Error thresholds and the constraints to RNA virus evolution

Edward C. Holmes

Department of Zoology, University of Oxford, South Parks Road, Oxford, UK OX1 3PS

RNA viruses are often thought of as possessing almost limitless adaptability as a result of their extreme mutation rates. However, high mutation rates also put a cap on the size of the viral genome by establishing an error threshold, beyond which lethal numbers of deleterious mutations accumulate. Herein, I argue that a lack of genomic space means that RNA viruses will be subject to important evolutionary constraints because specific sequences are required to encode multiple and often conflicting functions. Empirical evidence for these constraints, and how they limit viral adaptability, is now beginning to accumulate. Documenting the constraints to RNA virus evolution has important implications for predicting the emergence of new viruses and for improving therapeutic procedures.

Studies of RNA virus evolution are dominated by the idea that their high mutation rates, short replication times and often immense population sizes mean that these organisms harbour an enormous amount of genetic diversity, and as such can quickly adapt to changing environments. Indeed, it is the inherent adaptability of RNA viruses that hinders our attempts to develop broadly effective antiviral drugs and vaccines, and also means that they represent the most important cause of emerging diseases. However, although it is undoubtedly true that populations of RNA viruses are often remarkably diverse, it does not necessarily follow that they have little trouble responding to environmental challenges. There is mounting evidence for important constraints in RNA virus evolution, and that these constraints are a predictable outcome when mutation rates reach an error threshold.

Mutation and the error threshold

Mutation rates in RNA viruses are several orders of magnitude higher than those in DNA based life-forms. Such an inference can be made from the high levels of viral genetic variation seen within and among infected hosts [1], studies of long-term rates of nucleotide substitution [2] and most importantly from direct measurements of error frequency [3]. Together, these suggest an average mutation rate in the region of approximately one mutation per genome, per replication; although, slightly lower rates have been reported recently [4] and little is known about the variation in mutation rates among RNA viruses. A high mutation rate is a natural expectation given that the

RNA polymerase, or reverse transcriptase, used for replication lacks proofreading ability.

In addition to producing a great deal of genetic variation, the high mutation rate of RNA viruses is pivotal from an evolutionary perspective because it is close to the notional 'error threshold', which is set at approximately the reciprocal of the genome size [5,6]. Beyond this threshold, so many deleterious mutations occur during each replication cycle that 'fit' viral genomes are unable to reproduce themselves faithfully and the population eventually becomes extinct. RNA viruses are therefore permanently close to 'error catastrophe'. Although this might appear an inherently unstable evolutionary strategy, provided that viral population sizes are sufficiently large, life at the error threshold does allow RNA viruses to produce a myriad of potentially useful mutations within a very short time.

There is now sufficient empirical evidence to suggest that the error threshold is more than just a theoretician's whimsy. First, populations of RNA viruses often harbour numerous defective genomes, which is to be expected if deleterious mutations arise at a high frequency. A recent example is provided by dengue virus [7]. More dramatic evidence comes from experimental studies that show that increasing mutation rates beyond the error threshold by artificial means with the application of mutagens can result in the extinction of viral populations, a procedure known as 'lethal mutagenesis' [8–11]. Although it is possible that this could lead to a new generation of antiviral therapies, one important uncertainty concerns the effect of recombination. To date, most work on lethal mutagenesis has involved essentially clonal RNA viruses, where recombination plays little or no role in shaping genetic diversity. However, for viruses like human immunodeficiency virus (HIV), the rate of recombination per base can exceed that of mutation [12]. It might be expected that recombination will allow RNA viruses to avoid error catastrophe by regenerating fit genomes from defective ones, much in the same way that recombination can prevent the accumulation of deleterious mutations in small populations, a process known as Muller's Ratchet. However, simulation studies have revealed that recombination in RNA viruses is more likely to create combinations of deleterious mutations than purge them from genomes, and in doing so reduce the error threshold, so that dramatic fitness reductions occur at lower mutation rates [13]. If this also applies in nature, not only will recombination assist lethal mutagenesis, but it is possible

that recombination rates have been selectively reduced during RNA virus evolution. RNA viruses would then truly represent evolutionarily unique organisms. This is clearly an area that needs to be explored further.

Further evidence for the error threshold comes from comparative studies. In particular, a phylogenetic analysis of 50 RNA viruses showed a negative correlation between rates of nucleotide substitution (a surrogate of mutation rate) and genome size; the longer the RNA virus genome in question, the lower its substitution rate [2]. This is exactly the relationship predicted by the error threshold theory [6], although these studies will need to be reproduced with more direct estimates of mutation rate. Because substitution rates are largely a function of genome size, it can be argued that RNA viruses do not individually select their mutation rates to produce adaptively useful changes, as these would vary depending on the life-history of the virus in question, but rather that these rates are simply an inherent by-product of replication by error-prone RNA polymerases. Moreover, the existence of an error threshold does not prove that RNA viruses form the complex interconnected population structure referred to as the quasispecies [14]. Although error thresholds are certainly compatible with quasispecies theory, they can easily be produced by 'classical' population genetic models; every life-form probably faces a mutational meltdown if the rate of deleterious mutation is sufficiently high, although the exact limit is unknown because of the uncertain influence of sexual reproduction [15,16].

Another aspect of the error threshold that is of fundamental importance for understanding RNA virus evolution is that it sets an upper limit on genome size. As was recognized long ago by Manfred Eigen [5], it is impossible to replicate excessively long pieces of RNA because too many deleterious mutations accumulate. Notably, the vast majority of RNA viruses do not have genomes that are longer than ~15 kb, indicating that this is the genome size normally set by the error threshold, which is in stark contrast to the much larger genomes observed in many DNA viruses. The most notable exceptions are the coronaviruses with RNA genomes of ~30 kb, which should evolve measurably slower than other RNA viruses. Crucially, by limiting genome size, RNA viruses are automatically constrained; all viral functions must be encoded within a confined genomic space. This means that specific regions of genome sequence will often have multiple and sometimes very diverse functions, and that mutations that adapt a virus to one environment might have a negative affect in another. Therefore, the limitation to genome size, and the consequent lack of evolutionary elbow room, make it probable that RNA viruses will experience extensive pleiotropy, epistasis and negative fitness trade-offs, which together prevent them from adapting to all environmental conditions with equal success.

Constraints on RNA virus evolution

Paradoxically, although RNA viruses possess enormous mutational power that allows them to rapidly generate adaptively useful genetic variation, this high mutation rate also limits their adaptability by constraining genome size. Although the idea that RNA viruses are subject to strong constraints at first appears to be contradicted by the speed at which they can develop antigenic variation or resistance to antiviral agents, it is reflected in other key aspects of their evolution. These are discussed below. Although equivalent constraints can be observed in other organisms, RNA viruses are perhaps unique in the extent to which these constraints apply.

RNA virus genomes often use overlapping reading frames

One manifestation of the evolutionary constraints imposed by limited genome size is that RNA viruses frequently use overlapping reading frames to maximize the genetic information available to them. More than 120 of the RNA viruses for which complete genome sequence data are available on GenBank use overlapping reading frames, and in many cases the region of overlap extends to more than 500 bp. Mutations that occur within these regions are expected to be subject to complex fitness trade-offs, which might affect estimates of rates of nucleotide substitution and phylogenetic relationships.

Individual sequence regions have multiple functions

There are now a variety of examples where specific regions of viral genomes, or even individual mutations, have multiple and sometimes conflicting functions. In particular, the occurrence of mutations that affect both antigenicity and cell tropism has been documented in several viruses [17], with those in the V3 loop-region of the HIV-1 envelope gene particularly well studied [18]. Such functional overlap could have important evolutionary consequences. For example, the recognition of certain cell types could be 'out of reach' for viruses that are forced to fix conflicting immune-escape mutations, even though they might only be a few mutational steps away. Played out over evolutionary time, such a process could partially determine the host range of viruses.

Individual mutations can be subject to strong constraints

In some cases it has been possible to identify the evolutionary constraints acting on individual mutations. For example, there are instances where selectively advantageous mutations take far longer to appear than expected. One such case involves escape from cytotoxic T-lymphocyte (CTL) recognition in HIV-infected patients that express HLA-B27 + epitopes. Here, a single escape substitution at residue 264 in p24 *gag* might not occur until many years post-infection [19]. Given the rapid mutation and replication rates in HIV, the delayed appearance of a single escape mutation strongly suggests that its occurrence is inhibited by functional constraints. Indeed, residue 264 is highly conserved in HLA-B27 − patients, indicating that mutations at this site are normally deleterious, requiring compensatory amino acid changes to counteract any deleterious effects on capsid structure. Although drug-resistance mutations in HIV often appear more readily, the low fitness of many such mutants compared with wild-type in the absence of drug therapy [20] is testament to the high rate of deleterious mutation in RNA viruses.

Convergent and parallel evolution are frequent

High levels of convergent or parallel evolution have been reported in a variety of RNA viruses [21–24]. For example, *in vitro* analysis of vesicular stomatitis virus (VSV) revealed that 12 out of 25 variable sites contained convergent evolutionary changes, often with the same mutation appearing multiple times [24]. Convergent evolution is also frequently reported in the context of HIV drug-resistance [25]. Although there is currently no data to assess whether convergent evolution is more common in RNA viruses than in other organisms, it must be considered a natural expectation if RNA viruses are subject to strong evolutionary constraints. Specifically, if most mutations available to RNA viruses are deleterious, then there might only be a few pathways of evolutionary change that can easily be explored.

Fitness trade-offs are commonplace

I have already argued that evolutionary constraints mean that complex fitness trade-offs will be commonplace in RNA virus evolution. One important, although controversial, example concerns vector-borne RNA viruses (arboviruses). Because of the requirement to replicate in hosts that are as phylogenetically distinct as mammals and invertebrates, it might be expected that mutations that adapt the virus to the vector are sometimes deleterious for replication in a mammal (and vice versa). Support for this hypothesis is that arboviruses are subject to far less positive selective pressure in their surface-exposed structural genes than other RNA viruses [26], and also that they exhibit lower rates of non-synonymous substitution [2]. However, *in vitro* studies of arbovirus evolution have provided conflicting evidence for the existence of negative fitness trade-offs [27–29]. Whether these differing results reflect the complex nature of epistasis when adapting viruses to particular cell types, or indicate that experimental studies of viral diversification are not always an accurate reflection of natural evolution, remains to be seen.

The constrained evolution of arboviruses hints at a more general evolutionary rule; the greater the number and diversity (in terms of receptors) of cell types that a virus replicates in, the more it will be forced into fitness trade-offs and adaptive compromises, and the more constrained its evolution will be. An example might be provided by rabies virus, which displays lower rates of non-synonymous substitution than seen in many other RNA viruses [30]. Although this low rate could signify a lack of immune-driven positive selection, it might also reflect the very broad cell-tropism of rabies virus, in that many of the amino acid mutations that appear are deleterious in at least one cell type.

RNA secondary structure might limit genetic variation

The observation that some RNA viruses have complex RNA secondary structures and that these affect patterns of genetic variation suggests that RNA secondary structure might also be a factor in constraining sequence diversity [31]. In particular, constraints imposed by RNA secondary structure might mean that not all synonymous mutations are neutral and therefore able to accumulate in the monotonous fashion usually imagined. Although the importance of RNA secondary structure has been demonstrated in some RNA viruses, its overall role is uncertain and clearly needs to be explored further.

Major viral phenotypes rarely change

One striking observation from studies of long-term RNA virus evolution is that major phenotypic characteristics, such as the transmission mechanism, appear to change relatively infrequently. Therefore, genera or families of RNA viruses are usually comprised of individual members with a limited range of transmission mechanisms, indicating that these traits are rather difficult to evolve (although possessing multiple transmission mechanisms would clearly result in greater reproduction). For example, the genus *Flavivirus* only contains viruses that are vector-borne or directly transmitted, usually by unknown means, and within the vector-borne group there is a strong phylogenetic ordering that reflects whether the viruses are tick-borne or transmitted by *Aedes* or *Culex* mosquitoes [32]. Overall, the limited evolvability of these traits strongly suggests that they are constrained by important fitness trade-offs.

Implications for RNA virus evolution

The constraints faced by RNA viruses have important implications for their evolution and their eventual control. For example, one way in which DNA viruses adapt to new species is by host-gene capture [33]. However, this route will generally not be open to RNA viruses and has only been recorded occasionally [34]. Consequently, an increase in genome size would possibly result in the virus crossing the error threshold. For the same reasons, RNA viruses are rarely expected to undergo the process of gene duplication, which is so common in other life-forms. Not only will this limit adaptability, but it will have a major bearing on attempts to reconstruct all-encompassing RNA virus phylogenies using genome structure and organization [35]. The phylogenetic analysis of RNA viruses is also probably affected by constraints against individual base changes. As noted previously, if there are only a limited number of pathways that virus evolution can follow, without falling into the trap set by deleterious mutations, then it is possible that the same mutations will appear on unrelated branches. This will affect the accuracy of phylogenetic trees, as well as estimates of rates of nucleotide substitution and viral divergence times [36].

There are, however, more positive implications of the evolutionary constraints faced by RNA viruses. Most notably, it will be difficult for RNA viruses to acquire all of the characteristics that maximize their disease impact on human populations. For example, a lack of evolutionary space helps to explain the evolution of virulence in RNA viruses, especially the nature of the relationship between virulence and transmissibility [37,38]. As both traits might be encoded by the same gene, they might encounter negative fitness trade-offs, so that neither can increase without having an adverse affect on each other. Similarly, evolutionary constraints mean that the use of combinations of drugs will probably be more effective than single

drugs because of fitness trade-offs between resistance mutations. Finally, understanding the constraints to RNA virus evolution could be of fundamental importance for studies of viral emergence. Although RNA viruses are perhaps able to jump species boundaries more readily than other pathogens [39], it does not follow that all RNA viruses are equally well equipped in this respect. In particular, if evolutionary constraints mean that some RNA viruses are inhibited in their ability to jump species boundaries, for instance because the necessary change in cell-receptor recognition has a deleterious effect on some other component of fitness, then the study of constraints might help us predict what viruses could emerge in the future.

Although we can continue to marvel at the adaptability of RNA viruses, it is also the case that their evolution might be much more constrained than we usually imagine. High mutation rates do provide RNA viruses with the raw materials for adaptation, but also produce a small and complex genome, so that viral evolution will be characterized by intricate fitness trade-offs and epistatic interactions. RNA viruses might therefore only explore a small fraction of the total sequence space available to them. Although the constraints to RNA virus evolution have been hinted at previously [40,41], it is important that we now explore their mechanistic basis in far more detail. Therefore, rather than thinking about what RNA viruses can do in their evolution, we should concentrate on their limitations. RNA viruses might be more at the mercy of their mutation rates than we think.

References

1 Domingo, E. and Holland, J.J. (1997) RNA virus mutations for fitness and survival. *Annu. Rev. Microbiol.* 51, 151–178
2 Jenkins, G.M. *et al.* (2002) Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *J. Mol. Evol.* 54, 156–165
3 Drake, J.W. and Holland, J.J. (1999) Mutation rates among RNA viruses. *Proc. Natl. Acad. Sci. U. S. A.* 96, 13910–13913
4 Malpica, M.J. *et al.* (2002) The rate and character of spontaneous mutation in an RNA virus. *Genetics* 162, 1505–1511
5 Eigen, M. (1971) Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58, 465–523
6 Eigen, M. (1996) Steps Towards Life, Oxford University Press
7 Wang, W-K. *et al.* (2002) Dengue type 3 virus in plasma is a population of closely related genomes: quasispecies. *J. Virol.* 76, 4662–4665
8 Loeb, L.A. *et al.* (1999) Lethal mutagenesis of HIV with mutagenic nucleoside analogs. *Proc. Natl. Acad. Sci. U. S. A.* 96, 1492–1497
9 Sierra, S. *et al.* (2000) Response of foot-and-mouth disease virus to increased mutagenesis. Influence of viral load and fitness in loss of infectivity. *J. Virol.* 74, 8316–8323
10 Crotty, S. *et al.* (2001) RNA virus error catastrophe: direct test by using ribavirin. *Proc. Natl. Acad. Sci. U. S. A.* 98, 6895–6900
11 Ruiz-Jarabo, C.M. *et al.* (2003) Lethal mutagenesis of the prototypic arenavirus lymphocytic choriomeningitis (LCMV). *Virology* 308, 37–47
12 Jung, A. *et al.* (2002) Recombination – Multiply infected spleen cells in HIV patients. *Nature* 418, 144
13 Boerlijst, M.C. *et al.* (1996) Viral quasi-species and recombination. *Proc. R. Soc. Lond. B.* 263, 1577–1584
14 Domingo, E. (2002) Quasispecies theory in virology. *J. Virol.* 76, 463–465
15 Kondrashov, A.S. (1995) Contamination of the genome by very slightly deleterious mutations – why we have not died 100 times over. *J. Theor. Biol.* 175, 583–594
16 Kondrashov, A.S. (1988) Deleterious mutations and the evolution of sexual reproduction. *Nature* 336, 435–440
17 Baranowski, E. *et al.* (2001) Evolution of cell recognition by viruses. *Science* 292, 1102–1105
18 Chesebro, B. *et al.* (1992) Macrophage-tropic human immunodeficiency virus isolates from different patients exhibit unusual V3 envelope sequence homogeneity in comparison with T-cell tropic isolates: definition of critical amino acids involved in cell tropism. *J. Virol.* 66, 6547–6554
19 Kelleher, A.D. *et al.* (2001) Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27 restricted CTL responses. *J. Exp. Med.* 193, 375–385
20 Fan, N. *et al.* (1996) A drug resistance mutation in the inhibitor binding pocket of human immunodeficiency virus type 1 reverse transcriptase impairs DNA synthesis and RNA degradation. *Biochemistry* 35, 9737–9745
21 Cunningham, C.W. *et al.* (1997) Parallel molecular evolution of deletions and nonsense mutations in bacteriophage T7. *Mol. Biol. Evol.* 14, 113–116
22 Bull, J.J. *et al.* (1997) Exceptional convergent evolution in a virus. *Genetics* 147, 1497–1507
23 Fares, M.A. *et al.* (2001) Evidence for positive selection in the capsid protein-coding region of the foot-and-mouth disease virus (FMDV) subjected to experimental passage regimens. *Mol. Biol. Evol.* 18, 10–21
24 Cuevas, J.M. *et al.* (2002) Molecular basis of adaptive convergence in experimental populations of RNA viruses. *Genetics* 162, 533–542
25 Sato, H. *et al.* (2000) Convergent evolution of reverse transcriptase (RT) genes of human immunodeficiency virus type 1 subtypes E and B following nucleoside analogue RT inhibitor therapies. *J. Virol.* 74, 5357–5362
26 Woelk, C.H. and Holmes, E.C. (2002) Reduced positive selection in vector-borne RNA viruses. *Mol. Biol. Evol.* 19, 2333–2336
27 Novella, I.S. *et al.* (1999) Lack of evolutionary stasis during alternating replication of an arbovirus in insect and mammalian cells. *J. Mol. Biol.* 287, 459–465
28 Weaver, S.C. *et al.* (1999) Genetic and fitness changes accompanying adaptation of an arbovirus to vertebrate and invertebrate cells. *J. Virol.* 73, 4316–4326
29 Cooper, L.A. and Scott, T.W. (2001) Differential evolution of eastern equine encephalitis virus populations in response to host cell type. *Genetics* 157, 1403–1412
30 Holmes, E.C. *et al.* (2002) Genetic constraints and the adaptive evolution of rabies virus. *Virology* 292, 247–257
31 Simmonds, P. and Smith, D.B. (1999) Structural constraints on RNA virus evolution. *J. Virol.* 73, 5787–5794
32 Kuno, G. *et al.* (1998) Phylogeny of the genus *Flavivirus*. *J. Virol.* 72, 73–83
33 McGeoch, D.J. (2001) Molecular evolution of the gamma-Herpesvirinae. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356, 421–435
34 Meyers, G. *et al.* (1989) Ubiquitin in a togavirus. *Nature* 341, 491
35 Gorbalenya, A.E. (1995) Origin of RNA viral genomes; approaching the problem by comparative sequence analysis. In *Molecular Basis of Virus Evolution* (Gibbs, A. *et al.*, eds), pp. 49–66, Cambridge University Press
36 Holmes, E.C. (2003) Molecular clocks and the puzzle of RNA virus origins. *J. Virol.* 77, 3893–3897
37 Lipsitch, M. and Moxon, E.R. (1997) Virulence and transmittability of pathogens: what is the relationship. *Trends Microbiol.* 5, 31–37
38 Messenger, S.L. *et al.* (1999) Virulence evolution in a virus obeys a trade-off. *Proc. R. Soc. Lond. B. Biol. Sci.* 266, 397–404
39 Woolhouse, M.E.J. *et al.* (2001) Population biology of multihost pathogens. *Science* 292, 1109–1112
40 Domingo, E. *et al.* (1999) Viral quasispecies and fitness variations. In *Origin and Evolution of Viruses* (Domingo, E. *et al.*, eds), pp. 141–161, Academic Press
41 Sala, M. and Wain-Hobson, S. (1999) Drift and conservation in RNA virus evolution: are they adapting or merely changing? In *Origin and Evolution of Viruses* (Domingo, E. *et al.*, eds), pp. 115–140, Academic Press