Lessons learned from the dog genome

Robert K. Wayne[1] and Elaine A. Ostrander[2]

[1] Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 91302, USA
[2] Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA

Extensive genetic resources and a high-quality genome sequence position the dog as an important model species for understanding genome evolution, population genetics and genes underlying complex phenotypic traits. Newly developed genomic resources have expanded our understanding of canine evolutionary history and dog origins. Domestication involved genetic contributions from multiple populations of gray wolves probably through backcrossing. More recently, the advent of controlled breeding practices has segregated genetic variability into distinct dog breeds that possess specific phenotypic traits. Consequently, genome-wide association and selective sweep scans now allow the discovery of genes underlying breed-specific characteristics. The dog is finally emerging as a novel resource for studying the genetic basis of complex traits, including behavior.

> The more one gets to know of men, the more one values dogs. ~Alphonse Toussenel.

Teaching an old dog new tricks

The domestic dog (*Canis familiaris*) has long been a subject of scientific curiosity because of its great diversity in size, shape, coat color and texture, and behavior [1–3]. However, relative to model species such as mice, fruit flies and *Arabidopsis*, in which extensive crossing designs are possible, the genetic basis for specific phenotypic traits in the dog has been difficult to uncover. Nonetheless, the dog represents an important model species because of its phenotypic diversity and usefulness in the study of the origin of human genetic diseases. Indeed, the dog has a larger catalog of disease syndromes common to humans than any other laboratory or domestic species [4]. Moreover, although large-scale crossing designs are generally not feasible because of long generation times and the costs of maintaining a large mammal (for an interesting exception see [2]), the long history of dog domestication has uniquely segregated phenotypes into discrete breeds appropriate for genotype–phenotype association studies. Additionally, some phenotypes such as achondroplasia (foreshortened limbs) are commonly replicated in different breeds (e.g. Basset hounds, dachshunds, corgis) enhancing the statistical power of association analysis. Breeds can also share a high incidence of specific diseases [4]. Consequently, loci segregating with these phenotypes can potentially be identified through whole genome association studies across similarly affected breeds. With the

sequencing of the dog genome and related genomic resources, powerful new approaches for uncovering the genetic basis of phenotypic variation and disease are emerging. 'Selective sweep' approaches are particularly promising in the search for the genomic signal of positive selection [5–7]. Furthermore, canine genomic resources have a wide variety of applications including systematics, population genetics, endangered species conservation as well as gene mapping of specific phenotypic traits [4,8].

The evolutionary history of dog-like carnivores

The dog family is a phenotypically diverse group including 35 closely related extant species [9,10]. This recent radiation has resulted in an evolutionary branching pattern that is often difficult to resolve because many branches are closely spaced in time. Indeed, past phylogenetic studies focused on rapidly evolving mitochondrial genes, in an attempt to resolve this complicated phylogeny [9,11] or nuclear genes with relaxed evolutionary constraints [12,13]. Nonetheless, many recent and deep branching points remained poorly defined. Subsequently, the application of comparative genomics to the complete dog genome sequence enabled the identification of a suite of rapidly evolving nuclear genes [10] (Figure 1). The resulting phylogeny reaffirmed the close relationships between many phenotypically divergent taxa, such as the kit fox and Arctic fox, the bush dog and maned wolf, and the domestic dog and gray wolf (Figure 2) [10]. Further, the branching order of deep evolutionary divergences within the family was also revealed. For example, the topology of the three primary branches of the Canidae: wolf-like canids (blue), fox-like canids (red) and South-American canids (green) was deduced, with the latter being more closely related to the wolf- or dog-like canids (Figure 2). Similarly, the branching order within the wolf-like canids was resolved with the golden jackal as the clear outgroup to coyote, wolf and dog that were all phylogenetically distant from African jackals [10]. Molecular dating of branching events in the tree suggests that modern canids share a common ancestor ~ten million years ago. The modern canids are apparently a very recent radiation in a family that has an origin as much as 50 million years ago [14]. The wolf-like canids (Figure 2) have origins more than 6 million years ago and the group containing the closest relatives of the dog (gray wolf, coyote and golden jackal) share a common ancestor only ~three to four million years ago (Figure 2). Consequently, their divergence is more recent in time than chimps and humans and, likewise, comparative genomics of these species should reveal some of the changes that have uniquely evolved in each. These results
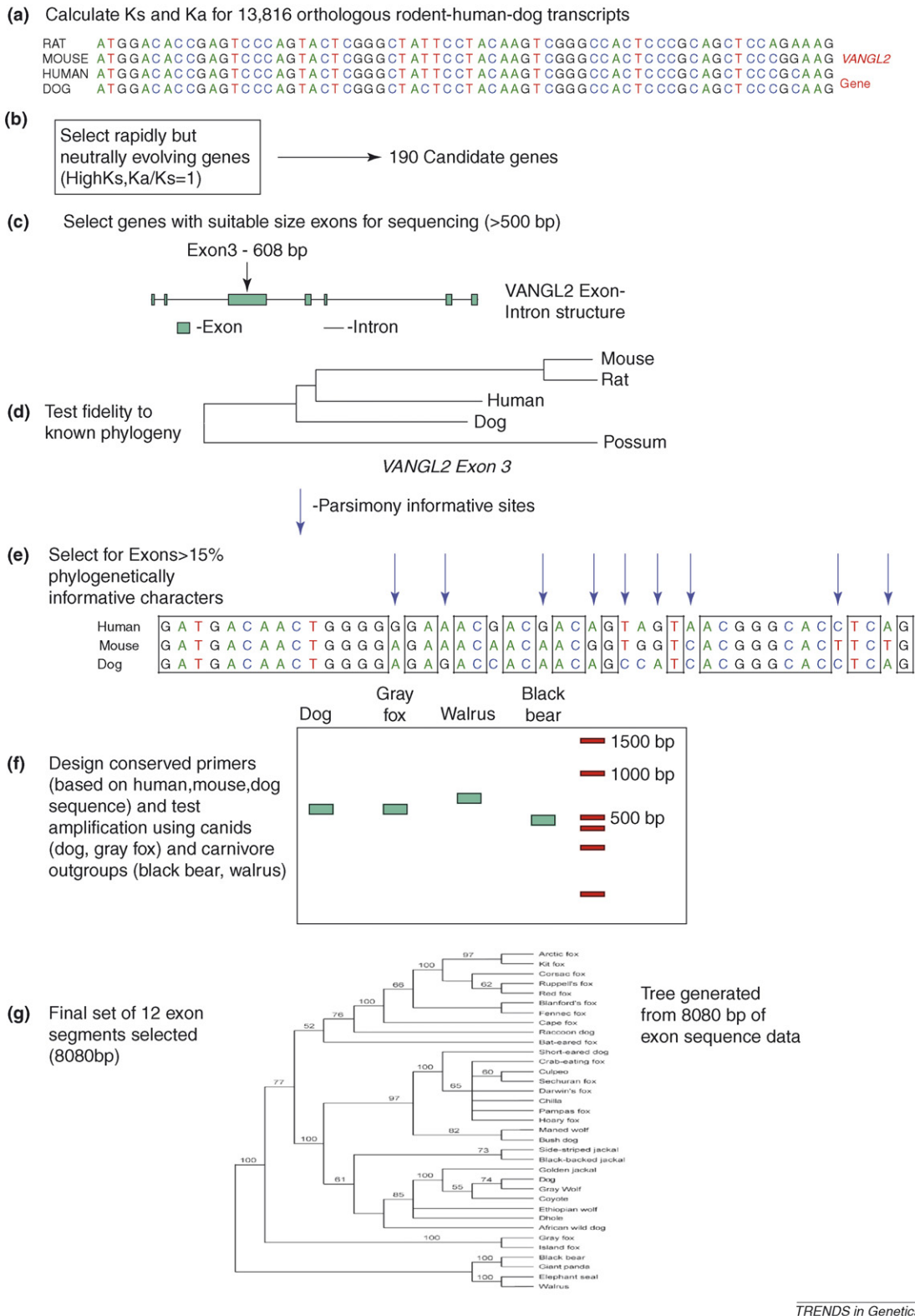
Figure 1. The discovery of rapidly evolving genes in the nuclear genome of canids. The protocol [10] for finding rapidly evolving nuclear genes involved the comparison of ~14 000 canine transcripts to their homologs in rat, human and mouse **(a)** and the subsequent identification of divergent canine exons that appeared to be evolving in a neutral manner ($Ks/Ka \approx 1$, where Ks and Ka represent the rates of synonymous and non-synonymous substitutions, respectively) **(b)**. From a list of 190 such exons, those of the appropriate size **(c)** that correctly passed a genealogical test with sequences from as all available mammal species **(d)** and were useful for phylogenetic reconstruction (only three of many *VANGL2* sequences are illustrated here) **(e)** were selected. Primers that amplified genes across the carnivore order (so that outgroups could be sequenced) were then designed **(f)** and used to amplify genes from a wide variety of species **(g)**. Similarly, rapidly evolving intron regions were identified as those with more than three polymorphic sites in a sliding window analysis of 500 bp. This resulted in a greater than a three- and fivefold enrichment of exons and introns, respectively, over background rates of sequence divergence. The resulting phylogeny based on ~15 000 base pairs of information was well resolved (Figure 2) [10].
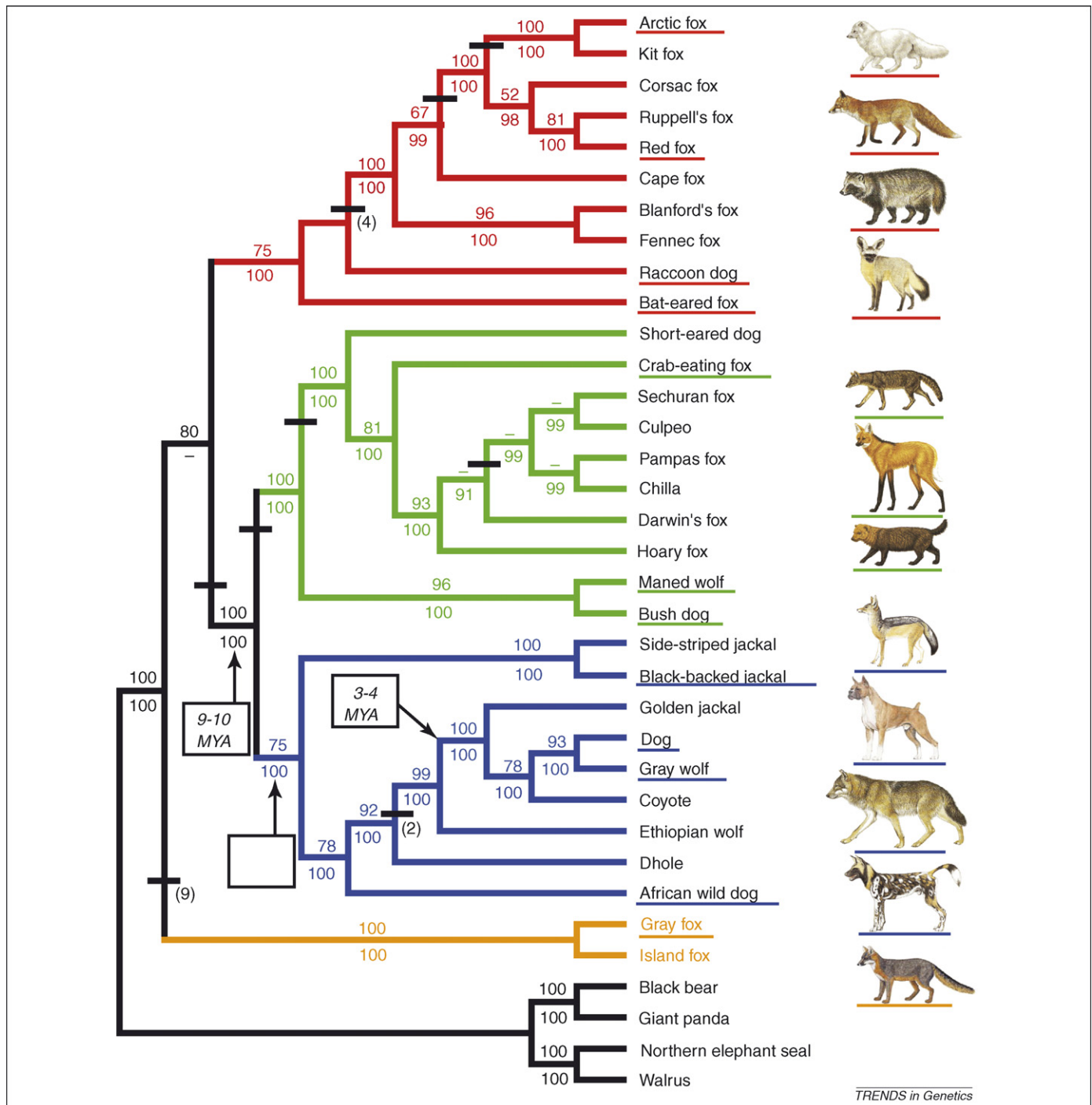
Figure 2. Molecular phylogeny of the dog family Canidae. The phylogeny is based on 14 948 bp of DNA sequence from 12 exons and four introns (Figure 1) [10]. Molecular dates are indicated in addition to the positions of indels (horizontal lines) that define specific lineages. The numbers above and below each branching point indicate two different measures of branching point support (bootstrapping and Bayesian posterior probability values) with values >90, which suggests strong support. The three primary phylogenetic groupings of canids are indicated by red (fox-like canids), green (South American canids) and blue (wolf-like canids) colors. Example species (underlined) are pictured. Scientific names of species mentioned in the text in order of appearance: kit fox (*Vulpes macrotis*), Arctic fox (*Alopex lagopus*), bush dog (*Speothos venaticus*), maned wolf (*Chrysocyon brachyurus*), domestic dog (*Canis familaris*), gray wolf (*Canis lupus*), golden jackal (*Canis aureus*), coyote (*Canis latrans*), African jackals (*Canis mesomelas and Canis adustus*) and gray fox (*Urocyon cinereoargenteus*). The tree shows good resolution of deep and shallow branches supporting the combined use of exons and introns, as well as defining the position of enigmatic canids, such as the bush dog, maned wolf and the topology of the wolf-like canids (see main text).

demonstrate that the use of rapidly evolving introns and more conserved exons produces a phylogeny resolved at deep and shallow levels of divergence. This approach can potentially be applied to other vertebrate families that have had at least one species, such as the domestic dog, with their complete genome sequenced.

Owing to the close relationships within wolf-like canids, molecular tools developed for the dog are likely to be applicable to several wild species, and hence these wild species are 'genome enabled' [8]. For example, about half of the microsatellite primers developed in the dog amplify DNA in other canids and have proven generally useful for

resolving variation at the population level [15] and for gene mapping [16]. Even microsatellite and sequencing primers developed for the dog major histocompatibility complex (MHC) have been successfully applied to the Island fox, an endangered species that belongs to the most divergent canid lineage (Figure 2) and is found on the six Channel Islands off the coast of Southern California [17]. The MHC is an assemblage of linked genes that have an important role in immunity and might be under balancing selection [18]. MHC variation in foxes on one island was high despite the absence of variation in selectively neutral microsatellites or multilocus DNA fingerprints [17]. This unexpected result suggests the action of intense natural selection and supports the role of balancing selection in the maintenance of variation at the MHC. Finally, single nucleotide polymorphisms (SNPs) identified in the domestic dog have also been used to characterize variation in wild wolves, thereby reaffirming previous conclusions based on mitochondrial DNA (mtDNA) sequencing and microsatellite loci [19]. This pioneering study represented one of the first SNP surveys on a wild population of a non-human vertebrate and showed that even small-scale SNP surveys can provide useful information. SNPs represent a new molecular marker for population genetics that might provide improved resolution of questions concerning historical population demography and gene flow, in addition to identifying genes that influence adaptation [20].

Dog origins, diversity and diversification

The evolution of the dog, its geographic origins, antiquity and backcrossing with wild canids, has long been a topic of debate. Phylogenetic studies of mtDNA sequences in domestic dogs and wolves suggested the potential for multiple domestication and backcrossing events and an origin perhaps from >100 000 to 15 000 years ago [21,22]. Higher mtDNA haplotype diversity was found in dogs from East Asia, which is suggested to be a locus of origin, assuming that the expansion of dogs from a domestication center resulted in the progressive loss of genetic diversity [21,22]. Furthermore, although dogs have occupied the New World for at least 9000 years based on the archeological record, they were not domesticated independently there. As shown by analysis of mtDNA sequencing from ancient remains, an evolutionarily distinct mtDNA lineage not closely related to New World wolves invaded North America over 10 000 years ago with the first humans [23]. This lineage was not found in a large survey of living breeds, which suggests that it was lost after European colonization along with the disintegration of Native American cultures [23]. In general, the domestic dog is an extremely close relative of the gray wolf, from which it differs by only ∼0.04% in nuclear coding-DNA sequence, and no dog mtDNA sequences have been found that show closer kinship to other canid species [10,21,22]. Therefore, the molecular genetic evidence does not support theories of non-wolf ancestry of domestic dogs. This result is consistent with the fossil record because the earliest dog remains are found alongside those of wolves [24].

In general, the usefulness of mtDNA sequence for historical inference is limited because, first, it represents the history of only the mitochondrial genome rather than that of the entire species; second, the lack of lineage-specific mutations and extensive crossing among breeds have constrained the resolution of phylogenetic analysis and mixed the history of various breeds [22]; third, the mitochondrial genome might be under relaxed evolutionary constraint and sustain more diversity than wild canids subjected to natural selection [25]. Consequently, multiple nuclear loci also need to be assayed to improve the documentation of genetic diversity and evolutionary patterns. An important locus in this regard is the MHC because the genes it contains have low mutation rates and, therefore, the levels of diversity might be a better reflection of founding size and subsequent backcrossing to wild wolves compared to other genes which have experienced mutations subsequent to domestication. MHC sequence variation in dogs has been found to be high [26–28]. For example, 42 distinct alleles were found in exon 2 of the *DRB1* genes [26]. Minimally, assuming no mutation, this suggests that 21 heterozygous individuals were involved in the domestication of dogs. However, the number of genomes is probably much larger given the limited samples size surveyed and the loss of alleles resulting from drift throughout the long history of domestication. Using these data and a specific demographic model, it was shown that the origin of dogs involved several populations and hundreds of individuals or, alternatively, a smaller number of founding individuals that were thereafter genetically augmented by backcrossing to wolves [26]. Past hybridization with wolves is also suggested by analysis of coat-color variants [29] and microsatellite loci [30]. The diversity of microsatellite loci across dog breeds and within indigenous dog populations is also high, and several diverse lineages of dogs have been identified [31–34]. Consequently, the model emerging from mtDNA, MHC and microsatellite data are that dogs have a diverse origin in East Asia that subsequently involved multiple contributions from several wolf populations through backcrossing [4,22,26]. Furthermore, once domesticated, dogs rapidly dispersed worldwide and, as a result, genetically divergent populations and breeds are found in Africa, Asia, the Arctic, Australia, the Middle East and historically, the New World [21,26,31,35,36].

Y-chromosome analysis has provided unique insights into the sex-specific patterns of evolution in the domestic dog. Although variability is high across breeds, Y-chromosome diversity within dog breeds is low and in a sample from 824 dogs from 50 breeds, 67 haplotype defined by six microsatellite markers has been found [37]. In this study, breeds shared haplotypes in accordance with known genealogical relationships but many haplotypes were also breed specific. Three African breeds and the Norwegian Elkhound did not share haplotypes with any breeds implying isolation, and phylogenetic analysis of Y-chromosome diversity suggested that these African breeds might be among the oldest and most divergent living dogs. However, sequence data are needed to confirm patterns based on this analysis and preliminary sequencing analysis in ten dogs suggested ample nucleotide diversity for phylogenetic analysis [38]. Samples sizes were too small for definitive conclusions about relationships but both African and some Asian dogs formed a divergent group in the phylogenetic tree based on these sequence data. Finally, the typing of

Y-chromosome markers and mtDNA sequence data allowed the comparison of male and female patterns of genetic diversity. A study of mtDNA sequence and Y-chromosome microsatellite loci found a strong sex bias in diversity patterns with males contributing less than females, a result in marked contrast to that observed in wolf societies, where both sexes contributed equally [39]. Moreover, paternal lineages are more differentiated than their maternal counterparts suggesting low rates of male gene flow. These finding are consistent with the 'popular sire' effect whereby a small subset of desirable males are favored for breeding and suggests that, in the evolution of most recent breeds, males have had a disproportionate effect on genetic diversity and evolution.

Modern breeds are closed gene pools

The evolution of the majority of dog breeds is a relatively recent phenomenon beginning with selective breeding practices in the Victorian era [1]. Nonetheless, these breeds demonstrate an immense variability in body size and form [40,41], in coat color, texture, length and thickness, and even tail shape and size [1]. Diversity is also evident in behavioral patterns, with breeds specialized for herding, guarding, agility, speed and companionship [1,3,42]. Phenotypic variation among dogs is partitioned into over 350 distinct breeds worldwide and are largely closed breeding populations that receive little genetic variation beyond that existing in the original founders [43,44]. These restrictive breeding practices reduce effective population size and increase overall genetic drift among domestic dogs, resulting in the loss of genetic diversity within breeds and greater divergence among them [31]. For example, variation among breeds accounts for 27% of total genetic variation as opposed to 5–10% among human populations [31]. In some breeds, genetic variation has been further reduced by catastrophic population declines during war or hard economic times, and are analogous to human populations of limited genetic variation used for disease-mapping studies. For example, the Leonberger dog breed population was reduced to five dogs in Europe following the second World War [45].

Because many breeds represent closed gene pools, they are likely to be genetically distinct [31,46]. For example, 99% of 414 dogs from 85 breeds were correctly assigned to their breed using data from 96 genetically unlinked microsatellite markers and individuals from the same breed were clustered together showing that they represent distinct gene pools [31]. Furthermore, breeds are organized into a distinct evolutionary hierarchy with only four primary groups, the most divergent of which contains nine ancient breeds from the Arctic, Asia, Africa and the Middle East, whereas the others correspond to mastiff, herding and modern European breeds (Figures 3,4) [31]. A recent study that increased the sample size to 132 breeds identified a fifth 'mountain' cluster containing large mountain dogs and a subset of spaniels (Figure 4) [47]. Most modern breeds, including all those from Europe, share a single exclusive common ancestor and show little phylogenetic structure, indicating a recent origin and limited interbreed hybridization [31]. Therefore, although breeds have heterogeneous origins, often involving a cross between distinct forms [1,45], and some interbreed crossing can occur thereafter, the genetic evidence clearly suggests nonetheless that they are genetically distinct. Finally, the study of individual genes has also provided new insight into breed history. Notably, a study of the multidrug resistance gene (*MDR1*) and four closely linked microsatellite markers were used to reconstruct the history of related breeds [48]. A single *MDR1* mutation and haplotype structure were shared in nine breeds that included seven herding breeds and two sight-hound subgroups suggesting identity by descent.

Varied but high levels of linkage disequilibrium

In the dog, several studies have assessed linkage disequilibrium (LD; i.e. the non-random association of genes at different loci) [10,49–52]. In a large scale study, 189 SNPs were typed in 20 unrelated dogs from each of five breeds for five unliked loci. A tenfold difference in LD among breeds having varied demographic histories was found [51] (Figure 5). LD extends for ~2 Mb in the dog compared with ~0.28 Mb in *Homo sapiens*. This difference suggests that as few as 10 000–30 000 SNPs in the dog, compared with 500 000 SNPs in *H. sapiens*, are required for whole-genome association studies, which assess the correspondence of a subset of markers with specific phenotypes [51]. Haplotype diversity in the dog was found to be more limited than expected with extensive haplotype sharing between breeds [51]. These exciting findings suggests that a single SNP map might generally be sufficient for genome mapping, regardless of the breed studied, and will greatly facilitate gene discovery. Recently, these results
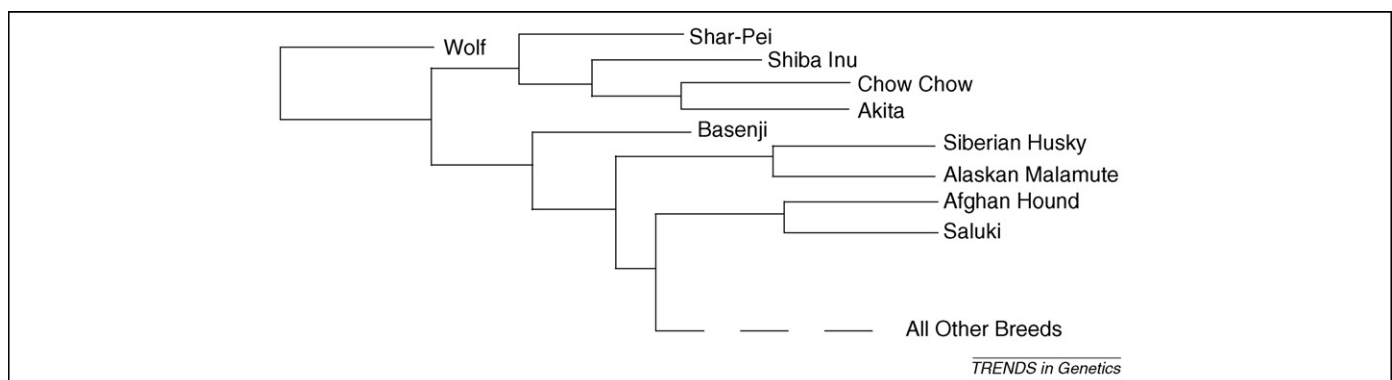

Figure 3. Genetic distance tree of dog groups. The genetic distance tree shows the early divergence of nine ancient breeds with the remaining, including those of European origin, being very closely related [31].
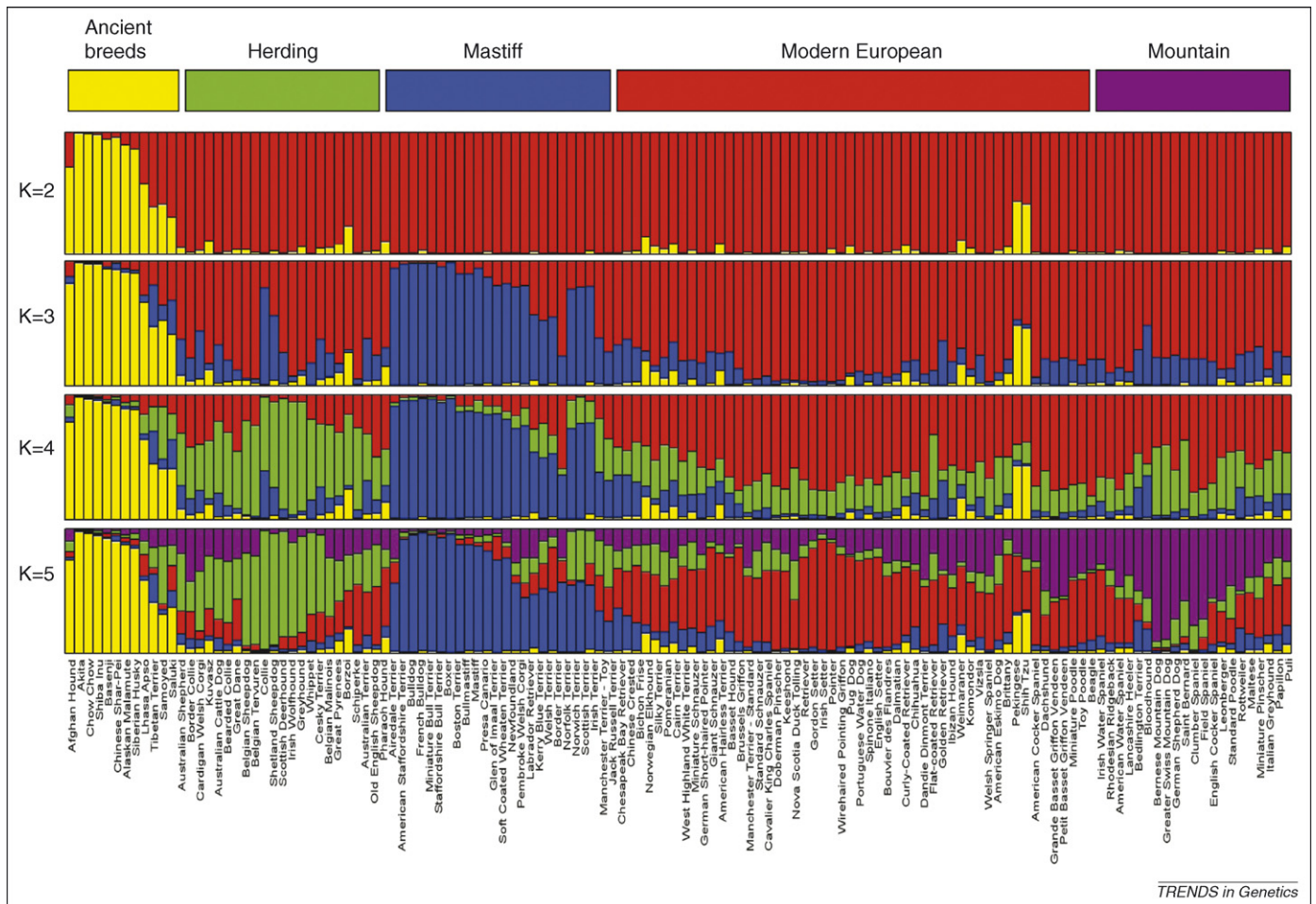
Figure 4. Genotype assignments for 96 dog breeds. Using the computer program Structure [19], the proportion of an individual's multilocus genotype belonging to 2 to 5 clusters (K) as indicated by colors was determined. Four well-distinguished groups of breeds are defined at K = 4, the most divergent of which contains nine ancient breeds from the Arctic, Asia, Africa and the Middle East (yellow) whereas the others correspond to mastiff (blue), herding (green) and modern European (red) breed groupings (Figure 3) [31] with a possible fifth group (purple) including the mountain dogs at K = 5 (Bernese Mountain Dog, Greater Swiss Mountain Dog, German Shepherd and Saint Bernard) [47].

have been corroborated and extended in a much larger study using ten breeds and 1200 SNPs [10,52] in which a pattern of long- and short-range LD were found, suggesting two distinct population bottlenecks. Short-range LD across breeds reflected an ancient bottleneck coincident with domestication whereas long-range patterns within breeds reflected more recent breed formation. Breeds with long-range LD will require fewer informative SNPs for mapping than those with short-range LD, however, the resolution will be greater in the latter. The identification of >2.1 million SNPs with high rates of polymorphism and the recent development of SNP genotype microarrays (http://www.broad.mit.edu/mammals/dog/caninearray.html) suggest that the dog is now an economical model for the mapping of common and complex diseases [4,10,52]. Because dog owners and breeders are highly motivated to improve the health of their breeds, and the American Kennel Club (AKC) is supportive of efforts to develop genetic tests for breed-specific diseases, collection of DNA samples and clinical data needed for mapping various disease traits has been a relatively easy task for the research community.

The discovery of new disease genes
With the availability of new genomic resources, many canine disease genes have now been mapped and, in some

cases, the underlying variant has been identified [53,54]. Specific examples include metabolic and endocrine disorders [55–57], blindness [58–60], cancer [61], neurological problems [62–64] in addition to skeletal and developmental disorders, including hip dysplasia, osteoarthritis [65–67] and others [47,68]. Genes implicated in cancer formation and progression are also an intense focus of research [4]. An exciting recent finding is the discovery that one type of cancer can be sexually transmitted between dogs [69] Based on analysis of microsatellite loci, the cancerous clone originated from 200 to 2500 years ago and derives from the insertion of a LINE, a kind of retrotransposon (see below), upstream of *c-myc*, a well-known oncogene. Also, expression of MHC molecules is downregulated or absent in tumors suggesting that they avoid immune detection. Although these canine tumors generally regress in a few months [70], a cancer with similar etiology in Tasmanian devils is fatal and threatens their survival [71]. Bioinformatic resources on inherited canine disease are now available, and the most complete listing is the IDID database (Inherited Disease in Dogs; see http://server.vet.cam.ac.uk/index.html) [72], which is modeled after OMIM (Online Mendelian Inheritance in Man). Approximately 360 genetic disorders found in humans are analogous to diseases in the dog
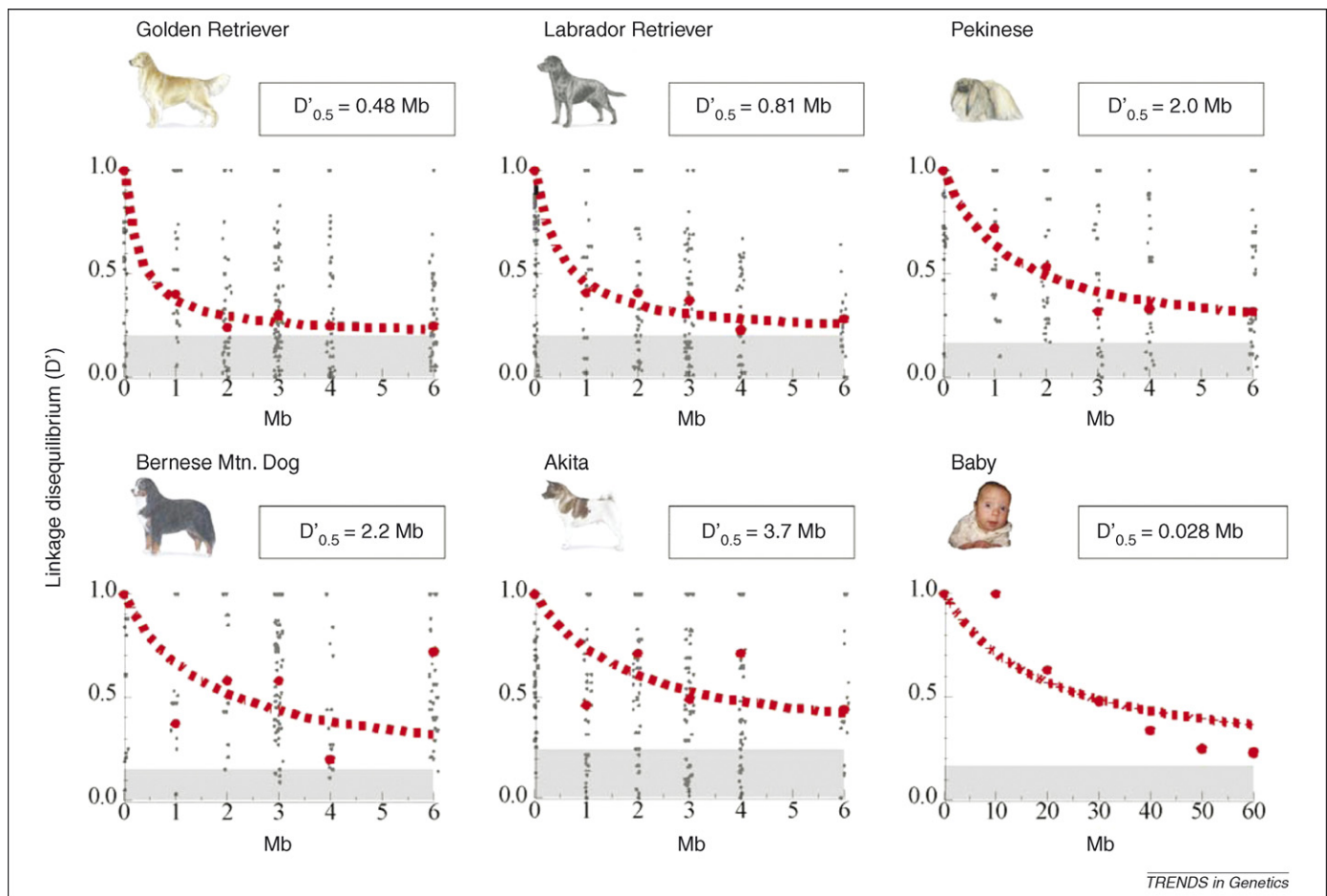
Figure 5. Linkage disequilibrium in five dog breeds. Linkage disequilibrium (LD) is the non-random association of alleles at different loci and, here, is measured by the statistic D' in five dog breeds for five loci [51]. Gray shading indicates background levels of LD. LD decay at the 50% level is indicated in the upper right of each panel. LD extends the farthest for the Akita, at 3.7 Mb, and is shortest for the Golden Retriever at 0.48 Mb. For the human, the comparable number is ∼0.028 Mb [51]. LD in dogs is ∼ten to 50 times greater than that observed in most human populations. Black dots are the D' statistic for each possible pair of points; the red dots are the median LD value within each bin of marker pair distances. The red dashes indicate the best-fitted curve to the median values [51].

[72] and often occur exclusively in a limited number of breeds.

Canine studies have revealed unique molecular mechanisms involved in genetic disease. For example, the genetic basis for progressive myoclonic epilepsy (PME) in miniature wirehaired dachshunds, a disorder that is analogous to Lafora disease in humans, involves a distinct genetic mechanism. Although mutations in the *EPM2B* gene are essential to both human and canine PME, the canine form of the disease stems from a bi-allelic expansion of a dodecamer (12-base pair) repeat found within the 5′ end of the single large exon of the gene [64]. Affected individuals carry 19 to 26 copies of the repeat sequence, rather than the expected two copies. This is the first example of a dodecamer repeat expansion associated with disease in any mammalian system and suggests a potential new mechanism for human disorders. It is not yet known whether repeat number correlates with the severity of the disease because of the small number of dogs typed.

Another example of a novel disease mechanism is the aberrant insertion of canine-specific short interspersed nuclear element (SINEs) in the dog genome. These elements are degenerate retrotransposons (self-replicating elements using an RNA intermediate) derived from a tRNA-Lys that occur frequently throughout the canine

genome and are often located in positions that affect gene expression [73]. For example, both canine narcolepsy and centronuclear myopathy (in the Labrador Retreiver) are caused by SINE insertions. A SINE insertion might also be responsible for the altered expression of the *SILV* gene that causes hearing and sight problems similar to those observed in Waardenburg syndrome in humans and is responsible for the merle phenotype found in several dog breeds [73,74]. SINEs appear to be highly mobile in the dog lineage suggesting a general role for retrotransposon insertions in canine disease and evolution [73].

The genetics of phenotypic diversity in dogs

Candidate gene and association studies are leading to a new understanding of the genetic basis for the differences in size and morphology between dog breeds and of sexual dimorphism [75]. These studies provide evidence for the existence of genetic 'tradeoffs' in which selection for one trait, such as forelimb length, leads to correlated changes in the length of other limb elements and in limb bone width [76]. One example of an early candidate gene study sequenced two genes, *MSX2* and *TCOF1,* which are expressed during craniofacial development, in ten different dog breeds that varied in both cranial and face shape [77]. A single amino-acid change in the TCOF1 protein

displayed a highly significant association with short and broad skulls. Negative results have also been published, such as a lack of association in dogs of dwarfism and a gene that causes this condition in humans [78]. A striking recent example of a successful candidate gene study is one involving a recessively acting mutation in the myostatin gene (*MSTN*) of whippets, which resulted in expression of a truncated protein producing a double muscled phenotype known as the 'bully' whippet [79]. These dogs are heavily muscled, similar to other mammals (e.g. mice and cattle) harboring *MSTN* mutations. Remarkably, individuals carrying only one copy of the mutation on average are more muscular and run significantly faster than wild-type individuals, demonstrating the potential functional significance of such mutations in the dog.

Non-classical genetic mechanisms have also been identified by candidate gene studies and might explain the high rate of breed differentiation. Analysis of skull measurements and repeat DNA in a panel of dog breeds found evidence for excessive purity and greater length polymorphism of simple two- and three-base pair repeats in dogs than humans suggesting that rapid repeat evolution in dogs accelerates the rate of morphologic change [80]. This hypothesis is supported by the observation that the size and the ratio of lengths of two tandem repeats in the *RUNX-2* gene, a regulator of osteoblast differentiation,

was correlated with the degree of dorsoventral nose bend (clinorhynchy) and midface length in a variety of breeds and among wild carnivore species [80,81]. This innovative proposal is also supported by the analysis of three-dimensional variation in the dog cranium and variation in several genes that are good candidates for influencing skull growth [82].

Body size is commonly regarded as a polygenic trait that is critical in the adaptive radiation of species and domesticated forms. A recent study used a three-pronged approach to identify a size gene in dogs [7]. First, a genomic scan of 460 Portuguese Water Dogs using 500 markers combined with an analysis of 91 skeletal measurements identified six quantitative trait loci (QTLs) that influenced size and shape; two of which mapped near *IGF-1* [83]. Second, genotyping of 104 SNPs near this gene in a large panel of small and large breeds showed evidence for a selective sweep near the *IGF-1* gene as small dog breeds displayed lower heterozygosity and higher allele frequency divergence at this locus (Figure 6). Third, all sampled small breeds displayed dominance of three closely related *IGF-1* haplotypes that were rare or absent in large breeds. Extensive resequencing and analysis of SNPs patterns near *IGF-1* suggested that the causative mutation was not located in the exons of the gene. One possibility is that the variation is caused by the insertion of a SINE within an 8.7-kb
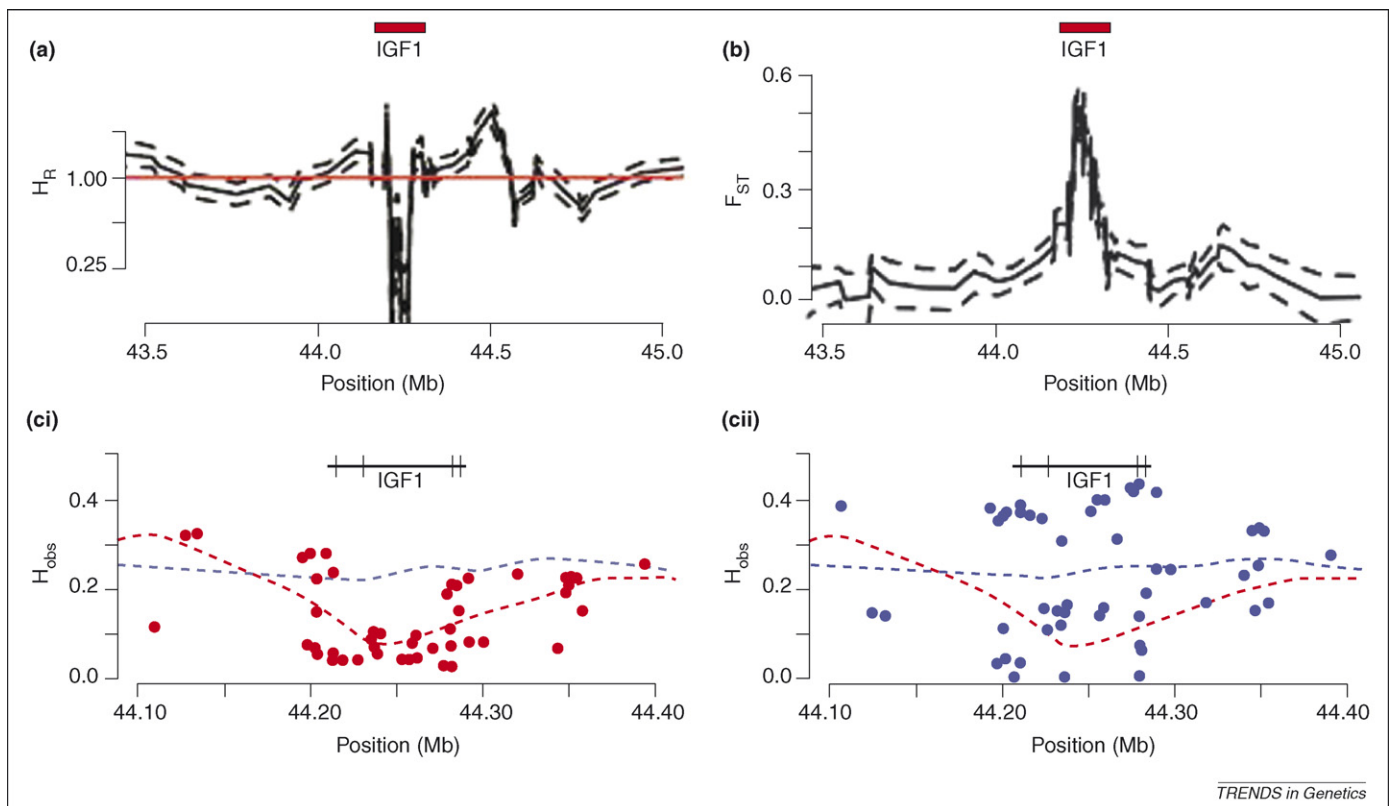

Figure 6. Evidence for a selective sweep near the *IGF-1* locus in small dog breeds. Represented are data for 104 SNPs typed near the *IGF-1* gene in 22 small (<9 kg) and 11 giant (>30 kg) dog breeds [7]. **(a)** The average ratio of heterozygosity (H_R) in small versus giant dogs for each SNP is shown. For example, if the proportion of heterozygotes was the same in large and small breeds, the ratio would equal one (indicated by the red line), whereas, if no heterozygotes were observed in small breeds, the ratio would equal zero. Small breeds have less than a quarter of the heterozygosity of large breeds near the *IGF-1* gene. **(b)** The difference in allele frequency between small and giant dogs for each SNP is shown. This is measured by the statistic F_{ST} (the proportion of genetic variation between populations). For both (a) and (b), a sliding 10-SNP window across *IGF-1* was used. The dashed lines delimit the 95% confidence intervals [7]. **(ci, cii)** The average observed heterozygosity (H_{Obs}) of SNPs very near *IGF-1* in small (red) breeds and giant (blue) breeds, shown in (ci) and (cii), respectively. The observed heterozygosity is lowest near *IGF-1* in the small breeds [as shown in (a) and (ci,cii)], whereas the allele frequency divergence is greatest near the *IGF-1* [shown in (b)], consistent with a selective sweep. Dashed lines are a best fit to the data. The *IGF-1* gene is shown as red or black bars with exons indicated by vertical lines. Reproduced, with permission, from Ref. [7].

region of *IGF-1*. The SINE was absent in a sample of 400 wolves and such insertions are known to cause differences in gene expression (see above). These results suggest that a single gene mutation occurred early in dog history and is common to all small breeds. Presumably, if this change had a large phenotypic effect, it might have been the 'gateway' change that led to increased selection for size diversification. As previously suggested by simulation studies [5], these results also provide empirical support for the use of selective sweep mapping to identify genes under positive selection.

The genetics of behavior

Dog breeds often display distinct behaviors and demonstrate behaviors not found in gray wolves [84]. However, our understanding of behavior has not advanced much beyond early pioneering work [3]. The development of reproducible techniques for assessing behavior and a focus on the pedigrees of dogs displaying aberrant behaviors will enable progress within this field. For example, some Bull Terriers have obsessive compulsive disease (OCD) phenotypes, such as tail chasing, which in other respects is similar to human OCD. Other breeds have distinctive aggressive traits, such as the 'Springer Rage Syndrome' [85].

Perhaps the most recent productive line of inquiry into the behavior of dogs concerns the comparative analysis of gene expression patterns. For example, analysis of gene expression patterns in 7762 genes from three regions of the brain in dogs, gray wolves and coyotes uncovered a unique pattern of gene expression in the hypothalamus of dogs whereas the amygdala and frontal cortex were less differentiated [86]. The hypothalamus controls specific emotional, endocrinological and autonomic responses of dogs, in particular behaviors related to survival, and consequently, these patterns of altered expression in the dog may represent an adaption to the novel challenges of human-altered environments [86]. Another recent expression study found associations between daytime sleepiness and neuropeptide precursor and cytokine signaling molecules [87]. Finally, behavioral traits in dogs have a genetic basis [3] and a recent study that scored 16 behavioral traits in 10 000 dogs from two breeds found a high degree of genetic correlation between traits ($r^2 = 0.79$) [88]. However, some specific traits, such as aggressiveness were less well correlated with other behavioral traits. At the genome level, only an association between canine dopamine receptors and aggression-related traits has been uncovered [89].

A remarkable parallel experiment in domestication of canids was conducted over the past few decades by Russian scientists led by D.K. Belyaev [90]. Beginning in the early 1960s, Belyaev and colleagues set out to develop a population of docile, or tame, red foxes from the Russian fur farm stocks. Progressive selection for tameness over many generations resulted in the creation of a domesticated fox, which eagerly accepts human companionship [90]. Surprisingly, among this population of foxes, phenotypic novelties common in the domestic dog such as mottling and depigmentation of the coat, floppy ears, curly tail and widened crania have emerged, albeit at low frequency [90]. These parallel changes can result from changes in the endocrine system that have generalized phenotypic effects [90]. Expression study of pooled mRNA from three areas of the brain in farm foxes (including those selected for tameness) relative to their wild conspecifics found highly significant differences, yet only 0.1% difference between tame and non-selected foxes [91]. Among the genes found to be differentially expressed, many are related to hemoproteins, which suggests that these genes might underlie behavioral differences [91]. However, probing the genetic basis of such differences has only begun with the advent of the meiotic linkage map [16].

What have we learned from the dog genome

The dog genome sequence and associated genomic resources will revolutionized the study of dog evolution, population structure and genetics. Indeed, a well-resolved phylogeny of the dog family Canidae was developed through the use of the genome sequence and tools of comparative genetics. Similarly, a more diverse analysis of dog genetic diversity, including the use of ancient DNA technology and new molecular markers, has shown that dogs are the oldest domesticated species and probably have genetic contributions from multiple gray wolf populations. The development of most modern breeds over the past few hundred years and limited interbreeding has resulted in pronounced genetic structure, such that nearly all breeds represent distinct gene pools. The advent of SNP markers and high-throughput genotyping methods has demonstrated that the dog genome is highly variable and is characterized by high levels of LD. In the future, these characteristics will facilitate gene identification by association in genome scans and selective-sweep mapping. Some traits, such as body size, might primarily reflect the fixation of genes of large phenotypic effect early in domestication. However, the underlying genetic mechanisms are diverse, ranging from simple point mutations to SINE insertions and the expansion and contraction of short nucleotide repeats. Because phenotypes and genetic disorders are common to subsets of breeds, the domestic dog offers unique evolutionary replicates than can be mined by molecular tools to undercover the genetic basis on phenotypic traits.

References

1 American Kennel Club Staff, eds (1998) *The Complete Dog Book*, Howell Book House
2 Stockard, C.R. ed. (1941) *The genetic and endocrinic basis for differences in form and behavior, as elucidated by studies of contrasted pure-line dog breeds and their hybrids*, The Wistar Institute of Anatomy and Biology
3 Scott, J.P. and Fuller, J.L., eds (1965) *Genetics and the Social Behaviour of the Dog*, University of Chicago Press
4 Ostrander, E.A. and Wayne, R.K. (2005) The canine genome. *Genome Res.* 15, 1706–1716
5 Pollinger, J.P. *et al.* (2005) Selective sweep mapping of genes with large phenotypic effects. *Genome Res.* 15, 1809–1819
6 Stinchcombe, J.R. and Hoekstra, H.E. (2007) Combining population genomics and quantitative genetics: approaches for finding the genes underlying ecologically important traits. *Heredity*, DOI: 10.1038/sj.hdy.6800937 (www.nature.com)

7 Sutter, N.B. *et al.* (2007) A single ancient *IGF1* allele causes small size in dogs. *Science* 316, 112–115

8 Kohn, M.H. *et al.* (2006) Genomics and conservation genetics. *Trends Ecol. Evol.* 21, 629–637

9 Wayne, R.K. *et al.* (1997) Molecular systematics of the Canidae. *Syst. Biol.* 46, 622–653

10 Lindblad-Toh, K. *et al.* (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438, 803–819

11 Geffen, E. *et al.* (1992) Phylogenetic relationships of the fox-like canids: mitochondrial DNA restriction fragment, site and Cytochrome *b* sequence analyses. *J. Zool. Lond.* 228, 27–39

12 Bardeleben, C. *et al.* (2005) A molecular phylogeny of the Canidae based on six nuclear loci. *Mol. Phylogenet. Evol.* 37, 815–831

13 Bardeleben, C. *et al.* (2005) Isolation and molecular evolution of the selenocysteine tRNA (CfTRSP) and RNase P RNA (CfRPPH1) genes in the dog family, Canidae. *Mol. Biol. Evol.* 22, 347–359

14 Wayne, R.K. (1989) Molecular and biochemical evolution of the Carnivora. In *Carnivore Behavior, Ecology and Evolution* (Gittleman, J.L., ed.), pp. 465–494, Cornell University Press

15 Wayne, R.K. and Morin, P.A. (2004) Conservation genetics in the new molecular age. *Front Ecol. Environ.* 2, 89–97

16 Kukekova, A.V. *et al.* (2007) A meiotic linkage map of the silver fox, aligned and compared to the canine genome. *Genome Res.* 17, 387–399

17 Aguilar, A. *et al.* (2004) High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *Proc. Natl. Acad. Sci. U. S. A.* 101, 3490–3494

18 Edwards, S.V. and Hedrick, P.W. (1998) Evolution and ecology of MHC molecules: from genomics to sexual selection. *Trends Ecol. Evol.* 13, 305–311

19 Seddon, J.M. *et al.* (2005) SNPs in ecological and conservation studies: A test in the Scandinavian wolf population. *Mol. Ecol.* 14, 503–511

20 Morin, P.A. *et al.* (2004) SNPs in ecology, evolution and conservation. *Trends Ecol. Evol.* 19, 208–216

21 Savolainen, P. *et al.* (2002) Genetic evidence for an East Asian origin of domestic dogs. *Science* 298, 1610–1613

22 Wayne, R.K. *et al.* (2005) Genetic analysis of dog domestication. In *Documenting Domestication: New Genetic and Archaeological Paradigms* (Zeder, M.E. *et al.*, eds), pp. 279–293, Smithsonian Institution Press

23 Leonard, J.A. *et al.* (2002) Ancient DNA evidence for old world origin of new world dogs. *Science* 298, 1613–1616

24 Sablin, M.V. and Khlopachev, G.A. (2002) The earliest ice age dogs: Evidence from Eliseevichi. *Curr. Anthropol.* 43, 795–799

25 Bjornerfeldt, S. *et al.* (2006) Relaxation of selective constraint on dog mitochondrial DNA following domestication. *Genome Res.* 16, 990–994

26 Vila, C. *et al.* (2005) Genes of domestic mammals augmented by backcrossing with wild ancestors. *Trends Genet.* 21, 214–218

27 Angles, J.M. *et al.* (2005) Frequency and distribution of alleles of canine MHC-II DLA-DQB1, DLA-DQA1 and DLA-DRB1 in 25 representative American Kennel Club breeds. *Tissue Antigens* 66, 173–184

28 Runstadler, J.A. *et al.* (2006) Dog leucocyte antigen class II diversity and relationships among indigenous dogs of the island nations of Indonesia (Bali), Australia and New Guinea. *Tissue Antigens* 68, 418–426

29 Schmutz, S.M. *et al.* (2007) Agouti sequence polymorphisms in coyotes, wolves and dogs suggest hybridization. *J. Hered.* 98, 351–355

30 Verardi, A. *et al.* (2006) Detecting introgressive hybridization between free-ranging domestic dogs and wild wolves (*Canis lupus*) by admixture linkage disequilibrium analysis. *Mol. Ecol.* 15, 2845–2855

31 Parker, H.G. *et al.* (2004) Genetic structure of the purebred domestic dog. *Science* 304, 1160–1164

32 Irion, D.N. *et al.* (2003) Analysis of genetic variation in 28 dog breed populations with 100 microsatellite markers. *J. Hered.* 94, 81–87

33 Irion, D.N. *et al.* (2005) Genetic variation analysis of the Bali street dog using microsatellites. *BMC Genet* 6, 6

34 Puja, I.K. *et al.* (2005) The Kintamani dog: genetic profile of an emerging breed from Bali, Indonesia. *J. Hered.* 96, 854–859

35 Leonard, J.A. *et al.* (2005) From wild wolf to domestic dog. In *The Dog and Its Genome* (Ostrander, E.A., Giger, U. and Lindblad-Toh, K., eds), pp. 95–118, Cold Spring Harbor Monograph Series

36 Savolainen, P. *et al.* (2004) A detailed picture of the origin of the Australian dingo, obtained from the study of mitochondrial DNA. *Proc. Natl. Acad. Sci. U. S. A.* 101, 12387–12390

37 Bannasch, D.L. *et al.* (2005) Y chromosome haplotype analysis in purebred dogs. *Mamm. Genome* 16, 273–280

38 Natanaelsson, C. *et al.* (2006) Dog Y chromosomal DNA sequence: identification, sequencing and SNP discovery. *BMC Genet.* 7, 45

39 Sundqvist, A-K. *et al.* (2006) Unequal contribution of sexes in the origin of dog breeds. *Genetics* 172, 1121–1128

40 Wayne, R.K. (1986) Cranial morphology of domestic and wild canids: the influence of development on morphological change. *Evolution* 40, 243–261

41 Wayne, R.K. (1986) Limb morphology of domestic and wild canids: the influence of development on morphologic change. *J. Morphol.* 187, 301–319

42 Svartberg, K. and Forkman, B. (2002) Personality traits in the domestic dog (*Canis familiaris*). *Appl. Anim. Behav. Sci.* 79, 133–155

43 Ostrander, E.A. *et al.* (2000) Canine genetics comes of age. *Trends Genet.* 16, 117–124

44 Sutter, N.B. and Ostrander, E.A. (2004) Dog star rising: the canine genetic system. *Nat. Rev. Genet.* 5, 900–910

45 Wilcox, B. and Walkowicz, C., eds (1995) *Atlas of Dog Breeds of the World* (5th edn), T.F.H. Publications

46 Koskinen, M.T. (2003) Individual assignment using microsatellite DNA reveals unambiguous breed identification in the domestic dog. *Anim. Genet.* 34, 297–301

47 Parker, H.G. *et al.* Breed relationships facilitate fine mapping studies: A 7.8-kb deletion segregates with collie eye anomaly across multiple dog breeds. *Genome Res*, DOI:10.1101/gr.6772807 (www.genome.org)

48 Neff, M.W. *et al.* (2004) Breed distribution and history of canine *mdr1-1Δ*, a pharmacogenetic mutation that marks the emergence of breeds from the collie lineage. *Proc. Natl. Acad. Sci. U. S. A.* 101, 11725–11730

49 Hyun, C. *et al.* (2003) Prospects for whole genome linkage disequilibrium mapping in domestic dog breeds. *Mamm. Genome* 14, 640–649

50 Lou, X.Y. *et al.* (2003) The extent and distribution of linkage disequilibrium in a multi-hierarchic outbred canine pedigree. *Mamm. Genome* 14, 555–564

51 Sutter, N.B. *et al.* (2004) Extensive and breed-specific linkage disequilibrium in *Canis familiaris*. *Genome Res.* 14, 2388–2396

52 Wade, C. *et al.* (2006) The dog genome: Sequence, evolution and haplotype structure. In *The Dog and Its Genome* (Ostrander, E.A., Giger, U. and Lindblad-Toh, K., eds), pp. 179–208, Cold Spring Harbor Monograph Series

53 Switonski, M. *et al.* (2004) The dog genome map and its use in mammalian comparative genomics. *J. Appl. Genet.* 45, 195–214

54 Parker, H.G. and Ostrander, E.A. (2005) Canine genomics and genetics: running with the pack. *PLoS Genet.* 1, e58

55 Safra, N. *et al.* (2006) Linkage analysis with an interbreed backcross maps Dalmatian hyperuricosuria to CFA03. *Mamm. Genome* 17, 340–345

56 van De Sluis, B. *et al.* (2002) Identification of a new copper metabolism gene by positional cloning in a purebred dog population. *Hum. Mol. Genet.* 11, 165–173

57 Chase, K. *et al.* (2006) Understanding the genetics of autoimmune disease: two loci that regulate late onset Addison's disease in Portuguese water dogs. *Int. J. Immunogenet.* 33, 179–184

58 Acland, G.M. *et al.* (1998) Linkage analysis and comparative mapping of canine progressive rod-cone degeneration (prcd) establishes potential locus homology with retinitis pigmentosa (RP17) in humans. *Proc. Natl. Acad. Sci. U. S. A.* 95, 3048–3053

59 Acland, G.M. *et al.* (1999) A novel retinal degeneration locus identified by linkage and comparative mapping of canine early retinal degeneration. *Genomics* 59, 134–142

60 Sidjanin, D.J. *et al.* (2002) Canine CNGB3 mutations establish cone degeneration as orthologous to the human achromatopsia locus ACHM3. *Hum. Mol. Genet.* 11, 1823–1833

61 Lingaas, F. *et al.* (2003) A mutation in the canine BHD gene is associated with hereditary multifocal renal cystadenocarcinoma and nodular dermatofibrosis in the German Shepherd dog. *Hum. Mol. Genet.* 12, 3043–3053

62 Lin, L. *et al.* (1999) The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene. *Cell* 98, 365–376

63 Lingaas, F. *et al.* (1998) Genetic markers linked to neuronal ceroid lipofuscinosis in English setter dogs. *Anim. Genet.* 29, 371–376

64 Lohi, H. *et al.* (2005) Expanded repeat in canine epilepsy. *Science* 307, 81

65 Chase, K. *et al.* (2004) Bilaterally asymmetric effects of quantitative trait loci (QTLs): QTLs that affect laxity in the right versus left coxofemoral (hip) joints of the dog (*Canis familiaris*). *Am. J. Med. Genet. A.* 124, 239–247

66 Chase, K. *et al.* (2005) Genetic regulation of osteoarthritis: A QTL regulating cranial and caudal acetabular osteophyte formation in the hip joint of the dog (*Canis familiaris*). *Am. J. Med. Genet. A.* 135, 334–335

67 Todhunter, R.J. *et al.* (2005) Quantitative trait loci for hip dysplasia in a cross-breed canine pedigree. *Mamm. Genome* 16, 720–730

68 Clark, L.A. *et al.* (2005) Linkage analysis and gene expression profile of pancreatic acinar atrophy in the German Shepherd Dog. *Mamm. Genome* 16, 955–962

69 Murgia, C. *et al.* (2006) Clonal origin and evolution of a transmissible cancer. *Cell* 126, 477–487

70 vonHoldt, B.M. and Ostrander, E.A. (2006) The singular history of a canine transmissible tumor. *Cell* 126, 445–447

71 Bostanci, A. (2005) Wildlife biology. A devil of a disease. *Science* 307, 1035

72 Sargan, D.R. (2004) IDID: inherited diseases in dogs: Web-based information for canine inherited disease genetics. *Mamm. Genome* 15, 503–506

73 Cordaux, R. and Batzer, M.A. (2006) Teaching an old dog new tricks: SINEs of canine genomic diversity. *Proc. Natl. Acad. Sci. U. S. A.* 103, 1157–1158

74 Clark, L.A. *et al.* (2006) Retrotransposon insertion in SILV is responsible for merle patterning of the domestic dog. *Proc. Natl. Acad. Sci. U. S. A.* 103, 1376–1381

75 Lark, K.G. *et al.* (2006) Genetic architecture of the dog: sexual size dimorphism and functional morphology. *Trends Genet.* 22, 537–544

76 Kharlamova, A.V. *et al.* (2007) Genetic regulation of canine skeletal traits: trade-offs between the hind limbs and forelimbs in the fox and dog. *Integr. Comp. Biol.* 47, 373–381

77 Haworth, K.E. *et al.* (2001) Canine TCOF1; cloning, chromosome assignment and genetic analysis in dogs with different head types. *Mamm. Genome* 12, 622–629

78 Young, A.E. *et al.* (2006) Deletions in the COL10A1 gene are not associated with skeletal changes in dogs. *Mamm. Genome* 17, 761–768

79 Mosher, D.S. *et al.* (2007) A mutation in the myostatin gene increases muscle mass and enhances racing performance in heterozygote dogs. *PLoS Genetics* 3, e79

80 Fondon, J.W., 3rd and Garner, H.R. (2004) Molecular origins of rapid and continuous morphological evolution. *Proc. Natl. Acad. Sci. U. S. A.* 101, 18058–18063

81 Laidlaw, J. *et al.* (2007) Elevated basal slippage mutation rates among the canidae. *J. Hered.* 98, 452–460

82 Fondon, J.W., 3rd and Garner, H.R. (2007) Detection of length-dependent effects of tandem repeat alleles by 3-D geometric deomposition of craniofacial variation. *Dev. Genes Evol.* 217, 79–85

83 Chase, K. *et al.* (2002) Genetic basis for systems of skeletal quantitative traits: principal component analysis of the canid skeleton. *Proc. Natl. Acad. Sci. U. S. A.* 99, 9930–9935

84 Hare, B. *et al.* (2002) The domestication of social cognition in dogs. *Science* 298, 1634–1636

85 Reisner, I.R. *et al.* (2005) National survey of owner-directed aggression in English Springer Spaniels. *J. Am. Vet. Med. Assoc.* 227, 1594–1603

86 Saetre, P. *et al.* (2004) From wild wolf to domestic dog: gene expression changes in the brain.. *Brain Res. Mol. Brain Res.* 126, 198–206

87 Linberg, J. *et al.* (2007) Reduced expression of TAC1, PENK and SOCS2 in Hcrtr-2 mutated narcoleptic dog brain. *BMC Neurosci.* 8, 34

88 Saetre, P. *et al.* (2006) The genetic contribution to canine personality. *Genes Brain Behav.* 5, 240–248

89 Ito, H. *et al.* (2004) Allele frequency distribution of the canine dopamine receptor D4 gene exon III and I in 23 breeds. *J. Vet. Med. Sci.* 66, 815–820

90 Trut, L.N. *et al.* (2004) An experiment on fox domestication and debatable issues of evolution of the dog. *Genetika* 40, 794–807

91 Lindberg, J. *et al.* (2005) Selection for tameness has changed brain gene expression in silver foxes. *Curr. Biol.* 15, R915–R916