

## The Evolutionary Genetics of Viral Emergence

E. C. Holmes<sup>1</sup> (✉) · A. J. Drummond<sup>2</sup> (✉)

<sup>1</sup>Center for Infectious Disease Dynamics, Department of Biology, Mueller Laboratory,  
The Pennsylvania State University, University Park, PA 16802 USA  
*ech15@psu.edu*

<sup>2</sup>Department of Computer Science, University of Auckland, Private Bag 92019  
Auckland New Zealand  
*alexei@cs.auckland.ac.nz*

1	Introduction .....	52
2	Are Certain Types of Virus More Likely to Emerge than Others? .....	53
3	Are Viruses from Phylogenetically Related Host Species More Likely to Experience Cross-Species Transmission? .....	56
4	Does Emergence Require Adaptation Within the New Host Species? .....	58
5	Is Recombination a Prerequisite for Viral Emergence? .....	61
6	Conclusions: Evolution and Emergence in RNA Viruses .....	63
	References .....	63

**Abstract** Despite the wealth of data describing the ecological factors that underpin viral emergence, little is known about the evolutionary processes that allow viruses to jump species barriers and establish productive infections in new hosts. Understanding the evolutionary basis to virus emergence is therefore a key research goal and many of the debates in this area can be considered within the rigorous theoretical framework established by evolutionary genetics. In particular, the respective roles played by natural selection and genetic drift in shaping genetic diversity are also of fundamental importance for understanding the nature of viral emergence. Herein, we discuss whether there are evolutionary rules to viral emergence, and especially whether certain types of virus, or those that infect a particular type of host species, are more likely to emerge than others. We stress the complex interplay between rates of viral evolution and the ability to recognize cell receptors from phylogenetically divergent host species. We also emphasize the current lack of convincing data as to whether viral emergence requires adaptation to the new host species during the early stages of infection, or whether it is largely a chance process involving the transmission of a viral strain with the necessary genetic characteristics.

## 1 Introduction

Until recently, studies of emerging viruses frequently involved compiling lists of pathogens that were considered new to human populations or that had increased in frequency and geographical range, and describing the ecological factors responsible for their appearance. Such studies gave particular emphasis to how changes in human ecology, notably increases in population size, modifications in land use, global travel, and political upheavals, had been responsible for an elevated burden of infectious disease, often by increasing the proximity and/or density of possible reservoir populations (Morse 1995). What was largely absent from these studies was a consideration of the evolutionary processes that underlie viral emergence (see the chapter by Childs et al., this volume). Indeed, the main role played by evolutionary biology in the first studies of emerging viruses was to reconstruct the origin and spread of new pathogens, largely through phylogenetic analysis (for example, Nichol et al. 1993). Although the focus on changing human ecology and phylogenetic history was an important and necessary first step, and it will often be impossible to disentangle ecology and genetics when explaining the emergence of a specific pathogen, it is also essential to ask what evolutionary processes are responsible for the appearance and spread of pathogens? Indeed, for the study of emerging viruses to come of age, it is critical to determine whether there are any general evolutionary rules governing the process of emergence.

Evolutionary genetics aims to understand the processes responsible for the origin and maintenance of genetic variation in populations. The great obsession of evolutionary genetics has been to reveal the respective roles of random fluctuations in allele frequencies—genetic drift—and the natural selection of advantageous mutations in shaping genetic diversity (Gillespie 1998). Although, at face value, evolutionary genetics may seem of little relevance to the problem of viral emergence, we suggest that it in fact provides an essential theoretical framework. For example, it might be that emergence simply requires the chance exposure of a virus to a new susceptible population, with little involvement of natural selection. Alternatively, it may be that viruses have to adapt to successfully spread in a new species. In this scenario, different species might represent different fitness peaks on an “adaptive landscape” and that traversing between these peaks is difficult because they are connected by valleys of low fitness. Moreover, because an emergent virus will only infect a small number of individuals when it first enters a new population, genetic drift is expected to play a major role in determining what viral mutations get fixed, because drift is more potent in small populations. Finally, the respective influences of drift and selection will also vary according to

the size of the population bottleneck that accompanies viral transmission among hosts, which will itself be a function of the mode of transmission.

In this paper we address some of the key questions surrounding the evolutionary genetics of viral emergence. Although evolutionary genetics relates equally well to hosts (such as differences in susceptibility and immune responses) as well as pathogens, we will concentrate on the latter where data is more abundant and their evolution can be tracked more readily. To focus our paper, we will consider four key questions:

1. Are certain types of virus more likely to emerge than others?
2. Are viruses from phylogenetically related host species more likely to undergo cross-species transmission than those viruses from distantly related host species?
3. Does viral emergence require adaptation to the new host species?
4. Is recombination a prerequisite for viral emergence?

Although definitive answers will not be forthcoming in all cases, we discuss the data required to resolve these issues. Finally, as well as shedding new light on the process of emergence in particular, answering these questions will provide more general insights into the nature of viral evolution.

## 2

### **Are Certain Types of Virus More Likely to Emerge than Others?**

The broadest division in virus classification is between those viruses in which the genome is composed of DNA (DNA viruses) and those where the genomic nucleic acid comprises RNA (RNA viruses), with the latter also usually considered to include retroviruses that make a DNA copy of the RNA genome through reverse transcription. Although most known viruses have RNA genomes, even accounting for this bias it is clear that RNA viruses are more often associated with emerging diseases than DNA viruses (Cleaveland et al. 2001; Woolhouse 2002; see the chapter by Cleaveland et al., this volume). In contrast, DNA viruses are often associated with a process of virus–host co-speciation that can extend many millions of years. This is perhaps because DNA viruses more often establish persistent infections than RNA viruses and so can more easily track host evolution (Holmes 2004).

That RNA viruses seem to possess inherently more “emergibility” than DNA viruses has usually been put down to their very rapid rates of evolutionary

change. RNA viruses are thought to evolve many logs faster than DNA viruses because of a combination of highly error-prone replication with RNA polymerase or reverse transcriptase, large population sizes, and rapid replication rates (Domingo and Holland 1997; Moya et al. 2004). In turn, a rapid rate of evolutionary change allows RNA viruses to quickly generate the mutations that might be required to adapt them to new environments, including new host species. Although this effect is broadly true, there is still substantial variation among RNA viruses in their ability to cause emergent diseases. Understanding the basis of this variation is critical to the development of an evolutionary model of viral emergence and for understanding the constraints on RNA virus evolution in general.

If the rate of evolutionary change is driven by the rate at which neutral variants are generated, then the most important factors determining rates of evolutionary change are the replication error rate and the generation time. The variation in error rates among viruses is a subject of considerable research activity (Malpica et al. 2002; Pugachev et al. 2004) and there is a growing body of data on how generation times vary within and among viruses (Markowitz et al. 2003; Whalley et al. 2001). However, despite a wealth of population genetic theory about the interplay between these factors and natural selection, we are still some way from producing an all-encompassing picture of what determines rates of evolutionary change in viruses. This is in part due to a lack of detailed and comprehensive experimental measurements of the critical parameters and partly because of the limited complexity of available analytical models when compared to the reality of viral evolution.

Moreover, not all RNA and DNA viruses neatly fit the picture of rapid and slow evolution, respectively. The most dramatic example is that of simian foamy virus (SFV), which has co-diverged with its primate hosts over many million of years and evolved at the remarkably low rate of approximately  $10^{-8}$  substitutions per site, per year, similar to that seen in that in host mitochondrial DNA (Switzer et al. 2005). The most likely explanation for this low rate of change is a combination of greatly reduced rates of replication and strong purifying selection, so that the vast majority of mutations that arise are deleterious and selectively eliminated. At the other end of the evolutionary spectrum, there is growing evidence that single-stranded (ss) DNA viruses evolve at rates approaching those seen in RNA viruses, most likely because of high intrinsic error rates (Sanz et al. 1999; Shackleton et al. 2005).

Although mutation provides the raw materials for evolutionary change, it is the dual processes of genetic drift and natural selection that are the ultimate arbiters over what genetic variation remains in the long term. Because all viruses have an absolute dependence on host cellular machinery for a productive life cycle, the interaction between viral proteins and the cellular receptors

of host cells makes up a large part of the viruses' fitness landscape. We would therefore expect the interaction between specific viral sequences and cellular receptors to be of particular importance in determining why some RNA viruses are more often associated with cross-species transmission than others (Baranowski et al. 2001). For example, avian influenza A viruses are usually unable to evolve human-to-human transmission because they lack the correct amino acids in a number of viral proteins (see the chapter by Webby et al., this volume). Most attention has been directed toward the viral haemagglutinin (HA), which requires specific amino acids to interact with the sialic receptors on human cells in the correct configuration (Scholtissek et al. 1993; see Sect. 4). The intimacy of the relationship between virus and cell receptor also predicts that generalist viruses, which infect a broad range of cellular receptors, are more able to cross species boundaries than specialist viruses that have a narrower tropism. Strikingly, provisional analyses suggest that this is indeed the case, as viruses that utilize conserved cell receptors are more able to jump species boundaries than viruses that use divergent cell receptors (Woolhouse 2002). If validated with more data, this result is of great importance because it allows predictions to be made as to the types of virus that are most likely to emerge in the future.

Another factor that might influence the ability of viruses to cross species boundaries is the mode of transmission. It is easy to imagine how certain types of transmission mechanism—particularly respiratory and vector-borne transmission—might more readily facilitate viral emergence than others. In both these cases, the probability of exposure to an emergent virus is relatively high compared to viruses that are blood-borne or sexually transmitted. Indeed, it is striking that many of the viruses described in the earliest lists of emerging viruses were transmitted by mosquito vectors (the arboviruses), which could potentially take blood meals from a range of mammalian hosts. However, although an increased probability of exposure will generally translate into an increased likelihood of emergence (see the chapters by Childs et al. and Real and Biek, this volume), there are a number of other factors that this simple calculus omits. For example, there appears to be an association between mode of transmission and the ability of a virus to successfully replicate in the cells of a new host species. This is best documented with the arboviruses where there is strong evidence from both comparative and *in vitro* studies that the necessity of replicating in very different host species, in this case arthropods and mammals, imposes strong constraints against sequence change (Holmes 2003b; Woelk and Holmes 2002; Zárata and Novella 2004). This effect is most likely attributable to an antagonistic fitness trade-off, such that mutations that increase fitness in one host species reduce it in another. Hence, the majority of amino acid changes that arise in either host are deleterious (or slightly deleterious)

and eventually removed by purifying selection. In particular, it seems especially difficult to establish productive infections in insect cells (Zárate and Novella 2004), which perhaps explains why vector species seems to be a key correlate of evolution in some animal and plant RNA viruses (Gaunt et al. 2001; Chare and Holmes 2004). However, perhaps the most striking of all observations in this context is that although arboviruses are frequently associated with sporadic disease in humans, few are able to sustain long-term transmission networks and dead-end infections are commonplace (M.E.J. Woolhouse, personal communication). Hence, the intricate adaptations to replicate in divergent hosts may act to prevent many arboviruses from successful emergence in new host species. Whether similar constraints apply to viruses transmitted by other mechanisms is unknown.

### **3 Are Viruses from Phylogenetically Related Host Species More Likely to Experience Cross-Species Transmission?**

Central to our discussion so far has been the assumption that nearly all emerging viruses have jumped to humans from another animal species, in a process of cross-species transmission. Indeed, one of the great successes of molecular epidemiology has been the identification, often very rapidly, of the reservoir species for a myriad of human viruses. The rapid discoveries of the nonhuman primate ancestry of the human immunodeficiency virus (HIV) (Huet et al. 1990), and of some bat species as the ultimate reservoir of SARS coronavirus (SARS-CoV) (Lau et al. 2005; Li et al. 2005), serve as important illustrations. However, there are exceptions. Perhaps the most notable are the Ebola virus, where many thousands of animal specimens have been surveyed in Africa without certain identification of the reservoir species (Breman et al. 1999; Peterson et al. 2004), although bats have been recently implicated (see the chapter by Gonzalez et al., this volume) and hepatitis C virus, cause of one of the most prevalent somewhat new diseases to be identified in humans but where no close relatives have been discovered (Simmonds 2004). However, it is likely that with an increased sampling of taxa the species reservoirs for these viruses will also be determined, if they still exist today.

The next question that arises is whether some viruses are more able to jump species boundaries than others. The most compelling idea in this context is that there are phylogenetic constraints to this process, such that the more closely related the host species in question, the greater the chance of successful cross-species transmission (DeFilippis and Villarreal 2000). This theory is supported

by some broad-brush observations. In particular, there is no evidence that the viruses that infect humans come from organisms as divergent as plants, fish, reptiles, or amphibians (Holmes and Rambaut 2004), even though in some cases, such as plant viruses, exposure might occur on a regular basis through the consumption of infected food. Rather, the majority of human viruses are of mammalian origin, with an occasional few coming from birds. Moreover, although insect viruses often infect human populations (that is, the arboviruses), these always jump from another mammalian species rather than directly from insects and, as described above, tend to cause dead-end infections in new hosts (see the chapter by Nel and Rupprecht, this volume, for discussion of the origin of rhabdoviruses infecting mammals from a potential insect source).

A more revealing question is whether there are any phylogenetic trends with respect to the mammalian viruses that also affect humans. Specifically, are those viruses from our closest relatives, the simian primates, more able to infect us than those from other mammalian orders? At present there is insufficient data to fully test this hypothesis, although it is clearly a research priority for the future. Additionally, it is difficult to fully disentangle probability of transmission from probability of exposure; for example, although we are clearly more closely related to other primates than to rodents, the global human population is more often exposed to the latter. There are, however, some tentative pieces of evidence to suggest that primate viruses are especially able to infect us as predicted from our close evolutionary relationship. As well as the obvious cases of HIV-1 and HIV-2, whose ultimate origins lie with chimpanzees (*Pan troglodytes troglodytes*) and sooty mangabey monkeys (*Cercocebus torquatus atys*), respectively, a variety of other major human viruses seem to have their origins in nonhuman primates. These include dengue virus, yellow fever virus, GB viruses A and C, hepatitis B virus, and HTLV-I and II. That some of these viruses seem to have appeared relatively recently in humans may be a consequence of changing ecological factors, most notably deforestation and linked activities, that have increased the rate of contact between humans and other primates (although an absence of retrospective diagnoses makes it difficult to determine whether emerging viruses are more common now than at previous times in our evolutionary history). Turning the tables, there are also examples of humans transmitting their viruses to other primates (often with serious consequences), as appears to be the case for measles (Ferber 2000) and TTV (Okamoto et al. 2000). This will doubtless be a continuing problem, as perhaps will be the movement of viruses from the populations of industrialized nations to indigenous peoples with naïve genetic backgrounds.

There are also good mechanistic reasons for believing that there is a relationship between phylogenetic distance and the likelihood of viral emergence. In particular, if, as argued above, the ability to recognize and infect host cells is

a key component of cross-species transmission, then phylogenetically related host species are more likely to share related cell receptors. Given the pace at which RNA viruses evolve, it is easy to see that highly dependent relationships between viruses and cell receptors will be quickly established, so that the probability of successful cross-species transmission will decrease with increasing phylogenetic distance. If true, this theory further predicts that the more slowly evolving DNA viruses should initially be able to jump wider phylogenetic boundaries but, when they do adapt to their host, will eventually find it much more difficult to make subsequent species jumps. This pattern may partly explain the tendency for slower-evolving DNA viruses to co-diverge rather than move horizontally among species (Holmes 2004).

There are, however, some complicating factors to this simple phylogenetic rule. The issue of exposure vs transmissibility has been discussed above. There are also numerous exceptions to the phylogenetic trend. In particular, a large number of the emerging viruses of humans appear to have arisen from rodents rather than primates (M.E.J. Woolhouse, personal communication). This implies that the high density of many rodent populations allows them to carry a greater diversity of pathogens and/or that rodents often live in close proximity to humans which increases the probability of exposure (see the chapters by Gonzalez et al. and Klein and Calisher, this volume). Another factor of potential importance is phylogenetically related immune responses. Specifically, closely related host species, such as humans and other primates, are also likely to share the alleles that determine immune responses to specific pathogens. This has been particularly well documented for the major histocompatibility complex (MHC) group of loci, in which certain allelic lineages have persisted for millions of years of evolutionary history (Figuerola et al. 1988). Consequently, although a species might be exposed to a novel pathogen, they might, through a combination of shared common ancestry and good fortune, already possess a sufficient immune response to prevent the infection from being established.

#### **4 Does Emergence Require Adaptation Within the New Host Species?**

Understandably, most definitions of emerging viruses focus on the issue of disease. This means that no distinction is drawn between those viruses that spread efficiently among us, and those that only cause sporadic disease, often with no human-to-human transmission. Indeed, it seems that many, if not the majority, of the emerging diseases of humans represent dead-end infections. This may represent the natural background dynamics of cross-species transmission.



For example, almost all avian-to-human transmissions of influenza A virus result in dead-end infections, yet occasional avian-to-human transmission can cause pandemics. Each time a cross-species transmission occurs, there is a small chance that it will take hold. The problem is identifying and quantifying the key factors that determine whether a particular initial infection will survive and grow into a full-fledged epidemic. To understand emergence, it is therefore crucial to understand why only some viruses are able to regularly establish long-term transmission networks (see the chapter by Childs et al., this volume).

Perhaps the central question in this respect is whether, following cross-species transmission, emergent viruses must adapt to replicate in their new species, or whether the process of emergence is essentially blind to natural selection? Arguments can be advanced on both sides and there is currently little good data to choose among them. For example, one model of viral emergence posits that adaptation to a new host species during the early period of an epidemic is of fundamental importance, because this raises the basic reproductive rate of the virus,  $R_0$ , to greater than 1, so that sustained transmission networks can be established (Anita et al. 2003). This adaptive process is thought to occur during the “stuttering chains of transmission” that might characterize the early stages of an epidemic (Anita et al. 2003). Hence, those viruses that have not evolved human-to-human transmission are simply those that have not yet fully adapted to our species as  $R_0 < 1$  (and human-to-human transmission would surely be favored by natural selection because it increases the number of secondary infections). Empirical evidence for this theory comes from one of the best-documented cases of emergence, that of the carnivore parvoviruses (ssDNA viruses). In this case, the feline parvoviruses that infected cats jumped to dogs in the early 1970s, therein giving rise to the canine parvoviruses, an event that was accompanied by strong positive selection and an extremely high rate of nucleotide substitution (Shackelton et al. 2005). Direct adaptation to a new host species also seems to have been central to the emergence of the Venezuelan equine encephalitis virus (Brault et al. 2002). Further, although there is no strong evidence to date that the cross-species transmission event from dengue virus in monkeys to dengue virus in humans involved adaptive evolution in the latter (Twiddy et al. 2002), experimental studies imply that adaptation to the principal vector of dengue virus in an urban human setting, the *Aedes aegypti* mosquito, is a crucial prerequisite for sustained human transmission (Moncayo et al. 2004).

An alternative model for viral emergence is that rather than the emergent virus adapting to the new host species following exposure, successful emergence will only occur if a virus that already possesses the necessary mutations (such as those for receptor-binding) is exposed to the recipient host. In other words, successful emergent strains are those that are in some sense *preadapted* to establish productive infections in the new host species (see the chapter by Childs et al.,

this volume), so that the probability of emergence then becomes a function of the frequency of exposure. Indeed, that the majority of emerging infections (in humans at least) result in dead-end infections implies that even short-term transmission chains are difficult to establish for most viruses because they lack the necessary mutations. Moreover, for the majority of emergent viruses it has been difficult to show that cross-species transmission is associated with adaptation in the recipient host. To take two high-profile examples, although some sequence analyses suggest that SARS-CoV was subject to adaptive evolution during its early spread through humans (Yeh et al. 2004), it is unclear whether this was adaptation to the new host or selection for immune escape. Similarly, while the transition from SIVcpz in chimpanzees to HIV in humans seems to have been associated with a change in selection pressure (Sharp et al. 2001), it is unclear whether this reflects adaptive evolution or a relaxation of selective constraints. Finally, viral exposure to hosts of the right genetic configuration may also be of critical importance in the establishment of new infections. For example, it might be that a particular host HLA type is a more willing recipient of an emergent virus than another. In these circumstances, it is the particular combination of viral sequence and host immune system that is necessary to start a successful infection.

The case of influenza A virus again provides a highly illustrative example (see the chapter by Webby et al., this volume). Central (although not sufficient) to whether this virus is able to productively infect hosts are the sialic acid cell receptors found on cell-surface oligosaccharides. All avian influenza viruses replicate in the gastrointestinal tract and bind to sialic acid in a  $\alpha$  2,3-linkage to galactose. In contrast, human influenza viruses replicate in the respiratory tract, producing the distinctive disease symptoms, and bind to sialic acid in a  $\alpha$  2,6-linkage. Hence, the shift from  $\alpha$  2,3- to  $\alpha$  2,6-linkage is critical in enabling the switch from birds to humans and often involves changes at two amino acid residues, although mutations in other genes also play key roles (Matrosovich et al. 1997; Taubenberger et al. 2005). The key question, therefore, is whether these mutations appear *de novo* in humans, in the short transmission networks of people who initially suffer avian influenza, or whether they preexist in the avian population, and if the appropriate strain is transmitted, then will emergence follow? Again, there is little data to determine the relative importance of these two aspects of cross-species transmission dynamics, although the population size of influenza virus in avian species must be orders of magnitude greater than that in the handful of human cases during most outbreaks.

We suggest that three important advances are required to fully elucidate the role of adaptive evolution in viral emergence. First, improvements are needed in the analytical methods available to measure selection pressures acting on genes. The methods most commonly used at present involve comparisons

of the numbers of synonymous ( $d_s$ ) and nonsynonymous ( $d_N$ ) substitutions per site. Although informative, these methods are highly conservative and are greatly limited when detecting positive selection at sporadic amino acid sites along a single lineage, which may be the form of adaptive evolution most often associated with viral emergence. Second, despite ever-expanding sequence databases, there are surprisingly few examples where viral sequence data is available from both donor and recipient species. As a case in point, although dengue is one of the most important emerging viruses of humans and a multitude of sequence data from humans are readily available, only a handful of samples have been collected from the most likely donor species, Old World monkeys (Wang 2000).

Third, models of evolution need to be developed that take into account not only the varying selective environment in which most viral pathogens exist but also that accurately reflect the often complex life cycle of emerging viruses. These models will probably not be tractable by analytical techniques and this will mean that computationally intensive simulation studies will become an increasingly important component of research into the evolutionary genetics of emerging disease.

## 5 Is Recombination a Prerequisite for Viral Emergence?

Although the engine of RNA virus evolution is undoubtedly their high mutation rate, there is mounting evidence that the genetic variability observed in RNA virus populations can be shaped, in part, by recombination. Further, because recombination is a process that potentially increases fitness by creating advantageous genotypes and removing deleterious mutations, it might also be supposed that it can assist the process of emergence. This is perhaps best shown in the case of the primate lentiviruses, such as HIV, which not only experience extremely high rates of recombination, with multiple template-switching events occurring during each replication cycle (Jung et al. 2002), but where recombinant viruses seem to be associated with many cases of cross-species transmission (Bailes et al. 2003). Similarly, the cross-species transmission of influenza A virus from birds to human is often associated with reassortment among hemagglutinin (HA) and neuraminidase (NA) subtypes (Webby and Webster 2001; see the chapter by Webby et al., this volume).

A more recent case in point concerns the emergence of SARS-CoV. In this case, it has been argued that human SARS-CoV is a recombinant of avian and other mammalian coronaviruses (Stavrinides and Guttman 2004), and that recombination may even have allowed the virus to acquire the critical

suite of amino acid changes required to cause infection in humans (Stanhope et al. 2004). However, a closer inspection of the pertinent sequence data casts serious doubt on this hypothesis (see the chapter by Wang and Eaton, this volume). First, the evidence for recombination in SARS coronavirus is weak at best, and it seems equally likely that the phylogenetic signal said to support recombination results from variation in substitution rate among lineages, such that different genes produce slightly different trees (Gibbs et al. 2004; Holmes and Rambaut 2004). Second, the proposed recombination in SARS-CoV would have involved such distantly related virus strains that it cannot be responsible for the very recent emergence of the virus in humans (Holmes and Rambaut 2004).

Another reason to doubt the role played by recombination in emergence in general is that, other than in the retroviruses, recombination is not a particularly common process in RNA viruses and there is no reason to suppose that it is any more than a mechanistic by-product. For example, recombination appears to be extremely rare in negative-sense RNA viruses (Chare et al. 2003), most likely because their RNA is always encapsidated, thereby greatly limiting the template-switching thought to be central to RNA recombination. As a number of emerging viruses have negative-sense RNA genomes, this automatically argues against recombination as a general process in viral emergence. Similarly, although recombination is more common in positive-sense RNA viruses, in most cases it appears to be a sporadic event that does not occur at a high enough frequency to make it a key evolutionary strategy, although, of course, rare events like recombination may sometimes be critical in kick-starting the process of viral emergence.

Most of the available evidence suggests that recombination rates in RNA viruses are controlled by two factors; the ability of the virus in question to undergo template switching and the frequency with which multiple infections occur. It is these factors that explain why HIV has such a high rate of recombination; the virus possesses two copies of the RNA genome, which means that template switching will occur readily, and the ease at which the virus has spread worldwide means that multiple infections are abundant. However, HIV appears to be the exception rather than the rule. For example, HTLV is a retrovirus like HIV and is also reported to have jumped species boundaries (between humans and other primates) during its evolutionary history. However, there is no evidence for recombination in HTLV. In fact, even in the case of HIV it is not certain that recombination has been critical to emergence, despite its frequency. Overall, as the rate of recombination, per base, will be very much lower than that of mutation for most RNA viruses, it seems logical to conclude that recombination is not a key requirement for emergence, but rather a happy coincidence.

## 6

### Conclusions: Evolution and Emergence in RNA Viruses

Much of this chapter has explored the issue of the role played by viral evolution in the process of emergence. The result, contrary to many perceptions of the inherent adaptability of RNA viruses, is that successful emergence, characterized by sustained human-to-human transmission may be a far more difficult process than might be expected given the remarkable rates at which RNA viruses mutate. Why might this be so? One probable answer lies in the theory that the genomes of RNA viruses are so small (usually less than 15 kb in length), and so multifunctional, that most mutations are likely to affect some critical aspect of virus biology (Holmes 2003a). As such, although mutations are abundant in RNA viruses, the vast majority are deleterious, or slightly deleterious, and will reduce viral fitness in the long term. Much of the genetic variation sampled within RNA virus populations, including those of emerging viruses in new host species, will therefore consist of short-lived deleterious mutations, which can be thought of as a form of mutational straightjacket. Unlike many other evolutionary systems that are dominated by a drift–selection balance, viral evolutionary genetics may be dominated by a mutation–selection balance (Domingo and Holland 1997). The requirements of a compact and highly pleiotropic genome, coupled with a high error rate, lead to a high mutational load (Elena and Moya 1999) that leaves little room for the limitless adaptability some have attributed to RNA viruses. Indeed, mutation rates are so high that it is possible that even highly beneficial mutations will not be able to spread through a viral population because they are, by chance, linked to deleterious mutations that arise in the same genome. This will evidently hinder their ability to emerge in new hosts and does much to explain why some RNA viruses are better able to cross species barriers than others. As such, we propose that the true importance of the nearly neutral (or slightly deleterious) theory of molecular evolution (Ohta 1992, 1998) to the study of RNA viruses has not been fully appreciated, yet may be crucial to fully understanding emergence.

**Acknowledgements** We thank The Wellcome Trust for financial support.

### References

- Antia R, Regoes RR, Koella JC, Bergstrom CT (2003) The role of evolution in the emergence of infectious diseases. *Nature* 426:658–610
- Bailes E, Gao F, Bibollet-Ruche F, Courgnaud V, Peeters M, Marx PA, Hahn BH, Sharp PM (2003) Hybrid origin of SIV in chimpanzees. *Science* 300:1713

- Baranowski E, Ruiz-Jarabo CM, Domingo E (2001) Evolution of cell recognition by viruses. *Science* 292:1102–1105
- Brault AC, Powers AM, Holmes EC, Woelk CH, Weaver SC (2002) Positively charged amino acid substitutions in the E2 envelope glycoprotein are associated with the emergence of Venezuelan equine encephalitis virus. *J Virol* 76:1718–1730
- Breman JG, Johnson KM, van der Groen G, Robbins CB, Szczeniowski MV, Ruti K, Webb PA, Meier F, Heymann DL (1999) A search for Ebola virus in animals in the Democratic Republic of the Congo and Cameroon: ecologic, virologic, and serologic surveys, 1979–1980. *J Infect Dis* 179:S139–S147
- Chare ER, Holmes EC (2004) Selection pressures in the capsid genes of plant RNA viruses reflect mode of transmission. *J Gen Virol* 85:3149–3157
- Chare ER, Gould EA, Holmes EC (2003) Phylogenetic analysis reveals a low rate of homologous recombination in negative-sense RNA viruses. *J Gen Virol* 84:2691–2703
- Cleaveland S, Laurenson MK, Taylor LH (2001) Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence. *Phil Trans R Soc Lond B* 356:991–999
- DeFilippis VR, Villarreal LP (2000) An introduction to the evolutionary ecology of viruses. In: Hurst CJ (ed) *Viral ecology*. Academic Press, New York, pp 126–208
- Domingo E, Holland JJ (1997) RNA virus mutations for fitness and survival. *Ann Rev Microbiol* 51:151–178
- Elena SF, Moya A (1999) Rate of deleterious mutation and the distribution of its effects on fitness in vesicular stomatitis virus. *J Evol Biol* 12:1078–1088
- Ferber D (2000) Human diseases threaten great apes. *Science* 289:1277–1278
- Figueroa F, Günther E, Klein J (1988) MHC polymorphism pre-dating speciation. *Nature* 355:265–267
- Gaunt MW, Sall AA, de Lamballerie X, Falconar AKI, Dzhivaniyan TI, Gould EA (2001) Phylogenetic relationships of flaviviruses correlate with their epidemiology, disease association and biogeography. *J Gen Virol* 82:1867–1876
- Gibbs AJ, Gibbs MJ, Armstrong JS (2004) The phylogeny of SARS coronavirus. *Arch Virol* 149:621–624
- Gillespie JH (1998) *Population genetics: a concise course*. Johns Hopkins University Press, Baltimore
- Holmes EC (2003a) Error thresholds and the constraints to RNA virus evolution. *Trends Microbiol* 11:543–546
- Holmes EC (2003b) Patterns of intra- and inter-host nonsynonymous variation reveal strong purifying selection in dengue virus. *J Virol* 77:11296–11298
- Holmes EC (2004) The phylogeography of human viruses. *Mol Ecol* 13:745–756
- Holmes EC, Rambaut A (2004) Viral evolution and the emergence of SARS coronavirus. *Phil Trans R Soc Lond B* 359:1059–1065
- Huet T, Cheynier R, Meyerhans A, Roelants G, Wain-Hobson S (1990) Genetic organisation of a chimpanzee lentivirus related to HIV-1. *Nature* 345:356–359
- Jung A, Maier R, Vartanian JP, Bocharov G, Jung V, Fischer U, Meese E, Wain-Hobson S, Meyerhans A (2002) Recombination: multiply infected spleen cells in HIV patients. *Nature* 418:144

- Lau SK, Woo PC, Li KS, Huang Y, Tsoi HW, Wong BH, Wong SS, Leung SY, Chan KH, Yuen KY (2005) Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc Natl Acad Sci U S A* 102:14040–14045
- Li W, Shi Z, Yu M, Ren W, Smith C, Epstein JH, Wang H, Crameri G, Hu Z, Zhang H, Zhang J, McEachern J, Field H, Daszak P, Eaton BT, Zhang S, Wang LF (2005) Bats are natural reservoirs of SARS-like coronaviruses. *Science* 310:676–679
- Malpica MJ, Fraile A, Moreno I, Obies CI, Drake JW, García-Arenal F (2002) The rate and character of spontaneous mutation in an RNA virus. *Genetics* 162:1505–1511
- Markowitz M, Louie M, Hurley A, Sun E, Di Mascio M, Perelson AS, Ho DD (2003) A novel antiviral intervention results in more accurate assessment of human immunodeficiency virus type 1 replication dynamics and T-cell decay in vivo. *J Virol* 77:5037–5038
- Matrosovich MN, Gambaryan AS, Teneberg S, Piskarev VE, Yamnikova SS, Lvov DK, Robertson JS, Karlsson KA (1997) Avian influenza A viruses differ from human viruses by recognition of sialyloligosaccharides and gangliosides and by a higher conservation of the HA receptor-binding site. *Virology* 233:224–234
- Moncayo AC, Fernandez Z, Ortiz D, Diallo M, Sall A, Hartman S, Davis CT, Coffey L, Mathiot CC, Tesh RB, Weaver SC (2004) Dengue emergence and adaptation to peridomestic mosquitoes. *Emerg Infect Dis* 10:1790–1796
- Morse SS (1995) Factors in the emergence of infectious diseases. *Emerg Infect Dis* 1:7–15
- Moya A, Holmes EC, González-Candelas F (2004) The population genetics and evolutionary epidemiology of RNA viruses. *Nat Rev Microbiol* 2:279–287
- Nichol ST, Spiropoulou CF, Morzunov S, Rollin PE, Ksiazek TG, Feldmann H, Sanchez A, Childs J, Zaki S, Peters CJ (1993) Genetic identification of a hantavirus associated with an outbreak of acute respiratory illness. *Science* 262:914–917
- Ohta T (1992) The nearly neutral theory of molecular evolution. *Ann Rev Ecol Syst* 23:263–286
- Ohta T (1998) Evolution by nearly-neutral mutations. *Genetica* 102/103:83–90
- Okamoto H, Fukuda M, Tawara A, Nishizawa T, Itoh Y, Hayasaka I, Tsuda F, Tanaka T, Miyakawa Y, Mayumi M (2000) Species-specific TT viruses and cross-species infection in nonhuman primates. *J Virol* 74:1132–1139
- Peterson AT, Carroll DS, Mills JN, Johnson KM (2004) Potential mammalian filovirus reservoirs. *Emerg Infect Dis* 10:2073–2081
- Pugachev KV, Guirakhoo F, Ocran SW, Mitchell F, Parsons M, Penal C, Girakhoo S, Pougatcheva SO, Arroyo J, Trent DW, Monath TP (2004) High fidelity of yellow fever virus RNA polymerase. *J Virol* 78:1032–1038
- Sanz AI, Fraile A, Gallego JM, Malpica JM, García-Arenal F (1999) Genetic variability of natural populations of cotton leaf curl geminivirus, a single-stranded DNA virus. *J Mol Evol* 49:672–681
- Scholtissek C, Ludwig S, Fitch WM (1993) Analysis of influenza A virus nucleoproteins for the assessment of molecular genetic mechanisms leading to new phylogenetic virus lineages. *Arch Virol* 131:237–250
- Shackelton LA, Parrish CR, Truyen U, Holmes EC (2005) High rate of viral evolution associated with the emergence of canine parvoviruses. *Proc Natl Acad Sci U S A* 102:379–384

- Sharp PM, Bailes E, Chaudhuri RR, Rodenburg CM, Santiago MO, Hahn BH (2001) The origins of acquired immune deficiency syndrome viruses: where and when? *Phil Trans R Lond B* 356:867–876
- Stanhope MJ, Brown JR, Amrine-Madsen H (2004) Evidence from the evolutionary analysis of nucleotide sequences for a recombinant history of SARS-CoV. *Infect. Genet Evol* 4:15–19
- Stavrinides J, Guttman DS (2004) Mosaic evolution of the severe acute respiratory syndrome coronavirus. *J Virol* 78:76–82
- Switzer WM, Salemi M, Shanmugam V, Gao F, Cong M-E, Kuiken C, Bhullar V, Beer B, Vallet D, Gautier-Hion A, Tooze A, Villinger F, Holmes EC, Heneine W (2005) Ancient co-speciation of simian foamy viruses and primates. *Nature* 434:376–380
- Taubenberger JK, Reid AH, Lourens RM, Wang R, Jin G, Fanning TG (2005) Characterization of the 1918 influenza virus polymerase genes. *Nature* 437:889–893
- Twiddy SS, Farrar JF, Chau NV, Wills B, Gould EA, Gritsun T, Lloyd G, Holmes EC (2002) Phylogenetic relationships and differential selection pressures among genotypes of dengue-2 virus. *Virology* 298:63–72
- Wang E, Ni H, Xu R, Barrett ADT, Watowich SJ, Gubler DJ, Weaver SC (2000) Evolutionary relationships of endemic/epidemic and sylvatic dengue viruses. *J Virol* 74:3227–3234
- Whalley SA, Murray JM, Brown D, Webster GJ, Emery VC, Dusheiko GM, Perelson AS (2001) Kinetics of acute hepatitis B virus infection in humans. *J Exp Med* 193:847–854
- Webby RJ, Webster RG (2001) Emergence of influenza A viruses. *Phil Trans R Soc Lond B* 356:1817–1828
- Woelk CH, Holmes EC (2002) Reduced positive selection in vector-borne RNA viruses. *Mol Biol Evol* 19:2333–2336
- Woolhouse MEJ, Taylor LH, Haydon DT (2001) Population biology of multihost pathogens. *Science* 292:1109–1112
- Woolhouse MEJ (2002) Population biology of emerging and re-emerging pathogens. *Trends Microbiol* 10:S3–S7
- Yeh S-H, Wang H-Y, Tsai C-Y, Kao C-L, Yang J-Y, Liu H-W, Su I-J, Tsai S-F, Chen D-S, Chen P-J, and the National Taiwan University SARS Research Team (2004) Characterization of severe acute respiratory syndrome coronavirus genomes in Taiwan: Molecular epidemiology and genome evolution. *Proc Natl Acad Sci U S A* 101:2542–2547
- Zárate S, Novella IS (2004) Vesicular stomatitis virus evolution during alternation between persistent infection in insect cells and acute infection in mammalian cells. Is dominated by the persistence phase. *J Virol* 78:12236–12242