

Unleashing the Canine Genome

Elaine A. Ostrander^{1,2} and Leonid Kruglyak^{1,3}

¹Division of Human Biology, ²Division of Clinical Research, and ³Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109–1024, USA

In humans, common diseases show complex modes of inheritance and, as a result, have been largely refractory to genetic analysis. Rodent systems are more tractable genetically, but the mutations typically represent induced rather than naturally arising alleles, and results are often of limited direct relevance to human disease because of profound differences in physiology. By comparison, the physiology, disease presentation, and clinical response of dogs often mimic human diseases closely. In addition, the modern dog offers key advantages over other animal systems for mapping genes relevant to human disease. In the following discussion, we highlight some of these advantages and provide specific examples where canine genetics is best suited to solve difficult problems in human genetics. In particular, we focus on the strong promise of linkage-disequilibrium (LD) mapping in dogs.

LD Mapping

LD mapping relies on population-level associations among alleles at disease loci and those at nearby markers. Such associations typically arise when all or most of the disease alleles in a population share a common ancestral origin. LD mapping has been applied with success in isolated human populations. For instance, LD methods have been used to fine-map a number of disease genes in Finnish populations, for which LD frequently extends over distances of several cM around the disease mutation (de la Chapelle and Wright 1998). Examples

of *de novo* mapping of human disease genes by the LD approach include the mapping of Hirschsprung's disease in a large, inbred Mennonite kindred (Puffenberger et al. 1994). The gene was mapped to chromosome 13q following analysis of a microsatellite screen of 10-cM density in only six affected individuals. Importantly, all six were related through multiple lines of descent to a single ancestral couple eight to 12 generations earlier. In addition, a gene for benign recurrent intrahepatic cholestasis (BRIC) was mapped to chromosome 18 by searching for chromosome segments shared by only three distantly related patients (Houwen et al. 1994). These individuals all came from an isolated Dutch community and shared common ancestry six to 12 generations back (Houwen et al. 1994). In a demonstration of the feasibility of LD mapping in domestic animals, Farnir and colleagues have shown that LD extends over tens of centimorgans in two different breeds of cattle (Farnir et al. 2000) and have used LD to successfully fine-map quantitative trait loci (Riquet et al. 1999).

The Unique History of Dogs

The domestic dog has a population history that makes it a particularly attractive model for LD mapping. The purebred dog population consists of >300 partially inbred genetic isolates called breeds. Over one million purebred dogs are newly registered in the United States each year. Gene flow between breeds is restricted by the pedigree barrier—registering a dog as a member of a particular breed requires that both of the dog's parents be registered members of that same breed. Most modern dog breeds are relatively young, with the ma-

ajority having been developed within the last 300 yr (Wilcox and Walkowicz 1995; Wayne and Ostrander 1999). Many of these were derived from a small number of founders—as few as six in the case of the modern Irish Wolfhound—that best represented the physical or behavioral traits breeders wished to feature in a given breed.

The natural history of some breeds has further restricted their genetic diversity over what is expected from breeding strategies alone. Catastrophic events in the last 100 yr, such as the two world wars and the American depression, have produced severe bottlenecks in many breeds, at times reducing the effective breeding stock to only a few dogs. At the end of World War I, for instance, only five dogs of the Leonburger breed remained alive in Europe (Wilcox and Walkowicz 1995). All Leonburgers alive today are believed to be descendents of those five. Diversity in some breeds is further reduced by the presence of popular sires. These dogs have physical features that make them particularly successful in the show ring and hunting or performance events, and as a result, they may produce >100 litters in their lifetime. For many breeds, therefore, the purebred dogs of today represent a limited genetic pool, with disease predispositions that derive from one or a small number of recent genetic founders (see below). Thus, the structure of dog breeds has the potential to dramatically reduce the problems associated with heterogeneity and genetic complexity of common disease inheritance and makes dog disease genes good targets for LD mapping. Dog breeds offer all the advantages of geographically isolated human populations but with a higher degree of isolation, narrower bottlenecks, and much better genealogical records.

²Corresponding author.

E-MAIL eostrand@fred.fhcrc.org; **FAX** (206) 667-6396.

Article and publication are at www.genome.org/cgi/doi/10.1101/gr.155900.

Canine Inherited Diseases

Over 360 genetic disorders in dogs have been described to date (Patterson 1980, 2000). This constitutes the largest set of naturally occurring genetic disorders in any nonhuman species (Patterson 2000). At least half of all described canine diseases resemble specific human disorders. Among those for which the mode of inheritance is known, over 70% are inherited as autosomal recessive, X-linked, or genetically complex traits. The high susceptibility to specific diseases in particular breeds, together with its much lower incidence or complete absence in other breeds, strongly suggests that some breeds are enriched for the presence of risk alleles. Such enrichment, caused by small founding populations, bottlenecks, and popular-sire effects, will only occur when the number of risk alleles is small, and they are relatively rare in the overall dog population (although possibly common within particular breeds). If many common alleles were involved in susceptibility to any disease, we would expect all breeds to show similar incidence patterns. They clearly do not. Indeed, 46% of genetic diseases reported in dogs are believed to occur predominantly or exclusively in one or a few breeds (Patterson 2000). Therefore, in any given breed we are likely to be dealing with a small number of (or even one) disease alleles of strong effect. Such alleles can be readily identified in samples of modest size.

The Canine Genome

Recently developed genomic resources make LD studies in dogs practical. The current version of the meiotic linkage map consists of 341 microsatellite markers with an average interval of 8.86 cM (Werner et al. 1999). In parallel with construction of a canine linkage map, a whole-genome radiation hybrid (RH) map of the dog is also being developed (Priat et al. 1998; Mellersh et al. 2000). Integration of the existing linkage and RH maps has recently produced a map of 724 markers (235 genes, 489 microsatellites), which has been completely aligned with the human genome map (Breen et al. 1999; Yang et al. 1999;

Mellersh et al. 2000). Thus, the more densely mapped human and mouse genomes can guide the selection of candidate gene analysis following initial findings of linkage.

Model Diseases: Cancer in Dogs

Cancer, deafness, heart disease, blindness, and epilepsy are among the many diseases well suited for LD mapping in dogs. Cancer appears particularly tractable. Malignancies are twice as common in dogs as in humans. At least a dozen distinct canine cancers are hypothesized to be appropriate models for human malignancies (Patterson et al. 1982; MacEwen 1989; Knapp and Waters 1997), with most types being more similar to human tumors in histopathological appearance, biological behavior, and response to therapy than the corresponding rodent tumors (MacEwen 1989; Knapp and Waters 1997). Unlike most rodent models, which often require specific carcinogenic exposures, cancer in dogs is a naturally occurring disease, with both inherited and sporadic forms noted for all common types of cancer.

Currently, lymphoma is the most commonly reported hematopoietic tumor in the dog (Dorn et al. 1968a, 1968b; Dorn and Priester 1987). A high breed-specific incidence is reported for Boxers and Pointers (Dorn and Priester 1987) and, more recently, for Golden Retrievers and Rottweilers. A role for viral factors in the etiology of the disease has been sought but not found. Osteosarcoma is another common cancer in dogs, with an estimated incidence rate 40 times that observed in humans (7.9/100,000 in dogs vs. 0.2/100,000 in humans; Dorn et al. 1968; Withrow et al. 1991). The clinical presentation, histopathology, and mode of treatments are virtually identical for humans and dogs (Hewitt 1978; Misdorp 1980; Withrow et al. 1991). Large and giant breeds such as Great Danes, St. Bernards, Irish Setters, Golden Retrievers, Rottweilers, and Doberman Pinschers, develop the disease frequently (Tjalma 1996). A genetic component to canine osteosarcoma has previously been described in St. Ber-

nards, but no gene has been cloned or even mapped (Bech-Nielsen et al. 1978).

Breed-Specific Strategies for Mapping Cancer Genes

Considerations of breed histories suggest that LD is likely to be present in Rottweilers and Golden Retrievers over distances of several cM around disease loci. In Rottweilers, this is due to a recent and very narrow population bottleneck. The Rottweiler was used as a popular drover, draft, and guard dog during the Middle Ages but nearly became extinct in the 1800s and early 1900s (Wilcox and Walkowicz 1995). A large part of the current U.S. population is drawn from a very small number of dogs that were brought to the United States from Germany in 1928 (Klem and Rademacher 1996). The first progeny from those dogs were registered in 1931 (Klem and Rademacher 1996; American Kennel Club 1999). Between 1931 and 1960 <100 new dogs were registered per year, with numbers in the single digits from 1931 to 1944 (American Kennel Club 1999). From 1960 to the mid-1970s the breed underwent only slight expansion, with no more than 200–500 new dogs registered per year (Klem and Rademacher 1996; American Kennel Club 1999). Beginning in 1982, however, the breed exploded in popularity; currently, some 55,000 new Rottweilers are registered each year (American Kennel Club 1999). Thus, the group of registered dogs alive today, estimated to be about 750,000, is derived almost exclusively from a few founders and popular sires and experienced a very long bottleneck of some 30 yr before the recent population explosion.

Using data provided by the American Kennel Club (AKC) Purebred Dog Registry Database, we have conducted simulation studies to determine the extent over which LD can be expected. We took into account the breed history and utilized yearly registration numbers from 1931 through 1998 (American Kennel Club 1999). In addition, we assumed that, on average, 10% of registered dogs will produce registered litters and that the average dog produces 1.5

litters. The mean litter size for Rottweilers is seven puppies (Klem and Rademacher 1996). From these numbers, we conservatively estimate that the effective population size of the breed was five dogs between 1931 and 1945, followed by expansion at a rate of 10% per year until the recent explosion in population size. Using these estimates, our simulations show that the breed history will lead to high levels of LD extending over 5–10 cM around a disease mutation. The variability derives, in part, from the degree to which we can collect distantly related dogs descended from popular sires. Since the AKC has tracked Rottweiler lineages from 1931, an opportunity exists to optimize sample collection accordingly. We estimate that screening a sample of 40 well-chosen affected dogs for regions of identity by descent will be sufficient for initial gene localization.

For Golden Retrievers, the considerations are somewhat different. This extremely popular breed was developed in 1868 by Sir Dudley Majoribanks of Guisachan, Scotland, who is reported to have initiated the breed by crossing a Tweed Water Spaniel (a now extinct breed) to a yellow retriever that likely carried the gene pool of several distinct types of retrievers from Labrador and Newfoundland (Rogers Clark and Brace 1995; Wilcox and Walkowicz 1995). The resulting progeny were bred for performance, and outcrosses were done with Labradors, Red Setters, Bloodhounds, and probably others, to allow for opportunities for selected variation in coat consistency (flat, wavy, and curly) and color (cream, golden, yellow to reddish gold). The breed was imported to the United States in the early 1900s and was officially recognized by the American Kennel Club in 1932. Because of its striking appearance, friendly temperament, and intense desire to please, the breed is consistently ranked as one of the most popular.

Unlike the Rottweiler, Golden Retrievers have not experienced any recent major bottlenecks. However, as this is a dog that tends to do well in dog shows, it exhibits several examples of strong popular-sire effects. The estimated relative risk (2.1) of all malignancies for

Golden Retrievers was significantly greater than the comparable rate for all other dogs combined (Priester and McKay 1980), and lines of Golden Retrievers with an unusually high incidence of cancer have been reported anecdotally. Because of the popular sire effect, LD is expected to be present over large distances within groups of Golden Retrievers related through a single sire. As a result, we predict that examination of haplotype sharing among affected dogs from a single line will easily identify the region containing the disease mutation. Screening of as few as 10 affected individuals has excellent power to detect regions of identity by descent under conditions of genetic homogeneity (which is expected within a single line), while screening 40 affecteds has high power to detect such sharing even under conditions of moderate heterogeneity (Houwen et al. 1994; Puffenberger et al. 1994). We hypothesize that different lines of dogs will share the same mutation because of an even older shared ancestor. In this case, the overlap of regions of haplotype sharing observed among the different lines of dogs would allow us to dramatically narrow the location of the disease gene.

False Positives and Breed-Specific Traits

LD mapping in highly inbred populations requires great care because chance occurrence of LD is expected to be commonplace. For instance, Farnir et al. (2000) observed LD between loci located on different chromosomes in cattle. This observation raised concerns about how frequent false-positives are, and prompted discussion regarding methods to identify and discard regions of genetic homogeneity that are unrelated to the trait under investigation (Farnir et al. 2000). Farnir et al. (2000) suggested using linkage information to distinguish real signals from false ones. In canines, a breed-specific assessment of genome-wide levels of background LD, as well as comparisons of LD levels between affected dogs and a control group with a similar degree of relatedness, may provide other means for controlling the false-positive rate.

A more subtle issue in canines is that of breed-specific traits. Breed-defining characteristics have been strongly selected for, and as a result, one expects LD in the regions of the genome that contain genes underlying these traits (e.g. Wang et al. 1999). Such regions must be identified in control studies to distinguish them from loci involved in disease. Although such screening will usually control for loci responsible for breed-specific traits, it may “mask” a disease gene linked to one of these loci.

The diseases for which it will be the most difficult to sort out the underlying genetics are those in which the disease is caused by peculiarities in the breed standard. Consider the example of the Shar-Pei, an ancient breed that has existed in China as a hunter, herder, and protector for centuries. The dog frequently provided entertainment to owners as a participant in dogfights. The breed standard for the Shar-Pei dictates that the dogs have extremely loose skin, giving it a wrinkled appearance. This feature enabled a dog to turn on an opponent even when firmly grasped between the opponent's teeth (Wilcox and Walkowicz 1995). As a result of the wrinkles, the skin is not well cleansed naturally, and comparatively higher levels of infectious disease in the skin and eyes can be observed in this breed. Thus, the underlying genetics responsible for the high incidence of disease in this breed is not a mutation in a disease gene or the presence of a particular MHC allele; rather, it is the result of physical characteristics associated with the breed standard.

Although the details are easy to sort out in the case of the Shar-Pei, one imagines they could be considerably harder in the case of other more common and complex diseases. For instance, hip dysplasia is frequently observed in many breeds. In a breed like the German Shepherd, where it is very common, we might theorize that it is either caused by germline mutations carried only in affected dogs or, alternatively, that poor hip development is a consequence of the breed standard, which dictates that the hindquarters for German Shepherds

be low and sloping. The high level of osteosarcoma in long-limbed breeds such as Scottish Deerhounds provides another example. It is not known whether there is a cancer susceptibility gene segregating in this breed or if the high frequency of osteosarcoma is simply the result of chance errors in mitosis that occur during the enormous number of cell divisions required for the dog to reach its full adult height of about 32 in.

Getting from Localizations to Genes

One drawback of genetic studies in populations with extensive LD is that it is difficult to localize a gene to a region smaller than a few centimorgans. Even with today's genomic technologies, sifting a region of this size for genes and mutations is a daunting task. Fortunately, the unique population structure of dogs may offer a solution. We propose that the unique histories and narrow population bottlenecks associated with many breeds will allow us to develop hierarchies of breeds such that younger breeds with extensive LD can be used for initial gene localization with a relatively sparse marker map, while older (related) breeds with the same mutations and more narrowly circumscribed regions of LD can be used for fine mapping and cloning.

Returning to the examples considered above, the analysis for Golden Retrievers, could include Flat-Coated Retrievers (which also have excess lymphoma), Irish Setters, and yellow retrievers because of their historical relationship to Golden. In the case of Rottweilers, Mastiffs, Bull Mastiffs, Great Danes, and Boxers, all of whom share an ancestral relationship with Rottweilers, could be analyzed (Pugnetti 1980; Wilcox and Walkowicz 1995). For other breeds, relationships may need to be reconstructed based on molecular markers, as anecdotal information and phenotypic similarity may prove unreliable. In some instances the mutations will be unique to a breed, having arisen subsequent to the breed's derivation. These caveats aside, interbreed comparative mapping provides a useful complement to the intrabreed mapping described above for refining localizations.

Coda

The difficulties encountered in mapping genes for complex traits in humans argue strongly for the development of new genetic systems. Once a system for using LD to map genes in dogs is established, we envision that it will have broad applications. The well-recorded manipulation of canine genetics by centuries of human intervention provides modern geneticists with unique opportunities to understand the genetics of disease susceptibility, morphology, and behavior.

ACKNOWLEDGMENTS

We thank David Waters, Deborah Lynch, Richard Dorn, Jenni Lowe, Cathryn Mellersh, and Ed Giniger for careful reading of this manuscript and their very helpful comments.

Supported by the American Kennel Club Canine Health Foundation (E.A.O.), the Burroughs Wellcome Fund (E.A.O. and L.K.), the Howard Hughes Medical Institute (L.K.), and the James S. McDonnell Foundation (L.K.).

REFERENCES

- American Kennel Club. 1999. *Breed registry data base*. American Kennel Club, New York.
- Bech-Nielsen, S., Haskins, M.E., Reif, J.S., Brodey, R.S., Patterson, D.F., and Spielman, R. 1978. *J. Natl. Cancer Inst.* **60**: 349–353.
- Breen, M., Thomas, R., Binns, M.M., Carter, N.P., and Langford, C.F. 1999. *Genomics* **61**: 145–155.
- de la Chapelle, A. and Wright, F.A. 1998. *Proc. Natl. Acad. Sci.* **95**: 12416–12423.
- Dorn, C.R., Taylor, D.O., Frye, F.L., and Hibbard, H.H. 1968a. *J. Natl. Cancer Inst.* **40**: 295–305.
- Dorn, C.R., Taylor, D.O., Schneider, R., Hibbard, H.H., and Klauber, M.R. 1968b. *J. Natl. Cancer Inst.* **40**: 307–318.
- Dorn, R.C. and Priester, W.A. 1987. *Veterinary cancer medicine* (ed. G.H. Theilen and B.R. Madewell), pp. 27–52. Lea and Febiger, Philadelphia.
- Farnir, F., Coppieters, W., Arranz, J.J., Berzi, P., Cambisano, N., Grisart, B., Karim, L., Marçq, F., Moreau, L., Mni, M., et al. 2000. *Genome Res.* **10**: 220–227.
- Hewitt, H.B. 1978. *Adv. Cancer Res.* **27**: 149–200.
- Houwen, R.H., Baharloo, S., Blankenship, K., Raeymaekers, P., Juyn, J., Sandkuijl, L.A., and Freimer, N.B. 1994. *Nat. Genet.* **8**: 380–386.

- Klem, J.R. and Rademacher, S.C. 1996. *The rottweiler experience*. Howell Book House, New York.
- Knapp, D.W. and Waters, D.J. 1997. *Mol. Med. Today* **3**: 8–11.
- MacEwen, E.G. 1989. *Clinical veterinary oncology* (ed. S.J. Withrow and E.G. MacEwen). Lippincott, Philadelphia.
- Mellersh, C.S., Hitte, C., Richman, M., Vignaux, F., Priat, C., Jouquand, S., Werner, P., C. André, C., DeRose, S., Patterson, D.F., et al. 2000. *Mamm. Genome* **11**: 120–130.
- Misdorp, W. 1980. *Am. J. Pathol.* **98**: 285–288.
- Patterson, D.F. 2000. *Canine genetic disease information system: A computerized knowledge base of genetic diseases in the dog*. Mosby-Harcourt, St. Louis, MO.
- Patterson, D.F. 1980. *Current veterinary therapy VII*. Saunders, Philadelphia.
- Patterson, D.F., Haskins, M.E., and Jezyk, P.F. 1982. *Adv. Hum. Genet.* **12**: 263–339.
- Priat, C., Hitte, C., Vignaux, F., Renier, C., Jiang, Z., Jouquand, S., Cheron, A., Andre, C., and Galibert, F. 1998. *Genomics* **54**: 361–378.
- Priester, W.A. and McKay, F.W. 1980. *Natl. Cancer Inst. Monogr.* **54**: 1–210.
- Puffenberger, E.G., Kauffman, E.R., Bolk, S., Matisse, T.C., Washington, S.S., Angrist, M., Weissenbach, J., Garver, K.L., Mascari, M., Ladda, R., et al. 1994. *Hum. Mol. Genet.* **3**: 1217–1225.
- Pugnetti, G. 1980. *Simon and Schuster's guide to dogs*. Simon and Schuster, New York.
- Riquet, J., Coppieters, W., Cambisano, N., Arranz, J.J., Berzi, P., Davis, S.K., Grisart, B., Farnir, F., Karim, L., Mni, M., et al. 1999. *Proc. Natl. Acad. Sci.* **96**: 9252–9257.
- Rogers Clark, A. and Brace, A.H. 1995. *The international encyclopedia of dogs*. Howell Book House, New York.
- Tjalma, R.A. 1966. *J. Natl. Cancer Inst.* **36**: 1137–1150.
- Wang, R.L., Stec, A., Hey, J., Lukens, L., and Doebley, J. 1999. *Nature* **398**: 236–239.
- Wayne, R.K. and Ostrander, E.A. 1999. *Bioessays* **21**: 247–257.
- Werner, P., Mellersh, C.S., Raducha, G.M., DeRose, S., Acland, G.M., Prociuk, U., Wiegand, N., Aguirre, G.D., Henthorn, P.S., Patterson, D.F., et al. 1999. *Mamm. Genome* **10**: 812–823.
- Wilcox, B. and Walkowicz, C. 1995. *Atlas of dog breeds of the world*. T.F.H. Publications, Neptune City, NJ.
- Withrow, S.J., Powers, B.E., Straw, R.C., and Wilkins, R.M. 1991. *Clin. Orthop.* **270**: 159–168.
- Yang, F., O'Brien, P.C., Milne, B.S., Graphodatsky, A.S., Solanky, N., Trifonov, V., Rens, W., Sargan, D., and Ferguson-Smith, M.A. 1999. *Genomics* **62**: 189–202.