Virology 446 (2013) 276–282

Contents lists available at ScienceDirect

Virology

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

Selectively maintained paleoviruses in Holarctic water fleas reveal an ancient origin for phleboviruses

Matthew J. Ballinger [a,*], Jeremy A. Bruenn [a], Alexey A. Kotov [b], Derek J. Taylor [a]

[a] Department of Biological Sciences, The State University of New York at Buffalo, Buffalo, NY 14260, USA
[b] A.N. Severtsov Institute of Ecology and Evolution, Leninsky Prospect 33, Moscow 119071, Russia

ARTICLE INFO

Keywords:
Paleovirology
Daphnia
Virus co-option
Purifying selection
Gene duplication
Pseudogenes

ABSTRACT

The ecological model, *Daphnia pulex* (Cladocera: *Daphniidae*), is broadly distributed in Holarctic freshwater habitats and has been the subject of multidisciplinary study for over half a century, but never has a natural RNA virus infection been reported in daphnids. Here we report on a group of paleoviruses related to RNA dependent RNA polymerase in the genome of *D. pulex*. Phylogenetic analysis suggests that these paleoviruses are derived from a viral lineage within the genus *Phlebovirus*. Comparison of the genomic sequences flanking individual paleoviruses reveal that some are orthologous viral insertions having been present in the common ancestor of the *D. pulex* species complex, which is millions of years old. Still, we detected some sites that have the signature of purifying selection. In contrast, other paleoviruses in this group seem to be unique to specific host lineages and even contain undisrupted open reading frames, suggesting either more recent acquisition, or selective maintenance.

Introduction

Our understanding of the deep evolutionary history of viruses and their host interactions is restricted by the absence of fossil records for viruses and the limited evolutionary reach of molecular clock estimates when based on the rapid nucleotide substitution rates of viruses (Holmes, 2003). Paleovirology, the study of endogenous viral elements (EVEs), including endogenous retroviruses (ERVs) and non-retroviral integrated RNA viruses (NIRVs), offers a needed glimpse into the deeper history of virus–host interactions. Analyses of individual ERVs and NIRVs in fungal, arthropod and vertebrate genomes have extended the minimum age of RNA viruses and their host associations, e.g. Filoviruses (Taylor et al., 2010), *Bornaviruses* (Horie et al., 2010) and *Lentiviruses* (Gifford et al., 2008), revealed genes of viral origin that have been selectively maintained or co-opted in host genomes (Malik et al., 2000; Mi et al., 2000; Taylor and Bruenn, 2009; Katzourakis and Gifford, 2010; Taylor et al., 2011; Fort et al., 2012), explained how viruses lacking reverse-transcriptase have integrated into DNA genomes (Katzourakis and Gifford, 2010; Ballinger et al., 2012), and supported a case of coevolution between viruses and hosts with a modified nuclear genetic code (Taylor et al., 2013). Yet, few attempts have been made to unravel the evolutionary history of what appear to be virally-derived gene families in eukaryotic genomes. In CTG-clade yeast, a family of totivirus capsid-like NIRVs are tandemly structured, suggesting host duplication, and some copies are expressed as proteins (Taylor and Bruenn, 2009; Taylor et al., 2013). Several arthropod genomes harbor a dozen or more NIRVs showing sequence similarity to a single viral gene (Katzourakis and Gifford, 2010; Fort et al., 2012), but many of these sequences are divergent pseudogenes, and little evidence remains to differentiate between an origin as a single ancient integration followed by duplication within the host, or multiple integrations of relatively closely related exogenous viruses. The ability to support one hypothesis over the other with statistical bioinformatics methods is further hindered by the incomplete representation of the ancient virosphere.

Phleboviruses (Bunyaviridae) are segmented, single-stranded, negative and ambisense RNA viruses. Bunyavirids display an extensive host range across vertebrates, invertebrates and plants. However, the arthropod-borne members of the genus *Phlebovirus*, appear to be curiously limited to unrelated blood-sucking dipterans and ticks. It is presently unknown if this distribution in arthropods is a sampling bias, or a real association with hematophagous arthropods. The arthropod-borne phleboviruses are subdivided into the Sandfly group, vectored by sand flies of the genera *Phlebotomus* and *Lutzomyia,* and mosquitoes (Tesh, 1988), and the Uukuniemi group, vectored by ticks (Saikku and Brummerk, 1973; Eley and Nuttall, 1984; Palacios et al., 2013). Some viruses in the Sandfly group, e.g. Toscana virus, have been shown to establish persistent, vertically transmitted infection in their vectors (Tesh and Modi, 1987; Bilsel et al., 1988). The bunyavirus genome is distributed across three segments (L, M, and S) and codes for up to five genes: a nucleoprotein (*N*), a glycoprotein (Gn-Gc), an RNA-dependent RNA polymerase (RdRp) and one or two nonstructural proteins, NSs and NSm, named

* Corresponding author.
E-mail address: ballinge@buffalo.edu (M.J. Ballinger).

0042-6822/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.virol.2013.07.032

for the segment on which they are coded (Elliott, 1990). As with most RNA viruses, little is known about the age of the phleboviruses. Several authors have used molecular clock and coalescent dating methods calibrated with empirically determined mutation rates to estimate the ages of extant lineages, e.g. severe fever with thrombocytopenia syndrome virus (Lam et al., 2013) and Rift Valley fever virus (Bird et al., 2007), and they have concluded that these viruses diversified within the past 200 years. The age of the genus has not been estimated, though there is some evidence to suggest that it is indeed ancient. For example, a clade of BEL retroelements (Cer7, 13 and 14) in *Caenorhabditis elegans* contains an envelope glycoprotein gene with unique ancestry compared to related BEL elements; it seems to have been co-opted from a phlebovirus glycoprotein (Malik et al., 2000), though the date of its acquisition is not known. Uukuniemi virus-like nucleoprotein and RdRp sequences have also been identified in the genome of the tick, *Ixodes scapularis*, but their age also remains unresolved (Katzourakis and Gifford, 2010).

The ecological model species, *Daphnia pulex* (Crustacea: Cladocera: *Daphniidae*), is a key member of freshwater communities worldwide. Cladocerans occupy a crucial trophic position, and have been extensively studied in ecotoxicology (Baird et al., 1991; Barata et al., 1998). Fungal and bacterial parasites of *Daphnia* are known, and their effects and coevolution have been described in detail (Ebert, 2008), while examples of natural viral infection in *Daphnia* are conspicuously absent. Indeed, until recently there have been no well-described cases of any viral infection in daphnids; the newly discovered ssDNA virus infection in *Daphnia mendotae* and *Daphnia retrocurva* populations represents the only confirmed case (Hewson, et al. 2013), to our knowledge, since the description of Chloriridovirus-like (Iridoviridae) particles in one population of *Daphnia magna* (Bergoin, et al. 1984). The sparsity of reported *Daphnia*-virus associations is especially surprising in light of the habitat overlap between *Daphnia* and one of the most widely-studied arbovirus vectors, mosquitoes.

The genus *Daphnia* is believed to have emerged at least 145 Mya (Colbourne and Hebert, 1996; Kotov and Taylor, 2011). Here, we describe an unexpected association between phleboviruses and daphnids based on the presence of phlebovirus RdRp-like NIRVs (PRNs) in the *D. pulex* genome. The PRNs form a monophyletic clade sister to the Uukuniemi virus group, firmly establishing this crustacean-infecting virus as a phlebovirus. We set the minimum age of this virus-host association as at least as ancient as the *D. pulex* species complex. We also consider evidence for co-option of these sequences by the host and we discuss the possible implications for the evolutionary history of the PRNs as well as this virus-host association.

Results

We discovered and assembled a dataset of 21 PRNs (Supplementary Table S1) by performing BLAST (Altschul et al., 1990) tBLASTn searches to the Joint Genome Institute's *D. pulex* (strain: The Chosen One [TCO]) genome assembly (Colbourne et al., 2011) available on GenBank using phlebovirus RdRp amino acid sequences as queries. An amino acid alignment of the 21 PRNs and four representative phleboviruses is available as Supplementary Fig. S1. In this alignment, the conserved motifs of exogenous phlebovirus RdRps (Muller et al., 1994) are labeled and are present in many of the PRNs. We used MAFFT 7 (Katoh and Standley, 2013) to create a codon alignment of the PRNs alone, and screened them for evidence of recombination using the single break point (SBP) and Genetic Algorithm Recombination Detection (GARD) methods (Pond et al., 2006) on the Datamonkey webserver (Delport et al., 2010). A single recombination breakpoint was identified with high support by SBP, but the GARD analysis identified a second breakpoint. We also performed a similar analysis for the same RdRp region of exogenous viruses in the Uukuniemi virus group and found significant support for at least one breakpoint, though KH tests did not support topological incongruence at any specific position (Supplementary Fig. S2).

Our phylogenetic analysis places all 21 PRNs in a well-supported monophyletic clade within the genus *Phlebovirus* (Fig. 1A). With regard to the evolutionary history of the PRNs, there are at least two interpretations of this tree topology. The first is that monophyly is the natural result of a single ancestral host integration event, which has subsequently undergone extensive duplication (illustrated by Fig. 1B), while the second is that this clade is the result of multiple, independent integrations of related, exogenous phleboviruses (Fig. 1C). We attempted to disqualify the multiple integration scenario by identifying homologous flanking sequences at the PRN sites in the *D. pulex* TCO genome, but the majority of such flanking regions lack evidence for homology.

To determine whether these *D. pulex* TCO PRNs are present in other species of *Daphnia*, we performed PCR on taxa throughout the genus. We consistently found taxa within the *D. pulex* species complex to be positive for PRNs, while those outside of this complex were universally negative (Fig. 2). We also blasted TCO PRNs to the *Daphnia pulicaria* hybrid genomic sequence database (dubbed the rejected one [TRO]) available at http://wfleabase.org/blast and found distinct matches for most PRNs that are present in *D. pulex* TCO (Supplementary Fig. S3). We tested whether individual PRNs are orthologous or lineage-specific by PCR amplifying from the PRN flanking regions. Importantly, some PRNs, e.g. PRN5 (Fig. 3), were successfully amplified from the flanking sequences across all *D. pulex* species complex taxa screened, indicating that such viral inserts were present in the common ancestor of the studied species. Other copies amplified only from reactions in which the primers targeted the interior PRN sequences. Still others, e.g. PRN1, could not be amplified in any lineage other than *D. pulex* TCO, regardless of the primer target.

We next performed tests for site-specific detection of selection in the *D. pulex* TCO PRN sequences using the Fast Unconstrained Bayesian Approximation (FUBAR) (Murrell et al., 2013) and the Mixed Effects Model of Evolution (MEME) (Murrell et al., 2012) methods of the HyPhy package (Pond et al., 2005) available on the Datamonkey webserver (Delport et al., 2010). We removed PRN1 from the alignments prior to performing these analyses as our inability to amplify any orthologous copy of this PRN suggested a more recent, independent integration. Of the 855 codon positions in the alignment, FUBAR identified 367 positions under purifying selection (43% of sites) and none under diversifying selection (Fig. 4), though MEME did detect evidence of episodic diversifying selection at 25 sites (Supplementary Fig. S4). We also performed these analyses on orthologous and putatively orthologous (i.e. those that amplified from internally-primed sequences only, but showed very high sequence identity) interspecific PRNs and found reduced, though still significant evidence of purifying selection at specific sites in these sequences (Fig. 4). We found no statistical support for pervasive or episodic diversifying selection between interspecific PRN sequences.

Discussion

Our results support the hypothesis that the association of phleboviruses with blood-sucking arthropods is a sampling artifact. Daphniid crustaceans are neither bloodsucking nor closely related to dipterans or to ticks. Yet, we find evidence of prior association of a unique clade of phleboviruses with *Daphnia* in the form of at least 21 paleoviruses. It is unknown if the phleboviruses that infected *Daphnia* were transmitted to other animals—certainly