Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

CHAPTER

21

The Widespread Evolutionary Significance of Viruses

Luis P. Villarreal

ABSTRACT

In the last 30 years, the study of virus evolution has undergone a transformation. Originally concerned with disease and its emergence, virus evolution had not been well integrated into the general study of evolution. This chapter reviews the developments that have brought us to this new appreciation for the general significance of virus evolution to all life. We now know that viruses numerically dominate all habitats of life, especially the oceans. Theoretical developments in the 1970s regarding quasispecies, error rates, and error thresholds have yielded many practical insights into virus–host dynamics. The human diseases of HIV-1 and hepatitis C virus cannot be understood without this evolutionary framework. Yet recent developments with poliovirus demonstrate that viral fitness can be the result of a consortia, not one fittest type, a basic Darwinian concept in evolutionary biology. Darwinian principles do apply to viruses, such as with Fisher population genetics, but other features, such as reticulated and quasispecies-based evolution distinguish virus evolution from classical studies. The available phylogenetic tools have greatly aided our analysis of virus evolution, but these methods struggle to characterize the role of virus populations. Missing from many of these considerations has been the major role played by persisting viruses in stable virus evolution and disease emergence. In many cases, extreme stability is seen with persisting RNA viruses. Indeed, examples are known in which it is the persistently infected host that has better survival. We have also recently come to appreciate the vast diversity of phage (DNA viruses) of prokaryotes as a system that evolves by genetic exchanges across vast populations (Chapter 10). This has been proposed to be the "big bang" of biological evolution. In the large DNA viruses of aquatic microbes we see surprisingly large, complex and diverse viruses. With both prokaryotic and eukaryotic DNA viruses, recombination is the main engine of virus evolution, and virus host co-evolution is common, although not uniform. Viral emergence appears to be an unending phenomenon and we can currently witness a selective sweep by retroviruses that infect and become endogenized in koala bears.

Origin and Evolution of Viruses
ISBN 978-0-12-374153-0

INTRODUCTION

Our understanding of virus evolution has reached a threshold in that it now appears to provide a much broader vista regarding its general significance and influence on host evolution. Several developments have brought us to this point. One has been the realization that viruses often evolve by processes involving the collective action of a consortia, or quasispecies. And the resulting adaptability and power of such evolution is unmatched by any other genetic entity. Much of this volume is dedicated to this issue. In such consortia, the concept of a "wild-type" virus is no longer considered to be the fittest type, as the quasispecies itself provides fitness (see Chapters 3 and 4). The quasispecies model resembles population genetics in some ways, but it has led to some significant departures from population genetics, and these departures are very well supported by experiments. Another development that has recalibrated our view of the overall significance of viruses is information on the scale and diversity of viruses. Viruses are present at a previously unappreciated global level and appear to have affected the evolution of all life on Earth. Much of this realization has been brought about by the development of metagenomic methods as applied to various habitats. Measurements of major habitats (the oceans, soil, extreme environments) have established that our biological world is predominantly viral, in terms of both numbers and diversity (Paul *et al.*, 2002; Breitbart *et al.*, 2003; Breitbart and Rohwer, 2005a, 2005b; Edwards and Rohwer, 2005; Comeau *et al.*, 2006).

These two developments would seem reason enough to consider virus evolution in a new light. However, there have also been numerous theoretical proposals suggesting viral involvement in some of the very earliest events and major transitions in the evolution of life. We no longer think of viruses as recent agents that escaped from the host chromosomes as run away replicons. Viruses now appear very old to us and they relate to and trace all branches of life. The last 30 years have been very active regarding virus evolution. Major developments in theory, technology, medicine, and the study of human disease with respect to virus evolution have all occurred. And as we seek to grow and manage various life forms for human use, virus evolution has also had major impact on such efforts.

As a science, virus evolution has benefitted greatly from traditional evolutionary biology. However, since viruses are molecular genetic parasites that are inscrutable by casual observation, our understanding of virus evolution has been dependent upon measuring sequence variation and sequence diversity in a large number of virus genomes. Because of this, these small genetic parasites have been the last domain of life to yield their secrets of evolution. And viruses harbor some clearly distinct evolutionary abilities. For one, they are polyphyletic. All major viral lineages have their own distinct origins. They are also difficult if not impossible to define as species and are able to exchange genetic information across normal boundaries. Even "dead" and defective viruses can participate in such exchanges, which confuses the definitions of fitness. We now know that viruses can evolve by a consortia process and also exchange information by recombination across vast genetic pools to assemble new mosaic combinations of genes. We thus no longer think of a specific genetic lineage in understanding virus evolution, but instead think of a cloud, matrix, or a population as the basis of virus evolution. Viruses are inherently fuzzy entities that can differ from their relatives in any specific feature. Yet even with such fuzziness, it is clear that common themes also link them. Patterns of evolution have become clear. Diversity and variation are often (but not always) observed. Stability and host congruence can also be observed. Nevertheless, the evolutionary power of viruses has been learned at a human cost.

The application of numerous analytical and phylogenetic tools have provided crucial insights into virus origin and evolution. Yet these methods struggle to incorporate the fuzzy nature of viruses and have clear limits, especially regarding quasispecies and

high recombination rates. Structural biology now also adds tools that extend our vision of virus evolution beyond what can be seen in the genetic sequence. For example, common structural motifs from phage to eukaryotic DNA viruses (T4 and herpesvirus) suggest very ancient links in virus evolution that span all domains of life (see below). Nevertheless, our analytical methods are currently lacking as we struggle to understand complex genetic mixtures that provide fitness, reticulated relationships, polyphyletic origins, and virus–host congruence. Virus evolution has, for the most part, been considered to be a specific, esoteric part of broader evolutionary biology and has been given limited attention in reference works on evolutionary biology (see Pagel, 2002). Historically, the focus has been on various RNA viruses and some DNA viruses that cause disease in humans and domesticated animals and plants (Domingo *et al.*, 1999). However, I have asserted that all life forms must be examined from the perspective of virus evolution (Villarreal, 2005), not just those pathogens that impact on us. How viruses evolve in a more general sense informs us of evolutionary paradigms that have not been previously well understood (especially the evolution of consortia or the dynamics of vast reticulated gene pools). This volume now extends these traditional topics of virus evolution to include the vast virology of the prokaryotic world. In so doing, it illuminates the global consequences viruses have had on all life forms.

Prior to the 1970s we saw some stunning successes in vaccine control for major viral human diseases, such as polio, measles, mumps, influenza, and especially smallpox virus. Due to this historic success, American health agencies and educators considered virus disease as a thing of the past, no longer a serious threat. The era of infectious disease that they represented was now one for the historical record, an unfortunate part of human history, or so we thought. In what now seems to have been a clear case of hubris and naivety, we have been humbled by the evolutionary power of viruses, which was woefully underappreciated, even by most virologists. By the end of that decade, the evolution and emergence of HIV-1 permanently changed our views (see Chapters 13 and 14). This has also been followed by a seemingly never-ending series of viral threats as newly emerging viral diseases have come to our attention. HIV-1 provides the only example of a public health situation that has reversed centuries of progress for extending human health and lifespan. It now limits human life expectancy in many parts of the world, especially sub-Saharian Africa. This development could not have been imagined in the 1970s. We are much less confident now about predicting the future of virus evolution and its potential impact on human health.

Diseases of domesticated microbiological, plant, and animal species have also experienced the trauma of the consequences of emerging viral diseases along with huge losses. However, the human HIV-1 story may not be a fluke of virology but may be telling us something basic about human and primate evolution. As we sought to understand the origin of this new virus, we have come to appreciate a much broader virus–host story which involves simian immunodeficiency virus (SIV), foamy viruses, and the speciation of Old World primates. We have also come to learn that the genomes of these primates show much evidence of past viral interaction and ongoing endogenous retrovirus colonization. The evolution of retroviral endogenization has taken on a much greater significance in basic evolutionary biology. Thus it is with great interest that we now study the ongoing endogenization of retroviruses in the koala bear genome (see below).

Historically, we are compelled to study viruses because they can cause serious disease. New viruses come to our attention also mainly because of disease. It is therefore understandable that most evolutionary biologists mainly think of viruses strictly as agents of disease. These are the products of run away replicons that provide negative selection to host survival. In this light, the application of predator–prey based mathematical models has seemed most appropriate. With such viral disease, variation has long been observed and

was initially used for the generation of most vaccines. However, this disease-centric view has also occluded another more prevalent virus–host relationship. For example the emergence of HIV-1 has led us conclude that it likely evolved from various versions of SIV. But SIV is not pathogenic in its native African primate host. Nor does it show the genetic diversification of HIV-1 in these native primate hosts. It is a silent, asymptomatic infection. Genomic and metagenomic analysis now allows us to identify many more silent, asymptomatic viruses that would not have previously been observed. We now know of many such viruses that are prevalent in a specific host. Evolutionary biology must escape the confines of disease-centric thinking and seek to understand these relationships as well.

A Role for Persistence in Viral Evolution

In the last ten years I have attempted to provide another view concerning virus–host evolution. I have argued that viruses often attain evolutionary stability by species-specific persistence and that such states apply to all domains of life, including prokaryotes. On an evolutionary time-scale, the majority of viral lineages tend to exist as species-specific persistent (aka temperate, latent, and chronic) infections in which individual hosts will be colonized by mostly silent (asymptomatic) viruses for the duration of their life (Villarreal *et al.*, 2000). Such persistence can have major consequences to the evolution of both virus and host, which also leads us to more directly link virus evolution to broader issues of host evolution. It is from this perspective that we start to clearly see that viruses indeed belong on the tree of life as major participants (Villarreal, 2005, 2006). Persisting viruses are not simply agents responsible for destruction of life, but are also agents that create genetic novelty on a vast scale that influences all life and promotes symbiosis (Marquez *et al.*, 2007; Ryan, 2007; Villarreal, 2007). The persistent lifestyle of such symbiotic virus–host relationships is not simply a less efficient, acute infection; nor is it simply a "reservoir" for acute virus (as epidemiologists are prone to assert). Neither can it be attributable to concepts of selfish DNA. Persistence represents a major virus life strategy that is both fundamental and highly adapted. It has distinct genetic, fitness, and evolutionary characteristics that require intimate, host (tissue)-specific viral strategies and precise gene functions to attain stable maintenance in the presence of immunity and to allow biologically controlled reactivation. Persistence also must resist displacement by similar viruses and competitors. It is virus–host persistence that provides the thread that allows us to link these polyphyletic viral lineages (and their clouds) with the entire tree of life. In turn, this link identifies a much more fundamental role for viruses in the evolution of host, visible from the very earliest to the most recent events in host evolution. It is from such species-specific persistent states that the large majority of acute diseases evolve and emerge by various mechanisms.

We know much about virus replication and disease. However, our understanding of the specific mechanisms of persistence is generally poor. Persistence is a generally silent and inscrutable state, it does not lend itself to *in vitro* or cell culture experimental models. We are left with but a few examples from which to attempt to extrapolate the possible existence of general relationships. The study of virus evolution thus struggles to incorporate concepts of persistence.

Viruses Mediating Innovation

Another recent and major development in virus evolution is the arrival of various proposals suggesting that viruses have been involved in some major innovation and transition in the evolution of life. In all these proposals, however, it is necessary that the virus in question has attained a stable genomic persistence with its host. These evolutionary events thus seem to be the products of viral-mediated symbiogenesis of host (Ryan, 2007; Villarreal, 2007). Proposals include the

possibility that viruses may have originated the DNA replication system of all three cellular domains (archaea, bacteria, eukarya) of life (Forterre, 1999, 2005, 2006a, 2006b; Filee *et al.*, 2003; Forterre *et al.*, 2007). The discovery and analysis of the largest DNA virus (1200 ORF mimivirus), a lytic cytoplasmic virus of amebae (a distant relative of phycodnavirus and poxviruses), has also led to proposals that this virus lineage may represent an ancient fourth domain of life (Raoult *et al.*, 2004; Desjardins *et al.*, 2005; Claverie *et al.*, 2006). It is interesting that in an initial structural analysis, the large complex replication centers for mimivirus were confused with the host nucleus (Suzan-Monti *et al.*, 2007). Thus, it seems relevant that others have proposed that a distant relative of phycodnaviruses and poxviruses may have originated the eukaryotic nucleus (Villarreal, 1999; Villarreal and DeFilippis, 2000; Bell, 2001, 2006; Takemura, 2001).

Such proposals, although consistent with various observations, however, remain outside of the consensus of most evolutionary biologists. Nevertheless, numerous other observations continue to suggest viral involvement in other major host innovations, such as a viral origin of RAG1/2 of the adaptive immune system (Dreyfus *et al.*, 1999 Kapitonov and Jurka, 2005; Fugmann *et al.*, 2006) or the role of endogenous retroviruses (ERVs) in the evolution of the placenta (Villarreal and Villareal, 1997; Harris, 1998; Blond *et al.*, 2000; Mi *et al.*, 2000; Dupressoir *et al.*, 2005; Caceres and Thomas, 2006). Such possible roles for viruses in host evolution are at odds with accepted views of virus–host relationships, but might be the products of viral symbiogenesis.

The Virosphere: a New Evolutionary Reality

The metagenomic viral measurements mentioned above for prokaryotic DNA viruses, along with the increasing realization that viruses and host can co-evolve, has led to various calls that a viral tree of life needs to be considered and developed (Forterre, 2003; Villarreal, 2006; Filee *et al.*, 2006b). A virosphere clearly exists but its nature and boundaries are not so clear. Multiple viral origins, their diversity and numerical dominance in distinct and sometimes harsh environments as well as their presence in host genomes suggest that any viral tree of life will be huge, multidimensional, and connected to the host tree of life. As discussed below regarding double-stranded (ds)DNA viruses of prokaryotes, they have all the above characteristics. Such viruses may represent the big bang of biological novelty. With their unmatched capacity to generate diversity they can function as the mass creators of biological novelty as well as destroyers of most species. Surely, such capacity must have had big influences on the evolution of life.

Symbiosis, simply defined, is the stable coexistence of two previously separate lineages of organisms. There can be little doubt that many temperate phage can stably colonize a bacterial cell resulting in a stable descendent from two lineages. This is clearly symbiotic. Endogenous retroviruses can similarly be found to persist in vertebrate genomes and also appear symbiotic. Yet studies of symbiosis seldom consider a role for virus (Villarreal, 2007).

How important are viruses in general to evolutionary biology? The core concepts of evolutionary biology were developed well before we had a modern understanding and definition of viruses (Luria, 1950). After all, the basic lysogenic model of phage integration was only clarified in 1962 when developed by Campbell (see Campbell, 2007). That cryptic and defective phage are ubiquitous in the genomes of all prokaryotes is generally considered uninteresting by many in the field of evolutionary biology. I suggest that the seemingly applicable concepts of selfish DNA effectively derailed any thinking that persisting genetic parasites might have a more germinal role in the evolution of life (Doolittle and Sapienza, 1980; Orgel and Crick, 1980). Yet as outlined above, virus footprints in major evolutionary transitions are clear and a direct role in such events now seems much more plausible. We therefore must seek to

defining the nature of the virosphere and how its evolution relates to the tree of life.

EXEMPLARS OF VIRUS EVOLUTION

This book represents the first integration of the entire field of virus evolution, including both prokaryotic and eukaryotic life forms. However, because our understanding of virus–host relationships remains uneven, the chapters necessarily focus on well-studied models (exemplars). These exemplars also tend to reflect a historic disease focus (i.e., *E. coli*, flowering plants, mouse, humans). It is unfortunate that the silent species-specific viruses that tend to exist in stable states with long evolutionary histories seldom provide our examplars. We understand these infections poorly and lack basic definitions concerning fitness or selective advantages. Only metagenomics tools now seem able to inform us of their presence, but not their biology.

Errors at the Start of Life: Quasispecies

Self-organization and the evolution of RNA molecules as an origin of biological information are discussed in Chapter 1. Autocatalytic chemical reactions, such as replication of RNA, presents issues such as how to optimize a rugged fitness landscape yet allow the study of evolution *in vitro* with RNA. RNA *in vitro* reduces genotype–phenotype issues to RNA secondary structure and minimal free energy states. This allows both continuity and discontinuity to be measured. These same issues are crucial for the study of RNA viruses, whose sites of secondary structure often define replicator identity. These models currently offer the best system to evaluate an early and simple biological world for evolutionary principles (see Chapter 3). Viruses and viroids with their RNA genomes may be the only extant survivors of this pre-DNA world. Since it was the consideration of error-prone replication that led to the development of the concept of quasispecies (see Chapter 2), such models have provided a conceptual foundation which led to several basic concepts. Chapter 4 discussed the foundations and various aspects of quasispecies theory. Viruses appear to operate close to the error threshold, thus allowing maximum evolutionary exploration (Biebricher and Eigen, 2006). However, as presented below, the loss of the "fittest" type concept has also led to clear experimental evaluations of consortia-based evolutionary behaviors. Such behaviors were not predicted by classical Darwinian models.

Although virologists were initially attracted to quasispecies models, many evolutionary biologists were initially hostile to the application of the quasispecies concept to evolution. It was thought that the classical mathematical models of population biology, as originally developed by Wright and Fisher, and later applied by Kimura and Maruyama to asexual haploid populations at the mutation–selection balance, had already fully developed the needed models and precluded additional need for the quasispecies concept. The classical models were thus argued to provide adequate mathematical coverage for viruses, including quasispecies and error threshold (Wilke, 2005). However, these two approaches differ fundamentally with regard to the significance of error-prone replication and it was the quasispecies approach that led to the clear experimental establishment that quasispecies selection, *per se*, is important for viral pathology and fitness (see below).

The development of quasispecies theory to virology does indeed demonstrate distinct differences with population genetics. Various phenomena, such as complementation, cooperation, competition, and even defective mediated extinction (Domingo *et al.*, 2001, Domingo 2006; Grande-Perez *et al.*, 2005) have been observed, all of which fall outside of the parameters of classical population genetics. Viral fitness has indeed been shown to be due to interaction within a diverse population, and not to the fittest or master type. And with RNA viruses, error threshold has become a central issue (Biebricher and Eigen, 2005).

The collective experience has thus made clear the value of theory to biology. Many working biologists understand that life seems overly complex and defies most generalizations. Thus they do not always appreciate attempts at general theory. Although in reality, biology may indeed often be too complex for accurate theoretical predictions, these theories have nevertheless clearly stimulated crucial concepts and experimental evaluation, and biologists should be encouraged by them. By providing new ways of thinking, entirely new experimental approaches can be developed.

The existence of error-prone replication and quasispecies also raises the issue of the conservation of information. How is information stability and higher fitness attained with such errors? How can genetic complexity be created in such a circumstance? Can cooperation (or consortia behavior) result from any of these models? These issues have yet to be resolved. Some interesting suggestions, however, have been proposed. One involved cooperative evolution that results from ligated genomes. A model was proposed in 2000 by Stadler and Schuster in which they considered the dynamics of replicator networks resulting from higher order reactions involving the templated ligation of smaller genomes (Stadler *et al.*, 2000). Although this was based on concepts such as triple-stranded nucleic acids, this clearly has some elements that also resemble the ligation of recombinational processes for DNA phage (presented below).

A most interesting outcome of these models is that, depending on initial replicator concentrations, permanent coexistence of replicators could result in a cooperative network. Such cooperation is a rare outcome for most models and given the conclusion that early DNA based life was a "horizontal" consortium, such models are of special interest. The issue of consortia will come up often. Consortia selection directly implies cooperation, but cooperation of selfish replicators presents dilemmas. Replicator networks with interaction functions that give highly non-linear dynamics can result in complex mixtures, with behaviors ranging from survival of the fittest to also including attainment of globally stable equilibrium tantamount to permanent coexistence. The fitness of populations, however, is inherent to current quasispecies concepts in virology. There may also be other ways to explain the genetic origin of cooperation (such as stable persistence involving addiction strategies). The stable persistence of a genetic parasite can compel cooperation and promote the conservation of information (see below).

FITNESS, CONSORTIA, AND PERSISTENCE

The definitions of fitness with respect to a virus in a natural habitat are far from clear. Although the concept of relative replicative fitness is often applied to lab experiments of virus growth, we know many situations in which virus replication is not maximized in natural settings and many viruses can exist in relatively non-replicative states for long periods. Even in the context of an acutely replicating virus in a host organism, the concept of fitness is clearly conditional, as the virus must replicate through various *in vivo* habitats that can have opposing selection. As presented below, *in vivo* models that study fitness and viral diversity have clearly indicated that diversity *per se* is important and fitness is the result of consortia. How do we define such fitness since the mixture clearly matters? Also, how do we define information content or integrity of a consortia? Currently, we cannot. In the lab, the viability of a virus is usually measured by the ability to produce plaques. This has been a crucial and main assay for many experimental systems that study virus population. Here, the definition of fitness seems direct: plaque formation equals fitness. Various highly useful models have thus been defined and developed that depend on these plaque assays (see Chapter 4). With this, populations and population growth are defined as relative growth of plaque-forming units. However, the concept has always been problematic when considered from a natural virology perspective. Plaque formation

is clearly not equal to fitness in natural habitats. There are many examples of highly successful viruses that either plaque poorly or not at all. Consider the roughly 100 types of human papillomavirus (HPV), a simple small circular DNA virus of epithelia; this does not form plaques in any known system (Chapter 18). HPV is clearly fit, well adapted to, and stable in its human host. In addition, HPV evolution is phylogenetically congruent with their primate host, as are most persistent viral infections. We have yet to understand the definition of fitness in this situation.

In some cases, it seems selection for plaque propagation has clearly resulted in loss of highly conserved genes; such as with the plaque-adapted laboratory strains of cytomegalovirus (CMV). The problem posed by viruses with inefficient plaque formation is not limited to DNA viruses. Many persisting RNA viruses also do not plaque well or at all, such as most RNA viruses of plants or many insect picorna-like viruses, such as those found in *Drosophila* and bees (which also conserve an extra ORF). Nor do most persistent infections make lots of virus. Low-level persistence, such as hantavirus in rodents, for example, is common (Hart and Bennett, 1999).

Clearly, our simplifying assumptions of viral fitness and population dynamics cannot apply to these stable evolutionary states. However, if we limit our definition of viral fitness to relative replication or plaque formation we can perform some clear and quantitative evaluations.

Fitness Theory for Persistence

Experimental evaluation forces us to study fitness by only those definitions that we can currently measure. As fitness appears to be a relativistic and transient concept, depending very much on the tissue, time, place, extant adaptive and innate immunity, and competition, it is likely that we can only measure with any accuracy one aspect of fitness at any one time. HIV infection of humans shows evidence of this in that the R5 virus is more fit for transmission and early disease whereas the X4 virus is fit later during the AIDS disease phase. Clearly conditional, time-dependent issues relate to fitness definitions. However, much more problematic is that we have no theory for viral persistence or its fitness. We lack specific or measurable parameters other than the simple maintenance of genetic material. Yet it seems clear that some distinguishing features of persistence can already be recognized. For example, the possible participation of viral defectives (normally considered unfit), which in numerous circumstances can modify or mediate persistence, would need to be included. Clearly, a defective role in persistence would also preclude them from being considered as genetic "junk," or selfish elements, since they would then matter in measurable ways to the biological outcome of virus persistent infection.

Persistence also requires an extended duration of infection, not simply maximized replication. In fact, persistence generally requires mechanisms to limit the replication of at least the same virus for at least some time. Thus, limited replication must be an essential element for this life strategy. In my judgment, and much like the quasispecies concept, the concept of persistence will eventually be recognized for the fundamental (symbiotic) force it represents in virus evolution.

The experimental work of Domingo and Holland spans the modern assessment of quasispecies theory that occurred in the 1980s and 1990s. These investigators were the chief proponents of this theory, bringing it to the attention of the broader virology community (Chapter 4). This work has transformed our thinking and laid the experimental foundations that we now build upon. This current volume is an extension from an earlier book on quasispecies (Domingo *et al.*, 1999) and now encompasses both prokaryotic and eukaryotic viruses. Since early experimental phage studies provided the foundations for quasispecies theory (Eigen *et al.*, 1988), using mathematical descriptions (differential equations) of mutation rates in T-even phage (Luria and Delbruck, 1943), this inclusion is appropriate.

Interestingly, a second early paper measuring replication rates by these same authors also noted the problems of viral interference and defectives (Delbruck, 1945). Other early experiments of phage RNA polymerase (Batschelet *et al.*, 1976), especially with RNA phage Qβ (Domingo *et al.*, 1978), helped set the stage for the subsequent experiments of the 1980s and 1990s. From the test tube to mouse models to the study of human disease, the work of Domingo and colleagues has spanned the entire history of viral quasispecies (Domingo and Gomez, 2007; Chapter 4).

Quasispecies deals with the products of error-prone replication. However, it is worth repeating that products of error-prone replication are not behaving in a simple "selfish DNA" capacity and are not devoid of biological relevance and phenotype. In their complex populations, they create clear and varied affects on viral adaptability, competition, and fitness. Since quasispecies necessarily involve defective and mutant virus, it is easy (and common) to think of these entities simply as genetic junk (Villarreal, 2005, 2006). Defective and even lethal or interfering variation in viral genomes can contribute to adaptability. Thus, viruses can clearly adapt as a cloud with a mutant spectra. In addition, unending competition and exclusion, consistent with the Red Queen hypothesis, has also been observed (Clarke *et al.*, 1994).

The poliovirus–mouse model (see below) in particular has provided a solid experimental system for evaluating the adaptive consequence of quasispecies. It is thus ironic that these same experiments have also made clear that the original simplifying assumptions of the quasispecies ordinary differential as presented by Eigen are violated by the resulting quasispecies. These products of error-prone replication do indeed strongly interact with each other in both positive and negative ways and such interactions contribute significantly to the observed fitness of the population.

Errors and interaction are important for fitness. For example, defectives have been reported to mediate extermination of a competing wild-type virus (Grande-Perez *et al.*, 2005). Complementation has also been observed (Garcia-Arriaza *et al.*, 2004), as has *trans*-dominant inhibition (Crowder and Kirkegaard, 2005). Genetic memory of past selection has been shown to be maintained in a minority of the population (Ruiz-Jarabo *et al.*, 2000; Briones *et al.*, 2006). Such cooperative (consortia) behavior, which can also depend on unfit or defective members, is at odds with classical Darwinian notions regarding survival of the fittest. Consider, for example, the fitness of a defective or mutant outside of its role in a quasispecies. Such a consideration ignores the very nature of a quasispecies yet it is an issue that has often been posed and experimentally evaluated. We should refrain from thinking of viruses simply as fit or non-fit individual types since they clearly exist in populations that provide population-based adaptability.

The selection of viral consortia or population raises some fundamental issues for evolutionary biology. This is essentially group selection in which a population, not the fittest individual, is selected. This view makes cooperation or interaction of individual genomes a significant component of selection, which is not commonly thought to be a general or accepted process in evolution. Yet population selection is no longer a contestable issue in RNA virus evolution (see below). I expect that many classical evolutionary biologists might interpret this as evidence that viruses really are an oddity in this feature and are not representative of broader processes of evolutionary biology. Furthermore, viral-based group selection may not be limited to quasispecies-based evolution. As presented below, persistent viral infections may also provide population-based selective advantage (see below for the P1 and mouse hepatitis virus (MHV) persistence exemplars; Villarreal *et al.*, 2000; Villarreal, 2006, 2007).

Since viruses are ancient, numerically dominant, and the most diverse biological entities on Earth, no life form can escape exposure to them. All extant life forms have evolved in a viral habitat. Thus we should expect that the viral footprints (including defectives) that we now find in all genomes have likely played an active role in their evolution; a role, I would

argue, that is fundamental, dynamic, and unending. If we can accept this assertion, we may start to see and appreciate the vast evolutionary power that viruses can bring to bear onto host evolution. We can start to attain a global perspective and appreciation for their ability to assemble genetic function from enormous, complex mixtures of genomes, and select gene sets needed to solve multivariant, temporally dynamic evolutionary problems. We can then seek evidence for the role of viral elements in fundamental host innovation and be open to evaluating the occurrence of viral entities from a constructive perspective and not instinctively dismiss such observations as due to coincidence, "junk," or selfish DNA.

The advantage of such a perspective is that it will promote the specific experimental evaluations that can better assess any constructive role genetic parasites might have played in host evolution. For example, there is much reason to think ERVs have played an active role in human evolution (for references see Ryan, 2007). The quasispecies concept has provided the foundation for us to understand virus evolution and informed us of the evolutionary power viruses possess. If that power also links to host evolution, then the tree of life becomes enriched by virus, much larger and more dynamic.

The Poliovirus–Mouse Exemplar: "Quasispecies *per se* Rather than the Selection of Individual Adaptive Mutations Correlates with Enhanced Pathogenesis"

The recent experimental studies from Andino and colleagues using poliovirus in the mouse model should, in my judgment, provide the keystone exemplar regarding the *in vivo* fitness of quasispecies (see Chapters 6 and 7; Vignuzzi *et al.*, 2006). These studies make clear the importance of quasispecies and error-prone replication. Such detailed *in vivo* experiments were made possible by a long and detailed history of poliovirus studies that has identified the nature of RNA polymerase fidelity as well as developed mouse models for the study of pathogenesis. Few other virus–host systems could have provided such potential for high resolution. These results also provide the experimental observations that distinguish quasispecies-based evolution from the classical Fisher-based population genetics. The general importance of this story for understanding virus evolution thus deserves special emphasis.

The very origins of modern animal virology stem from poliovirus studies with the need to develop *in vitro* cell culture technology in order to grow and evaluate poliovirus and generate variants. The live poliovirus vaccine is of special interest with regards to virus evolution and adaptability. The "live" oral Sabin vaccine can be considered to have been a miracle of the practical approach to virology developed in the 1950s (Horaud, 1993) in that it was used well before our understanding of the relevant evolutionary theory. The Sabin vaccine strain was the result of rodent-adapted virus and differs from the neurovirulent Mahoney strain by 56 point mutations (in the consensus sequence), although only a small number of these mutations were needed for neurovirulence (Christodoulou *et al.*, 1990). One of the important neurovirulent mutations was within the RNA polymerase gene (Tardy-Panit *et al.*, 1993). However, the significance of this observation took many years to unravel and exploit. In time it became apparent that 3Dpol mutants could affect replication fidelity. One poliovirus point mutant, 3DG64S, was shown to have enhanced high-fidelity replication and that selective pressure could be designed to increase fidelity in RNA polymerase (Pfeiffer and Kirkegaard, 2005).

Another major development was the molecular identification of the poliovirus receptor and the subsequent creation of transgenic mice expressing this receptor, making them susceptible to poliovirus infection. One of these transgenic lines allowed mouse brain infections with neurovirulent versions of poliovirus (Crotty *et al.*, 2002), and has provided a very useful animal model that allowed the evaluation of viral fitness in the context of *in vivo*

pathogenesis. Although 3DG64S replicates well in culture (with lowered error rate), it was less pathogenic in this mouse model and competed poorly with 3D wild-type virus. It seemed that the decreased viral diversity was less able to generate the variation needed to get past bottlenecks due to multiple selective differences presented *in vivo* in tissues in the host, such as brain infection (Pfeiffer and Kirkegaard, 2006). This experimental system also makes clear the greater complexity of fitness *in vivo* relative to that typically measured in culture.

Thus it seems that *in vivo* there may not be one fitness but several that cannot be distinguished or individually measured. It is likely that various *in vivo* barriers require distinct fitness solutions that tend to create bottlenecks and that the diversity *per se* is essential to get past such bottlenecks. A population, not a clone or a consensus, appeared more fit as higher titer infections of 3DG64S also failed to be pathogenic. Thus, higher levels of a consensus virus are not equivalent to higher diversity.

The relationship between RNA polymerase structure, error rates, and ribavirin action is discussed by Cameron in Chapter 6 and has been the subject of numerous studies (Crotty *et al.*, 2001; Crotty and Andino, 2002; Vignuzzi *et al.*, 2005). Knowledge of the structure and catalytic mechanism of RNA polymerase function has allowed a greatly enhanced level of detail to be considered into what affects error rate (see Castro *et al.*, 2007; Korneeva and Cameron, 2007; Marcotte *et al.*, 2007). This has provided insight into the likely action of ribavarin on product fidelity (Harki *et al.*, 2002). Thus, it appeared that even a mutant of RNA polymerase with increased fidelity could still generate elevated diversity by various methods. Such control of fidelity allowed for the design of control experiments in which the same consensus virus genome could be forced to generate either less or more diverse progeny populations. In no other virus–host system have we attained such detailed insight into issues of error rate as those that were put to such excellent use in the poliovirus–mouse system.

How generally important is this poliovirus *in vivo* quasispecies result? Although the poliovirus–mouse system provides us with a firm experimental result, it seems likely that the generality of this relationship will be questioned by evolutionary biologists for several reasons. For one, this was observed in a lab constructed model system, which, it could be argued, is not an accurate representation of *in vivo* virus–host fitness. Also, as mentioned above, group selection is a process that will not readily be accepted as representative by the broader community. Is there any evidence that this result with poliovirus indeed represents a general virus–host evolutionary relationship in natural settings? As presented in Chapters 13–15, retroviruses and also human hepatitis virus C clearly exist as quasispecies populations that affect disease outcome. In the case of the retroviruses, viral populations show diversity that far exceeds that seen for other RNA viruses. In both HIV-1 and HCV there is clear circumstantial evidence for the importance of quasispecies for *in vivo* disease outcome, drug resistance, and fitness. In addition, with HCV, CNS infection may sometimes result, and such brain infections appear to be mediated by distinct quasispecies (Forton *et al.*, 2004; Forton *et al.*, 2006), reminiscent of the poliovirus mouse model.

Quasispecies memory, as mentioned above, also seems to be an important issue with regard to failure of antiretroviral therapy (Kijak *et al.*, 2002) and it appears that pol gene mutations could also be involved in this (Carobene *et al.*, 2004). Measurements of HIV quasispecies in individual patients indicates that multiple evolutionary patterns can be found in typical individual patients (Casado *et al.*, 2001), thus mixtures of HIV exist in patients (Bello *et al.*, 2004, 2005). And HIV-1 recombination is clearly contributing to diversity (Kijak and McCutchan, 2005). Thus, with both HIV-1 and HCV, their capacity to cause human disease is clearly associated with quasispecies compositions that affect fitness in complex ways. The poliovirus mouse system therefore appears to reflect quasispecies issues as observed in natural virus–host situations.

Consideration of retrovirus–host evolution introduces another large issue in evolution: genomic viruses. Unlike poliovirus and most

RNA viruses, retroviruses (e.g. non-lentivirus) have colonized the genomes of animal species in large numbers and represent a large fraction of these genomes. Genomic retroviruses are present in vast numbers, most of which are defective and mutant copies. In this genomic colonization they resemble the dsDNA viruses of prokaryotes (discussed below) that also colonize all prokaryotes although at a much lower numbers. The human genome has fewer than 26 000 genes, but appears to have 500 000 retroviral-related LTR elements. Some of these elements are intact and conserved (human ERVs (HERVs)) and this genomic population has some clear characteristics of a viral quasispecies. Such large amounts of genetic material have previously been dismissed simply as selfish or junk DNA of no fitness consequences to the host. However, given the importance of quasispecies mutant genomes for viral fitness and persistence, we might need to re-evaluate this dismissal. Retroviruses are clearly part of the human ancestry thus we should seek to understand, not dismiss their role in human evolution.

Evolution of High Fidelity

In contrast to the story above in which polio infection of mouse brain was dependent on the quasispecies resulting from lowered fidelity replication, a different relationship has been proposed for the nidoviruses. These are also positive single-stranded polycistronic RNA viruses (Gorbalenya *et al.*, 2006). This group of virus includes the coronaviruses (e.g. mouse hepatitis virus and SARS-associated coronavirus), which are the largest RNA viruses known (26–32 kb). It has been proposed that such large genomes have required the adaptation of a high-fidelity RNA polymerase in order to increase the error threshold and accommodate large RNA genomes. Based on the phylogenetics of this polymerase and other RNA-processing enzymes, this group of viruses appears to be monophyletic and it is thought that the acquisition of a high-fidelity RNA replicase was central to the origin of this lineage. This type of replicase is unique to RNA viruses. The monophyletic view stems from an analysis of a small set of conserved genes. Overall, however, these larger genomes have many other genes that show no similarities to related viruses. The origins and evolution of these more diverse and numerous genes cannot be currently traced.

This is an inherent problem in the analysis of virus evolution: a small selected set of hallmark genes with some similarity are assumed to trace an apparently linear (tree-based) viral lineage whereas the larger number of genes are not included and cannot be traced. If most of RNA virus evolution is indeed mediated by a mixed cloud of genomes, any role for mutant mixtures thus becomes obscure. But perhaps there is little else we can currently do given the lack of information. How might we explain the increased fidelity and genome size of the nidoviruses? Was there some change in viral adaptation in which quasispecies and generation of mixtures was no longer as important for adaptation? Did the need and selection for a larger genome override the use of error to generate adaptability as seen in poliovirus and HIV-1? If so, what selective pressures might have changed this seemingly basic feature? What do we know about the natural biology of these viruses, which might provide some insight into this?

Unfortunately, the natural distribution and gene functions of the nidoviruses are generally poorly understood. In terms of coronaviruses, numerous mammal and avian species can be infected and the virus will cause acute disease. In several of these acute infections, the virus involved seems to have recently been adapted to the new host from other, often unknown sources. With the recent emergence of the SARS virus and human infections, however, much greater attention has been focussed on trying to understand the origin and evolution of this virus. It has recently become clear that there indeed appears to exist an evolutionary stable source of this virus from which adaptation to humans was possible.

Various bat species have been found to support persistent asymptomatic infections by specific versions of SARS viruses (Tang *et al.*, 2006; Wang *et al.*, 2006; Vijaykrishna *et al.*,

2007). These studies also indicate that there appear to be three different and independent groups of SARS viruses in bats. In fact six novel coronaviruses were isolated from six different bat species showing an astonishing diversity in bats. Furthermore, phylogenetic analysis indicates that all bat coronaviruses appear to have descended from a common ancestor. Only one of these bat groups includes SARS and SARS-like coronaviruses that adapted to acute human infections. Thus, a prevalent and species-specific persistence of SARS viruses is found in particular geographical populations. Why is this relationship stable? Could the adaptation to a host-specific persistence-based basal life strategy provide some explanations for the evolution of the higher fidelity RNA replicase of these coronaviruses? As I have argued, persistent viral infection represents the majority of evolutionary stable viral lineages (Villarreal, 2006). However, we have almost no knowledge regarding how these bat SARS viruses persist and escape elimination by innate and adaptive immunity and what, if any, role the high-fidelity replicase (or other genes) have in this life strategy.

MHV—Mouse Exemplar (A Case for Persistence and Virus Addiction)

Although we cannot yet evaluate natural SARS virus persistence in native bat hosts, another coronavirus may be more informative regarding the effects of persistence on host populations. Mouse hepatitis virus (MHV) may provide our best exemplar of virus–host relationships and show how the concept of virus addiction relates to population persistence. MHV is the best-studied coronavirus. As a natural and prevalent virus of rodents, MHV is our best natural model of persistent RNA virus–host relationships for any mammal. In general, rodents are the most studied non-domestic mammals with regard to natural virus distribution.

Overall, we know that wild-caught rodents seldom show signs of acute virus infection (Kashuba *et al.*, 2005). However, asymptomatic virus persistence is ubiquitous in wild rodents (Descoteaux *et al.*, 1977; Gannon and Carthew, 1980; Schoondermark-van de Ven *et al.*, 2006), including voles (Descoteaux and Mihok, 1986). Some field studies have evaluated broader patterns of virus persistence in mice (Singleton *et al.*, 1993; Becker *et al.*, 2007) which indicated that wild house mice are highly colonized with MHV (80–100% prevalence). In addition to MHV, mouse cytomegalovirus, mouse parvovirus, mouse thymic virus, and mouse adenovirus are also prevalent. Other well-studied mouse viruses, such as lymphocytic choriomeningitis virus (LCMV) and polyomavirus (PyV), were at low natural prevalence. Interestingly, some non-native house mice that have colonized isolated islands may lack MHV (Moro *et al.*, 1999), although most other isolated island populations retain MHV (Moro *et al.*, 2003).

Other small mammals have yet to show any viral disease whatsoever (hedgehogs, chinchillas, prairie dogs, gerbils, sugar gliders) (Kashuba *et al.*, 2005). Thus, asymptomatic persistent viral infection is clearly the norm in rodents. Yet, in spite of this usual asymptomatic viral persistence, historically, some zoonotic viral disease outbreaks have occasionally been documented in natural populations. One such early outbreak was an epizootic diarrhea that occurred in infant mice (Adams and Kraft, 1963). Later, it was established that one such infection was due to mouse hepatitis virus (Carthew, 1977; Ishida *et al.*, 1978). In spite of this disease outbreak, with MHV, it has since become clear that asymptomatic persistent infections are the norm and are highly stable. Yet MHV disease outbreaks, especially in virus-free mouse facilities, are also common and severe. How does MHV attain such stable and prevalent persistence in natural population yet retain the ability to cause disease in naive populations? What maintains the MHV fitness of natural persistence?

It is well known that once MHV is established in a mouse or rat colony it can be very difficult to eliminate (Gannon and Carthew, 1980; Lussier and Descoteaux, 1986), clearly

indicating that stability is rapidly attained and likely genetically programmed by the virus. I propose that these stable evolutionary states of viral persistence are due to a strategy we can call virus addiction (Villarreal, 2005) and that MHV can provide the exemplar of such a state. With MHV, only persistently infected mice colonies are protected from the disease that is otherwise caused by the virus. In wild asymptomatic mice, MHV is found mostly as an enteric infection. The CNS demyelinated disease that MHV can induce is most observed in newborn pups (Homberger, 1997; Nash *et al.*, 2001) and once in the brain, MHV can persist in CNS with recurring disease (Marten *et al.*, 2001). This recurring CNS disease is also associated with quasispecies (in the S gene) and recombination (Rowe *et al.*, 1998). The most serious CNS disease is in S-gene variant of MHV-4 (JHM), thus as with the polio–mouse model, pathogenic fitness with MHV is also associated with quasispecies.

Such MHV disease is the bane of all mouse colonies (Knobler *et al.*, 1982). However, once MHV persistence is attained, the problem to a mouse facility is not due to acute disease, but because immunological measurements are significantly affected by MHV persistence. Thus MHV alters mouse molecular identity regarding immunological (T-cell) reactions (Wilberz *et al.*, 1991). To establish stable asymptomatic persistence, however, MHV needs to infect newborns (Weir *et al.*, 1987), in which acute disease is prevented due to maternal passive immune antibody transfer (Gustafsson *et al.*, 1996). Being born to immune mothers thus protects against CNS disease and promotes enteric (not brain) virus colonization. In addition, it appears that persistence also promotes cross-species transfer (Baric *et al.*, 1999).

MHV persistence may involve genome stability and result in a distinct evolutionary dynamic. Asymptomatic persisting infections in a Lewis rat, for example, showed no variation in MHV S gene sequence, and no quasispecies as seen in brain infections (Stuhler *et al.*, 1997). The need to establish stable persistence could then be providing a strong selection for increased genome complexity and stability and might better explain the selection for the enhanced RNA polymerase fidelity in nidoviruses. How might such selection operate in natural populations? Evolutionary biologists often consider what might differentiate one group from another very similar group in a way that leads to two isolated and distinct populations.

Consider two hypothetical adjacent hay stacks harboring two *Mus musculus* colonies, one of which is persistently infected with MHV the other which is not. What is the fitness consequence to the colony harboring MHV relative to its uninfected neighbor? Our experience with MHV in mouse breeding colony provides a clear answer. The colony that is persistently infected with MHV will have a distinct advantage over its neighbor as MHV introduced into this uninfected colony will have severe effects on the offspring. Eventually, we can expect only the MHV-harboring colony will prevail in both hay stacks. This is a state I have called virus addiction. Only mice harboring persistent MHV are protected against the potential pathogenic consequence of acute MHV (or related virus) infection. The population is addicted to the virus.

Such a state, however, is clearly affecting colonies (or groups) of host, not individuals. An individual either quickly succumbs to the virus infection or, if infected, transmits it to others in the colony. A colony is thus under selection by MHV. To generalize this state, we expect that the persistence of SARS in specific bat populations would be expected to also affect the fitness of the corresponding specific bat populations. Persistence is a more demanding phenotype than acute replication. It requires greater gene complexity to counter host immunity and also to promote self-regulation. Thus the enhanced fidelity of RNA replication is selected in order to conserve this greater genetic complexity and stability.

We know that the high-fidelity RNA replication system (including RNA pol, helicase, endoribonuclease, and other activities) is also present in an ancient nidovirus relative of coronaviruses, such as fish-isolated white bream virus (26 kb RNA). I suggest there will also likely be species-specific persistent

infections with this virus that require this enhanced replication fidelity and maintain this virus in its natural habitat. Thus, I suggest, an ancient persistent life strategy could more easily explain the monophyletic character of the nidovirus virus lineage. It is particularly interesting that one of these unique and conserved replication proteins (ADP ribose-1-monophosphate) is dispensable for culture growth (Putics *et al.*, 2005). I suggest it will not be dispensable for persistence.

THE REAL WORLD OF VIRAL RNA IN HUMAN DISEASE

HIV-1

The HIV-1 pandemic is an unfinished story. HIV-1 represents a real-time biological event in human evolution that confirms for us the importance of quasispecies and retroviruses to human biology. However, even though its human toll is huge, modern medicine and culture has responded rapidly enough to limit the impact of HIV-1 to the point at which it will not likely be the cause of a selective evolutionary sweep that could have altered human genetic makeup (in contrast to the koala bear endogenization presented below). As described earlier, its amazing adaptability via quasispecies along with extensive recombination contribute directly to HIV-1's diversity (Charpentier *et al.*, 2006) and makes it the most dynamic genetic entity ever studied. Many studies track the dominant HIV population and fail to examine minority populations. Yet it is precisely these minority populations, which evolve independently of the majority population, that can determine drug resistance phenotype and biological outcome (Charpentier *et al.*, 2004; Briones *et al.*, 2006; Morand-Joubert *et al.*, 2006).

Clearly, the specific makeup of a complex HIV population matters. Furthermore, HIV defectives and variants can also have major consequences. In some cases, long-term non-progressors of HIV-1 have shown mixed populations and unusual polymorphism in the early phase of HIV infection, sometimes contributing to long-term non-progression (LTNP) (Alexander *et al.*, 2000). One population of LTNPs was reported to have been colonized by an HIV variant that showed low virus replication and slow or arrested evolution (Bello *et al.*, 2005). In another case, a stable non-progressor was colonized by a replication incompetent version of HIV-1 (Wang *et al.*, 2003). Some of these non-progressors also appear to resist super-infection (Zhu *et al.*, 2003). It seems clear that at least in these exceptional situations, non-majority HIVs are crucial to the outcome.

There is also reason to think that other retroviruses have had a major influence on recent primate and human evolution, such as apathogenic persisting foamy virus in primates (Switzer *et al.*, 2005; Murray and Linial, 2006). Human antiretroviral genes seem to have undergone recent adaptations, such as APOBEC3, which can interfere with exogenous retroviruses (such as MLV and SIV) and underwent an expansion in the hominid lineage (Esnault *et al.*, 2005). It thus seems clear that human and primate evolution has been significantly affected by earlier, prevalent primate retroviruses.

HCV

Another important human–virus quasispecies story that has long been recognized is with hepatitis C virus (HCV), (see Chapter 15; Domingo and Gomez, 2007). HCV seems to have adapted to humans in the recent past, possibly from asymptomatic enteric primate viruses currently found in Africa (Smith *et al.*, 1997). As HCV remains an infection predominantly transmitted by blood, it does not appear to have fully adapted to the tissues of and transmission within its human host. However, like HIV-1, HCV has long been recognized to generate quasispecies in chronically infected people (Martell *et al.*, 1992) and it soon became apparent that the viral quasispecies are affected by and affects the outcome of antiviral therapy (Enomoto *et al.*, 1994; Hohne *et al.*, 1994; Kurosaki *et al.*, 1994; Okamoto and Mishiro, 1994).

Thus, successful antiviral therapy is directly correlated with an initial dramatic reduction in genetic diversity. Unfortunately, it has become clear that only a minority of HCV-infected individuals will respond favorably to a combination of interferon and ribivarin. Thus it seems to be diversity *per se* and the resulting structure of an HCV quasispecies that has a direct consequence to human health. However, since HCV is less well-adapted to humans compared with HIV-1, it does not pose the same threat to potentially provoke an evolutionary event in human evolution.

VSV

VSV is a negative-stranded RNA virus that has been a very important experimental model and has provided many laboratory measurements regarding quasispecies theory (see Chapter 4). Using VSV, evidence supporting the Red Queen hypothesis, involving unending adaptation to greater competition and Mueller's ratchet has been presented (Clarke *et al.*, 1994; Novella *et al.*, 1995; Elena *et al.*, 1996). When VSV was evaluated as an arbovirus, requiring adaptation to alternating and opposing fitness of insect and mammalian host, it was also apparent that minority quasispecies populations were responsible for maintaining the apparently antagonistic phenotypes (Novella *et al.*, 1999). Thus here too, the consortia character of a quasispecies is clear. Yet in natural settings several very different virus–host relationships can be seen with rhabdoviruses.

A distant relative of VSV (VHSV) is also known to be responsible for mass die-off of commercially important fish (Marty *et al.*, 2003). This virus infects many teleost species and has shown 100% mortality in many experiments (i.e. with i.p. inoculation). In natural outbreaks, however, it has also shown surprising genetic stability (Einer-Jensen *et al.*, 2006). Clearly error-prone rhabdovirus replication must be kept in check by purifying selection in this situation. In contrast, another rhabdovirus, sigma virus of *Drosophila*, is associated with no mortality but is a vertically transmitted persisting virus in specific *Drosophila* populations (Fleuriet, 1996). Yet in some recent population measurements, sigma virus infected *Drosophila* are expanding for unknown reasons (Fleuriet, 1994). Clearly this particular virus–host persistent relationship has some undefined selective advantage that operates beyond the lab-based concepts as measured above. Other rhabdoviruses also have peculiar host-specific relationships, such as bats that tend to support many persistent infections (Badrane and Tordo, 2001; Li *et al.*, 2005), or birds that seem to be free of almost all rhabdoviruses. Clearly, although VSV lab results have been highly informative, we still have much to learn regarding natural settings that affect rhabdovirus adaptation and evolution.

Another major paradigm for the high rates of negative-strand virus evolution is found with influenza virus. Due to its history and potential for initiating great human epidemics, it has long held the special interest of evolutionary virologists (see Chapter 5; Nelson and Holmes, 2007). However, this research has not much emphasized the quasispecies character of influenza virus evolution. Instead, it concentrates on the evolution of the master template or clades of template for the purposes of vaccine development (Webster and Govorkova, 2006). The views stemming from this type of evolution have lent themselves well to master template-based phylogenetic analysis and have dominated how many researchers think of virus–host evolution. Thus it is curious, given the above emphasis, that the quasispecies character of influenza populations often seems of low relevance to issues of acute disease and vaccination, other then to provide a source of diversity. In some situations, viral competitive interference may contribute to drift variation and displacement in antigenic epitopes (Levin *et al.*, 2004). Yet outcomes of individual human and bird infections do not seem much affected by specific quasispecies structures, as we saw with HIV-1 and HCV.

With influenza, we are mainly concerned with epidemic human disease. However, by shear numbers of infections and deaths worldwide, it must be admitted that influenza virus is really a virus that affects mostly birds. For example, during the 2005 outbreak

in China, only 251 humans died whereas 230 million domestic birds died (Smith *et al.*, 2006). Although our concern on the large potential for human disease is understandable, these numbers should inform us of a more basic virus–host biology. In this case, influenza shows a high affinity for various birds; migratory water birds in particular can have high prevalence (Wallensten *et al.*, 2006). Some waterfowl, such as wild mallard ducks, have been called the stealth (asymptomatic) carriers of influenza H5N1 and free grazing ducks seem to introduce virus into domestic bird populations (Gilbert *et al.*, 2006). Thus waterfowl represent the well-accepted epidemiological concept of a reservoir species (Louz *et al.*, 2005). But these wild waterfowl, shorebirds, and gulls that are a natural host for avian influenza also seem to show a much slower rate of evolution (Spackman *et al.*, 2005). In contrast, the much higher rate of evolution as seen in chickens and turkeys indicates that these hosts should not be considered as natural reservoirs (Suarez, 2000). In waterfowl, influenza infections show several distinctions, such as virus co-infection or virus interference (Sharp *et al.*, 1997) as well as phylogenetically distinguishable waterfowl dendograms, including specific M lineages (Makarova *et al.*, 1998; Widjaja *et al.*, 2004). The diverse and stable avian pool of influenza virus appears to be ancestral to the influenza viruses that infected human populations.

THE ANALYTICAL PROBLEM OF QUASISPECIES: GROUP SELECTION, RETICULATION, AND RECOMBINATION

The Good

The phylogenetic methods that have been adapted from evolutionary biology have been tremendously helpful and have allowed us to trace the seemingly untraceable, virus evolution (see Chapter 5). Thus, we have often been able to make informed judgments concerning broader patterns of virus evolution and this has become the major tool for the current study of virus evolution, such as influenza virus (Nelson and Holmes, 2007). Influenza A, for example can be seen to show extended periods of stasis followed by periods of rapid adaptation that necessitates adaptations in vaccine strategy (Wolf *et al.*, 2006). However, the evolutionary variation between seemingly similar viruses can be surprisingly large (see above VSV section). For example, the very different phylogenetic behaviors between influenza A and measles virus, both acute human respiratory infections due to membrane-bound negative-stranded RNA viruses, are striking. The reasons for the maintained genetic stability of measles virus remain poorly understood, but may well involve more complex fitness associated with systemic infections.

Phylogenetic methods can also be highly informative regarding the likely origins of viral lineages and possible sources of emergence. For example, the studies of dengue virus by Holmes and colleagues suggest that this virus first entered its human host about 1000 years ago, and that sylvatic (African jungle) asymptomatic infection of primates may have provided the origin of this virus that later became a human pathogen (Holmes and Twiddy, 2003; Holmes, 2006). Such insight provides valuable clues concerning the likely selective pressures that may lead to the emergence of dengue virus. Phylogenetic methods are also highly informative regarding classification and taxonomy relationships and have allowed us to understand viral relationships across broad species definitions (Zanotto *et al.*, 1996).

The Bad

However, phylogenetic approaches necessarily assume the master template is the fittest type and that mutations or variants in the RNA populations are a source of genetic load that are deleterious and limiting to virus adaptation (Pybus *et al.*, 2007). Such variation is mostly due to "unfit" mutations, which indicates that a viral cloud is mostly and unfit consortia. It would seem that such conclusions go against the concept of quasispecies as being fit *per se* as described above. In this consideration we see a major weakness of extant

phylogenetic methods. They were not developed to access the evolutionary relationship and fitness of interacting mixtures. Nor were they designed to follow the evolution of systems with high rates of recombination between numerous parental templates. We currently lack the analytical tools for such a population analysis. Without such tools, however, it seems we can only evaluate those parameters we can define and will remain confused by those we cannot. Evolution of a consortia thus provides a new directions for theoretical and laboratory research. We should seek to investigate the mixture, not just its average.

Plants

Another major virus–host system that has been highly studied is the viruses of agricultural plants. Our understanding of plant viruses has also been highly influenced by disease associated with agricultural domestic species, thus natural virus–plant relationships are much less understood, although some recent field studies are starting to change this situation (see Chapter 12). We currently have a rather uneven understanding of broader virus–host relationships and evolution in plants. For example, viruses of the more ancient ferns, if they exist, are essentially unknown.

The prevalence and diversity of positive-stranded RNA viruses in plants is striking. In addition, we are starting to appreciate that virus–virus interactions are also frequently involved, although this issue remains poorly studied. One well-studied family of plant virus are the tobamoviruses of angiosperms (see Chapter 11; Gibbs, 1999). Progenitors of this virus family appear to also be found in algae and fungi consistent with a very long evolutionary history. Both high transmission between host and virus–host congruence are observed with these viruses. Virus–virus interactions also seem to be important. For example, tobacco mosaic virus (TMV) and tomato golden mosaic virus (TMGV) appear to have shown interactions in Australia which have apparently led to the extinction of TMV, but the retention of TMGV with no increase in genetic diversity (Fraile *et al.*, 1997).

Plant viruses have also been seen as quasispecies in some but not all settings (see Chapter 12; Roossinck, 2003; Roossinck and Schneider, 2006). Besides the interactions expected for typical viral quasispecies, plants often show evidence of more extensive mixed virus infections. There are, for instance, many examples of satellite viruses that must necessarily interact with other RNA viruses of plants. It is also clear that the subviral elements of even a single viral lineage can greatly affect the virus–host relationship. Such subviral elements (DIs) have been observed to both reduce and intensify disease, and also interact with satellite viruses (Qiu and Scholthof, 2001), thus virus–virus interactions are clearly crucial in many situations (Simon *et al.*, 2004) and viral interactions and synergism appear to have led to significant events in plant virus emergence (Fargette *et al.*, 2006).

Virus–virus interactions are not limited to plant RNA viruses. The ssDNA plant geminiviruses also display complex interactions with satellites as well as high diversity in field isolates of East Africa (Ndunguru *et al.*, 2005). Thus, plant viruses seem particularly prone to interactions. More recently, virus-mediated symbiosis with respect to host survival has been reported (Roossinck, 2005) (discussed below).

Phylogenetic methods also struggle to address the occurrence of high rates of recombination in viral lineages. Such a situation complicates the analysis, creating hard-to-define, reticulated trees, although these limitations can be partially overcome by using sliding windows for the analysis. Such approaches have allowed surveys of recombination in some viral lineages, such as with the plant potyviruses (Chare and Holmes, 2006). However, the rampant recombination and quasispecies generation of HIV-1 makes a quantitative assessment of the virus population problematic. One proposed solution is to use a composition vector method (Gao and Qi, 2007). The issue of measuring recombination and tracing evolution in large populations is especially a problem that applies to the DNA viruses (phage) of prokaryotes (see below).

THE BIG BANG OF BIOLOGY

Prokaryotic DNA Viruses, Mosaic Swarms, the Origin of DNA-Based Cells.

Our perception regarding the overall importance of DNA viruses of prokaryotes to the evolution of life on Earth has undergone a major shift in recent years. The main realization is that DNA phage are the numerically dominant genetic entity in most habitats on Earth (mentioned above). In addition, as discussed in Chapter 10, it is now clear that some of these viruses are surprisingly complex and that essentially the entire pool of dsDNA viruses of prokaryotes may be exchanging DNA via recombination at high rates. This would constitute by far the largest common gene pool on Earth. Historically, the evolution of the DNA viruses of prokaryotes has seldom been considered in the broader context of virus evolution or evolutionary biology. Although it has long been realized that there are many basic similarities between viruses of bacteria and eukaryotes (Luria *et al.*, 1959), not until structural studies solved the capsid genes of prokaryotic and eukaryotic viruses did the evolutionary relationships between these viruses become clear. In addition, there have been a number of striking proposals that suggest that DNA viruses of prokaryotes may be involved in the origin of several major systems used by cells and that viruses appear to be involved in several major transitions during host evolution. Thus we now consider the possibility that these DNA phage were fundamental to the origin and evolution of life on Earth.

The Big Bang of Phage and Cellular DNA

It now seems likely that some large DNA viruses infecting eubacteria, archaea, and eukaryotes share some common evolutionary histories. It also seems clear that such viruses can link all three domains of life. This realization was not apparent based on phylogenetic sequence conservation, which is absent. It stems from the structure and assembly of virion capsids in which T4 phage, halophage, and the herpesviruses all show clear similarity as well as similarity in replication strategies. In addition, phage PRD1 and adenoviruses show similar broad structural and strategic conservation. Some biochemical (DNA pol family) and genetic similarities (gene order, gene programming) are also apparent, which taken together supports the common origins of these viruses (Hendrix, 1999, 2002; Hendrix *et al.*, 1999, 2003). T4-like viruses in particular seem to represent a major source of global genetic diversity.

This giant genetic pool represents a huge potential to affect life (Filee *et al.*, 2005) and the viral genetic creativity represented by this pool would also be vast (Nolan *et al.*, 2006). Since T4-like phage that infect cyanobacteria also encode virus-specific type II photosynthetic core genes, viruses appear able to create the most complex of genes as well (Clokie *et al.*, 2006; Sullivan *et al.*, 2006).

As presented in Chapter 10, phage are now thought to evolve by distinct and highly mosaic "horizontal" processes of rampant recombination (Hendrix, 2002, 2003). Large DNA phage appear to be ancient, present before the split of the three main branches of cellular life: bacteria, archaea, and eukarya (Benson *et al.*, 2004). LUCA, the Last Universal Common Ancestor, would represent the putative cell ancestor prior to this split. However, phyogenetic analysis of common or conserved genes of LUCA identifies only about 325 or fewer genes in extant cellular genomes (Mushegian, 1999; Koonin *et al.*, 2001; Mirkin *et al.*, 2003). Ironically, the genes needed for DNA replication are not part of this conserved set, calling into question the nature of the first DNA-based cell. Large-scale "horizontal" transfer seems to have clearly prevailed early in the evolution of DNA-based cellular life and it has recently been asserted that LUCA existed in a highly horizontal "consortia" of cooperative genes that developed the common genetic code (Vetsigian *et al.*, 2006).

Since the DNA replication proteins in the extant three domains of life have distinct

compositions, it has been proposed by Forterre that DNA viruses and retroviruses were directly involved in the invention of the three extant cellular DNA replication systems (Forterre *et al.*, 2005). According to this view, early cellular life was completely entangled with viral (phage) lineages; hence cells must have evolved from an ancestral "virus"-mediated population not a single genetic lineage. Thus the evolution of early life would have clear similarity to the quasispecies (consortia) state of genetic information as seen in RNA viruses above. Thus the huge creative and adaptive potential of virus would have been directly involved in the very earliest evolution of life. Clearly, such conjectures regarding the most ancient events in the evolution of life are hard to substantiate. But, these theories are as viable as any other and deserve serious consideration.

In spite of this seemingly unending mosaic exchange in dsDNA phage, some phage isolates show surprisingly stable genetic makeup. We now accept that T4-related phage are an important source of the larger global phage genetic diversity (Liu *et al.*, 2006) and that most such viral genes are novel (Filee *et al.*, 2006b; Nolan *et al.*, 2006). Yet even with T4-like viruses, there can be clear barriers to horizontal gene transfer which promote the evolution of stable viral lineages (Filee *et al.*, 2006a). In T4-type phage, 24 similar core genes could be seen in all genomes, which seem to be inherited in gene blocks that preclude recombination. However, these blocks were not seen in the broader T-even and pseudo T-even genomes. Other phage also show surprising genetic stability when repeatedly isolated from similar habitats, such as soil phages of *Burkholderia* (Summer *et al.*, 2006) and Bam35 (Saren *et al.*, 2005) as well as some hot spring isolates (Khayat *et al.*, 2005). This Bam35 capsid also identifies another structural motif mentioned above that is broadly conserved in evolution and shows clear similarity to that capsids found in PRD1 and PBCV-1 (discussed below). SH1 also has a clear PRD1-related capsid, membrane, and genome; thus this halophilic euryarchaeon virus, although showing no sequence similarity to PRD1 or any other bacterial phage, is clearly structurally related (Bamford *et al.*, 2005a).

It is interesting that overall the viruses of hyperthermophilic Crenarchaeota generally show no sequence relationship to phage of bacteria. In addition, the use of the term phage for these viruses can also be questioned as most establish non-lytic chronic infections. Many of these Crenarchaeota viruses have unique morphologies not found in any other domain of life (Prangishvili *et al.*, 2006a, 2006b; Ortmann *et al.*, 2006). Some, however, have clear structural and genetic similarity to specific phage (i.e. T4).

GENOMIC STABILITY: PERSISTENCE AND TEMPERATE LIFESTYLE

Considerations of phage evolution and rampant recombination (especially with T4 and T-even phage) often emphasize the viral lytic lifestyle and host death. In fact this lytic relationship was argued by many early phage researchers to be the fundamental and only character of phage–host relationships in general. We now know, however, that persisting (temperate) phage are also common, some of which have no independent lytic phase. The fundamental model of phage persistence by unique integration into host chromosomes (temperate lysogeny) marks a major development in our understanding of molecular virology and virus–host relationships which was first clarified by Campbell in 1962 (see Campbell, 2007). All free-living prokaryotes show the presence of colonized phage in their genomes. Both complete and defective genomes of dsDNA viruses have been observed in the sequenced DNA of all free living prokaryotic genomes (Gelfand and Koonin, 1997) (exceptions are some intracellular parasites and plastids). Thus, the massive genetic diversity and novelty of phage evolution as presented above has a direct conduit into the genetic composition of all prokaryotes via lysogeny.

The fitness and evolutionary consequences of such colonization to the evolution of the host and its virus should be considerable but

is in need of theoretical development. Fitness of temperate phage, however, is more complicated then that of a lytic virus and, like fitness of persistence discussed above, cannot be simply described by relative replication or efficient virion production. Here too, successful phage colonization must inherently limit the replication of the same virus. Thus, a temperate lifestyle also requires an autoinhibitory capacity. This generally involves an immunity gene set that not only limits self-replication but can also affect replication of other temperate and lytic viruses, i.e. lambda (even as a defective) precludes T4 and other T-even phage. Uncolonized hosts are thus susceptible to lysis by highly prevalent acute tailed phage. Host fitness is thus strongly affected by a temperate phage due to its ability to preclude and survive other competing phage. I suggest this situation is similar to the MHV–mouse exemplar above, in that virus-colonized hosts are in a state of "virus addiction" in which persistence is needed to provide protection from the same or similar virus (Villarreal, 2005, 2006).

It is well established that most natural populations of bacteria have specific patterns of phage colonization, hence the utility of phage typing for strain identification. From this, we can infer that virus–virus competition is a prevalent and major issue regarding the prokaryotic fitness resulting from a symbiotic temperate phage–host combination. In addition, such virus–host symbiosis can also affect competition with other bacteria. This would be very much like the virus addiction concept outlined above for the MHV examplar.

The original observation of a lysogenic process and coining of this term occurred in the 1920s when two pure cultures of bacteria were grown together. It was observed that in some combinations, one strain would lyse the other strain (was lysogenic). Later, it became clear that such lysis was mediated by reactivation of temperate phage present in the lysogenic strain, but absent from the non-lysogenic susceptible strain. In this relationship, we see another example of group selection operating on bacterial populations harboring a persistent virus. Thus, what host is fit depends very much on the prevalent viruses it will encounter as well as the viruses that colonize it. Bacterial populations that are colonized by the same or similar phage express the appropriate immunity functions and are protected from lysis by the same or similar phage.

Such a situation has significant implication for the evolution of immunity and group identity for cells. Host stability becomes a major fitness issue for a persistent virus life strategy. It is generally thought that a temperate virus attains a stable colonization of its host by simply integrating into and become one with the host genome. However, there are also clear examples of stable phage persistence that does not integrate and uses other strategies to attain host stability (similar to eukaryotic DNA viruses; see below for the P1 phage exemplar of this). Like a temperate phage, a host that is colonized by episomal persisting viruses has also been much affected in its evolutionary potential.

EPISOMAL STABILITY: THE P1 EXEMPLAR OF PERSISTENCE

It is clear that phage can have complex effects on host populations, but these phage themselves often exist in complex and mixed states that can be difficult to unravel (Harcombe and Bull, 2005). It has been known for some time that the presence of otherwise silent phage can greatly affect the growth of other virus and susceptibility of host. One such silent and common phage that has long been studied is P1. P1 was initially discovered due to its effect on T4 and lambda. However, P1 has been a very interesting model, not because it causes disease or offers potential therapy against bacterial pathogens, but simply because it persists efficiently as an episome and competes effectively with many other phage (Yarmolinsky, 2004). Since it does so without integrating, P1 provides us with one of the only well-studied models that can inform us regarding the molecular strategies and details of how stability in non-genomic persistence is attained.

Curiously, a main strategy by which P1 attains this stability was inapparent and not

suspected after several decades of study. It became apparent only after replication mutations were made that induced self-destruction and uncovered the existence of what came to be called "addiction modules" (Lehnherr *et al.*, 1993). P1 encodes several gene pairs (toxins/antitoxins, such as the PhD/DOC pair) that protect bacteria harboring P1, but kill daughter bacteria that have lost the P1 genome (Gazit and Sauer, 1999). This strategy compels colonized *E. coli* to maintain P1 or die (DOC, death on curing). However, these very same addiction systems are also involved in protecting a P1-colonized colony from T4 and lambda infection and will also induce self-destruction when cells are infected by those viruses, protecting the colony (population). P1 also provides an exquisite level of molecular self-identification in that it will recognize a single second copy of its own genome (Yarmolinsky, 2000).

What then is the fitness and evolutionary consequence to *E. coli* harboring P1? Clearly it is major, but mostly host fitness is affected relative to other viruses. Accordingly, when contemplating the amazing complexity of the P1 immunity and how it evolved, Yarmolinsky posed the question; "Could the byzantine complexity of the controls at ImmI be the outcome, not of successive host–parasite accommodations, but of competition among related phages?" (Yarmolinsky, 2004). If we answer yes to this question, then we would also conclude that virus–virus interactions and competition in general are major forces in the adaptability and evolution of persisting phage and surviving colonized host. In this light, viral persistence takes on a major role in virus and host evolution. The P1 exemplar has thus provided us the concept of viral addiction that also promotes host group selection.

PROKARYOTES AND THEIR VIRUSES AS ONE EVOLUTIONARY POOL

Historically, we are biased to think of viruses (and phage) as agents that simply kill their host. Some have proposed that the prokaryotic global biomass is phage partitioned into those populations that live and those that die due to viral lysis. From such a perspective, viral novelity would seen of little relevance to host evolution. Metagenomic projects as noted above, have sequenced nearly 2 million phage genomes and report that most of these phage genes are unique, not in the database, and likely not derived from host (Edwards and Rohwer, 2005). The protein repertoire of sequenced phage indicates that 80% of conserved phage genes are specific to phage and show an evolutionary independence from genes of host (Liu *et al.*, 2006). This identifies a massive genetic novelty from virus, which is especially apparent in large DNA phage. As just discussed above, however, those hosts that live are also products of phage selection, and persisting temperate phage play a major role in this. Such phage colonization allows this massive phage novelty to find its way into host genomes, which allows viral complex gene sets to be applied to novel problems of host adaptation. Host novelty can thus be introduced by phage (Comeau and Krisch, 2005).

That persistence is a major life strategy of phage is confirmed by the large numbers of genes associated with persistence (i.e. integrases, immunity) observed in metagenomic screens. There is also much practical experience that supports the crucial role of prophage in host evolution. One particularly well-studied system that has been studied for over 50 years is the ongoing evaluation of phage evolution as observed in the dairy industry (Canchaya *et al.*, 2003, 2004; Brussow *et al.*, 2004). The temperate phage analysis of these bacteria follows a long tradition of lambda and *E. coli* studies (Campbell *et al.*, 1992; Canchaya *et al.*, 2003, 2004). Since lytic phage can severely disrupt dairy fermentation, it was of particular interest to understand and trace their evolution. These studies have led Brussow to conclude that much of the more recent dairy bacteria evolution can be considered to have resulted from the action of temperate phage.

A similar view applies to *E. coli* and cyanobacteria. In addition, the ECOR collection of 72 sequenced *E. coli* genomes of medical

interest shows that they differ from each other mainly due to patterns of genetic colonization, mostly by prophage, but they also show the presence tRNA-adjacent defective prophage and plasmid elements that differentiate these strains (Hurtado and Rodriguez-Valera, 1999; Mazel *et al.*, 2000; Nilsson *et al.*, 2004). Cyanobacteria (*Prochlorococcus*) is major model for the study of the origin of the type II (plant-like) photosynthetic system. Since such genes show much evidence of recent and massive horizontal movement, it seem quite likely that prophage are mediators of such transfers, especially as these phage encode their own version of these photosynthetic genes (Lindell *et al.*, 2004; Sullivan *et al.*, 2006).

Very similar *Prochlorococcus* strains exist in distinct oceanic populations in various habitats known as ecotypes. Some think that such ecotypes represent the initial type of genetic variation that leads to speciation. The sequencing of six ecotypes has shown that they are 99% similar to one another, but the genetic variation that distinguishes them is mostly due to patterns of prophage colonization (called phage islands) (Bouman *et al.*, 2006; Coleman *et al.*, 2006).

Thus in all these prokaryotic models, persisting viruses play a fundamental role in host evolution and host genetic novelty is mostly phage derived. Such observations have led some to propose that "war is peace" regarding virus–host evolution (Comeau and Krisch, 2005). Massive and complex innovation by phage appears to be a major force in the prokaryotic world.

Prokaryotes are the most adaptable of all cells. If we can accept the above conclusion concerning the role for viruses in the evolution of prokaryotes, we must then ask why such a successful evolutionary strategy was not apparently maintained in eukaryotes? In eukaryotes we see little evidence that large-scale integration by DNA viruses is an important evolutionary process (although the story with retroviruses is different). Why should prokaryotes and eukaryotes differ is such a fundamental way? Nevertheless, as noted at the start of this section, we do see good evidence that links the evolution of large DNA viruses of prokaryotes to the large DNA viruses of eukaryotes.

DNA Quasispecies

In case we were becoming comfortable with the apparently clear distinctions between RNA and DNA virus evolution as outlined above (quasispecies vs. domain recombination respectively), the evolution of the parvoviruses informs us that DNA viruses can also evolve by a quasispecies process. Parvovirus evolution (see Chapter 17) can show a sharp contrast to the evolutionary pattern displayed by other small dsDNA viruses above (HPV, Py). With the emergence of an acute pandemic in domestic dogs and cats (as well as other wild carnivore species), we see what is essentially evolution driven by single point mutations, mostly affecting the capsid genes and host cell receptor binding. This system provides us with one of the better studied examples of the evolutionary dynamics of an emergent viral disease. In addition, *in vivo* mouse studies with minute virus of mouse (MVM) now make it clear that parvoviruses can behave much like RNA viruses, generating quasispecies of diverse progeny that allow a high adaptability for the generation of fitness and disease *in vivo* (Lopez-Bueno *et al.*, 2006). This story is very reminiscent of the study of poliovirus in mice mentioned above. Human studies with B19 parvovirus are also consistent with high mutation rates (Parsyan *et al.*, 2007; Shackelton and Holmes, 2006).

THE TRANSITION TO EUKARYOTIC DNA VIRUSES

The Phycodnavirus Exemplar

Although not specifically addressed in this volume, the viruses of eukaryotic unicellular green algae are of special interest from the perspective of DNA virus evolution. These large, complex dsDNA membrane-containing icosahedral viruses are abundant in some water habitats (Van Etten, 2003; Ghedin and Claverie,

2005). The reason they deserve special attention is that they clearly have many features that are characteristic of both prokaryotic and eukaryotic viruses. They resemble prokaryotic viruses in that their life cycle is clearly phage-like, such as external virion attachment, injection of DNA and no pinocytosis. In addition, they also encode many phage-like genes, such as restriction-modification enzymes and homing endonucleases (Filee *et al.*, 2006c). They also resemble eukaryotic viruses in that they have eukaryotic DNA replication proteins (DNA polymerase beta and PCNA; Chen and Suttle, 1996; Nagasaki *et al.*, 2005; Villarreal and DeFilippis, 2000) as well as many genes associated with eukaryotic signal transduction (Van Etten *et al.*, 2002). Thus they represent a clear link between prokaryotic and eukaryotic DNA viruses. For example, the DNA polymerase of paramecium bursaria chlorella virus (PBCV-1) is the most conserved gene and most closely resembles that found in human herpesvirus and is distantly related to the similar family DNA pol encoded by T4. This polymerase is distinct from that of the poxviruses or PRD1/adenoviruses (associated with protein-primed DNA replication). However, numerous other genes of the phycodnaviruses are similar to some genes found in the mimiviruses (giant DNA virus of ameba), including the presence of conserved intenes in the DNA pol gene (Ogata *et al.*, 2005).

In view of this it is most curious that in structural similarity, polydnavirus capsids clearly resemble PRD1 capsid (Khayat *et al.*, 2005; Nandhagopal *et al.*, 2002). PRD1 contains the double-barrel trimer capsid structure that was first observed in adenovirus (for references see Saren *et al.*, 2005). Adenovirus also closely resembles PRD1 in DNA replication strategy (i.e. linear DNA with covalently closed ends (Benson *et al.*, 2004; Khayat *et al.*, 2005). The lineage of adenovirus-like DNA viruses, however, is thought to be distinct from that herpes and poxviruses and its DNA polymerase is clearly distinct from polyndavirus. It is clear that related elements of all these viruses can be found in phycodnaviruses. Overall, the phycodnaviruses, like phage, also appear to be creating genes in large numbers and they encode many genes unrelated to their host. What then is the evolutionary relationship that links all of these seemingly distinct viruses?

As outlined above, the pattern of evolution of dsDNA phage involves lots of exchange by recombination from a vast gene pool. This pool resembles a cloud from which various mosaic subelements and substrategies are assembled to allow viral gene acquisition and novelty (Blum *et al.*, 2001; Benson *et al.*, 2004). Does such a distributed pattern of evolution and gene novelty also apply to the phycodnaviruses?

Recently, another distinct phycodnavirus has been sequenced: coccolithovirus (EhV-86) (Allen *et al.*, 2006a, 2006b) conserves only 24 core genes in common with PBCV-1 and is unique to the phycodnaviruses in that it has acquired six DNAdep RNA polymerase subunit genes, which are absent in all other phycodnaviruses.

As RNA polymerase is considered a core viral gene function, it is clear that phycodnaviruses can alter some very basic molecular functions during their evolution. Oceanic phycodnaviruses are thought to have large influence on the free-living populations of eukaryotic algae, such as the termination of algal blooms reported for Emilian Huxley virus (Martinez *et al.*, 2007; Schroeder *et al.*, 2003). However, not all phycodnaviruses are lytic. Another lineage of phycodnaviruses is represented by two viruses of filamentous brown algae, EsV-1 and FirrV-1 (Delaroque *et al.*, 2003). Unlike the lytic phycodnaviruses noted above, these two viruses are "temperate phage" like. That is they exist as silent viruses whose DNA is integrated into the germlines of their host. In this, they are unique to all known eukaryotic DNA viruses; host chromosome integration is a normal part of their persistent life strategy. EsV-1 has a 335 593-bp genome and encodes 231 likely genes (Delaroque *et al.*, 2001). These genes are mostly unique and only 28 are clearly related to PBCV-1 genes. The gene differences include many replication genes and their gene order is completely different. Like the temperate phage–host evolutionary

relationship outlined above, it would be most interesting to understand how the integration of these large DNA viruses has affected host evolution.

Thus, the phycodnaviruses appear to represent a basal but diverse viral lineage that has both acute and persistent lifestyle and have some clear relationships to most large eukaryotic DNA viruses and many phage.

A Comment on the Proposed Monophylogeny of Nucleo-Cytoplasmic Large DNA Viruses (NCLDVs)

The phycodnavirus exemplar above should leave us with several impressions regarding the nature and evolution of these large and ubiquitous DNA viruses of algae, an early eukaryotic host. They show clear linkages by structure and function to both phage and various eukaryotic DNA viruses. They also show major variation and novelty in their own genetic composition, including their core genes. In addition, they show clear relationships to distinct and seemingly separate viral lineages (adenoviruses, herpesviruses, poxviruses, iridoviruses). The picture we are left with is that they seem to resemble phage evolution in that they appear to have evolved from a diverse pool that has exchanged many basic viral features and created many new genes.

This view, however, contrasts sharply with the work of Iyer *et al.* (2001, 2006). By considering the small number of conserved genes in four families of eukaryotic DNA viruses (poxviruses, asfarviruses, iridoviruses, phycodnaviruses), they suggest that these viruses are monophyletic, evolving from a common nucleo-cytoplasmic large DNA virus (NCLDV) with an icosahedral capsid.

Given the above information, I find this view unhelpful and possibly confusing. It has numerous problems. The main problem is that it fails to acknowledge the clear link between prokaryotic and eukaryotic viruses. Furthermore, by focussing on a small set of related genes, it represents a traditional perspective as found in evolutionary biology that assumes a common (fittest) linear lineage, not a cloud, cooperative, or mosaic pool as the main source of novelty resulting in the matrix pattern of virus evolution. The virosphere is clearly not disconnected from itself, but it is also clearly not a linear or tree-like evolutionary system as suggested above. We must learn to think of virus evolution in its own terms; fuzzy, mixed, reticulated, and cloud-like.

HERPESVIRUS; MOSTLY PERSISTING AND CO-SPECIATION

As mentioned in the phage section, there have been various publications that suggest a deep evolutionary relationship between the herpesviruses and dsDNA viruses of prokaryotes (Rice *et al.*, 2004; Khayat *et al.*, 2005; Duda *et al.*, 2006; Akita *et al.*, 2007). Such enormously distant relationships, however, cannot now be measured by any reliable metric. Although herpes-like viruses are found in invertebrates (such as ostreid herpesvirus 1 (OsHV-1)) in both lytic and asymptomatic states (Barbosa-Solomieu *et al.*, 2005), our interest in their evolution has been mainly focussed on the vertebrate herpesviruses. Vertebrate herpesvirus do tend to show clear sequence conservation that suggests broad patterns of evolution. One interesting feature of this evolution is the apparent link between the biology of the virus and its evolution. A common, but not universal pattern is that of virus and host co-evolution (McGeoch *et al.*, 2000, 2006; McGeoch and Gatherer, 2005). This trend has maintained several biological characteristics, such as highly species host- and tissue-specific persistence (i.e. neuronal and lymphoid persistence).

The discovery of HHV-8 has further stimulated studies of herpesvirus evolution in that HHV-8 appears to have undergone much recombination with herpesviruses of related primate lineages (McGeoch and Davison, 1999). Thus recombination seems prevalent in herpesviruses. The apparent link between herpesvirus evolution and recent human evolution, as well as an apparent link to primate retroviral evolution, is fascinating, but of

unknown significance (Kung and Wood, 1994; Lacoste *et al.*, 2000).

Herpesvirus Gene Acquisition

The herpesviruses lineages will often show the presence of lineage-specific genes. Many of these genes affect innate and adaptive host functions, whereas others affect host metabolism. When the source of such genes has been contemplated, in contrast to phage, phycodnaviruses, or baculoviruses (Herniou *et al.*, 2001), it is often proposed that most such herpes genes originate from the host. It is well accepted that the three major lineages of herpesviruses descended from a common ancestor in vertebrates (McGeoch *et al.*, 2006). There have been numerous proposals that most new lineage-specific herpesvirus genes have originated from host (see Becker and Darai, 2000). This includes herpesvirus dUTPase (Davison and Stow, 2005), and viral chemokines and viral Bcl-2 (Nicholas *et al.*, 1998). In my evaluation of such claims, however, it seems clear that the possibility that there was an ancient viral source of such genes was not considered and cannot now be dismissed.

We currently believe that ancient herpesvirus ancestors can be traced to tailed phage (Hendrix, 1999; Bamford, 2003; Baker *et al.*, 2005; Duda *et al.*, 2006; McGeoch *et al.*, 2006). Other phage lineages also appear to trace to eukaryotic viruses (Bamford *et al.*, 2005b). Within the herpesviruses, the same T-16 icosahedral structure, as well as invertable DNA regions are also present in the very distant but much more recognizable oceanic ostreid herpesvirus 1 (Davison *et al.*, 2005).

Given the highly diverse and mosaic nature of large DNA virus evolution in prokaryotes and lower eukaryotes described above, it seem quite possible that many other viral genes might also trace far back in virus evolution. Consider the example of dUTPase in avian and mammalian herpesvirus (Davison and Stow, 2005; McGeehan *et al.*, 2001). The current view requires very complicated gene rearrangements to account for the viral source of this gene from its host. Yet we know that diverse dUTPases are found in many ancient viral lineages. For example, the ERVs present in all vertebrate genomes also conserve dUTPase (Jern *et al.*, 2005), as do exogenous retroviruses (i.e. lentiviruses) (McIntosh and Haynes, 1996). In fact, since the herpesviruses genes are especially poor in introns, it would seem likely that any herpesviral gene acquisition would necessarily involve a retrovirus via a cDNA.

The oceans are especially filled with large complex DNA viruses (such as mimivirus and phycodnavirus, plus numerous relatives of OsHV-1) thought to be ancient ancestors of herpesvirus. The phycodnavirus (chlorella virus, PBCV-1) provides a clear bridge between phage and eukaryotic DNA viruses. PBCV-1 also encodes a dUTPase that has the highly conserved motif III (Zhang *et al.*, 2005). Many phage are also known to encode dUTPases of diverse types, such as *B. subtilis* (SPbeta) (Persson *et al.*, 2005), and a phage of *Thermus thermophilus* (Naryshkina *et al.*, 2006). This *Thermus* phage (phiYS40) is of special note since its dUTPase gene is clearly related to the dUTPases of eukaryotic viruses and has a version that has undergone multiple events of recombination from apparently distinct phage, exactly as expected for mosaic phage genes.

Thus, the origin of new herpesvirus genes might not be so different than that seen in other large DNA viruses and a potential ancient source of new genes from these ancestral viruses remains plausible.

Similar considerations apply to other possible examples of herpesvirus gene capture. For example, the herpes thymidylate synthase (TS) has also been considered to have originated by host gene capture (Chen *et al.*, 2001). Yet distinct versions of these genes are also found in different herpesviral lineages, which would necessitate multiple independent "capture" events of different version of host TS genes. TS genes are present in ancient virus sources. For example, *Bacillus* phage beta 22 encodes TS, which also has a self-splicing intron (Bechhofer *et al.*, 1994). Also, phage phiKZ has a highly conserved TS (Mesyanzhinov *et al.*, 2002), yet this virus

lacks a DNA polymerase or other replication proteins, clearly indicating that the viral TS genes has a basic viral role. Similarly, the cytokines-like genes (such as IL-10) as found in poxviruses and herpesviruses appear to have originated in at least three independent events prior to the divergence of mammalian eutherian orders. Yet it is still presupposed that they are necessarily the products of host gene capture (Hughes, 2002).

Comparative genomics supports the idea that the herpesviruse lineages are originating viral genes. A broader phylogenetic analysis of all herpesvirus genomes identified only 17 genes in common to all 30 taxa of herpesvirus (Wang *et al.*, 2006). Thus only 30 genes appear to be in common to all the herpesviruses. In this analysis, only a few genes of recent origin could be identified as possibly having been transferred between virus and host (e.g. new genes found at tips of phylogenetic dendograms). Thus, gene gain in the herpesviruses (as in DNA phage and phycodnavirus) is prevalent but the origination of such genes from the host is not prevalent. I suggest that our tendency to assume that new viral genes are usually "stolen" from the host should be revised (Moreira and Lopez-Garcia, 2005).

Poxvirus as Mostly Acute, with Frequent Species Shifts

In contrast to the herpesviruses, the poxviruses evolution tend to have little congruence to host evolution (see Chapter 19). Yet, they too show evidence of ancient linkages to other viruses. The replication of poxvirus DNA is distinct in that it involves a linear genome with inverted ends that have covalently closed "snapback" DNA. The resulting replication structures involve head-to-tail and tail-to-tail intermediates. This replication strategy is very different from that used by the host (and most other DNA viruses), but is clearly related to that found in other eukaryotic and prokaryotic viruses. Similar replication mechanisms are seen in all poxviruses, as well as African swine fever virus and phycodnaviruses (PBCV-1). This exact replication strategy is also present in archaeal lipothrixviruses (SIRV1 and SIRV2) which has been proposed to be ancestral to phycodnaviruses and poxviruses (Persson *et al.*, 2005). A similar replication strategy is also seen with N15 (Lobocka *et al.*, 1996), an unusual phage of *E. coli* that persists as a linear DNA (Casjens *et al.*, 2004).

Conservation of such replication similarities clearly suggests ancestral relationships, but no sequence similarity can be seen between these viruses. The similarity between poxvirus and PBCV-1 DNA replication deserves some additional comment. PBVC-1 and herpesvirus have very similar DNA polymerase genes, yet differ fundamentally in replication strategy. Furthermore, the poxviral DNA polymerase gene is very different from that found in the herpesviruses. Yet, the PBCV-1 capsid was clearly similar to that of adenoviruses and PRD1 phage (and iridovirus capsids). How then do we link poxvirus evolution to other more ancient DNA viruses, such as PBCV-1 which has the same DNA replication mechanism, but distinct replication proteins? Such observations might seem confusing, but they are clearly consistent with mosaic, reticulated evolution of DNA viruses. Various distinct phage lineages can link in multiple ways to various distinct eukaryotic DNA viruses. The concept of a net or matrix rather than a tree is thus a better way to describe the broad topology of DNA virus evolution.

The issue of gene gain and gene loss is also of central interest to orthopoxvirus evolution. Typically, we seek to understand poxviruses evolution from the perspective of pathogenesis, such as the origin of human-specific smallpox virus. With the comparative genomics of several orthopoxviruses now possible, we see curious overall patterns of gene loss in their evolution (Randall *et al.*, 2004). For example, comparing human smallpox to cowpox DNA (a rodent virus that is phylogenetically basal to smallpox), we observe an overall diminution of gene content in smallpox virus. Several poxviruses seem to have also lost genes relative to cowpoxvirus, especially genes that appear to affect immunity (Hughes and Friedman, 2005).

I suggest that this evolutionary tendency for gene reduction is associated with a switch from a more demanding species-specific persistent life strategy to a less demanding, acute life strategy in a new host. Cowpox is a naturally persistent infection in rodents (bank voles) (Feore *et al.*, 1997; Chantrey *et al.*, 1999), which has been called a natural virus reservoir (Hazel *et al.*, 2000). Smallpox is a strictly acute and human-specific disease. Such gene loss in association with lost persistence could be a general situation and might also explain why clinical isolates of human cytomegalovirus isolates show a strong tendency to delete genes with passage in culture (Davison *et al.*, 2003).

Most orthopoxviruses are not phylogenetically congruent with their vertebrate host. Host switching and acute replication seem to be relatively common but recent occurrences in their evolution (Babkin and Shchelkunov, 2006). The avian poxviruses are not as well studied in this context, but curiously have significantly more complex genomes than the orthopoxviruses (Jarmin *et al.*, 2006). The entomopoxviruses are even less well understood from both a biological and molecular perspective, although they do conserve 49 genes found in all poxvirus family members (Gubser *et al.*, 2004). Clearly these poxviruses share some degree of evolutionary history.

It is most curious that entomopoxviruses have even larger, more diverse and complex genomes than the other poxviruses. Why? As insects lack an adaptive immune system (the target of many orthopoxvirus genes), they would seem to present a simpler host for virus adaptation. This group appears to be the most basal phylogenetically, but evolutionary relationships between entomopoxvirus and insect evolution have not been studied. The enotomopoxviruses are particularly prevalent in grasshopper and locust species, often in unapparent states. Interestingly, within these viruses we can find examples of major shifts in core replication genes, such as the family of DNA pol gene that is used (a shift from DNA pol X to DNA pol B in two entomopoxvirus lineages). We can recall that the DNA pol B gene closely resembles that found in phycodnaviruses (and herpesvirus), but is distinct from that in orthopoxvirus (Zhu, 2003). We also see in the entomopoxviruses some clear links to phage genes, such as T4-like RNA ligase found in all entomopoxviruses (Ho and Shuman, 2002) as well as a lambda-like integrase seen in D1EPV (Hashimoto and Lawrence, 2005). This integrase in D1EPV implies possible integration and persistence, thus it is most significant that D1EPV also shows a clear persistent host infection as well as symbiosis and apparent phylogenetic congruence between virus and host.

This virus is symbiotic in its parasitoid wasp host in that virus is injected into larval host along with the wasp egg (and also along with a second D1RhV rhabdovirus) and virus is needed for successful host parasitization. This symbiosis is clearly very reminiscent of the genomic polydnaviruses of other parasitoid wasp species. DIEPV is also expressed in the male poison gland. However, it is unknown if DIEPV is integrating into the host DNA. Clearly, D1EPV it is part of a complex virus–virus–host symbiotic interaction.

Small DNA Viruses

The overall evolution of orthopoxviruses contrasts sharply with that of the papillomaviruses as presented by Bernard in Chapter 18. Here, highly species-specific and tissue-specific host infection are the norm and the viral evolution is typically highly congruent with the host (with some exceptions). The resolution between virus and host can be high, in that human racial and geographical populations, for example, can often be differentiated based on the type of HPV they harbor. Yet here too there is evidence of significant shifts in core gene usage early during papillomavirus evolution. In the human and rodent viruses, a highly conserved gene function associated with replication and cell control are the E6 and E7 early genes. In particular, the pRB-binding domain of the E7 gene is thought to be central to the biological strategy of the virus. Thus, it is most curious that the papillomaviruses of

lagomorphs, such as bovine and reindeer papillomavirus, lack an E7 Rb-binding domain and instead appear to use E5 or E9 genes for this regulatory function (Narechania *et al.*, 2004). It seems an early but significant and bifurcating shift occurred in the molecular strategy during the virus–host evolution of this group of viruses for unknown reasons.

Other small DNA viruses (JCV, BKV, Py) can also show similar high-resolution host congruence (Shadan and Villarreal, 1995). As well as similar curious shifts in basic molecular strategies. For example, the presence of a middle T-antigen in mouse virus (a third early gene), but its absence from primate viruses (Gottlieb and Villarreal, 2001), clearly differentiates these viral lineages. Although the origins of these entire small DNA viruses are obscure, and any links to prokaryotic viruses are unknown, it does appear they have tended to retain their overall biological strategy and show a strong tendency for tissue-specific (especially kidney) persistence and virus–host congruence.

Persistence as Symbiosis, another Foundation for Virus–Host Evolution

Since persistence requires the stable coexistence of a virus and its host, it also fits the simple definition of symbiosis (the stable living together of two distinct lineages of organisms). Viral involvement in symbiosis is a foreign idea to many and possibly presents a fundamentally different view of the role viruses may have in host evolution. A major role for persisting (temperate, cryptic) viruses in the evolution of prokaryotes is no longer a controversial idea. Thus, at least in the prokaryotic world, virus persistence can be accepted as adaptive. In eukaryotes, however, viral persistence is seldom considered adaptive. The MHV–mouse exemplar as presented above has suggested how persistence can directly affect host survival. Can this be considered an example of symbiosis in the accepted sense?

A crowning achievement in the field of symbiosis has been to explain the origin of plastids (chloroplasts, mitochondria) from symbiotic prokaryotes in eukaryotic cytoplasm (Margulis and Bermudes, 1985). This idea involves the high adaptability of prokaryotes to provide innovation but would seem not to involve virus in any way. Yet here too we can find viral footprints that suggest some involvement. For example, various plastid-specific RNA and DNA polymerases clearly resemble polymerases from T3/T7-like phage (Cermakian *et al.*, 1996; Shutt and Gray, 2006). Other models of symbiosis also show evidence of a viral role, such as the sexual isolation of *Buchnera* (Moran *et al.*, 2005).

Another very popular topic in the field of symbiosis is the symbiotic origin of the photosynthetic sea slug, *Elysa chlorotica*. What could be more fascinating than a green sea slug—an animal that can use light for photosynthesis? *E. chlorotica* eats photosynthetic eukaryotic algae (*Vaucheria litorea*) and retains the functional chloroplast from algae for months. Here too, however, there lies a viral footprint. This slug harbors an unusual endogenous retrovirus which is expressed in large numbers during sexual reproduction, following which all slugs die via synchronized apoptosis and in which the chloroplasts have accumulated numerous viral particles (Pierce *et al.*, 1999; Mondy and Pierce, 2003). Since there is reason to think gene movement from the algae to the slug genome is involved in this symbiosis, the presence of this retrovirus is a strong candidate to also be involved in symbiogenesis.

Clearly we should thus investigate retroviral elements as possible symbiotic participants and not dismiss them beforehand as irrelevant or "junk DNA" (as is automatically done in many database screens).

If viral persistence is a kind of symbiosis, viruses may also mediate the establishment of other symbiotic relationships (Villarreal, 2007). The recent studies by Roossinck and colleagues (see Chapter 12), in which a persisting virus, a plant, and a fungus were all symbiotically involved in altering the thermal tolerance of the plant could be an example of this (Marquez *et al.*, 2007). Many other virus–host relationships should also be examined for possible symbiosis. For example,

placental vertebrate evolution has involved various endogenous retroviruses (i.e. HERV-W, HERV-FRD). Intact HERV genomes, including *env* ORFs, are important for placental trophoblast fusion (for references see Ryan, 2007).

Some will dismiss this situation as the quirky usurping of a viral gene for host function which is of little general significance. The specific ERV involved is simply selfish and mostly defective genetic material of no general consequence. If so, why is it that in sheep a distinctly different lineage of retrovirus (enJSRV) was also selected to provide a related placental function to a another mammal with significantly different placental reproductive biology? It has been experimentally well established the enJSRV *env* is essential for sheep embryo implantation (Dunlap *et al.*, 2006a, 2006b). enJSRV is the endogenous version of JSRV, a problematic sheep-specific retrovirus that induces lung tumors (responsible for the death of Dolly, the famous first cloned sheep). The endogenous virus (enJSRV) is present in 20 copies in the sheep genome and all sheep have this virus. Sheep genomes also encoded a *trans*-dominant enJRSV *gag* that is inhibitory to exogenous JSRV (Mura *et al.*, 2004; Oliveira *et al.*, 2006; Murcia *et al.*, 2007). It seems clear that this situation can also be considered from the perspective of viral symbiosis and/or virus addiction in host evolution. We should thus seek to understand why colonization by an ERV population might generally provide a good solution to the evolutionary demands of placental biology.

Open Questions Regarding Virus in Human Evolution

There are many other opportunities to examine the potential role of persistent and symbiotic viruses in the evolution of viruses, animals, primates, and humans. For example, as we seek to understand the origins of the adaptive immune system we should pay attention to viral footprints. We can ask, for example, why the major histocompatibility complex (MHC) locus, the most polymorphic, diverse, and rapidly evolving gene set in our chromosome, is so densely colonized with retroviral elements (Andersson *et al.*, 1998). Why is a retrovirus also the basic element of the duplication unit that was thought to have been the progenitor for the expansion of the MHC class I (and II) genes (Gaudieri *et al.*, 1997; Kulski *et al.*, 1998, 2005)? Why do similar HERV element (L and 16) also differentiate between human and chimpanzee MHC I (Watkins, 1995; Kulski *et al.*, 1999)? What was the role for SIV in the evolution of primate MHC (Vogel *et al.*, 1999)?

Humans and primates appear to have undergone some significant and relatively recent evolution with regard to their endogenous and exogeneous retroviruses. Along these lines, APOBEC-like genes are basic component of the adaptive immune response but they are also antiretroviral genes that act on retroviral cDNA and *gag* (OhAinle *et al.*, 2006). The APOBEC3 antiviral system has expanded recently in humans, but not chimpanzees (Sawyer *et al.*, 2004; OhAinle *et al.*, 2006). Why? All African primates support unapparent foamy viruses (and also SIV co-infection), but not humans (Murray and Linial, 2006). APOBEC3C is active against foamy viruses (Delebecque *et al.*, 2006). Old World primates also underwent an expansion of ERVL colonization (a clear relative of foamy virus) (Sawyer *et al.*, 2004). Was this ERVL colonization of relevance to the ancient co-speciation of simian foamy virus and their primate host (Switzer *et al.*, 2005)? What exactly was the relevance of HERV endogenization to human survival and adaptations?

Curiously, human brain (neocortex) specifically expresses many of these more recent ERVs as transcripts (Nakamura *et al.*, 2003; Yi *et al.*, 2004). If we consider these situations as possible examples of virus-mediated symbiosis in human evolution, perhaps they may make more sense of the otherwise confusing role or HERVs.

REAL-TIME VIRUS–HOST EVOLUTION: KOALA BEAR EXEMPLAR

As noted, all primates, but especially humans show much evidence of recent endogenization

by retroviruses. But these events mostly occurred in our extinct ancestors and we do not see ongoing evidence that any HERVs remain active. However, we are currently witnessing a related virus–host evolutionary event of considerable interest. Koala bears, native marsupials of Australia, are currently experiencing a major epidemic caused by a leukemia-inducing retrovirus. As a consequence, they are undergoing massive endogenization by a gammaretrovirus (MLV-related). This virus is similar to Gibbon ape leukemia virus, but most likely originated from rodent ancestry (Tarlinton *et al.*, 2006; Fiebig *et al.*, 2006). The expectation is that extinction awaits those koalas that do not adapt or endogenize the retrovirus successfully (Stoye, 2006). This event has the appearances of a retroviral-driven addiction that will result in a genetic variant of koala bear that has acquired a new antiretroviral state. This seems equivalent to the expansion of human APOBEC3; or perhaps a closer analogy is the endogenization of a suppressive *gag* as occurred with enJSRV. The surviving koala bears will likely tolerate or be persistently infected with this retrovirus pool. The genome of the species will have undergone considerable (but unpredictable) genetic perturbations and likely contain a large pool of variant and defective retrovirus. However, in so doing, the descendent koalas will likely present a biological hazard to any koala species that remain virus-free (as in virus addiction). Currently, one island colony of koalas is sufficiently isolated to have remained virus-free. This population will henceforth be under persistent threat from populations of endogenized koalas, now favored by group selection.

From the very earliest events in evolution of prebiotic replicators to very recent events in human evolution, including the emergence of human-specific HIV, we expect viral evolution to show profound effects on the evolution of all life. Unlike accepted host evolution, viruses also employ consortia and mixed populations to evolve, sometimes at unprecedented rates. Thus viruses have informed us of quasispecies, group dynamics, and group selection in evolution. Virus evolution should now be considered as basic science, not just a medical concern. We must acknowledge that the tree of life cannot be properly understood without virus evolution. This book helps to lay the foundation for such understanding.

REFERENCES

Adams, W.R. and Kraft, L.M. (1963) Epizootic diarrhea of infant mice: indentification of the etiologic agent. *Science* **141**, 359–360.

Akita, F., Chong, K.T., Tanaka, H., Yamashita, E., Miyazaki, N., Nakaishi, Y. *et al.* (2007) The crystal structure of a virus-like particle from the hyperthermophilic archaeon *Pyrococcus furiosus* provides insight into the evolution of viruses. *J. Mol. Biol.* **368**, 1469–1483.

Alexander, L., Weiskopf, E., Greenough, T.C., Gaddis, N.C., Auerbach, M.R., Malim, M.H. *et al.* (2000) Unusual polymorphisms in human immunodeficiency virus type 1 associated with nonprogressive infection. *J. Virol.* **74**, 4361–4376.

Allen, M.J., Schroeder, D.C., Donkin, A., Crawfurd, K.J. and Wilson, W.H. (2006a) Genome comparison of two Coccolithoviruses. *Virol. J.* **3**, 15.

Allen, M.J., Schroeder, D.C., Holden, M.T. and Wilson, W.H. (2006b) Evolutionary history of the Coccolithoviridae. *Mol. Biol. Evol.* **23**, 86–92.

Andersson, G., Svensson, A.C., Setterblad, N. and Rask, L. (1998) Retroelements in the human MHC class II region. *Trends Genet.* **14**, 109–114.

Babkin, I.V. and Shchelkunov, S.N. (2006) [The time scale in poxvirus evolution]. *Mol. Biol. (Mosk.)* **40**, 20–24.

Badrane, H. and Tordo, N. (2001) Host switching in Lyssavirus history from the Chiroptera to the Carnivora orders. *J. Virol.* **75**, 8096–8104.

Baker, M.L., Jiang, W., Rixon, F.J. and Chiu, W. (2005) Common ancestry of herpesviruses and tailed DNA bacteriophages. *J. Virol.* **79**, 14967–14970.

Bamford, D.H. (2003) Do viruses form lineages across different domains of life?. *Res. Microbiol.* **154**, 231–236.

Bamford, D.H., Grimes, J.M. and Stuart, D.I. (2005a) What does structure tell us about virus evolution?. *Curr. Opin. Struct. Biol.* **15**, 655–663.

Bamford, D.H., Ravantti, J.J., Ronnholm, G., Laurinavicius, S., Kukkaro, P., Dyall-Smith, M. *et al.* (2005b) Constituents of SH1, a novel lipid-containing virus infecting the halophilic euryarchaeon *Haloarcula hispanica*. *J. Virol.* **79**, 1107–9097.

Barbosa-Solomieu, V., Degremont, L., Vazquez-Juarez, R., Ascencio-Valle, F., Boudry, P. and Renault, T. (2005) Ostreid herpesvirus 1 (OsHV-1) detection among three successive generations of Pacific oysters (*Crassostrea gigas*). *Virus Res.*. **107**, 47–56.

Baric, R.S., Sullivan, E., Hensley, L., Yount, B. and Chen, W. (1999) Persistent infection promotes cross-species transmissibility of mouse hepatitis virus. *J. Virol.* **73**, 638–649.

Batschelet, E., Domingo, E. and Weissmann, C. (1976) The proportion of revertant and mutant phage in a growing population, as a function of mutation and growth rate. *Gene* **1**, 27–32.

Bechhofer, D.H., Hue, K.K. and Shub, D.A. (1994) An intron in the thymidylate synthase gene of *Bacillus* bacteriophage beta 22: evidence for independent evolution of a gene, its group I intron, and the intron open reading frame. *Proc. Natl Acad. Sci. USA* **91**, 11669–11673.

Becker, S.D., Bennett, M., Stewart, J.P. and Hurst, J.L. (2007) Serological survey of virus infection among wild house mice (*Mus domesticus*) in the UK. *Lab. Anim.* **41**, 229–238.

Becker, Y. and Darai, G. (2000) *Molecular Evolution of Viruses, Past and Present: Evolution of Viruses by Acquisition of Cellular RNA and DNA*. Boston: Kluwer Academic.

Bell, P.J. (2001) Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus? *J. Mol. Evol.* **53**, 251–256.

Bell, P.J. (2006) Sex and the eukaryotic cell cycle is consistent with a viral ancestry for the eukaryotic nucleus. *J. Theor. Biol.* **243**, 54–63.

Bello, G., Casado, C., Garcia, S., Rodriguez, C., del Romero, J. and Lopez-Galindez, C. (2004) Coexistence of recent and ancestral nucleotide sequences in viral quasispecies of human immunodeficiency virus type 1 patients. *J. Gen. Virol.* **85**, 399–407.

Bello, G., Casado, C., Sandonis, V., Alonso-Nieto, M., Vicario, J.L., Garcia, S. *et al.* (2005) A subset of human immunodeficiency virus type 1 long-term nonprogressors is characterized by the unique presence of ancestral sequences in the viral population. *J. Gen. Virol.* **86**, 355–364.

Benson, S.D., Bamford, J.K., Bamford, D.H. and Burnett, R.M. (2004) Does common architecture reveal a viral lineage spanning all three domains of life?. *Mol. Cell.* **16**, 673–685.

Biebricher, C.K. and Eigen, M. (2005) The error threshold. *Virus Res.* **107**, 117–127.

Biebricher, C.K. and Eigen, M. (2006) What is a quasispecies?. *Quasispecies: Concept and Implications for Virology* **299**, 1–31.

Blond, J.L., Lavillette, D., Cheynet, V., Bouton, O., Oriol, G., Chapel-Fernandes, S. *et al.* (2000) An envelope glycoprotein of the human endogenous retrovirus HERV-W is expressed in the human placenta and fuses cells expressing the type D mammalian retrovirus receptor. *J. Virol.* **74**, 3321–3329.

Blum, H., Zillig, W., Mallok, S., Domdey, H. and Prangishvili, D. (2001) The genome of the archaeal virus SIRV1 has features in common with genomes of eukaryal viruses. *Virology* **281**, 6–9.

Bouman, H.A., Ulloa, O., Scanlan, D.J., Zwirglmaier, K., Li, W.K., Platt, T. *et al.* (2006) Oceanographic basis of the global surface distribution of Prochlorococcus ecotypes. *Science* **312**, 918–921.

Breitbart, M. and Rohwer, F. (2005a) Here a virus, there a virus, everywhere the same virus?. *Trends Microbiol.* **13**, 278–284.

Breitbart, M. and Rohwer, F. (2005b) Method for discovering novel DNA viruses in blood using viral particle selection and shotgun sequencing. *Biotechniques* **39**, 729–736.

Breitbart, M., Hewson, I., Felts, B., Mahaffy, J.M., Nulton, J., Salamon, P. and Rohwer, F. (2003) Metagenomic analyses of an uncultured viral community from human feces. *J. Bacteriol.* **185**, 6220–6223.

Briones, C., de Vicente, A., Molina-Paris, C. and Domingo, E. (2006) Minority memory genomes can influence the evolution of HIV-1 quasispecies in vivo. *Gene* **384**, 129–138.

Brussow, H., Canchaya, C. and Hardt, W.D. (2004) Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol. Mol. Biol. Rev.* **68**, 560–602.

Caceres, M. and Thomas, J.W. (2006) The gene of retroviral origin Syncytin 1 is specific to hominoids and is inactive in Old World monkeys. *J. Hered.* **97**, 100–106.

Campbell, A. (2007) Phage integration and chromosome structure. A personal history. *Annu. Rev. Genet.*

Campbell, A., Schneider, S.J. and Song, B. (1992) Lambdoid phages as elements of bacterial genomes. *Genetica* **86**, 259–267.

Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M.L. and Brussow, H. (2003) Phage as agents of lateral gene transfer. *Curr. Opin. Microbiol.* **6**, 417–424.

Canchaya, C., Fournous, G. and Brussow, H. (2004) The impact of prophages on bacterial chromosomes. *Mol. Microbiol.* **53**(1), 9–18.

Carobene, M.G., Rubio, A.E., Carrillo, M.G., Maligne, G.E., Kijak, G.H., Quarleri, J.F. and Salomon, H. (2004) Differences in frequencies of drug resistance-associated mutations in the HIV-1 pol gene of B subtype and BF intersubtype recombinant samples. *J. Acquir. Immune Defic. Syndr.* **35**, 207–209.

Carthew, P. (1977) Lethal intestinal virus of infant mice is mouse hepatitis virus. *Vet. Rec.* **101**, 465.

Casado, C., Garcia, S., Rodriguez, C., del Romero, J., Bello, G. and Lopez-Galindez, C. (2001) Different evolutionary patterns are found within human immunodeficiency virus type 1-infected patients. *J. Gen. Virol.* **82**, 2495–2508.

Casjens, S.R., Gilcrease, E.B., Huang, W.M., Bunny, K.L., Pedulla, M.L. *et al.* (2004) The pK02 linear plasmid prophage of *Klebsiella oxytoca*. *J. Bacteriol.* **186**, 1818–1832.

Castro, C., Smidansky, E., Maksimchuk, K.R., Arnold, J.J., Korneeva, V.S., Gotte, M. *et al.* (2007) Two proton transfers in the transition state for nucleotidyl transfer catalyzed by RNA- and DNA-dependent RNA and DNA polymerases. *Proc. Natl Acad. Sci. USA* **104**, 4267–4272.

Cermakian, N., Ikeda, T.M., Cedergren, R. and Gray, M.W. (1996) Sequences homologous to yeast mitochondrial and bacteriophage T3 and T7 RNA polymerases are widespread throughout the eukaryotic lineage. *Nucleic Acids Res.* **24**, 648–654.

Chantrey, J., Meyer, H., Baxby, D., Begon, M., Bown, K.J., Hazel, S.M. *et al.* (1999) Cowpox: reservoir hosts and geographic range. *Epidemiol. Infect.* **122**, 455–460.

Chare, E.R. and Holmes, E.C. (2006) A phylogenetic survey of recombination frequency in plant RNA viruses. *Arch. Virol.* **151**, 933–946.

Charpentier, C., Dwyer, D.E., Mammano, F., Lecossier, D., Clavel, F. and Hance, A.J. (2004) Role of minority populations of human immunodeficiency virus type 1 in the evolution of viral resistance to protease inhibitors. *J. Virol.* **78**, 4234–4247.

Charpentier, C., Nora, T., Tenaillon, O., Clavel, F. and Hance, A.J. (2006) Extensive recombination among human immunodeficiency virus type 1 quasispecies makes an important contribution to viral diversity in individual patients. *J. Virol.* **80**, 2472–2482.

Chen, F. and Suttle, C.A. (1996) Evolutionary relationships among large double-stranded DNA viruses that infect microalgae and other organisms as inferred from DNA polymerase genes. *Virology* **219**, 170–178.

Chen, H.H., Tso, D.J., Yeh, W.B., Cheng, H.J. and Wu, T.F. (2001) The thymidylate synthase gene of Hz-1 virus: a gene captured from its lepidopteran host. *Insect. Mol. Biol.* **10**, 495–503.

Christodoulou, C., Colbere-Garapin, F., Macadam, A., Taffs, L.F., Marsden, S. *et al.* (1990) Mapping of mutations associated with neurovirulence in monkeys infected with Sabin 1 poliovirus revertants selected at high temperature. *J. Virol.* **64**, 4922–4929.

Clarke, D.K., Duarte, E.A., Elena, S.F., Moya, A., Domingo, E. and Holland, J. (1994) The red queen reigns in the kingdom of RNA viruses. *Proc. Natl Acad. Sci. USA* **91**, 4821–4824.

Claverie, J.M., Ogata, H., Audic, S., Abergel, C., Suhre, K. and Fournier, P.E. (2006) Mimivirus and the emerging concept of "giant" virus. *Virus Res.* **117**, 133–144.

Clokie, M.R., Shan, J., Bailey, S., Jia, Y., Krisch, H.M., West, S. and Mann, N.H. (2006) Transcription of a 'photosynthetic' T4-type phage during infection of a marine cyanobacterium. *Environ. Microbiol.* **8**, 827–835.

Coleman, M.L., Sullivan, M.B., Martiny, A.C., Steglich, C., Barry, K., Delong, E.F. and Chisholm, S.W. (2006) Genomic islands and the ecology and evolution of Prochlorococcus. *Science* **311**, 1768–1770.

Comeau, A.M. and Krisch, H.M. (2005) War is peace—dispatches from the bacterial and phage killing fields. *Curr. Opin. Microbiol.* **8**, 488–494.

Comeau, A.M., Chan, A.M. and Suttle, C.A. (2006) Genetic richness of vibriophages isolated in a coastal environment. *Environ. Microbiol.* **8**, 1164–1176.

Crotty, S. and Andino, R. (2002) Implications of high RNA virus mutation rates: lethal mutagenesis and the antiviral drug ribavirin. *Microb. Infect.* **4**, 1301–1307.

Crotty, S., Cameron, C.E. and Andino, R. (2001) RNA virus error catastrophe: direct molecular test by using ribavirin. *Proc. Natl Acad. Sci. USA* **98**, 6895–6900.

Crotty, S., Hix, L., Sigal, L.J. and Andino, R. (2002) Poliovirus pathogenesis in a new poliovirus receptor transgenic mouse model: age-dependent paralysis and a mucosal route of infection. *J. Gen. Virol.* **83**, 1707–1720.

Crowder, S. and Kirkegaard, K. (2005) Trans-dominant inhibition of RNA viral replication can slow growth of drug-resistant viruses. *Nat. Genet.* **37**, 701–709.

Davison, A.J. and Stow, N.D. (2005) New genes from old: redeployment of dUTPase by herpesviruses. *J. Virol.* **79**, 12880–12892.

Davison, A.J., Dolan, A., Akter, P., Addison, C., Dargan, D.J., Alcendor, D.J. *et al.* (2003) The human cytomegalovirus genome revisited: comparison with the chimpanzee cytomegalovirus genome. *J. Gen. Virol.* **84**, 17–28.

Davison, A.J., Trus, B.L., Cheng, N., Steven, A.C., Watson, M.S., Cunningham, C. *et al.* (2005) A novel class of herpesvirus with bivalve hosts. *J. Gen. Virol.* **86**, 41–53.

Delaroque, N., Muller, D.G., Bothe, G., Pohl, T., Knippers, R. and Boland, W. (2001) The complete DNA sequence of the *Ectocarpus siliculosus* Virus EsV-1 genome. *Virology* **287**, 112–132.

Delaroque, N., Boland, W., Gerhard Müller, D. and Knippers, R. (2003) Comparisons of two large phaeoviral genomes and evolutionary implications. *J. Mol. Evol.* **57**, 613–622.

Delbruck, M. (1945) Interference between bacterial viruses: III. The mutual exclusion effect and the depressor effect. *J. Bacteriol.* **50**, 151–170.

Delebecque, F., Suspene, R., Calattini, S., Casartelli, N., Saib, A., Froment, A. *et al.* (2006) Restriction of foamy viruses by APOBEC cytidine deaminases. *J. Virol.* **80**, 605–614.

Descoteaux, J.P. and Mihok, S. (1986) Serologic study on the prevalence of murine viruses in a population of wild meadow voles (*Microtus pennsylvanicus*). *J. Wildl. Dis.* **22**, 314–319.

Descoteaux, J.P., Grignon-Archambault, D. and Lussier, G. (1977) Serologic study on the prevalence of murine viruses in five Canadian mouse colonies. *Lab. Anim. Sci.* **27**, 621–626.

Desjardins, C., Eisen, J.A. and Nene, V. (2005) New evolutionary frontiers from unusual virus genomes. *Genome Biol.* **6**, 212.

Domingo, E., Biebricher, C., Eigen, M. and Holland J.J. (2001) *Quasispecies and RNA Virus Evolution: Principles and Consequences*. Austin, TX: Landes Bioscience.

Domingo, E. (2006) Quasispecies: concept and implications for virology. In: *Curr. Top. Microbiol. Immunol*, Vol. 299. Berlin, New York: Springer.

Domingo, E. and Gomez, J. (2007) Quasispecies and its impact on viral hepatitis. *Virus Res.* **127**, 131–150.

Domingo, E., Sabo, D., Taniguchi, T. and Weissmann, C. (1978) Nucleotide sequence heterogeneity of an RNA phage population. *Cell* **13**, 735–744.

Domingo, E., Webster, R.G. and Holland, J.J. (eds) (1999) *Origin and Evolution of Viruses*. San Diego: Academic Press.

Doolittle, W.F. and Sapienza, C. (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**, 601–603.

Dreyfus, D.H., Jones, J.F. and Gelfand, E.W. (1999) Asymmetric DDE (D35E)-like sequences in the RAG proteins: implications for V(D)J recombination and retroviral pathogenesis. *Med. Hypoth.* **52**, 545–549.

Duda, R.L., Hendrix, R.W., Huang, W.M. and Conway, J.F. (2006) Shared architecture of bacteriophage SP01 and herpesvirus capsids. *Curr. Biol.* **16**, R11–R13.

Dunlap, K.A., Palmarini, M. and Spencer, T.E. (2006a) Ovine endogenous betaretroviruses (enJSRVs) and placental morphogenesis. *Placenta* **27**(Suppl A), S135–S140.

Dunlap, K.A., Palmarini, M., Varela, M., Burghardt, R.C., Hayashi, K., Farmer, J.L. and Spencer, T.E. (2006b) Endogenous retroviruses regulate periimplantation placental growth and differentiation. *Proc. Natl Acad. Sci. USA* **103**, 14390–14395.

Dupressoir, A., Marceau, G., Vernochet, C., Benit, L., Kanellopoulos, C., Sapin, V. and Heidmann, T. (2005) Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae. *Proc. Natl Acad. Sci. USA* **102**, 725–730.

Edwards, R.A. and Rohwer, F. (2005) Viral metagenomics. *Nat. Rev. Microbiol.* **3**, 504–510.

Eigen, M., McCaskill, J. and Schuster, P. (1988) Molecular quasi-species. *J. Phys. Chem.* **92**, 6881–6891.

Einer-Jensen, K., Ahrens, P. and Lorenzen, N. (2006) Genetic stability of the VHSV consensus sequence of G-gene in diagnostic samples from an acute outbreak. *Bull. Eur. Assoc. Fish Pathol.* **26**, 62–67.

Elena, S.F., Gonzalez-Candelas, F., Novella, I.S., Duarte, E.A., Clarke, D.K., Domingo, E. *et al.* (1996) Evolution of fitness in experimental populations of vesicular stomatitis virus. *Genetics* **142**, 673–679.

Enomoto, N., Kurosaki, M., Tanaka, Y., Marumo, F. and Sato, C. (1994) Fluctuation of hepatitis C virus quasispecies in persistent infection and interferon treatment revealed by single-strand conformation polymorphism analysis. *J. Gen. Virol.* **75**, 1361–1369.

Esnault, C., Heidmann, O., Delebecque, F., Dewannieux, M., Ribet, D., Hance, A.J. *et al.* (2005) APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. *Nature* **433**, 430–433.

Fargette, D., Konate, G., Fauquet, C., Muller, E., Peterschmitt, M. and Thresh, J.M. (2006) Molecular ecology and emergence of tropical plant viruses. *Annu. Rev. Phytopathol.* **44**, 235–260.

Feore, S.M., Bennett, M., Chantrey, J., Jones, T., Baxby, D. and Begon, M. (1997) The effect of cowpox virus infection on fecundity in bank voles and wood mice. *Proc. R. Soc. Lond. B Biol. Sci.* **264**, 1457–1461.

Fiebig, U., Hartmann, M.G., Bannert, N., Kurth, R. and Denner, J. (2006) Transspecies transmission of the endogenous koala retrovirus. *J. Virol.* **80**, 5651–5654.

Filee, J., Forterre, P. and Laurent, J. (2003) The role played by viruses in the evolution of their hosts: a view based on informational protein phylogenies. *Res. Microbiol.* **154**, 237–243.

Filee, J., Tetart, F., Suttle, C.A. and Krisch, H.M. (2005) Marine T4-type bacteriophages, a ubiquitous component of the dark matter of the biosphere. *Proc. Natl Acad. Sci. USA* **102**(35), 12471–12476.

Filee, J., Bapteste, E., Susko, E. and Krisch, H.M. (2006a) A selective barrier to horizontal gene transfer in the T4-type bacteriophages that has preserved a core genome with the viral replication and structural genes. *Mol. Biol. Evol.* **23**, 1688–1696.

Filee, J., Comeau, A.M., Suttle, C.A. and Krisch, H.M. (2006b) T4-type bacteriophages: ubiquitous components of the "dark matter" of the biosphere. *Med. Sci. (Paris)* **22**, 111–112.

Filee, J., Siguier, P. and Chandler, M. (2006c) I am what I eat and I eat what I am: acquisition of bacterial genes by giant viruses. *Trends Genet.* (in press).

Fleuriet, A. (1994) Female characteristics in the *Drosophila melanogaster* sigma-virus system in natural-populations from Languedoc (Southern France). *Arch. Virol.* **135**, 29–42.

Fleuriet, A. (1996) Polymorphism of the *Drosophila melanogaster* Sigma virus system. *J. Evol. Biol.* **9**, 471–484.

Forterre, P. (1999) Displacement of cellular proteins by functional analogues from plasmids or viruses could explain puzzling phylogenies of many DNA informational proteins. *Mol. Microbiol.* **33**, 457–465.

Forterre, P. (2003) The great virus comeback—from an evolutionary perspective. *Res. Microbiol.* **154**, 223–225.

Forterre, P. (2005) The two ages of the RNA world and the transition to the DNA world: a story of viruses and cells. *Biochimie* **87**, 793–803.

Forterre, P. (2006a) The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res.* **117**, 5–16.

Forterre, P. (2006b) Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: A hypothesis for the origin of cellular domain. *Proc. Natl Acad. Sci. USA* **103**, 3669–3674.

Forterre, P., Gribaldo, S. and Brochier, C. (2005) Luca: the last universal common ancestor. *Med. Sci. (Paris)* **21**, 860–865.

Forterre, P., Gribaldo, S., Gadelle, D. and Serre, M.C. (2007) Origin and evolution of DNA topoisomerases. *Biochimie* **89**, 427–446.

Forton, D.M., Karayiannis, P., Mahmud, N., Taylor-Robinson, S.D. and Thomas, H.C. (2004) Identification of unique hepatitis C virus quasispecies in the central nervous system and comparative analysis of internal translational efficiency of brain, liver and serum variants. *J. Virol.* **78**, 5170–5183.

Forton, D.M., Taylor-Robinson, S.D. and Thomas, H.C. (2006) Central nervous system changes in hepatitis C virus infection. *Eur. J. Gastroenterol. Hepatol.* **18**, 333–338.

Fraile, A., Escriu, F., Aranda, M.A., Malpica, J.M., Gibbs, A.J. and Garcia-Arenal, F. (1997) A century of tobamovirus evolution in an Australian population of *Nicotiana glauca*. *J. Virol.* **71**, 8316–8320.

Fugmann, S.D., Messier, C., Novack, L.A., Cameron, R.A. and Rast, J.P. (2006) An ancient evolutionary origin of the Rag1/2 gene locus. *Proc. Natl Acad. Sci. USA* **103**, 3728–3733.

Gannon, J. and Carthew, P. (1980) Prevalence of indigenous viruses in laboratory animal colonies in the United Kingdom 1978–1979. *Lab. Anim.* **14**, 309–311.

Gao, L. and Qi, J. (2007) Whole genome molecular phylogeny of large dsDNA viruses using composition vector method. *BMC Evol. Biol.* **7**, 41.

Garcia-Arriaza, J., Manrubia, S.C., Toja, M., Domingo, E. and Escarmis, C. (2004) Evolutionary transition toward defective RNAs that are infectious by complementation. *J. Virol.* **78**, 11678–11685.

Gaudieri, S., Kulski, J.K., Balmer, L., Giles, K.M., Inoko, H. and Dawkins, R.L. (1997) Retroelements and segmental duplications in the generation of diversity within the MHC. *DNA Seq.* **8**, 137–141.

Gazit, E. and Sauer, R.T. (1999) The Doc toxin and Phd antidote proteins of the bacteriophage P1 plasmid addiction system form a heterotrimeric complex. *J. Biol. Chem.* **274**, 16813–16818.

Gelfand, M.S. and Koonin, E.V. (1997) Avoidance of palindromic words in bacterial and archaeal genomes: a close connection with restriction enzymes. *Nucleic Acids Res.* **25**, 2430–2439.

Ghedin, E. and Claverie, J.M. (2005) Mimivirus relatives in the Sargasso sea. *Virol. J.* **2**, 62.

Gibbs, A. (1999) Evolution and origins of tobamoviruses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **354**, 593–602.

Gilbert, M., Chaitaweesub, P., Parakamawongsa, T., Premashthira, S., Tiensin, T., Kalpravidh, W. *et al.* (2006) Free-grazing ducks and highly pathogenic avian influenza, Thailand. *Emerg. Infect. Dis.* **12**, 227–234.

Gorbalenya, A.E., Enjuanes, L., Ziebuhr, J. and Snijder, E.J. (2006) Nidovirales: evolving the largest RNA virus genome. *Virus Res.* **117**, 17–37.

Gottlieb, K.A. and Villarreal, L.P. (2001) Natural biology of polyomavirus middle T antigen. *Microbiol. Mol. Biol. Rev.* **65**, 288–318.

Grande-Perez, A., Lazaro, E., Lowenstein, P., Domingo, E. and Manrubia, S.C. (2005) Suppression of viral infectivity through lethal defection. *Proc. Natl Acad. Sci. USA* **102**, 4448–4452.

Gubser, C., Hue, S., Kellam, P. and Smith, G.L. (2004) Poxvirus genomes: a phylogenetic analysis. *J. Gen. Virol.* **85**, 105–117.

Gustafsson, E., Blomqvist, G., Bellman, A., Holmdahl, R., Mattsson, A. and Mattsson, R. (1996) Maternal antibodies protect immunoglobulin deficient neonatal mice from mouse hepatitis virus (MHV)-associated wasting syndrome. *Am. J. Reprod. Immunol.* **36**, 33–39.

Hambly, E. and Suttle, C.A. (2005) The viriosphere, diversity and genetic exchange within phage communities. *Curr. Opin. Microbiol.* **8**, 444–450.

Harcombe, W.R. and Bull, J.J. (2005) Impact of phages on two-species bacterial communities. *Appl. Environ. Microbiol.* **71**, 5254–5259.

Harki, D.A., Graci, J.D., Korneeva, V.S., Ghosh, S.K., Hong, Z., Cameron, C.E. and Peterson, B.R. (2002) Synthesis and antiviral evaluation of a mutagenic and non-hydrogen bonding ribonucleoside analogue: 1-beta-D-Ribofuranosyl-3-nitropyrrole. *Biochemistry* **41**, 9026–9033.

Harris, J.R. (1998) Placental endogenous retrovirus (ERV): structural, functional and evolutionary significance. *Bioessays* **20**, 307–316.

Hart, C.A. and Bennett, M. (1999) Hantavirus infections: epidemiology and pathogenesis. *Microbes Infect.* **1**, 1229–1237.

Hashimoto, Y. and Lawrence, P.O. (2005) Comparative analysis of selected genes from *Diachasmimorpha longicaudata* entomopoxvirus and other poxviruses. *J. Insect Physiol.* **51**, 207–220.

Hazel, S.M., Bennett, M., Chantrey, J., Bown, K., Cavanagh, R., Jones, T.R. *et al.* (2000) A longitudinal study of an endemic disease in its wildlife reservoir: cowpox and wild rodents. *Epidemiol. Infect.* **124**, 551–562.

Hendrix, R.W. (1999) Evolution: the long evolutionary reach of viruses. *Curr. Biol.* **9**, R914–R917.

Hendrix, R.W. (2002) Bacteriophages: evolution of the majority. *Theor. Popul. Biol.* **61**, 471–480.

Hendrix, R.W. (2003) Bacteriophage genomics. *Curr. Opin. Microbiol.* **6**, 506–511.

Hendrix, R.W., Hatfull, G.F. and Smith, M.C. (2003) Bacteriophages with tails: chasing their origins and evolution. *Res. Microbiol.* **154**, 253–257.

Hendrix, R.W., Smith, M.C., Burns, R.N., Ford, M.E. and Hatfull, G.F. (1999) Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc. Natl Acad. Sci. USA* **96**, 2192–2197.

Herniou, E.A., Luque, T., Chen, X., Vlak, J.M., Winstanley, D., Cory, J.S. and O'Reilly, D.R. (2001) Use of whole genome sequence data to infer baculovirus phylogeny. *J. Virol.* **75**, 8117–8126.

Ho, C.K. and Shuman, S. (2002) Bacteriophage T4 RNA ligase 2 (gp24.1) exemplifies a family of RNA ligases found in all phylogenetic domains. *Proc. Natl Acad. Sci. USA* **99**, 12709–12714.

Hohne, M., Schreier, E. and Roggendorf, M. (1994) Sequence variability in the env-coding region of hepatitis C virus isolated from patients infected during a single source outbreak. *Arch. Virol.* **137**, 25–34.

Holmes, E.C. (2006) The evolutionary biology of dengue virus. *Novartis Found. Symp.* **277**, 177–187. discussion 187–192, 251–253.

Holmes, E.C. and Twiddy, S.S. (2003) The origin, emergence and evolutionary genetics of dengue virus. *Infect. Genet. Evol.* **3**, 19–28.

Homberger, F.R. (1997) Enterotropic mouse hepatitis virus. *Lab. Anim.* **31**, 97–115.

Horaud, F. (1993) Albert B. Sabin and the development of oral poliovaccine. *Biologicals* **21**, 311–316.

Hughes, A.L. (2002) Origin and evolution of viral interleukin-10 and other DNA virus genes with vertebrate homologues. *J. Mol. Evol.* **54**, 90–101.

Hughes, A.L. and Friedman, R. (2005) Poxvirus genome evolution by gene gain and loss. *Mol. Phylogenet. Evol.* **35**, 186–195.

Hurtado, A. and Rodriguez-Valera, F. (1999) Accessory DNA in the genomes of representatives of the *Escherichia coli* reference collection. *J. Bacteriol.* **181**, 2548–2554.

Ishida, T., Taguchi, F., Lee, Y.S., Yamada, A., Tamura, T. and Fujiwara, K. (1978) Isolation of mouse hepatitis virus from infant mice with fatal diarrhea. *Lab. Anim. Sci.* **28**, 269–276.

Iyer, L.M., Aravind, L. and Koonin, E.V. (2001) Common origin of four diverse families of large eukaryotic DNA viruses. *J. Virol.* **75**, 11720–11734.

Iyer, L.M., Balaji, S., Koonin, E.V. and Aravind, L. (2006) Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. *Virus Res.* **117**, 156–184.

Jarmin, S., Manvell, R., Gough, R.E., Laidlaw, S.M. and Skinner, M.A. (2006) Avipoxvirus phylogenetics: identification of a PCR length polymorphism that discriminates between the two major clades. *J. Gen. Virol.* **87**, 2191–2201.

Jern, P., Sperber, G.O. and Blomberg, J. (2005) Use of endogenous retroviral sequences (ERVs) and structural markers for retroviral phylogenetic inference and taxonomy. *Retrovirology* **2**, 50.

Kapitonov, V.V. and Jurka, J. (2005) RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol.* **3**, e181.

Kashuba, C., Hsu, C., Krogstad, A. and Franklin, C. (2005) Small mammal virology. *Vet. Clin. North Am. Exot. Anim. Pract.* **8**, 107–122.

Khayat, R., Tang, L., Larson, E.T., Lawrence, C.M., Young, M. and Johnson, J.E. (2005) Structure of an archaeal virus capsid protein reveals a common ancestry to eukaryotic and bacterial viruses. *Proc. Natl Acad. Sci. USA* **102**, 18944–18949.

Kijak, G.H. and McCutchan, F.E. (2005) HIV diversity, molecular epidemiology and the role of recombination. *Curr. Infect. Dis. Rep.* **7**, 480–488.

Kijak, G.H., Simon, V., Balfe, P., Vanderhoeven, J., Pampuro, S.E., Zala, C. *et al.* (2002) Origin of human immunodeficiency virus type 1 quasispecies emerging after antiretroviral treatment interruption in patients with therapeutic failure. *J. Virol.* **76**, 7000–7009.

Knobler, R.L., Lampert, P.W. and Oldstone, M.B. (1982) Virus persistence and recurring demyelination produced by a temperature-sensitive mutant of MHV-4. *Nature* **298**, 279–280.

Koonin, E.V., Makarova, K.S. and Aravind, L. (2001) Horizontal gene transfer in prokaryotes: quantification and classification. *Annu. Rev. Microbiol.* **55**, 709–742.

Korneeva, V.S. and Cameron, C.E. (2007) Structure-function relationships of the viral RNA-dependent RNA polymerase: Fidelity, replication speed and initiation mechanism determined by a residue in the ribose-binding pocket. *J. Biol. Chem.* **282**, 16135–16145.

Kulski, J.K., Gaudieri, S., Bellgard, M., Balmer, L., Giles, K., Inoko, H. and Dawkins, R.L. (1998) The evolution of MHC diversity by segmental duplication and transposition of retroelements. *J. Mol. Evol.* **46**, 734.

Kulski, J.K., Gaudieri, S., Inoko, H. and Dawkins, R.L. (1999) Comparison between two human endogenous retrovirus (HERV)-rich regions within the major histocompatibility complex. *J. Mol. Evol.* **48**, 675–683.

Kulski, J.K., Anzai, T. and Inoko, H. (2005) ERVK9, transposons and the evolution of MHC class I duplicons within the alpha-block of the human and chimpanzee. *Cytogenet. Genome Res.* **110**, 181–192.

Kung, H.J. and Wood, C. (1994) *Interactions between Retroviruses and Herpesviruses*. Singapore; River Edge, NJ: World Scientific.

Kurosaki, M., Enomoto, N., Marumo, F. and Sato, C. (1994) Evolution and selection of hepatitis C virus variants in patients with chronic hepatitis C. *Virology* **205**, 161–169.

Lacoste, V., Mauclere, P., Dubreuil, G., Lewis, J., Georges-Courbot, M.C. and Gessain, A. (2000) KSHV-like herpesviruses in chimps and gorillas. *Nature* **407**, 151–152.

Lehnherr, H., Maguin, E., Jafri, S. and Yarmolinsky, M.B. (1993) Plasmid addiction genes of bacteriophage P1: doc, which causes cell death on curing of prophage and phd, which prevents host death when prophage is retained. *J. Mol. Biol.* **233**, 414–428.

Levin, S.A., Dushoff, J. and Plotkin, J.B. (2004) Evolution and persistence of influenza A and other diseases. *Math. Biosci.* **188**, 17–28.

Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J.H. *et al.* (2005) Bats are natural reservoirs of SARS-like coronaviruses. *Science* **310**, 676–679.

Lindell, D., Sullivan, M.B., Johnson, Z.I., Tolonen, A. C., Rohwer, F. and Chisholm, S.W. (2004) Transfer of photosynthesis genes to and from Prochlorococcus viruses. *Proc. Natl Acad. Sci. USA* **101**, 11013–11018.

Liu, J., Glazko, G. and Mushegian, A. (2006) Protein repertoire of double-stranded DNA bacteriophages. *Virus Res.* **117**, 68–80.

Lobocka, M.B., Svarchevsky, A.N., Rybchin, V.N. and Yarmolinsky, M.B. (1996) Characterization of the primary immunity region of the *Escherichia coli* linear plasmid prophage N15. *J. Bacteriol.* **178**, 2902–2910.

Lopez-Bueno, A., Villarreal, L.P. and Almendral, J.M. (2006) Parvovirus variation for disease: a difference with RNA viruses?. *Curr. Top. Microbiol. Immunol.* **299**, 349–370.

Louz, D., Bergmans, H.E., Loos, B.P. and Hoeben, R.C. (2005) Cross-species transfer of viruses: implications for the use of viral vectors in biomedical research, gene therapy and as live-virus vaccines. *J. Gene Med.* **7**, 1263–1274.

Luria, S.E. (1950) Bacteriophage: an essay on virus reproduction. *Science* **111**, 507–511.

Luria, S.E. and Delbruck, M. (1943) Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28**, 491–511.

Luria, S.E., Kellenberger, E., Harrison, B.D., Schafer, W., Hirst, G.K., Isaacs, A. *et al.* (1959) *Virus growth and*

variation: Ninth symposium of the society for general microbiology. London: Cambridge University Press.

Lussier, G. and Descoteaux, J.P. (1986) Prevalence of natural virus infections in laboratory mice and rats used in Canada. *Lab. Anim. Sci.* **36**, 145–148.

Makarova, K.S., Wolf, Y.I., Tereza, E.P. and Ratner, V.A. (1998) Different patterns of molecular evolution of influenza A viruses in avian and human populations. *Genetika* **34**, 890–896.

Marcotte, L.L., Wass, A.B., Gohara, D.W., Pathak, H.B., Arnold, J.J., Filman, D.J. *et al.* (2007) Crystal structure of poliovirus 3CD protein: virally encoded protease and precursor to the RNA-dependent RNA polymerase. *J. Virol.* **81**, 3583–3596.

Margulis, L. and Bermudes, D. (1985) Symbiosis as a mechanism of evolution: status of cell symbiosis theory. *Symbiosis* **1**, 101–124.

Marquez, L.M., Redman, R.S., Rodriguez, R.J. and Roossinck, M.J. (2007) A virus in a fungus in a plant: three-way symbiosis required for thermal tolerance. *Science* **315**, 513–515.

Martell, M., Esteban, J.I., Quer, J., Genesca, J., Weiner, A., Esteban, R. *et al.* (1992) Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. *J. Virol.* **66**, 3225–3229.

Marten, N.W., Stohlman, S.A. and Bergmann, C.C. (2001) MHV infection of the CNS: mechanisms of immune-mediated control. *Viral Immunol.* **14**, 1–18.

Martinez, J.M., Schroeder, D.C., Larsen, A., Bratbak, G. and Wilson, W.H. (2007) Molecular dynamics of *Emiliania huxleyi* and cooccurring viruses during two separate mesocosm studies. *Appl. Environ. Microbiol.* **73**, 554–562.

Marty, G.D., Quinn, T.J., Carpenter, G., Meyers, T.R. and Willits, N.H. (2003) Role of disease in abundance of a Pacific herring (*Clupea pallasi*) population. *Can. J. Fisheries Aquat. Sci.* **60**, 1258–1265.

Mazel, D., Dychinco, B., Webb, V.A. and Davies, J. (2000) Antibiotic resistance in the ECOR collection: integrons and identification of a novel aad gene. *Antimicrob. Agents Chemother.* **44**, 1568–1574.

McGeehan, J.E., Depledge, N.W. and McGeoch, D.J. (2001) Evolution of the dUTPase gene of mammalian and avian herpesviruses. *Curr. Protein Pept. Sci.* **2**, 325–333.

McGeoch, D.J. and Davison, A.J. (1999) The descent of human herpesvirus 8. *Semin. Cancer Biol.* **9**, 201–209.

McGeoch, D.J. and Gatherer, D. (2005) Integrating reptilian herpesviruses into the family herpesviridae. *J. Virol.* **79**, 725–731.

McGeoch, D.J., Dolan, A. and Ralph, A.C. (2000) Toward a comprehensive phylogeny for mammalian and avian herpesviruses. *J. Virol.* **74**, 10401–10406.

McGeoch, D.J., Rixon, F.J. and Davison, A.J. (2006) Topics in herpesvirus genomics and evolution. *Virus Res.* **117**, 90–104.

McIntosh, E.M. and Haynes, R.H. (1996) HIV and human endogenous retroviruses: an hypothesis with therapeutic implications. *Acta Biochim. Pol.* **43**, 583–592.

Mesyanzhinov, V.V., Robben, J., Grymonprez, B., Kostyuchenko, V.A., Bourkaltseva, M.V., Sykilinda, N.N. *et al.* (2002) The genome of bacteriophage phiKZ of *Pseudomonas aeruginosa*. *J. Mol. Biol.* **317**, 1–19.

Mi, S., Lee, X., Li, X., Veldman, G.M., Finnerty, H., Racie, L. *et al.* (2000) Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403**, 785–789.

Mirkin, B.G., Fenner, T.I., Galperin, M.Y. and Koonin, E.V. (2003) Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. *BMC Evol. Biol.* **3**, 2.

Mondy, W.L. and Pierce, S.K. (2003) Apoptotic-like morphology is associated with annual synchronized death in kleptoplastic sea slugs (*Elysia chlorotica*). *Invertebrate Biol.* **122**, 126–137.

Moran, N.A., Degnan, P.H., Santos, S.R., Dunbar, H.E. and Ochman, H. (2005) The players in a mutualistic symbiosis: insects, bacteria, viruses and virulence genes. *Proc. Natl Acad. Sci. USA* **102**, 16919–16926.

Morand-Joubert, L., Charpentier, C., Poizat, G., Chene, G., Dam, E., Raguin, G. *et al.* (2006) Low genetic barrier to large increases in HIV-1 cross-resistance to protease inhibitors during salvage therapy. *Antivir. Ther.* **11**, 143–154.

Moreira, D. and Lopez-Garcia, P. (2005) Comment on 'The 1.2-megabase genome sequence of Mimivirus.' *Science* **308**, 1114.

Moro, D., Lloyd, M.L., Smith, A.L., Shellam, G.R. and Lawson, M.A. (1999) Murine viruses in an island population of introduced house mice and endemic short-tailed mice in Western Australia. *J. Wildl. Dis.* **35**, 301–310.

Moro, D., Lawson, M.A., Hobbs, R.P. and Thompson, R.C. (2003) Pathogens of house mice on arid Boullanger Island and subantarctic Macquarie Island, Australia. *J. Wildl. Dis.* **39**, 762–771.

Mura, M., Murcia, P., Caporale, M., Spencer, T.E., Nagashima, K., Rein, A. and Palmarini, M. (2004) Late viral interference induced by transdominant Gag of an endogenous retrovirus. *Proc. Natl Acad. Sci. USA* **101**, 11117–11122.

Murcia, P.R., Arnaud, F. and Palmarini, M. (2007) The transdominant endogenous retrovirus enJS56A1 associates with and blocks intracellular trafficking of Jaagsiekte sheep retrovirus Gag. *J. Virol.* **81**, 1762–1772.

Murray, S.M. and Linial, M.L. (2006) Foamy virus infection in primates. *J. Med. Primatol.* **35**, 225–235.

Mushegian, A. (1999) The minimal genome concept. *Curr. Opin. Genet. Dev.* **9**, 709–714.

Nagasaki, K., Shirai, Y., Tomaru, Y., Nishida, K. and Pietrokovski, S. (2005) Algal viruses with distinct intraspecies host specificities include identical intein elements. *Appl. Environ. Microbiol.* **71**, 3599–3607.

Nakamura, A., Okazaki, Y., Sugimoto, J., Oda, T. and Jinno, Y. (2003) Human endogenous retroviruses with

transcriptional potential in the brain. *J. Hum. Genet.* **48**, 575–581.

Nandhagopal, N., Simpson, A.A., Gurnon, J.R., Yan, X., Baker, T.S., Graves, M.V. *et al.* (2002) The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus. *Proc. Natl Acad. Sci. USA* **99**, 14758–14763.

Narechania, A., Terai, M., Chen, Z., DeSalle, R. and Burk, R.D. (2004) Lack of the canonical pRB-binding domain in the E7 ORF of artiodactyl papillomaviruses is associated with the development of fibropapillomas. *J. Gen. Virol.* **85**, 1243–1250.

Naryshkina, T., Liu, J., Florens, L., Swanson, S.K., Pavlov, A.R., Pavlova, N.V. *et al.* (2006) *Thermus thermophilus* bacteriophage phiYS40 genome and proteomic characterization of virions. *J. Mol. Biol.* **364**, 667–677.

Nash, A.A., Dutia, B.M., Stewart, J.P. and Davison, A.J. (2001) Natural history of murine gamma-herpesvirus infection. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **356**, 569–579.

Ndunguru, J., Legg, J.P., Aveling, T.A., Thompson, G. and Fauquet, C.M. (2005) Molecular biodiversity of cassava begomoviruses in Tanzania: evolution of cassava geminiviruses in Africa and evidence for East Africa being a center of diversity of cassava geminiviruses. *Virol. J.* **2**, 21.

Nelson, M.I. and Holmes, E.C. (2007) The evolution of epidemic influenza. *Nat. Rev. Genet.* **8**, 196–205.

Nicholas, J., Zong, J.C., Alcendor, D.J., Ciufo, D.M., Poole, L.J., Sarisky, R.T. *et al.* (1998) Novel organizational features, captured cellular genes and strain variability within the genome of KSHV/HHV8. *J. Natl Cancer Inst. Monogr.* **23**, 79–88.

Nilsson, A.S., Karlsson, J.L. and Haggard-Ljungquist, E. (2004) Site-specific recombination links the evolution of P2-like coliphages and pathogenic enterobacteria. *Mol. Biol. Evol.* **21**, 1–13.

Nolan, J.M., Petrov, V., Bertrand, C., Krisch, H.M. and Karam, J.D. (2006) Genetic diversity among five T4-like bacteriophages. *Virol. J.* **3**, 30.

Novella, I.S., Duarte, E.A., Elena, S.F., Moya, A., Domingo, E. and Holland, J.J. (1995) Exponential increases of RNA virus fitness during large population transmissions. *Proc. Natl Acad. Sci. USA* **92**, 5841–5844.

Novella, I.S., Hershey, C.L., Escarmis, C., Domingo, E. and Holland, J.J. (1999) Lack of evolutionary stasis during alternating replication of an arbovirus in insect and mammalian cells. *J. Mol. Biol.* **287**, 459–465.

Ogata, H., Raoult, D. and Claverie, J.M. (2005) A new example of viral intein in Mimivirus. *Virol. J.* **2**, 8.

OhAinle, M., Kerns, J.A., Malik, H.S. and Emerman, M. (2006) Adaptive evolution and antiviral activity of the conserved mammalian cytidine deaminase APOBEC3H. *J. Virol.* **80**, 3853–3862.

Okamoto, H. and Mishiro, S. (1994) Genetic heterogeneity of hepatitis C virus. *Intervirology* **37**, 68–76.

Oliveira, N.M., Farrell, K.B. and Eiden, M.V. (2006) In vitro characterization of a koala retrovirus. *J. Virol.* **80**, 3104–3107.

Orgel, L.E. and Crick, F.H. (1980) Selfish DNA: the ultimate parasite. *Nature* **284**, 604–607.

Ortmann, A.C., Wiedenheft, B., Douglas, T. and Young, M. (2006) Hot crenarchaeal viruses reveal deep evolutionary connections. *Nat. Rev. Microbiol.* **4**, 520–528.

Pagel, M.D. (2002) *Encyclopedia of Evolution*. Oxford and New York: Oxford University Press.

Parsyan, A., Szmaragd, C., Allain, J.P. and Candotti, D. (2007) Identification and genetic diversity of two human parvovirus B19 genotype 3 subtypes. *J. Gen. Virol.* **88**, 428–431.

Paul, J.H., Sullivan, M.B., Segall, A.M. and Rohwer, F. (2002) Marine phage genomics. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **133**, 463–476.

Persson, R., McGeehan, J. and Wilson, K.S. (2005) Cloning, expression, purification and characterisation of the dUTPase encoded by the integrated *Bacillus subtilis* temperate bacteriophage SPbeta. *Protein Expr. Purif.* **42**, 92–99.

Pfeiffer, J.K. and Kirkegaard, K. (2005) Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. *PLoS Pathog.* **1**, e11.

Pfeiffer, J.K. and Kirkegaard, K. (2006) Bottleneck-mediated quasispecies restriction during spread of an RNA virus from inoculation site to brain. *Proc. Natl Acad. Sci. USA* **103**, 5520–5525.

Pierce, S.K., Maugel, T.K., Rumpho, M.E., Hanten, J.J. and Mondy, W.L. (1999) Annual viral expression in a sea slug population: Life cycle control and symbiotic chloroplast maintenance. *Biol. Bull.* **197**, 1–6.

Prangishvili, D., Forterre, P. and Garrett, R.A. (2006a) Viruses of the Archaea: a unifying view. *Nat. Rev. Microbiol.* **4**, 837–848.

Prangishvili, D., Garrett, R.A. and Koonin, E.V. (2006b) Evolutionary genomics of archaeal viruses: unique viral genomes in the third domain of life. *Virus Res.* **117**, 52–67.

Putics, A., Filipowicz, W., Hall, J., Gorbalenya, A.E. and Ziebuhr, J. (2005) ADP-ribose-1″-monophosphatase: a conserved coronavirus enzyme that is dispensable for viral replication in tissue culture. *J. Virol.* **79**, 12721–12731.

Pybus, O.G., Rambaut, A., Belshaw, R., Freckleton, R.P., Drummond, A.J. and Holmes, E.C. (2007) Phylogenetic evidence for deleterious mutation load in RNA viruses and its contribution to viral evolution. *Mol. Biol. Evol.* **24**, 845–852.

Qiu, W. and Scholthof, K.B. (2001) Defective interfering RNAs of a satellite virus. *J. Virol.* **75**, 5429–5432.

Randall, A.Z., Baldi, P. and Villarreal, L.P. (2004) Structural proteomics of the poxvirus family. *Artif. Intell. Med.* **31**, 105–115.

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H. *et al.* (2004) The 1.2-megabase genome sequence of Mimivirus. *Science* **306**, 1344–1350.

Rice, G., Tang, L., Stedman, K., Roberto, F., Spuhler, J., Gillitzer, E. *et al.* (2004) The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. *Proc. Natl Acad. Sci. USA* **101**, 7716–7720.

Roossinck, M.J. (2003) Plant RNA virus evolution. *Curr. Opin. Microbiol.* **6**, 406–409.

Roossinck, M.J. (2005) Symbiosis versus competition in plant virus evolution. *Nat. Rev. Microbiol.* **3**, 917–924.

Roossinck, M.J. and Schneider, W.L. (2006) Mutant clouds and occupation of sequence space in plant RNA viruses. *Curr. Top. Microbiol. Immunol.* **299**, 337–348.

Rowe, C.L., Baker, S.C., Nathan, M.J., Sgro, J.Y., Palmenberg, A.C. and Fleming, J.O. (1998) Quasispecies development by high frequency RNA recombination during MHV persistence. *Adv. Exp. Med. Biol.* **440**, 759–765.

Ruiz-Jarabo, C.M., Arias, A., Baranowski, E., Escarmis, C. and Domingo, E. (2000) Memory in viral quasispecies. *J. Virol.* **74**, 3543–3547.

Ryan, F.P. (2007) Viruses as symbionts. *Symbiosis* **44**. (in press)

Saren, A.M., Ravantti, J.J., Benson, S.D., Burnett, R.M., Paulin, L., Bamford, D.H. and Bamford, J.K. (2005) A snapshot of viral evolution from genome analysis of the tectiviridae family. *J. Mol. Biol.* **350**, 427–440.

Sawyer, S.L., Emerman, M. and Malik, H.S. (2004) Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biol.* **2**, E275.

Schoondermark-van de Ven, E.M., Philipse-Bergmann, I.M. and van der Logt, J.T. (2006) Prevalence of naturally occurring viral infections, *Mycoplasma pulmonis* and *Clostridium piliforme* in laboratory rodents in Western Europe screened from 2000 to 2003. *Lab. Anim.* **40**, 137–143.

Schroeder, D.C., Oke, J., Hall, M., Malin, G. and Wilson, W.H. (2003) Virus succession observed during an *Emiliania huxleyi* bloom. *Appl. Environ. Microbiol.* **69**, 2484–2490.

Shackelton, L.A. and Holmes, E.C. (2006) Phylogenetic evidence for the rapid evolution of human B19 erythrovirus. *J. Virol.* **80**, 3666–3669.

Shadan, F.F. and Villarreal, L.P. (1995) The evolution of small DNA viruses of eukaryotes: Past and present considerations. *Virus Genes* **11**, 239–257.

Sharp, G.B., Kawaoka, Y., Jones, D.J., Bean, W.J., Pryor, S.P., Hinshaw, V. and Webster, R.G. (1997) Coinfection of wild ducks by influenza a viruses: Distribution patterns and biological significance. *J. Virol.* **71**, 6128–6135.

Shutt, T.E. and Gray, M.W. (2006) Bacteriophage origins of mitochondrial replication and transcription proteins. *Trends Genet.* **22**, 90–95.

Simon, A.E., Roossinck, M.J. and Havelda, Z. (2004) Plant virus satellite and defective interfering RNAs: new paradigms for a new century. *Annu. Rev. Phytopathol.* **42**, 415–437.

Singleton, G.R., Smith, A.L., Shellam, G.R., Fitzgerald, N. and Muller, W.J. (1993) Prevalence of viral antibodies and helminths in field populations of house mice (Mus domesticus) in southeastern Australia. *Epidemiol. Infect.* **110**, 399–417.

Smith, D.B., Pathirana, S., Davidson, F., Lawlor, E., Power, J., Yap, P.L. and Simmonds, P. (1997) The origin of hepatitis C virus genotypes. *J. Gen. Virol.* **78**, 321–328.

Smith, G.J., Fan, X.H., Wang, J., Li, K.S., Qin, K., Zhang, J.X. *et al.* (2006) Emergence and predominance of an H5N1 influenza variant in China. *Proc. Natl Acad. Sci. USA*, **103**, 16936–16941.

Spackman, E., Stallknecht, D.E., Slemons, R.D., Winker, K., Suarez, D.L., Scott, M. and Swayne, D.E. (2005) Phylogenetic analyses of type A influenza genes in natural reservoir species in North America reveals genetic variation. *Virus Res.* **114**, 89–100.

Stadler, B.M.R., Stadler, P.F. and Schuster, P. (2000) Dynamics of autocatalytic replicator networks based on higher-order ligation reactions. *Bull. Math. Biol.* **62**, 1061–1086.

Stoye, J.P. (2006) Koala retrovirus: a genome invasion in real time. *Genome Biol.* **7**, 241.

Stuhler, A., Flory, E., Wege, H., Lassmann, H. and Wege, H. (1997) No evidence for quasispecies populations during persistence of the coronavirus mouse hepatitis virus JHM: sequence conservation within the surface glycoprotein gene S in Lewis rats. *J. Gen. Virol.* **78**, 747–756.

Suarez, D.L. (2000) Evolution of avian influenza viruses. *Vet. Microbiol.* **74**, 15–27.

Sullivan, M.B., Lindell, D., Lee, J.A., Thompson, L.R., Bielawski, J.P. and Chisholm, S.W. (2006) Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *PLoS Biol.* **4**.

Summer, E.J., Gonzalez, C.F., Bomer, M., Carlile, T., Embry, A., Kucherka, A.M. *et al.* (2006) Divergence and mosaicism among virulent soil phages of the *Burkholderia cepacia* complex. *J. Bacteriol.* **188**, 255–268.

Suzan-Monti, M., Scola, B.L., Barrassi, L., Espinosa, L. and Raoult, D. (2007) Ultrastructural characterization of the giant volcano-like virus factory of *Acanthamoeba polyphaga* mimivirus. *PLoS ONE* **2**, e328.

Switzer, W.M., Salemi, M., Shanmugam, V., Gao, F., Cong, M.E., Kuiken, C. *et al.* (2005) Ancient co-speciation of simian foamy viruses and primates. *Nature* **434**, 376–380.

Takemura, M. (2001) Poxviruses and the origin of the eukaryotic nucleus. *J. Mol. Evol.* **52**, 419–425.

Tang, X.C., Zhang, J.X., Zhang, S.Y., Wang, P., Fan, X.H., Li, L.F. *et al.* (2006) Prevalence and genetic diversity of coronaviruses in bats from China. *J. Virol.* **80**, 7481–7490.

Tardy-Panit, M., Blondel, B., Martin, A., Tekaia, F., Horaud, F. and Delpeyroux, F. (1993) A mutation in the RNA polymerase of poliovirus type 1 contributes to attenuation in mice. *J. Virol.* **67**, 4630–4638.

Tarlinton, R.E., Meers, J. and Young, P.R. (2006) Retroviral invasion of the koala genome. *Nature* **442**, 79–81.

Van Etten, J.L. (2003) Unusual life style of giant chlorella viruses. *Annu. Rev. Genet.* **37**, 153–195.

Van Etten, J.L., Graves, M.V., Muller, D.G., Boland, W. and Delaroque, N. (2002) Phycodnaviridae—large DNA algal viruses. *Arch. Virol.* **147**, 1479–1516.

Vetsigian, K., Woese, C. and Goldenfeld, N. (2006) Collective evolution and the genetic code. *Proc. Natl Acad. Sci. USA* **103**, 10696–10701.

Vignuzzi, M., Stone, J.K. and Andino, R. (2005) Ribavirin and lethal mutagenesis of poliovirus: molecular mechanisms, resistance and biological implications. *Virus Res.* **107**, 173–181.

Vignuzzi, M., Stone, J.K., Arnold, J.J., Cameron, C.E. and Andino, R. (2006) Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* **439**, 344–348.

Vijaykrishna, D., Smith, G.J., Zhang, J.X., Peiris, J.S., Chen, H. and Guan, Y. (2007) Evolutionary insights into the ecology of coronaviruses. *J. Virol.* **81**, 4012–4020.

Villarreal, L.P. (1999) DNA virus contribution to host evolution. In: *Origin and Evolution of Viruses* (E. Domingo, R.G. Webster and J.J. Holland, eds), pp. 391–420. San Diego: Academic Press.

Villarreal, L.P. (2005) *Viruses and the Evolution of Life*. Washington, DC: ASM Press.

Villarreal, L.P. (2006) How viruses shape the tree of life. *Future Virol.* **1**, 587–595.

Villarreal, L.P. (2007) Virus-host symbiosis mediated by persistence. *Symbiosis* **44**, 1–9.

Villarreal, L.P. and DeFilippis, V.R. (2000) A hypothesis for DNA viruses as the origin of eukaryotic replication proteins. *J. Virol.* **74**, 7079–7084.

Villarreal, L.P. and Villareal, L.P. (1997) On viruses, sex and motherhood. *J. Virol.* **71**, 859–865.

Villarreal, L.P., Defilippis, V.R. and Gottlieb, K.A. (2000) Acute and persistent viral life strategies and their relationship to emerging diseases. *Virology* **272**, 1–6.

Vogel, T.U., Evans, D.T., Urvater, J.A., O'Connor, D.H., Hughes, A.L. and Watkins, D.I. (1999) Major histocompatibility complex class I genes in primates: co-evolution with pathogens. *Immunol. Rev.* **167**, 327–337.

Wallensten, A., Munster, V.J., Karlsson, M., Lundkvist, A., Brytting, M., Stervander, M. *et al.* (2006) High prevalence of influenza A virus in ducks caught during spring migration through Sweden. *Vaccine.* .

Wang, B., Mikhail, M., Dyer, W.B., Zaunders, J.J., Kelleher, A.D. and Saksena, N.K. (2003) First demonstration of a lack of viral sequence evolution in a nonprogressor, defining replication-incompetent HIV-1 infection. *Virology* **312**, 135–150.

Wang, L.F., Shi, Z., Zhang, S., Field, H., Daszak, P. and Eaton, B.T. (2006) Review of bats and SARS. *Emerg. Infect. Dis.* **12**, 1834–1840.

Wang, N., Baldi, P.F. and Gaut, B.S. (2006) Phylogenetic analysis, genome evolution and the rate of gene gain in the Herpesviridae. *Mol. Phylogenet. Evol.*

Watkins, D.I. (1995) The evolution of major histocompatibility class I genes in primates. *Crit. Rev. Immunol.* **15**, 1–29.

Webster, R.G. and Govorkova, E.A. (2006) H5N1 influenza—continuing evolution and spread. *N. Engl. J. Med.* **355**, 2174–2177.

Weir, E.C., Bhatt, P.N., Barthold, S.W., Cameron, G.A. and Simack, P.A. (1987) Elimination of mouse hepatitis virus from a breeding colony by temporary cessation of breeding. *Lab. Anim. Sci.* **37**, 455–458.

Widjaja, L., Krauss, S.L., Webby, R.J., Xie, T. and Webster, R.G. (2004) Matrix gene of influenza a viruses isolated from wild aquatic birds: Ecology and emergence of influenza A viruses. *J. Virol.* **78**, 8771–8779.

Wilberz, S., Partke, H.J., Dagnaes-Hansen, F. and Herberg, L. (1991) Persistent MHV (mouse hepatitis virus) infection reduces the incidence of diabetes mellitus in non-obese diabetic mice. *Diabetologia* **34**, 2–5.

Wilke, C.O. (2005) Quasispecies theory in the context of population genetics. *BMC Evol. Biol.* **5**, 44.

Wolf, Y.I., Viboud, C., Holmes, E.C., Koonin, E.V. and Lipman, D.J. (2006) Long intervals of stasis punctuated by bursts of positive selection in the seasonal evolution of influenza A virus. *Biol. Direct*, **1**, 34.

Yarmolinsky, M.B. (2000) A pot-pourri of plasmid paradoxes: effects of a second copy. *Mol. Microbiol.* **38**, 1–7.

Yarmolinsky, M.B. (2004) Bacteriophage P1 in retrospect and in prospect. *J. Bacteriol.* **186**, 7025–7028.

Yi, J.M., Kim, T.H., Huh, J.W., Park, K.S., Jang, S.B., Kim, H.M. and Kim, H.S. (2004) Human endogenous retroviral elements belonging to the HERV-S family from human tissues, cancer cells and primates: expression, structure, phylogeny and evolution. *Gene* **342**, 283–292.

Zanotto, P.M., Gibbs, M.J., Gould, E.A. and Holmes, E.C. (1996) A reevaluation of the higher taxonomy of viruses based on RNA polymerases. *J. Virol.* **70**, 6083–6096.

Zhang, Y., Moriyama, H., Homma, K. and Van Etten, J.L. (2005) Chlorella virus-encoded deoxyuridine triphosphatases exhibit different temperature optima. *J. Virol.* **79**, 9945–9953.

Zhu, T., Corey, L., Hwangbo, Y., Lee, J.M., Learn, G.H., Mullins, J.I. and McElrath, M.J. (2003) Persistence of extraordinarily low levels of genetically homogeneous human immunodeficiency virus type 1 in exposed seronegative individuals. *J. Virol.* **77**, 6108–6116.

Zhu, X.Y. (2003) [Phylogenetic reconstruction of DNA polymerase X family]. *Yi Chuan Xue Bao* **30**, 867–872.