Available online at www.sciencedirect.com
SCIENCE DIRECT
Virus Research 117 (2006) 156–184
Virus Research
www.elsevier.com/locate/virusres

Evolutionary genomics of nucleo-cytoplasmic large DNA viruses

Lakshminarayan M. Iyer, S. Balaji, Eugene V. Koonin, L. Aravind *

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

Available online 21 February 2006

Abstract

A previous comparative-genomic study of large nuclear and cytoplasmic DNA viruses (NCLDVs) of eukaryotes revealed the monophyletic origin of four viral families: poxviruses, asfarviruses, iridoviruses, and phycodnaviruses [Iyer, L.M., Aravind, L., Koonin, E.V., 2001. Common origin of four diverse families of large eukaryotic DNA viruses. J. Virol. 75 (23), 11720–11734]. Here we update this analysis by including the recently sequenced giant genome of the mimiviruses and several additional genomes of iridoviruses, phycodnaviruses, and poxviruses. The parsimonious reconstruction of the gene complement of the ancestral NCLDV shows that it was a complex virus with at least 41 genes that encoded the replication machinery, up to four RNA polymerase subunits, at least three transcription factors, capping and polyadenylation enzymes, the DNA packaging apparatus, and structural components of an icosahedral capsid and the viral membrane. The phylogeny of the NCLDVs is reconstructed by cladistic analysis of the viral gene complements, and it is shown that the two principal lineages of NCLDVs are comprised of poxviruses grouped with asfarviruses and iridoviruses grouped with phycodnaviruses-mimiviruses. The phycodna-mimivirus grouping was strongly supported by several derived shared characters, which seemed to rule out the previously suggested basal position of the mimivirus [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 (5700), 1344–1350]. These results indicate that the divergence of the major NCLDV families occurred at an early stage of evolution, prior to the divergence of the major eukaryotic lineages. It is shown that subsequent evolution of the NCLDV genomes involved lineage-specific expansion of paralogous gene families and acquisition of numerous genes via horizontal gene transfer from the eukaryotic hosts, other viruses, and bacteria (primarily, endosymbionts and parasites). Amongst the expansions, there are multiple families of predicted virus-specific signaling and regulatory domains. Most NCLDVs have also acquired large arrays of genes related to ubiquitin signaling, and the animal viruses in particular have independently evolved several defenses against apoptosis and immune response, including growth factors and potential inhibitors of cytokine signaling. The mimivirus displays an enormous array of genes of bacterial provenance, including a representative of a new class of predicted papain-like peptidases. It is further demonstrated that a significant number of genes found in NCLDVs also have homologs in bacteriophages, although a vertical relationship between the NCLDVs and a particular bacteriophage group could not be established. On the basis of these observations, two alternative scenarios for the origin of the NCLDVs and other groups of large DNA viruses of eukaryotes are considered. One of these scenarios posits an early assembly of an already large DNA virus precursor from which various large DNA viruses diverged through an ongoing process of displacement of the original genes by xenologous or non-orthologous genes from various sources. The second scenario posits convergent emergence, on multiple occasions, of large DNA viruses from small plasmid-like precursors through independent accretion of similar sets of genes due to strong selective pressures imposed by their life cycles and hosts.
Published by Elsevier B.V.

Keywords: DNA viruses; Evolution; Poxvirus; Capping enzyme; Triphosphatase; Iridovirus; Coccolithovirus; Phycodnavirus; Primase; Origin of viruses; DNA replication; Origin of DNA replication; Papain-like peptidase; Sugar metabolism; Ubiqutin signaling; Growth factors; Cytokine signaling; Apoptosis

1. Introduction

The origin(s) of viruses had been a topic of intense speculation and debate ever since their discovery (Gibbs et al., 1995; Koonin, 1992). With the first biochemical studies on viruses, it became clear that only two common features were shared by all viruses: (1) their obligate intracellular parasitism; and (2) their virion architecture comprised of a genomic nucleic acid, typically of a single type (either RNA or DNA), packaged into a protein capsid, which in some cases is further associated with outer or inner lipid membranes (Gibbs et al., 1995). Beyond these general features, viruses show tremendous diversity in every respect, including genome size and organization, capsid architecture, mechanisms of propagation, and interactions with host cells. Viruses infect organisms from all three superkingdoms of life (bacteria, archaea, and eukaryotes) and replicate in all known

* Corresponding author. Tel.: +1 301 594 2445; fax: +1 301 480 9241.
E-mail address: aravind@ncbi.nlm.nih.gov (L. Aravind).

0168-1702/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.virusres.2006.01.009

cell types (Wagner and Hewlett, 2003). The extreme diversity of viruses suggests that they must have had multiple evolutionary origins, and the common features observed in all viruses reflect convergences emerging from adaptations to intracellular parasitism. The times and the modes of origins of the various types of viruses and their relationships to cellular genomes remain major issues of debate among evolutionary biologists. Broadly, the early theories of viral origins could be placed in two categories. The first of these sought to place the viruses in the earliest phases of life's evolution and associated them with the primitive precursors of cellular systems (Alstein, 1992; Gibbs et al., 1995). The second group of theories saw viruses as secondary derivatives of cellular systems that underwent drastic degeneration as a consequence of extreme parasitism, or "break away" elements from cellular genomes that survived as minimal parasitic replicons (Gibbs et al., 1995). The two groups of theories are not mutually exclusive: conceivably, some classes of viruses could be primordial whereas others could be later derivatives of "break away" elements from cellular systems. The advent of the first complete genome sequences of viruses did not resolve these debates entirely, but threw considerable light on the actual diversity in the coding capacity of various viruses, the affinities between different viral groups and homologies between viral genes and those of cellular organisms.

The first decade of viral comparative genomics revealed several major assemblages of viruses that were unified on the basis of the evolutionarily conserved proteins of their replication apparatus. Firstly, it became clear that the retroviruses, together with their various relatives such as the hepadnaviruses, plant badnaviruses, and tungroviruses, and the diverse retroposons shared a common ancestor, which encoded a reverse transcriptase (RT) as their principal replication polymerase (Xiong and Eickbush, 1990). The RNA-dependent RNA polymerases (RDRP) of diverse positive strand RNA viruses and several double-stranded(ds) RNA viruses were likewise unified, indicating a common origin for this entire assembly of viruses (Kamer and Argos, 1984; Koonin et al., 1989). At a deeper level, the RTs and RDRPs have been shown to descend from an ancestral replicase that utilized an RNA template (Delarue et al., 1990; Kamer and Argos, 1984; Poch et al., 1989; Xiong and Eickbush, 1990), suggesting that at least these two major classes of viruses might have ultimately descended from an ancient replicon with an RNA genome. This unification also suggested that the diversification of these viruses might be linked to one of the fundamental evolutionary transitions from RNA genomes to the DNA genomes (Forterre, 2002; Leipe et al., 1999; Wintersberger and Wintersberger, 1987).

Similarly, certain assemblages sharing common replication systems also became apparent amongst the DNA viruses. In particular, many small DNA viruses and related plasmids and transposons were unified on the basis of a shared rolling circle replication endonuclease (RCRE), which initiates the eponymous form of replication of these elements (Ilyina and Koonin, 1992; Iyer et al., 2005; Kapitonov and Jurka, 2001). However, the relationships among large dsDNA viruses that have complex genomes with dozens or even hundreds of genes remained far more difficult to elucidate. Amongst the bacteriophages, several major monophyletic groups, such as the lambdoid phages, were identified (Hendrix, 2003). Among the animal large dsDNA viruses, the families *Herpesviridae*, *Baculoviridae*, and *Poxviridae* are obviously monophyletic. The common ancestors of each of these families have been partially reconstructed and, in each case, inferred to have had over 50 genes (Davison et al., 2005; Hughes and Friedman, 2005; Lauzon et al., 2005; McLysaght et al., 2003). Thus, the common ancestral forms of these viral families seem to have already attained considerable complexity—the salient features of replication, gene expression and virion architecture apparently emerged early in their evolution and were retained over vast evolutionary time spans. In contrast, higher-order relationships between various groups of large eukaryotic DNA viruses, if any, remained uncertain. In our previous work, we addressed this issue through comprehensive comparative analysis of the protein sequences encoded by large eukaryotic DNA viruses, followed by cladistic analysis using a character matrix based on the conserved features of these proteins (Iyer et al., 2001). This analysis produced evidence of common ancestry of several families of large eukaryotic DNA viruses, including the animal poxviruses, iridoviruses, and asfarviruses (with a single representative, the African Swine Fever Virus, ASFV), and the phycodnaviruses, which infect phylogenetically diverse algae.

We named this major, monophyletic assemblage of large eukaryotic DNA viruses the Nucleo-Cytoplasmic Large DNA Virus (NCLDV) clade as they either replicate exclusively in the cytoplasm of the host cell or start their life cycle in the host nucleus but complete it in the cytoplasm. Typically, the NCLDVs do not exhibit much dependence on the host replication or transcription systems for completing their replication because, even in viruses like *Paramecium bursaria* Chlorella virus (PBCV), which initiate replication in the nucleus, disruption of a functional host nucleus by irradiation does not abrogate replication (Van Etten et al., 1986). This relative independence of the NCLDVs from the host cells is consistent with the fact that all these viruses encode several conserved proteins performing most key life-cycle processes, such as DNA polymerases, helicases, and DNA clamps for DNA replication, Holliday junction resolvases and topoisomerases for genome manipulation, transcription factors involved in transcription initiation and elongation, ATPase pumps for DNA packaging, and chaperones involved in the capsid assembly (Iyer et al., 2001). In the original analysis, this conserved core was found to include 9 proteins shared by all families of NCLDVs and 22 additional proteins shared by at least three of the four families (Iyer et al., 2001). This suggested that all extant NCLDV families have descended from a common ancestor that already had a fairly complex gene repertoire and was capable of completing its replication cycle in relative autonomy from the cell.

Subsequent to the original description of the NCLDV group, several major developments have occurred, the chief among them being sequencing of the 1.2-megabase genome of the gigantic *Acanthamoeba polyphaga* Mimivirus (Raoult et al., 2004). Analysis of the mimivirus genome showed that it was a new branch of the NCLDV group. In addition, this largest known viral genome contains numerous multi-gene families as

well as genes that might have been accrued by the viral genome via extensive horizontal gene transfer (HGT) (Desjardins et al., 2005; Koonin, 2005; Raoult et al., 2004). Additionally, the genomes of several new vertebrate iridoviruses have been published and shown to contain many genes beyond those found in the originally sequenced isolate of fish lymphocystis disease virus (Do et al., 2004; He et al., 2001, 2002; Jancovich et al., 2003; Song et al., 2004; Tsai et al., 2005). Concomitantly, there have been several advances in the sequence analysis of the viral proteins, including the prediction of the replicative primase of the NCLDVs and its relationship to the herpesvirus primases (Iyer et al., 2005). The accumulating data on phage genomes have also provided additional material to compare diverse large DNA viruses.

In light of this new information, we herewith revisit the NCLDVs to address several major issues relevant for the evolution of this group of viruses: (i) new support for the monophyly of the NCLDV clade; (ii) reconstruction of key biological features of different NCLDV lineages using comparative genomics; (iii) contributions of lineage-specific expansions of gene families and gene accretion, via HGT from hosts and co-occurring symbionts and parasites, to the genomic growth of large DNA viruses; (iv) the relationship between NCLDVs and other large DNA viruses, phages, and plasmids; (v) the implications of the emerging picture of the evolution of NCLDVs and other large DNA viruses for the origins of cellular life.

2. Re-examination of the NCLDV phylogeny and derivation of core gene sets for different NCLDV clades

To re-evaluate the original results concerning the monophyly and evolutionary radiation of the NCLDVs in light of the new genome sequences, we performed a systematic analysis of the proteins encoded by the mimivirus and the following iridoviruses: the new Chinese isolate of lymphocystis disease virus, Singapore Grouper virus, Rock Bream iridovirus, Infectious spleen and kidney necrosis virus, Frog virus 3, *Ambystoma tigrinum stebbensi* virus, and Chilo iridescent virus. The new LDV isolate has 70–90 additional genes, which were not found in the originally sequenced LDV isolate, but are often present in other iridoviruses. Accordingly, we used this strain as it is a more representative form of this virus. The genome of a phycodnavirus infecting the prymnesiophyte (haptophyte) alga *Emiliania huxleyi* [EHV; (Wilson et al., 2005)] was released when the present manuscript was being finalized. Therefore, we could not include the EHV genome in the cladistic analyses; nevertheless, analysis of the predicted protein sequences of this virus was performed to identify interesting features relevant to the overall description of the NCLDVs. The *Feldmannia irregularis* virus, another phycodnavirus, was also analyzed but not used in any further comparisons because it is closely related to the *Ectocarpus siliculosus* virus and did not provide any additional useful characters. The viruses included in the analysis described here are: Chordopoxviruses (Vaccinia virus, VV; Molluscum Contagiosum virus, MCV; Fowlpox virus, FPV), Entomopoxviruses (*Amsacta moorei* Virus, AMV; *Melanoplus sanguinipes* Virus, MSV), Asfarviruses (African Swine Fever Virus, ASFV), Fish iridoviruses (lymphocystis disease virus Chinese isolate, LDV; Singapore Grouper virus, SGV; Rock Bream iridovirus, RBV), Amphibian iridoviruses (Frog virus 3, FV3; *Ambystoma tigrinum stebbensi* virus, ATSV), Insect iridoviruses (Chilo iridescent virus, CIV), Phycodnaviruses (Paramecium bursaria Chlorella Virus, PBCV; *Ectocarpus siliculosus* Virus, ESV; and *Emiliania huxleyi* virus-EHV), and mimivirus (*Acanthamoeba polyphaga* mimivirus). All conserved proteins shared by at least a pair of viruses were identified using a combination of clustering with the BLASTCLUST program (ftp://ftp.ncbi.nih.gov/blast/documents/README.bcl) and sequence profile searches with PSIBLAST (Altschul et al., 1997). Clustering of the entire set of NCLDV proteins was also carried out to identify lineage-specific expansions of protein families. All proteins were further investigated using PSI-BLAST position-specific scoring matrices to identify potential conserved domains; identification of such domains often leads to new insights into the functions of the respective proteins (Aravind and Koonin, 1999b). These features were used to develop a standard annotation of the protein complements of the NCLDVs. The evolutionary affinities of viral proteins with their homologs from other viruses and cellular organisms were assessed where feasible by using conventional phylogenetic trees constructed with neighbor-joining, minimum evolution, and maximum likelihood methods (Felsenstein, 2004).

The sequenced NCLDV genomes contain from ~150 (LDV) to ~900 (the mimivirus) predicted protein-coding genes. We re-evaluated the coding potential of the available NCLDV genomes by removing short open reading frames (ORFs) that did not have detectable homologs in the current databases, ORFs with completely biased amino acid composition, and ORFs which overlapped with well-defined genes but lacked homologs, and plotting the resulting number of predicted genes against the genome size. The numbers of genes identified by this conservative approach show a good linear fit ($R^2 = 0.94$; Fig. 1). Thus, current annotations of some of the NCLDV genomes are likely to be inflated as a result of inclusion of spurious ORFs. These obser-

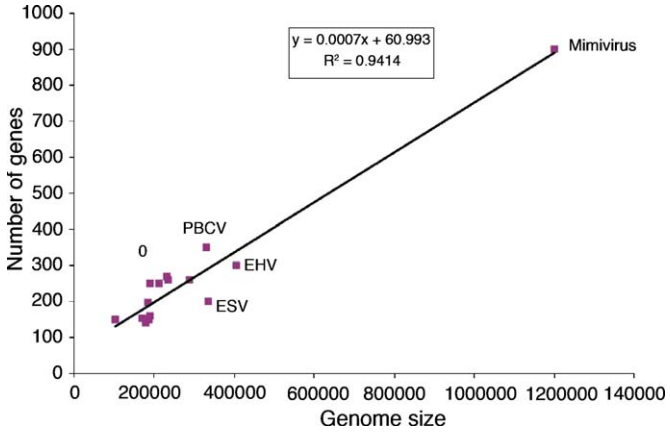

Fig. 1. Linear correlation between the number of predicted genes and the genome size (in nucleotides) in NCLDVs. The number of genes in each genome was corrected by removing short, compositionally biased ORFs, and predicted ORFs that overlap with well-defined genes but lack homologs.

vations suggest that the NCLDVs have a nearly constant gene density, which is indicative of similar selective forces affecting the gene organization of these viruses across a wide genome and proteome size range, and the enormous range of hosts infected by these viruses. The fundamental constraint of arranging intronless genes, with each having its own distinct promoters, and very few additional regulatory elements, seems to be the primary factor behind the observed pattern.

Having obtained, through the comprehensive protein sequence comparison, the conserved characters for the NCLDV assemblage, including the new sequences, we rebuilt the phylogenetic tree of NCLDVs using an exhaustive search for the most

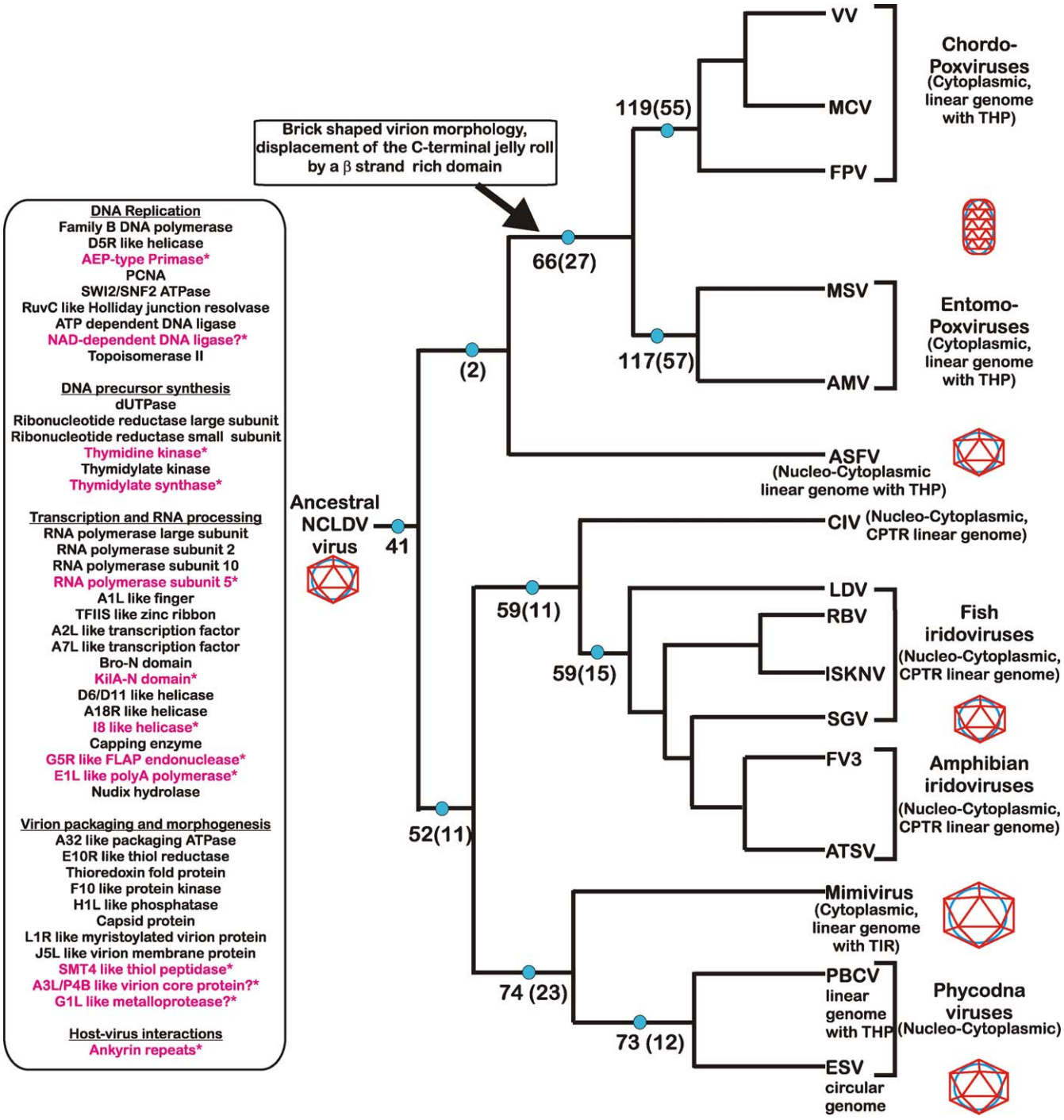

Fig. 2. A phylogenetic tree of the NCLDVs built on the basis of conserved gene set analysis. The tree topology is based on the consensus cladogram derived by cladistic analysis. The number of proteins reconstructed as being present in the ancestral core of a clade of viruses is shown next to the blue circles. Shown in brackets are the numbers of proteins that are unique to a particular clade. Proteins that are predicted to be part of the ancestral NCLDV genome are shown on the left. Protein names in red and marked with an asterisk represent members of the ancestral genes set that have not been identified in our previous reconstruction (Iyer et al., 2001). The transcription factor A7L protein was formerly called the ASFV-B385R-like protein (abbreviations: THP, Terminal hairpin; TIR, Terminal inverted repeats; CPTR, Circularly permuted terminally redundant).

parsimonious trees implemented in the PAUP program package (Swofford, 2000) (Fig. 2). The new data further strengthened the monophyly of the NCLDV assemblage, including the mimivirus, with at least 41 proteins (10 additional proteins beyond the originally defined set) traceable to their last common ancestor (Fig. 2). These proteins belong to a wide range of functional classes related to every aspect of the viral life cycle and morphogenesis and are discussed below in the context of individual functional systems. However, the addition of the mimivirus data substantially changed the internal relationships within the NCLDV class. A clade comprised of phycodnaviruses and mimivirus was strongly supported by 23 synapomorphies—unique proteins that are not found in any other NCLDV. Overall, at least 74 proteins were confidently assigned to the last common ancestor of the phycodnavirus-mimivirus clade. The monophyly of the phycodnaviruses within this clade was supported by another set of 12 proteins that are unique to these viruses (Fig. 2). This reconstruction suggests a large ancestral DNA virus, which appears to have had a number of specific adaptations that allowed it to spread widely and infect a range of phylogenetically diverse protists.

However, perhaps, unexpectedly, the new analysis did not recover any support for a clade uniting all animal NCLDV lineages. Instead, a set of 11 proteins was identified that supported a higher-order clade consisting of the iridoviruses and the mimivirus + phycodnavirus clade (Fig. 2). The alternative grouping of the three animal NCLDV lineages resulted in only three synapomorphies (unique proteins) supporting their monophyly; furthermore, there was no support for the grouping of either ASFV or the poxviruses with iridoviruses. As in the previous analysis (Iyer et al., 2001), the most parsimonious tree had ASFV and poxviruses as sister groups, but this node was only weakly supported and should be considered tentative in the absence of further evidence. This phylogeny supports an early radiation of the major NCLDV groups in protists as opposed to the emergence of all animal NCLDVs from a single, ancestral animal virus. Among the animal viruses, the evolution of the asfarviruses remains largely intractable as ASFV is the only representative of this family. In contrast, the evolution of poxviruses and iridoviruses can be reconstructed in greater detail thanks to the availability of genome sequences from several taxa infecting diverse animals. The common ancestor of all known poxviruses can be reconstructed as possessing at least 66 genes, while that of the known iridoviruses was inferred to contain at least 59 genes. Within both poxviruses and iridoviruses, there is strong support for the monophyly of the viruses infecting vertebrates, with the ancestor of the chordopoxviruses predicted to encode at least 119 proteins and that of the vertebrate iridoviruses at least 59 proteins. Likewise, the monophyly of the entomopoxviruses is strongly supported with a set of 117 proteins predicted for their last common ancestor. Anecdotal evidence from an incompletely sequenced insect iridovirus, the ascovirus, similarly supports a monophyletic arthropod iridovirus lineage. Within the vertebrates, the phylogenies of poxviruses seemed to recapitulate the phylogeny of the animal hosts (Fig. 2). However, the amphibian iridoviruses are closer to the fish SGV, to the exclusion of the other fish virus LDV. These observations are best compatible with a single invasion of the animals by the ancestors of poxviruses and iridoviruses each. Whereas the poxviruses appear to have subsequently followed a vertical co-evolution with the host, the vertebrate iridoviruses might have spread across different vertebrate hosts sharing the same aquatic environment.

Although these reconstructions assign a considerable number of genes to the last common ancestor of the extant NCLDVs, the reconstructed cores of both this progenitor and the ancestors of individual clades are considerably smaller than the proteomes of the extant viruses. Thus, evolution of the NCLDVs must have included either growth in the size of their genomes throughout their existence, or extensive turnover of some genes of an already gene-rich ancestor, except those belonging to the relatively stable core. These two trends are not mutually exclusive, and indeed, below we present evidence for gene accretion, lineage-specific expansions, as well as loss and displacement of many genes throughout the history of the NCLDV genomes.

3. Core functional systems of NCLDVs and their elaboration in different viral lineages

To place in a biological context the conserved proteins shared by all or most of the NCLDVs and inferred to have been present in their common ancestor, we discuss below the reconstructions and subsequent elaborations of the ancestral viral functional systems.

3.1. DNA replication

All NCLDVs share a DNA polymerase of the B family, which, like the principal replicative polymerases of archaea and eukaryotes, contains the polymerase catalytic domain fused to an N-terminal $3' \rightarrow 5'$ exonuclease domain (Leipe et al., 1999). The NCLDVs, except for the entomopoxviruses, also encode a PCNA-like DNA clamp. The poxvirus version of the clamp protein (vaccinia G8R) is extremely divergent (Iyer et al., 2001), which might be related to the additional role of this protein in transcription (Dellis et al., 2004; Iyer et al., 2001). In contrast to the DNA clamps, the clamp-loader ATPases, related to eukaryotic RF-Cs are present, often in multiple copies, only in the phycodnaviruses and the mimiviruses. The D5R-like ATPase, typified by the eponymous vaccinia protein essential for DNA replication, appears to be the replicative helicase that is conserved in all NCLDVs. The D5R family belongs to the helicase Superfamily (SF) III within the AAA+ ATPase class, which includes the primary replicative helicases of many other DNA and RNA viruses (Gorbalenya et al., 1990; Iyer et al., 2004b). Our recent analysis showed that the D5R helicases are distinguished from other SFIII family members by the presence of a unique N-terminal domain, the D5N domain (Iyer et al., 2005). The strict association between the D5N and the AAA+ ATPase domain in the D5R family suggests that D5N mediates recognition of the substrate and/or primer initiation sites by these proteins.

We recently showed that the N-terminal region of the poxvirus D5R contains a previously undetected primase domain related to the archaeo-eukaryote type primases (AEPs). Further analysis of the sequence of this domain led to the identification of

orthologous, even if highly diverged AEP domains in all other NCLDV and demonstrated a specific, probably monophyletic relationship between the (predicted) primases of NCLDV and herpesviruses (Iyer et al., 2005). These findings added a crucial component to the ancestral core set of proteins unifying the NCLDV group (Fig. 2). Fusions between these primases and D5R-like helicases are found in poxviruses, iridoviruses, and asfarviruses, while in the mimivirus, the D5R-like helicase and the primase are encoded by adjacent genes. This suggests that they function in close association with the D5R helicase in the initiation of DNA synthesis. The NCLDV-herpesvirus primase clade consists of two major families, namely, the iridovirus family and the herpes-pox family. The former family is characterized by the presence of the PriCT-2 domain C-terminal to the primase catalytic domain (Iyer et al., 2005) and is found in iridoviruses, ASFV, and the mimivirus. The herpes-pox family is typified by a cysteine-rich Zn-binding cluster in place of the PriCT-2 domain and is present in poxviruses, ASFV, phycodnaviruses, and the mimivirus. The simplest explanation for this unusual phyletic pattern (Fig. 3A) of two related families of primases amongst the NCLDV is that both versions existed in the ancestral virus, performing partially redundant roles. Subsequently, one or the other version was lost in most lineages, whereas ASFV and mimiviruses retained both versions. In addition to fusions with the D5R-like helicase, herpes-pox family primases are also fused to the UL9-like SFII helicases in ASFV and mimivirus, suggesting that they might perform distinct functions in conjunction with these mechanistically different helicases (Iyer et al., 2005).

Like the two distinct families of AEP-like primases, both the ATP- and NAD-dependent DNA ligases are found in different NCLDVs. The ATP-dependent ligases are seen in chordopoxviruses, ASFV, and phycodnaviruses, whereas entomopoxviruses, iridoviruses, and the mimivirus have the NAD-dependent ligase (Fig. 3A). Given the strictly complementary phyletic patterns of the two classes of ligases, they are likely to play equivalent roles in DNA replication. In phylogenetic analysis, all viral NAD-dependent versions form a well-supported monophyletic lineage to the exclusion of bacterial forms (data not shown). In contrast, the ATP-dependent ligases are far more divergent from each other and do not show strong affinities between the NCLDVs. These observations point to displacement of one or the other form through HGT between the viruses or from their hosts or co-occurring bacterial endosymbionts. The postulated existence of the ancestral NCLDV within the context of an early eukaryotic cell with its characteristic ATP-dependent ligases, as well as the high level of divergence of the ATP-dependent ligases among the NCLDVs favors them being the ancestral form.

A predicted Fen-1/FLAP-like $5' \rightarrow 3'$ endonuclease was detected in poxviruses (G5R), iridoviruses, EHV and the mimivirus, which makes it a likely part of the ancestral gene core of the NCLDVs (Fig. 2). The homologous nuclease in the phage T4 has been implicated in the removal of RNA primers synthesized by the primase during replication (Bhagwat and Nossal, 2001), suggesting a similar role for the NCLDV counterparts. However, recent experimental characterization of the poxvirus G5R gene failed to provide evidence for any such role in replication, and instead pointed to a role of this protein at an early stage of viral morphogenesis (da Fonseca et al., 2004). The homologous enzyme in herpesviruses (VHS) is involved in down-regulation of host genes through direct degradation of transcripts (Oroskar and Read, 1989; Taddeo et al., 2002), suggesting that this could be an alternative function of the Fen-1-like nucleases of NCLDVs.

3.2. DNA dynamics: recombination and repair

The segregation of replicated chromosomes and events related to packaging of viral DNA into capsids involve complex, dynamical processes, which include recombination. Consistent with the large genome size and relative autonomy, most of the NCLDVs possess multiple recombination enzymes. The RuvC-like Holliday junction resolvase (HJR) is encoded by poxviruses, iridoviruses, phycodnaviruses, and the mimivirus (MIMI_L451), indicating that it was the ancestral HJR of the NCLDV. The actual function of this protein in concatemer resolution has been demonstrated for vaccinia virus (Garcia et al., 2000). In ASFV, this resolvase is replaced by a ERCC4/Mus81-like nuclease (EP364R), which is related to the principal Holliday junction resolvase of the eukaryotes, Mus81 (Haber and Heyer, 2001). In addition to the RuvC-like HJR, iridoviruses, PBCV and the mimivirus also encode URI-family nucleases (Aravind et al., 1999). The mimiviral version is specifically related to Slx1p, the second eukaryotic HJR (Fricke and Brill, 2003). The role of this nuclease in the recombination processes of these viruses remains unclear. ASFV, phycodnaviruses, and the mimivirus also have a lambda-type exonuclease (Aravind et al., 2000; Iyer et al., 2001) that might be involved in processing DNA ends for strand exchange or single-strand annealing during recombination as observed in bacteriophages (Kovall and Matthews, 1997). The phyletic pattern of this nuclease suggests that it was definitely present in the common ancestor of the mimivirus-phycodnavirus clade, but whether or not it was present in the ancestral NCLDV remains uncertain.

Both Topoisomerase II (ASFV, iridoviruses, phycodnaviruses, and the mimivirus) and Topoisomerase IB (poxviruses and the mimivirus) are represented amongst the NCLDV. The phyletic patterns of these enzymes are roughly complementary, suggesting that they might have some degree of functional equivalence, and might have displaced each other in individual lineages (Fig. 3A). The wider distribution of TopoII implies that it was present in the ancestral NCLDV. Formally, the presence of TopoIB in the mimivirus might suggest that the ancestral NCLDV encoded this enzyme as well. However, the specific relationship between the mimivirus TopoIB and bacterial versions of this enzyme (data not shown) makes it equally plausible that TopoIB was acquired by NCLDVs on two independent occasions from co-occurring bacterial endosymbionts. ESV, *Feldmannia irregularis* virus and EHV have a protelomerase/ResT like enzyme, which are divergent members of the tyrosine-recombinase class of proteins which includes topoisomerase IB and transposon integrases (Huang et al., 2004b). Outside of the phycodnaviruses, they are widely observed

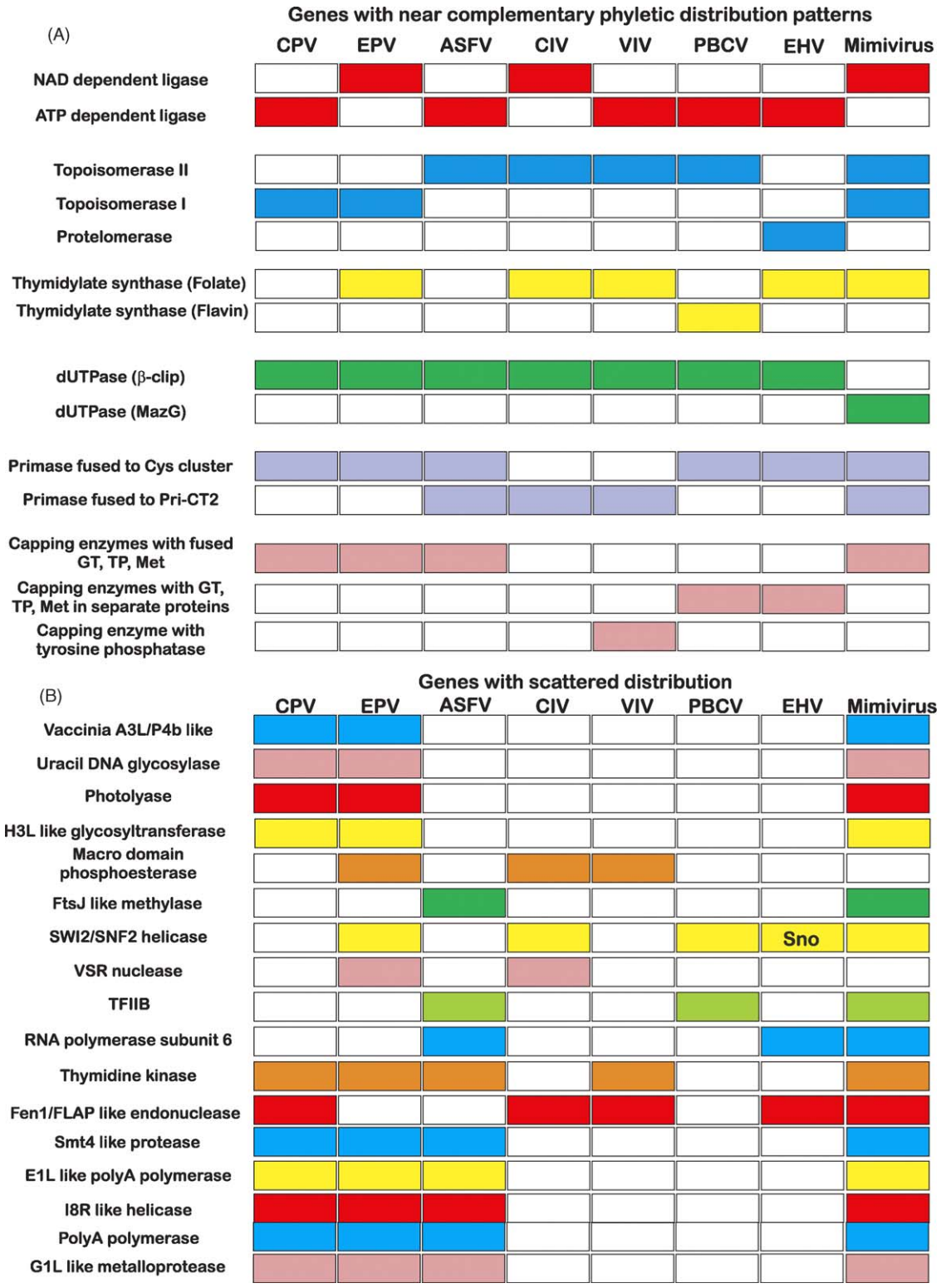

Fig. 3. Phyletic patterns of NCLDV genes. (A) Examples of complementary phyletic distributions among genes with the same function. Presence of a particular gene is denoted by a colored box, whereas absences are shown as empty boxes. (B) Examples of unexpected or sporadic distributions, which represent cases of either multiple gene losses or gains of the same gene. Absences are denoted by empty boxes. The set of genes chosen are only a subset of genes that show sporadic distributions between distant representatives of NCLDVs Sno: Strawberry notch like SWI2/SNF2 ATPase seen only in EHV. The thymidylate synthase of EHV is fused to dihydrofolate reductase. This fusion is characterisitic of its chromist host and might represent a recent case of xenologous gene displacement from the host. The Fringe-like glycosyltransferases present in all poxviruses (Vaccinia H3L) and mimivirus are required for adsorption to the cell surface heparan sulfate proteoglycan in poxviruses (da Fonseca et al., 2000; Lin et al., 2000). Phylogenetic analysis suggests that the ancestral poxvirus acquired this gene from a proteobacterium, but its rapid evolution makes it difficult to determine whether it was horizontally transferred into the ancestral NCLDV or once into the ancestral poxvirus and once into the mimivirus.

in bacteriophages, such as the coliphage N15, the *Klebsiella* phage phiKO2, and prophages from *Agrobacterium* and *Borrelia burgdorferii*. Hence, the NCLDV versions seem to have been acquired by a single recent transfer from a bacterial or phage source. In coliphage N15, the protelomerase cuts the circular phage DNA to form a linear genome with hairpin ends (Deneke et al., 2002; Ravin et al., 2001). Since all the NCLDVs that have the protelomerase have a circular genome, it is possible that this protein is involved in a linearization step, early in the infection process. Furthermore, these phycodnaviruses also lack the evolutionarily related enzyme, TopoIB (Fig. 3A), implying that they might potentially play a role equivalent to TopoIB in these viruses.

The SWI2/SNF2-like chromatin-remodeling ATPases of helicase SFII are sporadically encountered in representatives from most NCLDV lineages, suggesting that it might have even been present in the ancestral NCLDV. In EHV, there is a distinct version of the SWI2/SNF2 family belonging to the Strawberry Notch type of ATPases (Aravind et al., 2003a), which seems to have been acquired independently of the SWI2/SNF2s of the other NCLDVs, probably, from a bacterial source. Some iridoviruses (e.g., CIV) encode a RecN-like ATPase and the mimivirus has a SbcC-like ATPase (MIMI_R555), both of which belong to the SMC/Rad50 family of ABC ATPases involved in higher-order chromosome looping and condensation (Hirano, 2005). The evolutionary affinities of these proteins suggest that they were independently acquired from bacterial sources. Together, these observations support the notion that the ancestral NCLDV had a fairly large chromosome that required regulation of supercoiling, and, as the genome size grew, other global regulators of chromatin structure and chromosome organization also appear to have been recruited.

While there is no conserved DNA repair proteins seen in all the NCLDVs, several enzymes are shared by at least two lineages (Fig. 3B). Thus, photolyases and uracil DNA glycosylases are present in poxviruses and the mimivirus, whereas the Very Short Patch repair endonuclease is encoded by entomopoxviruses and CIV. The phyletic patterns and affinities of these enzymes do not support their presence in the ancestral NCLDV.

3.3. Nucleotide metabolism

All NCLDVs encode a nucleotide-metabolism apparatus centered on the pathway for providing dNTPs for DNA replication and repair. Four proteins could be traced back to the ancestral NCLDV, namely, the large and small subunits of the ribonucleotide reductase (RIR), thymidylate kinase, and dUTPase, which lowers the dUTP levels and provides dUMP for dTTP biosynthesis. Both RIR subunits are conserved in all NCLDVs, whereas the typical, β-clip fold dUTPase is present in all lineages except for the Mimivirus. A detailed analysis of the mimivirus genome revealed the presence of a MazG-type dUTPase (Moroz et al., 2005) suggesting that the ancestral β-clip fold dUTPase was displaced by the MazG-family enzyme, probably derived from a bacterial source (Fig. 3A). A similar displacement seems to have occurred in the case of thymidylate synthase (Fig. 3A): the classical, folate-dependent enzyme is present in some poxviruses and iridoviruses, and the mimivirus, whereas the recently characterized PBCV enzyme (Graziani et al., 2004) is the unrelated, flavin-dependent thymidylate synthase (Myllykallio et al., 2002) that is also seen in several bacteriophages. The poxviruses, ASFV, vertebrate iridoviruses and the mimiviruses encode thymidine kinases, which appear to be specifically related to each other, to the exclusion of other eukaryotic thymidine kinases in phylogenetic trees (data not shown). This suggests that the thymidine kinase was present in the ancestral NCLDV. The phycodnaviruses and the mimivirus also have a deoxycytidine deaminase suggesting that it was acquired by the ancestor of this clade. Smaller subsets of NCLDVs sporadically encode other nucleotide metabolism enzymes, such as the guanylate kinase of orthopoxviruses (a pseudogene in some viruses) and the deoxycytidine kinase of avian poxviruses. The mimivirus additionally has a more elaborate suit of enzymes for DNA precursor metabolism, including deoxynucleotide kinase (DONK), nucleoside diphosphate kinase, and GMP synthase. The mimivirus DONK appears to be specifically related to the equivalent enzymes of bacteriophages. Given that many of the large bacteriophages also encode similar nucleotide metabolism enzymes, it appears to be an ancient strategy of DNA viruses to subvert the host nucleotide precursor pool, including available ribonucleotides, for viral DNA synthesis. The frequent loss and acquisition of nucleotide metabolism enzymes during NCLDV evolution appear to reflect specific adaptations of viruses for the different types of cells in which they propagate. Thus, ESV and MCV replicate in actively dividing cells (Muller et al., 1998; Senkevich et al., 1996) where the viruses have easier access to cellular pools of deoxynucleotides than other NCLDVs, which might have triggered the loss of nucleotide metabolism enzymes in these viruses.

3.4. Transcription and chromatin modification

The conservation of the core of RNA polymerase subunits suggests that the ancestral NCLDV had a fairly complex RNA polymerase comparable to the cellular RNA polymerases of archaea and eukaryotes. In addition to the two largest catalytic subunits, the ancestral NCLDV also can be inferred to have encoded the archaeo-eukaryotic RNA polymerase subunits RPB10 and RPB5 (Fig. 2). All core RNA polymerase subunits are missing in PBCV and ESV, but are present in EHV and the mimivirus, the sister group of the phycodnaviruses. This suggests that PBCV and ESV made the transition from an entirely cytoplasmic replication cycle to one involving a transcriptionally active nuclear phase, thus obtaining access to the host RNA polymerases. This probably made the viral RNA polymerase subunits superfluous and triggered the loss of the respective genes as a group. The loss of RNA polymerase subunit genes in NCLDVs is not limited to PBCV and ESV; entomopoxviruses apparently have lost the two smaller subunits (RPB10 and RPB5), although, in this case, the biological underpinning is less clear. Several other orthologs of cellular RNA polymerase subunits are sporadically encountered in various lineages of NCLDVs, such as RPB6 in the mimivirus and ASFV, RPB3 in ASFV and EHV, RpoL in the mimivirus, and RPB10 and RPB11 in EHV.

Additionally, orthologs of several basal eukaryotic transcription factors are found in individual NCLDV lineages, including TBP and the TFIIB in the phycodnavirus-mimivirus clade, TFIIB in ASFV, $TAF_{II}18$ in ESV (ORF196), and HAP3 and TFIIF in LDV. Conceivably, the viruses sporadically acquired RNA polymerase subunits and basal transcription factors from their host, which might have increased their transcription efficiency in an entirely cytoplasmic niche.

Beyond the core transcription apparatus resembling that of the host, the NCLDV clade is defined by a unique set of three transcription factors (typified by Vaccinia proteins A1L, A2L, and A7L), of which A1L and A2L are conserved in all NCLDVs and A7L is missing in phycodnaviruses. These proteins have no detectable counterparts in eukaryotes or other viruses (Fig. 4A and B). These transcription factors might form a NCLDV-specific complex required for the recognition of viral promoters. The ancestral NCLDV also encoded a TFIIS-like Zn-ribbon protein that functions in transcription elongation. This gene has been lost in some phycodnaviruses (ESV and *Feldmannia irregularis* virus), which, like the loss of RNA polymerase subunits,

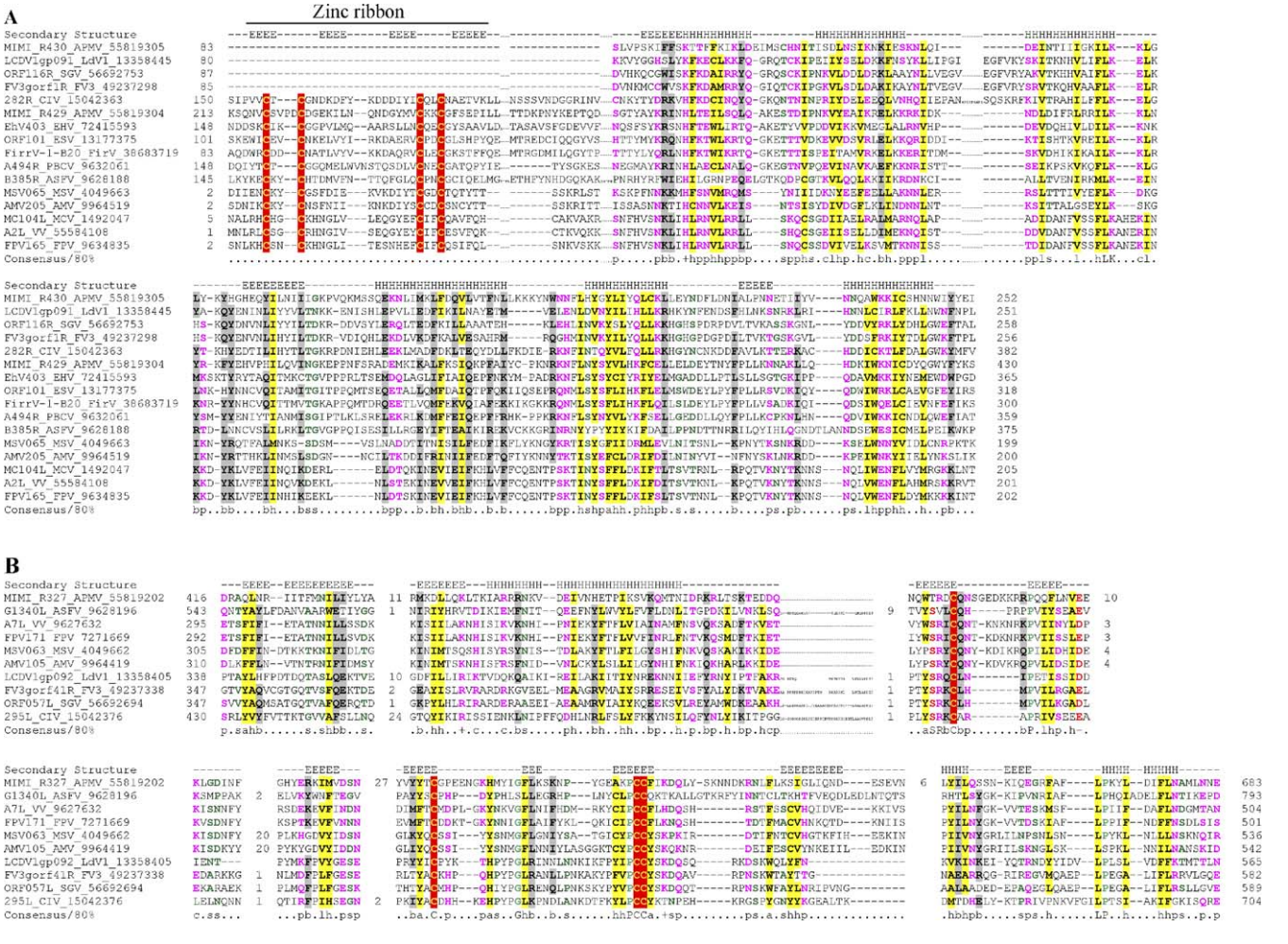

Fig. 4. (A) Multiple alignment of NCLDV-specific basal transcription factor A2L. The N-terminal Zn-ribbon domain shared by a subset of the A2L proteins is highlighted by reverse shading of the metal-chelating cysteines. (B) Multiple alignment of NCLDV-specific basal transcription factor A7L. The conserved cysteines that are characteristic of this transcription factor, which are predicted to chelate a metal ion in the central β-barrel domain of these proteins, are reverse-shaded. (C) The virus-specific 2-cysteine adaptor domain found fused to OTU/A20-like peptidases and S/T protein kinases. The two absolutely conserved cysteines are shown in reverse shading. (D) The small PBCV-specific basic adaptor domain found fused to S/T protein kinases and the 2-Cysteine domain. (E) A novel papain-like thiol peptidase found in the mimivirus, certain bacteria, and trypanosomes. The conserved cysteine and histidine residues characteristic of papain-like proteases is shown in reverse shading. Secondary structure predictions are shown above the alignment where E: strands and H: helix. Residues are colored based on their conservation in the alignment. Eighty percent was used to calculate the consensus. The coloring scheme and consensus abbreviations are as follows: h, hydrophobic residues (ACFILMVWY), l, aliphatic residues (VIL), and a, aromatic residues (FYWH) shaded yellow; b, big residues (LIYERFQKMW), shaded gray; s, small residues (AGSVCDN) and u, tiny residues (GAS), colored green; p, polar residues (STEDKRNQHC); +, basic residues (HRK) and −, acidic residues (DE), colored magenta. Species abbreviations are as follows: APMV: *Acanthamoeba polyphaga* mimivirus; ASFV: African swine fever virus; AMV: *Amsacta moorei* entomopoxvirus; Bant: *Bacillus anthracis*; Bcer: *Bacillus cereus*; Blic: *Bacillus licheniformis*; Bthu: *Bacillus thuringiensis*; CIV: Chilo iridescent virus; Cups: *Campylobacter upsaliensis*; Cwat: *Crocosphaera watsonii*; ESV: *Ectocarpus siliculosus* virus; EHV: *Emiliania huxleyi* virus; FPV: Fowlpox virus; FV3: Frog virus 3; FirV: *Feldmannia irregularis* virus a; Goxy: *Gluconobacter oxydans*; Gvio: *Gloeobacter violaceus*; IsknV: Infectious spleen and kidney necrosis virus; LDV1: Lymphocystis disease virus 1; Lmaj: *Leishmania major*; MCV: Molluscum contagiosum virus; MSV: *Melanoplus sanguinipes* entomopoxvirus; Npun: *Nostoc punctiforme*; PBCV: *Paramecium bursaria* Chlorella virus 1; SfAV1: *Spodoptera frugiperda* ascovirus 1a; SGV: Singapore grouper iridovirus; Tbru: *Trypanosoma brucei*; Tcru: *Trypanosoma cruzi*; VV: Vaccinia virus.

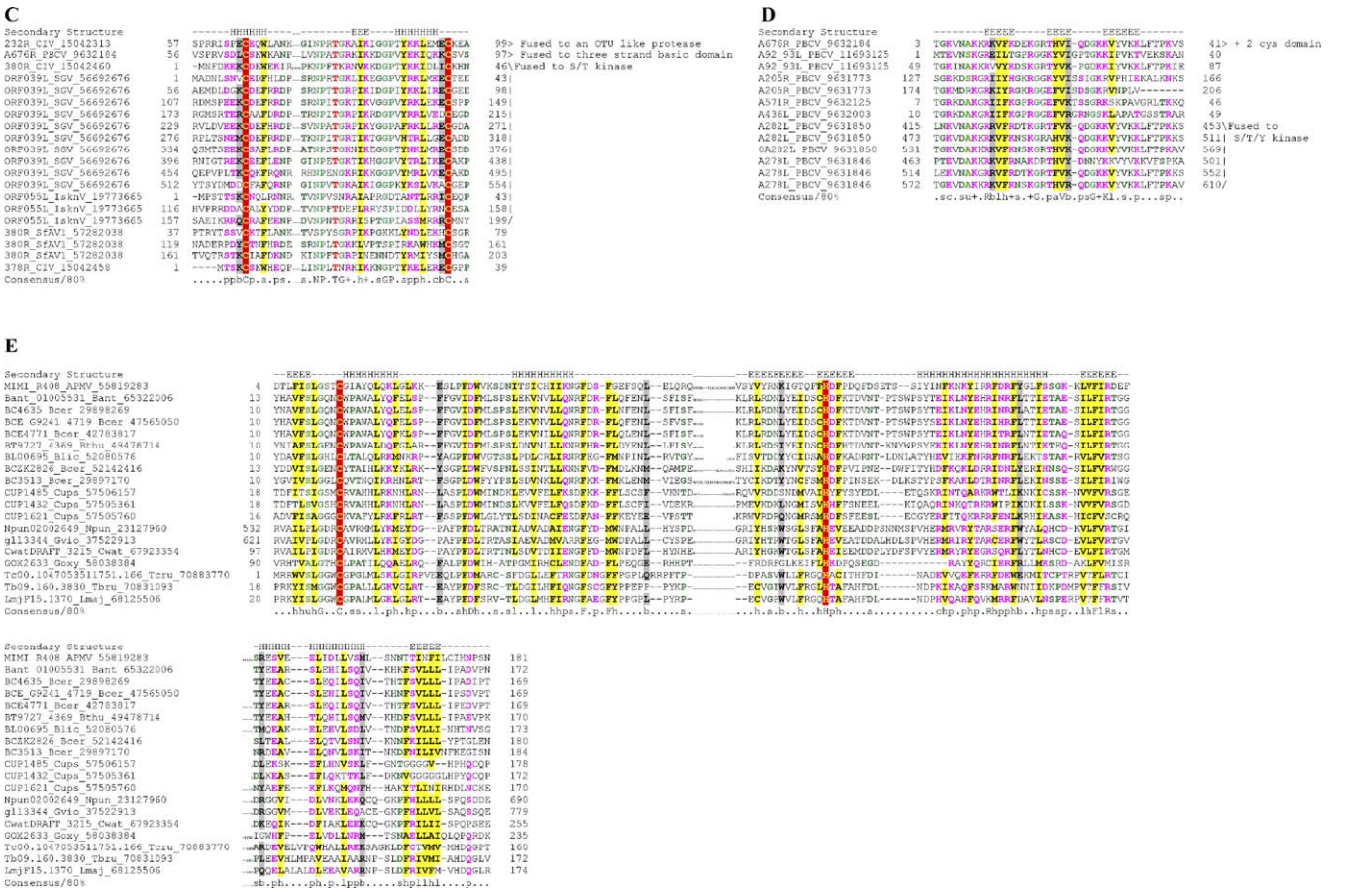

Fig. 4. (*Continued*).

seems to correlate with the nuclear phase of their life cycle. The sequence-specific DNA-binding domains KilA-N and Bro-N (Iyer et al., 2002) also most likely were encoded by the ancestral NCLDV and might function as specific transcription factors for certain viral genes. Three SFII helicases typified by the Vaccinia proteins I8, A18, and D6/D11 also are probable components of the transcription apparatus of the ancestral NCLDV. In particular, experimental studies suggest that A18 might be involved in the release of the transcript (Lackner and Condit, 2000). However, the contextual associations with primases (Iyer et al., 2005), suggest that the A18-like helicases might also play a role in replication by unwinding the RNA primer prior to its degradation.

Individual NCLDV lineages also appear to have acquired specific transcription factors from different sources through HGT. Most striking in this regard is the mimivirus, which appears to have acquired a homeodomain transcription factor (MIMI_R749) from the host and a cro-type repressor (MIMI_R559) from a bacterial or phage source. Transcription factors with the HMG domain and a SAZ (Stonewall, ADF1, Zeste)-type MYB domain, which is specifically expanded in insects, are present, respectively, in CIV (CIV-401R) and AMV (AMV266). A standalone SET domain protein, a histone methyltransferase, is conserved in the phycodnavirus-mimivirus clade. In several members of this clade, it also co-occurs with another protein containing the SWIB domain, which is found in several chromatin-remodeling proteins. Similar pairs of proteins with SWIB and SET domains are also found in several bacteria, such as *Chlamydia*, *Bordetella*, *Bdellovibrio*, *Rubrivivax*, *Magnetospirillum* and *Polaromonas*, suggesting that they form a functional association (LA and LMI unpublished observations). Furthermore, domain fusions in some of these bacteria and two-hybrid analysis of protein-protein interactions (Stephens et al., 1998) also suggest that the SWIB domain functions in cooperation with TopoI and related enzymes. Thus, the SET-SWIB pairs of NCLDVs might act in conjunction with the topoisomerases to modify histones and remodel chromatin associated with the host or the integrated virus genome as part of a strategy for global transcriptional regulation. Some iridoviruses also encode a protein with duplicated SWIB domains (e.g., CIV-306R) but we did not detect any SET domains in these genomes. The mimivirus also encodes a JOR domain demethylase (MIMI_L887) that is likely to catalyze the reverse reaction of histone demethylation (Aravind and Iyer, 2002). Other potential DNA-binding proteins, such as PBCV A437L homologous to the archaeal-type MC1 chromosomal proteins (Teyssier et al., 1994), ASFV A104R (a bacterial HU/IHF-like protein) (Swinger and Rice, 2004), and the LDV protein LDVICp062 containing a SAP domain, with a possible role in viral chromosome organization and packaging, are also encountered in some NCLDV lineages.

3.5. RNA metabolism and post-transcriptional regulation of gene expression

The core of the mRNA processing apparatus of the NCLDVs that can be traced back to their common ancestor consists of the capping and polyadenylation complexes. The capping enzyme complex contains three distinct catalytic domains, namely, a triphosphatase which removes the 5′ triphosphate, a guanylyl transferase which transfers the guanylate cap to the 5′-ends of viral mRNAs, and a methyltransferase that methylates the sugar of the cap (Shuman, 2001). Poxviruses, ASFV, and the mimivirus have a single polypeptide containing all three catalytic domains, whereas some phycodnaviruses, e.g., PBCV and EHV, encode only the triphosphatase and the guanylyltransferase as separate polypeptides, but apparently lack a cognate methyltransferase (Fig. 3A). The capping enzyme was observed only in certain iridoviruses (e.g., ISKNV and RBIV), and is closely related to the animal capping enzymes. Like the animal counterparts, the iridoviral capping enzyme contains a triphosphatase of the tyrosine phosphatase superfamily as opposed to the CYTH superfamily phosphoesterase (Iyer and Aravind, 2002) seen in other NCLDVs and all eukaryotic lineages other than plants and animals (Fig. 3A). The close relationship with the animal enzyme suggests a recent displacement of the ancestral NCLDV capping enzyme in the iridoviruses. A distinct polyA polymerase of the polymerase β class of nucleotidyltransferases (Aravind and Koonin, 1999a) is found in poxviruses, ASFV, and the mimivirus suggesting its presence in the ancestral NCLDV, followed by loss in the other NCLDV lineages, probably, associated with the use of the host mRNA processing apparatus as a consequence of the advent of the nuclear phase of the virus life cycle.

The cytoplasmic NCLDVs lack spliceosomal introns, and even viruses with a nuclear phase of reproduction like PBCV have very few small introns (Van Etten, 2003), suggesting that they were recently acquired. Accordingly, the NCLDVs have no enzymatic machinery for pre-mRNA processing. However, different lineages of NCLDVs encode a few enzymes that might regulate mRNA stability. These enzymes include nucleases, such as RNaseIII, found in PBCV (Zhang et al., 2003), the mimivirus (MIMI_R343), and some iridoviruses, the Xrn1p-like 5′–3′ exonuclease seen in CIV (CIV-012L) and the mimivirus (MIMI_R528), and the oligoribonuclease with duplicate RNaseIII domains in ESV (ORF139). Another RNA metabolism enzyme found in ASFV (EP424R) and the mimivirus is the FtsJ-like RNA methyltransferase (MIMI_L511, MIMI_R383) that potentially catalyzes the 2′-*O*-ribose methylation in rRNA and might play a role in stabilizing rRNA to facilitate viral translation (Bonnerot et al., 2003; Pintard et al., 2002). The NCLDVs, like the large dsDNA bacteriophages, also encode some proteins, which appear to play a role in RNA repair (Yin et al., 2003). These include the RNA ligase in AMV, the polynucleotide kinase/phosphatase in the mimivirus and the 2H phosphoesterase in FPV. We also detected a homolog of the phage T4 RNA ligase, which is highly conserved in all the iridoviruses (e.g., CIV-383L). However, this version does not have the key catalytic residues such as the conserved lysine required for nucleotidyl transfer, though they have their own set of unique conserved residues. This suggests that they might not be active RNA ligases, but either merely bind RNA or catalyze some other reaction. The presence of these multiple RNA repair proteins both in NCLDVs and in baculoviruses (Martins and Shuman, 2004), suggests that eukaryotic cells might counter viral infection through cleavage or modification of specific RNAs. Different lineages of NCLDVs also sporadically encode various RNA-binding proteins, such as the S1-domain-containing protein C3L in variola virus, the KH-domain protein MSV059 in MSV, and the CIV-132L and CIV-340R proteins, which contain, respectively, the CCCH and dsRBD domains in CIV. It is not currently clear if all these proteins act directly on viral transcripts or some might bind host transcripts.

3.6. Virion morphogenesis and DNA Packaging

The virion of the common ancestor of the NCLDVs appears to have possessed at least three distinct structural proteins that contributed to the protein capsid and lipid membrane. The capsid protein (cognate of Vaccinia D13L) is present in all NCLDVs and is predicted to contain β-jellyroll domains similar to those in the capsid proteins of many other DNA and RNA viruses (Iyer et al., 2001). The crystal structure of the PBCV ortholog (A622L) confirmed this prediction and showed that the capsid protein contained a tandem duplication of two jellyroll domains (Nandhagopal et al., 2002). In poxviruses, the C-terminal jellyroll domain appears to have been replaced by a distinct all β-strand domain (Iyer et al., 2001) and is probably correlated with the emergence of the atypical brick-shaped morphology of poxviruses. Consistent with this, the poxviral D13L has been shown to be a major component of the immature spherical capsid, and is subsequently shed with the formation of the mature virion (Szajner et al., 2005). The two conserved structural membrane proteins include the myristoylated membrane protein (Vaccinia L1R) and another membrane protein, typically, containing multiple disulfide bonds (paralogous Vaccinia proteins J5L, A16L, G9R (Senkevich et al., 1997)). These membrane proteins are found in all animal NCLDVs and the mimivirus, supporting their presence in the ancestral NCLDV (Iyer et al., 2001). However, both these membrane proteins were apparently lost in some or all phycodnaviruses (Iyer et al., 2001). These observations suggest that the ancestral NCLDV had a large icosahedral capsid, probably, with an inner lipid membrane, which played a role in virion assembly. The formation of disulfide bonds between the conserved cysteines in the structural proteins appears to be critical for the assembly and/or stability of the NCLDV virions. These reactions are mediated by a virus-specific disulfide-redox pathway that consists of two distinct thiol oxidoreductases, one of the thioredoxin fold (vaccinia G4L) and the other one related to the eukaryotic ERV1/2 flavin-dependent oxidoreductases (vaccinia E10R) (Senkevich et al., 2002). Both these redox enzymes can be traced back to the ancestral NCLDV and are conserved in all NCLDVs (Iyer et al., 2001), except for the loss of the G4L ortholog in ASFV. In poxviruses, the pathway includes a third redox protein (Vaccinia

A2.5L) that is detectable only in that lineage (Senkevich et al., 2002).

The P4B (Vaccinia A3L) is a virion core protein that has been shown to be processed by the SMT4-like thiol peptidase (Vaccinia I7L) in poxviruses (Ansarah-Sobrinho and Moss, 2004a; Byrd et al., 2002). Orthologs of P4B were detected in the mimivirus, whereas orthologs of the SMT4-like peptidase are conserved in poxviruses, ASFV, and the mimivirus. The ASFV peptidase also has been implicated in proteolytic processing during virion assembly (Alejo et al., 2003), implying that proteolytic processing of virion proteins by the SMT4-like peptidases might be an ancestral feature of the NCLDV. The vaccinia G1L and its apparent orthologs in the mimivirus (MIMI_L233) are peptidases of the mitochondrial protein maturation metallopeptidase family. While the poxviral enzyme has been implicated in virion maturation (Ansarah-Sobrinho and Moss, 2004b; Hedengren-Olcott et al., 2004), the exact step of its action remains unclear. As P4B and G1L orthologs are found in only two NCLDV lineages, it is not immediately obvious whether each of the respective genes descends from an ancestral virus gene or the respective genes spread via inter-viral HGT. A protein kinase (Vaccinia F10L) and, probably, a protein S/T/Y phosphatase (H1L) play a regulatory role in virion assembly (Betakova et al., 1999; Mossman et al., 1995). The F10L-like kinase has orthologs in all NCLDVs and the H1L-like phosphatase is missing only in ASFV and ESV, suggesting that both of them were present in the ancestral NCLDV. Almost all NCLDVs also encode one or several proteins with ankyrin repeats whose functions remain uncharacterized. Given that the ankyrins mediate formation of protein complexes via interactions with the repeats (Mosavi et al., 2004), we propose that they might play a structural role in the formation of the cytoplasmic virosome where the virions are assembled.

In the original analysis of the NCLDV, we showed that the ATPase required for DNA packaging (Vaccinia A32R (Cassetti et al., 1998; Koonin et al., 1993)) is conserved in all the NCLDVs (Iyer et al., 2001). Subsequent analysis of the A32R family showed that it belonged to the large FtsK-HerA superfamily of ATP-dependent DNA pumps that are required for pumping DNA during prokaryotic chromosome segregation and plasmid and transposon transport during conjugation (Iyer et al., 2005). It also has been shown that the DNA packaging enzymes of dsDNA phages, which do not depend on the terminase-portal protein pathway, and ssDNA phages, such as M13 and F1, belong to the same superfamily of ATPases (Iyer et al., 2005). The A32R family of the NCLDV is most closely related to the ATPases encoded by a distinctive class of transposons known as the Tlr transposons and then to the ATPases of dsDNA phages. Thus, DNA packaging in the ancestral NCLDV occurred via an ancient mechanism that had been widely used for chromosome segregation by DNA phages as well as cellular life forms (Iyer et al., 2004c). By analogy with cellular DNA segregation (Aussel et al., 2002), we speculate that the A32R family ATPases function along with the RuvC-like HJR or its equivalent to separate sister chromosomes and package them into virions. The identification of this constellation of 8–10 distinctive conserved proteins involved in virion maturation and DNA packaging suggests that the relatively complex, multi-step maturation process was already established in the ancestral NCLDV (Fig. 2).

Generally, the distribution patterns of NCLDV genes shared by two or more viral lineages show a striking tapestry of a few universally conserved genes and numerous genes with patchy distributions (Fig. 3B). On many occasions, unrelated or distantly related genes with the same function have complementary patterns, indicative of non-orthologous gene displacement (Fig. 3A). In other cases, the scattered gene distribution may be explained by HGT and lineage-specific gene loss (Fig. 3B).

4. Gene accretion in NCLDVs: role of lineage-specific expansion of gene families

The reconstructed common ancestor of the NCLDVs had at least 41 genes. While at this size the ancestral NCLDV had a larger genome than most of the other groups of DNA viruses, the extant NCLDVs have hundreds of genes, implying a massive growth in genome size since their last common ancestor. Of course, the possibility exists that the ancestral NCLDV had many more genes (perhaps, as many as some modern representatives of the clade), the majority of which have been lost in most surviving lineages. However, since two or more independent gene losses would be required for an ancestral gene to disappear from the horizon, it seems reasonable to assume that the number of such completely lost genes is relatively small. Consequently, it appears likely that evolution of the NCLDVs involved actual increase in genome size rather than just gene turnover. Unlike small viruses that optimize coding in every frame, the NCLDVs may show considerable gene redundancy as indicated by the existence of many lineage-specific multigene families. For example, MSV has a family of approximately 20 genes encoding a specific version of the leucine rich repeat (LRR) that is absent in the chordopox lineage (Table 1). This and other anecdotal evidence furnished by the NCLDV genome analysis indicates that lineage-specific expansion (LSE) of paralogous genes could have substantially contributed to the genome growth. In particular, the giant genome of the mimivirus, with its numerous multigene families, provided us with an opportunity to investigate this process of LSE systematically (Table 1).

Using BLAST-score-based clustering of the complete proteomes of NCLDVs, we identified 267 lineage-specific multigene families ranging in size from 2 to 115 members. A plot of the number of lineage specific families of each size (Fig. 5) showed that their distribution was best approximated ($R^2 = 0.97$) by the power law of the form $y = ax^b$ (where a and b are constants, $-3 < b \leq 2$). Among individual viruses, the best fit was observed for the largest available genome, that of the mimivirus, but wherever a sufficient number of data points were available, a good fit with the power law was observed as illustrated for PBCV (Fig. 5). Predictably, the mimivirus had a longer tail of very large families (Fig. 5). This power-law distribution closely recapitulated the size distribution of LSEs reported for eukaryotic genomes (Lespinet et al., 2002). This implies that the NCLDV genomes grow via lineage-specific gene family expansion similar to those operating in the evolution of eukaryotic nuclear

Table 1
Examples of lineage-specific expansions of globular domains in NCLDV viruses

Virus	Protein family (number of copies)	Comments
Mammalian pox viruses	Kelch repeats fused to POZ domains (4–6)	Lost in MCV
	Ankyrins (4–6)	Lost in MCV
MCV	IL-18-binding protein (3)	Inhibitor of interleukin 18 activity
FPV	Ankyrin repeats (30)	–
	KilA-N domain (10)	KilA-N is a DNA-binding domain found in the NCLDVs and several phages
	C-type lectin (6)	A cell surface adhesion molecule
	G-protein coupled receptor (3)	Potential chemokine receptors
MSV	Cytoplasmic LRR repeats (21)	LRR repeats related to the bacterial internalin-like actin-binding proteins
	KilA-N fused to a VSR nuclease (4)	Predicted DNA repair/recombination proteins
	Tryptophan repeats (4)	The tryptophan repeat domain is a small 23 aa α-helical motif, which is characterized by the presence of a conserved tryptophan residue. From 4 to 12 tandem copies may be present in a protein. The domain is present in the entomopoxviruses, baculoviruses and in the Cytophaga protein *Chut02002469*. It is also expanded in MSV
AMV	Cytoplasmic LRR repeats (5)	LRR repeats related to the bacterial internalin-like actin-binding proteins
	KilA-N fused to a BroC domain (4)	–
	BroN domain fused to a BroC domain (3)	–
	MSV199 domain fused to a T5 orf172 domain (3)	Unique domain fusion found only in entomopoxviruses and insect iridoviruses
ASFV	MGF 360/530 (15)	A divergent version of the ankyrin repeat module; are host range determinants and are required for macrophage infection
	L270L-like proteins (5)	A secreted protein with a β-strand rich domain with 8 conserved cysteines
CIV	MSV199 domain fused to a BroC domain (7)	Unique domain found only in entomopoxviruses and insect iridoviruses
LDV	S/T protein kinase (6)	A distinct family of viral serine-threonine kinases
SGV	Two extracellular Ig domains with a C-terminal transmembrane domain (8)	–
PBCV	Ankyrin repeats (10)	–
	Intron encoded endonuclease (4)	Contains a URI domain endonuclease; probably a mobile element
	Major Capsid protein (4)	
	Three stranded positively charged domain (7)	A small domain (∼35 aa) predicted to adopt an all β fold with three strands (Fig. 4D). Characterized by the presence of 4 conserved basic residues. Proteins may contain from 1 to 3 tandem repeats of the domain. PBCV A282L and A278L are fused to the C-terminus of a S/T kinase domain. A676L is fused to the two cysteine domain
ESV	Replication Factor C (5)	AAA+ ATPase required for loading of the DNA clamp PCNA
Mimivirus	Ankyrins (116)	–
	Leucine rich repeats (15)	A special class of intracellular LRRs with N-terminal F-box domains. Related LRRs are greatly expanded (>200 copies) in *Dictyostelium discoideum*
	Phage T5 yomD-like proteins (6)	A cysteine rich metal-chelating domain with 4 conserved cysteines that is found in several bacteriophages, eukaryotes and mimivirus
	Transposase (4)	This version of the transposase is also present in PBCV and is probably of bacterial origin
	POZ domains (36)	Of these at least 12 proteins are fused to WD40 repeats
	TPR repeats (5)	–
	KilA-N domain fused to BroC (5)	–
	Specialized S/T/Y kinases (3)	Multidomain protein where a type II Periplasmic-binding domain and an EGF repeat are at the N-terminus and are extracellular. C-terminal to these is a transmembrane helix, a cytoplasmic S/T/Y kinase domain and an adenylyl cyclase domain

genomes. Roughly, the genome growth can be approximated by a stochastic birth-and-death (gene duplication and loss) process superimposed on which is the adaptive proliferation of certain families that widely diversify into particular functional niches (Koonin et al., 2002); (Karev et al., 2003). Duplications within these large families may be fixed more often than duplications in other families because the new members add to the functional diversity and have adaptive value which results in self-accelerating growth of the family.

Some of the LSEs in the phycodnavirus-mimivirus clade appear to represent transposons with IS-element-type or HNH-type integrases (Table 1). However, the remaining LSEs of the NCLDVs, like their eukaryotic cellular counterparts, might have specialized adaptive functions, especially, those related to inter-

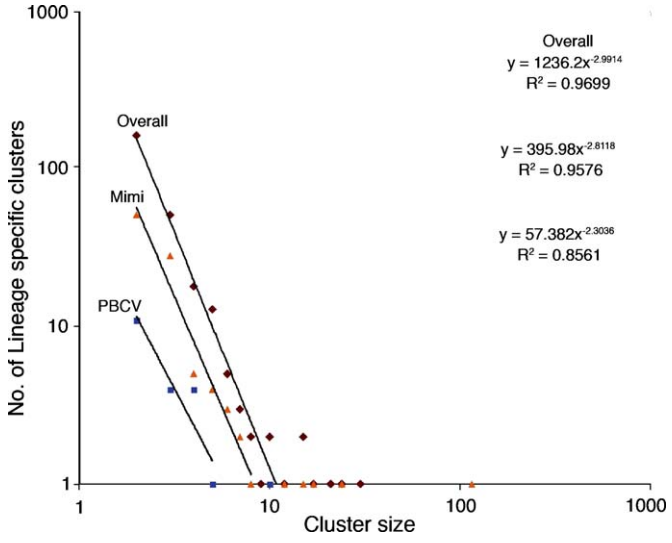

Fig. 5. Size distributions of lineage-specific expansions of gene families in NCLDVs. The graph is plotted in double logarithmic coordinates. The equation of the power law that best fits the linear part of the equation is shown along with the R^2 value for all NCLDVs taken together, the mimivirus, and PBCV. Points corresponding to PBCV and Mimivirus are colored blue (squares) and orange (triangles), respectively. Points corresponding to all the NCLDVs taken together are shown as brown diamonds.

actions with the host. A number of LSEs in different NCLDVs contain proteins with super-structure forming repeats (Table 1). Examples of these include the LRRs seen in entomopoxviruses and mimivirus, ankyrins in FPV and mimivirus, and β-propellers of the WD40 and Kelch types fused to N-terminal POZ domains in the mimivirus and orthopoxviruses. The LRRs in the entomopoxviruses appear to be similar to the cognate repeats found in the internalins of *Listeria*, which have been shown to bind actin filaments (Cossart et al., 2003). Likewise, the kelch proteins in animals are known to interact with actin (Robinson and Cooley, 1997), suggesting that virus-encoded kelch proteins might interact with the host cytoskeleton and probably form structural components of the cytoplasmic virus assembly centers. The domain architectures and evolutionary affinities of the LRR expansion in mimivirus suggest that it might play a role in interaction with the host (see below). We observed that the previously characterized ASFV multigene families 360 and 530, which are host-range determinants, required for proliferation in macrophages (Afonso et al., 2004), contain a divergent version of the ankyrin repeat. The available evidence suggests that these proteins might also form specific intracellular structures that protect the viral assembly centers from the host interferon-based defense response.

A number of smaller LSEs (3–10 members), particularly, in animal NCLDVs, appear to have been specifically selected for interactions on the host cell surface. Examples of these expansions include proteins with multiple Immunoglobulin (Ig) domains in the grouper iridoviruses, C-type lectins and G-protein-coupled 7TM receptors in FPV, the ASFV U104L-like secreted cysteine-rich proteins, and the interleukin-18-binding proteins in MCV (Xiang and Moss, 1999) (Table 1). These proteins are known or predicted to be expressed on the host cell surface and might provide mechanisms to modulate the host immune response by mimicking homologous adhesion molecules or cytokine receptors of the host immune system.

Other LSEs consist of proteins with no homologs outside of a restricted set of NCLDVs. In some of these cases, it is possible to glean the functions of these expansions using contextual inference from domain architecture (Table 1). For example, a small domain with two conserved cysteines is found in all iridoviruses and PBCV. This 2-cysteine domain forms LSEs in both SGV and LDV, and is often found in multiple repeats in the same polypeptide, or is fused to the N-termini of OTU/A20-like peptidases or protein kinases in iridoviruses (Fig. 4C). In PBCV, multiple protein kinases are fused to another small, highly positively-charged C-terminal domain, with 3 β-strands (Fig. 4D). In one PBCV protein, A676R, this module is fused to the 2-cysteine module suggesting that they might function in a common context. The domain associations of these proteins indicate that they might function as viral adaptors connecting the kinases and OTU/A20 peptidases to specific targets. The functions of other LSEs, such as the expansion of orthologs of Variola B22R in FPV, and 32 unique families in mimivirus remain mysterious. Given the predominantly α-helical, cysteine-supported or low-entropy structure of these proteins, we suspect that they might have emerged de novo, similarly to the lineage-specific α-helical and low-entropy structures observed in various eukaryotes (Lespinet et al., 2002).

5. Gene accretion in NCLDVs: role of HGT

Acquisition of genes from the host has been documented in all kinds of parasites, both cellular and viral (Hughes and Friedman, 2005; Koonin et al., 2001; Senkevich et al., 1997). Hence, it is not surprising that every NCLDV lineage has acquired a certain set of genes from the host at different points in evolution. These genes are incorporated into various functional systems of the viruses and, depending on the function for which they are recruited and how far it departs from the original cellular function, they show different degrees of modification in terms of domain architectures and sequence divergence. These trends in acquisition of host genes have been documented in some detail in previous analyses of poxvirus genomes (Hughes and Friedman, 2005; Senkevich et al., 1997). However, in addition to the genes apparently derived from the host, there are many genes in NCLDVs that appear to have specific phylogenetic affinities to homologs from bacteria or other viruses. The simplest explanation for these observations is that the NCLDVs have been swapping genes through HGT with endosymbiotic and pathogenic bacteria, and viruses that co-infect their hosts. In this section, we examine cases of such apparent HGT, with an emphasis on the role of transfers from bacteria.

Gene transfers from bacterial sources appear to be more common in the genomes of the Phycodnavirus-Mimivirus clade as compared to the animal NCLDVs. This may not be surprising in view of the greater abundance of phagocytosed and endosymbiotic bacteria in the protist hosts of these viruses compared to multicellular eukaryotes. Nevertheless, a number of horizontal transfers from bacteria seem to have substantially affected the

evolution of the animal NCLDVs as well. One example of this is the apparent transfer, to the poxviral lineage, of the G6R protein, which is a circularly permuted member of the NlpC/P60 superfamily of enzymes (Anantharaman and Aravind, 2003). This protein is predicted to function as an acyltransferase that might play a role in lipid metabolism during virion maturation (or less likely a peptidase). Phylogenetic analysis indicated that this gene was derived from a proteobacterial source (Anantharaman and Aravind, 2003). Similarly, ASFV also appears to have acquired a distinct array of bacterial genes, such as an EF-G-like GTPase (CP312R) and a NifS-like pyridoxal-phosphate-dependent enzyme (QP383R). Two distinctive prokaryotic chromatin proteins, MC1 and HU/IHF, are encoded, respectively, by PBCV and ASFV, suggesting that they have been acquired by these viral lineages from prokaryotic endosymbionts or parasites of their hosts. In a few cases, there is strong evidence of gene exchange between specific groups of endosymbiotic or endoparasitic bacteria and NCLDVs (Fig. 6). The SET and SWIB domain proteins of the phycodnavirus + mimivirus clade, a family of specialized LRR proteins (MIMI_R542-like) and a patatin-like α/β hydrolase in the mimivirus (MIMI_L620), and a membrane-associated prenyltransferase in ASFV (B318R) (Fig. 6C) all show specific phylogenetic affinity to homologs from *Chlamydia*. Likewise, *Chlamydia* are unique among the bacteria in having a SMT4-like peptidase similar to those of the NCLDVs (Stephens et al., 1998). While all the former examples appear to represent lineage-specific acquisitions of genes by NCLDVs from bacterial endosymbionts or parasites of the chlamydial lineage, the SMT4-like peptidase might be an acquisition by *Chlamydia* from an NCLDV.

The most dramatic instances of HGT are seen in the phycodnavirus-mimivirus clade. Using conservative clustering measures, we identified at least 75 proteins with bacterial affinity and 198 proteins with a eukaryotic affinity in the mimivirus. Likewise, PBCV has 40 proteins with bacterial affinity and 46 with eukaryotic affinity. The majority of the HGT cases detected in the phycodnavirus-mimivirus clade are unique to one of these lineages. A similar pattern, albeit involving far fewer proteins, is seen amongst the vertebrate poxviruses that appear to have acquired many host proteins with functions in immuno-regulation (see below). These observations suggest that the HGT events are, largely, lineage-specific and might have contributed to the growth of the NCLDV genomes. In the case of the mimivirus, conservatively, about 30 of the proteins with eukaryotic affinities are most closely related to homologs from *Dictyostelium discoideum*, an amoebozoan related to the host of mimivirus (for which no genome sequence is currently available). This suggests that numerous gene transfers from the host might have occurred after the mimivirus adapted to replicate in an amoebozoan (e.g., see Fig. 6). Generally, the horizontally transferred gene set appears to represent a volatile shell of the NCLDV genomes, which is in constant flux due to selective forces emerging from the virus–host co-evolution. In particular, in the mimivirus, the numerous horizontally acquired gene products have probably allowed the virus a higher degree of autonomy in its life cycle. This is strikingly illustrated by the presence of a variety of proteins with central functions in translation, the class of functions for which all other viruses rely almost entirely on the host cell. The translation system components encoded by the mimivirus include 4 aminoacyl tRNA synthetases, translation factors eIF1, eIF4A, eIF4E, EF1α, and eRF1, and tRNA- and rRNA-modifying methyltransferases (Raoult et al., 2004).

In PBCV, ∼18 genes mostly of apparent bacterial origin (some of them paralogous) encode various enzymes for polysaccharide metabolism. A similar situation is seen in the mimivirus, which encodes at least 33 proteins involved in various aspects of carbohydrate metabolism, most of them apparently of bacterial origin. Some of these PBCV enzymes have been shown to hydrolyze polysaccharides, probably, during virus entry and release (Markine-Goriaynoff et al., 2004; Sun et al., 2000). However, the roles of the others, which include proteins involved in polysaccharide biosynthesis and GPI anchor synthesis, e.g., hyaluronan synthase (PBCV A98) and a WcaK-like glycosyltransferase (Mimivirus MIMI_L143), suggest a more complex virus–host interaction (DeAngelis et al., 1997; Graves et al., 1999; Markine-Goriaynoff et al., 2004). The existence of such complex interactions is further suggested by the presence of 5 histidine kinases encoded by ESV and 3 novel signaling receptors encoded by the mimivirus, which contain an extracellular region with two type II periplasmic-binding protein domains and an EGF repeat, and an intracellular region with a protein kinase and an adenylyl cyclase domain (Table 1). These signaling proteins appear to have been uniquely acquired in these viral lineages from either their hosts or co-existing bacterial endosymbionts. This group of proteins could alter the responses of the infected host to various environmental inputs and metabolic perturbations caused by the infection. It should also be noted that protozoan viruses face a wide range of competitors including other viruses, symbiotic and parasitic bacteria and eukaryotic protist parasites. For example, there is some evidence that ESV might have to compete with a plasmodiophoran parasite *Maullinia ectocarpii* in its algal host *Ectocarpus* (Maier et al., 2000). Hence, some of the genes of apparent bacterial origin that are involved in complex metabolic and signaling functions might provide an advantage to the virus by making the host resistant to competing parasites, enabling it to survive particular environmental conditions or causing particular behaviors favoring viral survival or propagation. As an analogy such a phenomenon has recently been reported in cyanophages which alter the photosynthetic properties of their hosts (Mann et al., 2005).

The mimivirus has ∼10 laterally transferred genes encoding proteins that function in the eukaryotic mitochondrion, including molecular chaperones, peptidases, and mitochondrial membrane proteins. The virus might employ these proteins, perhaps in conjunction with additional host proteins, to form the virosome membrane in the host cytoplasm with structural features similar to mitochondrial membranes. Some horizontally transferred genes of the mimivirus define previously uncharacterized proteins families which might have interesting biochemical properties. One such protein, whose function we were able to predict using sequence-structure comparisons, is the protein MIMI_R408 that has homologs in several bacteria, such as *Bacillus anthracis* and *Nostoc*, and kinetoplastids. An alignment of this protein family revealed a conserved cysteine and histidine

Fig. 6. Phylogenetic trees showing evidence of horizontal gene transfer in NCLDVs. Phylogenetic trees are shown for the (A) iridovirus SOCS1 gene; (B) the collagenase like metalloprotease in AMV; (C) the prenyltransferase in ASFV; and (D) the AlkB superfamily protein in mimivirus. (A) and (D) represent instances of transfer of genes from the eukaryotic host; (B) represents an instance of transfer from baculoviruses that also infects insect hosts; and (C) represents a probable instance of lateral acquisition of a gene from a *Chlamydia*-like intracellular parasite. Phylogenetic trees were built using the least-square method with subsequent local rearrangement to obtain the maximum likelihood tree as previously described (Iyer et al., 2004a). The reliability of the tree topology was assessed using either of the following methods: (1) the RELL bootstrap method of MOLPHY, with 10,000 replications (Adachi and Hasegawa, 1992). (2) The construction of 1000 independent replicates of maximum-likelihood distance matrices with the PUZZLE program, followed by the determination of a consensus tree from these matrices. Nodes supported by bootstrap values >70% in these procedures marked with filled red circles. The blue colored branches highlight the clade where HGT has occurred. Proteins are denoted by their gene names, species abbreviations and genbank index (gi) numbers. Species abbreviations are as follows: Aae: *Aquifex aeolicus*, Aful: *Archaeoglobus fulgidus*, Afum: *Aspergillus fumigatus*, AhNPV: *Adoxophyes honmai* nucleopolyhedrovirus, AMV: *Amsacta moorei* entomopoxvirus,

pair, which occurred in a distinctive secondary structure context that is a conserved feature observed in papain-like peptidases (Fig. 4E). Thus, MIMI_R408 is likely to represent a previously uncharacterized family of papain-like peptidases that might be involved in mimivirus virion maturation.

Horizontal gene exchanges between viruses are harder to establish conclusively due to the rapid evolution of viral proteins; nevertheless, some clear-cut cases of such transfers became apparent from the comparative analysis of the NCLDV protein sequences. In particular, several genes of the entomopoxviruses, CIV, and baculoviruses are specifically related to one another, implying considerable gene swapping between unrelated insect viruses. Conceivably, once a gene providing increased fitness to a virus in a particular host or cell type emerges, there is a chance that it also will be useful for other viruses infecting the same host. Therefore, gene transfers between the different groups of insect viruses seem to have occurred more than once, spreading these survival-enhancing genes amongst them. This is supported by the several unique domain architectures that are shared between different virus groups, such as the Bro-N domain fused to the VSR nuclease present in MSV and CIV, and the Bro-N + BroC architecture shared by AMV and baculoviruses. In addition, AMV has a considerable complement of baculovirus-like genes, such as the p35-like caspase inhibitor (AMV010, AMV021), ubiquitin (AMV167), the matrix metalloprotease (AMV070), and the AMVITR10-like protein that shares a domain with the Viral Enhancing Factor of the granuloviruses. Similarly, CIV shares genes with both baculoviruses (CIV-162R, CIV030L and CIV-422L) and entomopoxviruses, which implies HGT between insect iridoviruses, entomopoxviruses, and baculoviruses. For many genes (e.g., the BROA-like proteins with a Bro-N + BroC domain fusion, the p35 like caspase inhibitor, the RNA ligase AMV019, the AMVITR10/VEF-like protein in AMV, and the superfamily-2 helicases CIV030L and CIV422L in CIV), the direction of transfer is clearly from the baculoviruses, as these genes are widely distributed in the baculoviruses, but are found in only a single or a few NCLDVs.

The FPV genome contains three genes that are possible horizontal transfers from unrelated viruses, two of which are particularly unusual. FPV063 is an ortholog of the NS-S protein of the bunyaviruses, a negative strand RNA virus, which may have been acquired by FPV via reverse transcription of the NS-S gene. FPV217, an uncharacterized protein, seems to have been acquired from a baculovirus-like source. No baculoviruses that infect chordates have been isolated to date, and the source for FPV217 is a mystery; in principle, such a transfer might have occurred through an intermediate virus that shares insect and vertebrate hosts.

6. Viral adaptations against host defenses

The majority of the NCLDVs show specific adaptations directed against the defenses of the host. These are most clearly understood in the case of the mammalian viruses, where the interactions of the viruses with the immune system of the host have been intensely studied (Bowie et al., 2004; Bugert and Darai, 2000; Moss et al., 2000; Seet et al., 2003). While the predicted proteins of the phycodnaviruses and mimiviruses strongly suggest virus-induced modification of host metabolism and behavior, next to nothing is known regarding the infection response in protists.

The presence of several enzymes of the ubiquitin (Ub) pathway in all the NCLDVs suggests that interference with Ub signaling may be a general mechanism to counter host defenses, by either targeting key proteins for degradation or modulating their activity (Table 2). This proposal is supported by the studies on the RING finger E3 Ub-ligase in poxviruses (N1R/p28), which is related to the cellular Makorin-like proteins, and is required for virulence (Huang et al., 2004a; Nerenberg et al., 2005; Senkevich et al., 1994). A wide diversity of RING finger E3 Ub-ligases are also seen across the NCLDVs; these proteins display a broad range of domain architectures, indicating multiple independent acquisitions of RING finger domains in different NCLDV lineages (Table 2). The mimivirus shows a lineage-specific expansion of 14 LRR proteins with N-terminal F-box domains, which are known to recruit E3 ligases to substrates associated with the LRRs. These mimiviral LRRs are specifically related to a class of LRRs, which are vastly expanded

Aper: *Aeropyrum pernix*, APMV: *Acanthamoeba polyphaga* mimivirus, ASFV: African swine fever virus, Asp.: *Acinetobacter* sp., Atha: *Arabidopsis thaliana*, Bant: *Bacillus anthracis*, Baph: *Buchnera aphidicola*, Bcep: *Burkholderia cepacia*, Bcer: *Bacillus cereus*, Bhal: *Bacillus halodurans*, Bmal: *Burkholderia mallei*, Bsub: *Bacillus subtilis*, Cace: *Clostridium acetobutylicum*, CcNPV: *Chrysodeixis chalcites* nucleopolyhedrovirus, Ccre: *Caulobacter crescentus*, Cele: *Caenorhabditis elegans*, Chis: *Clostridium histolyticum*, Cjej: *Campylobacter jejuni*, Csor: *Clostridium sordellii*, Ctet: *Clostridium tetani*, Ctra: *Chlamydia trachomatis*, Cvio: *Chromobacterium violaceum*, Ddis: *Dictyostelium discoideum*, Dmel: *Drosophila melanogaster*, Drer: *Danio rerio*, Erum: *Ehrlichia ruminantium*, Fsp.: *Frankia* sp., Ggal: *Gallus gallus*, Hasp: *Halobacterium* sp., Hpyl: *Helicobacter pylori*, Hsap: *Homo sapiens*, Hzea: *Helicoverpa zea* single nucleocapsid nucleopolyhedrovirus, IsknV: Infectious spleen and kidney necrosis virus, LdNPV: *Lymantria dispar* nucleopolyhedrovirus, Lint: *Leptospira interrogans*, McNPV: *Mamestra configurata* nucleopolyhedrovirus B, Mdeg: *Microbulbifer degradans*, Mjan: *Methanocaldococcus jannaschii*, Mmus: *Mus musculus*, Mthe: *Methanothermobacter thermautotrophicus*, Mtub: *Mycobacterium tuberculosis*, Ncra: *Neurospora crassa*, Nmen: *Neisseria meningitidis*, Npun: *Nostoc punctiforme*, OgIV: Orange-spotted grouper iridovirus, Paby: *Pyrococcus abyssi*, Paer: *Pseudomonas aeruginosa*, Pmar: *Prochlorococcus marinus*, Psp.: *Parachlamydia* sp., Rpro: *Rickettsia prowazekii*, Scer: *Saccharomyces cerevisiae*, Scoe: *Streptomyces coelicolor*, SlNPV: *Spodoptera litura* nucleopolyhedrovirus, Ssol: *Sulfolobus solfataricus*, Ssp: *Synechocystis* sp., Syn: *Synechococcus* sp., Taci: *Thermoplasma acidophilum*, Telo: *Thermosynechococcus elongatus*, Tmar: *Thermotoga maritima*, Tnig: *Tetraodon nigroviridis*, Vpar: *Vibrio parahaemolyticus*, Vvul: *Vibrio vulnificus*, Xaxo: *Xanthomonas axonopodis*, Xfas: *Xylella fastidiosa*, Xlae: Xenopus *laevis*, Xory: *Xanthomonas oryzae*.

Table 2
Domain architectures and ubiquitin pathway genes in NCLDVs

Domain architectures	Viruses (Genes)	Comments
Ubiquitin		
Ubiquitin	ESV (ORF 153), entomopoxviruses (MSV144, AMV167), SGV (ORF102L)	Several viruses are known to encode their own ubiquitins, which might be used to modify host proteins just as the endogenous ubiquitin
Ubiquitin conjugating enzymes and accessory subunits		
Ubiquitin-conjugating enzyme E2 domain	Mimivirus (L460, L709, L630, R521) ASFV (I215L), ISKNV (ORF099L, orf120R), RBV (ORF091L)	Transfers ubiquitin from the E1 enzyme onto a conserved cysteine in E2. The ubiquitin is subsequently transferred from the E2 enzyme to the protein substrate with the help of an E3 subunit
Cullin domain	Mimivirus (R337, L684)	Alpha alpha superhelical domain that is fused to a C-terminal winged HTH. This protein bridges the Skp1 protein and a RING finger in the SCF (Skp1, Cullin, F-box) E3 complex to the E2 UB ligase
Skp1	PBCV (A39L)	POZ domain protein; part of SCF E3 complex
E3 like ubiquitin ligases		
Solo Ring finger	Mimivirus (R311, R795), ESV (Orf 19, Orf 20, Orf 208), CIV (121R, 157L, 413R, 095L), PBCV (A481L)	Zinc chelating E3 ubiquitin ligase
Parkin like ring finger	MSV (MSV224), LDV (LCDV1gp056)	Ring finger similar to the one present in the Ariadne protein. This is present in all vertebrate iridoviruses. The MSV224 protein is additionally fused to a SWI2/SNF2 like helicase
Two Ring fingers fused to the C-terminus of vWA domain	Mimivirus (R811)	The vWA is a Mg^{2+} chelating domain with a Rossmann fold.
Ring finger fused to a divergent P-loop ATPase	ESV (ORF172)	The ATPase domain belongs to the ASCE division but is of unclear affinities
Ring finger fused C-terminal to BIR repeats	Entomopoxviruses (AMV021, MSV242), CIV (CIV-193R)	Inhibitor of apoptosis in Baculoviruses, the AMV protein has a duplication of the BIR domain. Possible caspase inhibitor
Ring finger fused C-terminal to Tryptophan repeats	MSV (MSV197, MSV027, MSV205)	The Tryptophan repeats probably function as protein-protein interaction motifs that recruit the RING finger to specific targets
RING finger fused N-terminal to ankyrin repeats	ESV (ORF142)	
Makorin like Ring finger fused C-terminal to a KilA-N domain	FPV (FPV150, FPV157)	The MCV ortholog has lost its KilA-N domain. Appear to be critical for virulence and poly-ubiquitination of proteins in the virus factories
RING finger fused to a SAP domain	CIV (175R, 332L)	The SAP domain is a DNA-binding domain, which tethers other domains to chromosomal scaffolds
RING finger fused to a Superfamily II helicase	AMV (AMV039)	–
RING finger fused to a TRAF domain	ISKNV	An E3 ligase that functions downstream of the TNF receptor in activating NFkB
F-box fused to LRRs	Mimivirus (L60, L414, L167, L165, R286, L166, L415, L162, L170, L168, R637, R638, R636, L281)	N-terminal F-box domain fused to C-terminal LRR repeats related to versions seen in *Dictyostelium*
Deubiquitinating enzymes		
USP superfamily protease	Mimivirus (R319) PBCV (A105L)	Deubiquitinating enzyme of the papain-like thiol protease fold
ULP1/Smt3 like protease	Feldmannia Virus (FirrV-1-E3)	Deubiquitinating enzyme of the papain like thiol protease fold. This protein contains a Sen2p-like ubiquitin fused to the N-terminus
OTU/A20-like cysteine protease fused to the two cysteine domain	CIV (232R)	The OTU/A20-like deubiquitinating enzymes belong to the papain-like thiol protease fold. The OTU domain of the Tipula iridescent iridovirus (L96) is fused C-terminal to a SAP domain

(∼200 paralogs) in the amoebozoan *Dictyostelium* that is related to *Acanthamoeba*. This expanded family of LRRs in *Dictyostelium* is often fused to serine/threonine kinases and B-Box zinc-binding domains. Analogous to a similar lineage-specific expansion of LRR-containing proteins involved in disease resistance in plants (Ellis et al., 2000), these amoebozoan proteins might play a role in recognizing intracellular parasites and initiating a defense response against them. In this light, it is possible that the mimiviral LRRs might counter this defensive mechanism of its host by targeting components for ubiquitination. Additionally, ASFV and the mimivirus also have one to four E2 Ub-ligases, PBCV has a Skp2-like E3 ligase subunit, and entomopoxviruses, vertebrate iridoviruses and ESV encode Ub-like proteins. Some of the NCLDVs also encode a range of de-ubiquitinating enzymes (DUBs), such as the SMT3-like de-SUMOylating enzyme in the *Feldmannia irregularis* Virus, USP-like DUBs in mimivirus and PBCV, and OTU/A20-like DUBs in iridoviruses (Table 2). These findings imply that ubiquitination of host proteins either by Ub or by virally encoded Ub-like proteins, as well as removal of Ub moieties from specific host proteins by viral DUBs is a strategy extensively used by the NCLDVs in their interactions with the hosts.

Most of the adaptations against host defenses in animal NCLDVs can be classified under two general themes that are not necessarily mutually exclusive (Table 3): (1) inhibition of apoptosis; and (2) modulation of immune response. Apoptosis is, probably, the simplest and most ancient defensive response of multicellular eukaryotes against intracellular parasites, and is a highly effective and generalized means of blocking viral propagation (Ameisen, 2002; Ameisen et al., 2003; Koonin and Aravind, 2002). Not surprisingly, large DNA viruses appear to have evolved numerous distinct adaptations, typically in the form of dominant negative regulators or inhibitors of apoptotic signaling (Barry et al., 2004; Clem, 2001). A common adaptation that is shared by all animal NCLDVs is the acquisition of a cognate of the cellular Bir domain protein, IAP, which is a caspase inhibitor. Given the current phylogenetic model (Fig. 2), it appears likely that an IAP-like protein was acquired at least twice independently by the animal NCLDVs (once in the poxvirus-ASFV lineage and once by iridoviruses). A very similar scenario seems to apply to the acquisition of the anti-apoptotic Bcl2 proteins by the NCLDVs. Homologs of Bcl2 have been detected in FPV (FPV039), ASFV (A179L), and the vertebrate iridoviruses LDV (LDVICp070) and SGV. We also identified a second Bcl2 homolog in LDV (LDVICp188). This suggests that the Bcl2-like proteins were acquired on three or four independent occasions during the evolution of the NCLDVs, namely, in FPV, ASFV, and probably more than once in iridoviruses. These multiple acquisitions are also consistent with the independent acquisition of Bcl2 by vertebrate herpesviruses.

Several other anti-apoptotic adaptations of large mammalian DNA viruses, such as the Death Effector Domain (DED) proteins, have been extensively discussed in the literature (Bertin et al., 1997; Senkevich et al., 1997; Shisler and Moss, 2001). Hence, we briefly survey here the under-appreciated anti-apoptotic and counter-immunity adaptations that we uncovered by genome analysis of entomopoxviruses and iridoviruses (Table 3). The entomopoxvirus AMV encodes an inactive caspase (AMV063) which probably acts as a dominant negative inhibitor of apoptosis in the insect, analogous to the DED, PYRIN and CARD domain inhibitors of caspases seen in chordopoxviruses and herpesviruses. Another entomopoxvirus, MSV, encodes a homolog of the pellino protein (MSV244) which functions in the IL-1/Toll signaling pathway downstream of the IRAK kinases and is required for innate immunity in insects and vertebrates (Jiang et al., 2003; Yu et al., 2002). We found that the pellino proteins contain an N-terminal domain, which adopts the same fold as the phosphopeptide-binding FHA domains (Hofmann and Bucher, 1995), fused to a C-terminal RING finger domain. Thus, pellino might bind phosphorylated proteins downstream of IRAK and regulate them through ubiquitination. In MSV244, however, the RING finger is disrupted, so this protein might inhibit ubiquitin-mediated regulation of components of the IL-1/Toll pathway by taking the place of the endogenous pellino.

Vertebrate iridoviruses display a striking array of adaptations against apoptosis and host immunity, mainly, in the form of proteins apparently acquired from the host. Specific anti-apoptotic proteins include a CARD-domain protein, a potential caspase inhibitor, seen in most vertebrate iridoviruses and a soluble version of the TNF receptor, analogous to the poxviral G4R protein (Hu et al., 1994), present in all fish iridoviruses. Vertebrate iridoviruses also encode a second protein containing TNFR-type cysteine-rich repeats fused to a CUB domain, which could bind TNF-α and modulate its action. Some of the fish iridoviruses encode an insulin-like growth factor and vascular endothelial growth factor-B, while both fish and amphibian iridoviruses encode a fibroblast growth factor, which might be deployed by the viruses to slow down cell death or augment growth of the tissues in which they proliferate. Some of the proteins encoded by the iridoviruses appear to interfere with the intracellular cytokine signaling. An entire array of proteins in the iridoviruses seems to have been recruited from the host for interfering with the immune response (Table 3). One striking example of these is a homolog of the SOCS-1 protein encoded by some fish iridoviruses, like SGV, which appears to have been recently acquired from the fish host (Fig. 6). The viral SOCS-1 homolog, however, only contains the SH2 domain and lacks the C-terminal SOCS domain that is required for signaling downstream of the JAK kinases in response to cytokines like interferon-γ (Larsen and Ropke, 2002). Similarly, other fish iridoviruses, like IKSNV, Red sea bream iridovirus (RSBV) and the Orange-spotted grouper iridovirus (OGV), encode a homolog of the host TRAF protein, which is an E3 UB-ligase required to initiate the activation of NFkB downstream of various cytokines (Pomerantz and Baltimore, 2002). The viral TRAF homologs lack either the N-terminal RING finger and CART domains (OGV) or just the CART domains (ISKNV and RBV). These proteins are likely to act as dominant negative regulators that directly block cytokine signaling immediately downstream of the receptor and thereby neutralize the host immune response.

Table 3
Examples of proteins predicted or known to be involved in virus–host interactions and immune evasion in NCLDVs

Virus–host interactions: apoptosis inhibition and immune evasion intracellular adaptations		
Protein/domain	Virus (gene name/s)	Comments
TGF-β like growth factor	FPV (FPV211)	Transforming growth factor like cystine knot
Pellino	MSV (MSV244)	Involved in Interleukin-1/Toll signaling by binding to the Pelle protein. Sequence analysis unifies this domain to the phosphopeptide-binding FHA domain. The MSV protein lacks the ring finger present in the eukaryotic versions.
Caspase	AMV (AMV063)	Cysteine protease with a distinct fold. The AMV protein is predicted to be inactive
Matrix metalloprotease	AMV (AMV070), MSV (MSV175, MSV176, MSV179), CIV (fuse ORF162R and 165R)	Zincin fold metalloprotease that is also found in baculoviruses
Baculovirus p35 like caspase inhibitor	AMV (AMV010, AMV021)	A distinct all beta strand sandwich that is widespread in the baculoviruses
Collagenase/metalloprotease	AMV (AMV003, AMV258)	Metalloprotease found in several baculoviruses. Probably aids in spreading of the virus
Enhancin	AMV (AMVITRIO)	Membrane associated domain that is shared with Baculovirus Enhancin protein. Enhancin is required for peritropic membrane disruption and fusion of nucleocapsids with midgut cells in Baculoviruses
Divergent SH2 domains	ASFV (DP141L, DP146L)	Very divergent SH2 domains, normally involved in protein-protein interaction during signaling by binding phosphotyrosine polypeptides
Two cysteine domain	CIV (232R, 378R, 380R), LDV (LDVICp184, LDVICp200, LDVICp013) and several orthologs in SGV and ISKNV, PBCV (A676R)	Small domain that may be present in 1–13 tandem copies. The domain is fused in several iridoviruses to a S/T kinase at their N-terminus. CIV 232R is fused to an OTU-like cysteine protease. The PBCV A676R protein is fused to a three stranded basic domain that in turn is fused to several S/T kinases in PBCV. This suggests that the two cysteine domain, the three stranded basic domain, the S/T kinase and the OTU like protease are involved in a viral-specific signaling pathway
TNFα receptor	VV (B28R), LDV (LDVICp101, LDVICp014, LDVICp016), SGV (ORF051L)	Contains characteristic TNFR repeats and binds the cytokine TNFα. The viral proteins are mainly soluble or secreted. The LDV proteins LDVICp101 and LDVICp014 are fused to a CUB domain at their C-terminus
CARD domain	LDV (LDVICp002), SGV (ORF048L), ATSV (ORF 40L), FV3 (FV3gorf64R)	Protein-protein interaction domain in the apoptosis signaling pathway
Fibroblast growth factor	SGV (ORF144R, ORF145R), ATSV (ATVp37)	A growth factor of the cystine knot fold
Vascular endothelial Growth factor-B-like	ISKNV (ORF048R)	A growth factor of the cystine knot fold
Insulin like growth factor	SGV (ORF062R), LDV1 (LCDV1gp068)	This protein is missing in the Chinese strain of LDV
Wnt	LDV (LDVICp126)	A signaling molecule that binds the Frizzled family of 7TM receptors and regulates cell differentiation
chemokine receptor	LDV (LDVICp058, LDVICp12)	7TM receptors
SOCS1-like	ISKNV (ORF103R)	Protein involved in cytokine signaling. Viral version lacks the C-terminal SOCS box domain
PAS domain protein	SGV (ORF075R)	The PAS domain is a small molecule ligand-binding domain. This iridoviral version is similar to PAS domains fused to bacterial histidine kinases
Semaphorin	SGV (ORF155R) VV (A39R)	Type I membrane protein with a SEMA domain, a TIG domain and an Immunoglobulin domain. Might inhibit migration of leukocytes. Independent acquired by poxviruses and iridoviruses
Other potential adaptations for virus–host interactions		
O-Glycosyltransferase	PBCV (A401R/A402R)	Nucleotide diphospho-sugar transferase fold protein of bacterial origin

Table 3 (*Continued*)

Virus–host interactions: apoptosis inhibition and immune evasion intracellular adaptations		
Protein/domain	Virus (gene name/s)	Comments
Fucosyltransferase	PBCV (A114R)	Protein of bacterial origin
Sugar transferase	PBCV (A111 R)	Composed of an N-terminal galactosyltransferase domain and a C-terminal glycosyltransferase domain
Tachycitin like chitin-binding domain	PBCV (A332L)	
Dps like protein	PBCV (A227L)	Iron chelating DNA-binding proteins of the ferretin fold that are involved in protecting DNA from oxidative damage. The PBCV version is related to the T4 phage rI.1A protein
PR5-like protein	ESV (ORF169)	Plant pathogenesis related protein involved in defense response in plants
O-methyltransferase	ESV (ORF226, ORF164)	Multidomain protein with a pectin methylesterase like single stranded beta helix at the N-terminus and a duplicated discoidin like double stranded jelly roll at the C-terminus
Rhomboid like protease	Mimivirus (L523)	Intramembrane serine proteases
CDC48	Mimivirus (R476)	AAA+ ATPase involved in vesicular transport
Predicted protease	Mimivirus (R408)	Predicted protease of the papain-like fold that is present in bacteria and kinetoplastids (Fig. 4E). The mimiviral version appears to be closer to the versions in Low GC gram positive bacteria.
Mitochondrial 18 kDa protein (MTP18)	Mimivirus (R740)	A protein localized to the mitochondria and the loss of which results in release of cytochrome *c* and activation of the caspase cascade
5′(3′)-deoxyribonucleotidase	Mimivirus (R824, 758)	Nucleotidase of the HAD fold. The mimivirus versions is closer to the ones targeted to the mitochondrion
BCS1 like AAA+ ATPase	Mimivirus (R776)	The protein is essential for proper protein folding in the mitochondrion
T4-like ADP ribosyltransferase	PBCV (A91L), Mimivirus (R217)	In the T4-like phages, this protein subverts a wide range of host proteins by modifying them with ADP-Ribose moieties

Notes: The above is only a small sampling of interesting proteins observed in NCLDV. As far as possible, we have tried to restrict our list to poorly characterized or poorly annotated proteins and protein orthologous groups.

The apparent absence of such counter-immunity and anti-apoptotic adaptations in the protist viruses, and their relative paucity of anti-apoptotic mechanisms in insect viruses (Table 3) are consistent with the late diversification of the apoptotic apparatus (Aravind et al., 2001) and emergence of adaptive immunity only in multicellular animals (Rinkevich, 2004).

7. Evolutionary implications of the relationships between core NCLDV proteins to proteins of other viruses

The above discussion shows that the NCLDV genomes form a complex web of evolutionary connections via HGT from their hosts, other endosymbionts, and unrelated viruses. Given that our reconstructions of the conserved gene core of the NCLDVs point to a common ancestor with a relatively large genome, the question arises as to how this ancestor arose in the first place. This immediately leads to the question of the relationship between the NCLDVs and other large eukaryotic DNA viruses, such as herpesviruses, baculoviruses, the White spot shrimp virus (WSSV), and polydnaviruses, and the prokaryotic viruses. A comparative analysis of the genomes of the large eukaryotic viruses reveals that WSSV and the polydnaviruses hardly share any homologous proteins with NCLDVs, suggesting that their origins might not be linked in any significant way.

The links to herpesviruses and baculoviruses are more complicated. These viruses differ from the NCLDVs most prominently in their virion assembly and DNA packaging systems. The baculoviruses have a unique set of conserved proteins that are required for virion maturation through budding (Pearson and Rohrmann, 2002), including components such as gp64, which, interestingly, is homologous to the envelope protein of insect-borne orthomyxoviruses (Ojala et al., 2001). We did not find any significant relationships between these baculoviral proteins and the proteins involved in NCLDV virion morphogenesis. The herpesviruses, in contrast, possess a terminase-portal protein system of DNA packaging, which they share with most of the large bacteriophages of the caudovirus class (Catalano, 2000; Newcomb et al., 2001). The only cognates of the NCLDV virion morphogenesis apparatus in other large DNA viruses are the homologs of the capsid protein that are present in a vast variety of viruses with icosahedral capsids (Hendrix, 1999; Nandhagopal et al., 2002) and the A32R-like ATPases detected in certain dsDNA phages of the tectivirus, corticovirus, and rudivirus families (Iyer et al., 2004c; Stromsten et al., 2005). These phages contain an internal lipid membrane similar to that present in several NCLDVs which

suggests significant parallels between their packaging mechanisms.

In other core systems, both baculoviruses and herpesviruses have several proteins that are homologous and functionally equivalent to the respective proteins of the NCLDVs. Like the NCLDVs, baculoviruses have RNA polymerases but their two large, catalytic subunits are distinct branches of the respective families and are only distantly related to the NCLDV RNA polymerase subunits (Iyer et al., 2003). Unlike the baculovirus versions, the NCLDV subunits are more closely related to the archaeo-eukaryotic cellular versions, and they also share additional RNA polymerase subunits exclusively with the archaeo-eukaryotic cellular polymerases (Fig. 1). A core DNA replication apparatus with a DNA polymerase, DNA helicase, primase, and PCNA-like DNA clamp is present in herpesviruses, baculoviruses, and the NCLDVs. The DNA polymerases of all these three viral clades, like the principal cellular DNA polymerases, belong to the B family. The PCNA homologs of baculoviruses appear to be recent acquisitions from their insect hosts and might not be essential components of their core replication system (Kool et al., 1994). In contrast, NCLDVs and herpesviruses contain distinctive versions of the DNA clamp that probably were parts of the replication apparatus from the early phases of their evolution. Like the NCLDV helicase, the baculovirus replicative helicase (P143) contains an AAA+ ATPase domain, which is a divergent member of the D5R family (Iyer et al., 2004b). It differs from the NCLDV helicases in possessing a large, unique N-terminal region and lacking a typical D5N domain. The replicative helicase of the herpesviruses (UL9) belongs to a specific family of SFII helicases. Among the NCLDVs, UL9-like helicases are encoded by ASFV and the mimivirus and are associated with primases related to the herpesviral primases. The two subunits of the baculovirus primases (Lef-1 and Lef-2) are, respectively, orthologous to the two subunits of the cellular primases of the archaeo-eukaryotic lineage (Iyer et al., 2005). As discussed above, the primases of the NCLDVs form two distinct, distantly related families, both of which were probably present in the ancestral NCLDV. The herpesvirus primase belongs to one of these families (the herpes-pox family) and appears to be specifically related to the NCLDV versions.

Looking beyond the large eukaryotic DNA viruses, homologs of family B DNA polymerases and the DNA clamp are seen in some large dsDNA bacteriophages of the caudovirus (tailed bacteriophage) class, e.g., T4 and its relatives. Likewise, both the UL9 and D5R-like helicases are widely distributed amongst the bacteriophages of the caudovirus class and several bacterial and archaeal plasmids (Iyer et al., 2005). Similarly, caudoviral counterparts can also be detected for the A18R-type helicases, ATP-dependent DNA ligases, the RuvC-like HJRs, KilA-N domains and SWI2/SNF2 ATPases. Representatives of the NCLDV-herpesvirus primase family so far have not been detected in phages or prokaryotes. However, members of the AEP superfamily of primases with the PriCT-2 domains, which are more distantly related to the NCLDV-herpesvirus primases, are seen in a wide range of prokaryotic plasmids and in the caudoviral group of phages (Iyer et al., 2005). The caudoviruses also encode homologs of several enzymes for deoxyribonucleotide metabolism that are found in large eukaryotic DNA viruses. For example, T4 encodes the two-subunit ribonucleotide reductase, thymidine kinase and thymidylate synthase, T5 has a dUTPase, and SPBc2 has a thymidylate kinase. Most of these phage nucleotide metabolism enzymes do not show evolutionary affinity with the NCLDV, herpesvirus or baculovirus counterparts, at least in conventional phylogenetic analyses. Although this implies parallel, independent acquisition, it is clear that eukaryotic and prokaryotic large DNA viruses employ very similar strategies for mobilizing the precursors for DNA synthesis. Likewise, many RNA repair enzymes such as RNA ligases, polynucleotide kinases/phosphatases and the 2H phosphoesterases are also encountered in several of the caudovirus group of phages and show specific relationships with the NCLDV and baculoviral versions. We also detected the T4-like ADP ribosyltransferase, which is widespread in the caudoviruses, in PBCV (PBCV A91L) and in mimivirus (MIMI_R217L). These enzymes are associated with the phage head and upon entry into the host they subvert a wide range of host proteins that include ribosomal and RNA polymerase subunits by modifying them with ADP-ribose moieties (Depping et al., 2005; Wilkens et al., 1997).

The distribution patterns of the key components of the DNA packaging, nucleotide metabolism, DNA replication and RNA repair machineries of the eukaryotic large DNA viruses suggest a complex set of connections with dsDNA phages. In the replication apparatus, there are several direct evolutionary links with the equivalent systems of the large bacteriophages, particularly, those of the caudoviruses. With respect to DNA packaging, herpesviruses group with caudoviruses in using a portal protein-terminase system, whereas the NCLDVs employ the same mechanism as phages of the tectivirus/corticovirus families that contain an internal lipid membrane and have a DNA-pumping ATPase of the HerA-FtsK superfamily. In terms of nucleotide metabolism, the large eukaryotic viruses and the phages of the caudovirus class have several functionally equivalent, homologous proteins, even if they are not necessarily monophyletic. The presence of nucleotide metabolism enzymes, in particular, distinguishes these viruses from the numerous prokaryotic and certain eukaryotic plasmids (e.g., the fungal linear mitochondrial plasmids), with which they may share common features in their replication proteins. The comparatively large gene complements of these large DNA viruses also distinguish them from the whole range of (relatively) small DNA viruses, such as the filamentous phages, geminiviruses, papovaviruses, and adenoviruses. In contrast to these smaller viruses, they appear to have a degree of autonomy in their life cycles, especially due to the above-described well-developed functional systems. This suggests that two fundamentally distinct life-cycle strategies, analogous to those of *K*-selected (large DNA viruses) and *r*-selected (small DNA viruses) species have been adopted in the viral universe (Maynard Smith, 1998). The choice between these strategies may have considerably affected the evolutionary trajectories of different DNA viruses.

Generally, despite the existence of indisputable evolutionary links, the current data do not allow us to conclude that either the major groups of eukaryotic large dsDNA viruses (NCLDV, herpesvirus, and baculoviruses), or any of these groups and any

particular family of large DNA bacteriophages evolved from a common ancestral virus. Accordingly, more complex evolutionary models are called for, and we consider two such models with somewhat overlapping scopes in terms of the explanatory capabilities:

(1) The large eukaryotic dsDNA viruses and large dsDNA phages are conceived as being derived independently from a pool of relatively small, plasmid-like elements which supplied a core ancestral set of replication proteins that are shared by different groups of large DNA viruses. These ancestral replicons then grew into larger and larger viruses by swapping genes with each other and gaining additional genes from the host. The sweeping spread of a relatively small set of high-fitness-conferring genes across these evolving viral replicons, and strong host-derived selective pressures favoring similar types of adaptations caused convergent evolution of similar gene content across diverse large dsDNA viruses. However, the intensity of gene exchanges affecting a certain core set of essential genes diminished with time, once the viral lineages diverged and stabilized.
(2) The majority of the large dsDNA viruses considered here, such as the NCLDVs, the herpesviruses, the baculoviruses, and various dsDNA phages, have descended from a small group of ancestral dsDNA viruses, that had already reached moderate genome sizes and encoded a number of distinct functional systems related to DNA replication, transcription, packaging and morphogenesis, and deoxyribonucleotide metabolism. As in the first model, these viruses converged to a degree due to inter-viral exchanges of certain advantageous genes. However, the vertical evolutionary relationships between them were chiefly eclipsed by non-orthologous gene displacements by functionally equivalent but unrelated or distantly related genes, or through xenologous displacement by cellular orthologs. Thus, this model posits that many of the homologous and functionally equivalent proteins shared by different groups of viruses are not due to convergence caused by similar selective pressures but rather reflects ancestral life-cycle constraints. Of course, this model prompts the question on the ultimate origin of the ancestral virus pool which may be interpreted within the framework of the first, plasmid model. However, the two models are distinct in the ways they stage virus evolution and reconstruct the immediate ancestry of the extant virus lineages.

There is considerable evidence for the common aspect in both models, namely, extensive gene transfers between various viral groups, including the genes for proteins involved in core functions like replication. For example, the herpesvirus primase and helicase might have been acquired from an NCLDV. The high rates of evolution of viral proteins make it hard, if possible at all, to decisively distinguish between the two models. However, several examples of non-orthologous gene displacement have been recovered from both eukaryotic and prokaryotic viruses, suggesting that the second model might, indeed, be viable (Fig. 3). Another, indirect argument also supports aspects of the second model, at least, with respect to the early stages of eukaryotic virus evolution. The eukaryotes emerged relatively late in evolution, well after the major bacterial and archaeal lineages had diversified (Doolittle et al., 1996). Bacteriophages seem to have played an important role in the formation of the mitochondrial replication apparatus, even before the divergence of the major eukaryotic lineages, with caudoviral-type DnaB helicase, DnaG-type primase, DNA polymerase and RNA polymerase replacing the bacterial counterparts (Aravind et al., 2003b). Thus, large phages, perhaps, those that replicated in the mitochondrial endosymbiont, seem to have been involved in the earliest stages of evolution of the eukaryotic cell. Extending this inference, it is not hard to imagine that some of these large phages evolved to propagate in eukaryotic cells. Given that, according to the reconstruction presented here, the common ancestor of the NCLDVs was already a fairly large, complex virus, it is conceivable that it evolved directly from a similar-sized phage. From the relationships described above, it is clear that such a precursor would combine features of the phages of the tectivirus/corticovirus and caudovirus groups. An initial phase of rapid sequence evolution, along with multiple non-orthologous and xenologous displacements, and gene transfers from eukaryotic sources would have helped this precursor to adapt to the eukaryotic cellular environment. These events would have eroded most of the traces of vertical evolutionary relationships between phages and the NCLDVs, with only a few strong connections between core genes remaining detectable.

8. Relationships between NCLDV core proteins and eukaryotic cellular systems

The precursor of the NCLDVs required a variety of specific adaptations to survive in the eukaryotic cytoplasmic environment. RNA polymerases, especially, homologs of the cellular large subunits containing the double-psi-β-barrel catalytic domains (Iyer et al., 2003), are rare in bacteriophages. Their appearance in the NCLDVs and baculoviruses seems to be a specific adaptation for virus reproduction in the eukaryotic cytoplasm. While the exact origins of the baculoviral RNA polymerase is unclear, the NCLDV polymerases appear to have been acquired directly from the eukaryotic cellular sources along with the additional subunits like Rpb2, Rpb5, Rpb10 and, possibly, the TFIIS homolog. Two other genes acquired by NCLDVs from the eukaryotes at the early stage of evolution of this virus class have been recruited for virion maturation. One of these is the ERV1-like disulfide-redox protein (Vaccinia E10) that was important for the formation of disulfide bonds required for virion assembly in the reducing environment of the eukaryotic cytoplasm, and the other is the SMT4-like peptidase that evolved from a eukaryotic DUB to participate in virion protein processing.

The 5′ cap is a distinct feature of the eukaryotic mRNAs, and the ability to synthesize caps would be a major adaptation of a relatively autonomous eukaryotic cytoplasmic virus. Both the guanylyl transferase and the triphosphatase, which are parts of the capping enzyme, appear to be major innovations in eukary-

otes (Anantharaman et al., 2002; Shuman, 2001). The guanylyl transferase is most closely related to the ATP-dependent DNA ligases, suggesting that it arose early in eukaryotic evolution through a duplication of the ligase (Anantharaman et al., 2002). Comparative analysis of the capping triphosphatases shows that the Ceg1p-like β-barrel triphosphatase is conserved in most eukaryotic lineages, such as fungi, *Dictyostelium*, apicomplexans, kinetoplastids, *Entamoeba* and *Giardia*. This was, probably, the ancestral capping triphosphatase, whereas, in plants and animals, it was displaced by a triphosphatase of the tyrosine phosphatase fold.

Structural comparisons showed that the β-barrel triphosphatases have the same fold as another class of triphosphatases, the CYTH domain enzymes (e.g., thiamin triphosphatase) (Iyer and Aravind, 2002). Both the Ceg1p-like capping triphosphatase and the CYTH domain have emerged from an ancestral duplication and circular permutation of a unit with 4 β-strands and 1 α-helix. They also share an unusual active site with 5 conserved acidic residues and 5 basic residues, which line opposite sides of the interior of the β-barrel formed by the duplicated 4 β-strand units. The CYTH enzymes are highly conserved in all archaea and widespread in most bacterial lineages, suggesting that a representative of this enzyme was already present in the last universal common ancestor (Iyer and Aravind, 2002). Thus, the eukaryotic capping triphosphatase appears to have been derived early in eukaryotic evolution from a CYTH domain precursor, which was present in the archaea. The capping triphosphatase of most NCLDVs, the baculoviruses and the linear fungal plasmids is more closely related to the Ceg1p-like proteins than to the CYTH domain triphosphatases seen in a few bacteriophages. Similarly, the viral guanylyl transferase is closer to the eukaryotic form than to RNA ligases or other members of this superfamily present in bacteriophages. Furthermore, the fusion of the capping triphosphatase, the guanylyl transferase, and the capping methylase in a single protein is a shared feature of most NCLDVs, baculoviruses, linear plasmids, and several eukaryotic lineages. These observations suggest that eukaryotic viruses and plasmids have acquired the capping enzyme from the eukaryotic cellular sources, after the eukaryotic capping apparatus had fully developed. The NCLDV polyA polymerase is extremely divergent and hardly shows any specific relationships with the cellular forms, beyond sharing the same fold of the catalytic domain (Aravind and Koonin, 1999a). The origin of this unusual viral polyA polymerase and the causes for its extreme divergence currently remain unclear.

In a few instances, there is clear evidence that the NCLDVs or other large eukaryotic DNA viruses might have also contributed genes to certain cellular systems of the eukaryotes. Members of the herpes-pox primase family are also present in crown group eukaryotes (plants, animals, *Dictyostelium*), apicomplexans, and kinetoplastids (Iyer et al., 2005), whereas members of the iridovirus primase family are seen only in the kinetoplastids (Iyer et al., 2005). Thus, these primases appear to have been acquired by the eukaryotes on two occasions, once prior to the divergence of major eukaryotic lineages, and the second time in the kinetoplastid lineage. Similarly, the RNA ligase of the kinetoplastid RNA-editing system also appears to have been acquired from the RNA-repair system of a large DNA virus (Aravind and Koonin, 1999b; Ho and Shuman, 2002).

9. General implications for the origins and evolution of viral and cellular life forms

Our reconstruction of the evolutionary history of the NCLDVs suggests that they are relatively late entries to the viral universe, which emerged in a recognizable form only after the eukaryotic cell with its entire complement of core structures was fully formed (Fig. 7). These viruses have subsequently acquired numerous genes from a variety of sources, such as the eukaryotic hosts and co-occurring endo-parasites and symbionts. The resulting gene complements of the NCLDVs are, in some respects, comparable to those of the eukaryotes. This is particularly evident in the gigantic genome of the mimivirus, which even led to the idea that it might represent a precursor for cellular life forms (Raoult et al., 2004). Given that the genes of the mimivirus are either of typical NCLDV provenance or, like in other viruses, show signs of HGT from different sources, this possibility appears untenable (Desjardins et al., 2005; Koonin, 2005). Nevertheless, the conserved NCLDV gene core includes distinct versions of proteins involved in several basic functions, such as replication, transcription and chromosome segregation, which are distantly related to their cellular homologs. Related versions of some of these proteins are also encountered elsewhere in the viral universe, i.e., in other eukaryotic dsDNA viruses, dsDNA phages, and some plasmids. It seems likely that these alternative viral replication enzymes, such as DNA polymerases, primases, DNA clamps, various types of helicases, and resolvases are remnants of the replication systems of ancient replicons that coexisted with the evolutionary precursors of the cellular replication systems (Forterre, 2001, 2002). This view is particularly consistent with the profound differences that exist between the bacterial and the archaeo-eukaryotic replication systems, which led to the hypothesis of two distinct origins of DNA replication (Leipe et al., 1999). Thus, it appears plausible that a variety of replicons with distinct replication systems co-existed in the pre-cellular phase of evolution, with some of them eventually growing in complexity and giving rise to the self-sufficient cellular systems (Fig. 7). The remaining replicons, with replication systems spanning a wide range of complexity, adapted to exploit the highly complex cellular systems as plasmids or viruses (Fig. 7).

The other major aspect of the virus life cycles is the segregation and packaging of their chromosome. The common origin of the capsid proteins of numerous RNA and DNA viruses with icosahedral capsids (Nandhagopal et al., 2002) suggests that they might have emerged early and were encoded by some of the ancient, pre-cellular replicons. The capsid would have provided a major selective advantage in the form of a protective coat for the nucleic acid and also favored their easy dispersal. We propose that these primordial replicons also developed at least two major systems for segregating and packaging their DNA: (1) the portal protein-terminase system; and (2) DNA pumping by HerA/FtsK superfamily ATPases. The second system is seen in viruses that typically contain an inner lipid membrane,

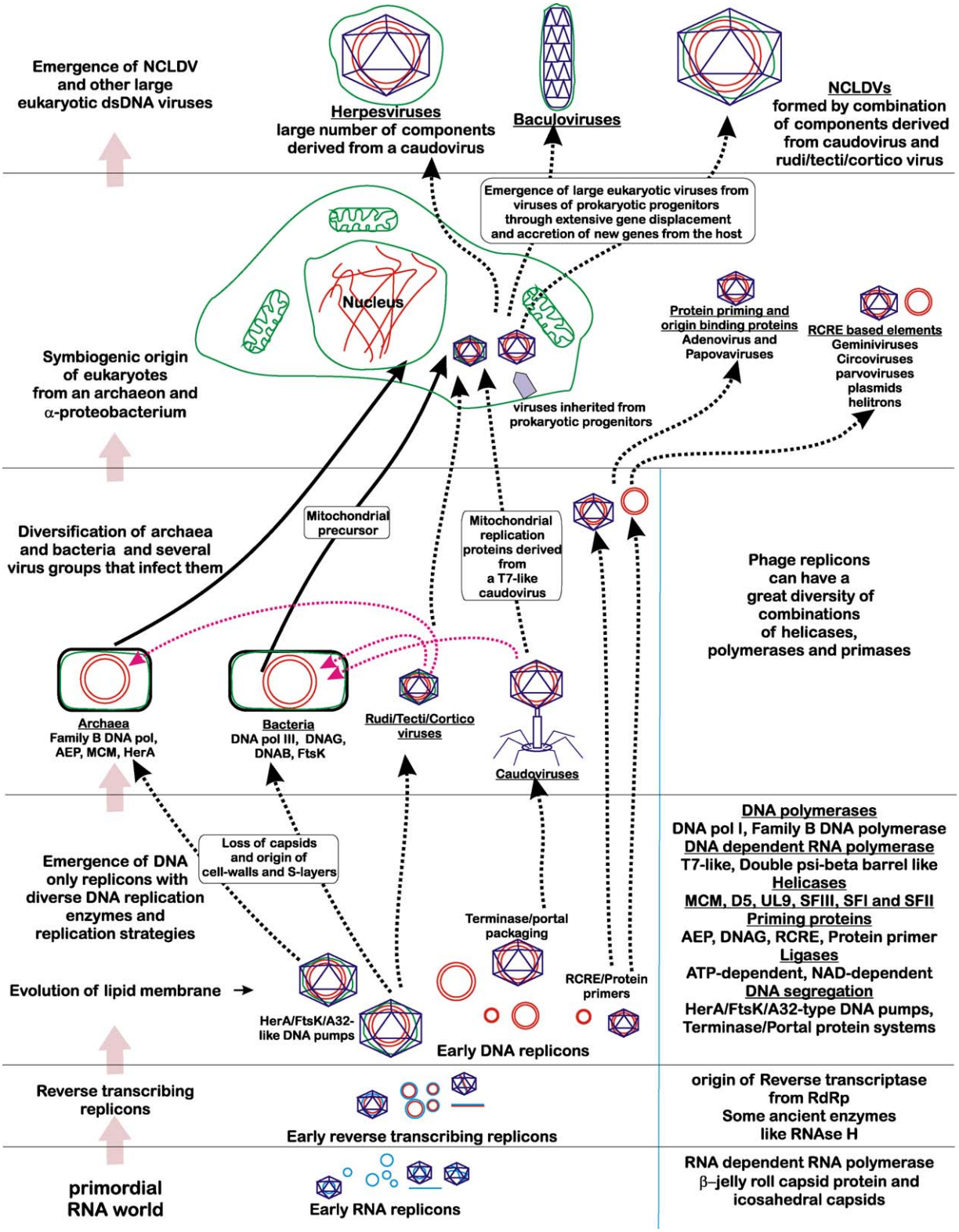

Fig. 7. A schematic representation of the probable scenario of evolution of DNA viruses and other DNA-based replicons. On the left, some of the major transitions in evolution are indicated. On the right the major innovations entailed by each transition are shown. Cartoons of various replicons and their associated envelopes corresponding to each evolutionary transition are also shown (not drawn to scale). DNA genomes are colored red, RNA genomes are colored blue, and lipid membranes are colored green.

like the NCLDVs, or in cells with lipid membranes. Hence, it might have originally operated in the context of lipid micelles that coated nucleic acids within protein capsids. The emergence of the secretory apparatus might have allowed these systems to break free from the capsid and stabilize or protect their lipid membrane more effectively. This might have ultimately led to the earliest cells, which continued to segregate their DNA using a DNA pump inherited from the ancestral virus-like elements. Alternatively, it is imaginable that the emerging cells (without specifying the scenario of cell origin) captured the HerA/FtsK-based pumping apparatus from the primordial virus-like entities and adopted it for chromosome segregation during cell division. Thus, at a basic level, comparative analysis of the functional systems of the NCLDVs and other DNA viruses might cast light on some of the early stages of life's evolution including the origin of replication systems and even of cells themselves.

The present re-investigation of the NCLDVs in light of the new information that has emerged since their original description as a monophyletic group (Iyer et al., 2001) resulted in greater clarity regarding their evolution and origins, and their evolutionary connections to other viruses and cellular systems. We also hope that this type of analysis may serve as a general model for future comparative-genomic and phylogenetic studies on various classes of large DNA viruses as more sequences become available.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.virusres.2006.01.009.

Acknowledgments

This research was supported by the intramural Research Program of the National Center for Biotechnology Information, National Library of Medicine, NIH.

References

Adachi, J., Hasegawa, M., 1992. MOLPHY: Programs for Molecular Phylogenetics. Institute of Statistical Mathematics, Tokyo.

Afonso, C.L., Piccone, M.E., Zaffuto, K.M., Neilan, J., Kutish, G.F., Lu, Z., Balinsky, C.A., Gibb, T.R., Bean, T.J., Zsak, L., Rock, D.L., 2004. African swine fever virus multigene family 360 and 530 genes affect host interferon response. J. Virol. 78 (4), 1858–1864.

Alejo, A., Andres, G., Salas, M.L., 2003. African Swine Fever virus proteinase is essential for core maturation and infectivity. J. Virol. 77 (10), 5571–5577.

Alstein, A.D., 1992. The protocellular concept of the origin of viruses. Semin. Virol. 3, 409–417.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25 (17), 3389–3402.

Ameisen, J.C., 2002. On the origin, evolution, and nature of programmed cell death: a timeline of four billion years. Cell Death Differ. 9 (4), 367–393.

Ameisen, J.C., Pleskoff, O., Lelievre, J.D., De Bels, F., 2003. Subversion of cell survival and cell death: viruses as enemies, tools, teachers and allies. Cell Death Differ. 10 (Suppl. 1), 3–6.

Anantharaman, V., Aravind, L., 2003. Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. Genome Biol. 4 (2), R11.

Anantharaman, V., Koonin, E.V., Aravind, L., 2002. Comparative genomics and evolution of proteins involved in RNA metabolism. Nucleic Acids Res. 30 (7), 1427–1464.

Ansarah-Sobrinho, C., Moss, B., 2004a. Role of the I7 protein in proteolytic processing of vaccinia virus membrane and core components. J. Virol. 78 (12), 6335–6343.

Ansarah-Sobrinho, C., Moss, B., 2004b. Vaccinia virus G1 protein, a predicted metalloprotease, is essential for morphogenesis of infectious virions but not for cleavage of major core proteins. J. Virol. 78 (13), 6855–6863.

Aravind, L., Anantharaman, V., Iyer, L.M., 2003a. Evolutionary connections between bacterial and eukaryotic signaling systems: a genomic perspective. Curr. Opin. Microbiol. 6 (5), 490–497.

Aravind, L., Dixit, V.M., Koonin, E.V., 2001. Apoptotic molecular machinery: vastly increased complexity in vertebrates revealed by genome comparisons. Science 291 (5507), 1279–1284.

Aravind, L., Iyer, L.M., 2002. The SWIRM domain: a conserved module found in chromosomal proteins points to novel chromatin-modifying activities. Genome Biol. 3 (8).

Aravind, L., Iyer, L.M., Wellems, T.E., Miller, L.H., 2003b. Plasmodium biology: genomic gleanings. Cell 115 (7), 771–785.

Aravind, L., Koonin, E.V., 1999a. DNA polymerase beta-like nucleotidyltransferase superfamily: identification of three new families, classification and evolutionary history. Nucleic Acids Res. 27 (7), 1609–1618.

Aravind, L., Koonin, E.V., 1999b. Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches [in process citation]. J. Mol. Biol. 287 (5), 1023–1040.

Aravind, L., Makarova, K.S., Koonin, E.V., 2000. Survey and summary: holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. Nucleic Acids Res. 28 (18), 3417–3432.

Aravind, L., Walker, D.R., Koonin, E.V., 1999. Conserved domains in DNA repair proteins and evolution of repair systems. Nucleic Acids Res. 27 (5), 1223–1242.

Aussel, L., Barre, F.X., Aroyo, M., Stasiak, A., Stasiak, A.Z., Sherratt, D., 2002. FtsK Is a DNA motor protein that activates chromosome dimer resolution by switching the catalytic state of the XerC and XerD recombinases. Cell 108 (2), 195–205.

Barry, M., Wasilenko, S.T., Stewart, T.L., Taylor, J.M., 2004. Apoptosis regulator genes encoded by poxviruses. Prog. Mol. Subcell. Biol. 36, 19–37.

Bertin, J., Armstrong, R.C., Ottilie, S., Martin, D.A., Wang, Y., Banks, S., Wang, G.H., Senkevich, T.G., Alnemri, E.S., Moss, B., Lenardo, M.J., Tomaselli, K.J., Cohen, J.I., 1997. Death effector domain-containing herpesvirus and poxvirus proteins inhibit both Fas- and TNFR1-induced apoptosis. Proc. Natl. Acad. Sci. U.S.A. 94 (4), 1172–1176.

Betakova, T., Wolffe, E.J., Moss, B., 1999. Regulation of vaccinia virus morphogenesis: phosphorylation of the A14L and A17L membrane proteins and C-terminal truncation of the A17L protein are dependent on the F10L kinase. J. Virol. 73 (5), 3534–3543.

Bhagwat, M., Nossal, N.G., 2001. Bacteriophage T4 RNase H removes both RNA primers and adjacent DNA from the 5′ end of lagging strand fragments. J. Biol. Chem. 276 (30), 28516–28524.

Bonnerot, C., Pintard, L., Lutfalla, G., 2003. Functional redundancy of Spb1p and a snR52-dependent mechanism for the 2′-*O*-ribose methylation of a conserved rRNA position in yeast. Mol. Cell 12 (5), 1309–1315.

Bowie, A.G., Zhan, J., Marshall, W.L., 2004. Viral appropriation of apoptotic and NF-kappaB signaling pathways. J. Cell. Biochem. 91 (6), 1099–1108.

Bugert, J.J., Darai, G., 2000. Poxvirus homologues of cellular genes. Virus Genes 21 (1–2), 111–133.

Byrd, C.M., Bolken, T.C., Hruby, D.E., 2002. The vaccinia virus I7L gene product is the core protein proteinase. J. Virol. 76 (17), 8973–8976.

Cassetti, M.C., Merchlinsky, M., Wolffe, E.J., Weisberg, A.S., Moss, B., 1998. DNA packaging mutant: repression of the vaccinia virus A32 gene results in noninfectious, DNA-deficient, spherical, enveloped particles. J. Virol. 72 (7), 5769–5780.

Catalano, C.E., 2000. The terminase enzyme from bacteriophage lambda: a DNA-packaging machine. Cell Mol. Life Sci. 57 (1), 128–148.

Clem, R.J., 2001. Baculoviruses and apoptosis: the good, the bad, and the ugly. Cell Death Differ. 8 (2), 137–143.

Cossart, P., Pizarro-Cerda, J., Lecuit, M., 2003. Invasion of mammalian cells by Listeria monocytogenes: functional mimicry to subvert cellular functions. Trends Cell Biol. 13 (1), 23–31.

da Fonseca, F.G., Weisberg, A.S., Caeiro, M.F., Moss, B., 2004. Vaccinia virus mutants with alanine substitutions in the conserved G5R gene fail to initiate morphogenesis at the nonpermissive temperature. J. Virol. 78 (19), 10238–10248.

da Fonseca, F.G., Wolffe, E.J., Weisberg, A., Moss, B., 2000. Effects of deletion or stringent repression of the H3L envelope gene on vaccinia virus replication. J. Virol. 74 (16), 7518–7528.

Davison, A.J., Trus, B.L., Cheng, N., Steven, A.C., Watson, M.S., Cunningham, C., Le Deuff, R.M., Renault, T., 2005. A novel class of herpesvirus with bivalve hosts. J. Gen. Virol. 86 (Pt 1), 41–53.

DeAngelis, P.L., Jing, W., Graves, M.V., Burbank, D.E., Van Etten, J.L., 1997. Hyaluronan synthase of chlorella virus PBCV-1. Science 278 (5344), 1800–1803.

Delarue, M., Poch, O., Tordo, N., Moras, D., Argos, P., 1990. An attempt to unify the structure of polymerases. Protein Eng. 3 (6), 461–467.

Dellis, S., Strickland, K.C., McCrary, W.J., Patel, A., Stocum, E., Wright, C.F., 2004. Protein interactions among the vaccinia virus late transcription factors. Virology 329 (2), 328–336.

Deneke, J., Ziegelin, G., Lurz, R., Lanka, E., 2002. Phage N15 telomere resolution. Target requirements for recognition and processing by the protelomerase. J. Biol. Chem. 277 (12), 10410–10419.

Depping, R., Lohaus, C., Meyer, H.E., Ruger, W., 2005. The mono-ADP-ribosyltransferases Alt and ModB of bacteriophage T4: target proteins identified. Biochem. Biophys. Res. Commun. 335 (4), 1217–1223.

Desjardins, C., Eisen, J.A., Nene, V., 2005. New evolutionary frontiers from unusual virus genomes. Genome Biol. 6 (3), 212.

Do, J.W., Moon, C.H., Kim, H.J., Ko, M.S., Kim, S.B., Son, J.H., Kim, J.S., An, E.J., Kim, M.K., Lee, S.K., Han, M.S., Cha, S.J., Park, M.S., Park, M.A., Kim, Y.C., Kim, J.W., Park, J.W., 2004. Complete genomic DNA sequence of rock bream iridovirus. Virology 325 (2), 351–363.

Doolittle, R.F., Feng, D.F., Tsang, S., Cho, G., Little, E., 1996. Determining divergence times of the major kingdoms of living organisms with a protein clock. Science 271 (5248), 470–477.

Ellis, J., Dodds, P., Pryor, T., 2000. Structure, function and evolution of plant disease resistance genes. Curr. Opin. Plant Biol. 3 (4), 278–284.

Felsenstein, J., 2004. Inferring Phylogenies. Sinauer Associates, Sunderland, MA.

Forterre, P., 2001. Genomics and early cellular evolution. The origin of the DNA world. C.R. Acad. Sci. III 324 (12), 1067–1076.

Forterre, P., 2002. The origin of DNA genomes and DNA replication proteins. Curr. Opin. Microbiol. 5 (5), 525–532.

Fricke, W.M., Brill, S.J., 2003. Slx1-Slx4 is a second structure-specific endonuclease functionally redundant with Sgs1-Top3. Genes Dev. 17 (14), 1768–1778.

Garcia, A.D., Aravind, L., Koonin, E.V., Moss, B., 2000. Bacterial-type DNA holliday junction resolvases in eukaryotic viruses. Proc. Natl. Acad. Sci. U.S.A. 97 (16), 8926–8931.

Gibbs, A., Calisher, C.H., Garcia-Arenal, F., 1995. Molecular Basis of Virus Evolution. Cambridge University Press, Cambridge, UK.

Gorbalenya, A.E., Koonin, E.V., Wolf, Y.I., 1990. A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses. FEBS Lett. 262 (1), 145–148.

Graves, M.V., Burbank, D.E., Roth, R., Heuser, J., DeAngelis, P.L., Van Etten, J.L., 1999. Hyaluronan synthesis in virus PBCV-1-infected chlorella-like green algae. Virology 257 (1), 15–23.

Graziani, S., Xia, Y., Gurnon, J.R., Van Etten, J.L., Leduc, D., Skouloubris, S., Myllykallio, H., Liebl, U., 2004. Functional analysis of FAD-dependent thymidylate synthase ThyX from Paramecium bursaria Chlorella virus-1. J. Biol. Chem. 279 (52), 54340–54347.

Haber, J.E., Heyer, W.D., 2001. The fuss about Mus81. Cell 107 (5), 551–554.

He, J.G., Deng, M., Weng, S.P., Li, Z., Zhou, S.Y., Long, Q.X., Wang, X.Z., Chan, S.M., 2001. Complete genome analysis of the mandarin fish infectious spleen and kidney necrosis iridovirus. Virology 291 (1), 126–139.

He, J.G., Lu, L., Deng, M., He, H.H., Weng, S.P., Wang, X.H., Zhou, S.Y., Long, Q.X., Wang, X.Z., Chan, S.M., 2002. Sequence analysis of the complete genome of an iridovirus isolated from the tiger frog. Virology 292 (2), 185–197.

Hedengren-Olcott, M., Byrd, C.M., Watson, J., Hruby, D.E., 2004. The vaccinia virus G1L putative metalloproteinase is essential for viral replication in vivo. J. Virol. 78 (18), 9947–9953.

Hendrix, R.W., 1999. Evolution: the long evolutionary reach of viruses. Curr. Biol. 9 (24), 914–917.

Hendrix, R.W., 2003. Bacteriophage genomics. Curr. Opin. Microbiol. 6 (5), 506–511.

Hirano, T., 2005. SMC proteins and chromosome mechanics: from bacteria to humans. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 360 (1455), 507–514.

Ho, C.K., Shuman, S., 2002. Bacteriophage T4 RNA ligase 2 (gp24.1) exemplifies a family of RNA ligases found in all phylogenetic domains. Proc. Natl. Acad. Sci. U.S.A. 99 (20), 12709–12714.

Hofmann, K., Bucher, P., 1995. The FHA domain: a putative nuclear signalling domain found in protein kinases and transcription factors. Trends Biochem. Sci. 20 (9), 347–349.

Hu, F.Q., Smith, C.A., Pickup, D.J., 1994. Cowpox virus contains two copies of an early gene encoding a soluble secreted form of the type II TNF receptor. Virology 204 (1), 343–356.

Huang, J., Huang, Q., Zhou, X., Shen, M.M., Yen, A., Yu, S.X., Dong, G., Qu, K., Huang, P., Anderson, E.M., Daniel-Issakani, S., Buller, R.M.L., Payan, D.G., Lu, H.H., 2004a. The poxvirus p28 virulence factor is an E3 ubiquitin ligase. J. Biol. Chem. 279 (52), 54110–54116.

Huang, W.M., Joss, L., Hsieh, T., Casjens, S., 2004b. Protelomerase uses a topoisomerase IB/Y-recombinase type mechanism to generate DNA hairpin ends. J. Mol. Biol. 337 (1), 77–92.

Hughes, A.L., Friedman, R., 2005. Poxvirus genome evolution by gene gain and loss. Mol. Phylogenet. Evol. 35 (1), 186–195.

Ilyina, T.V., Koonin, E.V., 1992. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. Nucleic Acids Res. 20 (13), 3279–3285.

Iyer, L.M., Aravind, L., 2002. The catalytic domains of thiamine triphosphatase and CyaB-like adenylyl cyclase define a novel superfamily of domains that bind organic phosphates. BMC Genomics 3 (1), 33–33.

Iyer, L.M., Aravind, L., Koonin, E.V., 2001. Common origin of four diverse families of large eukaryotic DNA viruses. J. Virol. 75 (23), 11720–11734.

Iyer, L.M., Koonin, E.V., Aravind, L., 2002. Extensive domain shuffling in transcription regulators of DNA viruses and implications for the origin of fungal APSES transcription factors. Genome Biol. 3 (3) (Research 0012).

Iyer, L.M., Koonin, E.V., Aravind, L., 2003. Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerises. BMC Struct. Biol. 3 (1), 1.

Iyer, L.M., Koonin, E.V., Aravind, L., 2004a. Evolution of bacterial RNA polymerase: implications for large-scale bacterial phylogeny, domain accretion, and horizontal gene transfer. Gene 335, 73–88.

Iyer, L.M., Koonin, E.V., Leipe, D.D., Aravind, L., 2005. Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. Nucleic Acids Res. 33 (12), 3875–3896.

Iyer, L.M., Leipe, D.D., Koonin, E.V., Aravind, L., 2004b. Evolutionary history and higher order classification of AAA+ ATPases. J. Struct. Biol. 146 (1–2), 11–31.

Iyer, L.M., Makarova, K.S., Koonin, E.V., Aravind, L., 2004c. Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging. Nucleic Acids Res. 32 (17), 5260–5279.

Jancovich, J.K., Mao, J., Chinchar, V.G., Wyatt, C., Case, S.T., Kumar, S., Valente, G., Subramanian, S., Davidson, E.W., Collins, J.P., Jacobs, B.L., 2003. Genomic sequence of a ranavirus (family *Iridoviridae*) associated with salamander mortalities in North America. Virology 316 (1), 90–103.

Jiang, Z., Johnson, H.J., Nie, H., Qin, J., Bird, T.A., Li, X., 2003. Pellino 1 is required for interleukin-1 (IL-1)-mediated signaling through its interaction with the IL-1 receptor-associated kinase 4 (IRAK4)-IRAK-tumor necrosis factor receptor-associated factor 6 (TRAF6) complex. J. Biol. Chem. 278 (13), 10952–10956.

Kamer, G., Argos, P., 1984. Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. Nucleic Acids Res. 12 (18), 7269–7282.

Kapitonov, V.V., Jurka, J., 2001. Rolling-circle transposons in eukaryotes. Proc. Natl. Acad. Sci. U.S.A. 98 (15), 8714–8719.

Karev, G.P., Wolf, Y.I., Koonin, E.V., 2003. Simple stochastic birth and death models of genome evolution: was there enough time for us to evolve? Bioinformatics 19 (15), 1889–1900.

Kool, M., Ahrens, C.H., Goldbach, R.W., Rohrmann, G.F., Vlak, J.M., 1994. Identification of genes involved in DNA replication of the Autographa californica baculovirus. Proc. Natl. Acad. Sci. U.S.A. 91 (23), 11212–11216.

Koonin, E.V., 1992. Introduction: virus evolution. Time for Schturm and Drang. Seminars in Virology 3, 311–313.

Koonin, E.V., 2005. Virology: Gulliver among the Lilliputians. Curr. Biol. 15 (5), R167–R169.

Koonin, E.V., Aravind, L., 2002. Origin and evolution of eukaryotic apoptosis: the bacterial connection. Cell Death Differ. 9 (4), 394–404.

Koonin, E.V., Gorbalenya, A.E., Chumakov, K.M., 1989. Tentative identification of RNA-dependent RNA polymerases of dsRNA viruses and their relationship to positive strand RNA viral polymerases. FEBS Lett. 252 (1–2), 42–46.

Koonin, E.V., Makarova, K.S., Aravind, L., 2001. Horizontal gene transfer in prokaryotes: quantification and classification. Annu. Rev. Microbiol. 55, 709–742.

Koonin, E.V., Senkevich, T.G., Chernos, V.I., 1993. Gene A32 product of vaccinia virus may be an ATPase involved in viral DNA packaging as indicated by sequence comparisons with other putative viral ATPases. Virus Genes 7 (1), 89–94.

Koonin, E.V., Wolf, Y.I., Karev, G.P., 2002. The structure of the protein universe and genome evolution. Nature 420 (6912), 218–223.

Kovall, R., Matthews, B.W., 1997. Toroidal structure of lambda-exonuclease. Science 277 (5333), 1824–1827.

Lackner, C.A., Condit, R.C., 2000. Vaccinia virus gene A18R DNA helicase is a transcript release factor. J. Biol. Chem. 275 (2), 1485–1494.

Larsen, L., Ropke, C., 2002. Suppressors of cytokine signalling: SOCS. APMIS 110 (12), 833–844.

Lauzon, H.A., Jamieson, P.B., Krell, P.J., Arif, B.M., 2005. Gene organization and sequencing of the Choristoneura fumiferana defective nucleopolyhedrovirus genome. J. Gen. Virol. 86 (Pt 4), 945–961.

Leipe, D.D., Aravind, L., Koonin, E.V., 1999. Did DNA replication evolve twice independently? Nucleic Acids Res. 27 (17), 3389–33401.

Lespinet, O., Wolf, Y.I., Koonin, E.V., Aravind, L., 2002. The role of lineage-specific gene family expansion in the evolution of eukaryotes. Genome Res. 12 (7), 1048–1059.

Lin, C.L., Chung, C.S., Heine, H.G., Chang, W., 2000. Vaccinia virus envelope H3L protein binds to cell surface heparan sulfate and is important for intracellular mature virion morphogenesis and virus infection in vitro and in vivo. J. Virol. 74 (7), 3353–3365.

Maier, I., Parodi, E., Westermeier, R., Muller, D.G., 2000. Maullinia ectocarpii gen. et sp. nov. (Plasmodiophorea), an intracellular parasite in Ectocarpus siliculosus (Ectocarpales, Phaeophyceae) and other filamentous brown algae. Protist 151 (3), 225–238.

Mann, N.H., Clokie, M.R.J., Millard, A., Cook, A., Wilson, W.H., Wheatley, P.J., Letarov, A., Krisch, H.M., 2005. The genome of S-PM2, a "photosynthetic" T4-type bacteriophage that infects marine Synechococcus strains. J. Bacteriol. 187 (9), 3188–3200.

Markine-Goriaynoff, N., Gillet, L., Van Etten, J.L., Korres, H., Verma, N., Vanderplasschen, A., 2004. Glycosyltransferases encoded by viruses. J. Gen. Virol. 85 (Pt 10), 2741–2754.

Martins, A., Shuman, S., 2004. Characterization of a baculovirus enzyme with RNA ligase, polynucleotide 5′-kinase, and polynucleotide 3′-phosphatase activities. J. Biol. Chem. 279 (18), 18220–18231.

Maynard Smith, J., 1998. Evolutionary Genetics. Oxford University Press, Oxford, UK.

McLysaght, A., Baldi, P.F., Gaut, B.S., 2003. Extensive gene gain associated with adaptive evolution of poxviruses. Proc. Natl. Acad. Sci. U.S.A. 100 (26), 15655–15660.

Moroz, O.V., Murzin, A.G., Makarova, K.S., Koonin, E.V., Wilson, K.S., Galperin, M.Y., 2005. Dimeric dUTPases, HisE, and MazG belong to a new superfamily of all-alpha NTP pyrophosphohydrolases with potential "house-cleaning" functions. J. Mol. Biol. 347 (2), 243–255.

Mosavi, L.K., Cammett, T.J., Desrosiers, D.C., Peng, Z.Y., 2004. The ankyrin repeat as molecular architecture for protein recognition. Protein Sci. 13 (6), 1435–1448.

Moss, B., Shisler, J.L., Xiang, Y., Senkevich, T.G., 2000. Immune-defense molecules of molluscum contagiosum virus, a human poxvirus. Trends Microbiol. 8 (10), 473–477.

Mossman, K., Ostergaard, H., Upton, C., McFadden, G., 1995. Myxoma virus and Shope fibroma virus encode dual-specificity tyrosine/serine phosphatases which are essential for virus viability. Virology 206 (1), 572–582.

Muller, D.G., Kapp, M., Knippers, R., 1998. Viruses in marine brown algae. Adv. Virus Res. 50, 49–67.

Myllykallio, H., Lipowski, G., Leduc, D., Filee, J., Forterre, P., Liebl, U., 2002. An alternative flavin-dependent mechanism for thymidylate synthesis. Science 297 (5578), 105–107.

Nandhagopal, N., Simpson, A.A., Gurnon, J.R., Yan, X., Baker, T.S., Graves, M.V., Van Etten, J.L., Rossmann, M.G., 2002. The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus. Proc. Natl. Acad. Sci. U.S.A. 99 (23), 14758–14763.

Nerenberg, B.T.H., Taylor, J., Bartee, E., Gouveia, K., Barry, M., Fruh, K., 2005. The poxviral RING protein p28 is a ubiquitin ligase that targets ubiquitin to viral replication factories. J. Virol. 79 (1), 597–601.

Newcomb, W.W., Juhas, R.M., Thomsen, D.R., Homa, F.L., Burch, A.D., Weller, S.K., Brown, J.C., 2001. The UL6 gene product forms the portal for entry of DNA into the herpes simplex virus capsid. J. Virol. 75 (22), 10923–10932.

Ojala, K., Mottershead, D.G., Suokko, A., Oker-Blom, C., 2001. Specific binding of baculoviruses displaying gp64 fusion proteins to mammalian cells. Biochem. Biophys. Res. Commun. 284 (3), 777–784.

Oroskar, A.A., Read, G.S., 1989. Control of mRNA stability by the virion host shutoff function of herpes simplex virus. J. Virol. 63 (5), 1897–1906.

Pearson, M.N., Rohrmann, G.F., 2002. Transfer, incorporation, and substitution of envelope fusion proteins among members of the *Baculoviridae, Orthomyxoviridae*, and *Metaviridae* (insect retrovirus) families. J. Virol. 76 (11), 5301–5304.

Pintard, L., Lecointe, F., Bujnicki, J.M., Bonnerot, C., Grosjean, H., Lapeyre, B., 2002. Trm7p catalyses the formation of two 2′-*O*-methylriboses in yeast tRNA anticodon loop. EMBO J. 21 (7), 1811–1820.

Poch, O., Sauvaget, I., Delarue, M., Tordo, N., 1989. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. EMBO J. 8 (12), 3867–3874.

Pomerantz, J.L., Baltimore, D., 2002. Two pathways to NF-kappaB. Mol. Cell 10 (4), 693–695.

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 (5700), 1344–1350.

Ravin, N.V., Strakhova, T.S., Kuprianov, V.V., 2001. The protelomerase of the phage-plasmid N15 is responsible for its maintenance in linear form. J. Mol. Biol. 312 (5), 899–906.

Rinkevich, B., 2004. Primitive immune systems: are your ways my ways? Immunol. Rev. 198, 25–35.

Robinson, D.N., Cooley, L., 1997. Drosophila kelch is an oligomeric ring canal actin organizer. J. Cell Biol. 138 (4), 799–810.

Seet, B.T., Johnston, J.B., Brunetti, C.R., Barrett, J.W., Everett, H., Cameron, C., Sypula, J., Nazarian, S.H., Lucas, A., McFadden, G., 2003. Poxviruses and immune evasion. Annu. Rev. Immunol. 21, 377–423.

Senkevich, T.G., Bugert, J.J., Sisler, J.R., Koonin, E.V., Darai, G., Moss, B., 1996. Genome sequence of a human tumorigenic poxvirus: prediction of specific host response-evasion genes. Science 273 (5276), 813–816.

Senkevich, T.G., Koonin, E.V., Bugert, J.J., Darai, G., Moss, B., 1997. The genome of molluscum contagiosum virus: analysis and comparison with other poxviruses. Virology 233 (1), 19–42.

Senkevich, T.G., Koonin, E.V., Buller, R.M., 1994. A poxvirus protein with a RING zinc finger motif is of crucial importance for virulence. Virology 198 (1), 118–128.

Senkevich, T.G., White, C.L., Koonin, E.V., Moss, B., 2002. Complete pathway for protein disulfide bond formation encoded by poxviruses. Proc. Natl. Acad. Sci. U.S.A. 99 (10), 6667–6672.

Shisler, J.L., Moss, B., 2001. Immunology 102 at poxvirus U: avoiding apoptosis. Semin. Immunol. 13 (1), 67–72.

Shuman, S., 2001. Structure, mechanism, and evolution of the mRNA capping apparatus. Prog. Nucleic Acid. Res. Mol. Biol. 66, 1–40.

Song, W.J., Qin, Q.W., Qiu, J., Huang, C.H., Wang, F., Hew, C.L., 2004. Functional genomics analysis of Singapore grouper iridovirus: complete sequence determination and proteomic analysis. J. Virol. 78 (22), 12576–12590.

Stephens, R.S., Kalman, S., Lammel, C., Fan, J., Marathe, R., Aravind, L., Mitchell, W., Olinger, L., Tatusov, R.L., Zhao, Q., Koonin, E.V., Davis, R.W., 1998. Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis. Science 282 (5389), 754–759.

Stromsten, N.J., Bamford, D.H., Bamford, J.K.H., 2005. In vitro DNA packaging of PRD1: a common mechanism for internal-membrane viruses. J. Mol. Biol. 348 (3), 617–629.

Sun, L., Gurnon, J.R., Adams, B.J., Graves, M.V., Van Etten, J.L., 2000. Characterization of a beta-1,3-glucanase encoded by chlorella virus PBCV-1. Virology 276 (1), 27–36.

Swinger, K.K., Rice, P.A., 2004. IHF and HU: flexible architects of bent DNA. Curr. Opin. Struct. Biol. 14 (1), 28–35.

Swofford, D.L., 2000. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Szajner, P., Weisberg, A.S., Lebowitz, J., Heuser, J., Moss, B., 2005. External scaffold of spherical immature poxvirus particles is made of protein trimers, forming a honeycomb lattice. J. Cell Biol. 170 (6), 971–981.

Taddeo, B., Esclatine, A., Roizman, B., 2002. The patterns of accumulation of cellular RNAs in cells infected with a wild-type and a mutant herpes simplex virus 1 lacking the virion host shutoff gene. Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17031–17036.

Teyssier, C., Laine, B., Gervais, A., Maurizot, J.C., Culard, F., 1994. Archaebacterial histone-like protein MC1 can exhibit a sequence-specific binding to DNA. Biochem. J. 303 (Pt 2), 567–573.

Tsai, C.T., Ting, J.W., Wu, M.H., Wu, M.F., Guo, I.C., Chang, C.Y., 2005. Complete genome sequence of the grouper iridovirus and comparison of genomic organization with those of other iridoviruses. J. Virol. 79 (4), 2010–2023.

Van Etten, J.L., 2003. Unusual life style of giant chlorella viruses. Annu. Rev. Genet. 37, 153–195.

Van Etten, J.L., Burbank, D.E., Meints, R.H., 1986. Replication of the algal virus PBCV-1 in UV-irradiated Chlorella. Intervirology 26 (1–2), 115–120.

Wagner, K.E., Hewlett, M.J., 2003. In: Fields, B.N., Howley, P.M., Griffin, D.E., Lamb, R.A., Martin, M.A., Roizman, B., Straus, S.E., Knipe, D.M. (Eds.), Basic Virology, second ed. Blackwell Publishers, Oxford, UK.

Wilkens, K., Tiemann, B., Bazan, F., Ruger, W., 1997. ADP-ribosylation and early transcription regulation by bacteriophage T4. Adv. Exp. Med. Biol. 419, 71–82.

Wilson, W.H., Schroeder, D.C., Allen, M.J., Holden, M.T.G., Parkhill, J., Barrell, B.G., Churcher, C., Hamlin, N., Mungall, K., Norbertczak, H., Quail, M.A., Price, C., Rabbinowitsch, E., Walker, D., Craigon, M., Roy, D., Ghazal, P., 2005. Complete genome sequence and lytic phase transcription profile of a Coccolithovirus. Science 309 (5737), 1090–1092.

Wintersberger, U., Wintersberger, E., 1987. Retroviruses and the origin of life. Trends Genet. 3, 198–202.

Xiang, Y., Moss, B., 1999. Identification of human and mouse homologs of the MC51L-53L-54L family of secreted glycoproteins encoded by the Molluscum contagiosum poxvirus. Virology 257 (2), 297–302.

Xiong, Y., Eickbush, T.H., 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. 9 (10), 3353–3362.

Yin, S., Ho, C.K., Shuman, S., 2003. Structure-function analysis of T4 RNA ligase 2. J. Biol. Chem. 278 (20), 17601–17608.

Yu, K.-Y., Kwon, H.-J., Norman, D.A.M., Vig, E., Goebl, M.G., Harrington, M.A., 2002. Cutting edge: mouse pellino-2 modulates IL-1 and lipopolysaccharide signaling. J. Immunol. 169 (8), 4075–4078.

Zhang, Y., Calin-Jageman, I., Gurnon, J.R., Choi, T.J., Adams, B., Nicholson, A.W., Van Etten, J.L., 2003. Characterization of a chlorella virus PBCV-1 encoded ribonuclease III. Virology 317 (1), 73–83.