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.
Coronaviruses: General Features

D Cavanagh and P Britton, Institute for Animal Health, Compton, UK
© 2008 Elsevier Ltd. All rights reserved.

Introduction

Coronaviruses are known to cause disease in humans, other mammals, and birds. They cause major economic loss, sometimes associated with high mortality, in neonates of some domestic species (e.g., chickens, pigs). In humans, they are responsible for respiratory and enteric diseases. Coronaviruses do not necessarily observe species barriers, as illustrated most graphically by the spread of severe acute respiratory syndrome (SARS) coronavirus among wild animals and to man, with lethal consequences. As a group, coronaviruses are not limited to particular organs; target tissues include the nervous system, immune system, kidney, and reproductive tract in addition to many parts of the respiratory and enteric systems. A great advance in recent years has been the development of systems (‘infectious clones’) for modifying the genomes of coronaviruses to study all aspects of coronavirus replication, and for the development of new vaccines.

Taxonomy and Classification

The genus Coronavirus together with the genus Torovirus form the family Coronaviridae; members of these two genera are similar morphologically. The Coronaviridae, Arteriviridae, and Roniviridae are within the order Nidovirales. Members of this order have a similar genome organization and produce a nested set of subgenomic mRNAs (nidus, Latin for nest). To date, coronaviruses have been placed into one of three groups (Table 1). Initially, this was on the basis of serological relationships which subsequently have been supported by gene sequencing.

Vorion Properties

Virions have a buoyant density of approximately 1.18 g ml⁻¹ in sucrose. Being enveloped viruses (Figure 1(a)), they are destroyed by organic solvents such as ether and chloroform.

Vorion Structure and Composition

All coronaviruses have four structural proteins in common (Figure 1(b)): a large surface glycoprotein (S; c. 1150–1450 amino acids); a small envelope protein (E; c. 100 amino acids, present in very small amounts in virions); integral membrane glycoprotein (M; c. 250 amino acids); and a phosphorylated nucleocapsid protein (N; c. 500 amino acids). Group 2a viruses have an additional structural glycoprotein, the hemagglutinin-esterase protein (HE; c. 425 amino acids). This is not essential for replication in vitro and may affect tropism in vivo.

Virions are c. 120 nm in diameter, although they can be up to twice that size, and the ring of S protein spikes is approximately 20 nm deep. When present, the HE protein forms a layer 5–10 nm deep. In some species, the S protein is cleaved into two subunits, the N-terminal S1 fragment being slightly smaller than the C-terminal S2 sequence. The S protein is anchored in the envelope by a transmembrane region near the C-terminus of S2. The functional S protein is highly glycosylated and exists as a trimer. The bulbous outer part of the mature S protein is formed largely by S1 while the stalk is formed largely by S2, having a coiled-coil structure. S1 is the most variable part of the S protein; some serotypes of IBV differ from one another by 40% of S1 amino acids. S1 is the major inducer of protective immune responses. Variation in the S1 protein enables one strain of virus to avoid immunity induced by another strain of the same species.

The M glycoprotein is the most abundant protein in virions. In most cases, only a small part (~20 amino acids) at the N-terminus protrudes at the surface of the virus. There are three membrane-spanning segments and the C-terminal half of the M protein is within the lumen of
the virus. In transmissible gastroenteritis virus (TGEV), a proportion of M molecules have four membrane-spanning segments, resulting in the C-terminus also being exposed on the outer surface of the virus (M in Figure 1(b)). The E protein is anchored in the membrane by a sequence near its N-terminus.

### Table 1  Species of coronavirus

| Group 1 | Group 2 | Group 3 |
|---------|---------|---------|
| **Subgroup 1a** | **Subgroup 2a** | **Infectious bronchitis virus (IBV)** |
| Transmissible gastroenteritis virus (TGEV) | Murine hepatitis virus (MHV) |  |
| Feline coronavirus (FCoV) | Bovine coronavirus (BCoV) | Turkey coronavirus (TCoV) |
| Canine coronavirus (CCoV) | Porcine haemagglutinating encephalomyelitis virus (HEV) | Pheasant coronavirus (PhCoV) |
| Ferret coronavirus (FeCoV) | Equine coronavirus (EqCoV) |  |
| **Subgroup 1b** |  | **Duck coronavirus** (DCoV) |
| Human coronavirus (HCoV) 229E | Human coronavirus (HCoV) OC43 | Goose coronavirus (GCoV) |
| Porcine epidemic diarrhoea virus (PEDV) | Human enteric coronavirus (HECoV) | Pigeon coronavirus (PiCoV) |
| Bat coronavirus-61 | Rat coronavirus (RCoV) |  |
| Bat coronavirus-HKU2 | Puffinosis coronavirus |  |
| Human coronavirus NL63 | **Subgroup 2b** |  |
|  | Severe acute respiratory syndrome coronavirus (SARS-CoV) |  |
|  | Bat-CoV-HKU9-1 |  |

*Recognized virus species.

The viruses duck coronavirus, goose coronavirus, and pigeon coronavirus have been recommended by the Coronavirus Study Group for recognition as species by the ICTV. Official virus species names are in italics. Coronaviruses that have not yet been recognized as distinct species have names that are not italicized. Rabbit coronavirus is considered as a tentative member of the genus. A coronavirus has been isolated from a parrot. On the basis of very limited sequence data, it is not clear in which, if any, of the three groups that this virus would be placed.

### Genome Organization and Expression

Coronaviruses have the largest known RNA genomes, which comprise 28–32 kb of positive sense, single-stranded RNA. The overall genome organization is being 5′ UTR–polymerase gene–structural protein genes–3′ UTR, where the UTRs are untranslated regions (Figure 2). The first 60–90 nucleotides at the 5′ end form a leader sequence. The coronavirus membrane proteins, S, E, M, and M′, are inserted into a lipid bilayer (MEM) derived from internal cell membranes.

(a) Reproduced from Gonzalez JM, Gomez-Puertas P, Cavanagh D, Gorbalenya AE, and Enjuanes L (2003) A comparative sequence analysis to revise the current taxonomy of the family Coronaviridae. Archives of Virology 148: 2207–2235, with permission from Springer-Verlag.
5 of murine hepatitis virus (MHV)), translation of the third and second open reading frame (ORF), respectively, is effected by the preceding ORFs acting as internal ribosome entry sites. The proteins encoded by these small ORFs are mostly not required for replication in vitro; some of them might function as antagonists of innate immune responses, though this has not yet been demonstrated.

Following entry into a cell and the release of the virus ribonucleoprotein (genome surrounded by the N protein) into the cytoplasm, ribosomes translate gene 1, which is approximately 20 kb, into two polyproteins (pp1a and pp1ab). These are cleaved by gene 1-encoded proteases, to generate 15 or 16 proteins (Figure 3). Translation of ORF 1b involves ribosomal frameshifting, which has two elements, a slippery site followed by an RNA pseudoknot. At the slippery site (UUUAAC in IBV), the ribosome slips one nucleotide backward and then moves forward, this time in a –1 frame compared with translation ORF 1a, resulting in the synthesis polyprotein 1ab. Translation of the genomic RNA results in the production of pp1a. However, the translating ribosomes undergo the –1 frameshift about 30% of the time resulting in pp1ab. The 5’ and 3’ UTR sequences are represented as single lines downstream of the leader and N gene sequences, respectively.

Replication Cycle

The N-terminal (S1) part of the S protein mediates that mediates attachment to cells. It is a determinant of host species specificity and, in some cases, pathogenicity, by determining susceptible cell range (tissue tropism) within a host. The C-terminal S2 part triggers fusion of the virus envelope with cell membranes (plasma membrane or endosomal membranes), which can occur at neutral or slightly acidic pH, depending on species or even strain. The virus glycoproteins (S, M, and HE, when present) are synthesized at the endoplasmic reticulum. Both subunits
of the S protein are multiply glycosylated, while the M protein has one or two glycans close to its N-terminus. Interestingly, glycosylation of the M protein can be either N- or O-linked, depending on the type of coronavirus, although experiments using reverse genetics showed that conversion of an O-linked glycosylated M protein to an N-linked version had no effect on virus growth.

Early and late in infection, formation of virus particles can occur in the endoplasmic reticulum–Golgi intermediate compartment (ERGIC) and endoplasmic reticulum, but most assembly occurs in the Golgi membranes. The M protein is not transported to the plasma membrane; its location at internal membranes determines the sites of virus particle formation. It interacts with the N protein (as part of the RNP) and C-terminal part of the S protein, retaining some, though not all, of the S protein at internal membranes. The E protein is essential for virus particle formation, though it is not known how it functions. It has a sequence that determines its accumulation at internal membranes, and its interaction with the M protein. The latter interacts with the C-terminus of the S protein, retaining some of it at internal membranes, and with the N protein (itself part of the ribonucleoprotein structure), enabling the formation of virus particles with spikes.

**Genome Replication and Recombination**

Following infection of a susceptible cell, the coronavirus genomic RNA is released from the virion into the cytoplasm and immediately recognized as an mRNA for the translation of the replicase pp1a and pp1ab proteins. These proteins are cleaved by ORF1a-encoded proteases, after which they become part of replicase complexes for the synthesis of either complete negative-sense copies of the genomic RNA or negative-sense copies of the sg mRNAs. The negative-sense RNAs are used as templates for the synthesis of genomic RNA and sg mRNAs (Figure 2). Following synthesis of the sg mRNAs, the structural proteins are produced for the assembly and encapsidation of the de novo-synthesized genomic RNA, resulting in the release of new infectious coronavirus virions. The release of new virions starts 3–4 h after the initial infection. As indicated above, the synthesis of the sg mRNAs is the result of a discontinuous process in which the synthesis of a negative-sense copy of an sg mRNA is completed by the addition of the negative-sense leader sequence by a recombination mechanism. If a cell is infected with two related coronaviruses, the polymerase may swap between two RNA templates, in a similar way to addition of the leader sequence. This ‘copy-choice’ mechanism of genetic recombination results in a chimeric RNA. Such RNAs may give rise to new viruses with modified genomes with a capacity to infect a different cell and, in some cases, new host species.

**Evolutionary Relationships among Coronaviruses**

Phylogenetic analyses of the structural proteins have resulted in the grouping of coronavirus species in accordance with earlier antigenic groups (Table 1 and Figure 4).
Members of subgroups have higher amino acid sequence identities to each other (≥60%) than to members of another group in the same group (with which they share ≤40% identity). Comparing one group with another, protein sequence identities are generally in the range 25–35%.

| Group 1 | Group 2 | Group 3 |
|---------|---------|---------|
| HCoV NL63 | TGEV | Goose-CoV |
| HCoV 229E | PEDV | Pigeon-CoV |
| HCoV OC43 | FCoV | Teal-CoV |
| HCoV HKU1 | MHV | Partridge-CoV |
| SARS-CoV | MHV A59 | IBV |
| Bat-CoV HKU3-1 | MHV JHM | IBV |
| MHV | MHV | IBV |

**Figure 4**  Phylogenetic relationship of aligned coronavirus-derived nucleoprotein amino acid sequences. The complete N protein sequences represent coronaviruses from each of the three groups (Table 1). The tree is unrooted and the three main coronavirus groups, 1–3, are highlighted as dark gray ellipsoids. Groups 1 and 2 are divided into two subgroups, a and b, representing some divergence of the sequences within their corresponding groups. Similar relationships are observed when comparing other structural proteins and replicase-derived proteins.

**Diseases and Host Range**

Probably all coronaviruses replicate in epithelial cells of the respiratory and/or enteric tracts, though not necessarily producing clinical damage at those sites. Avian IBV not only causes respiratory disease but can also damage gonads in both females and males, and causes serious kidney disease (dependent on the strain of virus, and to some extent on the breed of chicken). IBV is able to replicate at virtually every epithelial surface in the host. Some coronaviruses have their most profound effect in the alimentary tract (e.g., porcine TGEV causes ≥90% mortality in neonatal pigs). Human coronaviruses are known to be associated with enteric and respiratory diseases (e.g., diarrhea), in addition to respiratory disease. SARS-CoV was also associated with diarrhea in humans, in addition to serious lung disease. Other coronaviruses, for example, MHV and porcine HEV, spread to cells of the central nervous system, producing disease, for example, acute or chronic demyelination in the case of MHV.

Coronavirus replication and disease are not necessarily restricted to a single host species. Canine enteric CoV and feline CoV can replicate and cause disease in pigs; these two viruses have proteins with very high amino acid identity to those of porcine TGEV. Canine respiratory CoV has proteins, including the S protein (which is the attachment protein and a determinant of host range), with very high amino acid identity (≥95%) to other group 2 viruses Hu CoV-OC43 and BCoV. This raises the possibility of co-infection in these hosts. Bovine CoV causes enteritis in turkeys following experimental oral infection. There is evidence that pheasant CoV can infect chickens, and IBV infect teal (a duck), though without causing disease. The most dramatic demonstration that coronaviruses can have a wide host range was provided by SARS-CoV. This may have had its origin in bats, was transferred to various other species (e.g., civet cat) that were captured for trade, and then caused lethal disease in humans.

Persistent infections *in vivo* are well known for MHV, and less well known for other coronaviruses (e.g., IBV). Following infection of very young chickens, IBV is re-excreted when hens start to lay eggs. The trigger for release is probably the stress of coming into lay.

The S protein is a determinant of both tissue tropism within a host and host range. This has been elegantly demonstrated by genetic manipulation of the genome of MHV, which is unable to attach to feline cells. Replacement of the MHV S protein gene with that of CoV from feline coronavirus resulted in a recombinant virus that was able to attach, and subsequently replicate in, feline cells. However, other proteins can also affect pathogenicity. Research with genetically modified coronaviruses, using targeted recombination or ‘infectious clones’, has shown that modifications to proteins encoded in ORF1 and the small genes interspersed among the structural protein genes, result in attenuation of pathogenicity. Although the roles of these ‘accessory proteins’ are not known, this may offer a route to the development of a new generation of live vaccines. Currently, the most widely used prophylactics for control of IBV in chickens include killed vaccines and live vaccines attenuated by passage in embryonated eggs. However, disease control is complicated by extensive variation in the S1 protein which is the inducer of protective immunity.
Coronaviruses: General Features

Introduction

Coronaviruses (CoVs) were first identified during the 1960s by using electron microscopy to visualize the distinctive spike glycoprotein projections on the surface of enveloped virus particles. It was quickly recognized that CoV infections are quite common, and that they are responsible for seasonal or local epidemics of respiratory and gastrointestinal disease in a variety of animals. CoVs have been named according to the species from which they were isolated and the disease associated with the viral infection. Avian infectious bronchitis virus (IBV) infects chickens, causing respiratory infection, decreased egg production, and mortality in young birds. Bovine coronavirus (BCoV) causes respiratory and gastrointestinal disease in cattle. Porcine transmissible gastroenteritis virus (TGEV) and porcine epidemic diarrhea virus (PEDV) cause gastroenteritis in pigs. These CoV infections can be fatal in young animals. Feline infectious peritonitis virus (FIPV) and canine coronavirus (CCoV) can cause severe disease in cats and dogs. Depending on the strain of the virus and the site of infection, the murine CoV mouse hepatitis virus (MHV) can cause hepatitis or a demyelinating disease similar to multiple sclerosis. CoVs also infect humans. Human coronaviruses (HCoVs) 229e and OC43 are detected worldwide and are estimated to be responsible for 5–30% of common colds and mild gastroenteritis. Interestingly, HCoV-OC43 and BCoV share considerable sequence similarity, indicating a likely transmission across species (either from cows to humans or vice versa) and then adaptation of the virus to its host. In contrast to the relatively mild infections caused by HCoV-229e and HCoV-OC43, the CoV responsible for severe acute respiratory syndrome (SARS-CoV) causes atypical pneumonia with a 10% mortality rate. Two additional HCoVs, HCoV-NL63 and HCoV-HKU1, have been recently identified using molecular methods and are associated with upper and lower respiratory tract infections in children, and elderly