

Biology is in the grip of a genomics revolution – but Mark Viney thinks this may have taken a wrong turn. He says that just DNA sequencing genes isn't enough – we need to get back to finding out what each one does.

The failure of. genomics

Biology is the science of the 21st century, and genomics is its all-powerful master. At least that's how it might seem. Genomics – a catch-all phrase, but generally meaning any large-scale DNA sequencing of genomes – has come to dominate biological research because now it's possible to sequence DNA quickly and cheaply.

Impressive technical developments in DNA sequencing methods combined with sophisticated – but now routine – computer-based analysis of the data they produce have made this possible. The genomes of multicellular organisms are large – that of a 1mm-long nematode worm with less than a thousand cells has 100 million DNA bases; our own genome is more than 30 times bigger. The genomics revolution means that what ten years ago a small team of researchers could achieve in a year, a PhD student can now do in a week.

We have sequenced lots of DNA from many organisms. For many microbes we now know their whole genome, containing all of their genes. There are whole genome sequences of many animals and plants too, especially of model study species – intensively studied species where findings are expected to apply very widely – such as the nematode worm *C. elegans*, the fly

Drosophila, mice, and the plant *Arabidopsis*. There are also whole genome sequences of species that are important to humans, such as rice, wheat, chickens and pigs. For many other organisms we have large quantities of DNA sequence data, although not complete genome sequences.

But there's a problem: while we have lots of DNA sequences that we can read, we don't understand what we are reading. The genetic code that converts the four bases of DNA – A, C, G and T – into the amino acids that make up proteins is well known. We've used this code to read DNA sequences, particularly to find protein-coding genes.

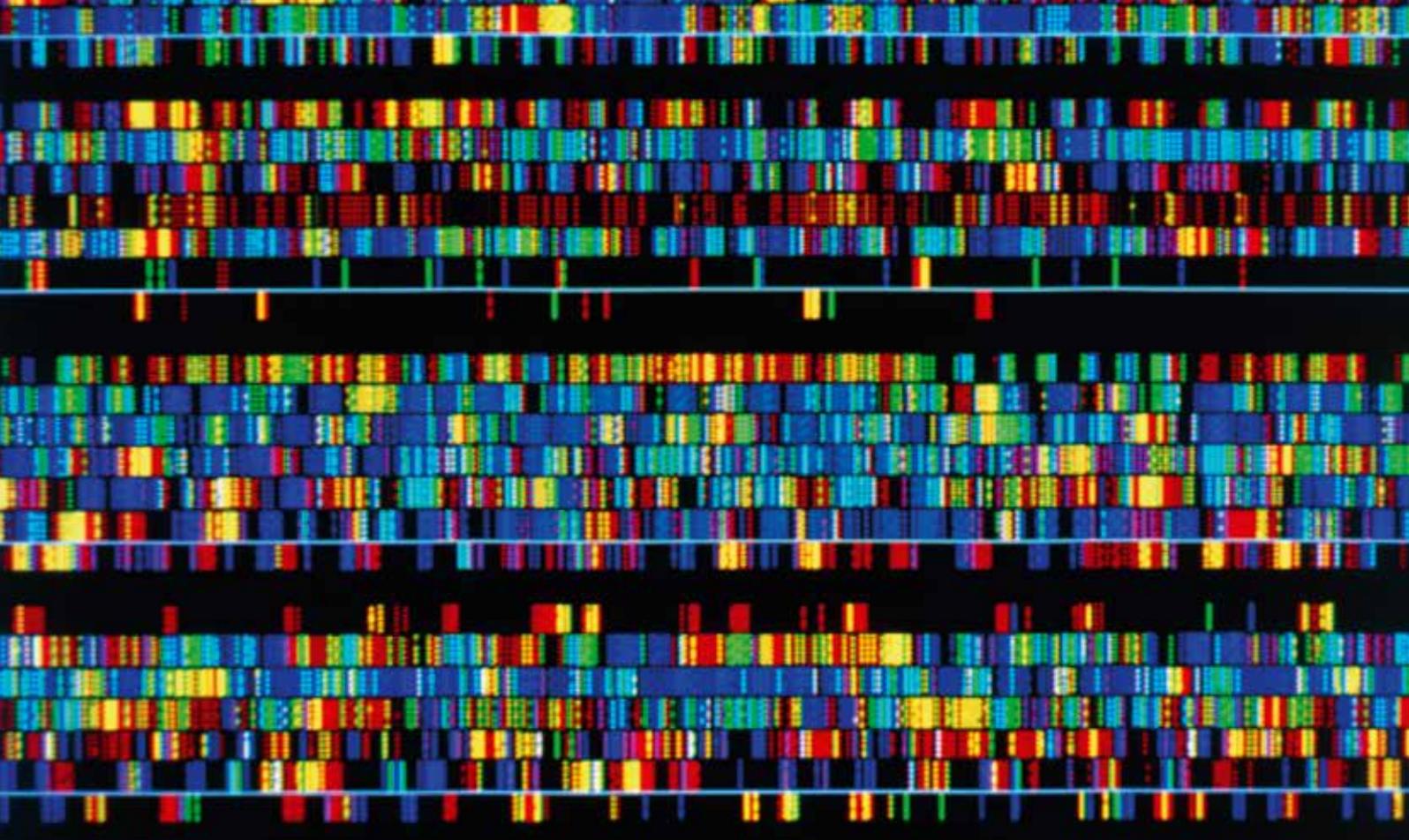
This has told us a lot. There's now a good idea of how genes are arranged along chromosomes – for many species genes are concentrated in the centre of chromosomes. We also now know how many genes organisms have – 20,000 for the *C. elegans* worm, 27,000 for the plant *Arabidopsis*. But it has also revealed how little else we know. While we can read the sequence of genes we do not understand what we are reading. In the same way, I know the Greek alphabet so I can (just about) read a Greek sentence, but I don't understand what I'm reading. Genomics hasn't added to our understanding of what

genes do, and we've largely stopped the basic science studies that can discover this.

Central to analysing DNA sequences is comparing sequences computationally. Any DNA sequence can now be compared to all other known DNA sequences in matters of seconds. If you have a gene's DNA sequence you can easily find 'hits' – DNA matches – to other genes. If you get a good 'hit' you might think you've discovered what your gene does, but actually you've only discovered that your gene is similar – perhaps very similar, perhaps only partially similar – to a gene in another organism.

From reading to understanding

Getting a database hit is never going to tell you what a gene does, because two partially similar DNA sequences may have very different roles and functions. Astonishingly, we don't know what most genes do – at best we know the function of perhaps a very few thousand. There are many thousands of sequenced genes that must have totally new and as yet undiscovered functions. Reading DNA sequences will never discover these new functions because gene–gene comparisons are necessarily conservative because they can only ever infer the same or similar function, never anything novel. But computationally



James King-Holmes/Science Photo Library

“While we can read the sequence of genes, we do not understand what we are reading.”

comparing DNA sequences and finding database hits can be important. For example, it has shown just how many genes are shared among different types of organisms, meaning that organisms' very different appearances and biology often belie their genetic similarity.

How do we work out the function of genes? In most cases this is done with 'knockout' studies. These mutate a gene to stop it working and then ask what this does to the organism, letting us then infer the gene's normal function. It's the same with my car: if I removed a small bit of its engine, the resulting malfunction would tell me what this part did when everything was working normally.

Of course, organisms have many vital genes without which an organism can't live, and so many gene knockouts are lethal to the organism. In the same way, lots of bits of a car's engine are also vital to make it go at all. As genomics has come to dominate, we've forgotten just how few genes' function we really know and understand.

Analysing DNA sequence data without understanding the sequence itself can still be very powerful. DNA sequence data have contributed to huge advances in understanding the evolutionary history of living organisms. From this we now know many of the deep, ancient relationships among major groups of animals and plants. Eventually we will work out the family tree of all life from genomic data. These phylogenetic advances have probably been genomics' greatest success in biology so far. DNA sequences can also be used to characterise the genetic differences among individuals, and so to work out the population genetic patterns and processes that occur in natural environments.

Variation is the raw material on which natural selection acts during evolution, and genomics has a key role in uncovering and understanding this genetic variation.

Comparing gene sequences is always conservative and can't discover gene function because DNA sequence analysis is reading without understanding. Our

lack of understanding is astonishing, and rarely talked about. But the times they are a-changin'. For many model species, genomics' failure is being turned towards a potential stunning success with systematic gene knockout programmes.

These involve systematically knocking out genes to work out what each one normally does. These are large and difficult studies – genomics almost looks trivial by comparison. The largest is the International Mouse Phenotyping Consortium, which is knocking out each mouse gene in turn, and then putting each knockout mouse through a full medical and health screen. Knockout studies like these show the way forward and are going to be genomics' salvation so that we'll finally be able to understand the DNA we've been reading for so long.

● Professor Mark Viney studies the biology of nematode worms at the University of Bristol.

■ Email: mark.viney@bristol.ac.uk
nematode.bio.bris.ac.uk

This article is based on a piece that first appeared in *Trends in Parasitology* in 2014.