Review

Endogenous viruses: Connecting recent and ancient viral evolution

Pakorn Aiewsakun, Aris Katzourakis *

Department of Zoology, University of Oxford, Oxford OX1 3PS, UK

Article info

Keywords:
Endogenous viral element
Paleoviromology
Virus evolution
Genomics
Endogenous viruses

ABSTRACT

The rapid rates of viral evolution allow us to reconstruct the recent history of viruses in great detail. This feature, however, also results in rapid erosion of evolutionary signal within viral molecular data, impeding studies of their deep history. Thus, the further back in time, the less accurate the inference becomes. Furthermore, reconstructing complex histories of transmission can be challenging, especially where extinct viral lineages are concerned. This problem has been partially solved by the discovery of viruses embedded in host genomes, known as endogenous viral elements (EVEs). Some of these endogenous viruses are derived from ancient relatives of extant viruses, allowing us to better examine ancient viral host range, geographical distribution and transmission routes. Moreover, our knowledge of viral evolutionary timescales and rate dynamics has also been greatly improved by their discovery, thereby bridging the gap between recent and ancient viral evolution.

© 2015 Elsevier Inc. All rights reserved.

Contents

Introduction ................................................................. 26
EVE discovery ............................................................. 27
EVE discovery has vastly expanded viral host range .................................................. 27
Dating EVEs and generating robust viral evolutionary timescales ................................. 29
EVEs reveal ancient viral–host interaction and transmission routes ............................. 30
Discrepancy between viral evolutionary timescales and rates inferred from EVEs and modern-day viral sequences .............................................................. 31
Conclusion ................................................................. 33
Acknowledgments ......................................................... 34
References ................................................................. 34

Introduction

Viruses are arguably the fastest-evolving biological entity on this planet. Many of them evolve at a rate in the order of $10^{-3}$ substitutions per site per year (s/n/y) (Hanada et al., 2004; Jenkins et al., 2002; Sanjuán, 2012). This is approximately a million times faster than the rates of evolution of cellular host organisms, typically around $10^{-9}$ s/n/y (Kumar and Subramanian, 2002). This extraordinarily high rate of evolution allows viruses to escape from host immunity and rapidly adapt to a new host when they cross species. This unique feature also causes viral genomic sequences to rapidly change at the molecular level, accumulating information that can be used to reconstruct their evolution (Holmes, 2004). Combined with the large availability of molecular data of some extant viruses, this enables their recent history to be reconstructed in great detail. For example, a reconstruction of the spread of the global pandemic HIV strain revealed that it originated in Kinshasa, Democratic Republic of Congo, in the 1920s, and shortly thereafter spread via three other highly-populated cities through the transportation network before emerging as a global pandemic (Faria et al., 2014). Such a detailed epidemiological history can be reconstructed only because HIV evolves extremely rapidly and there are many molecular sequences available for this virus.

Although molecular data of extant viruses allow us to reconstruct their recent history to a high degree of resolution, such data are often insufficient to inform us about how they evolved and interacted with their hosts in deep time. While their rapid rate of evolution causes...
viral genomic sequences to accumulate evolutionary signal rapidly, sequences can carry only a finite amount of information, and therefore the signal will also be overwritten at a rapid rate. This places a temporal limit on how far back in time we can accurately reconstruct evolutionary history from modern-day viral sequences, likely no more than a few million years into the past. Furthermore, extant viral sequences cannot shed light on the history of extinct viral relatives. The combination of these effects greatly hinders the investigation of deep viral evolution. Because of this, together with the high extinction rate of viruses and viral sampling biases (e.g. towards human- and livestock-infecting viruses) analyses of contemporary viral sequences are likely to miss many historical viral cross-species transmission events, for example. In order to obtain a more complete picture of how viruses evolved and interacted with their hosts in the distant past, it seems that more than just molecular data of extant viruses is needed.

Viruses occasionally leave long-lasting imprints within their host genomes, known as endogenous viral elements (EVEs). The discovery of these viral imprints allows us to investigate the past history of viruses in more detail. Some viruses, namely retroviruses, enter host chromosomes as an obligate step during their life cycle, and other viruses do so less often by accident (Katzourakis and Gifford, 2010). EVEs result from the process of endogenisation, in which viral DNA copies integrate into host germ-line chromosomes, and in turn are vertically transmitted from parent to offspring. In cases where EVEs do not impose significant deleterious effects upon the hosts, their frequency might increase within the host population and lead to fixation. Once endogenised, most EVEs retain similarity to the ancestral exogenous virus for many millions of years as they evolve at the host rate of evolution, which is typically several orders of magnitude slower than a virus. These EVEs give researchers unprecedented opportunities to directly observe ancient viruses in their host genomes, and examine how they might have evolved and interacted with these hosts. The recent availability of a large number of cellular whole genomes and ongoing genomic screening has led to a steady increase in the number of ancient viral sequences of all types (Table 1), enabling us to better examine the deep history of all viral groups.

The discovery of EVEs has markedly improved our knowledge of viral natural history; for example, how viruses were distributed across geographical space and among host species millions of years ago. By combining such information with the knowledge of contemporary viral host range and geographical distribution, the history of viral transmission sources and routes can be reconstructed like never before. Furthermore, EVEs also have profoundly impacted our understanding of viral evolutionary timescales and rate dynamics. Analyses of EVEs strongly suggest that all fast-evolving viruses are far older than previously thought and they actually evolve strikingly slowly in the long term. In this review, by exploring various detailed examples, we discuss how EVEs have advanced and challenged our understanding of viral evolutionary dynamics, and filled in the gap between recent and ancient viral evolution, improving our knowledge of viral natural history.

**EVE discovery**

A common approach researchers use to mine EVEs is to bioinformatically search for genomic elements in cellular organisms that exhibit similarity to known extant viral sequences. Owing to the advances in bioinformatics techniques and genomic sequencing technology, numerous EVEs have been uncovered to date. Retroviruses are the only known group of viruses that enter host chromosomes as an obligate part of their life cycle, and thus are predisposed to become endogenous. Surprisingly, although retroviral germ-line integration is expected to be rare, endogenous retroviruses (ERVs) are very common and make up a large portion of many eukaryotic genomes. For instance, it has been estimated that ERVs form ~7–8% of our genome (International Human Genome Sequencing Consortium, 2001; Smit, 1999) with >98,000 retroviral fragments (Paces et al., 2004, 2002), comprising >31 families (Katzourakis and Tristem, 2005). Nevertheless, it has been found that not only can retroviruses become endogenous, but every known group of viruses is capable of endogenisation (Table 1).

The route to endogenisation of retroviruses is clear due to their replication strategy. However, the means by which other viruses have endogenous are far less obvious. It has been documented that various exogenous DNA viruses can occasionally undergo host chromosome integration through the process of non-homologous double-stranded DNA end-joining (e.g. hepadnavirus, Bill and Summers, 2004), non-homologous DNA recombination (e.g. adeno-associated DNA virus, Kotin et al., 1992; Urcelay et al., 1995; Young and Samulski, 2001), and telomeric homologous recombination (e.g. herpesviruses, Morissette and Flamand, 2010). Thus, finding that DNA viruses are capable of becoming endogenous is not entirely unforeseeable (Holmes, 2011).

The discovery of endogenous non-retroviral RNA viruses came as a surprise to researchers however, as it requires three unusual steps that do not usually occur in their life cycle: (i) conversion of genomic RNA to DNA, (ii) nucleus entry and (iii) chromosomal integration. Structural analyses have suggested that many non-retroviral RNA viruses directly hijack reverse transcriptase and integrase activity encoded by long interspersed nuclear elements for their endogenisation (Belty et al., 2010a; Horie et al., 2010; Katzourakis and Gifford, 2010). Furthermore, it has been proposed that a non-retroviral RNA virus, namely potato virus Y, might have become endogenous in grapevine via the process of non-homologous recombination between an RNA of the virus and that of a retrotransposon, with the subsequent product inserted into the host genome by retrotransposition (Tanne and Sela, 2005). It has also been demonstrated that recombination between ERVs and exogenous non-retroviral RNA virus can result in viral chromosomal integration (Geuking et al., 2009). At present, the precise molecular details of how many non-retroviral viruses become endogenous are still poorly characterised.

To date, numerous EVEs have been identified from diverse organisms (Table 1) and across various timescales (Table 2). They offer us opportunities to study and reconstruct the evolutionary history of the viruses that they are descended from, some of which are long extinct. EVEs have proved to be extremely valuable in expanding our understanding of deep viral natural history and long-term evolutionary dynamics, thereby closing the gap between recent and ancient viral evolution in our knowledge.

**EVE discovery has vastly expanded viral host range**

One of the most direct ways in which EVEs can inform us about viral natural history is that they can tell us about the diversity of the hosts with which viruses are capable of interacting. This information has greatly broadened our knowledge of viral host range, in particular when they are identified in lineages that are not known to harbour exogenous representatives. The host range of extant viruses has been extensively surveyed among humans, livestock and companion animals, and this knowledge has aided the development of viral control and prevention strategies (Calisher et al., 2006; Chantrey et al., 1999; Daniels et al., 2007; Laminger and Prinz, 2010). Nevertheless, this information pertains largely to extant viral lineages that have survived to be sampled.
Table 1
Distribution of exogenous and endogenous viruses among organisms.

<table>
<thead>
<tr>
<th>Group/type*</th>
<th>Family (group)</th>
<th>Host range</th>
<th>Exogenous viruses</th>
<th>Endogenous viruses (ref.s.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I/dsDNA virus</td>
<td>Herpesviridae</td>
<td>Eutherians, birds, and reptiles</td>
<td>Tarsier (Aswad and Katzourakis, 2014)</td>
<td></td>
</tr>
<tr>
<td>I/dsDNA virus</td>
<td>Polydnaviridae</td>
<td>Wasp</td>
<td>Wasps (Belle et al., 2002; Bézier et al., 2009a, 2009b; Deng et al., 2000; Soltz, 1999)</td>
<td></td>
</tr>
<tr>
<td>I/dsDNA virus</td>
<td>Phycodnaviridae</td>
<td>Marine filamentous brown algae</td>
<td>Marine filamentous brown algae (Delaroque et al., 1999)</td>
<td></td>
</tr>
<tr>
<td>II/ssDNA virus</td>
<td>Paroviridae</td>
<td>Eutherians, birds, crustaceans and insects</td>
<td>Eutherians, marsupials, birds, fishes, tunicates, crustaceans and flatworms (Belyi et al., 2010b; Kapoor et al., 2010; Katzourakis and Gifford, 2010; Liu et al., 2011b; Tang and Lightner, 2006; Thézé et al., 2014)</td>
<td></td>
</tr>
<tr>
<td>II/ssDNA virus</td>
<td>Circoviridae</td>
<td>Eutherians, birds, fishes, and insects</td>
<td>Eutherians, marsupials, reptiles, amphibians, lancelets, insects, crustaceans, nematodes, gastropods, hydrozoans, and giardia (Belyi et al., 2010b; Gilbert et al., 2014; Katzourakis and Gifford, 2010; Liu et al., 2011a; Thézé et al., 2014)</td>
<td></td>
</tr>
<tr>
<td>II/ssDNA virus</td>
<td>Geminiviridae</td>
<td>Plants, and fungi</td>
<td>Plants, fungi, and amoeba (Ashby et al., 1997; Bejarano et al., 1996; Lefevre et al., 2011; Liu et al., 2011a)</td>
<td></td>
</tr>
<tr>
<td>II/ssDNA virus</td>
<td>Nanoviridae</td>
<td>Plants</td>
<td>Amphibians, crustaceans, molluscs, placozoans, algae, amoeba, diatoms, blastocystics, and giardia (Liu et al., 2011a; Thézé et al., 2014)</td>
<td></td>
</tr>
<tr>
<td>II/dsRNA virus</td>
<td>Reoviridae</td>
<td>Mammals, birds, and insects</td>
<td>Insects (Katzourakis and Gifford, 2010)</td>
<td></td>
</tr>
<tr>
<td>II/dsRNA virus</td>
<td>Puttiviridae</td>
<td>Plants, fungi, and protozoa</td>
<td>Plants, insects, ticks, and amoeba (Chiba et al., 2011; Liu et al., 2010)</td>
<td></td>
</tr>
<tr>
<td>III/dsRNA virus</td>
<td>Chrysoviridae</td>
<td>Plants, and fungi</td>
<td>Insects (Crochu et al., 2004; Katzourakis and Gifford, 2010; Roiz et al., 2009)</td>
<td></td>
</tr>
<tr>
<td>III/dsRNA virus</td>
<td>Endornaviridae</td>
<td>Plants, fungi, and protozoa</td>
<td>Plants (Tanne and Sela, 2005)</td>
<td></td>
</tr>
<tr>
<td>IV(+) ssRNA virus</td>
<td>Potyviridae</td>
<td>Mammals, birds, and insects</td>
<td>Insects (Katzourakis and Gifford, 2010)</td>
<td></td>
</tr>
<tr>
<td>IV(–) ssRNA virus</td>
<td>Orthomyxoviridae</td>
<td>Vertebrates, and insects</td>
<td>Insects, and crustaceans (Ballinger et al., 2013; Katzourakis and Gifford, 2010; Thézé et al., 2014)</td>
<td></td>
</tr>
<tr>
<td>V(+) ssRNA virus</td>
<td>Bornaviridae</td>
<td>Mammals, and birds</td>
<td>Mammals, marsupials, reptile, fish, and insects (Belyi et al., 2010a; Gilbert et al., 2014; Horie et al., 2013, 2010; Katzourakis and Gifford, 2010)</td>
<td></td>
</tr>
<tr>
<td>V(–) ssRNA virus</td>
<td>Filoviridae</td>
<td>Primates, bats, and pigs</td>
<td>Bats, rodents, shews, tenrecs and marsupials (Belyi et al., 2010a; Katzourakis and Gifford, 2010; Taylor et al., 2014, 2011, 2010)</td>
<td></td>
</tr>
<tr>
<td>V(–) ssRNA virus</td>
<td>Rhabdoviridae</td>
<td>Mammals, birds, and insects</td>
<td>Insects, ticks, and plants (Chiba et al., 2011; Katzourakis and Gifford, 2010)</td>
<td></td>
</tr>
<tr>
<td>V(–) ssRNA virus</td>
<td>Nyamiviridae</td>
<td>Ticks and birds</td>
<td>Fish, and crustacean (Belyi et al., 2010a; Thézé et al., 2014)</td>
<td></td>
</tr>
<tr>
<td>VI/RT-RNA virus</td>
<td>Alpharetroviruses</td>
<td>Birds</td>
<td>Birds (Astrin et al., 1980; Bolisetty et al., 2012; Frisby et al., 1979; Gifford et al., 2005; Herniou et al., 1998)</td>
<td></td>
</tr>
<tr>
<td>VI/RT-RNA virus</td>
<td>Betaretroviruses</td>
<td>Boreoeutherians</td>
<td>Boreoeutherians, marsupials, and birds (Arnaud et al., 2008; Baba et al., 2011; Baille and Wilkins, 2001; Baille et al., 2004; Bolisetty et al., 2012; Ericsson et al., 2001; Gifford et al., 2005; Hayward et al., 2013; Herniou et al., 1998; Lavie et al., 2004; Mayer et al., 2013; Patience et al., 2001; Van der Kuyl, 2011)</td>
<td></td>
</tr>
<tr>
<td>VI/RT-RNA virus</td>
<td>Gammaparetroviruses</td>
<td>Terrestrial boreoeutherians, marsupials, and birds</td>
<td>Marine and terrestrial boreoeutherians, marsupials, birds, reptiles, and amphibians (Anai et al., 2012; Banumunisinghe et al., 2013; Bolisetty et al., 2012; Cui et al., 2012a, 2012b; Eleder et al., 2012; Fiebig et al., 2006; Hanger et al., 2000; Jaratlerdsiri et al., 2009; Lamere et al., 2009; Martin et al., 1999; Patience et al., 2001; Tarlinton et al., 2013; Tristem et al., 1996; Wang et al., 2013)</td>
<td></td>
</tr>
<tr>
<td>VI/RT-RNA virus</td>
<td>Epsilonretroviruses</td>
<td>Fish</td>
<td>Primates, amphibians, and fish (Basta et al., 2009; Brown et al., 2014; Kambol et al., 2003; Szmiele et al., 2011)</td>
<td></td>
</tr>
<tr>
<td>VI/RT-RNA virus</td>
<td>Lentiviruses</td>
<td>Simians, felines, cattle, equines, and small ruminants</td>
<td>Lagomorphs, prosimians and weasels (Cui and Holmes, 2012a; Gifford et al., 2008; Gilbert et al., 2009; Han and Worobey, 2012b; Katzourakis et al., 2007; Keckesova et al., 2009; van der Loo et al., 2010; Kunii et al., 2004; Staginnus and Richert-Pöggeler, 2006)</td>
<td></td>
</tr>
<tr>
<td>VI/RT-RNA virus</td>
<td>Spumaviruses</td>
<td>Boreoeutherians</td>
<td>Xenarthrans, afrotherians, and prosimians. (Han and Worobey, 2012b; Katzourakis et al., 2014, 2009)</td>
<td></td>
</tr>
<tr>
<td>VII/RT-DNA virus</td>
<td>Hepadnaviridae</td>
<td>Mammals, birds, and reptiles</td>
<td>Birds, and reptiles (Cui and Holmes, 2012b; Gilbert and Feschtote, 2010; Gilbert et al., 2014; Katzourakis and Gifford, 2010; Liu et al., 2012b; Suh et al., 2014, 2013)</td>
<td></td>
</tr>
<tr>
<td>VII/RT-DNA virus</td>
<td>Caulimoviridae</td>
<td>Plants</td>
<td>Plants (Bertsch et al., 2009; Gayral et al., 2008; Iskra-Caruana et al., 2010; Kunii et al., 2004; Staginnus and Richert-Pöggeler, 2006)</td>
<td></td>
</tr>
</tbody>
</table>

* ds = Double-stranded.
* ss = Single-stranded.
* + = Positive-sense.
* – = Negative-sense.
* RT = Reverse transcribing.
Furthermore, there is also a tendency for non-agriculturally and medically important viruses to be overlooked. As a result, it could thus be that the actual viral host range is much broader than surveillance of existing viruses alone could reveal. This bias is less prominent for EVEs however, since searches for them are typically done across all cellular genomes that are publically available. Indeed, EVEs have revealed that many viruses have a much broader host range than previously appreciated.

By combining exogenous and endogenous viral host ranges together, our knowledge of the distribution of viruses among organisms has been vastly expanded. For example, all exogenous nanovirus EVEs in various animal genomes (Liu et al., 2011a,b; Kapoor et al., 2010; Katzourakis and Gifford, 2014) provided conclusive evidence that, at least in the past, they could also infect animals. Likewise, the timescale of endogenous begomoviruses has been estimated to be between 1.24 and 4.85 Myr old by this method (Lefèvre et al., 2011). This approach, however, can be greatly biased by viral cross-species transmission, which can give the erroneous impression that the considered viruses are ancient, especially when transmissions take place between distantly related hosts.

A more robust approach, which is unaffected by viral cross-species transmission, is dating using EVE orthologs. Given the low probability of two independent EVE insertions occurring at the same site, the presence of EVEs in two or more host lineages at the same locus is to calculate the distance between potentially orthologous paralogs and converting it back to time by assuming the host neutral rate of evolution. This method has been used to calculate the (minimum) timescales of many EVEs. For instance, analyses of orthologous bird hepadnavirus EVEs suggested that these viruses must be > 74 Myr old, originating in the Upper Cretaceous period of the Mesozoic era (Suh et al., 2013). The identification of orthologous hepadnavirus EVEs in turtles pushed the origin of these viruses even further back, to the late Triassic period of the Mesozoic era > 207 Myr ago (Ma) (Suh et al., 2014). Similarly, analyses of orthologous rabbit-hare endogenous lentiviruses have shown that they are > 12 Myr old (Keckesova et al., 2009), providing evidence for the origin of modern-day lentiviruses in the Miocene epoch or even earlier.

Another means to derive a minimum age for an EVE relies on cases where EVEs have been duplicated within the genome. This can be done by estimating the genetic divergence between duplicated EVE paralogs and converting it back to time by assuming the host neutral rate of evolution. This method has been used to calculate minimum concrete lower-bound age estimate for the viral group that they are descended from. Several approaches can be used to date EVEs (Fig. 1 and Table 2). The simplest approach is to assume that the time to most recent common ancestor (tMRCA) of all exogenous and endogenous viruses is comparable to that of the hosts that carry them. For example, by examining the distribution of endogenous betaretroviruses across bird genomes, it has been proposed that they are likely > 100 million years (Myr) old (Bolisetty et al., 2012).

\[ T \text{MRCA} = \frac{1}{2} (T \text{ divergence}) \]

**Table 2**

<table>
<thead>
<tr>
<th>Group/type</th>
<th>Family (group)</th>
<th>Timescale of evolution (Myr old)</th>
<th>Method of timescale estimation</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>I (ssDNA virus)</td>
<td>Herpesviridae (Betaherpesvirinae)</td>
<td>&gt; 76</td>
<td>TMR-dating*</td>
<td>(Aswad and Katzourakis, 2014)</td>
</tr>
<tr>
<td>II (ssDNA virus)</td>
<td>Paroviridae</td>
<td>&gt; 40</td>
<td>Ortholog-dating</td>
<td>(Belyi et al., 2010b)</td>
</tr>
<tr>
<td>II (ssDNA virus)</td>
<td>Circoviridae</td>
<td>&gt; 50</td>
<td>Ortholog-dating</td>
<td>(Belyi et al., 2010b; Katzourakis and Gifford, 2010)</td>
</tr>
<tr>
<td>II (ssDNA virus)</td>
<td>Gemviridae (Begomoviruses)</td>
<td>&gt; 20–30</td>
<td>Ortholog-dating</td>
<td>(Lefèvre et al., 2011)</td>
</tr>
<tr>
<td>III (ssDNA virus)</td>
<td>Partitiviridae</td>
<td>&gt; 10</td>
<td>Ortholog-dating</td>
<td>(Chiba et al., 2011)</td>
</tr>
<tr>
<td>V (−) ssRNA virus</td>
<td>Bunyaviridae (Phleboviruses)</td>
<td>&gt; 8</td>
<td>Ortholog-dating</td>
<td>(Ballinger et al., 2013)</td>
</tr>
<tr>
<td>V (−) ssRNA virus</td>
<td>Bornaviridae</td>
<td>&gt; 93</td>
<td>Ortholog-dating</td>
<td>(Katzourakis and Gifford, 2010)</td>
</tr>
<tr>
<td>V (−) ssRNA virus</td>
<td>Filoaviridae</td>
<td>&gt; 30</td>
<td>Ortholog-dating</td>
<td>(Katzourakis and Gifford, 2010; Taylor et al., 2010)</td>
</tr>
<tr>
<td>V (−) ssRNA virus</td>
<td>Rhabdoviridae</td>
<td>&gt; 8</td>
<td>Ortholog-dating</td>
<td>(Chiba et al., 2011)</td>
</tr>
<tr>
<td>VI RT-RNA virus</td>
<td>Retroviralidae (Bartetaretroviruses)</td>
<td>&gt; 100</td>
<td>LTR-dating</td>
<td>(Cui et al., 2012a)</td>
</tr>
<tr>
<td>VI RT-RNA virus</td>
<td>Retroviridae (Gammeraretroviruses)</td>
<td>&gt; 65</td>
<td>LTR-dating &amp; ortholog-dating</td>
<td>(Brown et al., 2014; Sinzelle et al., 2011)</td>
</tr>
<tr>
<td>VI RT-RNA virus</td>
<td>Retroviridae (Epsilonretroviruses)</td>
<td>&gt; 40</td>
<td>Ortholog-dating</td>
<td>(Gifford et al., 2008)</td>
</tr>
<tr>
<td>VII RT-DNA virus</td>
<td>Lentiviridae</td>
<td>&gt; 14</td>
<td>Co-speciation</td>
<td>(Katzourakis et al., 2014, 2009)</td>
</tr>
<tr>
<td>VII RT-RNA virus</td>
<td>Retroviridae (Spumaviruses)</td>
<td>~ 100</td>
<td>Ortholog-dating</td>
<td>(Suh et al., 2014)</td>
</tr>
<tr>
<td>VII RT-DNA virus</td>
<td>Hepadnaviridae</td>
<td>&gt; 207</td>
<td>Ortholog-dating</td>
<td>(Suh et al., 2014)</td>
</tr>
<tr>
<td>VII RT-DNA virus</td>
<td>Caulimoviridae</td>
<td>&gt; 4.6</td>
<td>Ortholog-dating</td>
<td>(Gayral and Iskra-Caruana, 2009)</td>
</tr>
</tbody>
</table>

* TMR = Telomeric repeats. LTR = Long terminal repeats.
Fig. 1. Inferring viral evolutionary history though endogenous viral element (EVE) analyses and dating. EVE timescales are generally inferred directly from their host timescales or based on the neutral rate of evolution. Their age represents the lower-bound age estimate of the corresponding viral family. Several approaches can be used to date viral timescales through EVE analyses. The simplest approach is to infer that the timescale of the (combined exogenous and endogenous) viruses is comparable to that of the hosts carrying them, without explicit phylogenetic analyses (1st panel from left). If EVE orthologs are available, the basal diversification date of the hosts carrying orthologous EVEs can be interpreted as the minimum age of the EVEs, and thus the minimum age of the corresponding viral family (2nd panel). In the case that EVEs duplicated, the date of the event can be calculated from genetic difference observed between EVE paralogs, assuming that they accumulate genetic changes at the host neutral rate of evolution after the duplication. Likewise, this date can be interpreted as an EVE minimum age (3rd panel). In the specific case of endogenous retroviruses (ERVs), their integration event can be dated directly. During the process of retroviral endogenisation, two identical stretches of paralogous sequences are produced, known as long terminal repeats (LTRs). Again, by assuming that once becoming endogenous the paired LTRs accumulate genetic changes independently from one another at the host neutral rate of evolution, the timescale required for them to be as different as observed can be calculated (4th panel). This is essentially the EVE timescale and thus the minimum age of the associated viral family. Alternatively, an EVE integration event can be dated by assuming an EVE ancestral state. This method is more general than LTR dating as it can be applied to any EVEs, and not just ERVs. By estimating the genetic difference between the current EVEs and the ancestral state, and assuming that the EVE evolutionary rate is equal to the host neutral rate of evolution, EVE integration can be dated (5th panel). Furthermore, if phylogenetic analyses suggest that viruses have been stably co-speciating with their hosts, the timescale of viruses can then be directly inferred from that of their hosts (6th panel). Lastly, by integrating multiple sources of information together, such as viral–host co-evolutionary history, selection imprints on antiviral genes, phylogenetic information, and bio-geographical distribution and timescale, our knowledge of viral evolutionary history can be improved even further (7th panel).

ages of endogenous rabbit lentiviruses and sloth foamy viruses to be 7 Myr old (Katzourakis et al., 2007), and 39 Myr old (Katzourakis et al., 2009), respectively. A special case of duplication event can also be used to directly estimate the date of EVE integration. In the case of retroviruses, two identical stretches of paralogous sequences flanking the viral genome are produced during the process of integration, known as long terminal repeats (LTRs). After endogenisation, these two sequences evolve independently from one another. By tracing the process backwards, and assuming host neutral evolution, it is possible to estimate the ERV integration date. This method has been used to date many ERV integration events, revealing that most retroviral families are many millions of years old (Brown et al., 2014; Cui et al., 2012a; Katzourakis and Tristem, 2005; Lavie et al., 2004; Martins and Villesen, 2011; Sinzelle et al., 2011; Tristem, 2000). These methods are based on a reasonable assumption that EVE sequences evolve at the host neutral rate after the duplication process. Some studies applied the host neutral rate to calibrate phylogenies derived from multi-copy EVE families (Lefeuvre et al., 2011). However, such approaches may overestimate the true EVE timescale. This is because some of the variations observed among EVEs would have been generated under the viral rate of evolution, which is much faster than the host neutral rate. Thus, the application of the host neutral rate would lead to the dates of some nodes, particularly the deep ones, being overestimated.

Another approach to dating EVE integration events is to compare current EVEs against an ancestral viral state. This method is more general than the LTR-dating approach as it can be applied to any EVE, and not just ERVs. For example, it has been used to estimate the integration date of a tarsier endogenous beta-herpesvirus to be >76 Myr old (Aswad and Katzourakis, 2014). In brief, herpesviruses possess satellite telomeric repeats (TMRs). By knowing that the viral TMRs are identical to those of the host at the time of endogenisation, the endogenisation date of the virus can be calculated, assuming that any changes within the repeats occur at the host neutral rate of evolution. Indeed, this finding is consistent with the results from viral–host co-speciation analyses that suggested that, as a whole, beta-herpesviruses are ~100 Myr old (McGeoch et al., 2006, 1995).

A related approach that is based on knowledge of the ancestral viral state is to estimate EVE integration dates by using the frequency of stop codons contained in their sequences. Briefly, by assuming that ancestral viral sequences did not contain stop codons at the time of integration, and that the stop codons observed in EVEs are the result of neutral mutations, the timescale required to accumulate the observed number of stop codons can be estimated. To do so, nucleotide composition may be assumed (Belyi et al., 2010a), or extant viral sequences may be used as model sequences (Katzourakis et al., 2014). This method has been used to estimate the integration dates of primate and rodent endogenous bornaviruses to be ~19–40 Myr ago (Ma) (Belyi et al., 2010a) and those of aye–aye and cape golden mole endogenous foamy viruses to be ~35 and ~65 Ma, respectively (Katzourakis et al., 2014).

One application of integration dates obtained from EVEs is that they can provide robust minimum estimates of the age of their respective viral lineages. These dates can also be used as calibration points in downstream analysis in order to obtain timelines of viral evolution. For example, in a study of begomoviruses (Lefeuvre et al., 2011), an ancestral sequence of endogenous begomoviruses was inferred and its integration date was estimated. This information was then used to calibrate the timescale of the whole viral family to be ~20–30 Myr old. In a study of nudiviruses and baculoviruses, a phylogeny comprising both exogenous and endogenous viruses was reconstructed and calibrated with the age of baculovirus EVEs, revealing that these large dsDNA viruses are ~310 Myr old (Thézé et al., 2011). Together with the consistently ancient dates obtained from EVEs (Table 2), such analyses suggest that many, if not all, modern viral families have ancient origins, being many millions of years old.

EVEs reveal ancient viral–host interaction and transmission routes

In addition to providing a means of estimating viral evolutionary timescales, EVEs have also proved instrumental in expanding our knowledge of ancient viral–host interaction. For example, phylogenetic analyses of extant simian foamy viruses suggested
that they have an extremely stable history of co-speciation with their hosts for > 30 Myr (Switzer et al., 2005). The origin of this stable foamy virus–host co-speciation history was then pushed further back to the origin of eutherians, > 100 Ma, by analyses of endogenous foamy viruses identified in xenarthran (sloth) and afrotherian (cape golden mole) genomes (Katzourakis et al., 2014, 2009). This example illustrates how EVEs can supplement and expand our knowledge of long-term viral–host interaction, and how this information can be used to date, and therefore improve our knowledge of, viral origin. Similarly, our knowledge of the interaction between hepadnaviruses and their hosts has also been greatly improved by the identification of ancient endogenous snake hepadnaviruses. Analyses suggested that they are basal to the clade of bird and mammal hepadnaviral relatives (Gilbert et al., 2014). Unlike in the case of foamy viruses however, this phylogenetic relationship is inconsistent with that of their hosts (where birds are more closely related to reptiles than mammals), indicative of deep viral cross-species transfers between vertebrate hosts of different classes (Gilbert et al., 2014). Since no extant xenarthran and afrotherian foamy viruses and reptile hepadnaviruses have been identified to date, it would have been impossible to obtain such information through analyses of currently available extant viral sequences alone.

In some cases, analyses of EVEs can reveal previously unknown reservoir host species and/or ancient viral geographical transmission routes, particularly where the hosts are limited geographically. Filoviruses consist of three viral subgroups, namely Ebola-, Cueva-, and Marburg-viruses, and are known to be currently circulating primarily among bats and primates in Africa and East Asia. They are characterised by frequent cross-species transmissions, and occasionally jump from animals to humans, causing haemorrhagic fever outbreaks. Very little is known about their origin and reservoirs, with some evidence suggesting that bats are the most likely candidates (Kuzmin et al., 2010; Leroy et al., 2008, 2005; Peterson et al., 2004a; Pourrut et al., 2005; Towner et al., 2009). Molecular analyses showed that South American and Australian marsupial endogenous filoviruses form a paraphyletic clade with, and are basal to, the clade of extant filoviruses (Katzourakis and Gifford, 2010; Taylor et al., 2010). Not only did these findings reveal that filoviruses are more geographically dispersed than previously thought, but also suggested that the Old World extant filoviruses might have originated in the New World (Taylor et al., 2010). Further insight into the natural history and reservoirs of extant filoviruses came from the identification of filovirus EVEs in small mammals (Katzourakis and Gifford, 2010; Taylor et al., 2014, 2010). Analyses have shown that the clade of rodent filovirus EVEs, which are > 18 Myr old, is a sister of the clade of Ebola-/Cueva-viruses, but not Marburg-viruses, indicating that modern filoviruses originated > 18 Ma, in the early Miocene Epoch (Taylor et al., 2014). Furthermore, more basal than endogenous marsupial filoviruses is the clade of filovirus EVEs identified in small bats, rodents, insectivores, and tenrecs, which are > 30 Myr old (Katzourakis and Gifford, 2010; Taylor et al., 2010). Combined, it is also possible that, rather than marsupials, it was these small mammals that had been acting as filovirus reservoirs, repeatedly transmitting viruses to the New World marsupials and Old World primates. Indeed, this hypothesis is consistent with the prediction that, in addition to bats, insectivores and rodents are one of the most probable filovirus host reservoirs, based on a series of assumptions about the natural history of a likely filovirus host (Peterson et al., 2004b). This detailed example clearly demonstrates the power of EVEs to further elucidate the deep history of their modern-day viral relatives.

By combining viral and host phylogenetic information together with bio-geographical distribution and timescales, even more detailed hypotheses of viral transmission sources and routes can be formed. For instance, although not strongly supported, analyses have shown that an endogenous Malagasy lemur lentivirus is most closely related to simian lentiviruses (Gifford et al., 2008; Gilbert et al., 2009). This finding is consistent with a viral–host co-speciation scenario, which, if indeed is true, would posit that primate lentiviruses are > 85 Myr old (Gifford et al., 2008). Nevertheless, this pattern could also be a result of lentiviral cross-species transmission between prosimians and simians subsequent to their speciation (Gifford et al., 2008). Under a terrestrial mode of viral transmission, it would require that the transmission was dated back > 14 Ma (Gifford et al., 2008), the most recent time terrestrial primate virus invasion occurred (Tattersall, 2006). Alternatively, if the transmission occurred via an aerial vector species, then the transmission could have happened more recently (Gifford et al., 2008). Although it is still unclear which hypothesis is correct, this example nevertheless illustrates how tying together the ancient and recent viral geographical distribution and timescale allows us to form detailed hypotheses of past viral transmission routes.

Additionally, not only can EVEs inform us about their past history, but the models derived from their analyses may have some implications for the history of their hosts as well. For example, amidst the stable interaction between foamy viruses and their hosts, a few viral cross-species transmissions were identified, involving exogenous Chinese bat and endogenous Masayan aye–aye foamy viruses (Katzourakis et al., 2014). Analyses have shown that the two form their own deep clade, branching very deeply towards the root of the tree (Katzourakis et al., 2014). Based on these findings and assuming the most parsimonious scenario (i.e. the least number of cross-species transmission events), it was postulated that the ancestor of these two viruses jumped from the African continent to the Madagascar–India landmass ~ 100 Ma, and were split into two groups ~ 80–90 Ma by the Madagascar–India landmass separation. The virus that remained on Madagascar later gave rise to the aye–aye endogenous foamy virus, and the other that was transferred to Asia then gave rise to the Chinese bat foamy virus. To date, this is the only documented bat (Yinpterochiroptera) foamy virus. However, if further analyses show that Yinpterochiroptera foamy viruses co-speciate with their hosts, it would then suggest that the cross-species transmission giving rise to the bat foamy virus likely occurred before the radiation of Yinpterochiroptera bats, ~ 63–53 Ma (Teeling et al., 2005). However, the Indian landmass did not come into contact with Asia until ~ 55–35 Ma (Aitchison et al., 2007). Such observations would thus posit that the radiation of the bats did not occur in Asia, but rather on the Indian landmass while drifting in the Indian Ocean. This is another clear example highlighting the value of geographical history in elucidating viral evolution, and how this knowledge might affect our understanding of host natural history.

Discrepancy between viral evolutionary timescales and rates inferred from EVEs and modern-day viral sequences

Before the emergence of paleoviology, viral evolutionary timescales have generally been estimated from extant viral sequences, using phylogenetic techniques (Fig. 2). In fact, most of the viral evolutionary timescales estimated available today have been calculated based on contemporary viral sequences that were collected at different times (i.e. heterochronous molecular datasets). In brief, this method uses the differences in sampling dates and terminal branch lengths to infer the rate of viral evolution and extrapolates the rate to other parts of the tree to convert the units of the remaining branch lengths from molecular substitutions to time (Drummond and Rodrigo, 2000; Drummond et al., 2001, 2002; Rambaut, 2000). This process allows us to infer the split-dates of all nodes, and thus the tMRCA of all viruses under investigation. Absolute divergence times of two or more viral lineages (i.e. node heights), usually inferred from the host geographical split or speciation dates, are another source of rate calibration information (Bernard, 1994; Lemey et al., 2005a; McGeoch and Cook, 1994;
McGeoch and Gatherer, 2005; McGeoch et al., 1995; Sakaoka et al., 1994; Switzer et al., 2005; Worobey et al., 2010). Alternatively, if the viral rate of evolution has already been established elsewhere, instead of inferring the rate from the investigated sequences, such rates can be used to directly calibrate the timescale of the tree (Sharp and Li, 1988; Yokoyama et al., 1988). This is particularly useful when rate calibration information is not available. Since viral evolutionary timescales estimated from these methods are independent from those derived from EVEs, they can be used to validate one another. Surprisingly, it appears that many EVE-derived age estimates are in conflict with those calculated from extant viral sequences.

There are many examples of such timescale estimate discrepancies. For example, analyses of positive selection footprints on anti-retroviral genes consistently supported ages of > 10 Myr for simian lentiviruses (Compton and Emerman, 2013; Laguette et al., 2012; Lim et al., 2012, 2010; Malfavon-Borja et al., 2013), in line with an analysis of an endogenous prosimian lentivirus, suggesting that lentiviruses likely have been infecting primates for > 14 Myr (Gifford et al., 2008). In contrast, analyses of extant viral sequences suggested that they are only a few hundred years (Sharp and Li, 1988; Sharp et al., 2000) to tens of thousands of years old (Worobey et al., 2010). The timescales of mammalian and bird hepatitisviruses estimated from EVEs and extant viral sequences are also very different. While analyses of EVEs suggested that they are > 74 Myr old (Suh et al., 2013), their origin was dated to be ~ 30,000 (Orito et al., 1989) to ~ 125,000 (Van Hemert et al., 2011) years ago by analyses of modern-day viruses. Likewise, while analyses of modern lentiviruses suggested that the tMRCA for modern lentiviruses ranges between 7,000 and 155,000 years (Carroll et al., 2013; Negredo et al., 2011; Suzuki and Gojobori, 1997), filovirus EVE analyses identified in hamster and vole genomes, suggested that they are > 18 Myr old (Taylor et al., 2014). Similar observations have also been made for circoviruses and phleboviruses. Their timescales of evolution were estimated to be < 500 years (Bird et al., 2007; Firth et al., 2009; Lam et al., 2013) whereas the discovery of their EVEs suggested that they are > 50 Myr (Belyi et al., 2010b; Katzourakis and Gifford, 2010) and > 8 Myr (Ballinger et al., 2013). Overall, analyses of extant viruses seem to suggest that

some modern viral families are young, whereas paleovirotical analyses strongly support that almost all of these viral groups are of great antiquity, being millions of years old (Table 2).

So, which timescale estimates are closer to the true dates: the ones estimated from EVEs or the ones derived from the rate estimates of modern-day viruses? Since none of them are self-evidently wrong, it is difficult to answer the question with certainty. However, EVE timescales have generally been estimated based on the host neutral rate of evolution or directly inferred from the host timescales. Therefore, if the deep viral origins derived from EVEs were to be wrong and the shallow dates were to be closer to the truth, it would posit that (i) the host neutral rate must be several orders of magnitude higher than what we currently know, and (ii) the hosts have come to existence very recently (e.g. < 1 Myr). These are extremely unlikely given that the host timescales and neutral rate of evolution are very well-established. Consequently, it is much more likely that the EVE-derived viral evolutionary timescales represent more accurate date estimates of viral origin. These observations have contributed to a radical shift in our knowledge of viral natural history, strongly suggesting that all viruses have a long co-evolutionary history with their hosts.

These findings have also challenged our understanding of viral evolutionary dynamics. Analyses of heterochronous extant viral sequences (collected over a time period of < 500 years) often suggested that dsDNA viruses and many RNA viruses evolve extremely quickly. Their rates are often reported to be in the order of 10^-7–10^-5 s/n/y (Davis et al., 2005; Duffy and Holmes, 2008; Jenkins et al., 2002; Lemey et al., 2006; Sanjuán, 2012; Shackelton and Holmes, 2006; Shackelton et al., 2005; Sharp and Li, 1988). As a result, these viruses are often said to be ‘fast-evolving’ viruses. Surprisingly, many ancient EVEs related to these ‘fast-evolving’ viruses appear to exhibit similarity higher than expected given their known rate of evolution (Feschotte and Gilbert, 2012). For instance, analyses of heterochronous extant lentiviral sequences have firmly established that their evolutionary rates are in the order of 10^-8 s/n/y (Aulicino et al., 2007; Bello et al., 2010; Biek et al., 2003; Lemey et al., 2005b). Assuming that the rate is homogenous and constant at a value of 1 × 10^-4 s/n/y, it is expected that, say, after a million years of evolution, each nucleotide position would have changed a hundred times on average, being fully saturated by nucleotide substitutions. This should render the ancestral and the evolved sequences to be no more similar than two random sequences. Despite this however, ancient endogenous lentiviruses that are millions of years old can still be easily recognised as relatives of modern lentiviruses (Cui and Holmes, 2012a; Gifford et al., 2008; Gilbert et al., 2009; Han and Worobey, 2012a; Keckesova et al., 2009; van der Loo et al., 2009). These findings mean that, in contrast to the conventional wisdom, lentiviruses evolve strikingly slowly in the long term. The same story can also be told for other ‘fast-evolving’ RNA and ssDNA viruses.

At the same time, viruses that appear to evolve slowly in the long term also seem to evolve extremely rapidly over short timescales. For example, the evolutionary rates of classic ‘slow-evolving’ dsDNA viruses, foamy viruses and deltaretroviruses (Duffy et al., 2008) are often reported to be in the range of ~ 10^-7–10^-9 s/n/y, typically estimated over timescales of millions of years under the viral–host co-speciation assumption (Bernard, 1994; Lemey et al., 2005a; McGeoch and Cook, 1994; McGeoch and Gatherer, 2005; McGeoch et al., 1995; Sakaoka et al., 1994; Switzer et al., 2005). In the case that the rates of evolution of viruses have already been established elsewhere, instead of inferring the rate from the investigated sequences, such rates can be directly used to calibrate the timescale of the tree (Sharp and Li, 1988; Yokoyama et al., 1988). The rates calibrated from serially sampled viral sequences are generally short-term rate estimates, and those estimated from viral speciation dates are normally long-term rate estimates.
3–4 orders of magnitudes higher than the typical reported rates of evolution of ‘slow-evolving’ foamy viruses. Deltaretroviruses have also been observed to evolve rapidly over short timescales. For instance, following a population of bovine leukaemia viruses in a single sheep for 1.5 years, Willems et al. (1993) estimated their rate of evolution to be as high as $\sim 3.43 \times 10^{-4}$ s/n/y. Analyses of human T-lymphotropic viruses type II circulating among intravenous drug users also suggested that their rate is $\sim 2.7 \times 10^{-4}$ s/n/y (Salemi et al., 1999; Vandamme et al., 2000). These rate estimates are 2–4 orders of magnitude higher than their conventional long-term evolutionary rate estimates. Combined with the results from paleovirological analyses, it seems that all documented viruses so far appear to evolve extremely rapidly in the short-term but much slower in the long-term.

One possibility is that the observed rate change genuinely reflects changes in viral biology. Indeed, it has been found that the error rates of viral replication vary greatly among viruses (Drake et al., 1998; Sanjuán, 2012). Therefore, it is conceivable that the fidelity of viral polymerases might have varied with time, and thus might explain some portion of the discovered short-term and long-term rate discrepancy (Gilbert and Feschotte, 2010). Changes in viral transmission mode and/or replication speed have also been put forward as candidate explanations for the observed rate variation in viruses; the higher the replication and transmission speed, the higher the chance of mutations being accumulated, and thus the higher the rate of evolution (Hanada et al., 2004; Salemi et al., 1999; Vandamme et al., 2000). Nevertheless, for these factors to account for the observed rate change, not only must they alter the rate of viral evolution, but the effect must be systematically associated with time. Furthermore, given the generality of the observations, this would also posit that the biology of all viruses has changed in a similar way. Combined, although it is still currently unknown to what extent changes in viral biology contribute to the observed changes in the rates of viral evolution, we believe that they are unlikely to be the main contributors.

Large-scale meta-analyses have shown that, in fact, viral evolutionary rate estimates are negatively correlated with the timescale of rate measurement (Duchêne et al., 2014). This relationship also stands within RNA and DNA viral groups, by combining rates across a range of measurement techniques (Duchêne et al., 2014). This ‘time-dependent rate phenomenon’ (TDRP) has been observed to be very common in nature, and was first noticed in cellular organisms (see review in Ho et al. (2011)). We suggest that all of the observed viral rate discrepancies are in fact parts of this phenomenon, where rate estimates have been sampled from two extreme ends of the measurement timescale spectrum. That is, evolutionary rates of extant exogenous viruses tend to be calculated over extremely short timescales ( $< 100$ years) and thus are extremely high, whereas EVE analyses typically span much longer timescales ( $> 1$ Myr) and hence yield rate estimates that are much lower.

Many explanations have been proposed to explain the TDRP (Fig. 3; see Ho et al. (2011) for review). One explanation is that short-term rate estimates tend to include transient deleterious variations that have not yet been purged by purifying selection, and thus appear to be very high (Ho et al., 2011; Jenkins et al., 2002; Loogváli et al., 2009). Viral adaptive changes resulting from positive selection or viral-host evolutionary arm races are also lineage-specific, and hence will tend to manifest themselves as recent changes, inevitably elevating the short-term rate of evolution. Similarly, molecular variations that result from sequencing errors will also appear as recent changes, and because of this, may contribute to the observed high short-term rates to some degree (Clark and Whittam, 1992; Ho et al., 2007, 2005; Johnson and Slatkin, 2008). These processes could all lead to an elevation of short-term rates. In contrast, over longer timescales, failing to properly account for superimposed substitutions and substitution saturation can cause underestimation of rates (Duchêne et al., 2014; Emerson, 2007; Garcia-Moreno, 2004; Ho et al., 2011; Sullivan and Joyce, 2005). Indeed, using inappropriate substitution models can lead to sequence divergence, and thereby evolutionary rate underestimation (Sullivan and Joyce, 2005). This bias can be alleviated by statistical model-selection procedures, in other words choosing the best model available. Nevertheless, given model simplifications and complex interactions between different parts of viral genomes, such as gene overlapping and epistasis, it is likely that even the best model currently available will fail to properly account for these complex patterns of substitution. Failing to account for rate heterogeneity among sites and the tendency for underestimating the true dates of calibrating nodes are other factors that have the potential to contribute to the TDRP (Ho et al., 2011; Soubrier et al., 2012). Alternatively, it is also possible that the observed TDRP can be caused by our tendency to use fast-evolving molecules for short-term evolutionary studies and slowly-evolving molecules for long-term evolutionary studies (Ho et al., 2011). At present, the extent to which each factor contributes to the TDRP is unclear. We believe that, to bridge the gap between long-term and short-term viral natural history and evolutionary dynamics, understanding how the TDRP is generated will be absolutely crucial.

**Conclusion**

The availability of modern-day viral sequences and their rapid rate of evolution allow us to decipher recent viral natural history to a high degree of resolution. In contrast, our understanding of how ancient viruses evolved and interacted with their hosts still
remains largely fragmented, especially where extinct viral lineages are concerned. The discovery of EVEs has partially solved this problem, providing us opportunities to study and reconstruct the natural history of the ancient viruses that they are descended from. EVEs have been instrumental in revealing aspects of viral sources and transmission routes, and enhancing our understanding of viral–host interaction. Furthermore, while conventional wisdom posits that some modern viral families are young, analyses of EVEs have instead provided conclusive evidence that almost all of them have ancient origins, being several millions of years old. These findings have subsequently led to the establishment of the short-term/long-term viral rate discrepancy, and in turn the TDRP as a possible explanation for this discrepancy in viruses. Improving our understanding of long-term and short-term viral rate dynamics and connecting the two will undoubtedly be a crucial step towards a better understanding of viral evolution.

Acknowledgments

PA is funded by the Royal Thai Government and AK is funded by the Royal Society of London.

References


