

REVIEW

# The domestic dog: man's best friend in the genomic era

Adam R Boyko\*

## Abstract

The domestic dog genome - shaped by domestication, adaptation to human-dominated environments and artificial selection - encodes tremendous phenotypic diversity. Recent developments have improved our understanding of the genetics underlying this diversity, unleashing the dog as an important model organism for complex-trait analysis.

In the 5 years since the publication of the genome sequence of the domestic dog (*Canis familiaris*) [1], our understanding of dog origins and evolution has improved considerably. Before this genome sequence was available (the breed chosen to sequence was a boxer, sequenced at 7.8 coverage), canine genomic analysis relied on linkage and radiation hybrid maps encompassing at most 3,000 to 4,000 markers, or approximately 1 Mb resolution of the genome [2,3]. Comparing the boxer genome to an earlier 1.5 coverage poodle genome [4] and low-coverage sequencing from nine other dog breeds and wolves, scientists have now cataloged over 2.5 million single nucleotide polymorphisms (SNPs) [1]. Genotyping technology has enabled tens of thousands of these SNPs to be typed at a modest cost (approximately US\$200 per sample for a 20,000 to 60,000 marker array), giving unprecedented resolution of canine population genetics [5-7] and leading to the rapid identification of loci underlying complex and Mendelian traits (see Additional file 1).

The phenotypic diversity of the world's 350 to 400 dog breeds is mirrored in their genetic diversity. Although most breeds have existed for less than two centuries, the level of diversity ( $F_{ST}$ ) in dogs is about twice that found in humans ( $F_{ST}$  averages 0.28 among dog breeds) [6,8]. In an effort to create a perfect companion, dog fanciers have embarked on an 'experiment', faithfully rearing, selecting,

breeding and adapting, generation after generation, millions of pedigreed animals with genetically based proclivities and susceptibilities awaiting genomic interrogation. The recent release of the new 170K Illumina HD canine SNP array coupled with an improved genome assembly (canFam3) and advances in targeted and high-throughput DNA and RNA sequencing will surely accelerate the pace of canine genomics in the near future, expanding our understanding of evolution in dogs and their utility as a model genetic system.

## From pack to pet

Because of the incredible diversity of modern dogs and the number of derived characteristics distinguishing dogs from their ancestors, determining the ancestor of dogs required genetic data. From Charles Darwin to Konrad Lorenz, early researchers believed that admixture with multiple canid species, including jackals, was necessary to explain domestic dog diversity [9-11]. However, modern mitochondrial DNA (mtDNA) analysis has instead shown that the gray wolf (*Canis lupus*) was the sole ancestor of modern dogs [12,13].

Shedding light on the specifics of the location and timing of dog origins has been difficult. The earliest definitive archeological evidence for dog burials is around 11,500 years ago in Israel [14,15], although evidence also exists for dog burials in Germany around 14,000 years ago [16]. Burials or artistic depictions of dogs before this are lacking from the archeological record, suggesting a relatively recent origin for dogs (less than 16,000 years ago) or at least a major shift in dog anatomy and/or human interaction at this time. Older dog-like canid fossil remains exist, but they are difficult to group unequivocally as being early dogs or small wolves (for example [17,18]).

In the same way, genetic studies have yet to provide a definitive account of dog origins. Initial estimates of a dog-wolf divergence of more than 100,000 years ago based on mtDNA sequence data relied on the flawed assumption that each modern dog mtDNA haplogroup descended from a single wolf mtDNA lineage [13]. Research by Savolainen and colleagues [19,20], however, gives a more recent estimate for dog domestication

\*Correspondence: aboyko@stanford.edu  
Department of Genetics, Stanford University, 300 Pasteur Drive, Stanford, CA 94305-5120, USA

(5,400 to 16,300 years ago) using a domestication model that allows for dozens of founding maternal lines. These studies conclude that dogs probably originated in East Asia south of the Yangtze River, on the basis of the high level of mtDNA diversity found in extant dogs in this area, but they make a crucial assumption that there was a single origin for dogs and that introgression from wolves in Asia has been minimal. If mtDNA haplogroups entered the dog lineage at different times and places, the true timing and location of domestication could differ substantially from those inferred from a single-origin model [21]. This seems to be the case: using whole-genome genotyping data from purebred dogs and wolves, Vonholdt *et al.* [7] found evidence for allele sharing between Asian dog breeds and Asian wolves (and between European dog breeds and European wolves), suggesting non-trivial levels of regional introgression or multiple founding events. More sophisticated models using whole-genome data from both dogs and wolves are needed to disentangle the complex demographic process underlying dog domestication.

In part because of the difficulties in resolving the location and timing of dog domestication, considerable debate surrounds the roles that humans, wolves and early dogs played in the process. Certain populations of gray wolves probably 'pre-adapted' themselves for domestication by scavenging from human settlements, an activity that not only placed wolves at close quarters to people, but also selected for reduced fearfulness. Individuals willing to forage in proximity to humans were better able to exploit their food sources, particularly if these individuals learned to read human cues (Box 1). Whether increased competition from bow-wielding human hunters or the emergence of rubbish dumps in the villages of these hunter-gatherers caused the shift to scavenging is unclear [22,23], but competition with humans, or persecution by them, certainly could have been an important isolating force keeping feral wolves from wiping out early proto-dog populations by swamping gene flow.

Notably, dog domestication occurred before the advent of agriculture [22]. The appearance of agriculture shortly after the origin of dogs suggests that dog domestication itself may have been an important precursor to the transformation of humans into agriculturalists. However, without knowing what roles early dogs played in human settlements, this hypothesis is highly speculative. Once agriculture was established, scavenging around human settlements became a highly profitable endeavor, and dogs rapidly spread throughout the world [24]. At some point, people began using these dogs as sentries, food sources and companions, but whether domestication was complete by this time (as proposed in Coppinger's model of dog 'self-domestication' [23]) or whether directed human selection was required to make dogs fully

### **Box 1. Dogs' innate ability to read human cues.**

One of the most remarkable aspects of dog cognition is their ability to 'read' people. Like humans, dogs seem naturally inclined to use cues like pointing or gazing to find hidden food sources in object-choice tests, unlike wolves and non-human primates. Dogs do not outperform these species in non-social tasks [65], suggesting that domestication itself has selected for human-like social responsiveness.

In a typical experimental set-up (reviewed in [66]), dogs are allowed to choose between two inverted cups, one of which has been randomly chosen to be baited with food hidden under the cup. A human experimenter stationed between the cups gives a signal (such as gazing at the baited cup while pointing), and the proportion of times the animal chooses the correct cup is recorded for each cue. Dogs of all breeds seem remarkably adept at these experiments [67].

Early studies concluded that making use of human social cues was a skill present in dogs at a very early age, regardless of upbringing, and absent in wolves [65]. Subsequent work, however, showed that both dog and wolf performance on this task is highly dependent on experience, environment and experimental set-up - pet dogs do not perform well outdoors [68], shelter dogs do better with obvious cues (point + gaze) than less obvious one (point or gaze separately) [69], and neither dogs nor wolves excel when they are separated from the cue-giving experimenter by a fence [68]. Furthermore, the evidence for cue use in puppies less than 16 weeks old is ambiguous [70,71].

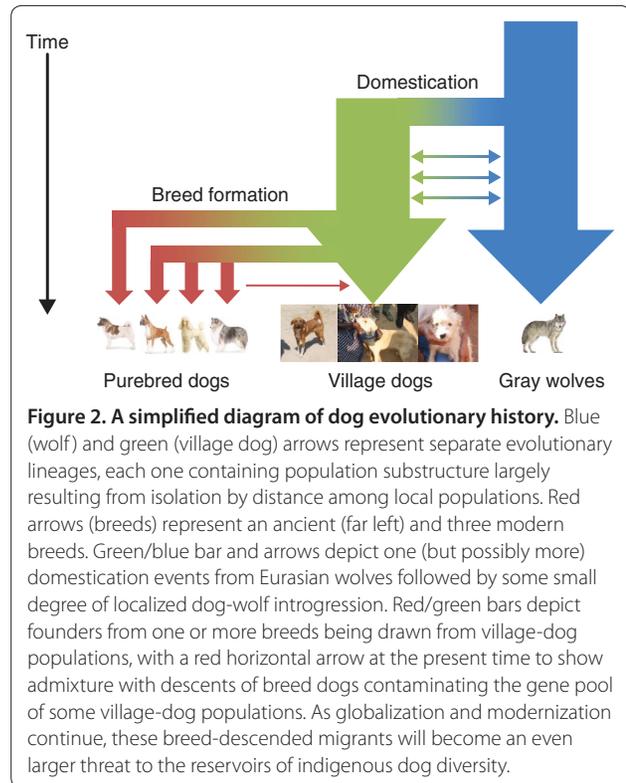
R Boyko and colleagues (personal communication) found differences in performance between shelter dogs in Western and non-Western countries. In Western countries where shelter dogs were raised in homes or shelters, dogs successfully followed human point and gaze cues whereas in non-Western countries where shelter dogs were raised on the streets, they did not. This supports the hypothesis that socialization depends on a critical developmental window with the process of domestication acting to lengthen the window [23]. Natural selection has clearly given dogs the cognitive abilities and temperament to excel at reading human signals, but early socialization is still critical for these skills to develop [72].

domesticated (as proposed in Crockford's model of 'classic domestication' [25]) is still debated.

In the Victorian era (approximately 200 years ago), the pace of breed selection expanded as hundreds of breeds were created, and registrations and pedigree tracking were used to ensure closed populations [26,27]. Much of the phenotypic variation present in modern dogs was driven by the whims of these fanciers and their artificial selection for distinctive phenotypes (Figure 1). As a consequence of this controlled breeding, the pace of both selection and drift have accelerated in these lineages, offering an excellent opportunity for geneticists to map regions underlying phenotypic variation. From an evolutionary standpoint, the genomes of modern



**Figure 1. Phenotypic diversity across 9 of the approximately 400 modern dog breeds.** From top left to bottom right: Basenji, Newfoundland, Chihuahua, Standard Poodle, Australian Cattle Dog, Afghan Hound, Bull Terrier, Greyhound, and English Mastiff. Photos are used under Creative Commons from fugzu, alicjap, Kjunstorm, greg westfall, 3Dobes, diveofficer, Just chaos, msmornington, and claudiogennari, respectively.



purebred domestic dogs are a mosaic consisting of regions of short selective sweeps resulting from domestication, longer regions underlying recent breed-specific sweeps, and the rest of the genome which, conversely, has undergone a relaxation of selective constraint as the forces of natural and sexual selection have weakened in purebred dogs [28,29].

In contrast to purebred dogs, most dogs today still live much as they have for millennia, as semi-feral human commensals known as ‘village dogs’ [23]. Importantly, most extant village-dog populations descend from ancient village-dog populations and are not significantly contaminated by recent admixture from modern breeds [30] (Figure 2). In fact, these populations of random-breeding village dogs provided the founders of ancient and modern dog breeds. The ease of trait mapping in purebred dogs coupled with the ability to study the history of these adaptive alleles in natural village-dog populations make dogs a uniquely powerful system for mapping mammalian phenotypes and understanding the genetic basis of adaptive evolution.

### Trait mapping in the domestic dog

Over the past two centuries, dog breeders have fortuitously generated a powerful system for mapping genes underlying phenotypic variation [31]. Variants with large, observable effects, like hairlessness or dwarfism, were strongly selected, reaching fixation within certain breeds.

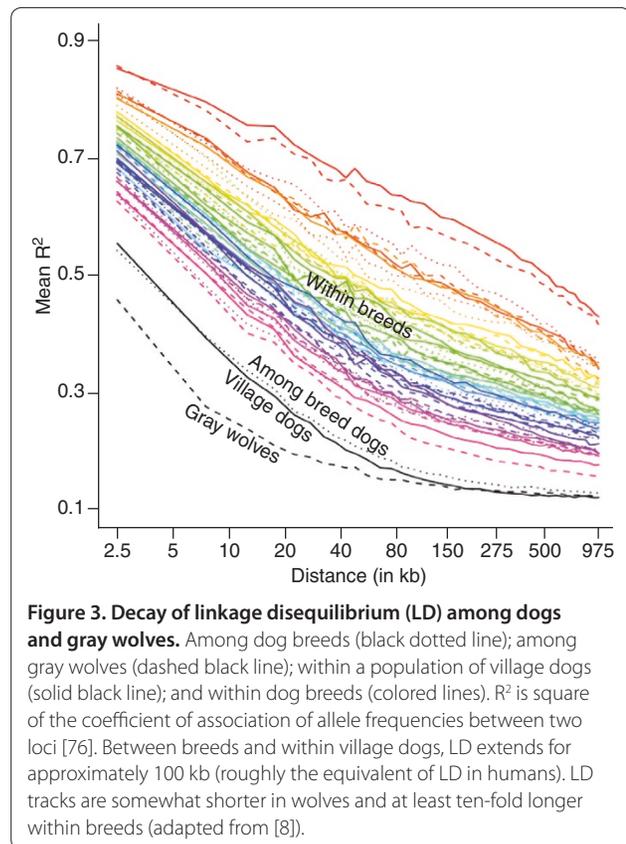
The canine genome itself can contribute to generating such variants. Canidae have been shown to exhibit elevated levels of DNA slippage contributing to microsatellite diversity [32] and have a highly active SINE (short interspersed nuclear element), SINE\_Cf, akin to Alu SINE repeats found in primates, that segregates at a rate ten-fold higher than the SINE rate in humans [1,33]. However, these types of variants underlie only a small proportion of the casual mutations that have been found to date in dogs [34], suggesting that population structure and selection for extreme phenotypes, not canine-specific mutational biases, are the main forces driving the rapid diversification of dogs [35].

Often, several breeds are characterized by a phenotype in which the causal genetic variant in each breed is identical because of shared descent or directed introgressive breeding for the trait [36]. For example, disproportionate dwarfism (chondrodysplasia) is a defining characteristic of at least 19 breeds, including dachshunds, pekingese, and basset hounds. A genome-wide association study (GWAS) using 797 dogs from eight chondrodysplastic breeds and 64 nonchondrodysplastic breeds found a region of canine chromosome 18 (CFA18) corresponding to a 5 kb expressed retrotransposon insert of fibroblast growth factor 4 (*FGF4*) unique to the chondrodysplastic breeds [37]. Multi-breed GWAS have likewise shown that short-snouted (brachycephalic) breeds share a haplotype near thrombospondin-2 (*THBS2*) on CFA1

[8,36] and floppy-eared breeds share a haplotype near methionine sulfoxide reductase B3 (*MSRB3*) on CFA10 [8], although the causal mutations in these regions remain undiscovered.

Many traits are fixed in some breeds and segregating in others, providing an opportunity for a multi-stage mapping approach. First, GWAS within a breed segregating for a trait can easily identify the genomic region underlying the trait because of long-range linkage disequilibrium (LD; the nonrandom association of alleles in a genomic region) within a breed, although this region will often be several megabases long and encompass several genes (Figure 3). Subsequent fine mapping can be done with GWAS across breeds to identify the smallest shared haplotype in the region, followed by sequencing across the region to reveal candidate causal variants [5,6]. Such an approach was successfully employed by Cadieu *et al.* [38] to find the missense transition in exon 2 of keratin-71 (*KRT71*) responsible for curly fur, the missense transversion in exon 1 of fibroblast growth factor-5 (*FGF5*) associated with long fur, and the 167-bp deletion in the 3' untranslated region of R-spondin-2 (*RSPO2*) believed to cause the 'furnished coat' phenotype (the presence of features such as a mustache or long eyebrows). Similar multi-stage mapping studies have identified, for example, a SINE\_Cf insertion in the gene for insulin-like growth factor 1 (*IGF1*) associated with small body size [39], a 7-bp frameshifting insert responsible for canine ectodermal dysplasia in hairless breeds [40], and a 3-bp in-frame insertion in the second exon of the canine beta-defensin 103 gene (*CBD103*) leading to black coat color [41].

It is perhaps surprising that for so many traits, the causal variants are identical across often distantly related breeds. In some cases (such as floppy ears or small body size), the causal variant is ancient and segregating at high frequency in natural village-dog populations. Thus, the variant was present in the founder population for several breeds and subsequently selected in parallel in a subset of them. In other cases, the causal variant might have been introduced directly from one breed to another (for example, the ridgeback phenotype caused by an identical 133-kb duplication of *CFA18* in Thai and Rhodesian ridgebacks [42]). Currently, little is known about the evolutionary age and history of most causal alleles, although haplotype analysis and/or the genotyping of diverse village-dog and wolf populations can be highly informative. Variants with a global distribution probably arose early in dog evolutionary history, whereas geographically restricted variants are expected to be more recent. Interestingly, two ancient causal haplotypes (the 'small dog' *IGF1* haplotype and the chondroplasia *fgf4* variant) both appear to have arisen on ancestral haplotypes associated with Middle Eastern or European gray wolves and not East Asian gray wolves [37,43].



**Figure 3. Decay of linkage disequilibrium (LD) among dogs and gray wolves.** Among dog breeds (black dotted line); among gray wolves (dashed black line); within a population of village dogs (solid black line); and within dog breeds (colored lines).  $R^2$  is square of the coefficient of association of allele frequencies between two loci [76]. Between breeds and within village dogs, LD extends for approximately 100 kb (roughly the equivalent of LD in humans). LD tracks are somewhat shorter in wolves and at least ten-fold longer within breeds (adapted from [8]).

For traits found in just one or a few breeds, linkage and GWAS have also been highly successful despite the increased difficulty of fine mapping across large linkage blocks (see [34] for a recent review). Furthermore, haplotype-based and  $F_{ST}$ -based methods to detect recent selection can improve the power of traditional GWAS and find genomic regions underlying selected phenotypes even if they are only present in a single breed. For example, excessive skin wrinkling (found almost exclusively in the Shar Peis breed) was mapped to hyaluronan synthase 2 (*HAS2*) using an  $F_{ST}$ -based approach [44]. The recent introduction of higher-density genotyping arrays (such as the 170K Illumina HD array) should further improve the power of these methods from what was possible with older (20 to 60K) SNP arrays.

### The dog as a model of human genetic disease

An important application of trait mapping in dogs is the discovery of variants underlying genetic disease. Traditional GWAS using well phenotyped cases and controls have proved highly successful in finding regions containing causal variants underlying more than 70 Mendelian diseases in dogs (see Additional data file 1). Many of these diseases have close human analogs: genetic mapping of narcolepsy [45], copper toxicosis [46,47] and ichthyosis (C André, personal communication) in dogs

led to the discovery of new causal variants in the human ortholog affecting the human disease. Because diseases that are complex and/or rare in humans are often monogenic and common in some dog breeds, and because of dogs' relatively large family sizes, dogs are particularly useful for identifying new candidate genes underlying such disorders.

In addition, dogs occupy a valuable intermediate position between the human and mouse genetic systems, increasing their utility as a model system [48]. Despite mice and humans last sharing a common ancestor more recently than dogs and humans (approximately 75 million years ago versus approximately 87 million years ago), the faster rate of evolution in the rodent lineage means that there is less sequence divergence between human and dogs than between humans and mice, and therefore approximately 650 Mb more human sequence can be syntetically aligned to the dog genome than to that of the mouse [1]. Furthermore, dogs are more similar to humans than are mice in terms of body size, longevity and behavior, which also leads to similarity in various genetic pathologies. Finally, dogs have co-habited with humans longer than any other domestic animal, sharing our nutritional and pathogenic environment during our species' unprecedented shift from a hunter-gatherer lifestyle to agriculture. Some human adaptations to this dramatic environmental shift that contribute through antagonistic pleiotropy to disease (such as highly reactive immune systems that protect from infectious disease but predispose individuals to autoimmune disorders [49,50]) might have evolved in parallel in dogs.

Strong artificial selection has contributed to the diversity of disorders exhibited in dogs. Independent, severe founder effects for each breed cause diseases that are at extremely low prevalence in natural dog populations to, by chance, reach appreciable frequency in one or a few breeds, either from the founder bottleneck itself or through the subsequent propagation of popular sires harboring the variant [51]. In particular, some recessive disorders caused by loss-of-function mutations and some cancers can be rare in humans but common in certain dog breeds (for example, osteosarcoma [52] and amyotrophic lateral sclerosis (ALS)-like canine degenerative myelopathy [53]). Diseases can also be associated with variants selected for a pleiotropic effect - for example, dermoid sinus, a neural tube defect in dogs, is caused by the same variant that produces the ridgeback coat phenotype [42]. Finally, the large selective sweeps containing artificially selected variants can also harbor linked disease variants that hitchhiked to high frequency during the sweep. It is perhaps not coincidental that the gene underlying lens dislocation in terriers [54] is adjacent to a gene implicated in controlling body size among small breed dogs (B Hoopes, personal communication).

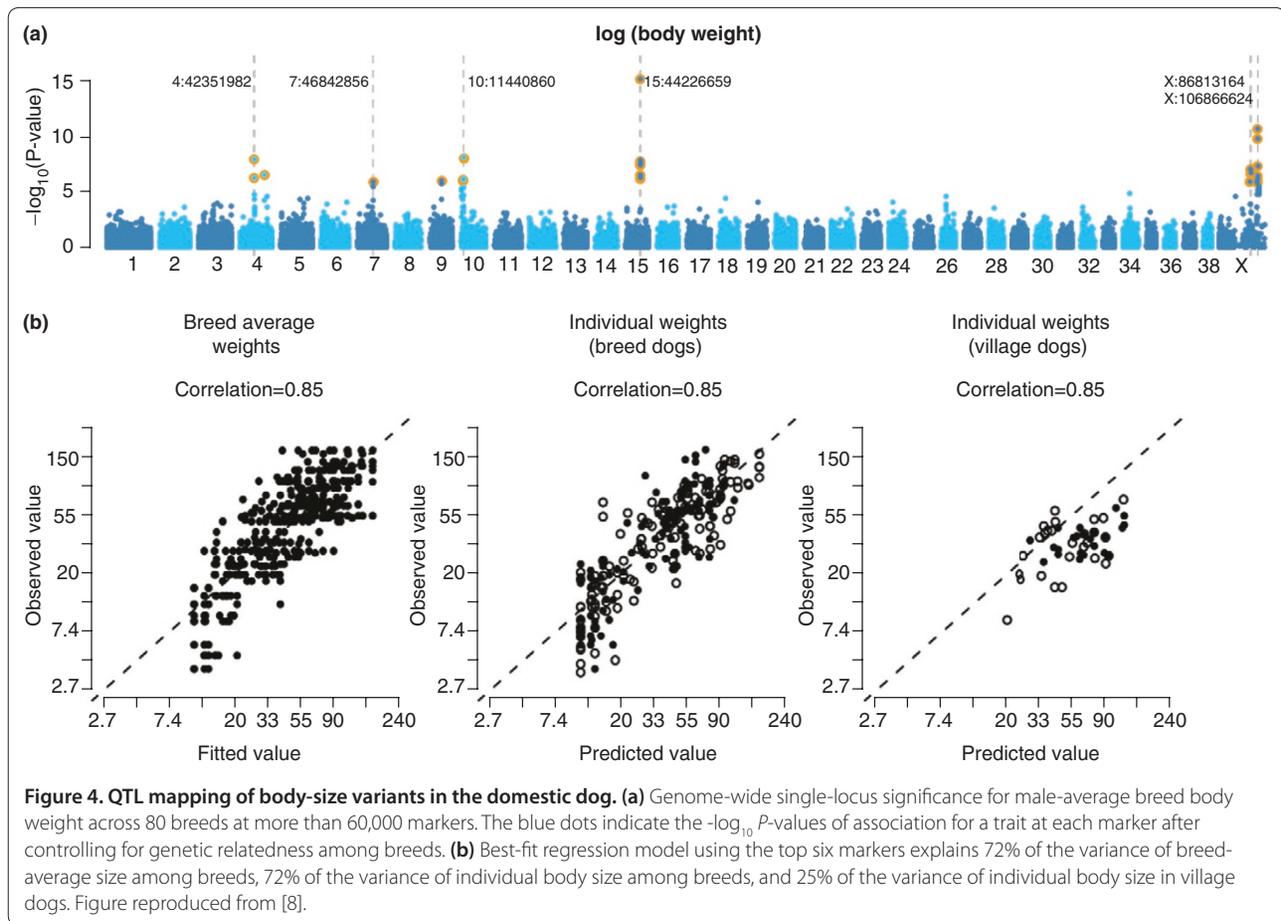
### The genetic architecture of canine phenotypic variation

Mapping causal variants for quantitative traits is generally more difficult than mapping monogenic traits, if only because accurate phenotyping and controlling for genetic background can be problematic. Nevertheless, GWAS have elucidated dozens of regions underlying quantitative variation in dogs, although most of the causal variants in these regions remain undiscovered. As next-generation sequencing costs decline and more and more canine genomes are sequenced, candidate loci in these regions and others should begin to emerge, improving our understanding of the genetic basis of phenotypic variation in this system. In particular, advances in targeted sequence capture (seq-cap) and DNA barcoding currently enable efficient sequencing and analysis of multiple individuals across candidate qualitative trait loci (QTL) regions [55,56].

Several studies have performed multi-breed GWAS for body weight and morphological measurements [8,44,57, 58]. Despite great differences in the breeds and marker sets used by each, the results are highly consistent for several traits. All studies identified *IGF1* as the primary locus affecting body weight and also consistently found other significant QTLs not yet associated with causal variants on *CFA7* near a SMAD family gene (*SMAD2*), *CFA10* near high-mobility group protein-A2 (*HMGGA2*) and *CFA34* near IGF2 mRNA-binding protein (*IGF2BP2*). After controlling for allometry, height was principally controlled by the *fgf4* retrotransposon on *CFA18* and also by an unknown variant near *RNF4* and *MXD* on *CFA3* in all four studies.

GWAS across 80 breeds for 50 body and skeletal dimensions has revealed strong evidence that each trait is primarily explained by a few loci of major effect [8]. Across all 50 traits, the top three QTLs for each explained, on average, 67% of the phenotypic variation (40% of the variation after controlling for allometry; Figure 4). In contrast, the top 180 QTLs for human height only explain approximately 10% of the variation in that trait [59]. Several non-mutually exclusive hypotheses can account for this simplification of the genetic architecture, including the reduction in allelic heterogeneity that also characterizes monogenic traits/disorders in dogs and the impact of strong, diversifying selection on genetic architecture (Box 2).

Does this simplified genetic architecture characterize other complex canine phenotypes, including those associated with behavior, longevity and common multifactorial diseases? Analyses of highly differentiated genomic regions among breeds (which might represent the 'low-hanging fruit' for across-breed mapping studies) show that, overwhelmingly, the highest differentiated regions correspond to known morphological traits involving



body size, proportion, coat characteristics and ear type [8]. Although these regions can also be associated with other traits - *IGF1*, for example, has also been implicated by GWAS as significantly affecting boldness, age of death and prevalence of several diseases [57, 58] - the most parsimonious explanation is that selection towards breed standards has led to stronger differentiation at loci affecting morphology than those affecting other traits. Perhaps this result is not surprising, as behavior can be difficult to quantify and disease prevalence and longevity, while highly breed-dependent, are not breed-defining characters undergoing direct diversifying selection.

Even for morphological variation, the influence of QTLs of major effect may be overstated. For example, although variation in *IGF1* explains 50% of the variation in size between breeds and nearly 50% of the variation within Portuguese water dogs where it segregates, it explains only 17% of the variation of body size within a population of village dogs where it segregates [8,60]. Portuguese water dogs were chosen to study body-size variation because they exhibit high intra-breed variation for size; other breeds also have intermediate *IGF1* frequencies and it is not clear if they exhibit more

intra-breed variation in body size than other breeds or if they possess loci that 'canalize' *IGF1* and other high-effect QTLs to reduce their effects.

In addition to the uncertainties regarding the simplified architecture underlying additive genetic variance in dogs, the degree to which non-additivity (dominance and recessiveness) and epistasis (interactions between loci) have an impact on complex traits in dogs is also uncertain. Although Lark [60] found a non-additive epistatic interaction between *IGF1* and a locus on the X chromosome controlling body size, most studies of complex traits in dogs have assumed additivity and ignored the X chromosome (Boyko *et al.* [8] did include the X chromosome, albeit only using an additive model, and found evidence for two body-size loci but not in the region reported in [60].) In addition, even the estimates for the proportion of variation explained by QTLs of large effect such as *IGF1* might be overstated. If diversifying selection is strong and several loci contribute to the trait but only the ones of largest effect are detected, then these large-effect loci will be associated with both the morphological effect they engender and also with the effects of the rare or small-effect loci that were also swept to some degree but not detected.

**Box 2. Some mechanisms to explain the simplified genetic architecture of morphological variation observed in purebred dogs.**

**Reduction of genetic diversity due to demographic forces.**

The founding and maintenance of dog breeds, and to a lesser extent bottlenecks associated with dog domestication, have reduced genetic diversity throughout the dog genome [73]. As a consequence, the genetic background on which a causal variant acts is less heterogeneous in dogs than it is in humans. Furthermore, allelic heterogeneity within a locus is significantly reduced, facilitating the discovery of candidate genes through GWAS. Whereas human populations may harbor several rare variants with varying (and sometimes opposite) effects at a locus, dogs have far fewer variants, often only one per locus. These single variants are more easily tagged in genotyping array studies, particularly since they are often relatively common in one or more breeds (and absent in others). In effect, breed structure significantly reduces the prevalence of rare variants that are believed to account for much of the missing heritability in human complex traits.

**Artificial selection favoring novelty.** Victorian dog fanciers actively selected for distinguishing characteristics in their animals. In fact, dog populations with distinguishing characteristics were probably more likely to gain official recognition as a sanctioned breed. Thus, among breeds, phenotypic variance in many traits has increased, favoring a simplified genetic architecture for the trait [74]. Furthermore, active selection for saltational mutations that would have been weeded out by natural selection further enhances the proportion of variance explained by large-effect QTLs.

**Short evolutionary history.** Most breeds were formed and adopted their breed standards little more than 100 generations ago. QTLs of minor effect necessarily have small fitness values and therefore have not had sufficient time to substantially differentiate in frequency in select breeds, a requirement if these alleles are to underlie a significant proportion of genetic variance. In contrast, QTLs of major effect can be very efficiently selected by breeders over short timescales. A corollary of this effect, however, is that most 'modifier loci', such as those that increase canalization of breed-defining characteristics or reduce recombination rates between epistatic loci, tend to be weakly selected, reducing the likelihood that such effects are a major part of the canine genetic architecture for complex traits.

**Relaxation of selective constraint.** Artificial selection by breeders dramatically reduces the efficacy of natural and sexual selection, allowing for genetic drift and phenotypic variation in traits that would otherwise be constrained by these forces. In general, selectively neutral traits exhibit simplified genetic architectures, as evidenced by the relatively large proportion of late- versus early-onset human diseases explained by just a few major QTLs (for example, Alzheimer's disease and macular degeneration [75]).

Nevertheless, dogs have clearly exhibited a shift in the genetic architecture of complex traits towards variants of large effect for several important phenotypes. This shift

facilitates complex-trait mapping, making the dog an extremely important model system. The extent to which this shift has affected non-morphological variation and whether this simplification extends to other aspects of genetic architecture is still unclear. Important insights into evolutionary biology, including the nature of standing genetic variation and the genomic consequences of adaptation to new environmental pressures, could be gained by determining the degree to which domestication, selective breeding and the genetic structure of the canine genome have each altered the genomic architecture of complex traits in this species.

**Concluding remarks and future perspectives**

Over the past decade, genetic analysis of dogs has not only enriched our understanding of their origins, but also led to the discovery of causal variants underlying myriad diverse phenotypes and diseases. As dogs became genome-enabled, candidate gene approaches gave way to GWAS, which have proved a particularly powerful method in this genetic system. Dogs have fewer allelic variants per locus and long tracts of linkage disequilibrium within breeds, meaning that far fewer markers are needed to find significant associations in dogs, perhaps an order of magnitude fewer than needed in human studies [61,62]. In addition to helping to identify regions associated with morphology and disease, this should make dogs particularly valuable in the near future when studies begin focusing on gene x gene interactions. In that case, the number of hypotheses to test scales by roughly the square of the number of markers; thus, depending on the breed, significant tracts of linkage disequilibrium could effectively reduce this number by a 100-fold or more.

Great progress is currently being made in mapping complex diseases in dogs, including cancer, diabetes, immune disorders, behavioral pathologies, osteoarthritis, and cardiac disease. Causal variants contributing to certain conditions in certain breeds will likely be identified in the near future, but it is less clear when these studies will begin to identify gene-gene and gene-environment interactions that could contribute even more to our understanding of the biological basis of the diseases. Reduced genetic (and perhaps environmental) heterogeneity compared with humans also makes expression QTL approaches to detecting complex traits promising in dogs. Previously, expression studies were limited by the power of microarrays and the reliability of gene annotation in the dog genome, but next-generation approaches such as direct RNA sequencing (RNA-seq) should greatly enhance functional genomics and expression profiling in dogs [63].

A significant impediment to fully realizing the potential of the dog as a model genomic system is the

lack of uniform data-release standards for published dog genomic studies. Making complete genetic and phenotypic data release *de rigueur* for canine genomics projects would provide the opportunity for meta-analyses utilizing many thousands of dogs that would be highly informative for many of these traits and interactions. Such meta-analyses are common in human genetic studies even though standards of subject confidentiality are much more rigorous, and have yielded valuable insights about the genetic basis of complex disease (see, for example [64]).

Future genomic studies may be able to unravel what it is that makes a dog a dog. For example, what are the genetic variants underlying the traits such as barking and neoteny (juvenilization) that became fixed very early in dog evolutionary history? How has the novel social, nutritional and disease environment of the dog affected its genome? The human genome is packed with strong signatures of selection at variants underlying phenotypic diversity of behavioral, metabolic and immune traits; is the dog genome littered with a parallel complement? The unique relationship between dogs and humans gives canine genomic studies the opportunity not only to flush out what it is that makes a dog a dog, but also to motivate comparative genomic approaches in their human companions.

## Additional files

**Additional file 1: Table S1.** Chronological listing of published known and putative genetic variants underlying Mendelian diseases in dogs. For canine disease with close human analogs or cases that could serve as a genetic model system for understanding a human disease, the human disease is also listed. Where more than five breeds are affected by the same variant, 'many' is listed. The dog chromosome that each variant is found on is listed according to the CanFam assembly (CFA) numbering.

## Competing interests

The author declares that he has no competing interests.

## Acknowledgements

I gratefully acknowledge the helpful comments of RH Boyko, JM Kidd and SM Myles on an earlier draft of this manuscript and insightful discussions with CD Bustamante. AR Boyko has been funded by NSF (DEB) 0948510 and (DBI) 0701382.

Published: 22 February 2011

## References

- Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ 3rd, Zody MC, Mauceli E, Xie X, Breen M, Wayne RK, Ostrander EA, Ponting CP, Galibert F, Smith DR, DeJong PJ, Kirkness E, Alvarez P, Biagi T, Brockman W, Butler J, Chin CW, Cook A, Cuff J, Daly MJ, DeCaprio D, Gnerre S, *et al*: **Genome sequence, comparative analysis and haplotype structure of the domestic dog.** *Nature* 2005, **438**:803-819.
- Guyon R, Lorentzen TD, Hitte C, Kim L, Cadieu E, Parker HG, Quignon P, Lowe JK, Renier C, Gelfenbeyn B, Vignaux F, DeFrance HB, Gloux S, Mahairas GG, André C, Galibert F, Ostrander EA: **A 1-Mb resolution radiation hybrid map of the canine genome.** *Proc Natl Acad Sci USA* 2003, **100**:5296-5301.
- Breen M, Hitte C, Lorentzen TD, Thomas R, Cadieu E, Sabacan L, Scott A, Evanno G, Parker HG, Kirkness EF, Hudson R, Guyon R, Mahairas GG, Gelfenbeyn B, Fraser CM, André C, Galibert F, Ostrander EA: **An integrated 4249 marker FISH/RH map of the canine genome.** *BMC Genomics* 2004, **5**:65.
- Kirkness EF, Bafna V, Halpern AL, Levy S, Remington K, Rusch DB, Delcher AL, Pop M, Wang W, Fraser CM, Venter JC: **The dog genome: survey sequencing and comparative analysis.** *Science* 2003, **301**:1898-1903.
- Parker HG, Kukekova AV, Akey DT, Goldstein O, Kirkness EF, Baysac KC, Mosher DS, Aguirre GD, Acland GM, Ostrander EA: **Breed relationships facilitate fine-mapping studies: A 7.8-kb deletion cosegregates with Collie eye anomaly across multiple dog breeds.** *Genome Res* 2007, **17**:1562-1571.
- Karlsson EK, Baranowska I, Wade CM, Salmon Hillbertz NH, Zody MC, Anderson N, Biagi TM, Patterson N, Pielberg GR, Kulbokas EJ 3rd, Comstock KE, Keller ET, Mesirov JP, von Euler H, Kämpe O, Hedhammar A, Lander ES, Andersson G, Andersson L, Lindblad-Toh K: **Efficient mapping of mendelian traits in dogs through genome-wide association.** *Nat Genet* 2007, **39**:1321-1328.
- Vonholdt BM, Pollinger JP, Lohmueller KE, Han E, Parker HG, Quignon P, Degenhardt JD, Boyko AR, Earl DA, Auton A, Reynolds A, Bryc K, Brisbin A, Knowles JC, Mosher DS, Spady TC, Elkahloun A, Geffen E, Pilot M, Jedrzejewski W, Greco C, Randi E, Bannasch D, Wilton A, Shearman J, Musiani M, Cargill M, Jones PG, Qian Z, Huang W, *et al*: **Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication.** *Nature* 2010, **464**:898-902.
- Boyko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, Lohmueller KE, Zhao K, Brisbin A, Parker HG, vonHoldt BM, Cargill M, Auton A, Reynolds A, Elkahloun AG, Castelhano M, Mosher DS, Sutter NB, Johnson GS, Novembre J, Hubisz MJ, Siepel A, Wayne RK, Bustamante CD, Ostrander EA: **A simple genetic architecture underlies morphological variation in dogs.** *PLoS Biol* 2010, **8**:e1000451.
- Morey DF: *Dogs: Domestication and the Development of a Social Bond.* New York, NY: Cambridge Univ Press; 2010.
- Townshend E: *Darwin's Dogs.* London, UK: Frances Lincoln Limited; 2009.
- Harting JE: **Dogs: ancient and modern.** *Zoologist* 1884, **8**:393-411.
- Wayne RK: **Molecular evolution of the dog family.** *Trends Genet* 1993, **9**:218-224.
- Vilà C, Savolainen P, Maldonado JE, Amorim IR, Rice JE, Honeycutt RL, Crandall KA, Lundeberg J, Wayne RK: **Multiple and ancient origins of the domestic dog.** *Science* 1997, **276**:1687-1689.
- Davis SJM, Valla FR: **Evidence for domestication of the dog 12,000 years ago in the Natufian of Israel.** *Nature* 1978, **276**:608-610.
- Tchernov E, Valla FF: **Two new dogs, and other Natufian dogs, from the southern Levant.** *J Hered* 1997, **24**:65-95.
- Nobis G: **Der älteste Haushunde lebte vor 14000 Jahren.** *Umschau* 1979, **79**:610.
- Germonpré M, Sablin MV, Stevens RE, Hedges REM, Hofreiter M, Stiller M, Després VR: **Fossil dogs and wolves from Palaeolithic sites in Belgium, the Ukraine and Russia: osteometry, ancient DNA and stable isotopes.** *J Archaeol Sci* 2009, **36**:473-490.
- Sablin MV, Khlopachev GA: **The earliest ice age dogs: evidence from Eliseevichi I.** *Curr Anthropol* 2002, **43**:795-799.
- Savolainen P, Zhang Y-p, Luo J, Lundeberg J, Leitner T: **Genetic evidence for an East Asian origin of domestic dogs.** *Science* 2002, **298**:1610-1613.
- Pang JF, Kluetsch C, Zou XJ, Zhang AB, Luo LY, Angleby H, Ardalan A, Ekström C, Sköllermo A, Lundeberg J, Matsumura S, Leitner T, Zhang YP, Savolainen P: **mtDNA data indicate a single origin for dogs south of the Yangtze River, less than 16,300 years ago, from numerous wolves.** *Mol Biol Evol* 2009, **26**:2849-2864.
- DeGiorgio M, Jakobsson M, Rosenberg NA: **Explaining worldwide patterns of human genetic variation using a coalascent-based serial founder model of migration outward from Africa.** *Proc Natl Acad Sci USA* 2009, **106**:16057-16062.
- Clutton-Brock J: **Origins of the dog: domestication and early history.** In *The Domestic Dog: Its Evolution, Behaviour, and Interactions with People.* Edited by Serpell J. Cambridge, UK: Cambridge Univ Press; 1995.
- Coppinger R, Coppinger L: *Dogs: A Startling New Understanding of Canine Origin, Behavior & Evolution.* New York, NY: Scribner; 2001.
- Leonard JA, Vilà C, Wayne RK: **From wild wolf to domestic dog.** In *The Dog and Its Genome.* Edited by Ostrander EA, Giger U, Lindblad-Toh K. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2005: 584.

25. Crockford SJ: *Rhythms of Life: Thyroid Hormone and the Origin of Species*. Victoria, BC: Trafford; 2006.
26. Parker HG, Kim LV, Sutter NB, Carlson S, Lorentzen TD, Malek TB, Johnson GS, DeFrance HB, Ostrander EA, Kruglyak L: **Genetic structure of the purebred domestic dog**. *Science* 2004, **304**:1160-1164.
27. Shearin AL, Ostrander EA: **Canine morphology: hunting for genes and tracking mutations**. *PLoS Biol* 2010, **8**:e1000310.
28. Björnerfeldt S, Webster MT, Vilà C: **Relaxation of selective constraint on dog mitochondrial DNA following domestication**. *Genome Res* 2006, **16**:990-994.
29. Cruz F, Vilà C, Webster MT: **The legacy of domestication: accumulation of deleterious mutations in the dog genome**. *Mol Biol Evol* 2008, **25**:2331-2336.
30. Boyko AR, Boyko RH, Boyko CM, Parker HG, Castelhanos M, Corey L, Degenhardt JD, Auton A, Hedimbi M, Kityo R, Ostrander EA, Schoenebeck J, Todhunter RJ, Jones P, Bustamante CD: **Complex population structure in African village dogs and its implications for inferring dog domestication history**. *Proc Natl Acad Sci USA* 2009, **106**:13903-13908.
31. Neff MW, Rine J: **A fetching model organism**. *Cell* 2006, **124**:229-231.
32. Laidlaw J, Gelfand Y, Ng K-W, Garner HR, Ranganathan R, Benson G, Fondon JW, III: **Elevated basal slippage mutation rates among the Canidae**. *J Hered* 2007, **98**:452-460.
33. Wang W, Kirkness EF: **Short interspersed elements (SINEs) are a major source of canine genomic diversity**. *Genome Res* 2005, **15**:1798-1808.
34. Parker HG, Shearin AL, Ostrander EA: **Man's best friend becomes biology's best in show: genome analyses in the domestic dog**. *Annu Rev Genet* 2010, **44**:309-336.
35. Hedrick PW, Andersson L: **Cause of extreme variation in dog morphology: mutation or selection? Are dogs genetically special?** *Heredity* 2010, In press; doi:10.1038/hdy.2010.98.
36. Bannasch D, Young A, Myers J, Truvé K, Dickinson P, Gregg J, Davis R, Bongcam-Rudloff E, Webster MT, Lindblad-Toh K, Pedersen N: **Localization of canine brachycephaly using an across breed mapping approach**. *PLoS One* 2010, **5**:e9632.
37. Parker HG, VonHoldt BM, Quignon P, Margulies EH, Shao S, Mosher DS, Spady TC, Elkahoulou A, Cargill M, Jones PG, Maslen CL, Acland GM, Sutter NB, Kuroki K, Bustamante CD, Wayne RK, Ostrander EA: **An expressed *fgf4* retrogene is associated with breed-defining chondrodysplasia in domestic dogs**. *Science* 2009, **325**:995-998.
38. Cadieu E, Neff MW, Quignon P, Walsh K, Chase K, Parker HG, Vonholdt BM, Rhue A, Boyko A, Byers A, Wong A, Mosher DS, Elkahoulou AG, Spady TC, André C, Lark KG, Cargill M, Bustamante CD, Wayne RK, Ostrander EA: **Coat variation in the domestic dog is governed by variants in three genes**. *Science* 2009, **326**:150-153.
39. Sutter NB, Bustamante CD, Chase K, Gray MM, Zhao K, Zhu L, Padhukasahasram B, Karlins E, Davis S, Jones PG, Quignon P, Johnson GS, Parker HG, Fretwell N, Mosher DS, Lawler DF, Satyaraj E, Nordborg M, Lark KG, Wayne RK, Ostrander EA: **A single IGF1 allele is a major determinant of small size in dogs**. *Science* 2007, **316**:112-115.
40. Drögemüller C, Karlsson EK, Hytönen MK, Perloski M, Dolf G, Sainio K, Lohi H, Lindblad-Toh K, Leeb T: **A mutation in hairless dogs implicates FOXI3 in ectodermal development**. *Science* 2008, **321**:1462.
41. Candille SI, Kaelin CB, Cattanach BM, Yu B, Thompson DA, Nix MA, Kerns JA, Schmutz SM, Millhauser GL, Barsh GS: **A  $\beta$ -defensin mutation causes black coat color in domestic dogs**. *Science* 2007, **318**:1418-1423.
42. Salmon Hillbertz NH, Isaksson M, Karlsson EK, Hellmén E, Pielberg GR, Savolainen P, Wade CM, von Euler H, Gustafson U, Hedhammar A, Nilsson M, Lindblad-Toh K, Andersson L, Andersson G: **Duplication of FGF3, FGF4, FGF19 and ORAOV1 causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs**. *Nat Genet* 2007, **39**:1318-1320.
43. Gray MM, Sutter NB, Ostrander EA, Wayne RK: **The IGF1 small dog haplotype is derived from Middle Eastern grey wolves**. *BMC Biol* 2010, **8**:16.
44. Akey JM, Ruhe AL, Akey DT, Wong AK, Connelly CF, Madeoy J, Nicholas TJ, Neff MW: **Tracking the footprints of artificial selection in the dog genome**. *Proc Natl Acad Sci USA* 2010, **107**:1160-1165.
45. Lin L, Faraco J, Li R, Kadotani H, Rogers W, Lin X, Qiu X, de-Jong PJ, Nishino S, Mignot E: **The sleep disorder canine narcolepsy is caused by a mutation in the Hypocretin (Orexin) Receptor 2 gene**. *Cell* 1999, **98**:365-376.
46. van de Sluis B, Rothuizen J, Pearson PL, van Oost BA, Wijmenga C: **Identification of a new copper metabolism gene by positional cloning in a purebred dog population**. *Hum Mol Genet* 2002, **11**:165-173.
47. Stuehler B, Reichert J, Stremmel W, Schaefer M: **Analysis of the human homologue of the canine copper toxicosis gene MURR1 in Wilson disease patients**. *J Mol Med* 2004, **82**:629-634.
48. Ostrander EA, Kruglyak L: **Unleashing the canine genome**. *Genome Res* 2000, **10**:1271-1274.
49. Nguyen DH, Hurtado-Ziola N, Gagneux P, Varki A: **Loss of Siglec expression on T lymphocytes during human evolution**. *Proc Natl Acad Sci USA* 2006, **103**:7765-7770.
50. Graham AL, Hayward AD, Watt KA, Pilkington JG, Pemberton JM, Nussey DH: **Fitness correlates of heritable variation in antibody responsiveness in a wild mammal**. *Science* 2010, **330**:662-665.
51. Ostrander EA, Giniger E: **Semper fidelis: what man's best friend can teach us about human biology and disease**. *Am J Hum Genet* 1997, **61**:475-480.
52. Vail DM, MacEwen EG: **Spontaneously occurring tumors of companion animals as models for human cancer**. *Cancer Invest* 2000, **18**:781-792.
53. Awano T, Johnson GS, Wade CM, Katz ML, Johnson GC, Taylor JF, Perloski M, Biagi T, Baranowska I, Long S, March PA, Olby NJ, Shelton GD, Khan S, O'Brien DP, Lindblad-Toh K, Coates JR: **Genome-wide association analysis reveals a *SOD1* mutation in canine degenerative myelopathy that resembles amyotrophic lateral sclerosis**. *Proc Natl Acad Sci USA* 2009, **106**:2794-2799.
54. Farias FH, Johnson GS, Taylor JF, Giuliano E, Katz ML, Sanders DN, Schnabel RD, McKay SD, Khan S, Gharahkhani P, O'Leary CA, Pettitt L, Forman OP, Boursnell M, McLaughlin B, Ahonen S, Lohi H, Hernandez-Merino E, Gould DJ, Sargan DR, Mellers C: **An ADAMTS17 splice donor site mutation in dogs with primary lens luxation**. *Invest Ophthalmol Vis Sci* 2010, **51**:4716-4721.
55. Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, Song X, Richmond TA, Middle CM, Rodesch MJ, Packard CA, Weinstock GM, Gibbs RA: **Direct selection of human genomic loci by microarray hybridization**. *Nat Methods* 2007, **4**:903-905.
56. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, Gabriel S, Jaffe DB, Lander ES, Nusbaum C: **Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing**. *Nat Biotechnol* 2009, **27**:182-189.
57. Jones P, Chase K, Martin A, Davern P, Ostrander EA, Lark KG: **Single-nucleotide-polymorphism-based association mapping of dog stereotypes**. *Genetics* 2008, **179**:1033-1044.
58. Chase K, Jones P, Martin A, Ostrander EA, Lark KG: **Genetic mapping of fixed phenotypes: disease frequency as a breed characteristic**. *J Hered* 2009, **100**:537-541.
59. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, Willer CJ, Jackson AU, Vedantam S, Raychaudhuri S, Ferreira T, Wood AR, Weyant RJ, Segre AV, Speliotes EK, Wheeler E, Soranzo N, Park JH, Yang J, Gudbjartsson D, Heard-Costa NL, Randall JC, Qi L, Vernon Smith A, Mägi R, Pastinen T, Liang L, Heid IM, Luan J, Thorleifsson G, et al.: **Hundreds of variants clustered in genomic loci and biological pathways affect human height**. *Nature* 2010, **467**:832-838.
60. Lark KG, Chase K, Sutter NB: **Genetic architecture of the dog: sexual size dimorphism and functional morphology**. *Trends Genet* 2006, **22**:537-544.
61. Sutter NB, Eberle MA, Parker HG, Pullar BJ, Kirkness EF, Kruglyak L, Ostrander EA: **Extensive and breed-specific linkage disequilibrium in *Canis familiaris***. *Genome Res* 2004, **14**:2388-2396.
62. Ostrander EA, Wayne RK: **The canine genome**. *Genome Res* 2005, **15**:1706-1716.
63. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L: **Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation**. *Nat Biotechnol* 2010, **28**:511-515.
64. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, de Bakker PI, Abecasis GR, Almgren P, Andersen G, Ardlie K, Boström KB, Bergman RN, Bonnycastle LL, Borch-Johnsen K, Burtt NP, Chen H, Chines PS, Daly MJ, Deodhar P, Ding CJ, Doney AS, Duren WL, Elliott KS, Erdos MR, Frayling TM, Freathy RM, Gianniny L, Grallert H, et al.: **Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes**. *Nat Genet* 2008, **40**:638-645.
65. Hare B, Brown M, Williamson C, Tomasello M: **The domestication of social cognition in dogs**. *Science* 2002, **298**:1634-1636.
66. Hare B, Tomasello M: **Human-like social skills in dogs?** *Trends Cogn Sci* 2005, **9**:439-444.
67. Dorey NR, Udell MA, Wynne CD: **Breed differences in dogs sensitivity to human points: A meta-analysis**. *Behav Processes* 2009, **81**:409-415.
68. Udell MAR, Dorey NR, Wynne CD: **Wolves outperform dogs in following human social cues**. *Anim Behav* 2008, **76**:1767-1773.
69. Udell MAR, Dorey NR, Wynne CD: **The performance of stray dogs (*Canis***

- familiaris*) living in a shelter on human-guided object-choice tasks. *Anim Behav* 2010, **79**:717-725.
70. Dorey NR, Udell MAR, Wynne CDL: **When do domestic dogs, *Canis familiaris*, start to understand human pointing? The role of ontogeny in the development of interspecies communication.** *Anim Behav* 2010, **79**:37-41.
71. Gácsi M, Györi B, Virányi Z, Kubinyi E, Range F, Belényi B, Miklósi Á: **Explaining dog wolf differences in utilizing human pointing gestures: selection for synergistic shifts in the development of some social skills.** *PLoS One* 2009, **4**:e6584.
72. Udell MAR, Wynne CDL: **Ontogeny and phylogeny: both are essential to human-sensitive behaviour in the genus *Canis*.** *Anim Behav* 2010, **79**:e9-e14.
73. Gray MM, Granka JM, Bustamante CD, Sutter NB, Boyko AR, Zhu L, Ostrander EA, Wayne RK: **Linkage disequilibrium and demographic history of wild and domestic canids.** *Genetics* 2009, **181**:1493-1505.
74. Koop M, Hermisson J: **The evolution of genetic architecture under frequency-dependent disruptive selection.** *Evolution* 2006, **60**:1537-1550.
75. Singleton AB, Hardy J, Traynor BJ, Houlden H: **Towards a complete resolution of the genetic architecture of disease.** *Trends Genet* 2010, **26**:438-442.
76. Hill WG, Robertson A: **Linkage disequilibrium in finite populations.** *Theor Appl Genet* 1968, **38**:226-231.

doi:10.1186/gb-2011-12-2-216

**Cite this article as:** Boyko AR: **The domestic dog: man's best friend in the genomic era.** *Genome Biology* 2011, **12**:216.