Characterization of *White Bream Virus* Reveals a Novel Genetic Cluster of Nidoviruses

Heike Schütze,1,2 Rachel Ulferts,3 Barbara Schelle,4 Sonja Bayer,4 Harald Granzow,2
Bernd Hoffmann,5 Thomas C. Mettenleiter,1 and John Ziebuhr1,6

Institutes of Molecular Biology,1 Infectology,2 and Diagnostic Virology,3 Friedrich-Loeffler-Institut, 17493 Greifswald-Insel Riems, Germany; Centre for Cancer Research and Cell Biology, School of Biomedical Sciences, The Queen’s University of Belfast, 97 Lisburn Road, Belfast BT9 7BL, United Kingdom;4 and Institute of Virology and Immunology, University of Würzburg, Versbacher Straße 7, 97078 Würzburg, Germany5

Received 14 August 2006/Accepted 12 September 2006

The order Nidovirales comprises viruses from the families *Coronaviridae* (genera *Coronavirus* and *Torovirus*), *Roniviridae* (genus *Okavirus*), and *Arteriviridae* (genus *Arterivirus*). In this study, we characterized *White bream virus* (WBV), a bacilliform plus-strand RNA virus isolated from fish. Analysis of the nucleotide sequence, organization, and expression of the 26.6-kb genome provided conclusive evidence for a phylogenetic relationship between WBV and nidoviruses. The polycistronic genome of WBV contains five open reading frames (ORFs), called ORF1a, -1b, -2, -3, and -4. In WBV-infected cells, three subgenomic RNAs expressing the structural proteins S, M, and N were identified. The subgenomic RNAs were revealed to share a 42-nucleotide, 5′-leader sequence that is identical to the 5′-terminal genome sequence. The data suggest that a conserved nonanucleotide sequence, CA(G/A)CACUAC, located downstream of the leader and upstream of the structural protein genes acts as the core transcription-regulating sequence element in WBV. Like other nidoviruses with large genomes (>26 kb), WBV encodes in its ORF1b an extensive set of enzymes, including putative polymerase, helicase, ribose methyltransferase, exoribonuclease, and endoribonuclease activities. ORF1a encodes several membrane domains, a putative ADP-ribose 1′-phosphatase, and a chymotrypsin-like serine protease whose activity was established in this study. Comparative sequence analysis revealed that WBV represents a separate cluster of nidoviruses that significantly diverged from toroviruses and, even more, from coronaviruses, roniviruses, and arteriviruses. The study adds to the amazing diversity of nidoviruses and appeals for a more extensive characterization of nonmammalian nidoviruses to better understand the evolution of these largest known RNA viruses.

---

* Corresponding author. Mailing address: The Queen’s University of Belfast, School of Biomedical Sciences, 97 Lisburn Road, Belfast BT9 7BL, United Kingdom. Phone: 44-28-90972089. Fax: 44-28-90972124. E-mail: j.ziebuhr@qub.ac.uk.

7 Published ahead of print on 20 September 2006.
(ADP-ribose)-binding and/or ADP-ribose 1'-phosphatase (ADRP) activities, and cyclic nucleotide phosphodiesterase activities (25, 40, 43, 54). The individual subunits of the nidovirus replicase machinery are released from pp1a/pp1b by autoproteolytic processing, involving viral proteases that themselves are part of the polyproteins (75).

Recently, significant progress was made in the functional and structural characterization of the nidovirus replication/transcription complex, with Severe acute respiratory syndrome coronavirus (SARS-CoV) being the most extensively studied nidovirus to date (1, 2, 4, 19, 30, 31, 34, 42, 43, 48, 57, 58, 68–70). It is now generally accepted that the enzymology involved in nidovirus replication is significantly more complex than that of other plus-strand RNA viruses. Most probably, these additional enzymes are required to replicate the exceptionally large genomes of nidoviruses and synthesize the nested set of sg RNAs from which the structural and, in some cases, several accessory proteins are expressed (37, 47). Previous studies identified conserved proteins and mechanisms mediating the replication cycle of viruses from the various nidovirus genera, but they also revealed interesting differences. Thus, for example, most, but not all, nidovirus sg RNAs contain a 5′ leader sequence derived from the 5′ end of the genomic RNA (13, 17, 56, 65). Furthermore, the number of replicase gene-encoded enzymes varies between different nidovirus genera or even between different groups of the same genus (23, 54), and also, the functional and structural properties of several key replicative proteins, including the main proteases of the various genera, are remarkably diverse (1, 3, 53, 72). Taken together, the studies have made it clear that more information is needed to understand the biological meaning of the differential conservation of specific proteins and/or enzymatic activities among nidoviruses (and other plus-strand RNA virus families). In this context, additional sequence information, particularly from nidoviruses prototyping previously unknown and distantly related genera, can be expected to provide new insights into the evolution of the Nidovirales and, potentially, even provide clues for a better understanding of the mechanisms and driving forces that have governed the separation of the Nidovirales from the bulk of plus-strand RNA viruses featuring smaller genomes and less-complex genome replication and expression strategies.

Here, we report the complete genome sequence of White bream virus (WBV), a novel plus-strand RNA virus isolated from fish (Blicca bjoerkna L.) (27). The sequence, organization, and expression of the 26.6-kb genome qualify WBV as a new member of the order Nidovirales. In common with other nidoviruses, the WBV genome contains two large ORFs (1a and 1b) in the 5′ region of the genome, with ORF1b probably being expressed by −1 ribosomal frameshifting, occurring at a putative UUUAAAC heptanucleotide “slippery” sequence located upstream of the ORF1a stop codon and a putative RNA pseudoknot structure. WBV is predicted to encode three structural proteins, spike (S) protein, membrane (M) protein, and nucleocapsid (N) protein, which are expressed from 5′ leader-containing sg RNAs as shown by Northern blotting and sequence analysis. Upstream of each of the S, M, and N ORFs, a conserved nonanucleotide sequence, CA(G/A)CACUAC, was identified which is also present near the 5′ end of the genome. By analogy with other nidoviruses, we predict this sequence to be the conserved core of the transcription-regulating sequence (TRS) elements that facilitate the template switch required to attach the complement of the 5′ leader sequence to the 3′ ends of subgenome-length minus strands, thereby producing the templates for the synthesis of 5′ leader-containing plus-strand RNAs. The phylogenetic relationship between WBV and nidoviruses is further corroborated by the presence of a conserved array of putative functional domains in pp1a, including an ADRP domain, three TMs, and the presumed viral main protease, 3CLpro, whose activity was established in this study. Furthermore, an RNA-dependent RNA polymerase domain featuring the nidovirus-specific SDD signature, as well as ZBD, helicase, exoribonuclease, NendoU, and ribose-2′-O-methyltransferase domains, were identified in the C-terminal part of pp1b. Sequence comparisons and phylogenetic studies lead us to conclude that toroviruses (followed by coronaviruses) are the closest relatives of WBV. In terms of classification, we think the phylogenetic position of WBV would be best reflected if the virus was assigned to a yet-to-be-established new genus.

**MATERIALS AND METHODS**

**Cells and virus.** WBV (isolate DF24/00) was propagated on epitheliumoma papulosum cyprini (EPC) cells at 20°C in a 5% CO₂ atmosphere as described previously (27). EPC cells (RIE 173) were obtained from the cell line collection of the Friedrich-Loeffler-Institut, Federal Research Institute for Animal Health (Insel Riems, Germany).

**Virus purification and RNA isolation.** WBV was harvested by two cycles of freezing and thawing of infected cells. Cell debris was removed by centrifugation at 4,000 rpm (SW28 rotor; Beckman) for 10 min, and the supernatant was layered onto a 15% sucrose cushion and centrifuged at 20,000 rpm for 90 min. The resulting virus pellet was resuspended in STE buffer (0.1 M NaCl, 0.01 M Tris-HCl, pH 8.3, 0.01 M EDTA, pH 8.0) and applied on a continuous gradient of 5 to 30% sucrose in STE. After centrifugation at 20,000 rpm (SW28 rotor; Beckman) for 30 min, the virus-containing band was isolated and dialyzed against STE. The virus was concentrated by sedimentation at 40,000 rpm (SW60; Beckman) for 4 h, and viral RNA was extracted using guanidinium thiocyanate, followed by centrifugation through a 5.7 M cesium chloride-EDTA cushion according to standard protocols (44).

**RNA transfection.** EPC cells (5 × 10⁵) were transfected with purified WBV genome RNA. To do this, 3 μg RNA was diluted in 50 μl diethyl pyrocarbonate-treated water containing 60 U RNasin (Promega) and mixed with 200 μl of serum-free OptiMEM (Invitrogen) containing 12.5 μl of Lipofectamine transfection reagent (Invitrogen). After incubation on ice for 5 min, the transfection mix was layered onto nearly confluent EPC cells maintained in OptiMEM. After incubation for 2 hours at 20°C, the medium was replaced by Eagle’s minimal essential medium supplemented with 10% fetal bovine serum. Transfected cells were incubated for 6 days at 20°C in a 5% CO₂ atmosphere.

**Electron microscopy.** For negative staining, infected cell cultures were scraped off from the cell culture dish, pelleted by low-speed centrifugation, and resuspended in phosphate-buffered saline. Formvar-coated grids were placed for 7 min on drops of cell culture supernatant or resuspended pellet. Negative staining was performed with 2% phosphotungstic acid (pH 7.4) for 7 min. Stained grids were examined with a Philips electron microscope 400T (Eindhoven, The Netherlands).

**Cloning and sequence analysis of the WBV genome.** Purified WBV genome RNA was used as a template for the construction of cDNA libraries. Viral RNA (0.5 to 1 μg) was incubated with 0.5 μg of oligo(dT)₁₂₋₁₄ primers or 0.037 μg of random pd(N)₃ hexa(ADP-ribose)-binding and/or ADP-ribose 1′-phosphatase (ADRP) activities, and cyclic nucleotide phosphodiesterase activities (25, 40, 43, 54). The individual subunits of the nidovirus replicase machinery are released from pp1a/pp1b by autoproteolytic processing, involving viral proteases that themselves are part of the polyproteins (75).
RESULTS AND DISCUSSION

WBV genome RNA is infectious. In a previous study, WBV was shown to be an enveloped RNA virus featuring a bacilliform shape and carrying coronavirus-like projections on its surface (27; Fig. 1). To further characterize the genome of this virus, we isolated viral RNA from purified WBV particles, transfected this RNA into EPC cells, and examined by electron microscopy whether virus particles were released from these cells. As Fig. 1 shows, rod-shaped virus particles identical to those produced in WBV-infected cells could be identified in tissue culture supernatants obtained from cells transfected with purified WBV genome RNA (Fig. 1C and D) but not in supernatants from mock-transfected cells (data not shown). The data provide conclusive evidence that, in the absence of viral proteins, WBV genome RNA is able to trigger a full replication cycle, implying that the genome RNA of WBV is infectious and, therefore, must be of mRNA (that is, positive) polarity.

Genome sequence analysis reveals that WBV is a nidovirus. Except for a 5′-proximal fragment (nts 3176 to 7881), which could not be stably propagated in E. coli, the complete sequence of the WBV genome was determined from a collection of cDNA clones derived from reverse-transcribed WBV genome RNA (see Materials and Methods) (Fig. 2, and data not shown). The genome ends were determined by using RACE methods, and the completeness and correctness of the obtained sequence were further corroborated by sequence analysis of a set of overlapping RT-PCR products covering the entire genome, including the 5′-proximal fragment. The WBV genome sequence was revealed to encompass 26,628 nucleotides [excluding the 3′ poly(A) tail] and has been deposited in the GenBank database (accession number DQ898157). It contains five major ORFs, ORF1a, -1b, -2, -3, and -4 (Fig. 2; Table 1), which are flanked by 5′- and 3′-terminal untranslated regions of 905 and 228 nucleotides, respectively. The nucleotide sequence and organization of the genome unambiguously identified WBV as a nidovirus. A more detailed sequence analysis showed that WBV is only distantly related to other nidoviruses and that the virus could not readily be assigned to any of the established nidovirus taxa. However, on the basis of its genome size and replicase domain structure, WBV is clearly a member of the large nidoviruses (Coronaviridae and Roniviridae) (24). Furthermore, in database searches for related sequences, a special relationship between WBV and toro- and coronaviruses became immediately evident when WBV

Other nidoviruses will be given below.

Putative structural proteins of WBV. Database searches and comparative sequence analysis revealed a remote sequence similarity between the WBV ORF2 gene product and the S proteins of corona- and toroviruses (data not shown), leading us to propose that WBV ORF2 encodes the viral S protein (1,220 residues), which probably mediates receptor binding
and fusion between viral and cellular membranes. Further studies (data not shown) suggest that the WBV S protein, like its homologs in corona- and toroviruses (14, 52), is a type I membrane glycoprotein, featuring (i) an N-terminal signal peptide (with a predicted cleavage site between residues Ala15 and Gln16), (ii) a putative C-terminal transmembrane anchor, and (iii) a short cytoplasmic tail.

Computer-aided sequence analysis further suggested that the WBV ORF3 gene product is a 227-residue, triple-membrane-spanning glycoprotein. Both the size of this protein and the predicted topology and length of the transmembrane regions (data not shown) link this protein to the triple-membrane-spanning M proteins of corona- and toroviruses (14, 52), even though at the primary structure level, the relationship is weak and the generation of reliable sequence alignments proved to be impossible. However, this distant relationship is not really surprising as the coronavirus and torovirus S and M proteins are only poorly conserved, with similarities being restricted mainly to general features, such as the numbers and positions of transmembrane domains, protease cleavage sites, and the general domain organization of these proteins (52).

Based on its 3′-terminal position, which in corona- and toroviruses is generally occupied by the N gene, and the quite reliable functional assignments for all of the other WBV gene products, it seemed reasonable to suggest that ORF4 specifies the WBV N protein (161 residues). The size of this protein corresponds well to that of the torovirus (but not coronavirus) N proteins, and also, some of the most conserved sequence signatures of torovirus N proteins appear to be partly conserved in the putative WBV N protein (data not shown). However, to unambiguously establish a structural relationship between the WBV and torovirus N proteins, further evidence has to be obtained. In support of this potential relationship, we note that the intracellular nucleocapsids formed by WBV and toroviruses share a rod-like structure (55) and also that the straight or slightly bent (kidney-like) rod-like morphology described for some (but not all) torovirus particles resembles the rod-like structure of WBV particles remarkably well (Fig. 1) (27). Taken together, the analysis suggests that WBV encodes three structural proteins, S, M, and N. The virus does not encode a homolog of the coronavirus E protein or a hemagglutinin esterase protein, the latter being conserved in toroviruses and several coronaviruses (52).

**WBV-specific RNAs.** All previously characterized nidoviruses produce a 3′-coterminal nested set of sg RNAs to express their structural and, in some cases, several accessory proteins. Therefore, the identification of three putative structural protein genes in the 3′-terminal region of the genome led us to predict that WBV produces sg RNAs to express these downstream ORFs. To confirm this hypothesis, we isolated poly(A)-

---

**FIG. 1.** Purified WBV genome RNA is infectious: evidence for viral particle formation and release. The cell culture supernatant from EPC cells transfected with purified WBV genome RNA was analyzed at 6 days posttransfection by electron microscopy (negative staining). (C and D). For comparison, electron micrographs taken from purified WBV virions (27) are shown in panels A and B. Representative pictures of both intact (A and C) and partially opened (B and D) virions are shown. Bar, 150 nm.
containing RNA from WBV-infected cells and used a 3′-terminal specific probe to detect WBV-specific genome-length and subgenome-length RNAs. The data we obtained in Northern blotting experiments (Fig. 3 and data not shown) demonstrate that WBV produces four RNAs (genome RNA and three sg RNAs) to express its genome. To determine the approximate sizes of the WBV-specific RNAs, we used two RNA markers, namely, (i) HCoV-229E genomic and sg RNAs and (ii) a 24.4-kb HCoV-229E-derived replicon RNA called Rep-1 (28). These RNA markers were detected by including an HCoV-229E-specific probe in the hybridization buffer. The observed sizes of the three sg RNAs strongly support the idea that RNAs 2, 3, and 4 are used to express the viral S, M, and N proteins, respectively. The size of RNA 1 was confirmed to be between 27.3 kb (HCoV-229E genome) and 24.4 kb (Rep-1 RNA), providing additional evidence for a WBV genome size of 26.6 kilobases which we had determined by genome sequence analysis.

As mentioned above, coronaviruses and arteriviruses pro-
produce a large number of sg RNAs that share a 5′ leader sequence identical to that of the 5′ end of the genome. The templates used for plus-strand sg RNA synthesis are sg minus strands carrying the complement of this leader sequence (an- tileader) at their 3′ ends. Attachment of this antileader re- quires a discontinuous step in minus-strand RNA synthesis (37, 47) and involves TRSs which are located upstream of each of the 3′ structural and accessory protein genes (body TRSs) as well as downstream of the 5′ leader sequence (leader TRS). Specific base-pairing interactions between the body TRS comple- ments and the leader TRS have been shown for coronaviruses and arteriviruses to guide the strand transfer of the nascent minus strand to the 5′ end of the genome, where minus-strand synthesis is completed by copying the leader sequence (38, 76). In contrast to corona- and arteriviruses, Echino- tovirus (strain Berne) produces only one leader-containing sg RNA (65) and the two ronivirus sg RNAs lack a 5′ leader sequence altogether (13). In view of these differences among the various nidovirus genera, we were interested in examining whether the three sg RNAs of WBV contained 5′ leader sequences. Like researchers of a previous study (60), we designed three sets of primers to amplify by RT-PCR the (potentially existing) junctions be-

![Diagram of WBV RNA lengths](image)

**FIG. 3.** Detection of WBV genome- and subgenome-length RNAs in virus-infected cells. Northern blot analysis of poly(A)-containing RNA isolated from WBV-infected EPC cells (lane 2). Poly(A) RNAs isolated from HCoV-229E-infected MRC-5 cells (lane 1) and HCoV-
229E-derived replicon RNA Rep-1 (lane 3) (28) were used as RNA size markers in this experiment. To detect both the HCoV-229E- and WBV-specific RNAs, a mixture of α-32P-multiprime-labeled probes specific for the 3′-terminal regions of HCoV-229E (nucleotides 26,857 to 27,277) and WBV (nucleotides 25,992 to 26,582) was used for hybridization. HCoV-229E genome- and subgenome-length RNAs and the in vitro-transcribed HCoV-229E Rep-1 RNA are indicated by black arrowheads, with sizes given in kilobases. White arrowheads indicate the four WBV-specific RNAs detected in this experiment. The longer exposure presented above shows the size of the WBV genomic RNA more clearly and allows its size to be compared with those of the 27.3- and 24.4-kb marker RNAs. The calculated sizes (Table 1 and Fig. 4) of the sg RNAs are 5,162 nts (RNA 2), 1,475 nts (RNA 3), and 774 nts (RNA 4) [including the 5′ leader but excluding the poly(A) tail].

tween 5′ leader and 3′ body sequences for the three WBV-
specific sg RNAs (see Materials and Methods). The upstream primer was specific for the 5′ end of the genome, and the downstream primers were specific for one of the ORFs, 2, 3, or 4. In each of the three reactions, we obtained specific ampli-
cons whose sizes were consistent with the presence of a short 5′ leader sequence (data not shown). Direct sequence analysis of the amplicons confirmed this conclusion and revealed the precise positions of the leader-body fusion sites (Fig. 4). Double peaks identified in two of the chromatograms suggested that two alternative fusion sites were used in the case of WBV RNAs 2 and 4. To address this possibility more rigorously, we cloned the RT-PCR products into pBluescript plasmid DNA and determined the sequences of individual cDNA clones. The data from this sequence analysis (summarized in Fig. 4) lead us to suggest the following: (i) WBV RNAs 2, 3, and 4 contain a 5′ leader sequence of generally 42 nucleotides whose sequence corresponds to that of the 5′ end of the WBV genome; (ii) in RNAs 2 and 4 (but not RNA 3), alternative leader-body fusion sites located three bases (RNA 2) and two bases (RNA 4) upstream of the major fusion site were occasionally used (Fig. 4); (iii) upstream of each of the ORFs, 2, 3, and 4, and down-
stream of the leader, a nonanucleotide sequence, CA(G/A)C ACUAC, is conserved, which, like in corona- and arteriviruses, might represent the core element of WBV TRSs. This putative core TRS could promote base pairing of as many as nine consecutive bases, thereby probably creating a very stable RNA structure. The actual fusion of leader and body se-
quences appears to occur outside of the base-paired region, namely, one (in a few cases, three or four) nucleotide(s) up-
stream of the core TRS element (plus-strand numbering). Similar observations have previously been made for several minor sg RNA species of arteriviruses (10, 15, 33). The WBV leader-
body fusion site data provide additional and very strong support for the sg RNA synthesis model originally introduced for coronaviruses by Sawicki and Sawicki (45, 46) and later ex-
tended to arteriviruses by van Marle and colleagues (64). Spec-
cifically, the data argue against a “free leader”-priming model wherein a free leader of a size greater than is found on the sg mRNA is annealed, trimmed with an exonuclease back to the fusion site, and extended.

The data further suggest that, after its relocation to the leader TRS, the nascent minus-strand RNA is extended by forming a phosphodiester bond with a mismatched 5′ nucleo-
tide, which, in most cases, is a uridylate. We also note that, in all leader-body fusion events, the first nucleotide to be incor-
porated by the minus-strand polymerase after the template switch is a guanylate. It remains to be seen whether this reflects specific requirements for initiation and/or reinitiation of RNA synthesis by the WBV RdRp. Notably, the 3′-terminal residue of the genome [preceding the poly(A) tail] is a cytidine, im-
plying that minus-strand RNA synthesis (also) starts by the incorporation of a guanylate.

Taken together, these data confirm and extend the previ-
ously established models for coronavirus and arterivirus sg RNA transcription (37, 47). At the same time, they make it clear that there is no simple relationship between the number of sg RNAs produced by specific groups of nidoviruses and the presence of a 5′ leader.
Identification of a putative ribosomal frameshifting element at the ORF1a/1b junction. We identified a putative slippery sequence, 14549UUUAAAC14555, just upstream of the WBV ORF1a translation stop codon, and the sequence downstream of the slippery sequence could be modeled into an RNA pseudoknot structure (Fig. 5). Over the past years, coronavirus RNA pseudoknot structures have been studied extensively, both structurally and functionally, and their critical role in mediating a shift into the /H11002 reading frame during translation has been firmly established (8, 35). Although our study does not provide formal evidence for that, it seems reasonable to predict that, as in other nidoviruses, WBV ORF1b expression (generating the viral RdRp and other key replicative proteins) is regulated at the translational level by ribosomal frameshifting.

Identification of a WBV ORF1a-encoded serine protease activity. The central and C-terminal portions of nidovirus replicase polyproteins are extensively processed by ORF1a-encoded 3C-like “main” proteases (75). Within the replicase polyprotein, nidovirus 3C-like proteases are generally flanked by membrane-spanning domains. Our sequence analysis of WBV ORF1a identified a putative 3CLpro domain in the C-terminal third of pp1a. The putative WBV 3CLpro domain contained the conserved GX(S/C)G signature of chymotrypsin-like proteases and was found to be flanked by hydrophobic domains at its N- and C-terminal borders. To confirm the
The bacteria were mock induced (lanes 1 and 3) or induced with 1 mM IPTG for 3 h (lanes 2 and 4). The positions of the fusion protein and cleavage product are indicated by arrowheads. The molecular masses of marker proteins are given in kDa to the left.

Proteolytic activity of this domain, we expressed in Escherichia coli the WBV pp1a/pp1ab residues Ser3424 to Gln3726 fused to MBP. The predicted size of the MBP-3CLpro fusion protein was 74.9 kDa. In lysates obtained from IPTG-induced cells transformed with pMal-WBV-3CL_559-560 (WT) (lanes 1 and 2) and pMal-WBV-3CL_S3589A (S3589A) (lanes 3 and 4) were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis in a 12.5% polyacrylamide gel and stained with Coomassie brilliant blue R-250. The positions of the fusion protein and cleavage product are indicated by arrowheads. The molecular masses of marker proteins are given in kDa to the left.

FIG. 6. Proteolytic activity of WBV pp1a/pp1ab amino acid residues Ser3424 to Gln3726. Total cell lysates from E. coli TB1 cells transformed with pMal-WBV-3CL_559-560 (WT) (lanes 1 and 2) and pMal-WBV-3CL_S3589A (S3589A) (lanes 3 and 4) were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis in a 12.5% polyacrylamide gel and stained with Coomassie brilliant blue R-250. The bacteria were mock induced (lanes 1 and 3) or induced with 1 mM IPTG for 3 h (lanes 2 and 4). The positions of the fusion protein and cleavage product are indicated by arrowheads. The molecular masses of marker proteins are given in kDa to the left.

encodes a 3C-like serine protease activity which, by analogy to other nidoviruses, is predicted to play a key role in the proteolytic processing of the WBV replicase polyproteins. The approximate positions of a limited number of cleavage sites can be inferred from the domain borders of the conserved ORF1b-encoded enzymes. However, we failed to identify strictly conserved sequence signatures at these interdomain junctions, suggesting that the substrate specificity of the WBV 3CLpro domain may be less well conserved than those of its coronavirus and torovirus counterparts. In order to make reliable predictions on potential 3CLpro cleavage sites, a minimum of biochemical evidence on the WBV 3CLpro substrate specificity needs to be obtained.

Domain structure of the WBV replicase. Comparative sequence analyses of the ORF1b-encoded part of the replicase polyproteins of WBV and other nidoviruses suggested that WBV is a large nidovirus (24) which is most closely related to members of the genera Torovirus and Coronavirus. This conclusion is supported by the following observations. First, unlike small nidoviruses (Arteriviridae), which encode only a limited set of domains (24), WBV encodes the full set of ORF1b domains conserved in previously characterized large nidoviruses, namely, RdRp, zinc-binding, helicase, exoribonuclease, NendoU, and putative ribose-2'-O-methyltransferase domains (Fig. 7A). Second, at least four putative membrane domains (two of them flanking 3CLpro) were identified in the ORF1a-encoded sequence. The relative positions and approximate sizes of the WBV pp1a membrane domains were found to correspond very well to those identified in toro- and coronaviruses, whereas the distributions of corovirus pp1a membrane domains (except for the two domains flanking 3CLpro) were clearly different (Fig. 7A). Third, the special affinity between WBV and toro- and coronaviruses was supported by the identification of an ADP-ribose 1'-phosphatase domain in WBV (Fig. 7B), the conservation of which has not been reported for Arteriviridae and Roniviridae. Thus far, we failed to obtain convincing evidence for the conservation of close homologs of the papain-like proteases of corona- and toroviruses in WBV. Experimental studies are under way to get insight into the expression of the WBV N-proximal pp1a/pp1ab regions.

Phylogeny of WBV. Taken together, the data provide strong evidence that WBV and other nidoviruses share a common ancestor. This hypothesis is based on multiple lines of evidence, including (i) the nidovirus-like polycistronic genome organization of WBV, (ii) the relationship between (some of) the WBV structural proteins and the respective homologs from toro- and (to a lesser extent) coronaviruses, and (iii) the use of common genome expression strategies involving regulation at both the transcriptional level (i.e., production of a nested set of sgRNAs), at the translational level (use of ribosomal frameshifting to express the replicase core domains), and at the posttranslational level (evidence for proteolytic processing by a chymotrypsin-like main protease). To define the relationship of WBV with other large nidoviruses more precisely, we produced sequence alignments of conserved ORF1b domains and used them to generate phylogenetic trees (Fig. 8 and data not shown). In these analyses, WBV consistently grouped together with members of the genus Torovirus. A moderate sequence similarity between WBV and toroviruses was also evident when the viral structural proteins were compared (see above), and
there are similarities between WBV and toroviruses with respect to virion morphology. It is also noteworthy that WBV and toroviruses share extremely long 5′-terminal nontranslated regions of more than 800 nucleotides which lack the small ORFs upstream of ORF1a that are conserved in coronaviruses and arteriviruses. Furthermore, the 3CL₂₆ domains of WBV and toroviruses are more related to each other than they are to other nidovirus main proteases (R. Ulferts and J. Ziebuhr, unpublished data). In several other respects, however, the two clusters have diverged significantly. For example, (i) WBV does not encode cyclic nucleotide phosphodiesterase, which is conserved in bovine and equine toroviruses; (ii) there is a poor sequence conservation between the N-proximal pp1a/pp1ab regions of WBV and toroviruses; (iii) unlike Equine torovirus...
VOL. 80, 2006 CHARACTERIZATION OF WHITE BREAM VIRUS

FIG. 8. Phylogenetic analysis of WBV helicase and polymerase core domains. Phylogenetic trees were generated from multiple-sequence alignments of the most conserved regions of nidovirus RNA-dependent RNA polymerase (residues Thr4723 to Gln5396 in the WBV pp1ab sequence) and helicase domains (residues Ala5644 to Cys5924 in the WBV pp1ab sequence), using the neighbor-joining algorithm as implemented in the ClustalX 1.8 program (for details, see Materials and Methods). The WBV sequences were compared with those from *Gill-associated virus* (GAV, accession no. AF227196), *Equine torovirus* Berne (EToV, X52374), and *Bovine torovirus* Breda-1 (BToV, AY427798) as well as from viruses representing the three coronavirus groups, including the recently introduced subgroups 1a, 1b, 2a, and 2b (26). Group 1a, *Transmissible gastroenteritis virus* Pursd-de-115 (TGEV, accession no. Z34093) and *Feline infectious peritonitis virus* WSU 79/1146 (FIPV, DQ010921); group 1b, HCoV-229E (NC_001451); group 2a, *Bovine coronavirus* LUN (BCoV, AF391542), *Human coronavirus* OC43 strain FC15 (HCoV-OC43, AY585229), *Mouse hepatitis virus* A59 (MHV, NC_001846), and *Human coronavirus* HKU1 (HCoV-HKU1, NC_006577); group 2b, *Severe acute respiratory syndrome coronavirus* Frankfurt 1 (SARS-CoV, AY291315); group 3, *Avian infectious bronchitis virus* Beaudette (IBV, NC_001451).

(strain Berne), in which three (out of four) sg RNAs lack a 5′ leader (65), all of the sg RNAs of WBV possess a 5′ leader sequence; (iv) in contrast to toroviruses, WBV does not encode a hemagglutinin esterase structural protein; and (v) the natural hosts of toroviruses (mammals) and WBV (fish) differ profoundly.

**Conclusion.** The study reports the complete genome sequence of WBV, the first nidovirus to be isolated from fish. A preliminary characterization of this virus and comparative sequence analysis identified toroviruses (followed by coronaviruses) as the closest known relatives of WBV. Given the limited number of published torovirus sequences (18, 53), the reconstruction of possible scenarios involved in the evolution of the WBV cluster remains speculative at the present time. Most likely, WBV (and its yet-to-be-identified close relatives) and the present-day toroviruses have evolved from a common ancestor that split off from the lineage that led to the present-day coronaviruses. Given the quite different hosts infected by WBV and toroviruses, it seems less likely that the ancestor of the WBV cluster split first from the common coronavirus-torovirus trunk to evolve then in parallel with the torovirus cluster. But clearly, additional studies and sequences of WBV- and torovirus-related viruses are required to (re)construct the most plausible scenario for the evolution of the WBV-like, coronavirus, and torovirus clusters. However, even in the absence of this information, we think the profoundly divergent evolution of the WBV cluster from the related genera *Torovirus* and *Coronavirus* and the different host ranges of these viruses would justify the introduction of a new genus, with WBV being its tentative type species. We propose the name *Bafinivirus* for this yet-to-be-approved nidovirus genus, referring to the bacilliform morphology of this cluster of fish nidoviruses. If approved, the new genus might then trigger a more general discussion and, possibly, revision of the current taxonomic structure of the *Nidovirales*.

The identification of nidoviruses in a very large number of mammalian species (39, 52, 59, 67) as well as in invertebrates (12) and fish (this study) suggests that nidoviruses, which in several respects are distinct from the huge variety of plus-strand RNA viruses (24), have managed to adapt to a remarkable diversity of biological niches. The continued sampling and characterization of nidoviruses, of which this study is a part, are anticipated to increasingly fill the major gaps that still exist between the individual clusters of nidoviruses. Additional sequence information and functional studies will help to identify the major forces and constraints that shape the evolution of nidoviruses. This information will also be required to unravel the basis for the differential conservation of specific replicase gene-encoded proteins among the various nidovirus families, genera, and species (24, 54) and help to relate these proteins to specific metabolic pathways and molecular mechanisms. More sequences and phylogenetic studies are also needed to determine the position of the WBV cluster within the nidovirus tree more precisely.

**ACKNOWLEDGMENTS**

We thank Sabine Feistkorn, Petra Meyer, Mandy Jörn, and Beatrix Loth for excellent technical assistance. The work of R.U., B.S., S.B., and J.Z. was supported in part by the Deutsche Forschungsgemeinschaft (Zi 618/3 and SFB 479-A8).

**REFERENCES**


base pairing between sense and antisense transcription-regulating sequences. 


