Virology ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Virology

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

Multiple occurrences of giant virus core genes acquired by eukaryotic genomes: The visible part of the iceberg?

Jonathan Filée

Laboratoire Evolution, Genomes, Spéciation (LEGS), CNRS UPR9034, Université Paris-Sud. Avenue de la Terrasse, 91190 Gif Sur Yvette, France

ARTICLE INFO

Keywords:
Giant virus
Genome evolution
Lateral gene transfer
Non homologous gene
Displacement

ABSTRACT

Giant Viruses are a widespread group of viruses, characterized by huge genomes composed of a small subset of ancestral, vertically inherited core genes along with a large body of highly variable genes. In this study, I report the acquisition of 23 core ancestral Giant Virus genes by diverse eukaryotic species including various protists, a moss and a cnidarian. The viral genes are inserted in large scaffolds or chromosomes with intron-rich, eukaryotic-like genomic contexts, refuting the possibility of DNA contaminations. Some of these genes are expressed and in the cryptophyte alga *Guillardia theta*, a possible non-homologous displacement of the eukaryotic DNA primase by a viral D5 helicase/primase is documented. As core Giant Virus genes represent only a tiny fraction of the total genomic repertoire of these viruses, these results suggest that Giant Viruses represent an underestimated source of new genes and functions for their hosts.

Introduction

Importance and significance of lateral gene fluxes between viruses and their hosts are still the matter of an intense debate. In a broad outline, the scientific community opinion is divided in two: (i) viruses are « bags of genes » or « genes robbers » frequently acquiring genetic material from cells (Moreira and Lopez-Garcia, 2009; Williams et al., 2011) and (ii) viruses and their hosts have evolved throughout a complex history in which viruses have significantly contributed to the proteome of the cells (Claverie, 2006; Forterre and Prangishvili, 2013; Villarreal and DeFilippis, 2000; Villarreal and Witzany, 2010). In fact, divergent results from a wide variety of viral families and cellular phyla support both views. In the prokaryotic domains, several studies have shown the accretion by phages of cellular genes (polarization cells to viruses) involved in various metabolic and informational functions (Dwivedi et al., 2013; Filee et al., 2006; Hendrix et al., 1999; Ignacio-Espinoza and Sullivan, 2012; Koonin and Dolja, 2006; Moreira, 2000). In eukaryotic viruses, numerous viral genomes carry cellular-originated genes (Bratke and McLysaght, 2008; Filee, Pouget, and Chandler, 2008; Monier et al., 2009; Moreira and Brochier-Armanet, 2008). In favor of the hypothesis of frequent gene transfers viruses to cells, prokaryotic genomes carry numerous inserted complete or partially deleted dsDNA prophages (Cortez et al., 2009) or ssDNA prophages (Krupovic and Forterre, 2011). Moreover, eukaryotic genomes harbor abundant derivatives of retroviruses (Herniou et al., 1998), non-retroviral RNA viruses (Horie et al., 2010) or DNA viruses (Pritham et al., 2007). However, few cases of recruitment of viral genes to perform actual cellular functions are well documented.

In fact, the orientation of the gene transfers (who gives and who receives) is critically dependent on the interpretation of the phylogenetic trees. Very often, viral sequences form a separate cluster, clearly distinguished from the cellular sequences, sometimes positioned as the base of the tree of the eukaryotic sequences [see for example the DNA replication phylogenies (Filee et al., 2002)]. This situation is rather inconclusive: a transfer from the viruses to the ancestors of the cells (viruses to cells) is possible but other explanations can be advocated: (i) vertical inheritance from a common ancestor or (ii) transfer from cells to the virus followed by a high level of sequence divergence in the virus that ultimately blurs the exact relationships. As a matter of fact, clear proofs of gene transfers viruses to hosts are encountered typically with genes widespread in viruses but that have no (or very distantly related) homologs in cellular genomes. Within the few examples reported till date, the syncytin genes promoting the formation of the placenta in mammals are clear cases of domestication of retroviral envelope genes (Dupressoir et al., 2012). Another clear example is the mitochondrial RNA polymerase, DNA polymerase and DNA helicase that are derived from T3/T7 phage genes (Filee and Forterre, 2005). In both cases, the domesticated viral genes have no or very distantly related counterparts in their host genomes. In addition, numerous genomes of retroviruses and tailed bacteriophages carrying these genes have been sequenced making easier the demonstration and the understanding

E-mail address: jonathan.filee@legs.cnrs-gif.fr

http://dx.doi.org/10.1016/j.virol.2014.06.004
0042-6822/

of the direction of the lateral transfers. Taken together these observations indicate that some biases probably lead to an underestimation of the gene flux polarized virus to cell and may have participated to minimize the role of viruses in providing new genes for cellular organisms. Fortunately, collections of related viral genomes become available opening opportunities to better understand the exact nature of the gene flux between viruses and their hosts. Among these collections, the discovery and subsequent genome sequencing of Giant Viruses (GV) all belonging to the Megavirus family (formerly Nucleo Cytoplasmic Large DNA Viruses, NCLDV) (Colson et al., 2013; Iyer et al., 2006) has generated an intense debate about the origin of the extraordinary genome complexity of these elements. Notably, the question of the role of lateral gene transfers between GVs and their diverse cellular hosts has focused on the discussion (Moreira and Lopez-Garcia, 2005). Indeed, Megavirus (NCLDVs) genomes are organized around a very small subset of 30–50 conserved genes called "core genes" encoding mainly DNA replication enzymes and structural functions (Yutin and Koonin, 2012). They are composed principally of viral hallmark genes *i.e.* viral genes with no cellular homolog such as capsid genes (Koonin et al., 2006), or genes that have distantly related homologs in cells (for example much of the components of the DNA replication machinery). With several exceptions (Filee et al., 2008) most of these cores genes are vertically inherited from a putative common ancestor. As GV genomes encode usually more than 1000 genes (Colson et al., 2013) the vast majority of them are not conserved, even in closely related viruses. This erratic gene distribution and the observation that a substantial part of these genes have homologs in cellular genomes have led several authors to propose that lateral acquisitions of genes from cellular organisms have played a decisive role during the evolution of GV genomes (Filee et al., 2007; Iyer et al., 2006; Moreira and Brochier-Armanet, 2008). Currently, there is no clear consensus on this point, mainly because estimation of the exact proportion of genes acquired laterally varies greatly with the methodologies used and the interpretation of the phylogenies. For example, for the mimivirus, approximately 100 genes with strong homologies with eukaryotic genes are detected, but less than ten display phylogenies indicating without ambiguities a gene transfer host to virus (Filee et al., 2008; Ogata et al., 2005). Conversely, gene transfers prokaryotes to GVs are better documented. The abundance of genes with prokaryotic origins are linked to ecology of their hosts: ameba or ciliates graze on microbes or harbor a large diversity of bacterial symbionts, providing a rich gene pool for the viruses (Filee et al., 2007).

Strikingly, there is only a single report in the literature of a gene transfer GV to host (Colson et al., 2011a). Despite their huge genomic repertoires, the contribution of GVs as gene reservoirs of domesticated genes for their hosts is still unknown. As described earlier, a major pitfall is the polarization of the gene transfers. However, we can overcome these difficulties using core genes as markers of gene transfers virus to host. Core genes have the advantages to be ancestral and vertically inherited in GVs for most of them. In addition, core genes are well conserved genes that have generally no (or distantly) related homologs in eukaryotes. Thus, presence of a core GV gene in an eukaryotic genome, displaying strong phylogenetic affinities with GV sequences, would inevitably document gene transfers viruses to hosts. In this study I systematically searched for core genes as defined by Yutin and Koonin in all available complete eukaryotic genomes (Yutin et al., 2013a, 2013b). Subsequent phylogenetic studies of the gene candidates show 23 cases of domestication of GV core genes in a large variety of organisms (amoeba, moss, hydrozoan, diverse alga…). These genes were located in large scaffolds or chromosomes, surrounded by typical intron-rich eukaryotic genes. Some of the cellular GV genes display typical exon/intron structure and the identification of several of these sequences in EST databases suggests that at least some of them are expressed. As core genes represent a tiny fraction of the GV proteome, these results suggest that the occurrences of gene transfers GV to host are probably much more important.

Results and discussion

Diverse core GV genes are present in eight eukaryotic genomes

In order to find core GV genes present in eukaryotic genomes, the 33 core genes presumably present in the ancestor with an apparent vertical inheritance were utilized (Yutin et al., 2013a, 2013b). BLAST searches with these sequences were performed

Table 1
List of GV core genes present in eukaryotic genomes.

Genes	NCVOG	Taxa	First BLAST hit and E-value	Number of introns	Remarks
Major Capsid Protein	0022		*Heterosigma aka.* virus -53	0	6 different genes
			Heterosigma aka. virus-61	3	
		Acanthamoeba castellanii	*Heterosigma aka* virus-47.	1	
			Heterosigma aka. virus-49	1	
			EhV 86-31	0	
			Heterosigma aka. virus-31	0	
			Mimivirus 77	0	3 different genes
		Hydra magnipapillata	Megavirus courdo11-128	0	
			Moumouvirus -111	0	
D5-like helicase/primase	0023	*Guillardia theta*	Mimivirus-51	2	
		Physcomitrella patens	Tunisvirus-52	1	
		Hydra magnipapillata	Moumouvirus-117	0	
Helicase II UL9	0024	*Ectocarpus siliculosus*	Moumouvirus-3	0	Localized outside the pro-virus
VLTF3 transcription factor	0262	*Guillardia theta*	*Ostreococcus* virus OLV3-61	0	
		Hydra magnipapillata	Mamavirus-24	0	
RNA helicase (COG1061)	0076	*Guillardia theta*	*Ostreoccocus* virus Olv5-67	3	
			Micromonas virus 12T-48	1	
		Emiliania huxleyi	*Bathycoccus* virus BpV1-72	3	3 different genes
			Bathycoccus virus BpV1-73	3	
Uracil DNA glycosylase	1115	*Hydra magnipapillata*	Moumouvirus -49	0	
mRNA capping enzyme	1117	*Guillardia theta*	*Phaeocystis* Virus 12T-58	0	
Nudix Hydrolase	0236	*Dictyostellium sp.*	*Chlorella* Virus-4	1	
		Polysphondylium pallidum	Mimivirus-7	1	

Please cite this article as: Filée, J., Multiple occurrences of giant virus core genes acquired by eukaryotic genomes: The visible part of the iceberg? Virology (2014), http://dx.doi.org/10.1016/j.virol.2014.06.004