Virology ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

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

Article history:
Received 4 December 2014
Returned to author for revisions
15 December 2014
Accepted 4 February 2015

Keywords:
Endogenous viral element
Paleovirology
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.

Contents

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

* Correspondence to: Department of Zoology, University of Oxford, South Parks Road, Oxford, UK OX1 3PS. Tel.: +44 1865 281 847.
E-mail address: aris.katzourakis@zoo.ox.ac.uk (A. Katzourakis).

http://dx.doi.org/10.1016/j.virol.2015.02.011
0042-6822/

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 endogenised 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 (Belyi 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.

Please cite this article as: Aiewsakun, P., Katzourakis, A., Endogenous viruses: Connecting recent and ancient viral evolution. Virology (2015), http://dx.doi.org/10.1016/j.virol.2015.02.011