

Genome project turns up evolutionary surprises

The latest studies of the instructions embedded in the human genome are revealing how evolution has shaped our species.

On page 799 of this issue^{1,2}, and in a themed issue of *Genome Research*³, scientists report the first findings from a project called ENCODE. This 'encyclopedia of DNA elements' attempts to discover how our cells make sense of the DNA sequence in the human genome. Already, ENCODE is up-ending one piece of conventional scientific wisdom: the idea that biologically relevant DNA resists change over evolutionary time.

ENCODE aims to catalogue all the "functional elements" in the genome — the DNA sequences that control how and when our cells use our genes. Most of these controls seem to be written into so-called non-coding DNA, which does not make a detectable protein product. Because organisms depend on functional elements working correctly, scientists have long thought that such elements should not change much over evolutionary time. So researchers have mostly looked for key functional elements in non-coding DNA that is the same across species, known as conserved or constrained DNA.

But ENCODE is the first project to compare long stretches of non-coding DNA across many mammals, from mice to monkeys to humans. This comparison suggests that evolutionary processes don't always freeze functional DNA in place.

"The fact that we found so much functional

sequence that did not seem to be evolutionarily constrained across all mammals is really surprising," says Elliott Margulies of the National Human Genome Research Institute in Bethesda, Maryland, who co-chaired one of the ENCODE analysis groups.

The finding comes from the ENCODE pilot project, which used multiple methods to collect and analyse data on just 1% of the human genome — not an easy task (see 'Scaling up to a monumental task'). In one part of the project, groups of experimental biologists used a suite of laboratory techniques to find out what portions of the genome might be functional. Meanwhile, groups of computational biologists compared the ENCODE sequences across humans and 28 other animals to find constrained regions of DNA that had changed little throughout evolution.

But when the different groups compared their results, they found that their predictions about key portions of the genome didn't always agree: the biologists' list of functional sequences didn't match the computational group's list of constrained sequences.

At first, many were sceptical of this result, says John Stamatoyannopoulos of the University of Washington in Seattle, a co-chair of one of the ENCODE analysis groups. "It raised some eyebrows," he says. "But eventually all the ENCODE groups started coming out with the same thing." Overall, biologists found no evidence of function for about 40% of the



constrained ENCODE regions. On the flipside, about half of the functional elements found in non-coding DNA were totally unconstrained.

The finding that many constrained regions weren't considered to be functional is not too surprising, because it is unlikely that ENCODE included enough tests on enough different types of cells to capture every major aspect of biology. But the idea that important DNA might also be unstable is newer, and intriguing, because it undermines the assumption that biological function requires evolutionary constraint.

"We're generalizing this principle over mammals, and over many functional elements," says Ewan Birney, head of genome annotation at the European Bioinformatics Institute in Cambridge, UK, and a leader of ENCODE. "We're coming out quite strongly that this is not merely a curiosity of our genome — it's a really important part of the way our genome works."

But how can major components of the mammalian genome change essentially randomly over time? That is not entirely clear. The authors of the ENCODE paper speculate that the unconstrained genomic regions are evolving "neutrally" — that is, they are

Scaling up to a monumental task

ENCODE, largely funded by the US National Human Genome Research Institute (NHGRI) in Bethesda, Maryland, began in 2003, the year the complete sequence of the human genome was first announced. But if scientists thought that that 13-year quest was difficult, ENCODE could prove even harder, says the NHGRI's Eric Green. "I think sequencing the genome is going to look so easy compared with what ENCODE is trying to accomplish," he says.

ENCODE is now scaling up from its pilot phase to cover

the entire human genome.

The pilot showed how challenging this task will be. It involved 308 scientists from 10 countries. But even more dauntingly, it unites scientists from two different scientific cultures — experimental and computational biology. And comparing the results of all the different experiments included in the project proved difficult, because not every lab performs them in exactly the same way.

ENCODE also revealed that the genome is far more

complex than scientists expected. Last month, the NHGRI announced a US\$62.4-million project — modENCODE — to perform ENCODE analyses on the model organisms fruitflies and roundworms⁴. It is hoped that this will help scientists make sense of the data when ENCODE is complete. The NHGRI has already spent \$42.5 million on ENCODE and plans to extend the analysis to the rest of the human genome at a cost of about \$20 million a year. **E.C.**



ARCTIC-IMAGES/CORBIS

constantly changing in ways that are neither good nor bad for the individual. This means that, on the whole, many genetic changes simply don't affect overall biology.

This has major consequences for understanding the relationship between genetics and biology, Birney says. "It means, for example, that if you look at some conserved piece of biology — say, how the kidneys work in mice and humans — not all of those bits of biology will be conserved or constrained at the level of the DNA bases, and that's quite a strong shift."

But not everyone agrees with that take. For example, John Mattick at the University of Queensland in Brisbane, Australia, argues that the widely accepted calculation of the baseline, or neutral, rate of mammalian evolution is flawed. Because measurements of constraint rely on a comparison with the neutral rate, it is possible that many of ENCODE's so-called unconstrained regions really aren't unconstrained, Mattick argues.

"I would have said that this finding suggests that many regions of our genome are evolving under weak selection pressure, or that our measurements of the neutral rate of evolution

The ENCODE project aims to catalogue all the 'functional elements' in the human genome.

are incorrect," says Mattick, who is an author on the ENCODE paper.

In fact, Mattick thinks scientists are vastly underestimating how much of the genome is functional. He and Birney have placed a bet on the question. Mattick thinks at least 20% of possible functional elements in our genome will eventually be proven useful. Birney thinks fewer are functional. The loser will buy the winner a case of the beverage of his choice.

Meanwhile, other scientists are gathering data to answer new questions raised by ENCODE. Many hope that other ongoing studies, such as comparable genome sequences from additional primate species, will help decide which parts of the ENCODE data to study first.

Erika Check

1. The ENCODE Project Consortium *Nature* **447**, 799–816 (2007).
2. Greally, J. M. *Nature* **447**, 782–783 (2007).
3. *Genome Res.* **17**, Issue 6 (2007).
4. *Nature* **447**, 361 (2007).

WORDWATCH

Cyber warming: The greenhouse effect caused by carbon dioxide emitted specifically by the manufacture, operation and disposal of computers. A UK government task force has been set up to tackle the issue.



STOCKBYTE SILVER/GETTY

ON THE RECORD

"We could not talk about a football World Cup, but about a World Cup at limited altitude."

Bolivian President Evo Morales slams a decision by FIFA, soccer's governing body, to ban international matches at elevations above 2,500 metres.

SCORECARD

Postage stamps



Dutch mail company TNT Post has launched a range of seed-impregnated stamps that can be planted after use, and will sprout a range of flowers.

US vacations



A survey shows that 20% of US holidaymakers take their laptops so they can stay in touch with the office — and some 40% check their work e-mail while they're supposed to be relaxing.

NUMBER CRUNCH

7 minutes 54 seconds

is the average duration of coitus, according to a US survey.

14,191 kilometres is the distance travelled by Earth around the Sun during that time, British astronomers calculate.

143 million kilometres

is the total amount the Earth moves during sex if you factor in the expansion of the Universe. For a precise, personalized distance based on your specific latitude (and performance), visit www.earthmove.info.

Sources: BBC News, FourFourTwo, inhabitat.com, ABC News, Cheltenham Science Festival