



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Chapter 2

Virion Structure, Genome Organization, and Taxonomy of Viruses

Gustavo Fermin

Universidad de Los Andes, Mérida, Venezuela

All living habitats . . . have and must operate in a virosphere.

Luis Villareal

WHAT IS A VIRUS?—PERHAPS THERE IS A SIMPLE ANSWER, NOT AN EASY ANSWER

The notion of what a virus is, and the nature of its relationship with the cellular world, has changed considerably over the past few decades. Viruses were initially distinguished from other organisms by their small size and their ability to pass through filters. They were described as “*filterable disease agents*”: in other words, infectious agents that are small enough to pass through bacterial filters. Since viruses are quintessential parasites that depend on a host for most of their sustaining functions, logic dictated that viruses also be defined and classified on the basis of the host they infect. However, the host range has proven to be a very complex biological phenomenon with intricacies that are yet to be fully understood. Additionally, not all viruses can be defined as strict parasites given the benefits derived from their association with certain hosts.

The recent discovery of giant viruses that infect protists and bacteria has initiated a major shift in the thinking on how to define and classify viruses. Indeed both the particle and genome sizes of these viruses overlap significantly with those of bacteria, some eukaryotes, and archaea. The discovery called for a definition of viruses based on their essential nature rather than on their size. To this end, one proposal called for the revision of the living world into two major groups of organisms, ribosome-encoding organisms

that include all archaea, bacteria and eukarya, and capsid-encoding organisms, the viruses. Clearly, this division distinguishes viruses from cellular organisms based on genome content; in particular, the genes that encode ribosomal proteins and ribosomal RNAs. These genes are among the few genes whose sequences are conserved in all cellular organisms. The last universal common ancestor probably possessed a sophisticated ribosome that contained at least 34 ribosomal proteins that are shared by all archaeal, bacterial, and eukaryotic organisms. Unlike cellular organisms, viruses lack ribosomes and must use the ribosomes of their host cells for the translation of their mRNA into proteins.

While there is no single protein that is common to all viruses the expression of a capsid is a necessary structure that is used to disseminate viruses between ribosome-encoding hosts. One striking feature of viral capsids lies in the folded topology of the protein monomers that make up the capsid. The jellyroll β barrel appears to be the most prevalent structural core motif among capsid proteins, but it has not been found in any cellular protein. However, numerous groups of viruses share a common evolutionary history with genetic elements that lack a capsid (and its coding gene) and are never encapsidated or, in some cases, encapsidated in the virions of “*host*” viruses. Hence another proposal that has been put forward is a scheme that does not focus on the presence of a capsid or any particular gene, but rather is rooted in the concept of genetic, informational parasitism. That is, viruses are parasitic genetic information; they possess a varied repertoire of replication strategies and establish a range of relationships with cellular hosts that exhibit various degrees of reliance on the information processing systems of the host. In this chapter we first examine the range of biological features of viruses. We later look at the taxonomic classification schemes for viruses, along with specific pitfalls. For a list of all recognized virus families and unassigned genera, the reader is referred to [Table 2.1](#).

VIRION MORPHOLOGIES

A virion is the physical entity that encompasses all that a virus represents in terms of its genome and the encapsidating protein shell, which maintains structural integrity and contribute to functional properties such as transmission. A virion is only one of the physical manifestations of a virus during a defined stage of its replication cycle; indeed, the most complex and complete of its physical manifestations. The most abundant forms of virions are based on spheres and rods with icosahedral and helical symmetries, respectively, and all their elegant variations. Some bacteriophages possess a unique morphology and are tadpole shaped. They possess a head-tail morphology consisting of a combination of icosahedral and helical symmetries. The recent discovery of archaeal viruses has widened our knowledge of viral morphologies. Although most resemble bacteriophages, some virions of

TABLE 2.1 List of Virus Families and Unassigned Genera as Recognized by the ICTV (2017), and Some Selected Features

Order	Viral Family	Genome Type	Genomic Molecule(s)	Host	Species	Env.	Virion Morphology
Unassigned	<i>Adenoviridae</i>	dsDNA	1 L	A	62	N	Icosahedral
<i>Herpesvirales</i>	<i>Alloherpesviridae</i>	dsDNA	1 L	A	12	Y	Spherical to pleomorphic
<i>Tymovirales</i>	<i>Alphaflexiviridae</i>	ssRNA(+)	1 L	F-P	50	N	Filamentous, flexible
Unassigned	<i>Alphatetraviridae</i>	ssRNA(+)	1 or 2 L	A	10	N	Icosahedral
Unassigned	<i>Alvernaviridae</i>	ssRNA(+)	1 L	Pr	1	N	Icosahedral
Unassigned	<i>Amalgaviridae</i>	dsRNA	1 L	P	4	NA	Capsid-less
Unassigned	<i>Ampullaviridae</i>	dsDNA	1 L	Ar	1	Y	Bottle-shaped, helical nucleoprotein
Unassigned	<i>Anelloviridae</i>	ssDNA(-)	1 C	A	68	N	Icosahedral
Unassigned	<i>Arenaviridae</i>	ssRNA(+ / -)	2 L	A	36	Y	Spherical to pleomorphic
<i>Nidovirales</i>	<i>Arteriviridae</i>	ssRNA(+)	1 L	A	17	Y	Spherical to pleomorphic
Unassigned	<i>Ascoviridae</i>	dsDNA	1 C	A	4	Y	Bacilliform
Unassigned	<i>Asfarviridae</i>	dsDNA	1 L	A	1	Y	Spherical to pleomorphic
Unassigned	<i>Astroviridae</i>	ssRNA(+)	1 L	A	22	N	Icosahedral
Unassigned	<i>Avsunviroidae</i>	ssRNA	1 C	P	4	Vrd	Capsid-less
Unassigned	<i>Baculoviridae</i>	dsDNA	1 C	A	66	Y	Rod-shaped, helical nucleocapsid
Unassigned	<i>Barnaviridae</i>	ssRNA(+)	1 L	F	1	N	Bacilliform
Unassigned	<i>Benyviridae</i>	ssRNA(+)	1 L	P	4	N	Rod shaped

(Continued)

TABLE 2.1 (Continued)

Order	Viral Family	Genome Type	Genomic Molecule(s)	Host	Species	Env.	Virion Morphology
<i>Tymovirales</i>	<i>Betaflexiviridae</i>	ssRNA(+)	1 L	F-P	89	N	Filamentous, flexible
Unassigned	<i>Bicaudaviridae</i>	dsDNA	1 C	Ar	1	N	Lemon shaped
Unassigned	<i>Bidnaviridae</i>	ssDNA	2 L	A	1	N	Icosahedral
Unassigned	<i>Birnaviridae</i>	dsRNA	2 L	A	6	N	Icosahedral
<i>Mononegavirales</i>	<i>Bornaviridae</i>	ssRNA(-)	1 L	A	8	Y	Spherical
Unassigned	<i>Bromoviridae</i>	ssRNA(+)	3 L	P	33	N	Icosahedral or Bacilliform
Unassigned	<i>Caliciviridae</i>	ssRNA(+)	1 L	A	7	N	Icosahedral
Unassigned	<i>Carmotetraviridae</i>	ssRNA(+)	1 L	A	1	N	Icosahedral
Unassigned	<i>Caulimoviridae</i>	dsDNA-RT	1 C	A-P	62	N	Bacilliform or Icosahedral
Unassigned	<i>Chrysoviridae</i>	dsRNA	4 L	F	9	N	Icosahedral
Unassigned	<i>Circoviridae</i>	ssDNA(+ / -)	1 C	A	70	N	Icosahedral
Unassigned	<i>Clavaviridae</i>	dsDNA	1 C	Ar	1	N	Bacilliform
Unassigned	<i>Closteroviridae</i>	ssRNA(+)	1-3 L	P	49	N	Filamentous, flexible
<i>Nidovirales</i>	<i>Coronaviridae</i>	ssRNA(+)	1 L	A	39	Y	Bacilliform or Spherical
Unassigned	<i>Corticoviridae</i>	dsDNA	1 C	B	1	N	Icosahedral
Unassigned	<i>Cystoviridae</i>	dsRNA	3 L	B	1	Y	Spherical
<i>Picornavirales</i>	<i>Dicistroviridae</i>	ssRNA(+)	1 L	A	15	N	Icosahedral
Unassigned	<i>Endornaviridae</i>	dsRNA	1 L	C-F-P	22	NA	Capsid-less

<i>Bunyavirales</i>	<i>Feraviridae</i>	ssRNA(-)	3 L	A	1	Y	Spherical, helical nucleocapsid
<i>Mononegavirales</i>	<i>Filoviridae</i>	ssRNA(-)	1 L	A	7	Y	Filamentous, flexible, helical nucleocapsid
<i>Bunyavirales</i>	<i>Fimoviridae</i>	ssRNA(-)	4-8 L	P	9	Y	Spherical, nucleocapsid
Unassigned	<i>Flaviviridae</i>	ssRNA(+)	1 L	A	82	Y	Spherical
Unassigned	<i>Fuselloviridae</i>	dsDNA	1 C	Ar	9	Y	Spindle-shaped
<i>Tymovirales</i>	<i>Gammaflexiviridae</i>	ssRNA(+)	1 L	F	1	N	Filamentous, flexible
Unassigned	<i>Geminiviridae</i>	ssDNA(+ / -)	1-2 C	P	369	N	Icosahedral
Unassigned	<i>Genomoviridae</i>	ssDNA(+ / -)	1 C	A-F-P	73	N	Icosahedral
Unassigned	<i>Globuloviridae</i>	dsDNA	1 L	Ar	2	Y	Spherical, pleomorphic
Unassigned	<i>Guttaviridae</i>	dsDNA	1 C	Ar	2	Y	Droplet shaped
<i>Bunyavirales</i>	<i>Hantaviridae</i>	ssRNA(-)	3 L	A	41	Y	Spherical, nucleocapsid
Unassigned	<i>Hepadnaviridae</i>	dsDNA-RT	1 C	A	12	Y	Spherical
Unassigned	<i>Hepeviridae</i>	ssRNA(+)	1 L	A	5	N	Icosahedral
<i>Herpesvirales</i>	<i>Herpesviridae</i>	dsDNA	1 L	A	89	Y	Spherical
Unassigned	<i>Hypoviridae</i>	ssRNA(+)	1 L	F	4	NA	Capsid-less
Unassigned	<i>Hytrosaviridae</i>	dsDNA	1 C	A	2	Y	Helical nucleocapsid
<i>Picornavirales</i>	<i>Iflaviridae</i>	ssRNA(+)	1 L	A	15	N	Icosahedral
Unassigned	<i>Inoviridae</i>	ssDNA(+)	1 C	B	31	N	Filamentous to rod shaped
Unassigned	<i>Iridoviridae</i>	dsDNA	1 L	A	12	Y/N	Icosahedral or Spherical
<i>Bunyavirales</i>	<i>Jonviridae</i>	ssRNA(-)	3 L	A	1	Y	Tubular or Spherical

(Continued)

TABLE 2.1 (Continued)

Order	Viral Family	Genome Type	Genomic Molecule(s)	Host	Species	Env.	Virion Morphology
Unassigned	<i>Lavidaviridae</i>	dsDNA	1 C	Pr	3	N	Icosahedral
Unassigned	<i>Leviviridae</i>	ssRNA(+)	1 L	B	4	N	Icosahedral
<i>Ligamenvirales</i>	<i>Lipothrixviridae</i>	dsDNA	1 L	Ar	8	Y	Filamentous
Unassigned	<i>Luteoviridae</i>	ssRNA(+)	1 L	P	34	N	Icosahedral
<i>Herpesvirales</i>	<i>Malacoherpesviridae</i>	dsDNA	1 L	A	2	Y	Icosahedral to pleomorphic
<i>Picornavirales</i>	<i>Marnaviridae</i>	ssRNA(+)	1 L	C	1	N	Icosahedral
Unassigned	<i>Marseilleviridae</i>	dsDNA	1 C	Pr	4	N	Icosahedral
Unassigned	<i>Megabirnaviridae</i>	dsRNA	2 L	F	1	N	Icosahedral
<i>Nidovirales</i>	<i>Mesoniviridae</i>	ssRNA(+)	1 L	A	7	Y	Spherical
Unassigned	<i>Metaviridae</i>	ssRNA-RT	1 L	A-F-P	39	Y	Ovoidal
Unassigned	<i>Microviridae</i>	ssDNA(+)	1 C	B	21	N	Icosahedral
Unassigned	<i>Mimiviridae</i>	dsDNA	1 L	Pr	2	N	Icosahedral (Pseudo)
<i>Mononegavirales</i>	<i>Myonaviridae</i>	ssRNA(-)	1 L	F	1	Y	Filamentous, flexible, helical nucleocapsid
<i>Caudovirales</i>	<i>Myoviridae</i>	dsDNA	1 L	Ar-B	271	N	Icosahedral head with tail
<i>Bunyavirales</i>	<i>Nairoviridae</i>	ssRNA(-)	3 L	A	12	Y	Spherical
Unassigned	<i>Nanoviridae</i>	ssDNA(+)	8 or 6 C	P	12	N	Icosahedral
Unassigned	<i>Narnaviridae</i>	ssRNA(+)	1 L	C-F-Pr	7	NA	Capsid-less

Unassigned	<i>Nimaviridae</i>	dsDNA	1 C	A	1	Y	Helical nucleocapsid
Unassigned	<i>Nodaviridae</i>	ssRNA(+)	2 L	A	9	N	Icosahedral
Unassigned	<i>Nudiviridae</i>	dsDNA	1 C	A	3	Y	Helical nucleocapsid
<i>Mononegavirales</i>	<i>Nyamiviridae</i>	ssRNA(-)	1 L	A	5	Y	Spherical, helical nucleocapsid
Unassigned	<i>Ophioviridae</i>	ssRNA(-)	3 or 4 L	P	7	N	Filamentous, flexible nucleocapsids
Unassigned	<i>Orthomyxoviridae</i>	ssRNA(-)	6-8 L	A	9	Y	Spherical, Filamentous, Pleomorphic
Unassigned	<i>Papillomaviridae</i>	dsDNA	1 C	A	116	N	Icosahedral
<i>Mononegavirales</i>	<i>Paramyxoviridae</i>	ssRNA(-)	1 L	A:	49	Y	Spherical, Filamentous, Pleomorphic
Unassigned	<i>Partitiviridae</i>	dsRNA	2 L	F-P	60	N	Icosahedral
Unassigned	<i>Parvoviridae</i>	ssDNA(+ / -)	1 L	A	62	N	Icosahedral
<i>Bunyavirales</i>	<i>Peribunyaviridae</i>	ssRNA(-)	3 L	A	52	Y	Spherical, nucleocapsid
Unassigned	<i>Permutotetraviridae</i>	ssRNA(+)	1 L	A	2	N	Icosahedral
<i>Bunyavirales</i>	<i>Phasmaviridae</i>	ssRNA(-)	3 L	A	6	Y	Helical nucleocapsid (?)
<i>Bunyavirales</i>	<i>Phenuiviridae:</i> <i>Goukovirus</i>	ssRNA(-)	3 L	A	4	Y	Spherical, icosahedral arrangement of glycoproteins
<i>Bunyavirales</i>	<i>Phenuiviridae:</i> <i>Phasivirus</i>	ssRNA(-)	3 L	A	3	Y	Spherical, icosahedral arrangement of glycoproteins
<i>Bunyavirales</i>	<i>Phenuiviridae:</i> <i>Phlebovirus</i>	ssRNA(+ / -)	3 L	A	10	Y	Spherical, icosahedral arrangement of glycoproteins

(Continued)

TABLE 2.1 (Continued)

Order	Viral Family	Genome Type	Genomic Molecule(s)	Host	Species	Env.	Virion Morphology
<i>Bunyvirales</i>	<i>Phenuiviridae:</i> <i>Tenuivirus</i>	ssRNA(-)	4–5 C	P	7	N	Helical nucleocapsid (also they are circular)
Unassigned	<i>Phycodnaviridae</i>	dsDNA	1 L	A-C-P	33	Y	Icosahedral
Unassigned	<i>Picobirnaviridae</i>	dsRNA	2 L	A	2	N	Icosahedral
<i>Picornavirales</i>	<i>Picornaviridae</i>	ssRNA(+)	1 L	A	80	N	Icosahedral
Unassigned	<i>Plasmaviridae</i>	dsDNA	1 C	B	1	Y	Quasispherical to pleomorphic
Unassigned	<i>Pleolipoviridae</i>	dsDNA; ssDNA	1 C	Ar	5	Y	Pleomorphic
Unassigned	<i>Pleolipoviridae</i>	dsDNA	1 C	Ar	3	Y	Pleomorphic
<i>Mononegavirales</i>	<i>Pneumoviridae</i>	ssRNA(-)	1 L	A	5	Y	Spherical, helical nucleocapsid
<i>Caudovirales</i>	<i>Podoviridae</i>	dsDNA	1 L	B	132	N	Icosahedral head with tail
Unassigned	<i>Polydnaviridae</i>	dsDNA	1 C	A	53	Y	Prolate ellipsoid; helical nucleocapsid
Unassigned	<i>Polyomaviridae</i>	dsDNA	1 C	A	80	N	Icosahedral
Unassigned	<i>Pospiviroidae</i>	ssRNA	1 C	P	28	Vrd	Capsid-less
Unassigned	<i>Potyviridae</i>	ssRNA(+)	1 or 2 L	P	195	N	Filamentous, flexible
Unassigned	<i>Poxviridae</i>	dsDNA	1 L	A	71	Y	Pleomorphic
Unassigned	<i>Pseudoviridae</i>	ssRNA-RT	1 L	A-F-P-Pr	34	N	Spherical (Isometric to quasiisometric)
Unassigned	<i>Quadriviridae</i>	dsRNA	4 L	F	1	N	Icosahedral
Unassigned	<i>Reoviridae</i>	dsRNA	9–12 L	A-F-P	90	N	Icosahedral, double layered

Unassigned	<i>Retroviridae</i>	ssRNA-RT	1 L	A	55	Y	Spherical to pleomorphic
<i>Mononegavirales</i>	<i>Rhabdoviridae</i>	ssRNA(-)	1 L	A-P	131	Y	Bullet-shaped or Bacilliform; helical nucleocapsid
<i>Nidovirales</i>	<i>Roniviridae</i>	ssRNA(+)	1 L	A	1	Y	Bacilliform, helical nucleocapsid
<i>Ligamenvirales</i>	<i>Rudiviridae</i>	dsDNA	1 L	Ar	3	N	Rod-shaped
Unassigned	<i>Sarothroviridae</i>	ssRNA(+)	1 L	A	1	Sat	Icosahedral
<i>Picornavirales</i>	<i>Secoviridae</i>	ssRNA(+)	1 or 2 L	P	81	N	Icosahedral
<i>Caudovirales</i>	<i>Siphoviridae</i>	dsDNA	1 L	Ar-B	553	N	Icosahedral head with tail
Unassigned	<i>Soliniviridae</i>	ssRNA(+)	1 L	A	2	N	Spherical
Unassigned	<i>Sphaerolipoviridae</i>	dsDNA	1 L or 1 C	Ar-B	7	N	Icosahedral
Unassigned	<i>Spiraviridae</i>	ssDNA(+)	1 C	Ar	1	N	Coil-shaped, helical nucleoprotein
<i>Mononegavirales</i>	<i>Sunviridae</i>	ssRNA(-)	1 L	A	1	Y	Helical nucleocapsid (?)
Unassigned	<i>Tectiviridae</i>	dsDNA	1 L	B	5	N	Icosahedral
Unassigned	<i>Togaviridae</i>	ssRNA(+)	1 L	A	32	Y	Spherical, icosahedral
Unassigned	<i>Toleucusatellitidae</i>	ssDNA	1 C	P	72	Sat	Icosahedral
Unassigned	<i>Tombusviridae</i>	ssRNA(+)	1 or 2 L	P	74	N	Icosahedral
<i>Bunyavirales</i>	<i>Tospoviridae</i>	ssRNA(+ / -)	3 L	P	11	Y	Spherical, nucleocapsid
Unassigned	<i>Totiviridae</i>	dsRNA	1 L	F-Pr	28	N	Icosahedral
Unassigned	<i>Tristromaviridae</i>	dsDNA	1 L	Ar	2	Y	Rod-shaped helical nucleocapsid
Unassigned	<i>Turriviridae</i>	dsDNA	1 L	Ar	2	N	Icosahedral
<i>Tymovirales</i>	<i>Tymoviridae</i>	ssRNA(+)	1 L	A-P	39	N	Icosahedral

(Continued)

TABLE 2.1 (Continued)

Order	Viral Family	Genome Type	Genomic Molecule(s)	Host	Species	Env.	Virion Morphology
Unassigned	<i>Virgaviridae</i>	ssRNA(+)	1-3 L	P	59	N	Rod-shaped
Unassigned	Unassigned: <i>Albetovirus</i>	ssRNA(+)	1 L	P	3	Sat	Icosahedral
<i>Mononegavirales</i>	Unassigned: <i>Anphevirus</i>	ssRNA(-)	1 L	A	1	Y	Helical nucleocapsid (?)
<i>Mononegavirales</i>	Unassigned: <i>Arlivirus</i>	ssRNA(-)	1 L	A	1	Y	Helical nucleocapsid (?)
Unassigned	Unassigned: <i>Aumaivirus</i>	ssRNA(+)	1 L	P	1	Sat	Icosahedral
Unassigned	Unassigned: <i>Bacilladnavirus</i>	ssDNA(+ / -)	1 C	C	1	N	Icosahedral
<i>Picornavirales</i>	Unassigned: <i>Bacillarnavirus</i>	ssRNA(+)	1 L	C	3	N	Icosahedral
Unassigned	Unassigned: <i>Blunervirus</i>	ssRNA(+)	4 L	P	1	N	Bacilliform or filamentous?
Unassigned	Unassigned: <i>Botybirnavirus</i>	dsRNA	2 L	F	1	?	Icosahedral
<i>Mononegavirales</i>	Unassigned: <i>Chengtivirus</i>	ssRNA(-)	1 L	A	1	Y	Helical nucleocapsid (?)
Unassigned	Unassigned: <i>Cilevirus</i>	ssRNA(+)	2 L	P	2	N	Bacilliform
<i>Mononegavirales</i>	Unassigned: <i>Crustavirus</i>	ssRNA(-)	1 L	A	1	Y	Helical nucleocapsid (?)
Unassigned	Unassigned: <i>Deltavirus</i>	ssRNA(-)	1 C	A	1	Y	Spherical, helical nucleocapsid
Unassigned	Unassigned: <i>Dinodnavirus</i>	dsDNA	1 C	C	1	?	Icosahedral
Unassigned	Unassigned: <i>Higreivirus</i>	ssRNA(+)	3 L	P	1	N	Bacilliform

Unassigned	Unassigned: <i>Idaeovirus</i>	ssRNA(+)	2 L	P	1	N	Icosahedral
<i>Picornavirales</i>	Unassigned: <i>Labyrnavirus</i>	ssRNA(+)	1 L	C	1	N	Icosahedral
Unassigned	Unassigned: <i>Ourmiavirus</i>	ssRNA(+)	3 L	P	3	N	Bacilliform or Icosahedral
Unassigned	Unassigned: <i>Papanivirus</i>	ssRNA(+)	1 L	P	1	Sat	Icosahedral
Unassigned	Unassigned: <i>Polemovirus</i>	ssRNA(+)	1 L	P	1	N	Icosahedral
Unassigned	Unassigned: <i>Rhizidiovirus</i>	dsDNA	1 L	F-C	1	N	Icosahedral
Unassigned	Unassigned: <i>Salterprovirus</i>	dsDNA	1 L	Ar	1	Y	Spindle shaped
Unassigned	Unassigned: <i>Sinaivirus</i>	ssRNA(+)	1 L	A	2	?	Icosahedral
Unassigned	Unassigned: <i>Sobemovirus</i>	ssRNA(+)	1 L	P	19	N	Icosahedral
Unassigned	Unassigned: <i>Tilapinevirus</i>	ssRNA(+)	10 L	A	1	Y	Spherical, Filamentous, Pleomorphic (?)
Unassigned	Unassigned: <i>Virtovirus</i>	ssRNA(+)	1 L	P	1	Sat	Icosahedral
<i>Mononegavirales</i>	Unassigned: <i>Wastrivirus</i>	ssRNA(-)	1 L	A	1	Y	Helical nucleocapsid (?)

Notes: (1) This table was organized in terms of virus family names (second column) ordered alphabetically. Since there are many genera not assigned to a particular family, they appear under the designation "Unassigned" in Column 2 after the family *Virgaviridae*. (2) Although the main criterion for listing virus families here was type of genome molecule, the families *Phenuiviridae* and *Pleolipoviridae* appear more than once since not all members of these families possess the same kind of genome molecule. (3) In Column 4 the number of genome molecules are listed followed by the kind of molecule, be linear (L) or circular (C). (4) In Column 5, we followed the proposal that all living beings can be classified into seven kingdoms: Animalia (A), Archea (Ar), Bacteria (B), Chromista (C), Fungi (F), Plantae (P) or Protozoa (Pr). We are aware that members in some virus families use as hosts a much defined group of species from a given kingdom, but for the purposes of simplicity, we decided not to be more specific than we inform here. (5) Finally we tried to be consistent with the designation of virion morphology, but we also understand that this is not an easy task given the extreme variability of virion forms. *Env.*, Envelope; *Sat*, satellite; *Vrd*, viroid.

archaeal viruses are bottle-, spindle-, or droplet-shaped. It is argued that there are at least 20 unrelated types of capsid proteins in all viruses (and hence, it might be assumed that they evolved independently on different occasions), and that most, if not all, originated from ancestral proteins of cellular organisms. A survey of the protein folds of capsid and nucleocapsid (protein associated with nucleic acids) structures reveals strong similarities in viruses infecting species of different domains of life, which is an indication of their antiquity and of the evolutionary connections among the viruses that encode them. For instance, seven major structural classes encompass more than 60% of the known virosphere. The most prevalent folds in viral capsid proteins are the single jellyroll and the double jellyroll. They can be found in more than a third of the capsid protein folds studied so far.

The unique folds of viral capsid proteins present a distinctive geometry and allow effective packaging and virion assembly. In general self-assembly guides the generation of new virions consisting of proteins coded in the virus genome that are synthesized *de novo* in and by the host. The virion can be formed by a defined number of the same protein (capsomers or nucleocapsid proteins), or by a few different proteins held together by hydrogen bonding and hydrophobic interactions. The virion can also be covered by a lipid envelope, derived from the very host. In some other cases the virion carries an internal membrane situated between the genome and the viral coat. Virions exhibit an amazing diversity in size, and range from 18 to 19 nm (*Nanoviridae*) to c. 900 nm (*Mimiviridae*), or even more (i.e., the unrecognized genera Pandoravirus and Pithovirus; the latter are more than 1000 nm in diameter). Fig. 2.1 provides a general, yet incomplete catalog of virion forms.

Icosahedral/Spherical Virions


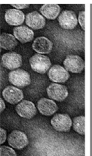

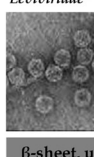
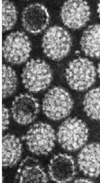
The majority of viruses adopt the form of a sphere. In this form the capsid is made up of 60 replicas of the same protein(s) that interact spontaneously through template domains at their edges, resulting in an icosahedral structure. This is the simplest icosahedral capsid with a diameter of 20–25 nm. The capsid is composed of 12 pentameric capsomeres, i.e., a total of 60 capsomeres are required to complete the capsid. In larger viruses, more than 60 capsomeres are required to completely encapsidate the genome. The icosahedral capsid of the *Rubella virus*, e.g., is composed of 12 pentameric and 30 hexameric capsomeres for a total of 240 capsomeres. The diameter of these virions is in the range of 65–70 nm. Some viruses are easily recognizable by the unique arrangements of their capsomeres at the vertices of the icosahedron. For example, virions of members in the family *Hepeviridae* show a slightly lumpy daisy shape, those of *Caliciviridae* cup-like indentations, while those of *Astroviridae* display a remarkable star shape. Other viruses exhibit a multilayered protein capsid morphology, among these

(A)

1

§ Nonenveloped virions §

Icosahedral

<i>Albetovirus</i>	<i>Luteoviridae</i>	<i>Adenoviridae</i>	<i>Botybirnavirus</i>	<i>Pseudoviridae*</i> , retro-like
<i>Alphatetraviridae</i>	<i>Marnaviridae</i>	<i>Corticoviridae*</i>	<i>Chrysoviridae</i>	
<i>Alvernnaviridae</i>	<i>Microviridae</i>	<i>Lavidaviridae</i>	<i>Partitiviridae</i> (PBL)	α-helix
<i>Anelloviridae</i>	<i>Nodaviridae</i>	<i>Marseilleviridae</i>	<i>Picobirnaviridae</i> (PBL)	
<i>Astroviridae</i>	<i>Papillomavirus</i>	<i>Sphaerolipoviridae</i> , single X2	<i>Totiviridae*</i>	Icosahedral with tails
<i>Aumavirus</i>	<i>Papillomaviridae*</i>	<i>Tectiviridae</i>		
<i>Bacilladnavirus</i>	<i>Parvoviridae</i>	<i>Turriviridae</i>	Reo-like	<i>Myoviridae*</i>
<i>Bidnaviridae</i>	<i>Permutotetraviridae</i>			Jellyroll, double
<i>Birnaviridae</i>	<i>Picomaviridae</i>	<i>Leviviridae*</i>	β-sheet, unique	
<i>Caliciviridae</i>	<i>Polemovirus</i>			Unknown
<i>Caliciviridae</i>	<i>Secoviridae</i>	<i>β-sheet, unique</i>	Unknown	
<i>Carmotetraviridae</i>	<i>Sinainvirus</i>			Unknown
<i>Circoviridae</i>	<i>Sobemovirus</i>		Unknown	
<i>Dicistroviridae</i>	<i>Tombusviridae</i>			Unknown
<i>Geminiviridae*</i>	<i>Tymoviridae</i>		Unknown	
<i>Hepeviridae</i>	<i>Virtovirus</i>			Unknown
<i>Iflaviridae</i>			Unknown	
<i>Labymavirus</i>				Unknown
	Twined icosahedra			
	Jellyroll, single			

(B)

2

§ Nonenveloped virions §

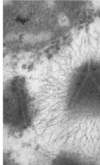


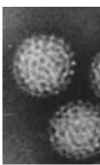


Icosahedral	Filamentous	Others
Pseudo icosahedral <i>Mimiviridae*</i>	Flexible <i>Alphaflexiviridae</i> <i>Betaflexiviridae</i> <i>Closteroviridae</i> <i>Gammaplexiviridae</i> <i>Potyviridae*</i>	Rigid rods <i>Benyviridae</i> (TMV-like) <i>Rudiviridae</i> (SIRV2-like) <i>Virgaviridae*</i> (TMV-like)
		
Jellyroll, double	Phlebo NC-like	α-helix
Double layered <i>Reoviridae*</i>	<i>Inoviridae*</i>	Bacilliform
		Barnaviridae <i>Ourniavirus*</i>
Reo-like	α-helix	<i>Cilevirus*</i> <i>Clavaviridae</i> <i>Higrevirus</i>
	<i>Ophiioviridae*</i>	
	Unknown	Unknown
		Jellyroll, single

FIGURE 2.1 Morphological variability among viruses. In representing the diversity in virion forms, viruses were first grouped according to the possession, or not, of an envelope. Then, hierarchical groups were formed following the most accepted criterion of form (*spherical/icosahedral, helical/bacilliform/filamentous, etc.*) and where known, the prevalent structural core motif among the capsid proteins (*jellyroll, α helix, β sheet, etc.*). The latter is indicated below or at the right side of the photographs; when a particular fold belongs to a class not yet defined, the word “unknown” is used. In instances where another protein fold is also found within the group, this information is added beside the family or genus name. For example, filamentous rigid rods of
(Continued)

are species belonging to the family *Reoviridae*. Here, the capsids consist of 1–3 layers of protein surrounding the genome core of 9–12 dsRNA segments. Quite a few viruses also have fiber-like structures extending from their surfaces. Adenoviruses have long fibers emanating from its vertices, Sputnik-like virions (e.g., *Adenoviridae*) have short fibers stemming from every capsomer, and mimiviruses possess a forest of fibers. These virions that lack an envelope are referred to as naked or nonenveloped viruses.

Many spherical viruses also have an outer envelope. That is, the capsid of these viruses is surrounded by a lipid bilayer that is derived from host cell membranes and contains viral proteins. These enveloped virions, by virtue of the host-derived lipid bilayer surrounding the virion, present particularities derived from the physical presence of the envelope. Arboviruses (viruses borne by arthropods like those belonging to the families *Flaviviridae*, *Togaviridae*, or *Phenuiviridae*, among others) present as fuzzy spheres possessing a peripheral fringe. HIV-1 carries prominent projections because of viral glycoprotein spikes that protrude from its envelope. The latter virus and other members of the *Retroviridae* are also described as pleomorphic. Pleomorphic viruses do not follow the rules of symmetry because the lipid envelope readily adopts different shapes and sizes, making each virion unique. Their virions are known to adopt spherical or polygonal structures. The influenza virion is an example. The virions generally take on an irregular shape or may appear spherical or elliptical in shape, ranging from approximately 80–120 nm in diameter, and are occasionally filamentous, reaching more than 20 μm in length. Each copy of its eight genomic ssRNA (-) is folded into a rod-shaped, helical nucleocapsid complex referred to as **ribonucleoprotein complex (RNP)**. Along with multiple copies of the viral-encoded **nucleoprotein (NP)**, each RNP contains a heterotrimeric viral polymerase

◀ members belonging to families grouped under nonenveloped viruses (Part 2 of the figure) may present coat proteins with α -helical motifs, that may be TMV-like or SIRV2-like. An asterisk indicates the specific taxon (family or genus) chosen to represent the morphology of the group. Unknown (*not in a gray box*) is also used when the presence or absence of an envelope, virion shape or kind of fold has not been reported for a particular family or genus of viruses. Part 6 of the figure presents viruses that do not possess a capsid or viruses for which limited information is available; i.e., families or genera with no further information in terms of form or kind of protein fold, the recently described unique group of viruses (e.g., *Tristromaviridae*), and a genus that is not fully accepted yet by ICTV (Faustovirus). *JR*, jellyroll; *NC*, nucleocapsid; *PBL*, picobirna-like. Source: Andrew King, Elliot Lefkowitz, Michael Adams and Eric Carstens. Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier 2012. *Apart from the pending genus Faustovirus, taxa in the groups represented in Part 6 are not described in Krupovic and Koonin (2017) <https://doi.org/10.1073/pnas.1621061114>.*

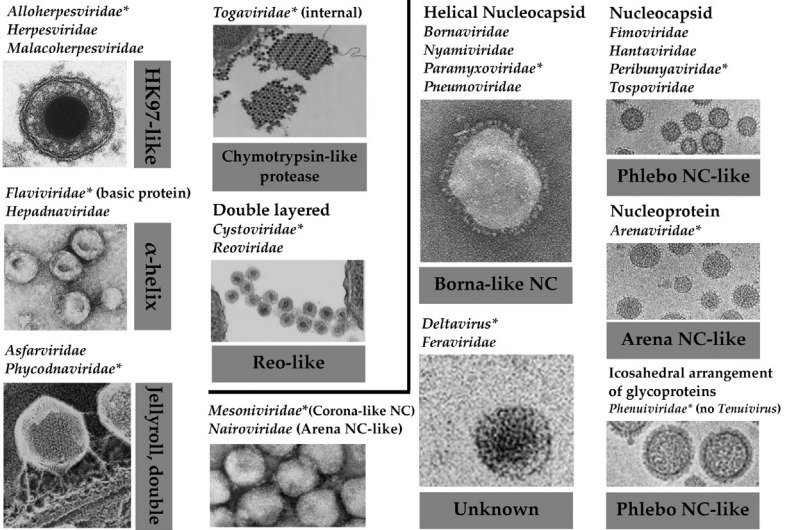
(C)

3

§ Enveloped virions §

Icosahedral

Spherical



(D)

4

§ Enveloped virions §

Spherical

Bacilliform

Helical NC

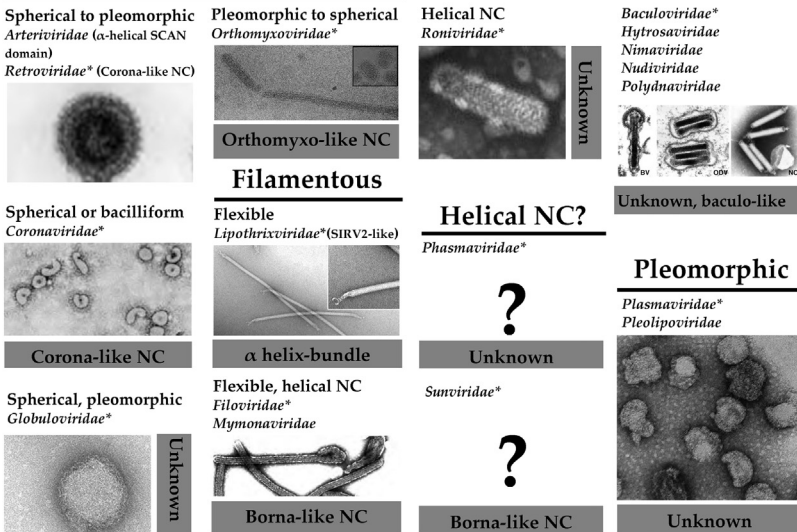
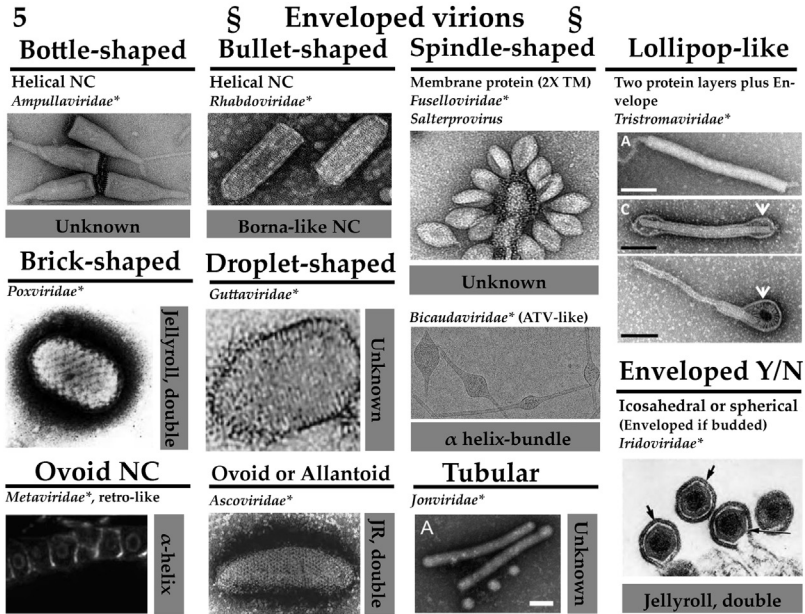


FIGURE 2.1 Continued

(E)



(F)

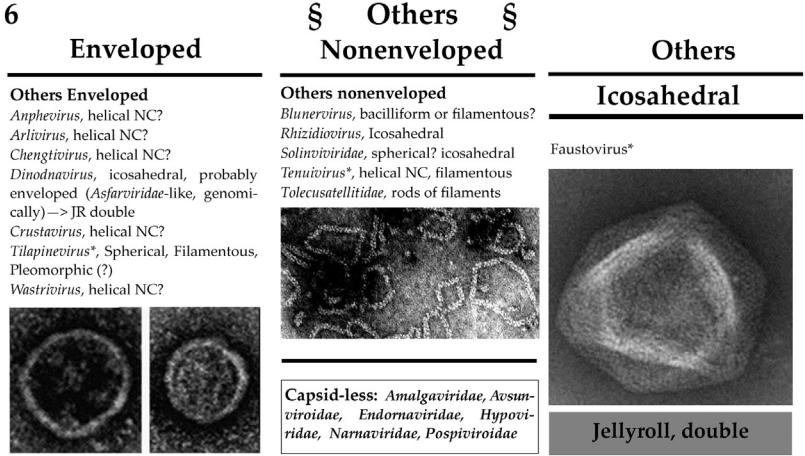


FIGURE 2.1 Continued

(consisting of PA, PB1, and PB2), and are themselves packaged in a lipid envelope that is derived from the host cell surface membrane. Two glycoproteins encoded by the virus are found in the envelope: hemagglutinin (HA) and neuraminidase (NA).

The most complex architecture of all spherical virions is shown by viruses of the families *Myoviridae*, *Podoviridae* and *Siphoviridae*, of the order *Caudovirales*, which infect bacteria and archaea. Some species of these three families show a head-tail structure where the icosahedral “head” is connected to a specialized host cell attachment structure, the tail, that may be contractile (*Myoviridae*) or not (the other two families). Three or four subterminal fibers, but more frequently six, may be present on these virions. Members belonging to the *Siphoviridae*, however, possess a noncontractile tail and never tail fibers. Virions with this type of structure are never enveloped.

Filamentous/Bacilliform Virions

The other simple way of constructing a virion is by making use of helical symmetry, in which the protein subunits and the nucleic acid are arranged in a helix. This gives a filament with flexibility that depends on the strength of the protein–protein interactions. That is, some filaments are rigid because of strong bonds between the protein monomers in successive turns of the helix (for instance *Tobacco mosaic virus*, *Virgaviridae*), while others are flexuous (*Potyviridae*) due to weaker bonds. Some filaments are so flexible that they can assume very different shapes. In *Filoviridae*, e.g., enveloped virions may be simple cylinders (the less frequent form) or form branches or loops. The loops can also vary in form and include the thread-like form (the most general morphotype) as well as U-, 9-, eye-bolt- or Shepperd’s crook-shapes, among others. In the case of the family *Ophioviridae*, the highly flexuous, nonenveloped virions can form open circles.

Other virions with helical symmetry have a wider diameter so they look a little thicker than the typical filamentous virions. These are bacilliform (hemispherical at both ends) or bullet-shaped (with one rounded and one flattened end) filaments. They are more frequently found in plant and animal viruses (*Caulimoviridae*, e.g., with hosts in both kingdoms), but some archaeal virions (*Clavaviridae*) are also bacilliform.

Amazing Forms Among Old and New Comers

Viruses that exist in extreme environments display a wide range in morphological variations that are not observed with those entities found in less harsh habitats. A filamentous dsDNA virus that infects members of the archaeal genus *Pyrobaculum*, e.g., has its linear genome enclosed in a tripartite shell consisting of two protein layers and an envelope, which is unusual for dsDNA viruses. Other structures found with archaeal viruses include, as mentioned earlier, bottle-, lemon-, droplet-, and spindle-shapes. Virions of unusual forms are not exclusive to archaea, though, since bullet-shaped virions can be found in rhabdoviruses. Ovoid forms, on the other hand, are

common among viruses belonging to the families *Ascoviridae* (which can also have a bacilliform or allantoid forms), *Metaviridae*, *Nimaviridae* (which also show a tail-like appendage at one end), and *Poxviridae* (which can also be brick shaped). Additionally, other virions (Pandoraviruses) resemble the shape of an amphora (Greek vase), have a prolate ellipsoid form (*Polydnaviridae*) or are twined icosahedra (*Geminiviridae*).

DO VIRUSES REALLY NEED A COAT?

The vast majority of viruses possess a capsid and form virions. The capsid plays multiple roles in the infection cycle of viruses:

1. A coat provides the virus genome protection against environmentally damaging agents. The virion serves as a physical barrier for the genome and proteins it may harbor. Indeed the structural integrity of the capsid is essential for viral replication and the expression of viral genes.
2. The coat facilitates interaction with host cells in terms of recognition, processing (mostly uncoating) and tropism—after host recognition, a virus displays its full biological potential in specific tissues of the host.
3. Coat proteins assist in virus movement between cells in multicellular organisms, thus allowing for systemic infection of the host.
4. Coat proteins, in addition, facilitate vector-mediated transmission of viruses between organisms. In the vector, coat proteins along with other viral proteins, allow for retention of virions to vector mouth parts, movement to other parts of the vector body, replication, and sometimes the production of more virions.

Nonetheless, viruses appear to have evolved from capsid-less selfish elements. We know of viruses that never have (nor code for) a capsid. For instance the families *Endornaviridae*, *Hypoviridae*, and *Narnaviridae* are comprised of members where a capsid is absent. Interestingly, members of the family *Amalgaviridae* encode for but do not form virions. The likely existence of capsid-less ancestors, along with the existence of capsid-less viruses today, suggests that the viral capsid was a late acquisition during their evolutionary history. Further, it appears that coat proteins not only accommodated the formation of capsids and their associated functions, but their acquisition seemingly widened opportunities for viruses to become the most numerous, conspicuous, and highly evolving biological entities known.

TYPES AND FUNCTION OF THE VIRUS ENVELOPE

As alluded to earlier, slightly less than a quarter of all viruses possess an envelope surrounding their protein capsid. The envelope typically consists of

a host-derived lipid bilayer membrane plus glycoproteins of viral origin. Its structure varies in terms of size, composition, morphology, and complexity. Viral-derived glycoproteins are embedded in the lipid bilayer and in some cases, other nonglycosylated proteins of viral origin form part of the envelope. The number of virus glycoproteins varies among viral groups; in members of the family *Herpesviridae*, e.g., more than 10 glycoproteins have been identified. In simpler cases (e.g., members of *Togaviridae* and *Orthomyxoviridae*), there are one or two multimeric proteins. Other viral-encoded glycoproteins are the ion channel proteins, viroporins. Viroporins have at least one transmembrane domain and sometimes an extracellular membrane region that interacts with viral or host proteins. Viroporins have been discovered in *Influenza A virus*, *Hepacivirus C* (previously known as Hepatitis C virus), *Human immunodeficiency virus 1*, and Coronaviruses. Enveloped viruses may also contain a layer of protein between the envelope and the capsid, known as the matrix (*Orthomyxoviridae*, *Retroviridae*), or there is no such layer and the capsid interacts directly with the internal tails of the membrane proteins (*Togaviridae*, *Bunyavirales*).

Recognition, attachment, and entry into cells through fusion with host cell membranes are probably the main roles played by the envelope and associated proteins. The envelope helps viruses to evade the mammalian host's immune system. Additionally, from an evolutionary point of view, virus envelope proteins, with variations, are well represented in vertebrate genomes and apparently were a potent force in placentation in different species. The syncytin genes represent a dramatic example of convergent evolution via the cooption of a retroviral envelope gene for a key biological function in placental morphogenesis.

GENOMES: THE NUCLEIC ACIDS SPACE EXPLOITED BY VIRUSES

All cellular organisms possess only one type of genomic molecule: double stranded DNA, along with a myriad of informational molecules represented by different types of RNA, and instances of ssDNAs. Viruses, on the other hand, have exploited the genomic space to its limits. Either DNA or RNA is used by viruses to encode their genomes, with many variations. In the case of RNA viruses, the genome can be single stranded in minus (–) or plus (+) orientations, or a combination of both (+/–), or double stranded. The vast majority of viruses have RNA genomes. When it comes to DNA genomes, more than a third of all recognized virus species possess a dsDNA genome. DNA viruses can also have ssDNA(–) or ssDNA(+) molecules; ssDNA (+/–) viruses also exist. In some cases the virus genome is only partially double stranded and partially single stranded (*Pleolipoviridae*), while other dsDNA genomes are described as open circular dsDNA due to nicks at

specific sites along the DNA (*Caulimoviridae* and others). In general, viruses' genomes are either in linear or circular conformations.

The genome of a virus exists in the form of one or more molecules of nucleic acid. The most common number of viral genomic molecule is one (monopartite viruses). Other viruses are bipartite or tripartite, and yet others are so complex that their complete genome is represented by up to 12 different molecules. Although one genome molecule per virion is the “rule,” some viruses can carry all the genomic molecules in one particle while other viruses have their segmented genomes portioned in different particles (*Nanoviridae*); yet others encapsidate replicas of the same genome in a single particle (e.g., virus species of the family *Polydnaviridae*). A very important family of viruses, *Geminiviridae*, consists of two incomplete icosahedra joined together to form twinned particles; they are monopartite or bipartite. Monopartite geminiviruses encode all their genes on a single circular ssDNA. Bipartite geminiviruses (genus *Begomovirus*), on the other hand, consist of two virions containing different genomic components that are required for productive infection (i.e., genes are distributed on two separate ssDNA molecules that are packaged separately). Additional details on the categorization of virus' nucleic acids are given in the following section. A summary of the virosphere, in terms of nucleic acids, is provided in [Table 2.2](#), and in [Fig. 2.2](#).

dsDNA viruses

Viruses with a dsDNA genome amount to more than a third (38.6%) of all recognized viruses. These genomes can be found as circular molecules (25.3%) or as single linear molecules (74.3%). Typically, viral genomic DNA of members of a dsDNA viral family all have the same topological form. However, species of the family *Sphaerolipoviridae* can have either a circular or a linear genome. Some members of the family *Pleolipoviridae* have a monopartite circular genome, while others have a bipartite genome composed of a circular dsDNA and another which is circular, but single stranded. In both cases the dsDNA genomic molecules is interrupted by short runs of ss linear DNA. Members of the family *Caulimoviridae* possess a monopartite, circular dsDNA genome with gaps (discontinuities) in each strand of the genomic molecule. Specific nicks can be found in the transcribed strand, along with 1–3 in the nontranscribed strand.

Roughly 77% of viruses with a dsDNA genome are nonenveloped. Of note, members of the family *Iridoviridae* can be enveloped or not, depending on their route of exit from the host cell (budding or lysis). Other dsDNA viruses may possess a lipid membrane between the coat and the genome (*Turriviridae*). Although dsDNA viruses can infect members of all kingdoms

TABLE 2.2 Summary of Virus Families (and Unassigned Genera) Based on Genomic Molecule (Regardless Strategies of Expression), 2017

Genome	Type: (#)	Host(s)	Env: (#)	Species
dsDNA, including dsDNA-RT viruses (74 species)	1 L: 1265	A, Ar, B, C, F, P, Pr	Y: 376	38 families (2 RT)
	1 C: 422		N: 1305	3 unassigned genera
	1 C or 1 L: 7		Y/N: 12	
			?: 1	1694 species
ssDNA, including a family of satellites	2 L: 1	A, P	N: 1	2 families
	1 C: 72		Sat: 72	73 species
ssDNA(+)	1 C: 53	Ar, B, P	N: 65	4 families
	6 or 8 C: 12			65 species
ssDNA(-)	1 C: 68	A	N: 68	1 family 68 species
ssDNA(+/-)	1 L: 62	A, C, F, P	N: 575	4 families
	1 C: 144			1 unassigned genus
	1-2 C: 369			575 species
dsRNA	1 L: 54	A, B, C, F, P, Pr	Y: 1	11 families
	2 L: 70		N: 197	1 unassigned genus
	3 L: 1		NA: 26	
	4 L: 10		?: 1	
	9-12 L: 90			225 species
ssRNA	1 C: 32	P	NA: 32	2 viroid families 32 species
ssRNA(-), excluding the members of the family <i>Phenuiviridae</i> (<i>Bunyavirales</i>)	1 L: 212	A, F, P	Y: 344	17 families
	3 L: 113		N: 7	6 unassigned genera
	3 or 4 L: 7			
	4-8 L: 9			
	6-8 L: 9			
	1 C: 1			

(Continued)

TABLE 2.2 (Continued)

Genome	Type: (#)	Host(s)	Env: (#)	Species
ssRNA(+), including ssRNA-RT viruses (120 species)	1 L: 723	A, B, C, F, P, Pr	Y: 273	39 families (3 RT)
	2 L: 12		N: 949	15 unassigned genera
	1 L or 2 L: 360		NA: 11	
	3 L: 37		Sat: 7	
	1–3 L: 108		?: 2	
	4 L: 1			
	10 L: 1			1242 species
ssRNA(+/-)	2 L: 36	A, P	Y: 47	2 families
Excluding <i>Phenuiviridae</i> (+/-)	3 L: 11			47 species
Mixed				
<i>Pleolipoviridae</i>				1 family
dsDNA	1 C: 3	Ar	Y:3	
<i>Pleolipoviridae</i>				
dsDNA; ssDNA	1 C: 5	Ar	Y: 5	8 species
<i>Phenuiviridae</i>				1 family
ssRNA(-)	3 L: 7	A, P	N: 7	
	4–5 C: 7		Y: 7	
<i>Phenuiviridae</i>				
ssRNA(+/-)	3 L: 10	A	Y: 10	24 species
Summary	Linear: 3209	A, Ar, B, C, F, P, Pr	N: 3174	122 families
DNA: 2483	1 L: 2316		Y: 1066	26 unassigned genera
dsDNA: 1702 ^a	2 L: 119		Y/N: 12	
ssDNA: 73	3 L: 179		NA: 69	
ssDNA(+): 65	1 or 2 L: 360		Sat: 79	
ssDNA(-): 68	1–3 L: 108		?: 4	
ssDNA(+/-): 575	4 L: 11			
	3 or 4 L: 7			

(Continued)

TABLE 2.2 (Continued)

Genome	Type: (#)	Host(s)	Env: (#)	Species
RNA: 1921	4–8 L: 9			
dsRNA: 225	6–8 L: 9			
ssRNA: 32	9–12 L: 90			
ssRNA(+): 1242 ^b	10 L: 1			
ssRNA(-): 365				
ssRNA(+/-): 57	Circular: 1188			
	1 C: 800			
	1 or 2 C: 369			
	4–5 C: 7			
	6 or 8 C: 12			
	1 L or C: 7			4404 species

Viral genomes can be linear (L) or circular (C); their hosts are coded here as A (Animalia), Ar (Archaea), B (Bacteria), C (Chromista), F (Fungi), P (Plantae), or Pr (Protozoa). Virions can be enveloped (Y) or not (N); when information is not available a “?” is used. NA, not applicable; Sat, satellite. Number of species (#) are indicated in each case, when appropriate.

^aIncluding all members of the family Pleolipoviridae, as well as all dsDNA-RT virus species.

^bIncluding all dsRNA-RT virus species.

of life, the vast majority are limited to species of Bacteria or Archaea (more than 60%), while roughly another third infect members of the kingdom Animalia. Virions of dsDNA viruses, particularly those able to infect Archaea, display extreme variability of forms. About half of these viruses produce virions with heads and tails, icosahedral viruses are also common, as are bacilliform, spherical, prolate ellipsoid as well as droplet-, lemon-, rod-, and spindle-shaped virions. Most dsDNA viruses are in the B form of the nucleic acid, and there is at least one report of an A form in a virus that infects a hyperthermophilic acidophile Archaea (*Sulfolobus islandicus rod-shaped virus 2*, family *Rudiviridae*).

Of note, among the 38 families and three unassigned genera of dsDNA viruses, members of the *Caulimoviridae* and *Hepadnaviridae* families are the only ones that generate new genome copies by retrotranscription (DNA to RNA to DNA) instead of canonical DNA replication (DNA to DNA).

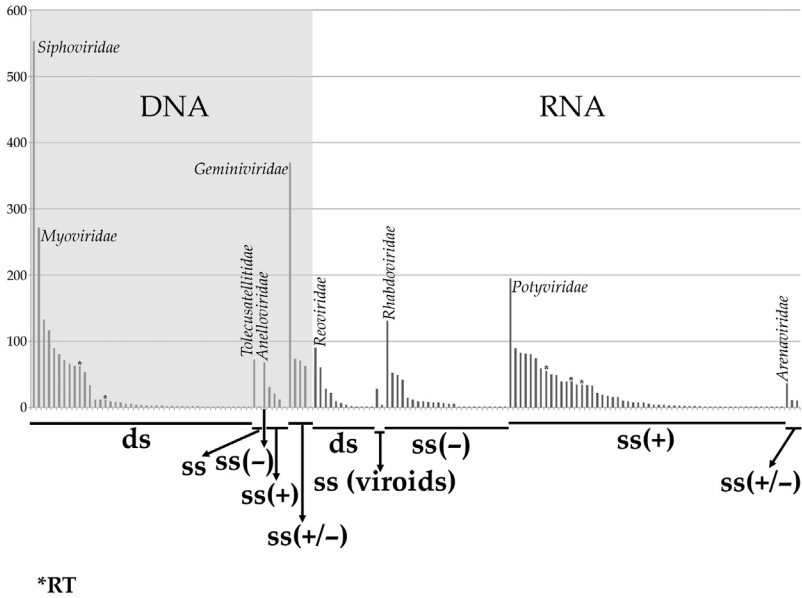


FIGURE 2.2 Schematic representation of virus families and unassigned genera as of April 2017. Number of species is indicated on the left, and all families and genera are presented according to the nature of their genome, DNA (gray) or RNA (white). In each case, families and genera have been subdivided in terms of the molecular type [ds, ss, ss(+), s(-), and ss(+/-)] and names of the largest families for each molecular type are indicated. The asterisk represents virus families that use a retrotranscribing (RT) mechanism. The figure has been drawn to scale.

ssDNA viruses—No Polarity or Mixed Polarity

These particular ssDNA viral entities belong to either one of two families: the newly proposed *Tolecusatellitidae* family that groups 72 satellite viruses—which are always associated with geminiviruses, and the monospecific family *Bidnaviridae* whose only member infects silkworms causing silkworm flacherie disease. Members of this class of viruses can have a bipartite, linear genome (bidnavirus), or a monopartite, circular genome (tolecusatellitids). Their virions are nonenveloped icosahedra.

Members of the family *Tolecusatellitidae* are satellites (i.e., they are non-independent entities) that rely on a helper virus for their spread (a begomovirus or a mastrevirus from the family *Geminiviridae*). On the other hand, the only member of the bidnaviruses, *Bombyx mori bidensovirus*, the bipartite, linear genome is composed of segments VD1 and VD2 that are encapsidated separately. Since equal amounts of positive (+) and negative (-) molecules are encapsidated, four different types of full particles are produced: VD1 + /VD2 +, VD1 + /VD2 -, VD1 - /VD2 +, and VD1 - /VD2 -. This bidnavirus is not related to tolecusatellitids; it is listed here in

order to emphasize the ss nature of its genome. ssDNA viruses convert into a double stranded form prior to transcription to mRNAs.

ssDNA(–) viruses

The only family of ssDNA(–) viruses is the family *Anelloviridae* that is comprised by 68 species. Anelloviruses are able to infect diverse groups of vertebrates, and are sometimes asymptomatic in humans. They are associated with various diseases that include hepatitis, lupus, and miopathy, among others. The virus can be transmitted sexually and by human excreta. The genome, represented by a single circular molecule of ssDNA of negative polarity, is protected in nonenveloped icosahedral virions. It is produced through dsDNA intermediates and contains four potential ORFs. Proteins are expressed by alternative splicing of a single premRNA.

ssDNA(+) viruses

Viruses with a ssDNA(+) genome are hosted mainly by bacteria (2 families, *Inoviridae* and *Microviridae*, encompassing 52 species), plants (one family, *Nanoviridae*, with 12 species) and archaea (with one monospecific family, *Spiraviridae*). The genome is circular, but the number of genomic molecules varies from 1 (in 53 species) to 6–8 (only for the nanoviruses). In nanoviruses, each genomic molecule is monocistronic, but in rare cases one gene can carry two overlapping ORFs. Each genomic molecule is encapsidated separately.

ssDNA(+/–) viruses

This group presents a different scenario than that explained for the bidnaviruses in the section on *ssDNA viruses*. In that case an assortment of ssDNA molecules gives rise to genomic combinations of + or – strands. In the case of ssDNA(+/–) viruses the genome is ambisense. That is, in the same, or different molecules, the genome is composed of DNA sequences coding in opposite directions. Most of these viruses are monopartite, but geminiviruses can have a monopartite or bipartite genomes. The genome is circular, except for members of the *Parvoviridae* family of animal viruses, which present a linear genome. Four families and one unassigned genus group all viruses with an ambisense ssDNA genome. This represents slightly more than 13% of all recognized viruses. Virions are icosahedral and nonenveloped. ssDNA(+/–) viruses are hosted only by eukaryotes, with the exception of members of the kingdom Protozoa at the time of this writing.

dsRNA viruses

All viruses with a dsRNA genome possess a linear molecule in variable numbers: 1, 2, 3, 4—or even between 9 and 12 (*Reoviridae*). Two unusual features of endornaviruses, not found in any other RNA viruses, is that their genome codes for an unusually long, single ORF and that the genomic molecule possesses a site-specific nick in the 5' region of the coding strand. Although all dsRNA viruses are nonenveloped, it is also true that not all seem to produce virions. For example, members of the *Amalgaviridae* capsid-less family of viruses do code for a coat protein. Amazingly, however, the RdRp of amalgaviruses is closely related to the same protein belonging to members of another family of dsRNA viruses (*Partitiviridae*), while the CP is homologous to the nucleocapsid proteins of ssRNA(−) viruses of the genera *Phlebovirus* and *Tenuivirus* (family *Phenuiviridae*). That is, the genome of amalgaviruses is chimeric (different origins), and probably evolved by recombination between a partitivirus and a tenuivirus. Amalgaviruses are transmitted by seeds and cannot be mechanically transmitted between plants. The other dsRNA virus family whose members do not produce virions is *Endornaviridae*. Members of this family infect chromists, fungi, and plants.

The only noneukaryote dsRNA virus is the tripartite *Pseudomonas virus phi6*, the lone species of the family *Cystoviridae*. It is also the only dsRNA virus with enveloped virions consisting of double layered icosahedrons. The genomic dsRNAs of this virus are never exposed in the cytoplasm. Instead they remain enclosed in the inner capsid where they are transcribed by the viral RdRp to create the corresponding mRNA for translation (ssRNA(+) molecules). Generation of new dsRNA occurs within assembled progeny capsids after the three different genomic sRNA(+) segments are translocated from the original capsids.

ssRNA (Viroids)

Members in this group are all viroids classified into two families; *Avsunviroidae* (four species) and *Pospiviroidae* (28 species). All infect only plant hosts. Viroids possess a circular genome that is not encapsidated in a protein coat. Biologically speaking, they behave like viruses, but their genomes do not code for proteins. Viroids depend on proteins of their hosts for their replication (via rolling circle model) and other functions. Curiously, viroid genomes of the family *Avsunviroidae* share some organization and replication characteristics with the human ssRNA(−) *Hepatitis delta virus*, from the unassigned, monospecific genus *Deltavirus*. Presumably, this virus and viroids have a common ancestor. It is not unreasonable to expect the discovery of more of these virus/viroids entities in the future, at least in the kingdoms Animalia and Plantae.

ssRNA(–) viruses

Viruses with a ssRNA(–) genome, classified into 18 families and 6 unassigned genera, include 365 viruses hosted almost exclusively by plants (two families, and all members of the genus *Tenuivirus* from the family *Phenuiviridae*) or animals (13 families, six unassigned genera, and the genera, *Goukovirus* and *Phasivirus* from the family *Phenuiviridae*). Notably the *Rhabdoviridae* is one of the most ecologically diverse families of RNA viruses, with 131 members identified in a range of plants and animals, including mammals, birds, reptiles, and fishes. The only exception to this two kingdom restricted host range is the virus *Sclerotinia sclerotimonavirus* which is able to infect a hypovirulent isolate of the fungus *Sclerotinia sclerotiorum*, and is classified in the monospecific family *Mymonaviridae*.

The vast majority of ssRNA(–) viruses possess a linear genome. Exceptions are found with members of the genus *Tenuivirus* (family *Phenuiviridae*) where there are 4–5 circular genomic molecules, and the only member of the genus *Deltavirus* that has a monopartite circular genome. Those with a linear genome can be monopartite, bipartite, tripartite, while some viruses have their genome proportioned into four, 4–8 or 6–8 molecules. Prior to translation, the genome of ssRNA(–) viruses is copied into the plus strand by the virus-encoded RdRp. Apart from members of the *Ophioviridae* family of plant viruses, ssRNA(–) virions are enveloped.

Negative-sense RNA viruses are the etiological agents of diseases in humans like rabies, influenza, hemorrhagic fever, and encephalitis. It has been hypothesized that arthropods host many (perhaps all) ssRNA(–) viruses that cause disease in plants and animals. This is because similar variation, in terms of number and type of genomic molecules, in negative-sense RNA viruses are also found in arthropods. We might expect that the study of these viruses in invertebrates will shed some light on the origin and radiation of this peculiar and somewhat specialized group of viruses in the future. Arthropods are a major reservoir of viral genetic diversity and it is quite possible that they were central to the evolution ssRNA(–) viruses.

ssRNA(+) viruses

After dsDNA viruses the positive-sense viruses are the most common in the virosphere—and for good reason. ssRNA(+) viruses, after reaching the cytoplasm, can be immediately translated giving rise to the proteins encoded in the viral genome, and serve, as well, as the template for the generation of more copies of the virus genome. From the point of view of economy and minimalism, there is no better option. ssRNA(+) viruses are classified into 39 families and 15 unassigned genera. Among these families we can find the three retrotranscribing families *Metaviridae*, *Pseudoviridae*, and *Retroviridae*—which add a layer of complexity to the realm of simplicity.

Viruses belonging to this group can infect members of virtually all kingdoms of life—except for Archaea. Only four viruses with ssRNA(+) genomes have been reported to infect bacteria, all belonging to the family *Leviviridae*.

Not surprisingly, all positive-sense RNA viruses have a linear genome (as no messenger RNAs in cellular organisms are circular). The most numerous family of this group of viruses is the *Potyviridae*—probably the most important plant virus family after *Geminiviridae*, not only in terms of the number of species but from the point of view of food security and economic losses they incur.

ssRNA(+/-) viruses

A few viruses possess an ambisense genome. That is, the genome has all the genetic information encoded in segments either in sense or antisense orientations. Two families (*Tospoviridae* and *Arenaviridae*) and the genus *Phlebovirus* of the family *Phenuiviridae* possess 57 species with ssRNA (+/-) genomes. All form enveloped, spherical virions that encase bipartite or tripartite linear genomes. Viruses of this group only infect animal or plant species. RdRp is packaged within virions, which makes the generation of plus segments from the minus segments of the genome possible early in the replication cycle. With more than 1000 species susceptible to infection, *Tomato spotted wilt orthotospovirus* (*Tospoviridae*) is reputed to be the virus with one of the widest host ranges of all.

VIRUS TAXONOMY: ORGANIZATION OF THE VAST DIVERSITY AMONG VIRUSES

The early 1990s saw a tremendous surge in the discovery of new viruses, and the need to classify viruses (and maintain a consistent naming system) was recognized as a necessity. Various classification schemes were proposed in an attempt to bring to order the apparent diversity, and to facilitate the study of these entities. Early schemes were based solely on size. Viruses were classified as distinct from cellular organisms by virtue of their ability to pass through unglazed porcelain filters known to retain the smallest of bacteria. But as the numbers of filterable agents increased, viruses were distinguished from each other by more measurable characteristics, namely the disease or symptoms caused in an infected host. Under this scheme, e.g., animal viruses that caused hepatitis or jaundice were grouped together as hepatitis viruses and viruses that induced mottling symptoms in plant hosts were grouped together as mosaic viruses. The advent of new technologies in virus purification, serology, and electron microscopy spurred the use of physical characteristics for distinguishing viruses.

Baltimore Classification

In 1971 the classification of viruses according to the genomic nature of viruses and method by which they replicate was proposed. It represented the easiest, and yet logical, way of grouping viruses. The central theme of the proposal, referred to as the Baltimore system of virus classification, is that all viruses must synthesize positive strand mRNAs from their genomes in order to produce proteins and replicate in their hosts. The proposal gave origin to six groups, and a seventh group was later added in order to accommodate the dsDNA-RT viruses (*Caulimoviridae*, of plant and animal hosts, and *Hepadnaviridae* of only animal hosts). The seven groups are as follows:

- Group I.** dsDNA viruses, in which mRNAs are produced by direct transcription using a host RNA polymerase. These mRNAs can be produced from the genome of infecting virus (early mRNAs) or from progeny viral dsDNAs (late mRNAs). The dsDNA-RT viruses are not included in this group, but in group VII (see below).
- Group II.** ssDNA viruses regardless of the genome polarity(ies) (+, −, +/−, and + plus −) are included in this group since the production of the mRNAs involves the step of generating dsDNA first. In eukaryotes, virus replication occurs in the nucleus, most probably by a rolling circle mechanism.
- Group III.** dsRNA viruses, as explained previously, produce mRNAs and replication templates by means of their own RdRP. This almost never occurs outside of the virus capsid.
- Group IV.** ssRNA(+) viruses with genomes that are recognized immediately by the host cell machinery as mRNA. Translated proteins subsequently direct the replication of the viral genome. This messenger can be polycistronic, in which case the polyprotein resulting from translation is later cleaved by proteases encoded in the viral genome to give rise to functional viral proteins. In other cases, transcription is more complex and may involve the generation of subgenomic mRNAs, ribosomal (+ or −) frameshifting or proteolytic processing of an initial precursor translation product. Retrotranscribing ssRNA(+) viruses are not included in this group, but in group VI (see later).
- Group V.** ssRNA(−) viruses that must be copied first into a sense RNA molecule in order to produce viral mRNAs. There are two main types of ssRNA(−) depending on where virus replication occurs: those that replicate in the cytoplasm are monopartite and are transcribed by the RdRp to produce the sense molecule that can be translated and serve also as the template for the generation of the minus strand (genomic molecule). And those with segmented genomes that are replicated in the nucleus, in which

case separate mRNAs are produced by the viral RdRp resulting in both messenger and replication template molecules. ssRNA (+/−) viruses are also included in this group.

Group VI. ssRNA(+)-RT viruses that, by means of a virus-encoded reverse transcriptase (RT), generate a dsDNA copy of the genomic molecule. This dsDNA molecule is integrated into the genome of the host (becoming a provirus) where it can be replicated and transcribed in the nucleus using the host cell's machinery to provide viral mRNAs as well as ssRNA(+) genomic molecules.

Group VII. dsDNA-RT viruses that possess a gapped genome (open circular, OC, dsDNA genome) that upon infection is filled to generate a Covalently Closed Circular (CCC) form. Transcription of the CCC genome in the nucleus generates viral mRNAs and a subgenomic RNA. Using a virus-encoded RT, the subgenomic RNA is converted to the genomic OC dsDNA molecule.

Altogether, the combination of virion architecture, variations in genome type, number of genomic molecules, and variations of genomic strategies, presence or not of an envelope, host range, and other particularities, makes every virus a unique, evolved construct of nature. The virion provides much information toward the characterization of viruses and the way they interact with cellular organisms, but should not be the main emphasis in distinguishing viruses. For one thing, virion morphology in different groups may be the result of convergent evolution. That is, there is no evolutionary relatedness involved, nor shared ancestry. The demonstration of chimerism in virus genomes also adds to the unreliable use of morphology when trying to analyze virus evolution and classification since we would be comparing features of different origins (replicase and capsid proteins, for example) in the same subject of study. Importantly also is the fact that not all viruses produce virions, as explained. Virion morphology, however, is still a very useful way to characterize (and even identify) viruses since the expression of their molecular features are identifiable and measurable, and derive directly from the information the virus encodes.

The International Committee on Taxonomy of Viruses

The first internationally organized initiative for developing a universal taxonomic scheme for viruses was the formation of the International Committee on Nomenclature of Viruses in 1966, which later became the International Committee on Taxonomy of Viruses (ICTV) in 1973. The system that was developed is essentially based on the familiar systematic taxonomy scheme of Order, Family, Subfamily, Genus, and Species. Levels higher than

Order were not included as such levels imply a common ancestry; multiple independent lineages for viruses now seem the more likely scenario. As of March 2017 ICTV subdivides viruses into 8 orders, 122 families, 35 subfamilies, 735 genera, and 4404 species.

The classification scheme unites viruses using a range of similar attributes. This way viruses are grouped by comparing numerous properties of individual viruses without assigning universal priority to anyone property. Orders and families are typically assigned by virus morphology, genome composition, orientation (+/– sense), segmentation, gene sets, and replication strategies. The further subdivision of families into genera generally segregates viruses into those possessing the same complement of homologous genes. Because of the congruence with evolutionary histories, taxon assignments are often recapitulated in virus nucleotide or amino acid sequence relationships. The lowest level of classification considered by the ICTV is species. According to the ICTV in 1991, “A *virus species* is a polythetic class of viruses that constitute a replicating lineage and occupy a particular ecological niche.” The advantage of defining virus species as polythetic is particularly relevant as viruses undergo continual evolutionary changes and show considerable variability. Of note, a type species is identified for each genus and is usually the virus that necessitated the creation of the genus and best defines or identifies the genus. Some criteria used in the classification of viruses are summarized in [Table 2.3](#).

Taxonomic levels lower than species (e.g., strains, variants) are not officially considered by the ICTV, but are left to specialty groups. It is widely accepted that a strain refers to isolates of the same virus from different geographical locations. Variants refer to viruses with phenotypes that differ from that of the original wild type strain, including serotypes and pseudotypes.

Further guidelines were formulated to facilitate the development of a reasonably uniform nomenclature for all viruses. [Table 2.4](#) highlights a few of the rules. Taxon names are designated with suffixes: Order (*-virales*), Family (*-viridae*), Subfamily (*-virinae*), Genus (*-virus*). In most cases the English common names of viruses have become the species names. Generally the virus name provides information regarding the host that the virus was originally isolated from (e.g., *Escherichia virus Lambda*, *Siphoviridae*, a virus originally isolated from *Escherichia coli*, that is regarded as the mother of Molecular Biology), a symptom associated with the disease it incites (*Papaya ringspot virus*, *Potyviridae*, first isolated from symptomatic *Carica papaya* plants showing characteristic ring spots on the fruits), or the disease itself (as mentioned previously, or *Yellow fever virus*, *Flaviviridae*, the causal agent of that disease), or the place where it was found for the first time (*Lake Sinai virus 1*, from the genus *Sinaivirus* belonging to as of yet an unassigned virus family, and detected in honey bee samples from an apiary

TABLE 2.3 Some Criteria Used for the Characterization, Description, and/or Classification of Viruses

General Properties	Specific Properties
Physical (Virions)	<ul style="list-style-type: none"> ● Molecular weight, buoyant density, sedimentation values ● Physical stability under pH, temperature, ions, irradiation, detergents, organic solvents ● Viability outside the host(s) ● Number of particles
Chemical (Virions)	<ul style="list-style-type: none"> ● Macromolecular composition: Content of proteins, lipids, carbohydrates, and nucleic acids
Biochemical (Virions)	<ul style="list-style-type: none"> ● Number, size, type and interactions of structural proteins, lipids, and carbohydrates ● Presence or not of internal layers and of an external envelope ● Type of nucleic acid (RNA or DNA), strandedness (ss or ds), linear or circular, and polarity (positive or negative sense, or ambisense) ● Chemical modifications of virions components
Structure (Virions)	<ul style="list-style-type: none"> ● Number and composition of capsomers ● Form, shape, and symmetry of virions (virion morphology); ornamentation ● Size and dimensions
Genomic, genetic, and biochemical	<ul style="list-style-type: none"> ● Genome length ● Number of genomic molecules (genome partition) ● Strategies followed to produce mRNAs ● Genome sequence and/or gene sequences ● Amino acid sequence of proteins (mostly deduced from the genomic/gene nt sequence) ● Posttranslational processing and/or protein modifications ● G + C content (total and along the virus genome) ● Presence or absence and type of capping molecule at the 5' end (true cap, modified tRNA, genome-linked protein) ● Presence or absence, and length, of the genomically encoded polyA tail at the 3' end ● Presence, type, and function (if any) of repetitive sequences in the genome ● Genome organization and replication strategies ● Number of ORFs and intended product(s) and their corresponding functions ● Genome expression strategies and site of expression ● Phylogenetical relationships with related viruses and group of viruses ● Capability of genomic insertion and heritability

(Continued)

TABLE 2.3 (Continued)

General Properties	Specific Properties
Immunological	<ul style="list-style-type: none"> ● Specific reactivity with monoclonal and polyclonal antibodies of virion proteins and genome-encoded proteins ● Cross-reactivity with related viruses ● Allergenicity (if any)
Cellular and organismal	<ul style="list-style-type: none"> ● Attachment and internalization (specific receptors and pathways) ● Site of virion accumulation ● Interactions with cell components ● Tissue tropism, cytotoxicity, and tissue pathogenicity ● Number, type, and localization of inclusion bodies ● Site and mode of replication and virion assemblage ● Characteristics of cell-to-cell movement and of systemic spread ● Symptomatology (if the virus causes a disease) ● Latency time (if any)
Others	<ul style="list-style-type: none"> ● Host range ● Mode of transmission, both vertical and horizontal ● If vectored, type and identity of vectors, and mode of transmission ● Geographical distribution ● Cooccurrence and interactions with other viruses, satellites, and/or viroids

near Lake Sinai, South Dakota, United States). Additionally, the rules are in line with the way in which other formal taxon species names are written; they are written in italic script and the first word in the taxon name begins with a capital letter.

Viruses can also be referred to by acronyms, sigla (singular: siglum) or abbreviations. An acronym of a virus name is created by using the initial letter in the words of a virus species name: *Cucumber mosaic virus* (*Bromoviridae*) is widely referred as CMV. Sigla, on the other hand, are constructed using letters or other characters taken from words of a compound term, like arbovirus for **arthropod-borne virus**. Most of the times, if not always, sigla are proposed by a group of experts. Abbreviations, finally, can be constructed by a combination of the aforementioned criteria: PiRV- (from 1 to 4) refers to RNA viruses hosted by the oomycete *Phytophthora infestans* (Chromista): the unrelated PiRV-2 and PiRV-3, and the still unclassified PiRV-4, and PiRV-1 of the family *Narnaviridae*. Acronyms, sigla and abbreviations of viruses or viral groups are never italicized.

TABLE 2.4 Selected Rules for the Classification and Nomenclature of Viral Ranks and Its Orthography

Number ^a	Summary of Rules	Notes and Examples
3.2	<p>Hierarchical levels of classification must include Order, Family, Subfamily, Genus, and Species (In general, most of the times, only the virus specific name is provided in the literature; sometimes, Family and Order). The suffixes used are:</p> <ul style="list-style-type: none"> ● <i>virales</i> for Orders (3.34) ● <i>viridae</i> for Families (-<i>viroidae</i> for viroids) (3.32; 3.36) ● <i>virinae</i> for Subfamilies (3.30) ● <i>virus</i> for Genera (-<i>viroid</i> for viroids) (3.27; 3.36) <p>In 2017 it was approved that the rules concerning the classification of viruses should also be applied to viroids and satellite nucleic acids. Accordingly, suffixes for satellite ranks are:</p> <ul style="list-style-type: none"> ● <i>satellitidae</i> for Families ● <i>satellitinae</i> for Subfamilies ● <i>satellite</i> for Genera 	<p>Not all virus families have been assigned to a specific order. Similarly, not all families are divided into subfamilies and, in some cases, a virus species does not belong to a defined genus (unassigned) although it does belong to a family. Some genera are also floating (unassigned) since they are not allocated to a specific family. The use of “<i>Virus name (Order: Family: Subfamily: Genus)</i>” is not required and it is used here just as an example</p> <p>Example 1 (all ranks):</p> <p><i>Human alphaherpesvirus 1 (Herpesvirales: Herpesviridae: Alphaherpesvirinae: Simplexvirus)</i></p> <p>Example 2 (unassigned genus in a family):</p> <p><i>Banana mild mosaic virus (Tymovirales: Betaflexiviridae: Quinvirinae)</i></p> <p>Example 3 (accepted genus, but unassigned family):</p> <p><i>Citrus leprosis virus C (Cilevirus)</i></p> <p>Example 4 (majority of viruses: no order assigned, no subfamilies defined):</p> <p><i>Chikungunya virus (Togaviridae: Alphavirus)</i></p>
3.8	<p>A name is valid if accepted by the ICTV given that (1) it has been published, (2) it is associated with descriptive material, and (3) it conforms to the Rules of the <i>The International Code of Virus Classification and Nomenclature</i></p>	<p>Of course the approval of a name must follow the recommendations and fulfill the requirements established to that effect (i.e., the paperwork related to the application). These are studied by special committees devoted to specific virus groups. Viruses described based on metagenomics studies only, generally lack descriptive material</p>

(Continued)

TABLE 2.4 (Continued)

Number ^a	Summary of Rules	Notes and Examples
3.11	Names for taxa should be easy to use and easy to remember (particularly for virus species names)	In general names in use are short and <i>probably</i> easy to remember. Nonetheless, cases exist in which the name of the rank is of complex spelling, or too long
		Example 1 (short name, easy to remember):
		<i>Aura virus</i>
		Example 2 (long name):
		<i>Choristoneura fumiferana DEF multiple nucleopolyhedrovirus</i>
		Example 3 (complex spelling, besides example 2): The Family <i>Toleucusatellitidae</i> , or the virus species <i>Thottapalayam orthohantavirus</i>
3.15	Sigla (singular: siglum) may be accepted as part of the name of a taxa	Sigla are constructed using letters or other characters taken from words of a compound term, like arbovirus for arthropod-borne virus (this is not a taxonomic rank)
		Example 1 (from an Order)
		<i>Bunyvirales</i> , sigil for <i>Bunyamwera orthobunyavirus</i> and the suffix -virales
		Example 2 (from a Family):
		<i>Tristromaviridae</i> , from the Greek words tria (three) and stroma (layer), plus the suffix – viridae
		Example 3 (from a Genus): <i>Anphevirus</i> , sigil for <i>Anopheles</i> and the suffix – virus

(Continued)

TABLE 2.4 (Continued)

Number ^a	Summary of Rules	Notes and Examples
3.23	The name of a virus species must contain few words, and never be constructed with only the name of the host and the word "virus"	<p>Many virus names include the vernacular or scientific name of the host from which the virus was first described. Sometimes a symptom name is included, while in other instances the name of the location where the virus was found is used instead</p> <p>Example 1 (vernacular host name and symptom included):</p> <p><i>Rice grassy stunt tenuivirus</i></p> <p>Example 2 (scientific name and symptom included):</p> <p><i>Echinochloa ragged stunt virus</i></p> <p>Example 4 (generalized host name included):</p> <p><i>Avian coronavirus</i></p> <p>Example 5 (Geographic name included):</p> <p><i>St Croix River virus</i></p> <p>Example 6 (no clue provided, except that is a <i>certain</i> virus of a specific genus):</p> <p><i>Rotavirus A</i></p>
3.39 and 3.40	Names of all accepted ranks should be italicized with the first letter of the first word in capitals. First capital letters can also be used if belonging to a proper noun	If the not yet approved rank (especially virus species) contains a scientific name of the host, this name should not be italicized. For instance, <i>Phytophthora infestans</i> virus 4 is used, although <i>Phytophthora infestans</i> is the scientific name of a widely known oomycete. On the other hand, when talking about a specific group of viruses (not species) in general, italics are not required. That is, "potyviruses comprise the largest group of viruses with a ssRNA(+) genome," or "the newly created order that accommodates bunyaviruses," or "a new coronavirus is described here"

^aFrom *The International Code of Virus Classification and Nomenclature (2002)*.

The Goal of Virus Taxonomy

Virus taxonomy can then be defined as the arrangement of viruses into related clusters, identification of the extent of relatedness within and among these clusters, and the assignment of names to the clusters (taxa). In other words the goal of taxonomy is to categorize the multitude of known viruses so as to maximize organization, stability, and predictivity. The first goal, probably the most basic, is easily attainable given that the rules of classification are clear and followed by all. That is, the vast diversity of viruses is organized by grouping all similar viruses that share the same denomination and, for the most part, established evolutionary relationships. If this preliminary classification is based on solid grounds, stability ensues. Nonetheless, taxonomy cannot be a static, never-changing tool for virologists. As our realization of viral diversity widens and new technologies become available, virus taxonomy has to be flexible enough to accommodate occasional revisions and reinterpretation of perceived relationships between viruses. As for predictability, this is more complex and challenging; i.e., being able to speculate on the characteristics of near relatives of known species. Whether this is even possible at this point is questionable, given that the current knowledge of virus biodiversity is both biased and fragmentary, reflecting a focus on culturable or disease causing agents.

However, the traditional approach to virus classification has been challenged by recent technological developments. Rapid advances in metagenomics have revealed a virosphere that is more phylogenetically and genomically diverse than that depicted in the current classification scheme. Indeed a number of studies have generated large sequence datasets of viral metagenomic sequences that cannot readily be classified using conventional criteria since these are based on phenotypic properties as well as genetic relationships. Classification based on sequence comparisons alone represents a pragmatic solution to the problem of classifying viral metagenomic sequences, but is a significant departure from current methods for classifying viruses. Viral metagenomic sequences may have sufficient defining characteristics to enable their classification as additional taxa in existing virus families and may even be used to justify the creation of new virus families. Further information on diversity and clustering may, however, be required to justify the formation of genera and species ranks within the family. Perhaps though, a reevaluation of the concept of viral species, genera, families, and higher levels of classification is needed; particularly a consensus on which diagnostic properties are the most useful for identifying individual members of a virus species. Just as microbiologists discarded dubious morphological traits in favor of more accurate molecular yardsticks of evolutionary change, virologists can gain new insight into viral evolution through the rigorous analyzes afforded by the molecular phylogenetics of viral genes, which are essential for the classification of viruses below the family taxon level. On the other

hand, others have gone as far as suggesting a new division of biological entities into two classes of organisms: ribosome-encoding organisms that include all archaea, bacteria and eukarya, and capsid-encoding organisms, the viruses. Doubts have been raised on the suitability of relying only on the capsid structure for resolving phylogenetic uncertainties. Some argue that a more integrated account of all three categories of containment, replication, and reliance on the information processing systems of the host, recognizes the origin and evolution of viruses. The last point is worth emphasizing; evolutionary processes have shaped the genetic structure and diversity of viruses and the viruses themselves have influenced the evolution of cellular organisms in all domains of life.

FURTHER READING

- Adams, M.J., Lefkowitz, E.J., King, A.M.Q., Harrach, B., Harrison, R.L., Knowles, N.J., et al., 2017. Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2017). *Arch. Virol.* 162, 2505–2538. Available from: <https://doi.org/10.1007/s00705-017-3358-5>.
- Koonin, E.V., Dolja, V.V., 2014. Virus world as an evolutionary network of viruses and capsidless selfish elements. *Microbiol. Mol. Biol. Rev.* 78, 278–303. Available from: <https://doi.org/10.1128/MMBR.00049-13>.
- Mateu, G.M. (Ed.), 2013. *Structure and Physics of Viruses. An Integrated Textbook*. Springer, Dordrecht, The Netherlands. Available from: <https://doi.org/10.1007/978-94-007-6552-8>.
- Paez-Espino, D., Eloe-Fadrosh, E.A., Pavlopoulos, G.A., Thomas, A.D., Huntemann, M., Mikhailova, N., et al., 2016. Uncovering Earth's virome. *Nature* 536, 425–430. Available from: <https://doi.org/10.1038/nature19094>.
- Peterson, T., 2014. Defining viral species: making taxonomy useful. *Virol. J.* 11, 131. Available from: <https://doi.org/10.1186/1743-422X-11-131>.
- Ruggiero, M.A., Gordon, D.P., Orrell, T.M., Bailly, N., Bourgoin, T., Brusca, R.C., et al., 2015. A higher level classification of all living organisms. *PLoS ONE* 10, e0119248. Available from: <https://doi.org/10.1371/journal.pone.0119248>.
- Simmonds, P., 2015. Methods for virus classification and the challenge of incorporating metagenomic sequence data. *J. Gen. Virol.* 96, 1193–1206. Available from: <https://doi.org/10.1099/jgv.0.000016>.