Evolutionary Virology at 40

Jemma L. Geoghegan* and Edward C. Holmes^{†,*,§,*,1}

*Department of Biological Sciences, Macquarie University, Sydney, New South Wales 2109, Australia and [†]Marie Bashir Institute for Infectious Diseases and Biosecurity, [‡]Charles Perkins Centre, [§]School of Life and Environmental Sciences, and ^{**}Sydney Medical School, The University of Sydney, New South Wales 2006, Australia

ORCID IDs: 0000-0003-0970-0153 (J.L.G.); 0000-0001-9596-3552 (E.C.H.)

ABSTRACT RNA viruses are diverse, abundant, and rapidly evolving. Genetic data have been generated from virus populations since the late 1970s and used to understand their evolution, emergence, and spread, culminating in the generation and analysis of many thousands of viral genome sequences. Despite this wealth of data, evolutionary genetics has played a surprisingly small role in our understanding of virus evolution. Instead, studies of RNA virus evolution have been dominated by two very different perspectives, the experimental and the comparative, that have largely been conducted independently and sometimes antagonistically. Here, we review the insights that these two approaches have provided over the last 40 years. We show that experimental approaches using *in vitro* and *in vivo* laboratory models are largely focused on short-term intrahost evolutionary mechanisms, and may not always be relevant to natural systems. In contrast, the comparative approach relies on the phylogenetic analysis of natural virus populations, usually considering data collected over multiple cycles of virus–host transmission, but is divorced from the causative evolutionary processes. To truly understand RNA virus evolution it is necessary to meld experimental and comparative approaches within a single evolutionary genetic framework, and to link viral evolution at the intrahost scale with that which occurs over both epidemiological and geological timescales. We suggest that the impetus for this new synthesis may come from methodological advances in next-generation sequencing and metagenomics.

KEYWORDS virus; evolution; phylodynamics; phylogeny; metagenomics; quasispecies

Introduction: Life at 40

THE year 2018 marks the 40th anniversary of the first published studies on the evolution of viruses. The field of evolutionary virology was inaugurated with two key papers that shaped the way virus evolution was studied in subsequent decades. The first was an experimental study by Domingo and colleagues that showed that individual populations of RNA viruses carried abundant genetic variation (Domingo *et al.* 1978). The second, by Palese and co-workers, considered variants of human influenza virus sampled from different patients to reveal the nature of genetic differences between RNA viruses at the interhost, epidemiological scale (Nakajima *et al.* 1978; and later Young *et al.* 1979). These studies shared a similar theme, understanding the extent of genetic variation within and between RNA virus populations, both utilized oligonucleotide fingerprinting, and both

highlighted that RNA viruses have an innate capacity to evolve rapidly. However, they initiated two very different avenues of investigation that have effectively run in parallel ever since (Figure 1).

The paper by Domingo *et al.* (1978) marks the beginning of experimental studies of RNA virus evolution, in which evolutionary processes in the short-term are analyzed by either *in vitro* or *in vivo* laboratory infections. Arguably the defining theme of this field is the idea that the exceptionally high mutation rate in RNA viruses means that they evolve according to a form of group selection known as the “quasispecies” (Domingo *et al.* 1978, 2012; Andino and Domingo 2015) (Box 1). Indeed, the quasispecies concept has become so widely adopted that it is often cited whenever genetic variation is encountered in a viral population, and has even been used in nonviral systems (Kuipers *et al.* 2000; Webb and Blaser 2002; Tannenbaum and Fontanari 2008). In contrast, the study by Palese and colleagues, with later work by Walter Fitch (Buonagurio *et al.* 1986; Yamashita *et al.* 1988; Fitch *et al.* 1991), pioneered comparative studies of RNA virus populations that involves the analysis of gene sequences (or

Figure 1 Approaches to studying RNA virus evolution. The Venn diagram illustrates the two historical, and largely parallel, strands of research in virus evolution—the experimental and the comparative—that arose in the late 1970s. They generally only overlap in the study of a limited number of interhost virus transmission events that often involve a substantial population bottleneck. Through the use of *in vitro* or *in vivo* model systems, experimental studies largely focus on evolution in the short-term, particularly that which occurs within individual hosts. In contrast, comparative approaches deal with interhost, epidemiological-scale dynamics that entail multiple rounds of interhost transmission and are usually based on phylogenetic analyses. We suggest that a new evolutionary genetics approach is required to bridge this divide.

other genetic markers) sampled from different individuals in a population. From this arose the modern science of molecular epidemiology, in which phylogenetic analysis is used to reveal evolutionary relationships among virus sequences sampled from different individuals, often during disease outbreaks, in turn leading to inferences on the underlying patterns and processes of virus evolution (Holmes 2009; Moratorio and Vignuzzi 2018).

An unfortunate by-product of this siloed approach has been the coexistence of two views of RNA virus evolution that are often more antagonistic than complementary. We believe that these differing world views are, in part, a reflection of their contrasting methodological perspectives. With the ability of next-generation sequencing and metagenomics to rapidly generate vast amounts of gene sequence data, from within individual hosts to global populations (Firth and Lipkin 2013; Willner and Hugenholtz 2013; Zhang *et al.* 2018), we suggest that the time is right to bring the experimental and the comparative approaches together. Herein, we set out a framework for this new synthesis, outlining some of the key outcomes of the last 40 years of virus evolution research, noting areas of agreement and continuing contention, and establishing a potential road map for future research.

Studying RNA Virus Evolution

As well as being major agents of infectious disease, RNA viruses are important model “organisms” capable of advancing our understanding of the evolutionary process (Holmes 2009). In particular, RNA virus evolution is characterized by

the generation and fixation of mutations over time periods amenable to direct human observation, in contrast to most evolutionary changes that occur in higher organisms. Hence, RNA viruses provide a useful natural laboratory to visualize evolutionary processes in real time, including during single-disease outbreaks (Gire *et al.* 2014). The utility of RNA viruses in experimental assays is enhanced by their small genomes, in which mutations often result in major phenotypic effects (Moya *et al.* 2000). It should therefore come as no surprise that RNA viruses have been used to test a variety of evolutionary theories (Turner and Chao 1999) and are powerful exemplars in the development of new methods of bioinformatic analysis (Lemey *et al.* 2009; Kühnert *et al.* 2014; To *et al.* 2016). Although there is also a large amount of literature on the evolution of DNA viruses, their usually lower rates of evolutionary change (Duffy *et al.* 2008) means that they are generally less suited for use as model systems and they will not be considered here.

To achieve a holistic understanding of RNA virus evolution it is important to bridge the divide between studies based on experimental approaches and those that utilize comparative, and usually phylogenetic, methods (Figure 1). Experimental approaches are strongly focused toward studying evolutionary change at the intrahost scale, which only represents a tiny, albeit hugely important, component of the overall evolutionary process. They also risk establishing inaccurate general rules for RNA virus evolution if they are founded on the analysis of a limited number of case studies. For example, while poliovirus has been one of the mainstays of experimental approaches to studying viral evolution [for example, Vignuzzi *et al.* (2006) and Stern *et al.* (2017)] and has provided a wealth of valuable biological data (Regoes *et al.* 2005), the evolution of poliovirus in the laboratory may not always reflect that in nature and it is mistaken to think that it is representative of all viruses. RNA viruses vary widely, having markedly different genome structures and replication cycles, infecting different hosts, possessing different propensities for disease, and experiencing variable rates of mutation and recombination.

There is a similar danger in generalizing results from experimental systems that do not reflect the natural host range of the virus in question. For example, the textbook example of the evolution of pathogen virulence involves the release of myxomavirus (MYXV; a double-stranded DNA virus) as a biological control against European rabbits in Australia (Kerr *et al.* 2012). Experimental approaches using cell culture have been used in determining which mutations in the MYXV genome might be responsible for the profound changes in virulence that have occurred in this virus since its release in 1950 (Mossman *et al.* 1996; Peng *et al.* 2016). However, these virulence determinants have often not been upheld when tested using reverse-genetic studies in laboratory-bred rabbits of the same species as infected in nature (Liu *et al.* 2017).

Drawbacks are also apparent in phylogenetic analyses that make use of viruses sampled from natural populations and are

the hallmark of comparative studies of virus evolution. Because observed phylogenetic patterns are the outcome of a variety of interacting evolutionary processes (mutation, genetic drift, natural selection, population growth and decline, and phylogeography) that occur at differing intensities and over different timescales, and are usually inferred between interhost comparisons performed many generations after they have occurred, it is inherently difficult to determine exactly which of these processes shape the phylogenetic patterns observed. Phylogenetic analyses are also limited by the availability of samples to inform on evolutionary patterns and processes, and are strongly impacted by sampling biases. As a consequence, phylogenetic analysis may sometimes be better used as a means to generate hypotheses that can then be tested experimentally, such as guiding the detection of virulence determinants in oral vaccine strains of poliovirus (Stern *et al.* 2017), rather than as a precision tool to reveal the history of actual evolutionary events.

An Evolutionary World Shaped by Mutation

The studies of Domingo *et al.* and Palese *et al.* both attempted to discern patterns in the genetic variation generated by frequent mutation in RNA viruses. However, they differ in the timescale over which the diversity considered is generated, and the way it is measured and visualized. Work over the last 40 years has established that the remarkable rapidity with which RNA viruses mutate is perhaps their defining characteristic. Such high mutation rates reflect erroneous genome replication in the absence of any error correction, with only sporadic instances of RNA repair in contrast with what is seen in double-stranded DNA-based organisms (Drake 1993; Drake *et al.* 1998; Bellacosa and Moss 2003). Across RNA viruses as a whole, estimated mutation rates fall within a range of 10^{-4} – 10^{-6} mutations per site per cell replication (Sanjuán *et al.* 2010; Sanjuán 2012; Peck and Lauring 2018), between different infected cells in the same culture or individual host (Combe *et al.* 2015). Evolutionary rates (that is, the number of fixed substitutions per unit time) range from $\sim 10^{-2}$ to 10^{-5} nucleotide substitutions per site per year (Duffy *et al.* 2008; Sanjuán 2012; Holmes *et al.* 2016), and hence are several orders of magnitude greater than those observed in double-stranded DNA organisms (Duffy *et al.* 2008; Sanjuán 2012). Despite the increasing accuracy of measures of mutation rate (Acevedo *et al.* 2014), truly slowly evolving RNA viruses, with rates of mutation/evolution that approach those of eukaryotes and bacteria, have yet to be identified.

High rates of background mutation have obvious consequences for virus evolution, quickly providing the raw material needed for adaptation to changing environments, including new hosts, immune responses, and antivirals. It is therefore no surprise that RNA viruses comprise the most important class of emerging viruses (Cleaveland *et al.* 2001). More difficult to determine are the selective forces that have shaped the evolution of mutation rates in RNA viruses

(Regoes *et al.* 2013; Peck and Lauring 2018). One suggestion is that the genetic diversity produced by frequent mutation is in itself selectively advantageous and may directly contribute to such features as viral pathogenesis (Vignuzzi *et al.* 2006). For example, the appearance of neurovirulent poliovirus infection in a mouse model system was associated with higher levels of virus genetic diversity (Vignuzzi *et al.* 2006). A contrary view, which recently received strong support from another experimental study involving poliovirus, is that the evolution of mutation rates in fact reflects an evolutionary trade-off between replication speed and fidelity; that is, rapid replication is selectively advantageous for a virus but comes at the cost of lower replication fidelity (Fitzsimmons *et al.* 2018).

Although RNA virus mutation rates are high, the majority of the mutations produced by faulty genomic replication are deleterious, and their removal from populations by purifying selection is perhaps the dominant process in viral evolution (Elena and Moya 1999). For example, deep sequencing studies of intrahost virus genetic diversity have revealed that most mutation variants present within a single host are present at low frequency, are short-lived, and are usually found only at a single sampling time point, suggesting that they represent transient deleterious mutations (Holmes 2009; McCrone *et al.* 2018). Similarly, experimental studies comparing the fitness of individual mutations against the wild-type have shown that deleterious mutations are commonplace (Sanjuán *et al.* 2004; Acevedo *et al.* 2014). It is possible that the very large intrahost population sizes of RNA viruses, which can be in the order of 10^{10} virions at any single time point (Piatak *et al.* 1993), mean that sufficient viable viral progeny are produced each generation to ensure evolutionary survival, so that RNA viruses experience a form of “population robustness” against the impact of deleterious mutations (Elena *et al.* 2006).

An important consequence of this process of gradual selective purging of low-fitness mutations is that evolutionary rates in RNA viruses are strongly “time-dependent” (Duchêne *et al.* 2014). That is, the highest inferred evolutionary rates are observed in comparisons involving closely related sequences (*i.e.*, within individual patients or from outbreaks), while lower rates are estimated from comparisons utilizing more divergent sequences. This pattern appears because short-term (*i.e.*, recent) evolutionary rates are inflated by the presence of transient deleterious mutations yet to be removed by purifying selection, while multiple substitutions at single sites mean that long-term rates from divergent taxa likely underestimate the true number of nucleotide substitutions (Duchêne *et al.* 2014). As well as providing insights into the nature of purifying selection (Wertheim and Kosakovsky Pond 2011), the time-dependent nature of virus evolution has important implications for the accuracy of the molecular clock dating of RNA viruses; for example, the inclusion of multiple sequences sampled within single disease outbreaks (short-term) may result in an underestimation of times to common ancestry of specific viruses as a whole (long-term) (Duchêne *et al.* 2014; Aiewsakun and Katzourakis 2016).

Mutation and the Quasispecies

As noted at the outset, arguably the most important idea in RNA virus evolution is that they form quasispecies (Andino and Domingo 2015). The concept of the quasispecies was originally developed by Eigen (1971), and was first applied to RNA viruses in earnest by Domingo and colleagues (Domingo *et al.* 1978). Since this time, it has been both popular and highly controversial (Domingo 2002; Holmes and Moya 2002). The quasispecies considers evolutionary behavior in RNA systems characterized by very high mutation rates. The core idea is that the evolutionary fate of an individual virus variant depends on both its own fitness *and* that of other variants in the population to which it is linked by mutation, and that natural selection acts on the population as a whole, maximizing average population fitness (Figure 2). A more detailed description of the quasispecies theory is provided in Box 1.

The idea that RNA viruses form quasispecies has almost become the default position in studies of viral evolution (Domingo *et al.* 2012). However, the term is often incorrectly applied as a simple surrogate for genetic diversity (Holmes 2009), quasispecies theory only applies to intrahost virus evolution, and there have been relatively few rigorous tests of whether RNA viruses constitute quasispecies as correctly defined (Sanjuán *et al.* 2007). The most commonly cited evidence for the existence of quasispecies is that populations of RNA viruses are genetically diverse (Eigen 1996; Llauro and Andino 2010), although this is an obvious outcome for any system characterized by frequent mutation. More compelling evidence for quasispecies behavior is that natural selection acts on populations of RNA viruses as a whole. While experimental studies have shown that viral populations can experience the form of group selection implied in quasispecies theory (Burch and Chao 2000; Bordería *et al.* 2015), particularly under artificially elevated mutation rates (Codoñer *et al.* 2006; Sanjuán *et al.* 2007), there is currently little evidence that this applies to viruses outside of the laboratory and hence uncertainty as to whether it is relevant for RNA viruses in nature. Indeed, the emerging picture from comparative analyses, especially the deep sequencing of natural populations of RNA viruses, is that they are often characterized by a dominant variant, presumably the fittest, together with an abundance of low-frequency variants, many of which are likely to represent transient deleterious mutations (Pybus *et al.* 2007; Holmes 2009; McCrone *et al.* 2018). Although natural selection undoubtedly operates at the intrahost scale, there is little definitive evidence for quasispecies dynamics, although it is possible that these are apparent at selection coefficients too low to easily measure. For example, the deep sequencing of intra- and interhost diversity in dengue virus provided strong evidence for host adaptation, with the same virus mutations appearing independently across multiple patients, seemingly because of similar immune pressures (Parameswaran *et al.* 2017). However, there was no evidence that mutational neighborhood impacted fitness and hence no

Figure 2 “Darwinian” vs. quasispecies models of RNA virus evolution. In the Darwinian virus population, natural selection favors the variant with the highest individual fitness (circle shown in red), with lower-fitness variants (blue, green, and yellow) produced by mutation at a relatively low rate. Under the quasispecies model, very high mutation rates lead to a mutational coupling among variants (of different colors). This, in turn, means that the viral population evolves as a single unit, with the mutational landscape greatly impacting virus evolution and natural selection acting on the population as a whole, maximizing mean fitness. In the top part of the figure, the circle sizes represent relative fitness values, whereas they are drawn to equivalent sizes in the bottom part of the figure for ease of visualization. See also Box 1.

evidence for quasispecies dynamics. In other cases, such as influenza virus, adaptive evolution appears to be of limited importance within hosts as stochastic processes, including genetic drift and large-scale population bottlenecks, play a more important role (McCrone *et al.* 2018), again in contrast to quasispecies models.

Often linked to quasispecies theory is the idea that viral populations can “cooperate” in a manner that enhances fitness (Vignuzzi *et al.* 2006; Ciota *et al.* 2012; Shirogane *et al.* 2012; Bordería *et al.* 2015; Díaz-Muñoz *et al.* 2017; Sanjuán 2017). For example, human H3N2 influenza A virus carries two different amino acid variants at a specific site in the neuraminidase protein that together increase fitness in cell culture compared to when these amino acids occur singly (Xue *et al.* 2016). While there was evidence for these evolutionary interactions in cell culture, no such evidence was apparent in analyses of natural populations as these two mutations very rarely cooccur in human clinical samples (Xue *et al.* 2018). This likely reflects the impact of major virus population bottlenecks both within and between hosts. Indeed, while there is some evidence from experimental systems that multiple viral variants can be transmitted between cells that could lead to cooperation-like interactions (Combe *et al.* 2015), experimental populations may often fail to mirror the natural situation. Most pointedly, it is uncertain how cooperation could be selectively maintained in the face of the severe population bottlenecks, particularly those that commonly occur when viruses transmit to new hosts (Geoghegan *et al.* 2016b; McCrone and Llauro 2018; McCrone *et al.* 2018). Transmission bottlenecks inevitably impinge on evolutionary processes that require groups of viruses to interact

Box 1 The Quasispecies

Quasispecies theory was developed by Manfred Eigen as a model of self-replicating macromolecules theoretically equivalent to those that characterized life's early evolution (Eigen 1971; Eigen and Schuster 1977). Mathematically, it has been defined as the “distribution of mutants that belong to the maximum eigenvalue of the system” (Eigen 1996). The quasispecies concept was first applied to RNA viruses by Esteban Domingo in the late 1970s, following the observation of genetic variation in the bacteriophage Q β (Domingo *et al.* 1978).

In simple terms, the quasispecies is a form of mutation–selection balance in which a distribution of variant viral genomes is ordered around the fittest, or “master,” sequence. Central to quasispecies theory is that mutation rates in RNA viruses are so high that the frequency of any variant is not only a function of its own replication rate (fitness), but also the probability that it is produced by mutation from other variants in the population that are linked to it in sequence space. This “mutational coupling” leads to a distribution of evolutionarily interlinked viral genomes, which in turn means that the entire mutant distribution behaves as a single unit, with natural selection acting on the mutant distribution as a whole rather than on individual variants (Figure 2). The quasispecies as a whole therefore evolves to maximize its *average fitness*, rather than that of individual variants.

One of the most interesting aspects of quasispecies is that variants with low individual fitness can reach a high frequency if they have mutational links to variants with higher fitness (Wilke 2005). In addition, the most common genotype is not necessarily the fittest within the quasispecies and the “wild-type” may only comprise a small proportion of the total population. Most notably, under particular mutant distributions, low-fitness variants can in theory out-compete those of higher fitness if they are surrounded by beneficial mutational neighbors. This has been termed the “survival of the flattest” (Wilke *et al.* 2001), although it is more correctly thought of as increased mutational robustness.

An important laboratory demonstration of quasispecies-like evolution was the observation that “evolvability” in the RNA bacteriophage $\phi 6$ *in vitro* was dependent on its mutational spectrum (Burch and Chao 2000). In particular, a high-fitness clone evolved to *lower* mean fitness because its mutational neighbors were of low fitness. However, as discussed in the main text, comparative studies of natural populations of RNA viruses have generally provided far less evidence for quasispecies behavior.

Although it has been claimed that quasispecies theory is qualitatively different from “classical” population genetic models (Eigen 1992), quasispecies dynamics can be framed within the mainstream of evolutionary theory, as a form of mutation–selection balance in a genetic system characterized by very high mutation rates, although its intellectual history is different. While quasispecies theory has been instrumental in introducing evolutionary ideas into virology and can shed new light on evolutionary dynamics when mutation rates are extremely high, it is still debatable whether it applies to RNA viruses in nature.

(Aaskov *et al.* 2006) and make it difficult to translate within-host evolution to that over epidemiological timescales (Figure 1). More generally, the quasispecies considers the joint effects of mutation and selective competition, and says nothing about cooperation *per se*, which is often poorly defined and described at a mechanistic level.

RNA Virus Phylogenies and Molecular Epidemiology

Phylogenetic studies of RNA virus evolution have come a long way since the late 1970s, and the science of molecular epidemiology has arguably been the most successful way in which evolutionary ideas have permeated into virology (Holmes 2009). With a sufficient sample of sequences, it is possible to reveal the origins, spread, and evolution of a diverse array of viruses, and phylogenetic studies are especially important whenever a novel virus emerges.

The speed at which viral diversity is created and genomic-scale phylogenetic analysis can be performed makes the latter a key tool in the response to outbreaks of infectious disease, as

demonstrated in the recent epidemics of Middle East respiratory syndrome coronavirus (MERS-CoV) (Dudas *et al.* 2018), Ebola (Dudas *et al.* 2017), Zika (Faria *et al.* 2017), and various forms of influenza virus (Bedford *et al.* 2014; Neher and Bedford 2015; Cui *et al.* 2016). More broadly, today's phylogenetic approaches can help reveal the patterns, processes, and rates of cross-species transmission (*i.e.*, host jumping) in viruses, as well as its determinants (Geoghegan *et al.* 2016a, 2017). Although the success of phylogenetics in virology in part stems from the rapidity of virus evolution, this also means that sequence similarity is quickly eroded in viral genomes and proteomes, greatly inhibiting studies of their origin and early evolution. The development of methods that accurately infer phylogenetic history from highly divergent virus sequences, perhaps utilizing elements of protein structure (Bamford *et al.* 2005), is clearly a research priority, although to date there has been relatively little movement in this space.

Although by far the most common use of phylogenies in virology is to simply infer the evolutionary relationships among gene sequences, should the data fit some form of

Figure 3 The different scales on which studies of RNA virus evolution can proceed from a comparative perspective. These scales range from the study of short-term intrahost evolution, through analysis of the initial host contact network within an infected population, and finally out to the meta-population scale, representing long-term virus evolution as often depicted in the fields of molecular epidemiology and phylogeography. At each scale, a variety of phylogenetic and phylodynamic inferences can be made. The R_0 estimate of HIV in the UK comes from Stadler *et al.* (2013).

molecular clock (Drummond *et al.* 2006), they can also be used to provide estimates of evolutionary rates and the time-scale over which viral evolution has occurred (Figure 3). If sampling is sufficiently dense and unbiased, clock-based phylogenetic methods also allow a range of epidemiological parameters to be estimated from genomic data, including the basic reproductive number, R_0 (the number of secondary infections caused by a single host in an entirely susceptible population), that is the cornerstone of mathematical epidemiology (Stadler *et al.* 2012, 2014; Boskova *et al.* 2014). These methods, combined with a new wealth of genome sequence data, have led to a blossoming of the field of “phylodynamics,” which attempts to marry phylogenetic studies of virus gene sequence data with epidemiological studies based on case (*i.e.*, incidence) data (Grenfell *et al.* 2004; Holmes and Grenfell 2009; Volz *et al.* 2013; Volz and Frost 2013).

Although the phylodynamic framework is usually applied at the epidemiological scale, it is possible, although complex, to link patterns of genetic variation observed at the intrahost scale to virus epidemics as a whole (Pybus and Rambaut 2009). This is of particular value when trying to infer chains of transmission (*i.e.*, who-infected-whom) during outbreaks and using this information to help manage disease control, for example by identifying the cause of outbreak “flare-ups” (Mate *et al.* 2015). Because virus transmission often occurs more rapidly than the speed with which mutations are fixed in virus populations, individuals from a transmission chain may harbor largely identical consensus sequences. In these cases, low-frequency variants (*i.e.*, variants present at lower frequency than the consensus sequence), may be central in establishing the links between patients if they survive the population bottleneck that routinely occurs when viruses transmit to new hosts (Stack *et al.* 2013; Hasing *et al.* 2016).

The related science of virus phylogeography has similarly made huge strides in recent years, such that with sufficient data the rates, patterns, and determinants of virus spatial spread can now be inferred easily and accurately (Figure 3)

(Lemey *et al.* 2009; Pybus *et al.* 2015). However, for both phylogeography and phylodynamics, it is critically important to consider the possible impact of sampling biases, especially as “convenience” sampling is rife. Although there have been important advances in this area using approaches like the structured coalescent to dampen the effect of sampling biases (Rasmussen *et al.* 2014; De Maio *et al.* 2015; Dudas *et al.* 2018), it is necessarily still the case that phylogenies can only link the geographic locations from which virus sequences have been sampled, which may not necessarily reflect the exact migration pathways of the virus. Detailed structured sampling would be an important means to overcome these biases, and there have been improvements in this area during recent disease outbreaks (Dudas *et al.* 2017).

One of the most useful recent applications of phylogenetics has been to help infer aspects of phenotypic evolution in viruses. At its most basic level, this involves using phylogenies as a scaffold on which to map traits like virulence and host range that are central to understanding disease emergence (Diehl *et al.* 2016; Stern *et al.* 2017). The location of key phenotypic mutations, such as virulence determinants, on phylogenetic trees provides insights into the evolutionary processes that led to their appearance. For example, mutations that fall at deeper nodes are more likely to be selectively advantageous, such as the A82V mutation in the glycoprotein of Ebola virus that seemingly increases replication in human cells (Diehl *et al.* 2016; Urbanowicz *et al.* 2016). In other cases, it is possible to directly combine phenotypic and the phylogenetic data. An important case in point is the melding of phylogenetics and antigenics to understand the process of seasonal antigenic drift in influenza A virus, which necessitates regularly updated vaccines (Bedford *et al.* 2014).

The Evolution of Recombination in RNA Viruses

One area in which experimental and comparative approaches have reached generally convergent viewpoints over the last

40 years is the frequency with which recombination occurs in RNA viruses (Holmes 2009). However, there is still considerable uncertainty over why recombination rates vary so much between viruses and hence the overall role played by recombination in RNA virus evolution (Simon-Loriere and Holmes 2011).

Some experimental studies have suggested that recombination is essential to virus fitness, allowing new and advantageous genomic configurations to be generated (Xiao *et al.* 2016). Although there is no doubt that recombination may create beneficial genotypic configurations, it is not necessarily the case that it evolved for this reason. Indeed, inferred recombination frequencies are highly variable: from cases like human immunodeficiency virus (HIV) where the recombination rate per base exceeds that of mutation (Shriner *et al.* 2004; Neher and Leitner 2010), or in influenza in which reassortment appears to be an almost an obligatory part of the replication (Lowen 2017), to viruses in which recombination rates are far, far lower and perhaps absent altogether. The most striking examples of the latter are those viruses with single-strand negative-sense genomes arranged as a single RNA molecule (*i.e.*, from the viral order Mononegavirales), within which only sporadic cases of recombination have been reported (Archer and Rico-Hesse 2002; Chare *et al.* 2003). Yet, although an effective lack of recombination may seem to be an important evolutionary constraint, this class of RNA viruses is clearly highly successful, being both abundant and able to infect multiple hosts.

Why, then, do RNA viruses exhibit such highly variable recombination rates? Although the evolution of RNA virus recombination has been treated in the same manner as the evolution of sex (Michod *et al.* 2008), a simpler explanation is that recombination reflects the evolution of strategies to better control gene expression in RNA viruses (Simon-Loriere and Holmes 2011). In particular, some virus genome structures are more receptive to recombination than others. For example, genome segmentation is an ancient evolutionary innovation that allows for recombination through genome reassortment. While reassortment undoubtedly assists in the generation of antigenic variation, as in the case of human influenza A virus (Young and Palese 1979; Lowen 2017), that segmented viruses are commonplace in invertebrates that lack adaptive immune systems (Li *et al.* 2015; Shi *et al.* 2016) strongly suggests that reassortment did not evolve for this purpose. Rather, it is possible that placing viral genomes into separate segments was the result of selection to enhance the control of gene expression, which is harder to achieve when genes are encoded by a single contiguous RNA molecule because the same amount of each protein product is produced. A fortuitous by-product of this was segmental reassortment following the mixed infection of single cells. Similarly, the existence of “multicomponent” viruses, in which different genomic segments are present in *different* virus particles, seems too convoluted an arrangement to evolve as a means of facilitating reassortment. A perhaps more reasonable idea is that multicomponent viruses (which mainly

infect plants) originated when individual segments from different viruses, which contributed different functions, co-infected a single cell and evolved to function together (Holmes 2009). Importantly, however, while the origin of recombination/reassortment may involve selection for reasons other than the generation of genetic diversity, once RNA viruses were able to recombine it is likely that natural selection optimized recombination rates to maximize other aspects of viral fitness (Xiao *et al.* 2016).

Finally, recent metagenomic studies of RNA virus diversity have revealed that interspecies recombination and lateral gene transfer across large (*i.e.*, interspecific) phylogenetic distances is far more common than previously realized. Invertebrate RNA viruses in particular appear to be mixing pots for virus genes (Li *et al.* 2015; Shi *et al.* 2016). Indeed, in some instances, RNA viruses may comprise genomic “modules” of differing function that can be placed in varying combinations to create evolutionary novelty through a “modular evolution” (Botstein 1980; McWilliam Leitch *et al.* 2010; Shi *et al.* 2016, 2018).

Metagenomics is Transforming Studies of Virus Evolution

We have only begun to scratch the surface of the biodiversity of RNA viruses in nature. Recent metagenomic studies using bulk shotgun sequencing have made it clear that far, far < 1% of the total universe of viruses, *i.e.*, the virosphere, has been sampled, and with a marked bias toward viruses associated with overt disease in hosts relevant to humans (Geoghegan and Holmes 2017; Shi *et al.* 2018; Zhang *et al.* 2018). This necessarily means that our understanding of RNA virus evolution is based on a tiny, and profoundly biased, subset of virus diversity.

It is trivial to predict that as we sample more of the virosphere through metagenomics, so too will new and perhaps unpredictable features of RNA virus evolution be unearthed. As hinted at throughout this paper, perhaps the most fundamental of these is whether RNA viruses exist that exhibit markedly lower rates of mutation and evolution than those characterized to date. Because there is a strongly inverse relationship between mutation rate per site and genome size (Drake *et al.* 1998; Gago *et al.* 2009), it is also reasonable to assume that those viruses with the lowest mutation rates will also have the largest genomes, although it will be interesting to see if any viruses break this relationship. At present, the maximum observed length of an RNA virus is < 45 kb. Longer genomes are assumed to result in an excessive number of deleterious mutations per replication, and this size-cap is one of the most characteristic features of RNA viruses (Belshaw *et al.* 2007). Although the size of the largest known RNA virus has gradually increased in recent years, all of these longer viruses fall into a single viral order, the Nidovirales (Gorbalenya *et al.* 2006), that uniquely (thus far) encode RNA-processing enzymes that may confer some form of RNA repair (Gorbalenya *et al.* 2006; Lauber *et al.* 2013). Of

course, it will be important to ascertain whether any newly discovered virus families with exceptionally long viral genomes also possess enzymes for RNA repair.

Although other explanations for the small genomes of RNA viruses have been proposed, the idea that they are limited by high mutation rates has gained the most traction (Belshaw *et al.* 2007; Cui *et al.* 2014). In support is the fact that single-stranded DNA viruses—which, like most RNA viruses, lack proof-reading—also experience rates of evolutionary change relatively close to those seen in some RNA viruses (Duffy *et al.* 2008), and similarly possess small genomes. Finally, it is noteworthy that there is a strong allometric relationship between genome and virion sizes in viruses, although what drives what is difficult to resolve (Cui *et al.* 2014). Again, the vast increase in sampling promised by metagenomics offers the chance to test these theories with empirical data.

As well as revealing an abundance of new virus taxa (species, genera, and families) and shedding light on the evolutionary processes that shape this diversity, it is likely that metagenomics will eventually document the existence of viruses in hosts that have not been regularly screened for RNA viruses (such as the Archaea). Similarly, it is highly likely that families of RNA viruses exist that are so divergent in sequence that they cannot readily be detected by the homology-based (e.g., Basic Local Alignment Search Tool- BLAST) detection methods that underpin metagenomics and that impose an arbitrary baseline similarity score (Zhang *et al.* 2018). Until we have a greater understanding of the true biodiversity of RNA viruses it is likely that many of the most vexing questions in RNA virus ecology and evolution will remain unanswered. For example, we know little of the processes that lead to the generation of new virus lineages, nor why some lineages proliferate and others go extinct. Likewise, the factors that shape virus diversity and evolution within ecosystems, and over long-term evolutionary scales, including how viruses emerge and adapt to new hosts, are unclear, as are the factors that dictate why hosts differ so profoundly in the abundance of RNA viruses they carry, and how virus evolution is shaped by intervirus and virus-microbial interactions (Zhang *et al.* 2018). Metagenomics will be central to producing the data that will enable us to address these questions, as well as raising new topics for study that are currently unforeseen.

Perspective

The study of virus evolution has made major advances over the last 40 years. Modern sequencing technologies enable us to describe the extent and pattern of virus genetic variation within and between hosts with remarkable speed and accuracy. The real-time sequencing of thousands of virus genomes during disease outbreaks can now be considered routine, and provides important real-time information for public health intervention. We are entering a new discovery phase in virology, spurred on by advances in deep next-generation sequencing within single hosts and during disease outbreaks, and metagenomic studies of diverse eukaryotic and

prokaryotic taxa. It will surely be the case that this deluge of new data will inspire new evolutionary ideas. An important lesson from the history of evolutionary genetics is that new methods for generating data commonly lead to new theory. As the electrophoretic studies of the 1960s revolutionized population genetics and oligonucleotide fingerprinting kick-started the study of virus evolution in the 1970s, so too will the metagenomics studies of the early 21st century surely lead to new theories on virus origins and evolution.

What, then, will be the role of evolutionary genetics in this new virology? Although it is assuredly the case that methodological advances will result in the continued discovery of novel viruses with hitherto unknown features, and that RNA viruses exhibit prodigious rates of mutation, this does not mean that their evolution needs to be understood outside of the framework of modern evolutionary genetics. As the neo-Darwinian synthesis of the 1930s and 1940s melded work on Mendelian genetics with that of natural selection (Huxley 1942), so too is a new synthesis required for the study of RNA virus evolution that harmonizes detailed and largely experimental studies of viral evolution at the intrahost scale with that occurring at the level of local and global populations, and over the evolutionary timescales inferred through comparative approaches (Figure 3).

Evolutionary genetics may play its most productive role in providing a framework to link evolution at these intra- and interhost scales. Despite the huge amount of viral genome sequence data now generated and our increasing knowledge of the fitness of individual mutations, there remains an important disconnect between evolution within individual hosts and evolution at the epidemiological scale following multiple rounds of virus-host transmission. For example, it is both difficult and dangerous to use short-term patterns to infer long-term evolutionary processes (and vice versa), not only because of time-dependent rates of evolution, but because environments and selection pressures differ markedly within and between hosts.

Although RNA viruses differ fundamentally in their underlying biology, experimental study has shown that the intrahost evolution of RNA viruses that cause short-term acute infections is generally characterized by frequent mutation, strong purifying selection, often limited adaptive evolution because of the short timescale of infection, the possible tissue compartmentalization of virus populations, variable rates of recombination, and relatively simple population dynamics (i.e., a virus population increases in size following initial infection and then sharply declines). In contrast, comparative studies have shown that interhost virus evolution is shaped by complex population dynamics incorporating epidemic peaks and troughs, a variety of epidemiological processes including variable patterns of spatial spread and the impact of “superspreaders,” selection to optimize transmission, differing levels of host immunity, and the recurrent population bottlenecks that accompany interhost transmission and play a major role in shaping genetic diversity. As a case in point, while the intrahost evolution of the influenza virus may be

dominated by stochastic processes (McCrone *et al.* 2018), the antigenic drift of the influenza virus hemagglutinin protein documented at the epidemiological scale is an exemplar of positive selection (Fitch *et al.* 1991).

A new framework for studying RNA virus evolution must therefore find consilience between research at the intra- and interhost scales, linking a variety of evolutionary processes and extending current evolutionary genetic models. Evolutionary genetics is central to bridging this gap because the issue of interest is how genetic diversity is generated and maintained within and among hosts, and understanding how microevolutionary processes combine with large-scale host and ecological phenomena to shape RNA virus macroevolution as depicted in phylogenetic data. Because genome sequence data naturally link these scales and are being increasingly used to provide precise parameter estimates, we believe that the increasing wealth of next-generation and metagenomic data will be central in the development of this new virology.

Acknowledgments

We thank our many colleagues who over the years have provided fruitful discussion on the nature of virus evolution. Special thanks go to Michael Turelli for the original invitation to write this article and his continual encouragement along the way. ECH is funded by an Australian Research Council Australian Laureate Fellowship (FL170100022).

Literature Cited

- Aaskov, J., K. Buzacott, H. M. Thu, K. Lowry, and E. C. Holmes, 2006 Long-term transmission of defective RNA viruses in humans and *Aedes* mosquitoes. *Science* 311: 236–238. <https://doi.org/10.1126/science.1115030>
- Acevedo, A., L. Brodsky, and R. Andino, 2014 Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature* 505: 686–690. <https://doi.org/10.1038/nature12861>
- Aiweisakun, P., and A. Katzourakis, 2016 Time-dependent rate phenomenon in viruses. *J. Virol.* 90: 7184–7195. <https://doi.org/10.1128/JVI.00593-16>
- Andino, R., and E. Domingo, 2015 Viral quasispecies. *Virology* 479–480: 46–51. <https://doi.org/10.1016/j.virol.2015.03.022>
- Archer, A. M., and R. Rico-Hesse, 2002 High genetic divergence and recombination in Arenaviruses from the Americas. *Virology* 304: 274–281. <https://doi.org/10.1006/viro.2002.1695>
- Bamford, D. H., J. M. Grimes, and D. I. Stuart, 2005 What does structure tell us about virus evolution? *Curr. Opin. Struct. Biol.* 15: 655–663. <https://doi.org/10.1016/j.sbi.2005.10.012>
- Bedford, T., M. A. Suchard, P. Lemey, G. Dudas, V. Gregory *et al.*, 2014 Integrating influenza antigenic dynamics with molecular evolution. *Elife* 3: e01914. <https://doi.org/10.7554/eLife.01914>
- Bellacosa, A., and E. G. Moss, 2003 RNA repair: damage control. *Curr. Biol.* 13: R482–R484. [https://doi.org/10.1016/S0960-9822\(03\)00408-1](https://doi.org/10.1016/S0960-9822(03)00408-1)
- Belshaw, R., O. G. Pybus, and A. Rambaut, 2007 The evolution of genome compression and genomic novelty in RNA viruses. *Genome Res.* 17: 1496–1504. <https://doi.org/10.1101/gr.6305707>
- Bordería, A. V., O. Isakov, G. Moratorio, R. Henningsson, S. Agüera-González *et al.*, 2015 Group selection and contribution of minority variants during virus adaptation determines virus fitness and phenotype. *PLoS Pathog.* 11: e1004838. <https://doi.org/10.1371/journal.ppat.1004838>
- Boskova, V., S. Bonhoeffer, and T. Stadler, 2014 Inference of epidemiological dynamics based on simulated phylogenies using birth-death and coalescent models. *PLoS Comput. Biol.* 10: e1003913. <https://doi.org/10.1371/journal.pcbi.1003913>
- Botstein, D., 1980 A theory of modular evolution for bacteriophages. *Ann. N. Y. Acad. Sci.* 354: 484–490. <https://doi.org/10.1111/j.1749-6632.1980.tb27987.x>
- Buonagurio, D. A., S. Nakada, J. D. Parvin, M. Krystal, P. Palese *et al.*, 1986 Evolution of human influenza A viruses over 50 years: rapid, uniform rate of change in NS gene. *Science* 232: 980–982. <https://doi.org/10.1126/science.2939560>
- Burch, C. L., and L. Chao, 2000 Evolvability of an RNA virus is determined by its mutational neighbourhood. *Nature* 406: 625–628. <https://doi.org/10.1038/35020564>
- Chare, E. R., E. A. Gould, and E. C. Holmes, 2003 Phylogenetic analysis reveals a low rate of homologous recombination in negative-sense RNA viruses. *J. Gen. Virol.* 84: 2691–2703. <https://doi.org/10.1099/vir.0.19277-0>
- Ciota, A. T., D. J. Ehrbar, G. A. Van Slyke, G. G. Willsey, and L. D. Kramer, 2012 Cooperative interactions in the West Nile virus mutant swarm. *BMC Evol. Biol.* 12: 58. <https://doi.org/10.1186/1471-2148-12-58>
- Cleaveland, S., M. K. Laurenson, and L. H. Taylor, 2001 Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence. *Philos. Trans. R. Soc. Lond., B* 356: 991–999. <https://doi.org/10.1098/rstb.2001.0889>
- Codóñer, F. M., J. A. Daròs, R. V. Sole, and S. F. Elena, 2006 The fittest versus the flattest: Experimental confirmation of the quasispecies effect with subviral pathogens. *PLoS Pathog.* 2: e136. <https://doi.org/10.1371/journal.ppat.0020136>
- Combe, M., R. Garijo, R. Geller, J. M. Cuevas, and R. Sanjuán, 2015 Single-cell analysis of RNA virus infection identifies multiple genetically diverse viral genomes within single infectious units. *Cell Host Microbe* 18: 424–432. <https://doi.org/10.1016/j.chom.2015.09.009>
- Cui, J., T. Schlub, and E. C. Holmes, 2014 An allometric relationship between the genome length and virion volume of viruses. *J. Virol.* 88: 6403–6410. <https://doi.org/10.1128/JVI.00362-14>
- Cui, H., Y. Shi, T. Ruan, X. Li, Q. Teng *et al.*, 2016 Phylogenetic analysis and pathogenicity of H3 subtype avian influenza viruses isolated from live poultry markets in China. *Sci. Rep.* 6: 27360. <https://doi.org/10.1038/srep27360>
- De Maio, N., C.-H. Wu, K. M. O'Reilly, and D. Wilson, 2015 New routes to phylogeography: a Bayesian structured coalescent approximation. *PLoS Genet.* 11: e1005421. <https://doi.org/10.1371/journal.pgen.1005421>
- Díaz-Muñoz, S. L., R. Sanjuán, and S. West, 2017 Sociovirology: conflict, cooperation, and communication among viruses. *Cell Host Microbe* 22: 437–441. <https://doi.org/10.1016/j.chom.2017.09.012>
- Diehl, W. E., A. E. Lin, N. D. Grubaugh, L. M. Carvalho, K. Kim *et al.*, 2016 Ebola virus glycoprotein with increased infectivity dominated the 2013–2016 epidemic. *Cell* 167: 1088–1098. <https://doi.org/10.1016/j.cell.2016.10.014>
- Domingo, E., 2002 Quasispecies theory in virology. *J. Virol.* 76: 463–465. <https://doi.org/10.1128/JVI.76.1.463-465.2002>
- Domingo, E., D. Sabo, T. Taniguchi, and C. Weissman, 1978 Nucleotide sequence heterogeneity of an RNA phage population. *Cell* 13: 735–744. [https://doi.org/10.1016/0092-8674\(78\)90223-4](https://doi.org/10.1016/0092-8674(78)90223-4)
- Domingo, E., J. Sheldon, and C. Perales, 2012 Virus quasispecies evolution. *Microbiol. Mol. Biol. Rev.* 76: 159–216. <https://doi.org/10.1128/MMBR.05023-11>

- Drake, J. W., 1993 Rates of spontaneous mutation among RNA viruses. *Proc. Natl. Acad. Sci. USA* 90: 4171–4175. <https://doi.org/10.1073/pnas.90.9.4171>
- Drake, J. W., B. Charlesworth, D. Charlesworth, and J. F. Crow, 1998 Rates of spontaneous mutation. *Genetics* 148: 1667–1686.
- Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut, 2006 Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 4: e88. <https://doi.org/10.1371/journal.pbio.0040088>
- Duchêne, S., E. C. Holmes, and S. Y. W. Ho, 2014 Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. *Proc. Biol. Sci.* 281: 20140732. <https://doi.org/10.1098/rspb.2014.0732>
- Dudas, G., L. M. Carvalho, T. Bedford, A. J. Tatem, G. Baele *et al.*, 2017 Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature* 544: 309–315. <https://doi.org/10.1038/nature22040>
- Dudas, G., L. M. Carvalho, A. Rambaut, and T. Bedford, 2018 MERS-CoV spillover at the camel-human interface. *Elife* 7: e31257 (erratum *Elife* 7: e37324). <https://doi.org/10.7554/eLife.31257>
- Duffy, S., L. A. Shackelton, and E. C. Holmes, 2008 Rates of evolutionary change in viruses: patterns and determinants. *Nat. Rev. Genet.* 9: 267–276. <https://doi.org/10.1038/nrg2323>
- Eigen, M., 1971 Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58: 465–523. <https://doi.org/10.1007/BF00623322>
- Eigen, M., 1992 *Steps Towards Life*. Oxford University Press, New York.
- Eigen, M., 1996 On the nature of viral quasispecies. *Trends Microbiol.* 4: 216–218. [https://doi.org/10.1016/0966-842X\(96\)20011-3](https://doi.org/10.1016/0966-842X(96)20011-3)
- Eigen, M., and P. Schuster, 1977 The hypercycle, a principle of natural self-organization. Part A: emergence of the hypercycle. *Naturwissenschaften* 64: 541–565. <https://doi.org/10.1007/BF00450633>
- Elena, S. F., and A. Moya, 1999 Rate of deleterious mutation and the distribution of its effects on fitness in vesicular stomatitis virus. *J. Evol. Biol.* 12: 1078–1088. <https://doi.org/10.1046/j.1420-9101.1999.00110.x>
- Elena, S. F., P. Carrasco, J. A. Daròs, and R. Sanjuán, 2006 Mechanisms of genetic robustness in RNA viruses. *EMBO Rep.* 7: 168–173. <https://doi.org/10.1038/sj.embor.7400636>
- Faria, N. R., J. Quick, I. M. Claro, J. Théze, J. G. de Jesus *et al.*, 2017 Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature* 546: 406–410. <https://doi.org/10.1038/nature22401>
- Firth, C., and W. I. Lipkin, 2013 The genomics of emerging pathogens. *Annu. Rev. Genomics Hum. Genet.* 14: 281–300. <https://doi.org/10.1146/annurev-genom-091212-153446>
- Fitch, W. M., J. M. E. Leiter, X. Li, and P. Palese, 1991 Positive Darwinian evolution in human influenza A viruses. *Proc. Natl. Acad. Sci. USA* 88: 4270–4274. <https://doi.org/10.1073/pnas.88.10.4270>
- Fitzsimmons, W. J., R. J. Woods, J. T. McCrone, A. Woodman, J. J. Arnold *et al.*, 2018 A speed-fidelity trade-off determines the mutation rate and virulence of an RNA virus. *PLoS Biol.* 16: e2006459. <https://doi.org/10.1371/journal.pbio.2006459>
- Gago, S., S. F. Elena, R. Flores, and R. Sanjuán, 2009 Extremely high mutation rate of a hammerhead viroid. *Science* 323: 1308. <https://doi.org/10.1126/science.1169202>
- Geoghegan, J. L., and E. C. Holmes, 2017 Predicting virus emergence amidst evolutionary noise. *Open Biol.* 7: 170189. <https://doi.org/10.1098/rsob.170189>
- Geoghegan, J. L., A. M. Senior, F. Di Giallardo, and E. C. Holmes, 2016a Virological factors that increase the transmissibility of emerging human viruses. *Proc. Natl. Acad. Sci. USA* 113: 4170–4175. <https://doi.org/10.1073/pnas.1521582113>
- Geoghegan, J. L., A. M. Senior, and E. C. Holmes, 2016b Pathogen population bottlenecks and adaptive landscapes: overcoming the barriers to disease emergence. *Proc. Biol. Sci.* 283: 20160727. <https://doi.org/10.1098/rspb.2016.0727>
- Geoghegan, J. L., S. Duchêne, and E. C. Holmes, 2017 Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. *PLoS Pathog.* 13: e1006215. <https://doi.org/10.1371/journal.ppat.1006215>
- Gire, S. K., A. Goba, K. G. Andersen, R. S. Sealfon, D. J. Park *et al.*, 2014 Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* 345: 1369–1372. <https://doi.org/10.1126/science.1259657>
- Grenfell, B. T., O. G. Pybus, J. R. Gog, J. L. N. Wood, J. M. Daly *et al.*, 2004 Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 303: 327–332. <https://doi.org/10.1126/science.1090727>
- Gorbalenya, A. E., L. Enjuanes, J. Ziebuhr, and E. J. Snijder, 2006 Nidovirales: evolving the largest RNA virus genome. *Virus Res.* 117: 17–37. <https://doi.org/10.1016/j.virusres.2006.01.017>
- Hasing, M. E., B. Hazes, B. E. Lee, J. K. Preiksaitis, and X. L. Pang, 2016 A next generation sequencing-based method to study the intra-host genetic diversity of norovirus in patients with acute and chronic infection. *BMC Genomics* 17: 480. <https://doi.org/10.1186/s12864-016-2831-y>
- Holmes, E. C., 2009 *The Evolution and Emergence of RNA Viruses*. Oxford University Press, Oxford.
- Holmes, E. C., and B. T. Grenfell, 2009 Discovering the phylodynamics of RNA viruses. *PLoS Comput. Biol.* 5: e1000505. <https://doi.org/10.1371/journal.pcbi.1000505>
- Holmes, E. C., and A. Moya, 2002 Is the quasispecies concept relevant to RNA viruses? *J. Virol.* 76: 460–462. <https://doi.org/10.1128/JVI.76.1.460-462.2002>
- Holmes, E. C., G. Dudas, A. Rambaut, and K. G. Andersen, 2016 The evolution of Ebola virus: insights from the 2013–2016 epidemic. *Nature* 538: 193–200. <https://doi.org/10.1038/nature19790>
- Huxley, J., 1942 *Evolution: The Modern Synthesis*, Vol. G. Allen and Unwin Ltd, London.
- Kerr, P. J., E. Ghedin, J. V. DePasse, A. Fitch, I. M. Cattadori *et al.*, 2012 Evolutionary history and attenuation of myxoma virus on two continents. *PLoS Pathog.* 8: e1002950. <https://doi.org/10.1371/journal.ppat.1002950>
- Kühnert, D., T. Stadler, T. G. Vaughan, and A. J. Drummond, 2014 Simultaneous reconstruction of evolutionary history and epidemiological dynamics from viral sequences with the birth-death SIR model. *J. R. Soc. Interface* 11: 20131106. <https://doi.org/10.1098/rsif.2013.1106>
- Kuipers, E. J., D. A. Israel, J. G. Kusters, M. M. Gerrits, J. Weel *et al.*, 2000 Quasispecies development of *Helicobacter pylori* observed in paired isolates obtained years apart from the same host. *J. Infect. Dis.* 181: 273–282. <https://doi.org/10.1086/315173>
- Lauber, C., J. J. Goeman, M. del C. Parquet, P. T. Nga, E. J. Snijder *et al.*, 2013 The footprint of genome architecture in the largest genome expansion in RNA viruses. *PLoS Pathog.* 9: e1003500. <https://doi.org/10.1371/journal.ppat.1003500>
- Lauring, A. S., and R. Andino, 2010 Quasispecies theory and the behavior of RNA viruses. *PLoS Pathog.* 6: e1001005. <https://doi.org/10.1371/journal.ppat.1001005>
- Lemey, P., A. Rambaut, A. J. Drummond, and M. A. Suchard, 2009 Bayesian phylogeography finds its roots. *PLoS Comput. Biol.* 5: e1000520. <https://doi.org/10.1371/journal.pcbi.1000520>
- Li, C. X., M. Shi, J. H. Tian, X. D. Lin, Y. J. Kang *et al.*, 2015 Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *Elife* 4: e05378. <https://doi.org/10.7554/eLife.05378>
- Liu, J., I. M. Cattadori, D. G. Sim, J. S. Eden, E. C. Holmes *et al.*, 2017 Reverse engineering field isolates of myxoma virus

- demonstrates that some gene disruptions or losses of function do not explain virulence changes observed in the field. *J. Virol.* 91: e01289-17. <https://doi.org/10.1128/JVI.01289-17>
- Lowen, A. C., 2017 Constraints, drivers, and implications of influenza A virus reassortment. *Annu. Rev. Virol.* 4: 105–121. <https://doi.org/10.1146/annurev-virology-101416-041726>
- Mate, S. E., J. R. Kugelman, T. G. Nysenswah, J. T. Ladner, M. R. Wiley *et al.*, 2015 Molecular evidence of sexual transmission of Ebola virus. *N. Engl. J. Med.* 373: 2448–2454. <https://doi.org/10.1056/NEJMoa1509773>
- McCrone, J. T., and A. S. Luring, 2018 Genetic bottlenecks in intraspecies virus transmission. *Curr. Opin. Virol.* 28: 20–25. <https://doi.org/10.1016/j.coviro.2017.10.008>
- McCrone, J. T., R. J. Woods, E. T. Martin, R. E. Malosh, A. S. Monto *et al.*, 2018 Stochastic processes constrain the within and between host evolution of influenza virus. *Elife* 7: e35962. <https://doi.org/10.7554/eLife.35962>
- McWilliam Leitch, E. C., M. Cabrerizo, J. Cardosa, H. Harvala, O. E. Ivanova *et al.*, 2010 Evolutionary dynamics and temporal/geographical correlates of recombination in the human enterovirus echovirus types 9, 11, and 30. *J. Virol.* 84: 9292–9300. <https://doi.org/10.1128/JVI.00783-10>
- Michod, R. E., H. Bernstein, and A. M. Nedelcu, 2008 Adaptive value of sex in microbial pathogens. *Infect. Genet. Evol.* 8: 267–285. <https://doi.org/10.1016/j.meegid.2008.01.002>
- Moratorio, G., and M. Vignuzzi, 2018 Monitoring and redirecting virus evolution. *PLoS Pathog.* 14: e1006979. <https://doi.org/10.1371/journal.ppat.1006979>
- Mossman, K., S. F. Lee, M. Barry, L. Boshkov, and G. McFadden, 1996 Disruption of M-T5, a novel myxoma virus gene member of the poxvirus host range superfamily, results in dramatic attenuation of myxomatosis in infected European rabbits. *J. Virol.* 70: 4394–4410.
- Moya, A., S. F. Elena, A. Bracho, R. Miralles, and E. Barrio, 2000 The evolution of RNA viruses: a population genetics view. *Proc. Natl. Acad. Sci. USA* 97: 6967–6973. <https://doi.org/10.1073/pnas.97.13.6967>
- Nakajima, K., U. Desselberger, and P. Palese, 1978 Recent human influenza A (H1N1) viruses are closely related genetically to strains isolated in 1950. *Nature* 274: 334–339. <https://doi.org/10.1038/274334a0>
- Neher, R. A., and T. Bedford, 2015 nextflu: real-time tracking of seasonal influenza virus evolution in humans. *Bioinformatics* 31: 3546–3548. <https://doi.org/10.1093/bioinformatics/btv381>
- Neher, R. A., and T. Leitner, 2010 Recombination rate and selection strength in HIV intra-patient evolution. *PLoS Comput. Biol.* 6: e1000660. <https://doi.org/10.1371/journal.pcbi.1000660>
- Parameswaran, P., C. Wang, S. B. Trivedi, M. Eswarappa, M. Montoya *et al.*, 2017 Intrahost selection pressures drive rapid dengue virus microevolution in acute human infections. *Cell Host Microbe* 22: 400–410.e5. <https://doi.org/10.1016/j.chom.2017.08.003>
- Peck, K. M., and A. S. Luring, 2018 Complexities of viral mutation rates. *J. Virol.* 92: e01031-17. <https://doi.org/10.1128/JVI.01031-17>
- Peng, C., S. L. Haller, M. M. Rahman, G. McFadden, and S. Rothenburg, 2016 Myxoma virus M156 is a specific inhibitor of rabbit PKR but contains a loss-of-function mutation in Australian virus isolates. *Proc. Natl. Acad. Sci. USA* 113: 3855–3860. <https://doi.org/10.1073/pnas.1515613113>
- Piatak, Jr., M., M. S. Saag, L. C. Yang, S. J. Clark, J. C. Kappes *et al.*, 1993 High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR. *Science* 259: 1749–1754. <https://doi.org/10.1126/science.8096089>
- Pybus, O. G., and A. Rambaut, 2009 Evolutionary analysis of the dynamics of viral infectious disease. *Nat. Rev. Genet.* 10: 540–550. <https://doi.org/10.1038/nrg2583>
- Pybus, O. G., A. Rambaut, R. Belshaw, R. P. Freckleton, A. J. Drummond *et al.*, 2007 Phylogenetic evidence for deleterious mutation load in RNA viruses and its contribution to viral evolution. *Mol. Biol. Evol.* 24: 845–852. <https://doi.org/10.1093/molbev/msm001>
- Pybus, O. G., A. J. Tatem, and P. Lemey, 2015 Virus evolution and transmission in an ever more connected world. *Proc. Biol. Sci.* 282: 20142878. <https://doi.org/10.1098/rspb.2014.2878>
- Rasmussen, D. A., M. F. Boni, and K. Koelle, 2014 Reconciling phylodynamics with epidemiology: the case of dengue virus in southern Vietnam. *Mol. Biol. Evol.* 31: 258–271. <https://doi.org/10.1093/molbev/mst203>
- Regoes, R. R., S. Crotty, R. Antia, and M. M. Tanaka, 2005 Optimal replication of poliovirus within cells. *Am. Nat.* 165: 364–373. <https://doi.org/10.1086/428295>
- Regoes, R. P., S. Hamblin, and M. M. Tanaka, 2013 Viral mutation rates: modelling the roles of within-host viral dynamics and the trade-off between replication fidelity and speed. *Proc. Biol. Sci.* 7: 280.
- Sanjuán, R., 2012 From molecular genetics to phylodynamics: evolutionary relevance of mutation rates across viruses. *PLoS Pathog.* 8: e1002685. <https://doi.org/10.1371/journal.ppat.1002685>
- Sanjuán, R., 2017 Collective infectious units in viruses. *Trends Microbiol.* 25: 402–412. <https://doi.org/10.1016/j.tim.2017.02.003>
- Sanjuán, R., A. Moya, and S. F. Elena, 2004 The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc. Natl. Acad. Sci. USA* 101: 8396–8401. <https://doi.org/10.1073/pnas.0400146101>
- Sanjuán, R., J. M. Cuevas, V. Furió, E. C. Holmes, and A. Moya, 2007 Selection for robustness in mutagenized RNA viruses. *PLoS Genet.* 3: e93. <https://doi.org/10.1371/journal.pgen.0030093>
- Sanjuán, R., M. R. Nebot, N. Chirico, L. M. Mansky, and R. Belshaw, 2010 Viral mutation rates. *J. Virol.* 84: 9733–9748. <https://doi.org/10.1128/JVI.00694-10>
- Shi, M., X.-D. Lin, J.-H. Tian, L.-J. Chen, X. Chen *et al.*, 2016 Redefining the invertebrate virosphere. *Nature* 540: 539–543. <https://doi.org/10.1038/nature20167>
- Shi, M., X. D. Lin, X. Chen, J. H. Tian, L. J. Chen *et al.*, 2018 The evolutionary history of vertebrate RNA viruses. *Nature* 556: 197–202 (erratum: *Nature* 561: E6). <https://doi.org/10.1038/s41586-018-0012-7>
- Shirogane, Y., S. Watanabe, and Y. Yanagi, 2012 Cooperation between different RNA virus genomes produces a new phenotype. *Nat. Commun.* 3: 1235. <https://doi.org/10.1038/ncomms2252>
- Shriner, D., A. G. Rodrigo, D. C. Nickle, and J. I. Mullins, 2004 Pervasive genomic recombination of HIV-1 *in vivo*. *Genetics* 167: 1573–1583. <https://doi.org/10.1534/genetics.103.023382>
- Simon-Loriere, E., and E. C. Holmes, 2011 Why do RNA viruses recombine? *Nat. Rev. Microbiol.* 9: 617–626. <https://doi.org/10.1038/nrmicro2614>
- Stack, J. C., P. R. Murcia, B. T. Grenfell, J. L. N. Wood, and E. C. Holmes, 2013 Inferring the inter-host transmission of influenza A virus using patterns of intra-host genetic variation. *Proc. Biol. Sci.* 280: 20122173. <https://doi.org/10.1098/rspb.2012.2173>
- Stadler, T., R. Kouyos, V. von Wyl, S. Yerly, J. Böni *et al.*, 2012 Estimating the basic reproductive number from viral sequence data. *Mol. Biol. Evol.* 29: 347–357. <https://doi.org/10.1093/molbev/msr217>
- Stadler, T., D. Kühnert, S. Bonhoeffer, and A. J. Drummond, 2013 Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc. Natl. Acad. Sci. USA* 110: 228–233. <https://doi.org/10.1073/pnas.1207965110>
- Stadler, T., D. Kühnert, D. A. Rasmussen, and L. du Plessis, 2014 Insights into the early epidemic spread of Ebola in Sierra

- Leone provided by viral sequence data. *PLoS Curr.* 6. <https://doi.org/10.1371/currents.outbreaks.02bc6d927ecee7bb-d33532ec8ba6a25f>
- Stern, A., M. T. Yeh, T. Zinger, M. Smith, C. Wright *et al.*, 2017 The evolutionary pathway to virulence of an RNA virus. *Cell* 169: 35–46.e19. <https://doi.org/10.1016/j.cell.2017.03.013>
- Tannenbaum, E., and J. F. Fontanari, 2008 A quasispecies approach to the evolution of sexual replication in unicellular organisms. *Theory Biosci.* 127: 53–65. <https://doi.org/10.1007/s12064-008-0023-2>
- To, T.-H., M. Jung, S. Lycett, and O. Gascuel, 2016 Fast dating using least-squares criteria and algorithms. *Syst. Biol.* 65: 82–97. <https://doi.org/10.1093/sysbio/syv068>
- Turner, P. E., and L. Chao, 1999 Prisoner's dilemma in an RNA virus. *Nature* 398: 441–443. <https://doi.org/10.1038/18913>
- Urbanowicz, R. A., C. P. McClure, A. Sakuntabhai, A. A. Sall, G. Kobinger *et al.*, 2016 Human adaptation of Ebola virus during the west African outbreak. *Cell* 167: 1079–1087.e5. <https://doi.org/10.1016/j.cell.2016.10.013>
- Vignuzzi, M., J. K. Stone, J. J. Arnold, C. E. Cameron, and R. Andino, 2006 Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* 439: 344–348. <https://doi.org/10.1038/nature04388>
- Volz, E. M., and S. D. Frost, 2013 Inferring the source of transmission with phylogenetic data. *PLoS Comput. Biol.* 9: e1003397. <https://doi.org/10.1371/journal.pcbi.1003397>
- Volz, E. M., K. Koelle, and T. Bedford, 2013 Viral phylodynamics. *PLoS Comput. Biol.* 9: e1002947. <https://doi.org/10.1371/journal.pcbi.1002947>
- Webb, G. F., and M. J. Blaser, 2002 Dynamics of bacterial phenotype selection in a colonized host. *Proc. Natl. Acad. Sci. USA* 99: 3135–3140. <https://doi.org/10.1073/pnas.042685799>
- Wertheim, J. O., and S. L. Kosakovsky Pond, 2011 Purifying selection can obscure the ancient age of viral lineages. *Mol. Biol. Evol.* 28: 3355–3365. <https://doi.org/10.1093/molbev/msr170>
- Willner, D., and P. Hugenholtz, 2013 From deep sequencing to viral tagging: recent advances in viral metagenomics. *BioEssays* 35: 436–442. <https://doi.org/10.1002/bies.201200174>
- Wilke, C. O., 2005 Quasispecies theory in the context of population genetics. *BMC Evol. Biol.* 5: 44. <https://doi.org/10.1186/1471-2148-5-44>
- Wilke, C. O., J. L. Wang, C. Ofria, R. E. Lenski, and C. Adami, 2001 Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature* 412: 331–333. <https://doi.org/10.1038/35085569>
- Xiao, Y., I. M. Rouzine, S. Bianco, A. Acevedo, E. F. Goldstein *et al.*, 2016 RNA Recombination enhances adaptability and is required for virus spread and virulence. *Cell Host Microbe* 19: 493–503 [corrigenda: *Cell Host Microbe* 22: 420 (2017)]. <https://doi.org/10.1016/j.chom.2016.03.009>
- Xue, K. S., K. A. Hooper, A. R. Ollodart, A. S. Diggins, and J. D. Bloom, 2016 Cooperation between distinct viral variants promotes growth of H3N2 influenza in cell culture. *Elife* 5: e13974. <https://doi.org/10.7554/eLife.13974>
- Xue, K. S., A. L. Greninger, A. Pérez-Osorio, and J. D. Bloom, 2018 Cooperating H3N2 influenza virus variants are not detectable in primary clinical samples. *mSphere* 3: e00552–17. <https://doi.org/10.1128/mSphereDirect.00552-17>
- Yamashita, M., M. Krystal, W. M. Fitch, and P. Palese, 1988 Influenza B virus evolution: co-circulating lineages and comparison of evolutionary pattern with those of influenza A and C viruses. *Virology* 163: 112–122. [https://doi.org/10.1016/0042-6822\(88\)90238-3](https://doi.org/10.1016/0042-6822(88)90238-3)
- Young, J. F., and P. Palese, 1979 Evolution of human influenza A viruses in nature: recombination contributes to genetic variation of H1N1 strains. *Proc. Natl. Acad. Sci. USA* 76: 6547–6551. <https://doi.org/10.1073/pnas.76.12.6547>
- Young, J. F., U. Desselberger, and P. Palese, 1979 Evolution of human influenza A viruses in nature: sequential mutations in the genomes of new H1N1. *Cell* 18: 73–83. [https://doi.org/10.1016/0092-8674\(79\)90355-6](https://doi.org/10.1016/0092-8674(79)90355-6)
- Zhang, Y. Z., M. Shi, and E. C. Holmes, 2018 Using metagenomics to characterize an expanding virosphere. *Cell* 172: 1168–1172. <https://doi.org/10.1016/j.cell.2018.02.043>

Communicating editor: A. S. Wilkins