Rates of evolutionary change in viruses: patterns and determinants

Siobain Duffy, Laura A. Shackelton* and Edward C. Holmes*‡*

Abstract | Understanding the factors that determine the rate at which genomes generate and fix mutations provides important insights into key evolutionary mechanisms. We review our current knowledge of the rates of mutation and substitution, as well as their determinants, in RNA viruses, DNA viruses and retroviruses. We show that the high rate of nucleotide substitution in RNA viruses is matched by some DNA viruses, suggesting that evolutionary rates in viruses are explained by diverse aspects of viral biology, such as genomic architecture and replication speed, and not simply by polymerase fidelity.

Generation time
The time between rounds of production of viral progeny, including any time required for virions to seek a susceptible host cell, followed by adsorption and infection of the susceptible cell, then viral replication and release.

**Center for Infectious Disease Dynamics, Department of Biology, Mueller Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16802, USA.*
‡Fogarty International Center, National Institutes of Health, Bethesda, Maryland 20892, USA.
Correspondence to E.C.H.
e-mail: ech15@psu.edu
doi:10.1038/nrg2323
Published online
4 March 2008

Mutation is the ultimate source of genetic variation, on which natural selection, genetic drift, gene flow and recombination act to shape the genetic structure of populations. Although there is a rich literature detailing how the patterns and processes of mutation differ among species, the biochemical and evolutionary determinants of mutation rate remain poorly understood in all but a few organisms[1,2]. Perhaps of more importance for our understanding of organismal evolution is how the rate at which mutations are generated per genome relates to the tempo of evolutionary change at the population level, as reflected in long-term rates of nucleotide substitution. When mutations are selectively neutral, or nearly so, there is a simple relationship between the rate at which they are generated in a genome (the mutation rate) and then fixed at the population level (the substitution rate)[3]. Deviations from this neutral expectation can reveal fundamental aspects of the evolutionary process, including the extent of natural selection and variation in replication dynamics. However, much of the research in this area has dealt with the history of individual genes, which can poorly reflect the history of the organism, particularly when rates of recombination are high[4,5]. Consequently, a genome-wide perspective of both mutation and substitution rates is crucial to obtain a full understanding of evolutionary dynamics.

Analyses on the scale of complete genomes are becoming increasingly viable[6] but, understandably, still favour organisms with small genomes. Similarly, evolutionary rates are most easily studied in rapidly evolving organisms, either those with high mutation rates or with short generation times, both of which facilitate the rapid generation of genetic diversity. Viruses, particularly those with RNA genomes, fulfil these criteria and therefore are ideal tools for studying fundamental aspects of molecular evolution (BOX 1).

Here we explore our current understanding of mutation and substitution rates (precise definitions of these terms are provided in BOX 2 and BOX 3, respectively), as well as their determinants, in viruses. We examine viruses with both RNA- and DNA-based genomes, encompassing infectious agents with diverse genome structures and replication dynamics. Not only do some viruses represent 'measurably evolving populations'[7], in which genetic diversity accumulates over a timescale that is observable by humans, but they also contain a large number of estimated mutation and substitution rates, allowing broad-scale comparisons. For the first time, we have sufficiently large genomic data sets and sophisticated analytical methods to rigorously estimate evolutionary rates in viruses of many genomic architectures, allowing us to test the widespread assumption that population substitution rate is solely a function of the underlying polymerase error rate.

We begin by reviewing our current knowledge of viral mutation rates, and how they vary within and among genomic architectures. We then summarize the evidence that supports the conventional wisdom that RNA-containing viruses have higher substitution rates than DNA-containing viruses, reflecting underlying differences in polymerase fidelity. Finally, we note the exceptions to this large RNA–DNA divide, and discuss the general features of viral life cycles and genome structures that are predictive of higher rates of evolutionary change.

Mutation rates in viruses

The range of mutation rates in RNA and DNA viruses. Our knowledge of mutation rates in viruses has expanded greatly since studies of mutational frequency in bacteriophages began in the 1940s. Our current understanding includes mutations that are induced by polymerase

Box 1 | Viruses as model systems to reveal evolutionary dynamics

Many features of small, rapidly evolving viruses (RNA based or single-stranded DNA based) make them ideal systems for studying rates of evolutionary change.

First, there is the wealth of knowledge on viral structure and function, coupled with an abundance of viral nucleotide sequences available on public databases such as GenBank[97], including an increasingly large number of whole genome sequences. These data have been acquired from economically and medically important pathogens as well as culture-independent environmental metagenomic studies[98]. Sequence data that are suitable for substitution rate studies therefore exist as a common resource for viruses with a wide range of genome structures, hosts and ecologies.

Second, a great deal is known about the molecular biology, replication and life cycles of particular viruses, which enables a connection to be made between processes at the evolutionary, physiological and structural levels.

Third, because they can be maintained easily in culture, some viruses represent powerful tools in studies of experimental evolution. This allows research on substitution and mutation rates to move beyond the purely comparative to direct hypothesis testing (for example, mutation rates in the double-stranded DNA bacteriophage RB69 (REF. 25)).

Finally, although viruses lack a fossil record, their evolution can often be recorded over the timescale of human observation, so that they are 'measurably evolving populations'[7]. Importantly, such measurable evolution allows the use of coalescent methods that are based on serial (time-structured) sampling to estimate rates of nucleotide substitution (BOX 3).

errors, nucleotide base modifications caused by other cellular enzymes, and how both replication and recombination can introduce insertion and deletion mutations into viral genomes[8,9]. There are various estimates of viral mutation rates, reflecting the changes in nucleotide sequence that occur during each round of viral replication (see BOX 2 for an explanation of mutation rates and how they are measured). These estimates span several orders of magnitude — from 1.5×10^{-3} mutations per nucleotide, per genomic replication (mut/nt/rep) in the single-stranded (ss) RNA phage Qβ[10], to 1.8×10^{-8} mut/nt/rep in the double-stranded (ds) DNA virus herpes simplex virus type 1 (HSV-1)[11]. This range approximately corresponds to the fidelity of the polymerases used in replication: RNA viruses (which utilize RNA-dependent RNA polymerases; RdRps) mutate faster than retroviruses (with RNA-dependent DNA polymerases (RdDps) or reverse transcriptases (RTs)), which mutate faster than DNA viruses (with DNA polymerases) (FIG. 1). Another relationship — between mutation rate and genome size — was noted by Drake, who proposed a 'universal' genomic mutation rate in DNA microorganisms of 3.4×10^{-3} mutations per genome, per genomic replication[2]. Although this rate clearly reflects the estimates obtained in some DNA viruses, its universality is open to question; the ssDNA phage ϕX174 (REF. 12) has a mutation rate an order of magnitude higher than predicted under this model (FIG. 1), whereas the papillomaviruses, with small dsDNA genomes, evolve more slowly than predicted (see below).

In vitro assays have confirmed that RdRps and RdDps are more error prone than DNA polymerases[13]. Notably, DNA polymerases can possess error-correcting domains, which further reduce the mutation rate during DNA replication by at least an order of magnitude[14]. No known RNA polymerase has this proofreading capability. Additionally, there are enzymes that perform base-excision repair on mispaired bases and mutated dsDNA, but that are unable to fix such mistakes in dsRNA or RNA–DNA heteroduplexes[14]. Finally, the fidelity of RTs is higher than that of RdRps (although it is still lower than that of DNA polymerases), resulting in lower mutation rates in retroviruses than in RNA viruses. For example, when not integrated into host genomes, rates of 0.1–0.2 mutations per genome, per generation have been estimated for viruses like the human immunodeficiency virus type-1 (HIV-1) and hepatitis B virus (HBV)[2,10], a rate that is approximately fivefold lower than observed in some viruses that replicate using RdRps[10]. Hence, the rapid evolutionary dynamics that are exhibited by HIV-1 also reflect high rates of recombination[15] and frequent positive selection[16].

Causes of variation in mutation rates. Studies of mutational dynamics *in vivo* have revealed that additional factors to those described above, including the oxidation and methylation of bases, also affect mutation rates[10]. One well-studied supplemental mechanism of viral mutation is the deamination of unpaired nitrogenous bases. Several cellular deaminating enzymes exist, including those in the apolipoprotein B-editing catalytic polypeptide-like subunit (APOBEC) family, which add further transition mutations to polymerase errors[9,17]. This enzymatic deamination is thought to be an intrinsic antiviral host defence mechanism[18] and can lead to long stretches of transitions, termed hypermutation, the products of which are usually non-functional because of the acquisition of multiple deleterious mutations[19]. Deamination can also be chemically induced and can occur spontaneously, especially if the genome spends a significant amount of time in a single-stranded state[20,21]. Viral mutation rates can also depend on the particular host species that is infected, although the mechanisms responsible for this rate variation are unknown[22].

Comparisons among *in vivo* mutation rates can be confounded by inconsistent measurement schemes (BOX 2). For example, lethal mutations are not counted in mutation assays that involve culturing, and many mutation-rate assays measure a phenotype that can be caused by mutations at multiple unique sites. Factors such as these can deflate or inflate reported per-nucleotide mutation rates[23,24]. Moreover, genomes that acquire multiple mutations per replication, which are more abundant than predicted in viral, bacterial and eukaryotic systems, might be recorded as having single mutations, further deflating the measured mutation rate[25–27]. It is also likely that mutation rates vary across the viral genome. Secondary structures in the genome can cause the polymerase to pause, increasing the chance of template slippage, which would lead to deletions[22,28]. Therefore, the existence of secondary structures might explain some of the variation in mutation rate within viral genomes[27]. Genomic context is also a factor: adjacent nucleotides affect mutation rate[28], and the precise location of a nucleotide in the genome can influence its mutation rate, as observed in some bacteria[29]. Although these factors will create *bona fide* 'hot spots' and 'cold spots' of mutation, these terms are often incorrectly

Coalescent
A population genetic theory that links the divergence times of a phylogenetic tree of individuals sampled from the same population with the demographic history (that is, rates of population growth and decline) of that population.

Positive selection
The fixation of advantageous alleles as a result of differential reproductive success.

Hypermutation
The long stretches of nucleotide transitions observed in RNA virus sequences (first noticed in human immunodeficiency virus with G-to-A transitions). This term can also be used to describe an elevated mutation rate of any kind, not necessarily in a run of adjacent nucleotides.

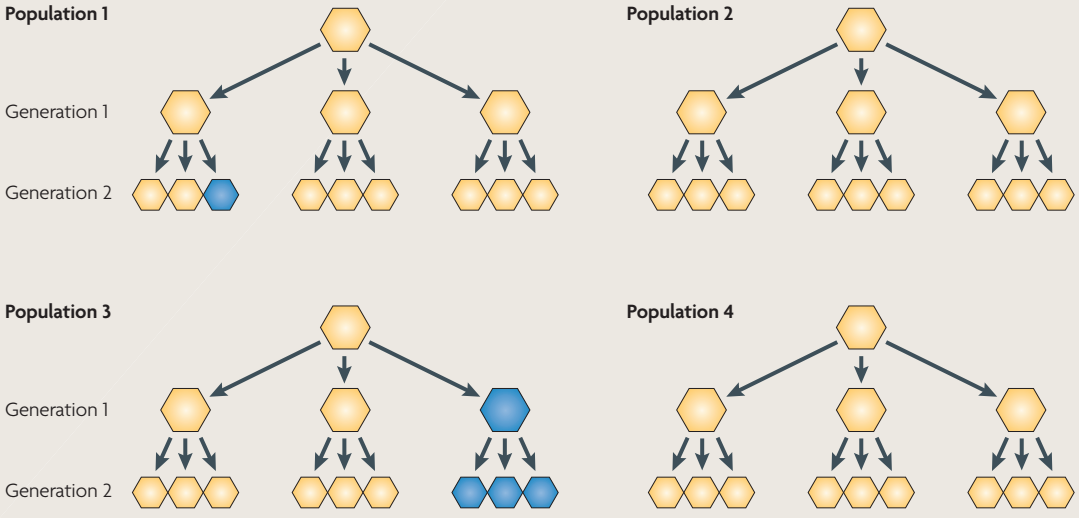
Box 2 | **Measuring mutation rates**

Mutation rate refers to the number of genetic errors (point mutations, insertions and deletions) that accumulate per unit time[62], or per generation[30] (for obligately lytic viruses, per burst), or per round of genomic replication[2,10,65]. The most useful of these measures is often the mutation rate per round of replication, although this can be difficult to determine as it requires knowledge of the details of viral replication, such as linear 'stamping machine' replication versus geometric genomic replication, or a mixture of the two replication modes[27]. Mutation rate is measured in one of two ways: through Luria–Delbrück fluctuation tests or mutation accumulation studies[65].

The classic Luria–Delbrück method involves measuring the frequency of mutations with a certain phenotype arising in replicate clonally expanding populations (see the figure). This frequency is then adjusted to account for the number of generations and the number of genome replications within each generation to obtain a mutation rate (per round of replication). These rates can also be calculated from the number of replicate populations that do not generate any mutant[10,65], assuming the number of mutations follows a Poisson distribution. Almost always, the Luria–Delbrück method detects point mutations in protein-coding genes, such as amber mutation reversions[10,30]. Although mutation-frequency assays are excellent measures of a limited number of mutations at one site, they do not inform the mutation frequency per nucleotide if mutations at multiple sites can cause the requisite mutant phenotype.

The figure shows replicate clonally expanding populations of a virus carrying an amber mutation, with a burst size of three, undergoing two generations of growth on an amber-suppressor host. Note the uneven distribution of mutants in each final population of viruses; reversion of the amber mutant phenotype occurred in two of the four populations (as indicated by the blue colour). The per-base, per-generation mutation rate is calculated by taking into consideration the frequency of reversion of the amber mutation at a particular site, given the number of generations of viral growth, adjusted for the multiple ways in which reversion is possible[11,65]. An estimate of the mutation rate per base, per round of genome replication can also incorporate differences between linear and geometric viral replication.

Mutation accumulation studies are defined in two ways. Phenotypic measures of mutation accumulation involve subjecting populations to bottlenecks and following the mutation frequency through these changes in population size[65]. Alternatively, sequence data can be obtained from genomes of replicate populations after a known number of generations[1]; however, because this method cannot exclude all the effects of natural selection, it can reflect the substitution rate rather than the mutation rate.



Amber mutaton reversion
The change of an amber stop codon (UAG) within a gene to a codon for an amino acid. This typically restores protein function in a gene that had been purposefully selected to contain an amber nonsense mutation as a genetic marker.

Stamping machine
Linear stamping machine replication is when the single virus that initiates an infection is the direct parent of all progeny genomes. That is, the parental genome (or its single complement) is the template for the semi-conservative replication of all the genomes that are produced in an infected cell. As there is only one template within the cell, progeny genomes accumulate linearly over time.

Geometric genome replication
A mode of viral replication in which the progeny genomes that are replicated early during infection can become templates for further genome replication. As the infection progresses, the number of templates for semi-conservative replication increases, and progeny genomes can be produced at an exponential, or geometric, rate.

applied to regions of frequently or infrequently assayed mutation. Such mutation-frequency assays usually require that the mutations have experienced selection to be functional (they cannot be lethal), so that the observed regions of high diversity should be more correctly referred to as hot spots of polymorphism or of substitution.

Finally, it is also likely that the precise mechanism of viral replication has a large effect on mutation rates. A single infecting virus could be the template (or the complement of the single template) for all progeny genomes — a so-called 'stamping machine' — in which case, mutations would accumulate linearly. Alternatively, some of the first progeny genomes can themselves become templates for further progeny[30] (BOX 2) and mutations should accumulate geometrically because a mutated template propagates the given error to all of its replicate copies[31]. Given the same intrinsic rate of polymerase error, stamping-machine replication results in lower overall mutation rates than geometric genome replication, although it is possible that some viruses can use both modes of replication[2]. For instance, the dsRNA bacteriophage ϕ6 seems to replicate mainly, but not exclusively, by a stamping-machine mechanism[30].

Substitution rates in viruses: the RNA–DNA divide

A strong prediction from studies of mutation rates is that viruses with an RNA stage should evolve quickly, whereas those with only DNA stages should evolve slowly (FIG. 2). Although most estimates of substitution

Box 3 | Measuring substitution rates in serially sampled viruses

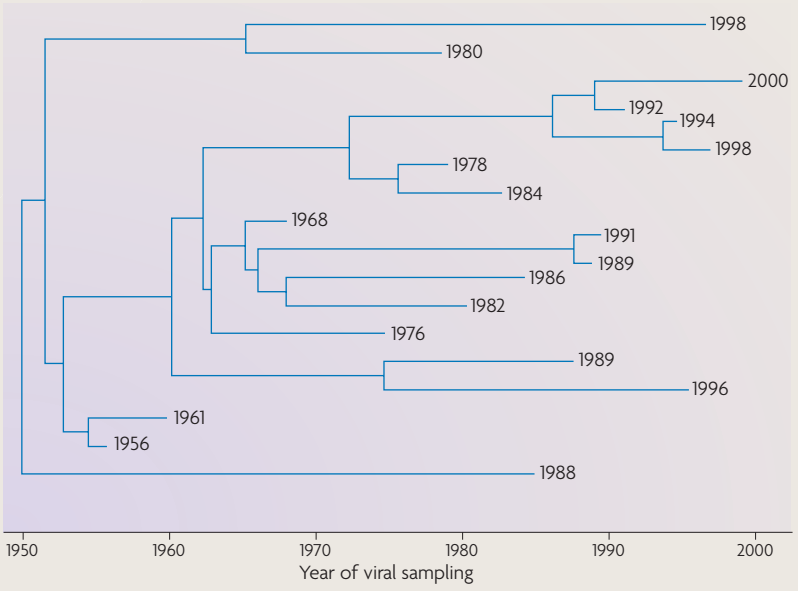
Accurately estimating rates of nucleotide substitution is an important research area in computational biology. The rate of nucleotide substitution is defined as the number of fixed (by natural selection or genetic drift) mutational changes per nucleotide site, per unit time (usually years). Substitution rates therefore reflect a complex product of four factors: underlying mutation rate, generation time, effective population size and fitness, with advantageous mutations fixed faster than neutral mutations.

Traditionally, estimates of substitution rate have been based on phylogenetic methods, in which the number of nucleotide differences is counted between sequences that are known to have diverged at a specific time, for instance, as determined by reliable fossil records for some cellular species. The signal of evolutionary rate in rapidly evolving viruses is encoded in the distribution of branch lengths in phylogenies of viruses that are sampled at different times (over the course of years, months or even days) — so-called serially sampled data (see the figure). Given this information, several methods can be used to estimate substitution rates.

The most commonly used method to estimate viral substitution rate is linear regression. Although this approach provides a useful overview[99], it suffers from two important limitations. First, because all sequences are compared, in a pairwise manner, with the oldest sequence, there is widespread pseudo-replication, such that deep branches are compared multiple times. Second, linear regression implicitly assumes a constant molecular clock, an assumption that only seems to fit a subset of RNA viruses[32]. Resolving the problem of phylogenetic non-independence was one of the principal motivations behind the development of maximum likelihood methods such as TipDate[100]. In these methods, a count is made of the number of substitutions on each branch of a phylogenetic tree with dated tips (although frequent recombination clearly compromises any analysis that is based on a single phylogeny).

The most recent class of methods that was developed to estimate substitution rates are set within a Bayesian Markov chain Monte Carlo (MCMC) coalescent framework, as manifest in the BEAST (Bayesian evolutionary analysis by sampling trees) package[101]. The power of this approach is that, as well as incorporating phylogenetic information, it accounts for variable substitution rates among lineages (through the incorporation of relaxed molecular clocks rather than constant molecular clocks[7,102]) and for differences in the demographic history of RNA viruses (that is, in rates of population growth). Furthermore, rate estimates are averaged across a large sample of plausible trees, therein accounting for uncertainty in tree estimation, providing a more rigorous statistical framework. However, these methods are computationally intensive and care must be taken with the choice of prior distributions.

Finally, an important caveat in all these methods is that they assume that the sampled sequences contain only fixed substitutions. Although nucleotide changes that fall on 'deep' branches of phylogenetic trees represent mutations that have reached high frequency in the population (including fixation), a subset of changes that fall on terminal branches will constitute transient polymorphisms that will ultimately be lost from the population. Consequently, the evolutionary rates that are estimated using serially sampled data in reality reflect a composite mutation and/or substitution rate parameter. This, in turn, leads to artificially high estimates of the long-term substitution rate when sampling has occurred over a short time-period (because in this case estimates approach the mutation rate).



The figure depicts a schematic representation of a time-structured viral phylogeny using serially sampled gene sequence data, in which tip times correspond to the year of viral sampling. Analysis of the distribution of tip times within a coalescent framework allows estimates of the rate of nucleotide substitution to be made.

Effective population size
The smallest theoretical population size that can evolve in the same way as the actual population under study. It is strongly influenced by population bottlenecks, such as those that occur during transmission of viruses between hosts, and therefore is often smaller than the total population size.

Linear regression
The estimation of a first-order relationship between two variables (for example, number of nucleotide substitutions and time), which involves fitting the best straight line to the data.

Constant molecular clock
The idea that nucleotide substitutions accumulate at a fixed (constant) rate over time, and that this can be used to estimate divergence times between sequences.

Maximum likelihood
A statistical method that selects the hypothesis (for example, the phylogenetic tree) that has the highest probability of explaining the data, under a specific model.

Bayesian Markov chain Monte Carlo
(MCMC). Bayesian methods incorporate prior information in assessing the probability of model parameters. Because the prior distribution (the users' belief about the probabilities of different parameter values before the data have been analysed) can have a large affect on the posterior distribution (the results) it must be chosen carefully. MCMC methods allow sampling from the posterior distribution to get an estimate of the distribution.

Relaxed molecular clock
A form of molecular clock in which rates of nucleotide substitution are allowed to vary among lineages.

rate support this basic division between DNA and RNA viruses, work in recent years has shown that the boundary between these groups of viruses with respect to rates of evolutionary change is more blurred than previously thought. Exceptions to the assumed dichotomy hint that additional aspects of viral genomic architecture and life cycle determine the overall tempo of viral evolution.

Substitution rates in RNA viruses. Evolutionary change in many RNA viruses can be easily observed in real time, providing strong evidence for their rapid evolution[7]. For nearly all RNA viruses examined, overall rates of nucleotide substitution fall in the range of 10^{-2} to 10^{-5} nucleotide substitutions per site, per year (subs/site/year)[32,33], with most of the viruses exhibiting rates within

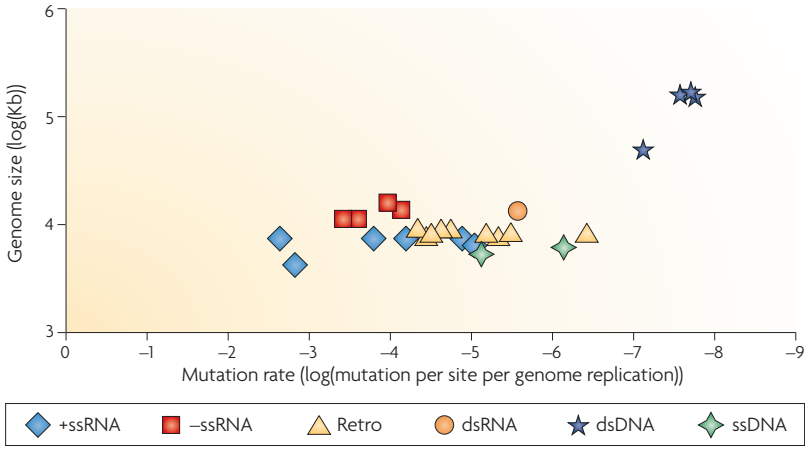


Figure 1 | **Average rates of spontaneous mutation in viruses, adjusted to the rate per genome replication.** Positive-sense single-stranded (ss) RNA viruses (+ssRNA; RNA phage Qβ, poliovirus[10] and Tobacco mosaic virus[103]), negative-sense ssRNA viruses (-ssRNA; vesicular stomatitis virus, influenza A virus[10] and measles virus[23]), the retrotranscribing viruses (Retro; spleen necrosis virus, murine leukaemia virus, Rous sarcoma virus[10], human immunodeficiency virus type-1 and bovine leukaemia virus[2]) and the double-stranded (ds) RNA virus (dsRNA; bacteriophage ϕ6 (REF. 30)) have RNA-dependent polymerases without any proofreading capabilities. Large dsDNA viruses (dsDNA; bacteriophages λ, T2 and T4 (REF. 65), and herpes simplex virus type 1 (REF. 11)) encode their own DNA polymerases. The two ssDNA viruses (ssDNA; bacteriophages M13 (REF. 65) and ϕX174 (REF. 12)) use the DNA polymerase of their host, *Escherichia coli*. There are multiple independent estimates for several of these viruses and all were included in this figure.

one order of magnitude of 1 x 10^{-3} subs/site/year. For an RNA virus with a genome of 10 kb, this is equivalent to the fixation of 10 substitutions per genome, per year, although there is a marked absence of estimates of substitution rate in dsRNA viruses. Although rates of non-synonymous substitution vary widely among RNA viruses (and among viral genes), these differences have only a small impact on overall substitution rates. For example, lower rates of non-synonymous change are found in viruses transmitted by arthropod vectors, reflecting the elevated purifying selection pressure that is associated with replication in diverse host species[34–37].

Similarly, in viruses that establish chronic infections, the contrasting modes of inter- and intra-host evolution might also influence evolutionary rates. For example, in HIV-1 there seems to be an inverse relationship between rates of viral transmission and rates of evolutionary change, with the highest rates observed at the intra-host level[38]. The elevated rate of nucleotide substitution within hosts might be because this part of the viral life cycle is dominated by the positive selection of amino-acid changes that facilitate immune escape[39]. Alternatively, it might be that many of the mutations that occur within hosts are purged at inter-host transmission owing to strong purifying selection, most notably because of a mismatch with host cytotoxic T-lymphocyte responses[40]. Therefore, even among rapidly evolving RNA viruses, there is marked variation among substitution rates that can only be explained by understanding factors such as generation time, the relative extent of inter- versus intra-host evolution, and population size — the selective fixation of advantageous mutations will occur with greater efficiency in larger populations (BOX 3).

Purifying selection
The purging of deleterious alleles as a result of differential reproductive success.

Cytotoxic T-lymphocyte
An antigen-specific T-cell of the vertebrate immune system that recognizes and destroys virus-infected cells.

Co-divergence
The parallel diversification or speciation of a parasite and its host, which is inferred when there is strong congruence between the phylogenetic trees of the host and parasite, and similar divergence times of corresponding nodes on the phylogenies.

Some RNA viruses evolve slowly. Importantly, a small number of RNA viruses experience anomalously low rates of nucleotide substitution. The best documented of these is the retrovirus simian foamy virus (SFV)[41]. Phylogenetic studies have revealed that the phylogeny of SFVs is largely congruent with that of its primate hosts. This widespread congruence allows rates of viral evolution to be estimated through calibration with the primate fossil record, resulting in values of only 1.7×10^{-8} subs/site/year[41]. However, genetic diversity has been observed among SFV isolates sampled within individual primates[41], indicating that the SFV RT has not evolved additional mechanisms of error correction. Consequently, the most likely explanation for the reduced substitution rates seen in SFV is that the virus is largely latent within hosts, primarily undergoing replication as integrated dsDNA within primate genomes. A reduced rate of replication associated with latency has also been proposed to explain the low rate of approximately 10^{-7} subs/site/year that is observed in another retrovirus — human T-cell lymphotropic virus type II (HTLV-II). In this case, low substitution rates are thought to be associated with low rates of inter-host transmission and correspondingly long periods of time within a single host, so that viruses largely spread through the clonal expansion of infected cells (in which the virus is integrated into host DNA), rather than active replication[42,43]. More recent and precise estimates of substitution rates in HTLV-II and its relative HTLV-I similarly revealed low rates of nucleotide substitution, although rates were higher when analyses were based on family-level transmissions compared with those that were based on virus–host co-divergence[44]. Indeed, far higher rates (>10^{-4} subs/site/year) are observed among HTLV-II populations that are experiencing rapid epidemic transmission because, in this case, active replication is required to initiate infection of each new host[42,43]. Together, both SFV and HTLV-I/II reveal how differences in viral generation time combine with mutation to shape overall rates of nucleotide substitution.

More controversial are those cases in which RNA viruses replicating with an RdRp, rather than an RT, are reported to evolve slowly. Three types of viruses fall into this class — the rodent-associated hantaviruses and the flavivirus GB virus C (GBV-C), in which substitution rates in the range of 10^{-7} subs/site/year have been inferred[45,46], and some RNA viruses that infect plants[47]. Although the two animal viruses differ in genome structure, both are associated with chronic rather than acute infections, and the low evolutionary rates proposed for these viruses rest on the common assumption that they have co-diverged with their mammalian hosts over millions of years[46,48,49]. However, given the small number of taxa that are involved in these co-divergence studies, especially for GBV-C, and that viruses can preferentially jump between related host species and those that live sympatrically[50], studies utilizing serially sampled data (BOX 3) are required to unequivocally show that these viruses evolve slowly.

There have been numerous suggestions that plant RNA viruses evolve more slowly than RNA viruses that infect animals. Both tobamoviruses and closteroviruses

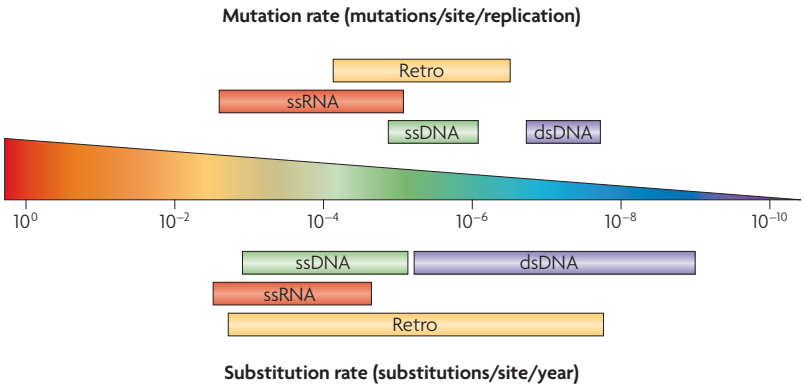


Figure 2 | **Comparison between viral mutation and substitution rates.** The ranges of mutation rates, given as mutations per site per round of replication, for viruses with different genomic architectures are summarized in the upper part of the figure (values from FIG. 1). The ranges of average substitution rates, given as substitutions per site per year, are shown in the lower part of the figure. The ranges are bound by the extremes of average substitution rates: for the retrotranscribing viruses (Retro), human immunodeficiency virus type-1 (REF. 104) and simian foamy virus[41]; for single-stranded (ss) RNA viruses, swine vesicular stomatitis virus[32] and European bat lyssaviruses[105]; for ssDNA viruses, the non-coding region of the Tomato yellow leaf curl virus[70] and canine parvovirus[68]; for double-stranded (ds) DNA viruses, the BK polyomavirus[95] and herpesvirus[57].

have shown few genetic changes between isolates that are separated in space[51] and time[52], even after decades[53] or centuries[54]. It has been proposed that these low rates are due to severe population bottlenecks[55], although changes in population size will not affect the rate of neutral substitution. Similarly, it has also been suggested that rates of non-synonymous substitution are reduced in plant viruses compared with their animal counterparts because of weaker immune-mediated positive selection[47]. However, the first estimate of the substitution rate of a plant RNA virus that was made using serially sampled data was recently calculated at >10^{-4} subs/site/year, which is firmly within the distribution of substitution rates of RNA viruses observed in animal RNA viruses[56].

Substitution rates in DNA viruses. DNA viruses are usually characterized by far lower rates of nucleotide substitution than RNA viruses, although with evidently more variation among taxa. For large dsDNA viruses of animals, rates of evolutionary change are often estimated under the assumption that these viruses have co-diverged with their hosts over millions of years. For the best documented of these cases, that of the gammaherpesviruses of vertebrates, this assumption translates into evolutionary rates in the range of 10^{-9} subs/site/year, and hence close to the values seen in multicellular species[57]. Similarly low rates (approximately 10^{-8} subs/site/year) have been estimated in some small dsDNA viruses that are also thought to have undergone host–virus co-divergence, such as the vertebrate-infecting papillomaviruses[58]. However, the notion that DNA viruses generally co-diverge with their hosts, and that this provides a molecular clock calibration point by which to estimate substitution rates, is evidently an assumption rather than a truism. For example, it has traditionally been assumed that the dsDNA JC polyomavirus (JCV) co-diverged with human populations since their migration out of Africa within the last 200,000 years[59]. Under this assumption, estimates of substitution rate are approximately 10^{-7} subs/site/year[60] and hence far lower than the rates observed in RNA viruses. However, there is no significant match between JCV and host phylogenies, and estimates of substitution rate that are based on serially sampled data (BOX 3) are generally higher[61]. Hence, the true rate and timescale of JCV evolution is unclear.

Some DNA viruses evolve rapidly. One of the most important developments in recent studies of viral evolutionary dynamics is the mounting evidence that ssDNA viruses evolve at rates approaching those observed in their RNA counterparts. Indeed, those working with ssDNA viruses have consistently observed relatively high levels of genetic diversity both within and among hosts[62–64]. Although the proposed universal genomic mutation rate for DNA microorganisms[2,65] requires that the ssDNA viruses (all of which have genomes smaller than ~13 kb) have high mutation rates, it was originally thought that these viruses had low mutation rates because of their reliance on host polymerases for replication[13]. Consequently, other explanations, such as frequent recombination, were sought for the high levels of intra-host genetic diversity[66,67].

The first precise estimates for the rate of ssDNA virus evolution came from a study of the well-documented emergence of canine parvovirus (CPV-2) from feline panleukopenia virus (FPV) in the last ~40 years[68]. In both CPV-2 and FPV, a substitution rate of approximately 10^{-4} subs/site/year was estimated, which is within the range that is seen in RNA viruses. Equally high rates of nucleotide substitution have now been determined in various other ssDNA viruses, including the human parvovirus B19 (REF. 61), the circovirus SEN-V[69] and the anellovirus SEN-V and the plant geminivirus Tomato yellow leaf curl virus[70]. Consequently, ssDNA viruses as a whole seem to have substitution rates closer to the similarly sized RNA viruses (FIG. 2) than to larger dsDNA viruses, and mutation rates that are intermediate between RNA and dsDNA viruses (FIG. 1).

What determines evolutionary rates in viruses?

In general, it is clear that differences between high-fidelity DNA polymerases, with error-correcting mechanisms, and low-fidelity RNA polymerases, without error correction, explain many of the differences in mutation rate among viruses. However, this broad observation cannot fully explain the range of evolutionary dynamics seen in viruses and does not consider the nature of the selection pressures that act to optimize these rates. To address these issues we concentrate on the factors affecting mutation rate, as this also determines the neutral substitution rate. The factors that increase or decrease the substitution rate relative to the rate that is expected under neutrality are summarized in FIG. 3.

Optimizing mutation rates. In any genetic system most mutations are deleterious[1,71] and viruses are no exception; for example, more than 60% of spontaneous mutations

Population bottleneck
The smallest size of a viral population at any point in viral propagation. A common bottleneck for viral populations occurs during transmission between hosts, when the population size can be as small as one virus or as large as several million virions.

Error threshold
The theoretical limit to the mutation rate of viruses, beyond which too many errors accumulate and populations of the virus become extinct. It is used to explain why it is difficult to generate RNA viruses with much higher mutation rates than those observed in natural isolates, and why RNA viruses have constrained genome sizes.

Burst size
The (average) number of progeny viruses produced from a single infected cell. This is more straightforward to measure for obligately lytic viruses than for viruses that can integrate into their host genomes.

Robustness
The constancy of a phenotype in the face of changing environments or changing genetics (mutations). Current research indicates that robustness is a trait that is under selection in viruses, and changes in viral robustness can be observed in laboratory experimental evolution.

in vesicular stomatitis virus (VSV) were found to be deleterious[72]. In theory, natural selection should favour a reduction in mutation rates in static environments to reduce this burden of deleterious mutation (although non-zero mutation rates might still be optimal[73]). However, viruses rarely experience a static environment, particularly as they often struggle against both innate and adaptive host immunity. As such, viruses probably always experience selection for non-zero mutation rates[74]. Therefore, the upper limit on mutation rate is a product of factors such as natural selection, genomic architecture[75] and the ability to avoid loss of viability[24] and/or genetic information[76].

We might also expect that random mutation will, on occasion, produce genotypes with higher (and lower) than optimal mutation rates, and that these might even be favoured by natural selection for a time. Indeed, transient periods with higher mutation rates are one explanation for the observation of larger than expected numbers of mutations within a single viral genome[25–27]. Furthermore, recent work has shown that replication fidelity can be altered by single base changes in viral polymerases or associated proteins[77,78], producing phenotypes that are often referred to as 'mutator' (high mutation rate) and 'antimutator' (low mutation rate), and which could facilitate rapid changes in mutation rate. However, the times when mutator phenotypes are favoured must be limited, as theoretical models predict a rapid return to lower mutation rates[79]. Intriguingly, whereas DNA viruses can evolve much higher and lower mutation rates, strong mutator phenotypes have not been observed for RNA viruses[77], perhaps because they already exist at an 'error threshold', beyond which high mutation rates would result in a large reduction in fitness and eventual extinction[35]. Similarly, recent work also suggests that antimutator phenotypes might suffer more costs than simply an increased time to sample beneficial mutations[80]. In particular, increased polymerase fidelity reduces fitness (as measured by burst size) in both T4 and VSV[77,81]. It also seems that when deleterious mutations do occur in antimutators they have a larger effect on fitness[82]. Importantly, this observation links the evolution of lower mutation rates to lower levels of robustness, a mutational buffering mechanism that reduces the effect of each new mutation[83].

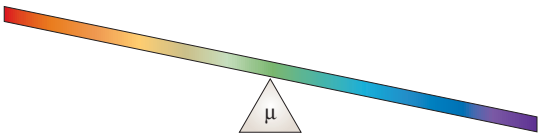
As noted above, evolutionary theory predicts that higher mutation rates should be favoured when viruses experience a changing environment, for example, in host immune defence[74,79,84]. However, experiments in which the mutation rate in RNA viruses is artificially increased are often associated with much lower viral titres[76]. For example, a 10-fold increase in mutation rate reduced poliovirus titres 1,000-fold[85]. Indeed, nucleoside analogues such as ribavirin have proven useful in greatly reducing viral loads in both laboratory and clinical settings[85,86], in theory because they increase the viral mutation rate to levels at which viable genomes are only able to produce unfit progeny[24,87]. Because even small increases in RNA-virus mutation rates have serious fitness effects, it is likely that many RNA viruses have mutation rates close to the highest rates they can tolerate[78,87].

Mutation rate and genome size: an error threshold. A further consequence of the idea that RNA viruses live at the edge of an error threshold is that there is a cap on their genome sizes; given the same rate of mutation per nucleotide, larger RNA genomes would suffer lethal mutations more frequently than smaller RNA genomes. This relationship is the basis of a long-standing theory that the maximum genome sizes of RNA viruses are set at approximately the reciprocal of their mutation rates[88]. Under this theory the upper limit on the genome size of a virus with a given genomic architecture will be determined by the lowest mutation rate that is sustainable by that architecture in nature. The fact that ssDNA viruses, like RNA viruses, evolve rapidly and possess similarly small genome sizes supports this hypothesis. Consequently, two distinct and unrelated viral genomic architectures that share high mutation rates — RNA viruses and ssDNA viruses — could be size limited by similar error thresholds. The error rates that are observed in these groups of viruses might therefore represent evolutionary optima that allow for a sufficient number of beneficial mutations to rapidly respond to changing environments, yet not so many as to breech the error threshold.

Alternatively, the limited genome size of RNA viruses could be due to biochemical processes that are independent of mutation rate. For example, the instability of RNA macromolecules could be one reason why RNA viruses do not exceed 33 kb in size. If faster replication of smaller molecules is favoured (see below), this could lead to high mutation rates as an inadvertent consequence[2].

Effects on substitution rate

Generation time	Shorter		Longer
Transmission	Direct		Vector borne
	Horizontal		Vertical
Selection	Positive		Purifying



Effects on mutation rate

Genomic architecture	Single stranded		Double stranded
	RNA		DNA
	Smaller		Larger
Replication speed	Faster		Slower
Viral enzymes	–		DNA repair
Host enzymes	Deamination, oxidation		–
Environmental effects	Deamination, oxidation, UV		–

Figure 3 | **Factors influencing mutation and substitution rates in viruses.** In the upper portion of the figure, the factors that increase viral substitution rates are shown in red and those that decrease it are shown in blue. The baseline substitution rate is determined by the neutral mutation rate, μ. Because increasing mutation rate can increase the substitution rate, the factors increasing mutation rate are also shown in red and those decreasing mutation rate in blue. It is important to note that many of these factors are not independent; for instance, smaller genomes tend to replicate faster, or vertically transmitted viruses sometimes become latent in the host genome and so experience fewer generations compared with obligately lytic, horizontally-transmitted viruses. UV, ultraviolet irradiation.

Fitness landscape
A metaphorical contour map of the varied fitness values that are experienced by different genotypes of an organism. As a genotype moves through genotype space, it can climb to a higher fitness peak, or stumble down to a less-fit genotype.

Mutation rate and genome size: speed versus accuracy. It is also possible that there is an evolutionary trade-off between replication speed and replication fidelity, and that high mutation rates are simply a consequence of selection for rapid replication[1] (FIG. 3). This trade-off is well established for the two replication modes of viruses (linear stamping machine and geometric genomic replication): the stamping machine replication, which has a higher fidelity, produces progeny genomes more slowly[31]. Similarly, selection for rapid replication would also favour viruses with the minimal genome sizes that their genome structure would allow.

Furthermore, if high mutation rates are not always beneficial, then one might expect viral traits to evolve that would compensate for the negative effects of high mutation rates, particularly the accumulation of deleterious mutations. Such traits have been proposed in the form of positive epistasis and the propensity to occupy flat regions (in which there are many possibilities for neutral movement) in fitness landscapes, manifested as sequence robustness[89,90].

Conclusions and future directions

Although recent work has shown that there is no strict or inherent divide between the evolutionary dynamics of RNA and DNA viruses, there is no coherent explanation for why this is so. This lack of clarity is largely a function of substantial gaps in our knowledge of the basic mechanisms by which mutations are generated and propagated in viral populations. For instance, little is known about the many factors that influence polymerase fidelity in viruses. Furthermore, existing measures of mutation rate focus only on mean rates, with little insight into the overall distribution of mutation rates during the replication process[25–27]. As a case in point, we still do not know the distribution of error rates within the replication cycle of a single RNA virus — what fraction of the progeny have multiple mutations and what fraction have no mutations at all — although understanding this variation is crucial to understanding adaptability. For example, the occurrence of multiple advantageous mutations in a single replication cycle might be required for viruses to successfully jump species boundaries[91]. Similarly, it is uncertain what portion of the *in vivo* mutation rate in DNA viruses is due to polymerase error, as opposed to the differing effects of host and viral error-correction enzymes[92]. It is possible that these enzymes are important in explaining the variation in rates of mutation per nucleotide together with similar rates of mutation per genome.

A thorough understanding of polymerase error rates in viruses with various genomic architectures is essential before the effects of generation time, mode of replication and polymerase-independent mutation on substitution rate can be determined. Small DNA viruses might be a particularly useful study system in this endeavour, as their genomes can be either single- or double-stranded. Furthermore, these viruses do not encode their own polymerases and, consequently, must use error-correcting host DNA polymerases. In these circumstances it will be important to determine how the small ssDNA viruses achieve mutation rates that are orders of magnitude higher than those of their hosts[65]. For instance, ssDNA viruses might not be able to take advantage of DNA correction and repair enzymes because the genomes might not be methylated[93], or the double-stranded intermediate might be inaccessible to host enzymes. Alternatively, viral proteins might interact with host factors to alter polymerase fidelity[93,94]. Unfortunately, however, rigorous analyses of substitution rates in small dsDNA viruses have proven difficult[95,96], and there are currently no estimates of mutation rates in these viruses to compare with those of ssDNA viruses.

Finally, despite our emphasis on the process of mutation in viral systems, it is important to recall the complex relationship between underlying mutational dynamics and the ability to generate antigenic variation, which, in turn, has important implications for large-scale epidemiological dynamics and vaccine efficacy. In particular, despite the overall similarity in substitution rates among RNA viruses, they differ radically in levels of antigenic diversity, as exemplified by measles virus (with low antigenic diversity, allowing the creation of a stable vaccine) and influenza A virus (with high antigenic diversity, such that vaccines need to be continually updated). Explaining this dichotomy represents one of the fundamental challenges in studies of RNA virus evolution.

1. Baer, C. F., Miyamoto, M. M. & Denver, D. R. Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nature Rev. Genet.* **8**, 619–631 (2007).
2. Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. F. Rates of spontaneous mutation. *Genetics.* **148**, 1667–1686 (1998).
3. Li, W.-H. *Molecular Evolution* (Sinauer Associates, Inc., Sunderland, Massachusetts, 1997).
4. Posada, D., Crandall, K. A. & Holmes, E. C. Recombination in evolutionary genomics. *Annu. Rev. Genet.* **36**, 75–97 (2002).
5. Philippe, H. & Douady, C. J. Horizontal gene transfer and phylogenetics. *Curr. Opin. Microbiol.* **6**, 498–505 (2003).
6. Clark, A. G. *et al.* Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**, 203–218 (2007).
7. Drummond, A. J., Pybus, O. G., Rambaut, A., Forsberg, R. & Rodrigo, A. G. Measurably evolving populations. *Trends Ecol. Evol.* **18**, 481–488 (2003).
8. Gibbs, A. J., Calisher, C. H. & García-Arenal, F. *Molecular Basis of Virus Evolution* (Cambridge University Press, New York, 1995).
9. Domingo, E. & Holland, J. J. RNA virus mutations and fitness for survival. *Annu. Rev. Microbiol.* **51**, 151–178 (1997).
10. Drake, J. W. Rates of spontaneous mutation among RNA viruses. *Proc. Natl Acad. Sci. USA* **90**, 4171–4175 (1993).
This paper was the first comprehensive survey of rates of mutation in RNA viruses.
11. Drake, J. W. & Hwang, C. B. C. On the mutation rate of herpes simplex virus type 1. *Genetics.* **170**, 969–970 (2005).
12. Raney, J. L., Delongchamp, R. R. & Valentine, C. R. Spontaneous mutant frequency and mutation spectrum for gene A of ϕX174 grown in *E. coli*. *Environ. Mol. Mutag.* **44**, 119–127 (2004).
13. Flint, S. J., Enquist, L. W., Racaniello, V. R. & Skalka, A. M. *Principles of Virology Molecular Biology, Pathogenesis and Control of Animal Viruses* (ASM, Washington, 2004).
14. Garcia-Diaz, M. & Bebenek, K. Multiple functions of DNA polymerases. *Crit. Rev. Plant Sci.* **26**, 105–122 (2007).
15. Jung, A. *et al.* Multiply infected spleen cells in HIV patients. *Nature* **418**, 144 (2002).
16. Bonhoeffer, S., Holmes, E. C. & Nowak, M. A. Causes of HIV diversity. *Nature* **376**, 125 (1995).
17. Walsh, C. P. & Xu, G. L. Cytosine methylation and DNA repair. *Curr. Topics Microbiol. Immunol.* **301**, 283–315 (2006).
18. Mangeat, B. *et al.* Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **424**, 99–103 (2003).
19. Caride, E. *et al.* Sexual transmission of HIV-1 isolate showing G–A hypermutation. *J. Clin. Virol.* **23**, 179–189 (2002).
20. Frederico, L. A., Kunkel, T. A. & Shaw, B. R. A sensitive genetic assay for the detection of cytosine deamination — determination of rate constants and the activation energy. *Biochemistry.* **29**, 2532–2537 (1990).
21. Xia, X. & Yuen, K. Y. Differential selection and mutation between dsDNA and ssDNA phages shape the evolution of their genomic AT percentage. *BMC Genet.* **6**, 20 (2005).

22. Pita, J. S., De Miranda, J. R., Schneider, W. L. & Roossinck, M. J. Environment determines fidelity for an RNA virus replicase. *J. Virol.* **81**, 9072–9077 (2007).
This paper describes an experimental test of *in vivo* rates of mutation caused by insertions and deletions. The study comprises a powerful comparison of RdRp fidelity of viruses that infect pepper and tobacco plants.
23. Schrag, S. J., Rota, P. A. & Bellini, W. J. Spontaneous mutation rate of measles virus: direct estimation based on mutations conferring monoclonal antibody resistance. *J. Virol.* **73**, 51–54 (1999).
24. Bull, J. J., Sanjuan, R. & Wilke, C. O. Theory of lethal mutagenesis for viruses. *J. Virol.* **81**, 2930–2939 (2007).
25. Drake, J. W., Bebenek, A., Kissling, G. E. & Peddada, S. Clusters of mutations from transient hypermutability. *Proc. Natl Acad. Sci. USA* **102**, 12849–12854 (2005).
26. Drake, J. W. Mutations in clusters and showers. *Proc. Natl Acad. Sci. USA* **104**, 8203–8204 (2007).
27. Drake, J. W. Too many mutants with multiple mutations. *Crit. Rev. Biochem. Mol. Biol.* **42**, 247–258 (2007).
28. Pathak, V. K. & Temin, H. M. 5-azacytidine and RNA secondary structure increase the retrovirus mutation rate. *J. Virol.* **66**, 3093–3100 (1992).
29. Hudson, R. E., Bergthorsson, U., Roth, J. R. & Ochman, H. Effect of chromosome location on bacterial mutation rates. *Mol. Biol. Evol.* **19**, 85–92 (2002).
30. Chao, L., Rang, C. U. & Wong, L. E. Distribution of spontaneous mutants and inferences about the replication mode of the RNA bacteriophage φ6. *J. Virol.* **76**, 3276–3281 (2002).
31. French, R. & Stenger, D. C. Evolution of wheat streak mosaic virus: dynamics of population growth within plants may explain limited variation. *Annu. Rev. Phytopathol.* **41**, 199–214 (2003).
32. Jenkins, G. M., Rambaut, A., Pybus, O. G. & Holmes, E. C. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *J. Mol. Evol.* **54**, 156–165 (2002).
This paper describes a large-scale comparative analysis of the rates of nucleotide substitution in RNA viruses, with the caveat that this study predates the use of more sophisticated analytical techniques.
33. Hanada, K., Suzuki, Y. & Gojobori, T. A large variation in the rates of synonymous substitution for RNA viruses and its relationship to a diversity of viral infection and transmission modes. *Mol. Biol. Evol.* **21**, 1074–1080 (2004).
34. Chare, E. R. & Holmes, E. C. Selection pressures in the capsid genes of plant RNA viruses reflect mode of transmission. *J. Gen. Virol.* **85**, 3149–3157 (2004).
35. Holmes, E. C. Error thresholds and the constraints to RNA virus evolution. *Trends Microbiol.* **11**, 543–546 (2003).
36. Woelk, C. H. & Holmes, E. C. Reduced positive selection in vector-borne RNA viruses. *Mol. Biol. Evol.* **19**, 2333–2336 (2002).
37. Weaver, S. C. & Barrett, A. D. Transmission cycles, host range, evolution and emergence of arboviral disease. *Nature Rev. Microbiol.* **2**, 789–801 (2004).
38. Berry, I. M. *et al.* Unequal evolutionary rates in the human immunodeficiency virus type 1 (HIV-1) pandemic: the evolutionary rate of HIV-1 slows down when the epidemic rate increases. *J. Virol.* **81**, 10625–10635 (2007).
39. Nielsen, R. & Yang, Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics.* **148**, 929–936 (1998).
40. Li, B. *et al.* Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution. *J. Virol.* **81**, 193–201 (2007).
41. Switzer, W. M. *et al.* Ancient co-speciation of simian foamy viruses and primates. *Nature* **434**, 376–380 (2005).
This study revealed that SFVs co-diverged with their primate hosts over many millions of years, leading to the inference that SFV evolves more slowly than any other RNA virus.
42. Salemi, M. *et al.* Different population dynamics of human T-cell lymphotropic virus type II in intravenous drug users compared with endemically infected tribes. *Proc. Natl Acad. Sci. USA* **96**, 13253–13258 (1999).
This paper is an important example of how different transmission routes (for example, needle-sharing versus mother-to-child transmission) can markedly affect the rate of nucleotide substitution.
43. Vandamme, A. M., Bertazzoni, U. & Salemi, M. Evolutionary strategies of human T-cell lymphotropic virus type II. *Gene* **261**, 171–180 (2000).
44. Lemey, P., Pybus, O. G., Van Dooren, S. & Vandamme, A. M. A Bayesian statistical analysis of human T-cell lymphotropic virus evolutionary rates. *Infect. Genet. Evol.* **5**, 291–298 (2005).
45. Suzuki, Y. *et al.* Slow evolutionary rate of GB virus C/hepatitis G virus. *J. Mol. Evol.* **48**, 383–389 (1999).
46. Plyusnin, A. & Morzunov, S. P. Virus evolution and genetic diversity of hantaviruses and their rodent hosts. *Curr. Topics Microbiol. Immunol.* **256**, 47–75 (2001).
47. Garcīa-Arenal, F., Fraile, A. & Malpica, J. M. Variability and genetic structure of plant virus populations. *Annu. Rev. Phytopathol.* **39**, 157–186 (2001).
48. Charrel, R. N., De Micco, P. & De Lamballerie, X. Phylogenetic analysis of GB viruses A and C: Evidence for cospeciation between virus isolates and their primate hosts. *J. Gen. Virol.* **80**, 2329–2335 (1999).
49. Hughes, A. L. & Friedman, R. Evolutionary diversification of protein-coding genes of hantaviruses. *Mol. Biol. Evol.* **17**, 1558–1568 (2000).
50. Charleston, M. A. & Robertson, D. L. Preferential host switching by primate lentiviruses can account for phylogenetic similarity with the primate phylogeny. *Syst. Biol.* **51**, 528–535 (2002).
51. Rodrīguez-Cerezo, E., Elena, S. F., Moya, A. & Garcīa-Arenal, F. High genetic stability in natural populations of the plant RNA virus tobacco mild green mosaic virus. *J. Mol. Evol.* **32**, 328–332 (1991).
52. Kim, T. *et al.* Molecular analysis of quasispecies of Kyuri green mottle mosaic virus. *Virus Res.* **110**, 161–167 (2005).
53. Marco, C. F. & Aranda, M. A. Genetic diversity of a natural population of Cucurbit yellow stunting disorder virus. *J. Gen. Virol.* **86**, 815–822 (2005).
54. Fraile, A. *et al.* A century of tobamovirus evolution in an Australian population of *Nicotiana glauca*. *J. Virol.* **71**, 8316–8320 (1997).
55. Li, H. Y. & Roossinck, M. J. Genetic bottlenecks reduce population variation in an experimental RNA virus population. *J. Virol.* **78**, 10582–10587 (2004).
56. Fargette, D. *et al.* Rice yellow mottle virus, an RNA plant virus, evolves as rapidly as most RNA animal viruses. *J. Virol.* 16 Jan 2008 (doi:10.1128/JVI.02506-07).
57. McGeoch, D. J. & Gatherer, D. Integrating reptilian herpesviruses into the family *Herpesviridae*. *J. Virol.* **79**, 725–731 (2005).
58. Bernard, H. U. Coevolution of papillomaviruses with human populations. *Trends Microbiol.* **2**, 140–143 (1994).
59. Pavesi, A. Utility of JC polyomavirus in tracing the pattern of human migrations dating to prehistoric times. *J. Gen. Virol.* **86**, 1315–1326 (2005).
60. Hatwell, J. N. & Sharp, P. M. Evolution of human polyomavirus JC. *J. Gen. Virol.* **81**, 1191–1200 (2000).
61. Shackelton, L. A. & Holmes, E. C. Phylogenetic evidence for the rapid evolution of human B19 erythrovirus. *J. Virol.* **80**, 3666–3669 (2006).
62. Ge, L. M., Zhang, J. T., Zhou, X. P. & Li, H. Y. Genetic structure and population variability of tomato yellow leaf curl China virus. *J. Virol.* **81**, 5902–5907 (2007).
63. Lopez-Bueno, A., Villarreal, L. P. & Almendral, J. M. Parvovirus variation for disease: a difference with RNA viruses? *Curr. Topics Microbiol. Immunol.* **299**, 349–370 (2006).
64. Isnard, M., Granier, M., Frutos, R., Reynaud, B. & Peterschmitt, M. Quasispecies nature of three maize streak virus isolates obtained through different modes of selection from a population used to assess response to infection of maize cultivars. *J. Gen. Virol.* **79**, 3091–3099 (1998).
65. Drake, J. W. A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl Acad. Sci. USA* **88**, 7160–7164 (1991).
This study was the first comprehensive survey of rates of mutation among DNA viruses. It predicts that smaller DNA viruses should have higher mutation rates per nucleotide.
66. Harrison, B. D. & Robinson, D. J. Natural genomic and antigenic variation in whitefly-transmitted geminiviruses (begomoviruses). *Annu. Rev. Phytopathol.* **37**, 369–398 (1999).
67. Hino, S. & Miyata, H. Torque teno virus (TTV): current status. *Rev. Med. Virol.* **17**, 45–57 (2007).
68. Shackelton, L. A., Parrish, C. R., Truyen, U. & Holmes, E. C. High rate of viral evolution associated with the emergence of carnivore parvovirus. *Proc. Natl Acad. Sci. USA* **102**, 379–384 (2005).
69. Umemura, T., Tanaka, Y., Kiyosawa, K., Alter, H. J. & Shih, J. W.-K. Observation of positive selection within hypervariable regions of a newly identified DNA virus (SEN virus). *FEBS Lett.* **510**, 171–174 (2002).
70. Duffy, S. & Holmes, E. C. Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus Tomato yellow leaf curl virus (TYLCV). *J. Virol.* **82**, 957–965 (2008).
71. Eyre-Walker, A. & Keightley, P. D. The distribution of fitness effects of new mutations. *Nature Rev. Genet.* **8**, 610–618 (2007).
72. Sanjuan, R., Moya, A. & Elena, S. F. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc. Natl Acad. Sci. USA* **101**, 8396–8401 (2004).
73. Sasaki, A. & Nowak, M. A. Mutation landscapes. *J. Theor. Biol.* **224**, 241–247 (2003).
74. Kamp, C., Wilke, C. O., Adami, C. & Bornholdt, S. Viral evolution under the pressure of an adaptive immune system: optimal mutation rates for viral escape. *Complexity.* **8**, 28–33 (2003).
75. Drake, J. W. Spontaneous mutation. *Annu. Rev. Genet.* **25**, 125–146 (1991).
76. Anderson, J. P., Daifuku, R. & Loeb, L. A. Viral error catastrophe by mutagenic nucleosides. *Annu. Rev. Microbiol.* **58**, 183–205 (2004).
77. Mansky, L. M. & Cunningham, K. S. Virus mutators and antimutators: roles in evolution, pathogenesis and emergence. *Trends Genet.* **16**, 512–517 (2000).
78. Pfeiffer, J. K. & Kirkegaard, K. A single mutation in poliovirus RNA-dependent RNA polymerase confers resistance to mutagenic nucleotide analogs via increased fidelity. *Proc. Natl Acad. Sci. USA* **100**, 7289–7294 (2003).
79. Sniegowski, P. D., Gerrish, P. J., Johnson, T. & Shaver, A. The evolution of mutation rates: separating causes from consequences. *Bioessays.* **22**, 1057–1066 (2000).
80. Vignuzzi, M., Stone, J. K. & Andino, R. Ribavirin and lethal mutagenesis of poliovirus: molecular mechanisms, resistance and biological implications. *Virus Res.* **107**, 173–181 (2005).
81. Furio, V., Moya, A. & Sanjuan, R. The cost of replication fidelity in an RNA virus. *Proc. Natl Acad. Sci. USA* **102**, 10233–10237 (2005).
82. Graci, J. D. *et al.* Lethal mutagenesis of poliovirus mediated by a mutagenic pyrimidine analogue. *J. Virol.* **81**, 11256–11266 (2007).
83. Elena, S. F., Carrasco, P., Daros, J. A. & Sanjuan, R. Mechanisms of genetic robustness in RNA viruses. *EMBO Rep.* **7**, 168–173 (2006).
This paper is a useful discussion of how genetic robustness is defined, measured and achieved in RNA viruses.
84. Ewald, P. W. Evolution of mutation rate and virulence among human retroviruses. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **346**, 333–341 (1994).
85. Crotty, S., Cameron, C. E. & Andino, R. RNA virus error catastrophe: direct molecular test by using ribavirin. *Proc. Natl Acad. Sci. USA* **98**, 6895–6900 (2001).
This study was an important test of the hypothesis that increasing the mutation rate of an RNA virus (with the application of an antiviral drug) can result in viral extinction, probably owing to error catastrophe and/or lethal mutagenesis.
86. Mansky, L. M. & Bernard, L. C. 3′-azido-3′-deoxythymidine (AZT) and AZT-resistant reverse transcriptase can increase the *in vivo* mutation rate of human immunodeficiency virus type 1. *J. Virol.* **74**, 9532–9539 (2000).
87. Summers, J. & Litwin, S. Examining the theory of error catastrophe. *J. Virol.* **80**, 20–26 (2006).
88. Eigen, M. *Steps Towards Life* (Oxford University Press, Oxford, 1996).
89. Codoner, F. M., Daros, J. A., Sole, R. V. & Elena, S. F. The fittest versus the flattest: experimental confirmation of the quasispecies effect with subviral pathogens. *PLoS Pathog.* **2**, e136 (2006).

90. Elena, S. F. & Sanjuán, R. Adaptive value of high mutation rates of RNA viruses: separating causes from consequences. *J. Virol.* **79**, 11555–11558 (2005).
91. Kuiken, T. *et al.* Host species barriers to influenza virus infections. *Science* **312**, 394–397 (2006).
92. Kunkel, T. A. DNA replication fidelity. *J. Biol. Chem.* **279**, 16895–16898 (2004).
93. Arguello-Astorga, G. R., Ascencia-Ibáñez, J. T., Dallas, M. B., Orozco, B. M. & Hanley-Bowdoin, L. High-frequency reversion of geminivirus replication protein mutants during infection. *J. Virol.* **81**, 11005–11015 (2007).
94. Gutierrez, C. Geminivirus DNA replication. *Cell. Mol. Life Sci.* **56**, 313–329 (1999).
95. Chen, Y. *et al.* Analysis of 15 novel full-length BK virus sequences from three individuals: Evidence of a high intra-strain genetic diversity. *J. Gen. Virol.* **85**, 2651–2663 (2004).
96. Shackelton, L. A., Rambaut, A., Pybus, O. G. & Holmes, E. C. JC virus evolution and its association with human populations. *J. Virol.* **80**, 9928–9933 (2006).
97. Ghedin, E. *et al.* Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature* **437**, 1162–1166 (2005).
98. Edwards, R. A. & Rohwer, F. Viral metagenomics. *Nature Rev. Microbiol.* **3**, 504–510 (2005).
99. Lukashov, V. V. & Goudsmit, J. Recent evolutionary history of human immunodeficiency virus type 1 subtype B: reconstruction of epidemic onset based on sequence distances to the common ancestor. *J. Mol. Evol.* **54**, 680–691 (2002).
100. Rambaut, A. Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. *Bioinformatics.* **16**, 395–399 (2000).
101. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
102. Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, 699–710 (2006).
This is an important methodological paper showing how rates of nucleotide substitution and times to common ancestry can be estimated from serially sampled data sets such as those from RNA viruses and ssDNA viruses.
103. Malpica, J. M. *et al.* The rate and character of spontaneous mutation in an RNA virus. *Genetics.* **162**, 1505–1511 (2002).
104. Aulicino, P. C., Holmes, E. C., Rocco, C., Mangano, A. & Sen, L. Extremely rapid spread of human immunodeficiency virus type 1 BF recombinants in Argentina. *J. Virol.* **81**, 427–429 (2007).
105. Davis, P. L. *et al.* Phylogeography, population dynamics, and molecular evolution of European bat lyssaviruses. *J. Virol.* **79**, 10487–10497 (2005).

Acknowledgements
This work was supported by a National Science Foundation grant, DBI 06-03070, to S.D. We thank A. Rambaut for his comments.

DATABASES

Entrez Genome: http://www.ncbi.nlm.nih.gov/sites/entrez?db=genome
bacteriophage RB69 | canine parvovirus | GB virus C | hepatitis B virus | herpes simplex virus type 1 | HTLV-I | human immunodeficiency virus type-1 | human T-cell lymphotropic virus type II | JC polyomavirus | parvovirus B19 | phage ϕX174 | RNA phage Qβ | simian foamy virus | Tomato yellow leaf curl virus | vesicular stomatitis virus

FURTHER INFORMATION

Edward Holmes homepage: http://www.cidd.psu.edu/people/bio_holmes.html
BEAST software: http://beast.bio.ed.ac.uk

ALL LINKS ARE ACTIVE IN THE ONLINE PDF