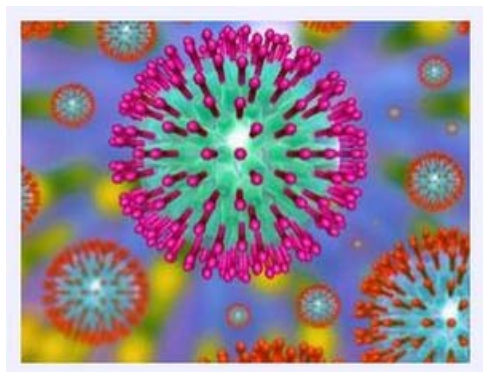


I, Virus: Why You're Only Half human

Frank Ryan
New Scientist
Fri, 29 Jan 2010 15:00 CST



© NewScientist

Part of our DNA (Image: Mehau Kulyk/SPL/Getty)

When, in 2001, the human genome was sequenced for the first time, we were confronted by several surprises. One was the sheer lack of genes: where we had anticipated perhaps 100,000 there were actually as few as 20,000. A bigger surprise came from analysis of the genetic sequences, which revealed that these genes made up a mere 1.5 per cent of the genome. This is dwarfed by DNA deriving

from viruses, which amounts to roughly 9 per cent.

On top of that, huge chunks of the genome are made up of mysterious virus-like entities called *retrotransposons*, pieces of selfish DNA that appear to serve no function other than to make copies of themselves. These account for no less than 34 per cent of our genome.

All in all, the virus-like components of the human genome amount to almost half of our DNA. This would once have been dismissed as mere "junk DNA", but we now know that some of it plays a critical role in our biology. As to the origins and function of the rest, we simply do not know.

The human genome therefore presents us with a paradox. How does this viral DNA come to be there? What role has it played in our evolution, and what is it doing to our physiology? To answer these questions we need to deconstruct the origins of the human genome - a story more fantastic than anything we previously imagined, with viruses playing a bigger part than you might care to believe.

Around 15 years ago, when I was researching my book *Virus X*, I came to the conclusion there was more to viruses than meets the eye. Viruses are often associated with plagues - epidemics accompanied by great mortality, such as smallpox, flu and AIDS. I proposed that plague viruses also interact with their hosts in a more subtle way, through symbiosis, with important implications for the evolution of their hosts. Today we have growing evidence that this is true ([New](#)

[Scientist, 30 August 2008, p 38](#)), and overwhelming evidence that viruses have significantly changed human evolution.

Symbiosis was defined by botanist Anton de Bary in 1878 as the living together of dissimilar organisms. The partners are known as symbionts and the sum of the partnership as the holobiont. Types of symbiotic relationships include parasitism, where one partner benefits at the expense of the other, commensalism, where one partner profits without harming the other, and mutualism, in which both partners benefit.

Symbiotic relationships have evolutionary implications for the holobiont. Although selection still operates on the symbionts at an individual level since they reproduce independently, it also operates at partnership level. This is most clearly seen in the pollination mutualisms involving hummingbirds and flowers, where the structure of flower and bill have co-evolved to accommodate each other and make a perfect fit. When symbiosis results in such evolutionary change it is known as symbiogenesis.

Viruses as partners

Symbiosis works at many different levels of biological organisation. At one end of the spectrum is the simple exchange of metabolites. Mycorrhizal partnerships between plant roots and fungi, which supply the plant with minerals and the fungus with sugars, are a good example. At the other end are behavioural symbioses typified by cleaning stations where marine predators line up to have their mouths cleared of parasites and debris by fish and shrimps.

Symbiosis can also operate at the genetic level, with partners sharing genes. A good example is the solar-powered sea slug [Elysia chlorotica](#), which extracts chloroplasts from the alga it eats and transfers them to cells in its gut where they supply the slug with nutrients. The slug's genome also contains genes transferred from the alga, without which the chloroplasts could not function. The slug genome can therefore be seen as a holobiont of slug genes and algal genes.

This concept of genetic symbiosis is crucial to answering our question about the origin of the human genome, because it also applies to viruses and their hosts. Viruses are obligate parasites. They can only reproduce within the cells of their host, so their life cycle involves forming an intimate partnership. Thus, according to de Bary's definition, virus-host interactions are symbiotic.

For many viruses, such as influenza, this relationship is parasitic and temporary. But some cause persistent infections, with the virus never leaving the host. Such a long-term association changes the nature of the symbiosis, making the evolution of mutualism likely. This process often follows a recognisable progression I have termed "aggressive symbiosis".

An example of aggressive symbiosis is the *myxomatosis* epidemic in rabbits in Australia in the 1950s. The European rabbit was introduced into Australia in 1859 as a source of food. Lacking natural predators, the population exploded, leading to widespread destruction of agricultural grassland. In 1950, rabbits infected with *myxoma* virus were deliberately released into the wild. Within three months, 99.8 per cent of the rabbits of south-east Australia were dead.

Although the myxomatosis epidemic was not planned as an evolutionary experiment, it had evolutionary consequences. The myxoma virus's natural host is the Brazilian rabbit, in which it is a persistent partner causing no more than minor skin blemishes. The same is now true of rabbits in Australia. Over the course of the epidemic the virus selected for rabbits with a minority genetic variant capable

of surviving infection. Plague culling was followed by co-evolution, and today rabbit and virus coexist in a largely non-pathogenic mutualism.

Now imagine a plague virus attacking an early human population in Africa. The epidemic would have followed a similar trajectory, with plague culling followed by a period in which survivors and virus co-evolved. There is evidence that this happened repeatedly during our evolution, though when, and through what infectious agents, is unknown ([*Proceedings of the National Academy of Sciences*, vol 99, p 11748](#)).

Even today viral diseases are changing the course of human evolution. Although the plague culling effect is mitigated by medical intervention in the AIDS pandemic, we nevertheless observe selection pressure on humans and virus alike. For example, the human gene HLA-B plays an important role in the response to HIV-1 infection, and different variants are strongly associated with the rate of AIDS progression. It is therefore likely that different HLA-B alleles impose selection pressure on HIV-1, while HLA-B gene frequencies in the population are likely to be influenced by HIV ([*Nature*, vol 432, p 769](#)). This is symbiogenesis in action.

How does that move us closer to understanding the composition of the human genome? HIV-1 is a retrovirus, a class of RNA virus that converts its RNA genome into DNA before implanting it into host chromosomes. This process, known as endogenisation, converts an infectious virus into a non-infectious endogenous retrovirus (ERV). In humans, ERVs are called HERVs.

Germline invaders

Endogenisation allows retroviruses to take genetic symbiosis to a new level. Usually it is an extension of the normal infectious process, when a retrovirus infects a blood cell, such as a lymphocyte. But if the virus happens to get incorporated in a chromosome in the host's germ line (sperm or egg), it can become part of the genome of future generations.

Such germ-line endogenisation has happened repeatedly in our own lineage - it is the source of all that viral DNA in our genome. The human genome contains thousands of HERVs from between 30 and 50 different families, believed to be the legacy of epidemics throughout our evolutionary history. We might pause to consider that we are the descendents of the survivors of a harrowing, if brutally creative, series of viral epidemics.

Endogenisation is happening right now in a retroviral epidemic that is spreading

among koalas in Australia. The retrovirus, KoRv, appeared about 100 years ago and has already spread through 75 per cent of the koala's range, culling animals on a large scale and simultaneously invading the germ line of the survivors.

Retroviruses don't have a monopoly on endogenisation. Earlier this month researchers reported finding genes from a bornavirus in the genomes of several mammals, including humans, the first time a virus not in the retrovirus class has been identified in an animal genome. The virus appears to have entered the germ line of a mammalian ancestor around 40 million years ago ([*Nature*, vol 463, p 84](#)). Many more such discoveries are anticipated, perhaps explaining the origin of some of that mysterious half of the genome.

The ability of viruses to unite, genome-to-genome, with their hosts has clear evolutionary significance. For the host, it means new material for evolution. If a virus happens to introduce a useful gene, natural selection will act on it and, like a beneficial new mutation, it may spread through the population.

Could a viral gene really be useful to a mammal? Don't bet against it. Retroviruses have undergone a long co-evolutionary relationship with their hosts, during which they have evolved the ability to manipulate host defences for their own ends. So we might expect the genes of viruses infecting humans to be compatible with human biology.

This is also true of their regulatory DNA. A virus integrating itself into the germ line brings not just its own genes, but also regulatory regions that control those genes. Viral genomes are bookended by regions known as long terminal repeats (LTRs), which contain an array of sequences capable of controlling not just viral genes but host ones as well. Many LTRs contain attachment sites for host hormones, for example, which probably evolved to allow the virus to manipulate host defences.

Retroviruses will often endogenise repeatedly throughout the host genome, leading to a gradual accumulation of anything up to 1000 ERVs. Each integration offers the potential of symbiogenetic evolution.

Once an ERV is established in the genome, natural selection will act on it, weeding out viral genes or regulatory sequences that impair survival of the host, ignoring those that have no effect, and positively selecting the rare ones that enhance survival.

Most ERV integrations will be negative or have no effect. The human genome is littered with the decayed remnants of such integrations, often reduced to fragments, or even solitary LTRs. This may explain the origin of retrotransposons. These come in two types: long and short interspersed repetitive elements (LINEs and SINEs), and it now appears likely that they are heavily degraded fragments of ancient viruses.

As for positive selection, this can be readily confirmed by looking for viral genes or regulatory sequences that have been conserved and become an integral part of the human genome. We now know of many such sequences.

The first to be discovered is the remnant of a retrovirus that invaded the primate genome a little less than 40 million years ago and gave rise to what is known as the W family of ERVs. The human genome has roughly 650 such integrations. One of these, on chromosome 7, contains a gene called syncytin-1, which codes for a protein originally used in the virus's envelope but now critical to the functioning of the human placenta. Expression of syncytin-1 is controlled by two LTRs, one derived from the original virus and another from a different retrovirus called MaLR. Thus we have a quintessential viral genetic unit fulfilling a vitally

important role in human biology.

Virus genes

There are many more examples. Another gene producing a protein vital to the construction of the placenta, syncytin-2, is also derived from a virus, and at least six other viral genes contribute to normal placental function, although their precise roles are poorly understood.

There is also tentative evidence that HERVs play a significant role in embryonic development. The developing human embryo expresses genes and control sequences from two classes of HERV in large amounts, though their functions are not known ([Virology, vol 297, p 220](#)). What is more, disrupting the action of LINE retrotransposons by administration of the drug nevirapine causes an irreversible arrest in development in mouse embryos, suggesting that LINEs are somehow critical to early development in mammals ([Systems Biology in Reproductive Medicine, vol 54, p 11](#)).

It also appears that HERVs play important roles in normal cellular physiology. Analysis of gene expression in the brain suggests that many different families of HERV participate in normal brain function. Syncytin-1 and syncytin-2, for example, are extensively expressed in the adult brain, though their functions there have yet to be explored.

Other research groups have found that 25 per cent of human regulatory sequences contain viral elements, prompting suggestions that HERVs make a major contribution to gene regulation ([Trends in Genetics, vol 19, p 68](#)). In support of that, HERV LTRs have been shown to be involved in the transcription of important proteins. For example, the beta-globin gene, which codes for one of the protein components of haemoglobin, is partly under the control of an LTR derived from a retrovirus.

The answer to our paradox is now clear: the human genome has evolved as a holobiontic union of vertebrate and virus. It is hardly surprising that researchers who have made these discoveries are now calling for a full-scale project to assess the contribution of viruses to our biology ([BMC Genomics, vol 9, p 354](#)).

It is also probable that this "virolution" is continuing today. HIV belongs to a group of retroviruses called the lentiviruses. Until recently virologists thought that lentiviruses did not endogenise, but now we know that they have entered the germ lines of rabbits and the grey mouse lemur. That suggests that HIV-1 might have the potential to enter the human germ line ([Proceedings of the National Academy of Sciences, vol 104, p 6261](#) and [vol 105, p 20362](#)), perhaps taking our evolution in new and unexpected directions. It's a plague to us - but it could be vital to the biology of our descendants.

Frank Ryan is a writer, medical doctor and biologist based in Sheffield, UK. His book [Virolution](#) is published by HarperCollins. He is the author of a series of five review articles on the impact of viral symbiosis on medical genetics, published in the Journal of the Royal Society of Medicine (vol 102, [p 272](#), [p 324](#), [p 415](#), [p 474](#) and [p 530](#)),