Demographic history, selection and functional diversity of the canine genome

Elaine A. Ostrander[1], Robert K. Wayne[2], Adam H. Freedman[3] and Brian W. Davis[1,4]

Abstract | The domestic dog represents one of the most dramatic long-term evolutionary experiments undertaken by humans. From a large wolf-like progenitor, unparalleled diversity in phenotype and behaviour has developed in dogs, providing a model for understanding the developmental and genomic mechanisms of diversification. We discuss pattern and process in domestication, beginning with general findings about early domestication and problems in documenting selection at the genomic level. Furthermore, we summarize genotype–phenotype studies based first on single nucleotide polymorphism (SNP) genotyping and then with whole-genome data and show how an understanding of evolution informs topics as different as human history, adaptive and deleterious variation, morphological development, ageing, cancer and behaviour.

[1]*Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, 50 South Drive, Building 50, National Institutes of Health, Bethesda, Maryland 20892, USA.*
[2]*Department of Ecology and Evolutionary Biology, 601 Charles E Young Drive South, University of California at Los Angeles, Los Angeles, California 90095, USA.*
[3]*Informatics Division, Department of Organismic and Evolutional Biology, Harvard University, Cambridge, Massachusetts 02138, USA.*
[4]*Current Address: Department of Veterinary Integrative Biosciences, Texas A and M University, College Station, Texas 77843, USA.*

Correspondence to E.A.O.
eostrand@mail.nih.gov

doi:10.1038/nrg.2017.67
Published online 25 Sep 2017

The domestic dog has long been a leading example of extreme phenotypic diversification under domestication and consists of nearly 400 breeds of the species *Canis lupus familiaris* (also known as *Canis familiaris*)[1–4]. Although the origin and timing of dog domestication are controversial[5–11], there is consensus from several studies regarding three general aspects of dog domestication. Specifically, the common ancestor of all dogs is the grey wolf, and no other species has left a genetic record in its genome. Additionally, the dog originated in the Old World, and appears to be sister to Eurasian grey wolves in phylogenetic analysis of whole-genome data[10]. Finally, the dog was first domesticated >15 thousand years ago, before the development of modern agriculture; however, 'modern' breeds are much more recent in origin and were developed over the past few hundred years by selective breeding[5,12–14].

The unique demographic history of the dog, which is characterized by an ancient population contraction common to all dogs, followed by much more recent breed-specific population bottlenecks, has left distinct genetic signatures in dog genomes[5,7,14–20]. Furthermore, post-divergence gene flow between dogs and grey wolves, as well as between ancient and modern dog lineages, is a marked feature of dog history[10,16,17]. Therefore, modern breeds, to varying degrees, carry genomic signatures from early progenitors. Natural and artificial selection have also modified the genomes of dog breeds at specific sites, reflecting the varying intensity and direction of selection[18]. This modification occurred first in the context of early interactions with human hunter-gatherers, and their movements to track and consume large prey, and later through association with agrarian societies and the development of starch-based agricultural industries. Lastly, recent intense selection associated with generating breed-specific phenotypes has altered genomic patterns of variation (see below). Throughout these evolutionary processes, population bottlenecks associated with domestication, local population history and breed creation have also left an imprint on the dog genome, manifesting as genome-wide decreases in genetic variation, increased haplotype homozygosity and linkage disequilibrium (LD)[15,20,21]. By contrast, directed selection is site specific, resulting in local increases in homozygosity and associated hitchhiked variation surrounding the locus that is the target of selection[5,7,18,22,23–28].

These results, based mostly on genome-wide single nucleotide polymorphism (SNP) array studies, demonstrate a new frontier in the use of dogs as an evolutionary model for domestication and are now augmented by sequencing-based approaches that have resulted in the accumulation of a large whole-genome sequence database from a wide diversity of breeds[9,19,20,29] (see the Dog Genome SNP Database (DoGSD) and the Dog 10K Genomes Project). Such new genomic resources provide a rich source of information to directly test predictions regarding history and selection, casting the domestic dog into a unique role as a model species rivalling the human.

In this Review, we first discuss the demographic history of the dog and the genomic changes associated with early domestication to provide a historical context, albeit only briefly, as this topic has recently been reviewed elsewhere[30]. Next, we examine specific problems in the analysis of genes under selection during domestication and the need to formally consider demography. We then focus on the genomic alterations associated with the creation of modern breeds, as assessed by genome-wide genotyping and whole-genome sequencing (WGS). The resulting analyses provide insights into human history and migration and identify candidate genes associated with a variety of phenotypes. Finally, we discuss the relevance of recent advances to our understanding of cancer and behavioural genetics.

Demography and selection

One crucial concern with inferring the effects of selection on the genome is the influence of demographic history on selective sweep analyses and the problem of false positives. The effects of increased genetic drift, mediated by reductions in effective population size (N_e), on genome-wide patterns of genetic heterozygosity and divergence result in some sites displaying a pattern of polymorphism that mimics those caused by artificial or natural selection[18,31–33]. Similarly, population subdivision can also produce signals that are difficult to distinguish from selection[34,35]. Most analytical models that are used to identify the action of selection on the genome assume a hypothetical demographic history, such as an expanding or stable population, to develop inferences.

A more realistic approach involves two steps. First, a specific demographic model is inferred from genetic data. Second, via simulations of the inferred demographic history, the neutral expectations for selection scan statistics are generated that can be used to calculate a false positive rate (also known as false discovery rate; FDR) for each SNP or genomic window. These feature-specific rates are estimates of the probability that an entirely neutral process produced the observed values of selection scan statistics. The first use of this two-pronged approach for dog evolutionary inferences was reported in a pair of recently published studies[17,18]. Demographic inference was performed with three dog genomes, three wolf genomes and one golden jackal genome, using the Generalized Phylogenetic Coalescent Sampler (*G-PhoCS*), which simultaneously solves for N_e, divergence time and admixture among lineages[17] (FIG. 1). This analysis produced some surprising results and other findings that were consistent with previous genetic studies. The authors found a strong bottleneck associated with initial domestication, which was nearly coincident

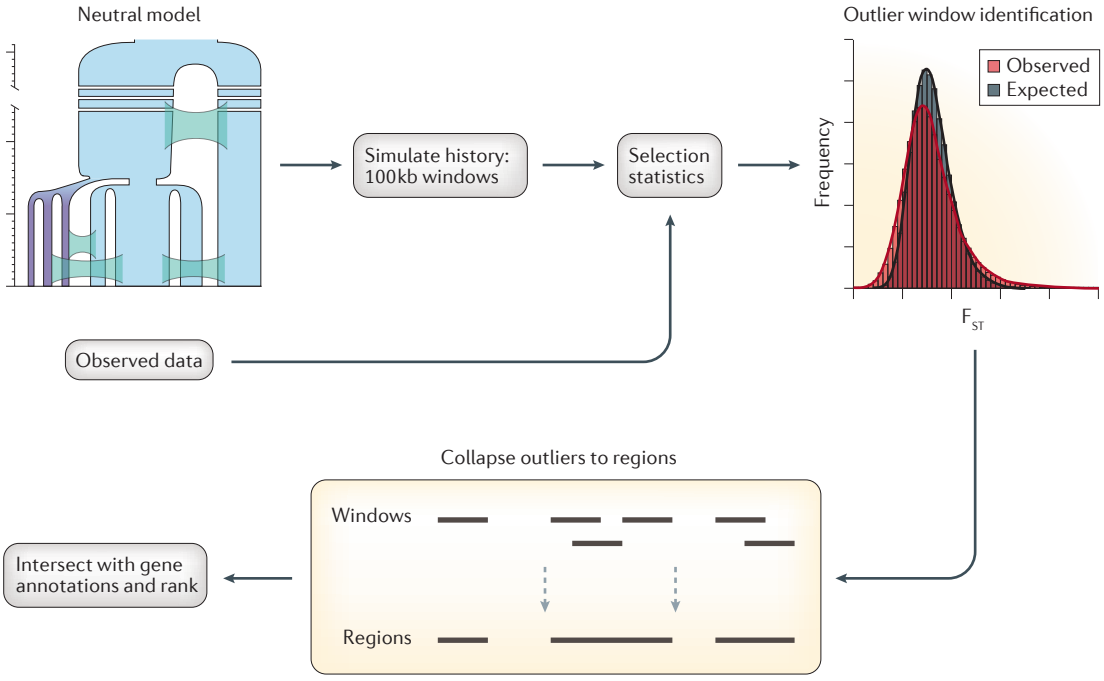

Figure 1 | **Analytical approach to correct for demographic history in selection scans.** First, a model (such as the Generalized Phylogenetic Coalescent Sampler (*G-PhoCS*)) is used to reconstruct demographic history based on genetic data. In the schematic demographic model shown[17], the history of dogs is traced chronologically from top to bottom. Branching represents lineage divergence, branch widths are proportional to inferred effective population size (N_e), and curved green connections between branches represent migration bands. Second, sequence data are simulated under this model assuming neutrality and compared to the observed data. Specifically, statistics (in our example, F_{ST}) for individual windows calculated from the empirical data are compared to distributions for those statistics obtained from the simulations. This comparison allows identification of empirical windows containing patterns of polymorphism that are unlikely to have arisen under the inferred neutral demographic history — that is, those that are putatively under positive selection on the dog lineage[18]. Candidate windows within a certain distance from each other are then merged into candidate regions. These regions are then analysed regarding all the genes they contain (see REF. 18). Adapted from REFS 17,18.

Population bottlenecks
Reduction in the size of a population due to any of a variety of factors (for example, natural disasters, disease or human intervention) that in turn reduces genetic variation in the population.

Haplotype
A group of variants or markers on a chromosome that are inherited together from one generation to the next. It can also refer to a pattern of variation observed across members of a population.

Linkage disequilibrium
(LD). Nonrandom association of alleles located at distinct loci; measured by determining if the frequency of two loci co-occurring is higher than expected by chance.

Selective sweep
A decrease in genomic variation surrounding a mutation due to positive selection for the mutation.

Genetic drift
Allele frequency changes in a population due to random mating of members of the population.

Effective population size
(N_e). The number of individuals that contribute equally to inherited genetic variation to the next generation within a given population.

Population subdivision
The relational structure of species with multiple subpopulations that exist either in total isolation or with minimal gene flow between them.

Admixture
The process by which isolated populations initiate previously non-existent gene flow.

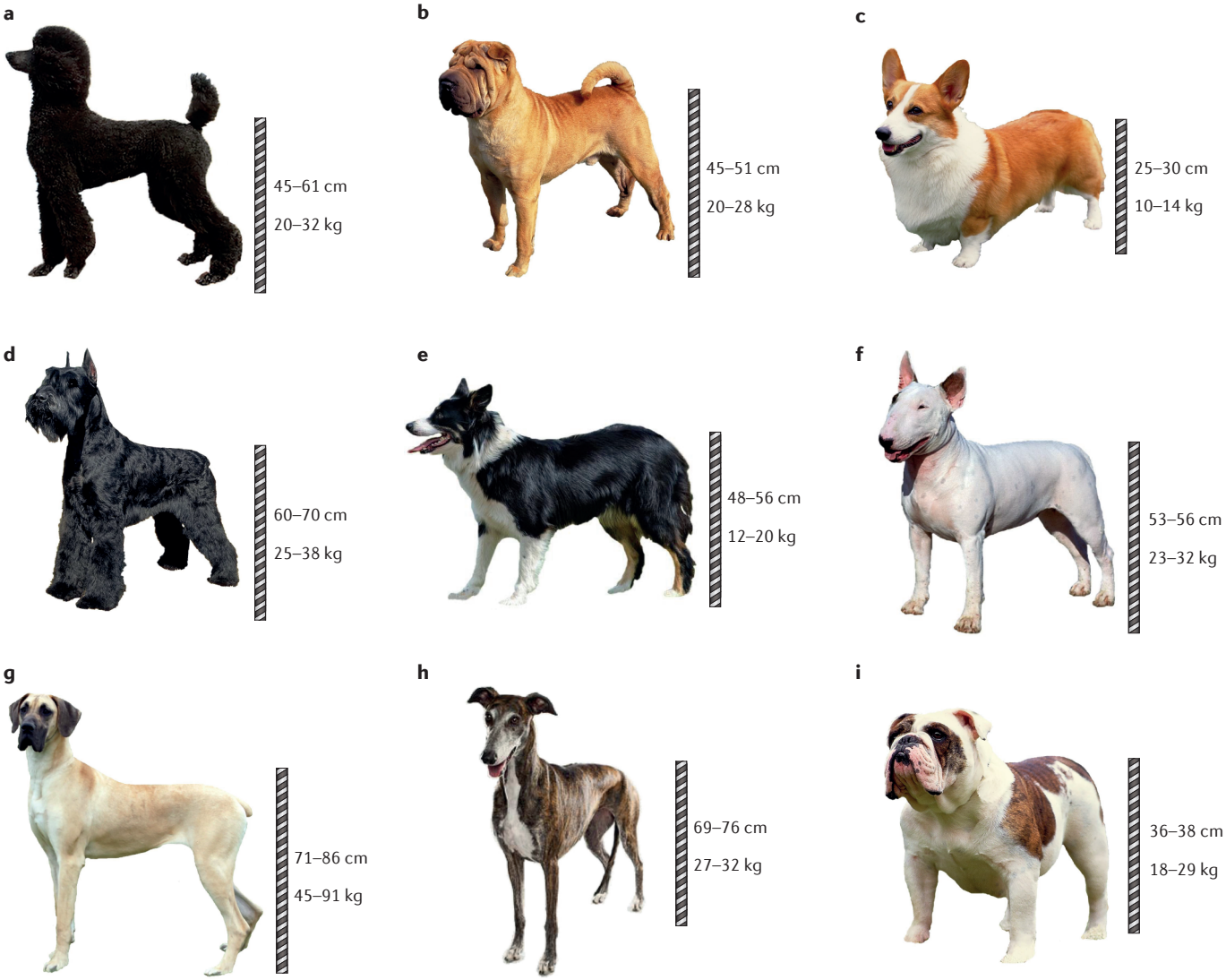

Figure 2 | **Phenotypic diversity of dog breeds.** A variety of dog breeds are shown with a range of heights and body weights noted for comparison (data from the American Kennel Club (http://www.akc.org)). **a** | The Standard Poodle is a popular dog of varying coat colours. **b** | The Shar-Pei has excess skin wrinkling. **c** | The Pembroke Welsh Corgi displays chondrodysplasia, meaning it has disproportionately short legs. **d** | The Giant Schnauzer has distinctive facial hair. **e** | The Border Collie is a standard herding dog with white spotting. **f** | The Bull Terrier demonstrates white spotting as well. **g** | The Great Dane is large in body size. **h** | The Greyhound is a standard sight hound. **i** | The Bulldog features a distinctive skull shape.

with a bottleneck in the wolf ancestor. Consequently, the loss of genetic variation in dogs relative to the wolf ancestor is nearly an order of magnitude more than was previously estimated by comparison to living wolves alone[17,36]. They also determined that modern breeds have also experienced severe bottlenecks that have altered their genetic landscape, consistent with historical records and previous analyses not utilizing demographic modelling that suggest they were of varying intensity[5,12–14,17,22,37] (see below). In the second study, the authors used the model inferred with *G-PhoCS* to control for FDR and to robustly identify candidate regions of the genome that had undergone positive selection early in dog history[18]. One of the top hits was centred on carbon catabolite repression 4-like (*CCRN4L*; also known as *NOCT*), a gene important for lipid metabolism, consistent with a change in dietary composition as the wolves from which dogs would arise began associating with hunter-gatherers.

In addition to the bottleneck associated with the initial domestication, two additional demographic features make it essential that future studies of selection in dogs explicitly incorporate a demographic model. First, there is a wide range of diversity levels among dog populations, such that dogs cannot be considered as a single group when conducting demographic inference. Many modern breeds have probably been, and continue to be, founded by highly related individuals sharing a specific phenotype (for example, FIG. 2). Furthermore, breeding to a predetermined phenotypic standard and the 'popular-sire effect' further diminish variation, leaving a signature of reduced variation in the genomes of many dog breeds[5,14,20].

Popular-sire effect
A reduction in genetic diversity in a population due to nonrandom and excess mating of a sire with desirable traits.

However, many free-living or semi-feral populations, such as village dogs, may retain high genetic variation because their breeding is less constrained[7,22]. Second, admixture has been common among dog populations and breeds and with grey wolves throughout the history of domestication[7,10,16–18]. Even the common ancestral root of dogs and grey wolves shows substantial gene flow with the ancestor of modern Eurasian golden jackals[17]. The amount of admixture between dogs and grey wolves is substantial, with as much as 15% of dog and 25% of wolf variation derived from interbreeding after dogs began to diverge[10,17], suggesting the dog and wolf genomes have been enriched through admixture that may confound analysis of selection and demographic history. In particular, admixture may increase diversity in dogs and obscure the signature of past episodes of positive selection.

Overall, the complicated demographic history of dogs has probably led to a high FDR[17] (and probably also false negatives) in studies searching for signals of selection in the dog genome that do not account for demography, explaining why candidate gene lists based upon demography-informed and demography-agnostic approaches have little overlap[18]. The most convincing case studies of selection relate specific genes to function or phenotype (see below), as demonstrated by amylase (*AMY2B*) gene copy-number increases in response to a high-starch diet during the agricultural revolution[18,36], the transfer of black coat colour from dogs to wolves[38], and hypoxia adaptations from high-altitude wolves to Tibetan dogs[9,39,40]. Ultimately, candidate genes will need to be verified by resequencing and functional studies, including the use of genome-editing technology to introduce genes and specific mutations to dogs or model organisms to verify phenotypic effects (for example REF. 41) (BOX 1). Until such confirmatory data appear, previous results based on selective sweep analyses should be interpreted cautiously.

Recent history of dogs and humans

Research and theory have focused on the relationship between the early phase of dog domestication and human demographic changes such as population increase and dispersal[7,9,11,16,42]. However, few studies have addressed the same question in modern dog breeds. A notable exception has been the study of *AMY2B* copy number in modern breeds, which defined a pattern that was consistent with the geographic origin of agriculture[43]. In both dogs and humans, a copy-number expansion of *AMY2B* accompanied the rise of agriculture, enabling more effective processing of complex carbohydrates[44]. This co-evolution continues in modern populations, with a decrease in amylase copy number in both human and dog populations with less carbohydrate-rich diets[45,46], and is also consistent with ancient DNA analysis of pre-agricultural Neolithic dog remains from Europe, which did not possess *AMY2B* copy-number expansions[47].

Recently, Dreger *et al.*[48] hypothesized that existing dog populations could be used to track the movement of human populations over the past 500 years, and they tested this hypothesis using dogs from Sardinia. Although most modern breeds have arisen through strong selection for physical and behavioural traits, the phenotypically heterogeneous Fonni's Dog, a niche population that has been present on the island of Sardinia for over 150 years, is not a registered breed. Yet, these dogs have distinct behavioural patterns, reflecting their skill as flock guardians and protectors[49,50]. In this study[48], WGS and SNP data from 155 modern canids, including Mediterranean breeds, were used to reconstruct the genomic architecture of the Fonni's Dog, revealing that the breed originated from the interbreeding of sight hounds and mastiff breeds, probably Saluki-type coursing hounds and Komondor-type flock guardians, respectively. As such, the genetic history of Fonni's Dog parallels known human demographic events[48] (FIG. 3). Specifically, the relationship of the Fonni's Dog to breeds from the Eastern and Southern Mediterranean, the Komondor and Saluki, respectively, parallels the migration of human populations to Sardinia, whose people are genetically similar to natives of Hungary, Egypt, Israel and Jordan[51]. Additionally, relationships of the Fonni's Dog with breeds originating in the Middle East and North Africa (for example, Anatolian Shepherd and Sloughi) mirror documented secondary trade and migration routes[51]. Thus, where humans travel, so go their dogs, demonstrating the value of studying canine populations, particularly in regions where humans and dogs have developed complementary skills for survival and may show parallel adaptation[39,52].

Finally, the movement of dogs alongside humans also increases the opportunity for genetic exchange with wild wolf populations and even the opportunity for enhancing adaptation. For example, *EPAS1*, a gene linked to hypoxic adaptation in humans, is clearly under selection in the Tibetan Mastiff, a breed native to the Tibetan plateau, and Tibetan grey wolves, which are native to the mountains of Tibet[53,54] (FIG. 4a). A recent analysis has demonstrated that the beneficial mutations originated in high-altitude Tibetan wolves and were transferred to dogs, perhaps about 20,000 years ago, paralleling an analogous transfer of beneficial *EPAS1* variants from Denisovans, an ancient, now-extinct hominin, to modern humans in the Tibetan plateau[54].

Deleterious variation

Theory and empirical data support the notion that selection should be inefficient in small or bottlenecked populations such that deleterious variants can drift

Box 1 | CRISPR–Cas9 genome editing in dogs

A technology that is certain to affect dog breeding is the genome-editing tool CRISPR–Cas9 (REFS 163–167), which was recently and successfully applied to dogs. In 2015, a knockout mutation in the myostatin (*MSTN*) gene, which has been shown previously to increase overall body muscle mass in Whippets, was introduced into Beagles using CRISPR–Cas9 (REF. 168). The resulting experiment produced two *MSTN*$^{-/-}$ Beagles, albeit with a weaker phenotype than observed in Whippets[168,169], suggesting that additional modifications to the procedure are necessary and/or that other factors in the genetic background are relevant. The idea of creating dogs using genome editing with specific traits is interesting, but it raises the question of how to control the rise in deleterious alleles that is almost certain to occur if the genomes of CRISPR–Cas9 dogs with new desirable traits become over-represented in the population, due to popular-sire effects, with single founders over-contributing to the subsequent generation.

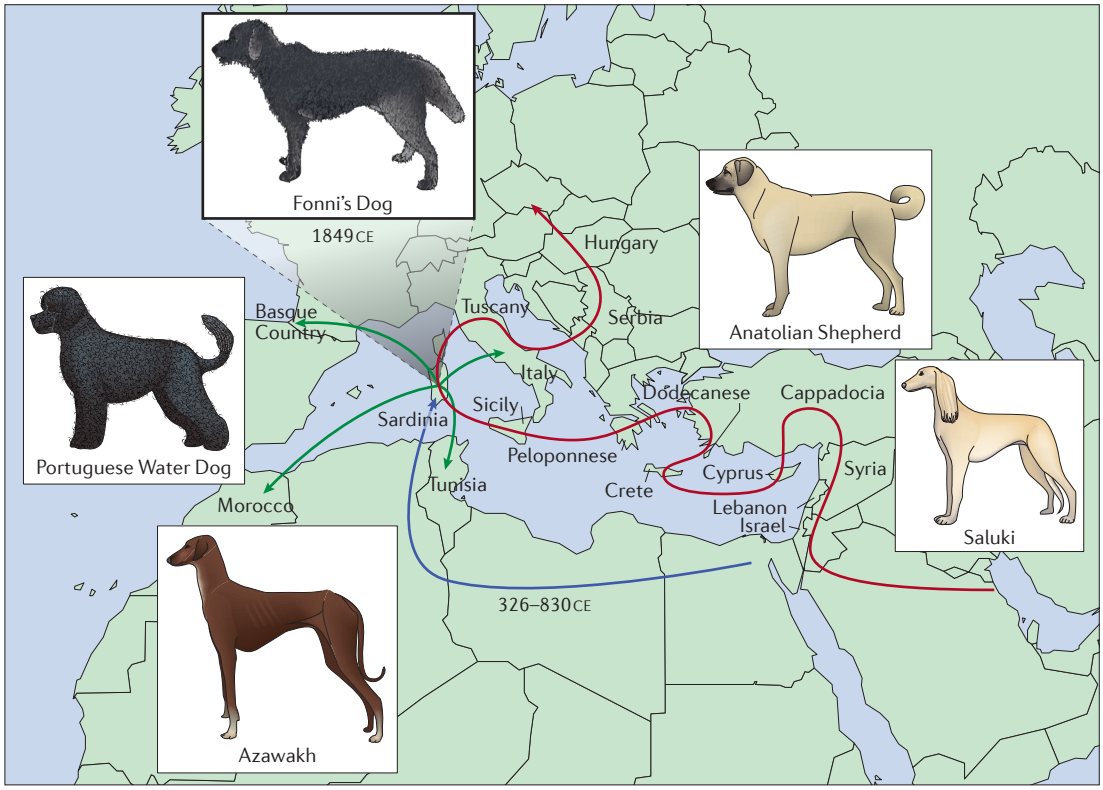

Figure 3 | **Migration of the Fonni's Dog mimics human population migration.** Molecular data suggest that the Fonni's Dog shares most recent genetic ancestry with the livestock guardian breeds from Eastern Europe (Komondor and Anatolian Shepherd) and sighthound breeds from the Middle East and Northern Africa (Saluki, Azawakh, and Sloughi). Humans are believed to have migrated across the Mediterranean Sea via island-hopping from the Middle East, westward through the islands of Greece and Italy (red arrow). Humans in Sardinia experienced admixture from Northern Africa (blue arrow). Therefore, the proposed human migration route through the Mediterranean overlaps directly with the geographic origins of the source populations that eventually contributed to the formation of the Fonni's Dog. Conversely, the human population of Sardinia is expected to have contributed to surrounding areas (green arrows), including the Gibraltar region (bottom left green arrow), which likewise mirrors the evidence that the Fonni's Dog has strong external influence on the development of the Portuguese Water Dog.

to high frequency and potentially become fixed[55–61]. These findings are particularly germane to understanding deleterious variation in dogs, as domestication involved two bottleneck phases. The first resulted from the initial domestication process and involved at least a 16-fold reduction in N_e (REF. 17). Over the past few hundred years, breed-specific bottlenecks occurred as a result of breed formation, typically from a limited number of founder individuals, and subsequent strong artificial selection for phenotypic traits deemed desirable. Furthermore, persistent artificial selection during the development and maintenance of specific traits is likely to have captured loci containing deleterious alleles. The accumulation of deleterious variation in dogs thus reflects the combined effects of drift in small bottlenecked populations and of selection. The higher frequencies of genetic disease and anatomical abnormalities in modern dog breeds support this idea[19,62–65]. The increase in genetic load resulting from population bottlenecks and selection probably has an effect on health, but it is one that varies across breeds given different demographic histories and intensity of artificial selection.

To quantitatively assess the contribution of early and breed-specific bottlenecks to deleterious variation, Marsden *et al.*[19] examined heterozygosity in non-synonymous versus synonymous SNPs, given that the former are more likely to be deleterious. Using whole-genome data from purebred dogs, village dogs, and grey wolves analysed relative to golden jackal (*Canis aureus*), they showed a substantial increase in the proportion of non-synonymous heterozygosity in dogs in general relative to wolves, demonstrating the strong effect of the initial domestication bottleneck rather than subsequent breed-specific bottlenecks or inbreeding. On average, the genetic load from the ancestral dog bottleneck was increased by a substantial 2%. This value is similar to that found in humans from Eurasia, which all derived from an out-of-Africa bottleneck that was not shared with those of contemporary African descent, and suggests that ancient bottlenecks leave a lasting and pronounced signature in descendent genomes. Marsden *et al.* also detected a higher fraction of Mendelian disease genes in sweep regions associated with artificial selection, further supporting the notion that persistent selection in breeds enhances the incidence of disease[19].

Genetic load
A reduction in the mean individual fitness of a population due to the presence of deleterious alleles or allelic combinations relative to a genotypically ideal population.

Non-synonymous
A change in DNA sequence that alters the encoded amino acid, thus altering the encoded protein.

Synonymous
A change in DNA sequence which, if it occurs in a coding region, does not alter the resultant amino acid.

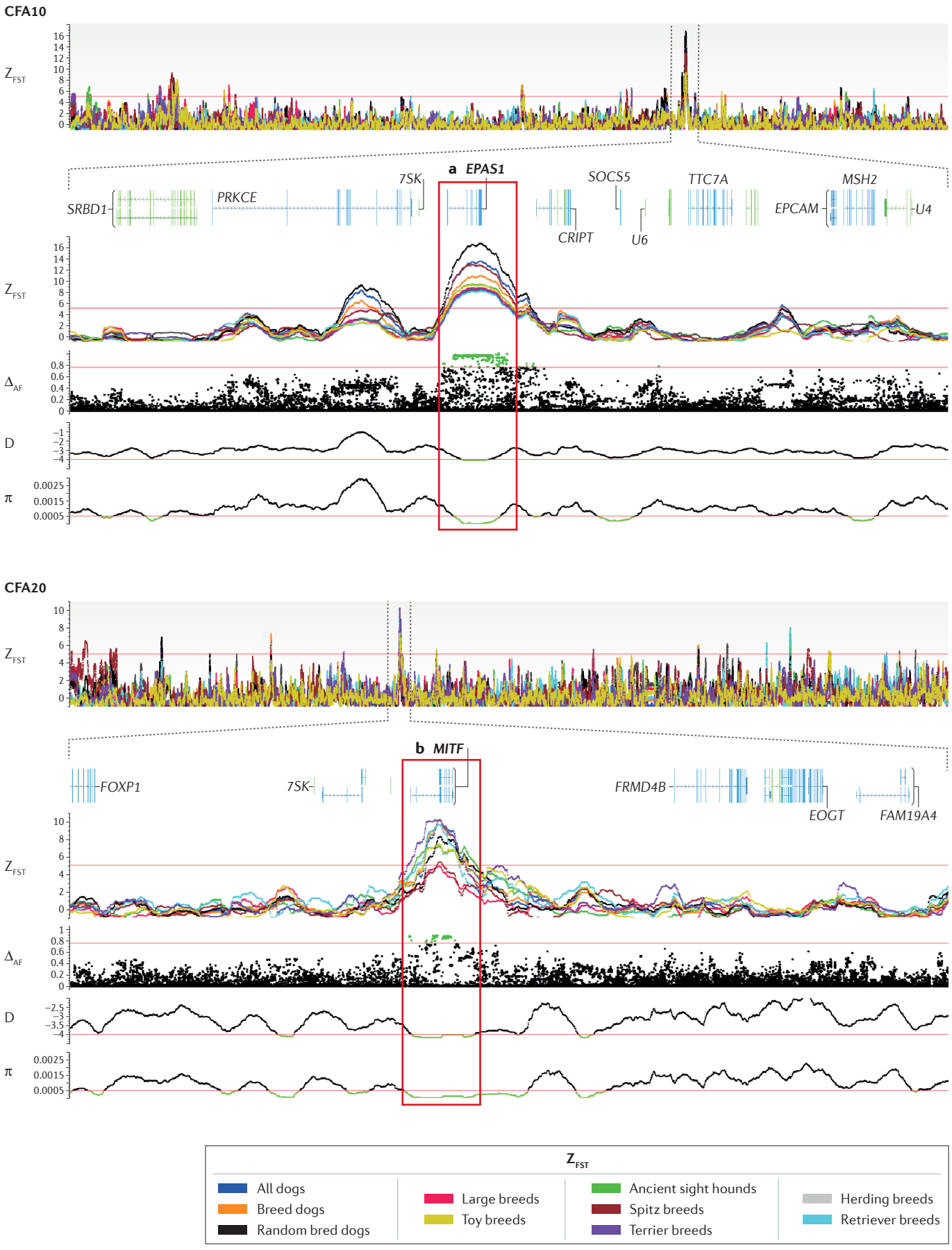

Figure 4 | **Selective sweeps from whole-genome sequencing analysis.** Z_{FST} represents the outlier score for F_{ST}, indicating a region with high divergence from others. DeltaAF (Δ_{AF}) is the shift in allele frequency in the population compared to all others, indicating private or enriched variation. Tajima's D (D) is a measure of the nonrandom segregation of alleles, with negative values indicative of selective or demographic influence in the region. Pi (π) represents nucleotide diversity, which is reduced in regions undergoing selection. **a** | The region surrounding *EPAS1*, a gene linked to hypoxic adaptation at high altitudes in Tibetan Mastiffs[174], on *Canis familiaris* chromosome 10 (CFA10). This gene has also been shown to be under selection in human populations and grey wolves living in the same environment. **b** | A sweep surrounding the *MITF* gene on CFA20 in herding breeds such as the Border Collie and Shetland Sheepdog that is linked to the degree of white colouration present in the coat[77].

Numerous targeted studies demonstrate the consequences of the accumulation of deleterious variation in dogs. One example is squamous cell carcinoma (SCC) of the digit, where copy number expansion in an enhancer-like element strongly associated with the KIT ligand (*KITLG*) gene dramatically increased risk of disease in the black but not white Standard Poodle[64] (FIG. 2a). The result was directly linked to selection for increased intensity of dark black coat colour, a prized trait among conformation breeders and one in which *KITLG* plays a major role. Subsequent studies show that regions of the genome identified as having a breed-specific selective sweep are enriched in known Mendelian disease alleles[19]. In another example, selection for reduced craniofacial length induces susceptibility to respiratory tract disorders[66]. Similarly, the desired excess skin wrinkling driven by hyaluronan synthase 2 (*HAS2*) gene mutations in the Shar-Pei (FIG. 2b) is unfortunately accompanied by excess mucus production. Together the mucus levels and skin folds lead to an unusually high level of skin infections in the breed[23]. In addition, an unstable 16 kb duplication upstream of *HAS2* in Shar-Pei has been linked to increased levels of hyaluronic acid, which induces a potentially severe autoimmune reaction and periodic fever[67]

In general, persistent selection for breed-defining traits, small N_e and inbreeding should lead to increases in the frequency of deleterious alleles and in the occurrence of Mendelian disorders[63,64,68]. This conclusion is supported by a large study of >90,000 purebred and mixed-breed dogs whose electronic medical records were evaluated[62], with 10 of 24 disorders found to be of higher incidence in purebreds, although no differences in prevalence were found for 13 of the 24 disorders, including common diseases such as cancer and hip dysplasia. The authors suggest that these disease alleles, which are spread throughout the dog population, may have occurred either multiple times or only once early in domestication. Some of these alleles may either represent or be linked to desirable traits and were brought to higher incidence by recent artificial selection or breed demography. Thus, understanding the demographics of breed formation and subsequent selection is pivotal to predicting the health of modern breeds, as strong and direct selection for desirable and extreme phenotypes, such as those associated with appearance, can drive the formation and retention of deleterious traits[19,69].

Chondrodysplastic
A state of abnormal cartilaginous growth resulting in disproportionate dwarfism. In dogs, this affects only the limbs, with minimal other observed effects.

Identical-by-descent
A haplotype shared between individuals that is inherited from a recent common ancestor without intervening recombination.

Quantitative trait locus
(QTL). A defined region of DNA that correlates with variation in a phenotype. Quantitative traits, by comparison, are phenotypes that vary in degree or presentation due to the joint effects of multiple genes.

Genomic studies of modern breeds

Breed-associated morphology. Dogs are unique compared to other species in both the purpose and the strength of artificial selection during breed formation and maintenance. Breeding of agricultural species is primarily focused on profit-motivated disease tolerance, reproductive and production traits, with little interest in behaviour or aesthetics (reviewed in REF. 70). Although other companion species such as the domestic cat have experienced moderate amounts of recent selection for coat length[71], hair type[72] and shortened muzzle[73], the establishment of isolated breed gene pools is more recent than for dogs, and selection for extreme phenotypes is less pronounced[74].

The diversification of modern dog breeds occurred predominantly during the past few hundred years, as modern breeding practices isolated specific phenotypes in discrete populations[5,12,14]. Genetic analyses of breeds suggest that distinct phenotypes were transferred among dogs through admixture, often early in breed formation, followed by phenotypic selection[20]. For example, the gene responsible for chondrodysplastic limbs in 19 dog breeds, such as the Corgi (FIG. 2c), is identical-by-descent, yet it spans at least four distinct breed groups or clades, suggesting that the mutation is old and occurred early in breed formation, rather than appearing independently in different breeds[75]. Similarly, the mutation responsible for the development of a moustache and eyebrows typified by the Schnauzer (FIG. 2d) is in the R-spondin 2 (*RSPO2*) gene, which, along with the surrounding haplotype, is identical across multiple diverse breeds, suggesting a single origin of the variant followed by admixture among breeds[76]. In the newly created Black Russian Terrier, haplotype sharing suggests that this trait was contributed by either the Schnauzer or the Airedale Terrier during breed formation (FIG. 5a). Using phylogeny to infer breed history facilitates the tracking of trait exchange between breeds. For instance, variation in the *MITF-M* promoter is believed to influence 'white spotting' (white belly and white collar) and is observed in several breeds in the herding group, such as the Border Collie (FIGS 2e,4b) and Shetland Sheepdog, as well as in the distantly related, non-herding Bull Terrier[77] (FIG. 2f).

The recent shared history of breeds has facilitated successful inter-breed mapping studies[22,24,78], as genome-wide association studies (GWAS) for common traits often reveal selectively swept shared haplotypes, permitting tentative genotype–phenotype links to be inferred[22,23,24,78]. Breed GWAS and quantitative trait locus (QTL) studies of coat colour, baldness, body size, ear shape, leg length, skull shape, hair growth patterns and other traits have all identified candidate genes in specific haplotypes that contribute to phenotypic variation and are reviewed in REFS 4,69,80 (BOX 2).

Body size provides several interesting lessons regarding how subtle levels of variation are responsible for large changes in phenotype[81,82]. Large (Great Dane; FIG. 2g) and small (Chihuahua) dog breeds differ in size by nearly 40-fold, a claim that does not apply to any other land mammal. In 2013, four new body-size genes — growth hormone receptor (*GHR*), high mobility group

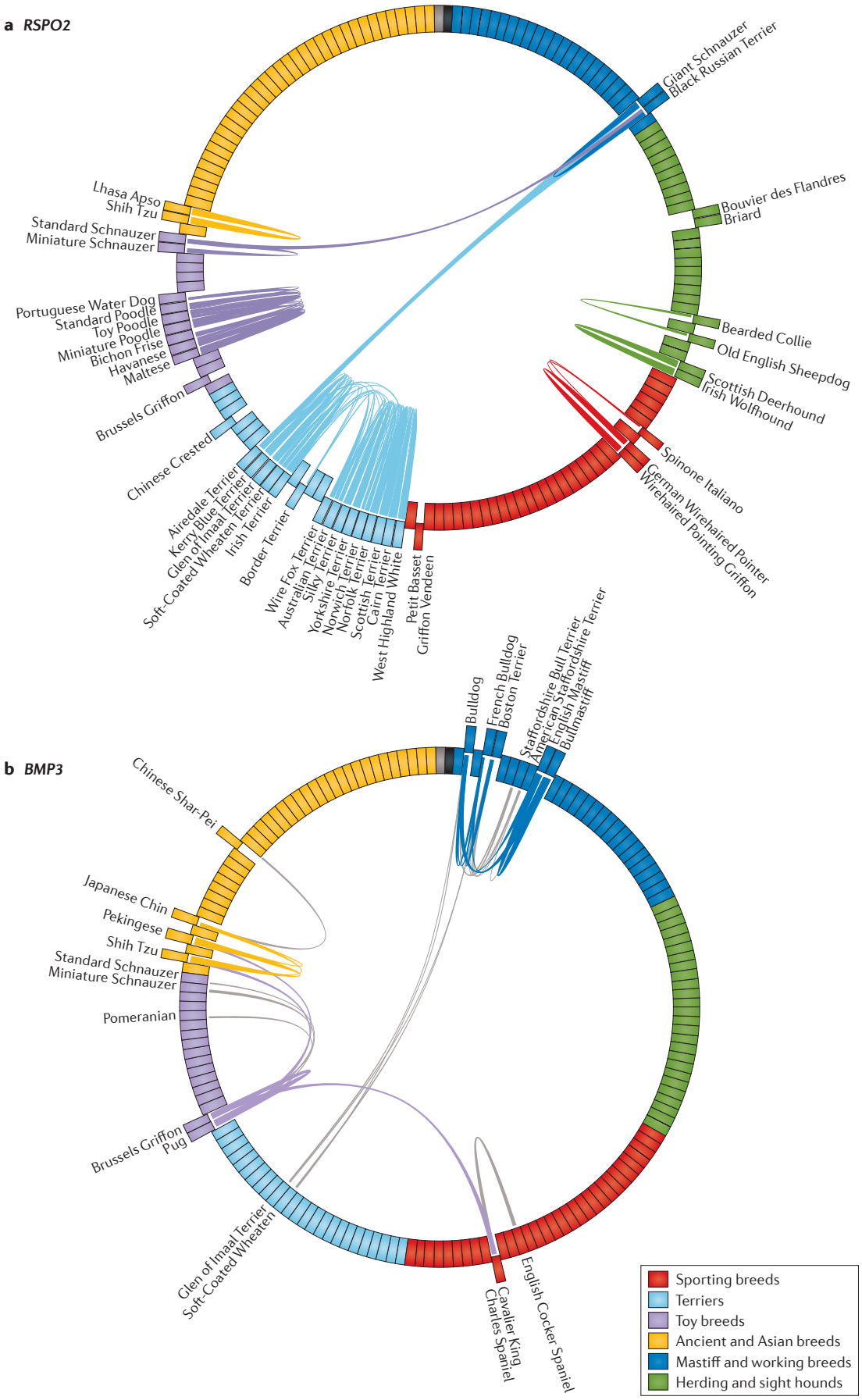

Figure 5 | **Haplotype sharing between breeds indicates the source of a common phenotype for a given trait.** Figures are derived from data published by Parker *et al.*[79]. **a** | An R-spondin 2 (*RSPO2*) mutation causes the 'furnishings' phenotype, typified by a 'moustache and eyebrows' that is observed in the Schnauzer (FIG. 2d) and in ~40 known dog breeds. Excessive haplotype sharing, indicated by the ribbons connecting breeds across the circle, show that the furnishings trait, which appears in the recently created Black Russian Terrier, came from either the Airedale Terrier or the Schnauzer. Breeds with the *RSPO2* mutation are set out from the main circle of breeds. The distribution of breeds with the phenotype suggests the mutation occurred early in the development of breeds. **b** | A nonsense mutation in bone morphogenetic protein 3 (*BMP3*) is associated with the brachycephalic phenotype in multiple breeds. The mutation is also carried by breeds with less severe head shapes. Breeds with the mutation and phenotype are set out from the circle. Coloured ribbons show extensive haplotype sharing between those breeds. Grey ribbons indicate haplotype sharing between brachycephalic breeds and breeds that do not have the phenotype. These breeds do not display the mutation, perhaps due to incomplete penetrance, and the mutation passes throughout the population.

AT-hook 2 (*HMGA2*), *SMAD2*, and stanniocalcin 2 (*STC2*) — were identified as major contributors to breed standard body weight (BSW)[81]. Combined with data from earlier studies of insulin-like growth factor 1 (*IGF1*) and IGF1 receptor (*IGF1R*)[25,83], derived alleles from these six genes account for 46–52.5% of the variance in overall body size across breeds, with most explaining variation in small-to-medium-sized breeds <41 kg (90 lb). This study[81] highlights a now-recurring theme in dog genetics whereby a small number of genes of large effect control complex phenotypes, as opposed to many genes of small effect, which is typical for humans (over 180 human body-size loci identified[84–86]). Three new genes on the X chromosome contributing to breeds with a BSW >41 kg (90 lb.), hereafter known as 'large breeds', have recently been found[82]. These include insulin receptor substrate 4 (*IRS4*) and immunoglobulin superfamily member 1 (*IGSF1*), both genes involved in the thyroid hormone pathway, and associated with IGF1R signalling, obesity and body mass[87,88,89–91]. Also included are variants in acyl-CoA synthetase long-chain family member 4 (*ACSL4*), which, in pigs, controls muscling and back-fat thickness[92,93]. A similar role in dogs probably explains why the derived variant is homozygous only in large 'bulky' dogs (for example, the English Mastiff and Saint Bernard), whereas the ancient allele is homozygous in large lean breeds (for example, the Irish Wolfhound and Greyhound) (FIG. 2h).

One other notable complex trait is skull shape, particularly brachycephaly, which is characterized by fore-shortening and a change in angulation of the muzzle as well as a globular skull, as observed in the Bulldog (FIG. 2i) and other breeds. Although multiple loci contribute to this phenotype[94,95], there is a particularly strong signal for the trait on canine chromosome one, which was recently shown to affect the SPARC-related modular calcium-binding protein 2 (*SMOC2*) gene[96]. The underlying mutation is an intronic transposable element that drastically reduces gene expression, accounting for 36% of facial length variation. An additional gene that affects skull shape, promoting a brachycephalic appearance, encodes bone morphogenetic protein 3 (*BMP3*)[97] (FIG. 5b). This mutation and the associated haplotype are carried by multiple breeds with brachycephalic head shapes such as the Shih Tzu, Pekingese and Japanese Chin (FIG. 5b). However, haplotype sharing also exists between brachycephalic and non-brachycephalic breeds including the Pomeranian and Schnauzer, in which the effects are less penetrant, but they have passed the haplotype to or received it from breeds such as the Brussels Griffon and Pug, which are in turn brachycephalic. This interesting observation highlights how knowledge of breed relationships facilitates the tracking of haplotypes, which may only substantially influence phenotype when present on specific genetic backgrounds. In this case, dogs offer a unique mechanism for disentangling the genetics of a complex trait, as one can reasonably assume that closely related breeds share variant alleles for specific traits. Thus, increasing the number of individuals in a GWAS for a specific trait by combining SNP-chip data from related breeds both increases power and decreases false positives.

The dog has already proven to be an invaluable model for the identification of basic genetic mechanisms influencing biology, as many causative and influential variants have been identified for obvious phenotypes. The combination of accurate and discrete phenotyping with high-resolution genomic data will also make it tenable to explore more complex traits with impacts on human health, such as differential disease susceptibility between breeds, subtle variation in morphological traits and the persistence of heritable behaviours.

Breeds and lifespan. Dog breeds differ substantially and reproducibly in lifespan. Smaller breeds almost uniformly live longer than large breeds, and the relationship between lifespan and body size in dog breeds is surprisingly linear[98]. Although deleterious variation is driven largely by population bottlenecks[19], intense selection for extreme phenotypes appears to diminish longevity. This is biased towards large breeds, where body size is almost two orders of magnitude greater than small breeds and is accompanied by a twofold decrease in life expectancy[52,53]. Interestingly, association studies using age of death as an end point identified loci with significant *P* values on six chromosomes, many of which are located near body-size loci, such as *IGF1* (REF. 98).

Although there is some correlation between decreased lifespan and increased serum levels of IGF1 (REF. 56), there is no obvious biological reason why increased body size is strongly associated with decreased lifespan. In fact, among other mammals, the reverse is often true, with smaller species having shortened lifespans. An argument can be made that disease incidence plays a role. For instance, in large dogs, selection for mass may increase the risk of some forms of dilated cardiomyopathy (reviewed in REF. 58), as well as gastric dilatation and volvulus (bloat)[59]. Some studies suggest that factors such as thymus function, which declines more slowly in long-lived breeds, play a major role[99]. Small dogs also have longer average telomere length, and Fick *et al.*[100] have shown that this correlates with lifespan in dogs. Alternatively, the difference may reflect breed incidence of common disorders, such as heart disease or cancer[101]. Kraus *et al.*[102] have followed a large multibreed cohort for over 20 years and suggested

Box 2 | **GWAS and follow-up studies in dogs**

By calculating genomic homozygosity, demography, and molecular measures of genome diversity, the following findings are clear. Each breed has a unique profile of genome diversity that varies in the amount of total homozygosity as well as the number, distribution and size of homozygous regions[5,20,79]. Additionally, multiple individuals from a single breed can be combined to obtain an accurate reflection of breed-specific homozygosity, a result which is supported by multibreed mapping studies (reviewed in REF. 170). Finally, when attempting to ascertain the total single nucleotide polymorphism (SNP) variation within a breed, rarely are >10 individuals needed[20]. Not surprisingly therefore, the optimal cost/benefit ratio for whole-genome sequencing (WGS) as a follow-up to genome-wide association studies (GWAS) is largely achieved with 30× data from two individuals, as a third sequence improves variant detection by only 8%[20].

These results dovetail well with findings from previous studies by Lindblad-Toh *et al.*[15], who have shown that for any 10 kb region, 80% of haplotypes observed with a frequency of at least 5% in one breed are found in other breeds as well. Thus, they argue that dog GWAS should include rather than exclude SNPs that are rare in a single breed, as the resulting haplotypes may be more common than expected in the general canine population, a finding that seems to be validated by canine WGS. This is further supported by their observation that, on average, the probability of sampling the same haplotype on two chromosomes chosen from different breeds is only twofold lower than for the same chromosome in dogs of a single breed. Given the limited genomic structure in the dog phylogenetic breed tree, this would be expected[14]. Dog breeds are far more related than their appearance would suggest[14]. However, the early estimates that a SNP chip of approximately 15,000 evenly spaced SNPs would be sufficient for identifying a given trait-associated locus 99% of the time if one had a collection of 100 cases and controls is probably an underestimate[15]. Although obvious issues such as penetrance, phenocopy, and the ability to accurately determine phenotypes contribute, so do issues of breed structure, including popular-sire effects. Therefore, while many studies have been undertaken successfully with the current SNP chip of 170,000 markers, a denser chip is needed and is expected for the community soon. Regardless, the task of moving from a linked or associated marker for a single trait to finding the causal gene mutation will still be difficult, as deleterious mutations can be expected in non-coding regions such as general regulatory regions[171], long intergenic non-coding RNAs (lincRNAs)[172], and splicing regulatory sequences[173].

a relationship among all of these factors; for example, large dogs die earlier because once senescence initiates, which happens earlier in giant breeds, they then experience an increased rate of ageing. That is, once large dogs begin to age, the rate at which they do so accelerates compared to small dogs. Selman and colleagues[103] have argued that this is supported by IGF1 serum protein data, as IGF1 levels are higher for longer periods of time in large dogs. Multiple studies show that pathways related to metabolism and growth, such as IGF1 signalling, affect both lifespan and cancer, which is itself a disease of older age.

The potential impact of dissecting how genomic differences between breeds contribute to differential lifespan is far-reaching. The genomic, metabolic and developmental correlates to lifespan in the dog model can not only inform on the longevity of humans and other mammals but also hold the potential to greatly increase the time humans have with their closest companions.

Canine cancer

Breed susceptibility. Dogs are second only to humans in medical surveillance and preventive health care. Cancer is the leading cause of disease-associated death in dogs, affecting one in four of the 77 million dogs living in the United States, with 50% of dogs >10 years old developing the disease[104–106]. Dogs are diagnosed with many of the same cancers as humans[107,108], with similar underlying presentation, pathology and treatment response[68,109,110]. Multiple studies show that the genes and pathways involved in canine cancer development and progression are the same as those found in humans[111,112]. However, the compressed lifespan of dogs compared to humans means that cancers that take 15–20 years to mature in humans can be studied in the dog in two to three years[4,113]. Perhaps most importantly, canine cancers are spontaneous, a fact that distinguishes the dog from other mammalian cancer models, such as the mouse, where many cancers must be induced[68,108,110,114,115]. The high incidence of breed-specific cancers offers opportunities to identify germline sequence variants leading to disease susceptibilities that have been difficult to uncover in humans[64,116–119] and to promote understanding of biological processes such as tumour heterogeneity, resistance to chemotherapy and the role of the immune system in tumour evolution[53,10].

For instance, selection for black coat colour, and therefore increased susceptibility for SCC of the digit, offers an explanation for mutation persistence in the Poodle population, and the disease is hence a growing concern[64]. SCC is also found in Giant Schnauzers and Briards, though in both light- and dark-coated dogs, which was initially puzzling given the previously identified links between SCC and black coat colour intensity. However, in Giant Schnauzers and Briards, dark coat colour is controlled by β-defensin 103 (*CBD103*; also known as the K locus), which acts as an alternative ligand for the melanocortin 1 receptor (*MC1R*) gene in some dog breeds and wolves[38]. Alternatively, *MC1R* is inactivated in white and cream-coloured Poodles by a common nonsense mutation, suggesting that the *KITLG* disease allele requires a functional *MC1R* gene to function, a fact previously unknown for this disorder, which further suggests that gene–gene interactions may play a role in the pathogenesis of both the human and canine disease.

Penetrance
The proportion of individuals in a population who display a given phenotype in the presence of a specific genotype.

One particularly common cancer observed in a limited number of breeds is histiocytic sarcoma (HS), which affects 15–25% of Bernese Mountain Dogs (BMDs)[120–122] and Flat-Coated Retrievers (FCRs)[122,123]. In a study of cancer death conducted in the 1990s among Swedish dogs <10 years of age, the BMDs and FCRs, together with the Rottweiler, Irish Wolfhound and Saint Bernard breeds, possessed the highest rate of cancer-related death, with the BMDs most at risk[124]. HS is a 'catch-all' term for a dendritic neoplasm, which can present in a localized (largely in FCRs) or disseminated (largely in BMDs) form. It is highly aggressive and uniformly lethal[122]. Phylogenetic studies demonstrate that FCRs and BMDs are not closely related, and it is therefore not surprising that the associated risk loci do not overlap[117]. A recent GWAS in BMDs identified a 35 kb susceptibility locus spanning the cyclin-dependent kinase inhibitor 2A (*CDKN2A*) and *CDKN2B* loci and the adjacent methylthioadenosine phosphorylase (*MTAP*) gene, although no causative mutation was found[117].

The most comprehensive study of susceptibility to a common canine cancer has been that of osteosarcoma[125,126]. The study identified 33 distinct osteosarcoma loci, which the authors argued explained 55 to 85% of the osteosarcoma phenotypic variance observed in Rottweilers, Greyhounds and Irish Wolfhounds[126]. The highest correlation was for Rottweilers, in which 15 loci explain a striking 85% of variance. The idea that a complex cancer can be explained by such a small number of loci is remarkable and is not a theme observed in most human cancer studies. Most of the loci found in the dog study were either previously identified osteosarcoma-associated loci or genes that regulate bone differentiation and growth. This study may have been more successful at identifying genetic factors than previous work because the investigators took advantage of the generally late onset of osteosarcoma, enrolling aged controls and young cases, thus increasing the likelihood that individuals were assigned the correct case/control status. The authors wisely considered the three breeds separately — even though the Greyhound and Irish Wolfhound could have been considered together as hounds — as they argued that a deleterious allele might be fixed in one or more breeds, making it impossible to detect the mutation via a GWAS in that breed or combination of breeds. That was found to be the case: some mutations were fixed in one or another breed and were only found because the authors compared data between breeds to find the disease alleles. This scenario is one way in which human and canine genetics differ. Few disease alleles are fixed in human populations, yet they are tolerated in dogs because of the strong selection for other traits.

Tumour biology. Many studies highlight the value of canine cancer models for the development of clinical therapies[108,115,127,128]. A thorough summary of canine tumour biology and advances in human cancer related to dog studies has been recently published[127] and will not be replicated here except for two salient points. First, a major advantage to the dog model is that cancers are spontaneous and thus recapitulate the growth and pathological features of analogous human cancers more often than cancers in rodents[127]. A recent success story is that of spontaneous canine invasive transitional cell carcinoma of the bladder (invTCC) in which gene expression studies based on RNA sequencing (RNA-seq) demonstrated that about 85% of tumours carry a somatic V595E mutation in the *BRAF* oncogene[128]; this is homologous to the V600E mutation that is present in 8% of all human tumours[129], including 45% of human melanomas[130]. Functional experiments showed that in canine tumours, as in human tumours, activating mutations in *BRAF* stimulate the MAPK pathway and that this effect can be reduced through treatment with BRAF(V600E) inhibitors[128]. The development of urine-based diagnostics will prove extremely useful to the 20,000 pet owners whose dogs are diagnosed with invTCC each year, allowing them to circumvent invasive and expensive tests. Of at least equal importance, this result provides a new system for the study of human BRAF-driven cancers[128].

A second key point is that continued maturation of the dog cancer model is directly tied to the development of several consortia. Most important is the Comparative Oncology Trials Consortium (COTC), which facilitates preclinical trials of both therapy and diagnosis across the United States, with a goal of advancing canine studies relevant to human disease[127]. Cancer drug development is an expensive endeavour, and rarely do drugs advance from phase three clinical studies to regulatory approval in the United States (REFS 127,131 and references therein) leading to speculation that "less than one drug is approved for each billion dollars of research and development" (REF. 127). Introduction of the dog as a spontaneous preclinical model is likely to reduce these costs and speed the time to drug approval[108].

Canine transmissible venereal tumours. Human and most canine tumours arise through somatic mutations that occur in a tissue, coupled with germline variants that contribute to greater or lesser degrees to the disease onset, progression, and treatment response. There are rare cases in humans where viruses play a role in cancer susceptibility, such as that of human papillomavirus (HPV). Canine transmissible venereal tumours (CTVTs) are an entirely new class of tumours that are unlike any of the above. CTVTs are clones of a cancer that has been propagated for thousands of years via sexual transfer of malignant cells between dogs[132,133]. These tumours are endemic throughout the world, and host survival is high[134]. Data from mitochondrial DNA, major histocompatibility loci, and nuclear microsatellite analyses suggest that CTVTs are derived from one neoplastic clone that arose in a single dog[133,135,136]. Thus, considerable research is now aimed at understanding how the tumour has escaped host immune detection while passing through hundreds of different dog populations from all over the world[29,136]. A recent WGS comparison of two CTVT genomes from different parts of the world and 186 normal dog genomes generated nearly a million high-confidence somatic substitutions for analysis[29]. Functional annotation and gene set enrichment analysis identified several pathways with

multiple somatic mutations in the tumours. By far the most compelling results identified genes that play roles in immune surveillance, particularly those for self- versus non-self-recognition, antigen processing, initiators and executors of apoptosis, and DNA repair. The large number of mutations per gene indicated that thousands of years of cryptic selective pressure to avoid the host immune system have heavily mutated the genome of the tumour. The identification of mutations in 100 such genes that are specifically altered early in tumour evolution, including additional mismatch repair genes, tumour suppressors and oncogenes, provides useful general guidance on how tumours, in general, can avoid the host immune system.

Canine behavioural genetics

Anomalous behaviours. While cancer represents one end of a phenotypic spectrum, as it is a comparatively easy-to-define trait (although cancer subtypes can be very difficult to discern), behaviour represents the opposite end of the continuum, as nearly all behavioural observations fall victim to some degree of subjectivity. The greatest success has been achieved in studying canine behavioural diseases that meet the criteria of phenotypic similarity to the corresponding human disorder and responsiveness to the same therapies as the human disease and that have the potential to reveal underlying biology[137]. Canine compulsive disorder (CCD) meets these criteria. The phenotype can include licking, tail chasing, spinning, self-mutilation, and blanket sucking[138]. It has been observed in multiple breeds including Doberman Pinschers and Bull Terriers[138,139] and is similar to human obsessive-compulsive disorder (OCD) in both presentation[138] and response to therapy[140]. A GWAS of CCD in Doberman Pinschers identified four synaptic-function genes: neuronal cadherin (*CDH2*), catenin α2 (*CTNNA2*), ataxin-1 (*ATXN1*), and plasma glutamate carboxypeptidase (*PGCP*; also known as *CPQ*). The *CDH2* result has been replicated and is compelling[141,142], as the gene is involved in synaptic plasticity, and reduced expression of *CDH2* is associated with Gilles de la Tourette syndrome in humans, which includes elements of OCD[143].

Behavioural scientist Karen Overall has argued that beyond CCD, additional canine conditions are dog models for human disorders, including canine dominance aggression and panic disorder as models for human impulse control disorder and panic disorder, respectively[137]. Of these, aggressiveness is among the most interesting, although there is no standard measure for the trait[144–146]. Hence, Van den Berg *et al.* developed a questionnaire to assess aggressive behaviour phenotypes in the Golden Retriever based on the previously published Canine Behavioral Assessment and Research Questionnaire (C-BARQ)[145,147]. The study showed that aggressive behaviour in Golden Retrievers can be divided into three phenotypes (stranger-, owner- and dog-directed aggression). However, the behaviour is not static, decreasing by 50% over a period of 4.3 years, indicating that ageing and/or owner interactions can alter behaviour. Using the same behavioural assessment tool, Zapata *et al.* conducted a GWAS for fear and aggression across 150 dogs of 11 breeds[148]. They identified statistically significant loci, but the same loci (*IGF1*, *HMGA2* and a locus at 105 Mb on the canine X chromosome) have previously been identified as relevant for body size, suggesting either a historical relationship between selection for body size and fear/aggression behaviours or that the tool failed to correctly phenotype the dogs[22,24,78]. In general, these results suggest that additional and refined assays are needed to classify phenotypes for behaviour and that behaviour is likely to be affected by genetic background and environment and is probably controlled by multiple genes. Yet, recent advances suggest that there is hope that these experiments can produce meaningful results. In one of the largest studies on aggressive behaviour conducted to date, 10,000 German Shepherd Dogs and Rottweilers were scored for 16 behavioural traits[149]. The authors found that over 50% of the additive genetic variance for the phenotyped traits could be explained by one principal component, which suggests that these two breeds share much of the genetics controlling their observed behaviour. These results are encouraging, as they suggest there is sufficient heritability to map aggression and that the trait can be accurately phenotyped.

Overall also reported that dogs could serve as a good model for Alzheimer disease (AD)[137]. The claim is based on observations that AD-like symptoms are observed in aged dogs (>10 years), including cognitive changes, anxiety, circadian rhythm disturbances, and reduced social interactions and activity levels[150,151]. Brains of aged dogs suffering cognitive disorders also share pathology with human patients with AD, suggesting that canine cognitive disorders and AD may be different presentations of the same disease[152,153]. Although the dog brains lacked classic dense neuritic plaques, amyloid-β deposits were present and showed similarities to the diffuse plaques and cerebrovascular deposits observed in early stages of human AD[154]. The claim that dogs experience AD is also supported by the observation that young and middle-aged dogs that performed poorly on standard cognition tests had high levels of amyloid-β in the cerebrospinal fluid, which is a marker for amyloid-β deposition in the brain[153]. Dogs also respond to some of the same therapies as human patients with AD[155]. In aggregate, these data provide compelling and exciting evidence that dogs are a viable model for studies of early AD in humans.

Breed-specific behaviours. Although genetic studies of truly anomalous behaviour have advanced, understanding the genetic underpinning of stereotypical breed behaviours has proven more difficult to study. Yet, breed behaviour is perhaps the most obvious phenotype that defines dog breeds. Since domestication, the number of required functions that dogs must perform has both grown and diversified, which has provided numerous independent models for the study of canine behaviour and neural function. Many behaviours such as hunting game, managing livestock and guarding property exist across many cultures and

geographic locations; hence, there are many breeds with common personality traits or behavioural proclivities that are slightly modified to meet specific needs[156,157]. For instance, 'pointing', the tendency for a dog to assume a motionless stance when in direct line of sight with a quarry to indicate its presence to a hunter, is not observed only in Pointers[158]; rather, it is partnered with retrieving in the Brittany, swimming in the German Shorthaired Pointer, and tracking in the English Setter. This presents a challenge for phenotype assessment and may render the cross-breed mapping approach ineffective. Other breed skills such as herding or hunting are not only difficult to score objectively with high reproducibility but also heavily influenced by training and environment, providing additional complexity.

Despite these challenges, recent mapping studies of breed behaviour show promising results. In a GWAS of 46 breeds, testing for association with dozens of traits, Vaysse *et al.*[24] identified a putative locus for canine hunting on chromosome 22. A previous study by Chase *et al.*[159], which attempted to map boldness, also identified a locus on chromosome 22, albeit about 22 Mb distant from the candidate locus highlighted in the Vaysse *et al.* study. However, Chase *et al.* used fewer genetic markers, as did a similar study by Jones *et al.*[98].

In the most recent and complete mapping study for boldness, Akkad *et al.*[160] used SNP-chip data and genome sequencing to perform homozygosity mapping, comparing the genomes of the Large Munsterlander and Weimaraner, which are pointing breeds, to the Berger des Pyrenees and Schapendoes, which are herding breeds. They found non-synonymous variants with likely functional consequences in the coding regions of the SET domain bifurcated 2 (*SETDB2*) gene, which is associated with the establishment of left–right asymmetry and may be particularly important for pointing dogs, and cysteinyl leukotriene receptor 2 (*CYSLTR2*), which is a G-protein-coupled receptor[160]. These genes are located within the region proposed previously by Vaysse *et al.* as important in boldness[24]. Additional putative behavioural loci include a boldness locus on canine chromosome 10 and a locus on the X chromosome[24] that is hypothesized to contribute to sociability. Finally, vonHoldt *et al.*[161] recently showed that selection during dog domestication targeted copy number variants (CNVs) in a locus that, in humans, is associated with Williams–Beuren Syndrome, which includes hyper-sociability as a major feature. Interestingly, both the chromosome 10 and X loci span regions containing body-size genes[22,78,81]. The pattern of body size and behavioural traits mapping together in behavioural scans was also observed by Chase *et al.*[159], and Jones *et al.*[98], who found an additional boldness locus on canine chromosome 15 near *IGF1* (REF. 98), although the meaning of this result is unclear. The locus on chromosome 10 is under selection for several traits[22,24,78] in addition to behaviour, perhaps explaining this association. However, the general theme of colocalization of body-size and behavioural loci elsewhere in the genome remains unexplained.

Introgression
Gene flow from one population or individual into the gene pool of another by repeated crosses between related individuals, resulting in individuals with genetic components from both initial populations.

Conclusions and future perspectives

Over the past several years, the number of sequenced canid genomes has rapidly increased, leading to fundamental advances in our understanding of dog domestication, many of which we have reviewed above. Comparative analyses of wolf and dog genomes have highlighted the importance of explicitly incorporating demographic models into tests for positive selection and have revealed the distribution of deleterious variation in dogs, arising both early and recently in dog history, including those that accumulated during breed-specific bottlenecks. Complementary to these broad surveys, functional studies focused on small numbers of genes have dissected the genetic architecture of phenotypic traits arising in breeds, including heritable disease.

Nevertheless, there remain large gaps in our understanding of evolutionary process and trait architecture. First, there is now abundant evidence that post-divergence gene flow between canid species is a common event. Yet, with the exception of introgression of black coat colour from dogs into wolves and hypoxic adaptation from Tibetan wolves, we have only a poor idea of the proportion of these introgression events that enhanced adaptation, let alone the mechanism by which fitness is increased, as we increasingly understand for admixture between humans and Neanderthals[162]. Second, demographically informed selection scans have now identified loci that are putatively involved in the early phenotypic divergence of dogs from wolves, but these candidates have yet to be functionally validated. Such validation will be practically and ethically challenging because their effects would, ideally, be assessed against a genomic background that is as wolf-like as possible. Creative approaches, some involving model organisms or genetically altered dog and wolf cell lines, will probably be necessary to make progress in this area. To date, functional studies have investigated traits with relatively simple genetic architecture, for example, dominant-effect deletions, transposable element insertions into genes, and non-synonymous substitutions altering protein function. We know far less about polygenic, quantitative traits, the interactions of multiple genes (that is, epistasis), and the gene-by-environment interactions that influence phenotypic outcomes. Similarly, we know little about pleiotropy, the multifarious effects on phenotypes caused by individual genes. At a broader, genomic scale, there is a need to collect, from individual samples, data on diverse genomic features beyond genetic polymorphism, including DNA methylation, gene expression, and chromatin interactions. Such assays will allow researchers to test hypotheses about the domestication process (for example, that mutations selected for early in domestication were variants that altered gene expression more than protein structure), as well as obtain more precise inferences about the architecture of specific phenotypic traits. As with humans, integration of such diverse functional data will help to identify the molecular basis of genetic diseases and aid in designing appropriate therapies. In addition, it will enrich our understanding of the history we have shared with dogs since they first began accompanying our hunter-gatherer ancestors.

1. Wayne, R. K. Cranial morphology of domestic and wild canids the influence of development on morphological change. *Evolution* **40**, 243–261 (1986).
2. Wayne, R. K. Limb morphology of domestic and wild canids: the influence of development on morphologic change. *J. Morphol.* **187**, 301–319 (1986).
3. Drake, A. G. & Klingenberg, C. P. Large-scale diversification of skull shape in domestic dogs: disparity and modularity. *Am. Nat.* **175**, 289–301 (2010).
4. Parker, H. G., Shearin, A. L. & Ostrander, E. A. Man's best friend becomes biology's best in show: genome analyses in the domestic dog. *Annu. Rev. Genet.* **44**, 309–336 (2010).
5. vonHoldt, B. M. *et al.* Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* **464**, 898–902 (2010).
The authors present a genetic similarity tree of dog breeds that suggests breed grouping by form and function and implies that cross breeding between these breed groups transfers key genetic mutations that are responsible for discrete phenotypic traits. They identify genes that may have been under selection in the origin of dogs.
6. Thalmann, O. *et al.* Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science* **342**, 871–874 (2013).
For the first time, ancient DNA of dogs and wolves was used to illuminate the timing and geographic origin of dogs, and it suggests that dogs originated from an extinct wolf population that existed in Europe about 27,000 years ago.
7. Shannon, L. M. *et al.* Genetic structure in village dogs reveals a Central Asian domestication origin. *Proc. Natl Acad. Sci. USA* **112**, 13639–13644 (2015).
8. Skoglund, P., Ersmark, E., Palkopoulou, E. & Dalen, L. Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. *Curr. Biol.* **25**, 1515–1519 (2015).
9. Wang, G. D. *et al.* Out of southern East Asia: the natural history of domestic dogs across the world. *Cell Res.* **26**, 21–33 (2016).
10. Fan, Z. *et al.* Worldwide patterns of genomic variation and admixture in gray wolves. *Genome Res.* **26**, 163–173 (2016).
11. Frantz, L. A. *et al.* Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science* **352**, 1228–1231 (2016).
12. Ash, E. C. *Dogs: Their History and Development* (E. Benn Limited, 1927).
13. American Kennel Club. *The Complete Dog Book* 20th edn (Ballantine Books, 2006).
14. Parker, H. G. *et al.* Genetic structure of the purebred domestic dog. *Science* **304**, 1160–1164 (2004).
15. Lindblad-Toh, K. *et al.* Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803–819 (2005).
This paper announced the first whole-genome sequence of a dog, the Boxer, together with a description of unique elements within the canine genome and predictions of ancestral sample size and SNP density needed for whole-genome association studies.
16. Boyko, A. R. *et al.* Complex population structure in African village dogs and its implications for inferring dog domestication history. *Proc. Natl Acad. Sci. USA* **106**, 13903–13908 (2009).
17. Freedman, A. H. *et al.* Genome sequencing highlights the dynamic early history of dogs. *PLoS Genet.* **10**, e1004016 (2014).
In this paper, the authors compare whole-genome sequences from several canine lineages, demonstrating that a severe bottleneck occurred in wolves soon after their divergence from dogs, implying that the pool of diversity that gave rise to dogs was substantially larger than that represented by modern wolf populations and narrowing the time frame for initial dog domestication to 11–16 thousand years ago.
18. Freedman, A. H. *et al.* Demographically-based evaluation of genomic regions under selection in domestic dogs. *PLoS Genet.* **12**, e1005851 (2016).
Explicit demographic models based on genetic data were used to develop a null distribution and better reduce the false positives for selective sweeps. The work shows selection on genes related to neurological, metabolic and phenotypic traits, but it does not confirm many of the genes proposed to be under selection in previous studies.
19. Marsden, C. D. *et al.* Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proc. Natl Acad. Sci. USA* **113**, 152–157 (2016).
20. Dreger, D. L. *et al.* Whole-genome sequence, SNP chips and pedigree structure: building demographic profiles in domestic dog breeds to optimize genetic-trait mapping. *Dis. Model. Mech.* **9**, 1445–1460 (2016).
21. Sutter, N. B. *et al.* Extensive and breed-specific linkage disequilibrium in Canis familiaris. *Genome Res.* **14**, 2388–2396 (2004).
22. Boyko, A. *et al.* A simple genetic architecture underlies morphological variation in dogs. *PLoS Biol.* **8**, e1000451 (2010).
23. Akey, J. M. *et al.* Tracking footprints of artificial selection in the dog genome. *Proc. Natl Acad. Sci. USA* **107**, 1160–1165 (2010).
24. Vaysse, A. *et al.* Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLoS Genet.* **7**, e1002316 (2011).
25. Hoopes, B. C., Rimbault, M., Liebers, D., Ostrander, E. A. & Sutter, N. B. The insulin-like growth factor 1 receptor (IGF1R) contributes to reduced size in dogs. *Mamm. Genome* **23**, 780–790 (2012).
26. Axelsson, E. *et al.* The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* **495**, 360–364 (2013).
These investigators used whole-genome resequencing of wolves and dogs to identify genes under selection during domestication, showing for the first time that amplifications of the amylase gene in early dogs provided a mechanism to survive on a starch-based diet.
27. Haywood, S. *et al.* Copper toxicosis in non-COMMD1 Bedlington terriers is associated with metal transport gene ABCA12. *J. Trace Elem. Med. Biol.* **35**, 83–89 (2016).
28. Cagan, A. & Blass, T. Identification of genomic variants putatively targeted by selection during dog domestication. *BMC Evol. Biol.* **16**, 10 (2016).
29. Decker, B. *et al.* Comparison against 186 canid whole-genome sequences reveals survival strategies of an ancient clonally transmissible canine tumor. *Genome Res.* **25**, 1646–1655 (2015).
This work describes the use of the largest multibreed whole-genome sequence data set to date, for filtering potential deleterious variants from millions to thousands and subsequent reduction to cancer pathways that are responsible for transmissible tumours.
30. Freedman, A. H. & Wayne, R. K. Deciphering the origin of dogs: from fossils to genomes. *Annu. Rev. Anim. Biosci.* **5**, 281–307 (2017).
31. Barton, N. H. The effect of hitch-hiking on neutral genealogies. *Genet. Res.* **72**, 123–133 (1998).
32. Barton, N. H. Genetic hitchhiking. *Phil. Trans. R. Soc. Lond. B* **355**, 1553–1562 (2000).
33. Crisci, J. L., Poh, Y. P., Mahajan, S. & Jensen, J. D. The impact of equilibrium assumptions on tests of selection. *Front. Genet.* **4**, 235 (2013).
34. Santiago, E. & Caballero, A. Variation after a selective sweep in a subdivided population. *Genetics* **169**, 475–483 (2005).
35. Slatkin, M. & Wiehe, T. Genetic hitch-hiking in a subdivided population. *Genet. Res.* **71**, 155–160 (1998).
36. Arendt, M., Cairns, K. M., Ballard, J. W., Savolained, P. & Axelsson, E. Diet adaption in dog reflects spread of prehistoric agriculture. *Heredity (Edinb.)* **117**, 301–306 (2016).
37. Parker, H. *et al.* Breed relationships facilitate fine-mapping studies: a 7.8-kb deletion cosegregates with Collie eye anomaly across multiple dog breeds. *Genome Res.* **17**, 1652–1571 (2007).
38. Anderson, T. M. *et al.* Molecular and evolutionary history of melanism in North American gray wolves. *Science* **323**, 1339–1343 (2009).
39. Li, Y. *et al.* Population variation revealed high-altitude adaptation of Tibetan mastiffs. *Mol. Biol. Evol.* **31**, 1200–1205 (2014).
40. vonHoldt, B., Fan, Z., Ortega-Del Vecchyo, D. & Wayne, R. K. *EPAS1* variants in high altitude Tibetan wolves were selectively introgressed into highland dogs. *PeerJ* **5**, e3522 (2017).
41. Lin, L., Vad-Nielsen, J. & Luo, Y. CRISPR-mediated multiplexed genetic manipulation. *Oncotarget* **7**, 80103–80104 (2016).
42. Shipman, P. *The Invaders: How Humans and Their Dogs Drove Neanderthals to Extinction* 3rd edn (Belknap Press, 2015).
43. Tonoike, A. *et al.* Copy number variations in the amylase gene (AMY2B) in Japanese native dog breeds. *Anim. Genet.* **46**, 580–583 (2015).
44. Ollivier, M. *et al.* Amy2B copy number variation reveals starch diet adaptations in ancient European dogs. *R. Soc. Open Sci.* **3**, 160449 (2016).
This paper presents ancient DNA data mapping the temporal origin of the increased amylase gene copy number using DNA from ancient dogs and shows that it first appeared over 7,000 years ago, concurrent with the beginning of agriculture.
45. Inchley, C. E. *et al.* Selective sweep on human amylase genes postdates the split with Neanderthals. *Sci. Rep.* **6**, 37198 (2016).
46. Reiter, T., Jagoda, E. & Capellini, T. D. Dietary variation and evolution of gene copy number among dog breeds. *PLoS ONE* **11**, e0148899 (2016).
47. Botigue, L. *et al.* Ancient European dog genomes reveal continuity since the early Neolithic. *Nat. Commun.* **8**, 16082 (2017).
48. Dreger, D. L. *et al.* Commonalities in development of pure breeds and population isolates revealed in the genome of the sardinian Fonni's Dog. *Genetics* **204**, 737–755 (2016).
49. Cetti, F. *I quadrupedi di Sardegna*. (ed. Edoardo Perino) (Gia Sassari, 1885).
50. Tyndale, J. W. *The Island of Sardinia, including pictures of the manners and customs of the Sardinians, and notes on the antiquities and modern objects of interst in the island: to which is added some account of the House of Savoy.* Vol. 2 (Richard Bentley, 1849).
51. Fiorito, G. *et al.* The Italian genome reflects the history of Europe and the Mediterranean basin. *Eur. J. Hum. Genet.* **24**, 1056–1062 (2015).
52. Gou, X. *et al.* Whole-genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-altitude hypoxia. *Genome Res.* **24**, 1308–1315 (2014).
53. Zhang, J. E. *et al.* Polymorphisms in the prion protein gene (PRNP) in the Tibetan Mastiff. *Anim. Genet.* **40**, 1001–1002 (2014).
54. Miao, B., Wang, Z. & Li, Y. Genomic analysis reveals hypoxia adaptation in the Tibetan mastiff by introgression of the grey wolf from the Tibetan Plateau. *Mol. Biol. Evol.* **34**, 734–743 (2017).
55. Akashi, H., Osada, N. & Ohta, T. Weak selection and protein evolution. *Genetics* **192**, 15–31 (2012).
56. Boyko, A. R. *et al.* Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet.* **4**, e1000083 (2008).
57. Keightley, P. D. & Eyre-Walker, A. Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics* **177**, 2251–2261 (2007).
58. Ohta, T. Role of very slightly deleterious mutations in molecular evolution and polymorphism. *Theor. Popul. Biol.* **10**, 254–275 (1976).
59. Eyre-Walker, A., Woolfit, M. & Phelps, T. The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics* **173**, 891–900 (2006).
60. Tennessen, J. A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64–69 (2012).
61. Nelson, M. R. *et al.* An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* **337**, 100–104 (2012).
62. Bellumori, T. P., Famula, T. R., Bannasch, D. L., Belanger, J. M. & Oberbauer, A. M. Prevalence of inherited disorders among mixed-breed and purebred dogs: 27,254 cases (1995–2010). *J. Am. Vet. Med. Assoc.* **242**, 1549–1555 (2013).
63. Schoenebeck, J. J. & Ostrander, E. A. Insights into morphology and disease from the dog genome project. *Annu. Rev. Cell Dev. Biol.* **30**, 535–560 (2014).
64. Karyadi, D. M. *et al.* A copy number variant at the KITLG locus likely confers risk for canine squamous cell carcinoma of the digit. *PLoS Genet.* **9**, e1003409 (2013).
65. Ostrander, E. A., Davis, B. W. & Ostrander, G. K. Transmissible tumors: breaking the cancer paradigm. *Trends Genet.* **32**, 1–15 (2016).
66. O'Neill, D. G. *et al.* Epidemiological associations between brachycephaly and upper respiratory tract disorders in dogs attending veterinary practices in England. *Canine Genet. Epidemiol.* **2**, 10 (2015).
67. Olsson, M. *et al.* A novel unstable duplication upstream of HAS2 predisposes to a breed-defining skin phenotype and a periodic fever syndrome in Chinese Shar-Pei dogs. *PLoS Genet.* **7**, e1001332 (2011).

68. Davis, B. W. & Ostrander, E. A. Domestic dogs and cancer research: a breed-based genomics approach. *ILAR J.* **55**, 59–68 (2014).
69. Boyko, A. R. The domestic dog: man's best friend in the genomic era. *Genome Biol.* **12**, 216 (2011).
70. Mirkena, T. *et al.* Genetics of adaptation in domestic farm animals: a review. *Livestock Sci.* **132**, 1–12 (2010).
71. Bertolini, F. *et al.* Evidence of selection signatures that shape the Persian cat breed. *Mamm. Genome* **27**, 144–155 (2016).
72. Gandolfi, B. *et al.* A splice variant in KRT71 is associated with curly coat phenotype of Selkirk Rex cats. *Sci. Rep.* **3**, 2000 (2013).
73. Lyons, L. A. *et al.* Aristaless-Like Homeobox protein 1 (ALX1) variant associated with craniofacial structure and frontonasal dysplasia in Burmese cats. *Dev. Biol.* **409**, 451–458 (2016).
74. Lipinski, M. J. *et al.* The ascent of cat breeds: genetic evaluations of breeds and worldwide random-bred populations. *Genomics* **91**, 12–21 (2008).
75. Parker, H. G. *et al.* An expressed fgf4 retrogene is associated with breed-defining chondrodysplasia in domestic dogs. *Science* **325**, 995–998 (2009).
76. Cadieu, E. *et al.* Coat variation in the domestic dog is governed by variants in three genes. *Science* **326**, 150–153 (2009).
77. Baranowska Körberg, I. *et al.* A simple repeat polymorphism in the MITF-M promoter is a key regulator of white spotting in dogs. *PLoS ONE* **9**, e104363 (2014).
78. Hayward, J. J. *et al.* Complex disease and phenotype mapping in the domestic dog. *Nat. Commun.* **7**, 10460 (2016).
Using the largest SNP-based data set to date, which encompassed 4,200 dogs genotyped with 180,000 SNPs, this paper summarizes selection data for dozens of morphological, behavioural and disease features.
79. Parker, H. G. *et al.* Genomic analyses reveal the influence of geographic origin, immigration and cross-breed introgression on modern dog breed development. *Cell Rep.* **19**, 697–708 (2017).
80. Shearin, A. L. & Ostrander, E. A. Canine morphology: hunting for genes and tracking mutations. *PLoS Biol.* **8**, e1000310 (2010).
81. Rimbault, M. *et al.* Derived variants at six genes explain nearly half of size reduction in dog breeds. *Genome Res.* **23**, 1985–1995 (2013).
82. Plassais, J. *et al.* Analysis of large versus small dogs reveals three genes on the canine X chromosome associated with body weight, muscling and back fat thickness. *PLoS Genet.* **13**, e1006661 (2017).
83. Sutter, N. B. *et al.* A single IGF1 allele is a major determinant of small size in dogs. *Science* **316**, 112–115 (2007).
84. Vrieze, S. I. *et al.* An assessment of the individual and collective effects of variants on height using twins and a developmentally informative study design. *PLoS Genet.* **7**, e1002413 (2011).
85. Lango, A. H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
86. N'Diaye, A. *et al.* Identification, replication, and fine-mapping of loci associated with adult height in individuals of African ancestry. *PLoS Genet.* **7**, e1002298 (2011).
87. Qu, B. H., Karas, M., Koval, A. & LeRoith, D. Insulin receptor substrate-4 enhances insulin-like growth factor-I-induced cell proliferation. *J. Biol. Chem.* **274**, 31179–31184 (1999).
88. Melkersson, K. & Persson, B. Association between body mass index and insulin receptor substrate-4 (IRS-4) gene polymorphisms in patients with schizophrenia. *Neuro Endocrinol. Lett.* **32**, 634–640 (2011).
89. Sun, Y. *et al.* Loss-of-function mutations in IGSF1 cause an X-linked syndrome of central hypothyroidism and testicular enlargement. *Nat. Genet.* **44**, 1375–1381 (2012).
90. Joustra, S. D. *et al.* IGSF1 variants in boys with familial delayed puberty. *Eur. J. Pediatr.* **174**, 687–692 (2015).
91. Asakura, Y. *et al.* Combined growth hormone and thyroid-stimulating hormone deficiency in a japanese patient with a novel frameshift mutation in IGSF1. *Horm. Res. Paediatr.* **84**, 349–354 (2015).
92. Ma, J. *et al.* Fine mapping of fatness QTL on porcine chromosome X and analyses of three positional candidate genes. *BMC Genet.* **14**, 46 (2013).
93. Cepica, S., Bartenschlager, H. & Geldermann, H. Mapping of QTL on chromosome X for fat deposition, muscling and growth traits in a wild boar x Meishan F2 family using a high-density gene map. *Anim. Genet.* **38**, 634–638 (2007).
94. Schoenebeck, J. & Ostrander, E. A. The genetics of canine skull shape variation. *Genetics* **193**, 317–325 (2013).
95. Bannasch, D. *et al.* Localization of canine brachycephaly using an across breed mapping approach. *PLoS ONE* **5**, e9632 (2010).
96. Marchant, T. W. *et al.* Canine brachycephaly is associated with a retrotransposon-mediated missplicing of SMOC2. *Curr. Biol.* **27**, 1573–1584.e6 (2017).
97. Schoenebeck, J. J. *et al.* Variation of BMP3 contributes to dog breed skull diversity. *PLoS Genet.* **8**, e1002849 (2012).
98. Jones, P. *et al.* Single-nucleotide-polymorphism-based association mapping of dog stereotypes. *Genetics* **179**, 1033–1044 (2008).
99. Holder, A., Mella, S., Palmer, D. B., Aspinall, R. & Catchpole, B. An age-associated decline in thymic output differs in dog breeds according to their longevity. *PLoS ONE* **11**, e0165968 (2016).
100. Fick, L. J. *et al.* Telomere length correlates with life span of dog breeds. *Cell Rep.* **2**, 1530–1536 (2012).
101. Bonnett, B. N., Egenvall, A., Hedhammar, A. & Olson, P. Mortality in over 350,000 insured Swedish dogs from 1995-2000: I. Breed-, gender-, age- and cause-specific rates. *Acta Vet. Scand.* **46**, 105–120 (2005).
102. Kraus, C., Pavard, S. & Promislow, D. E. The size-life span trade-off decomposed: why large dogs die young. *Am. Nat.* **181**, 492–505 (2013).
This work puts forth and defends the hypothesis that small dogs live longer than large dogs because large dogs simply age at an accelerated pace once senescence is initiated.
103. Selman, C., Nussey, D. H. & Monaghan, P. Ageing: it's a dog's life. *Curr. Biol.* **23**, R451–R453 (2013).
104. Dobson, J. M. Breed-predispositions to cancer in pedigree dogs. *ISRN Vet. Sci.* **2013**, 941275 (2014).
105. Vail, D. M. & MacEwen, E. G. Spontaneously occurring tumors of companion animals as models for human cancer. *Cancer Invest.* **18**, 781–792 (2000).
106. Adams, V. J., Evans, K. M., Sampson, J. & Wood, J. L. N. Methods and mortality results of a health survey of purebred dogs in the UK. *J. Small Anim. Pract.* **51**, 512–524 (2010).
107. Merlo, D. F. *et al.* Cancer incidence in pet dogs: findings of the Animal Tumor Registry of Genoa. *Italy. J. Vet. Intern. Med.* **22**, 976–984 (2008).
108. Khanna, C. *et al.* The dog as a cancer model. *Nat. Biotechnol.* **24**, 1065–1066 (2006).
109. Dorn, C. R. Epidemiology of canine and feline tumors. *Comp. Cont. Educ. Pract. Vet.* **12**, 307–312 (1976).
110. Cadieu, E. & Ostrander, E. A. Canine genetics offers new mechanisms for the study of human cancer. *Cancer Epidemiol. Biomarker Prev.* **16**, 2181–2183 (2007).
111. Ranieri, G. *et al.* A model of study for human cancer: spontaneous occurring tumors in dogs. Biological features and translation for new anticancer therapies. *Crit. Rev. Oncol. Hematol.* **88**, 187–197 (2013).
112. Dhawan, D. *et al.* Comparative gene expression analyses identify luminal and nasal subtypes of canine invasive urothelial carcinoma that mimic patterns in human invasive bladder cancer. *PLoS ONE* **10**, e0136688 (2015).
113. Knapp, D. W., Dhawan, D. & Ostrander, E. "Lassie, " "Toto, " and fellow pet dogs: poised to lead the way for advances in cancer prevention. *Am. Soc. Clin. Oncol. Educ. Book.* http://dx.doi.org/10.14694/EdBook_AM.2015.35.e667 (2015).
114. Ostrander, E. A. Franklin, H. Epstein Lecture. Both ends of the leash — the human links to good dogs with bad genes. *N. Engl. J. Med.* **367**, 636–646 (2012).
115. Rowell, J. L., McCarthy, D. O. & Alvarez, C. E. Dog models of naturally occurring cancer. *Trends Mol. Med.* **17**, 380–388 (2011).
116. Jónasdóttir, T. J. *et al.* Genetic mapping of a naturally occurring hereditary renal cancer syndrome in dogs. *Proc. Natl Acad. Sci. USA* **97**, 4132–4137 (2000).
117. Shearin, A. *et al.* The MTAP-CDKN2A locus confers susceptibility to a naturally occurring canine cancer. *Cancer Epidmiol Biomarkers Prev.* **21**, 1019–1027 (2012).
118. Thomas, R., Smith, K. C., Ostrander, E. A., Galibert, F. & Breen, M. Chromosome aberrations in canine multicentric lymphomas detected with comparative genomic hybridisation and a panel of single locus probes. *Br. J. Cancer* **89**, 1530–1537 (2003).
119. Phillips, J. C., Lembcke, L. & Chamberlin, T. A novel locus for canine osteosarcoma (OSA1) maps to CFA34, the canine orthologue of human 3q26. *Genomics* **96**, 220–227 (2010).
120. Affolter, V. K. & Moore, P. F. Localized and disseminated histiocytic sarcoma of dendritic cell origin in dogs. *Vet. Pathol.* **39**, 74–83 (2002).
121. Moore, P. F., Affolter, V. K. & Vernau, W. Canine hemophagocytic histiocytic sarcoma: a proliferative disorder of CD11d+ macrophages. *Vet. Pathol.* **43**, 632–645 (2006).
122. Fulmer, A. K. & Mauldin, G. E. Canine histiocytic neoplasia: an overview. *Can. Vet. J.* **48**, 1041–1050 (2007).
123. Dobson, J., Hoaher, T., McKinley, T. J. & Wood, J. L. Mortality in a cohort of flat-coated retrievers in the UK. *Vet. Comp. Oncol.* **7**, 115–121 (2009).
124. Bonnett, B. N., Egenvall, A., Olson, P. & Hedhammar, A. Mortality in insured Swedish dogs: rates and causes of death in various breeds. *Vet. Rec.* **141**, S40–S44 (1997).
125. Mueller, F., Fuchs, B. & Kaser-Hotz, B. Comparative biology of human and canine osteosarcoma. *Anticancer Res.* **27**, 155–164 (2007).
126. Karlsson, E. *et al.* Genome-wide analysis implicate 33 loci in heritable dog osteosarcoma, including regulatory variants near CDKN2A/B. *Genome Biol.* **14**, R132 (2013).
127. Gardner, H. L., Fenger, J. M., & London, C. A. Dogs as a model for cancer. *Annu. Rev. Anim. Biosci.* **4**, 199–222 (2016).
128. Decker, B. *et al.* Homologous mutation to human BRAF V600E is common in naturally occurring canine bladder cancer — evidence for a relevant model system and urine-based diagnostic test. *Mol. Cancer Res.* **13**, 993–1002 (2015).
129. Davies, H. *et al.* Mutations of the BRAF gene in human cancer. *Nature* **417**, 949–954 (2002).
130. Puntervoll, H. E., Molven, A. & Akslen, L. A. Frequency of somatic BRAF mutations in melanocytic lesions from patients in a CDK4 melanoma family. *Pigment Cell Melanoma Res.* **27**, 149–151 (2014).
131. DiMasi, J. A. & Grabowski, H. G. Economics of new oncology drug development. *J. Clin. Oncol.* **25**, 209–216 (2007).
132. Nowinsky, M. Zur Frage ueber die Impfung der krebsigen Geschwuelste. *Zentralbl Med. Wissensch* **14**, 790–791 (1876).
133. Murgia, C., Pritchard, J. K., Kim, S. Y., Fassati, A. & Weiss, R. A. Clonal origin and evolution of a transmissible cancer. *Cell* **126**, 477–487 (2006).
134. Strakova, A. & Murchison, E. P. The changing global distribution and prevalence of canine transmissible venereal tumour. *BMC Vet. Res.* **10**, 168 (2014).
135. Katzir, N., Arman, E., Cohen, D., Givol, D. & Rechavi, G. Common origin of transmissible venereal tumors (TVT) in dogs. *Oncogene* **1**, 445–448 (1987).
136. Murchison, E. P. *et al.* Transmissible dog cancer genome reveals the origin and history of an ancient cell lineage. *Science* **343**, 437–440 (2014).
137. Overall, K. L. Natural animal models of human psychiatric conditions: assessment of mechanism and validity. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **24**, 727–776 (2000).
138. Moon-Fanelli, A. A. & Dodman, N. H. Description and development of compulsive tail chasing in terriers and response to clomipramine treatment. *J. Am. Vet. Med. Assoc.* **212**, 1252–1257 (1998).
139. Moon-Fanelli, A. A., Dodman, N. H. & Cottam, N. Blanket and flank sucking in Doberman Pinschers. *J. Am. Vet. Med. Assoc.* **231**, 907–912 (2007). 140. Rapoport, J. L., Ryland, D. H. & Kriete, M. Drug treatment of canine acral lick. An animal model of obsessive-compulsive disorder. *Arch. Gen. Psychiatry* **49**, 517–521 (1992).
141. Tang, R. *et al.* Candidate genes and functional noncoding variants identified in a canine model of obsessive-compulsive disorder. *Genome Biol.* **15**, R25 (2014).
142. Dodman, N. H. *et al.* A canine chromosome 7 locus confers compulsive disorder susceptibility. *Mol. Psychiatry* **15**, 8–10 (2010).
143. Nazaryan, L. *et al.* Association study between CDH2 and Gilles de la Tourette syndrome in a Danish cohort. *Psychiatry Res.* **228**, 974–975 (2015).
144. Goodloe, L. P. & Borchelt, P. L. Companion dog temperament traits. *J. Appl. Anim. Welf. Sci.* **1**, 303–338 (1998).

145. van den Berg, L., Schilder, M. B., de Vries, H., Leegwater, P. A. & van Oost, B. A. Phenotyping of aggressive behavior in golden retriever dogs with a questionnaire. *Behav. Genet.* **36**, 882–902 (2006).
146. van den Berg, L., Schilder, M. B. & Knol, B. W. Behavior genetics of canine aggression: behavioral phenotyping of golden retrievers by means of an aggression test. *Behav. Genet.* **33**, 469–483 (2003).
147. Hsu, Y. & Serpell, J. A. Development and validation of a questionnaire for measuring behavior and temperament traits in pet dogs. *J. Am. Vet. Med. Assoc.* **223**, 1293–1300 (2003).
148. Zapata, I., Serpell, J. A. & Alvarez, C. E. Genetic mapping of canine fear and aggression. *BMC Genomics* **17**, 572 (2016).
149. Saetre, P. *et al.* The genetic contribution to canine personality. *Genes Brain Behav.* **5**, 240–248 (2006).
150. Schutt, T., Toft, N. & Berendt, M. Cognitive function, progression of age-related behavioral changes, biomarkers, and survival in dogs more than 8 years old. *J. Vet. Intern. Med.* **29**, 1569–1577 (2015).
151. Fast, R., Schutt, T., Toft, N., Moller, A. & Berendt, M. An observational study with long-term follow-up of canine cognitive dysfunction: clinical characteristics, survival, and risk factors. *J. Vet. Intern. Med.* **27**, 822–829 (2013).
152. Rofina, J. E. *et al.* Cognitive disturbances in old dogs suffering from the canine counterpart of Alzheimer's disease. *Brain Res.* **1069**, 216–226 (2006).
153. Borghys, H. *et al.* Young to middle-aged dogs with high amyloid-beta levels in cerebrospinal fluid are impaired on learning in standard cognition tests. *J. Alzheimers Dis.* **56**, 763–774 (2016).
154. Schutt, T. *et al.* Dogs with cognitive dysfunction as a spontaneous model for early alzheimer's disease: a translational study of neuropathological and inflammatory markers. *J. Alzheimers Dis.* **52**, 433–449 (2016).
155. Bosch, M. N., Gimeno-Bayon, J., Rodriguez, M. J., Pugliese, M. & Mahy, N. Rapid improvement of canine cognitive dysfunction with immunotherapy designed for Alzheimer's disease. *Curr. Alzheimer Res.* **10**, 482–493 (2013).
156. Neff, M. W. & Rine, J. A fetching model organism. *Cell* **124**, 229–231 (2006).
157. Spady, T. C. & Ostrander, E. A. Canine behavioral genetics: pointing out the phenotypes and herding up the genes. *Am. J. Hum. Genet.* **82**, 10–18 (2008).
158. American Kennel Club. Sporting Group. *American Kennel Club* http://www.akc.org/dog-breeds/groups/sporting/ (2017).
159. Chase, K., Jones, P., Martin, A., Ostrander, E. A. & Lark, K. G. Genetic mapping of fixed phenotypes: disease frequency as a breed characteristic. *J. Hered.* **100** (Suppl. 1), 37–41 (2009).
160. Akkad, D. A., Gerding, W. M., Gasser, R. B. & Epplen, J. T. Homozygosity mapping and sequencing identify two genes that might contribute to pointing behavior in hunting dogs. *Canine Genet. Epidemiol.* **2**, 5 (2015).
161. vonHoldt, B. M. *et al.* Structural variants in genes associated with human Williams-Beuren syndrome underlie stereotypical hypersociability in domestic dogs. *Sci. Adv.* **3**, e1700398 (2017).
162. Slatkin, M. & Racimo, F. Ancient DNA and human history. *Proc. Natl Acad. Sci. USA* **113**, 6380–6387 (2016).
163. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
164. Barrangou, R. Diversity of CRISPR-Cas immune systems and molecular machines. *Genome Biol.* **16**, 247 (2015).
165. Hendel, A. *et al.* Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat. Biotechnol.* **33**, 985–989 (2015).
166. Mali, P., Esvelt, K. M. & Church, G. M. Cas9 as a versatile tool for engineering biology. *Nat. Methods* **10**, 957–963 (2013).
167. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
168. Zou, Q. *et al.* Generation of gene-target dogs using CRISPR/Cas9 system. *J. Mol. Cell. Biol.* **7**, 580–583 (2015).
169. Mosher, D. S. *et al.* A mutation in the myostatin gene increases muscle mass and enhances racing performance in heterozygote dogs. *PLoS Genet.* **3**, e79 (2007).
170. Karlsson, E. K. *et al.* Efficient mapping of mendelian traits in dogs through genome-wide association. *Nat. Genet.* **39**, 1321–1328 (2007).
171. Whiteley, M. H., Bell, J. S. & Rothman, D. A. Novel allelic variants in the canine cyclooxgenase-2 (Cox-2) promoter are associated with renal dysplasia in dogs. *PLoS ONE* **6**, e16684 (2011).
172. Plassais, J. *et al.* A point mutation in a lincRNA upstream of GDNF is associated to a canine insensitivity to pain: a spontaneous model for human sensory neuropathies. *PLoS Genet.* **12**, e1006482 (2016).
173. Biesiadecki, B. J., Elder, B. D., Yu, Z. B. & Jin, J. P. Cardiac troponin T variants produced by aberrant splicing of multiple exons in animals with high instances of dilated cardiomyopathy. *J. Biol. Chem.* **277**, 50275–50285 (2002).
174. Wang, G. D. *et al.* Genetic convergence in the adaptation of dogs and humans to the high-altitude environment of the tibetan plateau. *Genome Biol. Evol.* **6**, 2122–2128 (2014).

Acknowledgements
We gratefully acknowledge the many individuals who provided comments and edits for this paper. We thank Dayna Dreger, Heidi Parker and Andrew Hogan for providing figures and valuable suggestions. E.A.O. and B.W.D. are supported by the Intramural Program of the US National Human Genome Research Institute. B.W.D. also acknowledges support from Texas A&M University. R.K.W. acknowledges support from the US National Science Foundation grants DEB 1021397 and 1257716 and seminal insights from John Novembre, Kirk Lohmueller and Claire Marsden, all of which have greatly enhanced this perspective on canine evolution.

Author contributions:
All authors contributed to researching data for the article and to writing and editing the manuscript. E.A.O., R.K.W. and A.H.F. substantially contributed to discussion of the content.

Competing interests statement
The authors declare no competing interests.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

DATABASES
Broad Institute CanFam3 Improved Annotation Data v1: https://data.broadinstitute.org/vgb/dog/dog/canFam3/tracksDb.txt
Dog 10K Genomes Project: http://www.dog10kgenomes.org
Dog Genome SNP Database (DoGSD): http://dogsd.big.ac.cn
VCF data from the whole genome sequence: https://data.broadinstitute.org/vgb/435_dog_data/

ALL LINKS ARE ACTIVE IN THE ONLINE PDF