

Prisoners of war – host adaptation and its constraints on virus evolution

5 Peter Simmonds^{1*} Pakorn Aiewsakun^{1,2} and Aris Katzourakis^{3*}

10 ¹*Nuffield Department of Medicine, University of Oxford, Oxford, UK.*

²*Department of Microbiology, Faculty of Science, Mahidol University, Bangkok, Thailand.*

³*Department of Zoology, University of Oxford, Oxford, UK.*

15

*Correspondence to: Peter Simmonds Peter.Simmonds@ndm.ox.ac.uk
Aris Katzourakis aris.katzourakis@zoo.ox.ac.uk

20

Recent discoveries of contemporary genotypes of hepatitis B virus and parvovirus B19 in ancient human remains demonstrate that little genetic change has occurred in these viruses over 4,500–6,000 years. Endogenised virus elements in host genomes provide separate evidence that viruses similar to many major contemporary groups circulated 100 million years ago or earlier. In this Opinion article, we argue that the extraordinary conservation of virus genome sequences is best explained by a niche-filling model in which fitness optimisation is rapidly achieved in their specific hosts. Whereas short-term substitution rates reflect the accumulation of tolerated sequence changes within adapted genomes, longer-term rates increasingly resemble those of their hosts as the evolving niche moulds and effectively imprisons the virus in co-adapted virus / host relationships. Contrastingly, viruses that jump hosts undergo strong and stringent adaptive selection as they maximise their fit to their new niche. This adaptive capability may paradoxically create evolutionary stasis in long-term host relationships. While viruses can evolve and adapt rapidly, their hosts may ultimately shape their longer-term evolution.

45 **Table of contents blurb.**

Studies of ancient DNA and endogenous virus elements have revealed extraordinary conservation of virus genome sequences over thousands of years. In this Opinion article, Simmonds, Aiewsakun and Katzourakis describe a niche-filling model in which viruses rapidly adapt to their new niche while their longer term evolution is driven by their hosts.

50

[H1] Introduction

Viruses, with their often small genomes and error-prone replication mechanisms, possess
55 extraordinary adaptive abilities and can display rates of sequence change that are orders of
magnitude greater than those of the hosts they infect. They display evolution in real time as
they acquire antiviral drug resistance, mediate persistent infection through escape from T
and B cell immune system responses to infection, or at the experimental level, rapidly adapt
to different cell culture conditions, new receptors and new hosts. Although biologists since
60 the time of Darwin have convincingly inferred the existence of natural selection from the
current species distributions of animals and plants, and their genetic relationships, this
evidence is almost always indirect and observational. By contrast, virologists have access to
the remarkable field of experimental evolution, such that adaptive processes that may occur
over centuries or millennia in larger organisms can be observed in viruses over days or
65 weeks.

The paradox that this Opinion article aims to address is the increasing evidence for extreme
genetic conservation of viruses over longer periods of evolution. Newly developed methods
to characterise viruses from ancient DNA (aDNA) samples has revealed that viruses that
70 circulated in ancient times do not substantially differ genetically from those that currently
circulate in humans. Furthermore, the discovery of endogenous virus elements (EVEs) in the
genomes of mammals, birds and other eukaryotes shows that viruses similar to
contemporary virus species existed tens of millions of years ago.

75 In this Opinion article, we describe a niche-filling model of virus evolution that aims to
reconcile these conflicting aspects of virus evolutionary histories over different evolutionary
timescales – a framework in which the host represents the primary driver of the
longer term evolution of viruses.

80 [H1] Rapid virus sequence change

A large body of literature documents remarkably high nucleotide substitution rates in virus
genomes, up to 0.1–1% per year in human immunodeficiency virus type 1 (HIV-1) and a wide
range of mammalian RNA viruses^{1–3} and small DNA viruses such as parvoviruses^{4–7}. Virus
sequence change occurs so quickly that phylogenetic trees of their genes are often
85 temporally structured – viruses from older samples show systematically less divergence from

the most recent common ancestor (MRCA) than those collected more recently. As an early example of this phenomenon, the distance from the tree root of sequences of enterovirus 70 isolates collected through the 1970s and 1980s showed a linear relationship with collection date, the calculated nucleotide substitution rate of 5×10^{-3} substitutions/site/year (SSY) allowed the start of the outbreak to be dated to 1967 (REF. ⁸). The availability of virus samples collected over relatively wide date ranges, often stretching back to the 1950s or 1960s, has enabled more sophisticated Bayesian methods (for example, Bayesian evolutionary analysis by sampling trees; BEAST^{9,10}) to estimate dates of origins and substitution rates for a wide range of viruses associated with recent outbreaks. Furthermore, these evolutionary timescales have often been linked to historical events. Amongst many examples, a nucleotide substitution rate of 5×10^{-4} SSY in the hepatitis C virus (HCV) genome was used to calculate dates of emergence of various genotype 2 subtypes to 1470, a finding that might explain the association of genotype 2 infections in areas where the slave trade operated several hundred years ago¹¹. Based on its substitution rate and geography of currently circulating genotypes, hepatitis B virus (HBV) was proposed to have originated in native South American populations and spread into Europe and elsewhere after contact with the Europeans in the 1500s¹². Likewise, the origin of the four genotypes of hepatitis E virus (HEV) infecting humans was estimated to be between 536 and 1344 years ago¹³, and this was suggested to be associated with the spread of pig farming; HEV strains of genotypes 3 and 4 in Japan apparently originated from the 1900s when pigs were first imported from Yorkshire in England¹⁴.

Virus sequence change is often dominated by synonymous substitutions in coding regions that leave sequences of the encoded proteins unaltered. Fixation of these changes may be facilitated by repeated transmission bottlenecks that reduce effective population sizes which occur as the virus transmits between hosts¹⁵. Sequence change may be augmented by adaptive changes. For example, influenza A virus shows rapid, antibody-driven antigenic drift of the haemagglutinin gene that enables it to escape from neutralising antibodies¹⁶. HIV and HCV both fix several amino acid changes in immunodominant T cell epitopes during primary infection that prevents antigen presentation to cytotoxic T cells, contributing to their ability to replicate and transmit^{17,18}.

These observations contribute to a general perception of the ephemeral nature of RNA viruses and a broader idea that viruses are rapidly evolving entities with perhaps frequent

recent origins¹⁹⁻²¹. This appears particularly applicable to those emerging viruses responsible for the numerous recent and often severe disease outbreaks that have afflicted humans, animals and plants. This impression is reinforced by what we know about the origins of particular viruses; the emergence of HIV is indeed documented to be recent, originating from multiple cross-species transmissions of a chimpanzee lentivirus into humans in the late 19th century in Gabon and the Congo²². This was followed by various genomic changes associated with human adaptation and increases in human-to-human transmissibility in the subsequent decades that enabled its spread out of Africa in the 1970s to become a global pandemic^{23,24}. Recent outbreaks of influenza A virus, Nipah virus, Hendra virus, Middle East respiratory syndrome coronavirus and severe acute respiratory syndrome-related coronavirus similarly have zoonotic origins with the associated public health concern of host adaptation and the permanent establishment of these viruses in human populations²⁵.

[H1] A darkening cloud of uncertainty over viral evolutionary rates. Methods that predict the temporal dynamics and phylogeography of recent virus emergence have been remarkably effective in reconstructing recent virus evolutionary histories. Although extrapolation of these substitution rates to longer periods seemingly provides the means to reconstruct much deeper evolutionary histories of viruses, a series of recent developments challenges the applicability of such methods to viruses and, more disturbingly, the widely accepted concepts of the evolutionary timescales of viruses.

An early and convincing example of potential problems with extrapolating substitution rates was found in estimates of the dates of divergence of simian immunodeficiency virus (SIV) strains that were the source of HIV-1 and HIV-2 infections in humans, and of SIV variants infecting various monkey species^{22,26}. Relatively rapid substitution rates, such as the 1.38×10^{-3} SSY (range $1.03 \times 10^{-3} - 1.73 \times 10^{-3}$) calculated for SIV strains infecting African green monkeys²⁶ predicted time spans of hundreds of years for these divergence events and strengthened concepts of their relatively recent origins. However, a subsequent study of SIV strains infecting isolated populations of Old World monkeys on the island of Bioko, Equatorial Guinea, 32 km off the coast of Africa, was entirely incompatible with this recent origin hypothesis²⁷. Although post-glacial sea level rises separated the island from the African landmass over 10,000 years ago, SIV strains were found to be minimally divergent from those infecting the same species in mainland Africa monkey populations. These observations lowered the minimum substitution rates of each of the SIV strains by over two

orders of magnitude and, extrapolated back, predicted an MRCA for SIVs infecting different
155 host species to around 80,000 years before the present (BP).

This isolated (literally) geological separation event provided a single opportunity to look at
longer timescales for virus evolution. However, systematic investigation has been hampered
by the general unavailability of suitably stored (that is, frozen) samples dating back to much
160 before the 1960s or 1970s from which viruses can be reliably recovered. Without the
opportunity to investigate long-term substitution rates, the paradigm of RNA viruses being
highly mutable emerged and has dominated much of the thinking about their evolution over
many decades. Many have noted the depiction of what looks like poliomyelitis in a man on
an Ancient Egyptian stele that dates to the 18th dynasty (reviewed with other possible
165 depictions in Ancient Egypt in REF. ²⁸) but could poliovirus have existed in the 14th century
BC? By conventional extrapolation, the emergence of the enterovirus C species (to which
poliovirus belongs) would be dated to only a few hundred years ago²⁹, not >3,000 years ago.

Two recent developments have provided the means to look further back into virus
170 evolutionary histories. These challenge current thoughts about virus nucleotide substitution
rates and the time depths for their evolution.

[H2] Findings from ancient DNA (aDNA) and archaeovirology. DNA degrades after the death
of the host, but it can be effectively sequenced by next generation sequencing methods.

175 These newly developed methods have allowed the genomes of ancient human populations
to be sequenced and enabled direct analyses of genetic relationships between
contemporary humans, Neanderthals and Denisovans, and other archaic human population
groups over the last hundred thousand years³⁰⁻³². aDNA-based studies have also contributed
to investigations of the longer-term evolution of viruses over historical timescales, including
180 the analysis of parvovirus B19 (B19V) in human remains dating from World War 2 in Russia³³,
the pandemic 1918 influenza A virus H1N1 strain from Alaskan permafrost³⁴, and HBV and
smallpox in mummified material from the 1600s^{35,36}. The timescales over which aDNA
sequences can be recovered have now been extended by three recent reports of the
detection of viruses in human samples dating back to the early Neolithic (5,000 BC)³⁷⁻³⁹.

185 Two recent studies report the detection of HBV in several individuals in European and
Central Asian populations as early as the Bronze Age and Neolithic (2,500 – 3,000 BC^{37,38}).

Viruses circulating in these prehistoric times in many cases matched currently circulating HBV genotypes (genotypes A, B and D), only 1.3–3% divergent from modern strains. This indicates a long term substitution rate of 8.04×10^{-6} – 1.51×10^{-5} SSY, which is 100-fold lower than that measured in contemporary samples (7.72×10^{-4} SSY⁴⁰). Similar samples also provided evidence for the circulation of B19V in humans from Central Asia 5,000 BC and in Vikings from Sweden 1,000 AD³⁹. These strains closely matched contemporary genotypes (type 1 and 2) and a similarly revised lower substitution rate estimate was observed.

Whereas an early study of sequence change in B19V (REF. 7) predicted a time of origin of current genotype 1 strains to 1960s–1970s, the aDNA study indicated that this genotype was actually alive and kicking in Eurasia in the early Neolithic era, nearly 7,000 years ago. Further analyses of progressively older aDNA sequence libraries will undoubtedly reveal more insights into the pace of virus evolution for ever-widening collections of human, animal and plant viruses.

[H2] Findings from endogenised virus elements and paleovirology. A second and again, entirely unanticipated opportunity to study virus evolution over even longer periods was provided by the discovery that copies of DNA and RNA viruses can become integrated in the genomes of animals and plants^{41–45} (Box 1 and Supplementary Figure 1). Once endogenised, EVEs are genetically stable and preserve information about the circulation of ancient viruses that is impossible to infer from examination of contemporary virus populations. For example, lentiviruses were originally considered as a recently-emerged group of viruses, based on the very recent origins of HIV-1 itself and measured substitution rates that place the origins of lentiviruses to a few thousand years ago²². However, endogenous lentiviruses in rabbits⁴⁶, ferrets, Madagascan lemurs and Colugos demonstrate the circulation of lentiviruses over almost the entire time span of mammalian evolution^{47–49}. In addition to retroviruses, other RNA and DNA viruses have also adventitiously integrated into host germlines and created records of ancient infections. Based on their distribution in descendant species, filoviruses^{45,50}, parvoviruses, circoviruses and bornaviruses must have all circulated over long periods during mammalian evolution⁴⁵. In addition, the detection of reptilian hapadnaviruses provides evidence for the circulation of this virus in the early Mesozoic >200 million years (Myr) ago, long before the radiation of mammals⁵¹.

220 The presence of EVEs in contemporary host genomes provides irrefutable evidence that viruses recognisably similar to contemporary strains have been continuously infecting their hosts over timescales spanning tens of millions of years.

[H2] Virus–host co-evolution. Predictions on the longevity of virus lineages from the EVE
225 ‘fossil’ record are further supported by observations of the apparent co-speciation of viruses and hosts⁵²; these observations can inform predictions about the even earlier origin of specific viral groups. For example, the phylogeny of spumaviruses closely follows that of their mammalian, amphibian and piscine hosts, consistent with virus–host co-speciation over 450 Myr^{53,54}. The proposed co-evolution of papillomaviruses with their hosts suggests
230 their similarly ancient origins 400–600 Myr⁵⁵. Increasingly divergent homologues of HBV have been observed as EVEs in birds and reptiles⁵¹, and exogenous hepadna-like viruses have recently been found in fish genomic libraries⁵⁶. The authors of the latter study propose a co-evolutionary scenario in which the ancestor of currently extant HBV-like viruses may have existed >400 Myr. In a similar but even more extreme example, homologues of
235 polyomaviruses have been detected in DNA libraries of vertebrates and scorpions and spiders⁵⁷, implying a pre-Cambrian origin before the common ancestor of deuterostomes and protostomes ~650 Myr.

In the following sections, we aim to clarify how the remarkable similarity of ancient viruses
240 discovered through archaeovirology and palaeovirology to contemporary sequences can be explained, given the extraordinary rates of evolutionary change that viruses can undergo.

[H1] Short and long-term rates of viral evolution. When viral evolution is measured over short timescales, rapid rates of sequence change are typically observed. However, over
245 longer timescales, viral evolutionary rates are several orders of magnitude slower, approaching those of their hosts. Rather than a simple dichotomy between short and long timescales, viral evolutionary rates appear to decrease continuously with the timescale of measurement⁵⁸, with a decay rate that is strikingly consistent with a power law relationship between substitution rate and observational period⁵⁴ (Fig. 1). Over the longest timescales
250 (100 million–1 billion years), substitution rates for DNA and RNA viruses of any configuration were remarkably similar: rates of $1\text{--}5 \times 10^{-9}$ SSY; these in turn closely match the 2.2×10^{-9} SSY mean substitution rate calculated for mammalian genes⁵⁹. At the other end of the scale, short-term substitution rates varied by virus group with slower rates for dsDNA viruses ($4 \times$

10⁻⁴ SSY) than RNA viruses (8 x 10⁻³ SSY for those with positive-strand RNA genomes), with a
255 degree of virus lineage-specific variability in short term rates within each Baltimore group
(discussed in REF. ⁵⁸). However, for each Baltimore group, rate decay over time was
comparable. Remarkably, the recently obtained substitution rates from aDNA studies
superimpose directly upon the regression line inferred from other methods (Fig. 1; blue
dots).

260 Several hypotheses have been proposed to account for the time-dependent rate
phenomenon (TDRP)^{42,60}, many of which have been developed to account for substitution
rate variability in other organisms (reviewed in REF. ⁶⁰). Using inappropriate substitution
models frequently leads to underestimations of age through, for example, the effects of
265 saturation⁶¹. However, it is unlikely that even the most complex currently available models
can accurately capture nuances of viral genome evolution (for example, the effects of gene
overlap, epistasis and nucleotide biases) and reconcile these disparities in age estimations.
Sequencing errors, now rare in next-generation sequencing data, could also elevate recent
rate estimates, but this effect cannot scale over the longer timescales, over which rate
270 variation is observed. Explanations positing changes in biology over time have also been put
forward, such as variance in the fidelity of viral polymerases⁶², but it is difficult to see how
such features could explain the wide-ranging observation of the phenomenon across taxa
and over time. Perhaps the most widely accepted explanation is that short-term rate
measurements capture population level processes including transient deleterious mutations
275 and transient beneficial but ‘shortsighted’ adaptations for their current host^{63,64}, but which
do not survive in the longer term, whereas long-term rates more closely represent the ‘true’
fixation rate of mutations over macroevolutionary timescales^{58,65}. Although this explanation
could account for the TDRP over short timescales, it is not clear whether deleterious
mutations persist for long enough to explain the effect over timescales spanning millions of
280 years.

Although these explanations have been of considerable value in accounting for the TDRP
phenomenon in hosts⁶⁰, none appear to provide an adequate explanatory framework for the
>1 million-fold range in virus substitution rates over different observation periods (Fig. 1)
285 and the long-term extreme conservation of virus genomes. These findings beg the question:
what prevents viruses with their seemingly unlimited evolutionary potential from forever
diversifying? An overarching model that reconciles both the high rates of sequence change

over short timescales and what appear to be implausibly early origins for many virus groups at the other extreme is currently lacking. Although the wide-ranging existence of the TDRP across viral groups and timescales provides an observational description of how apparent viral evolutionary rates vary over time⁵⁸, we lack a biologically realistic functional model that could account for the apparent ubiquity of this phenomenon.

[H1] Host-driven virus evolution. As an alternative explanatory model, we developed ideas originating from niche-filling models⁶⁶⁻⁶⁹ that emphasise the role of host interactions in shaping virus evolution. This approach contrasts with the typically virus-centric accounts of their evolution in the literature and provides the means to account for the remarkably different trajectories of their evolution at different ends of the observational timescale. Including the host in our model does, however, place unfamiliar constraints on the concept of progressive and diversifying virus evolution.

In this model, high error rates and large population sizes achieved on infection of macroscopic hosts provide viruses with extraordinary adaptive abilities that enable them to maximise fitness in whatever host environments they find themselves (Fig. 2 and Box 2). As viruses can rapidly evolve to a fitness peak in a given host environment, this may have the paradoxical effect of restricting sequence change rather than accelerating it in any period other than the short-term. Infection of the same host over tens or hundreds of years, or perhaps even millennia may drive the evolution of each host-adapted virus to evolutionary stasis – an optimised genome that is maximised in those aspects of its fitness that maintain infections in the host population (Fig. 3). This idea is consistent with the model proposed many years ago that close cooperation between RNA virus proteins and host proteins requires their coevolution and thus limits their divergence⁷⁰. However, this stasis may extend much further, not just to the amino acid co-variation within virus genes but also to the preservation of sites at synonymous coding positions and non-coding regions that preserve codon choices, RNA secondary structures and replication elements. Once fully adapted to their niche, the intensity of peer competition may create virus genomes with few genuinely phenotypically neutral sites.

[H2] Host adaptation. The process of host adaptation generates viruses that are primarily shaped by the constraints of the niche and less by the ancestry of the virus. If we take parvovirus B19V and HBV as examples of viruses showing evidence for long term presence in

their host populations, their genotypes typically show diversity in the 10%–15% nucleotide sequence divergence range, which is represented figuratively as the blue area of potential sequence ‘wobble’ in the virus niche (Fig. 2). This pattern of within-species diversity typifies a wide range of other human, veterinary and plant viruses; examples of the former include individual serotypes of alphaviruses, flaviviruses, measles virus, mumps virus, most of the paramyxoviruses and coronaviruses, etc. This pattern is also the norm for the vast range of virus species infecting arthropods and fungi, and represents the fraction of genome sites not under selection for fitness optimisation. Variation at this level represents the majority of what is captured in temporal sampling and may underlie the generally rapid substitution rates reported for RNA and small DNA viruses over short observation periods. However, the sequence space is small and restrictive — changes at those few neutral sites may saturate at much lower divergence levels than evolutionary models typically expect. We might describe this constraint as a ‘cage’ — not in the sense of the limited genome size of RNA viruses⁷¹, but reflecting those host-imposed constraints on virus sequence change that create the appearance of much less sequence divergence and hence temporal depth than is actually present.

Over much longer periods, virus genome sequence change driven by the host change resembles niche-filling models developed for phenotypic trait evolution in cellular organisms^{68,69}; traits evolve adaptively to fit the niche in which a viral species finds itself, rather than for example, a random-walk model in which traits evolve continuously and progressively over time, and lead to clock-like sequence change. The niche is defined by the host organism in which the virus infects, the viral sequence defines the phenotype, and changes are primarily adaptive. Short-term substitution rates simply reflect a virus exploring the limits of its ‘cage’ at rates linked to their error rates and demography; longer term diversification of RNA and DNA viruses calculated from aDNA and EVE data (Fig. 1) reflects how viruses adapt as niche ‘shapes’ change (Fig. 3). These changes ultimately drive the long-term evolution of viruses, and explain why their nucleotide substitution rates ultimately approach those of their hosts.

[H2] Host jumps. The model equates virus jumps with the occupancy of a new niche, and hence a rapid adaptation of trait values to fit this niche (Fig. 4). Host jumps are associated with periods of accelerated sequence change as the virus re-models and regains fitness in an altered environment, very much as conceptualised in bacterial evolution⁷². Host adaptation

after cross-species transmission is associated with rapid amino sequence changes of viral genes, typically those associated with receptor interactions and the evasion of innate immunity⁷³⁻⁷⁷ but often pervasive throughout the entire virus genome⁷⁸. Larger scale gene modifications, such as the re-purposing of the HIV-1 accessory protein Vpu to antagonise the cellular antiviral protein tetherin was a key adaptive change that enhanced its replication ability in humans following its zoonotic transfer from chimpanzees²³. The diversification of HIV-1 populations in the 100 or more years since its zoonotic introduction might indeed be interpreted as an on-going process of fitness optimisation. The gradual attenuation of disease severity in HIV-1 infections⁷⁹ perhaps anticipate a time when HIV-1 diversity is substantially lessened following niche adaptation and the evolution of fitness optimised, less pathogenic and fully host-adapted HIV-1 strains. HIV-1 population structures and diversity may ultimately match the endemic and tolerated SIV strains that have infected and adapted to many Old World monkey species over much longer periods.

In vertebrates, further adaptive change is driven by their highly polymorphic adaptive immune system. The heterogeneity of the major histocompatibility complex (MHC) between individual hosts defines virus epitope recognition and hence the adaptive changes required to avoid antibody or T cell recognition^{17,18}. Immediately after infection, immune escape of viruses in different individuals may drive rapid antigenic diversification. However, the sequential transit of a virus through dozens or many hundreds of individuals may lead to a static cycle of adaptation on infection and reversion on transmission through different MHC repertoires. At the population level, there may be no net sequence change, an interesting variant of the Red Queen hypothesis^{80,81}. This larger adaptive space (but still a cage) feeds into a complex dynamic of population susceptibility, transmission rates, neutralisation escape and changes in receptor use that perpetuate infections in hosts with adaptive immunity. The elaborate serotype and antigenic shift/drift population structures of mammalian viruses in particular may be its direct consequence.

[H1] Conclusions. In this Opinion article, we present a model of virus sequence change that links substitution rates to those of their long-term hosts, providing an alternative paradigm for understanding virus evolution and adaptation, and the associated TDRP. Although it is known that viruses evolve under constraints and adapt to hosts on transmission, the perspective we offer casts viruses and their genetic relationships to each other as being primarily conditioned by hosts they infect. Their own genetic history that is emphasised so

much in virus-centric accounts of their evolution over short periods is quite subservient to the shaping forces of host-driven evolution. Similarly, although existing accounts of virus sequence change are so much focused on their seemingly unlimited evolutionary potential and adaptability, the range of viruses that are able to successfully infect and maintain transmission in their hosts appears limited and is more a function of the host niches a virus can exploit⁶⁶. The wide range of viruses that infect humans possess specific tissue tropisms, pathologies and transmission routes. However, homologues of these viruses in other mammalian species typically reproduce these virus–host relationships very closely. As further evidence of host-induced constraints, virus replication ability, transmissibility and successful establishment of zoonoses is predicated, at least in part, on the degree of relatedness of the hosts involved in the host jump^{82–85}. Host relatedness indeed underpins the distribution and pathogenicity of lentiviruses infecting primates and humans^{86,87}. If viruses were genuinely able to adapt and innovate in any host environment, these regularities and apparent niche restrictions across viruses infecting different hosts would not occur.

Although this moulding process equates ultimate virus evolutionary rates to those of their hosts, the niche perspective is also fully consistent with the hypothesis of neutral evolution of viruses over the much shorter periods of virus evolution observed in contemporary virus samples (as discussed in REF. ⁸⁸). Indeed, more than any other factor, the idea that host-adapted viruses are exploring space around a small cage of tolerated substitutions accounts best for the absurdly different short and long-term substitution rates they display over differing evolutionary timescales. That small cage and the consequent isolation of virus populations from each other may frequently underpin what are classified as virus species in virus taxonomy^{89,90}, which we may now regard as constrained, separate virus populations with often highly demarcated host ranges. The model of host-driven virus evolution thus places viruses as long-term residents of the hosts they infect, perhaps over millions of years or longer, a concept that accords with the general host specificity that virus species display. The majority of their differences from each other are driven by their host adaptation; niche-filling models accord with the growing evidence of the role of selection and adaptation as the driving forces behind longer-term evolution and speciation elsewhere in biology⁹¹.

There seems to be a beautiful paradox in virus evolution — the same remarkable ability of viruses to rapidly adapt to new hosts, and escape from innate and adaptive immune

responses may also help to create the evolutionary stasis of viruses in long-term host
425 relationships. It is the viruses in their niches that are conservative and it is their hosts that
force them to change.

430 **Acknowledgements**

We would like to thanks Jessica Metcalf, Princeton University, for reviewing and providing helpful comments on the manuscript before submission.

Author contributions

435 P.S., P.A. and A.K. researched the data for the article. P.S., P.A. and A.K. substantially contributed to discussion of content. P.S., and A.K. wrote the article. P.S., P.A. and A.K. reviewed and edited the manuscript before submission.

Competing interests

The authors declare no competing interests.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

440 **Supplementary information**

Supplementary information is available for this paper at <https://doi.org/10.1038/s415XX-XXX-XXXX-X>

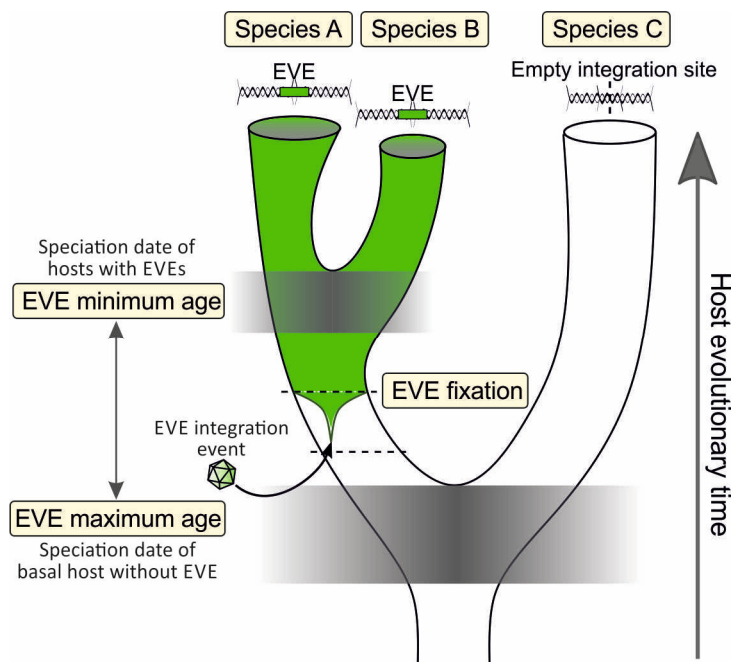
445

Box 1: Endogenised virus elements.

Genome sequencing of animals and plants has revealed the existence of large numbers of integrated copies of DNA and RNA viruses in host genomes corresponding to all known major virus groups⁴¹⁻⁴⁵. As part of host genomes, endogenised virus elements (EVEs) are inherited, vertically passing from parents to offspring to create a genomic 'fossil record' stretching back millions of years (Figure). These preserve information about ancient viruses that would have been impossible to reconstruct from contemporary virus populations (Supplementary Figure 1).

The timing of integration and thus the dates when exogenous forms of the virus circulated can be estimated by examination of the distribution of EVEs in descendant host species (Figure). The endogenous lentiviruses in rabbits⁴⁶ integrated over 12 million years ago, based on the presence of unambiguous orthologous copies of this virus in lagomorph species that diverged after this time⁹². Integration times calculated for endogenous lentiviruses detected in ferrets, lemurs, and Colugos further demonstrate the circulation of lentiviruses in the range of tens of millions of years⁴⁷⁻⁴⁹.

The EVE record formed by retroviruses provides the richest datasets because of their obligate genome integration step in their replication cycle. However, other viruses can be adventitiously reverse transcribed after which the cDNA can integrate into the host cell germline and form EVEs (Supplementary Figure 1). Recent characterisation of genome sequences of a wide range of mammals and birds has revealed the existence of integrated copies of all known major virus groups⁴⁵. Examples include a filovirus, similar to Ebola virus that integrated over 30 million years ago into the genomes of rodents^{45,50}. Similar integration events include parvoviruses (>30 Myr), circoviruses (>60 Myr), and bornaviruses in elephants, hyraxes and tenrecs (>93 Myr)⁴⁵. The times of integration events must be regarded as conservative minimum estimates — viruses dated from their presence as orthologues may have circulated long before germline integration in the most recent ancestor of their current hosts.



The Figure depicts an integration event of an exogenous virus into a host germline and its subsequent inheritance in two descendant species, A and B, and its absence in species C, that split before the EVE integration event. As the approximate timescale for vertebrate evolution is known from the fossil record, the distribution of EVEs in contemporary species provides fixed minimum and maximum dates for their integrations. This in turn provides strong evidence of when the virus circulated.

Box 2: What is a host niche?

A niche is effectively the whole environment in which a virus replicates, inside a cell and between cells during cell–cell spread and host transmission (Fig. 2). Although depicted as a spatial fit, the nature of the virus–host interaction and its adaptation involves both virus interactions with host factors that enable replication and specific adaptations to counter innate cellular defence mechanisms. Virus fitness is further determined by broader host interactions, most crucially its choice of either an acute or persistent lifestyle strategy for evasion of host systemic and adaptive immune responses. Control of virus replication, modulation of their pathogenicity, effective transmission routes and ultimately the existence of reservoirs of new hosts to infect are all factors that determine the evolutionary success of a virus.

Niche evolution

Host factors that delimit a niche are themselves subject to continuous change (Fig.), as hosts diversify and speciate over longer evolutionary periods. The dynamics and pace of their evolution differ between cellular features exploited by the virus for replication, and host defence factors that are specifically purposed to protect the host. The former include cell surface receptors, translational mechanisms and the nuclear or cytoplasmic structural elements that are parasitized by the virus to build replication and virus assembly sites. The latter include elements of the innate, cellular and systemic response elements that directly interact with viruses to limit or clear infections. Genes associated with host antiviral mechanisms frequently show elevated evolutionary rates and evidence for positive selection once engaged in intricate arms races with their virus targets that aim to counter their antiviral functions⁹³⁻⁹⁶. Accelerated niche-associated evolution in such genes may reproduce the power law relationships between observation period and virus substitution rates (Fig. 1).

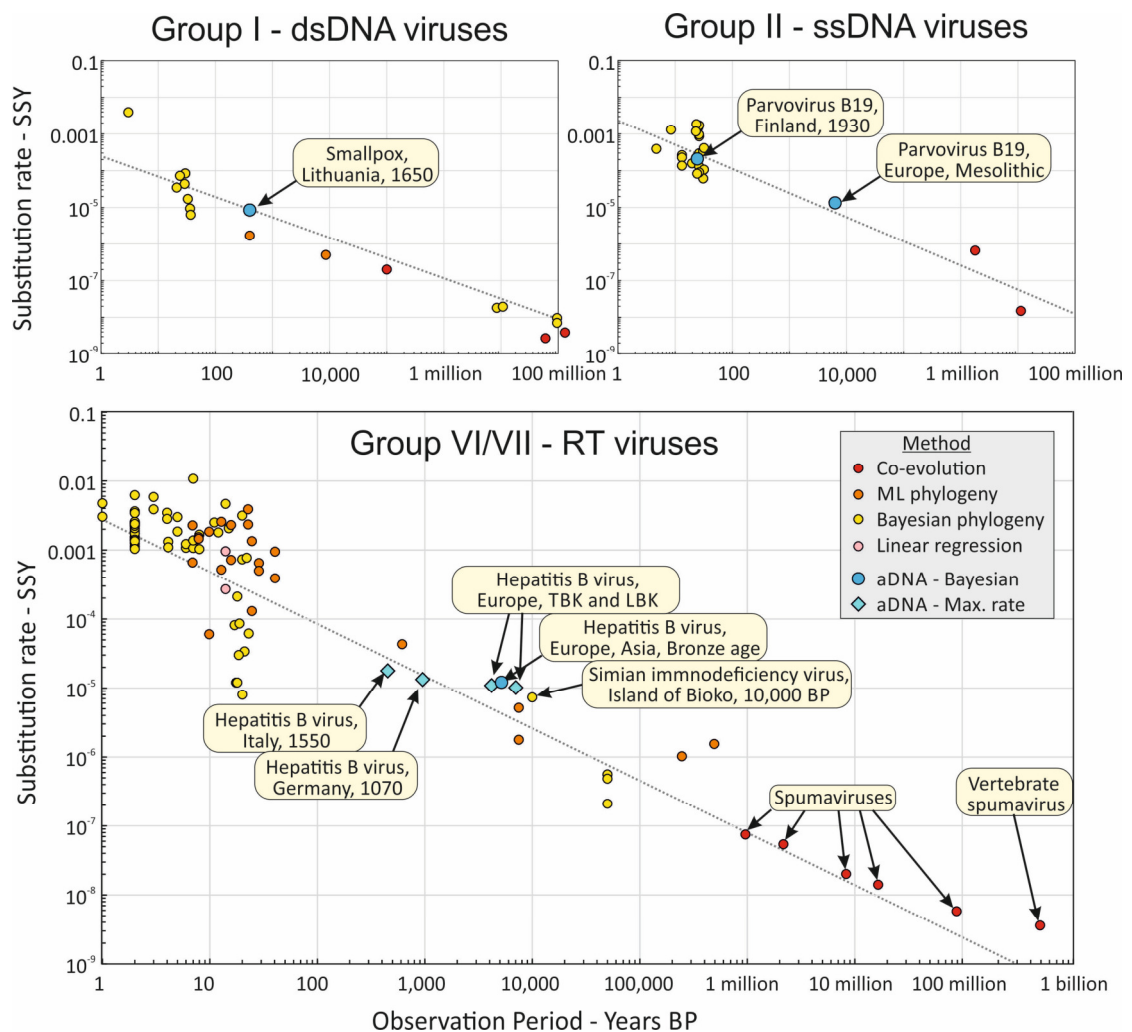


Fig. 1. Virus genome nucleotide substitution rates of different observation periods

Substitution rates of DNA and RNA viruses calculated over different time periods using different methods. These include Bayesian evolutionary reconstructions and rates inferred from instances of virus–host co-evolution (see figure key). Data used in the figure was based on a previous analysis of published virus substitution rates with different genomic configurations⁵⁸ and expanded with more recent published data (listed in full in Table S3; Suppl. Data).

Three groups are depicted: dsDNA viruses in Baltimore Group I (part a), ssDNA in Baltimore Group II (part b) and reverse transcribing (RT) viruses in Baltimore Groups IV and V⁵⁸ (part c). These groups showed a remarkably similar relationship between substitution rate (y-axis) and observation times over which substitution rates were

calculated (plotted on a log transformed scale on the x-axis), despite their intrinsic differences in replication error rates and evolutionary histories. The regression line is based on substitution rates calculated from co-evolution and phylogeny methods. Rates inferred from very ancient co-evolutionary scenarios among reverse transcribing (RT) viruses show a potential flattening of substitution rates as they approach those of host genes (mean value 2.2×10^{-9} substitutions/site/year (SSY) REF.⁵⁹.

Evolutionary rates estimated from ancient DNA (aDNA) sequences of variola virus³⁵, hepatitis B virus (HBV)³⁷ and parvovirus B19³⁹ (blue circles) superimpose directly onto rates calculated by other methods. Maximum substitution rates (aDNA – max. rate) for other HBV sequences^{36,38} were calculated from their divergence to the most closely related contemporary HBV strains (blue diamonds).

Abbreviations used in figure labels: ML: maximum likelihood, TBK, LBK: pottery derived terms Trichterbecher (funnel beaker) and Linearbandkeramik (linear band ware) used to describe European Neolithic populations.

Evolutionary rates estimated from aDNA sequences of smallpox⁴⁶, HBV⁴⁸ and parvovirus B19⁵¹ (blue circles) superimpose directly onto rates calculated by other methods. Maximum substitution rates for other HBV sequences^{47,49} were calculated from their divergence to the most closely related contemporary HBV strains (blue diamonds).

A) Niche adapted virus

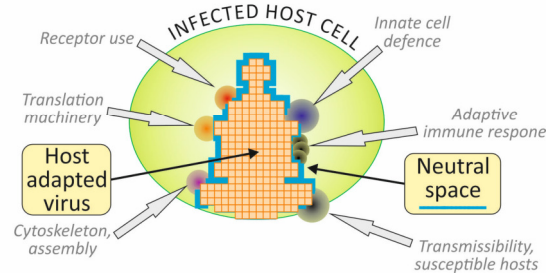


Fig. 2. A spatial representation of a virus infecting a cell

The host niche is depicted as simplified, spatial representation of the host environment that a virus occupies (see Box 2 for an outline of the typical host elements defining a niche). The range of host factors exploited by the virus and those associated with host response are depicted as pressure points (filled circles) on the virus that restrict divergence in virus regions involved in these cellular interactions. The blue area represents variable extents of sequence space in which sequence change may occur without phenotypic cost (neutral space).

B) Host -driven virus evolution

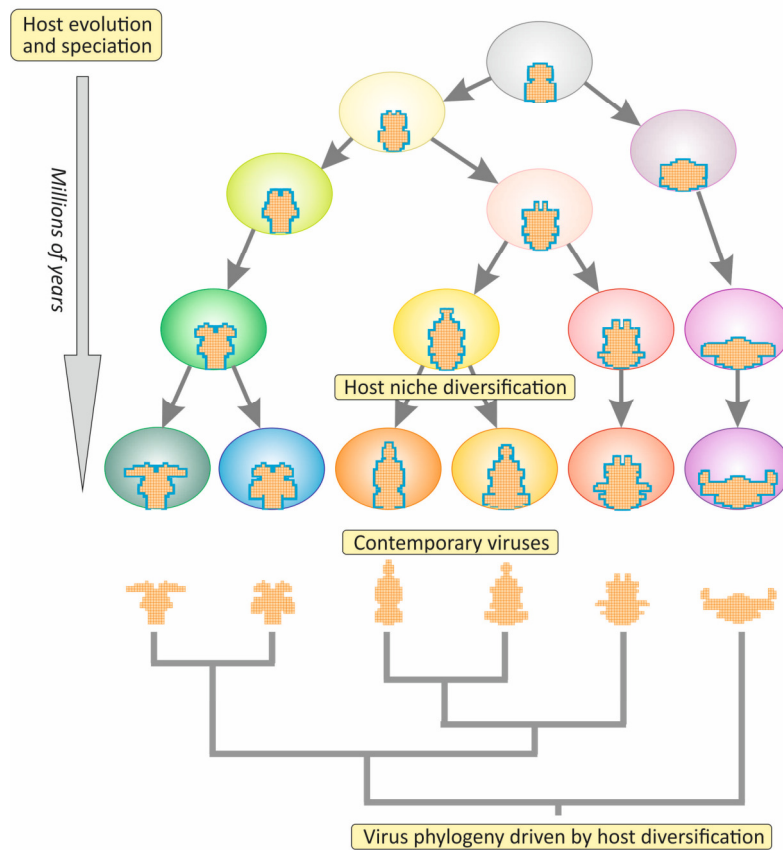


Fig. 3. Host-driven virus evolution.

Viruses remain associated and highly adapted to their host, even as the hosts themselves evolve and speciate over long periods (tens of millions or potentially hundreds of millions of years). Viruses continue to infect cells in each host lineage but they themselves must evolve in concert with their host to retain fitness and host adaptation as the niche they occupy gradually changes. After a prolonged period of co-evolution, viruses acquire very different virus 'shapes' and a phylogeny that resembles in part that of their host. Viruses involved in this co-evolutionary process display long term substitution rates that approach those of their hosts.

530

FIGURE 4

C) Adaptive sequence change - transmission to a new host

535

540

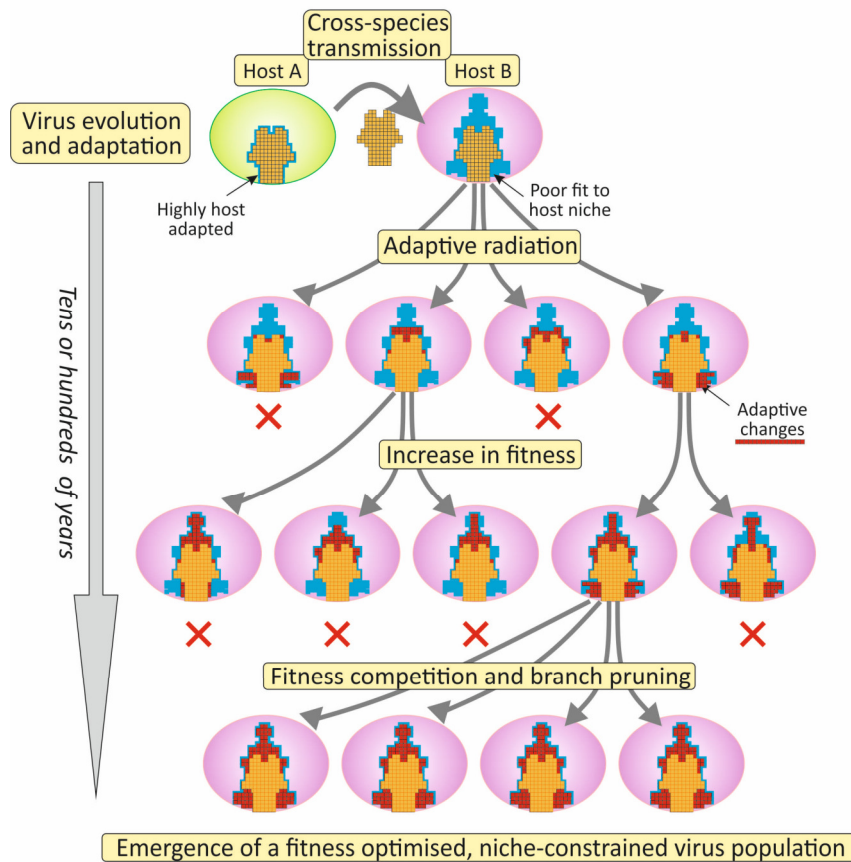


Fig. 4. Cross-species transmission and niche adaptation.

545

A virus adapted to host A may be able to infect an alternative host (host B) but it may be initially poorly adapted to any available niches. Rapid fixation of adaptive changes improves virus fitness associated with sequence diversification. Fitness competition leads to subsequent branch pruning and the emergence of a highly adapted virus strain that is genetically distinct from the founder virus over relatively short evolutionary periods. The red crosses label lineages that have extinct over the period of virus / host adaptation.

550

REFERENCES

- 1 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).
- 555 2 Lemey, P., Rambaut, A. & Pybus, O. G. HIV evolutionary dynamics within and among hosts. *AIDS reviews* **8**, 125-140 (2006).
- 3 Sanjuan, R. From molecular genetics to phylodynamics: evolutionary relevance of mutation rates across viruses. *PLoS Pathog* **8**, e1002685, doi:10.1371/journal.ppat.1002685 (2012).
- 560 4 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, doi:10.1093/molbev/msh109 (2004).
- 565 5 Duffy, S., Shackelton, L. A. & Holmes, E. C. Rates of evolutionary change in viruses: patterns and determinants. *Nat.Rev.Genet.* **9**, 267-276 (2008).
- 6 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.U.S.A.* **102**, 379-384 (2005).
- 570 7 Shackelton, L. A. & Holmes, E. C. Phylogenetic evidence for the rapid evolution of human B19 erythrovirus. *Journal of Virology* **80**, 3666-3669 (2006).
- 8 Takeda, N., Tanimura, M. & Miyamura, K. Molecular evolution of the major capsid protein VP1 of enterovirus 70. *Journal of Virology* **68**, 854-862 (1994).
- 575 9 Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* **29**, 1969-1973, doi:10.1093/molbev/mss075 (2012).
- 10 Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC.Evol.Biol.* **7**, 214 (2007).
- 11 Markov, P. V. *et al.* Phylogeography and molecular epidemiology of hepatitis C virus genotype 2 in Africa. *Journal of General Virology* **90**, 2086-2096 (2009).
- 580 12 Bollyky, P. L., Rambaut, A., Grassly, N., Carman, W. F. & Holmes, E. C. Hepatitis B virus has a new world evolutionary origin. *Hepatology* **26**, 765-765 (1997).
- 13 Purdy, M. A. & Khudyakov, Y. E. Evolutionary history and population dynamics of hepatitis E virus. *Plos One* **5**, e14376, doi:10.1371/journal.pone.0014376 (2010).
- 585 14 Tanaka, Y. *et al.* Molecular tracing of Japan-indigenous hepatitis E viruses. *J Gen Virol* **87**, 949-954, doi:10.1099/vir.0.81661-0 (2006).
- 15 McCrone, J. T. & Luring, A. S. Genetic bottlenecks in intraspecies virus transmission. *Current opinion in virology* **28**, 20-25, doi:10.1016/j.coviro.2017.10.008 (2018).

- 16 Fitch, W. M., Leiter, J. M., Li, X. Q. & Palese, P. Positive Darwinian evolution in human influenza A viruses. *Proc Natl Acad Sci U S A* **88**, 4270-4274 (1991).
- 590 17 Leslie, A. J. *et al.* HIV evolution: CTL escape mutation and reversion after transmission. *Nat.Med.* **10**, 282-289 (2004).
- 18 Timm, J. *et al.* CD8 epitope escape and reversion in acute HCV infection. *J Exp Med* **200**, 1593-1604, doi:10.1084/jem.20041006 (2004).
- 19 Holland, J. J. Transitions in understanding of RNA viruses: a historical perspective.
595 *Curr Top Microbiol Immunol* **299**, 371-401 (2006).
- 20 Duffy, S., Shackelton, L. A. & Holmes, E. C. Rates of evolutionary change in viruses: patterns and determinants. *Nature reviews. Genetics* **9**, 267-276, doi:10.1038/nrg2323 (2008).
- 21 Pybus, O. G. & Rambaut, A. Evolutionary analysis of the dynamics of viral infectious
600 disease. *Nature reviews. Genetics* **10**, 540-550, doi:10.1038/nrg2583 (2009).
- 22 Sharp, P. M. *et al.* Origins and evolution of AIDS viruses: estimating the time-scale. *Biochem.Soc.Trans.* **28**, 275-282 (2000).
- 23 Sauter, D. *et al.* Tetherin-driven adaptation of Vpu and Nef function and the
605 evolution of pandemic and nonpandemic HIV-1 strains. *Cell host & microbe* **6**, 409-421, doi:10.1016/j.chom.2009.10.004 (2009).
- 24 Wain, L. V. *et al.* Adaptation of HIV-1 to its human host. *Mol Biol Evol* **24**, 1853-1860, doi:10.1093/molbev/msm110 (2007).
- 25 Woolhouse, M. & Gaunt, E. Ecological origins of novel human pathogens. *Crit Rev.Microbiol.* **33**, 231-242 (2007).
- 610 26 Wertheim, J. O. & Worobey, M. Dating the age of the SIV lineages that gave rise to HIV-1 and HIV-2. *PLoS computational biology* **5**, e1000377, doi:10.1371/journal.pcbi.1000377 (2009).
- 27 Worobey, M. *et al.* Island biogeography reveals the deep history of SIV. *Science.* **329**, 1487 (2010).
- 615 28 Galassi, F. M., Habicht, M. E. & Ruhli, F. J. Poliomyelitis in Ancient Egypt? *Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology* **38**, 375, doi:10.1007/s10072-016-2720-9 (2017).
- 29 Lukashev, A. N. & Vakulenko, Y. A. Molecular evolution of types in non-polio
620 enteroviruses. *J Gen Virol* **98**, 2968-2981, doi:10.1099/jgv.0.000966 (2017).
- 30 Olalde, I. *et al.* The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* **555**, 190-196, doi:10.1038/nature25738 (2018).
- 31 Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222-226, doi:10.1126/science.1224344 (2012).

- 625 32 Noonan, J. P. *et al.* Sequencing and analysis of Neanderthal genomic DNA. *Science* **314**, 1113-1118, doi:10.1126/science.1131412 (2006).
- 33 Toppinen, M. *et al.* Bones hold the key to DNA virus history and epidemiology. *Scientific reports* **5**, 17226, doi:10.1038/srep17226 (2015).
- 630 34 Reid, A. H., Fanning, T. G., Hultin, J. V. & Taubenberger, J. K. Origin and evolution of the 1918 "Spanish" influenza virus hemagglutinin gene. *Proc Natl Acad Sci U S A* **96**, 1651-1656 (1999).
- 35 Duggan, A. T. *et al.* 17(th) Century Variola Virus Reveals the Recent History of Smallpox. *Current biology : CB* **26**, 3407-3412, doi:10.1016/j.cub.2016.10.061 (2016).
- 635 36 Patterson Ross, Z. *et al.* The paradox of HBV evolution as revealed from a 16th century mummy. *PLoS Pathog* **14**, e1006750, doi:10.1371/journal.ppat.1006750 (2018).
- 37 Muhlemann, B. *et al.* Ancient hepatitis B viruses from the Bronze Age to the Medieval period. *Nature*, doi:10.1038/s41586-018-0097-z (2018).
- 640 38 Krause-Kyora, B. *et al.* Neolithic and Medieval virus genomes reveal complex evolution of Hepatitis B. *Elife* **7**, doi:10.7554/eLife.36666 (2018).
- 39 Muhlemann, B. *et al.* Ancient human parvovirus B19 in Eurasia reveals its long-term association with humans. *Proc Natl Acad Sci U S A*, doi:10.1073/pnas.1804921115 (2018).
- 645 40 Zhou, Y. & Holmes, E. C. Bayesian estimates of the evolutionary rate and age of hepatitis B virus. *J Mol Evol* **65**, 197-205, doi:10.1007/s00239-007-0054-1 (2007).
- 41 Kryukov, K., Ueda, M. T., Imanishi, T. & Nakagawa, S. Systematic survey of non-retroviral virus-like elements in eukaryotic genomes. *Virus Res*, doi:10.1016/j.virusres.2018.02.002 (2018).
- 650 42 Aiewsakun, P. & Katzourakis, A. Endogenous viruses: Connecting recent and ancient viral evolution. *Virology* **479-480**, 26-37, doi:10.1016/j.virol.2015.02.011 (2015).
- 43 Belyi, V. A., Levine, A. J. & Skalka, A. M. Sequences from ancestral single-stranded DNA viruses in vertebrate genomes: the parvoviridae and circoviridae are more than 40 to 50 million years old. *J Virol* **84**, 12458-12462, doi:10.1128/jvi.01789-10 (2010).
- 655 44 Horie, M. *et al.* Endogenous non-retroviral RNA virus elements in mammalian genomes. *Nature* **463**, 84-87, doi:10.1038/nature08695 (2010).
- 45 Katzourakis, A. & Gifford, R. J. Endogenous viral elements in animal genomes. *PLoS genetics* **6**, e1001191, doi:10.1371/journal.pgen.1001191 (2010).
- 660 46 Katzourakis, A., Tristem, M., Pybus, O. G. & Gifford, R. J. Discovery and analysis of the first endogenous lentivirus. *Proc Natl Acad Sci U S A* **104**, 6261-6265, doi:10.1073/pnas.0700471104 (2007).
- 47 Han, G. Z. & Worobey, M. Endogenous lentiviral elements in the weasel family (Mustelidae). *Mol Biol Evol* **29**, 2905-2908, doi:10.1093/molbev/mss126 (2012).

- 48 Gifford, R. J. *et al.* A transitional endogenous lentivirus from the genome of a basal primate and implications for lentivirus evolution. *Proc Natl Acad Sci U S A* **105**, 20362-20367, doi:10.1073/pnas.0807873105 (2008).
- 49 Hron, T., Farkasova, H., Padhi, A., Paces, J. & Elleder, D. Life History of the Oldest Lentivirus: Characterization of ELVgv Integrations in the Dermopteran Genome. *Mol Biol Evol* **33**, 2659-2669, doi:10.1093/molbev/msw149 (2016).
- 50 Taylor, D. J., Leach, R. W. & Bruenn, J. Filoviruses are ancient and integrated into mammalian genomes. *BMC evolutionary biology* **10**, 193, doi:10.1186/1471-2148-10-193 (2010).
- 51 Suh, A. *et al.* Early mesozoic coexistence of amniotes and hepadnaviridae. *PLoS genetics* **10**, e1004559, doi:10.1371/journal.pgen.1004559 (2014).
- 52 Sharp, P. M. & Simmonds, P. Evaluating the evidence for virus/host co-evolution. *Curr Opin Virol.* **1**, 436-441 (2011).
- 53 Katzourakis, A., Gifford, R. J., Tristem, M., Gilbert, M. T. & Pybus, O. G. Macroevolution of complex retroviruses. *Science* **325**, 1512, doi:10.1126/science.1174149 (2009).
- 54 Aiewsakun, P. & Katzourakis, A. Marine origin of retroviruses in the early Palaeozoic Era. *Nat Commun* **8**, 13954, doi:10.1038/ncomms13954 (2017).
- 55 Van Doorslaer, K. *et al.* Unique genome organization of non-mammalian papillomaviruses provides insights into the evolution of viral early proteins. *Virus evolution* **3**, vex027, doi:10.1093/ve/vex027 (2017).
- 56 Lauber, C. *et al.* Deciphering the Origin and Evolution of Hepatitis B Viruses by Means of a Family of Non-enveloped Fish Viruses. *Cell host & microbe* **22**, 387-399.e386, doi:10.1016/j.chom.2017.07.019 (2017).
- 57 Buck, C. B. *et al.* The Ancient Evolutionary History of Polyomaviruses. *PLoS Pathog* **12**, e1005574, doi:10.1371/journal.ppat.1005574 (2016).
- 58 Aiewsakun, P. & Katzourakis, A. Time-Dependent Rate Phenomenon in Viruses. *J Virol* **90**, 7184-7195, doi:10.1128/jvi.00593-16 (2016).
- 59 Kumar, S. & Subramanian, S. Mutation rates in mammalian genomes. *Proc Natl Acad Sci U S A* **99**, 803-808, doi:10.1073/pnas.022629899 (2002).
- 60 Ho, S. Y. *et al.* Time-dependent rates of molecular evolution. *Molecular ecology* **20**, 3087-3101, doi:10.1111/j.1365-294X.2011.05178.x (2011).
- 61 Wertheim, J. O. & Kosakovsky Pond, S. L. Purifying selection can obscure the ancient age of viral lineages. *Mol Biol Evol* **28**, 3355-3365, doi:10.1093/molbev/msr170 (2011).
- 62 Gilbert, C. & Feschotte, C. Genomic fossils calibrate the long-term evolution of hepadnaviruses. *PLoS.Biol.* **8**, e1000495 (2010).

- 700 63 Lythgoe, K. A., Gardner, A., Pybus, O. G. & Grove, J. Short-Sighted Virus Evolution and a Germline Hypothesis for Chronic Viral Infections. *Trends Microbiol* **25**, 336-348, doi:10.1016/j.tim.2017.03.003 (2017).
- 64 Lythgoe, K. A., Pellis, L. & Fraser, C. Is HIV short-sighted? Insights from a multistrain nested model. *Evolution* **67**, 2769-2782, doi:10.1111/evo.12166 (2013).
- 705 65 Duchene, S., Holmes, E. C. & Ho, S. Y. Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. *Proceedings. Biological sciences / The Royal Society* **281**, doi:10.1098/rspb.2014.0732 (2014).
- 66 Cooper, N., Jetz, W. & Freckleton, R. P. Phylogenetic comparative approaches for studying niche conservatism. *Journal of evolutionary biology* **23**, 2529-2539, doi:10.1111/j.1420-9101.2010.02144.x (2010).
- 710 67 Holt, R. D. Bringing the Hutchinsonian niche into the 21st century: ecological and evolutionary perspectives. *Proc Natl Acad Sci U S A* **106 Suppl 2**, 19659-19665, doi:10.1073/pnas.0905137106 (2009).
- 68 Price, T. Correlated evolution and independent contrasts. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **352**, 519-529, doi:10.1098/rstb.1997.0036 (1997).
- 715 69 Harvey, P. H. & Rambaut, A. Comparative analyses for adaptive radiations. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **355**, 1599-1605, doi:10.1098/rstb.2000.0721 (2000).
- 720 70 Koonin, E. V. & Gorbalenya, A. E. Evolution of RNA genomes: does the high mutation rate necessitate high rate of evolution of viral proteins? *J Mol Evol* **28**, 524-527 (1989).
- 71 Belshaw, R., Gardner, A., Rambaut, A. & Pybus, O. G. Pacing a small cage: mutation and RNA viruses. *Trends in ecology & evolution* **23**, 188-193, doi:10.1016/j.tree.2007.11.010 (2008).
- 725 72 Sheppard, S. K., Guttman, D. S. & Fitzgerald, J. R. Population genomics of bacterial host adaptation. *Nature reviews. Genetics*, doi:10.1038/s41576-018-0032-z (2018).
- 73 Sawyer, S. L., Emerman, M. & Malik, H. S. Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS biology* **2**, E275, doi:10.1371/journal.pbio.0020275 (2004).
- 730 74 Taubenberger, J. K. & Kash, J. C. Influenza virus evolution, host adaptation, and pandemic formation. *Cell host & microbe* **7**, 440-451, doi:10.1016/j.chom.2010.05.009 (2010).
- 75 Li, W. *et al.* Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *Embo j* **24**, 1634-1643, doi:10.1038/sj.emboj.7600640 (2005).
- 735 76 Urbanowicz, R. A. *et al.* Human Adaptation of Ebola Virus during the West African Outbreak. *Cell* **167**, 1079-1087.e1075, doi:10.1016/j.cell.2016.10.013 (2016).

- 77 Allison, A. B. *et al.* Host-specific parvovirus evolution in nature is recapitulated by in
740 vitro adaptation to different carnivore species. *PLoS Pathog* **10**, e1004475,
doi:10.1371/journal.ppat.1004475 (2014).
- 78 Bhatt, S. *et al.* The evolutionary dynamics of influenza A virus adaptation to
mammalian hosts. *Philosophical transactions of the Royal Society of London. Series*
B, Biological sciences **368**, 20120382, doi:10.1098/rstb.2012.0382 (2013).
- 79 Blanquart, F. *et al.* A transmission-virulence evolutionary trade-off explains
745 attenuation of HIV-1 in Uganda. *Elife* **5**, doi:10.7554/eLife.20492 (2016).
- 80 Clarke, D. K. *et al.* The red queen reigns in the kingdom of RNA viruses. *Proc Natl*
Acad Sci U S A **91**, 4821-4824 (1994).
- 81 Tedder, R. S., Bissett, S. L., Myers, R. & Ijaz, S. The 'Red Queen' dilemma--running to
750 stay in the same place: reflections on the evolutionary vector of HBV in humans.
Antivir Ther **18**, 489-496, doi:10.3851/imp2655 (2013).
- 82 Faria, N. R., Suchard, M. A., Rambaut, A., Streicker, D. G. & Lemey, P. Simultaneously
reconstructing viral cross-species transmission history and identifying the underlying
constraints. *Philosophical transactions of the Royal Society of London. Series B,*
Biological sciences **368**, 20120196, doi:10.1098/rstb.2012.0196 (2013).
- 755 83 Streicker, D. G. *et al.* Host phylogeny constrains cross-species emergence and
establishment of rabies virus in bats. *Science* **329**, 676-679,
doi:10.1126/science.1188836 (2010).
- 84 Longdon, B., Hadfield, J. D., Webster, C. L., Obbard, D. J. & Jiggins, F. M. Host
760 phylogeny determines viral persistence and replication in novel hosts. *PLoS Pathog*
7, e1002260, doi:10.1371/journal.ppat.1002260 (2011).
- 85 Davies, T. J. & Pedersen, A. B. Phylogeny and geography predict pathogen
community similarity in wild primates and humans. *Proceedings. Biological sciences /*
The Royal Society **275**, 1695-1701, doi:10.1098/rspb.2008.0284 (2008).
- 86 Nyamweya, S. *et al.* Comparing HIV-1 and HIV-2 infection: Lessons for viral
765 immunopathogenesis. *Rev Med Virol* **23**, 221-240, doi:10.1002/rmv.1739 (2013).
- 87 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, doi:10.1080/10635150290069940 (2002).
- 88 Frost, S. D. W., Magalis, B. R. & Kosakovsky Pond, S. L. Neutral Theory and Rapidly
770 Evolving Viral Pathogens. *Mol Biol Evol* **35**, 1348-1354, doi:10.1093/molbev/msy088
(2018).
- 89 Van Regenmortel, M. H. Virus species and virus identification: past and current
controversies. *Infection, genetics and evolution : journal of molecular epidemiology*
and evolutionary genetics in infectious diseases **7**, 133-144,
775 doi:10.1016/j.meegid.2006.04.002 (2007).

- 90 Simmonds, P. A clash of ideas - the varying uses of the 'species' term in virology and their utility for classifying viruses in metagenomic datasets. *J Gen Virol*, doi:10.1099/jgv.0.001010 (2018).
- 91 Kern, A. D. & Hahn, M. W. The Neutral Theory in Light of Natural Selection. *Mol Biol Evol* **35**, 1366-1371, doi:10.1093/molbev/msy092 (2018).
- 92 Keckesova, Z., Ylinen, L. M., Towers, G. J., Gifford, R. J. & Katzourakis, A. Identification of a RELIK orthologue in the European hare (*Lepus europaeus*) reveals a minimum age of 12 million years for the lagomorph lentiviruses. *Virology* **384**, 7-11, doi:10.1016/j.virol.2008.10.045 (2009).
- 93 Daugherty, M. D., Young, J. M., Kerns, J. A. & Malik, H. S. Rapid evolution of PARP genes suggests a broad role for ADP-ribosylation in host-virus conflicts. *PLoS genetics* **10**, e1004403, doi:10.1371/journal.pgen.1004403 (2014).
- 94 van der Lee, R., Wiel, L., van Dam, T. J. P. & Huynen, M. A. Genome-scale detection of positive selection in nine primates predicts human-virus evolutionary conflicts. *Nucleic Acids Res* **45**, 10634-10648, doi:10.1093/nar/gkx704 (2017).
- 95 Sawyer, S. L., Emerman, M. & Malik, H. S. Discordant evolution of the adjacent antiretroviral genes TRIM22 and TRIM5 in mammals. *PLoS Pathog* **3**, e197, doi:10.1371/journal.ppat.0030197 (2007).
- 96 Munk, C., Willemsen, A. & Bravo, I. G. An ancient history of gene duplications, fusions and losses in the evolution of APOBEC3 mutators in mammals. *BMC evolutionary biology* **12**, 71, doi:10.1186/1471-2148-12-71 (2012).