ONCE, we could only marvel at the wonder of life. Like movie audiences not so long ago, we had little idea of what went on behind the scenes.

How times have changed. As the genomes of more and more species are sequenced, geneticists are piecing together an extraordinarily detailed "Making of..." documentary. Nowadays, we can not only trace how the bodies of animals have evolved, we can even identify the genetic mutations behind these changes.

Most intriguing of all, we can now see how genes - which are the recipes for making proteins, the building blocks of life - arise in the first place. And the story is not unfolding quite as expected.

The most obvious way for a new gene to evolve is through the gradual accumulation of small, beneficial mutations. Less obvious is how an existing gene that already does something important can evolve into a different gene. The scope for such a gene to change tack without capsizing the organism that carries it is very limited. However, as biologists realised a century ago, this constraint no longer applies when mutations produce an entire extra copy of a gene.

Trillions of copies

According to the textbooks, the process by which new genes form starts with gene duplication. In the vast majority of cases one of the copies will acquire harmful mutations and will be lost. Just occasionally, though, a mutation will allow a duplicate gene to do something novel. This copy will become specialised for its new role, while the original gene carries on performing the same task as before.

Surprisingly, gene duplication has turned out to be nearly as common as mutations that change a
single "letter" of DNA code. During the exchange of material between chromosomes prior to
sexual reproduction, mistakes can create extra copies of long DNA sequences containing anything
from one gene to hundreds. Entire chromosomes can be duplicated, as happens in Down's
syndrome, and sometimes even entire genomes.

Since duplication can throw up trillions of copies for evolution to work with, it is not surprising that
over hundreds of millions of years, a single original gene can give rise to many hundreds of new
ones. We humans have around 400 genes for smell receptors alone, all of which derive from just
two in a fish that lived around 450 million years ago.

Not the whole story

This classical view of gene evolution is far from the whole story, however. A decade ago, Michael
Lynch at Indiana University in Bloomington and a colleague outlined an alternative scenario.
Genes often have more than one function, and Lynch considered what might happen after such a
gene is duplicated. If a mutation knocks out one of the two functions in one of the copies, an
organism can cope fine because the other copy is still intact. Even if another mutation in this other
copy knocks out the second function, the organism can carry on as normal. Instead of having one
gene with two functions, the organism will now have two genes with one function each - a
mechanism Lynch dubbed subfunctionalism (Genetics, vol 154, p 459). This process can provide
the raw material for further evolution. "A gene preserved by subfunctionalisation can later pick up
a new function," Lynch says.

Some theoretical biologists think gene copies can also be preserved by other, more subtle,
mechanisms, but the real challenge to the classical model comes from actual studies of new
genes in various organisms. Earlier this year, in the most comprehensive study of its kind yet, a
team led by Wen Wang of Kunming Institute of Zoology in Yunnan, China, looked at several
closely related species of fruit fly. By comparing their genomes, Wang was able to identify new
genes that have evolved in the 13 million years or so since these species split from a common
ancestor.

One of Wang's surprise discoveries was that around 10 per cent of the new genes had arisen
through a process called retroposition. This occurs when messenger RNA copies of genes - the
blueprints sent to a cell's protein-making factories (see diagram) - are turned back into DNA that is
then inserted somewhere else in the genome. Many viruses and genetic parasites copy
themselves through retroposition, and the enzymes they produce sometimes accidentally retropose
the RNA of their host cells.

Dead on arrival?

The gene copies created by retroposition are not the same as the original, as genes consist of
more than just the sequence coding for a protein. There are also "promoter" regions in the front of
the coding part, to which other proteins bind, and this determines when and in which tissues the
gene is turned on. Since retroposed gene copies lose their promoters, which are not transcribed
into RNA, it used to be assumed that these partial copies were never expressed and gradually
disappeared as mutations accumulated. Retroposed gene copies were dismissed as "dead on
arrival", says Henrik Kaessmann of the University of Lausanne, Switzerland.

However, it has become apparent that a retroposed copy can sometimes get inserted in the
genome near an existing promoter, making it active. Crucially, though, with a different promoter, it
will be turned on at different times or in different tissues or both. In this way a retroposed gene
can immediately acquire a new function.
This process may have created many of the recently evolved genes in us apes. A burst of retroposition in our ancestors, peaking around 45 million years ago, gave rise to many thousands of gene duplicates, of which at least 60 or 70 evolved into new genes, according to a 2005 study led by Kaessmann. The burst was probably due to a new genetic parasite invading our genome.

**Brainy genes**

Kaessmann's team is now studying some of these genes in more detail. Their work suggests that at least two, called CDC14Bretro and GLUD2, could be related to apes' increased cognitive abilities.

The evolution of new genes often involves even more drastic changes. In his fruit fly survey, Wang found that a third of new genes were significantly different from their parent genes, having lost parts of their sequences or acquired new stretches of DNA.

Where do these extra sequences come from? In complex cells, the DNA coding for a protein is broken into several parts, separated by non-coding sequences. After an RNA copy of the entire gene is made, the non-coding bits - the introns - are cut out and the coding parts - called exons - are spliced together. This edited RNA copy is then sent to a cell's protein-making factory. The modular form of genes greatly increases the chances of mutations reshuffling existing genes and generating novel proteins. There are all sorts of ways in which it can happen: exons within a gene can be lost, duplicated or even combined with exons from different genes to create a new, chimeric gene.

**Variations on a theme**

For instance, most monkeys produce a protein called TRIM5, which protects them from infection by retroviruses. In one macaque in Asia around 10 million years ago, an inactive copy of a gene called CypA, produced by retroposition, was inserted near the TRIM5 gene. Further mutation resulted in cells producing a chimeric protein that was part TRIM5, part CypA. This protein provides better protection against some viruses. Although it might seem an unlikely series of events, in fact the TRIM5-CypA gene has evolved not once but twice - much the same thing happened in owl monkeys in South America.

Given enough time - or rather enough mutations - gene duplication and reshuffling can produce new genes that are very different from the ancestral ones. But are all new genes variations on a theme, or can evolution throw up new genes unlike any that already exist?

A couple of decades ago, it was suggested that unique genes could arise from what is called a frameshift mutation. Each amino acid in a protein is specified by three DNA "letters", or bases - the triplet codon. If a mutation shifts the starting point for reading codons - the "reading frame" - by one base, or by two, the resulting protein sequence will be completely different. Since DNA is double-stranded, any given piece can be "read" in six different ways.

**Gibberish**

The vast majority of mutations that alter the reading frame of a gene produce nonsense, usually dangerous nonsense. Many genetic diseases are the result of frameshift mutations wrecking proteins. It's a bit like swapping every letter for the next one along in the alphabet: the result is usually gibberish.

But not always. In 2006, Stephen Scherer of the University of Toronto in Canada and his
colleagues searched the human genome for new genes that had evolved by duplication followed by frameshift mutations affecting at least part of the original gene. They found 470 examples, suggesting that the process is surprisingly common (Genomics, vol 88, p 690).

Another source of unique new genes could be the "junk" DNA littering most genomes. An early hint this might be so came a decade ago when a team at the University of Illinois revealed the genesis of the antifreeze protein produced by one Antarctic fish. The gene involved originally coded for a digestive enzyme. Then, around 10 million years ago, as the world's climate cooled, part of one of the introns - a piece of junk DNA, in other words - got turned into an exon and subsequently duplicated many times, generating the characteristic repetitive structure of antifreeze proteins. From a random bit of DNA evolved a gene vital to the fish's survival.

From scratch

Still, the antifreeze gene evolved from a pre-existing gene. What are the chances of mutations in junk DNA generating an entire new gene from scratch? Practically zero, most biologists thought until very recently. As Lynch points out, it takes a whole set of unlikely conditions for a piece of random DNA to evolve into a gene. First, some of the DNA must act as a promoter, telling the cell to make RNA copies of the rest. Next, these RNA copies must have a sequence that can be edited into a viable messenger RNA blueprint for the protein-making factories. What's more, this messenger RNA must encode a relatively long protein - the average length is 300 amino acids - which is unlikely because in a random stretch of DNA, on average 1 in 20 every codons will be a "stop" codon. Finally, of course, the new protein must do something useful. The obstacles seemed insurmountable.

Then, in 2006, David Begun of the University of California, Davis, and colleagues identified several new genes in fruit flies with sequences unlike any of the older genes. They suggested that these genes, which code for relatively small proteins, have evolved from junk DNA in the past few million years. Begun quotes Sherlock Holmes: "When you have eliminated the impossible, whatever remains, however improbable, must be the truth."

Junk DNA

This year, during his hunt for new genes in fruit flies, Wang found another nine genes that appear to have evolved from scratch out of junk DNA. For eight of the nine, Wang has identified the non-coding sequences from which the genes evolved in related species, ruling out the possibility that these genes were somehow acquired ready-made from another organism.

Altogether, an astonishing 12 per cent of recently evolved genes in fruit flies appear to have evolved from scratch. And Wang suspects this rate is low compared with other animals. "My gut feeling is it may be higher in vertebrates because they have more junk DNA," he says.

It looks as if Wang could be right. A team at Trinity College Dublin in Ireland has found evidence that at least six new human genes have arisen from non-coding DNA in the 6 million years or so since humans and chimps diverged. The work is ongoing but the preliminary findings were presented at a meeting in Barcelona, Spain, in June. "We're very excited," says team leader Aoife McLysaght.

First hurdle

How can the number be so high when the likelihood of a gene evolving in this way is so rare? Part of the answer could be the recent discovery that even though at least half of our genome is junk, as much as 90 per cent of it can be accidentally transcribed into RNA on occasion. "The first
hurdle has already been overcome," McLysaght says.

This means it might not be that uncommon for random bits of junk DNA to get translated into a protein. Since most random proteins will probably be harmful, natural selection will eliminate these DNA sequences, but just occasionally one will strike it lucky, Begun says. A sequence that does something beneficial will spread through a population and rapidly evolve into a new gene, becoming optimised for whatever role it plays.

It might not be that uncommon for random bits of junk DNA to get translated into a protein

It will be many years yet before we have a clear picture of the relative importance of the various mechanisms by which genes can evolve. What is certain, though, is that the classical view of how they evolve is far from complete. Evolution isn't fussy - it'll take new genes wherever it can get them. "Natural selection is aggressively opportunistic," says Begun. "The source of the raw material is irrelevant."

And as sequence data continues to pour in, biologists are well on their way to working out how every one of our 20,000 or so genes evolved. Better stock up on the popcorn and make sure your sofa is comfy: this is going to be one epic "Making of..." documentary.

Who needs new genes?
To do new things or make new body parts, organisms don't necessarily need to evolve entire new genes. Identical proteins often take on different roles in different parts of the body, while a single gene can produce many proteins. Alternative splicing of RNAs - including some parts of a gene but not others (see diagram) - can generate a huge variety of different proteins. A study out this month found that alternative splicing is far more common than thought in people, with most genes producing at least two variants. One human gene, bn2, can generate more than 2000 different proteins, some of which have no similarity at all. The record is held by a fruit fly gene, Dscam, which can produce an astonishing 38,000 variants. That's not all. In September, a team at Yale University School of Medicine showed that the RNAs from two different human genes can be edited together to generate a new protein. Cells lining the uterus produce a protein by a fusion of the JAZF1 gene found on chromosome 7 and the JJAZ1 gene on chromosome 17. The JAZF1-JJAZ1 protein appears to promote cell growth. This phenomenon, known as trans-splicing, was known to occur in nematode worms but was thought to happen only by accident in vertebrates. The Yale team speculate that it could in fact be fairly common, greatly increasing the number of potential proteins (Science, vol 321, p 1357).

From issue 2683 of New Scientist magazine, page 44-47. Subscribe and get 4 free issues.
Browse past issues of New Scientist magazine

If you would like to reuse any content from New Scientist, either in print or online, please contact the syndication department first for permission. New Scientist does not own rights to photos, but there are a variety of licensing options available for use of articles and graphics we own the copyright to.

Back to article

http://www.newscientist.com/article/mg20026831.800-recipes-for-life-how-genes-evolve.html?full=true&print=true