

## DNA could safely store data being generated worldwide

23 January 2013

The material that holds the codes for life could be the answer to storing the ever-increasing volume of data being generated around the world safely and cheaply.

According to researchers at the European Bioinformatics Institute (EBI), a cupful of custom-built strands of DNA can encode and store 100 million hours of high-definition video in a form that requires no energy to maintain and would be stable for tens of thousands of years.

The ever-growing volume of new data is becoming a pressing problem. As more and more forms of content are created, especially in the sciences, storing this information and keeping it safe is difficult. Hard disks require power and mechanical maintenance, while magnetic tape degrades. Ironically, DNA is at the heart of the problem — sequencing genes, which is becoming increasingly vital to the health science sector, generates massive amounts of data.

However, it is possible to synthesise DNA with a defined sequence of bases — the chemical groups whose order along the DNA strand encodes information. 'We already know that DNA is a robust way to store information because we can extract it from bones of woolly mammoths and make sense of it,' said research leader Nick Goldman of the EBI, which is part of the European Molecular Biology Laboratory.

Using DNA as the basis of an information storage system poses problems, however, as Goldman and co-author Ewan Birney describe in a paper in *Nature*.

It's only possible to manufacture DNA in short strings, and both writing it and reading it are prone to errors, especially when the same base is repeated in the sequence. 'We figured, let's break up the code into lots of overlapping fragments going in both directions, with indexing information showing where each fragment belongs in the overall code, and make a coding scheme that doesn't allow repeats,' explained Birney. 'That way, you would have to have the same error on four different fragments for it to fail, and that would be very rare.'

Goldman and Birney went to a Californian biotech company, Agilent Technologies, which specialises in synthesising custom DNA sequences. Agilent took an MP3 of Martin Luther King's 'I have a dream' speech, a photo of EBI in jpeg form, a text file of all of William Shakespeare's sonnets and a pdf of James Watson and Francis Crick's paper on the structure of DNA and transformed them into DNA strands using the EBI team's encoding strategy.

The resulting DNA, the size of a dust fragment, was posted back to EBI, where the sequences were read and decoded — with no errors. 'As long as someone knows what the code is, you'll be able to read it back if you have a machine that reads DNA,' Goldman said.

The researchers need to do more work on the coding scheme and the practicalities of coding, but believe they now have the basis of a commercially viable system.



Sign up for **FREE**  
email services



Advertise  
your vacancy