Remarkable Diversity of Endogenous Viruses in a Crustacean Genome

Julien Théze1, Sébastien Leclercq1,2, Bouziane Moumen1, Richard Cordaux1, and Clément Gilbert1,*

1Université de Poitiers, UMR CNRS 7267 Ecologie et Biologie des Interactions, Equipe Ecologie Evolution Symbiose, Poitiers, France
2State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China
*Corresponding author: E-mail: clement.gilbert@univ-poitiers.fr.

Accepted: July 25, 2014

Data deposition: The sequences generated in this study have been deposited in GenBank under the accession numbers KM034067–KM034115.

Abstract

Recent studies in paleovirology have uncovered myriads of endogenous viral elements (EVEs) integrated in the genome of their eukaryotic hosts. These fragments result from endogenization, that is, integration of the viral genome into the host germline genome followed by vertical inheritance. So far, most studies have used a virus-centered approach, whereby endogenous copies of a particular group of viruses were searched in all available sequenced genomes. Here, we follow a host-centered approach whereby the genome of a given species is comprehensively screened for the presence of EVEs using all available complete viral genomes as queries. Our analyses revealed that 54 EVEs corresponding to 10 different viral lineages belonging to 5 viral families (Bunyaviridae, Circoviridae, Parvoviridae, and Totiviridae) and one viral order (Mononegavirales) became endogenized in the genome of the isopod crustacean Armadillidium vulgare. We show that viral endogenization occurred recurrently during the evolution of isopods and that A. vulgare viral lineages were involved in multiple host switches that took place between widely divergent taxa. Furthermore, 30 A. vulgare EVEs have uninterrupted open reading frames, suggesting they result from recent endogenization of viruses likely to be currently infecting isopod populations. Overall, our work shows that isopods have been and are still infected by a large variety of viruses. It also extends the host range of several families of viruses and brings new insights into their evolution. More generally, our results underline the power of paleovirology in characterizing the viral diversity currently infecting eukaryotic taxa.

Key words: paleovirology, viral diversity, viral host range, isopod crustacean, Armadillidium vulgare.

Introduction

Endogenous viral elements (EVEs) are pieces of (or entire) viral genomes that became integrated in the germline genome of their hosts and inherited vertically over host generations (Katzourakis and Gifford 2010; Feschotte and Gilbert 2012). The bulk of known EVEs are retroviruses (Belshaw et al. 2004; Katzourakis et al. 2009). These viruses encode proteins involved in the integration of the viral genome into host chromosomes (the integrated virus is then called a provirus), a step that is necessary to the completion of the retroviral replication cycle. In addition to integrating into host somatic genomes upon infections, retroviruses have recurrently colonized their host’s germline genomes during evolution, spawning dozens of thousands of EVEs that now make up a substantial fraction of vertebrate genomes (e.g., 8% of the human genome). Replication of all other known viruses does not go through a proviral stage, and as such, integration in their host genome of viruses other than retroviruses is rare. This explains why only very few nonretroviral EVEs had been reported until recently (Horie et al. 2010; Katzourakis and Gifford 2010). However, thorough searches of the numerous whole-genome sequences produced at an increasing pace during the last 5 years have led to the discovery of many nonretroviral EVEs in the genomes of a large diversity of eukaryotes (Katzourakis and Gifford 2010; Liu et al. 2010, 2011a, 2011b; Chiba et al. 2011). A major conclusion of these studies is that any type of virus can become endogenous via accidental integration into its host germline genome and that much like retroviruses, some families of nonreverse transcribing viruses have been endogenized recurrently over long periods of time and sometimes independently in various taxa.

The discovery and analysis of recently uncovered nonretroviral EVEs has yielded new insights on both host biology and virus evolution. Unlike endogenous retroviruses, nonretroviral EVEs are typically few in a given genome, such that their impact on global eukaryote genome architecture is unlikely
to be profound. However, some studies have suggested that some nonretroviral EVEs copies have been domesticated and are now fulfilling a new, beneficial function that may have been linked to immunity against circulating viruses (Maori et al. 2007; Flegel 2009; Katzourakis and Gifford 2010; Taylor et al. 2011; Aswad and Katzourakis 2012; Ballinger et al. 2012; Fort et al. 2012). In some instances, it is even clear that nonretroviral EVE domestication has been the basis of a new function that is crucial to the development of the host (Herniou et al. 2013). In terms of viral evolution, the study of EVEs has revealed that many currently circulating families of viruses are much older than previously thought (Katzourakis et al. 2009; Belyi et al. 2010; Gilbert and Feschotte 2010; Théze et al. 2011) and that viral long-term substitution rates calculated using EVE sequences are orders of magnitude slower than rates inferred using only extant viruses (Gilbert and Feschotte 2010). Another interesting outcome of EVE discovery is that it often extends the known host range of viral families, and it may help to uncover species likely to be reservoirs of circulating zoonotic viruses (Taylor et al. 2010).

So far, most studies of nonretroviral EVEs have conducted searches of endogenous copies of a specific virus or group of viruses in a large number of whole-genome sequences. Here, we adopted a host-centered approach in which we thoroughly searched all EVEs present in the genome of one species—the common pillbug Armadillidium vulgare (Crustacea, Isopoda). We show that a large diversity of viruses was endogenized during the evolution of crustacean isopods and that close relatives of many of these viruses are likely to be still circulating in A. vulgare populations. Our analysis shows that in addition to increasing the known host range of viruses, searching for EVEs can also yield numerous information on the viral flora infecting a particular taxonomic lineage.

**Materials and Methods**

**Genome Screening**

The EVEs from the isopod crustacean A. vulgare were identified from data generated as part of the ongoing A. vulgare genome project in our laboratory. Briefly, total genomic DNA was extracted from a single A. vulgare individual. A paired-end library with approximately 370 bp inserts was prepared and sequenced on an Illumina HiSeq2000. Reads were filtered with FastQC and assembled using the SOAP de novo software version 1.05. The best assembly (obtained with a k-mer size of 49) was composed of approximately 3.5 million scaffolds and contigs totaling approximately 1.5 Gb (at 40× average coverage). An in-house pipeline of in silico analyses was developed to search for EVEs in the A. vulgare genome sequences. We first constructed a comprehensive library of all nonretroviral virus nucleotide sequences available in public databases (GenBank and EMBL), including genomes from small RNA and DNA viruses, as well as large dsDNA viruses, often not considered in paleovirology screenings. This library was used as a query to perform TBLASTX searches (Altschul et al. 1997) (e value ≤ 1) to screen for A. vulgare genome sequences exhibiting similarity to virus sequences. This analysis aimed at selecting a subset of A. vulgare genome sequences that matched with viral sequences before further in-depth analyses. Then, we performed reciprocal BLASTX searches (Altschul et al. 1997) using the selected subset of A. vulgare genome sequences as queries to screen for homologous coding sequences in the whole set of nonredundant protein sequences of the National Center for Biotechnology Information (NCBI) database. Armadillidium vulgare genome sequences were considered of viral origin if they unambiguously matched viral proteins in the reciprocal best hits (e value ≤ 0.001).

From these sequences, putative viral open reading frames were inferred through a combination of automated alignments, using the exonerate program (Slater and Birney 2005) and manual editing, based on the most closely related exogenous viral sequences in the nonredundant protein database. For each putative resulting A. vulgare viral peptides, we retrieved the function and predicted the taxonomic assignation by comparison to the best reciprocal BLASTX hit viral proteins.

**Polymerase Chain Reaction Validation of Endogenization**

We verified by polymerase chain reaction (PCR) and Sanger sequencing that the viral genome fragments we uncovered computationally in the A. vulgare whole-genome sequences were endogenous and did not result from contamination by exogenous viruses that would have been coextracted together with A. vulgare genomic DNA. For this, we designed primer pairs for eight EVEs loci representing five of the six viral groups identified in this study (supplementary table S2, Supplementary Material online). For each pair, one primer was anchored in the upstream or downstream region flanking the EVE locus, and the other primer was anchored within the EVE sequence. We also used these primers to screen for presence/absence of orthologous EVEs in two other isopod crustacean species (A. nasatum and Cylisticus convexus). PCRs were conducted using the following temperature cycling: Initial denaturation at 94 °C for 5 min, followed by 30 cycles of denaturation at 94 °C for 30 s, annealing at 54–58 °C (depending on the primer set) for 30 s, and elongation at 72 °C for 1 min, ending with a 10-min elongation step at 72 °C. Purified PCR products were directly sequenced using ABI BigDye sequencing mix (1.4 ml template PCR product, 0.4 ml BigDye, 2 ml manufacturer supplied buffer, 0.3 ml primer, and 6 ml H2O). Sequencing reactions were ethanol precipitated and run on an ABI 3730 sequencer. Presence and sequences of all selected A. vulgare EVEs were confirmed as predicted in silico. Altogether, we conclude that our
final set of 54 EVEs sequences is highly unlikely to result from contamination by exogenous viruses.

**Phylogenetic Analyses**

Using ClustalOmega (Sievers et al. 2011) and manual edition, multiple amino acid (aa) alignments were performed for each inferred *A. vulgare* EVE peptide, including closely related exogenous and endogenous viral proteins resulting from the reciprocal BLASTX analysis and closely related proteins of representative virus species recognized by the International Committee on Taxonomy of Viruses (ICTV; King et al. 2011).

In addition to *A. vulgare* EVEs, we included several previously unknown EVEs uncovered in other taxa as a result of the reciprocal BLASTX. These EVEs correspond to proteins of viral origin that have been annotated as host genes and are therefore present in the nonredundant protein database of NCBI because they are devoid of nonsense mutation.

Maximum likelihood (ML) inferences were performed on each multiple aa alignment using RAxML (Stamatakis 2006) with the substitution model and parameters WAG + G + I. Support for nodes in ML trees were obtained from 100 non-parametric bootstrap iterations, and the root of ML trees was determined by midpoint rooting.

Based on the trees we obtained using this approach, we tentatively propose that some of the EVEs we have discovered in the *A. vulgare* genome may be considered new viral species, genus, or family. Basically, when an EVE is as or more distant from its closest known virus a than another known virus b is from the virus a, we consider that the EVE could be given the same taxonomic rank as viruses a and b. We acknowledge that this criterion alone may not be sufficient for the ICTV to follow our proposition and to recognize and give a name to these various new EVE lineages. However, we believe that the various taxonomical aspects we address in the article are important for the reader to fully appreciate the breadth of our results and the extent to which a paleovirological study can further our understanding of the viral fauna infecting a given eukaryotic host species.

**Nucleotide Sequence Accession Numbers**

The nucleotide sequences produced in this study have been deposited in GenBank under the accession numbers KM034067–KM034115 (see supplementary table S1, Supplementary Material online, for details).

**Results**

**EVE Diversity in the *A. vulgare* Genome**

To identify EVEs in whole-genome sequences of the isopod crustacean *A. vulgare*, we first performed a TBLASTX search using all complete viral genomes publicly available in GenBank and EMBL (January 2014) as queries (n = 2,048). We then used all hits resulting from this search (n = 10,727) as queries to carry out a reciprocal BLASTX on the nonredundant protein database of the NCBI. This approach yielded a total of 54 *A. vulgare* genome sequences of unambiguous viral origin, ranging from 42 to 588 aa in length (average = 173 aa) and showing 46–78% aa similarity (average = 58%) to their most closely related exogenous viral protein sequences (fig. 1 and supplementary table S1, Supplementary Material online). The 54 EVEs were assigned to four different families (*Bunyaviridae*, *Circoviridae*, *Paroviridae*, and *Totiviridae*) and one order (*Mononegavirales*), representing three of the seven types of viral genomes (−ssRNA, dsRNA, and ssDNA). Among those families/order, the *Circoviridae* and *Totiviridae* families are not currently reported by the ICTV (King et al. 2011) to infect arthropods (but see e.g., Wu et al. 2010; Rosario et al. 2012). The diversity of EVEs discovered in the *A. vulgare* genome is remarkable in that most previously published paleovirology studies have reported less than 20 EVEs and/or less than 4 different viral families in a given genome (Feschotte and Gilbert 2012).

It is noteworthy that we also detected two fragments of 42 and 43 aa showing, respectively, 73% and 74% similarity to the wsv209 gene of the Shrimp white spot syndrome virus (WSSV, Nimaviridae family of dsDNA viruses). We could not reconstruct a phylogeny for these two *A. vulgare* nimavirus-like fragments, because wsv209 is only present in the Shrimp WSSV (Yang et al. 2001). Here in fact, we cannot firmly assess whether the presence of two wsv209 homologs in the *A. vulgare* genome results from viral endogenousization or whether this gene is present in the WSSV genome because of a horizontal transfer that would have taken place from an unsequenced host to the WSSV.

Several lines of evidence indicate that the virus-like sequences we found in the *A. vulgare* whole-genome sequences are integrated in the isopod genome and do not correspond to circulating viruses, the genome of which could have been coextracted and sequenced with that of the host. First, we used standard protocols to extract and sequence DNA, which do not involve any reverse transcription step and therefore could not allow the sequencing of RNA viruses. Second, if the virus-like sequences uncovered in the *A. vulgare* genomic contigs/scaffolds were from circulating viruses, one would expect to find entire viral genomes or at least fragments containing several viral ORFs. Yet, all EVEs characterized in this study correspond to partial single ORFs (except *A. vulgare* sequence nb 15, which contains two partial ORFs), and for each family/order, we found only one or two different ORFs but never did we recover a complete genome (fig. 1 and supplementary table S1, Supplementary Material online). Finally, our PCR tests using primers anchored in EVEs, and their flanking regions yielded positive products of the expected size for all seven EVEs we screened in *A. vulgare*, which encompass five of the six virus families/order we uncovered computationally.
FIG. 1.—Mapping of the 54 Armadillidium vulgare EVEs on representative virus genomes. Light gray rectangles represent virus genes with their genomic positions, including conserved domains in dark gray. Numbered black lines represent A. vulgare EVEs. Numbers below these black lines indicate the position (continued)
In terms of the mechanisms underlying endogenization, it has been proposed that integration of viral sequences into host genomes could be facilitated by transposable element encoded enzymes (Geuking et al. 2009; Taylor and Bruenn 2009; Horie et al. 2010), DNA repair mechanisms (Bill and Summers 2004) or viral proteins, (Belyi et al. 2010). Inspection of the regions flanking the EVEs reported in this study did not reveal any obvious target site duplications, which are molecular signatures typically generated upon retrotransposition. Thus, although the A. vulgare genome contains a large proportion of transposable elements (including various non-LTR and LTR retrotransposons; unpublished), our data indicate that these elements are unlikely to have been involved in endogenization. Finally, we did not detect any similarity between the various EVE flanking regions, suggesting that most or all EVEs result from multiple independent

---

**FIG. 1.**—Continued

of A. vulgare EVEs on the above viral genome. Numbers in red correspond to EVEs that were PCR amplified and sequenced. Numbers in blue correspond to EVEs for which recognizable flanking regions were identified (4: 3'-flanking region contains a host gene of unknown function approximately 700 bp away from the EVE, 19 and 42: 3'-flanking regions contain nonlong terminal repeat retrotransposon-like reverse transcriptases approximately 8,600 bp and 250 bp away from the EVEs, respectively). Dots and vertical bars represent stop codons and frameshifts found in A. vulgare EVEs, respectively. The Rift Valley fever virus (NC_014395, NC_014396, NC_014397; Bunyaviridae), Midway virus (NC_012702; Mononegavirales), Armigeres subalbatus virus SaX06-AK20 (NC_014609; Totiviridae), Dragonfly orbiculatus virus (NC_023854; Circoviridae like), and infectious hypodermal and hematopoietic necrosis virus (NC_002190; Parvoviridae) were the representative virus genomes used for the mapping.

**FIG. 2.**—Phylogeny of the Bunyaviridae family. The tree was obtained from ML analysis of the RNA-dependent RNA polymerase multiple aa alignment, including Armadillidium vulgare EVE sequences, sequences of closely related exogenous and endogenous viruses, and representative virus species of the Bunyaviridae family. ML nonparametric bootstrap values (100 replicates) are indicated at each node. Associated host vectors are indicated by branch colors and silhouettes at the bottom.
events of endogenization rather than from segmental duplication of one or a few EVE loci.

**Phylogeny and Evolution of A. vulgare EVEs**

To better understand the evolutionary history of A. vulgare EVEs, we aligned these sequences together with representative viral species of each viral family/order recognized by the ICTV and with other closely related exogenous and endogenous viruses, and representative virus species of the Mononegavirales order. ML nonparametric bootstrap values (100 replicates) are indicated at each node. Associated hosts are indicated by branch colors and silhouettes at the bottom.

**Bunyaviridae**

Within bunyaviruses, the seven A. vulgare EVEs fall in two distinct lineages. The first one (A. vulgare sequence number 2 in fig. 2 and number 5 in supplementary fig. S1, Supplementary Material online) falls within an extended Phlebovirus genus that in addition to well-characterized exogenous viruses (e.g., Uukuniemi, Rift Valley, and Toscana viruses; Palacios et al. 2013) has recently been proposed to include endogenous viruses from three water flea species (Daphnia genus, Cladocera, Crustacea) (Ballinger et al. 2013). The second one (fig. 2; A. vulgare sequence numbers 1, 3, 4, and 6–8) forms a well-supported clade (bootstrap = 100) sister to a large clade including the Phlebovirus and Tenuivirus genera and unclassified Bunyaviridae viruses infecting insects (Cumuto and Gouleako viruses; Marklewitz et al. 2011; Auguste et al. 2014) (bootstrap = 75).

**Mononegavirales**

The seven A. vulgare mononegaviruses also fall into two distantly related lineages. The first one (fig. 3; A. vulgare sequences 11–16 and supplementary fig. S2, Supplementary Material online; A. vulgare sequences 10 and 15) forms a mildly supported clade (bootstrap = 63) sister to a large clade including the Phlebovirus and Tenuivirus genera and unclassified Bunyaviridae viruses infecting insects (Cumuto and Gouleako viruses; Marklewitz et al. 2011; Auguste et al. 2014) (bootstrap = 75).
the closest well-characterized family (Bornaviridae), the nyaviruses + Soybean cyst nematode midway virus + A. vulgare EVEs probably deserves recognition as an entirely new family that has been tentatively named Nyamiviridae by Kuhn et al. (2013). The remaining A. vulgare sequence (fig. 3; sequence 9) forms a well-supported clade together with closely related EVEs newly discovered in various zebrafish BAC clones (CU694452.16, CR759863.7, CR846102.12, BX323595.8, BX855590.3, BX248129.5, CR847797.8, CU207259.10, and BX284614.8), and a more distantly related EVE previously found in the Aedes mosquito (bootstrap = 100; Katzourakis and Gifford 2010).

**Totiviridae**

The three A. vulgare totiviruses (fig. 4; sequences 24–26 and supplementary fig. S3, Supplementary Material online; sequences 17–23) form a clade (bootstrap = 76) with the Penaeid shrimp infectious myonecrosis virus, Armigeres subalbatus virus, Drosophila melanogaster totivirus, and Omono River, which belong to an unassigned Totiviridae genus of arthropod-infecting viruses (suggested Artivirus genus; Poulos et al. 2006; Wu et al. 2010; Zhai et al. 2010; Isawa et al. 2011). Given that these various viruses are all infecting arthropod hosts and that their grouping is relatively well supported, we believe A. vulgare totiviruses should be included in the Artivirus genus.

**Circoviridae**

The 20 A. vulgare circovirus-like sequences (fig. 5; 27–46) seemingly form a monophyletic clade that appears to be most closely related to the unclassified Dragonfly orbiculatus virus (Rosario et al. 2012), though this position is not well supported.
supported (bootstrap = 51). Overall, the phylogeny indicates that *A. vulgare* circovirus-like EVEs likely belong to a new lineage of circular Rep-dependent ssDNA viruses (CRESS-DNA according to Rosario et al. 2012) distantly related to the *Circoviridae* and *Nanoviridae* families. In addition, the circovirus-like sequences we found in two mollusc species (the oyster *Crassostrea gigas* and *Lottia gigantea*) likely correspond to new nonplant *Nanoviridae* lineages.

**Paroviridae**

The eight *A. vulgare* parovirus sequences (fig. 6; sequences 47–54) fall into a large clade that includes the penaeid shrimp infectious hypodermal and hematopoietic necrosis virus (Bonami et al. 1990), its endogenous relative found in the shrimp *Penaeus monodon* genome (Tang and Lightner 2006), and densoviruses from *Aedes* mosquitoes (Boublik et al. 1994; Sivaram et al. 2009). Given that these exogenous viruses all belong to the *Brevidensovirus* genus and that their grouping with *A. vulgare* EVEs is relatively well supported (bootstrap = 83), *A. vulgare* EVEs should be considered new arthropod-infecting lineages of brevidensoviruses.

**Discussion**

Paleovirology and metagenomics studies are gradually changing our global understanding of viral evolution, which has long been heavily based on pathogenic viruses isolated from model species or from species of economical or medical interest. The ongoing characterization of myriads of new viral genes and genomes from various environments (including host genomes) is revealing that the viral diversity is extremely large and that current families of viruses are much older and have a larger host tropism than previously thought (Katzourakis and Gifford 2010; Rosario and Breitbart 2011; Feschotte and Gilbert 2012; Roossinck 2012). Only one virus belonging to the *Iridoviridae* family (large dsDNA) is currently known to infect isopod crustaceans (Cole and Morris 1980; Federici 1980). This virus was detected and isolated because of the iridescent blue color of infected individuals, which is due to...
paracrystalline arrays formed by virions inside parasitized cells (Lupetti et al. 2013). Together our results indicate that isopod crustaceans have been additionally exposed to a remarkable diversity of viruses. Not only the A. vulgare EVEs belong to or are related to five major groups of known viruses, but within four of these groups (bunyaviruses, Mononegavirales, totiviruses, and parvoviruses), the EVEs also belong to at least two distinct lineages. Each of these lineages is most closely related to different known exogenous or endogenous viruses or to newly discovered endogenous viruses that are distantly related to each other and separated by long branches. In total, we have uncovered no less than ten new viral lineages, some of which may be new genera (one in the Bunyaviridae, one in the Densovirus, and one in the Mononegavirales) or new families (one in the Mononegavirales and one family of CRESS-DNA viruses).

Though more information on the morphology/replication of these viruses are required to propose names for these new lineages, we believe it is important to place new viral genes or genomes discovered by paleovirology or metagenomics studies in a comprehensive phylogenetic and taxonomical framework (King et al. 2011). This will facilitate their inclusion in future classifications based on mechanistic studies of closely related viruses that we anticipate are awaiting discovery, fostering our global understanding of viral diversity and evolution.

Interestingly, 30 of the 54 A. vulgare EVEs are devoid of nonsense mutations (fig. 1 and supplementary table S1, Supplementary Material online) suggesting either that they have a recent origin or that they may be ancient but would have been exapted and evolved under purifying selection since endogenization (e.g., Lavialle et al. 2013). To test which of

---

**Fig. 6.**—Paroviridae family phylogeny. The tree was obtained from ML analysis of the nonstructural protein 1 multiple aa alignment, including Armadillidium vulgare EVE sequences, sequences of closely related exogenous and endogenous viruses and of representative species of the Paroviridae family. ML nonparametric bootstrap values (100 replicates) are indicated at each node. Associated hosts are indicated by branch colors and silhouettes at the bottom.
these two scenarios was the most likely, we carried out cross-
species PCR screenings of eight EVE loci, all from distinct viral
lineages, in a species closely related to *A. vulgare* (*A. nasatum*)
and a more distantly related one (*C. convexus*) (Michel-Salzat
and Bouchon 2000). None of the loci amplified in both
species, seven amplified in three *A. vulgare* individuals (the
one for which we sequenced the genome and two others),
and the last one amplified only in the *A. vulgare* individual
for which we sequenced the genome. We acknowledge
the fact that the absence of amplification for some of
these loci may be due to insufficient sequence conservation
for the PCR primers to bind properly and not necessarily
imply absence of the orthologous EVE locus in other species.
However, together with the fact that we find intact
*A. vulgare* EVEs in each of the six viral groups and that
exaptation of nonretro EVEs appears to be relatively
rare (Kobayashi et al. 2011; but see also Taylor et al.
2011; Ballinger et al. 2012; Fort et al. 2012), we believe
these results tend to support recent or even ongoing
degenzonalization of at least some of these EVEs. This further
suggests that the very exogenous viruses that produced
these EVEs or closely related ones may still circulate in
extant populations of *A. vulgare* and other isopod
crustaceans.

Our study is a clear illustration of the potential of the paleo-
virology approach in furthering our understanding of viruses
and host–virus interactions. In addition to the large diversity
of EVEs we uncovered in *A. vulgare*, our comprehensive
mapping of host lineages on the viral trees reveals multiple
incongruences between host and viral phylgenies (figs. 2–6;
supplementary figs. S1–S3, Supplementary Material online).
In fact, for each of the five viral groups, crustacean viruses
are clearly polyphyletic. Other clear examples of polyphyly
include insect paroviruses, insect and arachnid bunyaviruses,
insect rhabdoviruses, and fungus totiviruses. This pattern
suggests that the evolution of the various viral families
found in *A. vulgare* has involved multiple host switches be-
tween widely divergent taxa, which likely took place over a
large evolutionary timescale. Furthermore, we extend the
known host range of the six viral groups to isopod crusta-
ceans, as well as to molluscs for the family Nanovirinae
(fig. 5) and flatworms for the Densovirinae (fig. 6). The finding
of phleboviruses in *A. vulgare* is intriguing given that all
known viruses from this genus were isolated from various
mammalian species (including humans in which they are the
cause of various diseases) and from arthropod vectors such as
ticks, sandflies, and mosquitoes (Elliott and Brennan 2014).
Whether the *A. vulgare* phleboviruses have developed a strat-
egy allowing them to replicate and persist only in a single
(arthropod) host or whether isopods, that are cosmopolitan
and often in contact with humans, can act as vectors of
these viruses and transmit them to mammals is an interesting
question that deserves further investigation.

Finally, the large viral diversity we uncovered in *A. vulgare*
using a paleovirology approach is surprisingly as high as that
detected in recent metagenomic studies of exogenous viruses
targeting a given host species (e.g., Granberg et al. 2013;
Rosario et al. 2014). Given the fact that endogenization of
nonretro EVEs results from accidental—thus relatively
infrequent—recombination between host and viral genomes,
we speculate that *A. vulgare* EVEs represent only a fraction
of the total viral diversity that is circulating in these animals
today. These findings provide a solid ground justifying the
inclusion of viruses in studies considering eukaryotic organisms
as holobionts, that is, organisms harboring and interacting
with a diverse microbial community (Zilber-Rosenberg
and Rosenberg 2008), which have so far focused only on commu-
nities of bacteria. We anticipate that in addition to the role
of viruses in pathogenesis and their likely involvement in hor-
izontal transfer of DNA (Piskurek and Okada 2007; Routh
et al. 2012; Gilbert et al. 2014), such studies will uncover
a wide range of novel types of interactions with their hosts
(Roossinck 2011), further emphasizing the major influence
of viruses on the evolution of their hosts.

**Supplementary Material**

Supplementary data S1–S8, tables S1 and S2, and figure
S1–S3 are available at Genome Biology and Evolution
(http://www.gbe.oxfordjournals.org).

**Acknowledgments**

The authors acknowledge all the technical staff of UMR EBI
7267 for their assistance in the laboratory. This work was
supported by a European Research Council Starting Grant
(FP7/2007-2013, grant 260729 EndoSexDet) to R.C.

**Literature Cited**

Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation
of protein database search programs. Nucleic Acids Res. 25:
3389–3402.

Aswad A, Katszourakis A. 2012. Paleovirology and virally derived immunity.
Trends Ecol Evol. 27:627–636.

Auguste AJ, et al. 2014. Characterization of a novel Negevirus and a novel
Bunyavirus isolated from *Culex* (Culex) declarator mosquitoes in
Trinidad. J Gen Virol. 95:481–485.

Ballinger MJ, Bruenn JA, Kotov AA, Taylor DJ. 2013. Selectively maintained
paleoviruses in Holarctic water fleas reveal an ancient origin for phle-
oboviruses. Virology 446:276–282.

Ballinger MJ, Bruenn JA, Taylor DJ. 2012. Phylogeny, integration and ex-
pression of sigma virus-like genes in *Drosophila*. Mol Phylogenet Evol.
65:251–258.

Belshaw R, et al. 2004. Long-term reinfection of the human genome by
endogenous retroviruses. Proc Natl Acad Sci U S A. 101:4894–4899.

Belý VA, Levine AJ, Skalka AM. 2010. Sequences from ancestral
stranded DNA viruses in vertebrate genomes: the *Paroviridae* and
*Circoviridae* are more than 40 to 50 million years old. J Virol.
84: 12458–12462.

**GBE**
Gilbert C, Feschotte C. 2012. Endogenous viruses: insights into viral evolution and impact on host biology. Nat Rev Genet. 13:283–296.

Flegel TW. 2009. Hypothesis for heritable, anti-viral immunity in crustaceans and insects. Biol Direct. 4:32.

Fort P, et al. 2012. Fossil rhabdoviral sequences integrated into arthropod genomes: ontogeny, evolution, and potential functionality. Mol Biol Evol. 29:381–390.

Geuking MB, et al. 2009. Recombination of retrotransposon and exogenous RNA virus results in nonretroviral cDNA integration. Science 323:393–396.

Gilbert C, Feschotte C. 2010. Genomic fossils calibrate the long-term evolution of hepadnaviruses. PLoS Biol. 8:e1000495.

Gilbert C, et al. 2014. Population genomics supports baculoviruses as vectors of horizontal transfer of insect transposons. Nat Commun. 5:3348.

Granberg F, et al. 2013. Metagenomic detection of viral pathogens in Spanish honeybees: co-infection by Aphid Lethal Paralysis, Israel Acute Paralysis and Lake Sinai Viruses. PLoS One 8:e73459.

Hemiotu EA, et al. 2013. When parasitic wasps hijacked viruses: genomic and functional evolution of polydnaviruses. Philos Trans R Soc Lond B Biol Sci. 368:20130005.

Horie M, et al. 2010. Endogenous non-retroviral RNA virus elements in mammalian genomes. Nature 463:84–87.

Isava H, et al. 2011. Identification and molecular characterization of a new nonsegmented double-stranded RNA virus isolated from Culex mosquitoes in Japan. Virus Res. 155:147–155.

Katrouzakis A, Gifford RJ. 2010. Endogenous viral elements in animal genomes. PLoS Genet. 6:e1001191.

Katrouzakis A, Gifford RJ, Tristem M, Gilbert MTP, Pybus OG. 2009. Macroevolution of complex retroviruses. Science 325:1512.

King AM, Lefkowitz E, Adams MJ, Carstens EB. 2011. Virus taxonomy: proposal for a new family in the order Mononegavirales. Arch Virol. 156:2209–2226.

Kohn HJ, et al. 2013. Nyamiviridae: proposal for a new family in the order Mononegavirales. Arch Virol. 158:2209–2226.

Lavielle C, et al. 2013. Paleovirolology of “syncytins”, retroviral env genes exapted for a role in placentation. Philos Trans R Soc B. 368:20120507.

Liu H, et al. 2010. Widespread horizontal gene transfer from double-stranded RNA viruses to eukaryotic nuclear genomes. J Virol. 84:11876–11887.

Liu H, et al. 2011a. Widespread horizontal gene transfer from circular single-stranded DNA viruses to eukaryotic genomes. BMC Evol Biol. 11:276.

Liu H, et al. 2011b. Widespread endogenization of densivirus and parvoviruses in animal and human genomes. J Virol. 85:9863–9876.

Lupetti P, et al. 2013. Iridovirus infection in terrestrial isopods from Sicily (Italy). Tissue Cell. 45:321–327.

Maori E, Tanne E, Sela I. 2007. Reciprocal sequence exchange between non-retro viruses and hosts leading to the appearance of new host phenotypes. Virology 362:342–349.

Marklevitz M, et al. 2011. Gouleako virus isolated from West African mosquitoes constitutes a proposed novel genus in the family Bunyaviridae. J Virol. 85:9227–9234.

Michel-Salzat A, Bouchon D. 2000. Phylogeographic analysis of mitochondrial LSU rRNA in oniscids. C R Acad Sci III. 323:827–837.

Mihindukulasuriya KA, et al. 2009. Nyamamini and midway viruses define a novel taxon of RNA viruses in the order Mononegavirales. J Virol. 83:5109–5116.

Palacios G, et al. 2013. Characterization of the Uukuniemi virus group (Phlebovirinae: Bunyaviridae): evidence for seven distinct species. J Virol. 87:3187–3195.

Piskurek O, Okada N. 2007. Poxviruses as possible vectors for horizontal transfer of retroelements from reptiles to mammals. Proc Natl Acad Sci U S A. 104:12046–12051.

Poulos BT, Tang KFJ, Pantora CR, Bonami JR, Lightner DV. 2006. Purification and characterization of infectious myonecrosis virus of penaeid shrimp. J Gen Virol. 87:987–996.

Roossinck MJ. 2011. The good viruses: viral mutualistic symbioses. Nat Rev Microbiol. 9:99–108.

Roossinck MJ. 2012. Plant virus metagenomics: biodiversity and ecology. Annu Rev Genet. 46:359–369.

Rosario K, Breibart M. 2011. Exploring the viral world through metagenomics. Curr Opin Virol. 1:289–297.

Rosario K, Capobianco H, Ng TFF, Breibart M, Polston JE. 2014. RNA viral metagenome of whiteflies leads to the discovery and characterization of a whitefly-transmitted caravirus in North America. PLoS One 9: e86748.

Roosendaal M, et al. 2012. Diverse circular ssDNA viruses discovered in dragonflies (Odonata: Euphlebia). J Gen Virol. 93:2668–2681.

Routh A, Dominovic T, Johnson JE. 2012. Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus. Proc Natl Acad Sci U S A. 109:1907–1912.

Siewers F, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 7:529.

Sivaram A, et al. 2009. Isolation and characterization of densonucleosis virus from Aedes aegypti mosquitoes and its distribution in India. Intervirology 52:1–7.

Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6:31.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

Taylor DJ, Bruenn J. 2009. The evolution of novel fungal genes from non-retroviral RNA viruses. BMC Biol. 7:88.

Taylor DJ, Dittmar K, Ballinger MJ, Bruenn JA. 2011. Evolutionary maintenance of filovirus-like genes in bat genomes. BMC Evol Biol. 11:336.
Taylor DJ, Leach RW, Bruenn J. 2010. Filoviruses are ancient and integrated into mammalian genomes. BMC Evol Biol. 10:193.

Thézé J, Bézier A, Periquet G, Drezen J-M, Herniou EA. 2011. Paleozoic origin of insect large dsDNA viruses. Proc Natl Acad Sci U S A. 108: 15931–15935.

Wu Q, et al. 2010. Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. Proc Natl Acad Sci U S A. 107: 1606–1611.

Yang F, et al. 2001. Complete genome sequence of the shrimp white spot bacilliform virus. J Virol. 75:11811–11820.

Zhai Y, et al. 2010. Isolation and full-length sequence analysis of Armigeres subalbatus totivirus, the first totivirus isolate from mosquitoes representing a proposed novel genus (Artivirus) of the family Totiviridae. J Gen Virol. 91: 2836–2845.

Zilber-Rosenberg I, Rosenberg E. 2008. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. FEMS Microbiol Rev. 32:723–735.

**Associate editor:** Purificación López-García