

DECEMBER 01, 2004

Bringing to Life the Genome of an Ancient Mammal



Image Title: "I am undertaking this project out of scientific curiosity," said HHMI investigator David H. Haussler. "I want to know in molecular detail how we evolved from a furry, nocturnal, shrew-like creature, and now is the time to find out." - Tim Archibald

A group of researchers has re-created with remarkable accuracy part of the genome of the common ancestor of almost all placental mammals, a small shrew-like creature that prowled the forests of what is now Asia more than 80 million years ago. By comparing DNA sequences of 19 species of existing mammals, including humans, the researchers have reconstructed a large segment of DNA in the species from which all of today's placental mammals arose. They estimate that the reconstruction is 98 percent accurate.

The project, which was led by David Haussler, a Howard Hughes Medical Institute investigator at the University of California, Santa Cruz, based their reconstruction efforts around a region of the genome that covers about 1.1 million bases flanking the cystic fibrosis transmembrane conductance regulator (CFTR) gene. That region of the genome has been sequenced in a large number of species as part of a comparative sequencing program being conducted by the National Institutes of Health. Coauthors of the article,

which will be published in the December 2004 issue of *Genome Research*, are Mathieu Blanchette of McGill University, Eric Green of the National Human Genome Research Institute, and Webb Miller of Pennsylvania State University.

"It sounds implausible. But there's enough information to reconstruct the ancestral genome on the basis of mammals that live today."

— **David Haussler**

When geneticists hear that most DNA from the genome of a species extinct for many millions of years can be re-created with 98 percent accuracy, “jaws occasionally drop,” said Haussler. “It sounds implausible. But there's enough information to reconstruct the ancestral genome on the basis of mammals that live today. We just need to sequence the genomes of these living mammals.” The reconstructed ancestral genome will offer an invaluable vantage point from which to watch evolution at work.

According to Rasmus Nielsen, a geneticist at Cornell University who is familiar with the work, the paper is guaranteed to turn heads. “Previously it was thought that we would never really know what our ancestors looked like at the genetic level, but now it appears that we'll be able to tell,” he said. “And now that we know it is possible, I think we'll see many more attempts to do this.”

Efforts to extract DNA from fossils generally have been disappointing because DNA molecules break down over time. “Ancient DNA decays faster than [science fiction writers] would like,” said Haussler. After a maximum of about 50,000 years, DNA sequences typically are too fragmented to be pieced together. Geneticists therefore have turned to a technique that has been called “[computerized] paleogenomics” to infer the DNA sequences of past organisms.

All of the placental mammals living today are descended from an early species that lived tens of millions of years before the final demise of the dinosaurs. This species underwent a rapid diversification, splitting into the evolutionary lineages that have led to today's placental mammals. Because all of these species are descended from a common ancestral species, they all have inherited specific DNA sequences from that ancestor.

Reconstructing the DNA sequence of this ancestor is comparable to drawing conclusions about the first automobile by observing the many different kinds of automobiles existing today. Though the separate makes of automobile have changed and diversified over time, they share features that were present in their conceptual ancestor: four rubber tires, a windshield, and an internal combustion engine, for example.

The challenge for Haussler and his colleagues was to determine how the DNA sequence of the common ancestor changed in each of the evolutionary

lineages leading to current mammals. This task is not so complicated where individual nucleotides changed in a particular lineage, because the original nucleotide often was retained in other lineages. It is much more difficult where stretches of DNA were inserted or deleted in the genomes of particular species.

“DNA comes and goes,” said Haussler. “Some DNA gets deleted, and new DNA gets inserted. Tracking the history of these insertions and deletions is essential.”

Haussler's research team built a computer program that looked both for individual nucleotide changes and for insertions and deletions in the DNA sequences of a number of mammalian species, including species of pig, horse, cat, dog, bat, mouse, rabbit, gorilla, chimpanzee, and human. “Although this project took about two years, we were building on a solid foundation constructed by many other researchers over a much longer time scale,” said Haussler. What they learned they incorporated into a program that provides a detailed simulation of the evolution of DNA in these mammalian lineages. By running repeated simulations, they were able to test the accuracy of their DNA reconstruction method.

“I was very pessimistic at the beginning because I thought we weren't going to be able to do a very good job [of reconstructing the ancestral genome],” said first-author Blanchette, who was a postdoctoral fellow in Haussler's lab when work on the project began. But the research team was able to estimate the accuracy of the reconstruction, using simulations and comparisons across existing species, and they were astonished to find that the accuracy rates averaged about 98 percent. “We looked carefully for bugs in our program to see why the accuracy was so high, but we couldn't find any,” said Blanchette. Comparisons using DNA from additional species, not used in the reconstruction itself, confirmed the high accuracy.

The technique works particularly well for placental mammals, because the rapid radiation of lineages after the time of the common ancestor produced many different versions of the original sequence that can be compared. For common ancestors with fewer living descendant species, comparisons would be more difficult, as would also be the case if the speciation process were spread out over a longer time than was the case for placental mammals.

Knowing about the genome of the common ancestor of placental mammals creates tremendous scientific opportunities, according to Haussler. Most important, it reveals how DNA sequences have changed in each of the lineages leading to one of today's mammalian species. “You can feel the DNA evolving,” said Haussler, who has helped build a browser that compares DNA sequences nucleotide by nucleotide across multiple species (<http://genome.ucsc.edu>). “You can see where DNA was inserted, where it was deleted, where substitutions happened, and where they didn't happen along the evolutionary path toward humans.”

For example, even though there is currently DNA sequence data from only eight species in the *FOXP2* region of the genome in the Santa Cruz database,

more extensive comparisons of that region undertaken during independent studies by Svante Paabo and his colleagues clearly show that changes in *FOXP2* may have contributed to the evolution of fluent speech in the human lineage, Haussler said. “The nucleotide is *C* for more than 300 million years, and then suddenly it's *A*, just in the human lineage,” he said. “You can see it. That's the excitement of documenting these dramatic events that can change the nature of an organism over evolutionary time.”

But since the Santa Cruz database includes extensive information about eight species for all genes, genome-wide, not just for special, well-studied regions of the genome, like *FOXP2* or *CFTR*, Haussler said that it should be an exciting tool for researchers. “Now all researchers worldwide can do evolutionary analysis on the genes they are most interested in,” he said.

The tool is already proving valuable to Haussler's group, which has done comparisons of modern species with their common ancestor, and turned up differences in the rates of genetic change among lineages. For instance, about 22 percent of the human genome consists of new DNA insertions since the time of the common ancestor, and of the remaining DNA about 9 percent of the bases have undergone changes. In rodents, about 55 to 60 percent of nucleotides are new since the common ancestor—a heightened rate of DNA change that results partly from the shorter generation time of rodents but appears to be due to other factors as well. Haussler said that the recent development of new mathematical approaches, such as those created by HHMI investigator Philip Green at the University of Washington, should aid greatly in understanding how subtle patterns of natural mutation in DNA contribute to evolutionary change in mammals.

By examining which individual nucleotides and sections of DNA have changed and which have remained the same, geneticists can draw conclusions about which parts of our DNA are essential to the functioning of life. “Evolution is our laboratory,” said Haussler. “Regions of the genome that have very important functional roles are under stronger negative selection, so you see greater DNA conservation in these regions.”

These observations apply both to the DNA regions that code for proteins and to other parts of the genome. In addition to the approximately 1.5 percent of the human genome that codes for proteins, several percent more has been strongly conserved and must be playing some role in DNA's function. “Someday we are going to understand what some of these other noncoding elements in the middle of introns and upstream from genes are doing,” Haussler said. “But most of this is terra incognita today.”

Haussler said he is confident that significant medical benefits will accrue from a better understanding of the genome, this particular project was motivated by pure scientific curiosity. “I want to know in molecular detail how we evolved from a furry, nocturnal, shrew-like creature, and now is the time to find out.”