

MAY 06, 2004

Analysis Uncovers Critical Stretches of Human Genome

Hundreds of stretches of DNA may be so critical to life's machinery that they have been ultra-conserved throughout hundreds of millions of years of evolution. Researchers have found precisely the same sequences in the genomes of humans, rats, and mice; sequences that are 95 to 99 percent identical to these can be found in the chicken and dog genomes, as well.

Most of these ultra-conserved regions do not appear to code for proteins, but may instead play a regulatory role. Evolutionary theory suggests these sequences may be so central to mammalian biology that even small changes in them would compromise the animal's fitness.

"The most exciting thing for me is that the ultra-conserved regions we have identified do represent evolutionary innovations that must have happened sometime during vertebrate development."

— David Haussler

Led by Howard Hughes Medical Institute investigator David Haussler, at the University of California at Santa Cruz, the researchers published their findings online May 6, 2004, in *Science Express*, the Web counterpart of the journal *Science*. The lead author on the paper was Gill Bejerano in Haussler's laboratory. Also co-authoring the paper were John Mattick and his colleagues from the University of Queensland in Australia.

It's extraordinarily exciting to think that there are these ultra-conserved elements, so many of which are near well-studied genes, that weren't noticed by the scientific community before because we didn't have the comparative data that highlighted these regions, said Haussler. The real credit goes to the prodigious efforts in sequencing these multiple genomes, which have given us this tremendous opportunity, opening our eyes to these very unusual genomic elements, he said.

According to Haussler, the researchers were launched on their analysis when initial studies hinted at major regions of conserved DNA sequences. When

we had compared the human and mouse genomes, we found that about five percent of each of these showed some kind of evolutionary selection that partially preserved the sequence, he said. We got excited about this because only about 1.5 percent of the human genome codes for protein. So five percent was about three times as much as one might expect from the standard model of the genome, in which it basically codes for proteins, with a little bit of regulatory information on the side, and the rest is nonfunctional or junk DNA.

These initial findings suggested that quite a lot of the genome was performing some kind of regulatory or structural role - doing something important other than coding for proteins, said Haussler.

When the rat genome sequence became available, the researchers decided to search for the most extreme cases of conservation among the three mammalian species. They looked for long stretches of DNA, at least 200 base-pairs in a row, that were identical among humans, rats and mice. Statistically, the likelihood that a sequence of this length would appear unchanged among all three genomes by chance was infinitesimally small.

The results, said Haussler, were startling. The comparison of the three genomes revealed 481 such elements that they called ultra-conserved. What really surprised us was that the regions of conservation stretched over so many bases. We found regions of up to nearly 800 bases where there were absolutely no changes among human, mouse and rat.

Although 111 of these ultra-conserved elements overlapped with genes known to code for proteins, 256 showed no evidence that they overlap genes, and another 114 appeared inconclusively related to genes. In the 111 that overlapped genes, relatively small portions were actually in coding regions. Many were either in untranslated regions of the gene's messenger RNA transcript or in regions that are spliced out before the message is translated into protein.

Ultra-conserved regions were often found overlapping genes that specified proteins involved in binding RNA and regulating its splicing. One of these genes is known to regulate its own splicing so as to either include or not include an ultra-conserved section, depending on conditions. There is also evidence for regulatory 'crosstalk' with another member of the same gene family at this point. We may want to investigate further to see if these ultra-conserved elements that overlap RNA-processing genes are part self-regulating networks of RNA-processing activity, said Haussler.

As to the function of the conserved regions that don't overlap genes, Haussler said, there are hints that they may be involved in regulating transcription, but if so, it's a complete mystery how they work. What people find most interesting and exciting about these results is that they raise more questions than they answer.

For example, said Haussler, the many conserved elements that are not in genes still tend to cluster in groups at certain places on the chromosomes.

These clusters are often next to or surrounding genes that are known to play a role in regulating the activity of other genes in embryonic development. The conserved elements in the cluster can be up to a million bases away from the gene, however. The fact that conserved elements are hanging around the most important development genes suggests that they have some role in regulating the process of development and differentiation, said Haussler, even though they are often far away from the gene itself.

What really surprised us was that when we included the chicken genome in this comparison, we found that nearly all these regions still showed amazingly high levels of conservation, he said. In 29 cases it was 100%. This, despite the fact that the common ancestor of chickens, rodents, and humans is thought to have lived about 300 million years ago, he said.

However, the researchers found these regions to be significantly less conserved in the genome of the fish called fugu. And when they extended their comparisons to the even more ancient genomes of the sea squirt, fruit fly and roundworm, they found very little evidence of these conserved elements. The sea squirt exhibits a simple spinal cord early in its life cycle, and so it is more closely related to vertebrates than are flies or worms.

The most exciting thing for me is that the ultra-conserved regions we have identified do represent evolutionary innovations that must have happened sometime during vertebrate development, because we see such large pieces that no longer match in fish, and almost nothing in sea squirt. They must have evolved rather rapidly while our ancestors were still in the ocean, with some further evolution when animals first started to colonize land; after that they must have essentially frozen evolutionarily.

This suggests that these were foundational innovations that were very important to the species, and since the conserved elements are different from one another, that each one was important in some particular way. It is possible that further innovations in other interacting elements created so many dependencies that these foundational elements couldn't be mutated any more without disrupting something vital, said Haussler.

Besides the fact that the purpose of the non-coding ultra-conserved elements remains unknown, said Haussler, the researchers also do not understand the molecular mechanism of their action that requires them to be so faithfully preserved. A major question is what molecular mechanism would demand such a relentless conservation over hundreds of bases, he said. There is still the possibility that these regions are not so vital to the function of the organism, but in fact change very slowly for some other reason, such as lack of susceptibility to mutation, or hyper-repair. But it is even harder to imagine a mechanism for that.

Further studies, said Haussler, will involve not only more detailed comparisons of the conserved elements, but also laboratory studies exploring their functionality.