. *Aust. J. Zool.* **28**, 615–625 (1980).
38. D. Schübeler, Function and information content of DNA methylation. *Nature* **517**, 321–326 (2015).
39. A. Seko, S. H. Kuge, K. Yamashita, Molecular cloning and characterization of a novel human galactose 3-O-sulfotransferase that transfers sulfate to Galβ1→ 3GalNAc residue in O-glycans. *J. Biol. Chem.* **276**, 25697–25704 (2001).
40. K. Honke, M. Tsuda, Y. Hirahara, A. Ishii, A. Makita, Y. Wada, Molecular cloning and expression of cDNA encoding human 3′-phosphoadenylylsulfate: Galactosylceramide 3′-sulfotransferase. *J. Biol. Chem.* **272**, 4864–4868 (1997).
41. M. Arendt, T. Fall, K. L. 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).
42. J. M. Ridlon, S. C. Harris, S. Bhowmik, D.-J. Kang, P. B. Hylemon, Consequences of bile salt biotransformations by intestinal bacteria. *Gut Microbes* **7**, 22–39 (2016).
43. J. B. J. Ward, N. K. Lajczak, O. B. Kelly, A. M. O'Dwyer, A. K. Giddam, J. N. Gabhann, P. Franco, M. M. Tambuwala, C. A. Jefferies, S. Keely, A. Roda, S. J. Keely, Ursodeoxycholic acid and lithocholic acid exert anti-inflammatory actions in the colon. *Am. J. Physiol. Gastrointest. Liver Physiol.* **312**, G550–G558 (2017).
44. E. Y. Purwani, T. Purwadaria, M. T. Suhartono, Fermentation RS3 derived from sago and rice starch with *Clostridium butyricum* BCC B2571 or *Eubacterium rectale* DSM 17629. *Anaerobe* **18**, 55–61 (2012).
45. N. Molinero, L. Ruiz, B. Sánchez, A. Margolles, S. Delgado, Intestinal bacteria interplay with bile and cholesterol metabolism: Implications on host physiology. *Front. Physiol.* **10**, 185 (2019).
46. A. C. Poole, J. K. Goodrich, N. D. Youngblut, G. G. Luque, A. Ruaud, J. L. Sutter, J. L. Waters, Q. Shi, M. El-Hadidi, L. M. Johnson, H. Y. Bar, D. H. Huson, J. G. Booth, R. E. Ley, Human salivary amylase gene copy number impacts oral and gut microbiomes. *Cell Host Microbe* **25**, 553–564.e7 (2019).
47. E. A. Choi, H. C. Chang, Cholesterol-lowering effects of a putative probiotic strain *Lactobacillus plantarum* EM isolated from kimchi. *LWT* **62**, 210–217 (2015).
48. B. J. Bizjan, T. Katsila, T. Tesovnik, R. Šket, M. Debeljak, M. T. Matsoukas, J. Kovač, Challenges in identifying large germline structural variants for clinical use by long read sequencing. *Comput. Struct. Biotechnol. J.* **18**, 83–92 (2020).
49. K. Honke, Biological functions of sulfoglycolipids and the EMARS method for identification of co-clustered molecules in the membrane microdomains. *J. Biochem.* **163**, 253–263 (2018).
50. A. Husby, On the use of blood samples for measuring DNA methylation in ecological epigenetic studies. *Integr. Comp. Biol.* **60**, 1558–1566 (2020).
51. U. Ravnskov, High cholesterol may protect against infections and atherosclerosis. *QJM* **96**, 927–934 (2003).
52. M. Yoshikawa, T. Tsujii, K. Matsumura, J. Yamao, Y. Matsumura, R. Kubo, H. Fukui, S. Ishizaka, Immunomodulatory effects of ursodeoxycholic acid on immune responses. *Hepatology* **16**, 358–364 (1992).
53. T. W. H. Pols, T. Puchner, H. I. Korkmaz, M. Vos, M. R. Soeters, C. J. M. de Vries, Lithocholic acid controls adaptive immune responses by inhibition of Th1 activation through the vitamin D receptor. *PLOS ONE* **12**, e0176715 (2017).
54. Y. Feng, K. Siu, N. Wang, K.-M. Ng, S.-W. Tsao, T. Nagamatsu, Y. Tong, Bear bile: Dilemma of traditional medicinal use and animal protection. *J. Ethnobiol. Ethnomed.* **5**, 2 (2009).
55. E. D. Sonnenburg, J. L. Sonnenburg, Starving our microbial self: The deleterious consequences of a diet deficient in microbiota-accessible carbohydrates. *Cell Metab.* **20**, 779–786 (2014).
56. Y. Liu, B. Liu, C. Liu, Y. Hu, C. Liu, X. Li, X. Li, X. Zhang, D. M. Irwin, Z. Wu, Z. Chen, Q. Jin, S. Zhang, Differences in the gut microbiomes of dogs and wolves: Roles of antibiotics and starch. *BMC Vet. Res.* **17**, 112 (2021).
57. G. Bosch, E. A. Hagen-Plantinga, W. H. Hendriks, Dietary nutrient profiles of wild wolves: Insights for optimal dog nutrition? *Br. J. Nutr.* **113**, S40–S54 (2015).
58. I. Mathieson, I. Lazaridis, N. Rohland, S. Mallick, N. Patterson, S. A. Roodenberg, E. Harney, K. Stewardson, D. Fernandes, M. Novak, K. Sirak, C. Gamba, E. R. Jones, B. Llamas, S. Dryomov, J. Pickrell, J. L. Arsuaga, J. M. B. de Castro, E. Carbonell, F. Gerritsen, A. Khokhlov, P. Kuznetsov, M. Lozano, H. Meller, O. Mochalov, V. Moiseyev, M. A. R. Guerra, J. Roodenberg, J. M. Vergès, J. Krause, A. Cooper, K. W. Alt, D. Brown, D. Anthony, C. Lalueza-Fox, W. Haak, R. Pinhasi, D. Reich, Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
59. A. K. Hewson-Hughes, V. L. Hewson-Hughes, A. Colyer, A. T. Miller, S. J. McGrane, S. R. Hall, R. F. Butterwick, S. J. Simpson, D. Raubenheimer, Geometric analysis of macronutrient selection in breeds of the domestic dog, *Canis lupus familiaris*. *Behav. Ecol.* **24**, 293–304 (2013).
60. M. T. Roberts, E. N. Bermingham, N. J. Cave, W. Young, C. M. Mckenzie, D. G. Thomas, Macronutrient intake of dogs, self-selecting diets varying in composition offered *ad libitum*. *J. Anim. Physiol. Anim. Nutr.* **102**, 568–575 (2017).
61. B. L. Allen, L. R. Allen, G. Ballard, S. M. Jackson, P. J. Fleming, A roadmap to meaningful dingo conservation. *Canid. Biol. Conserv.* **20**, 45–56 (2017).
62. A. N. Wilton, DNA methods of assessing dingo purity, in *A Symposium on the Dingo*, C. R. Dickman, D. Lunney, Eds. (Royal Zoological Society of New South Wales, 2001), pp. 49–56.
63. N. C. Durand, M. S. Shamim, I. Machol, S. S. P. Rao, M. H. Huntley, E. S. Lander, E. L. Aiden, Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
64. O. Dudchenko, S. S. Batra, A. D. Omer, S. K. Nyquist, M. Hoeger, N. C. Durand, M. S. Shamim, I. Machol, E. S. Lander, A. P. Aiden, E. L. Aiden, *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
65. O. Dudchenko, M. S. Shamim, S. S. Batra, N. C. Durand, N. T. Musial, R. Mostofa, M. Pham, B. G. St. Hilaire, W. Yao, E. Stamenova, M. Hoeger, S. K. Nyquist, V. Korchina, K. Pletch, J. P. Flanagan, A. Tomaszewicz, D. M. Aloose, C. P. Estrada, B. J. Novak, A. D. Omer, E. L. Aiden, The Juicebox assembly tools module facilitates *de novo* assembly of mammalian genomes with chromosome-length scaffolds for under $1000. bioRxiv 254797 [Preprint]. 28 January 2018. https://doi.org/10.1101/254797.
66. J. T. Robinson, D. Turner, N. C. Durand, H. Thorvaldsdottir, J. P. Mesirov, E. L. Aiden, Juicebox.js provides a cloud-based visualization system for Hi-C data. *Cell Syst.* **6**, 256–258.e1 (2018).
67. H. Li, Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
68. Institute JG. BBMap. A DOE office of science user facility [Internet]. 1997-2021; https://sourceforge.net/projects/bbmap/.
69. M. J. Roach, S. A. Schmidt, A. R. Borneman, Purge haplotigs: Allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinform.* **19**, 460 (2018).
70. J. Keilwagen, F. Hartung, J. Grau, GeMoMa: Homology-based gene prediction utilizing intron position conservation and RNA-seq data. *Methods Mol. Biol.* **1962**, 161–177 (2019).
71. W. Wang, A. Das, D. Kainer, M. Schalamun, A. Morales-Suarez, B. Schwessinger, R. Lanfear, The draft nuclear genome assembly of *Eucalyptus pauciflora*: A pipeline for comparing de novo assemblies. *Gigascience* **9**, giz160 (2020).
72. G. Marcais, A. L. Delcher, A. M. Phillippy, R. Coston, S. L. Salzberg, A. Zimin, MUMmer4: A fast and versatile genome alignment system. *PLOS Comput. Biol.* **14**, e1005944 (2018).
73. G. Glazko, A. Gordon, A. Mushegian, The choice of optimal distance measure in genome-wide datasets. *Bioinformatics* **21** (Suppl. 3), iii3–iii11 (2005).
74. K. Schliep, A. J. Potts, D. A. Morrison, G. W. Grimm, Intertwining phylogenetic trees and networks. *Meth. Ecol. Evol.* **8**, 1212–1220 (2017).
75. O. Hammer, D. A. Harper, P. D. Ryan, PAST: Paleontological software package for education and data anaylysis. *Palaeontol. Electron.* **9**, 9 (2001).
76. L. Burger, D. Gaidatzis, D. Schübeler, M. B. Stadler, Identification of active regulatory regions from DNA methylation data. *Nucleic Acids Res.* **41**, e155 (2013).

77. T. Sugita, K. Amano, M. Nakano, N. Masubuchi, M. Sugihara, T. Matsuura, Analysis of the serum bile acid composition for differential diagnosis in patients with liver disease. *Gastroenterol. Res. Pract.* **2015**, 717431 (2015).
78. R. C. Edgar, Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
79. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
80. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2—Approximately maximum-likelihood trees for large alignments. *PLOS ONE* **5**, e9490 (2010).
81. S. Koren, B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman, A. M. Phillippy, Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
82. PacificBiosciences, GenomicConsensus: Genome polishing and variant calling. https://github.com/PacificBiosciences/gcpp.
83. A. C. English, S. Richards, Y. Han, M. Wang, V. Vee, J. Qu, X. Qin, D. M. Muzny, J. G. Reid, K. C. Worley, R. A. Gibbs, Mind the gap: Upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLOS ONE* **7**, e47768 (2012).
84. 10X Genomics linked-read alignment,variant calling, phasing, and structural variant calling (2020); https://support.10xgenomics.com/genome-exome/software/pipelines/latest/what-is-long-ranger.
85. B. J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C. A. Cuomo, Q. Zeng, J. Wortman, S. K. Young, A. M. Earl, Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLOS ONE* **9**, e112963 (2014).
86. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin; 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
87. F. J. Sedlazeck, P. Rescheneder, M. Smolka, H. Fang, M. Nattestad, A. von Haeseler, M. C. Schatz, Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods* **15**, 461–468 (2018).
88. M. A. Urich, J. R. Nery, R. Lister, R. J. Schmitz, J. R. Ecker, MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nat. Protoc.* **10**, 475–483 (2015).
89. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
90. H. F. Chen, A. D. Smith, T. Chen, WALT: Fast and accurate read mapping for bisulfite sequencing. *Bioinformatics* **32**, 3507–3509 (2016).
91. G. T. Selvarajah, F. A. S. Bonestroo, E. P. M. T. Sprang, J. Kirpensteijn, J. A. Mol, Reference gene validation for gene expression normalization in canine osteosarcoma: A geNorm algorithm approach. *BMC Vet. Res.* **13**, 354 (2017).
92. M. Ollivier, A. Tresset, F. Bastian, L. Lagoutte, E. Axelsson, M. L. Arendt, A. Bălăşescu, M. Marshour, M. V. Sablin, L. Salanova, J. D. Vigne, C. Hitte, C. Hänni, *Amy2B* copy number variation reveals starch diet adaptations in ancient European dogs. *R. Soc. Open Sci.* **3**, 160449 (2016).
93. A. Klindworth, E. Pruesse, T. Schweer, J. Peplies, C. Quast, M. Horn, F. O. Glöckner, Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* **41**, e1 (2013).
94. C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, F. O. Glöckner, The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).
95. P. J. McMurdie, S. Holmes, phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLOS ONE* **8**, e61217 (2013).
96. Z. S. Foster, T. J. Sharpton, N. J. Grünwald, Metacoder: An R package for visualization and manipulation of community taxonomic diversity data. *PLOS Comput. Biol.* **13**, e1005404 (2017).
97. J. Oksanen, F. Guillaume Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Haran, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, H. Wagner, The vegan package: Community ecology package. R package ver 20-2 (2011).
98. F. Wemheuer, J. A. Taylor, R. Daniel, E. Johnston, P. Meinicke, T. Thomas, B. Wemheuer, Tax4Fun2: A R-based tool for the rapid prediction of habitat-specific functional profiles and functional redundancy based on 16S rRNA gene marker gene sequences. *Environ. Microb.* **15**, 11 (2020).
99. A. Rhie, Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 245 (2020).
100. S. Mandal, W. Van Treuren, R. A. White, M. Eggesbø, R. Knight, S. D. Peddada, Analysis of composition of microbiomes: A novel method for studying microbial composition. *Microb. Ecol. Health Dis.* **26**, 27663 (2015).

Acknowledgments: B. Eggleton and L. Eggleton rescued Sandy from outback Australia. We wish to thank all those who voted in the PacBio 2017 World's Most Interesting Genome Competition and E. Hetas for running the show. D. Kudrna completed PacBio sequencing at Arizona Genomics Institute. v1.0 of the genome assembled by C. Drescher at Computomics. The 10X and PromethION sequencing was completed at the Garvan Institute, Sydney. V. M. Hayes at the Garvan Institute funded the BioNano data generation used in the v1 Sandy genome assembly. Hi-C data and chromosome-length Hi-C scaffolding were generated by the DNA Zoo Consortium (www.dnazoo.org). DNA Zoo is supported by Illumina Inc., IBM, and the Pawsey Supercomputing Center. For the experimental study, dingoes were made available by Dingo Sanctuary Bargo and Pure Dingo, and German shepherds were made available by Kingvale and Allendell Kennels. A. Brown gave insight into the cholesterol results, S. Towarnicki contributed to the biochemical assays, and A. Shaw collected scat. S Ho, M. McGeoch, W. Grant, S. Gopalakrishnan, and K. Ballard provided comments. Mass spectrometry was obtained at the Bioanalytical Mass Spectrometry Facility within the Mark Wainwright Analytical Centre of the University of New South Wales. This work was undertaken using infrastructure provided by NSW Government co-investment in the National Collaborative Research Infrastructure Scheme (NCRIS), and subsidized access to this facility is gratefully acknowledged. **Funding:** This work was funded by Australian Research Council Discovery Project DP150102038 (to J.W.O.B.), Danish National Research Foundation DNRF143 (to M.T.P.G. and J.A.R.), National Health and Medical Research Council APP5121190 (to M.A.F.), NIH CEGS RM1HG011016-01A1 (to E.L.A.), Welch Foundation Q-1866 (to E.L.A.), McNair Medical Institute Scholar Award (to E.L.A.), NIH Encyclopedia of DNA Elements Mapping Center Award UM1HG009375 (to E.L.A.), U.S.-Israel Binational Science Foundation Award 2019276 (to E.L.A.), Behavioral Plasticity Research Institute NSF DBI-2021795 (to E.L.A.), and NSF Physics Frontiers Center Award NSF PHY-2019745 (to E.L.A.). E.L.A. was also supported by NIH 4D Nucleome Grants (U01HL130010 and U01HL156059). **Author contributions:** J.W.O.B. coordinated, designed, and funded the project. M.A.F. performed variation analyses. S.Y. compiled the data, gene ontology, and gene expression analysis. R.A.Z. provided the samples. B.D.R. and T.P.L.S. performed and assisted with the ONT sequencing, genome assembly, and polishing. The DNA Zoo initiative, including O.D., A.O., and Z.C., performed and funded the Hi-C experiment. O.D. and E.L.A. conducted the Hi-C analyses. B.C. constructed the phenograms. K.S. and O.B. funded and conducted the DNA methylation analyses. R.J.E. performed the final polishing, final assembly cleanup, and KAT analysis. J.K. performed the genome annotation. E.K.F.C. and B.D.R. performed the *AMY2B* analyses. M.E., T.T., and J.A.R. performed microbiome analysis. S.B., B.J.C., and B.M. performed biochemical analysis. R.G.M., A.E.M., T.P.L.S., and M.T.P.G. commented on the manuscript. J.W.O.B. and M.A.F. wrote the initial draft. All authors edited and approved the final manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The complete assembled genome generated during the study is available at NCBI (ASM325472v2; GenBank assembly accession no. GCA_003254725.2). DNA methylation data accession numbers are GSE119099 (dingo, this study) and GSE136348 [GSD (*13*)].

Submitted 28 September 2021
Accepted 9 March 2022
Published 22 April 2022
10.1126/sciadv.abm5944

ScienceAdvances

The Australian dingo is an early offshoot of modern breed dogs

Matt A. FieldSonu YadavOlga DudchenkoMeera EsvaranBenjamin D. RosenKsenia SkvortsovaRichard J. EdwardsJens KeilwagenBlake J. CochranBikash ManandharSonia BustamanteJacob Agerbo RasmussenRichard G. MelvinBarry ChernoffArina OmerZane ColaricEva K. F. ChanAndre E. MinocheTimothy P. L. SmithM. Thomas P. GilbertOzren BogdanovicRobert A. ZammitTorsten ThomasErez L. AidenJ. William O. Ballard

Sci. Adv., 8 (16), eabm5944. • DOI: 10.1126/sciadv.abm5944

Science Advances (ISSN) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.