



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.

## Enveloped, Positive-Strand RNA Viruses (*Nidovirales*)

**Luis Enjuanes**, National Center for Biotechnology – Spanish National Research Council (CNB-CSIC), Madrid, Spain

**AE Gorbalenya**, Leiden University Medical Center, Leiden, The Netherlands

**Raoul J de Groot**, Utrecht University, Utrecht, The Netherlands

**Jeff A Cowley**, CSIRO Livestock Industries, Brisbane, QLD, Australia

**John Ziebuhr**, The Queen's University of Belfast, Belfast, United Kingdom

**Eric J Snijder**, Leiden University Medical Center, Leiden, The Netherlands

© 2008 Elsevier Ltd. All rights reserved.

This is a reproduction of L. Enjuanes, A.E. Gorbalenya, R.J. de Groot, *et al.*, Nidovirales, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00775-5.

### Glossary

**3CL<sup>pro</sup> or M<sup>pro</sup>** 3C-like proteinase, or main proteinase.

**ADRP** ADP-ribose-1''-phosphatase.

**CS** TRS Core sequence.

**ExoN** 3' to 5' exoribonuclease.

**NendoU** Nidovirus endoribonuclease.

**O-MT** Ribose-2'-O-methyltransferase.

**PL<sup>pro</sup>** Papain-like cysteine proteinase.

**TRS** Transcription-regulating sequence.

### Taxonomy and Phylogeny

The order *Nidovirales* includes the families *Coronaviridae*, *Roniviridae*, and *Arteriviridae* (Fig. 1). The *Coronaviridae* comprises two well-established genera, *Coronavirus* and *Torovirus*, and a tentative new genus, *Bafinivirus*. The *Arteriviridae* and *Roniviridae* include only one genus each, *Arterivirus* and *Okavirus*, respectively. All nidoviruses have single-stranded RNA genomes of positive polarity that, in the case of the *Corona*- and *Roniviridae* (26–32 kbp), are the largest presently known RNA virus genomes. In contrast, members of the *Arteriviridae* have a smaller genome ranging from about 13 to 16 kbp. The data available from phylogenetic analysis of the highly conserved RNA-dependent RNA polymerase (RdRp) domain of these viruses, and the collinearity of the array of functional domains in nidovirus replicase polyproteins, were the basis for clustering coronaviruses and toroviruses (Fig. 2). The more distantly related roniviruses also group with corona- and toroviruses, thus forming a kind of supercluster of nidoviruses with large genomes. By contrast, arteriviruses must have diverged earlier during nidovirus evolution. The current taxonomic position of coronaviruses and toroviruses as two genera of the family *Coronaviridae* is currently being revised by elevating these virus groups to the taxonomic rank of either subfamily or family.

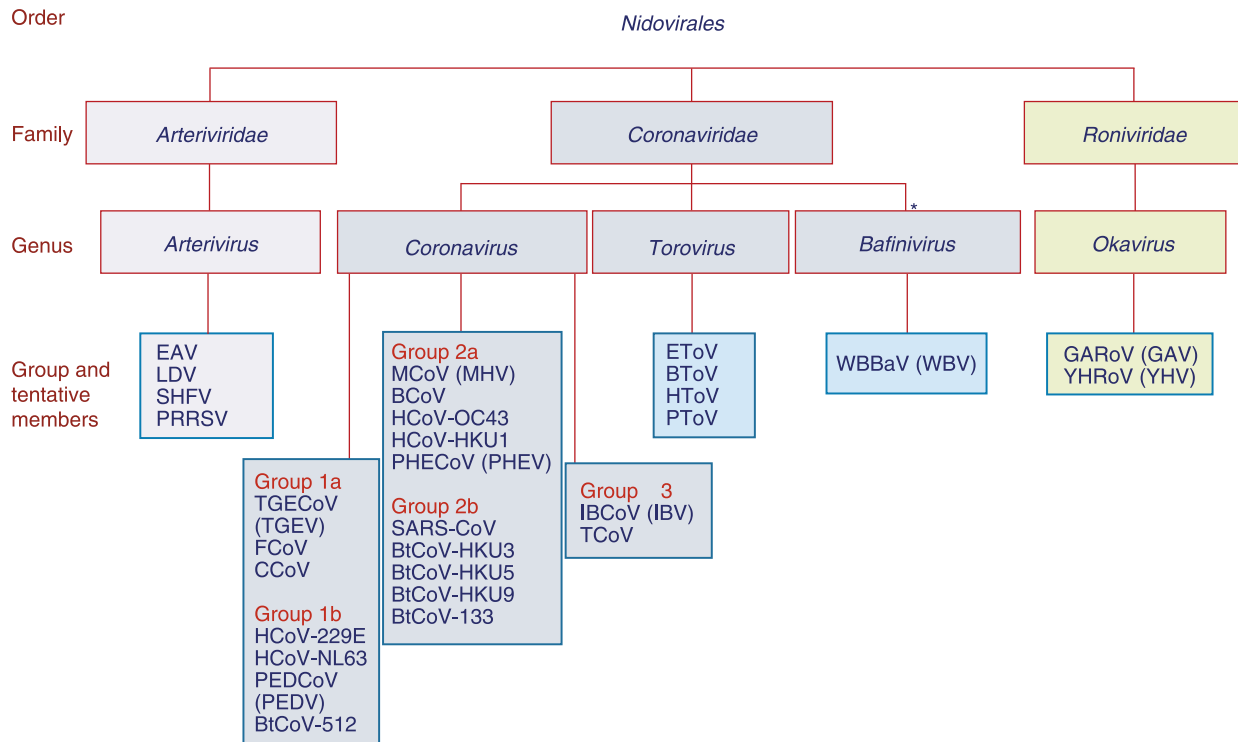
A comparative sequence analysis of coronaviruses reveals three phylogenetically compact clusters: groups 1, 2, and 3. Within group 1, two subsets can be distinguished: subgroup 1a that includes transmissible gastroenteritis virus (TGEV), canine coronavirus (CCoV), and feline coronavirus (FCoV), and subgroup 1b that includes the human coronaviruses (HCoV) 229E and NL63, porcine epidemic diarrhoea virus (PEDV), and bat coronavirus (BtCoV) 512 which was isolated in 2005. Within group 2 coronaviruses, two subsets have been recognized: subgroup 2a, including mouse hepatitis virus (MHV), bovine coronavirus (BCoV), HCoV-OC43, and HCoV-HKU1; and subgroup 2b, including severe acute respiratory syndrome coronavirus (SARS-CoV) and its closest circulating bat coronavirus relative, BtCoV-HKU3. A growing number of other bat viruses has been recently identified in groups 1 and 2. It is currently being debated whether some of these viruses (e.g., BtCoV-HKU5, BtCoV-133 (isolated in 2005), and BtCoV-HKU9) may in fact represent novel subgroups or groups. Avian infectious bronchitis virus (IBV) is the prototype of coronavirus group 3, which also includes several other bird coronaviruses. In arteriviruses, there are four comparably distant genetic clusters, the prototypes of which are equine arteritis virus (EAV), lactate dehydrogenase-elevating virus (LDV) of mice, simian hemorrhagic fever virus (SHFV) infecting monkeys, and porcine reproductive and respiratory syndrome virus (PRRSV) which infects pigs and includes European and North American genotypes.

Roniviruses are the only members of the order *Nidovirales* that are known to infect invertebrates. The family *Roniviridae* includes the penaeid shrimp virus, gill-associated virus (GAV), and the closely related yellow head virus (YHV).

More than 100 full-length coronavirus genome sequences and around 30 arterivirus genome sequences have been documented so far, whereas only very few sequences have been reported for toroviruses, bafiniviruses, and roniviruses. Therefore, information on the genetic variability of these nidovirus *taxa* is limited.

### Diseases Associated with Nidoviruses

Coronavirus infections are mainly associated with respiratory, enteric, hepatic, and central nervous system diseases. In humans and fowl, coronaviruses primarily cause upper respiratory tract infections, while porcine and bovine coronaviruses establish enteric



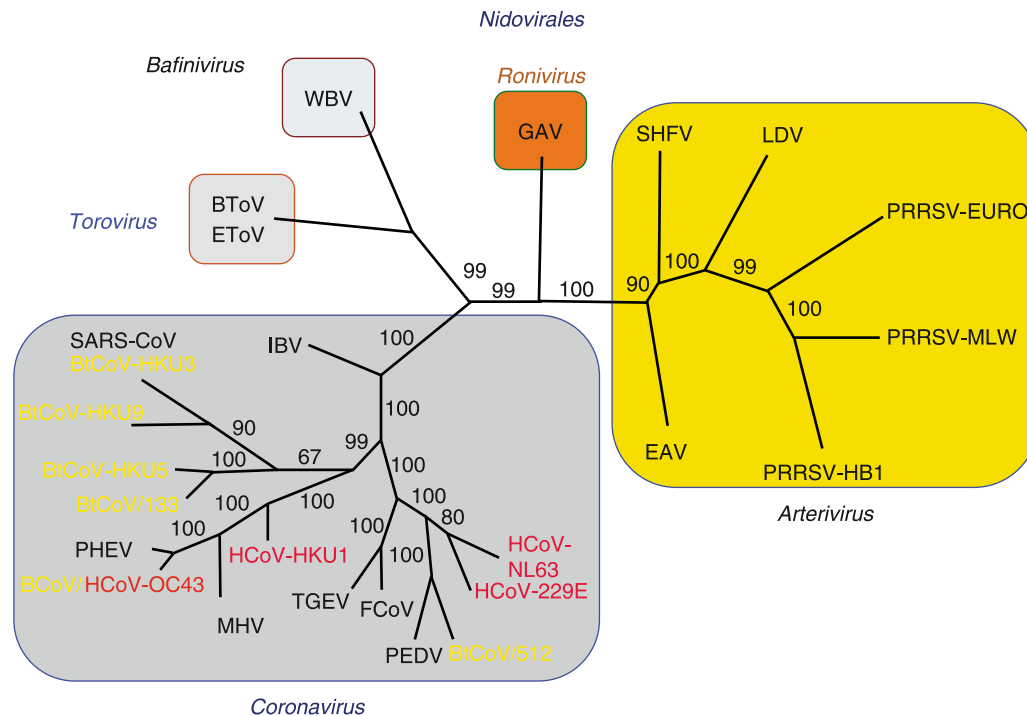
**Fig. 1** Nidovirus classification and prototype members. The order *Nidovirales* containing the families *Coronaviridae* (including the established genera *Coronavirus* and *Torovirus*, and a new tentative genus *Bafinivirus*), *Arteriviridae*, and *Roniviridae*. Phylogenetic analysis (see Fig. 2) has confirmed the division of coronaviruses into three groups. In arteriviruses, four comparably distant genetic clusters have been differentiated. To facilitate the taxonomy of the different virus isolates, the types Co, To, Ba, Ro, standing for coronavirus, torovirus, bafinivirus, or ronivirus, respectively, have been included. The following CoVs are shown: human coronaviruses (HCoV) 229E, HKU1, OC43 and NL63, transmissible gastroenteritis virus (TGEV), feline coronavirus (FCoV), porcine epidemic diarrhoea virus (PEDV), mouse hepatitis virus (MHV), bovine coronavirus (BCoV), bat coronaviruses (BtCoV) HKU3, HKU5, HKU9, 133 and 512 (the last two isolated in 2005), porcine hemagglutinating encephalomyelitis virus (PHEV), avian infectious bronchitis virus (IBV), and severe acute respiratory syndrome coronavirus (SARS-CoV); ToV: equine torovirus (EToV), bovine torovirus (BToV), human torovirus (HToV), and porcine torovirus (PToV); BaV: white bream virus (WBV); Arterivirus: equine arteritis virus (EAV), simian haemorrhagic fever virus (SHFV), lactate dehydrogenase-elevating virus (LDV), and three (Euro, HB1, and MLV) porcine reproductive and respiratory syndrome viruses (PRRSV); RoV: gill-associated virus (GAV) and yellow head virus (YHV). Human viruses are highlighted in red. Some nodes are formed by a pair of very closely related viruses (e.g., SARS-CoV and BtCoV-HKU3). Asterisk indicates tentative genus.

infections, often resulting in severe economic losses. In 2002, a previously unknown coronavirus that probably has its natural reservoir in bats crossed the species barrier and caused a major outbreak of SARS, which led to more than 800 deaths worldwide.

Toroviruses cause gastroenteritis in mammals, including humans, and possibly also respiratory infections in older cattle. Bafiniviruses have been isolated from white bream fish but there is currently no information on the pathogenesis associated with this virus infection. Roniviruses usually exist as asymptomatic infections but can cause severe disease outbreaks in farmed black tiger shrimp (*Penaeus monodon*) and white pacific shrimp (*Penaeus vannamei*), which in the case of YHV can result in complete crop losses within a few days after the first signs of disease in a pond. Infections by arteriviruses can cause acute or persistent asymptomatic infections, or respiratory disease and abortion (EAV and PRRSV), fatal age-dependent poliomyelitis (LDV), or fatal hemorrhagic fever (SHFV). Arteriviruses, particularly PRRSV in swine populations, cause important economic losses.

## Virus Structure

In addition to the significant variations in genome size among the three nidovirus families mentioned above, there are also major differences in virion morphology (Fig. 3) and host range. Nidoviruses have a lipid envelope which protects the internal nucleocapsid structure and contains a number of viral surface proteins (Fig. 3). Whereas coronaviruses and the significantly smaller arteriviruses have spherical particle structures, elongated rod-shaped structures are observed in toro-, bafini-, and ronivirus-infected cells. The virus particles of the *Corona*- and *Roniviridae* family members carry large surface projections that protrude from the viral envelope (peplomers), whereas arterivirus particles possess only relatively small projections on their surface. Coronaviruses have an internal core shell that is formed by a nucleocapsid featuring a helical symmetry. The nucleocapsid (N) protein interacts with the carboxy-terminus of the envelope membrane (M) protein. The intracellular forms of torovirus, bafinivirus, and ronivirus



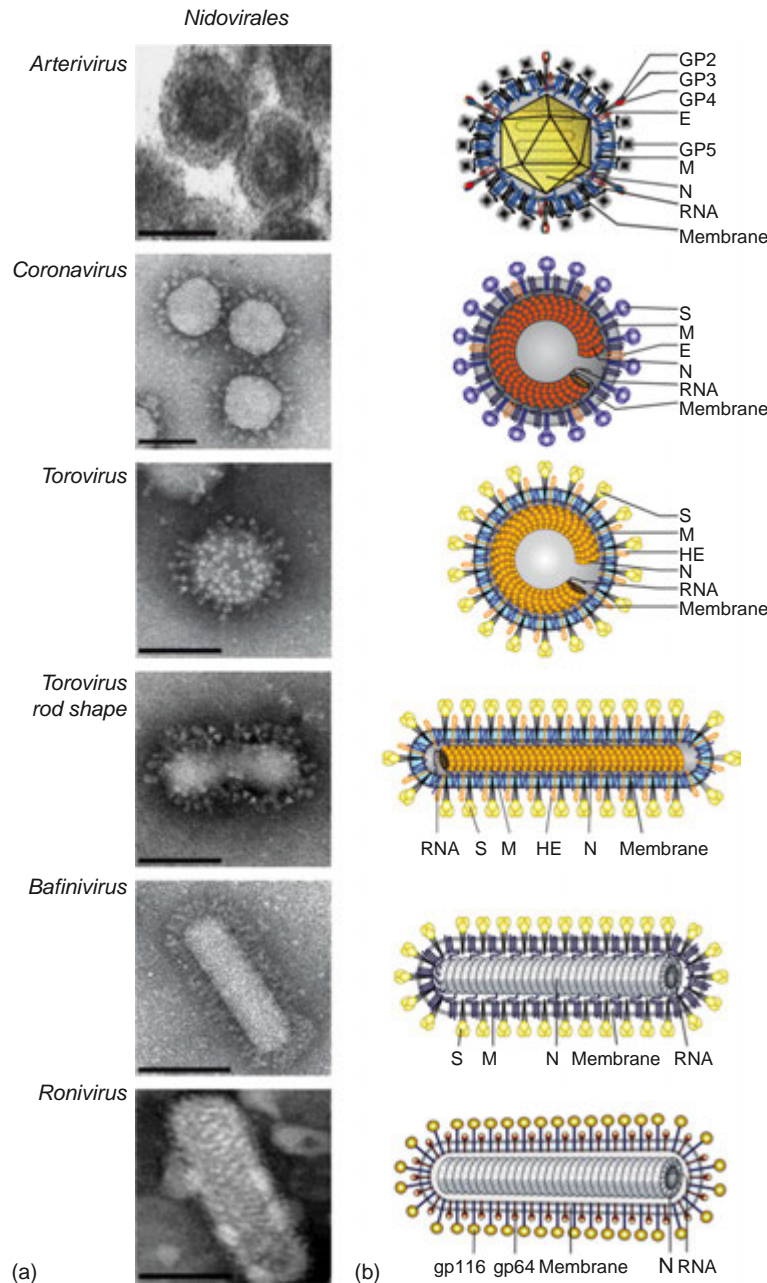
**Fig. 2** Nidovirus phylogeny. Tree depicting the evolutionary relationships between the five major groups of nidoviruses as shown in Fig. 1 (Coronavirus, Torovirus, Bafinivirus, Ronivirus, and Arterivirus). This unrooted maximum parsimonious tree was inferred using multiple nucleotide alignments of the RdRp-HEL region of a representative set of nidoviruses with the help of the PAUP<sup>+</sup>v.4.0b10 software (AEG, unpublished). Support for all bifurcations from 100 bootstraps performed is indicated. The phylogenetic distances shown are approximate. For acronyms, see Fig. 1.

nucleocapsids have extended rod-shaped (helical) morphology. By contrast, mature (extracellular) toroviruses (but not bafini- and roniviruses) feature a remarkable structural flexibility, which allows them to adopt crescent- and toroid-shaped structures also. Unlike other nidoviruses, arteriviruses have an isometric core shell. In all nidoviruses, the nucleocapsid is formed by only a single N protein that interacts with the genomic RNA.

Both the number and properties of structural proteins vary between viruses of the three families of the *Nidovirales* and may even vary among viruses of the same family. Nidoviruses usually encode at least three structural proteins: a spike (S) or major surface glycoprotein, a trans-membrane (M) or matrix protein, and the N protein (Fig. 3). Ronivirus particles are unique in that they possess two envelope glycoproteins, gp116 (S1) and gp64 (S2), but no M protein. Coronavirus and arterivirus particles possess another envelope protein called E that is not conserved in toroviruses, bafini-, and roniviruses. Toroviruses and subgroup 2a coronaviruses, such as MHV, have a hemagglutinin esterase (HE) as an additional structural protein, whereas the SARS-CoV has at least four additional proteins that are present in the viral envelope (encoded by ORFs 3a, 6, 7a, and 7b). The proteins may promote virus growth in cell culture or *in vivo*, but they are dispensable for virus replication.

The major envelope proteins are the S and M proteins in coronaviruses and toroviruses, the GP<sub>5</sub> and M proteins in arteriviruses, and the S1 and S2 proteins in roniviruses. Among these, only the corona-, toro-, and bafinivirus S and M proteins share limited sequence similarities, possibly indicating a common origin. Whereas S proteins can differ in size, they share an exposed globular head domain and, with the exception of roniviruses, a stem portion containing heptad repeats organized in a coiled-coil structure. The S proteins of corona- and toroviruses (and most likely those of bafiniviruses) form trimers that bind the cell surface receptor whereas receptor binding in roniviruses is probably mediated by gp116 (S1). The arterivirus envelope proteins form two higher-order complexes: one is a disulfide-linked heterodimer of GP<sub>5</sub> and the M protein; and the other is a heterotrimer of the minor structural glycoproteins GP<sub>2</sub>, GP<sub>3</sub>, and GP<sub>4</sub>. Except for the E and M proteins, all arterivirus structural proteins are glycosylated. By contrast, the M proteins of corona- and toroviruses (and, most likely, bafiniviruses) are glycosylated, and they share a triple-spanning membrane topology with the amino-terminus exposed on the outside of the virions and the carboxy-terminus facing the nucleocapsid. In TGEV, a proportion of the M proteins has a tetra-spanning membrane topology leading to the exposure of both termini on the virion surface.

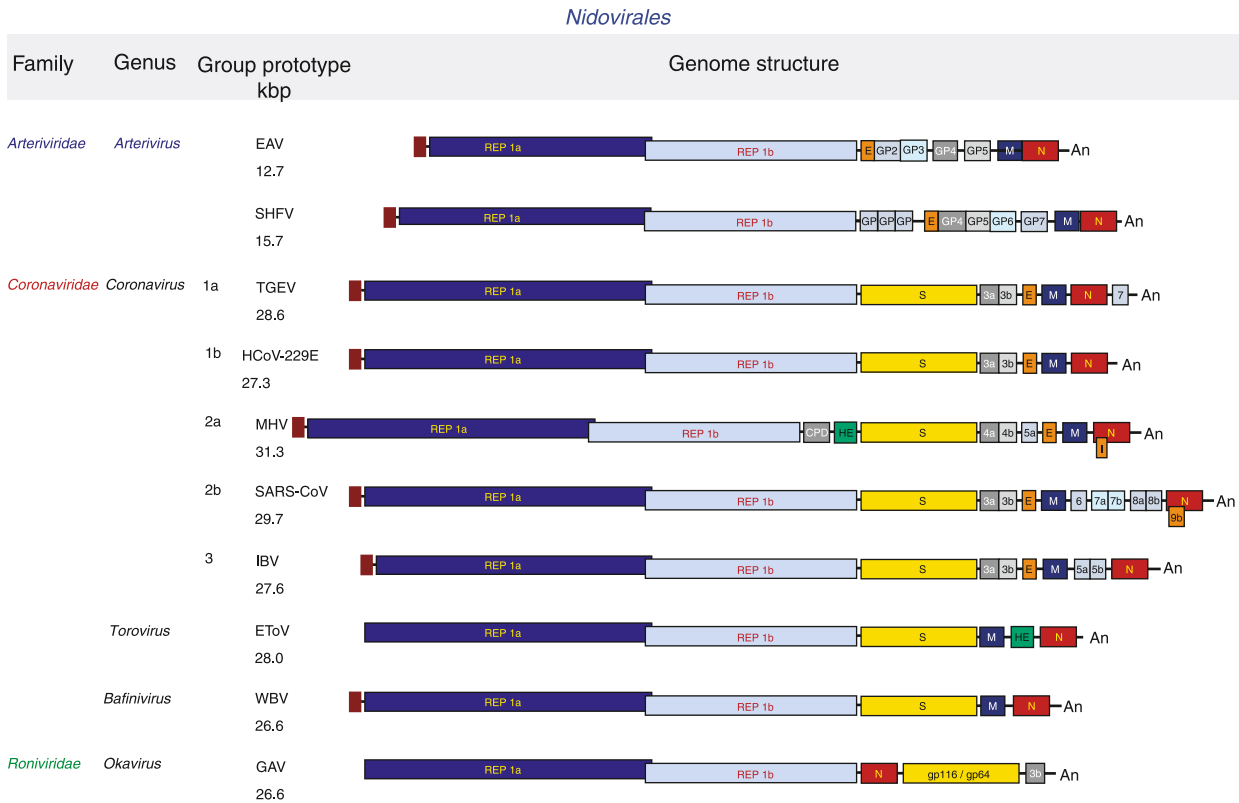
In the virion, the coronavirus E protein has a low copy number (around 20) and deletion of the E protein gene from the genome of the group 1 coronavirus TGEV blocks virus maturation, preventing virus release and spread. In the group 2 coronaviruses MHV and SARS-CoV, deletion of the E protein results in a dramatic reduction, of up to 100 000-fold, of virus infectivity. The coronavirus E and SARS-CoV 3a proteins are viroporins, that is, they belong to a group of proteins that modify membrane permeability by forming ion channels in the virion envelope.



**Fig. 3** Nidovirus structure. Architecture of particles of members of the order *Nidovirales*: electron micrographs (a) and schematic representations (b) are shown. N, nucleocapsid protein; S, spike protein; M, membrane protein; E, envelope protein; HE, hemagglutinin-esterase. Coronavirus M protein interacts with the N protein. In arterivirus, GP<sub>5</sub> and M are major envelope proteins, while GP<sub>2</sub>, GP<sub>3</sub>, GP<sub>4</sub>, and E are minor envelope proteins. Toro-, bafini-, and roniviruses lack the E protein present in corona- and arteriviruses. Proteins gp116 and gp64, ronivirus envelope proteins. Different images were reproduced with permission from different authors: arterivirus, E. Snijder (Leiden, The Netherlands); ronivirus, P. J. Walker (CSIRO, Australia); bafinivirus, J. Ziebuhr (Queen's University, Belfast) torovirus, D. Rodriguez (CNB, Spain); coronavirus, L. Enjuanes (CNB, Spain).

### Genome Organization

Nidovirus genomes contain variable numbers of genes, but in all cases the 5' terminal two-thirds to three-quarters of the genome is dedicated to encoding the key replicative proteins, whereas the 3' proximal genome regions generally encode the structural and, in some cases, accessory (group- and virus-specific) proteins (Fig. 4). Nidovirus genome expression is controlled at the translational and post-translational levels. Thus, for example, ribosomal frameshifting is required for the expression of ORF1b, and the two replicase polyproteins (pp1a and pp1ab) are proteolytically processed by viral proteases. The proteolytic processing occurs in a



**Fig. 4** Nidovirus genome structure. Genome organization of selected nidoviruses. The genomic ORFs of viruses representing the major nidovirus lineages are indicated and the names of the replicase and main virion genes are given. References to the nomenclature of accessory genes can be found in the text. Genomes of large and small nidoviruses are drawn to different scales. Red box at the 5' end refers to the leader sequence. Partially overlapping ORFs have been drawn as united boxes. Spaces between boxes representing different ORFs do not mean noncoding sequences.

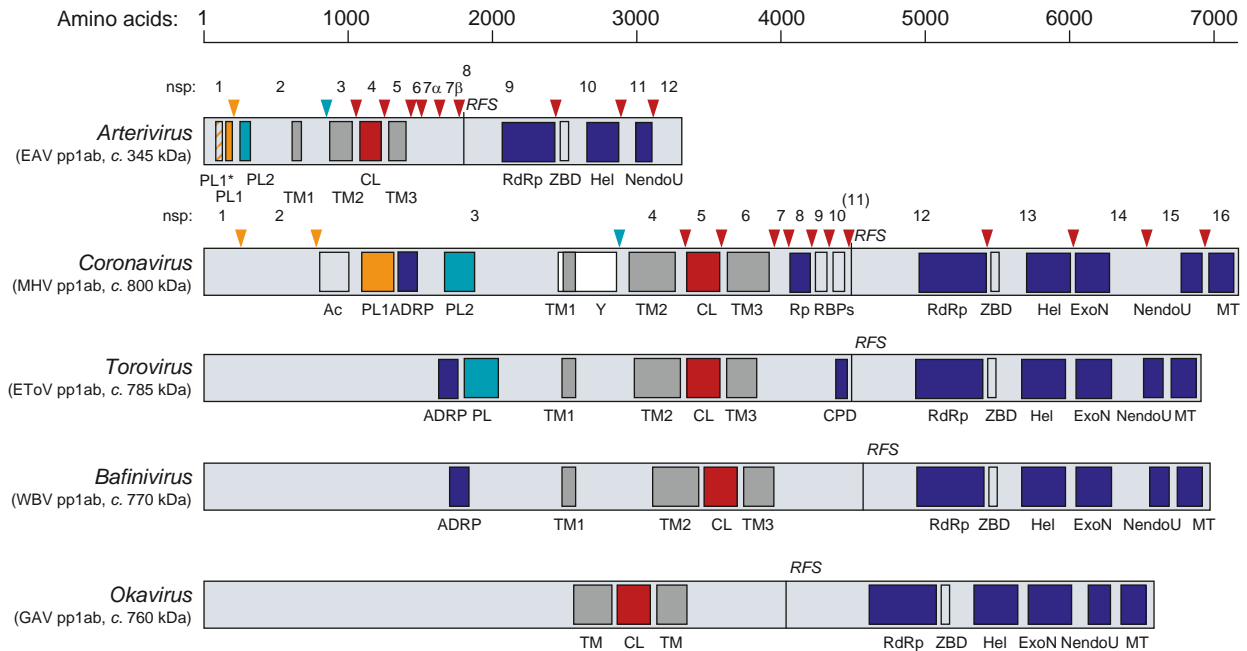
coordinated manner and gives rise to the functional subunits of the viral replication–transcription complex. By contrast, the expression of the structural and several accessory proteins is controlled at the level of transcription. It involves the synthesis of a nested set of 3' co-terminal sg mRNAs that are produced in nonequimolar amounts. As in cellular eukaryotic mRNAs, in general only the ORF positioned most closely to the 5' end of the sg mRNA is translated.

### The Replicase

The nidovirus replicase gene is comprised of two slightly overlapping ORFs, 1a and 1b. In corona-, toro-, bafini-, and roniviruses, ORF1a encodes a polyprotein (pp1a) of 450–520 kDa, whereas a polyprotein of 760–800 kDa (pp1ab) is synthesized from ORF1ab. Expression of the ORF1b-encoded part of pp1ab involves a ribosomal frameshift mechanism that, in a defined proportion of translation events, directs a controlled shift into the –1 reading frame just upstream of the ORF1a stop codon (Fig. 4). In arteriviruses, pp1a (190–260 kDa) and pp1ab (345–420 kDa) are considerably smaller in size. Proteolytic processing of coronavirus pp1a and pp1ab generates up to 16 nonstructural proteins (nsps 1–16), while processing of the arterivirus replicase polyproteins generates up to 14 nsps. It is generally accepted that most of the replicase nsps assemble into a large protein complex, called the replication–transcription complex. The complex is anchored to intracellular membranes and likely also includes a number of cellular proteins. Nidoviruses replicase genes include a conserved array of protease, RNA-dependent RNA polymerase (RdRp), helicase (HEL), and endoribonuclease (NendoU) activities. In contrast to other positive-strand RNA viruses, they employ an RdRp with a characteristic SDD rather than the usual GDD active site.

The vast majority of proteolytic cleavages in pp1a/pp1ab are mediated by an ORF1a-encoded chymotrypsin-like protease that, because of its similarities to picornavirus 3C proteases, is called the 3C-like protease (3CL<sup>Pro</sup>). Also the term 'main protease' (M<sup>Pro</sup>) is increasingly used for this enzyme, mainly to refer to its key role in nidovirus replicase polyprotein processing (Fig. 5). Nidovirus–M<sup>Pro</sup> share a three-domain structure. The two N-terminal domains adopt a two- $\beta$ -barrel fold reminiscent of the structure of chymotrypsin. With respect to the principal catalytic residues, there are major differences between the main proteases from different nidovirus genera. The presence of a third, C-terminal domain is a conserved feature of nidovirus main proteases, even though these domains vary significantly in both size and structure. The C-terminal domain of the coronavirus M<sup>Pro</sup> is involved in protein dimerization that is required for proteolytic activity *in trans*. Over the past years, a large body of structural and functional information has been obtained for corona- and arterivirus main proteases which, in the case of coronaviruses, has also been used





**Fig. 5** Nidovirus replicase genes. Polyprotein (pp) 1ab domain organizations are shown for representative viruses from the five nidovirus genera. Acronyms as in Fig. 1. Arterivirus and coronavirus pp1ab processing pathways have been characterized in considerable detail and are illustrated here for EAV and MHV. N-proximal polyprotein regions are cleaved at two or three sites by viral papain-like proteases 1 (PL1) and 2 (PL2), whereas the central and C-terminal polyprotein regions are processed by the main protease,  $M^{pro}$ . PL1 domains are indicated by orange boxes and cognate cleavage sites are indicated by orange arrowheads. PL2 domains and PL2-mediated cleavages are shown in green and CL domains and CL-mediated cleavages are shown in red. Note that EAV encodes a second, but proteolytically inactive PL1 domain (PL1\*; orange-striped box). For the genera *Torovirus*, *Bafinivirus*, and *Okavirus*, the available information on pp1ab proteolytic processing is limited and not shown here. Other predicted or proven enzymatic activities are shown in blue: ADRP, ADP-ribose 1'-phosphatase; Rp, noncanonical RNA polymerase ('primase') activity; RdRp, RNA-dependent RNA polymerase; HEL, NTPase/RNA helicase and RNA 5'-triphosphatase; ExoN, 3'-to-5' exoribonuclease; NendoU, nidoviral uridylyate-specific endoribonuclease; MT, ribose-2'-O methyltransferase; CPD, cyclic nucleotide phosphodiesterase. Regions with predicted transmembrane (TM) domains are indicated by gray boxes. Other functional domains are shown as white boxes: Ac, acidic domain; Y, Y domain containing putative transmembrane and zinc-binding regions; ZBD, helicase-associated zinc-binding domain; RBPs, RNA-binding proteins. Expression of the C-terminal part of pp1ab requires a ribosomal frameshift, which occurs just upstream of the ORF1a translation stop codon. The ribosomal frameshift site (RFS) is indicated.

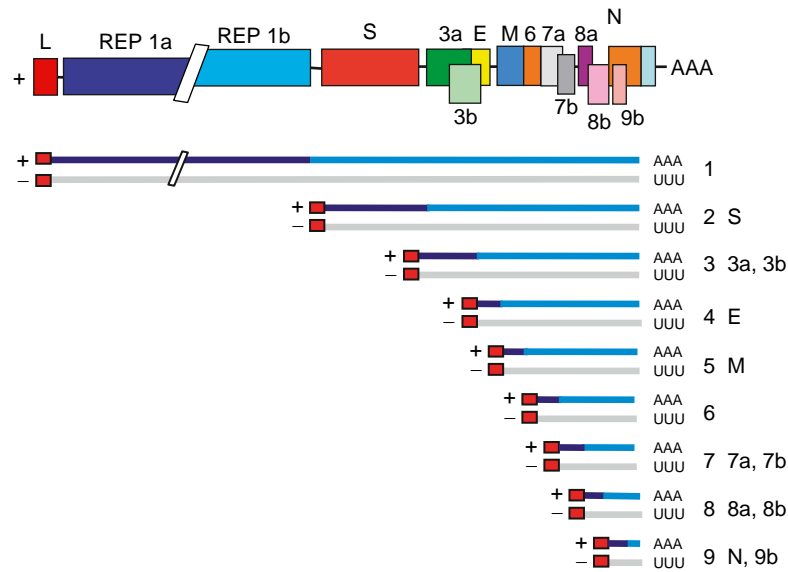
to develop selective protease inhibitors that block viral replication, suggesting that nidovirus main proteases may be attractive targets for antiviral drug design.

In arteri-, corona-, and toroviruses, the  $M^{pro}$  is assisted by 1–4 papain-like ('accessory') proteases (PL $^{pro}$ ) that process the less well-conserved N-proximal region of the replicase polyproteins (Fig. 5). Nidovirus PL $^{pro}$  domains may include zinc ribbon structures and some of them have deubiquitinating activities, suggesting that these proteases might also have functions other than polyprotein processing. Bafini- and roniviruses have not been studied in great detail and it is not yet clear if these viruses employ papain-like proteases to process their N-terminal pp1a/pp1ab regions (Fig. 5).

The replicase polyproteins of 'large' nidoviruses with genome sizes of more than 26 kb (i.e., corona-, toro-, bafini-, and roniviruses) include 3'-5' exoribonuclease (ExoN) and ribose-2'-O-methyltransferase (MT) activities that are essential for coronavirus RNA synthesis but are not conserved in the much smaller arteriviruses (Fig. 5). The precise biological function of ExoN has not been established for any nidovirus but the relationship with cellular DEDD superfamily exonucleases and recently published data suggest that ExoN may have functions in the replication cycle of large nidoviruses that, like in the DEDD homologs, are related to proofreading, repair, and recombination mechanisms.

NendoU is a nidovirus-wide conserved domain that has no counterparts in other RNA viruses. It is therefore considered a genetic marker of the *Nidovirales*. The endonuclease has uridylyate specificity and forms hexameric structures with six independent catalytic sites. Cellular homologs of NendoU have been implicated in small nucleolar RNA processing whereas the role of NendoU in viral replication is less clear. Reverse genetics data indicate that NendoU has a critical role in the viral replication cycle.

Two other RNA-processing domains, ADP-ribose-1''-phosphatase (ADRP) and nucleotide cyclic phosphodiesterase (CPD), are conserved in overlapping subsets of nidoviruses (Fig. 5). Except for arteri- and roniviruses, all nidoviruses encode an ADRP domain that is part of a large replicase subunit (nsp3 in the case of coronaviruses). The coronavirus ADRP homolog has been shown to have ADP-ribose-1'-phosphatase and poly(ADP-ribose)-binding activities. Although the highly specific phosphatase activity is not essential for viral replication *in vitro*, the strict conservation in all genera of the *Coronaviridae* suggests an important (though currently unclear) function of this protein in the viral replication cycle. This may be linked to host cell functions and,



**Fig. 6** SARS-CoV genome organization and sg mRNA expression. A diagram of CoV structure using SARS-CoV as a prototype. Below the top bar a set of positive- and negative-sense mRNA species synthesized in infected cells is shown. Dark and light blue lines (+), mRNA sequences translated and nontranslated into viral proteins, respectively. Light gray lines (-), RNAs complementary to the different mRNAs. L, leader sequence. Poly(A) and Poly(U) tails are indicated by AAA or UUU, respectively. Rep 1a and Rep 1b, replicase genes. Other acronyms above and below the top bar indicate structural and nonstructural proteins. Numbers and letters to the right of the thin bars indicate the sg mRNAs.

particularly, to the activities of cellular homologs called ‘macro’ domains which are thought to be involved in the metabolism of ADP-ribose and its derivatives.

The CPD domain is only encoded by toroviruses and group 2a coronaviruses. In toroviruses, the CPD domain is encoded by the 3' end of replicase ORF1a (Fig. 5), whereas in group 2a coronaviruses, the enzyme is expressed from a separate subgenomic RNA. The enzyme's biological function is not clear. Coronavirus CPD mutants are attenuated in the natural host whereas replication in cell culture is normal, suggesting some function *in vivo*. The available information suggests that nidovirus replicase polyproteins (particularly, those of large nidoviruses) have evolved to include a number of nonessential functions that may provide a selective advantage in the host.

ORF1a of all nidoviruses encodes a number of (putative) transmembrane proteins, like the coronavirus nsps 3, 4, and 6 and the arterivirus nsps 2, 3, and 5. These have been shown or postulated to trigger the modification of cytoplasmic membranes, including the formation of unusual double-membrane vesicles (DMVs). Tethering of the replication–transcription complex to these virus-induced membrane structures might provide a scaffold or subcellular compartment for viral RNA synthesis, possibly allowing it to proceed under conditions that prevent or impair detection by cellular defense mechanisms, which are usually induced by the double-stranded RNA intermediates of viral replication.

Finally, recent structural and biochemical studies have yielded novel insights into the function of a set of small nsps encoded in the 3'-terminal part of the coronavirus ORF1a. For example, nsp7 and nsp8 were shown to form a hexadecameric supercomplex that is capable of encircling dsRNA. The coronavirus nsp8 was also shown to have RNA polymerase (primase) activity that may produce the primers required by the primer-dependent RdRp residing in nsp12. For nsp9 and nsp10, RNA-binding activities have been demonstrated and crystal structures have been reported for both proteins. Nsp10 is a zinc-binding protein that contains two zinc-finger-binding domains and has been implicated in negative-strand RNA synthesis.

### Structural and Accessory Protein Genes

In contrast to the large genome of *Coronaviridae*, which can accommodate genes encoding accessory proteins (i.e., proteins called ‘nonessential’ for being dispensable for replication in cell culture (Fig. 6)), the smaller genomes of arteriviruses only encode essential proteins (Fig. 4). Coronaviruses encode a variable number of accessory proteins (2–8), while the torovirus genome contains a single accessory gene encoding a hemagglutinin-esterase (HE). Coronavirus accessory genes may occupy any intergenic position in the conserved array of the four genes encoding the major structural proteins (5'-S-E-M-N-3'), or they may reside upstream or downstream of this gene array. Roniviruses are unique among the presently known nidoviruses in that the gene encoding the N protein is located upstream rather than downstream of the gene encoding the glycoproteins. Several members of the coronavirus group 1a are exceptional in that they contain genes downstream of the N protein gene, which has not been reported for other coronaviruses. The ronivirus glycoprotein gene is also unique in that it encodes a precursor polyprotein with two



internal signal peptidase cleavage sites used to generate the envelope glycoproteins S1 and S2 as well as an amino-terminal protein with an unknown function.

The accessory genes are specific for either a single virus species or a few viruses that form a compact phylogenetic cluster. Many proteins encoded by accessory genes may function in infected cells or *in vivo* to counteract host defenses and, when removed, may lead to attenuated virus phenotypes. Group 1 coronaviruses may have 2–3 accessory genes located between the S and E genes and up to two other genes downstream of N gene. Viruses of group 2 form the most diverse coronavirus cluster, and they may have between three and eight accessory genes. In this cluster, MHV, HCoV-OC43, and BCoV form the phylogenetically compact subgroup, 2a, that is characterized by the presence of (1) two accessory genes located between ORF1b and the S gene encoding proteins with CPD and HE functions, (2) two accessory genes located between the S and E protein genes, and (3) an accessory gene, I, that is located within the N protein gene. Of this set of five accessory proteins, only three homologs are encoded by the recently identified HCoV-HKU1, which is the closest known relative of the cluster formed by MHV, HCoV-OC43, and BCoV. In contrast, the most distant group 2 member, SARS-CoV, has seven or eight unique accessory genes, two between the S and E protein genes, four to five between the M and N protein genes, and ORF9b which entirely overlaps with the N protein gene in an alternative reading frame. In group 3 avian coronaviruses, of which IBV is the prototype, several accessory genes, which are expressed from functionally tri- or bicistronic mRNAs, have been identified in the region between the S and E protein genes (gene 3) and between the M and N protein genes (gene 5).

Some functionally dispensable ORF1a-encoded replicase domains may also be considered as accessory protein functions. For instance, MHV and SARS-CoV nsp2 turned out to be nonessential for replication in cell culture.

## Replication

Like in all other positive-stranded RNA viruses, nidovirus genome replication is mediated through the synthesis of a full-length, negative-strand RNA which, in turn, is the template for the synthesis of progeny virus genomes. This process is mediated by the viral replication complex that includes all or most of the 14–16 nsps derived from the proteolytic processing of the pp1a and pp1ab replicase polyproteins of arteriviruses and coronaviruses. The replication complex, which is likely to include also cellular proteins, is associated with modified intracellular membranes, which may be important to create a microenvironment suitable for viral RNA synthesis as well as for recruitment of host factors. Electron microscopy studies of cells infected with arteriviruses (EAV) and coronaviruses (MHV and SARS-CoV) have shown that RNA synthesis is associated with virus-induced, DMVs. The origin of DMVs is under debate and different intracellular compartments including the Golgi, late endosomal membranes, autophagosomes, and the endoplasmic reticulum have been implicated in their formation.

Studies of *cis*-acting sequences required for nidovirus replication have mainly relied on coronavirus defective-interfering (DI) RNAs replicated by helper virus. Genome regions harboring minimal *cis*-acting sequences have been mapped to around 1 kb domains of the genomic 5' and 3' ends. Studies with MHV DI RNAs have indicated that both genome ends are necessary for positive-strand synthesis, whereas only the last 55 nt and the poly (A) tail at the genomic 3' end are required for negative-strand synthesis. It has been postulated that the 5' and 3' ends of the genome may interact directly during RNA replication, as predicted by computer-aided simulations of MHV and TGEV genomic RNA interactions in protein-free media. There is, however, some experimental evidence supporting protein-mediated cross-talk between both genome ends in the form of RNA–protein and protein–protein interactions.

Several experimental approaches have implicated, in addition to the nsps encoded by the replicase gene, the N protein in coronavirus RNA synthesis. Early in infection, the coronavirus N protein colocalizes with the site of viral RNA synthesis. In addition, the N protein can enhance the rescue of various coronaviruses from synthetic full-length RNA, transcribed *in vitro* or from cDNA clones. In contrast, arterivirus RNA synthesis does not require the N protein.

Host factors that may participate in nidovirus RNA synthesis have been identified mainly from studies of coronaviruses and arteriviruses. In coronaviruses (MHV and TGEV), heterogeneous nuclear ribonucleoprotein (hnRNP) A1 has been identified as a major protein binding to genomic RNA sequences complementary to those in the negative-strand RNA that bind another cellular protein, polypyrimidine tract-binding protein (PTB). hnRNP A1 and PTB bind to the complementary strands at the 5' end of coronavirus RNA and could mediate the formation of an RNP replication complex involving the 5' and 3' ends of coronavirus genomic RNA. The functional relevance of hnRNP A1 in coronavirus replication was supported by experiments showing that its overexpression promotes MHV replication, whereas replication was reduced in cells expressing a dominant-negative mutant of hnRNP A1. There is also experimental evidence to suggest that the poly(A)-binding protein (PABP) specifically interacts with the 3' poly(A) tail of coronavirus genomes, and that this interaction may affect their replication. Other cellular proteins found to bind to coronavirus genomic RNA, such as aconitase and the heat shock proteins HS40 and HS70, might be involved in modulating coronavirus replication. Similarly, interactions of cellular proteins such as transcription cofactor p100 with the EAV nsp1, or of PTB or fructose biphosphate aldolase A with SHFV genomic RNA, suggest that, in arterivirus replication also, a number of cellular proteins may be involved.

## Transcription

RNA-dependent RNA transcription in some members of the *Nidovirales* (coronaviruses, bafiniviruses, and arteriviruses), but not in others (roniviruses), includes a discontinuous RNA-synthesis step. This process occurs during the production of subgenome-length

negative-strand RNAs that serve as templates for transcription and involves the fusion of a copy of the genomic 5'-terminal leader sequence to the 3' end of each of the nascent RNAs complementary to the coding (body) sequences (Fig. 6). The resulting chimeric sg RNAs of negative polarity are transcribed to yield sg mRNAs that share both 5'- and 3'- terminal sequences with the genome RNA. Genes expressed through sg mRNAs are preceded by conserved 'transcription-regulatory sequences' (TRSs) that presumably act as attenuation or termination signals during the production of the subgenome-length negative-strand RNAs. In arteriviruses and coronaviruses, the TRSs preceding each ORF are presumed to direct attenuation of negative-strand RNA synthesis, leading to the 'jumping' of the nascent negative-strand RNA to the leader TRS (TRS-L). This process is guided by a base-pairing interaction between complementary sequences (leader TRS and body TRS complement) and it has been proposed that template switching only occurs if the free energy ( $\Delta G$ ) for the formation of this duplex reaches a minimum threshold. This process is named 'discontinuous extension of minus strands' and can be considered a variant of similarity-assisted template-switching that operates during viral RNA recombination. The genome and sg mRNAs share a 5'-leader sequence of 55–92 nt in coronaviruses and 170–210 nt in arteriviruses.

Toroviruses are remarkable in that they employ a mixed transcription strategy to produce their mRNAs. Of their four sg mRNA species, the smaller three (mRNAs 3 through 5) lack a 5' common leader and are produced via nondiscontinuous RNA synthesis. In contrast, sg mRNA2 has a leader sequence that matches the 5'-terminal 18 nt of the genomic RNA and its production requires a discontinuous RNA-synthesis step reminiscent of, but not identical, to that seen in arteri- and coronaviruses.

Synthesis of torovirus mRNAs 3 through 5, and possibly of the two mRNAs in roniviruses, is thought to require the premature termination of negative-strand RNA synthesis at conserved, intergenic, TRS-like sequences to generate subgenome-length negative-strand RNAs that can be used directly as templates for sg mRNA synthesis. In the case of torovirus mRNA2, a TRS is lacking. Fusion of noncontiguous sequences seems to be controlled by a sequence element consisting of a hairpin structure and 3' flanking stretch of 23 residues with sequence identity to a region at the 5' end of the genome. It is thought that during negative-strand synthesis, the hairpin structure may cause the transcriptase complex to detach, prompting a template switch similar to that seen in arteri- and coronaviruses.

In addition to regulatory RNA sequences, viral and host components involved in protein–RNA and protein–protein recognition are likely to be important in transcription. For example, the arterivirus nsp1 protein has been identified as a factor that is dispensable for genome replication but absolutely required for sg RNA synthesis. The identification of host factors participating in nidovirus transcription is a field under development and specific binding assays have recently identified a limited number of cellular proteins that associate with *cis*-acting RNA regulatory sequences. For example, differences in affinity of such factors for body TRSs might regulate transcription in nidoviruses by a mechanism similar to that of the DNA-dependent RNA-polymerase I termination system, in which specific proteins bind to termination sequences.

## Origin of Nidoviruses

The complex genetic plan and the replicase gene of nidoviruses must have evolved from simpler ones. Using this natural assumption, a speculative scenario of major events in nidovirus evolution has been proposed. It has been speculated that the most recent common ancestor of the *Nidovirales* had a genome size close to that of the current arteriviruses. This ancestor may have evolved from a smaller RNA virus by acquiring the two nidovirus genetic marker domains represented by the helicase-associated zinc-binding domain (ZBD) and the NendoU function. These two domains may have been used to improve the low fidelity of RdRp-mediated RNA replication, thus generating viruses capable of efficiently replicating genomes of about 14 kbp. The subsequent evolution of much larger nidovirus genomes may have been accompanied by the acquisition of the ExoN domain. This domain may have further improved the fidelity of RNA replication through its 3'–5' exonuclease activity, which might operate in proofreading mechanisms similar to those employed by DNA-based life forms. It has been suggested that the ORF1b-encoded HEL, ExoN, NendoU, and O-MT domains may provide RNA specificity, whereas the relatively abundantly expressed CPD and ADRP might control the pace of a common pathway that could be part of a hypothetical oligonucleotide-directed repair mechanism used in the present coronaviruses and roniviruses. The expansion of the replicase gene may have been associated with an increase in replicase fidelity, thus also supporting the further expansion of the 3'-proximal genome region to encode the structural proteins required to form complex enveloped virions.

## Effect of Nidovirus Infection on the Host Cell

Compared to other viruses, the interactions of nidoviruses with their hosts have not been studied in great detail. In many cases, information is based on relatively few studies performed on a limited number of viruses from the families *Coronaviridae* and *Arteriviridae*. Also, most studies have been performed with viruses that have been adapted to cell culture and therefore may have properties that differ from those of field strains. Coronaviruses and arteriviruses are clearly the best-studied members of the *Nidovirales* in terms of their interactions with the host.

Coronavirus infection affects cellular gene expression at the level of both transcription and translation. Upon infection, host cell translation is significantly suppressed but not shut off, as is the case in several other positive-RNA viruses. The underlying mechanisms have not been characterized in detail, but data obtained for MHV and BCoV suggest that they may involve the 5'-leader sequences present on coronavirus mRNAs. The viral N protein was reported to bind to the 5'-common leader sequence and

it has been speculated that this might promote translation initiation, leading to a preferential translation of viral mRNAs. Furthermore, host mRNAs were reported to be specifically degraded in MHV- and SARS-CoV-infected cells, further reducing the synthesis of cellular proteins. Another mechanism affecting host cell protein synthesis may be based on specific cleavage of the 28S rRNA subunit, which was observed in MHV-infected cells.

Studies on cellular gene expression following nidovirus infections have mainly focused on the coronaviruses MHV and SARS-CoV. For example, SARS-CoV infection was reported to disrupt cellular transcription to a larger extent than does HCoV-229E. Differences in cellular gene expression have been proposed to be linked to differences in the pathogenesis caused by these two human coronaviruses. Apart from the downregulation of genes involved in translation and cytoskeleton maintenance, genes involved in stress response, proapoptotic, proinflammatory, and procoagulating pathways were significantly upregulated. Both MHV and SARS-CoV induce mitogen-activated phosphate kinases (MAPKs), especially p38 MAPK. In addition, activation of AP-1, nuclear factor kappa B (NF- $\kappa$ B), and a weak induction of Akt signaling pathways occur after SARS-CoV infection and the N and nsp1 proteins were suggested to be directly involved in inducing these signaling pathways.

Nidoviruses have also been reported to interfere with cell cycle control. Infection by the coronaviruses TGEV, MHV, SARS-CoV, and IBV was reported to cause a cell cycle arrest in the G0/G1 phase and a number of cellular proteins (e.g., cyclin D3 and hypophosphorylated retinoblastoma protein) and viral proteins (MHV nsp1, SARS-CoV 3b 7a, and N proteins) have been proposed to be involved in the cell cycle arrest in G0/G1.

Many viruses encode proteins that modulate apoptosis and, more generally, cell death, which allows for highly efficient viral replication or the establishment of persistent infections. Infection by coronaviruses (e.g., TGEV, MHV, and SARS-CoV) and arteriviruses (e.g., PRRSV and EAV) have been reported to induce apoptosis in certain cell types. Apoptosis has also been reported in shrimp infected with the ronavirus YHV and is thought to be involved in pathogenesis. Both apoptotic and antiapoptotic molecules have been found to be upregulated, suggesting that a delicate counterbalance of pro- and antiapoptotic molecules is required to ensure cell survival during the early phase of infection, and rapid virus multiplication before cell lysis occurs. Coronavirus-induced apoptosis appears to occur in a tissue-specific manner, which obviously has important implications for viral pathogenesis. For instance, SARS-CoV was shown to infect epithelial cells of the intestinal tract and induce an antiapoptotic response that may counteract a rapid destruction of infected enterocytes. These findings are consistent with clinical observations of a relatively normal endoscopic and microscopic appearance of the intestine in SARS patients. Furthermore, SARS-CoV causes lymphopenia which involves the depletion of T cells, probably by apoptotic mechanisms that are triggered by direct interactions of the SARS-CoV E protein with the antiapoptotic factor Bcl-xL. Also the MHV E protein has been reported to induce apoptosis. The SARS-CoV 7a protein was found to induce apoptosis in cell lines derived from lung, kidney, and liver, by a caspase-dependent pathway. Apoptosis has also been associated with arterivirus infection but information on underlying mechanisms and functional implications is limited.

Coronavirus and arterivirus infections trigger proinflammatory responses that often are associated with the clinical outcome of the infection. Thus, for example, there seems to be a direct link between the IL-8 plasma levels of SARS patients and disease severity, similar to what has been described for pulmonary infections caused by respiratory syncytial virus. In contrast, despite the upregulation of IL-8 in intestinal epithelial cells, biopsy specimens taken from the colon and terminal ileum of SARS patients failed to demonstrate any inflammatory infiltrates, which may be the consequence of a virus-induced suppression of specific cytokines and chemokines, including IL-18, in the intestinal environment.

Innate immunity is essential to control vertebrate nidovirus infection *in vivo*. The induction of type I IFN (IFN- $\alpha/\beta$ ) varies among different coronaviruses and arteriviruses. Whereas some coronaviruses such as TGEV are potent inducers of type I IFN, other coronaviruses (MHV and SARS-CoV) or arteriviruses (PRRSV) do not stimulate its production, thus facilitating virus escape from innate immune defenses. Type I interferon is a key player in innate immunity and in the activation of effective adaptive immune responses. Upon viral invasion, IFN- $\alpha/\beta$  is synthesized and secreted. IFN- $\alpha/\beta$  molecules signal through the type I interferon receptor (IFNR), inducing the transcription of several antiviral mediators, including IFN- $\gamma$ , PKR, and Mx. IFN- $\gamma$  is critical in resolving coronavirus (MHV and SARS-CoV), and also arterivirus (EAV, LDV, and PRRSV) infections. Like many other viruses, coronaviruses have developed strategies to escape IFN responses. For example, it has been shown that the SARS-CoV 3b, 6, and N proteins antagonize interferon by different mechanisms, even though all these proteins inhibit the expression of IFN by interfering with the function of IRF-3.

In arteriviruses such as PRRSV, IFN- $\gamma$  is produced soon after infection to promote Th1 responses. However, PRRSV infections or vaccination with attenuated-live PRRSV vaccines cause only limited IL-1, TNF- $\alpha$ , and IFN- $\alpha/\beta$  responses. This then leads to IFN- $\gamma$  and Th1 levels that fail to elicit strong cellular immune responses.

*See also:* Coronaviruses: General Features (*Coronaviridae*). Coronaviruses: Molecular Biology (*Coronaviridae*). Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS) (*Coronaviridae*)

## Further Reading

- de Groot, R.J., 2007. Molecular biology and evolution of toroviruses. In: Snijder, E.J., Gallagher, T., Perlman, S. (Eds.), *The Nidoviruses*. Washington, D.C.: A.S.M. Press, pp. 133–146.
- Enjuanes, L. (Ed.), 2005. *Current Topics in Microbiology and Immunology*, vol. 287: Coronavirus Replication and Reverse Genetics. Berlin: Springer.
- Enjuanes, L., Almazan, F., Sola, I., Zuniga, S., 2006. Biochemical aspects of coronavirus replication and virus–host interaction. *Annual Review of Microbiology* 60, 211–230.
- Gorbalenya, A.E., Enjuanes, L., Ziebuhr, J., Snijder, E.J., 2006. *Nidovirales: Evolving the largest RNA virus genome*. *Virus Research* 117, 17–37.

- Masters, P.S., 2006. The molecular biology of coronaviruses. *Advances in Virus Research* 66, 193–292.
- Sawicki, S.G., Sawicki, D.L., Siddell, S.G., 2007. A contemporary view of coronavirus transcription. *Journal of Virology* 81, 20–29.
- Siddell, S.G., Ziebuhr, J., Snijder, E.J., 2005. Coronaviruses, toroviruses, and arteriviruses. In: Mahy, B.W.J., ter-Meulen, V. (Eds.), *Virology*, 10th edn., vol. 1. London: Hodder-Arnold, pp. 823–856.
- Snijder, E.J., Siddell, S.G., Gorbalenya, A.E., 2005. The order *Nidovirales*. In: Mahy, B.W.J., ter-Meulen, V. (Eds.), *Virology*, 10th edn., vol. 1. London: Hodder-Arnold, pp. 390–404.
- Snijder, E.J., Spaan, W.J.M., 2007. Arteriviruses. In: Knipe, D.M., Howley, P.M., Griffin, D.E., *et al.* (Eds.), *Fields Virology*, vol. 1. Philadelphia: Lippincott Williams and Wilkins, pp. 1205–1220.
- Spaan, W.J.M., Cavanagh, D., de Groot, R.J., *et al.*, 2005. Nidovirales. In: Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A. (Eds.), *Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses*. San Diego, CA: Elsevier Academic Press, pp. 937–945.
- van Vliet, A.L.W., Smits, S.L., Rottier, P.J.M., de Groot, R.J., 2002. Discontinuous and non-discontinuous subgenomic RNA transcription in a nidovirus. *EMBO Journal* 21 (23), 6571–6580.
- Walker, P.J., Bonami, J.R., Boonsaeng, V., *et al.*, 2005. Roniviridae. In: Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A. (Eds.), *Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses*. San Diego, CA: Elsevier Academic Press, pp. 975–979.
- Ziebuhr, J., Snijder, E.J., 2007. The coronavirus replicase: Special enzymes for special viruses. In: Thiel, V. (Ed.), *Molecular and Cellular Biology: Coronaviruses*. Norfolk, UK: Caister Academic Press, pp. 31–61.
- Zuñiga, S., Sola, I., Alonso, S., Enjuanes, L., 2004. Sequence motifs involved in the regulation of discontinuous coronavirus subgenomic RNA synthesis. *Journal of Virology* 78, 980–994.