

# FINE SCALE ANALYSIS OF HAPLOTYPE STRUCTURE IN THE DOMESTIC DOG POPULATION

Elinor K. Karlsson<sup>1,2</sup> Claire M. Wade<sup>1,3</sup>, E.J. Kulbokas<sup>1</sup>, Tara Biagi<sup>1</sup>,  
Mark J. Daly<sup>1</sup>, Michael C. Zody<sup>1</sup>, Eric S. Lander<sup>1</sup>, Kerstin Lindblad-Toh<sup>1</sup>  
and the Dog Genome Sequencing Consortium

**Keywords:** *Canis familiaris*, dog, genome, haplotype, single nucleotide polymorphism, linkage disequilibrium, disease mapping

## 1 Introduction.

The modern domesticated dog has a unique population structure consisting of hundreds of purebred, genetically isolated breeds, the vast majority less than 300 years old. Each breed features its own defining constellation of morphologic and behavioral traits, often generated by deliberate breeding of closely related individuals. Small founder populations and inbreeding have led to the entrapment of disease alleles, and many dog breeds suffer increased susceptibility to diseases that also occur in humans, including cancers and diabetes. Understanding the genetic basis of canine diseases as well as phenotypic variation across dog breeds offers unique opportunities for understanding human biology and improving human health.

We examined the extent of linkage disequilibrium (LD), the number of haplotypes, and the degree of haplotype sharing within and between 10 dog breeds (20 individuals each) and single individuals from 24 additional breeds representing the dog population as a whole. Our observations have important implications for disease gene mapping, demonstrating that detecting association within a breed through whole genome LD analysis requires just a fraction of the SNPs needed in human studies.

## 2 Results.

The Canine Genome Sequencing Consortium has produced a high-quality draft assembly (N50 contig size 180kb and N50 supercontig size of 42Mb) of a female boxer. We identified >2.5 million single nucleotide polymorphisms (SNPs) discovered in nine breeds representing the seven major groups of domestic dogs in the United States and five canids, in addition to those identified by comparison of the two boxer haplotypes as well as comparison of the boxer to the poodle 1.5x draft reported previously. To examine haplotype structure within and between dog breeds, roughly 6% of the genome contained within ten random region was carefully surveyed using complete resequencing of 10 kb regions and genotyping of ~1,300 SNPs covering 15 Mb in each region.

We observed complete homozygosity over at least 10kb in 36% of dogs. Conditional on total homozygosity at 10kb, homozygosity extended past 15Mb in ~20% of individuals. This is consistent with data from the boxer genome assembly, where we found large regions of

---

<sup>1</sup> Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, MA, USA. E-mail: webmaster@broad.mit.edu

<sup>2</sup> Bioinformatics Program at Boston University, 48 Cummington Avenue, Boston, MA, USA  
E-mail: bioinfo@bu.edu

<sup>3</sup> Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA, USA  
E-mail: info@wi.mit.edu

homozygosity (N50=11 Mb) mixed with slightly smaller regions of heterozygosity, suggesting the presence of haplotypes many megabases in size within breeds. In addition, haplotypes and linkage disequilibrium (LD) extend over several megabases within breeds (both in the sequenced boxer and in ten other breeds studied), which is almost 100-fold longer than in the human population, but was much shorter across the whole dog population.

Using a dense marker set and the haplotype analysis software *Haploview*, we examined the fine-scale structure across 100kb regions using a modified four-gamete rule to identify haplotype blocks. We discerned strikingly different haplotype structures within individual breeds compared with the all-breed population. Within breeds, the haplotype blocks typically extended across the full 100kb windows. Across many breeds, represented by a dataset of 24 dogs of 24 different breeds encompassing the full spectrum of breed physiology, we discovered short haplotype blocks of around 4-6 kb, smaller than in the human population (15-25 kb). Across the whole dog population, we estimate 2-5 haplotypes per block (minor allele frequency > 0.04), presumably originating from the ancestral wolf chromosomes. Closer examination reveals that the long haplotypes identified within each breed (a handful per breed) represents various combinations of the shorter, ancestral haplotypes identified across many breeds, supporting the relatively recent creation of these purebred populations from a large and freely interbreeding domestic dog population. However, we note that haplotypes as large as 100 kb can be shared between breeds, suggesting that some breeds are more closely related than others.

### 3 Conclusion.

We propose to utilize this remarkable haplotype structure to implement a two-tiered mapping strategy. We will first establish haplotype and disease associations within a breed using a low-density genome-wide scan. By adding additional related breeds, associated regions can be significantly shortened, potentially identifying a segment containing only a few genes and facilitating rapid disease mutation identification.

### 4 Figures.

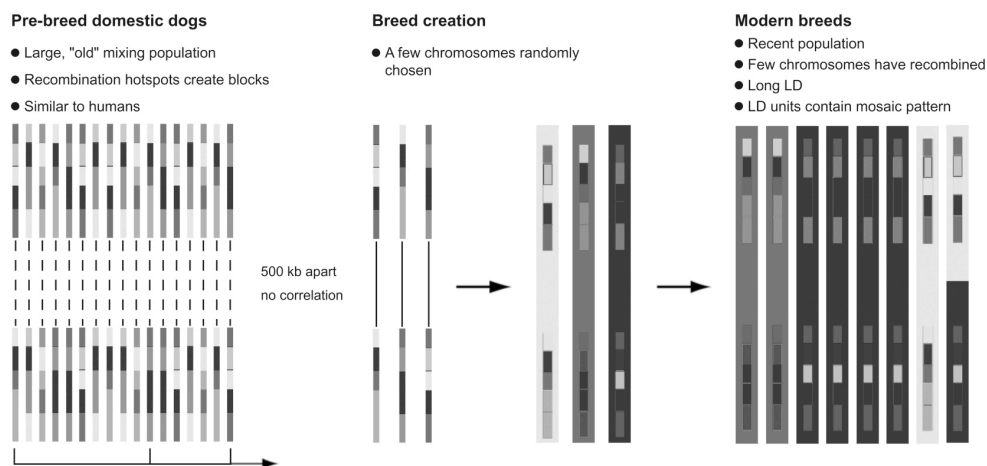


Figure 1: Haplotype structure reflects population history of domesticated dog breeds.