Can Viruses Make Us Human?[1]

LUIS P. VILLARREAL
Director, Center for Virus Research
University of California at Irvine

THIS QUESTION WILL SEEM preposterous to most. Viruses are molecular genetic parasites and are mostly recognized for their ability to induce disease in their host. Their effect on host evolution has long been thought to be like that of a predator on its prey, eliminating the host with weakened defenses. How can we propose any constructive role for viruses? Many viruses, however, can infect their host in a stable and persisting manner, generally with no disease, often for the life of the host. Such viruses can bring to bear onto their host the viral seeds of genetic creation. For such persisting viruses to successfully colonize their host, they must superimpose a complex viral molecular genetic identity onto their host. This essay will develop and present the argument that such stable persisting viruses represent a major creative force in the evolution of the host, driving the host to acquire new, and accumulate ever more complex, molecular identities. Based on this premise, this essay will examine the possible role of viruses in the evolution of complexity, including the evolution of human-specific attributes. This view of human evolution is part of a larger idea, that stable persisting viruses (genetic parasites) can allow the host to acquire complicated functions (complex phenotype) in one punctuated event of colonization. Such a process can now be considered as a possible explanation for several major dilemmas in evolutionary biology. All these dilemmas involve the origin of various host lineages that have acquired a complex and interacting set of functions in a relatively short time frame. Such acquisitions of complexity have always been difficult to explain by a simple Darwinian process. These dilemmas include the origin of the eukaryotic nucleus, the origin of flowering plants, the origin of the adaptive immune system in animals, and the origin of live birth (viviparous) placental mammals. In this essay I will only briefly consider the origin of the eukaryotic nucleus as

[1] Read 15 November 2003.

an example of how persisting viruses can contribute to the evolution of complex host. I direct the attention of readers interested in these other issues to my book *Viruses and the Evolution of Life*, soon to be published by the American Society for Microbiology.

How Do Humans Differ From Their Closest Relatives? Retroposons

With the completion of the human genome project and evaluation, we can now evaluate global differences between our human genome and that of our closest relative, the chimpanzees. It is expected that such an evaluation should identify the evolutionary process that allowed human and chimpanzee genomes to evolve and diverge from a common ancestor, about five million years ago. The genetic changes that have occurred in these two lineages should allow us to identify at a genetic level what makes us human. One of the most surprising observations made with the comparison of the human and chimpanzee genomes was how similar these two genomes were to each other. The similarity is so great (98.5% in coding regions) that it would be difficult to distinguish human from chimpanzee genes only via their coding regions. Recently, it has been reported that those genes that do differ often correspond to those involved in smell or diet. Thus, at the level of genes, it is not at all clear how or why humans and chimpanzees diverged from each other. Yet our human genomes do show significant differences from those of our chimpanzee relatives. The most obvious of these differences, however, is not in the coding regions, but can be found in their corresponding Y (and X) chromosomes. The human Y chromosome differs substantially (in size and sequence) from that of the chimpanzee. But the Y chromosome encodes few genes and most of this difference corresponds to "retroposon" (DNA that moved via the action of reverse transcriptase, a retroviral replication enzyme) or other repeat-derived DNA that has colonized the Y chromosome. Because much of this "retroposon" DNA is non-coding, it has been considered a "junk" or "selfish" DNA that has no phenotype or consequence to host evolution, but is simply able to accumulate. Yet it is precisely such genetic material that not only distinguishes human DNA from chimpanzee DNA, but also distinguishes the genomes of all higher life forms from each other. In fact, a genome-wide comparison of the human mouse and marsupial mammalian genomes establishes that these non-genetic sequences not only are more distinguishing amongst these organisms, but, paradoxically, are also more conserved than are the coding sequences. This observation appears to be contradictory. How can non-coding sequences as a group be both more distinguishing, and more conserved, than coding sequences? All mammals have their own unique collections of such sequences. Yet all mammals have conserved old versions of such sequences. Thus, such sequences appear to have been acquired during species diversification. Yet all mammals also appear to conserve both the more recently acquired and the older non-coding sequences. Why is this, and how can we understand this seemingly paradoxical behavior?

Retroposons Are Retrovirus-Derived

It is now clear that the majority of genomic non-coding DNA mostly represents variations of, defectives of, and/or products of endogenous or genomic retroviruses. An endogenous retrovirus (ERV) is a retrovirus whose DNA is found embedded in the genome of its host. These ERV sequences are present in substantial numbers in vertebrates. However, even more abundant are derivatives of ERVs. These defective and derivative ERVs have been given names, such as LINES and SINES (long and short interspersed sequences), but are clearly related to retroviruses. For example, human LINES retain some sequences of the HERV pol gene, whereas human SINES retain some sequences of HERV LTR and env regions. These sequence elements have no obvious function and are generally incapable of expressing genes. Because of this, they have often been called both junk and selfish DNA. Can such apparently “derelict” viral-derived DNA have any role in what makes us human? The instinctive response of most evolutionary biologists would be negative, and no compelling experimental evidence would currently appear able to refute this instinct. Yet relevant observations do exist that can support the view that such viral-derived sequences play a big role in evolution. In this essay I will develop the theory that such retroposons represent products of past colonization events by persisting genetic parasites that endowed the host with major creative acquisitions in the evolution of life. From this view, we will examine evidence for viral footprints in the genomic human record that may be clues to how genetic parasites may have led to the evolution of some very human-specific characteristics, such as associative (social-based) learning and the acquisition of the cognitive capacity for human language.

The Two Life Strategies of Viruses: Acute and Persistent

The main thesis of this essay will seem highly counter-intuitive to most readers. How can genetic parasites, such as viruses, contribute to the evolution of complex characteristics of their host? We know very well that viruses are destructive, not constructive, entities, able to cause mass epidemics and kill humans and other host in very large numbers.

Our collective perception of viruses as agents of disease is thus well justified and is measured in large numbers of human deaths. In the last century alone millions of humans have succumbed to the ravages of pandemic influenza virus, measles virus, poliovirus, and now HIV. In prior centuries, especially in the New World, entire cultures were destroyed by smallpox virus epidemics. How can we think that such agents might contribute in any positive way to their host evolution? Due to the risk they pose, we know much about the dynamics of such epidemic viruses and their effects on host populations. These agents of acute disease can reproduce in huge numbers and have been measured to evolve at rates up to one million times that of their host. Viruses such as HIV and influenza A can evolve so quickly, evolutionary genetic changes are observed during the short duration of individual infections. Viruses clearly represent the leading edge of all evolving biological entities. However, viruses have more than one strategy for living. Not all viruses are rapidly replicating, highly evolving agents of disease. We know, from our mathematical models, for example, that acute, disease-causing viruses did not likely co-evolve with their human host. Pre-agricultural human populations were simply too small to sustain the rates of transmission and host recovery for any of these acute infectious diseases. The relationship these viruses have with their host is not stable in small populations or in an evolutionary time scale that could date back to the evolution of their host. Yet viruses are highly related to their host. Even on the broadest scale, viruses show strong and distinct relationships to their specific host. The viruses of bacteria, archaea, algae, fungi, aquatic invertebrates, insects, plants, amphibians, and mammals all have distinct patterns and relationships with their host. For example, in bacteria, the great majority of viruses are large dsDNA viruses. Similar dsDNA viruses are also found in algae, but absent from fungi and plants. Instead, fungi harbor dsRNA viruses, whereas flowering plants mostly harbor ss+RNA viruses. In contrast, mammals show a strong tendency to infection with endogenous retroviruses. Even at the species level, some viruses will often be highly specific to their host. There is also a curious link with respect to hosts and viral diversity. Hosts that are species-diverse tend also to support diverse viruses, whereas hosts that are not diverse, but successful, tend to support few viruses. However, as will be developed below, viruses that are highly species-specific tend to be persistent viruses.

Persistence: Genetic Conduit to Host Genomes

It is here proposed that the stable persistence of genetic parasites in their host can channel the tremendous genetic creativity of a viral entity into the evolution of its host. Persistent parasites can thus contribute, in a cumulative and punctuated way, to the evolution of host complexity. Persistence is thus crucial to the creative process of host evolution. As the term is herein used, persistence can be defined as the capacity of a virus or genetic parasite to maintain a continued presence in its host such that it allows transmission of the genetic parasite to subsequent generations of the host. Viral persistence can be thought of as a distinct viral life strategy for those viruses that replicate mainly by an acute infection. Acute viruses are characterized by being able to generate tremendous genetic complexity, and produce large numbers of progeny. These viruses are responsible for most of the disease-based epidemics that are familiar to us and do not show co-speciation with their host. To maintain their epidemic or endemic replication cycle, these viruses are highly dependent on the population structure and size of their host. They require large communicating host populations. This feature makes such viruses less stable during evolution, which can often involve bottlenecks of host populations. Persistence, on the other hand, is much less dependent on host population structure and can work well in non-gregarious host populations.

Persistence, Fitness, and "King of the Hill" Phenotype

Before further developing the case for the role of persisting genetic parasites in host evolution, it is important to better understand how the concept of fitness relates to persistence. In this essay, I define persistence as an infection that results in the capacity of an individually infected host to maintain the viral genome or reproduce that genome in an episodic way. Although we can accept the general idea that fitness drives evolution, the actual measurement of fitness in a laboratory or field setting poses some major problems. Laboratory measurements will usually use *relative fitness* or *reproductive value* as a lab-suitable measurement. Relative fitness is the number of reproductively successful progeny virus divided by the average number of progeny for the population. Reproductive value is measured as the net reproductive output (R_0), which is the viable offspring per individual per lifetime. Such fitness measurements, applied to acute viral agents, result in differential equations that appear to accurately describe most viral epidemics. Thus, these fitness measurements of acute viruses have been well supported by experimental observations. However, these fitness measurements are dimensionless metrics solely dependent on reproduction, not time. They do not account for differences in survival time or persistence. In a more general and comprehensive sense, we can see a problem with the above definition. If we instead define fitness as the genetic contribution to an individual's *continued life*, and the *continued* *life* of its decedents, we have a definition that is more similar to that originally used by Darwin. Such a definition would encompass fitness resulting from high rates of reproduction as well as long life. In the context of a persisting genetic parasite, we can reason that the parasite must increase the survival time of the virus or host to allow the attainment of a high probability of viral continuation or transmission. This does not mean that persistence must maximize reproductive rates. Persistence will actually inhibit reproductive rates during establishment and maintenance of the persistent state. For persistence to attain a high probability of maintenance or transmission, the persisting agent must be able to compete with other parasitic agents in a more static sense that does not depend on replication rates. A persistent parasite must "hunker down" and take on all comers (resulting from both host immunity and competition) that might seek to displace or destroy the persisting virus. This is the familiar "King of the Hill" childhood game strategy in which fitness is defined by who is left standing at the end, not how many attempts were made by the competition to displace the winner. For this, persistence must sense and respond to competitors, the environment, and the time for maintenance and maximized transmission. Thus, persisting genetic parasites need to incorporate a phenotype or strategy that assures maintenance. This feature is the crucial difference between the concept of persistence and the concept of selfish genes, which had been previously proposed to explain defective genetic parasites (transposons). A selfish gene simply seeks its own maintenance. Selfish DNA has no phenotype associated with it and hence no direct consequence to host competition or evolution. A persisting genetic parasite, in contrast, must superimpose onto its host a new molecular genetic identity that compels persistence and precludes competition and displacement. This competition may come from the very same genetic parasite as the persisting agent itself. Hence, as mentioned, persisting genetic parasites will generally be inhibitory to their own replication, thus allowing the establishment of the persistent state. A "defective" genetic parasite (missing genes, unable to replicate by itself) is an effective mediator of persistence and will often be able to superimpose a state of persistence onto the fully infectious parasite. Thus, defectives represent a common strategy to attain persistence, while resembling selfish DNA.

Persistence Requires Host Intimacy, Co-evolution

The above paragraph argues that persistence has a distinct fitness relationship with its host compared with acute viral agents. In terms of host evolution, persisting viruses are much more intimately associated with their host species and have generally co-evolved with them. Persistence is often stable in an evolutionary time scale, and numerous instances exist in which persisting viruses and their host appear to be co-evolving. Persisting viruses are not dependent on host population structures, and are effectively maintained in small populations. They can be ubiquitous, in some cases found in 100 percent of the specific natural host population. Both large and small host populations can be effectively colonized. Persistence is often transmitted from old to young, often during sex or birth. Persistent infections are usually lifelong and generally show little if any disease in colonized host. These infections are typically highly species-specific. In terms of human viruses, many persisting viruses are commonly known. These tend especially to include various types of DNA viruses, such as herpesvirus type I/II, Epstein Bar virus, cytomegalovirus, various types of adenovirus, human polyomaviruses (JCV, BKV), human papillomaviruses, and the small TT virus. All of these viruses tend to establish lifelong inapparent infections. Furthermore, they are all highly human-specific (in some cases distributed in congruence with human racial and geographical patterns). All of them also show some degree of co-evolution with human and other primate host.

Persistence, Immunity, and Addiction Modules

We can now ask, in general terms, what is the consequence to the evolution of its host, when an organism has been stably colonized by a persisting virus? What difference will this make to the evolutionary potential or trajectory of this host? One clear consequence is that these colonized host must control the replication of the persisting virus itself to establish persistence. Generally, the virus will have some immunity or control functions that will prevent virus replication. Thus, persistence must provide some form of immunity against similar viruses. In addition, stable persisting viruses will often express genes or functional strategies that insure the maintenance of the viral genome. In lower organisms, these maintenance functions can often be in the form of addiction modules. Addiction modules are matched sets of genes or functions that are harmful for host that lose the persisting agent, but beneficial for host that maintain the persisting agent. Typically, the harmful function is stable but the beneficial function may be unstable, requiring the constant presence of the viral genome. Often the harmful function is a toxin and the beneficial function is an antitoxin. Sometimes the harm and benefit can come from the replication of the virus itself, which will kill uncolonized host. A defective virus that prevents replication of the infectious virus can serve such a function if acute viruses are prevalent. Such viral-based immunity modules are inherently complex genetic systems. They involve matched sets of functions that are both harmful and beneficial to the colonized host, but typically also affect the ability of the host to compete with or preclude other viral or genetic parasites. Can such persisting agents provide the origin of host immunity systems?

Persistence, Addiction, and the Origin of Prokaryotic Immunity

We can consider this question from the context of bacteria and archaea. These prokaryotes are the most adaptable organisms on earth. How do they protect themselves from infectious agents and how did this system originate? All free-living bacteria and archaea use restriction-modification enzymes as a system of immunity, as well as other systems. That their genomes all avoid palindromic sequences supports the general evolutionary importance of the restriction-modification systems. These restriction-modification systems represent the most diverse of all genes in prokaryotes. They also represent the simplest version of a "complex phenotype," requiring the simultaneous creation of two matched enzyme functions. The restriction enzyme, which can degrade unmodified DNA, is stable, but the modification enzyme normally acts only during DNA replication. The likely origin of these systems is from viruses and parasitic plasmids. For example, persisting phage such as P1 and P7 both code for restriction-modification enzymes that are part of an addiction module (*phd/doc*). These enzymes compel the colonized host to maintain the virus in order to prevent host self-destruction. However, as a result of this virus colonization, the host cell is now immune to a whole array of other viruses (including the lytic T4-like phage). If this parasite acquisition becomes stable, the host has acquired a new complex phenotype of immunity in one infectious colonization event. This argument can be extended to include numerous other complex gene sets (such as fitness or pathogenic islands whose genes number in the hundreds) that also result from stable host colonization. In prokaryotes, it is now clear that host evolution is occurring mainly by an infectious process. However, many would feel that this process is "horizontal" in nature—that is, involving the transfer of genes from one host species into another using viruses as a vehicle. I disagree with that view. As presented below, viruses are not merely vehicles that transmit genes from one host to another. Instead, I argue that viruses represent the ultimate genetic creators, inventing new genes in large numbers, some of which find their way into host lineages following stable viral colonization.

Host Lineages Are Marked by Their Acquired Genetic Parasites

We are now ready to consider a global dilemma in evolutionary biology—why do lineages of host tend to evolve to higher genetic complexity and why is this evolution associated with the acquisition of non-coding "derelict" DNA? These non-coding sequences mainly stem from various types of parasitic genetic elements, such as retroposons Why is there a linkage between the evolution of genetic complexity and the acquisition of parasitic elements? Bacteria have gene counts in the order of 2,000 to 3,000 and genomes of about 4 million base pairs (BP). Very little of their genomes corresponds to parasitic genetic elements. With yeast, gene counts are about 6,000 and genomes about 1.3 billion bp with about ten times more parasitic DNA (mostly DNA transposons such as Ty). In Drosophila, about 8,800 genes exist in a 1.4×10^8 BP genome, which contains about 10% of its genome as parasitic elements. In mammals, such as humans, about 40,000 genes exist in a genome of about 2.9×10^9 BP, but 97% of this DNA is non-coding. Some of the non-coding elements (LINES and SINES) are present in more than 100,000 copies. Endogenous retroviral sequences are also found in these genomes in numbers well in excess of the total gene count. How can we account for this pattern? What forces lead to host genome colonization such as this?

Was a DNA Virus the Last Universal Ancestor?

The specific question we would like to evaluate along this line of thinking relates to the possible viral origin of the eukaryotic nucleus. Could a stable persisting virus have been the origin of the eukaryotic nucleus? On the face of it, this seems like a highly implausible proposition. Given the above gene counts and genome sizes of free living cells, it seems clear that the net genetic capacity of the first eukaryote should be well in excess of even the largest known DNA virus (about 800 genes). At best, a large virus would seem to correspond to about one tenth the needed genetic capacity. However, more detailed analysis indicates that this viral gene count is more than adequate for the earliest likely eukaryote. Comparison of the genes found in common to all three domains of life (bacteria, archaea, eucharia) indicates that only a surprisingly small set of genes has been conserved across these domains. It is assumed that there likely existed a progenitor cell, which was the Last Universal Common Ancestor (LUCA) to all life. However, only 324 genes are sufficiently conserved to be considered as having been all derived from LUCA. Surprisingly, the DNA replication genes, so crucial for genome identity and maintenance, are not part of this conserved set! Thus, in terms of gene number, a DNA virus could clearly have provided this quantity of genes. For example, T4 phage has 274 genes, CMV has 220 genes. Other viruses, such as DNA viruses of microalgae described below, have about twice this number. Viruses can do the job.

Horizontal Transfer: False Accusations and Viral Creativity

There is another point that merits emphasis at this time: that point has to do with the idea of horizontal gene transfer mentioned above. As we evaluate the gene count of various large DNA viruses, we can see that they have many genes that are unique. For example, WSSV of shrimp has 184 genes, of which only 11 have similarity to any genes in the Genbank database, and these are mostly related to DNA replication proteins. Other DNA viruses, such as AFV1, a chronic linear DNA phage of hyperthermophilic archaea, have almost no genes, including replication genes, that are similar to any in GenBank. Various viral lineages have an abundance of genes that are unique to the specific DNA viral lineages. The implication is that viral lineages frequently evolve by creating entirely unique genes. In the case of the entire baculovirus family (DNA viruses of insects), the whole family tree has been evaluated, and it is clear that most viral lineages differ by having acquired new and unique genes. Some 80% of their genomes are in fact unique. Some gene loss and gene conversion are also seen, but compare only 12 gene losses to 255 gene acquisitions, in this tree. In some cases, unique viral genes are highly complicated, interacting with a very large set of cellular proteins, such as the large T-Ag of SV40. Yet there is no host analogue for T-Ag, a gene that is highly conserved in the viral lineage. Frequently, the viral version of a gene represents the simplest example of any related protein in a functional protein family. For example, the potassium transport genes found in algal DNA viruses are only about a hundred amino acids compared with the much larger host versions. The conclusion is clear. DNA viruses invent genes, in large numbers, both complex and simple genes. Yet ironically, whenever it is observed that some viral genes show similarity to host genes, it is usually argued that the viruses must have “stolen” the gene from the host genome. Frequently such “stolen” genes are rather simple (less than 100 a.a.), such as immunoregulators like cytokines and chemokines. I submit that such views are misplaced. Viruses have been given a “bum rap,” accused of not contributing genes, but mainly deriving them from host and delivering them to another host. The evidence supports the opposite view. Even in most of the cases in which similarity can be seen between host and viral genes, a properly conducted phylogenetic analysis will usually show that the viral version is basal to the version found in the host. The viral version appears to be older, often simpler. Thus I argue that the concept of frequent horizontal transfer of genes by viruses is generally flawed and vastly overstated. It presumes that viruses must not be originating genes, but simply moving them between host. I argue that it is much more likely that when host genes show similarity to viral genes, it is the viral gene that is ancestral to the host gene.

The Accepted View, a Prokaryote Ancestor of the Nucleus, is a Dilemma

We are now ready to consider the evidence that a virus was the ancestor of the eukaryotic nucleus. Our focus will be to consider the origin of the proteins involved in the replication of eukaryotic DNA, since, as mentioned above, these proteins are not universally conserved or conserved between prokaryotes and eukaryotes. However, before delving into this consideration, we should review the more accepted views for the origin of the nucleus as well as the major dilemmas these views pose. The most accepted current view for the origin of the eukaryotic nucleus is that it represents the descendant organelle of a symbiotic prokaryote that stably colonized the predecessor to the eukaryotic cell (a mycoplasma-like cell lacking a cell wall). This idea was first proposed in 1905 by Mereschowsky. In the 1970s, Margulis convincingly argued that symbiosis with prokaryotes could explain the origins of both mitochondria and chloroplast organelles; thus the concept of symbiotic evolution gained much support. However, as has been noted by several authors (Pool and Penny 2001, Smith and Szathmary 1999), the theory that a symbiotic bacteria may have been the predecessor to the nucleus solves few dilemmas. The differences between prokaryotic chromosomes and replication systems and those of eukaryotes are simply too great to be explained by such symbiosis. Prokaryotes have circular chromosomes, with unique origins of replication. Their chromosomes are only loosely associated with chromatin proteins and they have distinct origin control and structures for initiating and completing DNA replication. All the proteins involved in these processes are distinct between prokaryotes and eukaryotes. Prokaryotes have mixed transcription and translation systems and the transcription of their genes uses distinct RNA polymerase enzymes. On top of that, eukaryotes have many nuclear features that are not found in any prokaryote. These include the use of linear chromosomes with repeat ends and multiple origins of replication. Chromosomes that are tightly associated with stoichiometricly bound basic proteins, the separation of transcription from translation by multiple membranes, the processing of RNA by 5′ capping, splicing, and 3′ polyadenylation, the existence of complex nuclear pore structures that actively transport RNA, and the existence of a tubulin-based system for the segregation of duplicated chromosomes—all of these features represent examples of complex phenotypes, involving the coordination of numerous protein functions. Yet none of them can be identified to exist in a prokaryotic cell that might have been the predecessor to the nucleus. Thus we can better understand why the issue concerning the origin of the eukaryotic nucleus represents the largest dilemma in all of evolutionary biology.

Viral and Eukaryotic Replication Proteins Are Similar

The observation that replication proteins from some bacterial viruses are more similar to eukaryote proteins than those found in prokaryotic cells is not recent. The most studied DNA polymerase in all of molecular biology was that of the phage T4, the very first characterized DNA polymerase in the 1950s. When T4 was first sequenced in the late 1980s, it was observed that this phage polymerase was, curiously, much more like eukaryotic DNA polymerase than it was like any prokaryotic DNA polymerase. The similarities included the existence of six functional domains of the protein as well as the sensitivity of the polymerase to various inhibitors (aphidicolin, PAA, etc.). The overall functionality of the prokaryotic cellular DNA replication apparatus is essentially the same as the eukaryotic replication apparatus. Yet none of the proteins are conserved between these orders, nor can any sequence similarity be identified as was done with the T4 DNA polymerase. The implication of this result was that the eukaryotic DNA polymerase and T4 DNA polymerase shared a common ancestor. Yet the T4 phage represent a very large family of phage that are known to infect both bacteria and archaea that may predate the divergence of these host orders. This phage-like DNA polymerase can thus be found represented in all three domains of life. Several years ago, Victor DeFilippis (then a graduate student) and I became interested in this issue. We wanted to better evaluate the idea that a DNA virus might have been the origin of the eukaryotic replication system (and the nucleus). Based on prior reasoning, we felt that a large DNA virus, able to persist, would be the best candidate for such a proto-nucleus. Possible protonuclear viruses would likely be found in both prokaryotes and early eukaryotes. In terms of viral prokaryotic candidates, the cyanophage

CPS-1, CPS-2, S-PM2 were all attractive since they have dsDNA genomes that encode DNA and RNA polymerases. Phage of archaea, such as SIRV1, TTV1,2,3,4 were also attractive since they are chronic non-lytic infections that use chromatin-bound dsDNA that have internal membranes in the virions. Of special interest was the AFV1 phage of hyperthermophiles, which in addition uses TATA (eukaryote-like) promoters and has linear genomes with eukaryotic-like telomeres. Also noteworthy, the AFV1 DNA polymerase appears to be the only known DNA polymerase that is basal to the DNA polymerases found in chlorellaviruses, African swine fever virus, and the poxviruses (Prangishvili, personal communication). Other interesting viruses of prokaryotes include P1 (N15) of bacteria and various P1-like phage that encode addiction modules and persistently infect spores of B. subtillis. However, for our analysis, we chose a well-studied large DNA virus known to infect micro algae, chlorella species virus (CSV-1).

Algae and its Virus: An Early Eukaryote

The evolution of algae marks a major transition in the evolution of higher life forms. Algae, as found in the oceans, represents the first eukaryote that can clearly be documented in the fossil records. Thus we reasoned that unlike the above viruses of prokaryotes, whose decedents might not have contributed directly to the evolution of the nucleus, viruses that infect micro algae must have adapted to the great changes that occurred during the evolution of eukaryotes and thus might retain features of the putative proto-nuclear virus. These viruses represent a family known as phycodnaviruses. A related family of viruses, called the phaeoviruses, is known to persistently infect the filamentous brown algae. In our analysis, we began by determining, in the entire genetic database, all amino acid sequences that showed significant similarity to the DNA polymerase found in CSV-1. This analysis identified large sets of related sequences, all of which appeared to be DNA polymerases of the B family. The sets included the replicative (extension) polymerases of all higher eukaryotes and all the large DNA virus families of eukaryotes, as well as the primer polymerase of eukaryotes, the repair polymerase of archaebacteria and bacteria, and some phage polymerase. The DNA polymerases that appeared to be the most similar to this CSV-1 algal viral polymerase were the replicative (extension) polymerases found in various fungi (yeast). The next most similar were the polymerases of the herpesvirus family. We then aligned all of these sequences to determine the regions that were most highly similar or conserved. As had been observed in prior studies of DNA polymerase, we identified four distinct domains that were most conserved. When we aligned all the sequences to the most highly conserved of these domains (as an anchor), we could see that the different sequences maintained the relative positions of these domains, but varied considerably in their overall length. However, the CSV-1 polymerase represented the simplest version of all these polymerases. This alignment allowed us to eliminate those sequences that were highly variable to the specific polymerase (a major source of noise to our analysis) and restrict our subsequent analysis to the more conserved regions. By doing this, we were able to construct a genetic tree of all the DNA polymerases, which was very well supported by statistical analysis.

The Algal Virus Has the Basal Eukaryotic DNA Polymerase and Other Genes

The genetic tree that was derived from the virus of algae showed some very interesting features. The most crucial is that it supported the idea that the DNA polymerase of this virus appears to be basal to all the replicative DNA polymerases found in eukaryotes. In other words, this viral polymerase appears to be the best candidate to represent the ancestor to all the polymerases known to replicate the genomes of eukaryotes. There were no other viral or prokaryotic cellular DNA polymerases that showed this characteristic in being at the base of the host branch. The DNA polymerases of the other DNA viruses (herpes, pox, baculovirus) formed their own branches and had no host genes within these sets. Another point was that the link between these prokaryote and eukaryote genes was via DNA viruses. This result added strong support to the idea that viruses could have provided the origin of the eukaryotic replication proteins.

We then became interested in expanding this analysis, which had been based solely on the DNA polymerase. It turns out that the CSV-1 also codes for several other genes that are not only characteristic of eukaryotic replication systems, but also characteristic of other eukaryotic-specific functions. CSV-1 codes for two versions of PCNA (proliferating cell nuclear antigen), a eukaryotic-specific replication-fork protein. A similar evaluation of one of these genes also shows that it is basal to all the PCNA genes of eukaryotes, as well as being basal to the PCNA gene found in eukaryotic organisms. In this case, related genes from various archaea bacteria were also observed to share similarity, consistent with reports by others that the archaea share greater similarity with eukaryotes in some of their replication proteins. This result further supports the idea that the chlorella viruses may represent basal versions of eukaryotic replication proteins. A similar analysis of the CSV-1 superoxide dismutase (SOD, a eukaryotic gene involved in protection against oxygen radicals), also showed that this viral gene was basal to most of the similar genes in eukaryotes. In this analysis, it was interesting that many baculoviruses (insect DNA viruses) also had a similar SOD gene. Of particular interest, however, was the observation that no prokaryotic version of the SOD gene was identified. In fact, the only prokaryotic representative of prokaryotes that was observed was the Fels-1 gene of a bacterial virus, which is also a component of phage immunity. Thus, not only was the CSV-1 version of SOD basal to all those found in eukaryotes, but the only likely prokaryotic ancestral version of the gene was also found in a virus. These results add further support to the idea that a virus can be ancestral to host gene function.

Others Have Also Proposed a Viral Nuclear Origin

My colleagues and I are not the only investigators who have proposed that viruses might be contributing to the evolution of host nuclear components. In 2002, P. Bell performed an analysis of the eukaryotic-specific guanylyltransferase enzyme (the enzyme that adds the eukaryotic-specific cap to mRNA). This enzyme is absent for any prokaryote. However, consistent with our observations, P. Bell reported that the enzyme encoded by CSV-1 was basal to all those found in eukaryotes. Another enzyme found in CSV-1 has a similar characteristic. The CSV-1 version of HAS (hyaluronan synthase), which modifies the surfaces of eukaryotic cells, was basal to all three versions of this enzyme found in higher (vertebrate) eukaryotes, but absent from lower eukaryotes. In this case, prokaryotic versions of this enzyme were also identified, but these prokaryotic versions were less similar to those versions found in vertebrates. Ironically, this study was a part of the landmark paper on the completion of the human genome project, and the authors used this example to examine the possibility that bacterial genes might have moved directly into the human genome, bypassing early eukaryotes. The irony is that although they clearly identified the CSV-1 version as basal, they did not consider any interpretation of this result, let alone that the viral version could be the ancestor. Another investigator who has been very active in the study of the possible origin of eukaryotic replication proteins is P. Forterre in France. In a series of studies, he has concluded that the eukaryotic replication proteins have resulted from gene displacement. These replication proteins do not for the most part originate from prokaryotic cells, but viruses clearly appear to have contributed. Thus there is a substantial body of literature that supports the concept that viruses have contributed to the origin of basic eukaryotic-specific functions.

Possible Viral Origin of Other Nuclear Functions

What about the other aspects of the eukaryotic nucleus—can viruses be identified that might have led to these additional highly complex functions? Can viruses allow us to resolve the many dilemmas, mentioned above, concerning the origin of these complex nuclear functions? A strong case can, in fact, be made that all these dilemmas have potential virus-based solutions. For example, the multi-membrane bound separation of transcription from replication is a characteristic found in poxviruses, such as vaccinia virus, as well as other DNA viruses (ASFV, TTV1 of thermophiles). Also, these same viral systems have simple pore structures that actively transport the RNA from the membrane-bound core into the host cytoplasm. Similarly, the tight association of the DNA genome with small basic chromatin proteins and linear chromosomes with repeat telomer ends is a characteristic of various cytoplasmic DNA viruses as well as TTV1 and phycodnaviruses. In addition, a viral version of DNA-dependent RNA polymerase (from ASFV) is known to be phylogenetically basal to all three versions of the DNA-dependent RNA polymerase found specifically in eukaryotes, but absent from any prokaryote. As mentioned above, the enzymes that modify mRNA in a eukaryotic-specific way (5′-capping, splicing, and 3′-polyadenylation) can all be found in various types of DNA viruses, and these viral versions are generally simpler than and basal to those found in eukaryotes. Even the situation with respect to the complex role of tubulin in the process of separation of eukaryotic daughter chromosomes and the dissolution and reformation of the nuclear membrane has a viral analogue as cytoplasmic DNA viruses perform all these same functions. We can thus propose that most all of the characteristics of a eukaryotic nucleus could have been derived from a stable, persisting, membrane-bound large DNA virus with linear chromosomes and viral-specific replication and transcription proteins. This proto-nuclear virus must have stably colonized a prokaryotic host that lacked cell walls. Thus, according to this scenario, there is no cellular ancestor to the eukaryotic nucleus, nor is there a prokaryotic source for many of the essential eukaryotic proteins, since they originate from viral sources. Such a viral-origin proposal would also allow us to resolve another major problem in evolutionary biology: the issue of the Last Universal Common Ancestor (LUCA), which should be common to all three existing domains of cellular life. There does not currently exist a cellular life form that has conserved the genetic characteristics common to all three domains of life: bacteria, archaea, and eucharia. This problem, however, is resolved if we propose that there never was such an ancestral cell, but rather that the common linkage is derived from viral sources. Thus there neither is, nor ever was, a LUCA. Therefore, we can now consider the main question of this essay. If viruses can contribute to the evolution of the most complex molecular structures in cells, can they also have contributed other types of complex characteristics, such as those found in mammals or specific to humans?

The concept that viruses might play a fundamental role in the evolution of the complexity of cellular life, as here proposed, may seem novel to many, especially to evolutionary biologists. However, even this idea is not entirely new. The modern definition of a virus, as a molecular genetic parasite, was first clearly put forward by S. E. Luria in an essay published in *Science* in 1950. Later in that decade, in his book *Virus Growth and Variation*, when considering the role viruses might play in cellular evolution, he wrote, ". . . may we not feel that [in] the virus, in their merging with the cellular genome and re-emerging from them, we observe the units and process which in the course of evolution, have created the successful genetic patterns that underlie all living cells?" This view, however, did not gain a following, and has been completely overlooked or forgotten by most evolutionary biologists. Too often we could observe only the destructive nature of viruses and think of them as non-living toxic entities that do not contribute to the web of life. Seldom did we see any constructive role for the viruses. Only with the ability to sequence entire genomes and only after understanding how persistence of genetic parasites drives host evolution may we now finally come to accept the view that viruses are indeed the origin of many of the complex patterns we see in their hosts.

With this viro-centric view, we can examine anew how higher life forms have evolved and diverged from the perspective of the patterns of acquisition of genetic parasites. With this view we can now also try to understand how humanity has diverged from its closest relative, the chimpanzee. If we limit our re-examination to the evolution of mammals, we do indeed see many species-specific patterns of acquisition of genetic parasites. The first mammals are rather old, about 210×10^6 years before the present time (YBP). These early mammals (such as multituberculates) were small, furry, egg-laying, monotreme-like insect predators. They were present prior to the advent of the dinosaurs, but became extinct about 35×10^6 YBP. We know nothing about their genomes, but two other mammalian lines developed from them, the marsupials and the eutharian placentals. Of these, the placental mammals are much more successful, radiating to about 2,000 genera, relative to about 140 marsupial genera. Thus the invention of the trophectoderm, the placenta, and live birth are all associated with the success of the placental mammals. With these placental mammals, we do indeed see clear and compelling patterns of acquisition of genetic parasitic elements, mainly in the context of LINES, SINES, and ERVS.

Mammals Have Acquired Their Own Characteristic Endogenous Retroviruses and Retroposons

As we have said, all of these elements (related to MLV retrovirus) were present in the genomes of early vertebrates. However, only with the evolution of the placental mammals do we see a large-scale and lineage-specific expansion of these elements. In addition, each placental lineage has its own peculiar version of these elements, which is distinguished from that found in other placental species. For example, mice have lAPs ERV elements. Rats and hamsters are clearly related to mice, but have their own specific version of lAP elements. Ungulates have JSAV sequences in their genomes. Felines have RDl14 and pigs PERV elements. All simians have HERVIP elements, but humans in particular have specific versions of HERVK10. For reasons we have not until now been able to give, each mammalian lineage has its own species-specific version of these ERVs, which is not shared with other species. We can now ask why there is a linkage between the advent of a specific mammalian lineage and the colonization of that lineage with specific versions of ERVs. It appears that at the origin of each placental lineage, a high-level colonization by genetic parasites occurred and that the resulting lineage stably maintains these mostly inactive parasitic ERV genomes.

The placental lineage-specific pattern of parasitic colonization is even more extensive than might be apparent from a simple examination of elements that are clearly related to endogenous retroviruses (ERVs). This is because these genomes all have much higher levels of elements known as LINES, SINES, and retro-transcripts such as the human Alu element. For example, in the human genome, Line-1 is present at about 10,000 copies. Sine (R) is even more abundant, present at about 100,000 copies per cell, whereas Alu elements (derived from a steroid receptor) are present at nearly 1,000,000 copies. Yet Line-1 is clearly derived from the pol sequence of HERV K, whereas the Sine (R) is derived from the LTR and env sequence of HERV K. Thus, all of these highly numerous elements show some relationship to the human-specific HERV K family. A related pattern of genetic similarity to ERVs applies to the Lines and Sines found in other placental lineages (i.e., mouse Line and lAPs). Why are the placental genomes so highly colonized by both intact and derivative elements of peculiar ERVs? What is the relationship between this colonization and the evolution of the placental life style?

One way to approach the question of the possible ERV role in mammals is to determine if these parasitic elements are ever expressed (transcribed or translated) in any specific manner in the placental host.

Evaluating the specific expression of ERVs, however, gets into a problem of nomenclature. For historic reasons, the ERVs present in a particular host genome have often acquired an array of names, many of which are confusing. These early names were based on criteria such as the morphology of the virus particle, the tissue tropisms, and the sequence similarity of various genes (pol, env, LTR). However, in some cases, it was clear that some ERVs appeared to be mosaic elements when evaluated by these criteria. This led researchers to focus on the intact ERV elements found in particular genomes as well as use the t-RNA primer sequence employed by that element to prime reverse transcriptase. Hence the name HERV -K (K lysine tRNA) for the common human-specific element. With this nomenclature, a whole alphabet (E, F, H, I, K, L, R, W) of HERVs could be distinguished and evaluated for expression. It will surprise many to learn that most of these HERVs are indeed expressed as transcripts and some also as proteins and sometimes as virion particles in specific tissues, although they seldom, if ever, make infectious virus. Curiously, reproductive tissue is by far the most common site of ERV expression. It is of particular interest to this discussion that the placental trophoblasts were especially prone to ERV expression. Given the importance of the invention of the placenta to mammalian speciation, this observation was particularly intriguing. Also of interest was the observation that many of these elements are highly repeated on the Y chromosome. In fact, this observation of placental HERV expression should not have come as a big surprise. Many years ago (1970s), J. Levy and his colleagues had reported that normal human placenta was producing large quantities of particles that clearly resembled retrovirus. These particles contained RT activity characteristic of a retrovirus, but they could not be shown to be infectious. However, antibodies against them were sometimes seen during pregnancy.

Some years ago, my own lab became interested in evaluating the role the placentally expressed ERVs might play in placental biology. I and others had proposed that these ERVs might be crucial for the normal biology of the placenta of viviparous mammals. The trophectoderm surrounds the egg and mediates egg implantation, feeding, and immune evasion in the mother. In a sense, a placental egg resembles a parasite, which must invade the mother host tissue, manipulate the mother's physiology to feed itself, and escape detection by the mother's immune system. None of these characteristics was present in monotreme mammals or marsupials; they all appeared to have been acquired in one complex evolutionary event. It seemed likely to us that ERVs were somehow involved in these complex placental characteristics. However, this posed a daunting experimental problem. Given the large number and complexity of the genomic ERVs, there were no experimental systems that were known to affect genome-wide expression of a particular ERV set. In addition, the trophectoderm is a very difficult tissue to manipulate in a live embryo, as it is the very first tissue to differentiate, prior to implantation. We therefore sought a surrogate mouse-based system that would allow us to globally affect ERV expression and examine implantation. Mouse cell lines do exist that are clearly able to differentiate into trophectoderm. Using such cells, one can make "embryoid bodies" in culture that closely resemble pre-implantation embryos. We sought to inactivate the expression of an env containing mouse ERV, then evaluate its effect on the ability of the embryoid bodies to implant. Using the gene of another virus (SV40 T-Ag), we were able to globally suppress mouse ERV expression in differentiated trophectoderm and show that this indeed prevented any implantation. This experiment is a bit too complicated to be completely clear in its significance, but the results did support the idea that ERV expression is important for the normal function of the trophectoderm.

HERVs Provide Normal Gene Function for the Placenta

Since our study, however, various other investigations have continued to evaluate the possible function of ERVs in placental biology. Most compelling have been the identification of the HERV W env gene as the molecule (Syncytin) used by the host to fuse the trophoblastic cells into syncytia, which form a tissue used to feed the embryo. Thus it has been clearly established that at least some of these HERV sequences are indeed important for normal placental biology. Such observations lend weight to the theory that ERV colonization was an important and creative event in the origin and evolution of placental species.

HERVs and Human Attributes

Can we now apply any part of the above rationale to consider the evolution of human-specific characteristics? This issue can be developed via two lines of reasoning. One, we can first consider the differences that have occurred during the divergence of human genomes from chimpanzee genomes to try to understand how these changes relate to human attributes. Another approach is to consider known human attributes, such as language acquisition, then seek to understand the underlying genetic basis for such characteristics. We will start by outlining the first of these lines of reasoning. Although, as mentioned at the start of this essay, human genes (coding regions) differ little from those of our primate relatives (98% conserved), other genetic differences are much more pronounced. Genomic analysis has established that the African primate genomes have undergone frequent rounds of colonization by lineage-specific types of retroviruses. These ERVs have names such as Fclenv, Fc2ma.\′fer, Fc2d env, and BabFcenv. Of these, Fc2master, Fc2d env are both ERV acquisitions that distinguish the great apes from the other African primates. However, as we have said, the genetic feature that most distinguishes humans from chimpanzees is the Y chromosome. Current estimates are that about 30 million years ago, the African primates underwent a substantial colonization of ERVs that currently distinguish those primates from the New World primates. These ERV acquisitions, along with some ERV deletions and LINE/SINE-based elements, are especially evident in the corresponding Y chromosomes (and to a lesser extent also the X chromosome). Blocks of sequence now distinguish the various primate Y chromosomes. These sequence blocks are mainly derived from retroposons that are most commonly related to HERV K.

HERV K and Africa

It is significant that a distinguishing characteristic of all the African primates (including humans) is that they have conserved intact copies of HERV K that also encode for a functional dUTPase enzyme. This UTPase conservation has major implications for the ability of African primates to support infection with the lentivirus family of retroviruses. Lentiviruses are the substantially more complex family of retroviruses, such as HIV-1, that have several additional genes relative to simple retroviruses. The lentivirus also have a distinct capacity for high-level replication in infected cells, such as HIV-1 in human CTLs or SIV in African monkey CTLs, a feature not seen in other retrovirus families. In fact, a curious observation has been that the lentiviruses of the African primates are distinct from the lentiviruses of all other animal species in that the primate lentiviruses lack dUTPase genes, whereas the lentiviruses of all other species (ungulates, felines) retain dUTPase genes. It would seem that the conservation of the genomic HERV K dUTPase is providing this function for the primate lentiviruses. The dUTPase appears to "detoxify" dUTP, which due to the high UTP concentrations of the cytoplasm would normally poison the RT catalyzed high-level cytoplasmic synthesis of cDNA viral genome. Thus, the ironic implication is that the African primates are prone to high-level lentivirus replication due to the conservation of HERV K dUTPase. This implication would seem the opposite of what many would consider to be best for the fitness and survival of these primate species. Why would African primates maintain selection for HERV K?

It is worth noting that the basal African primates are the most numerous species of African monkeys. Unlike the African great apes, African monkey species can support persistent high-level SIV infection with no disease. These same SIV infections can be lethal in Asian monkeys. Thus the basal African primates underwent a colonizing event that allowed them to persistently support nonpathogenic high-level lentivirus (SIV). I would suggest that this acquired capacity to support lentivirus persistence also created a circumstance that put African primate evolution into "fast forward," continually driven under the pressure of either stable ERV colonization or virus-induced disease. The result is that primates, such as the great apes and humans, that are not persistently SIV-infected, are propelled into higher rates of evolution by this viral pressure. HIV-1 would be the latest example of such pressure, but HIV-2, also restricted to Africa and more related to SIV, would be another example.

This view concerning the role of HERVs in human and primate evolution is consistent with the recently completed sequencing of the human Y chromosome. The major distinctions between human and chimpanzee Y chromosome have now been reported. From the early evolution of non-simian placentals, through the divergence of New World monkeys, to the development of African monkeys, and the emergence from African great apes of humans—all these transitions can be seen as distinct ERV-colonizing events on the Y chromosome. And the large blocks of sequence that distinguish human and primate species have been identified. Yet, curiously, the Y chromosomes of marsupials and monotremes have remained tiny (10,000 BP) and don't show these same ERV-colonizing events, indicating that this colonization process was associated with placental species. One surprise was how few coding sequences were found in the human Y chromosome (possibly as few as 20 genes). In addition, most of these sequence differences were retroposon-related, including coding sequences. Many of these HERV K/Sine R related sequences are also known to be expressed in reproductive tissue. Thus, the acquisition of such parasitic DNA marks us as human. In the past, evolutionary biologists considered the presence of such sequences to be simply due to accumulation of "selfish" or "junk" DNA that had no phenotypic consequence to the host. However, as I have argued above, we can expect important consequences with respect to the host-virus relationship, hence host survival. But the simple fact that we can identify these colonization events by genetic parasites does not obviously help us to understand how they might contribute to human characteristics.

What are those human characteristics, and can we see evidence of viral footprints in the genetic record that relates to their acquisition?

We should preface this discussion by acknowledging that in fact no definitive experimental evidence currently exists on this issue. We are left to consider some coincidental observations as well as some studies that are indirect at best. Yet there remains some very intriguing information that seems relevant. Cognition, especially the ability to think abstractly and the development of the capacity to learn recursive human language, is considered as a uniquely human characteristic. Another apparently uniquely human characteristic is the capacity for associative learning, which underlies the formation of human bonds and much of human behavior, but may also relate to other learning. With respect to human language, many neuroscientists feel that brain lateralization is a crucial aspect of the human capacity for language.

To approach the possible genetic basis of such a complex human capacity, it is often fruitful to consider disease states of the same process. In this respect, humans also appear to be uniquely prone to schizophrenia, a disease that many neuroscientists feel is associated with altered brain lateralization. Schizophrenia is a disease that is found in all human populations at low rates, and has also been associated with abstraction. Surprisingly to many, for some years now there has been a line of investigation that has pursued the possible role of endogenous virus in schizophrenia. These studies come about from two initial observations. One set of observations is the genetic one described above. That is, by considering what genetic changes distinguish chimpanzee from human, candidate genetic elements can be identified. Another line of investigation seeks to study directly the molecules of diseased regions of the brain with molecular approaches. Using a process known as differential display, several research groups have sought to isolate RNA that is specifically expressed in schizophrenic regions of the brain. These groups have reported isolating a cDNA that is very similar to the Sine R family of retroposons. Others have reported isolating and characterizing the HERV W related sequence (including an *env* sequence) that was expressed in these affected brain regions. However, since these are endogenous sequences, these observations can only be considered coincidental at this time, since it is difficult to establish whether expression of such HERVs is a consequence of, or causal for, the schizophrenia. Nor is it clear how the expression of such elements could affect basic brain function. It is, however, highly intriguing that these elements also represent recently acquired HERV elements that distinguish humans from chimpanzees. Thus, the surprising possibility remains open, that HERV could be involved in the development of human cognition and language.

The other human attribute we wanted to consider was the acquisition of associative learning, especially in the context of human social bonding. Such associative learning and bonding are thought to underlie the family, social, and cultural structure of all human societies. How might we evaluate the possible role of viral agents in such a process? On the surface, such a question almost seems preposterous. A virus involved in associative learning? Currently, there is no way to address this issue by human experimentation. We have essentially no data from human studies that would seem relevant. However, there are some animal models that would seem to be useful, if indirectly. Of these, the prairie vole is of special interest. In contrast to their close relatives, such as the montane vole, prairie voles form life-long monogamous pair bonds with their partners following mating, during which associative learning has occurred. They also learn to provide shared life-long care to their offspring. Prairie and montane voles are very similar genetically (99% in coding region), and differ mostly in the Y chromosomes. Thus, they provide a useful model to evaluate the genetic basis of associative learning. Brain-imaging studies of these voles suggested that the density of vasopressin receptor in the ventral pallidium might be directly involved in this learning process. Interestingly, an artificial recombinant virus was made that expressed this receptor and was injected into the brains of male mice. Such injected infected mice, when placed in the presence of females, would establish a pair bonding without the usual requirement for mating. Thus, a synthetic virus was able to induce associative learning in these voles.

Now the fact that a synthetic virus can do such a thing is not evidence that this is in fact the way it evolved. Yet from the context of a persistent virus, such a phenotype is logical and might be expected. This is because persistence is often transmitted via sexual activity, and must also be supported in the offspring of the infected host. Thus, mother-to-offspring transmission is common for many persisting viruses. Such a virus-host situation might require that the persisting virus be capable of manipulating the sexual and parenting behavior of the infected host in order to maximize virus survival. As will be described below, it is in fact already established that some viruses are indeed capable of such complex behavioral host modification. However, there is another aspect of persistent infections that may be relevant. As previously described, many persisting viruses achieve stability by using addiction modules that compel the host to maintain the parasite. It is interesting that the very same regions of the vole brain responsible for this associative learning are also involved in drug addiction. The expression of high levels of this receptor appears to create both a pleasurable and a painful (i.e., anxiety) learned state. Such a two-component situation clearly resembles an addiction module. Some, however, have questioned the relevance of this vole model to human behavior. Surely, it is argued, human love and bonding cannot be so simple as it is in these rodents. Yet recent genomic sequencing results suggest that at the level of gene complexity, we are not all that different from our mammalian relatives, including rodents. Additionally, recent PET-based brain-imaging studies of college couples in a newly acquired state of romantic love have found that the caudate nucleus regions of the brain are affected, and that these regions are also involved in the reward system. These human results are not inconsistent with the vole studies.

Some might also question the idea that a virus can manipulate complex sexual behavior of its host, as suggested above. It will probably surprise them to know that there are in fact several good examples of exactly such manipulation. The most compelling example is found in the polydnaviruses of parasitoid wasps. These viruses are endogenous DNA viruses of various wasp species that are produced by the wasp and injected along with the wasp egg into parasitized host larvae. These viruses alter larval physiology and behavior in various ways, involving expression of receptor molecules in nervous tissues, and expression of developmental hormones, in ways that strongly influence the host development and behavior. Recently, another virus has been identified in parasitoid wasps that directly affects the sexual behavior of infected wasps. Normally, the parasitoid wasp in this study will not lay an egg into a host larva that has already been parasitized by another parasitoid wasp egg. However, when infected by this newly identified virus, the wasp behavior is modified so that it now lays eggs into the already parasitized host, thus ensuring additional wasp eggs will be present for virus transmission to occur. It is, therefore, clear that persisting viruses can indeed manipulate host behavior in complex ways. The question that remains to be answered, however, is whether this process might also have been involved in the acquisition of complex human attributes that affect behavior.

We cannot now provide compelling evidence that the endogenous viruses that have colonized human DNA have directly contributed to the evolution of human attributes. The source of our human complexity remains to be identified; it will be the work of future generations of scientists. However, a much stronger case can be made with respect to the source of other complex phenotypes that have led to the evolution of complexity in other living lineages. This includes the eukaryotic nucleus, flowering plants, the adaptive immune system, and viviparous mammals. All of these examples represent dilemmas of host complexity, acquired in a seemingly punctuated evolution, that defy simple answers based on Darwinian selection of point changes. The arguments and relevant observations that support a viral role in the origin of such genetic complexity are too numerous and detailed to be summarized here. However, they have been collected into a book soon to be published by ASM (*Viruses and the Evolution of Life*). The point to be made here, however, is that the stable colonization of cells by persisting viruses offers an explanation of the source of complex genetic creativity that can be acquired in a punctuated manner. Host evolution and species diversification may actually depend on such viral parasitization. Thus, we can start to explain broad order-based patterns of virus-host relationships. That this same process of viral-based genetic creativity could also apply to the acquisition of human attributes does not then seem so farfetched.

Viruses are inherently invisible and are normally observed only as a consequence of diseases they can cause. Yet such disease states are not the most common mode of viral existence. Unapparent persistent infections are more common. It is our inability to perceive viruses, especially the silent virus, that has limited our understanding of the role they play in all of life. Only now, in the era of genomics, can we more clearly see their ubiquitous footprints in the genomes of all life. We need to expand our consciousness with respect to the vast creative potential of the viruses: It is vast, almost beyond comprehension. Viruses must now be considered as the leading edge or the points of genesis in the growth of the tree of life. As recently defined by Frank Ryan, viruses are inherently symbiotic. They add new genetic identity to their host, thus endlessly seeking to add further complexity as well as culling those host that fail to prevent competition with other genetic parasites.

Persistence allows the vast viral creative potential to contribute to the genesis of new host. Let us consider the oceans, that vast cauldron from which all life has evolved. Few of us are aware that the oceans are also a vast and ancient viral cauldron. In fact, recent estimates suggest that the combined oceans contain about 10^{31} viral particles, mostly consisting of large icosahedral dsDNA viruses (both lytic and persistent) of prokaryotes, such as cyanobacteria, but also including viruses of algae. In order to get a physical sense of this scale of quantity, we can put this number in physical terms. For example, we know that the average diameter of these viral particles is about 100 nm. If this quantity of virus were laid side by side, it would span 10^{24} meters. That corresponds to the estimated diameter of the known universe. In addition, most of this viral mass has been measured to turn over every day (mostly from UV irradiation). Thus, every day in the oceans, virus genomes are regenerated on an astronomical scale. This turnover is likely to have been continuing since the oceans were first colonized by prokaryotes (3~ billion years). Furthermore, we know that substantial quantities of viral genetic information stably colonize host cells as prophage, so there is a well-established conduit from virus into cellular life. We can add to this calculation the vast numbers of viruses in terrestrial habitats (microorganisms, plants, animals), to get a global sense of the scale of viral creativity. This viral-mediated process of evolution is unending and continues before our eyes, if we would only acknowledge it.

Ever since Darwin introduced his concepts of common descent and survival of the fittest from genetic variation, some have wondered why we generally fail to see the acquisition of complexity by any living entity. Why have we not witnessed the evolution of new super-species, such as other monkeys able to talk or acquire other human attributes? Clearly, we have been observing evolution only for a very short time. Yet we can witness what current viruses such as HIV-1 can and might do to the human population. Without our current culture (a product of associative learning), we can project what HIV-1 would do to a human population as is currently occurring in an unrelenting pandemic in Africa. The projections are simple. Without social or technological intervention, essentially all Africans would become HIV-1 infected. AIDS would sweep the entire population and most would succumb to disease. However, we can also expect at least a few humans to survive. These survivors will be those who fail to progress in disease and are persistently HIV-1 infected. Such individuals have in fact already been observed. These survivors would thus be left to repopulate the continent. However, the resulting human population would be distinct, and would have acquired some new and complex characteristics. This surviving population would now differ from the previously existing human population in several biologically important ways. For one, it would be sexually incompatible with any pre-existing human population since sexual relationships between the two populations would result in HIV-1 infection and AIDS in the non-selected population. This would tend to provide a selective setting for the separation of interbreeding. Another important biological outcome is that the now resistant HIV-1 infected African population would have acquired a new set of complex genes that regulated various aspects of cellular molecular biology and immunology. These would be the HIV-1 lentivirus genes. Such a gene set would now be available for Darwinian selection to operate on, applying the potential functions they provide to create a more fit human population. If this were to be the outcome, we would see a new species of human, marked by its newly acquired endogenous viruses, much like the differences we see between human and chimpanzee genomes. Thus viruses may well be the unseen creator that most likely did contribute to making us human.

References

Bromham, L. 2002. The human zoo: Endogenous retroviruses in the human genome. Trends in Ecology and Evolution 17:91–97.

Griffiths, D. J. 2001. Endogenous retroviruses in the human genome sequence. Genome Biology 2001, 2(6):reviews1017.1–1017.5.

Katsanis, N., K. C. Worley, and J. R. Lupski. 2001. An evaluation of the draft human genome sequence. Nat. Genet. 29:88–91.

Olivier, M., A. Aggarwal, J. Allen, A. A. Almendras, E. S. Bajorek, E. M. Beasley, S. D. Brady, J. M. Bushard, V. I. Bustos, A. Chu, T. R. Chung, A. De Witte, M. E. Denys, R. Dominguez, N. Y. Fang, B. D. Foster, R. W. Freudenberg, D. Hadley, L. R. Hamilton, T. J. Jeffrey, L. Kelly, L. Lazzeroni, M. R. Levy, S. C. Lewis, X. Liu, F. J. Lopez, B. Louie, J. P. Marquis, R. A. Martinez, M. K. Matsuura, N. S. Misherghi, J. A. Norton, A. Olshen, S. M. Perkins, A. J. Perou, C. Piercy, M. Piercy, F. Qin, T. Reif, K. Sheppard, V. Shokoohi, G. A. Smick, W. L. Sun, E. A. Stewart, J. Fernando, Tejeda, N. M. Tran, T. Trejo, N. T. Vo, S. C. Yan, D. L. Zierten, S. Zhao, R. Sachidanandam, B. J. Trask, R. M. Myers, and D. R. Cox. 2001. A high-resolution radiation hybrid map of the human genome draft sequence. Science 291:1298–302.

Ryan, F. 2002. *Darwin's blind spot: evolution beyond natural selection.* Boston: Houghton Mifflin Company.

Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, et al. 2001. The sequence of the human genome. Science 291:1304–51.