

NOVEMBER 05, 2009

Scientists Launch Effort to Sequence the DNA of 10,000 Vertebrates

Scientists have an ambitious new strategy for untangling the evolutionary history of humans and their biological relatives: a genetic menagerie made of the DNA of more than 10,000 vertebrate species. The plan, proposed by an international consortium of scientists, is to obtain, preserve, and sequence the DNA of approximately one species for each genus of living mammals, birds, reptiles, amphibians, and fish.

“Understanding the evolution of the vertebrates is one of the greatest detective stories in science,” said David Haussler, a Howard Hughes Medical Institute investigator at the University of California, Santa Cruz (UCSC). “No one has ever really known how the elephant got its trunk, or how the leopard got its spots. This project will lay the foundation for work that will answer those questions and many others.”

Known as the Genome 10K Project, the approximately \$50 million initiative is “tremendously exciting science that will have great benefits for human and animal health,” Haussler said. “Within our lifetimes, we could get a glimpse of the genetic changes that have given rise to some of the most diverse life forms on the planet.”

"Understanding the evolution of the vertebrates is one of the greatest detective stories in science."

- David Haussler

Haussler is one of the lead authors of an article, published online November 5, 2009, in the *Journal of Heredity*, that outlines the project. The other lead authors include Stephen J. O'Brien, chief of the Laboratory of Genomic Diversity at the National Cancer Institute, and Oliver A. Ryder, director of genetics at the San Diego Zoo's Institute for Conservation Research and adjunct professor of biology at the University of California, San Diego. Coauthors and additional authors, who together make up a group called the Genome 10K Community of Scientists (G10KCOS), include geneticists,

paleontologists, ecologists, conservationists, and other scientists representing major zoos, museums, research centers, and universities around the world.

The proposal originated at a meeting Haussler hosted at UCSC in April 2009. More than 50 scientists came together to discuss the merits of the project and its daunting logistic and financial challenges. "Some of the people at the meeting were initially skeptical," Haussler said. "But they quickly recognized the many advantages of a shared infrastructure and data analysis system."

The primary impetus behind the proposal is the rapidly expanding capability of DNA sequencers and the associated decline in sequencing costs. "We'll soon be in a situation where it will cost only a few thousand dollars to sequence a genome," Haussler said. "At that point, most of the cost will be getting samples, managing the project, and handling data."

All living vertebrates descend from a single marine species that lived 500-600 million years ago. Paleontologists do not know much about the physical appearance of that species, but because all of its descendants share certain characteristics, they know that it had segmented muscles, a forebrain, midbrain, and hind brain attached to spinal cord structures, and a sophisticated innate immune system.

That primitive vertebrate gave rise to what Haussler calls "one of the most spectacularly malleable branches of life." Vertebrates spread throughout the oceans, conquered land, and eventually took to the air. Over the course of time they produced stunning innovations, including multichambered hearts, bones and teeth, an internal skeleton that has supported the largest aquatic and terrestrial animals on the planet, and a species of primate -- *Homo sapiens* -- that has produced sophisticated language, culture, and technology.

By sequencing the DNA of 10,000 vertebrates -- roughly one-sixth of the 60,000 species estimated to be living today -- biologists will be able to reconstruct the genetic changes that gave rise to this astonishing diversity. Some parts of our DNA are very similar to the DNA of other vertebrates, reflecting our descent from a common ancestor, while other parts are markedly different. "We can understand the function of elements in the human genome by seeing what parts of the genome have changed and what parts have not changed in humans and other animals," said Haussler.

The project also will help conservation efforts by documenting the genomes and genetic diversity of threatened and endangered vertebrate species. By helping scientists predict how species will respond to climate change, pollution, emerging diseases, and invasive competitors, it will support the assessment, monitoring, and management of biological diversity.

The G10KCOS consortium has been developing guidelines for the collection, preservation, and documentation of cell lines and DNA samples. It also has been discussing potential public and private sources of funding for the project

-- estimated at \$50 million if the price of handling and sequencing each DNA sample eventually falls to \$5,000. Said Haussler: "How do you raise \$50 million? Ask nicely and make a strong case."

In planning the project, the G10KCOS group has used the Human Genome Project as a model. For example, the consortium plans to release sequencing data immediately according to standards developed for the sequencing of the human genome. Haussler also cited that project, which began before needed sequencing technologies were available, as evidence that it is worthwhile to begin planning for the Genome 10K Project before the cost of sequencing falls enough to make it feasible. "The time to start is now, or the job will get away from us," said Haussler. "The sequencing machines will be waiting, but the samples won't be ready."