

Easy access to genetic testing

December 4 2013, by Heather Vincent



It's not just about hardware, software matters too. Credit: Argonne

[Frederick Sanger](#), who died recently at the age of 95, won two Nobel prizes in chemistry for his methods for sequencing proteins and DNA. Proteins were of more direct interest to many people because many disease-causing mutations are observed as changes in proteins. But we can find the protein sequence from the DNA sequence, and it turned out to be faster too, eventually playing a part in the [Human Genome Project](#).

Sanger was a chemist who wanted to understand biological polymers, so

biology and chemistry are two strands leading to the success of the Human Genome Project. The third, newer, strand is computer science.

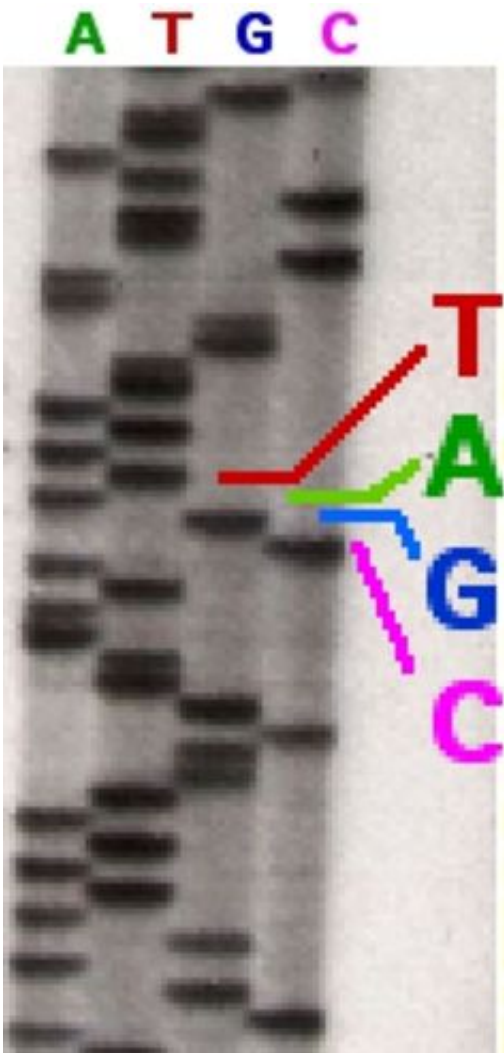
Alan Turing's Automatic Computing Engine, the ACE, ran its first program in 1950, just three years before the landmark publication of the structure of DNA. In 1970, EF Codd published [a data model](#) that, although not obviously significant to biologists at the time, has proved to be critical to the organisation and management of large amounts of data. By the time Sanger's DNA sequencing method was published, in 1977, computer scientists were ready to speed the way towards the [announcement](#) of the completion of the first draft of the [human genome](#).

Computers are faster

DNA sequence codes for the amino acids that form proteins, and the code had been worked out earlier. At first, DNA sequence was read from a gel and then translated to amino acids. This was slow and tedious.

By 1976, computers were appearing in labs, so computer scientists could work more closely with chemists and biologists producing DNA sequence data. Translating the triplet DNA code into an [amino acid sequence](#) and printing it out is a set of tasks easily converted into a computer program.

Over the next few years labs around the world began to produce more sequence data. Scientists were keen to get hold of data from other labs to compare sequences. For example, sequences from different beetles can be compared to see how closely the beetles are related



Part of a radioactively labelled sequencing gel. Credit: John Schmidt

The earliest sequence records were printed in journals, but as labs switched to using computer storage for sequence data, sequences could be shared through early networks.

By 1980, Michael Ashburner, a geneticist at the University of Cambridge, was ready to compare his sequence data with data held in Stanford University. He describes the problems that he encountered in using an early version of the internet. The whole process was

complicated, partly because the protocols used in the UK and in the US were different.

The only place that had an interface between the two protocols was University College, London, and they were very helpful, giving us 5 kb of disk space.

The way you did it was to dial up your local packet switching exchange at the Post Office and connect to the Rutherford Appleton Laboratory. You then typed in some code which connected you to UCL where you could use TCP/IP. I had a dumb terminal, that is a box with no memory, so everything had to be captured by a printer in parallel.

A shared data repository was clearly a better solution to the data sharing problem so, in 1981, the European Molecular Biology Laboratory (EMBL) electronic library of nucleotide sequence data was founded in Heidelberg. The repository grew rapidly, so a database management system was also needed. The reorganised data could be easily managed using Codd's data model.

The sequence data were now freely available over the new internet, and new sequences could be deposited to the database. Sister databases were also established in the US and in Japan, so users worldwide can now share data from their own laptops.

Finding meaning in the data

Ashburner had been keen to contact Stanford to compare sequence data because he had been studying a gene coding for the enzyme alcohol dehydrogenase. Alcohol dehydrogenases break down alcohols, and so help to protect cells from the toxic effects. For example, the enzyme is important to the fruitfly as it enables it to feed on fermenting fruit. Ashburner wanted to look for possible variations in the gene found in

different species of fruit fly. By looking at small differences in the sequences he could work out how damage to a gene can have an important effect on a protein.

The methods used by Ashburner and others for comparing sequences are now used routinely in biology, agriculture and medicine. For example, genome sequencing can be used to find out which type of bacteria are responsible for outbreaks of food poisoning. Another common use is in genetic tests, to see whether a patient has a damaged form of a gene. Angelina Jolie decided to have a test to find out whether she had inherited a damaged form of the BRCA1 gene. People with the faulty gene may have a high risk of developing breast cancer.

New challenges

Since the completion of the Human Genome Project sequencing machines have become much faster. This creates new problems for scientists who need to handle huge amounts of data. Just moving all the computer files is a difficult task, so we need new ways to compress the data, to make the files smaller and easier to move. Programs need to run faster too. New hardware can help here, but programmers are also thinking up new shortcuts to getting at the results.

Biologists learned a lot from comparing sequences to find damage to single genes, such as BRCA1, but we need to do more to find the causes of many rare diseases. We can learn much more by comparing whole genome sequences. The government recently announced the launch of the Personal Genome Project UK, and now there will be additional funding for the improved programs and strategies that will be needed to handle and find meaning in the new [sequence data](#) that will be generated. The next few years will bring exciting challenges for computational biologists.

This story is published courtesy of [The Conversation](#) (under Creative Commons-Attribution/No derivatives).

Source: The Conversation

Citation: Easy access to genetic testing (2013, December 4) retrieved 20 April 2024 from <https://medicalxpress.com/news/2013-12-easy-access-genetic.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.