

Endogenous viral sequences and their evolution

John M. Coffin

Tufts University, Boston MA

Throughout evolution, all living things have had to face the challenges of infectious agents – viruses, bacteria, parasites – and invent ways to deal with them. To survive, infectious agents must ways to counter these defenses, resulting in an eternal arms race. A full understanding of this arms race will help us to better understand the microbial enemy as well as provide important clues to development of new strategies for prevention and therapy of infection. With most infectious diseases, obtaining this understanding is hampered by the complete absence of a fossil record from which to construct an evolutionary history. Retroviruses, however, are a major exception to this general rule. Unlike all other entities infecting animals, retroviruses have the ability (and need) to integrate their genetic information, in the form of a DNA copy called the provirus, into the DNA of the host cell at more or less random locations. An important consequence of this unique property is that occasional infection of the germ line leads to proviruses – called endogenous proviruses –, which are then inherited as part of the genetic composition of all descendants of the animal in which the integration occurred. Endogenous proviruses have been found in the DNA of all animals where they have been sought, including mammals and all other vertebrates, insects, mollusks, and many others. In humans, it has been estimated that there are about 80,000 endogenous proviruses, comprising about 8% of our total genetic makeup. In some species, including mice, chickens, cats, koalas, and others, endogenous viruses are still active and continue to be inserted into the genome, and can be important causes of disease. In humans, no active proviruses are known, their role in disease is uncertain, and they may all be extinct. In all cases, however, they provide an invaluable fossil record of the evolution of this large and important group of infectious pathogens.

Because integration is at nearly random sites, and our genome is so large, if two proviruses are found at exactly the same location, they must be descendants of the same original integration event in a common ancestor of that species. Almost all proviruses in human DNA are found in the same place in chimpanzee DNA, meaning that they must be at least 5 million years old. Indeed, many are much older, occupying the same site in humans and new world monkeys, implying that the virus that gave rise to them must have existed at least 45 million years ago, and is long extinct as an infectious entity. Remarkably, even these very old fossils look very much like modern day viruses, meaning that all important events in their evolution took place in the very distant past. With the exceptions mentioned below, we know this for no other infectious agent, and many evolutionary biologists believe (incorrectly, I think) others to be much younger, because of the very rapid rate of evolution that some of them display over short time intervals.

By revealing important biologic properties of the ancestors of modern viruses, endogenous viruses can illuminate important events in the host-virus arms race. For example, many endogenous proviruses of mice can infect all mammals, except mice. The explanation for this apparent paradox is that the presence of a pathogenic virus in the genome strongly selected for a mutation in the mouse gene coding for the receptor – the cell surface protein that mediates viral entry, leaving the virus in the DNA no longer able to infect and spread in that species. In response, variants of the virus evolved that are now able to use the mutant form of the protein for entry. Remarkably, a virus called XMRV, which must have been derived from one of these endogenous mouse viruses, has recently been found in humans and may be associated with a variety of diseases, including prostate cancer and chronic fatigue syndrome.

A few endogenous proviruses have also been co-opted to do something useful for their hosts. For example, in humans, the formation of the syncytiotrophoblast, a fused layer of cells that forms the surface of the placenta and prevents the passage of potentially harmful substances from mother to fetus, is mediated by the action of a gene found in an endogenous provirus. This gene had the original function of mediating virus entry by fusing virus and cell membranes together, quite similar to the function it now serves in human development. Other endogenous proviruses can help protect their host from infection by similar viruses from outside by blocking access to specific receptors.

Until recently, retroviruses were thought to be the only entities to have generated this kind of a fossil record. Recently, however, a few sequences derived from genes of a group of viruses called Bornaviruses have been found in the DNA of humans and a few other species. As with retroviruses, these sequences demonstrate that this group of viruses is far older than evolutionary biologists have thought, and they also may provide some important function to their host, although we don't yet have any idea what that function may be.