



ANNUAL REVIEWS **Further**

Click [here](#) to view this article's online features:

- Download figures as PPT slides
- Navigate linked references
- Download citations
- Explore related articles
- Search keywords

Deciphering the Origin of Dogs: From Fossils to Genomes

Adam H. Freedman¹ and Robert K. Wayne²

¹Informatics Group, Faculty of Arts & Sciences, Harvard University, Cambridge, Massachusetts 02138

²Department of Ecology and Evolutionary Biology, University of California, Los Angeles, California 90095; email: rwayne@ucla.edu

Annu. Rev. Anim. Biosci. 2017. 5:281–307

First published online as a Review in Advance on November 28, 2016

The *Annual Review of Animal Biosciences* is online at animal.annualreviews.org

This article's doi:
10.1146/annurev-animal-022114-110937

Copyright © 2017 by Annual Reviews.
All rights reserved

Keywords

dog domestication, genomics, fossil record, phylogeny, human history

Abstract

Understanding the timing and geographic context of dog origins is a crucial component for understanding human history, as well as the evolutionary context in which the morphological and behavioral divergence of dogs from wolves occurred. A substantial challenge to understanding domestication is that dogs have experienced a complicated demographic history. An initial severe bottleneck was associated with domestication followed by postdivergence gene flow between dogs and wolves, as well as population expansions, contractions, and replacements. In addition, because the domestication of dogs occurred in the relatively recent past, much of the observed polymorphism may be shared between dogs and wolves, limiting the power to distinguish between alternative models of dog history. Greater insight into the domestication process will require explicit tests of alternative models of domestication through the joint analysis of whole genomes from modern lineages and ancient wolves and dogs from across Eurasia.

INTRODUCTION

The Focus and Organization of This Review

The domestic dog (*Canis lupus familiaris*) is both the first species and the only large carnivore ever to have been domesticated. Long after the initial domestication process and mostly within the last 200 years, dogs underwent a rapid phenotypic radiation characterized by the formation and maintenance of breeds through strong artificial selection and closed breeding systems imposed by humans (1, 2). Compared with other wild and domestic vertebrate species, modern dogs display a stunning phenotypic diversity. Dogs vary by two orders of magnitude in size, from one-pound teacup poodles to giant mastiffs weighing over 200 pounds. This would be the equivalent of a one-pound human being appropriately scaled to have the typical locomotor functions of an average adult. Beyond sheer size, body, limb, and skull proportions differ markedly among breeds, from the elongate, short-limbed, dolichocephalic dachshund to the similarly sized, brachycephalic pug, to the gracile whippet and the more wolflike Alaskan malamute. Given this variation, it is perhaps not surprising that quantitative comparisons to other carnivore groups suggest dogs display more phenotypic diversity than found in the entire carnivore order (3–5). Similarly, functional and behavioral diversity is striking among dogs, with certain breeds excelling in traits such as herding, retrieving, scent detection, and guarding (1). Over the last few decades, much attention has been devoted to elucidating the genetic basis of phenotypic variation among breeds for traits such as furnishings (elongated eyebrows and moustache), hair length and coat curling (6), body size (7–11), tail curl (9), skull shape (12), limb length (13), and behavior (11). Although the earliest dogs were certainly wolflike, it is less clear which phenotypic changes were the ones that coincided with the earliest stages of dog–wolf divergence (14).

Despite the many advances in elucidating the genetic basis of phenotypic traits and the increasing availability of genomic data for dogs (and other wild canids), the timing, geographic origins, and ecological context of dog domestication continue to remain controversial. First and foremost among reasons for the ongoing debate is that the divergence between the modern wolf population and dogs occurred over a very narrow time period (2, 15), such that the ancient branching events are poorly resolved. In particular, analysis is confounded by (a) incomplete lineage sorting (ILS), which is the discordance between a population phylogeny and individual loci owing to coalescence occurring before species divergence, and (b) postdivergence gene flow (16–18). Given the relatively recent origin of dogs, it is perhaps not surprising that ILS between dogs and wolves is common (15). Moreover, modern wolves may not be the direct ancestor of dogs. Rather, an extinct Late Pleistocene wolf may be their nearest common ancestor (15, 19). Further confounding phylogenetic reconstruction is reticulation caused by admixture between dogs and wolves through the history of domestication. Lastly, with only thousands to tens of thousands of generations having passed since domestication, the number of phylogenetically informative substitutions that can be used in dating or assigning ancestral populations is limited.

As confirmed by numerous independent genetic studies, the closest living relatives of dogs are gray wolves, with no evidence of other canid species having contributed to the genetic legacy of the domestic dog (19–22). However, beginning in the pregenomic era with single-locus mitochondrial DNA (mtDNA) and Y-chromosome studies and then genome-scale investigations, the evidence for the number and timing of domestication episodes and the geographic origins of dogs was conflicting. In assessing the current status of genetic evidence for competing hypotheses, we first review evidence from fossil and archeological remains. Not only do data on ancient canids provide a line of evidence independent from genetics, but genetic studies frequently rely on the fossil record to make temporal inferences. For example, first fossil appearance dates are used to estimate mutation rates

and divergence times. After reviewing evidence from fossils, we then focus on three main subject areas: (a) the geographic origin of dogs, including whether there were single or multiple domestication events; (b) the timing of dog domestication; and (c) the influence of dog–wolf admixture on dog history, and how it complicates inferences about timing and geographic origins. Within each of these research questions, we organize our discussion and evaluation of evidence by the scale of genomic sampling, starting with pregenomic-era studies of single (typically nonrecombining) loci and ending with analyses of whole-genome sequence data. Although the questions of geography and timing are intertwined, we first consider each separately in the discussion of genetic data. We then provide an integrated critical assessment of our current state of knowledge, focusing on broad issues that may complicate efforts to infer the timing of domestication and geographic origins. Finally, we discuss future research avenues necessary to address key gaps in our knowledge. As the oldest domesticated species, both the context and causes of dog domestication are obscure. Adding a definitive timetable and geographic context will likely allow more precise conclusions about the conditions that allowed for the early association of a large carnivore with hunter-gatherers. This early event was unique in the history of domesticated species, as all other domestication events were coincident with or postdated the development of agriculture approximately 10 kya. The early association of dogs and humans potentially allowed dogs to have a profound influence on the course of early human history and the development of civilization (23, 24).

Evidence from Fossils

Although fossil remains attributed to wolves have been found in association with hominids as early as 400 kya (25, 26), the appearance of dogs in the fossil record is much more recent, and current debates focus on whether dogs first appeared in the late Paleolithic or early Neolithic period. Larson et al. (14) highlighted three key issues with respect to the interpretation of ancient canid remains. First, the earliest dogs were likely very similar to wolves morphologically; thus, they cannot readily be distinguished from each other. Second, wolves once had a much broader distribution, making it difficult to classify remains solely on the basis of geography. Third, canid fossil remains are rare, and as a result, the record is temporally and geographically fragmented. Absent DNA evidence, another potential confounding factor is whether putative dog remains are in fact ancestral to modern dogs, or whether they might represent a domestication process that eventually failed (19). Similarly, as discussed above, doglike bones might originate from a smaller, morphologically distinct lineage of wolves that is now extinct (15, 19). Fossils classified as dogs are present on nearly every continent (**Figure 1**). The earliest dog remains are found in Europe at the Předmosti site in the Czech Republic (27), estimated to be ~27,000 years old, and at the Goyet site in Belgium, estimated to be ~36,000 years old (28). In the Levant, the earliest fossils are from Iraq, estimated to be 13,000 years old, and show evidence for tooth size reduction and crowding consistent with domestication (29). In Asia, the oldest remains are found in the Altai mountains in Russia, dating to approximately 33 kya (30). Unfortunately, East Asia has a poor record of ancient canines, with the oldest remains dating to 12–13 kya (**Figure 1**). Likewise, in the New World, where genetic evidence now indicates dogs originated from Old World dog lineages (31), fossil specimens are considerably more recent, with the oldest remains from Chile dated at ~9,000 years old (14).

Not surprisingly, subsequent analyses of many putative dog fossils have suggested they were misclassified and should instead be assigned to wolves (see tables S2 in 14 and S7 in 32). Re-analysis of fossils from the Goyet site concluded that they were not Paleolithic dogs but more likely wolves, perhaps from an extinct wolf lineage (33, 34; but see reply to these criticisms in 35, 36).

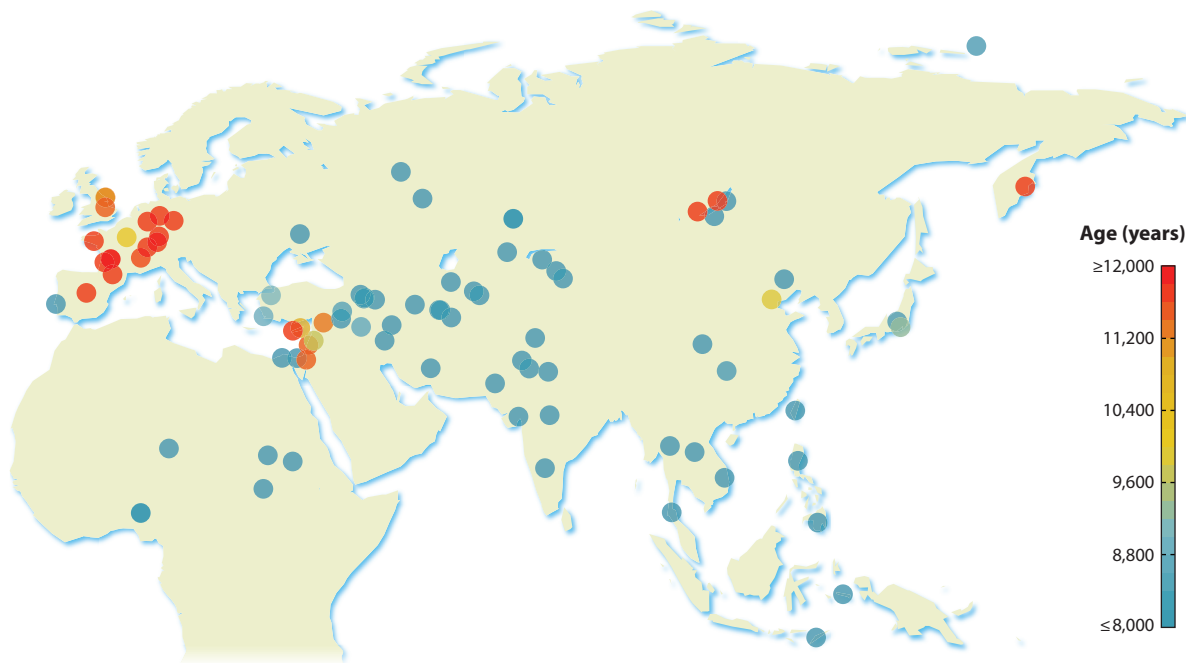


Figure 1

Geographic origins and age of the oldest validated fossil dog specimens in Eurasia. Points are dog fossil sites, with colors indicating their age. Modified from Frantz et al. (32) with permission from AAAS.

Similarly, a recent 3D geometric morphometric analysis (37) indicated significant morphospace overlap between wolves and putative dogs from the Goyet and Eliseevichi sites (Russia, dated to ~17,000 BP), supporting their assignment to wolves. These putative misclassifications may be due to the diversification of wolves into sympatric environment- and diet-based ecotypes (38). In contrast, analysis of ancient DNA has in at least one case confirmed the original classification of ancient remains as a dog. In this case, although Larson et al. (14) disputed the assignment of Altai Mountains specimens to dogs (30) and suggested they might be extinct wolves, a genetic analysis of mtDNA is consistent with the claim that they are, in fact, dogs (39). Outside of the Altai Mountains, the oldest remains in East Asia are from approximately 12–13 kya, found in Kamchatka, Russia, and northern China (**Figure 1**). These remains are younger than the oldest undisputed fossils in Europe (8–19 kya) and comparable in age to the oldest remains found in the region including the Levant, Anatolia, and Central Asia (5–13 kya) (**Figure 1**, reviewed in 14, 32). Collectively, these studies suggest dogs likely originated in Eurasia as early as 33 kya. In addition, the fossil record suggests a complicated evolutionary history, one potentially containing both morphologically doglike wolves and wolflike dogs arranged along a phenotypic continuum. Understanding how these diverse ancient canids contributed to the genetic legacy of contemporary dogs will require analyses of DNA from fossil specimens used for morphological studies. In fact, if the earliest dogs followed modern humans and scavenged the carcasses they left behind (23, 24), early selection may have favored a wolflike morphology. Perhaps not until humans became more sedentary, and dogs more closely associated with them, was there selection for smaller, phenotypically distinct canids, even if the occurrence of reduced body size in dogs may have predated agriculture (fossil evidence reviewed in Reference 14).

GEOGRAPHIC ORIGINS

MtDNA and Y Chromosome Studies

Although not inferring a geographic origin, an important benchmark for subsequent work was Vilà et al.'s (21) phylogenetic analysis of dogs and wolves. They constructed a global phylogeny of both species using mtDNA control region (CR) sequence data and identified four major clades containing dog haplotypes (**Figure 2a**). Notably, a clade containing the majority of dog sequences was monophyletic (clade I) to the exclusion of wolves, but the three other clades (II–IV) contained both dog and wolf haplotypes. These data implied either multiple independent origins or, as the authors suggest, an initial domestication event forming clade I, followed by repeated rounds of hybridization and selection for doglike phenotypic variation.

Evidence for specific geographic scenarios of domestication has relied on arguments concerning genetic diversity and phylogenetic relationships. The rationale is that a region with the greatest diversity and that includes the largest proportion of observed haplotypic diversity is likely to be a center of domestication. A corollary to this argument is that centers of origin should also contain a greater proportion of geographically restricted genetic variation. The first study to infer a geographic origin for dogs relied primarily on a diversity argument. Sequencing 582 bp of mtDNA CR in 654 domestic dogs from Europe, Asia, Africa, and Arctic America, as well as 38 Eurasian wolves, Savolainen et al. (40) constructed a phylogeny that recovered clades I–IV of Vilà et al. (21), designating them clades A–D, as well as two additional clades (E, F) containing dog haplotypes. The frequencies of clades A–C across geographic regions were used to argue that they constituted a common gene pool from which dogs originated, a claim subsequently reiterated by Pang et al. (41). Unlike in Vilà et al.'s work, clade A was found to contain wolf haplotypes from China and Mongolia, suggesting an origin of this clade in East Asia. However, given the high mobility and minimal global population structure in wolves, the authors relied on other evidence to argue for an East Asian origin. First, genetic diversity was higher in East Asia than in western Eurasia. In particular, mean sequence divergence of clade-A individuals was higher in East Asia compared with in Southwest Asia and Europe (**Figure 2b**). Second, after correcting for sample size differences, dogs in East Asia contained more haplotypes than found elsewhere. However, no significant difference was found between the number of haplotypes found in Southwest Asia and the number in Europe. Third, East Asia contained a greater number of unique haplotypes. Fourth, unique haplotypes in the east, on average, differed by more mutations from western haplotypes than unique western haplotypes did from eastern ones (see figure 2b in Reference 40). These findings suggested a greater antiquity of haplotypes in East Asia.

Using Savolainen et al.'s (40) 582-bp CR, Pang et al. (41) assessed patterns of mtDNA diversity for 1,543 dogs from across the Old World, sampling either indigenous village dogs or breeds with known geographic origins. In this and subsequent studies, village dog genetic variation is sampled because they presumably contain indigenous ancestry more germane to disentangling dog origins than modern breed dogs. They found that 97.4% of all dogs had haplotypes from A, B, or C clades and that the frequency of these three clades was similar across geographic regions within the Old World, indicating a common-source gene pool. They inferred a geographic origin for dogs in Asia southeast of the Yangtze River (ASY) based on comparisons of unique genetic diversity among regions.

Pang et al. (41) found that the majority of individuals had a universal CR haplotype (i.e., found in all geographic regions, hereafter designated as UT). However, they found that ASY had a higher proportion of region-specific haplotypes (**Figure 3**). They also found a cline in the frequency of UTs, highest in western Eurasia, decreasing in East Asia, and lowest in ASY

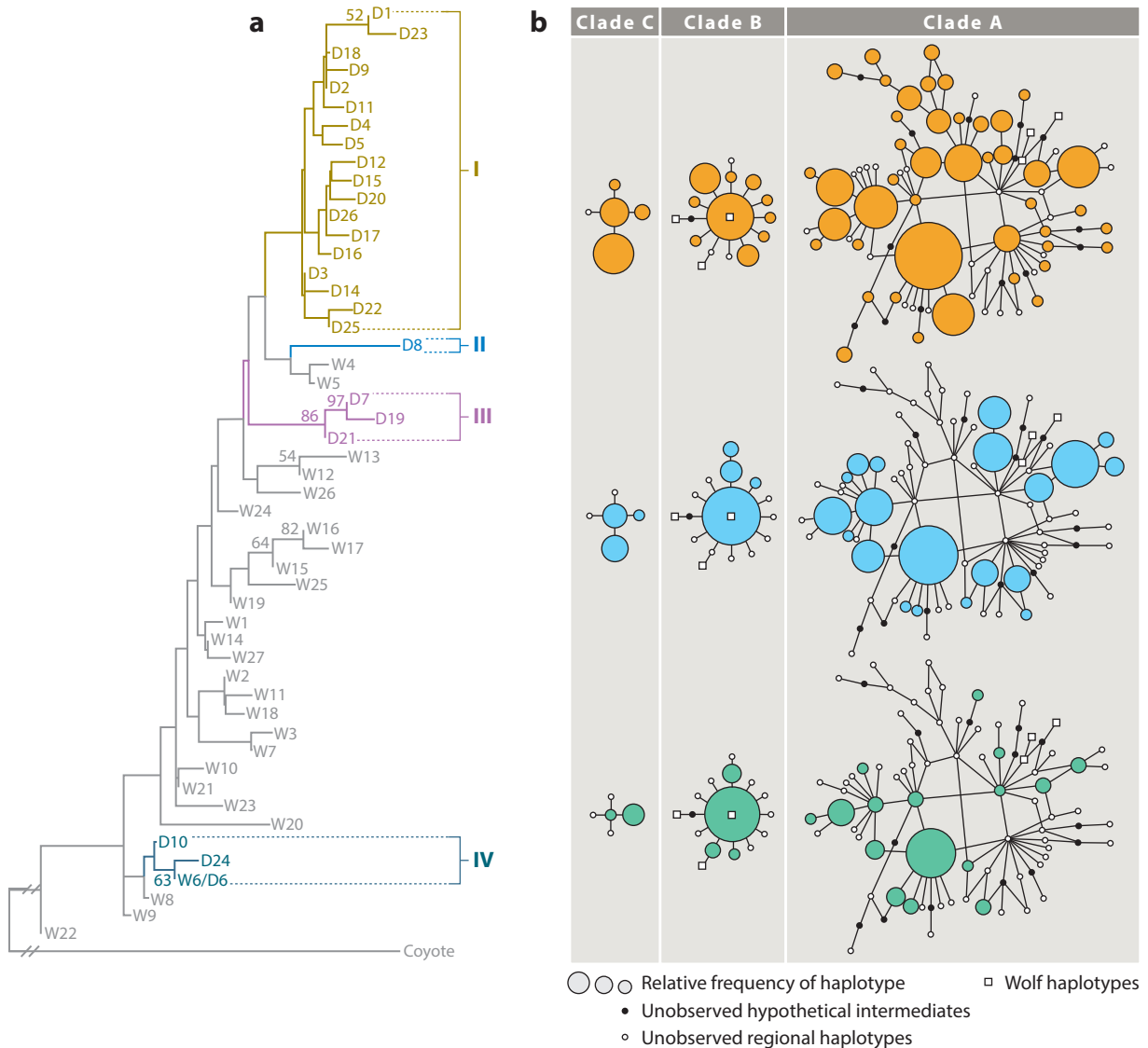


Figure 2

(a) Neighbor-joining tree of wolf and dog mitochondrial DNA (mtDNA) control region haplotypes (21), modified with permission from AAAS. (b) Minimum-spanning network of mtDNA control region haplotypes from dogs for the three major identified clades (40). Haplotypes found in East Asia, Europe, and Southwest Asia are displayed in separate networks, colored orange, blue, and green, respectively. Clades I, II, and III in panel a are analogous to A, B, and C in panel b. Modified with permission from AAAS.

(Figure 3). Furthermore, they found that most dogs in Europe and western Asia contained either a UT or a haplotype separated from a UT by one mutation (dUT), and that the dUT frequencies were lowest in ASY. These data, and the interpretation that every haplotype in western Eurasia can be traced to one in East Asia, was used to argue for an origin of dogs in ASY. Expanding this analysis by sampling whole mtDNA genomes from 169 dogs from CR clades A, B, and C, they identified 10 major haplogroups, all of which were found in ASY. In contrast, six haplogroups

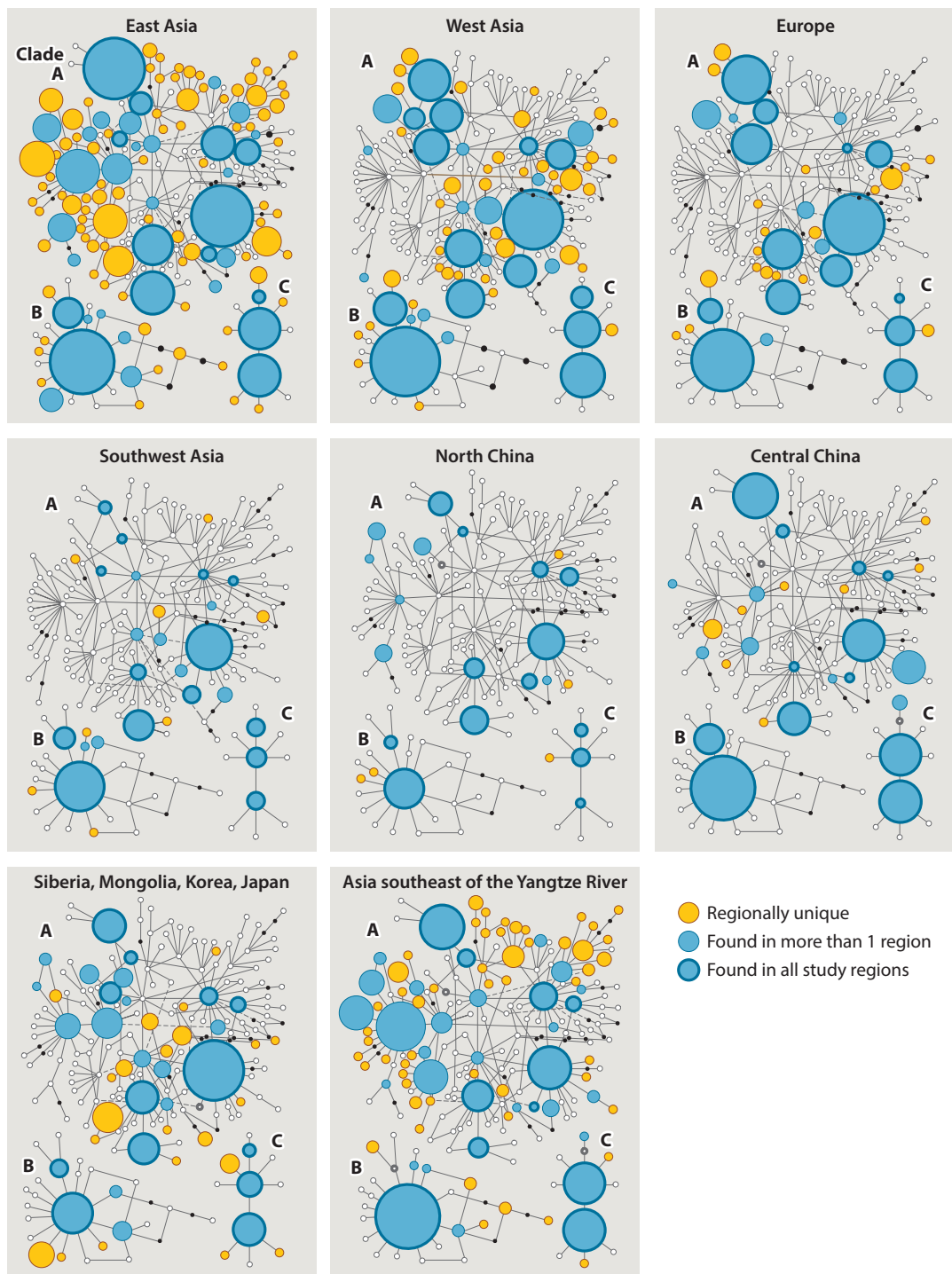


Figure 3

Minimum-spanning networks for mitochondrial DNA (mtDNA) haplotypes in the three major dog clades, A, B, and C, by region. Modified from Reference 41 under a Creative Commons Open Access license.

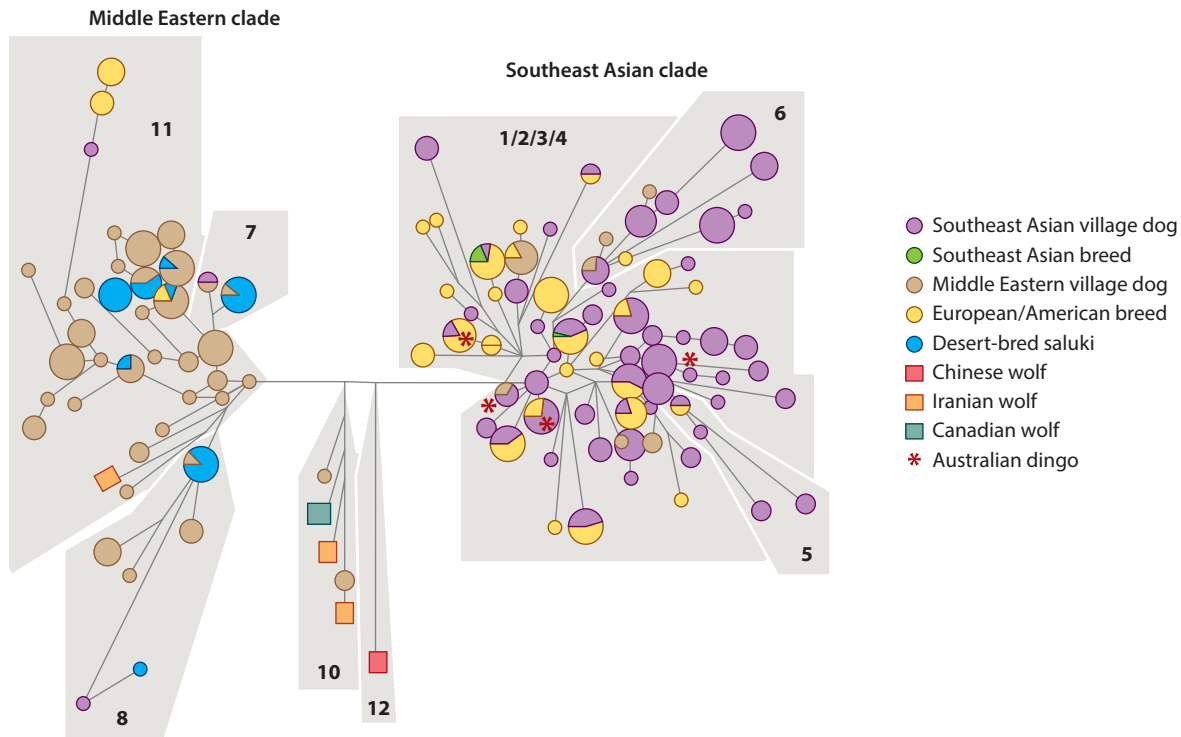


Figure 4

Village and breed dog Y chromosome SNP-STR haplotype network for 300 village dogs/dingoes, 124 breed dogs, and 7 wolves. Adapted from Brown et al. (42) under a Creative Commons Open Access license.

were virtually absent from western Eurasia. In sum, these diversity-based findings led the authors to conclude that dogs originated in ASY.

In a follow-up study, Brown et al. (42) conducted intensive sampling of village dogs from Southeast Asia (including islands that had once been connected to the mainland) and the Middle East, as well as a smaller sample of breed dogs, to evaluate evidence for an ASY origin versus Middle East origin as proposed by vonHoldt et al. (22). Brown et al. (42) not only analyzed mtDNA hypervariable region I (D-loop) but also included 5 single-tandem repeats and 11 single-nucleotide polymorphisms (SNPs) from the nonrecombining region of the Y chromosome. Their analyses were aimed at comparing patterns of genetic diversity between the two regions, under the same assumption made by Pang et al. (41) that higher diversity likely reflects the geographic source of dogs. Specifically, they investigated patterns of genetic diversity in three ways, two of which echoed previous methods (41). First, they calculated gene diversity summary statistics. Their D-loop sequences produced a pattern of gene diversity nearly identical to that of Pang et al. (41), despite different geographic sampling, and showed that gene diversity was lower in the Middle East compared with Southeast Asia (Middle East = 0.87, $n = 199$; Southeast Asia = 0.92, $n = 253$) (**Figure 4**). Second, they compared the frequency of trans-regional universal Y chromosome haplotypes between regions. The proportion of universal Y chromosome haplotypes represented a higher proportion of Middle East haplotypes than in Southeast Asia, and there were three times as many novel haplotypes discovered in Southeast Asia as in the Middle East, despite extensive sampling of the former in previous studies. Third, for regional subclusters of

Y chromosome haplotypes within a minimum-spanning network, they quantified the average number of mutational steps (ρ) between ancestral and descendant nodes for each region. In this analysis, the authors equate greater ρ with higher diversity. Moreover, they indirectly infer geographic origins based on timing, suggesting that the dog haplotype subnetwork with the greatest ρ not only is more diverse but also must be older. Following this logic, and to the extent that it is geographically localized, the oldest subnetwork represents the geographic origin for dogs. Thus, as with their other diversity-based arguments, the authors interpreted the greater ρ for dogs from Southeast Asia than the Middle East, as consistent with an origin in the former, despite the fact that the genetic distances of the regional dog clades to any wolf clade are nearly identical, and the 95% confidence intervals for ρ for the two regions are overlapping. This issue aside, the authors present a novel interpretation of the observation that most European dog diversity is contained within Southeast Asia. Rather than interpret this as further evidence for a Southeast Asian origin of dogs, they instead suggest that this reflects a replacement of indigenous European dog ancestry by more recent immigrant lineages from Southeast Asia. This replacement perhaps accompanied a general geographic expansion of East Asian dog populations or, alternatively, human-mediated transport owing to trade. In fact, admixture and lineage replacement figure prominently in subsequent genome-scale investigations (see section titled Genotyping Arrays and Whole-Genome Sequencing below). Subsequently, an ASY origin of dogs was advanced using Y chromosome data (43) by applying the same diversity-based arguments described above for mtDNA CR sequence data.

Genomic Approaches

SNPs and microsatellites. In the first effort to survey genome-wide genetic variation in dogs and assess population structure, Boyko et al. (44) not only sampled mtDNA D-loop sequence variation but typed 223 African village dogs from Uganda, Namibia, and Egypt for 89 microsatellites and 300 SNPs. In addition, they genotyped a smaller sample of Puerto Rican street dogs presumed to represent highly admixed lineages. Microsatellite loci matched those typed in 126 breed dogs used previously to quantify breed population structure. Principal component analysis (PCA) and STRUCTURE analyses found that >84% of village dogs outside of central Namibia showed little to no evidence for admixture. Admixture in central Namibia and near Giza, Egypt, presumably reflected mixing that occurred as a result of the arrival of nonindigenous dogs with white settlers south of the agricultural cordon dividing the country. These indigenous village dogs had higher genetic diversity than breed dogs. More germane to the domestication origin question, mtDNA diversity in African village dogs was found to be comparable to that in East Asia, particularly when controlling for sample size. As the domestic dog was known to not originate in Africa, this finding was used to call into question the Southeast Asian origin of dogs.

VonHoldt et al. (22) expanded genomic sampling further, analyzing polymorphism in 48,036 autosomal SNPs for 912 breed dogs, 225 globally distributed wolves, and 60 coyotes. Several breeds thought to be ancient, whose lineages extended well before the formation of modern breeds, were found to be basal in both allele-sharing and haplotype-based phylogenies (**Figure 5a**). More importantly, in assessing the fraction of unique haplotypes shared between dogs and wolves, Middle Eastern wolves consistently shared a greater fraction of dog haplotypes than any other wolf population (**Figure 5b**), including a panel of Chinese wolves that would have been the putative source of dogs if they had first been domesticated in Southeast Asia. Although not explicitly claiming that dogs were first domesticated in the Middle East, the authors concluded that Middle Eastern wolves contributed a significant amount of genomic diversity to modern dog breeds, and that contributions from European and Chinese wolves were more breed specific. In a related study examining the geographic origin of the interleukin growth factor 1 haplotype associated with small dog

breeds, Gray et al. (45) found the haplotype to be phylogenetically grouped with Middle Eastern wolves. As the haplotype conferred reduced size in the majority of small dogs, it was argued that it must have originated early in dog history and in the Middle East, although this finding does not exclude an early origin for dogs elsewhere. However, the inferred ancient origin of the small dog haplotype in the Middle East is consistent with origination and subsequent diversification there.

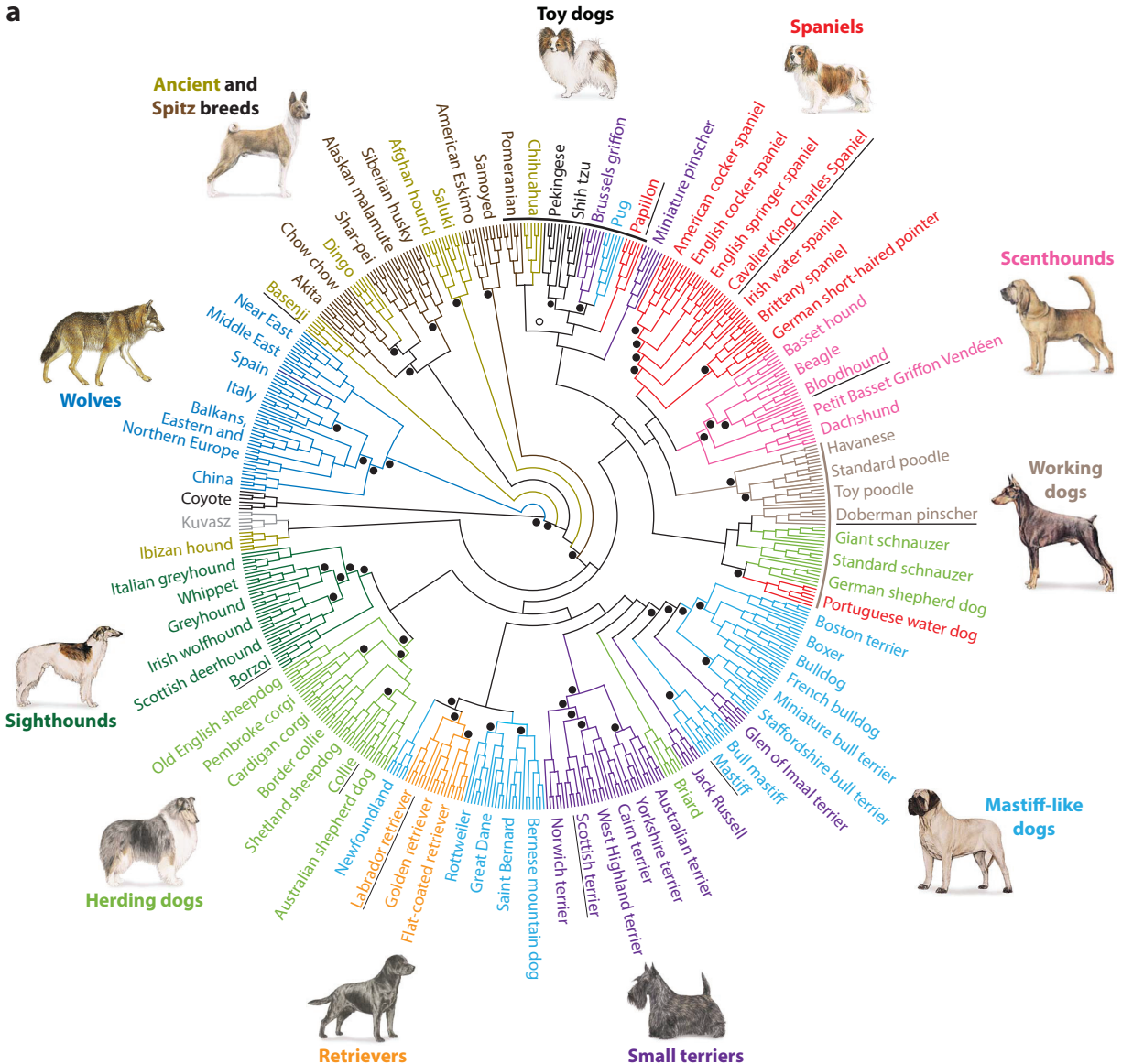


Figure 5

(a) Haplotype-sharing cladogram for 10-SNP windows and (b) fraction of unique haplotypes shared between 64 dog breeds and wolf populations for 5-SNP (left) and 15-SNP (right) windows. Diamonds indicate significant sharing ($P < 0.05$); from vonHoldt et al. (22), with permission from Nature Publishing Group.

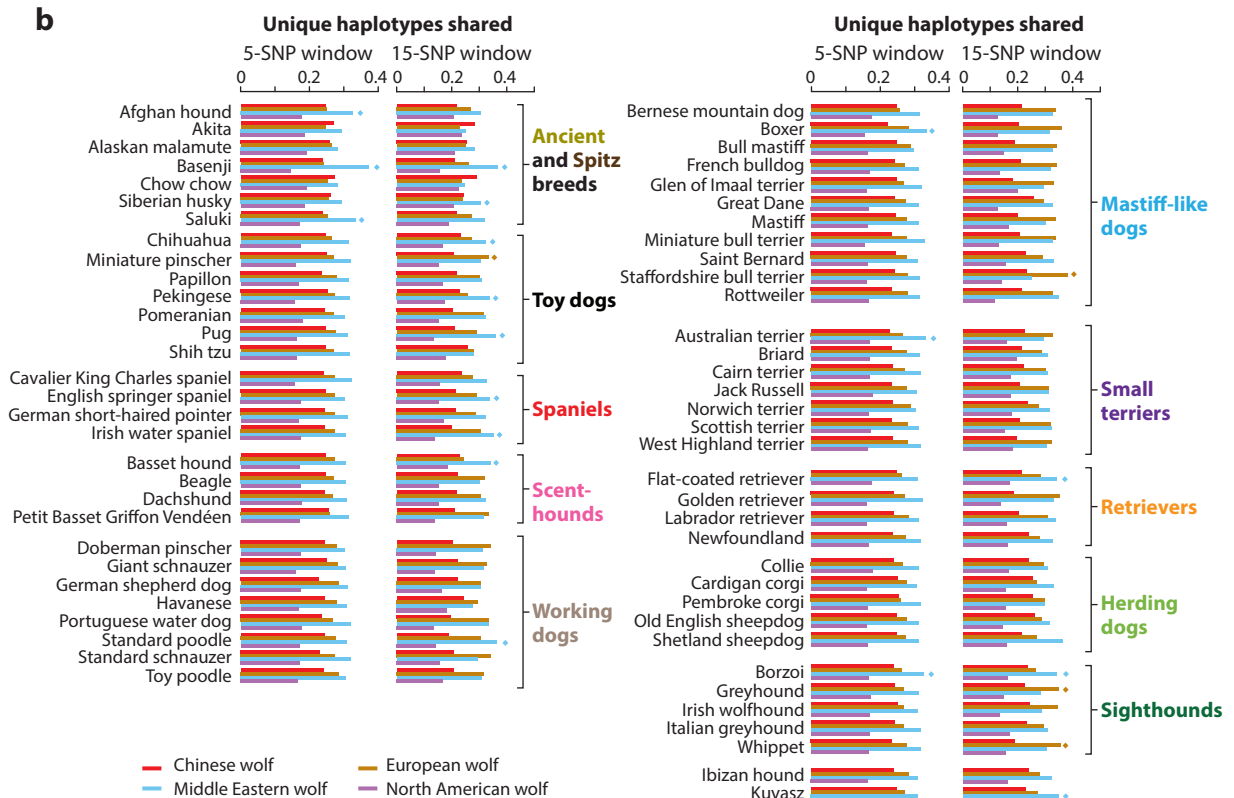


Figure 5

(Continued)

Whole mitochondrial genomes. As we discuss in detail below, perhaps the best approach for clarifying the timing and geographic origin of dogs is by interrogating genetic variation of both contemporary and ancient dog and wolf lineages. An ancient DNA perspective can capture domestication events lost owing to extinction of ancestral wolf populations and may clarify the taxonomic status of fossils originally classified on the basis of dating and morphology alone. The first comprehensive assessment of genetic variation from contemporary and ancient samples analyzed whole mtDNA genomes (19). This study shed substantial insight on two fronts relevant to the geography of dog domestication. First, this research highlighted the need for validation of inferences based on morphologic analysis of fossils alone. The ancient (>20 kya) fossil Belgian canids, including the Goyet dog, were found to be a sister group to all contemporary dogs and wolves, suggesting they represent an aborted domestication event or morphologically distinct form of wolf. Second, although dogs and wolves were not reciprocally monophyletic clades, the A, C, and D clades encompassed 78% of the dog mtDNA genomes sampled and were each sister to an ancient wolf or dog lineage found in Europe (Figure 6). From this observation, and the fact that clade-B dogs are sister to contemporary European dogs, the authors concluded that dogs likely originated in Europe, probably from a now-extinct wolf lineage more than 20 kya.

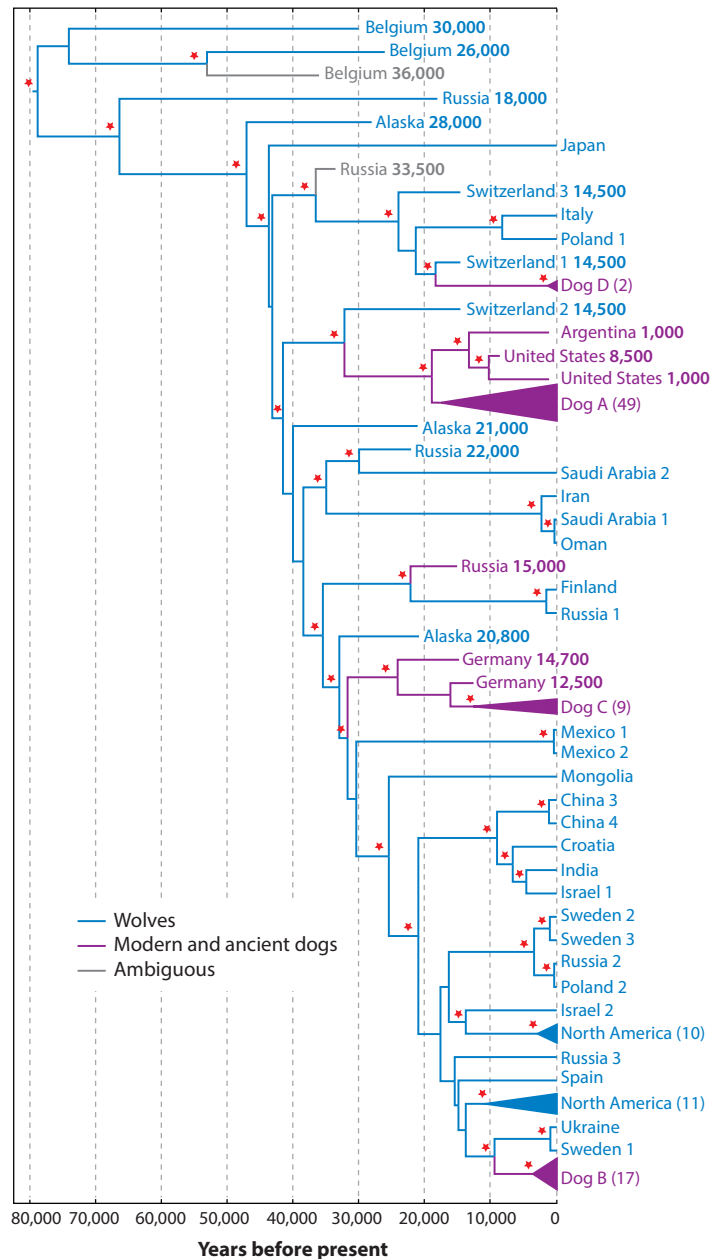


Figure 6

Dated phylogeny of modern and ancient dogs (*violet*) and wolves (*blue*) based upon whole mitochondrial DNA (mtDNA) genomes, from Thalmann et al. (19). For ancient samples, fossil age estimates are provided in years before present. For monophyletic clades, numbers in parentheses indicate the number of samples. Fossils with ambiguous classification are colored gray. Red stars indicate maximum likelihood bootstrap support of >90% and Bayesian posterior support of >0.9. Reprinted with permission from AAAS.

Genotyping arrays and whole-genome sequencing. Freedman et al. (15) employed an experimental design to avoid geographic sampling bias that was a departure from previous model-free studies of mtDNA, Y chromosome, or genome sequence data. For example, the study of Wang et al. (46) based on genome sequence data assumed a Southeast Asian origin rather than treating it as a hypothesis to be tested. Freedman et al. (15) sequenced two divergent and basal dog breeds, basenji and dingo, whose geographic origins trace to the Middle East/Africa and Southeast Asia, respectively, representing two of the major hypothesized centers of dog domestication. They also used the genome sequence of the reference boxer, which originated in Europe, a third hypothesized center of domestication, based upon the fossil record and analysis of whole mtDNA genomes (19). Genomes from a Croatian wolf (Europe), Chinese wolf (Southeast Asia), and Israeli wolf (Middle East) were also sequenced. As an outgroup, they sequenced a golden jackal (also from Israel). With the exception of the basenji (12×), all other canids were sequenced to >20× coverage, depth sufficient to call heterozygous sites. The authors tested five alternative models of domestication: (a) reciprocal monophyly of dogs and wolves [reflecting the phylogeny reconstructed from so-called autosomal SNPs, and consistent with vonHoldt et al. (22)], (b) a regional model in which each dog arose from the geographically corresponding wolf population (e.g., dingo from Chinese wolf, basenji from Israeli wolf), and (c) three single-wolf population origin models in which all dogs originated from one of the three wolf populations (**Figure 7**). To fit the polymorphism data, the regional model required unusually high amounts of gene flow between nearly all dog–dog, wolf–dog, and wolf–wolf pairs (**Figure 7b**); the single-origin models all compressed the lineage divergences to be nearly instantaneous (**Figure 7c**). In addition, to evaluate the relative support for the different models, the authors used a method representing an extension of the ABBA-BABA tests of Durand et al. (47). The original method tests for post-divergence gene flow by considering situations where three samples (P1, P2, and P3) are bi-allelic, where an outgroup is in state A, and where P3 is a source lineage in state B that can contribute an allele via gene flow to lineages P1 and P2. Asymmetries of ABBA and BABA allelic configurations among P1, P2, and P3 are indicative of asymmetric gene flow. Freedman et al. (15) extended this method by examining these two configurations as well as BBAA among quartets of samples that contain the jackal outgroup and quantified discordance between the frequencies of these allelic configurations and those simulated under the three tested models of domestication. Although all models were consistent with the data, the discordance rate was the lowest for the model in which dogs and wolves are reciprocally monophyletic. This finding, and the empirical values of gene flow and the distribution of divergence times, led the authors to conclude that reciprocal monophyly was the best-fitting model. Under this model, there were modest amounts of gene flow between dog and wolf lineages that were regionally co-occurring (**Figure 7a**). Based upon this model, they concluded that none of the extant wolf populations from the three hypothesized centers of origin was the source of dog domestication and, instead, that dogs may have initially diverged from a now-extinct wolf population. Although they did not specify a particular geographic origin, this conclusion was consistent with the previous mtDNA genome study of ancient and contemporary dogs and wolves (19), which suggested a now-extinct European wolf population as the source for the majority of dog lineages.

As part of a broader study examining patterns of genetic diversity and admixture in wolves, Fan et al. (20) also reevaluated the findings of Freedman et al. (15) using an expanded set of canine genomes sequenced at 9–28× coverage. Including the samples of Freedman et al. (15), they analyzed 34 genomes, including 20 wolves from across Eurasia, 4 from North America, 6 dogs, an Israeli golden jackal, a Kenyan golden jackal, and a coyote. Their canid phylogeny based upon autosomal SNPs recovered reciprocal monophyly of Old World wolves and dogs observed by Freedman et al. (15) and was consistent with the inference that no extant wolf population

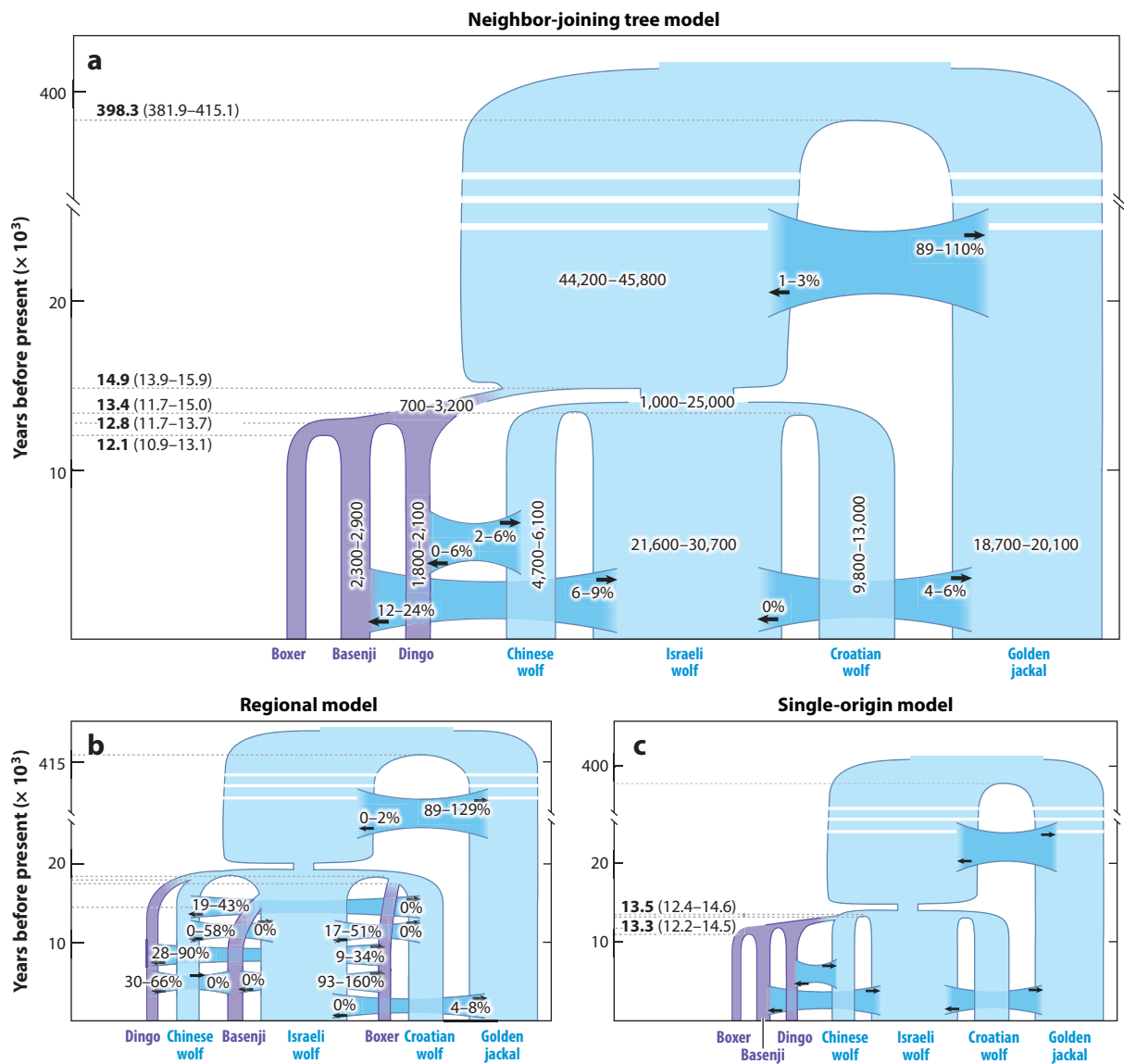


Figure 7

Demographic models of domestication evaluated using whole-genome sequence data from dogs and wolves, from Freedman et al. (15). Divergence times, population sizes, and migration bands are inferred using *G-PhoCS*. (a) Model based upon neighbor-joining tree reconstructed with autosomal SNPs, in which dogs and wolves are reciprocally monophyletic and no modern wolf population is the putative source of all dogs. (b) A regional model in which boxer, basenji, and dingo originate from geographically proximate wolf populations, requiring excessive migration rates to explain data. (c) A single-origin model with dogs originating from the Israeli wolf, which requires divergence times to occur close together, i.e., a star phylogeny; similar results obtained for single origins for the Chinese wolf and the Croatian wolf (see 15 for detail). Adapted under a Creative Commons Open Access license.

is the source of dogs. This phylogeny was also consistent with the known time course for wolf colonization of the New World, as a clade composed of Yellowstone and Mexican gray wolves diverged first from one including Eurasian wolves and dogs (figure 3 in Reference 20). This finding also supports assigning dogs and gray wolves to the same species, as to do otherwise would make wolves a paraphyletic group. Consistent with the previous sequence study (15), both ABBA-BABA tests and *G-PHoCS* used consistently found evidence for postdivergence gene flow between regionally cooccurring populations of dogs and wolves. This globally complex pattern of admixture has the potential to confound inference of a specific region as being the source of dogs (see section titled The Admixture Problem below).

Although Freedman et al. (15) showed both the feasibility and the benefits to be gained from testing specific domestication models, two subsequent studies reverted to diversity and summary statistic-based types of arguments and logic originally employed with single-locus, mtDNA, or Y chromosome studies. Wang et al. (48) sequenced the genomes of 12 Eurasian gray wolves, 11 indigenous dogs from Southeast Asia, 12 indigenous dogs from northern East Asia, 4 village dogs from Nigeria, and 19 breed dogs from across the Old World and the Americas. They argued for a Southeast Asian origin of dogs on the basis of three lines of evidence. First, echoing earlier arguments, they found that Southeast Asian village dogs have the highest genetic diversity among the limited assayed dog groups. Second, population structure, phylogenetic, and admixture analyses indicate that the deepest division among dogs is between Southeast Asian dogs, which are the most basal, with the remaining dogs (including modern breeds) containing varying degrees of Southeast Asian ancestry. Their third argument was based on analyses of the degree of nonrandom associations or linkage disequilibrium (LD) between variants, which are richer sources of information about genetic diversity than simple summary statistics. LD was lower for Southeast Asian dogs than other dog groupings (Figure 8a). The authors conclude that dogs originated from Southeast Asia, expanding outward and contributing to the gene pools of Northeast Asian dogs and then subsequently to increasingly far-flung populations in Europe and the Americas.

Using similar LD-based arguments, Shannon et al. (49) genotyped 185,805 markers for an unprecedented number of purebred dogs (4,676 from 161 breeds) and village dogs (549 from 38 countries) and combined this with previously generated array and mtDNA data from dogs and wolves. Because village dogs likely contain more indigenous genetic diversity, they compared

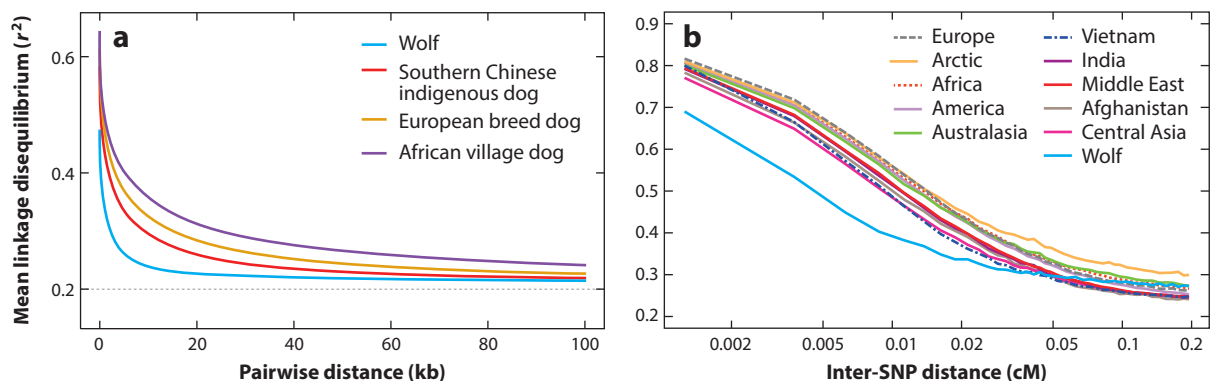


Figure 8

Linkage disequilibrium decay curves for (a) wolves, southern Chinese indigenous dogs, European breed dogs, and African village dogs [modified with permission from Wang et al. (48) under a Creative Commons Open Access license] and (b) geographic village dog populations [from Shannon et al. (49), with permission from PNAS].

patterns of LD decay among different geographic village dog populations. They found that short-range LD was lowest in village dogs from Afghanistan and Central Asia and at intermediate inter-SNP distances in Vietnam, with increasingly higher LD observed in populations with distance from Central Asia (**Figure 8b**). Solely on the basis of this observation, the authors concluded that dogs likely originated in Central Asia. These LD patterns were called into question by an analysis that added LD data from East Asian village dogs based upon whole-genome sequencing, which found they had lower LD than Central Asian village dogs (50). However, this finding was attributed to negative bias in LD estimates arising from the low-coverage nature of these genome sequences (51).

Two recent studies have used whole-genome data from both extant and fossil canid genomes to further our understanding of dog domestication. With evidence mounting for an origin of dogs from an extinct wolf lineage (15, 19), one possibility is that dogs derived from a genetically distinct population of wolves that occupied the steppe-tundra biome during the Late Pleistocene but was subsequently replaced by modern wolf lineages (14). This hypothesis is consistent with fossil evidence for a morphologically distinct, now-extinct wolf population that was adapted to hunting megafauna in Late Pleistocene Beringia (52). To investigate this idea further, Skoglund et al. (53) sequenced the genome of the Taimyr wolf, a 35,000-year-old fossil specimen found on the Taimyr Peninsula of northern Siberia. Although the autosomal genome was sequenced to only 1× coverage, the mtDNA genome was sequenced to 182×. Rather than being basal to all dogs, the mtDNA phylogeny indicated that the Taimyr wolf was reciprocally monophyletic to a clade composed of other ancient dogs, modern Eurasian and North American wolves, and modern dog clades B and C. This interleaving of wolf and dog lineages, both ancient and extant, was similar to that found by Thalmann et al. (19) and suggested an early divergence of modern wolves and dogs from a now-extinct ancient wolf population.

The second study sequenced the genome of the ancient Newgrange dog (~4.8 kya) to 28× coverage and an additional 59 mtDNA genomes from ancient European dogs (~3–14 kya). They combined the Newgrange genome sequence with that from previously published whole-genome data from 80 modern wolves and dogs and 605 modern dogs (including village dogs and 48 breeds) genotyped on the Illumina 170K HD SNP array (32). Because the authors used a combination of arguments concerning timing and geography, we present their findings in this section. They concluded that (a) dogs originated twice, once in East Asia and once in western Eurasia, and (b) dog lineages expanding out of East Asia replaced those of indigenous western Eurasian ancestry. First, their phylogenetic reconstruction placed the Newgrange dog within a western Eurasian dog clade, with East Asian and western Eurasian dogs found to be reciprocally monophyletic. Second, after calibration of the mutation rate using the age of the Newgrange dog, the multiple sequentially Markovian coalescent (MSMC) method (54) was used to infer the divergence times by comparing the ratio of between-population cross-coalescence events (between dogs and wolves) to that within populations (within dogs). For lineages that come from the sample population, this ratio approaches one, and conversely, it approaches zero when lineages have diverged and gene flow has ceased. Assuming that a ratio of 0.25–0.75 was indicative of divergence, they estimated that dogs and wolves diverged ~60–20 kya, although this might be an overestimate if the Russian wolves used in the analysis did not represent the founding population of dogs. The divergence between East Asian and western Eurasian dogs was ~14–6.4 kya (**Figure 9a**). Because the timing of divergence between East Asian and western Eurasian dogs postdated the known occurrence of fossil dogs in both regions based upon a reanalysis of the fossil record, they argued that dogs must have been present in both regions prior to this divergence. Their finding that on a PCA the Newgrange dog was found to be distinct led them to further suggest that it must contain mixed ancestry from the indigenous dogs of the region, and from the lineages that expanded out of East Asia and replaced western Eurasian dogs.

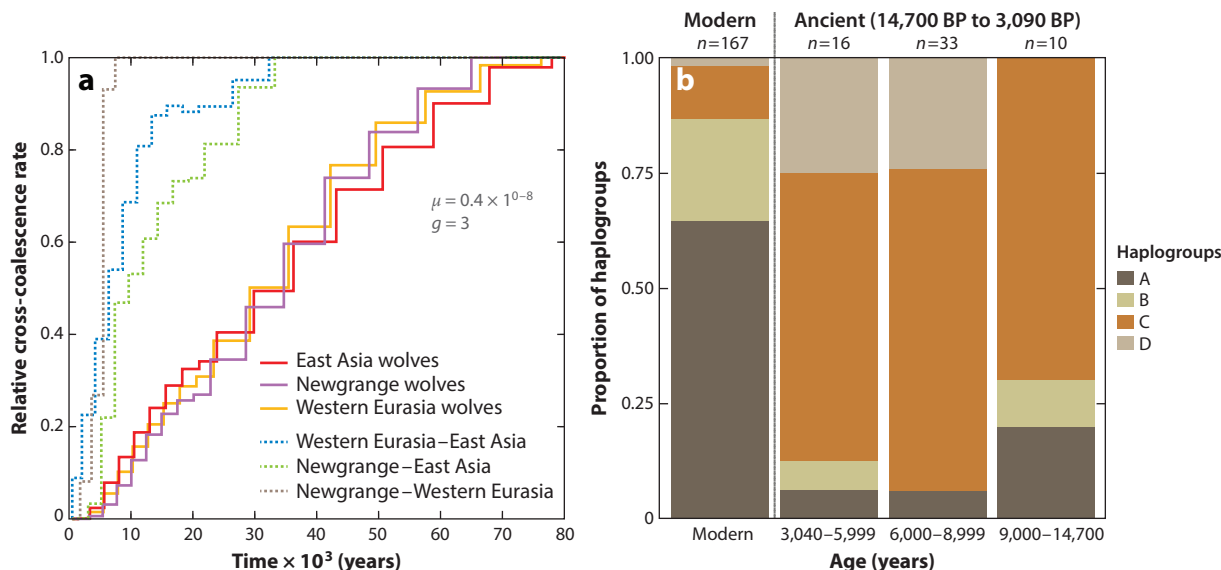


Figure 9

(a) Relative cross-coalescence rates showing evidence for a divergence between western Eurasian and East Asian dogs that postdates the oldest known fossil dogs from the regions. (b) Turnover in mitochondrial DNA (mtDNA) haplotypes between ancient and modern European dogs, consistent with a lineage replacement model. Both panels modified with permission from Frantz et al. (32).

To support this claim, they pointed to three lines of evidence. First, from mtDNA CR sequences for ancient and extant dogs, they demonstrated that there was extensive turnover of mtDNA haplotypes between ancient and modern dogs, during which clade A rose to high frequency and clades B, C, and D all decreased (**Figure 9b**). Second, using simulations, they found this pattern could not be explained by genetic drift. Third, and also using simulations, they found that lineage replacement had a larger effect on the ratio of between- to cross-coalescence rates in MSMC than postdivergence gene flow (although it is unclear why a larger effect favors one scenario over another).

TIMING OF DOG DOMESTICATION

MtDNA Studies

Timing estimates for dog domestication using mtDNA are based on two kinds of analyses. First, phylogenies are constructed, and clades are used to define subsequent groups for analysis. Then, the number of mutational steps, either between clusters of dog and wolf haplotypes or among central and peripheral haplotypes of dogs in haplotype networks, is used to infer a divergence time, typically conditioning on a mutation rate inferred from wolf–coyote divergence. Vilà et al. (21) constructed a phylogeny of dogs and wolves with coyotes as outgroup based upon CR mtDNA sequence (**Figure 2a**). They determined that the majority of dog haplotypes belonged to a single clade (clade I). Assuming the origin of this clade closely corresponds to the timing of the origin of all dogs, they transformed the observed maximum sequence divergence between clade I dog haplotypes into units of time by equating the sequence divergence observed between wolves and coyotes to a fossil-based estimate of wolf–coyote divergence time of ~ 1 Mya (55). This transformation indicated that dogs could have been domesticated as early as 135 kya. Although this divergence time well predates the fossil estimates available at the time (~ 15 kya), the authors

suggested that early dogs and wolves may have been morphologically indistinguishable. Presumably, the demographic divergence of wolves that followed humans and were ancestral to dogs could have been much earlier than suggested by morphology. However, recently the divergence time of wolves and coyotes was found to be more than an order of magnitude too large, suggesting that coyotes and wolves share common ancestry less than 100 kya (56). The more recent data are consistent with recent fossil evidence finding a great range in the temporal size variation of early coyote-like forms, which confounds simple reconstructions of ancestry and descent (57).

Savolainen et al. (40) made similar assumptions about wolf–coyote divergence to date sequence divergence in their mtDNA-based phylogeny but focused on mean sequence divergence within the A, B, and C clades [analogous to clades I, II, and III of Vilà et al. (21)] (**Figure 2b**). They assumed that haplotype networks showing signals of population expansion must reflect demographic events postdating domestication from a wolf founder haplotype. Thus, for each clade, mean sequence divergence was calculated from a central, presumed wolf haplotype in a network and the peripheral dog haplotypes. Given the complicated structure of the network for clade A, this analysis was conducted for all of clade A assuming one founding haplotype and three separate subclusters of haplotypes. For clade A (one founder), three separate clade-A clusters, clade B, and clade C, the estimates for the timing of dog origins were $14,000 \pm 4,000$ years ago, $11,000 \pm 4,000$ years ago, $16,000 \pm 3,000$ years ago, $26,000 \pm 8,000$ years ago, $13,000 \pm 3,000$ years ago, and $17,000 \pm 3,000$ years ago, respectively. Their conclusion was that dogs either originated from clade A ~ 40 kya or originated ~ 15 kya from a pool of founders drawn from all three clades. Obviously, this exercise assumes a correctly calibrated clock (see comments above) and single origin followed by limited crossing with wolves; at the very least, limited gene flow between dogs and wolves is not supported by genomic evidence (15, 20).

Genomic Studies

The first published study to use whole-genome sequence data from dogs and wolves based its analysis on the assumption, derived from previous studies of mtDNA and Y chromosome diversity, that dogs originated from Southeast Asia (46). These authors sequenced the genomes of Chinese indigenous dogs and wolves from China and eastern Russia to moderate coverage ($\sim 9\text{--}14\times$). From the resulting polymorphism data, they then estimated demographic parameters from the joint site frequency spectra of indigenous Chinese dogs and wolves from China and eastern Russia under an isolation-with-migration model. Using a generic mammalian mutation rate of 2.2×10^9 , they inferred the timing of domestication to be on the order of 32 kya. In a nearly identical analysis assuming the same mutation rate with an expanded set of genome sequences, including those from Wang et al. (46), a demographic history was inferred from the site-frequency spectrum, again assuming an East Asian origin for dogs (48). They estimated the divergence time between wolves and Chinese indigenous dogs was 33 kya (48).

Subsequent to Wang et al. (46), Freedman et al. (15) sequenced six dog, wolf, and golden jackal genomes (see above) to evaluate alternative demographic models of dog domestication and, using the most parsimonious model, inferred a timing of dog–wolf divergence of 11–16 kya (**Figure 7a**). As with all studies that rescale coalescent time units to real time units, this estimate is highly dependent upon the specified mutation rate. To account for this uncertainty, they considered the range of reported mutation rates for canids, or more generally mammals, and expanded this interval to 11–34 kya.

Using the 35,000-year-old Taimyr wolf sample from northern Siberia, Skoglund et al. (53) estimated the canid mutation rate as closer to the lower end of the range explored by Freedman et al. (15). To do so, they implemented a novel method that accounted for the low coverage

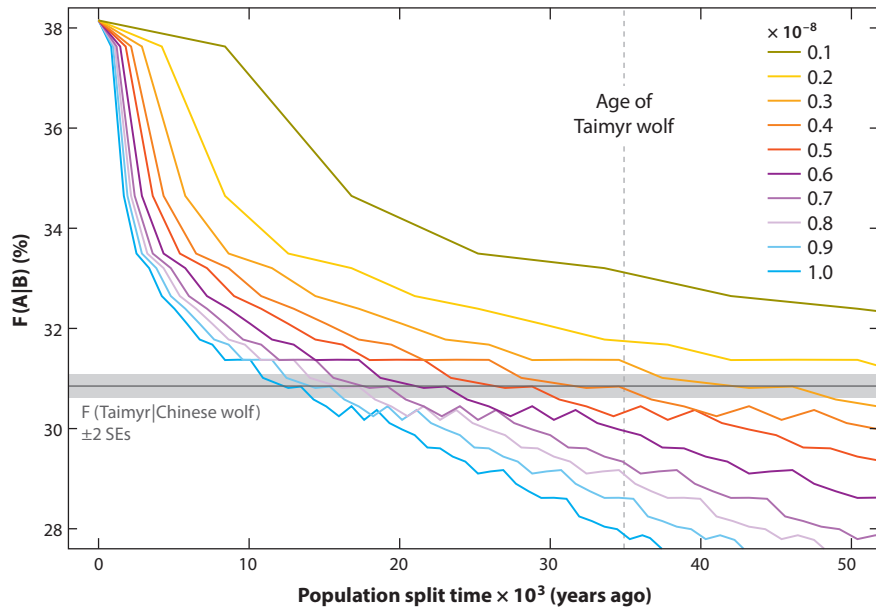


Figure 10

Calibration curves for $F(A|B)$, the probability that the Taimyr wolf carries the derived allele at positions where the Chinese wolf is heterozygous. Modified from Skoglund et al. (53) with permission from Elsevier. Curves are calculated with varying mutation rates, using a demographic history for Chinese wolf inferred with pairwise sequentially Markovian coalescent method (58) and a Chinese wolf–Taimyr wolf divergence time of 35 kya.

(1 ×) of their ancient wolf sample. They first fitted a demographic model to the Taimyr wolf and modern canid samples, allowing for postdivergence gene flow. Their data were consistent with a trifurcation of the Taimyr wolf, modern wolf, and dog lineages and suggested a divergence time greater than the previously reported dog–wolf divergence of ~11–16 kya (15) that was based upon a mutation rate of 1.0×10^{-8} . Specifically, they calculated $F(A_{\text{derived}} | B_{\text{heterozygous}})$, the probability of observing a derived allele in the Taimyr wolf (A) that was present as a heterozygote in a modern sample (B), to estimate mutation rate. They then used a summary statistic to estimate the divergence time of the Taimyr wolf from the Chinese wolf, given a demographic history of the latter inferred using the pairwise sequentially Markovian coalescent (PSMC) method (58). Because of the trifurcation of the Taimyr wolf, extant wolves (including the Chinese wolf), and dogs, calibration curves of $F(A|B)$ against estimated Taimyr wolf–Chinese wolf split times can be calculated under a range of mutation rates (**Figure 10**). The rate that most closely matches 35 kya can then be used to recalibrate previous estimates of wolf–dog divergence time based on different rates. They estimated the mutation rate to be 0.4×10^{-8} and thus concluded that the timing of domestication based on the genome sequences in Freedman et al. (15) should be shifted earlier, to ~27–40 kya (**Figure 10**).

THE ADMIXTURE PROBLEM

Admixture at two levels, specifically, admixture between dogs and other wild canid lineages and between resident indigenous dogs and dogs that originated from elsewhere, can complicate attempts to infer the evolutionary history of dogs from genomic data. The canine genome is particularly porous with regard to admixture and contains signals of interbreeding on varying timescales across

past and present geographic distributions. A notable example includes the historical introgression of the black coat color allele into North American wolves (59). Additionally, coyotes and wolves have hybridized to produce red wolves, which are dominated by coyote ancestry, and Great Lakes wolves, which have heterogeneous proportions of coyote ancestry (60). Recent studies analyzing whole-genome sequence data have repeatedly found evidence for geographically structured gene flow between dogs and wolves, with admixture that postdates domestication common within local populations of wolves and dogs (15, 20). Surprisingly, evidence was presented for more ancient and substantial gene flow between the dog–wolf ancestor and golden jackals (15). The clear implication from these findings is that a component of the genome diversity found in dogs may not reflect descent so much as admixture and traces to interbreeding with local wolf populations. Consequently, inferences regarding dog origins may be confounded by admixture. For example, allele and haplotype sharing based on genome-wide SNP data between dogs and Middle Eastern wolves suggested an origin there (22). Subsequent genome studies revealed strong signals of gene flow between Middle Eastern wolves and basenjis. Thus, an alternative explanation for high haplotype sharing is that admixture transferred dog haplotypes to wolves in the Middle East and enhanced overall genetic similarity between them and dogs (15). Similarly, arguments for a Southeast Asian origin of dogs based on levels of genetic diversity may have been influenced by geographically localized gene flow between Chinese wolves and Southeast Asian dog lineages. This idea is supported by a signature of gene flow between Chinese wolves and dingoes, a dog population known to have originated in mainland Southeast Asia, and with other Chinese dogs (15, 20, 61). Notably, there is substantial geographic structure in Chinese wolves, which can add to diversity of dogs that are geographically overlapping with them (15, 20). Admixture from Asian wolves into dogs would elevate metrics of genetic diversity in ways that do not reflect their genetic ancestry and may be a problem for studies that do not use explicit demographic models that account for admixture (e.g., 48, 51).

Overlaying patterns of genetic diversity for ancient breeds that are basal on dog phylogenies with the distribution of putative dog fossils results in a paradox first noted by Larson et al. (14). Specifically, the oldest dog fossils that are all found in Europe and Central Asia overlap geographically with modern breeds with relatively recent origins, whereas ancient breeds are often found in regions without fossil evidence of ancient dogs. This pattern holds even when excluding some of the oldest samples from Europe for which the dog classification has been disputed (14). One possible explanation for this pattern is that the richer record in Europe reflects more favorable opportunities for preservation of dog remains, or more intensive archeological research. An alternative explanation, increasingly supported by genetic data, is that in regions containing the oldest evidence for fossil dogs, indigenous ancestry may have been eliminated by admixture with immigrant lineages from elsewhere, particularly from East Asia. As noted by Brown et al. (42), and now supported by genome-level surveys of current and ancient genetic diversity (32), these two taxonomic levels of admixture perhaps represent the greatest challenge to inferring the evolutionary history of dogs. Although computational methods exist to jointly infer divergence times, effective population sizes, and gene flow (reviewed in 62), their precision when effective rates of gene flow are high is uncertain. For example, given surprisingly high gene flow between the dog–wolf ancestor and golden jackals from the Middle East, the estimated date of divergence between these two lineages was ~350 kya, nearly a third of that previously inferred from fossil divergence times (55).

CRITICAL ISSUES

Inference of an evolutionary process from patterns of genetic polymorphism typically does not employ deductive reasoning. More often than not, it relies on an inductive approach, in which the

feasibility of observed patterns is evaluated under a model of neutral evolution. Such models are convenient oversimplifications of biological reality that are thought to capture the most important biological features of a system. As such, these analyses are typically accompanied by a set of assumptions about reality that may hold to varying degrees. In the case of dog domestication, reality is complicated, with known bottlenecks occurring throughout the history of domestication and breed formation, population expansion and replacement, and postorigin dog–wolf admixture. Below, we discuss the major issues that arise when making inferences about the geography and timing of domestication given these complications. Rather than focusing too deeply on potential issues with any one study, our goal is to highlight broad themes that should be considered when drawing inferences from the current body of work we have reviewed.

Demography, Genetic Diversity, and Linkage Disequilibrium

Temporal and geographic variation in genetic diversity can be influenced by a diversity of evolutionary processes. For example, secondary contact between previously isolated populations followed by admixture (see section titled The Admixture Problem, above) will lead to an increase in diversity as measured by summary statistics, as will regional population structure (assuming sampling for a region spanning differentiated subpopulations). Conversely, genetic drift accompanying a population bottleneck will tend to eliminate rare alleles that would contribute to metrics such as the number of haplotypes. Likewise, population expansions can also reduce diversity as a result of the successive founder events that occur at the expansion front, and when a particular haplotype by chance rises to high frequency (63, 64). Thus, although higher genetic diversity observed in some geographic regions might suggest locales for dog origins, it may also reflect demographic stability or higher levels of admixture relative to another source population that has either undergone more severe population bottlenecks or not experienced diversity-augmenting admixture to the same degree. The early mtDNA and Y chromosome studies create the greatest interpretive challenge in this regard. Because these studies examined variation in nonrecombining regions, patterns of admixture are difficult to trace, and the demographic fluctuations that may have shaped diversity can be inferred only indirectly, for example, by assessing a signal of population expansion using Tajima's *D* (65) or Fu's *F* (66). Nonetheless, even genome-scale studies using diversity-based arguments (e.g., 48) may be confounded by differences in demographic history.

Interpretation of patterns of LD is similarly complicated by population structure, admixture, and demographic expansions or contractions. For example, admixture will elevate LD at larger inter-SNP distances because of the presence of divergent long-range haplotypes in the parent populations. Extreme demographic fluctuations will also tend to elevate LD. As a result, variation in rates of LD decay between populations might contain some signal of proximity to a source population for dogs, but it may also represent demographic heterogeneity. The recent counterarguments for Central Asian (49) and East Asian (48) origins based upon LD patterns are noteworthy in this regard, as sampling bias with respect to demography can confound inference. Wang et al. (48) sampled only indigenous dogs from eastern Asia and compared diversity patterns to breed dogs from other geographic regions. As bottlenecks associated with breed formation strongly reduce genetic diversity (2, 22, 44, 67, 68) and increase LD, this is not an appropriate comparison. Sequencing strategies can also influence LD patterns, such as a negative bias when inferences are based on low-coverage data (51). Finally, there is the problem of biological versus statistical significance. Although analyses of LD patterns have confirmed a clearly observable reduced LD in wolves relative to dogs, the differences between LD decay curves of dog populations are typically very subtle (Figure 8). Aside from the demographic complexities that might contribute to these differences, no formal tests have been conducted to determine whether such

curves are significantly different from each other. Even if these curves prove to be statistically distinguishable, inferring biological significance is challenging, and specific demographic models need to be used to explore the demographic causes and timing of demographic events.

Population Sampling

Studies of dog domestication have, to date, employed heterogeneous strategies for sampling populations. To the extent that these populations are not those most germane to the domestication process, they may confound inferences of dog origins. In fact, differences in sampling have provoked debate. Boyko et al. (44) questioned an earlier inference of higher genetic diversity in East Asia (40) owing to greater sampling of dogs from that region. However, a subsequent study indicated the patterns held even with more even sampling (41). Freedman et al. (15) sought to avoid geographic bias by using the boxer genome and two divergent, putatively basal dog breeds (basenji and dingo) to sample the three putative domestication centers. However, these breeds may not represent the ancestral, indigenous diversity of their respective regions. Instead, the distinctiveness of the basenji and dingo may reflect geographic isolation from the admixture that may have largely supplanted indigenous diversity, such as is thought to have occurred in European dogs (14, 32). In general, the extent to which dog populations derive from ancient populations from the same region is poorly understood. Without including multiple ancient dog and wolf genome sequences in formal tests of alternative models of domestication, robust conclusions concerning the timing and geography of dog domestication will remain elusive. In this regard, it is notable that two studies using whole-genome data from ancient canids have primarily used sequence data from them only to infer a mutation rate (32, 53).

Mutation Rate Uncertainty

Although more sophisticated model-based approaches have now led to increasingly precise estimates for the timing of domestication (15, 20, 53), the greatest uncertainty surrounding these estimates is, as with studies of human evolution (69), due to differences in the assumed mutation rate. Assuming a rate of 1×10^{-8} per site per generation, Freedman et al. (15) inferred domestication as having occurred ~ 11 – 16 kya. Using a slower rate increased the upper limit to ~ 30 kya (also see Reference 20). Subsequent research using fossil samples to infer a mutation rate (32) suggested that the rate is in fact slower, consistent with a Late Palaeolithic origin of dogs. As parameter estimates for genetically based demographic models are rescaled into time units, the ordering of divergence events will not be impacted. More challenging is comparing parameter estimates from genome sequence data with patterns inferred from the distribution of fossils. For example, substantially slower mutation rates could push the timing of coalescent events between western Eurasian and East Asian dogs further back in time such that they would predate the occurrence of known fossils in both regions. As the argument for separate domestication events in each region relies on genetic estimates of divergence being contemporary to or more recent than fossil evidence (32), a much slower inferred mutation rate would call this conclusion into question.

Incomplete Lineage Sorting and Admixture

The effects of ILS are most pronounced for recent divergence events, such as those between dogs and wolves, for which so few generations have passed that a proportion of the genome may exhibit genealogies discordant with species history owing to deep coalescence events that predate divergence. Even for divergences that are older (in the number of generations) than dog–wolf divergence, ILS can still be substantial. For example, for the genealogy of the chimpanzee, gorilla,

and human, 30% of nucleotide positions show evidence of ILS (70). When ILS is not explicitly accounted for, discordant genealogies can be mistaken for postdivergence gene flow (18) and can influence phylogenetic reconstructions and the divergence times associated with particular nodes (17, 18). In the pregenomic era, single-locus studies based upon mtDNA or Y chromosomes could not assess the impact of ILS. Even as genome-scale studies of dog domestication have accelerated in the last three years, few studies have employed methods that disentangle the effects of gene flow and ILS (although see 15, 20). Inferences about the location of centers of domestication that rely on summary statistics (genetic diversity, LD patterns) do not take into account the joint effects of ILS and gene flow, which may obscure the true pattern of lineage divergence. Additionally, model-based approaches that do not address ILS directly will also be prone to incorrect inferences concerning divergence times.

SYNTHESIS AND FUTURE DIRECTIONS

Despite the rapid advances in sequencing technology and, concurrently, the rapid reduction in costs of whole-genome sequencing, studies of whole-genome sequence data necessary to unravel the history of dog domestication have commenced only in the last three years, starting with the work of Wang et al. (46). Although, as highlighted above, questions can be raised concerning the robustness of particular statistical approaches, we can draw a few tentative conclusions. First, dogs likely originated from one or more now-extinct populations of wolves. Second, these domestication events likely occurred in the last 11,000–35,000 years and were associated with hunter-gatherer cultures rather than agrarian societies. Third, the history of dog domestication has involved a complex demography, including population bottlenecks, expansions, local extinctions and replacements, and regionally restricted postdivergence gene flow between dogs and wolves and ancient and modern dogs. These complexities substantially complicate efforts to understand where, when, and how many times dogs have been domesticated.

Nonetheless, these general conclusions suggest an early domestication process perhaps beginning more than 25 kya, in one or more wolf populations, perhaps in Europe, the High Arctic, and/or East Asia. This early association may have been facilitated by the presence of large carcasses left by humans on the landscape and imposed on some wolves a migratory, human-niche orientation, which led to their demographic divergence from resident territorial wolves. This process may have been followed by an increasingly close, mutualistic association between wolves or protodogs and humans that might have involved coordinated activities, such as hunting and defense from other humans and carnivores. A second stage involved the development of agriculture beginning approximately 10 kya and a more sedentary lifestyle, where selection for modified phenotypes occurred, resulting in marked phenotype divergence from wolf ancestors and the appearance of dogs in a wide range of body sizes. Finally, the Victorian era of directed selection for fancy or novelty resulted in a vast range of dog phenotypes, which are the dominant feature of dog diversity today. These periods of domestication have left dramatic imprints of selection and genetic diversity on the dog genome. The population bottlenecks associated with first domestication and breed formation have reduced genetic diversity and the efficacy of selection, leading to an excess of deleterious variation, long regions of autozygosity, and elevated LD (22, 44, 49, 68). Nevertheless, positive selection has left clear signals with respect to traits experiencing selection under domestication. For example, selection appears to have acted on metabolic functions, to cope first with changes in dietary fat composition (71) and later with a dietary increase in starch associated with an increasingly commensal lifestyle (72).

Two significant challenges suggest needed future work. First, there is a trade-off between the ability to perform robust tests of alternative demographic models of dog domestication and

the number of sequences that can be analyzed. Jointly inferring divergence times, population size changes, and gene flow estimates (for all possible population pairs) for a phylogeny of all sampled individuals is computationally intensive, such that it is not feasible to analyze genome-scale data from the hundreds (soon to be thousands) of samples for which genome data are now available. Other recently developed methods for model-based demographic inference using whole-genome data are also limited in the number of samples/populations that can be jointly analyzed (reviewed in 61). Although these approaches permit quantitative assessments of the fit of different models, the computational limit also places a ceiling on the generality of conclusions that can be made from them. The less-than-ideal alternative is to analyze data from a much larger set of samples, increasing both geographic and taxonomic coverage, but resorting to inferences of demographic processes based on their consistency with general expectations under competing neutral demographic models, rather than formal model testing (e.g., 48, 49). Enhanced inference in the future rests on the success of statistical geneticists working on coalescent inference methods for a large number of populations that incorporate ILS, gene flow, and heterogeneous variations in population size over time. Recent studies using the PSMC (58) and MSMC (54) methods have demonstrated heterogeneity among dog breeds and wolf populations in their demographic trajectories (15, 20, 32), highlighting the need to build this variation into future analyses.

A final critical issue is the importance of sampling ancient remains across a diversity of geographic regions and time points. Collecting whole-genome data from fossils both poses substantial technical challenges because of the degradation of ancient DNA and is constrained by the temporal and spatial distribution of fossil material (**Figure 1**). Inferences of demographic parameters in any joint history of wolves and dogs will most likely reflect reality when current patterns of genetic diversity can be linked directly to that of ancient samples. Without such data, failed domestication attempts cannot be readily distinguished from early dog lineages that contributed to the ancestry of extant dogs. Critically, fossil data will allow a better understanding of lineage replacement over time (14, 32). To date, whole-genome studies of dog domestication have focused on a single ancient sample [one wolf (53), one dog (32)], which may have only tenuous connections to the populations that eventually led to domestic dogs. For example, without genome sequences for a geographically and temporally broader set of ancient samples, the observation that the divergence of East Asian and western Eurasian dogs appears to postdate the appearance of dogs in the fossil record provides little insight into the actual lineage from which dogs ultimately arose. With the launch of a recent initiative to sequence more genome sequences of ancient canids (73; <http://www.dog10kgenomes.org>), researchers are better poised than ever to disentangle the complex evolutionary history that led to the only known transformation of a large carnivore into the working animal and companion that is the domestic dog.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

1. Am. Kennel Club. 1992. *The Complete Dog Book: The Photograph, History, and Official Standard of Every Breed Admitted to AKC Registration, and the Selection, Training, Breeding, Care, and Feeding of Pure-Bred Dogs*. New York/Toronto: Howell Book House
2. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, et al. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438:803–19

3. Drake AG, Klingenberg CP, Heard AESB, McPeck EMA. 2010. Large-scale diversification of skull shape in domestic dogs: disparity and modularity. *Am. Nat.* 175:289–301
4. Wayne RK. 1986. Limb morphology of domestic and wild canids: the influence of development on morphologic change. *J. Morphol.* 187:301–19
5. Wayne RK. 1986. Cranial morphology of domestic and wild canids: the influence of development on morphological change. *Evolution* 40:243–61
6. Cadieu E, Neff MW, Quignon P, Walsh K, Chase K, et al. 2009. Coat variation in the domestic dog is governed by variants in three genes. *Science* 326:150–53
7. Sutter NB, Bustamante CD, Chase K, Gray MM, Zhao K, et al. 2007. A single *IGF1* allele is a major determinant of small size in dogs. *Science* 316:112–15
8. Boyko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, et al. 2010. A simple genetic architecture underlies morphological variation in dogs. *PLOS Biol.* 8:e1000451
9. Hayward JJ, Castelano MG, Oliveira KC, Corey E, Balkman C, et al. 2016. Complex disease and phenotype mapping in the domestic dog. *Nat. Commun.* 7:10460
10. Rimbault M, Ostrander EA. 2012. So many doggone traits: mapping genetics of multiple phenotypes in the domestic dog. *Hum. Mol. Genet.* 21:R52–57
11. Vaysse A, Ratnakumar A, Derrien T, Axelsson E, Rosengren Pielberg G, et al. 2011. Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLOS Genet.* 7:e1002316
12. Schoenebeck JJ, Hutchinson SA, Byers A, Beale HC, Carrington B, et al. 2012. Variation of *BMP3* contributes to dog breed skull diversity. *PLOS Genet.* 8:e1002849
13. Parker HG, vonHoldt BM, Quignon P, Margulies EH, Shao S, et al. 2009. An expressed *Fgf4* retrogene is associated with breed-defining chondrodysplasia in domestic dogs. *Science* 325:995–98
14. Larson G, Karlsson EK, Perri A, Webster MT, Ho SYW, et al. 2012. Rethinking dog domestication by integrating genetics, archeology, and biogeography. *PNAS* 109:8878–83
15. Freedman AH, Gronau I, Schweizer RM, Ortega-Del Vecchyo D, Han E, et al. 2014. Genome sequencing highlights the dynamic early history of dogs. *PLOS Genet.* 10:e1004016
16. Maddison WP. 1997. Gene trees in species trees. *Syst. Biol.* 46:523–36
17. Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.* 24:332–40
18. Rogers J, Gibbs RA. 2014. Comparative primate genomics: emerging patterns of genome content and dynamics. *Nat. Rev. Genet.* 15:347–59
19. Thalmann O, Shapiro B, Cui P, Schuenemann VJ, Sawyer SK, et al. 2013. Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science* 342:871–74
20. Fan Z, Silva P, Gronau I, Wang S, Armero AS, et al. 2016. Worldwide patterns of genomic variation and admixture in gray wolves. *Genome Res.* 26:163–73
21. Vilà C, Savolainen P, Maldonado JE, Amorim IR, Rice JE, et al. 1997. Multiple and ancient origins of the domestic dog. *Science* 276:1687–89
22. vonHoldt BM, Pollinger JP, Lohmueller KE, Han E, Parker HG, et al. 2010. Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* 464:898–902
23. Shipman P. 2015. How do you kill 86 mammoths? Taphonomic investigations of mammoth megasites. *Quat. Int.* 359–360:38–46
24. Shipman P. 2015. *The Invaders: How Humans and Their Dogs Drove Neanderthals to Extinction*. Cambridge, MA: Belknap. 288 pp. 3rd ed.
25. Clutton-Brock J. 1995. Origins of the dog: domestication and early history. In *The Domestic Dog: Its Evolution, Behaviour and Interactions with People*, ed. J Serpell, pp. 7–20. Cambridge, UK: Cambridge Univ. Press
26. Thurston ME. 1996. *The Lost History of the Canine Race: Our 15,000-Year Love Affair With Dogs*. Kansas City, MO: Andrews & McMeel. 301 pp.
27. Germonpré M, Lázníková-Galetová M, Sablin MV. 2012. Palaeolithic dog skulls at the Gravettian Předmostí site, the Czech Republic. *J. Archaeol. Sci.* 39:184–202

28. Germonpré M, Sablin MV, Stevens RE, Hedges REM, Hofreiter M, et al. 2009. Fossil dogs and wolves from Palaeolithic sites in Belgium, the Ukraine and Russia: osteometry, ancient DNA and stable isotopes. *J. Archaeol. Sci.* 36:473–90
29. Zeder MA. 2012. Pathways to animal domestication. In *Biodiversity in Agriculture: Domestication, Evolution, and Sustainability*, ed. P Gepts, TR Famula, RL Bettinger, SB Brush, AB Damania, et al., pp. 227–58. Cambridge, MA: Cambridge Univ. Press
30. Ovodov ND, Crockford SJ, Kuzmin YV, Higham TFG, Hodgins GWL, et al. 2011. A 33,000-year-old incipient dog from the Altai mountains of Siberia: evidence of the earliest domestication disrupted by the last glacial maximum. *PLOS ONE* 6:e22821
31. Leonard JA, Wayne RK, Wheeler J, Valadez R, Guillen S, Vilà C. 2002. Ancient DNA evidence for Old World origin of New World dogs. *Science* 298:1613–16
32. Frantz LAF, Mullin VE, Pionner-Capitan M, Lebrasseur O, Ollivier M, et al. 2016. Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science* 352:1228–31
33. Crockford SJ, Kuzmin YV. 2012. Comments on Germonpré et al., *Journal of Archaeological Science* 36, 2009 “Fossil dogs and wolves from Palaeolithic sites in Belgium, the Ukraine and Russia: osteometry, ancient DNA and stable isotopes,” and Germonpré, Láznicková-Galetová, and Sablin, *Journal of Archaeological Science* 39, 2012 “Palaeolithic dog skulls at the Gravettian Předmostí site, the Czech Republic.” *J. Archaeol. Sci.* 39:2797–801
34. Morey DF. 2014. In search of Paleolithic dogs: a quest with mixed results. *J. Archaeol. Sci.* 52:300–7
35. Germonpré M, Sablin MV, Després V, Hofreiter M, Láznicková-Galetová M, et al. 2013. Palaeolithic dogs and the early domestication of the wolf: a reply to the comments of Crockford and Kuzmin 2012. *J. Archaeol. Sci.* 40:786–92
36. Germonpré M, Sablin MV, Láznicková-Galetová M, Després V, Stevens RE, et al. 2015. Palaeolithic dogs and Pleistocene wolves revisited: a reply to Morey 2014. *J. Archaeol. Sci.* 54:210–16
37. Drake AG, Coquerelle M, Colombeau G. 2015. 3D morphometric analysis of fossil canid skulls contradicts the suggested domestication of dogs during the Late Paleolithic. *Sci. Rep.* 5:8299
38. Perri A. 2016. A wolf in dog’s clothing: initial dog domestication and Pleistocene wolf variation. *J. Archaeol. Sci.* 68:1–4
39. Druzhkova AS, Thalmann O, Trifonov VA, Leonard JA, Vorobieva NV, et al. 2013. Ancient DNA analysis affirms the canid from Altai as a primitive dog. *PLOS ONE* 8:e57754
40. Savolainen P, Zhang Y, Luo J, Lundeberg J, Leitner T. 2002. Genetic evidence for an East Asian origin of domestic dogs. *Science* 298:1610–13
41. Pang J-F, Kluetsch C, Zou X-J, Zhang A, Luo L-Y, et al. 2009. MtDNA data indicate a single origin for dogs south of Yangtze River, less than 16,300 years ago, from numerous wolves. *Mol. Biol. Evol.* 26:2849–64
42. Brown SK, Pedersen NC, Jafarishorijeh S, Bannasch DL, Ahrens KD, et al. 2011. Phylogenetic distinctiveness of Middle Eastern and Southeast Asian village dog Y chromosomes illuminates dog origins. *PLOS ONE* 6:e28496
43. Ding Z-L, Oskarsson M, Ardalan A, Angleby H, Dahlgren L-G, et al. 2012. Origins of domestic dog in southern East Asia is supported by analysis of Y-chromosome DNA. *Heredity* 108:507–14
44. Boyko AR, Boyko RH, Boyko CM, Parker HG, Castelano M, et al. 2009. Complex population structure in African village dogs and its implications for inferring dog domestication history. *PNAS* 106:13903–8
45. Gray MM, Sutter NB, Ostrander EA, Wayne RK. 2010. The *IGF1* small dog haplotype is derived from Middle Eastern grey wolves. *BMC Biol.* 8:16
46. Wang GD, Zhai W, Yang HC, Fan RX, Cao X, et al. 2013. The genomics of selection in dogs and the parallel evolution between dogs and humans. *Nat. Commun.* 4:1860
47. Durand EY, Patterson N, Reich D, Slatkin M. 2011. Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* 28:2239–52
48. Wang G-D, Zhai W, Yang H-C, Wang L, Zhong L, et al. 2016. Out of southern East Asia: the natural history of domestic dogs across the world. *Cell Res.* 26:21–33
49. Shannon LM, Boyko RH, Castelano M, Corey E, Hayward JJ, et al. 2015. Genetic structure in village dogs reveals a Central Asian domestication origin. *PNAS* 112:13639–44
50. Wang G-D, Peng M-S, Yang H-C, Savolainen P, Zhang Y-P. 2016. Questioning the evidence for a Central Asian domestication origin of dogs. *PNAS* 113:E2554–55

51. Shannon LM, Boyko RH, Castelhan M, Corey E, Hayward JJ, et al. 2016. Reply to Wang et al.: Sequencing datasets do not refute Central Asian domestication origin of dogs. *PNAS* 113:E2556–57
52. Leonard JA, Vilà C, Fox-Dobbs K, Koch PL, Wayne RK, Van Valkenburgh B. 2007. Megafaunal extinctions and the disappearance of a specialized wolf ecomorph. *Curr. Biol.* 17:1146–50
53. Skoglund P, Ersmark E, Palkopoulou E, Dalén L. 2015. Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. *Curr. Biol.* 25:1515–19
54. Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* 46:919–25
55. Wayne RK, Van Valkenburgh B, O'Brien SJ. 1991. Molecular distance and divergence time in carnivores and primates. *Mol. Biol. Evol.* 8:297–319
56. vonHoldt BM, Cahill J, Fan Z, Gronau I, Robinson J, et al. 2016. Whole-genome sequence analysis shows that two endemic species of North American wolf are admixtures of the coyote and gray wolf. *Sci. Adv.* 2(7):e1501714
57. Meachen JA, Samuels JX. 2012. Evolution in coyotes (*Canis latrans*) in response to the megafaunal extinctions. *PNAS* 109:4191–96
58. Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475:493–96
59. Anderson TM, vonHoldt BM, Candille SI, Musiani M, Greco C, et al. 2009. Molecular and evolutionary history of melanism in North American gray wolves. *Science* 323:1339–43
60. vonHoldt BM, Pollinger JP, Earl DA, Knowles JC, Boyko AR, et al. 2011. A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. *Genome Res.* 21:1294–305
61. Sacks BN, Brown SK, Stephens D, Pedersen NC, Wu J-T, Berry O. 2013. Y chromosome analysis of dingoes and Southeast Asian village dogs suggests a Neolithic continental expansion from Southeast Asia followed by multiple Austronesian dispersals. *Mol. Biol. Evol.* 30:1103–18
62. Schraiber JG, Akey JM. 2015. Methods and models for unravelling human evolutionary history. *Nat. Rev. Genet.* 16:727–40
63. Le Corre V, Kremer A. 1998. Cumulative effects of founding events during colonisation on genetic diversity and differentiation in an island and stepping-stone model. *J. Evol. Biol.* 11:495–512
64. Excoffier L, Ray N. 2008. Surfing during population expansions promotes genetic revolutions and structuration. *Trends Ecol. Evol.* 23:347–51
65. Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–95
66. Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. *Genetics* 133:693–709
67. Gray MM, Granka JM, Bustamante CD, Sutter NB, Boyko AR, et al. 2009. Linkage disequilibrium and demographic history of wild and domestic canids. *Genetics* 181:1493–505
68. Marsden CD, Ortega-Del Vecchyo D, O'Brien DP, Taylor JF, Ramírez O, et al. 2016. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *PNAS* 113:152–57
69. Scally A, Durbin R. 2012. Revising the human mutation rate: implications for understanding human evolution. *Nat. Rev. Genet.* 13:745–53
70. Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, et al. 2012. Insights into hominid evolution from the gorilla genome sequence. *Nature* 483:169–75
71. Freedman AH, Schweizer RM, Ortega-Del Vecchyo D, Han E, Davis BW, et al. 2016. Demographically-based evaluation of genomic regions under selection in domestic dogs. *PLOS Genet.* 12:e1005851
72. Axelsson E, Ratnakumar A, Arendt M-L, Maqbool K, Webster MT, et al. 2013. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495:360–64
73. Grimm D. 2015. How the wolf became the dog. *Science* 348:277