

October 29, 2009

Contact: Tim Stephens (831) 459-2495; stephens@ucsc.edu

Scientists propose a "genome zoo" of 10,000 vertebrate species

EMBARGOED: Not for release until 11 a.m. U.S. Pacific Standard Time (1 p.m. EST) on Wednesday, November 4.

SANTA CRUZ, CA--In the most comprehensive study of animal evolution ever attempted, an international consortium of scientists plans to assemble a genomic zoo--a collection of DNA sequences for 10,000 vertebrate species, approximately one for every vertebrate genus.

Known as the Genome 10K Project, it involves gathering specimens of thousands of animals from zoos, museums, and university collections throughout the world, and then sequencing the genome of each species to reveal its complete genetic heritage.

Launched in April 2009 at a three-day meeting at the University of California, Santa Cruz, the project now involves more than 68 scientists. Calling themselves the Genome 10K Community of Scientists (G10KCOS), the group outlined its proposal to create a collection of tissue and DNA specimens for the project in a paper to be published online November 5 in the *Journal of Heredity*.

The project was conceived by the paper's three lead authors: David Haussler, professor of biomolecular engineering at UC Santa Cruz and a Howard Hughes Medical Institute investigator; 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 UC San Diego.

"For the first time, we have a chance to really see evolution in action, caught in the act of changing whole genomes," Haussler said. "This is possible because the technology to

sequence DNA is thousands of times more powerful now than it was just a decade ago, and is poised to get even more powerful very soon."

According to O'Brien, the cost of genome sequencing has been dropping steadily over the past decade, making the sequencing of 10,000 genomes a realistic possibility. "The original cost of sequencing the human genome by a major international consortium was over a billion dollars," he said. "With the latest sequencing technology, it now costs \$50,000 to \$100,000 per genome. The price only needs to drop down another log or so to make the sequencing of 10,000 genomes possible."

Among the authors is pioneering geneticist Sydney Brenner, a Nobel Laureate and senior distinguished fellow at the Salk Institute. "The most challenging intellectual problem in biology for this century will be the reconstruction of our biological past so we can understand how complex organisms such as ourselves evolved," Brenner said. "Genomes contain information from the past--they are molecular fossils--and having sequences from vertebrates will be an essential source of rich information."

At the UCSC meeting, 55 leading scientists representing major zoos, museums, research centers, and universities around the world hashed out the challenging logistics involved in carrying out this ambitious project.

"These are scientists who have devoted their lives to biology, evolution, and the preservation of animals, and now they see an opportunity for deeper study," Haussler said. "We came away from that meeting with a plan for moving forward and an extraordinary online database of samples from more than 16,000 different species of vertebrate animals compiled from more than 50 institutions."

After recruiting a few more key scientists to broaden the depth of the collection, the participants collaborated on the database (available at <http://sampledb.genome10k.org/>) and drafted the proposal now published in the *Journal of Heredity*.

As part of the project, the genomic database will be analyzed to reveal the evolutionary changes it records, and it will be annotated with experimental findings related to specific

sites of change. "Analysis of these data will be a far greater challenge than anything yet attempted in comparative genomics," Haussler said.

O'Brien explained that the team hopes this project will integrate genomic inference into nearly every aspect of vertebrate biological enquiry.

"Biological science is about understanding how species work, so having the genomes available for 10,000 species will give us a new sense to understand biology," he said. "The genome is like a sixth sense, adding to what we can see, smell, taste, hear, and feel. If we have this information for species that are not generally studied, it will be a particularly strong arrow in the quiver of students of biology."

The scientists have identified specimens that span a broad range of evolutionary diversity. The species include living mammals, birds, reptiles, amphibians, and fishes, many of which are threatened or endangered, as well as some recently extinct species, O'Brien said.

"We are capturing what evolution left us with before the human population started impacting species--a set of genomes inclusive of the biota that a magnificent evolutionary process has produced," he said.

Participants expect the Genome 10K Project to lay a foundation for understanding the genetic basis of recent and rapid adaptive changes within vertebrate species and between closely related species. The results can help conservation efforts by enabling scientists to predict how species will respond to climate change, pollution, emerging diseases, and invasive competitors.

"The risk of extinction is lessened for species for which we have a genome sequence, because it enables studies that can provide important information relevant to conservation," Ryder said.

Genome sequences will be particularly valuable in efforts to assess genetic diversity in endangered populations. "Any tool used so far for evaluating genetic diversity and genetic variation is overshadowed by the resolving power of genomic information," Ryder said.

The consortium agreed to a set of guidelines for sample collection, including the types and volumes of tissues, recommendations for preservation and documentation, and adherence to national and international statutes regulating the collection, use, and transport of biological specimens. Where possible, specimens for each species include both males and females and reflect geographic diversity or diversity within localized populations.

The collection will include more than a thousand frozen samples of fibroblast cells derived from 602 different vertebrate species. These cell samples, maintained by the San Diego Zoo, the National Cancer Institute, and the world's cell repositories, are a valuable resource for genetic studies, according to Ryder.

"When you sequence a whole genome, it may be 3 billion bases, of which only a few percent code for genes. If you want to quickly learn something about the genes, you can sequence the RNA transcripts of the genes. These cells are robust sources of high-quality RNA," Ryder said.

Because the evolution of species living today involved ancient genetic changes still preserved in their DNA, the Genome 10K project can help uncover answers to longstanding questions about the history of evolution. Having full genomes at hand will enable detailed studies of base-by-base evolutionary changes throughout the genome.

"Differences in the DNA that makes up the genomes of the animals we find today hold the key to the great biological events of the past, such as the development of the four-chambered heart and the magnificent architecture of wings, fins and arms, each adapted to its special purpose," Haussler said.

But now the most challenging parts of the project begin, O'Brien said. "The first challenge is to bring this whole promise into reality by actually getting samples, characterizing them, doing quality control on them, and delivering them to sequencing centers that can

accomplish the goal," he said. "The second is to very quickly raise the money to pay for sequencing and analysis and annotation of the sequences."

In addition to corresponding authors Haussler, O'Brien, and Ryder, coauthors of the paper who also served as committee chairs include F. Keith Barker of the University of Minnesota; Michele Clamp of the Broad Institute of MIT and Harvard; Andrew J. Crawford of Universidad de los Andes, Bogotá, Colombia; Robert Hanner of the Biodiversity Institute of Ontario, University of Guelph; Olivier Hanotte of the University of Nottingham; Warren E. Johnson of the National Cancer Institute, Laboratory of Genomic Diversity; Jimmy A. McGuire of the Museum of Vertebrate Zoology, University of California, Berkeley; Webb Miller of Pennsylvania State University; Robert W. Murphy of the Royal Ontario Museum, Toronto; William J. Murphy of Texas A&M University; Frederick H. Sheldon of the Museum of Natural Science, Louisiana State University; Barry Sinervo of UC Santa Cruz; Byrappa Venkatesh of the Institute of Molecular and Cell Biology, Singapore; and Edward O. Wiley of the Natural History Museum and Biodiversity Research Center, University of Kansas.

The April 2009 Genome 10K meeting was supported in part by the American Genetic Association, Gordon and Betty Moore Foundation, NHGRI Intramural Sequencing Center, and UCSC Alumni Association. Individual researchers received funding from Howard Hughes Medical Institute; Gordon and Betty Moore Foundation; Assembling the Euteleost Tree of Life; National Science Foundation; The Global Viral Forecasting Initiative; Biomedical Research Council of A*STAR, Singapore; Natural Sciences and Engineering Research Council Discovery Grant; National Basic Research Program of China, the National Natural Science Foundation of China, and Bureau of Science and Technology of Yunnan Province; MCB SB RAS Programs; Portuguese-American Foundation for Development; CIBIO; UP; University of Montana; and Portuguese Science Foundation.

#####

Note to reporters: You may contact David Haussler at (831) 459-2105 or haussler@soe.ucsc.edu; Stephen O'Brien at (301) 846-1296 or steveobrien@mail.nih.gov; and Oliver Ryder at (858) 663-9337 or oryder@ucsd.edu.

For more UCSC news, visit www.ucsc.edu/news_events.