Dog genetics and genomics

C. Drögemüller[*]

Introduction

The dog, as a favoured companion of humans, is unique among domestic animal species in providing new insights into mammalian genetics and genomics. The major advantages dogs offer for genetic studies are the structure of dog populations consisting of >300 partially inbred genetic isolates with genetic traits or diseases predominantly or exclusively in one or a few breeds, the high degree of medical surveillance, and the excellent arsenal of dog genome resources (Ostrander *et al.* (2000)). Each of these breeds is defined by specific phenotypes that have been driven to exceptionally high frequency by population bottlenecks and strong human artificial selection and partial inbreeding (Karlsson and Lindblad-Toh (2008)).

Just after the establishment of the first set of linked canine microsatellite markers (Lingaas *et al.* (1997)) the first meiotic linkage map of the whole dog genome was published (Mellersh *et al.* (1997)). Considerable effort in the canine mapping was made by the development of an integrated high quality radiation hybrid (RH) map (Guyon *et al.* (2003)) and a high density resolution dog-human chromosomal comparative map (Breen *et al.* (2004)). Finally, after the availability of a 1.5x Poodle sequence (Kirkness *et al.* (2003)), the first high-quality draft (7.5x) sequence of the Boxer dog was made publicly available in July of 2004 (Lindblad-Toh *et al.* (2005)).

Family based linkage studies using micorsatellites were primarily conducted to map monogenic diseases leading to the development of gene tests. In some cases linkage based positional cloning also assigned previously unknown gene functions, e.g. the unexpected finding that *CBD103*, a gene encoding a beta-defensin and previously studied for its role in immunity, influence black coat color in domestic dogs (Candille *et al.* (2007)). Compared to livestock species, the QTL mapping especially for complex traits in dogs is limited due to the inability to perform larger crosses of divergent dog breeds. However, one QTL that was linked to size was mapped to the *IGF1* gene, at which multiple small breeds show evidence of a selective sweep of a single shared allele (Sutter *et al.* (2007)). Within a dog breed, linkage disequilibrium is extensive, enabling genome-wide association (GWA) with only around 15,000 single nucleotide polymorphisms (SNP) and fewer individuals than in human studies (Lindblad-Toh *et al.* (2005)). Proof-of-principle studies of GWA in dogs have identified the genes underlying two monogenic traits, white spotting (Karlsson *et al.* (2007)) and the hair ridge in ridgeback dogs (Salmon Hillbertz *et al.* (2007)). In both cases the initial GWA mapping was achieved with only 10 cases and 10 controls which in most cases are sufficient to give genome-wide significance for a simple Mendelian recessive trait

[*] University of Berne, Institute of Genetics, 3001 Berne, Switzerland

(Andersson (2009)). Monogenic traits and disorders in dogs are mostly due to a single causative allele present in all cases giving a strong signal in genome-wide association analysis. Selection for a favoured allele or genetic drift in populations with a limited effective population size are the explanations for this genetic homogeneity. Related examples using microarray-based SNP genotyping and GWA detected the responsible genes for three breed-specific coat color and type pattern (Cadieu *et al.* (2009)), and breed-defining chondrodysplasia in short-leg dog breeds (Parker *et al.* (2009)).

Incidences of specific diseases are elevated in different breeds, indicating that a few genetic risk factors might have accumulated through drift or selective breeding. Recent results indicate that the homogeneity of strong genetic risk factors within dog breeds allows polygenic disorders to be mapped with fewer than 100 cases and 100 controls, making dogs an excellent model in which to identify pathways involved in human complex diseases (Wibe *et al.* (2010)). The recent development of a canine high-density SNP genotyping array with approximately 170,000 SNPs will probably improve the mapping resolution significantly. Currently, the international collaborative research project "Dog genetics to unravel human diseases – LUPA" funded by the European Commission focus on the molecular analysis of different canine genetic diseases as model for human medicine (www.eurolupa.org).

Taken together, the outstanding dog genome resources, together with the specific haplotype structure due to the remarkable intrabreed homogeneity, coupled with striking interbreed heterogeneity, currently facilitate analyses of well-characterized canine inherited diseases as well as the various breed specific morphological and behavioral traits. Recent examples of successful positional cloning studies in dogs like the identification of a *FOXI3* mutation in hairless dogs and a *SERPINH1* mutation in osteogenesis imperfecta affected Dachshunds reflect the enormous potential of dogs to gain further insights into mammalian biology and human genetics, respectively.

Example 1

FOXI3 mutation in hairless dogs.

Chinese Crested, Mexican and Peruvian hairless dogs are characterized by a very sparse hair coat and deficient or abnormally shaped teeth, a special phenotype summarized under the classification canine ectodermal dysplasia (CED; O'Brien *et al*. (2005)). CED is inherited as a monogenic autosomal semi-dominant trait. Heterozygous dogs exhibit the CED phenotype and homozygous mutant animals die during embryonic development.

Genotyping of 49,663 SNPs and GWA mapping using 20 hairless and 19 coated Chinese Crested dogs identified a single region of strong association on chromosome 17 containing an eight SNP haplotype with complete phenotypic concordance. Subsequent linkage disequilibrium fine-mapping across three breeds narrowed the critical interval to 102 kb. This interval contains only one gene, a previously uncharacterized member of the forkhead box transcription factor family (*FOXI3*). Whole mount *in situ* hybridization of mouse embryos demonstrated a specific expression of *Foxi3* in hair and whisker placodes and in

developing teeth. Mutation analysis revealed a 7 bp duplication leading to a frameshift within the *FOXI3* coding sequence that co-segregates with the CED phenotype in all three investigated dog breeds (Drögemüller *et al.* (2008)).
In summary, this two-stage genome-wide mapping strategy identified *FOXI3* as the causative gene for CED in dogs, and thereby discovered the crucial role this novel gene plays during ectodermal development. This gene may act as a downstream target of the ectodysplasin signalling pathway, which seems plausible given the more restricted hair and tooth phenotype in CED compared to the broader hair-tooth-gland phenotype seen in the anhidrotic ectodermal dysplasia mutants of the ectodysplasin pathway (Drögemüller *et al.* (2001)). On the other hand, *FOXI3* is obviously also expressed in very specific cell populations during early development, and a function upstream of the ectodysplasin signalling cascade or even in additional developmental processes outside the ectoderm cannot be ruled out. This study reflects the enormous potential of dogs to gain further insights into mammalian biology.

Example 2

***SERPINH1* mutation in Dachshunds with osteogenesis imperfecta.**

Osteogenesis imperfecta (OI) is a hereditary disease occurring in humans and dogs. It is characterized by extremely fragile bones and teeth. In humans the large majority of OI cases is caused by defects in one of two collagen genes (*COL1A1* and *COL1A2*). Recently, it was discovered that mutations in two other genes (*CRTAP* and *LEPRE1*) related to collagen maturation also lead to OI in some human patients.

A pedigree of Dachshunds with an autosomal recessive form of OI (Seeliger *et al.* (2003)) were initially screened for the two known collagen genes which did not revealed a mutation in OI affected dogs. A subsequent genome-wide search for shared identical-by-descent segments across the entire genome in five affected Dachshunds localized the causative mutation to a 5.82 Mb interval on chromosome 21 by homozygosity mapping. The *SERPINH1* gene which encodes a serine protease inhibitor, also called heat shock protein 47 (HSP47) or collagen binding protein 1, known to be involved in collagen maturation was identified as positional candidate. Sequencing of the *SERPINH1* gene in healthy and affected Dachshunds revealed a single mutation exclusively shared by all affected dogs but not by healthy controls (Drögemüller *et al.* (2009)). The missense mutation is located in an evolutionary conserved domain and was perfectly associated with the OI phenotype.

The knowledge of this mutation enables genetic testing and will allow breeders to eradicate the deleterious allele from the Dachshund breeding population. The study also provides a defined animal model and a novel genetic mechanism for a lethal or severely debilitating human hereditary disease, as *SERPINH1* mutations might also be responsible for some human OI forms, where the causative mutation has not yet been identified. Finally, this study has identified a candidate causative mutation for OI in Dachshunds and *SERPINH1* as a fifth OI gene.

Conclusion

Application of high-throughput SNP genotyping and GWA analysis facilitates forward genetics in dogs. The dog genome sequence represents the key element for successful positional cloning and the identification of causative mutations. The variety of canine diseases and desired attributes, like morphological traits, offers unique opportunities to understand the genetic basis of natural variation in mammals.

References

Andersson, L. (2009) *Genetica* 136:341-349.

Breen, M., *et al*. (2004) *BMC Genomics* 5:65.

Cadieu, E., *et al*. (2009) *Science* 326:150-153.

Candille, S.I., *et al*. (2007) *Science* 318:1418-1423.

Drögemüller, C., *et al*. (2001) *Genome Res.* 11:1699-1705.

Drögemüller, C., *et al*. (2008) *Science* 321:1462.

Drögemüller, C., *et al*. (2009) *PLoS Genetics* 5: e1000579.

Guyon, R., *et al*. (2003) *Proc. Nat. Acad. Sci. USA* 100:5296-5301.

Karlsson, E.K., *et al*. (2007) *Nat. Genet.* 39:1321-1328.

Karlsson, E.K., and Lindblad-Toh, K. (2008) *Nat. Genet. Rev.* 9:713-725.

Kirkness, E.W., *et al*. (2003) *Science* 301:1898-1903.

Lindblad-Toh, K., *et al.* (2005) *Nature* 438:803-819.

Lingaas, F., *et al*.(1997*) Mamm. Genome* 8:218-221.

Mellersh, C.S., *et al*. (1997) *Genomics* 46:326-336.

O'Brien, D.P. *et al*. (2005) *J. Hered.* 96:727-734.

Ostrander, E.A. (2000) *Trends Genet*., 16:117-124.

Parker, H.G., *et al*. (2009) *Science* 325:995-998.

Salmon Hillbertz, N.H.C., *et al*. (2007) *Nat. Genet*. 389:1318-1320.

Seeliger, F., *et al*. (2003) *Vet Pathol* 40:530-539.

Sutter, N.B., *et al*. (2007) *Science* 316:112-115.

Wibe, M., *et al*. (2010) *Nat. Genet*. (Epub ahead of print).