Title

New evolutionary frontiers from unusual virus genomes

Journal

Genome Biology, 6(3)

ISSN

1474-760X

Authors

Desjardins, Christopher
Eisen, Jonathan A
Nene, Vishvanath

Publication Date

2005

DOI

10.1186/gb-2005-6-3-212

Peer reviewed

Minireview

New evolutionary frontiers from unusual virus genomes

Christopher Desjardins, Jonathan A Eisen and Vishvanath Nene

Address: The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20854, USA.

Correspondence: Vishvanath Nene. E-mail: nene@tigr.org. Jonathan Eisen. E-mail: jeisen@tigr.org

Published: 2 March 2005

Genome ***Biology*** 2005, **6:**212

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2005/6/3/212

Abstract

The sequences of two giant viral genomes, Mimivirus and a polydnavirus, have recently been published. Mimivirus has the largest known viral genome and encodes an unprecedented number of proteins, whereas the polydnavirus genome has an extremely low coding density and does not encode DNA-replication proteins. These and other unusual features challenge the way we view the evolution and definition of viruses.

If an alien landed on Earth and studied the biology here, it might justifiably conclude that viruses run the planet. They are numerically the most abundant biological entities [1], and they are profoundly important in shaping the ecology and evolution of just about every species on Earth [2]. Yet viruses are not considered to be alive by most biologists, and they have arguably fallen by the wayside in the genomics revolution [3]. The recent publication of the genome sequences of two unusual viruses, however, highlights the wealth of information that remains to be discovered through viral genomics. Here, we discuss Mimivirus [4] and *Cotesia congregata* Bracovirus [5] (CcBV) and the interesting questions they raise concerning the biology and evolution of viruses.

Both Mimivirus and CcBV are classified as double-stranded DNA (dsDNA) viruses, and some of their features are summarized in Table 1. Mimivirus was discovered in amoebae [6], and it has a cycle of viral transmission and replication that is typical of many dsDNA viruses (Figure 1a). Its name is derived from 'mimicking microbe,' in reference to the bacterium-like appearance of its large particle (400 nm in diameter) and its Gram-positive staining. Mimivirus has the largest known viral genome (1.18 megabase-pairs) and encodes an unprecedented number of components of the transcriptional, translational and replication machinery, many of which have not previously been identified in viruses [4]. In addition, the genome encodes a large number of genes associated with metabolic pathways. Although the size and content of the Mimivirus genome might rival those of some obligate intracellular prokaryotes, it still appears to be absolutely dependent on its host cell for synthesis of proteins.

CcBV differs from Mimivirus and other viruses in many fundamental aspects. As a member of the Polydnaviridae, the transmission and replication cycle of this Bracovirus is unconventional [7]. The Polydnaviridae - pronounced polyd-na-viridae by the research community and named after the unique segmented structure of the packaged genome - consists of two subgroups, Bracoviruses and Ichnoviruses, which associate with braconid and ichneumonid wasps, respectively [7]. These wasps are parasitoids (parasites that kill their hosts) that attack caterpillars and are of particular interest for their use as biological control agents. In the wasp host, polydnaviruses exist in a benign state, integrated into the wasp genome as a provirus. Amplification of segments from the provirus and production of virions (particles containing viral DNA encased within a capsid) occurs only in the ovaries of a female wasp, and virions are co-injected with eggs during parasitization of caterpillars. The viral particles are replication-deficient in both hosts; the virus can increase in number only through genome amplification in wasp ovaries but is transmitted from wasp to wasp by vertical transmission of the provirus. Viral gene expression in caterpillars interferes with the latter's immune response and developmental cycle, promoting

Table 1

Characteristics of Mimivirus and CcBV and their genomes

Feature	Mimivirus	CcBV (in caterpillar host)	CcBV (in wasp host)
Genome size (base-pairs)	1,181,404	567,670	NK
Genome structure	Linear	30 closed circles	Provirus (linear)*
G + C composition (%)	28.0	33.95	NK
Coding density (%)	90.5	26.9	NK
Number of genes	1,262	156	NK
Genes containing introns	4†	107	NK
Genes with assigned function	298	42	NK
Obligate intracellular parasite	Yes	Yes	No‡
Viral DNA replication in host	Yes	No	No*
Virion assembly in host	Yes	No	Yes
Transmission via virion	Yes	No	No*
Viral gene expression dependent on host cellular machinery	Yes	Yes	Yes

NK, not known; *circles are produced from amplified proviral DNA in female wasps at the time of egg laying, resulting in vertical transmission; †genes contain self-splicing introns; ‡ CcBV is not pathogenic to the wasp host and exhibits a mutualistic association with it (see text for details).

survival of the parasitoid and therefore of the provirus. Thus, polydnaviruses depend on vertical transmission in a tripartite relationship that includes both mutual and parasitic symbioses.

The genome of CcBV - whose wasp host is *C. congregata* - totals 568 kilobase-pairs (kbp) and is composed of 30 circles ranging in size from about 5 kbp to 40 kbp [5]. Although the cumulative genome size of CcBV would place it in the category of a giant virus, segments appear to be packed into individual capsids, with several capsids being enveloped by a single membrane [7] (Figure 1b). In contrast to the high coding density of most viruses, the CcBV genome encodes very few proteins, and the smallest segment consists entirely of non-coding DNA [5]. Almost 70% of the protein-coding genes are predicted to contain introns dependent on spliceosomal excision; it is unusual for viruses to have introns. This high rate of intron prediction remains to be confirmed by cDNA sequence data, however. About 40% of the proteins with assigned functions fall into four gene families: protein tyrosine phosphatases, inhibitors of NF-κB, cystatins and cysteine-rich proteins; these proteins may modulate the responses of lepidopteran caterpillars to infection. In contrast with Mimivirus and other viruses, the CcBV genome contains very few recognizable homologs of components of the transcriptional, translational and replication machinery, although it does encode a homolog of the chromatin protein histone H4.

Mimivirus is the sole member of Mimiviridae and is classified as a nucleocytoplasmic large DNA virus (NCLDV) [6] on the basis of the presence/absence pattern of 'core' genes defined for NCLDVs [8]. By contrast, the viral origins of polydnaviruses are less certain. It has been hypothesized from virion morphology that bracoviruses may be related to baculoviruses [7,9], but only three genes in the CcBV genome are similar to genes found in free-replicating viruses, two to a baculovirus and one to an ascovirus [5]. Thus, even though we have a genome sequence the origin of bracoviruses remains unclear. Many genes that are typical of viruses are absent from the CcBV genome and may have been transferred to the wasp genome, as is the case for a gene coding for a major structural protein in *Campoletis sonorensis* Ichnovirus [10].

Inferring viral phylogenies is often difficult, as high rates of viral evolution make it difficult to identify conserved genes between viruses. The ultimate origin of viruses - where viruses should be placed on the Tree of Life [11] - is even more vexing. Many theories abound [2,12-14]: that they evolved before the first cells; that because they infect all domains they arose from cellular life before the last universal common ancestor; or that they evolved from cells at a later point in evolution. In principle, it should be possible to distinguish among these theories through careful genome analysis. For example, if viruses have a separate origin from 'living' organisms, their gene content should overlap very little, if at all, with that of bacteria, archaea and eukaryotes. In contrast, if viruses evolved from a bacterial parasite, their content should resemble bacteria more than that of archaea and their genes should branch in evolutionary trees with genes from bacteria, as is the case for organelles [15].

But there are several complications in determining viral origins. First, it is possible that different types of viruses arose independently. Second, and more confounding, there has unquestionably been gene flow between viruses and their hosts, which means that any one gene might not reflect the phylogeny of the virus itself. The authors of the Mimivirus genome sequence paper [4] try to address this by making a concatenated alignment of multiple genes shared by Mimivirus and living organisms. They report that the viral genes branch as a sister group to the eukaryotes, potentially identifying Mimivirus as the basal member of a major branch on the Tree of Life. But in a separate phylogenetic analysis of the RNA polymerase β' subunit (one of the proteins used in the concatenated analysis and also encoded in other NCLDVs), Mimivirus did not group with the NCLDVs. The ramifications of this conflict are still unclear: is Mimivirus an NCLDV that acquired many of its genes through lateral transfer, although differences in codon usage between viruses and amoebae would indicate otherwise? Is it a sister-group to the eukaryotes and not an NCLDV, despite the presence of NCLDV core genes? Resolving such questions will have to wait until other members or close relatives of Mimiviridae are discovered and their genome sequences analyzed.

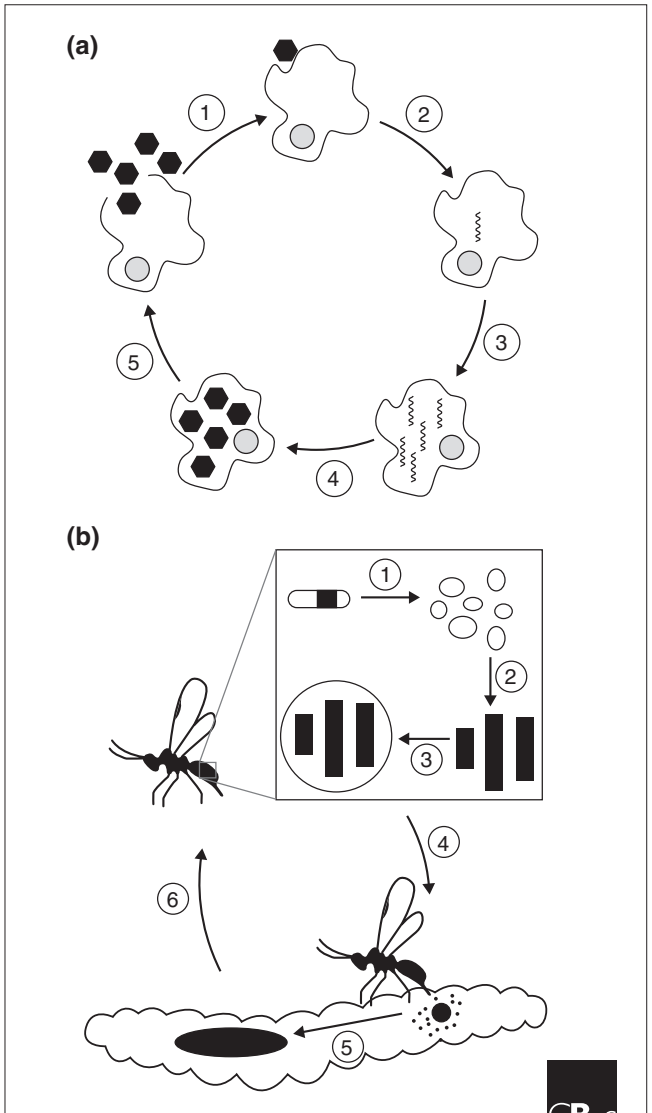

Figure 1
The transmission and replication cycles of Mimivirus and CcBV. **(a)** Mimivirus. At the beginning of the life cycle, 1, the virus enters the amoeba; 2, the viral genome is released; 3, viral proteins are expressed and whole virus genomes are replicated; 4, viral genomes are packaged into capsids; 5, viral particles are released from the amoeba. The grey circle represents the nucleus. **(b)** CcBV. Steps 1-3 take place in specialized ovarian cells of the wasp: 1, the provirus integrated in the wasp genome is amplified; 2, amplified viral DNA is packaged into capsids; 3, several capsids are enveloped by membrane(s); 4, the wasp oviposits eggs (only one is shown) and virions into a caterpillar; 5, viral gene expression promotes survival of the wasp progeny (such as the larva shown), but virions do not replicate; 6, the wasp larva (oval) emerges from the caterpillar and metamorphoses into an adult, and the caterpillar dies.

As CcBV has very few features associated with other viruses, and its coding content and gene structure resemble that of the wasp host more than that of viruses, Espagne *et al.* [5] raise questions about the viral ancestry of bracoviruses. They propose that bracoviruses may have evolved from mobile DNA that acquired the ability to be packaged into capsids (perhaps from a virus); the existence of remnants of transposon and retrovirus-like elements in the CcBV genome provides additional support for such an argument. Genome sequencing and phylogenetic analysis of additional polydnaviruses, their proviral sequences, and genes associated with virion formation will be required to shed light on the question of whether bracoviruses are the product of reductive viral evolution or not.

Mimivirus expands our definition of viruses quantitatively to accommodate bigger genomes and larger particle size. Although Raoult *et al.* [4] point out that Mimivirus has more components of the cellular machinery than any other virus, in our opinion it does not yet seem to stretch the definition of viruses in any fundamental way. It is just a more complicated virus. By contrast, CcBV appears to differ qualitatively from many definitions of a virus (Table 1), but it could still be classified as a highly defective one. Clearly, these two viruses present some interesting problems regarding viral phylogeny and classification that remain to be resolved. Considering the importance of viruses in evolution, we believe that we need to direct more effort to systematically characterize the genomes and biology of diverse viruses, as this will further our understanding of how and where they fit into the Tree of Life [11].

References

1. Bergh O, Borsheim KK, Bratbak G, Heldal M: **High abundance of viruses found in aquatic environments.** *Nature* 1989, **340:**467-468.
2. Filée J, Forterre P, Laurent J: **The role played by viruses in the evolution of their hosts: a view based on informational protein phylogenies.** *Res Microbiol* 2003, **154:**237-243.
3. Mindell DP, Villarreal L: **Don't forget about viruses.** *Science* 2003, **302:**1677.
4. Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, La Scola B, Suzan M, Claverie J-M: **The 1.2-megabase genome sequence of Mimivirus.** *Science* 2004, **306:**1344-1350.
5. Espagne E, Dupuy C, Huguet E, Cattolico L, Provost B, Martins N, Poirie M, Periquet G, Drezen JM: **Genome sequence of a polydnavirus: insights into symbiotic viral evolution.** *Science* 2004, **306:**286-289.
6. La Scola B, Audic S, Robert C, Jungang L, de Lamballerie X, Drancourt M, Birtles R, Claverie J-M, Raout D: **A giant virus in Amoebae.** *Science* 2003, **299:**2033.
7. Beckage N (Ed): **Polydnaviruses of insects: evolution and physiological functions.** In *J Insect Physiology* 2003, **49:**395-543.
8. Iyer LM, Aravind L, Koonin EV: **Common origin of four diverse families of large eukaryotic DNA viruses.** *J Virology* 2001, **75:**11720-11734.
9. Turnbull M, Webb, B: **Perspectives on polydnavirus origins and evolution.** *Advances Virus Res* 2002, **58:**203-254.
10. Deng L, Stoltz DB, Webb BA: **A gene encoding a polydnavirus structural protein is not encapsidated.** *Virology* 2000, **269:**440-450.
11. **Tree of Life** [http://tolweb.org/tree/phylogeny.html]
12. Forterre P: **The great virus comeback—from an evolutionary perspective.** *Res Microbiol* 2003, **154:**223-225.
13. Krisch HM: **The view from Les Treilles on the origins, evolution and diversity of viruses.** *Res Microbiol* 2003, **154:**227-229.
14. Bamford DH: **Do viruses form lineages across different domains of life?** *Res Microbiol* 2003, **154:**231-236.
15. Gray MW: **Evolution of organellar genomes.** *Curr Opin Genet Dev* 1999, **9:**678-687.