

A brief history of the genome

Le Page, M. (2012), *New Scientist*, September 15, 2012 (2882), pp30-35.

From the first cells to the dawn of humans, **Michael Le Page** takes you on a whirlwind tour through 3 billion years of evolution

GTGCCAGCAGCCGCGGTAATTCCAGCTCCAATA GCGTATATTAAGTTGCTGCAGTTAAAAAG

It looks like gibberish, but this DNA sequence is truly remarkable. It is present in all the cells of your body, in your cat or dog, the fish on your plate, the bees and butterflies in your garden and in the bacteria in your gut. In fact, wherever you find life on Earth, from boiling hot vents deep under the sea to frozen bacteria in the clouds high above the planet, you find this sequence. You can even find it in some things that aren't technically alive, such as the giant viruses known as mimiviruses.

This sequence is so widespread because it evolved in the common ancestor of all life, and as it carries out a crucial process, it has barely changed ever since. Put another way, some of your DNA is an unimaginable 3 billion years old, passed down to you in an unbroken chain by your trillions of ancestors.

Other bits of your DNA are brand new. You have around 100 mutations in your genome that are not present in your mother or father, ranging from one or two-letter changes to the loss or gain of huge chunks of DNA.

We can tell which bits of our DNA are old or new by comparing genomes. Comparing yours with those of your brother or sister, for instance, would reveal brand new mutations. Contrasting the genomes of people and animals reveals much older changes.

Our genomes, then, are not just recipes for making people. They are living historical records. And because our genomes are so vast, consisting of more than 6 billion letters of DNA – enough to make a pile of books tens of metres high – they record our past in extraordinary detail. They allow us to trace our evolution from the dawn of life right up to the present.

While we have only just begun to decipher these records, we have already discovered that our ancestors didn't just face a harsh struggle for survival in a world red in tooth and claw. There were also epic battles going on in our genomes, battles that transformed the way our genome works and ultimately made us what we are today.

The universal ancestor

In the beginning there was RNA. This multitalented molecule can store information and catalyse reactions, which means some RNAs can replicate themselves. As soon as one RNA molecule, or set of molecules, began replicating itself, the first genome was born.

The downside of RNA is that it isn't particularly stable, so very early on life switched to storing information in a molecule with a slightly different chemical backbone that is less likely to break apart – DNA. Proteins also replaced RNA as catalysts, with RNA relegated to the role of a go-between. DNA stored the recipes for making proteins, sending out RNA copies of the recipes to the protein-making machinery.

Many traces of the ancient RNA-dominated world remain in our genome. The ubiquitous sequence at the beginning of this article, for instance, codes for part of an RNA enzyme that still plays a key role in the synthesis of proteins.

By around 3.5 billion years ago, a living entity had evolved with a genome that consisted of recipes for making RNAs and proteins – [the last universal common ancestor of all life](#). At least 100 genes can confidently be traced all the way back to LUCA, says [Eugene Koonin](#) of the National Institutes of Health in Bethesda, Maryland, who studies the evolution of life, and LUCA probably had more than 1000 genes in total.

LUCA had a lot of the core machinery still found in all life today, including that for making proteins. Yet it may have been quite unlike life as we know it today. Some researchers believe that LUCA wasn't a discrete, membrane-bound cell at all but rather a mixture of virus-like elements replicating inside [some non-living compartment](#), such as the pores of alkaline hydrothermal vents.

Split and reunion

One possible scenario for the next stage is that subsets of LUCA's virus-like elements broke away on two separate occasions, acquiring cell membranes and becoming simple cells. This would explain why there are two kinds of simple cell – bacteria and archaea – each with a completely different cell membrane. “It's a very appealing hypothesis,” Koonin says. What is certain is that life split into two major branches very early on.

Bacteria and archaea evolved some amazing molecular machinery and transformed the planet, but they remained little more than tiny bags of chemicals. It wasn't until an extraordinary event reunited the two great branches of life that complex cells, or eukaryotes, emerged – an event that transformed the genome and paved the way for the evolution of the first animals.

Around a billion years ago, a bacterium ended up inside an archaeon. Instead of one killing the other, the two forged a symbiotic relationship, with the descendants of the bacterium gradually evolving to take on a crucial role: they became mitochondria, the power factories inside cells that provide our energy.

Without this union, complex life might never have evolved at all. We tend to assume that it is natural for simple organisms to evolve into more complex ones, but individual bacteria and archaea have never evolved beyond a certain level of complexity. Why?

According to Nick Lane of University College London, it's because they hit an energy barrier. All simple organisms generate energy using their cell membranes. As they get bigger, the ratio of surface area to volume falls, making it harder to produce enough energy. The upshot is that simple cells have to stay small – and small cells don't have room for big genomes. Mitochondria eliminated this barrier by providing modular, self-contained power sources. Cells could now get bigger simply by producing more mitochondria, allowing them to expand their genomes and so their information-storing capacity.

Besides freeing cells from this energy constraint, the ancestor of mitochondria was also the source of up to three-quarters of our genes. The original bacterium probably had 3000 or so genes, and over time most were either lost or transferred to the main genome, leaving modern mitochondria with just a handful of genes.

Despite the obvious benefits, the forging of this alliance was fraught with peril. In particular, the genome of the ancestral mitochondrion was [infested with pieces of parasitic DNA](#), or transposons, that did nothing except create copies of themselves. They sometimes landed in the middle of genes, leaving them with big chunks of irrelevant DNA known as introns. It's the equivalent of sticking a recipe for soup into the middle of a cake recipe.

Yet the result was not always a recipe for disaster, because these introns were “self-splicing”: after an RNA copy of a gene was made – the first step of the protein-making process – they cut themselves out. This didn't always happen, though, so their presence was a disadvantage. Most bacteria have no introns in their genes, because in large populations with a lot of competition between individuals, natural selection is strong and weeds them out. But the population of the ancestral eukaryote was very small, so selection was weak. The genetic parasites that arrived with the ancestor of the mitochondrion began to replicate like crazy, littering the main genome with hundreds of introns.

Today, each of our genes typically contains about eight introns, many of which date back to the very first eukaryotes – our ancestors never did manage to get rid of most of them. Instead, they evolved ways of dealing with them that altered the structure of our genes and the way that cells reproduce. One was sex.

The benefits of sex

The crucial thing about sex is not just the mingling of genes from different individuals, important as this is for bringing together evolutionary advances made in separate lineages. Simple cells had long been swapping genes without bothering with sex.

It's also a process known as recombination, in which pairs of chromosomes swap corresponding pieces before being divided into sperm or eggs. Recombination helps solve a fundamental problem with having a genome consisting of many genes linked together like beads on a necklace.

Imagine a necklace with a truly magnificent pearl right next to a flawed one. If you can't swap one pearl for another, you either have to get rid of the whole thing or take the necklace as it is. Similarly, if a beneficial mutation ends up next to a harmful one, either the beneficial mutation will be lost or the harmful mutation will spread through a population, dragged along by its neighbour.

Recombination gives you the opportunity to swap pearls. Just as you can produce one perfect necklace and one with defects, so some offspring will get a disproportionate number of good genes, while others get lots of bad ones, perhaps with disruptive introns. The unlucky individuals are likely to die out while those with the good genes thrive.

In large populations, so many mutations arise that some will counteract the effects of the harmful genes, so there is no need to resort to recombination. But in a small population, sex wins out. This is why it became the norm for the first eukaryotes and thus for most of their descendants. So next time you make love, remember to thank the genetic parasite harboured by your ancient bacterial ancestor for the joy of sex.

You can thank the genetic parasite harboured by our ancient bacterial ancestor for the joy of sex

By the time sex had evolved, there were too many introns to get rid of them all. So early eukaryotes soon faced another serious problem: as introns acquired more and more mutations, the self-splicing mechanisms began to fail. In response, these early eukaryotes evolved special machines, called spliceosomes, that could cut out the introns from the RNA copies of genes.

Spliceosomes are the kind of mindless solution typical of evolution: cutting the junk out of the RNA copies of genes, rather than out of the original DNA, is very inefficient. What's more, spliceosomes are slow. Many RNAs would have reached the protein-making factories before their introns were spliced out, leading to defective proteins.

This is [why the nucleus evolved](#), Koonin has proposed. Once a cell's DNA was enclosed in a compartment separate from the protein-making machinery, only spliced RNAs could be allowed out, preventing cells from wasting energy by producing useless proteins.

Even this didn't solve all the problems, though. Spliceosomes often cut out coding sections of genes – known as exons – by mistake, resulting in mutant versions of the proteins. “Alternative splicing was not an adaptation,” says Koonin. “It was something that organisms had to deal with.”

So our ancient ancestors evolved layer upon layer of complex machinery to cope with the proliferation of introns, yet still hadn't solved all the problems they caused. But unlike simple cells, they could afford this wastefulness because they were flush with energy – and in the long run all this extra complexity led to new opportunities.

Versatility and control

The presence of introns, and thus exons, in effect made genes modular. In an uninterrupted gene, mutations that add or remove sections usually change the way the rest of the gene is read, producing gibberish. Exons, by contrast, can be moved around without disrupting the rest of the gene. Genes could now evolve by shuffling exons within and between them.

Suppose, for instance, that random mutations add an extra exon to a gene. Thanks to alternative splicing, the original version of the protein can still be made, but it also means a new protein can come from the same gene (see “The cutting room”). The mutation might have little effect and so wouldn't be eliminated by selection, but over time, the new protein might take on a new function. Quite by accident, eukaryotes' mindless efforts to deal with introns had made their genes more versatile and more evolvable.

If this view of the evolution of complex cells is correct, many of the key features of our genome, from modular genes to sex, evolved as a direct result of the acquisition of parasite-bearing mitochondria. Alternative ideas cannot be ruled out, but none provides such a beautiful explanation. “It's my favourite scenario,” says Koonin.

All these novel features led to a burst of evolutionary innovation, and eukaryotes thrived and soon began to diversify. Even so, they still faced a relentless onslaught from the invasion of new kinds of parasitic DNA and viruses. Having transcended the size constraints on simple cells, however, complex cells were free to evolve more sophisticated defence mechanisms.

One was to “silence” the transposons' parasitic genes by adding tags to the DNA that stop RNA copies being made – a process called methylation. Another was to destroy the RNAs of invading

viruses to stop them replicating themselves. These defences were only partly successful. Today, around 5 per cent of the human genome consists of the mutated and mostly inert remains of viruses, and an astonishing 50 per cent consists of the remnants of transposons, a testament to the many occasions on which these parasites somehow got into the genomes of our ancestors and ran rampant.

Such defence mechanisms were soon co-opted for another purpose: to control the activity of a cell's own genes. "Mechanisms for controlling transposons became mechanisms for controlling genes," says [Ryan Gregory](#) of the University of Guelph, Canada, who studies the evolution of genomes.

Building bodies

The stage was now set for the next big step in evolution, roughly 800 million years ago, when cells began to cooperate more closely than ever before. Although a few bacteria are multicellular, the constraints on their complexity have never allowed them to go far down this road. Eukaryotes, by contrast, have evolved multicellularity on dozens of occasions, giving rise to hugely complex organisms such as fungi, seaweeds, land plants and, of course, animals.

One reason was their bigger repertoire of genes, which could be co-opted for new purposes such as binding cells together and communicating with other cells. Even more importantly, the modular nature of their genes allowed more rapid evolution. The proteins that join cells together, for instance, consist of a part that straddles the cell membrane and a part that protrudes outwards. With [modular genes](#), all kinds of different protruding bits can be tacked onto to the membrane-straddling part, like different attachments on a vacuum cleaner. Many crucial genes for multicellularity evolved via exon shuffling.

In addition, eukaryotes' more sophisticated mechanisms for controlling genes could be used to allow cells to specialise. By switching different sets of genes on or off, different groups of cells could take on distinct roles. As a result, organisms could begin to develop different types of tissue, allowing early animals to evolve from simple sponge-like creatures to animals with increasingly sophisticated bodies.

The next great leap forward was the result of a couple of genetic accidents. When things go wrong during reproduction, the entire genome can occasionally be duplicated – and this happened not once but twice in the ancestor of all vertebrates.

These genome duplications produced lots of extra copies of genes. Many were lost but others took on new roles. In particular, the duplications produced four clusters of the master genes that establish body plans during development – the *Hox* genes – and these clusters are thought to have played a crucial role in the evolution of an internal skeleton.

Whole-genome duplications are rare, and most new genes arise from smaller duplications, or from exon shuffling, or both. Evolution is shameless – it will exploit any DNA that does something useful regardless of where it comes from. Some crucial genes have evolved from bits of junk DNA, whereas others have been acquired from elsewhere.

About 500 million years ago, for instance, the genome of our ancestors was invaded by a genetic parasite called a hAT transposon, which copies itself using a "cut and paste" mechanism. The cutting is done by two enzymes that bind to specific DNA sequences.

At some point in an early vertebrate, the sequences bound to by the DNA-cutting enzymes ended up near or in a gene involved in recognising invading bacteria and viruses. The result was that during the course of an individual's life, as their cells multiplied, the hAT enzymes cut bits out of the gene. Crucially, different bits got cut out in different cell lines, generating lots of mutant versions of the protein.

In some cases, this turned out to be a lifesaver, because the mutant proteins were better at latching onto invading pathogens. Soon a mechanism evolved for recognising the cells producing the most effective versions and encouraging them to multiply – the adaptive immune system. The human immune system is now mind-bogglingly complex, but the two enzymes that cut up and rearrange genes – the crucial process that allows it to target invaders – are [direct descendants of the hAT enzymes](#). So we have an ancient parasite to thank for our most effective weapon against disease.

The human genome

Armed with these advanced defences, and with a genetic toolkit that could be tweaked to produce a huge variety of body shapes, early vertebrates were extremely successful. They conquered the seas, colonised the land, took to the trees and then came back down and started walking on two legs.

What made us so different from other apes? There is one apparently big difference between us: [we have 23 chromosomes](#) rather than the 24 of our ape ancestors. But chromosomes are essentially bags of genes: it makes little difference if they split apart or fuse together as long as we still have the genes that we need. Rather, it seems a long series of smaller changes gradually altered our brains and bodies. We've identified a few key mutations already ([New Scientist, 9 June, p 34](#)), but there may be many thousands involved.

Looking back at the bigger picture, it is clear that increases in the complexity of cells and bodies began with increases in the complexity of genomes. What is striking, though, is that many of the initial increases in complexity were due to a lack of evolutionary selection, rather than being driven by it. “Most of what's going on at the genomic level is probably neutral,” says Gregory.

In other words, mutations arise that have little if any effect, such as a duplicate gene. In a large population, such mutations would soon be lost. But in a tiny population, they can spread by chance, through genetic drift. “This is an inevitable consequence of population genetics,” says Koonin. It is only later that such complexity is selected for, such as when a duplicate gene acquires a new role.

Many key events in our history, such as the genome duplications that produced our *Hox* genes, may be a result of relaxed selection in a tiny population. Indeed, a population bottleneck right at the beginning of human evolution might explain the spread of some of the mutations that make us so different to other apes, such as our loss of muscle strength.

The other striking thing is that viruses and parasites have played a huge role. Many of the main features of our genome, from sex to methylation, evolved in response to their attacks. What's more, a fair number of our genes and exons, like the immune enzymes, derive directly from these attackers. “Viruses have been necessary parties to cellular life from the very beginning,” says Koonin.

Necessary but not pleasant. Our evolution has come at a tremendous cost. They say history is written by the victors – well, our genome is a record of victories, of the experiments that succeeded or least didn't kill our ancestors. We are the descendants of a long line of lottery winners, a lottery in which the prize was producing offspring that survived long enough to reproduce themselves. Along the way, there were uncountable failures, with trillions of animals dying often horrible deaths.

They say history is written by the victors. Our genome is a record of victories, of successful experiments

Our genome is far from a perfectly honed, finished product. Rather, it has been crudely patched together from the detritus of genetic accidents and the remains of ancient parasites. It is the product of the kind of crazy, uncontrolled experimentation that would be rejected out of hand by any ethics board. And this process continues to this day – go to any hospital and you'll probably find children dying of horrible genetic diseases. But not as many are dying as would have happened in the past. Thanks to methods such as embryo screening, we are starting to take control of the evolution of the human genome. A new era is dawning.

glossary

Archaeon – one of two kinds of simple organism

Bacterium – one of two kinds of simple organism

Eukaryote – a complex cell with intricate internal structures

Exon – one of the parts of a gene that codes for a protein

Gene – a recipe for making a protein or functional RNA

Intron – a part of a gene that does not code for a protein. Introns are usually cut out of a gene's RNA copy before it reaches the protein-making factory

LUCA – last universal common ancestor

Splicing – the process of removing introns from RNA

Transposon – a genetic parasite. Contains code for enzymes that allow it to copy and paste itself into other parts of the genome

Links to relevant studies can be found in the online version of this article

Michael Le Page is biology features editor at *New Scientist*.