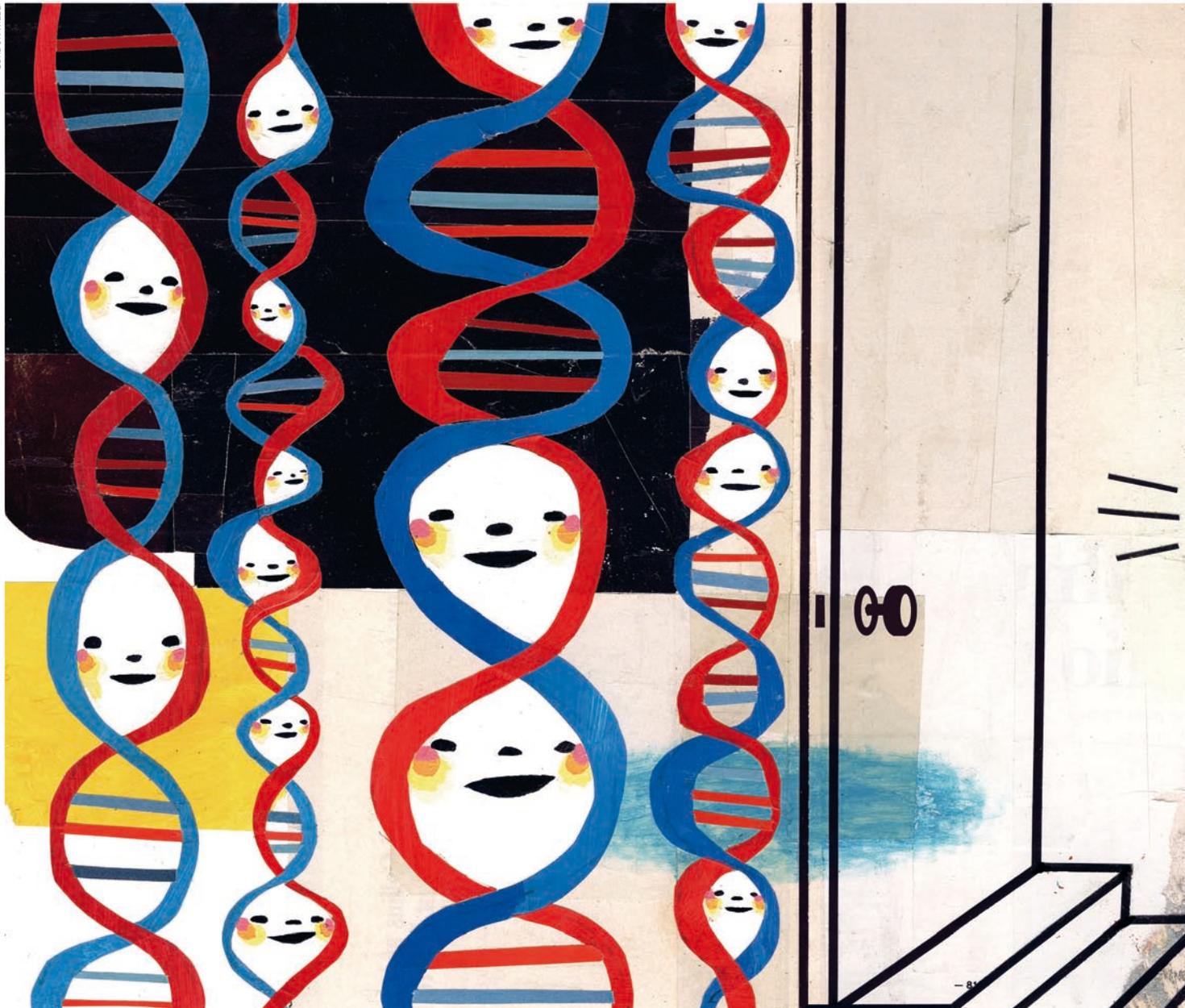


Many of our genes have no obvious relatives or evolutionary history. So where did they come from, wonders Helen Pilcher?

All alone

GORDON WIEBE





NOT having any family is tough. Often unappreciated and uncomfortably different, orphans have to fight to fit in and battle against the odds to realise their potential. Those who succeed, from Aristotle to Steve Jobs, sometimes change the world.

Who would have thought that our DNA plays host to a similar cast of foundlings? When biologists began sequencing genomes, they discovered that up to a third of genes in each species seemed to have no parents or family of any kind. Nevertheless, some of these “orphan genes” are high achievers, and a few even seem have played a part in the evolution of the human brain.

But where do they come from? With no obvious ancestry, it was as if these genes had appeared from nowhere, but that couldn't be true. Everyone assumed that as we learned more, we would discover what had happened to their families. But we haven't – quite the opposite, in fact.

Ever since we discovered genes, biologists have been pondering their origins. At the dawn of life, the very first genes must have been thrown up by chance. But life almost certainly began in an RNA world, so back then, genes weren't just blueprints for making enzymes that guide chemical reactions – they themselves were the enzymes. If random processes threw up a piece of RNA that could help make more copies of itself, natural selection would have kicked in straight away.

As living cells evolved, though, things became much more complex. A gene became a piece of DNA coding for a protein. For a protein to be made, an RNA copy of the DNA has to be created. This cannot happen without “DNA switches”, which are actually just extra bits of DNA alongside the protein-coding bits saying “copy this DNA into RNA”. Next, the RNA has to get to the protein-making factories. In complex cells, this requires the presence of yet more extra sequences, which act as labels saying “export me” and “start making the protein from here”.

The upshot is that the chances of random mutations turning a bit of junk DNA into a new gene seem infinitesimally small. As the French biologist François Jacob famously wrote 35 years ago, “the probability that a functional protein would appear de novo by random association of amino acids is practically zero”.

Instead, back in the 1970s it was suggested that the accidental copying of genes can result in a single gene giving rise to a whole family of

genes, rather like the way animals branch into families of related species over time. It's common for entire genes to be inadvertently duplicated. Spare copies are usually lost, but sometimes the duplicates come to share the function of the original gene between them, or one can diverge and take on a new function.

Take the light-sensing pigments known as opsins. The various opsins in our eyes are not just related to each other, they are also related to the opsins found in all other animals, from jellyfish to insects. The thousands of different opsin genes found across the animal kingdom all evolved by duplication, starting with a single gene in a common ancestor living around 700 million years ago (see diagram, page 40).

Most genes belong to similar families, and their ancestry can be traced back many millions of years. But when the yeast genome was sequenced around 15 years ago, it was discovered that around a third of yeast genes appeared to have no family. The term orphans (sometimes spelt ORFans) was used to describe individual genes, or small groups of very similar genes, with no known relatives.

“If you see a gene and you can't find a relative you get suspicious,” says Ken Weiss, who studies the evolution of complex traits at Penn State University. Some suggested orphans were the genetic equivalent of living fossils like the coelacanth, the last surviving members of an ancient family. Others thought they were nothing special, just normal genes whose family hadn't been found yet. After all, the sequencing of entire genomes had only just begun.

So many orphans

But as the genomes of more and more organisms were sequenced, genetic family reunions proved to be the exception rather than the rule. Orphan genes have since been found in every genome sequenced to date, from mosquito to man, roundworm to rat, and their numbers are still growing.

The study of orphan genes is still in its infancy, and we know very little about most of them. Those we do know about are a mixed bag. Some are involved with the repair and organisation of DNA, or in controlling the activity of other genes. The insect orphan *flightin*, which encodes a muscular wing protein, evolved to aid flight. And in a study published last year, Manyuan Long of the University of Chicago and his team showed that two recently evolved young insect

orphans help shape foraging behaviour in the fruit fly *Drosophila*.

In corals, jellyfish and polyps, orphan genes guide the development of explosive stinging cells, sophisticated structures that launch toxin-filled capsules to stun prey. In the freshwater polyp *Hydra*, orphans guide the development of feeding tentacles around the organism's mouth. And the polar cod's orphan antifreeze gene enables it to survive life in the icy Arctic.

However improbable...

Curiously, orphan genes are often expressed in the testes – and in the brain. Lately, some have even dared speculate that orphan genes have contributed to the evolution of the biggest innovation of all, the human brain. In 2011, Long and his colleagues identified 198 orphan genes in humans, chimpanzees and orang-utans that are expressed in the prefrontal cortex, the region of the brain associated with advanced cognitive abilities. Of these, 54 were specific to humans. In evolutionary terms, the genes are young, less than 25 million years old, and their arrival seems to coincide with the expansion of this brain area in primates. "It suggests that these new genes are correlated with the evolution of the brain," says Long.

Critics argue that most genes, new or old, are somehow involved with the workings of the brain, and that correlation does not prove causation. But Long cites a recent animal study that lends credence to the theory. Expressing one of the human orphan genes, *SRGAP2C*, in the neurons of developing mice doesn't make the animals' brains grow bigger. But it does encourage the nerve cells to grow denser arrays of dendritic spines, the tiny protrusions that enable neurons to connect with their neighbours. Having more connections might increase computing power, he argues. So these recently evolved human genes may have helped shape the human brain. "I think we've underestimated orphan genes," says geneticist Diethard Tautz of the Max-Planck Institute for Evolutionary Biology in Plön, Germany.

But where did they come from? In 2003, Tautz and a colleague suggested that orphan genes are formed by duplication, but then evolve so rapidly that any similarity to the original is obliterated. And they did have evidence that seemed to support this idea. They showed that orphans in fruit flies evolve three times more quickly than non-orphans.

Orphan genes were thus crowbarred into the old model of genes arising by duplication. Later studies, however, suggest this can only explain the origins of a minority of orphans. So while the process is clearly important, it's not the whole story. "The idea seemed reasonable at the time," says Tautz,

"because the alternative seemed so unlikely."

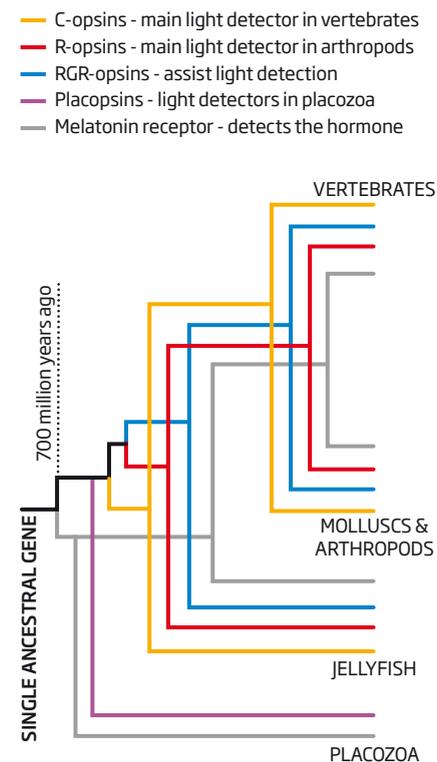
The alternative? The only other possibility was that genes really can evolve from scratch, from random stretches of non-coding DNA. This is the idea long dismissed as completely implausible, because the leap from non-coding DNA to a gene with a useful protein product was considered so huge as to be impossible. But nature hasn't read the textbooks. A few years ago, evidence began to emerge of genes created "de novo" in yeast, rice, mice and fruit flies. Then in 2009, David Knowles and Aoife McLysaght of the University of Dublin in Ireland showed that three orphan genes in people had indeed been created from scratch.

They found that DNA sequences nearly identical to the genes existed in several other primates but were non-coding, meaning the genes must have arisen sometime after the human-chimpanzee split. They also showed that the orphan genes are transcribed to RNA and then translated into protein in multiple tissues, though the functions of the genes are not known yet.

In 2011, another team described a further 60 human orphan genes created from scratch. McLysaght thinks this might be an overestimate – she believes that de novo gene synthesis is a rare phenomenon.

A family affair

The genes for light-detecting pigments in animals all evolved by duplication from a **single ancestral gene**. It was thought that almost all genes arose this way and thus belonged to families



"The idea seemed reasonable at the time because the alternative was so unlikely"

Some other researchers, however, are starting to think it may be surprisingly common. A study of 270 primate orphan genes, led by M. Mar Albà and Macarena Toll-Riera of the Municipal Foundation Institute for Medical Research in Barcelona, Spain, found that only a quarter could be explained by rapid evolution after duplication (*Molecular Biology and Evolution*, vol 26, p 603). Instead, around 60 per cent appeared to be new. "De novo evolution is clearly a strong force – constantly generating new genes over time," says Tautz. "It seems possible that most orphan genes have evolved through de novo evolution."

But how can it be possible? Knowles and McLysaght showed that the orphan genes they found sit next to and slightly overlap existing, older genes, so the orphans might be able to "borrow" their switches. Similarly, Albà and Toll-Riera found that half of the 270 primate orphans had acquired sequences from the genes of "transposable elements", genetic parasites that can jump around in the genome. Meanwhile, the ENCODE study of the human genome published earlier this year showed that our DNA is littered with millions of potentially useful short switching sequences, and that single switches can interact with many genes.

All this suggests that it is relatively easy for non-coding DNA to acquire the switches needed for RNA copies of it to be made. Indeed, the ENCODE study found that as much as



RICHARD HERRMANN/INDEN PICTURES/LEA

80 per cent of DNA is copied into RNA at least occasionally. Some argue that all this RNA is functional, but another interpretation is that most of this activity is just noise, and that junk DNA is routinely transcribed into RNA.

Proto-genes

If so, we are basically experimenting with thousands of potential new genes all the time – and Anne-Ruxandra Carvunis of the University of California, San Diego, has shown that this is indeed the case, at least in yeast. Last year, her team analysed 108,000 short, unknown but potentially protein-coding sequences in the yeast genome (*Nature*, vol 487, p 370). More than 1000 were interacting with the cell's protein-making factory, suggesting that they were being converted to proteins. "These may just be the tip of the iceberg," says Carvunis.

Her findings suggest that the protein-making factories in yeast are constantly churning out new proteins, allowing them to be "tested", and she suspects that same is happening in all complex organisms. In between non-coding DNA and fully fledged genes, Carvunis thinks there is a whole continuum of "proto-genes". Most will code for proteins that are neutral or harmful, so there will be no selection and the vast majority of proto-genes will revert to non-coding DNA sooner or later. But a few proto-genes that are neutral or maybe even helpful will sometimes persist, and start to gather

beneficial mutations. Over millions of years of natural selection, they can become a proper gene – and thus is an orphan born.

All this helps explain why orphan genes are often expressed in the testis. In most cells, DNA is tightly packaged, which reduces the chances of RNA copies being made. In certain immature sperm cells, however, the structure is more open, making it easier for proto-genes to be copied into RNA. Over time, the gene may come to be expressed in other tissues and evolve new functions.

New discoveries about the nature of proteins also make the idea of genes arising de novo seem far more plausible. It was once thought proteins must be folded into a delicate, precise 3D structure to work properly, but it now seems many exist in a state of intrinsic disorder, flitting through thousands of different possible conformations, all the while remaining perfectly functional. About half of human proteins have at least one long intrinsically disordered segment, while 10 per cent are disordered from beginning to end.

Peter Tompa of the Flanders Institute of Biotechnology in Brussels, who studies intrinsically disordered proteins, suspects that

"Between non-coding DNA and fully fledged genes there could be a continuum of 'proto-genes'"

Most of our genes have relatives in animals like jellyfish, but a surprising number are orphans

new orphan genes are likely to code for disordered proteins because they are easier to make than folded proteins. And disordered proteins often play a role in cell signalling and regulation. "I wouldn't be surprised if orphan genes turn out to have regulatory functions," says Tompa.

Perhaps this helps explain why orphan genes can become essential very quickly. In 2010, Long's team used RNA interference to switch off evolutionarily old and new genes in fruit flies. They found that new genes, including orphans, were just as likely to be essential for life as old genes (*Science*, vol 330, p 1682). "This goes against the textbooks, which say the genes encoding essential functions were created in ancient times," says Long.

We still have a lot to learn about orphan genes, but we are now starting to trace their ancestry. And it looks as if we couldn't find the families of most orphans because they don't really have families. The raw DNA from which they sprung can be traced, but as genes they are the first of their kind. In this sense, the term orphans may be a misnomer. Perhaps they ought to be renamed Pinnocchio genes – non-genes carved by chance and natural selection into proper, "living" genes. ■

Helen Pilcher is a freelance science writer based in the UK