

The number of genes encoding repeat domain-containing proteins positively correlates with genome size in amoebal giant viruses

Avi Shukla, Anirvan Chatterjee, and Kiran Kondabagil^{*,†}

Department of Biosciences and Bioengineering, Indian Institute of Technology Bombay, Powai, Mumbai, Maharashtra 400076, India

*Corresponding author: E-mail: kirankondabagil@iitb.ac.in

†<http://orcid.org/0000-0002-7942-023X>

Abstract

Curiously, in viruses, the virion volume appears to be predominantly driven by genome length rather than the number of proteins it encodes or geometric constraints. With their large genome and giant particle size, amoebal viruses (AVs) are ideally suited to study the relationship between genome and virion size and explore the role of genome plasticity in their evolutionary success. Different genomic regions of AVs exhibit distinct genealogies. Although the vertically transferred core genes and their functions are universally conserved across the nucleocytoplasmic large DNA virus (NCLDV) families and are essential for their replication, the horizontally acquired genes are variable across families and are lineage-specific. When compared with other giant virus families, we observed a near-linear increase in the number of genes encoding repeat domain-containing proteins (RDCPs) with the increase in the genome size of AVs. From what is known about the functions of RDCPs in bacteria and eukaryotes and their prevalence in the AV genomes, we envisage important roles for RDCPs in the life cycle of AVs, their genome expansion, and plasticity. This observation also supports the evolution of AVs from a smaller viral ancestor by the acquisition of diverse gene families from the environment including RDCPs that might have helped in host adaptation.

Key words: giant virus; repeat domain-containing proteins; genome expansion; genome plasticity.

1. Introduction

Allometry, the study of the relationship between biological size and function, is considered as an important readout of evolutionary processes (Klingenberg, 2016). In the case of viruses, an allometric exponent of 1.5 between the length of the viral genome and the volume of the virion particle suggests a significant positive correlation between virion and genome size (Cui, Schlub and Holmes 2014). An increase in the virion volume was strongly attributed to an increase in the genome length rather than protein content and capsid morphology (Cui, Schlub and Holmes 2014). Consistent with this observation, genomes of

giant viruses that infect amoeba [amoebal viruses (AVs)] are large, despite being intracellular parasites (Koonin and Wolf 2010; Colson and Raoult 2012; Yutin, Wolf and Koonin 2014). If the amount of DNA is assumed to be a predominant factor in the virion volume (Cui, Schlub and Holmes 2014), amoeba-infecting megaviruses emerge as the bellwethers of large genomes driving the size of the virion. Interestingly, amoeba-resistant bacteria (ARBs) adapted to intra-amoeba lifestyle such as *Legionella pneumophila* and *Rickettsia bellii* also harbor unusually large genomes (Moliner, Fournier, and Raoult 2010). This seemingly contradicts the evolution of intracellular organisms

from their free-living ancestors by genome reduction (Andersson and Kurland 1998; Sakharkar, Kumar, and Chow 2004; Merhej et al. 2009; Darmon and Leach 2014; McNally et al. 2016). In ARBs, genome expansion has been linked to the horizontal acquisition of mobile elements and genes encoding repeat domain-containing proteins (RDCPs) with functions analogous to the immune system and anti-host secretory system (Moliner, Fournier, and Raoult 2010). The genomes of AVs also harbor genes encoding RDCPs such as ankyrin, FNIP, and WD40 repeat domain-containing proteins (Suhre 2005). Both ARBs and AVs are internalized via phagocytosis, resist digestion, and exhibit many similar genomic features (Moliner, Fournier, and Raoult. 2010). Unlike other intracellular pathogens that are known to undergo genome reduction, ARBs and AVs maintain large genomes and acquire genes via horizontal gene transfer (HGT) (Boyer et al. 2009; Colson and Raoult 2010). In a complex evolutionary path, AVs and ARBs emerge as competitors (Slimani et al. 2013) for an amoebal host that also facilitates the horizontal transfer of genes. The cytoplasmic life cycle within amoeba emerges as a key evolutionary force driving the genomic content of both ARBs and AVs. The shared ‘mobilome’ among ARBs and AVs enable both to succeed in subverting the host predation/immune system. Here, we have identified an association between lineage-specific genome size expansion and acquisition and duplication of repeat domain proteins/multigene family in AVs.

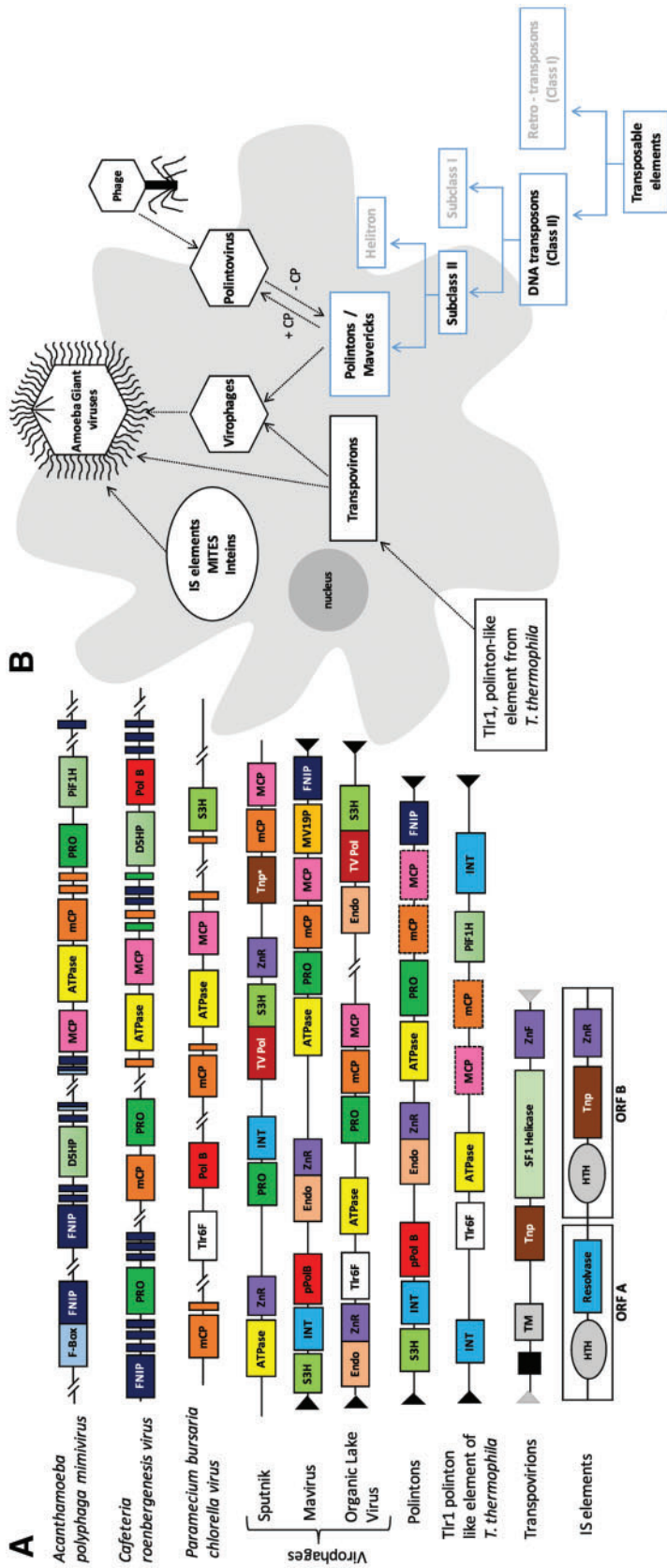
1.1 Classes of RDCPs in AVs and their functions in cellular homologs

Amoebal giant viruses are replete with proteins containing repeating amino acid sequences and are classified as RDCPs. These include ankyrin (ANK) repeat (Boyer et al. 2011; Herbert, Squire and Mercer 2015), Kelch repeat (Suhre 2005), leucine-rich repeat (LRR) (Suhre 2005), Tetratricopeptide (TPR) repeat (Sobhy et al. 2015), membrane occupation, and recognition nexus (MORN) repeat (Boyer et al., 2009), phenylalanine-asparagine-isoleucine-proline (FNIP/IP22) repeat (Suhre 2005), tryptophan-aspartic acid (WD40) repeat (Suhre 2005), and Sel 1 repeat. Proteins containing these repeat motifs regulate various intracellular processes through protein-protein interactions (Kobe and Deisenhofer, 1994; Sedgwick et al. 1999; Adams, Kelso, and Cooley 2000; Voronin and Kiseleva 2007; Catalano et al. 2010; Zeytuni and Zarivach, 2012). In plants, genes encoding RDCPs and their duplication have been associated with adaptation to rapid environmental variations (Richard, Kerrest, and Dujon 2008; Sharma and Pandey 2015). These proteins are thought to be the result of intragenic tandem duplication via recombination and are more commonly found in eukaryotes and metazoans, than prokaryotes (Marcotte et al. 1999; Andrade, Perez-Iratxeta, and Ponting 2001). AVs encode many of these RDCPs that are either integrated into the functional genes or present as stand-alone repeats. Motif length and

Box 1. HGT and the mobilome of AVs

Polintons (also known as mavericks) are the large DNA transposons (9–22 kb long) that are widely distributed in eukaryotes (Kapitonov and Jurka 2006; Fischer and Suttle 2011; Krupovic and Koonin, 2015). Recently, it was shown that virophages (parasitic viruses of large DNA viruses) and polintons, in addition to encoding several key homologous proteins including major and minor capsid proteins, FtsK-type packaging ATPase, protein-primed DNA polymerase B, retroviral-like family integrase and cysteine protease, exhibit similar genomic architecture (Fig. A). These observations imply that Polintons and virophages are evolutionarily linked (Filee, Pouget and Chandler 2008). Although Polintons encode two capsid proteins, their ability to form virion has not been demonstrated. Although an earlier study suggested the evolution of polintons from a virus (Benson et al. 1999), more recently, Polintons were hypothesized to have evolved from bacteriophages to become the first eukaryotic DNA viruses from which most of the extant NCLDVs have evolved (Krupovic and Koonin, 2015). Mavirus, a virophage of the *Cafeteria roenbergensis* virus (CroV) that infects the marine flagellate *C. roenbergensis*, possesses terminal inverted repeats that are characteristic of Polintons and other transposons (Filee, Pouget and Chandler 2008) and can integrate at multiple sites within the host (*C. roenbergensis*) genome and get reactivated in a CroV-infection dependent manner (Fischer and Hackl 2016). Furthermore, Polintons are thought to be one of the major components of the complex genetic network that include NCLDVs, adenoviruses, virophages, bacteriophages, naked DNA elements (Koonin, Krupovic, and Yutin 2015; Krupovic and Koonin, 2015). Similar to Class 2 DNA transposons, polintons transfer genetic material by a replicative or a cut-paste mechanism (Wicker et al. 2007) (Fig. B) and augment the number of shared genes across the network in the mobilome (Desnues et al. 2012; Colson et al. 2017). Another key member of this mobilome is the transpovirons found in Mimiviridae (Desnues et al. 2012; Yutin, Raoult, and Koonin 2013; Yutin et al. 2013). ORFs found in transpovirons have diverse evolutionary histories (Desnues et al. 2012) with origins in bacteria and their phages, and eukaryotes such as *Tetrahymena thermophila* (Yutin, Raoult, and Koonin 2013; Yutin et al. 2013). With the ability to integrate non-specifically into any part of the host (Mimiviridae family) chromosome (Desnues et al. 2012), transpovirons, along with virophages and polintons, are speculated to drive gene transfer within the mobilome (Boyer et al. 2011). Consequently, homologs of several hallmark genes of AVs have been found to be present in the polintons, virophages, and transpovirons (Fig. A), along with genetic elements (integras and terminal repeats) reminiscent of TEs (Fig. A). Thus, polintons and transpovirons frequently introduce genetic material from other branches of life (bacteria and eukarya) into the mobilome which is then transferred to AVs by virophages (Fig. B).

Insertion sequences, a major component of HGT are also commonly found in giant viruses, specifically in Mimiviridae and Phycodnaviridae with two overlapping ORFs (Filee, Siguier, and Chandler 2007). Interestingly, identical elements are also found to be part of *A. Castellanii* genome suggesting a route for gene transfer either from prokaryotes via giant viruses or from proto-eukaryotic ancestors (Gilbert and Cordaux 2013). These elements can manipulate the downstream gene expression (Siguier, Gourbeyre, and Chandler 2014) and play a major role in gene inactivation, deletion, duplication and genetic rearrangement in the genome via homologous/illegitimate recombination (Filee, Siguier, and Chandler 2007). In an extreme case, about 30 non-autonomous transposable elements commonly known as MITEs (10 are integrated into the coding regions) have ‘colonized’ (Sun et al. 2015) the genome of Pandoravirus salinus, but were undetectable in Pandoravirus dulcis (Sun et al. 2015). Akin to their role in prokaryotes, they promote gene deletion and genetic rearrangement (Feschotte, Zhang, and Wessler, 2002). A conceivable outcome of such genome plasticity would be the loss and/or gain of function, accelerating host-switching and adaptation. Apart from these family-specific mobile elements, the genomes of NCLDV also contain self-splicing introns (Azza et al. 2009) and inteins along with HNH endonuclease which might aid in the mobility of genetic elements (Filee and Chandler 2010). All three are known to influence genome evolution in all forms of life through their splicing and nuclease activity (Darmon and Leach 2014).



Box Figure A. Organization of various genes and their homologs in AVs, virophages, Polintons, transposons, and IS elements constituting the predicated mobilome network (Desnues et al. 2012; Yutin, Raouf, and Koonin 2013; Yutin et al. 2013).

Despite limited synteny, the mobilome exhibits genetic and functional conservation. AVs, virophages, and polintons encode four core genes, viz. packaging ATPase, major and minor capsid protein, and cysteine protease. The presence of different types of helicases across the mobilome illustrates functional conservation. All the members of the mobilome have one or more genes encoding transposase, integrase, and endonuclease which facilitate genetic exchange. Although the inverted repeats are encoded in all the genomes, they have not been reported in the terminal regions of CroV and Mamavirus.

Box Figure B. Probable evolutionary routes for the exchange of mobile domains of mobilome are from non-viral system suggesting, their acquisition from different microbial sources sharing the same niche. (Dotted line shows the probable transmission while solid blue lines shows classification) +CP, with capsid protein; -CP, without capsid protein.

Table 1. The basic composition, structure, and functions of different repeat domain proteins in diverse forms of life excluding Megavirales

Multigene repeat families	Composition	Structural unit	Tertiary structure	Participates in	Commonly found in	References
Ankyrin repeats (ANK)	33 aa	Two antiparallel α -helices joined by β -hairpin at 90° forming L-shaped structure	Cupped hand shape solvent accessible groove formed by repeating protomers	Cell cycle regulation, cytoskeletal binding, protein trafficking across membrane, acquired resistance.	Prokaryotes and eukaryotes	Nguyen, Liu and Thomas 2014, Alkhdor et al. 2009, Voronin and Kiseleva 2007, Sedgwick and Smerdon 1999, Cao et al. 1997, Shchelkunov, Blinov and Sandakhchiev 1993
Leucine rich repeats (LRR)	20 to 29 aa	A β -sheet and an α -helix arranged in an anti-parallel manner	Multiple repeats are oriented parallel to the axis forming horse-shoe like structure	Protein -protein interaction, signal transduction and formation of protein complexes	Prokaryotes and eukaryotes	Kobe and Deisenhofer 1994, Sharma and Pandey 2015
FNIP/IP22 repeats	22 aa	A β -sheet and an α -helix arranged in an anti-parallel manner	Horse shoe like structure (like LRR)	Interaction of calmodulin binding proteins, increases cell motility and chemotaxis	Dictyostelium and NCLDV	Catalano et al. 2010, O'Day et al. 2006
Tetratricopeptide repeats (TPR)	34 aa	Multiple array of α -helix turn α -helix unit packaged in parallel	A right-handed super-helix that provide concave groove for molecule binding	Cell cycle regulation, chaperone functioning, protein translocation, bacterial pathogenesis, and biogenesis of multi-functional pilli	Prokaryotes and eukaryotes including humans	Cervený et al. 2013, Zeytuni and Zarivach 2012
Sel1 repeats	33 to 44 aa	Multiple array of α -helix turn α -helix unit packaged in parallel	A right-handed super-helix	ER-associated protein ubiquitination, regulation of mitosis and septum formation, host-pathogen interaction	Bacteria and eukaryotes	Newton et al. 2007, Mittl and Schneider-Brachert 2007
WD 40 repeats	40 aa	Four anti-parallel β -sheet arranged radially with flanking dipeptide	Propeller like structure	Gene regulation, chromatin modelling, transmembrane signalling, mRNA modification, vesicle fusion and adhesion complex of malarial parasites	Eukaryotes	Suganuma, Pattenden, and Workman 2016, von Bohl et al. 2015, Neer et al. 1994
Kelch repeats	44 to 56 aa	Four anti-parallel β -sheet arranged radially with flanking dipeptide	Propeller like structure	Actin binding, manipulates cell organization and morphology	Prokaryotes, eukaryotes and viruses	Prag and Adams 2003, Adams, Kelso and Cooley 2000
MORN repeats	23 aa	Not known	Not known	Parasites' budding, protein translocation, flagellum biogenesis, form junctional complex between plasma membrane to endoplasmic reticulum, promotes phagocytosis of bacterium	Prokaryotes and eukaryotes	Morriswood and Schmidt 2015, Abnave et al. 2014, Cuttel et al. 2008, Gubbels et al. 2006, Hui Ma et al. 2006, Takeshima et al. 2000

structure of RDCPs found in AVs and their known functions in prokaryotes and eukaryotes are summarized in Table 1.

1.2 Effect of genes encoding RDCPs on AV genome size

We compared the frequency of occurrence of genes encoding RDCPs and core viral functions and their association with

genome size (Fig. 1A and B) as well as their genomic location in the representative genome from the thirteen giant virus families (Fig. 1C). A near-linear relationship was observed between the number of genes encoding RDCPs and the genome size of most large viruses (Fig. 1A). The trend is most evident in AVs, where the number of genes encoding RDCPs correlated with an increase in the genome size ($r^2 = 0.87$). No such correlation was

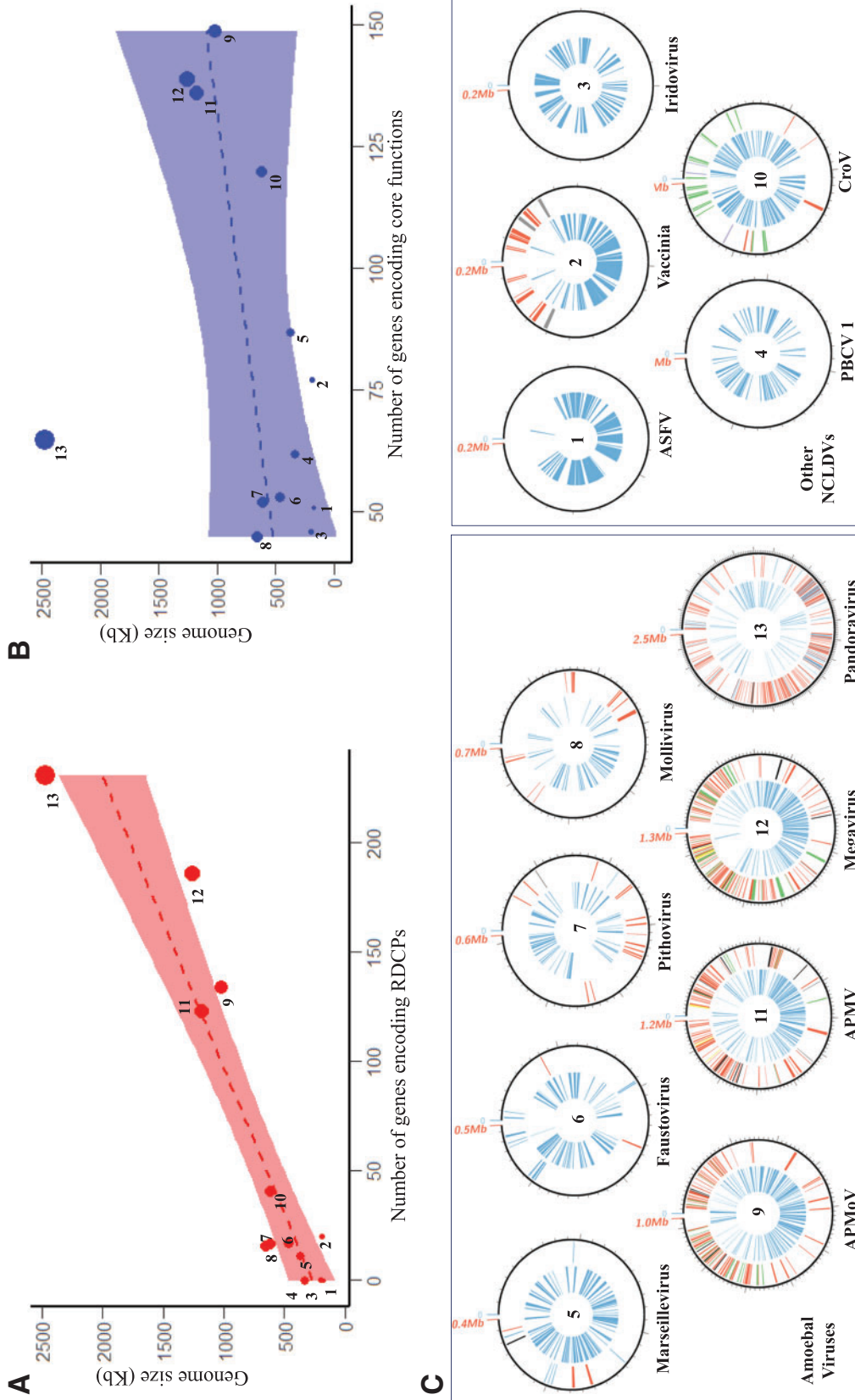


Figure 1. A near-linear relationship between the genome size and the number of genes encoding RDCPs in AVs. Core genes and RDCPs were manually curated from 13 published genomes. Core function definitions were chosen as per the previous reports (Raoult et al. 2004; Yutin et al. 2009; Yutin, Raoult, and Koonin 2013; Yutin et al. 2013). These included genes encoding DNA replication, recombination and repair, transcription and RNA processing, translation, and post-translation modifications, nucleotide metabolism, virion packaging, and morphogenesis. Genes encoding these functions in the 13 representative NCLDV families were retrieved as per annotations in the public databases. Genes for which annotations were not updated, but yielded significant alignment matches in Interpro, CDD, Pfam and Smart servers, were also included. (A) Scatterplot of the number of repeat protein families plotted against genome size. A high correlation between number of repeat protein families and genomes size ($r^2 = 0.87$) was observed. (B) Scatterplot of the number of genes encoding core viral functions plotted against genome size, which shows a poor association between the two ($r^2 = 0.11$). In (A) and (B), the shaded area indicates the standard error as per a linear regression model. The size of the data label (solid dot) representing genomes is proportional to the genome size. Number alongside the data label corresponds to the number inside the ideograms shown in Figure 1C. (C) Circos-generated ideograms of giant viral genomes. Outer concentric represents the clusters of repeat domain proteins/multigene families that include proteins containing ANK repeats = red, FNIP repeats = green, MORN repeats = yellow, TPR = purple, WD40 repeats = black, LRR and kelch repeat = gray, and the inner concentric denotes the core genome. APMV, *Acanthamoeba polyphaga mimivirus*; APMoV, *Acanthamoeba polyphaga momovirus*; PBCV 1, *Paramecium bursaria chlorella virus 1*; ASFV, *African swine fever virus*.

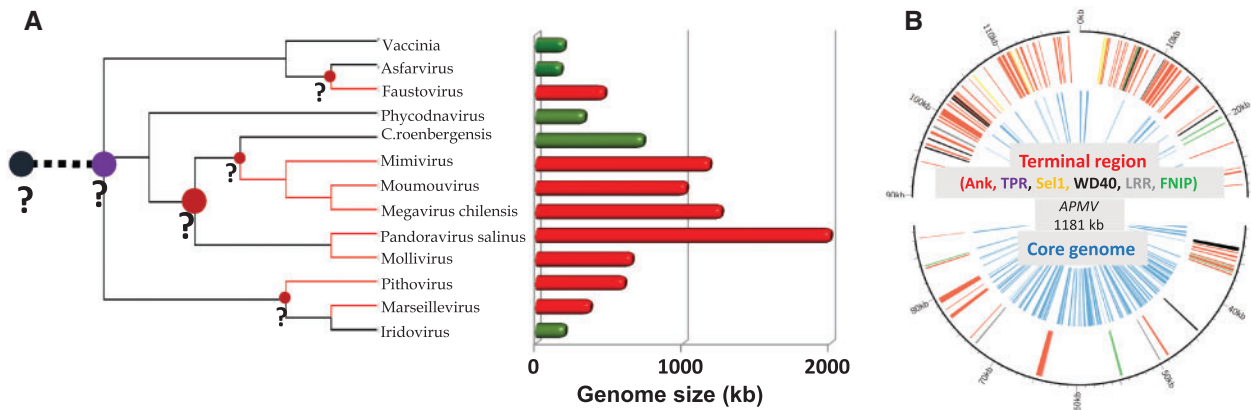


Figure 2. A speculative hypothesis on the RDCP driven lineage-specific genome expansion in AVs. (A) Genome size distribution and B family DNA polymerase phylogeny. ML Tree of B family DNA polymerase amino acid sequence was constructed using FastTree with default settings using a representative sequence from 13 NCLDV families. A large red circle on the internal node of the AV lineage indicates a more recent ancestor from which we believe genome expansion has ensued, especially in the amoebal milieu. Smaller red circles indicate a much recent ancestor from which independent genome expansion strategy might have led to larger genomes in Faustoviruses and Pithovirus. Black and purple circles indicate ancestors of unknown genome size and nature. More genome sequences are needed to resolve the genome size distribution pattern and its evolutionary link to the nature of the ancestor in large DNA viruses. (B) Circos ideogram of Mimivirus genome. Three concentrics, labeled as 1, 2, and 3 represent RDCPs, core and hypothetical genes, and mobile elements, respectively. The bipartite AV genome consists of a conserved core region derived from a common ancestor, and the RDCPs that are clustered in the genomic termini of the AVs. In addition to aiding in genome expansion, the RDCPs may also help in survival in the competitive environment (see Fig. 3 for details). In an allopatric condition, most of these RDCPs are lost causing a reduced genome size (Boyer et al., 2011; Colson and Raoult 2012).

observed between the genome size and the number of genes encoding core viral functions ($r^2 = 0.11$) (Fig. 1B). The correlation was less evident in other giant viruses, viz. Asfarviridae, Poxviridae, Iridoviridae, and Phycodnaviridae, which are not known to infect amoeba, suggesting genome expansion via the acquisition of RDCPs is specific to AVs (Moliner, Fournier and Raoult 2010). Interestingly, genes encoding RDCPs are concentrated towards the termini on either side of the core genes (Fig. 1C). This arrangement is most apparent in Mimiviridae family members. AVs with significantly smaller genomes have fewer RDCPs, spread across the genome (Mollivirus and Faustovirus), and in AVs with larger genomes (Pandoravirus), RDCPs appear to have spread throughout the genome. Proteins with repeat domain play important roles in protein-protein interactions (Table 1; Brüggemann, Cazalet, and Buchrieser, 2006). Interestingly, when Mimivirus was propagated repeatedly under a competition-free axenic environment, genes present in the termini region were lost (Boyer et al. 2011; Colson and Raoult 2012). The lost patches include the genes encoding proteins participating in the fiber formation and its glycosylation, and ANK repeat proteins (Boyer et al. 2011). But, in a competitive environment, the presence of fibers increases the virion size and may facilitate efficient phagocytosis. And in addition, the genomic-termini regions populated with RDCPs might aid survival in a sympatric environment but are under low selection pressure in an allopatric environment and can afford deletions (Boyer et al. 2011). This ensures the protection of the centrally located core genome that is also thought to be less recombinogenic than the termini (Filee, Signier, and Chandler 2007; Boyer et al. 2011). We suggest that in a competitive environment, accumulation of RDCPs in the termini provide a selective advantage over other viruses and bacteria.

1.3 HGT of RDCPs in AVs

AVs might have acquired genes encoding RDCPs from amoeba and bacteria by various HGT mechanisms resulting in genome expansion. Virophages, polintoviruses, and transpovirons, associated with AVs, facilitate HGT between AVs and their host environment (Desnues et al. 2012; Yutin, Raoult, and Koonin 2013; Yutin et al.

2013). Akin to mobile genetic elements (MGEs), these three drive HGT in AVs and have contributed to a shared gene pool, consisting of a variety of genes encoding essential functions and transposition, giving rise to the mobilome of AVs (Yutin, Raoult, and Koonin 2013; Yutin et al. 2013) as discussed in Box 1. The presence of multiple MGEs in nucleocytoplasmic large DNA virus (NCLDV) genomes along with proteins known for DNA transport connote the makings of a genome populated with agents for large-scale genomic insertion, deletion, and rearrangements. The presence of self-splicing intronic regions in several genes including capsid proteins and DNA polymerase in some AVs, not reported in other viruses, also suggests their acquisition from eukaryotic genomes (Arslan et al. 2011). This is analogous to HGT in bacteria and eukaryotes that facilitate the development of drug resistance (Novais et al. 2010), defense systems (Makarova et al. 2011; Krupovic et al. 2014), regulatory roles in transcriptional and signaling mechanisms (Negi, Rai, and Suprasanna 2016), and immunological variation (Huang et al., 2016). The genome expansion ensuing from this plasticity could be crucial for enabling the evolutionary success of AVs as seen in ARBs with similar MGE architecture.

1.4 RDCPs and the lineage-specific genome expansion in AVs

Initial studies indicated the apparent monophyly of AVs (Yutin and Koonin 2012; Zade, Sengupta, and Kondabagil 2015) and recent comparative genomics of diverse AVs have provided more robust phylogenies suggesting a probable lineage-specific expansion in AVs (Iyer et al. 2006; Filée 2009). Tracing the genome size over a phylogeny based on the B family DNA polymerase amino acid sequence, conserved across NCLDVs suggests the presence of larger genomes in the AV lineages (Fig. 2A). This expansion in AV lineages could be primarily attributed to the acquisition of RDCPs which shows a positive correlation with genome size (Fig. 1A). However, in the case of Pandoravirus, the genome expansion may be free from geometrical constraints, unlike other AVs, such as Mimivirus and Faustovirus where the viral morphology may limit genome size expansion.

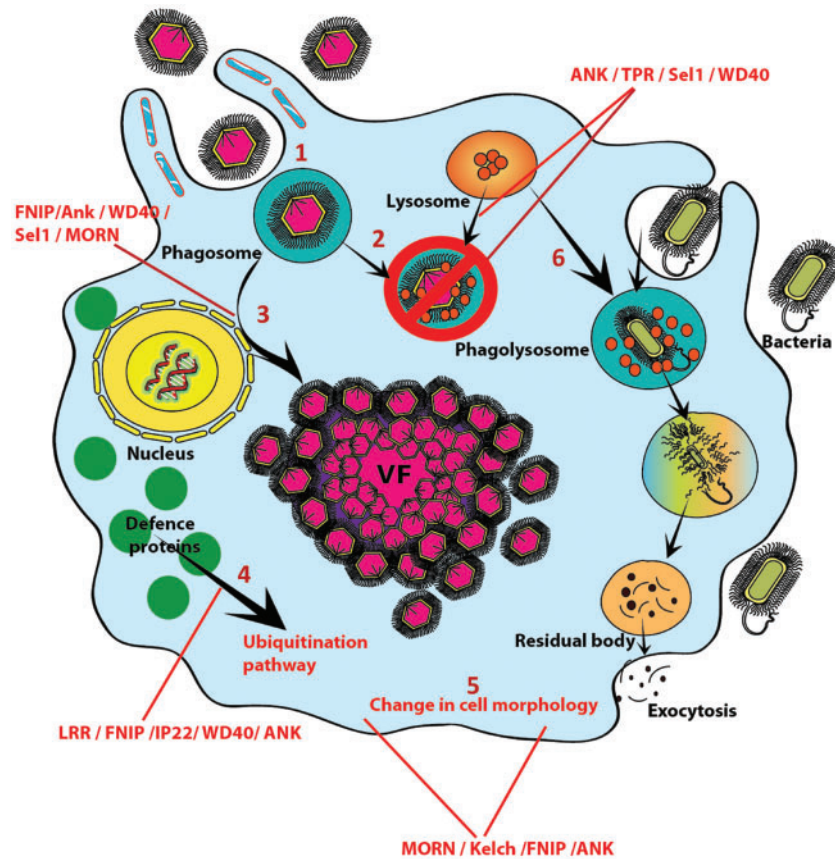


Figure 3. Putative roles of various RDCPs in the AV infection cycle. Giant capsid mimics the size of bacteria for promoting phagocytosis in a sympatric environment prohibiting the host encystment. Once inside, it suppresses the host immune system by interfering with various host defense mechanisms via repeat domain-containing protein (that also mimic some of the host proteins) or/and deviating them to ubiquitination. The distinct phases of the intra-amoebal life cycle of a virus involve: (1) Particle size plays an important role in the mode entry on viruses. As seen in other viruses (Cui et al. 2014), the large particle size may be driven by genome expansion, caused by accumulation of RDCPs. (2) Once phagocytosed, the encystment of the trophozoite is arrested and the fusion of the phagosome to the lysosome is inhibited by ankyrin, TPR, WD40, and Sel1 repeat domains proteins, as has been reported in intra-amoebal parasitic bacteria (Shchelkunov, Blinov, and Sandakhchiev 1993; Newton et al. 2007; Cerveny et al. 2013; Nguyen, Liu, and Thomas 2014). Some of the RDCPs have been reported to be packaged in the virion indicating their role in the initiation of the viral replication cycle (Renesto et al. 2006). (3) The viral genome is released into the cytoplasm from the phagosome and the formation of a replication center is initiated by the recruitment of various cytoplasmic membranes, mitochondria, and cytoskeletal components. This formation requires a number of complex interactions and signaling pathways, that are probably mediated by FNIP, ANK repeats, Sel1, WD40, or/and MORN repeats domain proteins. (4) During infection, RDCPs such as LRR, FNIP, IP22, WD40, ANK repeats, and F-box proteins might interfere with host defense mechanisms. They have been shown to modify/regulate the host gene expression and subvert the host proteins to ubiquitination or mimics some of the inhibitory molecules to suppress the immune pathways (Sharma and Pandey 2015). (5) During the infection cycle, host cell morphology changes to avoid superinfection. This morphological change is brought about by MORN, Kelch, FNIP and ANK repeat domain proteins (Table 1). In addition, MORN repeat containing protein might also promote the degradation of other internalized microorganisms. (6) Unlike AVs, bacteria are unable to interfere with the formation of the phagolysosome, and are consequently digested by the hydrolytic enzymes in the lysosome (Cosson and Soldati 2008; Aky, Pointon and Thomas 2009). Although phagocytosis of AVs and bacteria is primarily driven by particle size, they have distinct fates. The RDCPs emerge as crucial drivers of both, the particle size and a successful viral life cycle.

As seen in plants and some pathogenic bacteria, RDCPs are characterized by frequent duplications and deletions (Siozios et al. 2013; Sharma and Pandey 2015) which confer plasticity to their genomes. Genome plasticity imparted by genes encoding RDCPs in AVs could be a major contributor to their 'accordion'-like evolution (Filee 2013; Filee 2015). An accretion scenario considers a smaller virus as an ancestor of giant viruses (Yutin, Wolf, and Koonin 2014; Koonin, Krupovic, and Yutin 2015) that got bigger in some lineages by gene acquisition leading to both genome and particle size expansion (Rodrigues et al. 2016). On the other hand, a genome reduction scenario considers evolution from an ancestor with a larger genome (Claverie and Abergel 2013; Filee 2013). Although the presence of HGT-derived genes and MGEs (Filee, Siguier, and Chandler 2007) has been used as evidence for the former argument, the

presence of some translation-related genes and lack of cellular homologs of giant viral genes (Jeudy et al. 2012; Abrahão et al. 2017) have been used to support the latter. Genes related to key processes such as transcription, nucleotide metabolism, translation, virion assembly, and DNA packaging, are part of the Nucleo-Cytoplasmic Virus Orthologous Genes (NCVOGs) (Yutin et al. 2009) and are believed to be vertically transferred from a common ancestor (Raoult et al. 2004; Iyer et al. 2005, 2006; Chelikani et al. 2014; Zade, Sengupta, and Kondabagil 2015). Isolation of several novel NCLDV and their genomic characterization has reduced the number of conserved genes to nine (Iyer, Aravind, and Koonin 2001; Yutin et al. 2009), with a conceivable diversity in other core genes arising from replacement of essential genes by unrelated ones with similar function (Forterre 2006; Iyer et al., 2006; Filee, Pouget and Chandler 2008).

Further, it was also suggested that the common ancestor encoded several genes in addition to the basal machinery, indicating that the NCLDV ancestor was relatively complex (Yutin et al. 2009; Koonin and Yutin 2010). A majority of the other (non-core) NCVOGs are coded by two or more of the NCLDV family members. The core genomic landscape of the vertically transferred genes with lineage-specific diversification is reminiscent of *gene reservoirs* of pathogenic bacteria, which facilitate rapid adaptation to host (Hannan 2012; Andam and Hanage 2015; McNally et al. 2016). Based on the location of the genes encoding RDCPs and genes with known viral functions, the AV genomes could be thought of as bipartite, the central core genome flanked by the genomic termini (Fig. 2B). The core genes that are under high selection pressure predominates the central part. The peripheral segments on either side harbor genes encoding RDCPs, which confer plasticity and are under relatively less selection pressure. This bipartite genome may undergo lineage-specific expansion, primarily through accumulation and duplication of genes encoding RDCPs, resulting in a large genome size. This is consistent with the view that the members of *Mimiviridae* might have undergone genomic expansion from a common ancestor, as against a probable genome reduction scenario in some members of *Phycodnaviridae* family, which infect algae (Maruyama and Ueki 2016). Although the list of sequenced large DNA viral genomes from wider geographies is growing (Hingamp et al. 2013; Aherfi et al. 2016; Chatterjee et al. 2016a,b), isolation and sequencing of more large DNA viruses enable the description of phylogenetic intermediates that are critical for a parsimonious explanation of particle and genome size evolution. Despite missing the probable clade and lineage-specific ancestors, we observed genomic arrangement patterns in AVs which may enable their intra-amoebal lifestyle (Figs. 2A and 3).

1.5 Competitive advantage of large particle size driven by gene accretion including RDCPs

The capsid that harbors the giant genome plays a major role in the entry of these viruses into their respective hosts (Rodrigues et al. 2016). The mode of entry of metazoan and algal viruses differ from the AVs. Asfarvirus, Iridovirus, and Poxvirus enter the multicellular host by an actin-dependent macropinocytosis or a receptor-mediated endocytosis (Rodrigues et al. 2016). Poxviruses also enter the host by their membrane fusion to the plasma membrane (Moss 2012; Rizopoulos et al. 2015). Phycodnavirus generally enter their algal host by degradation of the host cell membrane (Wilson, Van Etten, and Allen 2009). Giant viruses such as Mimivirus, Pandoravirus, Pithovirus, and Mollivirus undergo phagocytosis (Fig. 3) (Rodrigues et al. 2016), which is predominantly a function of the size of the particle; the threshold size for entry is ~500 nm (Korn and Weisman 1967). The importance of particle size in the mode of entry is further exemplified in the case of Marseillevirus, which is phagocytosed when present as a 'parcel' (many particles) in a vesicle (>1 μm). However, when present as a solitary particle of ~220 nm, Marseillevirus undergoes endocytosis or macropinocytosis (Arantes et al. 2016). Amoeba generally grazes on particles of the general size of a bacterium (Korn and Weisman 1967) and digest it via phagolysosome pathway (Fig. 3; Khan 2001; Akya, Pointon and Thomas 2009; Raoult and Boyer, 2010)]. Thus, the giant size, largely driven by the acquisition and duplication of RDCPs, is critical for infecting amoeba via phagocytosis (Rodrigues et al., 2016). Once phagocytized, giant viruses must subvert encystment and hijack the host cellular machinery to

initiate the formation of the viral replication center (Fig. 3, see figure legend for details). The rapidity of the hijack necessitates a multipronged approach of naturalization into the host via gene products adapted to the host pathway and infectiousness which directs cellular process towards the synthesis of viral proteins. Many of these are mediated by RDCPs. Some of these, such as WD40 and ANK repeat containing proteins are packaged in the Mimivirus particle indicating their imminent role in initiating the viral replication cycle (Renesto et al. 2006).

2. Conclusion: repeat domain proteins are essential for intra-amoebal adaptation

Acquired vertically or horizontally, the genomic composition of AVs exhibit an exceptional variability. Genomes of AVs could be thought of as bipartite, with genes encoding core functions populating the center and genes encoding RDCPs frequenting the termini. Unsurprisingly, RDCPs, which are considered to be the hotspots of protein evolution (Persi et al. 2016), emerge as one of the key genetic elements responsible for the lineage-specific genome expansion of AVs. With most genes in AVs found to be under purifying selection (Doutre et al. 2014), RDCPs are also expected to contribute to virus fitness. However, as in Ohno's dilemma (Bergthorsson, Andersson, and Roth. 2007), strong purifying selection on RDCPs would reduce diversity. Consequently, as seen in repeat domain proteins across cellular organisms, RDCPs of AVs might undergo cycles of relaxed and strong purifying selection (Persi et al. 2016) to provide increased fitness in a competitive host environment, such as amoeba. This is expected to lead to the evolution of new functions and/or establishment of existing functions. We suggest that the acquisition of RDCPs in AVs facilitated both genome expansion and host adaptation. The later probably led to an allometric increase in the particle size. Finally, similar to a 'telomeric strategy', these elements are concentrated towards the termini protecting the core genes. This genomic arrangement of RDCPs in the termini may be crucial for AVs to adapt to a wide variety of hosts and outcompete prokaryotes and other viruses in the prokaryote-grazing protozoan milieu.

Data availability

Data are available through Dryad.

Conflict of interest: None declared.

Acknowledgements

Research in KK lab was supported by grants from DST (SR/SO/BB-0031/2012) and DBT (BT/PR4808/BRB/10/1029/2012). A.S. was supported by Senior Research Fellowship by CSIR and A.C. was supported by IIT Bombay post-doctoral fellowship.

References

- Abnave, P. et al. (2014) 'Screening in planarians identifies MORN2 as a key component in LC3-associated phagocytosis and resistance to bacterial infection', *Cell Host and Microbe*, 16: 338–50.
- Abrahão, J. S. et al. (2017) 'The Analysis of Translation-Related Gene Set Boosts Debates around Origin and Evolution of Mimiviruses', *PLOS Genetics*, 13: e1006532.

- Adams, J., Kelso, R., and Cooley, L. (2000) 'The Kelch Repeat Superfamily of Proteins: Propellers of Cell Function', *Trends in Cell Biology*, 10: 17–24.
- Aherfi, S. et al. (2016) 'Giant Viruses of Amoebas: An Update', *Frontiers in Microbiology*, 7: 349.
- Akya, A., Pointon, A., and Thomas, C. (2009) 'Mechanism Involved in Phagocytosis and Killing of *Listeria Monocytogenes* by *Acanthamoeba Polyphaga*', *Parasitology Research*, 105: 1375–83.
- Al-Khodor, S. et al. (2010) 'Functional diversity of ankyrin repeats in microbial proteins', *Trends in Microbiology*, 18:132–9.
- Andam, C. P., and Hanage, W. P. (2015) 'Mechanisms of Genome Evolution of *Streptococcus*', *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, 33: 334–42.
- Andersson, S.G. e., and Kurland, C. G. (1998) 'Reductive Evolution of Resident Genomes', *Trends in Microbiology*, 6: 263–8.
- Andrade, M. A., Perez-Iratxeta, C., and Ponting, C. P. (2001) 'Protein Repeats: Structures, Functions, and Evolution', *Journal of Structural Biology*, 134: 117–31.
- Arantes, T. S. et al. (2016) 'The Large Marseillevirus Explores Different Entry Pathways by Forming Giant Infectious Vesicles', *Journal of Virology*, 90: 5246–55.
- Arslan, D. et al. (2011) 'Distant Mimivirus Relative with a Larger Genome Highlights the Fundamental Features of Megaviridae', *Proceedings of the National Academy of Sciences of the United States of America*, 108: 17486–91.
- Azza, S. et al. (2009) 'Revised Mimivirus Major Capsid Protein Sequence Reveals Intron-Containing Gene Structure and Extra Domain', *BMC Molecular Biology*, 10: 39.
- Benson, S. D. et al. (1999) 'Viral Evolution Revealed by Bacteriophage PRD1 and Human Adenovirus Coat Protein Structures', *Cell*, 98: 825–33.
- Berghthorsson, U., Andersson, D. I., and Roth, J. R. (2007) 'Ohno's Dilemma: evolution of New Genes under Continuous Selection', *Proceedings of the National Academy of Sciences of the United States of America*, 104: 17004–9.
- Boyer, M. et al. (2009) 'Giant Marseillevirus Highlights the Role of Amoebae as a Melting Pot in Emergence of Chimeric Microorganisms', *Proceedings of the National Academy of Sciences of the United States of America*, 106: 21848–53.
- et al. (2011) 'Mimivirus Shows Dramatic Genome Reduction after Intraamoebal Culture', *Proceedings of the National Academy of Sciences of the United States of America*, 108: 10296–301.
- Brüggemann, H., Cazalet, C., and Buchrieser, C. (2006) 'Adaptation of *Legionella Pneumophila* to the Host Environment: role of Protein Secretion, Effectors and Eukaryotic-like Proteins', *Current Opinion in Microbiology*, 9: 86–94.
- Cao, H. et al. (1997) 'The Arabidopsis NPR1 gene that controls systemic acquired resistance encodes a novel protein containing ankyrin repeats', *Cell*, 88: 57–63.
- Catalano, A. et al. (2010) 'Synthesis and Biological Activity of Peptides Equivalent to the IP22 Repeat Motif Found in Proteins from *Dictyostelium* and Mimivirus', *Peptides*. Elsevier Inc, 31: 1799–805.
- Cervený, L. et al. (2013) 'Tetratricopeptide Repeat Motifs in the World of Bacterial Pathogens: Role in Virulence Mechanisms', *Infection and Immunity*, 81: 629–35.
- Chatterjee, A. et al. (2016a) 'Complete Genome Sequence of a New Megavirus Family Member Isolated from an Inland Water Lake for the First Time in India', *Genome Announcements*, 4: e00402-16.
- et al. (2016b) 'Isolation and Complete Genome Sequencing of Mimivirus Bombay, a Giant Virus in Sewage of Mumbai, India', *Genomics Data*, 9: 1–3.
- Chelikani, V. et al. (2014) 'Genome Segregation and Packaging Machinery in *Acanthamoeba Polyphaga* Mimivirus Is Reminiscent of Bacterial Apparatus', *Journal of Virology*, 88: 6069–75.
- Claverie, J.-M., and Abergel, C. (2013) 'Open Questions about Giant Viruses', *Advances in Virus Research*, 85: 25–56.
- Colson, P. et al. (2017) 'Mimivirus: leading the Way in the Discovery of Giant Viruses of Amoebae', *Nature Reviews. Microbiology*, 15: 243–54.
- and Raoult, D. (2010) 'Gene Repertoire of Amoeba-Associated Giant Viruses', *Intervirology*, 53: 330–43.
- and — (2012) 'Lamarckian Evolution of the Giant Mimivirus in Allopatric Laboratory Culture on Amoebae', *Frontiers in Cellular and Infection Microbiology*, 2: 91.
- Cosson, P., and Soldati, T. (2008) 'Eat, Kill or Die: when Amoeba Meets Bacteria', *Current Opinion in Microbiology*, 11: 271–6.
- Cui, J., Schlub, T. E., and Holmes, E. C. (2014) 'An Allometric Relationship between the Genome Length and Virion Volume of Viruses', *Journal of Virology*, 88: 6403–10.
- Cuttell, L. et al. (2008) 'Undertaker, a *Drosophila* Junctophilin, Links Draper-Mediated Phagocytosis and Calcium Homeostasis', *Cell*, 135: 524–34.
- Darmon, E., and Leach, D. R. F. (2014) 'Bacterial Genome Instability', *Microbiology and Molecular Biology Reviews*, 78: 1–39.
- Desnues, C. et al. (2012) 'Provirophages and Transpovirons as the Diverse Mobilome of Giant Viruses', *Proceedings of the National Academy of Sciences of the United States of America*, 109: 18078–83.
- Doutre, G. et al. (2014) 'Genome Analysis of the First Marseilleviridae Representative from Australia Indicates That Most of Its Genes Contribute to Virus Fitness', *Journal of Virology*, 88: 14340–9.
- Feschotte, C., Zhang, X., and Wessler, S. R. (2002) 'Miniature inverted-repeat transposable elements and their relationship to established DNA transposons', in Craig, N. L. (ed) *Mobile DNA II*, Chapter 50, pp. 1147–58. ISBN:9781555812096.
- Filee, J. (2015) 'Genomic Comparison of Closely Related Giant Viruses Supports an Accordion-like Model of Evolution', *Frontiers in Microbiology*, 6: 593.
- (2009) 'Lateral Gene Transfer, Lineage-Specific Gene Expansion and the Evolution of Nucleo Cytoplasmic Large DNA Viruses', *Journal of Invertebrate Pathology*, 101: 169–71.
- (2013) 'Route of NCLDV Evolution: The Genomic Accordion', *Current Opinion in Virology*, 3: 595–9.
- and Chandler, M. (2010) 'Gene Exchange and the Origin of Giant Viruses', *Intervirology*, 53: 354–61.
- , Pouget, N., and Chandler, M. (2008) 'Phylogenetic Evidence for Extensive Lateral Acquisition of Cellular Genes by Nucleocytoplasmic Large DNA Viruses', *BMC Evolutionary Biology*, 8: 320.
- , Siguier, P., and — (2007) 'I Am What I Eat and I Eat What I Am: Acquisition of Bacterial Genes by Giant Viruses', *Trends in Genetics*: Tig, 23: 10–5.
- Fischer, M. G., and Hackl, T. (2016) 'Host Genome Integration and Giant Virus-Induced Reactivation of the Virophage Mavirus', *Nature*, 540: 288–91.
- , and Suttle, C. A. (2011) 'A Virophage at the Origin of Large DNA Transposons', *Science (New York, N.Y.)*, 332: 231–4.
- Forsterre, P. (2006) 'The Origin of Viruses and Their Possible Roles in Major Evolutionary Transitions', *Virus Research*, 117: 5–16.

- Gilbert, C., and Cordaux, R. (2013) 'Horizontal Transfer and Evolution of Prokaryote Transposable Elements in Eukaryotes', *Genome Biology and Evolution*, 5: 822–32.
- Gubbels, M.-J. et al. (2006) 'A MORN-repeat protein is a dynamic component of the *Toxoplasma gondii* cell division apparatus', *Journal of Cell Science*, 119: 2236–45.
- Hannan, A. J. (2012) 'Tandem Repeat Polymorphisms: Mediators of Genetic Plasticity, Modulators of Biological Diversity and Dynamic Sources of Disease Susceptibility', *Advances in Experimental Medicine and Biology*, 769: 1–9.
- Herbert, M., Squire, C., and Mercer, A. (2015) 'Poxviral Ankyrin Proteins', *Viruses*, 7: 709–38.
- Hingamp, P. et al. (2013) 'Exploring Nucleo-Cytoplasmic Large DNA Viruses in Tara Oceans Microbial Metagenomes', *The ISME Journal*, 7: 1678–95.
- Huang, S. et al. (2016) 'Discovery of an Active RAG Transposon Illuminates the Origins of V(D)J Recombination', *Cell*, 166: 102–14.
- Iyer, L. M. et al. (2005) 'Origin and Evolution of the Archaeo-Eukaryotic Primase Superfamily and Related Palm-Domain Proteins: structural Insights and New Members', *Nucleic Acids Research*, 33: 3875–96.
- et al. (2006) 'Evolutionary Genomics of Nucleo-Cytoplasmic Large DNA Viruses', *Virus Research*, 117: 156–84.
- , Aravind, L., and Koonin, E. V. (2001) 'Common Origin of Four Diverse Families of Large Eukaryotic DNA Viruses', *Journal of Virology*, 75: 11720–34.
- Jeudy, S. et al. (2012) 'Translation in Giant Viruses: A Unique Mixture of Bacterial and Eukaryotic Termination Schemes', *PLoS Genetics*, 8: e1003122.
- Kapitonov, V. V., and Jurka, J. (2006) 'Self-Synthesizing DNA Transposons in Eukaryotes', *Proceedings of the National Academy of Sciences of the United States of America*, 103: 4540–5.
- Khan, N. A. (2001) 'Pathogenicity, Morphology, and Differentiation of *Acanthamoeba*', *Current Microbiology*, 43: 391–5.
- Klingenberg, C. P. (2016) 'Size, Shape, and Form: concepts of Allometry in Geometric Morphometrics', *Development Genes and Evolution*, 226: 113–37.
- Kobe, B., and Deisenhofer, J. (1994) 'The Leucine-Rich Repeat: A Versatile Binding Motif', *Trends in Biochemical Sciences*, 19: 415–21.
- Koonin, E. V., and Wolf, Y. I. (2010) 'Constraints and Plasticity in Genome and Molecular-Phenome Evolution', *Nature Reviews. Genetics*, 11: 487–98.
- , and Yutin, N. (2010) 'Origin and Evolution of Eukaryotic Large Nucleo-Cytoplasmic DNA Viruses', *Intervirology*, 53: 284–92.
- , Krupovic, M., and Yutin, N. (2015) 'Evolution of Double-Stranded DNA Viruses of Eukaryotes: From Bacteriophages to Transposons to Giant Viruses', *Annals of the New York Academy of Sciences*, 1341: 10–24.
- Korn, E. D., and Weisman, R. A. (1967) 'Phagocytosis of Latex Beads by *Acanthamoeba*: II. Electron Microscopic Study of the Initial Events', *The Journal of Cell Biology*, 34: 219–27.
- Krupovic, M. et al. (2014) 'Casposons: A New Superfamily of Self-Synthesizing DNA Transposons at the Origin of Prokaryotic CRISPR-Cas Immunity', *BMC Biology*, 12: 36.
- and Koonin, E. V. (2015) 'Polintons: A Hotbed of Eukaryotic Virus, Transposon and Plasmid Evolution', *Nature Reviews Microbiology*, 13: 105–15.
- Ma, H. et al. (2006) 'MORN motifs in plant PIPKs are involved in the regulation of subcellular localization and phospholipid binding', *Cell Research*, 16: 466–78.
- Makarova, K. S. et al. (2011) 'Defense Islands in Bacterial and Archaeal Genomes and Prediction of Novel Defense Systems', *Journal of Bacteriology*, 193: 6039–56.
- Marcotte, E. M. et al. (1999) 'A Census of Protein Repeats', *Journal of Molecular Biology*, 293: 151–60.
- Maruyama, F., and Ueki, S. (2016) 'Evolution and Phylogeny of Large DNA Viruses, Mimiviridae and Phycodnaviridae Including Newly Characterized Heterosigma Akashiwo Virus', *Frontiers in Microbiology*, 7: 1942.
- McNally, A. et al. (2016) 'Add, Stir and Reduce": *Yersinia* Spp. as Model Bacteria for Pathogen Evolution', *Nature Reviews. Microbiology*, 14: 177–90.
- Merhej, V. et al. (2009) 'Massive Comparative Genomic Analysis Reveals Convergent Evolution of Specialized Bacteria', *Biology Direct*, 4: 13.
- Mittl, P. R. E. and Schneider-Brachert, W. (2007) 'Sel1-like repeat proteins in signal transduction', *Cellular Signalling*, 19: 20–31.
- Moliner, C., Fournier, P. E., and Raoult, D. (2010) 'Genome Analysis of Microorganisms Living in Amoebae Reveals a Melting Pot of Evolution', *FEMS Microbiology Reviews*, 34: 281–94.
- Morriswood, B. and Schmidt, K. (2015) 'A morn repeat protein facilitates protein entry into the flagellar pocket of *Trypanosoma brucei*', *Eukaryotic Cell*, 14: 1081–93.
- Moss, B. (2012) 'Poxvirus Cell Entry: How Many Proteins Does It Take?' *Viruses*, 4: 688–707.
- Neer, E. J. et al. (1994) 'The ancient regulatory-protein family of WD-repeat proteins', *Nature*, 371: 297–300.
- Negi, P., Rai, A. N., and Suprasanna, P. (2016) 'Moving through the Stressed Genome: Emerging Regulatory Roles for Transposons in Plant Stress Response', *Frontiers in Plant Science*, 7: 1448.
- Newton, H. J. et al. (2007) 'Sel1 Repeat Protein LpnE Is a *Legionella pneumophila* Virulence Determinant That Influences Vacuolar Trafficking', *Infect Immun*, 75: 5575–85.
- Nguyen, M. T. H. D., Liu, M., and Thomas, T. (2014) 'Ankyrin-Repeat Proteins from Sponge Symbionts Modulate Amoebal Phagocytosis', *Molecular Ecology*, 23: 1635–45.
- Novais, A. et al. (2010) 'Evolutionary Trajectories of Beta-Lactamase CTX-M-1 Cluster Enzymes: predicting Antibiotic Resistance', *PLoS Pathogens*, 6: e1000735.
- O'Day, D. H. et al. (2006) 'Isolation, characterization, and bioinformatic analysis of calmodulin-binding protein cmbB reveals a novel tandem IP22 repeat common to many *Dictyostelium* and Mimivirus proteins', *Biochemical and Biophysical Research Communications*, 346: 879–88.
- Persi, E. et al. (2016) 'Positive and Strongly Relaxed Purifying Selection Drive the Evolution of Repeats in Proteins', *Nature Communications*, 7: 13570.
- Prag, S. and Adams, J. C. (2003) 'Molecular phylogeny of the kelch-repeat superfamily reveals an expansion of BTB/kelch proteins in animals', *BMC bioinformatics*, 4: 42.
- Raoult, D. et al. (2004) 'The 1.2-Megabase Genome Sequence of Mimivirus', *Science (80-)*, 306: 1344–50.
- and Boyer, M. (2010) 'Amoebae as Genitors and Reservoirs of Giant Viruses', *Intervirology*, 53: 321–9.
- Renesto, P. et al. (2006) 'Mimivirus Giant Particles Incorporate a Large Fraction of Anonymous and Unique Gene Products', *Journal of Virology*, 80: 11678–85.
- Richard, G.-F., Kerrest, A., and Dujon, B. (2008) 'Comparative Genomics and Molecular Dynamics of DNA Repeats in Eukaryotes', *Microbiology and Molecular Biology Reviews*, 72: 686–727.
- Rizopoulos, Z. et al. (2015) 'Vaccinia Virus Infection Requires Maturation of Macropinosomes', *Traffic*, 16: 814–31.

- Rodrigues, R. A. L. et al. (2016) 'Giants among Larges: How Gigantism Impacts Giant Virus Entry into Amoebae', *Curr Opin Microbiol*, 31: 88–93.
- Sakharkar, K. R., Kumar, D. P., and Chow, V. V. T. K. (2004) 'Genome Reduction in Prokaryotic Obligatory Intracellular Parasites of Humans: A Comparative Analysis', *Int J Syst Evol Microbiol*, 54: 1937–41.
- Sedgwick, S. G. et al. (1999) 'The Ankyrin Repeat: A Diversity of Interaction on a Common Structural Framework', *Trends in Biochemecial Sciences*, 24: 311–6
- Sharma, M., and Pandey, G. K. (2015) 'Expansion and Function of Repeat Domain Proteins during Stress and Development in Plants', *Frontiers in Plant Science*, 6: 1218.
- Shchelkunov, S. N., Blinov, V. M., and Sandakhchiev, L. S. (1993) 'Ankyrin-like Proteins of Variola and Vaccinia Viruses', *FEBS Letters*, 319: 163–5.
- Siguiet, P., Gourbeyre, E., and Chandler, M. (2014) 'Bacterial Insertion Sequences: Their Genomic Impact and Diversity', *FEMS Microbiology Reviews*, 38: 865–91.
- Siozios, S. et al. (2013) 'The Diversity and Evolution of Wolbachia Ankyrin Repeat Domain Genes', *PLoS One*, 8: e55390.
- Slimani, M. et al. (2013) 'Amoebae as Battlefields for Bacteria, Giant Viruses, and Virophages', *Journal of Virology*, 87: 4783–5.
- Sobhy, H. et al. (2015) 'Identification of Giant Mimivirus Protein Functions Using RNA Interference', *Frontiers in Microbiology*, 6: 345.
- Suhre, K. (2005) 'Gene and Genome Duplication in Acanthamoeba Polyphaga Mimivirus', *Journal of Virology*, 79: 14095–101.
- Suganuma, T., Pattenden, S. G. and Workman, J. L. (2008) 'Diverse functions of WD40 repeat proteins in histone recognition', *Genes and Development*, 4: 1265–68.
- Sun, C. et al. (2015) 'DNA Transposons Have Colonized the Genome of the Giant Virus Pandoravirus Salinus', *BMC Biology*, 13: 38.
- Takeshima, H. (2000) 'Junctophilins A Novel Family of Junctional Membrane Complex Proteins', *Molecular Cell*, 6: 11–22.
- von Bohl, A. et al. (2015) 'A WD40-repeat protein unique to malaria parasites associates with adhesion protein complexes and is crucial for blood stage progeny', *Malaria Journal. BioMed Central*, 14: 435.
- Voronin, D. A., and Kiseleva, E. V. (2007) 'Functional Role of Proteins Containing Ankyrin Repeats', *Tsitologiya*, 49: 989–99.
- Wicker, T. et al. (2007) 'A Unified Classification System for Eukaryotic Transposable Elements', *Nature Reviews Genetics*, 8: 973–82.
- Wilson, W. H., Van Etten, J. L., and Allen, M. J. (2009) 'The Phycodnaviridae: The Story of How Tiny Giants Rule the World', *Current Topics in Microbiology and Immunology*, 328: 1–42.
- Yutin, N. et al. (2009) 'Eukaryotic Large Nucleo-Cytoplasmic DNA Viruses: clusters of Orthologous Genes and Reconstruction of Viral Genome Evolution', *Virology Journal*, 6: 223.
- et al. (2013) 'Mimiviridae: clusters of Orthologous Genes, Reconstruction of Gene Repertoire Evolution and Proposed Expansion of the Giant Virus Family', *Virology Journal*, 10: 106.
- et al. (2015) 'A Novel Group of Diverse Polinton-like Viruses Discovered by Metagenome Analysis', *BMC Biology*, 13: 95.
- and Koonin, E. V. (2012) 'Hidden Evolutionary Complexity of Nucleo-Cytoplasmic Large DNA Viruses of Eukaryotes', *Virology Journal*, 9: 161.
- , Raoult, D., and Koonin, E. V. (2013) 'Virophages, Polintons, and Transpovirons: A Complex Evolutionary Network of Diverse Selfish Genetic Elements with Different Reproduction Strategies', *Virology Journal*, 10: 158.
- , Wolf, Y. I., and —— (2014) 'Origin of Giant Viruses from Smaller DNA Viruses Not from a Fourth Domain of Cellular Life', *Virology*, 466–467: 38–52.
- Zade, A., Sengupta, M., and Kondabagil, K. (2015) 'Extensive in Silico Analysis of Mimivirus Coded Rab GTPase Homolog Suggests a Possible Role in Virion Membrane Biogenesis', *Frontiers in Microbiology*, 6: 929.
- Zeytuni, N., and Zarivach, R. (2012) 'Structural and Functional Discussion of the Tetra-Trico-Peptide Repeat, a Protein Interaction Module', *Structure. Elsevier Ltd*, 20: 397–405.