Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Virology

Double-stranded DNA viruses: 20 families and only five different architectural principles for virion assembly

Mart Krupovic[1] and Dennis H Bamford[2]

The number of viral particles in the biosphere is enormous. Virus classification helps to comprehend the virosphere and to understand the relationship between different virus groups. However, the evolutionary reach of the currently employed sequence-based approaches in virus taxonomy is rather limited, producing a fragmented view of the virosphere. As a result, viruses are currently classified into 87 different families. However, studies on virion architectures have unexpectedly revealed that their structural diversity is far more limited. Here we describe structures of the major capsid proteins of double-stranded DNA viruses infecting hosts residing in different domains of life. We note that viruses belonging to 20 different families fall into only five distinct structural groups, suggesting that optimal virus classification approach should equally rely on both sequence and structural information.

Addresses
[1] Institut Pasteur, Unité Biologie Moléculaire du Gène chez les Extrêmophiles, Department of Microbiology, 25 rue du Dr Roux, 75015 Paris, France
[2] Department of Biosciences and Institute of Biotechnology, Viikki Biocenter 2, P.O. Box 56 (Viikinkaari 5), University of Helsinki, FIN-00014 Helsinki, Finland

Corresponding author: Krupovic, Mart (krupovic@pasteur.fr)

Current Opinion in Virology 2011, **1**:118–124

This review comes from a themed issue on
Virus structure and function
Edited by Richard Kuhn and Mavis Agbandje-McKenna

Available online 2nd July 2011

1879-6257/$ – see front matter

DOI 10.1016/j.coviro.2011.06.001

Introduction

Our comprehension of the virosphere has changed dramatically during the past decade or so. Numerous studies have illuminated the previously unsuspected abundance of viruses in any given ecological niche, be it aquatic or terrene [1]. Viruses are believed to outnumber their host cells by at least an order of magnitude and the estimated number of viruses in the biosphere is indeed mind-boggling. With 10^{31} virions present in the biosphere [2], viruses constitute the most copious group of biological entities on our planet. As a result, viruses are a major factor controlling the number, diversity and evolution of their hosts and have a significant impact on global biogeochemical cycles [3–5].

Human mindset seeks order in every system it encounters, from subatomic particles to biological entities all the way to astronomical bodies. Among biological systems, the virosphere is the least understood one. The reason for this may lie within our way of approaching the task; viruses are currently classified mainly based on the sequence data [6]. As of today (April 2011), there are less than 4000 complete viral genome sequences available at the NCBI database. And yet, these viruses fall into 87 distinct families currently recognized by the International Committee on Taxonomy of Viruses (ICTV; URL: http://www.ictvonline.org). Whereas all cellular organisms are neatly grouped into three domains and are believed to share a common ancestor, the relationship between viruses belonging to different families is not so clear. Furthermore, viruses with available complete genome sequences constitute a truly miniscule fraction of those present in the biosphere. Therefore, it is obvious that unless we stop isolating new viruses (clearly not an option), many more viral families will have to be created in the next few years.

It is obvious that the number of established virus families does not reflect the number of independent evolutionary events that led to emergence of novel types of viruses. Certain viruses, despite belonging to different families, are evolutionarily related and almost surely have evolved from a common ancestor (see below). However, insights that reach further back in time are rarely obtained from sequence-based analyses, and are therefore poorly reflected in the current taxonomy of viruses [7••]. In addition, high mutation rate [8], extensive horizontal gene exchange and mosaic nature of viral genomes [9] suggest that it will never be possible to organize the virosphere in a biologically relevant manner based on sequence information alone.

Genes for genome replication proteins are often shuffled between unrelated viruses, plasmids and their hosts [10,11,12•,13]. The ability to build virions, on the other hand, is a hallmark of viruses that distinguishes them from other mobile genetic elements [7••]. Whereas possible protein sequence diversity is enormous (theoretically, 22^{100} different 100 amino acid-long polypeptides can be synthesized using the 22 proteinogenic amino acids), the number of biologically relevant structural folds is far more modest. According to the latest release of the SCOP (Structural Classification of Proteins) database (v1.75, February 2009) [14], the 38 221 proteins for which high resolution structures are available possess 1195 distinct folds. Obviously, only a handful of these folds can be

utilized to construct infectious virions. Consequently, the entire virosphere is likely to be comprised of only a limited number of viral groups that rely on fundamentally different architectural principles for the assembly of their virions. In this review, we focus on virion architectures of double-stranded (ds) DNA viruses infecting hosts from different domains of life.

Viruses with dsDNA genomes

Viruses with dsDNA genomes are classified by the ICTV into 28 distinct families (ICTV; URL: http://www.ictvonline.org), while some are still awaiting taxonomic assessment. These viruses differ tremendously in their genome lengths as well as virion sizes and complexity (Table 1) [6]. Information on the structure of the major virion proteins obtained by X-ray crystallography, high-resolution electron microscopy (EM) or bioinformatic analyses is currently available for dsDNA viruses belonging to 20 different families. Notably, based on the structures of the capsid proteins, viruses belonging to these families can be grouped into five classes or lineages (Table 1), which we discuss below.

Viruses with the jelly-roll capsid proteins

The jelly-roll fold is one of the most commonly found structural folds in viral proteins. It consists of two four-stranded β-sheets facing one another, where the end strands almost form a closed circular barrel structure (reviewed in [15,16]). Currently, viruses from 16 different families, including those with RNA and DNA genomes, are known to build their capsids using the jelly-roll fold [17•]. However, among dsDNA viruses only those belonging to families *Papillomaviridae* and *Polyomaviridae* utilize this structural fold as the main building block for

Table 1

Overview of dsDNA viruses.

	Family	Virion morphology	Host	MCP fold	PDB
	Papillomaviridae	Icosahedral	E	jelly-roll	1DZL
	Polyomaviridae	Icosahedral	E	jelly-roll	1SVA
NCLDVs	*Asfarviridae*	Icosahedral	E	double jelly-roll	
NCLDVs	*Iridoviridae*	Icosahedral	E	double jelly-roll	
NCLDVs	*Mimiviridae*	Icosahedral	E	double jelly-roll	
NCLDVs	*Phycodnaviridae*	Icosahedral	E	double jelly-roll	1J5Q
NCLDVs	*Poxviridae*	Oval	E	double jelly-roll	
	Adenoviridae	Icosahedral	E	double jelly-roll	1P30
	Ascoviridae	Oval	E	doublejelly-roll	
	Corticoviridae	Icosahedral	B	double jelly-roll	2VVF
	Tectiviridae	Icosahedral	B	double jelly-roll	1HX6
	STIV (unassigned)	Icosahedral	A	double jelly-roll	2BBD
Caudo-Herpes-virales	*Alloherpesviridae*	Icosahedral	E	HK97-like	
Caudo-Herpes-virales	*Herpesviridae*	Icosahedral	E	HK97-like	
Caudo-Herpes-virales	*Malacoherpesviridae*	Icosahedral	E	HK97-like	
Caudo-Herpes-virales	*Myoviridae*	Icosahedral heads, contractile tails	B, A	HK97-like	1YUE
Caudo-Herpes-virales	*Podoviridae*	Icosahedral heads, short tails	B	HK97-like	2XYY
Caudo-Herpes-virales	*Siphoviridae*	Icosahedral heads, non-contractile tails	B, A	HK97-like	1OHG
	Lipothrixviridae	Flexible filaments	A	four-helix bundle	3FBL, 3FBZ
	Rudiviridae	Stiff rod-shaped	A	four-helix bundle	3F2E
	Bicaudaviridae	Spindle-shaped with two tails	A	mainly α-helical	3FAJ
	Ampullaviridae	Bottle-shaped	A	TBD	
	Baculoviridae	Rod-shaped	E	TBD	
	Fuselloviridae	Spindle-shaped/pleomorphic	A	TBD	
	Globuloviridae	Spherical, enveloped	A	TBD	
	Guttaviridae	Droplet-shaped	A	TBD	
	Nimaviridae	Ovoid/bacilliform with a tail	E	TBD	
	Plasmaviridae	Pleomorphic	B	TBD	
	Polydnaviridae	Rod-shaped, fusiform	E	TBD	

MCP, major capsid protein; NCLDVs, Nucleo-cytoplasmic large DNA viruses; STIV, *Sulfolobus* turreted icosahedral virus; A, Archaea; B, Bacteria; E, Eukarya; TBD, to be determined.