Human Genome Project: computer power maps the code of life

Video transcript: clip from *Megabits*, first broadcast on BBC Two on 5 Jul 2012

NARRATOR:

In recent years biologists have made a massive scientific breakthrough in discovering, for the first time, how to sequence species' entire genetic codes.

DR NICHOLAS THOMSON:

We can see DNA – you can see it by eye. But you can't decode it, you can't understand what it's saying.

And it's only by sequencing that DNA – you can decode it, you can understand the messages, the sequences, the genes that are encoded by that DNA.

NARRATOR:

What sequencing a species' entire DNA into a genome means, is that scientists are effectively now capable of examining the building blocks of life.

DR NICHOLAS THOMSON:

Sequencing gives you a glimpse – it's actually a very privileged view – of the instruction manual for an organism.

NARRATOR:

To give you some idea of the challenge computer programmers face – a single genome can be composed of over 3 billion base pairs of data.

In computing terms, that would equate to over 4 gigabytes of storage space. Or if you were to read it out loud, it would take you over nine years of continuous reading.

DR NICHOLAS THOMSON:

Using the current sequencing technologies, we generated in 6 months more data than we had produced in the previous ten years.

This tells you the pace – the development of both sequencing and it also gives you an idea of how computing has had to keep up with that pace of sequencing.

NARRATOR:

But to explore all the possibilities that reading this data gives them, scientists have needed to be able to compare thousands and thousands of separate individual codes.

That's so much data it's making my head hurt just thinking about it.

DR NICHOLAS THOMSON:

Both the processing power and the memory for storing all the data we are generating is absolutely huge.

And without that we wouldn't even be able to entertain many of the projects that we are now initiating. We wouldn't even be able to start those projects.

NARRATOR:

So whilst it's biologists who initially have been able to create the data, it's people with a knowledge and understanding of how to manipulate some serious computer hardware that has allowed them to analyse it efficiently and effectively.