

Genome Insider

A conversation with Sean Eddy

Having helped build the rough draft of the human genome, the U.S. genome-sequencing centers are starting to make inroads into the genetics of other animals as well. Determining which animal genomes are most worthy of sequencing, though, is the task of a panel of scientists convened by the National Human Genome Research Institute (NHGRI) of the National Institutes of Health (NIH). Sean R. Eddy, an HHMI investigator and computational biologist who studies the evolution of genomes at Washington University in St. Louis, is on the panel.

Why is the panel necessary? Why can't the genome-sequencing centers decide for themselves?

Eddy: If the genome centers had to plan which organisms to sequence next, they'd spend too much time reading zoology literature and not enough time sequencing. NHGRI's idea was an "open public contract" mechanism: If you think your dog should be sequenced as a service to the community, you can submit a 10-page white paper and explain why. If the panel agrees, the dog genome gets placed on NHGRI's list of high-priority items. Then a genome center can look at the list and say "Ah! The dog! We can do the dog." The idea is to use this constant amount of sequencing capacity and fill the pipeline.

What are your criteria for selection?

Eddy: One criterion is whether the organism has long served as a model system for molecular-genetics research. Some of these, like *Drosophila*, *C. elegans* and *E. coli*, were no-brainers, and those genomes got done early. A second criterion involves comparative genomics. Once I have, say, the *C. elegans* sequence, I want to find features in that

sequence that are important. One of the best ways to do that is to find another nematode sequence that's closely related. And a third criterion is "weird biology," so to speak, where the genome is inherently interesting in its own right. An example, which didn't get approved but probably will be before long, is the *Oxytricha* genome sequence. It's a ciliate—distantly related to paramecium. The weird thing about *Oxytricha* is that each gene is basically its own little chromosome. If we had the *Oxytricha* genome, we would have one in which a lot of the gene finding had already been done for us,



Sean Eddy, shown with Muggins, is waiting to see a white paper on the cat genome.

since the ends of the chromosomal DNA pretty much define the ends of the genes.

Is there much lobbying from researchers or other groups to get something placed on the high-priority list?

Eddy: As the genome project expands outward to cover more animals, no one wants to be in a backwater that doesn't have a genome yet. You want your genome to be up at the top of the list, or at least part of the club. Everyone is pushing an agenda, but not particularly strongly. You'll be at a pub or a meeting and someone will say, "I think we should do the platypus." And someone else will say, "Oh no, the platypus is a stupid genome. We should do the koala." So there's

that level of buzz or arguing, but there actually isn't a lot of serious lobbying.

Has the panel seen anything that caused members to laugh?

Eddy: Sure, but I'd hesitate to name names. For instance, we'll see an out-of-left-field white paper written on an organism that no one has thought hard about—they don't even know the genome size. This is an immediate killer because there are genomes out there that are big—much bigger than the human genome, like the lily genome, which is about 100 gigabases (Gb), 30 times the size of the human.

There's no way we would sequence something whose proposers hadn't asked the basic how-heavy, how-much-DNA questions.

You're a self-professed "cat person." How soon should the cat genome be sequenced?

Eddy: Well, clearly, cats are much more important than dogs [chuckling], so cats have priority. We haven't seen a white paper for the cat yet, but it's perfectly justifiable. Still, the dog's behavioral characteristics are extremely well-studied; if we had a dog genome, the rate at which people could clone genes involved in behavioral traits might be accelerated. So even though

I'm a cat person, I'm leaning toward the dog project. I want to see the dog people put in a white paper.

How many genomes can the centers get through?

Eddy: If you do everything as rough draft, you'd have the ability to do maybe 5 Gb of assembled sequence per year at the three main NHGRI centers combined, not counting the capacity of other large genome centers, such as the Wellcome Trust Sanger Institute in England or the Department of Energy's Joint Genome Institute. That means the NHGRI centers can do one or two large mammalian genomes a year, plus many other smaller genomes. —BRIAN B. REID