Sequencing the Dog & Cat Genome and Iits Implications

Kerstin Lindblad-Toh
Broad Institute of MIT and Harvard, Cambridge, MA

OVERVIEW OF THE ISSUE

Two main objectives exist for sequencing the genomes of dogs and cats:

1. Even through the human genome has been sequenced, we still have a hard time identifying all the functional elements, such as genes and regulatory elements (sequences turning on and off genes) in the human genome. Fortunately, genes and regulatory elements are highly similar between mammals, whereas non-functional sequence has changed between species. Thus, by sequencing and comparing the genomes of many mammals we can find the functional elements. An initiative is now ongoing to sequence as many as 20 mammals in the next year or two.
2. Identification of disease mutations becomes much less laborious if one already has access to the genome sequence and knowledge of the genes and other functional elements. Since disease gene mapping has been ongoing for many years in both dogs and cats, this resource will greatly facilitate canine and feline genetics. Particularly the dog has a breed structure that suggests that identifying genes for common diseases such as cancer and diabetes might be much easier than in humans. The genome sequence is accompanied by a map of the variation (markers) within and between breeds, which can be used to identify disease genes. Understanding of the patterns of variation in breeds will facilitate effective use of these markers and will lead to disease gene identification important for both canine and human health.

THE DOG GENOME

- The genome sequence of a female boxer has been generated. Each position in the genome was sampled ~7.5 times, which means that the genome is relatively complete (~99%).
- The genome has been compared to other mammalian genomes such as human, mouse and rat.
- The genes have been identified and the majority of these have a corresponding gene in other mammalian genomes. Thus, a disease gene identified in one species can be studied also in other mammals, where it is likely to cause similar disease.
- A number of possible regulatory elements have been identified by comparison of the mammals. Mutations in these elements are more likely to cause disease than those found elsewhere in the genome.
- A SNP map of several million single base variants (markers) has been generated. Small amounts of sequence from 10 breeds, 4 wolves and one coyote was compared to the boxer sequence to generate this map.
- The structure of variation within and across breeds suggests that mapping disease genes in breeds that have a high incidence of a particular disease will be feasible.
- The Broad Institute is currently collecting samples from dogs affected with osteosarcoma, hemangiosarcoma, melanoma, mammary carcinoma, lymphoma and mast cell tumors at this point. We also need older unaffected dogs to use as controls (see www.broad.mit.edu/mammals/dog/ for more details). Now that the genomic tools are in place sample collection is the key step towards identifying canine disease genes.

THE CAT GENOME:

- The genome sequence of a female cat will be ready in early fall of 2005. For this project each position has been sampled only ~2 times. Thus we expect the sequence to cover only ~80% of the cat genome.
- The genome will be compared to the other mammals to identify genes and other functional elements.
- A small effort to study the variation within the cat population is being discussed.
- The sequence will permit easier identification of disease genes.

Kerstin Lindblad-Toh
Broad Institute of MIT and Harvard
Cambridge, MA