

RESEARCH

Open Access



Comparative paleovirological analysis of crustaceans identifies multiple widespread viral groups

Gabriel Metegnier¹, Thomas Becking¹, Mohamed Amine Chebbi¹, Isabelle Giraud¹, Bouziane Moumen¹, Sarah Schaack², Richard Cordaux¹ and Clément Gilbert^{1*}

Abstract

Background: The discovery of many fragments of viral genomes integrated in the genome of their eukaryotic host (endogenous viral elements; EVEs) has recently opened new avenues to further our understanding of viral evolution and of host-virus interactions. Here, we report the results of a comprehensive screen for EVEs in crustaceans. Following up on the recent discovery of EVEs in the terrestrial isopod, *Armadillidium vulgare*, we scanned the genomes of six crustacean species: a terrestrial isopod (*Armadillidium nasatum*), two water fleas (*Daphnia pulex* and *D. pulicaria*), two copepods (the salmon louse, *Lepeophtheirus salmonis* and *Eurytemora affinis*), and a freshwater amphipod (*Hyalella azteca*).

Results: In total, we found 210 EVEs representing 14 different lineages belonging to five different viral groups that are present in two to five species: *Bunyaviridae* (–ssRNA), *Circoviridae* (ssDNA), *Mononegavirales* (–ssRNA), *Parvoviridae* (ssDNA) and *Totiviridae* (dsRNA). The identification of shared orthologous insertions between *A. nasatum* and *A. vulgare* indicates that EVEs have been maintained over several millions of years, although we did not find any evidence supporting exaptation. Overall, the different degrees of EVE degradation (from none to >10 nonsense mutations) suggest that endogenization has been recurrent during the evolution of the various crustacean taxa. Our study is the first to report EVEs in *D. pulicaria*, *E. affinis* and *H. azteca*, many of which are likely to result from recent endogenization of currently circulating viruses.

Conclusions: In conclusion, we have unearthed a large diversity of EVEs from crustacean genomes, and shown that four of the five viral groups we uncovered (*Bunyaviridae*, *Circoviridae*, *Mononegavirales*, *Parvoviridae*) were and may still be present in three to four highly divergent crustacean taxa. In addition, the discovery of recent EVEs offers an interesting opportunity to characterize new exogenous viruses currently circulating in economically or ecologically important copepod species.

Keywords: Paleovirology, Endogenous viral elements, Virus, *Bunyaviridae*, *Circoviridae*, *Mononegavirales*, *Parvoviridae*, *Totiviridae*, Copepoda, Crustacea

Background

Our perception of viruses has shifted drastically during the last ten years owing to the rapid development of viral metagenomics methods [1]. Sequencing viral metagenomes from various environments has revealed that viruses are the most numerous and diverse organisms on Earth [2–4] and that, likely, only a small proportion of them are

harmful pathogens. The results of these studies, coupled with the finding of many “good viruses”, suggest viruses could now often be considered mutualistic symbionts, fully integrated in holobionts, which have been defined as organisms harboring and interacting with a diverse microbial community [5, 6]. Viruses are thought to be at least as old as cellular organisms and it is becoming increasingly clear that they have had a strong, long-lasting, and ongoing influence on the evolution of their hosts and on ecosystem function [7–10].

* Correspondence: clement.gilbert@univ-poitiers.fr

¹Université de Poitiers, UMR CNRS 7267 Ecologie et Biologie des Interactions, Equipe Ecologie Evolution Symbiose, building B8-B35, 6 rue Michel Brunet, TSA 51106 F-86073 Poitiers, Cedex 9, France

Full list of author information is available at the end of the article

The recent discovery that many viral genomes integrate into the genome of their eukaryotic hosts has shed new light on our understanding of viral evolution and on the evolution of host-virus interactions [11, 12]. Paleovirology, the study of these endogenous viral elements (EVEs), has produced several major breakthroughs. First, we have learned that many extant viral families are much older than what was previously thought and that fast rates of evolution inferred from currently circulating viruses cannot be extrapolated over long evolutionary periods of time [12]. Other interesting outcomes of paleovirology studies include, much like viral metagenomics, the dramatic expansion of the host ranges for viral families. For example, in a recent study, we performed a comprehensive bioinformatic screen for EVEs in the genome of a terrestrial crustacean isopod, the pill bug *Armadillidium vulgare* [13]. We uncovered 54 EVEs from 10 diverse lineages belonging to the *Bunyaviridae*, *Circoviridae*, *Parvoviridae* and *Totiviridae* families as well as to the *Mononegavirales* order, indicating that isopods have been and may still be exposed to a remarkable diversity of viruses. These findings extended the host range of all five viral groups to isopod crustaceans, and led to the question of whether *A. vulgare* is unique in terms of abundance and diversity of EVEs among crustaceans or if a diverse EVE biota is characteristic of the group as a whole. In order to address this question, and to shed new light on the dynamics of viral endogenization more generally, we extended our screen to another species of terrestrial isopod (*A. nasatum*) and to five additional crustacean species (two species of water flea [*Daphnia pulex* and *Daphnia pulex*], a marine copepod [*Eurytemora affinis*], a freshwater amphipod [*Hyaella azteca*], and the salmon louse [*Lepeophtheirus salmonis*; Copepoda]; Additional file 1: Figure S1).

Results

EVE abundance and diversity in crustacean genomes

Overall, our comprehensive screening for EVEs in six crustacean genomes led to the discovery of a total of 210 EVEs belonging to five viral groups (*Bunya*-, *Circo*-, *Parvo*-, *Toti-viridae* and *Mononegavirales*; Figs. 1 and 2). All EVEs are provided in Additional file 2: Dataset S1. This search revealed 69 EVEs in *A. nasatum*, 22 in *D. pulex* (most of which correspond to the phlebovirus-like EVEs reported by Ballinger et al. [14]), 74 in *D. pulex*, 10 in *E. affinis*, 22 in *H. azteca* and 13 in *L. salmonis* (Fig. 1). Among these 210 EVEs, 103 showed the highest amino acid identity to members of the *Bunyaviridae* (best blastp hits range from 24 to 73 % identities; average length = 242 aa), 46 were most similar to members of the *Circoviridae* (best blastp hits are 29 to 74 % identities; average length = 128 aa), 32 to members of the *Mononegavirales* (25 to 51 % identities; average length = 745 aa), 21 to members of *Parvoviridae* (best blastp hits are 28 to

100 % identities; average length = 118 aa) and 8 to *Totiviridae* (best blastp hits are 27 to 49 % identities; average length = 126 aa) (Additional file 3: Table S2).

Several lines of evidence indicate that the viral genome fragments detected in this study are integrated in the genome of their host, rather than circulating as free viruses. First, assuming that exogenous viruses were sequenced and assembled together with the targeted crustacean genomes, we should have been able to uncover entire viral genomes. Yet, our search only revealed pieces of viral genomes (Fig. 2). Secondly, the method used to sequence the six crustacean genomes did not involve a reverse transcription step, and thus did not allow sequencing of any RNA molecule. Yet, many of the EVEs we found originate from exogenous RNA viruses (*Bunyaviridae*, *Mononegavirales* and *Totiviridae*). Thirdly, the presence of all 12 EVEs we targeted in isopods was confirmed by PCR amplification and Sanger sequencing (see below).

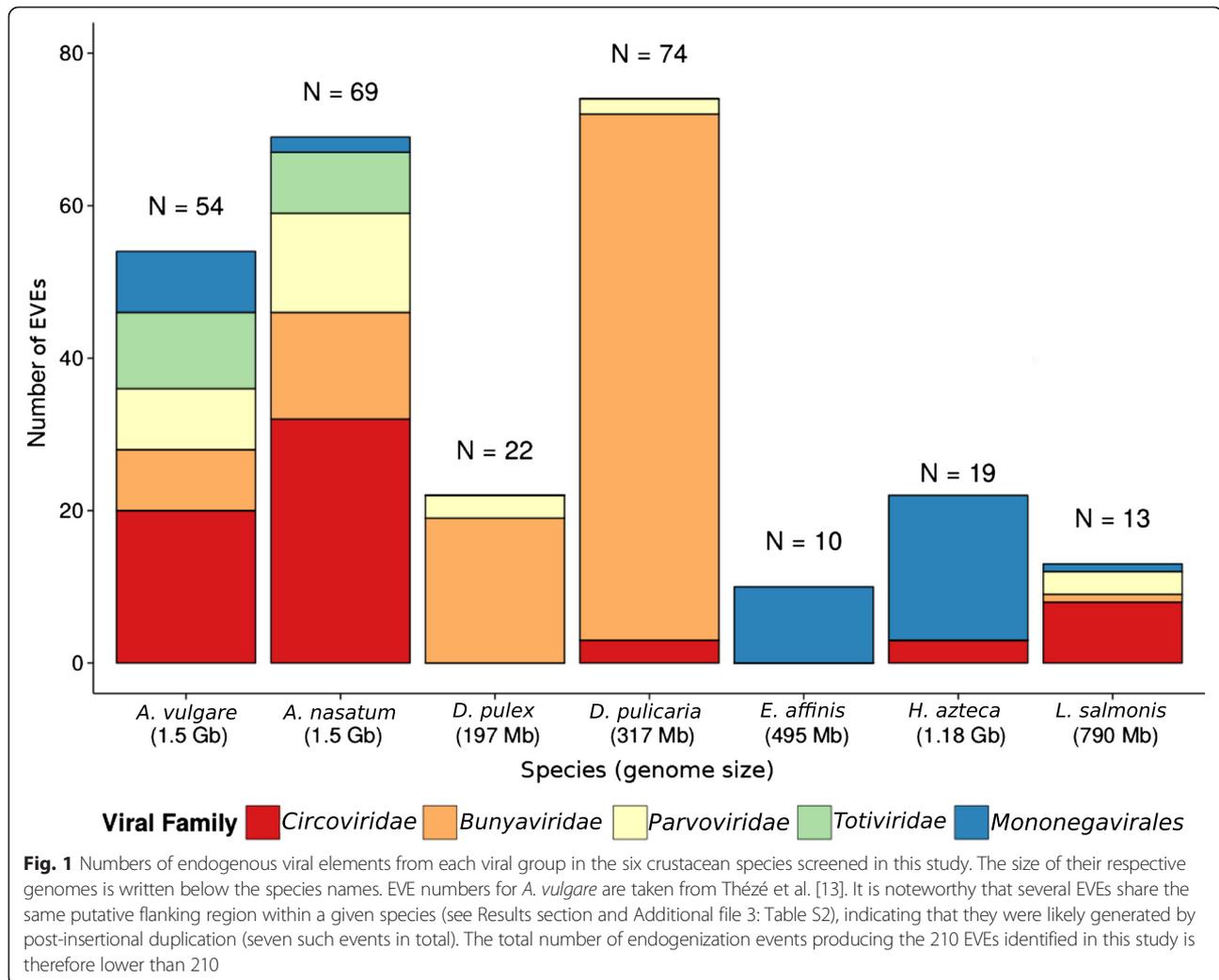
The number of EVEs detected in the various crustacean species varies substantially, from 10 in the copepod *E. affinis* to 74 in the water flea *D. pulex*. Though these differences may have biological underpinnings, they may also be in part explained by the varying quality of the genome assemblies (suggested by Geering et al. [15] and Zhuo et al. [16]), which varies greatly between species (Additional file 4: Table S1). Regarding the mechanisms underlying integration of viral genomes into crustacean genomes, we could not detect any sequence signature indicative of transposition-mediated or microhomology-mediated insertion. However, we found that several EVEs share the same putative flanking region within a given species (Additional file 3: Table S2), indicating that they were likely generated by post-insertional duplication (one such duplication in *E. affinis*, in *D. pulex*, in *H. azteca* and in *L. salmonis*; three such events in *D. pulex*; Additional file 3: Table S2).

Phylogenies of crustacean endogenous viral elements

To better characterize the diversity of crustacean EVEs uncovered in this study and to decipher their evolutionary history, we aligned these EVEs with several representative exogenous (and sometimes endogenous) viruses for each family, including the *A. vulgare* EVEs described in Thézé et al. [13], and carried out phylogenetic analyses. All resulting trees are overall congruent with the trees described by the International Committee on Taxonomy of Viruses [17].

Bunyaviridae

In the RNA-dependent RNA polymerase (RdRp) phylogeny (Fig. 3), the 12 *A. nasatum* (sequences 1–11 and 13) *Bunyaviridae*-like EVEs are all closely related to *A. vulgare* *Bunyaviridae*-like EVEs described in Thézé et al. [13], forming a relatively well-supported cluster with



recently described unclassified exogenous ssRNA viruses infecting arthropods [18]. *Daphnia* Bunyviridae-like EVEs fall into two distinct lineages: one that includes all *D. pulex* and all but one *D. pulicaria* sequences, which is not closely related to any previously known Bunyavirus, and one corresponding to a single *D. pulicaria* sequence (*D. pulicaria* 48) that is related to the Nairovirus genus.

In the nucleocapsid phylogeny (Additional file 5: Figure S2), the *A. nasatum* sequence (*A. nasatum* 12) belongs to the same lineage as the *A. vulgare* sequence reported by Thézé et al. [13]. Given the global differences in the topology of the RdRp and nucleocapsid phylogenies, we cannot conclude whether the RdRp and nucleocapsid EVEs found in isopods originate from the same virus (or same viral lineage) or not. In the discussion, we conservatively assume that they come from the same exogenous virus. Finally, the *L. salmonis* nucleocapsid EVE fragment (*L. salmonis* 1) falls near Orthobunyaviruses and the unclassified Wuhan Fly ssRNA virus [18] but we cannot

determine if this sequence belongs to one of the lineages described on the RdRp phylogeny.

Mononegavirales

In the *Mononegavirales* RdRp phylogeny (Fig. 4), the newly described crustacean EVEs fall into three distinct lineages (without considering *A. vulgare* EVEs reported in Thézé et al. [13]). The first one includes the *H. azteca* EVEs, which cluster with the recently described unclassified exogenous Wenzhou crab virus [18] (bootstrap value = 100) and one of the *A. vulgare* EVE lineages described in Thézé et al. [13]. The second lineage corresponds to the two *A. nasatum* EVEs which group with unclassified ssRNA exogenous viruses infecting arthropods reported by Li et al. [18]. The third new lineage of crustacean *Mononegavirales*-like EVEs groups the two sequences from *E. affinis* which fall within the *Rhabdoviridae* family (bootstrap value = 100). In the nucleocapsid phylogeny (Additional file 6: Figure S3), the EVEs newly discovered in *E. affinis*, *L. salmonis*

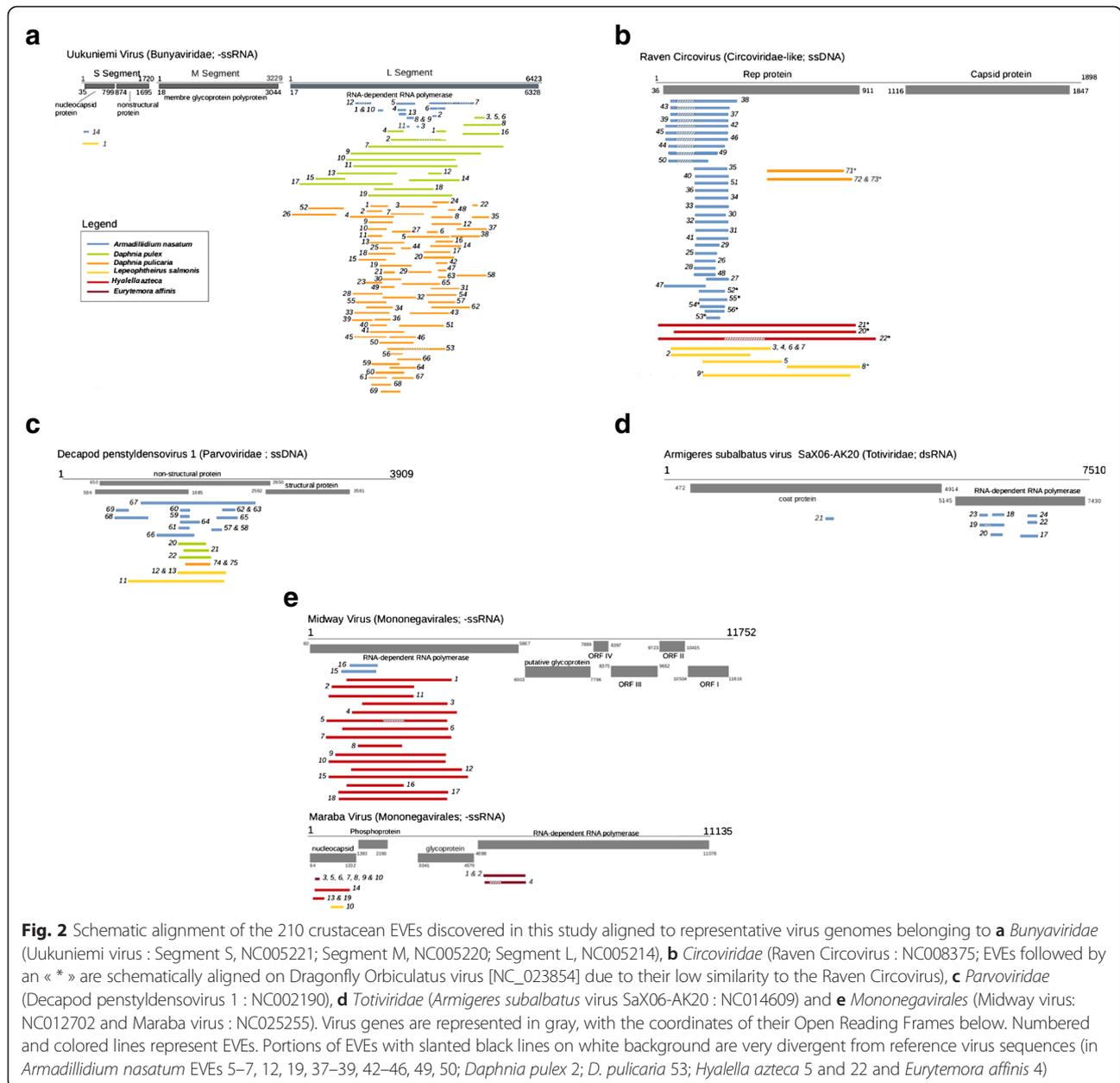


Fig. 2 Schematic alignment of the 210 crustacean EVEs discovered in this study aligned to representative virus genomes belonging to **a** *Bunyaviridae* (Uukuniemi virus : Segment S, NC005221; Segment M, NC005220; Segment L, NC005214), **b** *Circoviridae* (Raven Circovirus : NC008375; EVEs followed by an « * » are schematically aligned on Dragonfly Orbiculatus virus [NC_023854] due to their low similarity to the Raven Circovirus), **c** *Parvoviridae* (Decapod penstyldenovirus 1 : NC002190), **d** *Totiviridae* (*Armigeres subalbatus* virus SaX06-AK20 : NC014609) and **e** *Mononegavirales* (Midway virus: NC012702 and Maraba virus : NC025255). Virus genes are represented in gray, with the coordinates of their Open Reading Frames below. Numbered and colored lines represent EVEs. Portions of EVEs with slanted black lines on white background are very divergent from reference virus sequences (in *Armadillidium nasatum* EVEs 5–7, 12, 19, 37–39, 42–46, 49, 50; *Daphnia pulex* 2; *D. pulex* 53; *Hyalella azteca* 5 and 22 and *Eurytemora affinis* 4)

and *H. azteca* fall within the *Rhabdoviridae* family of exogenous viruses.

Given their similar placement in the two trees, it is likely that the *E. affinis* RdRp and nucleocapsid EVEs originate from the same *Rhabdoviridae*-like exogenous virus (or viral lineage). Interestingly, the placement of the *H. azteca* nucleocapsid EVEs (within *Rhabdoviridae*; Additional file 6: Figure S3) clearly differs from that of the RdRp EVEs found in this species (distantly related to *Rhabdoviridae*; Fig. 4), suggesting that the two types of sequences originated from two distinct viral lineages. It is also noteworthy that the *L. salmonis* sequence clusters tightly with the two sequences of exogenous *Rhabdoviridae* reported

in this same host species by Okland et al. [19] (bootstrap value = 80), suggesting that it must result from a relatively recent endogenization event.

Circoviridae

In the *Circoviridae* phylogeny (Fig. 5), crustacean EVEs fall into three distinct lineages: (1) corresponding to the *A. nasatum* EVEs and *A. vulgare* EVEs (described in Thézé et al. [13]), (2) a large cluster including *L. salmonis* and *Daphnia* EVEs together with Nanoviruses, unclassified exogenous circovirus-like sequences obtained from environmental metagenomics [20, 21], and endogenous viruses from mollusks (included in the Thézé et al. [13] phylogeny), and (3)

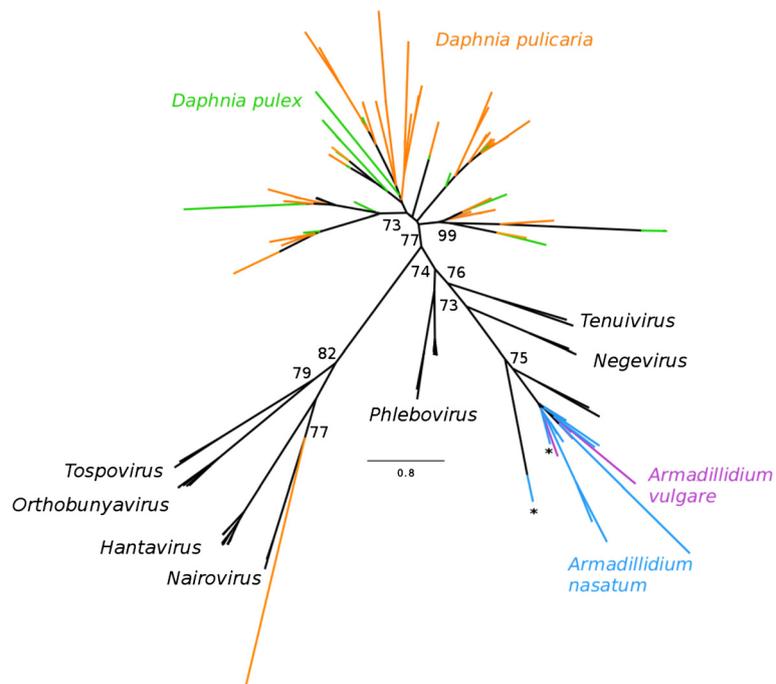


Fig. 3 Phylogeny of the *Bunyviridae* family, based on a multiple amino acid alignment and ML analysis of the RdRp. In addition to the EVEs discovered in this study, we added sequences from endogenous or exogenous viruses from the *Bunyviridae* family. ML nonparametric bootstrap values (100 replicates) are indicated when > 70

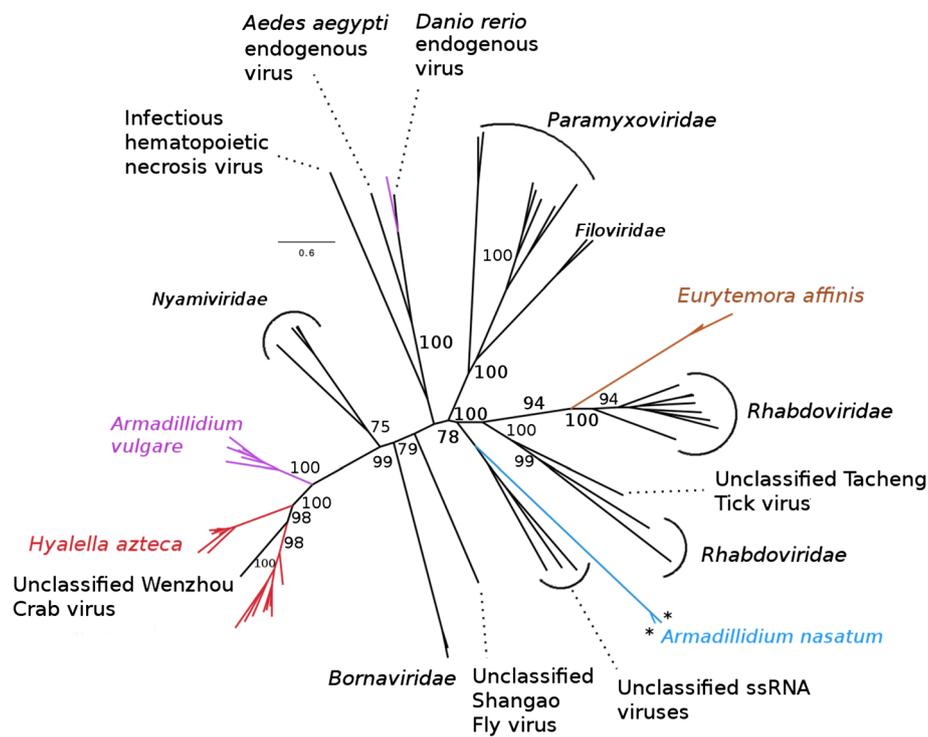
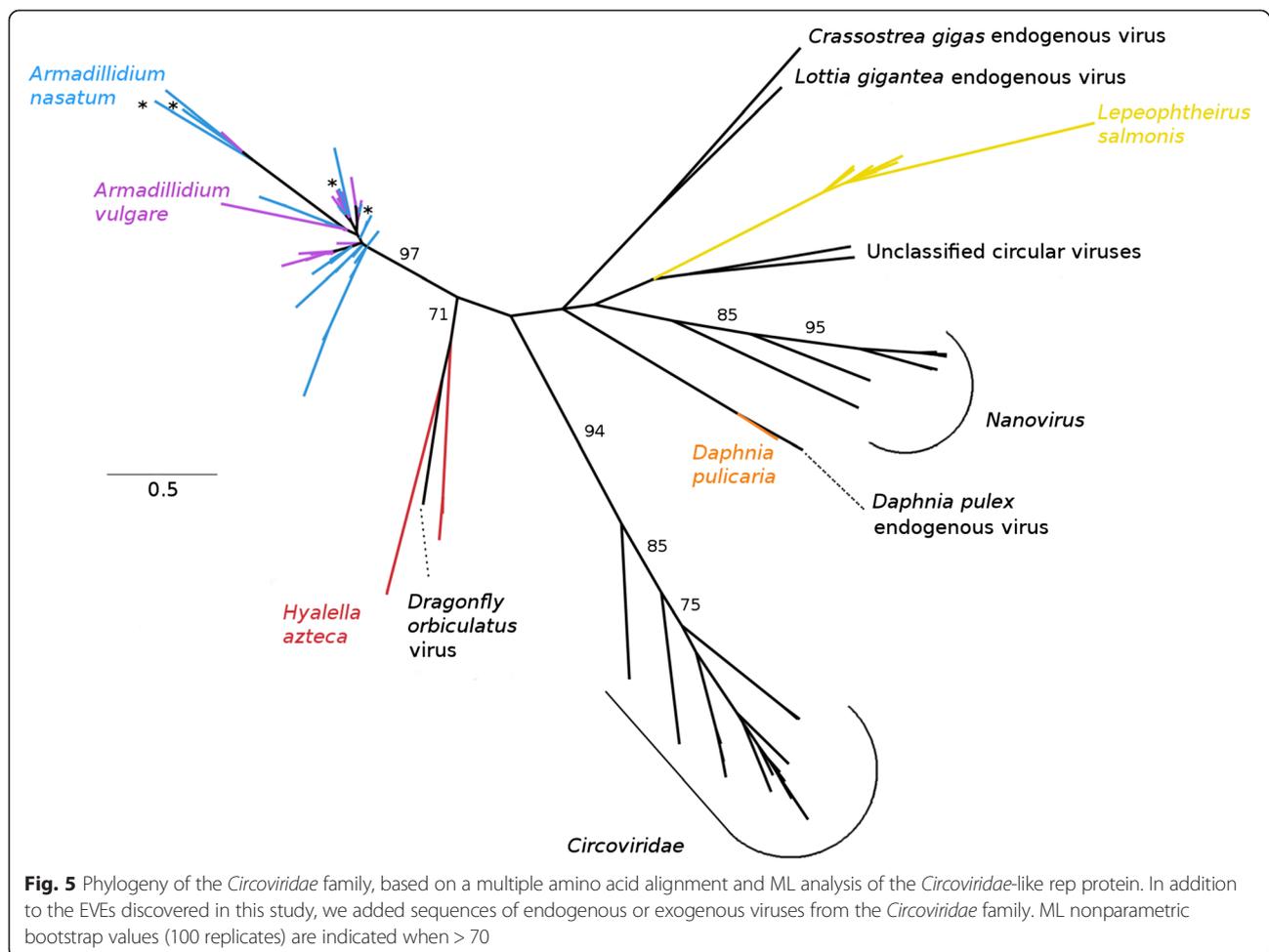


Fig. 4 Phylogeny of the *Mononegavirales* group, based on a multiple amino acid alignment and ML analysis of the *Mononegavirales*-like RdRp. In addition to the EVEs discovered in this study, we added sequences of endogenous or exogenous viruses from the *Mononegavirales* group. ML nonparametric bootstrap values (100 replicates) are indicated when > 70



a group linking *H. azteca* EVEs to the Dragonfly orbiculatus exogenous virus reported in Rosario et al. [22].

Parvoviridae

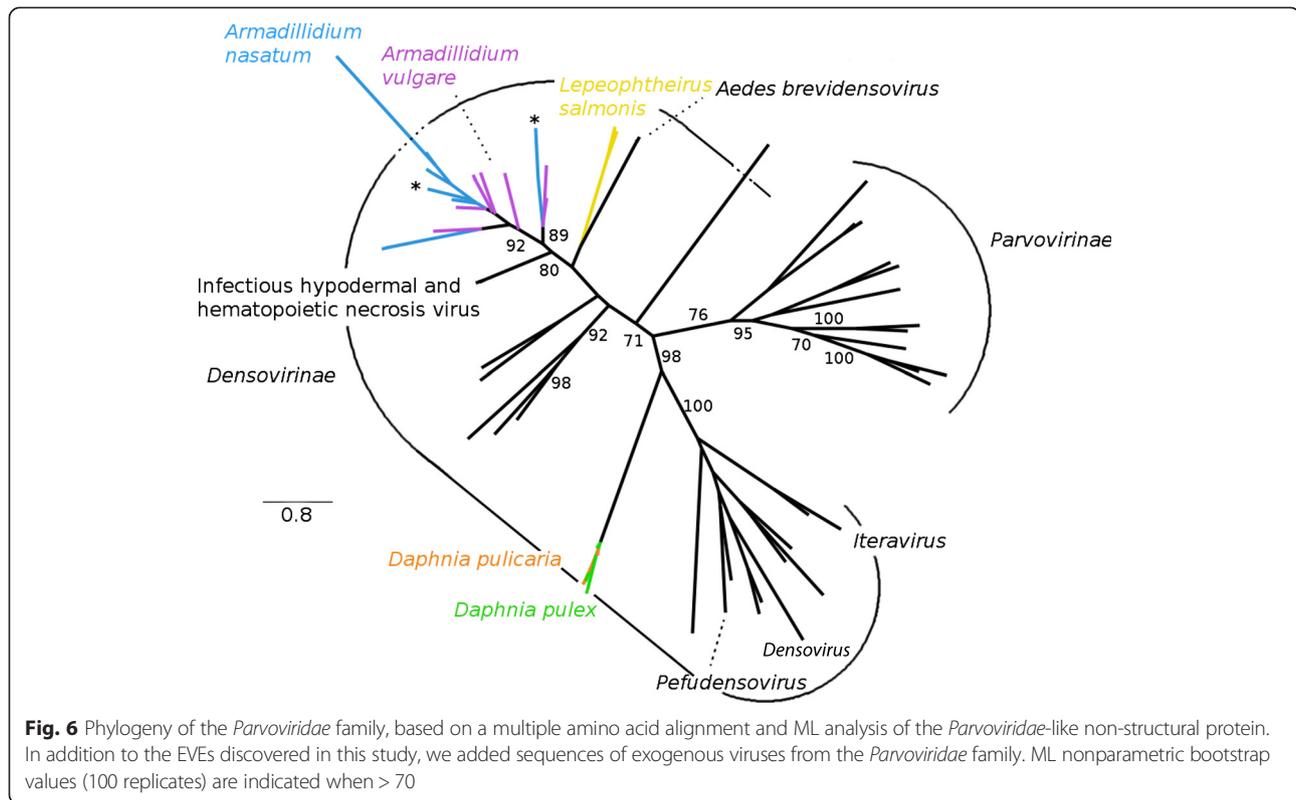
Crustacean *Parvoviridae*-like EVEs fall into at least two distinct lineages within the *Densovirinae* (Fig. 6). The first one corresponds to *Daphnia* EVEs and is related to the *Densovirus*, *Pefudensovirus* and *Iteravirus* genera. The second one includes *Armadillidium* and *L. salmonis* EVEs, as well as exogenous *Brevidensoviruses* from *Aedes* mosquitoes [23, 24], and the exogenous and endogenous versions of the *Infectious Hypodermal and Hematopoietic Necrosis Virus* [25, 26]. Though *L. salmonis* EVEs seem more closely related to mosquito *brevidensoviruses* than to *Armadillidium* EVEs, we conservatively assume that all these sequences are part of the same large lineage (bootstrap value = 80) because there is no large phylogenetic gap between them, *i.e.*, the distance separating each branch are relatively homogeneous.

Totiviridae

The only *Totiviridae*-like EVEs we found were in *A. nasatum*. Both RdRp and coat protein fragments (Fig. 7 and Additional file 7: Figure S4) were found in this species, and they cluster with the lineage of *A. vulgare* *Totiviridae*-like EVEs described in Thézé et al. [13]. This is most closely related to exogenous viruses from the *Artivirus* genus and to a virus from the unicellular eukaryote *Giardia* [13, 27–30].

Orthologous endogenous viral elements

We obtained positive PCR products for all 12 EVEs screened using the genomic DNA sample that served for sequencing the *A. nasatum* genome. Most of these 12 EVE loci were also amplified and Sanger sequenced in the other two *A. nasatum* DNA samples (Table 1), except the *Bunyaviridae*-like EVE 7 (negative PCR in *A. nasatum* sample 2 [An2]), *Mononegavirales*-like EVEs 15 and 16 (negative PCR in An2 and An3), and the *Parvoviridae*-like EVE 67 (negative PCR in An3). Our *in silico* search for



orthologous EVEs revealed three loci shared between *A. vulgare* and *A. nasatum* for which the host origin of the flanking region could be identified unambiguously (Additional file 8: Figure S5): *A. nasatum* Bunyavirus-like EVE 12, *A. nasatum* Circovirus-like 50 and *A. nasatum* Circovirus-like EVE 44. Not only are the flanking regions of these loci not similar to any known viral sequence, but they are characterized by the presence of interspersed and/or long microsatellite repeats (repeated at least six times or more). Such repeats are absent from the genome of all viruses belonging to the *Circoviridae* and *Bunyaviridae*, indicating that what we have identified as flanking regions indeed correspond to the eukaryotic host (*Armadillidium*) rather than the viral genome (Additional file 8: Figure S5). In addition to the sequences obtained from the genome sequence of *A. nasatum* and *A. vulgare*, we were able to PCR/Sanger sequence the three loci in *A. tunisiense* and one of them in *A. depressum* (Additional file 8: Figure S5). All EVEs identified computationally in *A. nasatum* or using PCR/sequencing were deposited in Genbank under accession numbers KT713978 – KT714035.

Discussion

Until recently, most of the knowledge available on crustacean viruses derived from studies of disease-causing viruses in shrimp farming, such as the white spot syndrome virus (WSSV; *Nimaviridae*), the taura syndrome virus

(TSV; *Picornaviridae*), and the yellowhead virus (YHV; *Roniviridae*) [31, 32], as well as the infectious hypodermal and hematopoietic virus (IHHN; *Parvoviridae*) [33] and the infectious myonecrosis virus (IMNV; *Totiviridae*) [28]. In addition, invertebrate iridescent viruses (*Iridoviridae*; dsDNA) have been observed in four species of decapods, two species of maxillopods, two species of branchiopods and 18 species of isopods [34]. These viruses are relatively easy to detect because of the iridescent blue or red color of infected individuals [35–37]. Dunlap et al. [38] also described a circovirus infecting two ecologically important copepod species, and two new species of rhabdoviruses were recently characterized in the salmon louse [19].

We recently discovered EVEs in *Armadillidium vulgare* and showed that terrestrial crustacean isopods have been and may still be exposed to a large variety of viruses, many of which belong to viral lineages that had never been reported in crustaceans before [13]. Here, we show that members of all five viral groups found in *A. vulgare* (*Bunyaviridae*, *Circoviridae*, *Mononegavirales*, *Parvoviridae*, *Totiviridae*) have also become endogenized in another terrestrial isopod, *A. nasatum*, and that four other crustacean species each harbor a viral flora composed of a subset of these five viral groups as well. Interestingly, all but one EVE lineage found in *A. nasatum* group with those previously identified in *A. vulgare*, which suggest that the two species are infected by the same viruses, an observation

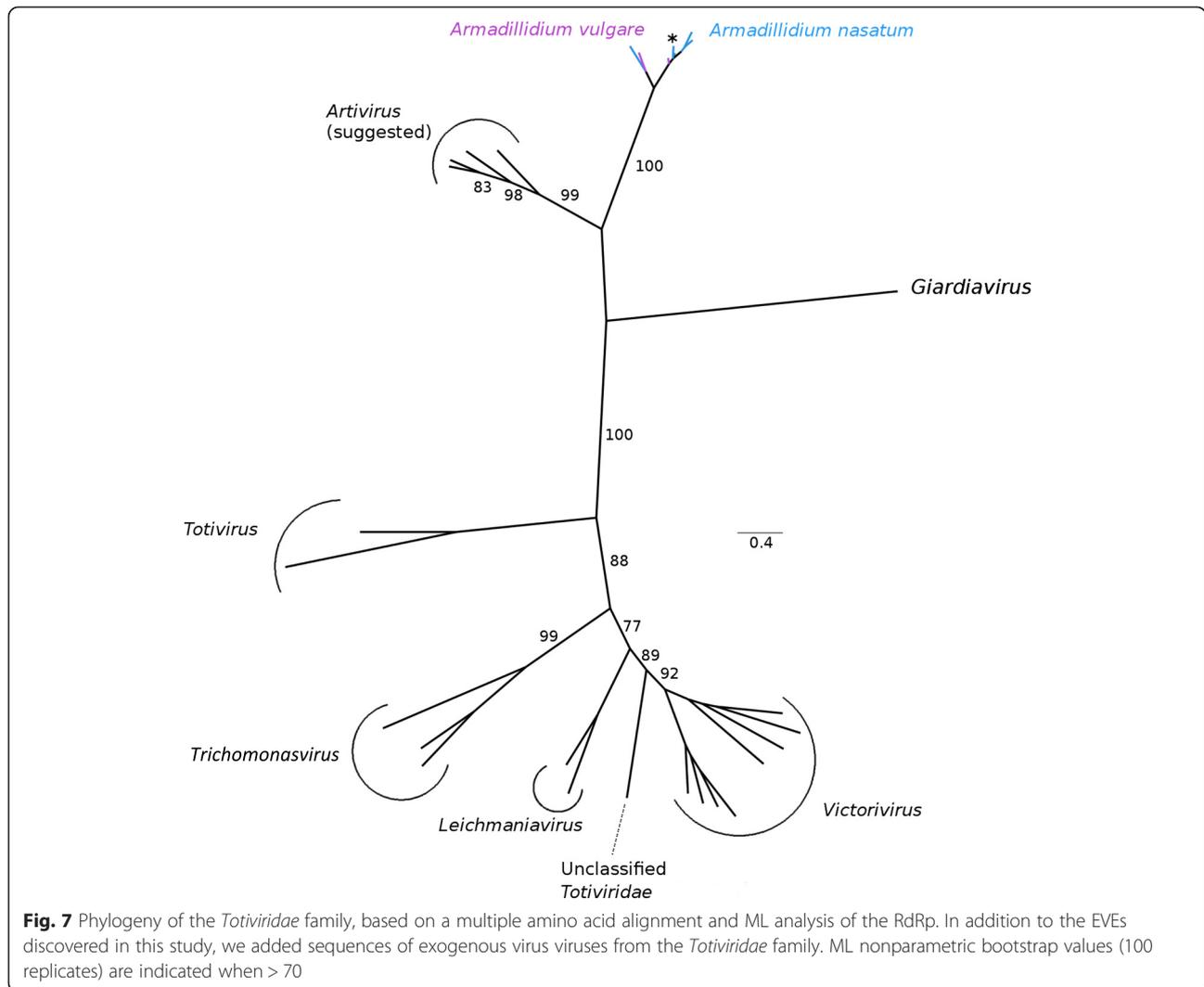


Table 1 EVEs PCR amplifications in 4 species of terrestrial isopods

Viral Family	EVE ^a	An1	An2	An3	At	Ad	Av
<i>Bunyaviridae</i>	7	+	-	+	-	-	-
<i>Bunyaviridae</i>	12	+	+	+	+	+	-
<i>Mononegavirales</i>	15	+	-	-	-	-	-
<i>Mononegavirales</i>	16	+	-	-	-	-	-
<i>Totiviridae</i>	21	+	+	+	-	-	+
<i>Totiviridae</i>	23	+	+	+	-	-	-
<i>Circoviridae</i>	44	+	+	+	+	-	-
<i>Circoviridae</i>	45	+	+	+	-	-	-
<i>Circoviridae</i>	46	+	+	+	-	-	+
<i>Circoviridae</i>	50	+	+	+	+	-	-
<i>Parvoviridae</i>	67	+	+	-	+	+	-
<i>Parvoviridae</i>	69	+	+	+	-	-	-

An: *Armadillidium nasatum* (three individuals were screened); At : *A. tunisiense*; Ad : *A. depressum*; Av : *A. vulgare*

+: Amplified; - : Not amplified

^aEach EVE is identified by a number as in Fig. 2

which is consistent with the fact that the distribution of the two species largely overlaps in Europe and that they are often found in the same habitats [39]. Overall, our phylogenetic analyses revealed that crustacean EVEs tend to group by taxa in distinct, well supported clusters across no fewer than 14 distinct viral lineages: four *Bunyaviridae*, five *Mononegavirales* (including a new *Armadillidium* lineage in *A. nasatum*), two *Circoviridae*, two *Parvoviridae* and one *Totiviridae*, 10 of which were also found in the initial screen of *A. vulgare* [13].

Given the tremendously large diversity of viruses known to infect eukaryotes and the fact that we screened species that are widely divergent from each other and from *A. vulgare*, it is perhaps surprising that all new EVEs detected here belong to the same viral groups than those detected in *A. vulgare* (no additional viral family was detected) and that only four of the lineages reported here were not found in *A. vulgare*. This leads to three non-mutually exclusive hypotheses: (1) that these five viral groups are

simply the most widespread in crustaceans, (2) these viral groups are more likely to endogenize than other viruses without being more prevalent as exogenous viruses, or (3) that crustacean genomes are uniquely vulnerable to endogenization by these 5 groups, relative to other host genomes. We note that a member of at least one other viral family (*Nimaviridae*) has been unearthed from a crustacean genome [40], and we believe that as more metagenomics and paleovirology studies are conducted, comparing global patterns of endogenization and global viral flora of extant viruses in a given taxonomic group will yield interesting insights into the ecology of host/virus interactions. But our current knowledge in this area is still too limited to draw any firm conclusion on this aspect of our results.

Our alignment of crustacean EVEs to representative exogenous viruses from each of the five viral groups revealed that most EVE fragments (83 %) are from the polymerase, with the remaining fragments being derived from different open reading frames such as coat or nucleocapsid protein (Fig. 2). Because this pattern is consistent throughout all five viral groups, we believe it is most likely explained by the strong purifying selection pressures acting on viral polymerases [41–44], leading to a high degree of conservation of such proteins between viruses that were endogenized and extant exogenous viruses. The other structural proteins (coat or capsid proteins) tend to be involved in more direct interactions with host factors (such as cell receptors) and are key to the entry of the virus in the cell. Thus, they are more likely to be engaged in an evolutionary arms race with the host and to evolve under rapid positive selection (e.g. [45, 46]). The level of similarity of such proteins between endogenized viruses and extant ones is therefore expected to be lower than that observed for polymerases.

The crustacean EVEs detected in this study show various levels of degradation when compared to their closest exogenous virus relatives, some being intact or disrupted by just one or a few mutations inducing a stop codon and/or a frameshift and others being heavily degraded by more than 10 nonsense mutations (Additional file 3: Table S2). This pattern indicates that viral endogenization has been recurrent during the evolution of the taxa included in this study. Further suggesting recurrent endogenization over time, we identified three EVEs shared at orthologous loci between *A. nasatum* and two or three other *Armadillidium* species (Additional file 8: Figure S5), and we were unable to amplify three other EVEs by PCR in one or two *A. nasatum* individuals sampled from a different population than the one used for genome sequencing (Table 1). These data indicate that, while some EVEs are old and were endogenized before the split between *A. nasatum* and the other *Armadillidium* species, others are more recent and are likely to still be polymorphic (with respect to presence/absence patterns) in *A. nasatum*. The phylogenetic relationships of the three *Armadillidium*

species included in our study have yet to be robustly resolved, but the mitochondrial COI gene from the two most distantly relatives (*A. nasatum* and *A. vulgare*, according to Dupeyron et al. [47]) differ by 16 %. Considering the proposed COI substitution rate of 1.4 % per million years in decapods [48], we can infer that the EVEs detected in isopods result from recurrent endogenization events that took place over several millions of years during the evolution of terrestrial isopods. The three EVEs that became endogenous in the ancestor of the *A. nasatum* + *A. vulgare* clade are all disrupted by one to four nonsense mutations and we did not find evidence for their transcription in the transcriptome of *A. nasatum* and *A. vulgare* [49]. Thus unlike previously described examples in non-crustacean taxa (e.g. [50, 51]), these three isopod EVEs do not appear to evolve under purifying selection and to fulfill a cellular function. Their maintenance in isopod genomes over several millions of years is therefore either completely neutral or due to initial exaptation, followed by loss of function and ongoing degradation as proposed for *Syncytin* genes in primates [52].

Finally, this study is the first to report viruses in the water flea *D. pulicaria*, the amphipod *H. azteca*, and the copepod *E. affinis*. The latter species is a major component of the mesozooplankton found in various saline and freshwater environments of the northern hemisphere [53]. Viruses have an important impact on the structure and ecology of phytoplankton communities [54], and it has recently been suggested they may play an important role in shaping mesozooplankton communities as well [38]. In addition, there is evidence suggesting that copepods can serve as vectors for transmitting viruses to fish and shrimp, causing important economic losses [55, 56], and to phytoplankton, with possible consequences on global biogeochemical cycling [57]. Despite these major consequences, only one study has characterized viral infections in copepods so far [38]. Interestingly, many of the copepod EVEs are devoid of nonsense mutation (Additional file 3: Table S2), suggesting they were endogenized very recently and may still be very similar at the nucleotide level to currently circulating viruses.

Conclusions

In conclusion, we characterized a large diversity of EVEs in crustacean genomes resulting from recurrent events of endogenization taking place over several millions of years. Most EVEs correspond to non-structural viral proteins, likely reflecting the slower rate of change of these proteins as compared to structural proteins. Interestingly, we found that four viral groups (*Bunyaviridae*, *Circoviridae*, *Mononegavirales*, *Parvoviridae*) are widespread in crustaceans, being present in three to four highly divergent taxa (amphipods, copepods, isopods, branchiopods) and that all viral groups found in non-isopod crustaceans are present

in isopods. We anticipate that further large scale paleovirology and metagenomics studies will shed light on the factors shaping global patterns of viral endogenizations and the composition of the viral communities currently circulating in a given taxonomic group. Finally, the sequences of recent EVEs that we identified in this study could facilitate the discovery of new exogenous viruses through targeted searches. The characterization of EVEs not only serves to provide a catalog of paleoviral events shedding light on past host-virus interactions but it can also help discovering new viruses in ecologically and/or economically important taxa (e.g. the copepods *E. affinis* and *L. salmonis*).

Methods

Genome screening for endogenous viral elements

The genomes of *E. affinis*, *H. azteca*, *L. salmonis* and *D. pulex* were downloaded from the GenBank database under accession numbers AZAI000000000, JQDR000000000, ADND000000000 and ACJG000000000 respectively. The genome of *D. pulicaria* was downloaded from the *wFlea-Base* Internet repository. The whole genome sequences of *A. nasatum* used in this study were generated as part of the ongoing *A. nasatum* genome project in our laboratory. Briefly, total genomic DNA was extracted from two *A. nasatum* individuals. A paired-end library with ~230 bp inserts was prepared and sequenced on an Illumina HiSeq2000. Reads were filtered with FastQC and assembled using the SOAP de novo software version 2.04 [58]. The best assembly was obtained with a *k*-mer size of 61. Genome statistics are available for all species in Additional file 4: Table S1.

To search for endogenous viral elements in crustacean genomes, we first removed low complexity repeats from the six genomes using RepeatMasker 4.0.5 [59]. We then carried out tblastx similarity searches [60] on these genomes using all available viral genomes ($n = 5678$ as of April 2015) as queries. Crustacean sequences yielding tblastx hits were then parsed from the tblastx output and converted into a fasta file using a custom script. Many tblastx hits were false positives corresponding to repeated sequences or to eukaryote genes that are present in viruses following host-to-virus horizontal transfers, known to be common in large dsDNA viruses [13, 61–63]. In order to remove these false-positive sequences, a reciprocal blastp was carried out using the tblastx fasta output as query on the “nr” (non-redundant) Genbank protein database, to eliminate any sequences for which the best reciprocal blastp hit was not a virus. The remaining sequences were manually aligned to a reference viral genome in BioEdit 7.1.9 [64] in order to draw schematic maps to illustrate the viral genome fragments endogenized in crustacean genomes.

Phylogenetic analyses

To better evaluate the diversity of newly discovered crustacean EVEs and to shed light on their evolutionary history, we carried out phylogenetic analyses of viral sequences including EVEs from Thézé et al. [13] and exogenous viral sequences obtained from Genbank. This phylogenetic analysis included closely related viral sequences selected following the BLAST analysis with the addition of closely-related proteins of representative virus species (International Committee on Taxonomy of Viruses, [17]), as well as recently published sequences [18] for two viral groups (*Bunyaviridae* and *Mononegavirales*). Triple iteration amino acid sequence multiple alignments were generated using ClustalOmega software (version 1.2.1; [65]). Maximum-Likelihood inferences were then performed on each alignment using the WAG empirical model of protein evolution [66] implemented in the RAXML software V. 7.4.6 [66]. Non-parametric bootstrap support values were obtained using parameters optimized for small datasets [67] after 100 iterations.

PCR and *in silico* screening of orthologous crustaceans endogenous viral elements

We used PCR and Sanger sequencing to verify the presence of some of the EVEs identified computationally in the *A. nasatum* genome and to assess whether some of them are polymorphic in terms of presence/absence at orthologous genomic sites in *A. nasatum* individuals sampled from three different populations available in our laboratory. We also investigated whether these EVEs are present at orthologous sites in three other closely related isopod species (*A. depressum*, *A. tunisiense* and *A. vulgare*). For this analysis, we selected the two EVEs of each viral group (four for *Circoviridae* because we found many more EVEs for this viral group) with the longest flanking regions. We then designed PCR primers to the flanking regions (Additional file 9: Table S3) and conducted a series of PCR screens on three *A. nasatum* DNA samples, one *A. vulgare* sample, one *A. depressum* sample, and one *A. tunisiense* sample. Genomic DNA extraction followed the Wilson protocol [68] which involved 3 h incubation of the tissue sample in proteinase-K at 56 °C, centrifugation (8000 g, 2 min), and an RNase treatment (30 min at 37 °C). DNA samples were then purified using spin columns from the DNeasy Blood & Tissue Kit (Qiagen). PCR reactions were carried out in 25 µl with 5 µL Buffer 5X, 0.5 µL dNTPs (2.15 mM), 1 µL of each primer (100 µM), 0.25 µL Taq polymerase 5 u/µL, 1 µL DNA. Thermocycling consisted of a 94 °C phase for 4 min, then 30 cycles of 30 s at 94 °C, 30 s at 55 °C and 50 s at 72 °C, followed by a final extension step of 5 min at 72 °C. PCR products resulting from amplifications in species other than *A. nasatum* were systematically purified and Sanger sequenced.

We also carried out an *in silico* screen to detect EVEs that are orthologous between *A. nasatum* and *A. vulgare*. For this we used all *A. nasatum* EVEs flanked on one or both sides by at least 150 bp of sequence showing no similarity to any virus as queries to perform blastn searches on *A. vulgare* sequences (using the sequences published in Thézé et al. [13]). Our identification of EVE flanking regions first relies on the fact that these regions are not similar to any known virus. In order to verify that they correspond to the host genome, we searched for the presence of known proteins motifs using these regions as queries to perform blastx searches against the Genbank non-redundant protein database. We also searched for the presence of interspersed repeats, which are typically abundant in eukaryotic genomes but very rare in viruses in general and absent from the genomes of *Bunyaviridae* and *Circoviridae* (the two viral families to which belong the three EVEs for which we found repeats in their flanking regions). For this we used each EVE flanking region as a query to perform blastn searches against the host genome it was extracted from. We considered interspersed repeats as regions longer than 100 bp repeated at least 10 times in the *Armadillidium nasatum* genome. The CENSOR searches we ran in Repbase [69] on the two interspersed repeats that we identified did not reveal any similarity to any known transposable element.

Additional files

Additional file 1: Figure S1. Phylogenetic relationships of the species studied in this project. The species targeted are in red. Divergence times are from www.timetree.org, except for *Daphnia* [70]. (PDF 122 kb)

Additional file 2: Dataset S1. Nucleotide sequence of all endogenous viral elements identified in this study. (TXT 201 kb)

Additional file 3: Table S2. Characteristics of endogenous viral elements in six crustacean genomes. In the "PCR test" column, hyphens indicate that we did not attempt to amplify the locus by PCR, while the "✓" sign indicates that we successfully amplified the locus by PCR in *A. nasatum*. The EVEs sharing the same letter (A, B, C, D, E or F) in the "Post insertional duplications" column have identical flanking regions, suggesting they were generated by post insertional duplication. In the same column, hyphens indicate either no flanking region or no similarity of the flanking region to any other EVE locus. (ZIP 139 kb)

Additional file 4: Table S1. Quality assessments of the 6 crustacean genomes used in this study. (DOCX 17 kb)

Additional file 5: Figure S2. Phylogeny of the *Bunyaviridae* family, based on a multiple amino acid alignment and ML analysis of the nucleocapsid protein. In addition to the EVEs discovered in this study, we added sequences of exogenous viruses from the *Bunyaviridae* family. ML nonparametric bootstrap values (100 replicates) are indicated when > 70. (PDF 89 kb)

Additional file 6: Figure S3. Phylogeny of the *Mononegavirales* group, based on a multiple amino acid alignment and ML analysis of the *Mononegavirales*-like nucleocapsid protein. In addition to the EVEs discovered in this study, we added sequences of exogenous viruses from the *Mononegavirales* group. ML nonparametric bootstrap values (100 replicates) are indicated when > 70. (PDF 92 kb)

Additional file 7: Figure S4. Phylogeny of the *Totiviridae* family, based on a multiple amino acid alignment and ML analysis of the nucleocapsid

protein. In addition to the EVEs discovered in this study, we added sequences of exogenous viruses from the *Totiviridae* family. ML nonparametric bootstrap values (100 replicates) are indicated when > 70. (PDF 74 kb)

Additional file 8: Figure S5. Schematic representation of the three EVE loci that are orthologous between the various *Armadillidium* species. The plain green portion of the loci are similar to a virus. a) *A. nasatum* Bunyavirus-like EVE 12 is most similar to the Wuhan insect virus 1 RdRp (AJG39261). Its 3' flank contains a 103-bp interspersed repeat (IR in blue) which is repeated at least 66 times in the *A. nasatum* genome (average similarity between repeats = 86 %) and a partial ORF similar to a hypothetical protein from *Helobdella robusta* (in grey). b) *A. nasatum* Circovirus-like EVE 44 is most similar to the Dragonfly orbiculatus virus rep protein (AFS65301). Its 5' and 3' flank contain a dinucleotide microsatellite (in orange) repeated at least 15 and 6 times respectively, that are shared at the exact same position with *A. vulgare*. c) *A. nasatum* Circovirus-like EVE 50 is most similar to the rep protein of an Uncultured marine virus (GAC77817). Its 3' flank contains a 130-bp interspersed repeat which is repeated at least 13 times in the *A. nasatum* genome (average similarity between repeats = 91 %), as well as a trinucleotide microsatellite repeated at least 17 times and shared with *A. vulgare*. The green portions of the loci with slanted black stripes correspond to the rest of the flanking regions, which are not similar to any known sequence. Red arrows indicate the position of forward and reverse PCR primers. (PDF 393 kb)

Additional file 9: Table S3. PCR primers used to confirm the presence of EVEs uncovered by the bio-informatic analysis and to screen for orthologous insertions in 3 *Armadillidae* species : *Armadillidium vulgare*, *A. tunisiense* and *A. depressum*. (DOCX 17 kb)

Abbreviations

RdRp: RNA-dependent RNA polymerase; PCR: Polymerase chain reaction; EVE: Endogenous viral element.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

GM participated in the design of the study, performed research and drafted the manuscript. TB, MAC, IG and BM generated the *Armadillidium nasatum* genome scaffolds. CG, RC and SS participated in the design of the study and in the writing of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors acknowledge Julien Thézé for discussions regarding phylogenetic analyses and 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 RC.

Author details

¹Université de Poitiers, UMR CNRS 7267 Ecologie et Biologie des Interactions, Equipe Ecologie Evolution Symbiose, building B8-B35, 6 rue Michel Brunet, TSA 51106 F-86073 Poitiers, Cedex 9, France. ²Department of Biology, Reed College, Portland, OR, USA.

Received: 3 August 2015 Accepted: 10 September 2015

Published online: 16 September 2015

References

- Willner D, Hugenoltz P. From deep sequencing to viral tagging: Recent advances in viral metagenomics. *Bioessays*. 2013;35:436–42.
- Suttle C. Viruses in the sea. *Nature*. 2005;437:356–61.
- Breitbart M, Rohwer F. Here a virus, there a virus, everywhere the same virus? *Trends Microbiol*. 2005;13:278–84.
- Kristensen DM, Mushegian AR, Dolja VV, Koonin EV. New dimensions of the virus world discovered through metagenomics. *Trends Microbiol*. 2010;18:11–9.
- Zilber-Rosenberg I, Rosenberg E. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiol Rev*. 2008;32:723–35.

6. Roossinck MJ. Move over, bacteria! Viruses make their mark as mutualistic microbial symbionts. *J Virol.* 2015;89:6532–5.
7. Bamford DH, Grimes JM, Stuart DI. What does structure tell us about virus evolution? *Curr Opin Struct Biol.* 2005;15:655–63.
8. Forterre P, Prangishvili D. The great billion-year war between ribosome- and capsid-encoding organisms (cells and viruses) as the major source of evolutionary novelties. *Ann N Y Acad Sci.* 2009;1178:65–77.
9. Koonin E, Senkevich T, Dolja V. The ancient virus world and evolution of cells. *Biol Direct.* 2006;1:29.
10. Nasir A, Sun FJ, Kim KM, Caetano-Anollés G. Untangling the origin of viruses and their impact on cellular evolution. *Ann N Y Acad Sci.* 2015;1341:61–74.
11. Katzourakis A, Gifford RJ. Endogenous viral elements in animal genomes. *PLoS Genet.* 2010;6:e1001191.
12. Feschotte C, Gilbert C. Endogenous viruses: insights into viral evolution and impact on host biology. *Nat Rev Genet.* 2012;13:283–96.
13. Thézé J, Leclercq S, Moumen B, Cordaux R, Gilbert C. Remarkable diversity of endogenous viruses in a crustacean genome. *Genome Biol Evol.* 2014;6:2129–40.
14. Ballinger MJ, Bruenn JA, Kotov AA, Taylor DJ. Selectively maintained paleoviruses in Holarctic water fleas reveal an ancient origin for phleboviruses. *Virology.* 2013;446:276–82.
15. Geering ADW, Maumus F, Copetti D, Choise N, Zwickl DJ, Zytnicki M, et al. Endogenous florendoviruses are major components of plant genomes and hallmarks of virus evolution. *Nat Commun.* 2014;5:5269.
16. Zhuo X, Rho M, Feschotte C. Genome-wide characterization of endogenous Retroviruses in the bat *Myotis lucifugus* reveals recent and diverse infections. *J Virol.* 2013;87:8493–501.
17. King AMQ, Lefkowitz E, Adams MJ, Carstens EB. Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses. Boston (MA): Elsevier; 2011.
18. Li C-X, Shi M, Tian J-H, Lin X-D, Kang Y-J, Chen L-J, et al. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *eLife.* 2015;4.
19. Økland AL, Nylund A, Øvergård A-C, Blindheim S, Watanabe K, Grotmol S, et al. Genomic characterization and phylogenetic position of two new species in *Rhabdoviridae* infecting the parasitic copepod, salmon louse (*Lepeophtheirus salmonis*). *PLoS One.* 2014;9:e112517.
20. Zawar-Reza P, Argüello-Astorga GR, Kraberger S, Julian L, Stainton D, Broady PA, et al. Diverse small circular single-stranded DNA viruses identified in a freshwater pond on the McMurdo Ice Shelf (Antarctica). *Infect Genet Evol.* 2014;26:132–8.
21. Dayaram A, Goldstien S, Argüello-Astorga GR, Zawar-Reza P, Gomez C, Harding JS, et al. Diverse small circular DNA viruses circulating amongst estuarine molluscs. *Infect Genet Evol.* 2015;31:284–95.
22. Rosario K, Dayaram A, Marinov M, Ware J, Kraberger S, Stainton D, et al. Diverse circular ssDNA viruses discovered in dragonflies (Odonata: Epiprocta). *J Gen Virol.* 2012;93(Pt 12):2668–81.
23. Boublik Y, Jousset F-X, Bergoin M. Complete nucleotide sequence and genomic organization of the *Aedes albopictus* Parvovirus (AaPV) pathogenic for *Aedes aegypti* larvae. *Virology.* 1994;200:752–63.
24. Sivaram A, Barde P, Kumar SRP, Yadav P, Gokhale M, Basu A, et al. Isolation and characterization of densovirus from *Aedes aegypti* mosquitoes and its distribution in India. *Intervirology.* 2009;52:1–7.
25. Bonami J-R, Trumper B, Mari J, Brehelin M, Lightner DV. Purification and characterization of the infectious hypodermal and haematopoietic necrosis virus of penaeid shrimps. *J Gen Virol.* 1990;71:2657–64.
26. Tang KFJ, Lightner DV. Infectious hypodermal and hematopoietic necrosis virus (IHHNV)-related sequences in the genome of the black tiger prawn *Penaeus monodon* from Africa and Australia. *Virus Res.* 2006;118:185–91.
27. Isawa H, Kuwata R, Hoshino K, Tsuda Y, Sakai K, Watanabe S, et al. Identification and molecular characterization of a new nonsegmented double-stranded {RNA} virus isolated from *Culex* mosquitoes in Japan. *Virus Res.* 2011;155:147–55.
28. Poulos BT, Tang KFJ, Pantoja CR, Bonami JR, Lightner DV. Purification and characterization of infectious myonecrosis virus of penaeid shrimp. *J Gen Virol.* 2006;87:987–96.
29. Wu Q, Luo Y, Lu R, Lau N, Lai EC, Li W-X, et al. Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. *Proc Natl Acad Sci.* 2010;107:1606–11.
30. Zhai Y, Attoui H, Mohd Jaafar F, Wang H, Cao Y, Fan S, et al. 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.* 2010;91:2836–45.
31. Overstreet RM, Jovonovich J, Ma H. Parasitic crustaceans as vectors of viruses, with an emphasis on three penaeid viruses. *Integr Comp Biol.* 2009;49:127–41.
32. Stentiford GD, Bonami J-R, Alday-Sanz V. A critical review of susceptibility of crustaceans to Taura syndrome, Yellowhead disease and White Spot Disease and implications of inclusion of these diseases in European legislation. *Aquaculture.* 2009;291:1–17.
33. Lightner DV, Redman RM, Bell TA. Infectious hypodermal and hematopoietic necrosis, a newly recognized virus disease of penaeid shrimp. *J Invertebr Pathol.* 1983;42:62–70.
34. Williams T. Natural invertebrate hosts of iridoviruses (*Iridoviridae*). *Neotrop Entomol.* 2008;37:615–32.
35. Cole A, Morris TJ. A new iridovirus of two species of terrestrial isopods, *Armadillidium vulgare* and *Porcellio scaber*. *Intervirology.* 1980;14:21–30.
36. Federici BA. Isolation of an iridovirus from two terrestrial isopods, the pill bug, *Armadillidium vulgare*, and the sow bug, *Porcellio dilatatus*. *J Invertebr Pathol.* 1980;36:373–81.
37. Lupetti P, Montesanto G, Ciolfi S, Marri L, Gentile M, Paccagnini E, et al. Iridovirus infection in terrestrial isopods from Sicily (Italy). *Tissue and Cell.* 2013;45:321–7.
38. Dunlap DS, Ng TFF, Rosario K, Barbosa JG, Greco AM, Breitbart M, et al. Molecular and microscopic evidence of viruses in marine copepods. *Proc Natl Acad Sci.* 2013;110:1375–80.
39. Vandel A. Isopodes terrestres (Deuxième partie). *Faune de France.* 1962; 66:
40. Rozenberg A, Brand P, Rivera N, Leese F, Schubart C. Characterization of fossilized relatives of the White Spot Syndrome Virus in genomes of decapod crustaceans. *BMC Evol Biol.* 2015;15:142.
41. Shangjin C, Cortey M, Segalés J. Phylogeny and evolution of the NS1 and VP1/VP2 gene sequences from porcine parvovirus. *Virus Res.* 2009;140(1–2):209–15.
42. Zhang Z, Jia R, Lu Y, Wang M, Zhu D, Chen S, et al. Identification, genotyping, and molecular evolution analysis of duck circovirus. *Gene.* 2013;529(2):288–95.
43. Huang X, Liu L, Du Y, Wu W, Wang H, Su J, et al. The evolutionary history and spatiotemporal dynamics of the fever, thrombocytopenia and leukocytopenia syndrome virus (FTLSV) in China. *PLoS Negl Trop Dis.* 2014;8:e3237.
44. Park DJ, Dudas G, Wohl S, Goba A, Whitmer SL, Andersen KG, et al. Ebola virus epidemiology, transmission, and evolution during seven months in Sierra Leone. *Cell.* 2015;161(7):1516–26.
45. Trinidad L, Blasdel KR, Joubert DA, Davis SS, Melville L, Kirkland PD, et al. Evolution of bovine ephemeral fever virus in the Australian epistystem. *J Virol.* 2014;88:1525–35.
46. Bidokhti MRM, Träven M, Krishna NK, Munir M, Belák S, Alenius S, et al. Evolutionary dynamics of bovine coronaviruses: natural selection pattern of the spike gene implies adaptive evolution of the strains. *J Gen Virol.* 2013;94:2036–49.
47. Dupeyron M, Leclercq S, Cerveau N, Bouchon D, Gilbert C. Horizontal transfer of transposons between and within crustaceans and insects. *Mobile DNA.* 2014;5:4.
48. Knowlton N, Weigt LA. New dates and new rates for divergence across the Isthmus of Panama. *Proceedings of the Royal Society of London B: Biological Sciences.* 1998;265:2257–63.
49. J. Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, et al. Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature.* 2014;515:261–3.
50. Herniou EA, Huguet E, Thézé J, Bézier A, Periquet G, Drezen J-M. When parasitic wasps hijacked viruses: genomic and functional evolution of polydnviruses. *Philos Trans R Soc Lond B Biol Sci.* 2013;368:20130051.
51. Lavielle C, Cornelis G, Dupressoir A, Esnault C, Heidmann O, Vernochet C, et al. Paleovirology of "Syncytins", retroviral env genes exapted for a role in placentation. *Philos Trans R Soc Lond B Biol Sci.* 2013;368:20120507.
52. Esnault C, Cornelis G, Heidmann O, Heidmann T. Differential evolutionary fate of an ancestral primate endogenous Retrovirus envelope gene, the EnvV *Syncytin*, captured for a function in placentation. *PLoS Genet.* 2013;9:e1003400.
53. Lee CE. Rapid and repeated invasions of freshwater by the copepod *Eurytemora affinis*. *Evolution.* 1999;53:1423–34.
54. Gustavsen JA, Winget DM, Tian X, Suttle CA. High temporal and spatial diversity in marine RNA viruses implies that they have an important role in

- mortality and structuring plankton communities. *Frontiers in microbiology*. 2014;5:703.
55. Jakob E, Barker DE, Garver KA. Vector potential of the salmon louse *Lepeophtheirus salmonis* in the transmission of infectious haematopoietic necrosis virus (IHNV). *Dis Aquat Organ*. 2011;97:155–65.
 56. Mendoza-Cano F, Sánchez-Paz A, Terán-Díaz B, Galván-Alvarez D, Encinas-García T, Enríquez-Espinoza T, et al. The endemic copepod *Calanus pacificus californicus* as a potential vector of White Spot Syndrome Virus. *J Aquat Anim Health*. 2014;26:113–7.
 57. Frada MJ, Schatz D, Farstey V, Ossolinski JE, Sabanay H, Ben-Dor S, et al. Zooplankton may serve as transmission vectors for viruses infecting algal blooms in the ocean. *Curr Biol*. 2014;24:2592–7.
 58. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience*. 2012;1:18.
 59. Smit AFA, Hubley R. RepeatModeler Open-1.0. 2008–2015. <http://www.repeatmasker.org> 2015.
 60. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25:3389–402.
 61. Filée J. Route of NCLDV evolution: the genomic accordion. *Current Opinion in Virology*. 2013;3:595–9.
 62. Filée J, Chandler M. Gene exchange and the origin of giant viruses. *Intervirology*. 2010;53:354–61.
 63. Holzerlandt R, Orengo C, Kellam P, Albà MM. Identification of new herpesvirus gene homologs in the Human genome. *Genome Res*. 2002;12:1739–48.
 64. Hall T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser*. 1999;41:95–8.
 65. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011;7:539.
 66. Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a Maximum-Likelihood approach. *Mol Biol Evol*. 2001;18:691–9.
 67. Stamatakis A. RAXML-VI-HPC: Maximum Likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006;22:2688–90.
 68. Kocher TD, Thomas WK, Meyer A, Edwards SV, Pääbo S, Villablanca FX, et al. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc Natl Acad Sci*. 1989;86:6196–200.
 69. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA*. 2015;6:11.
 70. Kotov AA, Taylor DJ. Mesozoic fossils (145 Mya) suggest the antiquity of the subgenera of *Daphnia* and their coevolution with chaoborid predators. *BMC Evol Biol*. 2011;11:129.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

