Molecular Phylogenetics and Evolution 84 (2015) 44–52

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Evolutionary relationships of iridoviruses and divergence of ascoviruses from invertebrate iridoviruses in the superfamily Megavirales

Benoît Piégu [a], Sassan Asgari [a,b], Dennis Bideshi [c,d], Brian A. Federici [d,e], Yves Bigot [a,*]

[a] UMR INRA-CNRS 7247, PRC, Centre INRA de Nouzilly, 37380 Nouzilly, France
[b] School of Biological Sciences, The University of Queensland, Brisbane, QLD 4072, Australia
[c] California Baptist University, Department of Natural and Mathematical Sciences, 8432 Magnolia Avenue, Riverside, CA 92504, USA
[d] Department of Entomology and Developmental Biology, University of California, Riverside, CA 92521, USA
[e] Interdepartmental Graduate Programs in Microbiology and Cell, Molecular and Developmental Biology, University of California, Riverside, CA 92521, USA

ARTICLE INFO

Article history:
Received 8 April 2014
Revised 9 October 2014
Accepted 19 December 2014
Available online 3 January 2015

Keywords:
Ascoviridae
Iridoviridae
Marseilleviridaek
Evolution
Taxonomy

ABSTRACT

The family *Iridoviridae* of the superfamily Megavirales currently consists of five genera. Three of these, *Lymphocystivirus*, *Megalocytivirus* and *Ranavirus*, are composed of species that infect vertebrates, and the other two, *Chloriridovirus* and *Iridovirus*, contain species that infect invertebrates. Until recently, the lack of genomic sequence data limited investigation of the evolutionary relationships between the invertebrate iridoviruses (IIVs) and vertebrate iridoviruses (VIVs), as well as the relationship of these viruses to those of the closely related family *Ascoviridae*, which only contains species that infect insects. To help clarify the phylogenetic relationships of these viruses, we recently published the annotated genome sequences of five additional IIV isolates. Here, using classical approaches of phylogeny via maximum likelihood, a Bayesian approach, and resolution of a core protein tree, we demonstrate that the invertebrate and vertebrate IV species constitute two lineages that diverged early during the evolution of the family *Iridoviridae*, before the emergence of the four IIV clades, previously referred to as Chloriridoviruses, Polyiridoviruses, Oligoiridoviruses and Crustaceoiridoviruses. In addition, we provide evidence that species of the family *Ascoviridae* have a more recent origin than most iridoviruses, emerging just before the differentiation between the Oligoiridoviruses and Crustaceoiridovirus clades. Our results also suggest that after emergence, based on their molecular clock, the ascoviruses evolved more quickly than their closest iridovirus relatives.

1. Introduction

The Megavirales is a newly erected superfamily of viruses consisting of a monophyletic group of viruses with large double-stranded DNA (dsDNA) genomes, most of which infect and replicate in the cytoplasm of animals and diverse unicellular eukaryotes (Iyer et al., 2006; Colson et al., 2013). Until this new superfamily was constructed, these viruses were commonly referred to as the nucleocytoplasmic large DNA viruses (NCLDVs). Currently, the superfamily Megavirales includes seven virus families, *Ascoviridae*, *Asfarviridae*, *Iridoviridae*, *Marseilleviridae*, *Mimiviridae*, *Poxviridae*, and *Phycodnaviridae*. A new family within the Megavirales consists of the recently discovered pandoraviruses, the largest viruses known, but additional studies are needed before concluding that these represent a new family (Philippe et al., 2013). For example, a common feature of all viruses belonging to the Megavirales is the presence of five core proteins: a major capsid protein, a D5 helicase-primase, a DNA polymerase of the family B, a DNA-packaging ATPase (A32), and viral late transcription factor 3 (A2L/VLTF3) (Colson et al., 2013). However, in the two pandoravirus genomes sequenced, no gene encodes a major capsid protein or a D5 helicase-primase. Nevertheless, about 30 pandoravirus genes encode proteins directly related to Megavirales protein orthologues. Most of the various species of Megavirales share approximately 40–50 genes (of more than 100), that were assigned with high confidence in the reconstruction of the gene set contained in the putative genome of their ancestor (Iyer et al., 2006; Yutin et al., 2009). These genes encode the key proteins required for genome replication, expression and virion morphogenesis.

Among Megavirales, two lineages can be distinguished. The first comprises animal viruses belonging to the families *Asfarviridae* and *Poxviridae*. The second contains viruses of unicellular eukaryotes, those of the families *Marseilleviridae*, *Mimiviridae* and

* Corresponding author. Fax: +33 (0)2 47 42 77 43.
E-mail address: yves.bigot@tours.inra.fr (Y. Bigot).

http://dx.doi.org/10.1016/j.ympev.2014.12.013
1055-7903/

Phycodnaviridae, and a sub-lineage of animal viruses belonging to the families *Ascoviridae* and *Iridoviridae*. All viruses in this second lineage have capsids composed primarily of a major capsid protein that assembles forming icosahedral virions (Colson et al., 2013). The exception is the ascoviruses, which have allantoid or bacilliform virions (Stasiak et al., 2003; Bigot et al., 2011). The host ranges distinguish the viruses of the latter five families noted above into the following three major lineages: the phycodnaviruses, which infect unicellular plants (algae); the marseilleviruses and the mimiviruses, which infect algae or unicellular eukaryotes (amoebae); and the iridoviruses and ascoviruses, which infect, respectively, vertebrates [i.e., vertebrate iridoviruses (VIV), fish, amphibians, frogs, or invertebrates, (i.e., invertebrate iridoviruses (IIVs), insects or crustaceans), or only insects (the ascoviruses, lepidopterans and certain parasitic wasps).

The configurations of their genomes are also distinctive. The genomes of phycodnaviruses are composed of a linear dsDNA molecule ranging from 100 to more than 550 kbp that have a closed hairpin at each end, as in PBCV-1, or a single strain extension or inverted repeats at each end that facilitates circularization, as respectively in EhV and EsV-1 (Van Etten et al., 2010). Mimivirus virions, alternatively, contain linear genomes of about 730 to more than 1200 kbp (Raoult et al., 2004; Fischer et al., 2010). These genomes are terminated by inverted repeats, but it is unknown whether they have sequence properties that facilitate genome circularization (Raoult et al., 2004). Iridovirus virions encapsidate a linear dsDNA molecule ranging from 100 to 230 kbp, with the size determined by the specific species (Jancovich et al., 2011). IIV genomes have a circularly permuted configuration, wherein the ends of each encapsidated genome vary from one virion to another. Ascovirus virions contain a circular genome having a species-specific size ranging from 100 to 200 kbp, which in some virions appears to be superhelical (Cheng et al., 1999; Bigot et al., 2011; Xue and Cheng, 2011). To date, the genome configuration of marseilleviruses has not been fully elucidated. Indeed, the genome map appears to be circular but their genomic configuration appears to be a linear DNA molecule terminated by repeats (Boyer et al., 2009; Thomas et al., 2011; Legendre et al., 2014). These features suggest their genome configuration might be similar to that of iridoviruses, i.e., circularly permuted linear dsDNA molecules.

Several recent phylogenetic studies of core genes have established the evolutionary relationships among these seven virus families, as well as among the genera within the families *Phycodnaviridae* and *Mimiviridae*. These studies infer that the families *Marseilleviridae*, *Iridoviridae* and *Ascoviridae* share a common virus ancestor, with the latter two families having evolved after the *Marseilleviridae* emerged (Boyer et al., 2009; Fischer et al., 2010; Thomas et al., 2011). Ascoviruses were previously proposed to share a common ancestor with invertebrate iridoviruses (IIVs, Stasiak et al., 2003; Bigot et al., 2009). However, the availability at that time of only two IIV genomes, IIV3 (Delhon et al., 2006) and IIV6 (Jakob et al., 2001), hampered a more refined resolution of the evolutionary relationships of these viruses and their taxonomy, thus leaving several questions unresolved. The first is whether ascoviruses originated before the IIV emergence or among IIVs. The second is whether emergence of the ascoviruses occurred before or after the divergence between IIVs and vertebrate iridoviruses (VIVs). And the third is whether the current taxonomy of IIVs within the family *Iridoviridae* is suitable, or needs to be revised. At present, the taxonomic structure of the family *Iridoviridae* is organized into five genera: *Chloriridovirus*, *Iridovirus*, *Lymphocystivirus*, *Megalocytivirus* and *Ranavirus*. As noted above, members of the first two genera have a host range restricted to invertebrate species, whereas the three others only infect poikilothermic vertebrates (fish, amphibians, and reptile species). The type species for the genus *Chloriridovirus* is IIV3, the only species reported in this genus. The type species for the genus *Iridovirus* is IIV6, and only two species, IIV1 and IIV6, have been recognized by the International Committee for Virus Taxonomy (ICTV) as representatives of this genus (Jancovich et al., 2011). A proposition to divide the genus *Iridovirus* into three species complexes, the Polyiridovirus (type species: IIV9; Wong et al., 2011), the Oligoiridovirus (type species: IIV6) and the Crustaceoiridovirus (type species: IIV31), has been made (Williams, 1994; Williams and Cory, 1994). In the absence of arguments supporting this old proposal, ICTV has not acted on the proposal (Jancovich et al., 2011).

In the present study, we use data from five recently published IIV genomes, namely IIV22, IIV22A, IIV25, IIV30 and IIV31 (Piégu et al., 2013, 2014a,b,c,d), to investigate the evolutionary relationships among vertebrate iridoviruses, invertebrate iridoviruses, and ascoviruses. Our analyses provide answers to the three questions raised above. First, we show that the taxonomy proposed by ICTV for IIVs needs to be revised. Second, our data provide strong evidence that ascoviruses have a "recent" origin anchored within IIVs, their closest relatives being IIV6 and IIV31. And third, our analyses demonstrate that VIVs emerged early during evolution, before the evolutionary differentiation in the four clades: Chloriridovirus (IIV3), Polyiridovirus (IIV9), Oligoiridovirus (IIV6) and Crustaceoiridovirus (IIV31).

2. Materials and methods

2.1. Sequence data

Sequences used to calculate phylogenies were first determined using BLAST results from databases (Dereeper et al., 2010). The inventory is available in Supplementary material online (Tables S1–S3). Protein sequences that were used originate from 8 VIVs including 3 ranaviruses [*Ambystoma tigrinum* virus (ATV); Frog virus 3 (FV3); Grouper iridovirus (GIV)], two megalocytiviruses [Infectious spleen and kidney necrosis virus (ISKNV); Orange-spotted grouper iridovirus (OSGIV); Rock bream iridovirus, isolate C1 (TSBIV)] and two lymphocystiviruses [Lymphocystis disease virus 1 (LCDV); Lymphocystis disease virus, isolate china (LCDV-C)], but also 8 IIVs [IIV3 (*Aedes taeniorhynchus* iridescent), 6 (*Chilo* iridescent virus), 9 (*Wiseana* iridescent virus), 22 (*Simulium* sp. iridescent virus A), 22A (*Simulium* sp. iridescent virus A), 25, 30 (*Helicoverpa zea* iridescent virus) and 31 (*Armadillidium vulgare* iridescent virus), 5 ascoviruses (*Diadromus pulchellus* ascovirus 4a (DpAV4a), *Heliothis virescens* ascovirus 3a and 3g (HvAV3e and HvAV3g), *Trichoplusia ni* ascovirus 6a (TnAV6a), *Spodoptera frugiperda* ascovirus 1a (SfAV1a)] and two species of *Marseilleviridae*, the Lausannevirus and the Cannes 8 virus.

2.2. Phylogenetic analyses from aligned protein sequences

Alignments were performed with Mafft and Muscle (for review, Dessimoz and Gil, 2010). They were then qualified with the evaluation mode of M-Coffee. Alignments were manually inspected and conserved blocks were reviewed using SeaviewV4 (Gouy et al., 2010). The conserved positions were selected using Gblocks (parameters were: -t = p, -e = -gb1, -b2 = N, -b3 = 40, -b4 = 2, -b5 = a, -v = 120; Castresana, 2000). Phylogenetic trees based on maximum likelihood were calculated with PhyML at http://www.phylogeny.fr/ (Guindon et al., 2010). Parameters used were WAG (substitution matrix), 0 (proportion of invariable sites), 7 in a, and 5 in f (number of relative substitution rate categories), and F (substitution model). The protein substitution model, the proportion of invariable sites, the number of relative substitution, number of rate categories and substitution model for ML trees were selected and evaluated by ProtTEST 3 (Darriba et al., 2011). The

best model was chosen on the basis of the Akaike Information Criterion (AIC, Akaike, 1973).

2.3. Phylogenetic analyses from non-aligned protein sequences using Baliphy

Baliphy was used to calculate phylogenetic relationships for each protein using non-aligned full-length sequences (Suchard and Redelings, 2006). This software used a Bayesian approach in which the multiple alignment of full length sequences and the phylogenetic trees are optimized using simple or mixed models for residue substitution or mutation frequencies, and which, to a large extent, consider deletions and insertions. This approach is of interest because the alignment and the phylogeny are constructed simultaneously, and a guide tree is not required (Suchard and Redelings, 2006). Parameters used were: –alphabet Amino-Acids –smodel LG –imodel RS07 –iterations = 100,000.

2.4. Core protein tree

Core gene trees were calculated using Mafft or Muscle alignments curated with Gblock as described above, except for the RNAse III orthologues, for which the complete sequence alignment was used. Alignments were concatenated and trees based on maximum likelihood were calculated with PhyML using parameters similar to those supplied above.

2.5. Graphic representation of trees

Trees were drawn with FigTree v1.4.0 at http://tree.bio.ed.ac.uk/ and SplitTree 4 at http://www.splitstree.org/.

3. Results

3.1. Definition of the datasets of core genes for the analyses

Previously, we showed that only twenty protein sequences were conserved throughout the genomes of all VIVs and IIVs, and ascoviruses (Table 1; Eaton et al., 2007; Bigot et al., 2009). Protein sequences from the Lausannevirus and the Cannes 8 virus (Family *Marseilleviridae*) were used as outgroups for phylogenetic calculations.

Several genes in Table 1 were eliminated for our investigation. The first is the protein involved in the genome packaging in SfAV1a and encoded by the open reading frame (ORF) 048 (Tan et al., 2009a,b), which is composed of variable numbers of motifs repeated in tandem and rich in SPRKT residues along with vs2C-ad motifs known to be found fused to OTU/A20-like peptidases and S/T protein kinases. Orthologues are present in IIV genomes, and those of some VIVs, however they cannot rationally be aligned to each other because the number of both motifs and their location varies significantly among proteins of different VIV and IIV species. Moreover, proteins with vs2C-ad motifs are absent in marseillevirus genomes. The second, third and forth proteins eliminated (encoded by IIV6 ORFs 118L, 143R and 337L) are a thymidine kinase and a myristylated membrane like-protein that are conserved in all VIVs, IIVs and ascoviruses, but absent in marseillevirus genomes. The fifth protein is coded for by IIV6 ORF 343L. This protein is highly conserved in sequence and size (~450 amino acid residues) among ascoviruses and in IIV6 (ORF343L) and IIV31 (ORF152R), but it appears to be encoded as a N-terminal fusion with a peptide of about 800–900 amino acid residues in other iridoviruses (IIV3 ORF 090L and IIV9 ORF 165R). Moreover, it is absent in marseilleviruses.

To analyze phylogenetic results, an annotation that can be important is the presence of each protein in the virion (Table 1). Indeed, marseilleviruses and iridoviruses on one hand, and ascoviruses on the other, have virions that differ in shape and symmetry, with the former two virus types being icosahedral and the latter allantoid or bacilliform. The reason for this important evolutionary change is not known, but it may correlate with ascovirus biology in that the virions of all these viruses are mechanically or vertically transmitted by parasitic wasps. As a result of this morphological change, it can be expected that at least some of the virion structural proteins would show convergence among the ascoviruses and a corresponding divergence between the ascoviruses and IIVs if involved in capsid assembly and structure. Indeed, their sequences may have co-evolved with those of other protein components, under selection constraints that favored, respectively,

Table 1
Proteins common to ascoviruses and iridoviruses and their presence/absence in their virions.

Core gene function	Orthologous ORFs			Core protein in virions[a]		
	SfAV1a	IIV6	IIV9	SfAV1a	IIV6	IIV9
DNA polymerase B	001	037L	116R	–	–	+
RNase III	022	142R	034L	–	+	+
Protein involved in genome packaging in SfAV1a[b]	048	232R	129R	+	+	+
Major capsid protein	041	274L	010R	+	+	+
Swi2/Snf2 ATP-dependent DExD Helicase	009	022L	055L	+	+	+
Serine/threonine kinase	104	098R	065R	–	–	–
Myristylated lipid membrane protein of large eukaryotic DNA viruses	035	118L	005R	+	+	+
Thiol oxidoreductase	061	374L	134L	+	–	+
RNA ligase	065	393L	059R	–	–	+
Thymidine kinase	040	143R	098R	–	–	
Protein with a PHA03307 domain	056	067R	001R	+	–	–
Myristylated membrane protein-like	054	337L	031R	+	+	+
DNA-directed RNA polymerase, subunit A′	067	343L	165R	–	–	–
DNA-directed RNA polymerase II subunit RPB2	052	428L	113R	–	–	–
Dynein-like beta chain	084	295L	143L	+	+	+
ATPase3 (A32)	110	075L	079R	–	+	+
DNA-directed RNA polymerase II subunit RPB1	008	176R	165R	–	–	–
D5 helicase-primase	093	184R	120R	–	–	–
Viral late transcription factor 3 (A2L/VLTF3)	029	282R	074L	–	–	–
NIF/NLI interacting factor/ctd-like phosphatase	109	355R	036L	+	+	+

[a] Tan et al. (2009b), Ince et al. (2010), Wong et al. (2011).
[b] Tan et al. (2009a).

either an icosahedral or allantoid virion. The significant change in ascovirus virion shape has, therefore, probably led to a corresponding significant modification of the molecular clock for these proteins, thus modifying their genetic distances with their closest IIV relatives with respect to protein orthologues that compose the above icosahedral virions.

3.2. Phylogenetic analyses with PhyML and Baliphy

A first round of analyses was performed with PhyML using alignments of each of the sixteen sequence datasets using Mafft and Muscle, then a curation with Gblock. Three kinds of trees were obtained. The first topology gathered proteins encoded by IIV6 ORFs 022L, 037L, 067R, 075L, 176R, 295L, 347L, 393L and 428L. In these trees (e.g. in the DNA polymerase and the Swi/snf2 helicase in Fig. 1a and b), the bootstrap values ranged from 70% to 100%, values that vary widely depending on the alignment method used. It is relevant that four of these proteins are found in virions, but none of them has yet been described as being a capsid component. These trees suggest that the VIV lineage and that of the origin of the four clades of IIVs diverged early in evolution from a common ancestor. In such a scenario, it must be noted that it is still unclear whether the common ancestor was already an iridovirus or a virus closer to a marseillevirus. The trees also suggested that

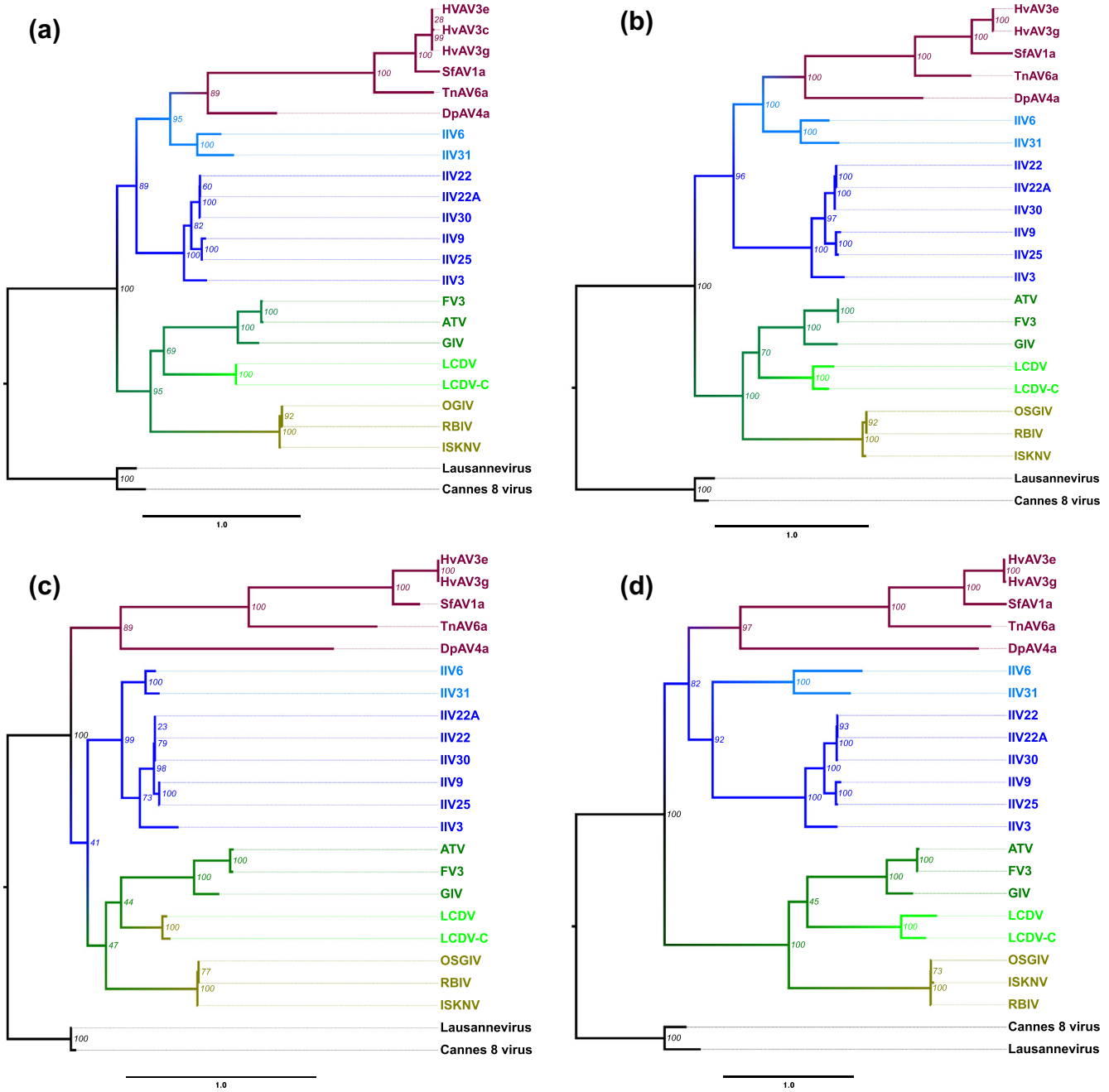

Fig. 1. Phylogenetic trees obtained with sequence alignments curated with Gblock and corresponding to orthologues of the DNA polymerase (a), the Swi2/Snf2 ATP-dependent DExD Helicase (b), the major capsid protein (c) and the D5-helicase primase (d). Numbers in italics at nodes indicate bootstrap values (%) retrieved from 1000 replicates. Branch lengths were proportional to genetic distances. Vertebrate iridovirus (VIV) proteins are in dark green for the three ranaviruses, light green for the two lymphocystiviruses, sepia for the three megalocytivirus, ascoviruses in purple, the clades of iridoviruses (IIV) 6 and 31 in light blue, and those of IIV3 and close relative IIV9 in dark blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ascoviruses emerged recently, and have common ancestors with two of the IIVs, specifically, IIV6 and IIV31.

The second set of trees deal with two proteins, the major capsid protein and the D5-helicase primase. Regardless of the alignment method used (Fig. 1c and d), the results indicated that IIVs were more closely related to each other than they were to ascoviruses, for which the branch lengths suggested that these viruses were distant relatives.

Trees of the third kind deal with proteins encoded by IIV6 ORFs 098R, 142R, 184R, 282R, 355R. These have low bootstrap values at most key nodes (less than 50%) and/or have inconsistent topologies with respect to what is known about the vertebrate iridovirus and invertebrate relationships. Moreover, the orthologue of the NIF/NLI interacting factor/ctd-like phosphatase (IIV6 ORF 355R) in TnAV6a was found to be much less similar to its relatives in the other ascoviruses (Fig. 2). This suggests that this gene may have arisen by a recent lateral transfer or by an event of partial inner recombination with a related gene.

Since insertions and deletions are important markers of protein evolution, as illustrated by previous research on parasites (Redelings and Suchard, 2007; Dessimoz and Gil, 2010), phylogenies were calculated from alignments not curated with Gblock. Under these conditions, trees obtained with IIV6 proteins encoded by ORFs 022L, 037L, 067R, 075L, 176R, 295L, 347L, 393L and 428L were similar and at least as consistent as those obtained after Gblock curation. In contrast, this solution did not produce consistent trees with IIV6 proteins encoded by ORFs 142R, 098R and 282R, even when complementary sequences were added. However, the addition of complementary sequences of Megavirales and bacterial origin to drive the alignments enabled us to obtain consistent trees similar to those obtained with IIV6 proteins encoded by ORFs 142R, even if certain bootstrap values were still low at key nodes (Fig. 3). Interestingly, BLAST searches using proteins encoded by IIV6 ORF 282R indicated that if the presence of this gene in the viruses we investigated was conserved during evolution, the origin of a few of these was different. Such a situation has been previously encountered with orthologues of the protein encoded by IIV6 ORF 179R, for which BLAST searches indicated that it is ubiquitous among these viruses, but had at least three different origins among VIVs, IIVs as well as DpAV4a and other ascoviruses (Bigot et al., 2009).

To confirm these results, complementary calculations were monitored with Baliphy. At the threshold of 95% confidence, the 16 trees obtained with Baliphy after 100,000 cycles of optimization fully confirmed the results obtained with PhyML.

3.3. Core protein tree

Alignments curated with Gblock for protein sequences encoded by IIV6 ORFs 022L, 037L, 067R, 075L, 176R, 295L, 347L, 393L and 428L were fused. The uncurated alignment of RNase III sequences was also added and a PhyML tree was calculated. Bootstrap values indicated that this tree was robust since the values were all over 99% at key nodes for virus taxonomy (Fig. 4). The genetic distances (i.e., branch lengths) between ascoviruses and the IIV ancestor that they share with IIV6 and IIV31 are significantly greater than those of IIV6 and 31, suggesting that ascoviruses evolved faster since their evolutionary "divergence" from the insect IIVs. In this regard, we note that while the mechanism in general for how IIVs are transmitted is not understood, it is well known that ascoviruses are readily transmitted via contamination of the ovipositors of female parasitic wasps with ascovirus virions during the process

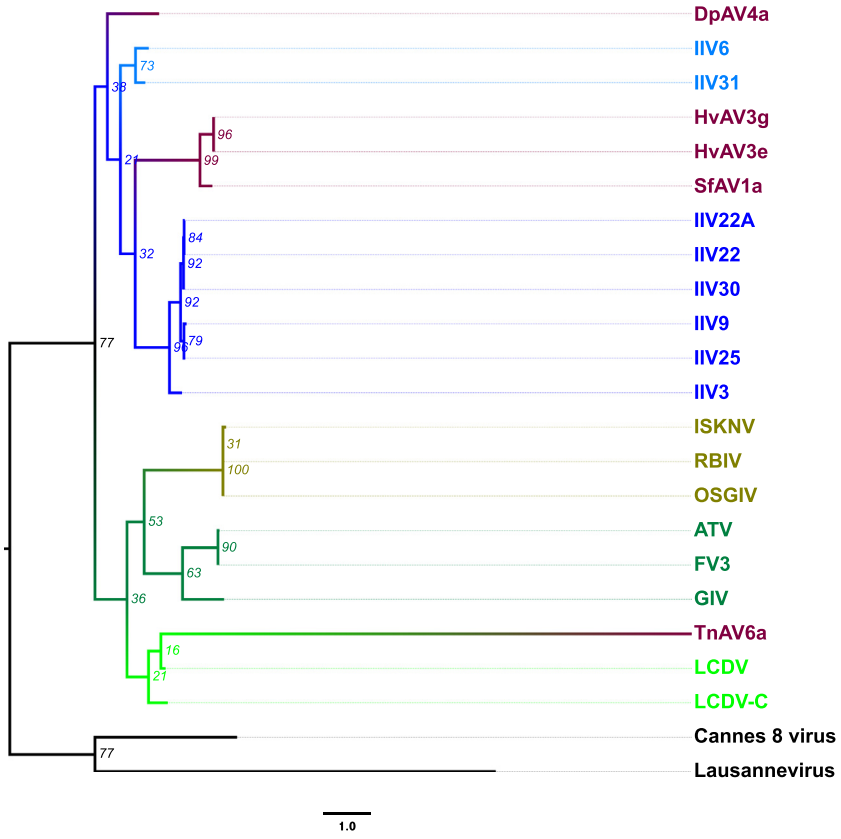

Fig. 2. Phylogenetic tree obtained with orthologues of the protein encoded by IIV6 ORF 355R, the NIF/NLI interacting factor/ctd-like phosphatase. Numbers in italics at nodes indicate bootstrap values (%) retrieved from 1000 replicates. Branch lengths were proportional to genetic distances. Color codes are the same as those used in Fig. 1 (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

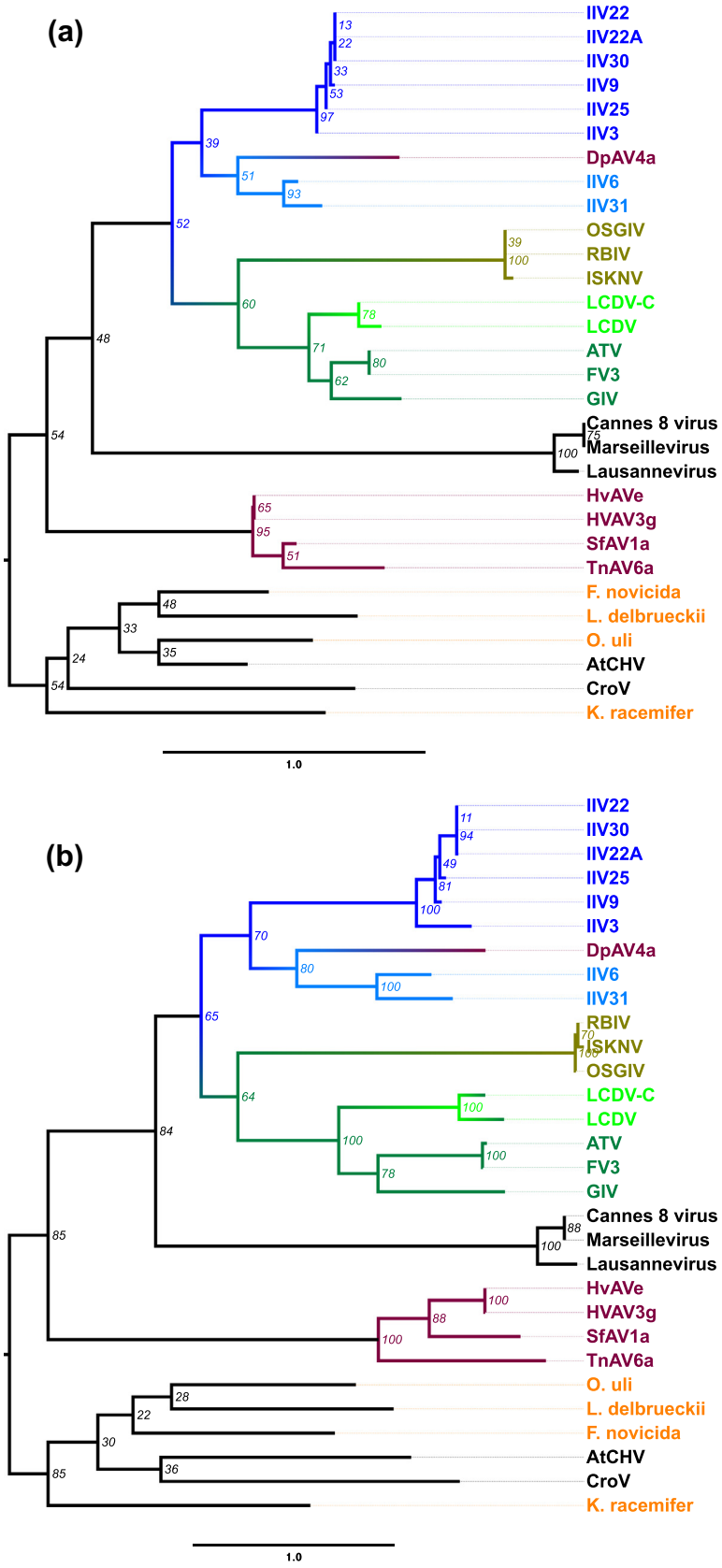

Fig. 3. Phylogenetic trees obtained with (a) an alignment of RNase III sequences curated with Gblock or (b) an alignment of RNase III sequences not curated with Gblock, and in which were added RNase III sequences from the phycodnavirus, *Acanthocystis turfacea*, Chlorella virus OR0704.3 (AGE59304.1; AtCHV), the mimivirus, *Cafeteria roenbergensis* virus BV-PW1 (YP_003969814.1; CroV), the bacteria, *Lactobacillus delbrueckii* (WP_002879727.1), *Francisella cf. Novicida* 3523 (YP_005824584), *Ktedonobacter racemifer* (WP_007911808.1), *Olsenella uli* DSM 7084 (YP_003801039.1) and the duplicated ORF 022 and 023 in SfAV1a. Numbers in italics at nodes indicate bootstrap values (%) retrieved from 1000 replicates. Branch lengths were proportional to genetic distances. Color codes are the same as those used in Fig. 1. Proteins of bacterial origins are in orange. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

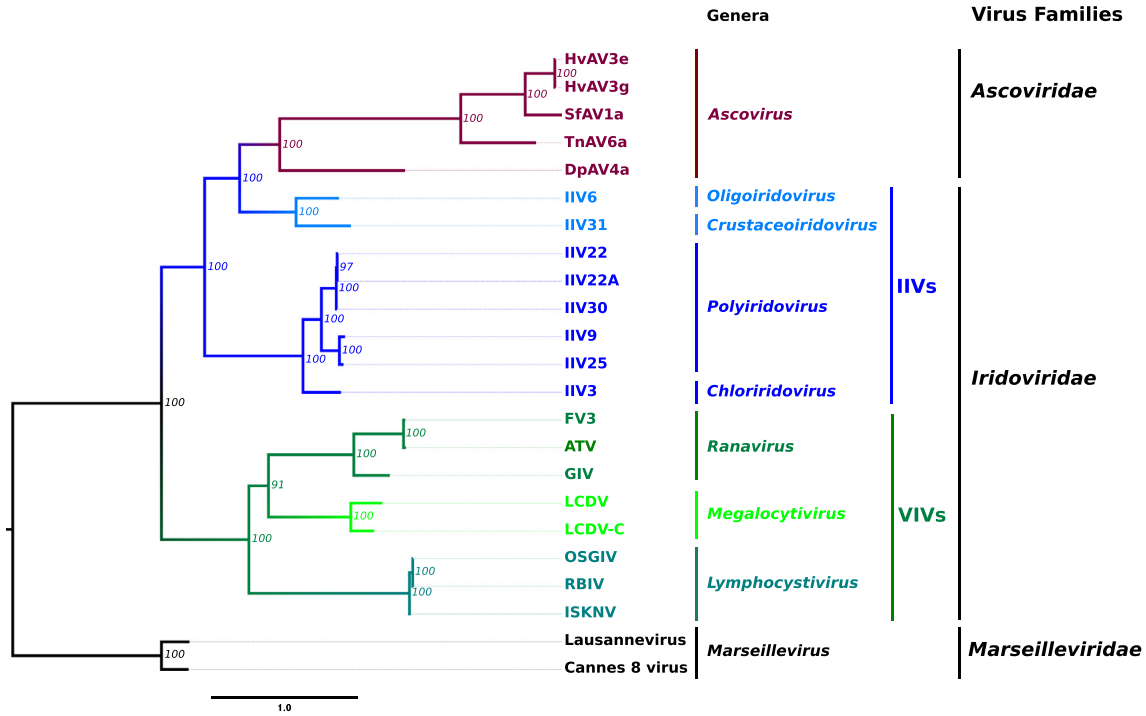

Fig. 4. Core protein tree. Numbers in italics at nodes indicate bootstrap values (%) retrieved from 1000 replicates. Branch lengths were proportional to genetic distances. Color codes are the same as those used in Fig. 1. The taxonomic levels from the genera to the families are indicated in the right margin. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of depositing eggs in their caterpillar hosts. This is a highly effective means of transmission that could in part well account for the more rapid evolution of ascoviruses. It is worth mentioning that IIV (in particular IIV6 and IIV31) and ascoviruses have overlap of host range and duration of replication cycle.

3.4. Complementary support

From above, we noted that some core genes might have been exchanged during evolution. Others may have been lost during evolution in certain lineages. This is the case for one set of six proteins encoded by IIV6 ORFs 050L, 244L, 350L, 359L, 361L and 454R (Table 2; Supplementary material online, Table S2) that do not occur in the VIV lineage, except for the orthologue of IIV6 ORF 361L, which is present in members of the genus *Lymphocystivirus*. This is likely the result of a lateral gene transfer. The use of these six datasets in PhyML analyses enabled us to confirm that ascoviruses emerged more recently and have a common ancestor with two IIVs, namely, IIV6 and IIV31. There are also two conserved genes among the marseilleviruses, and vertebrate and invertebrate iridoviruses, which have an orthologue encoded by DpAV4a ORFs 025 and 066 (Table 3; Supplementary material online, Table S3), but which have been lost in other ascoviruses. Their phyML analyses permitted confirmation, as noted above, that ascoviruses emerged recently and have common ancestors with two IIVs, IIV6 and IIV31, but also that the vertebrate iridovirus lineage diverged from an invertebrate iridovirus ancestor prior to the emergence of four clades of this virus type.

4. Discussion

The present analysis has enabled us to provide much clearer insights into the evolutionary relationships between the two main lineages of iridoviruses, i.e., vertebrate and invertebrate, as well as the relationship of these viruses to the ascoviruses, and to refine the taxonomy of all three. Secondly, the methods we employed have permitted us to identify core genes that can be used to establish the taxonomy of new viral isolates that belong to these groups. In addition to the confirmation that all iridoviruses and ascoviruses have evolved from a common ancestor that they share with members of the family *Marseilleviridae*, our phylogenetic studies have provided substantial evidence that (1) the vertebrate iridovirus

Table 2
Genes absent in the vertebrate iridovirus lineage but present in invertebrate iridoviruses, ascoviruses, and marseilleviruses – and their presence/absence in their virions.

Gene function	Orthologous ORFs			Core protein in virions[a]		
	SfAV1a	IIV6	IIV9	SfAV1a	IIV6	IIV9
Putative ubiquitin thioesterase	057	359L	037R	–	–	–
DNA-directed RNA polymerase II subunit RPB5	089	454R	154L	–	–	–
DNA repair exonuclease SbcCD D subunit	059	244L	138R	–	–	–
DNA double-strand break repair rad50 ATPase-like	103	050L	106R	–	–	–
Papain-like proteinase	114	361L	177R	–	+	+
Putative zinc finger/DNA-binding protein	113	350L	189R	–	–	–

[a] Tan et al. (2009b), Ince et al. (2010), Wong et al. (2011).

Table 3
Genes conserved in marseilleviruses, iridoviruses, and ascovirus DpAV4a, but absent in other ascoviruses - and their presence/absence in their virions.

Gene function	Orthologous ORFs			Core protein in virions[a]	
	DpAV4a	IIV6	IIV9	IIV6	IIV9
Proliferating cell nuclear antigen	025	346R	053R	–	–
Ribonucleoside-diphosphate reductase, beta subunit	066	376L	187R	–	+

[a] Ince et al. (2010), Wong et al. (2011).

lineage diverged early during evolution of the family *Iridoviridae*, (2) that the invertebrate iridoviruses likely evolved in four sub-lineages and (3) that from the perspective of the scale of iridovirus evolution, the ascoviruses comprise the most recent lineage to diverge, and they evolved from the invertebrate iridovirus lineage.

Our first and second findings above are supported by phylogenies of the 9 cores genes along with those of protein orthologues encoded by DpAV4a ORFs 025 and 066. These results make it highly likely that the vertebrate iridovirus lineage emerged early during the evolution of iridoviruses, before IIVs diverged into the four current clades so far identified. Our second finding also is supported by the phylogenies of the same 9 core genes, and those of protein orthologues encoded by IIV6 ORFs 050L, 244L, 350L, 359L, 361L and 454R. With respect to the divergence of the four invertebrate iridovirus lineages, our results also indicate that these four clades are more related by pairs, IIV6 and IIV31 on one hand, and IIV3 and IIV9 on the other (Wong et al., 2011; Piégu et al., 2013, 2014a,b,c,d). Thus, the current ICTV classification for IIVs (Jancovich et al., 2011) needs to be revised, with the alternative improvements to be considered. The first is to keep the two existing genera, *Iridovirus* and *Chloriridovirusis*, and to transfer IIV9 and its close relatives (IIV22, IIV22A, IIV25, and IIV30) from the genus *Iridovirus* into the genus *Chloriridovirusis*. The second is to resurrect a former 1994 proposition (Williams, 1994; Williams and Cory, 1994), which distinguished four groups of IIVs and to raise each of them at the level of genus, i.e. *Chloriridovirus* (IIV3), *Oligoridovirus* (IIV6), *Polyiridovirus* (IIV9), and *Crustaceoiridovirus* (IIV31). Whether the length of branches in the trees, including in the core protein tree, indicate that the genetic distances between VIV genera are more important than between the four IIV clades, the level of synteny (i.e. collinearity) in the gene organizations among the four IIV groups are as low as between the VIV genera. Indeed, it is limited to five genes between IIV3 and IIV9 (ORFs 028R to 032R versus 097R to 101R; Wong et al., 2011; Piégu et al., 2013, 2014a,b) and to three genes between IIV6 and IIV31 (ORFs 196R to 198R versus ORFs 042R to 044R, Piégu et al., 2014c). Our opinion is that the solution with four IIV genera would be the best integration of the data for illustrating the diversity of IIVs.

With respect to our third finding noted above, that is that the ascoviruses emerged recently from an invertebrate ancestor iridovirus lineage and have been undergoing more rapid evolution than the earlier lineages, our observations are in agreement with the adaptive requirements related to their change of virion shape and genome configuration. Interestingly, the classification of SfAV1a, HvAV3e, HvAV3g and TnAV6a into the single genus *Ascovirus* can be considered suitable from a taxonomic standpoint, but the distance of DpAV4a from these ascoviruses is much greater than distances observed between the vertebrate and invertebrate iridovirus genera, or between DpAV4a and IIV6 and IIV31. Moreover, it must also be noted that ascoviruses present an elevated level of synteny (i.e. collinearity) in the gene organization, expect for DpAV4a. Thus, we think it would be appropriate to create a new genus within the family *Ascoviridae*, with DpAV4a as type species (Bigot et al., 2011).

From a methodological standpoint, our three key findings are also of interest to the use of phylogenies for taxonomic purposes in virology for the following reasons. The first is that Gblock is useful when conserved regions in sequences are sufficiently rich in phylogenetic information. However, it can be detrimental when insertions and deletions in some genes are very informative, as previously shown with parasites such as some mobile genetic elements and viruses (Redelings and Suchard, 2007; Dessimoz and Gil, 2010; Bigot et al., 2013). Associated with the use of Baliphy to confirm the results, we therefore think that the use of Gblock is useless for numerous cases of molecular taxonomy in virology. Second, definition of a group of core genes is something that must be considered carefully because it does not only depend on the ubiquity of the group within a virus taxon. Indeed, we found that while the presence of orthologues of DpAV4 ORFs 048, 103 and 105 was strongly conserved in viral genomes of *Marseilleviridae*, *Iridoviridae* and *Ascoviridae*, these genes have likely been subjected to lateral transfers during evolution. The success of these conservative replacements means that these genes cannot be considered and used as core genes, since their evolution does not follow those of the rest of the core genes in their viral genomes. Third, we found two consistent results with our dataset of seventeen core genes. The first is that the majority of genes used here yield trees that have a topology that matches that of the core protein tree (Fig. 4). The second is that proteins encoded by orthologues of IIV6 ORFs 184R and 274L give consistent trees with a different topology, i.e., indicating that ascoviruses emerged before the differentiation of IIVs into four lineages. We suggested above that this might be because these proteins are virion components co-evolving under constraints related to the capsid shape. However, this explanation does not refer to all virion proteins, since phylogenetic trees calculated with virion proteins encoded by orthologues of IIV6 ORFs 295L and 347L have a topology matching that of the core protein tree.

Finally, we also found one ubiquitous protein, orthologues of IIV6 ORF 098R, for which we failed to obtain consistent trees, perhaps owing to the properties of these proteins being large, about 510 and 1050 amino acid residues, respectively. Our hypothesis is that these multi-domain proteins might have been affected by lateral domain exchanges of variable sizes during evolution that were conservative and retained the activity of the chimeric proteins. The variability of the extent of such events in affected genes would prevent their location when phylogenetic trees are calculated. However, they are known to exist in families of gene paralogues that are widespread in the superfamily Megavirales [for example, the *bro*-like family (Bideshi et al., 2003), ORF paralogues of IIV6-006L and IIV6-261R], and also exemplified herein with protein orthologues encoded by IIV6 ORF 343L, and with orthologues of SfAV1a ORF 048. Overall, the definition of a dataset of core genes for virus taxonomy is a subjective exercise, since it will always depend both on the sequences available and the polythetic properties of viruses (Van Regenmortel et al., 2013).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2014.1 2.013.

References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), Proceedings of 2nd International Symposium on Information Theory, Budapest, Hungary, Akademiai-Kiado, pp. 267–281.

Bideshi, D.K., Renault, S., Stasiak, K., Federici, B.A., Bigot, Y., 2003. Phylogenetic analysis and possible function of bro-like genes, a multigene family widespread among large double-stranded DNA viruses of invertebrates and bacteria. J. Gen. Virol. 84, 2531–2544.

Bigot, Y., Renault, S., Nicolas, J., Moundras, C., Demattei, M.V., Samain, S., Bideshi, D.K., Federici, B.A., 2009. Symbiotic virus at the evolutionary intersection of three types of large DNA viruses; iridoviruses, ascoviruses, and ichnoviruses. PLoS ONE 4, e6397.

Bigot, Y., Asgari, S., Bideshi, D.K., Cheng, X., Federici, B.A. & Renault, S., 2011. Family ascoviridae. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), Viral Taxonomy, IX Report of the International Committee on the Taxonomy of Viruses, third ed., Elsevier – Academic Press, London.

Bigot, Y., Piégu, B., Casteret, S., Gavory, F., Bideshi, D.K., Federici, B.A., 2013. Characteristics of inteins in invertebrate iridoviruses and factors controlling insertion in their viral hosts. Mol. Phylogenet. Evol. 67, 246–254.

Boyer, M., Yutin, N., Pagnier, I., Barrassi, L., Fournous, G., Espinosa, L., Robert, C., Azza, S., Sun, S., Rossmann, M.G., Suzan-Monti, M., La Scola, B., Koonin, E.V., Raoult, D., 2009. Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. Proc. Natl. Acad. Sci. USA 106, 21848–21853.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552.

Cheng, X.W., Carner, G.R., Brown, T.M., 1999. Circular configuration of the genome of ascoviruses. J. Gen. Virol. 80, 1537–1540.

Colson, P., De Lamballerie, X., Yutin, N., Asgari, S., Bigot, Y., Bideshi, D.K., Cheng, X.W., Federici, B.A., Van Etten, J.L., Koonin, E.V., La Scola, B., Raoult, D., 2013. "Megavirales", a proposed new order for eukaryotic nucelocytoplasmic large DNA viruses. Arch. Virol. 158, 2517–2521.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27, 1164–1165.

Delhon, G., Tulman, E.R., Afonso, C.L., Lu, Z., Becnel, J.J., Moser, B.A., Kutish, G.F., Rock, D.L., 2006. Genome of invertebrate iridescent virus type 3 (mosquito iridescent virus). J. Virol. 80, 8439–8449.

Dereeper, A., Audic, S., Claverie, J.M., Blanc, G., 2010. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. BMC Evol. Biol. 10, 8.

Dessimoz, C., Gil, M., 2010. Phylogenetic assessment of alignments reveals neglected tree signal in gaps. Genome Biol. 11, R37.

Eaton, H.E., Metcalf, J., Penny, E., Tcherepanov, V., Upton, C., Brunetti, C.R., 2007. Comparative genomic analysis of the family Iridoviridae: re-annotating and defining the core set of iridovirus genes. Virol. J. 4, 11.

Fischer, M.G., Allen, M.J., Wilson, W.H., Suttle, C.A., 2010. Giant virus with a remarkable complement of genes infects marine zooplankton. Proc. Natl. Acad. Sci. USA 107, 19508–19513.

Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. 27, 221–224.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.

Ince, I.A., Boeren, S.A., van Oers, M.M., Vervoort, J.J., Vlak, J.M., 2010. Proteomic analysis of Chilo iridescent virus. Virology 405, 253–258.

Iyer, L.M., Balaji, S., Koonin, E.V., Aravind, L., 2006. Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. Virus Res. 117, 156–184.

Jakob, N.J., Müller, K., Bahr, U., Darai, G., 2001. Analysis of the first complete DNA sequence of an invertebrate iridovirus: coding strategy of the genome of Chilo iridescent virus. Virology 286, 182–196.

Jancovich, J.K., Chinchar, V.G., Hyatt, A., Miyazaki, T., Williams, T., Zhang, Q.Y., 2011. Family iridoviridae. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), Viral Taxonomy IX Report of the International Committee on the Taxonomy of Viruses, third ed. Elsevier – Academic Press, London.

Legendre, M., Bartoli, J., Shmakova, L., Jeudy, S., Labadie, K., Adrait, A., Lescot, M., Poirot, O., Bertaux, L., Bruley, C., Couté, Y., Rivkina, E., Abergel, C., Claverie, J.M., 2014. Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. Proc. Natl. Acad. Sci. USA 111, 4274–4279.

Philippe, N., Legendre, M., Doutre, G., Couté, Y., Poirot, O., Lescot, M., Arslan, D., Seltzer, V., Bertaux, L., Bruley, C., Garin, J., Claverie, J.M., Abergel, C., 2013. Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. Science 341, 281–286.

Piégu, B., Guizard, S., Spears, T., Cruaud, C., Couloux, A., Bideshi, D.K., Federici, B.A., Bigot, Y., 2013. Complete genome sequence of invertebrate iridovirus IIV22 isolated from a blackfly larvae. J. Gen. Virol. 94, 2112–2116.

Piégu, B., Guizard, S., Spears, T., Cruaud, C., Couloux, A., Bideshi, D.K., Federici, B.A., Bigot, Y., 2014a. Complete genome sequence of invertebrate virus IIV25 isolated from a blackfly larvae. Arch. Virol. 159, 1181–1185.

Piégu, B., Guizard, S., Spears, T., Cruaud, C., Couloux, A., Bideshi, D.K., Federici, B.A., Bigot, Y., 2014b. Complete genome sequence of invertebrate iridovirus IIV30 isolated from the corn earworm, *Helicoverpa zea*. J. Invertebr. Pathol. 116, 43–47.

Piégu, B., Guizard, S., Tan, Y., Cruaud, C., Couloux, A., Bideshi, D.K., Federici, B.A., Bigot, Y., 2014c. Complete genome sequence of invertebrate iridovirus IIV22A, a variant of IIV22, isolated originally from a blackfly larva. Stand. Genomic Sci. 9, 940–947.

Piégu, B., Guizard, S., Tan, Y., Cruaud, C., Asgari, S., Bideshi, D.K., Federici, B.A., Bigot, Y., 2014d. Genome sequence of a crustacean iridovirus, IIV31, isolated from the pill bug, *Armadillidium vulgare*. J. Gen. Virol. 95, 1585–1590.

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., Claverie, J.M., 2004. The 1.2-megabase genome sequence of Mimivirus. Science 306, 1344–1350.

Redelings, B.D., Suchard, M.A., 2007. Incorporating indel information into phylogeny estimation for rapidly emerging pathogens. BMC Evol. Biol. 7, 40.

Stasiak, K., Renault, S., Demattei, M.V., Bigot, Y., Federici, B.A., 2003. Evidence for the evolution of ascoviruses from iridoviruses. J. Gen. Virol. 84, 2999–3009.

Suchard, M.A., Redelings, B.D., 2006. BAli-Phy: simultaneous Bayesian inference of alignment and phylogeny. Bioinformatics 22, 2047–2048.

Tan, Y., Spears, T., Bideshi, D.K., Johnson, J.J., Hice, R., Bigot, Y., Federici, B.A., 2009a. P64, a novel major virion DNA-binding protein potentially involved in condensing the *Spodoptera frugiperda* ascovirus 1a genome. J. Virol. 83, 2708–2714.

Tan, Y., Bideshi, D.K., Johnson, J.J., Bigot, Y., Federici, B.A., 2009b. Proteomic analysis of the *Spodoptera frugiperda* ascovirus 1a virion reveals 21 proteins. J. Gen. Virol. 90, 359–365.

Thomas, V., Bertelli, C., Collyn, F., Casson, N., Telenti, A., Goesmann, A., Croxatto, A., Greub, G., 2011. Lausannevirus, a giant amoebal virus encoding histone doublets. Environ. Microbiol. 13, 1454–1466.

Van Etten, J.L., Lane, L.C., Dunigan, D.D., 2010. DNA viruses: the really big ones (giruses). Annu. Rev. Microbiol. 64, 83–99.

Van Regenmortel, M.H., Ackermann, H.W., Calisher, C.H., Dietzgen, R.G., Horzinek, M.C., Keil, G.M., Mahy, B.W., Martelli, G.P., Murphy, F.A., Pringle, C., Rima, B.K., Skern, T., Vetten, H.J., Weaver, S.C., 2013. Virus species polemics: 14 senior virologists oppose a proposed change to the ICTV definition of virus species. Arch. Virol. 158, 1115–1119.

Williams, T., 1994. Comparative studies of iridoviruses: further support for a new classification. Virus Res. 33, 99–121.

Williams, T., Cory, J.S., 1994. Proposals for a new classification of iridescent viruses. J. Gen. Virol. 75, 1291–1301.

Wong, C.K., Young, V.L., Kleffmann, T., Ward, V.K., 2011. Genomic and proteomic analysis of invertebrate iridovirus type 9. J. Virol. 85, 7900–7911.

Xue, J.L., Cheng, X.W., 2011. Comparative analysis of a highly variable region within the genomes of Spodoptera frugiperda ascovirus 1d (SfAV-1d) and SfAV-1a. J. Gen. Virol. 92, 2797–2802.

Yutin, N., Wolf, Y.I., Raoult, D., Koonin, E.V., 2009. Eukaryotic large nucleocytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. Virol. J. 6, 223.