Canine Genomics: Identifying Genes Behind Diseases

G.R. Pielberg[*] and K. Lindblad-Toh[*,†]

Modern dog breeds have been bred for size, shape and behaviour in a unique way, suggesting that genes underlying these traits have been enriched in specific breeds. With these desirable traits, often comes the enrichment of specific diseases within certain breeds suggesting that enrichment of disease-causing mutations within breeds also has occurred. Historically, large canine families were collected and linkage mapping was used to identify single gene traits. With the dog genome sequence and associated tools we are now able to map both physiological and disease traits more easily using unrelated cases and controls.

Once disease genes have been identified, the ability to identify genetic susceptibilities by DNA tests and to use these in breeding strategies, for disease diagnosis as well as for guidance to treatment strategies is likely to become more and more important for veterinary medicine in the coming years. We will describe the principles of disease mapping and some success stories as well as the applications of gene identification in veterinary medicine.

Dogs (*Canis familiaris*) have a number of unique features that have the potential to make them a superior genetic system to study the molecular basis of disease. The population history of dog breeds results in a number of features that make them uniquely suited for the detection of susceptibility genes (Figure 1). These include: (i) the fact that the genetic complexity of inherited diseases is considerably reduced within the partially inbred dog breeds. This often manifests itself as an increased risk for a disease within a few specific breeds; for some complex traits a ten fold increased risk can be seen between breeds. (ii) the fact that – as a result of the recent breed creation, long regions in the genome that are inherited together (haplotypes) within breeds – the number of genetic markers (SNPs) needed to effectively tag a disease causing mutation by performing whole genome scans is reduced by an order of magnitude from hundreds to tens of thousands of markers.

We have proposed a two-stage approach, combining within-breed and subsequent between-breed analyses, for very precise localization of genetic risk factors (Figure 2) (Lindblad-Toh et al. (2005)). The canine genome sequencing project has produced a high-quality draft sequence of a Boxer covering ~99% of the genome and identified ~19,000 canine genes, virtually the same gene set as in humans and other mammals (Lindblad-Toh et al. (2005)). In addition, the sequencing project has identified ~2.5 million SNP markers common to many breeds. All these resources are publicly available. The SNP markers have been utilized to assemble highly informative genome-wide SNP arrays that can be used to genotype cost-effectively using high throughput platforms. For complex phenotypes, when mulitple genes influence a certain trait, simulations demonstrate that 100-300 cases + 100-300 controls provide adequate power to detect alleles conferring 2 to 5-fold multiplicative risk (Lindblad-Toh et al. (2005)).

[*] Uppsala University, Uppsala, Sweden
[†] Broad Institute of MIT and Harvard, Cambridge, MA, USA

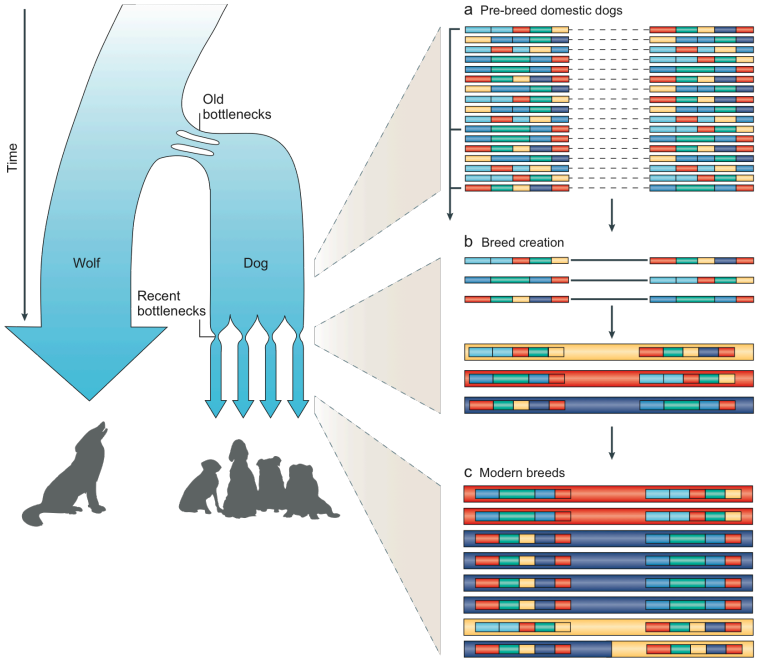

Figure 1: Two bottlenecks in the history of dog breeds shape the genome Two population bottlenecks in dog population history, one old and one recent, shaped haplotype structure (pieces of chromosome inherited together without recombination) in modern dog breeds. First, the domestic dog diverged from wolves ~15,000 years ago (Wayne and Ostrander (1999)). Within the past few hundred years, modern dog breeds were created. Both bottlenecks influenced the haplotype pattern of current breeds. (a) Before the creation of modern breeds, the dog population had the short-range haplotypes expected given its large size and long time since the domestication bottleneck. (b) In the creation of modern breeds, a small subset of chromosomes was selected from the pool of domestic dogs. The long-range patterns carried on these chromosomes became common within the breed. (c) In the short time since breed creation, these long-range patterns have not yet been substantially broken down by recombination. Long breed haplotypes, however, still retain the underlying short ancestral haplotype blocks from the domestic dog population, and these are revealed when one examines chromosomes across many breeds (Lindblad-Toh et al. (2005)).

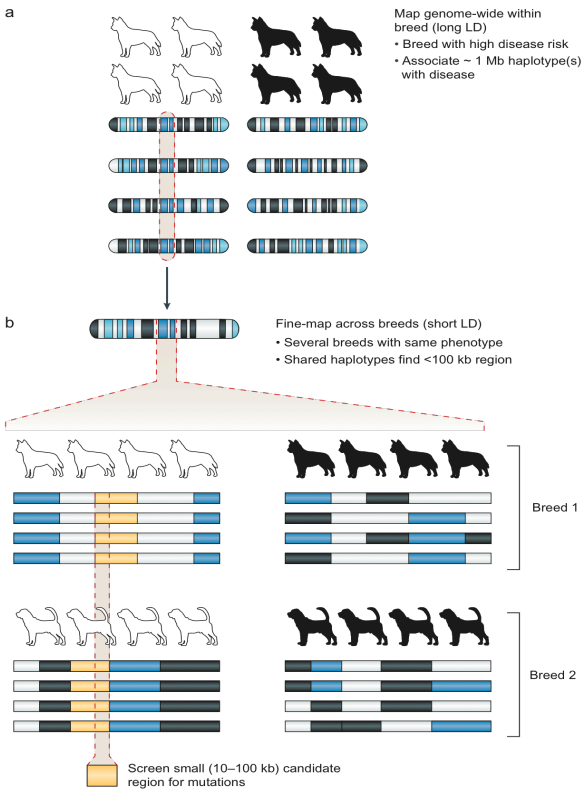

Figure 2: A two-stage approach for mapping susceptibility genes. A two-stage approach takes full advantage of the long LD within breeds and the short ancestral haplotypes shared across breeds, allowing traits to be mapped with relatively few samples (Lindblad-Toh et al. (2005)). In the first step whole-genome wide association mapping using >15,000 SNPs is employed in a breed with a high disease risk. Secondly, fine-mapping of the disease-associated region will be performed in the initial breed together with several related breeds. This permits rapid narrowing of the region to enable mutation screening of part of a gene or a few genes at most.

We originally identified four traits with a clear monogenic inheritance pattern to be used as proof-of-principle for the mapping two-stage strategy derived from the genome project. These projects include the white spotting pattern (Karlsson et al. (2007)), the hair-ridge (Salmon Hillbertz et al. (2007)), canine ectodermal dysplasia (Drögemüller et al. (2008)) and cone-rod dystrophy (Wiik et al. (2008)). These have now been published and demonstrate the efficacy of our gene mapping approach. We have also identified a major gene for degenerative myelopathy (Awano et al. (2009)). This gene, *SOD1*, is frequently mutated in human Amyotrophic Lateral Sclerosis (ALS). The canine mutation is very common in multiple breeds including in boxers and Pembroke Welsh corgis, while only some dogs in these breeds get the disease. We are therefore in the process of performing mapping in early onset cases and old healthy dogs with the *SOD1* mutation to identify modifier loci. Such loci might be more useful for limiting the disease prevalence by breeding against the trait.

Several complex traits are also now being mapped in the dog using only small sample numbers. Many of these are being mapped as part of the LUPA project, a large European collaboration (http://www.eurolupa.org/). For most of these we expect to find multiple loci with non-coding mutations that primarily changes the amount of protein produced rather than change/destroy the actual proteins. Here I will give examples of recently mapped dog diseases, finding multiple highly associated genes affecting specific pathways.

While some monogenic traits, where the mutant alleles are rare within a breed, will likely be depleted from the breed by careful breeding, some monogenic traits will be frequent within specific breeds and it will be harder to breed away from them. In addition, complex traits will be dependent on many genes and it will be much more difficult to breed away from them effectively. Given these complexities of interpreting the tests, we expect that in the future veterinarians will be asked to advice on how the genetic tests are used for breeding.

However, we envision that the most positive effects from gene mapping of disease genes will come from three major applications. 1) By identifying disease genes and pathways we will gain a better understanding of the disease mechanisms and therefore be in a better position to develop novel therapeutics. 2) In the future genetic tests will likely also aid in the diagnosis of disease, as well as sub-typing. Such sub-typing could likely be used for selection of treatment strategies when different outcomes have been connected to different genetic risk factors. 3) Due to the similarities between dogs and humans, any findings in dogs will likely also inform human medicine.

References

Awano, T., Johnson, G.S., Wade C.M. et al. (2009). *Proc. Natl. Acad. Sci. U S A*, 106:2794–2799.

Drögemüller, C., Karlsson, E.K., Hytönen, M.K. et al. (2008). *Science*, 321:1462.

Karlsson, E.K., Baranowska, I., Wade, C.M. et al. (2007*). Nat. Genet.*, 39:1321–1328.

Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S. et al. (2005). *Nature,* 438:803–819.

Salmon Hillbertz, N.H., Isaksson, M., Karlsson, E.K. et al. (2007). *Nat. Genet.*, 39:1318–1320.

Wiik, A.C., Wade, C., Biagi, T. et al. (2008). *Genome Res.*, 18:1415–1421.

Wayne, R.K. and Ostrander, E.A. (1999*). Bioessays*, 21:247–257.