Analysis on Genetic Diversity and Molecular Evolution of Human Group B Rotaviruses Based on Whole Genome Segments

Dai Yamamoto*[1], Souvik Ghosh[1], Balasubramanian Ganesh[2], Triveni Krishnan[2], Mamta Chawla-Sarkar[2], Mohammed Mahbub Alam[3], Tin Sabai Aung[4], Nobumichi Kobayashi[1]

[1] Department of Hygiene, Sapporo Medical University School of Medicine, Sapporo, Japan
2 National Institute of Cholera and Enteric Diseases, Kolkata, India
3 Bangladesh Agricultural University, Mymensingh, Bangladesh
4 National Health Laboratory, Yangon, Myanmar

Running title : full genomic analysis of human group B rotavirus

:

:

:

* Corresponding author

Dai Yamamoto
Department of Hygiene
Sapporo Medical University School of Medicine
S-1 W-17, Chuo-ku, Sapporo 060-8556, Japan
Tel: +81-11-611-2111(ext.2733)
e-mail: yamadai@sapmed.ac.jp

SUMMARY

Group B rotavirus (GBR) is a rare enteric pathogen which causes severe diarrhea primarily in adults. Nearly full-length sequences of all the 11 RNA segments were determined for human GBRs detected recently in India (IDH-084 in 2007, IC-008 in 2008), Bangladesh (Bang117 in 2003), and Myanmar (MMR-B1 in 2007), and analyzed phylogenetically with the sequence data of GBRs reported previously. All the RNA segments of GBR strains from India, Bangladesh, and Myanmar showed more than 95% nucleotide sequence identities. Among the 11 RNA segments, the VP6 and NSP2 genes showed the highest identities (>98%), while the lowest identities were observed in the NSP4 gene (96.1%), NSP5 gene (95.6%), and VP8*-coding region of the VP4 gene (95.9%). Divergent or conserved regions in the deduced amino acid sequences of GBR VP1-VP4, NSP1-NSP5 were similar to those in group A rotaviruses (GARs), and the functionally important motifs and structural characteristics in viral proteins known for GAR were conserved in all the human GBRs. These findings suggest that while the degree of genetic evolution may be dependent on each RNA segment, human GBR may have been evolving in a similar manner to GAR, associated with the similar functional roles of individual viral proteins.

INTRODUCTION

Rotavirus, a member of the family *Reoviridae*, is the most important viral pathogen causing gastroenteritis in humans. Rotavirus has 11 segments of double-stranded RNA as a genome, and the viral particle is composed of three concentric layers, i.e. the outer capsid, inner capsid, and core (Estes & Kapikian, 2007). The outer capsid consists of two structural proteins VP4 and VP7, which are neutralization antigens. The inner capsid consists of structural protein VP6. Rotavirus is classified into five groups, i.e., groups A-E, and two putative groups F and G, based on the antigenicity of the inner capsid protein VP6 and genomic characteristics (Kapikian *et al.*, 2001). In humans, groups A, B, C have been detected so far. Group A Rotavirus (GAR) is the most prevalent throughout the world and is recognized as the leading viral pathogen of acute gastroenteritis in children.

Group B Rotavirus (GBR) is genetically and antigenically distinct from GAR and has been detected in humans, mice, calves, pigs and sheep. In humans, GBR has been noted because it causes severe cholera-like diarrhea, mostly in adults (Mackow, 1995). GBR was first identified as adult diarrhea rotavirus (ADRV) in nationwide outbreaks in China in 1982-1983 (Hung *et al.*, 1983, 1984; Wang *et al.*, 1985), and the detection of this virus has been limited to China (Dai *et al.*, 1987; Fang *et al.*, 1989). They were subsequently detected in sporadic cases in India in 1997 and in Bangladesh in 2000, demonstrating the distribution of GBRs in other Asian countries outside China (Krishnan *et al.*, 1999; Sanekata *et al.*, 2003). Thereafter, GBRs have been again detected in these countries in sporadic cases of diarrhea (Barman *et al.*, 2006; Rahman *et al.*, 2007). Furthermore, a human GBR was detected in Myanmar in 2007 (Aung *et al.*, 2009). In contrast, recently in China, there has been only a 2002 report of detection of GBR in sporadic cases of diarrhea in Wuhan (Yang *et al.*, 2004). Despite the extensive epidemiologic surveillance on rotaviruses worldwide, the detection of human GBRs has

been extremely rare and limited to only the four countries mentioned above.

With the limited genetic information, human GBRs were known to be quite distinct from bovine, porcine, ovine and murine GBRs. Recently, Kuga and coworkers (2009) proposed a classification scheme of GBR genotypes in terms of VP7 gene sequence, and GBRs were divided into five genotypes (G1-G5), among which human strains were assigned to a single genotype G2. The human GBRs detected were classified genetically into two lineages within the genotype G2 based on VP7 gene, the Chinese lineage and the Indian-Bangladeshi lineage (Ahmed *et al.*, 2004; Yang *et al.*, 2004; Aung *et al.*, 2009). The GBR strains of these two lineages were genetically closely related, suggesting that these lineages diverged from a common ancestral origin several decades ago (Yang *et al.*, 2004; Rahman *et al.*, 2007). However, genetic diversity in RNA segments other than the VP7 gene has been scarcely analyzed, and thus accurate status of molecular evolution of the whole genome of GBR is still unknown. Similar to GAR, six RNA segments of GBR encode structural proteins (VP1-VP4, VP6 and VP7) and five segments encode non-structural proteins (NSP1-5). However, it is unique to human GBR that NSP1 gene contains two ORFs encoding putative two protein products (Mackow, 1995; Kobayashi et al., 2001).

Full genomic sequence of rotaviruses has been determined for many GAR strains to date. Accordingly, based on the findings on diversity of individual RNA segments, a full-genome based genotyping system composed of genotypes of 11 RNA segments has been proposed (Matthijnssens *et al.*, 2008). However, genetic data of GBR is extremely limited. Full genomic sequence of GBR has been determined for only three human strains (CAL-1 in India, Bang373 in Bangladesh, and WH-1 in China) and a murine strain IDIR. Genetic information available for human and animal GBRs is mostly for VP7 gene sequences.

In the present study, nearly full-length sequences of all the gene segments were determined for human GBRs detected recently in India, Bangladesh, and Myanmar. The

obtained sequence data were analyzed and compared with those reported previously, to understand difference in genetic diversity among the 11 RNA segments and divergent regions in individual RNA segments. The results have provided fundamental information about the genomic evolution of GBR, including the relatedness of genetic diversity to the function of each viral protein.

RESULTS

Phylogenetic analysis of the VP7 gene

The VP7 gene sequences of GBRs determined in the present study were analyzed phylogenetically with those of human, bovine, porcine, and murine GBR strains. Fig. 1(A) represents a phylogenetic tree of the VP7 genes. Sequence identities of VP7 genes among GBRs are shown in supplementary Table S1. The Indian strains IC-008 and IDH-084, and Bangladeshi strain Bang117 were located in the human Indian-Bangladeshi lineage of the GBR genotype G2. Within this lineage, IC-008 and IDH-084 clustered with MMR-B1 in Myanmar, while Bang117 clustered with the Bangladeshi strains (Bang373 and Bang544) reported previously. Strains IC-008 and IDH-084 having 99.5% nucleotide sequence identity to each other, showed extremely high sequence identities to GBRs in the Indian-Bangladeshi lineage (97.8-99.4%), with the highest identity to MMR-B1 (99.4%) (Table S1). The strain Bang117 exhibited extremely high sequence identities to Bangladeshi strains Bang373 (99.9%) and Bang544 (99.6%), while slightly lower identities (97.9-98.3%) to GBRs in India and Myanmar. All the GBRs in the Indian-Bangladeshi lineage showed 91.3-92.6% identities to Chinese strains ADRV and WH-1, and considerably lower identities to bovine GBRs (62.2-66.6%), porcine GBRs (62.9-63.9%) and a murine GBR (58.0-59.8%).

Genetic analysis of other viral protein genes

Phylogenetic analysis of GBR genes encoding VP1, VP2, VP4, VP6, NSP1, NSP2, and NSP5 revealed the presence of murine, human, and bovine GBR clusters corresponding to GBR genotypes G1, G2, G3, respectively, which had been classified for VP7 gene (Kuga et al., 2009) (Fig.1(B), (C), (E)-(H), (K)). For the genes of VP3, NSP3, and NSP4, two genotypes G1 and G2 were discriminated (Fig.1(D), (I), (J)). Furthermore, all the gene segments of human GBRs were discriminated into two lineages, i.e., Indian-Bangladeshi lineage and Chinese lineage, within the GBR genotype G2. Range of sequence identities of individual gene segments within a lineage and between the two lineages are summarized in Table 1 (Sequence identities of each gene segment among GBRs are shown in supplementary Tables S2-S11). Within the same lineage, gene sequences showed extremely high identities, i.e., 95.6-100.0% (94.7-100% at amino acid level) in the Indian-Bangladeshi lineage, and 98.0-98.9% (98.0-100% in amino acid level) in the Chinese lineage, while lower identities (89.5-94.9% at nucleotide level) were found between the two lineages. Throughout the RNA segments, the strain MMR-B1 showed the highest sequence identities (>99%) to IC-008 and IDH-084. In contrast, within GBRs of the Indian-Bangladeshi lineage, the lowest levels of sequence identities were observed mostly between the recent Indian strains (IC-008 or IDH-084) and CAL-1.

Among the 11 RNA segments of GBR strains from India, Bangladesh, and Myanmar, the VP6 and NSP2 genes showed the highest identities (>98%), while the lowest identities were observed in the NSP4 gene (96.1%) and NSP5 gene (95.6%). The sequence identities of the VP8*-coding region of the VP4 gene were lower than those of the VP5* region. Similarly, between the two lineages, VP6 and NSP2 gene sequences were most conserved (93.1-94.9% identity); in contrast, NSP3, NSP5, and VP8* genes exhibited the highest diversity (88.4-92.4% identity). Between strains IC-008 and IDH-084, derived from a child and an adult, respectively, sequence identity was 98.9-99.7% throughout the gene segments.

Divergent regions in viral proteins

To investigate whether genetic diversity among GBR genes occurred randomly, presence of divergent or conserved region(s) within a viral protein was analyzed by sequence alignment of the deduced amino acid sequences. Amino acid sequences of GBR VP1 and NSP3, and partial VP3 and NSP2 sequences were aligned with those of GAR and group C Rotavirus (GCR) (Figs. 2-4). Except for VP1 and NSP3, primary amino acid sequence alignment of human GBRs are shown in supplementary Figs. S1-S9. Due to the extremely conserved nature of VP7 (Fig. S5) and VP6 (Fig. S4), the divergent region of these proteins was not specified,

The RNA polymerase domain of VP1 (Finger I, II, Palm I, II, and Thumb) located in the central portion (469 amino acids) (McDonald et al., 2009), including functionally critical motifs shared by different rotavirus groups as well as other RNA viruses (Cohen, *et al.*, 1989; Nagashima *et al.*, 2008), was highly conserved among GBRs (Fig. 2). In the catalytic region (182 amino acids from position 546 to 727), 54 amino acids (29.7%) were conserved among GAR, GBR, and GCR, and the consensus motif of RNA polymerase (SG, T, N, T, GDD) was commonly found in these three rotavirus groups. Most of the amino acid differences in GBR VP1 were found in the N-terminal region (376 amino acids) and C-terminal region (317 amino acids) where GAR and GCR had quite distinct sequences.

Sequence divergence of VP2 among GBRs was found in the 80 amino acid-sequence from the N-terminus (Fig. S1). Amino acid differences of VP3 sequences among the GBR strains were found throughout the sequence. However, a conserved motif (ALYSLSNXXN) (Ito et al., 2001) was found in all the GBRs as well as GAR and GCR (Fig. 3(A)). Some sequences similar to the possible active sites of guanylyltransferase in VP3 (Cook & McCrae, 2004) were conserved among human GBRs (Fig. S2). In the VP4 sequence, some hydrophobic regions located mostly in the VP5* portion were highly conserved, while the sequence in the N-terminal hydrophilic region in the VP8* portion

had more amino acid diversity than any other regions in VP4 (Fig. S3).

In the peptide 2 of the NSP1 gene, while the divergent amino acids were located over the sequence, a cysteine and histidine-rich region was highly conserved (Fig. S6). A cysteine-rich region was found in the peptide 1. Although NSP2 sequences of human GBRs showed much less diversity, they were considerably distinct from those of GAR and GCR (Fig. S7, Fig. 3(B)). However, some amino acids which had been known to be required for nucleoside triphosphatase activity in GAR, including the conserved histidine (H225 for GAR) (Kumar et al., 2007), were also conserved among human GBR, as well as GCR (Fig. 3(B)).

Among the NSP3 sequences of GBRs, the N-terminal 120 amino acid-region was highly conserved and corresponded to the RNA-binding region revealed for GAR NSP3 (Deo et al., 2002) (Fig. 4). Remarkably, amino acids in a motif RNXXW in the alpha-helix 4 (H4) which are essential for RNA binding by GAR NSP3 (Vende et al., 2000) were conserved in NSP3 of GBR as well as GCR. In contrast, the remaining two-thirds portion was divergent and included a region associated with eIF4G binding.

Two hydrophobic regions and two putative enterotoxin regions in NSP4 (Ishino et al., 2006) were also highly conserved among the GBR strains (Fig. S8). In the NSP5 sequence, the C-terminal 60 amino acid-sequence was more conserved than the remaining N-terminal portion comprising 110 amino acids (Fig. S9).

DISCUSSION

To date, human GBRs have been detected only in China, India, Bangladesh, and Myanmar, and have been classified genetically into two lineages, i.e., Chinese lineage and Indian-Bangladeshi lineage. However antibodies to human GBRs have been detected in humans from several other countries such as Kenya, Thailand, Canada, the USA, and UK

at low prevalence, suggesting wide distribution of this virus (Brown *et al.*, 1987; Nakata *et al.*, 1987). Also, GBRs have been detected in rats, pigs, cows, and goats, and have been suggested to be highly prevalent in these animals through seroepidemiologic evidences (Brown *et al.*, 1987; Tsunemitsu *et al.*, 2005). Animal GBRs are genetically distinct from human GBRs and contain divergent populations of viruses. According to the classification of GBR in terms of the VP7 gene by Kuga and coworkers (2009), GBRs were divided into five genotypes (G1-G5) containing a single genotype G2 of human strains and other four genotypes of murine, porcine, and bovine strains. Although genetic and molecular epidemiologic study of GBR has been based primarily on the VP7 gene, genetic information of other viral gene segments is extremely limited. So far, full-genomic information has been available for only a murine strain and three human strains, and eight gene segments have been sequenced for a bovine strain (Ghosh et al., 2007, 2009). For ovine and porcine GBRs, sequence data of only a few gene segments are available (Shen et al., 1999; Kuga et al., 2009). In the present study, nearly full-length sequences of four human GBRs were determined, which enabled us to analyze substantially the genetic diversity of GBR at a full genome level.

Sequence analysis of whole gene segments of the human GBR provided two epidemiologically significant findings. First, the Myanmarese strain MMR-B1 is closer genetically to the recent Indian strains than the old Indian strain (CAL-1) and Bangladeshi strains for all the gene segments, suggesting that the GBR in Myanmar and current Indian GBRs may have been derived from the same origin. Second, the two latest GBR strains from a child (IC-008) and an adult (IDH084) detected in Kolkata were genetically almost identical, which indicated distribution of the same GBR in a child and an adult. Similarly, distribution of genetically identical GAR in both children and adults was reported in China and Bangladesh (Wang et al., 2007; Paul et al., 2008). Together with the report that GBR was detected by RT-PCR in 18.5% of diarrheal specimens from children in Kolkata (Barman et al., 2006), GBR is believed to infect both children and

adults and may be maintained among them.

Although the human GBRs analyzed in the present study were genetically highly similar, sequence identities were different depending on gene segments. The VP6 and NSP2 genes showed the highest identities, while NSP3-NSP5 genes were more divergent. Similarly, within a single genotype of GAR, the VP6 gene showed the highest identity among structural proteins (Matthijnssens *et al.*, 2008). However, sequence variation in nonstructural proteins seems to be different from GAR; that is, in GAR, NSP5 is the most conserved while NSP1 is the most divergent. Such difference in GBR may be in part due to insufficient numbers of strains analyzed in the present study.

In some RNA segments, genetic divergence or conservation was detected in specific regions which are correlated to the function of the protein. The RNA polymerase domain in VP1 and N-terminal RNA-binding domain in NSP3 were evidently conserved compared with other regions in individual proteins, as found also in cognate proteins of GAR (Heiman et al., 2008; McDonald et al., 2009; Rao et al., 1995). In these domains, some amino acids/motifs are commonly found among GAR, GBR, and GCR, despite high sequence diversity among them. Furthermore, some functionally important motifs and structural features of other rotavirus proteins known for GAR were also conserved in GBR, including the conserved motif in VP3 of which the function is unknown (Ito et al., 2001; Nagashima et al., 2008), and that in NSP2 related to nucleoside triphosphatase activity (Kumar et al., 2007). The cysteine and histidine-rich region which was noted for GAR NSP1 (Hua et al., 1993) was found also in the peptide 1 and 2 of GBR NSP1. The C-terminal region of NSP5 which is conserved among GAR and critical for its function for viroplasm-like structure formation in cells (Sen et al., 2007) was also conserved in GBR NSP5.

In contrast, the VP8* portion in VP4 and the N-terminal region in VP2 were more divergent than other regions in these proteins, in both GAR and GBR. The VP8* is located as the outermost portion of the VP4 spike protein on the rotavirus virion and is

associated with the antigenic specificity and genotype of VP4 (Dormitzer et al., 2004; Kapkian, 2001). The amino termini of GAR VP2 are predicted to lie inside core shell and to bind viral enzyme-RNA complex (VP1/VP3/RNA) (McDonald & Patton, 2008). It can be suggested that the interaction with protein and/or RNA causes the sequence variation in the N-terminal region of VP2. The above findings suggested that viral proteins of GBR and GAR have similar structural and functional features, subjected to similar molecular evolution, despite a considerable genetic divergence between the two rotavirus groups. Although the degree of genetic evolution may be dependent on each RNA segment, human GBR genes are suggested to have evolved in association with functional roles of individual viral proteins.

In the present study, GBR was first characterized for its genomic diversity and evolution at a level of full genome. Further accumulation of genetic data with more GBRs may be necessary to understand ecological features and epidemiologic dynamics of GBRs.

MATERIALS AND METHODS

Rotavius strains

Four human GBR strains, IC-008, IDH-084, Bang117, and MMR-B1 were analyzed. The GBR strains IC-008 and IDH-084 were detected as a sole pathogen of diarrhea in stool specimens from a child (12-month-old female) and an adult (35-year-old male), respectively, who visited the Infectious Disease Hospital in Kolkata, India, in January 2008 and November 2007, respectively. The strain Bang117 was found in a 32-year-old male patient with severe diarrhea admitted to SK hospital in Mymensingh, Bangladesh, in March 2002. The strain MMR-B1 was detected in the diarrhea of an adult patient in Yangon, Myanmar, and its VP7, VP4, VP6, and NSP4 genes had been sequenced and reported previously (Aung *et al.*, 2009). Therefore, in the present study,

sequences of the remaining seven viral genes were determined for this strain. The presence of GBRs in stool specimens was determined by detection of the typical migration pattern (4-2-1-1-1-1-1 pattern) of 11 dsRNA segments in polyacrylamide gel electrophoresis, and further confirmed by RT-PCR as described previously (Gouvea *et al.*, 1991). Stool specimens collected from patients were stored at -80℃ until analyzed.

Sequencing, Phylogenic analysis

Nucleotide sequences of GBR genes were determined directly with the amplified cDNA products by RT-PCR. As a template for RT-PCR, dsRNA was extracted from stool suspension with a commercially available kit (RNAID kit, BIO101, Inc., La Jolla,CA) according to manufacturer's instructions. RT-PCR was performed with reverse transcriptase (AMV) (Seikagaku Co., Tokyo), thermostable DNA polymerase (Expanded High Fidelity PCR System, Roche, Mannheim, Germany) with the primers described previously for rotavirus genes encoding VP2, VP4, VP6, VP7, NSP1-NSP5 (Ahmed *et al.*, 2004). The cDNA for VP1 and VP3 genes were amplified by primers prepared in the present study based on the sequences of Bang373 strain (supplementary Table S12). For all the gene segments, full-length sequences except for primer binding regions at 5'- and 3'-end were amplified and sequenced.

PCR products were purified by Wizard^{R} SV GEL and PCR Clean-Up System (Promega, Inc., Madison,WI). Sequencing reaction was performed with fluorescent dideoxy chain termination chemistry using the BigDye Terminator version 3.1 cycle sequencing kit (Applied Biosystems, Foster City, CA). Sequence was determined by ABI Prism 3100 genetic analyzer (Applied Biosystems). GENETYX-Win version 5.1 (Software Development, Tokyo, Japan) was used to perform pairwise alignment and calculate the identity of gene segments among GBRs. Multiple alignment of GBR sequences were performed by the neighbor-joining method using the CLUSTAL W program. Phylogenetic analysis was performed with MEGA software version4.1 based on the neighbor-joining

method and the Kimura two-parameter model. Phylogenetic trees were supported statistically by bootstrapping with 1,000 replicates.

Accession numbers of sequences

The nucleotide sequences of GBR strains determined in this study were deposited in the GenBank database under following accession numbers : GU377213-GU377223 (IC-008), GU377224-GU377234 (IDH-084), GU391301-GU391311 (Bang117), and GU370054-GU370060 (MMR-B1).

ACKNOWLEDGEMENT

This study was supported in part by the Program of Founding Research Centers for Emerging and Reemerging Infectious Diseases (Okayama University - National Institute of Cholera and Enteric Diseases, India) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, and a Grant-in-Aid of The Japan Medical Association.

REFERENCES

Ahmed, M.U., Kobayashi, N., Wakuda, M., Sanekata, T., Taniguchi, K., Kader, A., Naik, T.N., Ishino, M., Alam, M.M. & other authors (2004). Genetic analysis of group B human rotaviruses detected in Bangladesh in 2000 and 2001. J Med Virol 72, 149-155.

Aung, T.S., Kobayashi, N., Nagashima, S., Ghosh, S., Aung, M.S., Oo, K.Y. & Win, N. (2009). Detecton of group B rotavirus in an adult with acute gastroenteritis in yangon, Myanmar. J Med Virol 81, 1968-1974.

Barman, P., Ghosh, S., Samajdar, S., Mitra, U., Dutta, P., Bhattacharya, S.K., Krishnan, T., Kobayashi, N. & Naik, T.N. (2006). RT-PCR based diagnosis revealed importance of human group B rotavirus infection in childhood diarrhoea. J Clin Virol. 36,222-227.

Brown, D.W., Beards, G.M., Chen, G.M. & Flewett, T.H. (1987). Prevalence of antibody to group B (atypical) rotavirus in humans and animals. J Clin Microbiol 25,316-319.

Cohen, J., Charpilienne, A., Chilmonczyk, S. & Estes, M.K. (1989). Nucleotide sequence of bovine rotavirus gene 1 and expression of the gene product in baculovirus. Virology. 171,131-140.

Cook, J.P. & McCrae, M.A. (2004) Sequence analysis of the guanylyltransferase (VP3) of group A rotaviruses. J Gen Virol 85, 929-932.

Dai, G-Z., Sun, M-S., Liu, S-Q., Ding, X-F., Chen, Y-D., Wang, L-C., Du, D-P., Zhao, G., Su, Y., & other authors. (1987). First report of an epidemic of diarrhoea in human neonates involving the new rotavirus and biological characteristics of the epidemic virus strain

(KMB/R85). J Med Virol 22, 365-373.

Deo, R.C., Groft, C.M., Rajashankar, K.R. & Burley, S.K. (2002). Recognition of the rotavirus mRNA 3' consensus by an asymmetric NSP3 homodimer. Cell. 108,71-81.

Dormitzer, P.R., Nason E.B., Prasad, B.V. & Harrison, S.C. (2004). Structural rearrangements in the membrane penetration protein of a non-enveloped virus. Nature. 430, 1053-1058.

Estes, M.K. & Kapikian, A.Z. (2007). Rotaviruses. In Fields Virology, 5th ed,pp.1917-1974. Edited by D.M.Knipe & P.M.Howley. Philadelphia, PA: Lippincott, Williams & Wilkins Co.

Fang, Z-Y., Ye, Q., Ho, M-S., Dong, H., Qing, S., Peneranda, M.E., Hung, T., Wen, L. & Glass, R.I. (1989). Investigation of an outbreak of adult diarrhea rotavirus in China. J Infect Dis 160, 948-953.

Ghosh, S., Varghese, V., Sinha, M., Kobayashi, N. & Naik, T.N. (2007). Evidence for interstate transmission and increase in prevalence of bovine group B rotavirus strains with a novel VP7 genotype among diarrhoeic calves in Eastern and Northern states of India. Epidemiol Infect 135, 1324-1330

Ghosh, S., Kobayashi, N., Nagashima, S., Chawla-Sarkar, M., Krishnan, T., Ganesh, B. & Naik, T.N. (2009). Molecular characterization of the VP1, VP2, VP4, VP6, NSP1 and NSP2 genes of bovine group B rotaviruses: identification of a novel VP4 genotype. Arch Virol. [Epub ahead of print]

Gouvea, V., Allen, J.R., Glass, R.I., Fang, Z.Y., Bremont, M., Cohen, J., McCrae. M.A., Saif, L.J., Sinarachatanant, P. & other author (1991). Detection of group B and C rotaviruses by polymerase chain reaction. J Clin Microbiol. 29,519-523.

Heiman, E.M., McDonald, S.M., Barro, M., Taraporewala, Z.F., Bar-Magen, T. & Patton, J.T. (2008). Group A human rotavirus genomics: evidence that gene constellations are influenced by viral protein interactions. J Virol. 82, 11106-11116.

Hua, J., Mansell, E.A. & Patton, J.T. (1993) Comparative analysis of the rotavirus NS53 gene: conservation of basic and cysteine-rich regions in the protein and possible stem-loop structures in the RNA. Virology 196:372-378.

Hung, T., Chen, G., Wang, C., Chou, Z., Chao, T., Ye, W., Yao, H. & Meng, K. (1983). Rotavirus-like agent in adult non-bacterial diarrhoea in China. Lancet 2, 1078-1079.

Hung, T., Wang, C., Fang, Z., Chou, Z., Chang, X., Liong, X., Chen, G., Yao, H., Chao, T. & other authors (1984). Waterborne outbreak of rotavirus diarrhoea in adults in China caused by a novel rotavirus. Lancet 1, 1139-1142.

Ishino, M., Mise, K., Takemura, H., Ahmed, M.U., Alam, M.M., Naik, T.N. & Kobayashi, N. (2006). Comparison of NSP4 protein between group A and B human rotaviruses: detection of novel diarrhea-causing sequences in group B NSP4. Arch Virol 151,173-182.

Ito, H., Sugiyama, M., Masubuchi, K., Mori, Y., & Minamoto N. (2001). Complete nucleotide sequence of a group A avian rotavirus genome and a comparison with its counterparts of mammalian rotaviruses. Virus Res 75:123-138.

442 Kapikian, A.Z., Hoshino, Y. & Chanock, R.M. (2001). Rotaviruses, In Fields Virology, 4 th
443 edn, pp.1787-1833. Edited by D.M. Knipe & P.M. Howley. Philadelphia, PA: Lippincott,
444 Williams & Wilkins Co.
445
446 Kobayashi, N., Naik, T.N., Kusuhara, Y., Krishnan, T., Sen, A., Bhattacharya, S.K.,
447 Taniguchi, K., Alam, M.M. & Urasawa, S. (2001). Sequence analysis of genes encoding
448 structural and nonstructural proteins of a human group B rotavirus detected in Calcutta,
449 India. J Med Virol 64, 583-588.
450
451 Krishnan, T., Sen, A., Choudhury, J.S., Das, S., Naik, T.N. & Bhattacharya, S.K. (1999).
452 Emergence of adult diarrhoea rotavirus in Calcutta, India. Lancet 353, 380-381.
453
454 Kuga, K., Miyazaki, A., Suzuki, T., Takagi, M., Hattori, N., Katsuda, K., Mase, M.,
455 Sugiyama, M. & Tsunemitsu, H. (2009). Genetic diversity and classification of the outer
456 capsid glycoprotein VP7 of porcine group B rotaviruses. Arch Virol. 154,1785-1795.
457
458 Kumar, M., Jayaram, H., Vasquez-Del, C. R., Jiang, X., Taraporewala, Z.F., Jacobson,
459 R.H., Patton, J.T. & Prasad, B.V. (2007). Crystallographic and biochemical analysis of
460 rotavirus NSP2 with nucleotides reveals a nucleoside diphosphate kinase-like activity. J
461 Virol. 81,12272-12284.
462
463 Mackow, E.R. (1995). Group B and C rotaviruses. In Infections of the gastrointestinal
464 tract, 1st edn, pp. 983-1008. Edited by M. J. Blaser, P.D. Smith, J.I. Ravdin, H.B.
465 Greenberg & R.L. Guerrant. New York: Raven Press.
466
467 Matthijnssens, J., Ciarlet, M., Heiman, E., Arijs, I., Delbeke, T., McDonald, S.M.,
468 Palombo, E.A., Iturriza-Gómara, M., Maes, P. & other authors (2008). Full genome-based

classification of rotaviruses reveals a common origin between human Wa-Like and porcine rotavirus strains and human DS-1-like and bovine rotavirus strains. J Virol. 82,3204-3219.

McDonald, S.M. & Patton, J.T. (2008). Molecular characterization of a subgroup specificity associated with the rotavirus inner capsid protein VP2. J Virol. 82, 2752-2764.

McDonald, S.M., Aguayo, D., Gonzalez-Nilo, F.D. & Patton, J.T. (2009). Shared and group-specific features of the rotavirus RNA polymerase reveal potential determinants of gene reassortment restriction. J Virol. 83, 6135-6148

Nagashima, S., Kobayashi, N., Ishino, M., Alam, M.M., Ahmed, M.U., Paul, S.K., Ganesh, B., Chawla-Sarkar, M., Krishnan, T. & other authors (2008). Whole genomic characterization of a human rotavirus strain B219 belonging to a novel group of the genus Rotavirus. J Med Virol 80:2023-2033.

Nakata, S., Estes, M.K., Graham, D.Y., Wang, S.S., Gary, G.W. & Melnick, J.L. (1987). Detection of antibody to group B adult diarrhea rotaviruses in humans. J Clin Microbiol. ;25, 812-818.

Paul, S.K., Kobayashi, N., Nagashima, S., Ishino, M., Watanabe, S., Alam, M.M., Ahmed, M.U., Hossain, M.A. & Naik, T.N. (2008). Phylogenetic analysis of rotaviruses with genotypes G1, G2, G9 and G12 in Bangladesh: evidence for a close relationship between rotaviruses from children and adults. Arch Virol. 153,1999-2012.

Rahman, M., Hassan, Z.M., Zafrul, H., Saiada, F., Banik, S., Faruque, A.S., Delbeke, T., Matthijnssens, J., Van, Ranst, M. & other author (2007). Sequence analysis and evolution of group B rotaviruses. Virus Res 125,219-25.

Rao, C.D., Das, M., Ilango, P., Lalwani, R., Rao, B.S. & Gowda, K. (1995). Comparative nucleotide and amino acid sequence analysis of the sequence-specific RNA-binding rotavirus nonstructural protein NSP3. Virology. 207,327-333.

Sanekata, T., Ahmed, M.U., Kader, A., Taniguchi, K. & Kobayashi, N. (2003). Human group B rotavirus infections cause severe diarrhea in children and adults Bangladesh. J Clin Microbiol 41, 2187-2190.

Sen, A., Sen, N. & Mackow, E.R. (2007). The formation of viroplasm-like structures by the rotavirus NSP5 protein is calcium regulated and directed by a C-terminal helical domain. J Virol. 81, 11758-11767.

Shen, S., McKee, T.A., Wang, Z.D., Desselberger, U. & Liu, D.X. (1999) Sequence analysis and in vitro expression of genes 6 and 11 of an ovine group B rotavirus isolate, KB63: evidence for a non-defective, C-terminally truncated NSP1 and a phosphorylated NSP5. J Gen Virol. 80:2077-2085.

Tsunemitsu, H., Kamiyama, M., Kawashima, K., Katsuda, K., Kohmoto, M., Saif, L.J., Shouji, T. & Onodera, T. (2005). Molecular characterization of the major capsid protein VP6 of bovine group B rotavirus and its use in seroepidemiology. J Gen Virol. 86, 2569-2575.

Vende, P., Piron, M., Castagné, N., & Poncet D (2000). Efficient translation of rotavirus mRNA requires simultaneous interaction of NSP3 with the eukaryotic translation initiation factor eIF4G and the mRNA 3' end. J Virol. 74:7064-7071.

524 Wang, S., Cai, R., Chen, J., Li, R. & Jiang, R. (1985). Etiologic studies of the 1983 and
525 1984 outbreaks epidemic diarrhea in Guangxi. Intervirology 24, 140-146.
526
527 Wang, Y.H., Kobayashi, N., Zhou, D.J., Yang, Z.Q., Zhou, X., Peng, J.S., Zhu, Z.R., Zhao,
528 D.F., Liu, M.Q. & other author (2007). Molecular epidemiologic analysis of group A
529 rotaviruses in adults and children with diarrhea in Wuhan city, China, 2000-2006. Arch
530 Virol. 152, 669-685.
531
532 Yang, J.H., Kobayashi, N., Wang, Y.H., Zhou, X., Li, Y., Zhou, D.J., Hu, Z.H., Ishino, M.,
533 Alam, M.M. & other authors (2004). Phylogenetic analysis of a human group B rotavirus
534 WH-1 detected in China in 2002. J Med Virol74, 662-667.
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550

551 FIGURE LEGENDS

552

553 Fig .1

554 Phylogenetic dendrograms (A-K) of group B rotavirus genes (RNA segments encoding
555 VP7, VP1-VP4, VP6, NSP1-NSP5, respectively) constructed by neighbor-joining method
556 with MEGA.4 program. Dendrogram is rooted with human rotavirus strain B219
557 belonging to a novel rotavirus group. Variation scale is described at the bottom. Percent
558 bootstrap support is indicated by the values at each node (the values <80 are omitted).
559 Closed circles indicates strains of which the genes were determined in the present study
560 (IC-008, IDH-084, Bang 117, and MMR-B1). Genotypes (G1-G5) of GBR strains are
561 indicated on the left. Vertical solid line and dotted line represent human GBRs belonging
562 to Indian-Bangladeshi lineage and Chinese lineage, respectively.

563

564 Fig.2

565 The primary amino acid sequence alignment of VP1 from the human GBRs, GAR (Wa),
566 and GCR (Bristol). Structural subdomains are indicated in the box above the sequence
567 alignment. Dot indicates identical amino acid to that of strain WH-1, and amino acids
568 numbers based on WH-1 are indicated above the sequences. A dash denotes gap, and an
569 asterisk indicates identical amino acid among all the rotavirus strains. A colon shows
570 conserved amino acid between GBR and GAR, or GBR and GCR. A catalytic region of
571 RNA polymerase predicted by ScanProsite program is indicated by a line above the
572 sequence, and the consensus motif of RNA polymerase (SG, T, N, T, GDD) (Cohen J, et al.,
573 1989; Nagashima et al., 2008) is shown by underlines.

574

575 Fig.3

576 Alignment of partial VP3 (A) and NSP2 (B) amino acid sequences of from the human
577 GBRs, GAR (Wa), and GCR (Bristol). Dot indicates identical amino acid to that of strain

ADRV, and amino acids numbers based on ADRV are indicated above the sequences. A dash denotes gap, and an asterisk indicates identical amino acid among all the rotavirus strains. A colon shows conserved amino acid between GBR and GAR, or GBR and GCR. A consensus motif (ALYSLSNXXN) of VP3 found in all the rotavirus groups (Nagashima et al., 2008) is shown by a line above the sequence alignment A. Active region of the nucleotide triphosphatase of NSP2 (Kumar et al., 2007) is indicated by a line above the sequences, and a catalytic residue H225 for GAR is shown by an arrowhead.

Fig.4

The primary amino acid sequence alignment of NSP3 from the human GBRs, GAR (Wa), and GCR (Bristol). Dot indicates identical amino acid to that of strain ADRV, and amino acids numbers are indicated above the sequences. A dash denotes gap, and an asterisk indicates identical amino acid among all the rotavirus strains. A colon shows conserved amino acid between GBR and GAR, or GBR and GCR. RNA binding domain (Deo et al., 2002) is indicated by a line above the sequences. Shaded regions indicate alpha helices (H1-H8) and beta strands (S1-S3) in the RNA-binding domain.

Table 1. Sequence identities (%) of individual gene segments among human group B rotaviruses

Gene segment	Indian, Bangladeshi, Myanmarese strains[a](A)		Chinese strains[b](B)		Between A and B	
	Nucleotide	Amino acid	Nucleotide	Amino acid	Nucleotide	Amino acid
VP1 gene	97.4-99.8	98.7-99.9			90.3-91.7	95.4-95.9
VP2 gene	97.2-99.4	98.9-99.9	97.963	98.501	90.4-91.8	97.4-98.1
VP3 gene	97.5-99.5	98.6-100.0			90.3-91.2	94.0-94.5
VP4 gene	97.3-99.4	97.5-99.6	98.439	98.133	90.7-92.4	93.5-96.1
VP5∗ region	97.6-99.7	98.7-99.8	98.545	98.694	91.5-93.2	94.8-96.8
VP8∗region	95.9-99.5	93.2-99.5	98.113	96.618	88.4-90.4	89.9-94.7
VP6 gene	98.0-99.9	98.7-100.0	98.582	98.721	93.1-94.6	96.7-98.5
VP7 gene	97.7-99.9	98.0-100.0	98.649	99.197	91.3-92.6	94.0-95.2
NSP1 gene	97.6-99.6		98.824		91.6-92.8	
peptide1 region		97.2-100.0		100		97.2-99.1
peptide2 region		97.2-99.4		98.754		91.6-92.8
NSP2 gene	98.0-99.4	98.0-99.7	98.373	99.003	93.6-94.9	96.0-98.0
NSP3 gene	97.0-99.7	96.8-99.7	98.137	97.983	89.8-91.2	89.0-89.9
NSP4 gene	96.1-99.9	97.3-100.0	98.4	98.63	90.8-92.9	93.2-95.9
NSP5 gene	95.6-100.0	94.7-100.0	98.891	99.412	89.5-92.4	90.6-93.5

GenBank accession nos. of genes from human GBRs published previously and used in this analysis are as follows : CAL-1, EU490414,EU490418, AB037931- AB037932, AF184083-AF184084, AF230974-AF230975, AY238383, AY238386-AY238387 ; Bang373, EU490415, EU490418

[a] CAL-1, Bang373, IC-008, IDH-084, Bang117, MMR-B1

[b] **ADRV, WH-1(VP1 and VP3 sequences are available only for WH-1)**

Fig.1

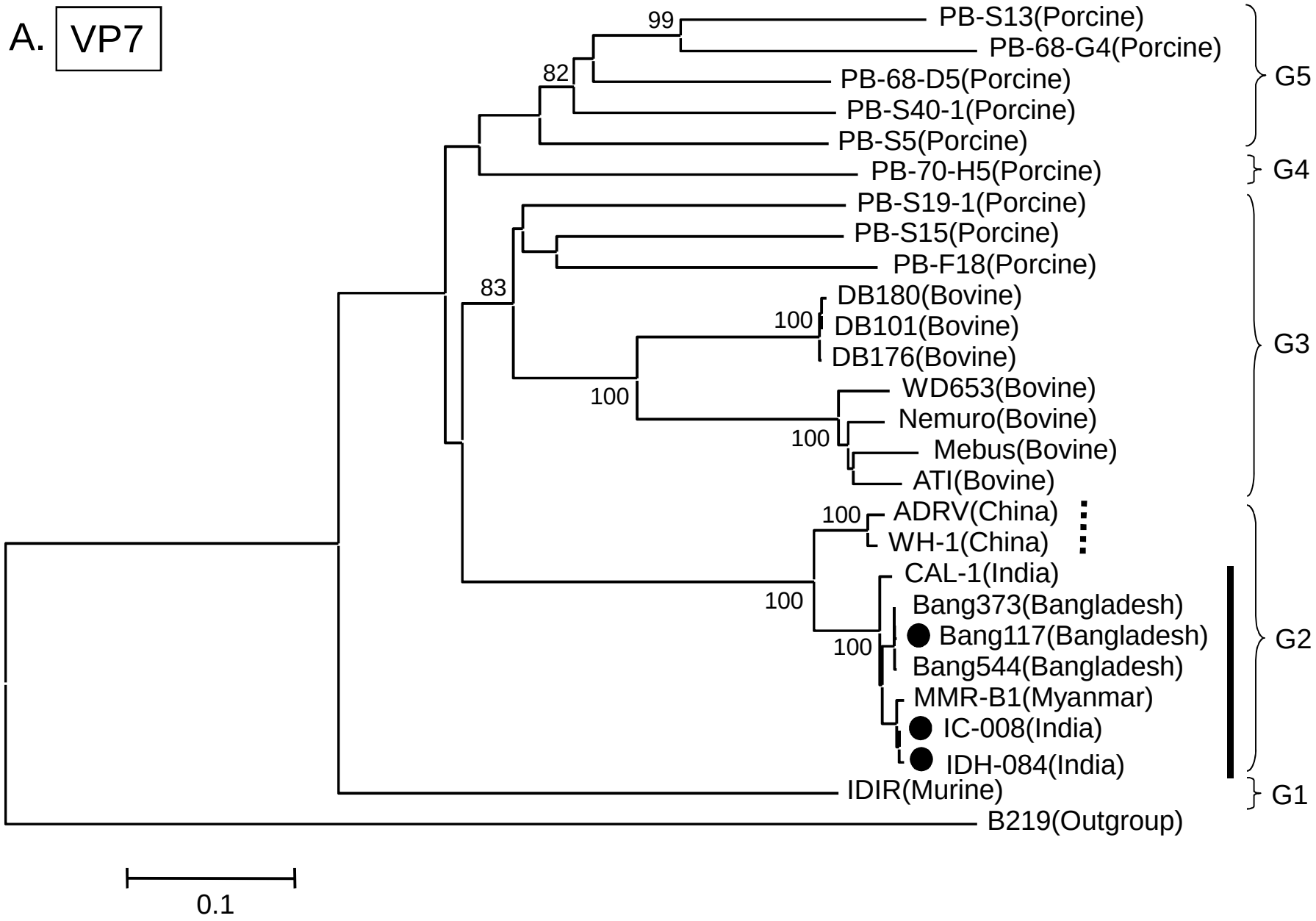

A.
VP7
99
82
83
100
100
100
100
100
PB-S13(Porcine)
PB-68-G4(Porcine)
PB-68-D5(Porcine)
PB-S40-1(Porcine)
PB-S5(Porcine)
PB-70-H5(Porcine)
PB-S19-1(Porcine)
PB-S15(Porcine)
PB-F18(Porcine)
DB180(Bovine)
DB101(Bovine)
DB176(Bovine)
WD653(Bovine)
Nemuro(Bovine)
Mebus(Bovine)
ATI(Bovine)
ADRV(China)
WH-1(China)
CAL-1(India)
Bang373(Bangladesh)
Bang117(Bangladesh)
Bang544(Bangladesh)
MMR-B1(Myanmar)
IC-008(India)
IDH-084(India)
IDIR(Murine)
B219(Outgroup)
G5
G4
G3
G2
G1
0.1

Fig.1

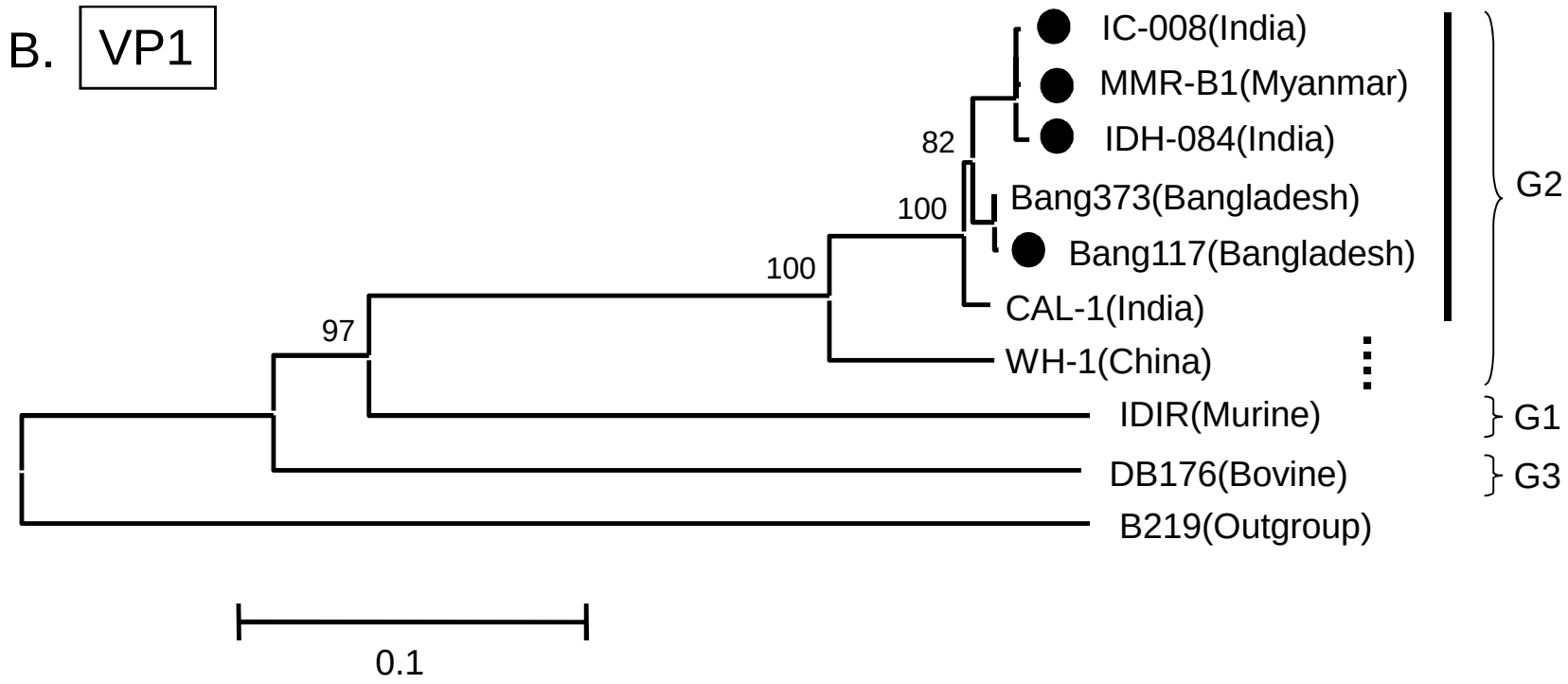

B.
VP1
IC-008(India)
MMR-B1(Myanmar)
IDH-084(India)
Bang373(Bangladesh)
Bang117(Bangladesh)
CAL-1(India)
WH-1(China)
IDIR(Murine)
DB176(Bovine)
B219(Outgroup)
82
100
100
97
G2
G1
G3
0.1

Fig.1

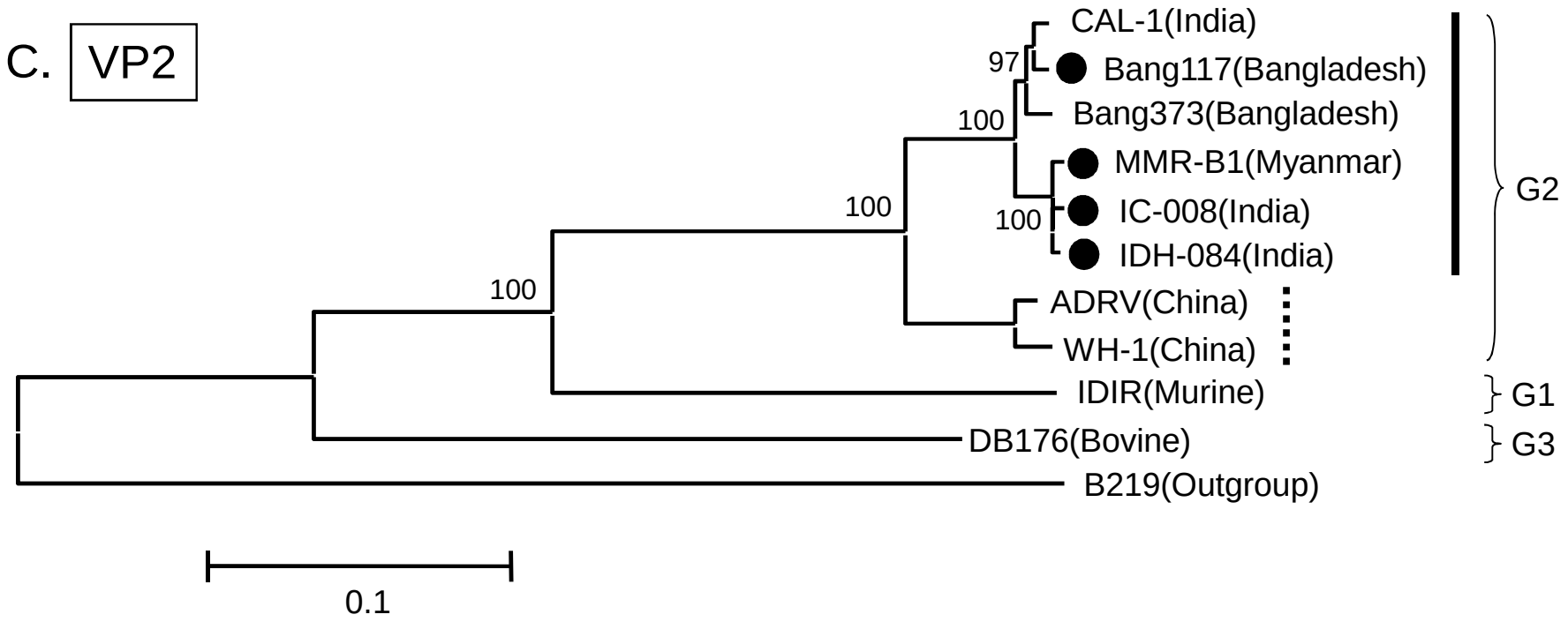

C.
VP2
CAL-1(India)
Bang117(Bangladesh)
Bang373(Bangladesh)
MMR-B1(Myanmar)
IC-008(India)
IDH-084(India)
ADRV(China)
WH-1(China)
IDIR(Murine)
DB176(Bovine)
B219(Outgroup)
97
100
100
100
100
G2
G1
G3
0.1

Fig.1

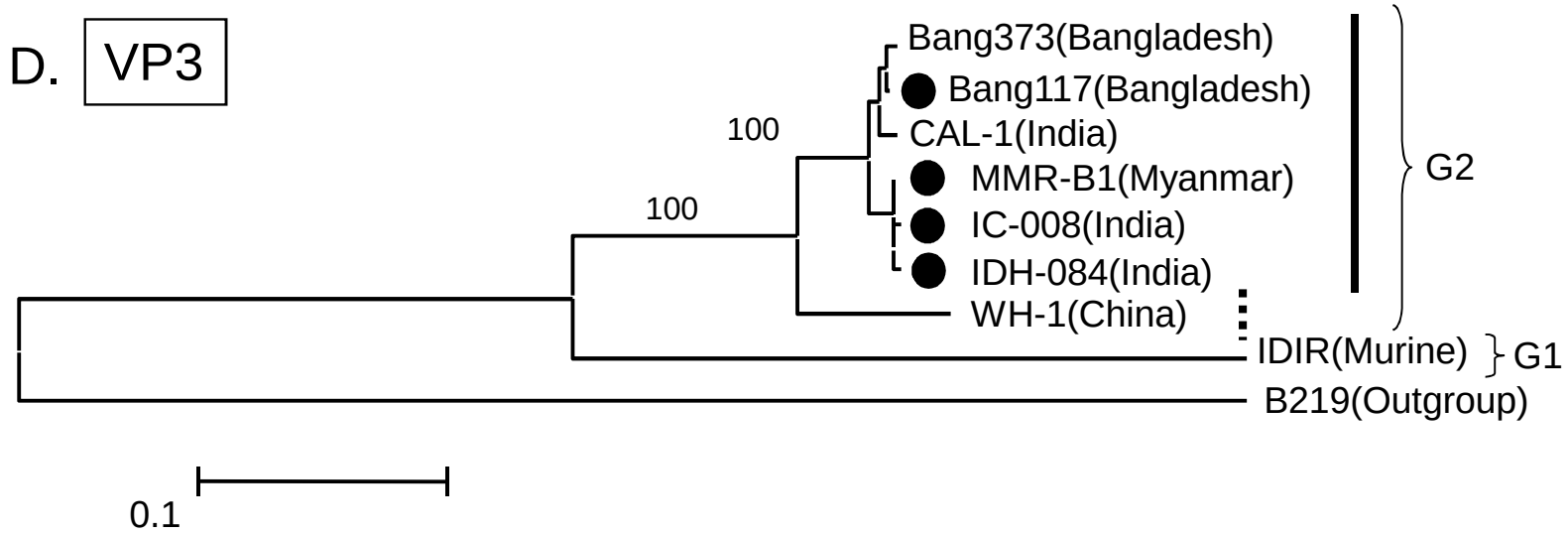

D.
VP3
Bang373(Bangladesh)
Bang117(Bangladesh)
CAL-1(India)
MMR-B1(Myanmar)
IC-008(India)
IDH-084(India)
WH-1(China)
IDIR(Murine)
B219(Outgroup)
100
100
G2
G1
0.1

Fig.1

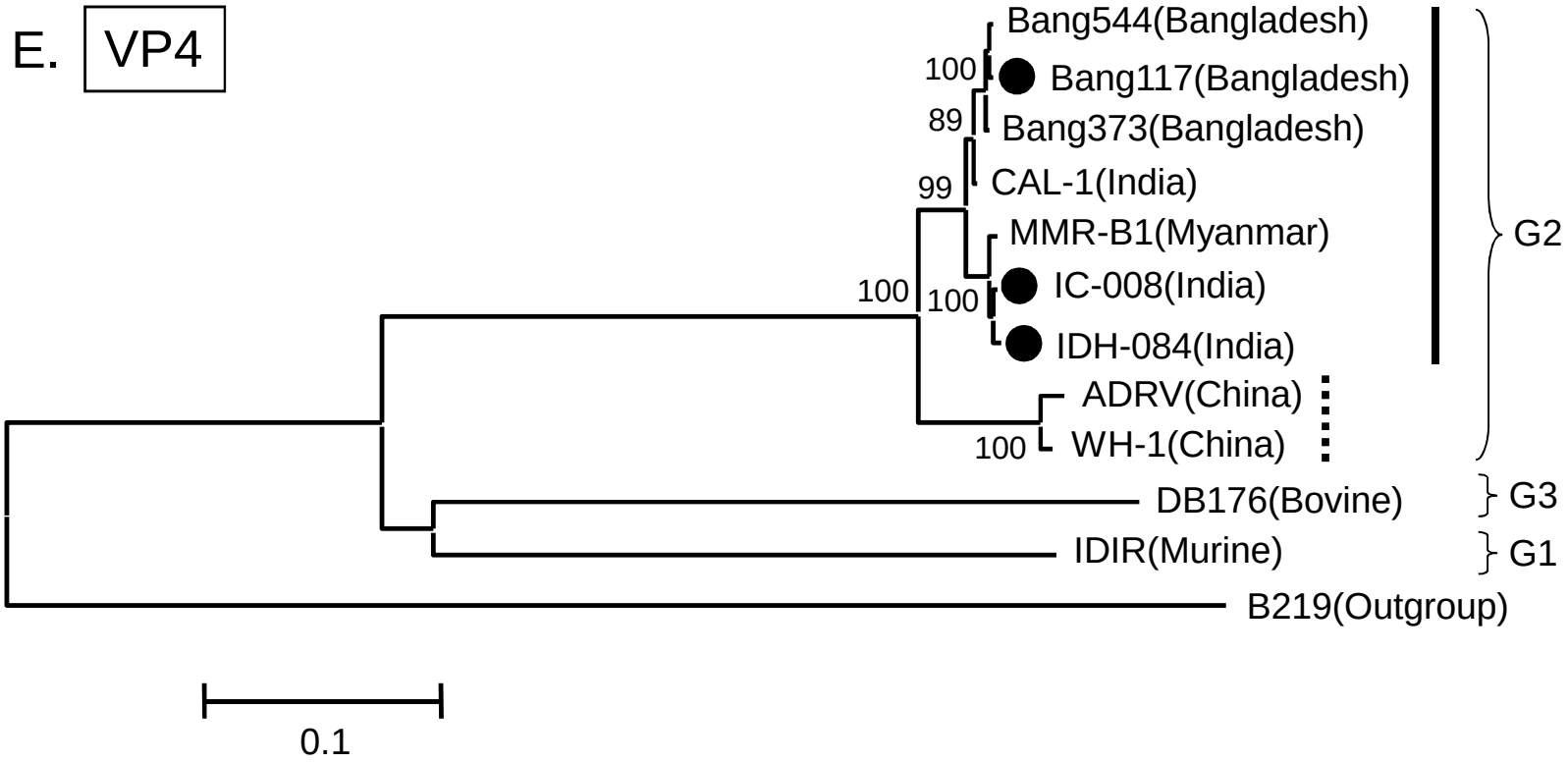

E.
VP4
Bang544(Bangladesh)
100
Bang117(Bangladesh)
89
Bang373(Bangladesh)
99
CAL-1(India)
MMR-B1(Myanmar)
100
100
IC-008(India)
IDH-084(India)
ADRV(China)
100
WH-1(China)
G2
DB176(Bovine)
G3
IDIR(Murine)
G1
B219(Outgroup)
0.1

Fig.1

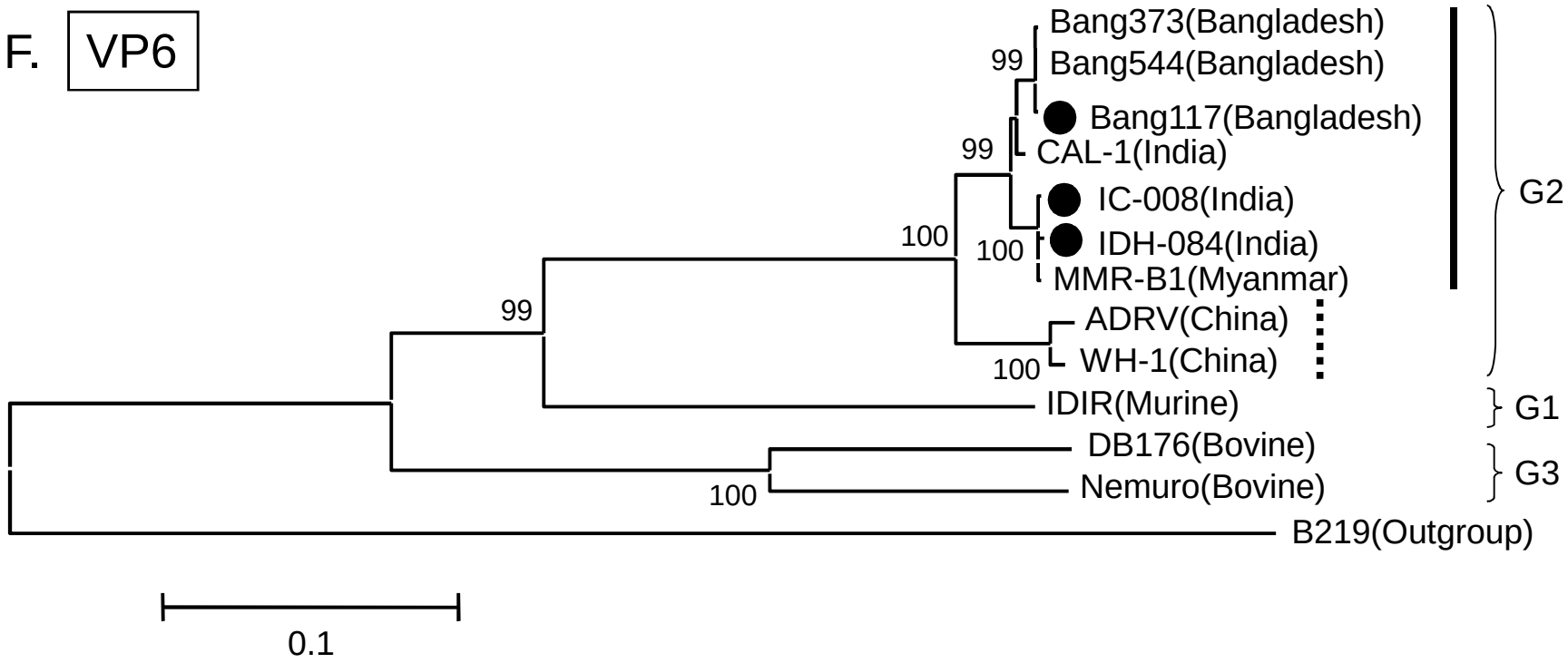

F.
VP6
Bang373(Bangladesh)
Bang544(Bangladesh)
Bang117(Bangladesh)
CAL-1(India)
IC-008(India)
IDH-084(India)
MMR-B1(Myanmar)
ADRV(China)
WH-1(China)
IDIR(Murine)
DB176(Bovine)
Nemuro(Bovine)
B219(Outgroup)
99
99
100
100
99
100
100
G2
G1
G3
0.1

Fig.1

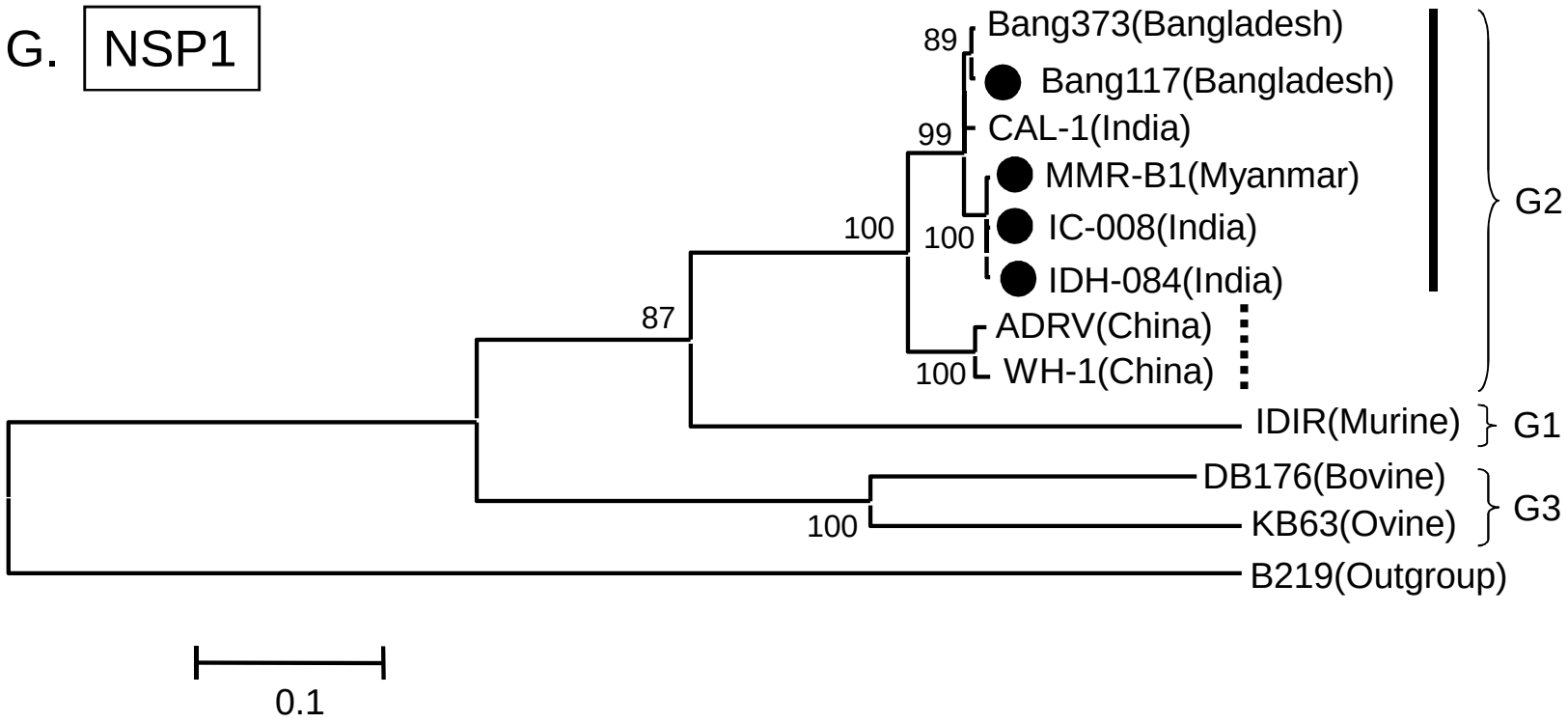

G.
NSP1
Bang373(Bangladesh)
Bang117(Bangladesh)
CAL-1(India)
MMR-B1(Myanmar)
IC-008(India)
IDH-084(India)
ADRV(China)
WH-1(China)
IDIR(Murine)
DB176(Bovine)
KB63(Ovine)
B219(Outgroup)
89
99
100
100
87
100
100
G2
G1
G3
0.1

Fig.1

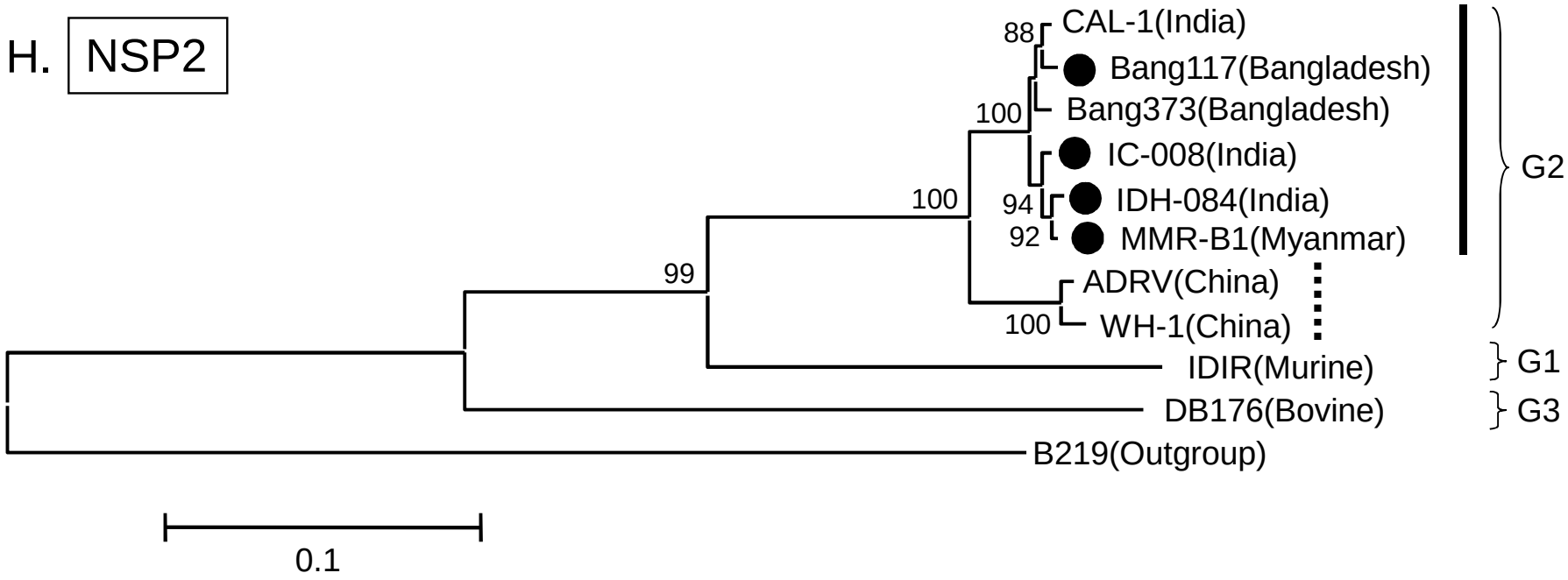

H.
NSP2
CAL-1(India)
Bang117(Bangladesh)
Bang373(Bangladesh)
IC-008(India)
IDH-084(India)
MMR-B1(Myanmar)
ADRV(China)
WH-1(China)
IDIR(Murine)
DB176(Bovine)
B219(Outgroup)
88
100
100
94
92
99
100
G2
G1
G3
0.1

Fig.1

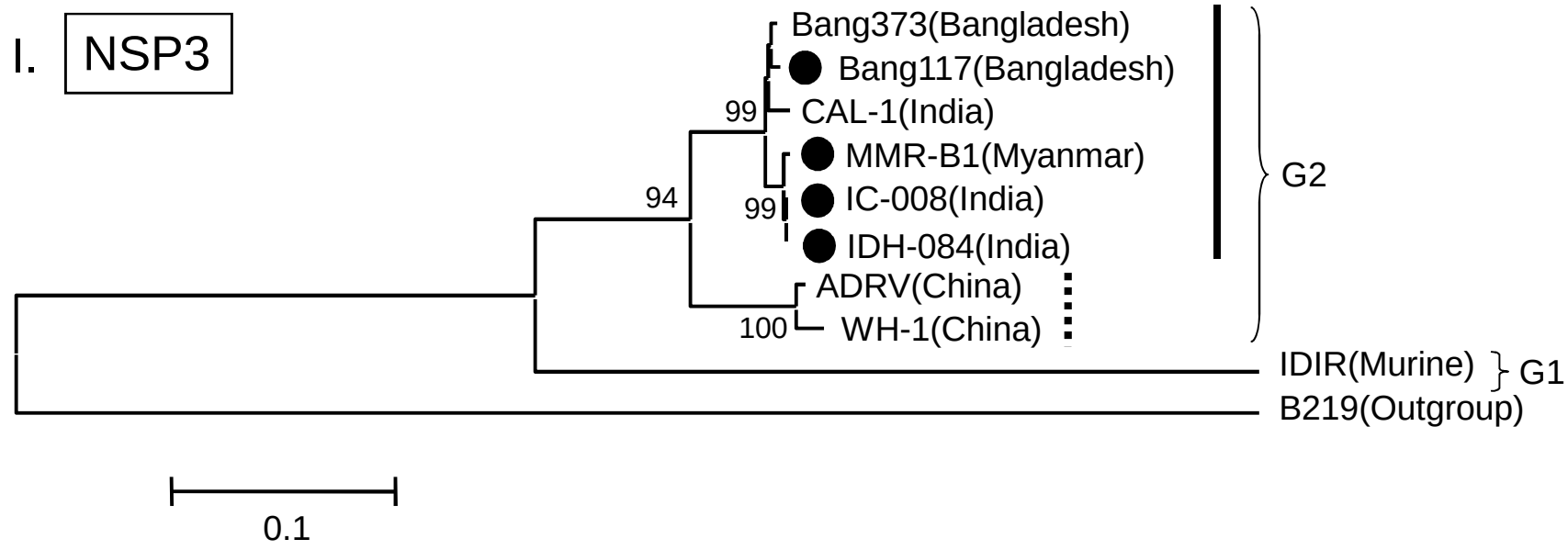

I.
NSP3
Bang373(Bangladesh)
Bang117(Bangladesh)
99
CAL-1(India)
MMR-B1(Myanmar)
94
99
IC-008(India)
IDH-084(India)
ADRV(China)
100
WH-1(China)
G2
IDIR(Murine)
G1
B219(Outgroup)
0.1

Fig.1

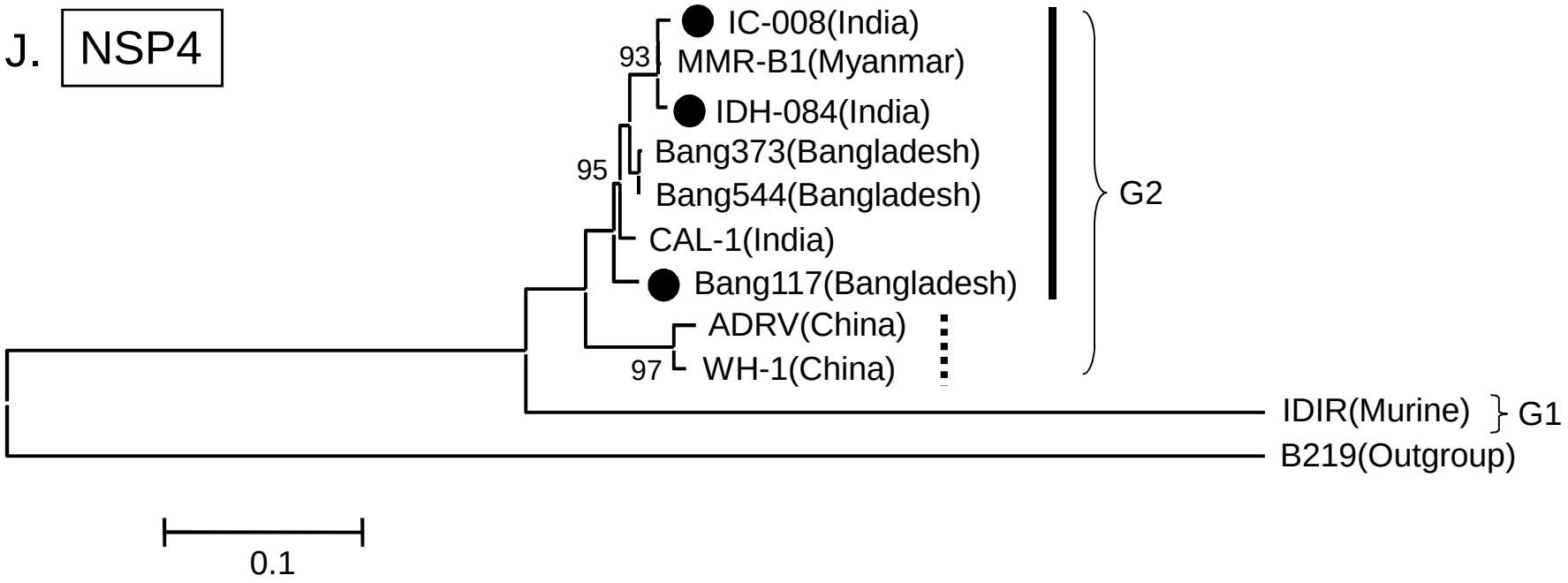

J.
NSP4
IC-008(India)
93
MMR-B1(Myanmar)
IDH-084(India)
Bang373(Bangladesh)
95
Bang544(Bangladesh)
CAL-1(India)
Bang117(Bangladesh)
ADRV(China)
97
WH-1(China)
G2
IDIR(Murine)
G1
B219(Outgroup)
0.1

Fig.1

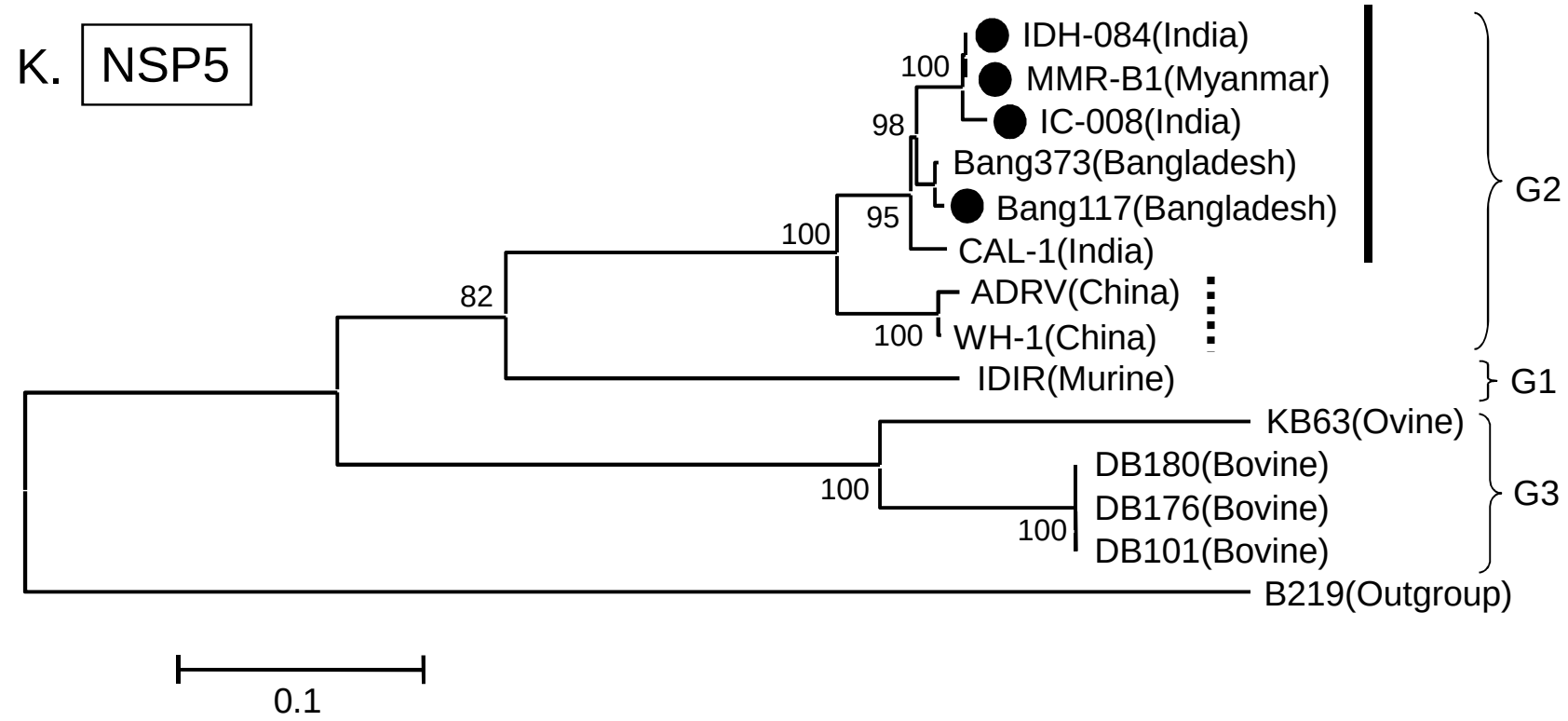

K.
NSP5
IDH-084(India)
MMR-B1(Myanmar)
IC-008(India)
Bang373(Bangladesh)
Bang117(Bangladesh)
CAL-1(India)
ADRV(China)
WH-1(China)
IDIR(Murine)
KB63(Ovine)
DB180(Bovine)
DB176(Bovine)
DB101(Bovine)
B219(Outgroup)
100
98
95
100
82
100
100
100
G2
G1
G3
0.1

Fig. 2

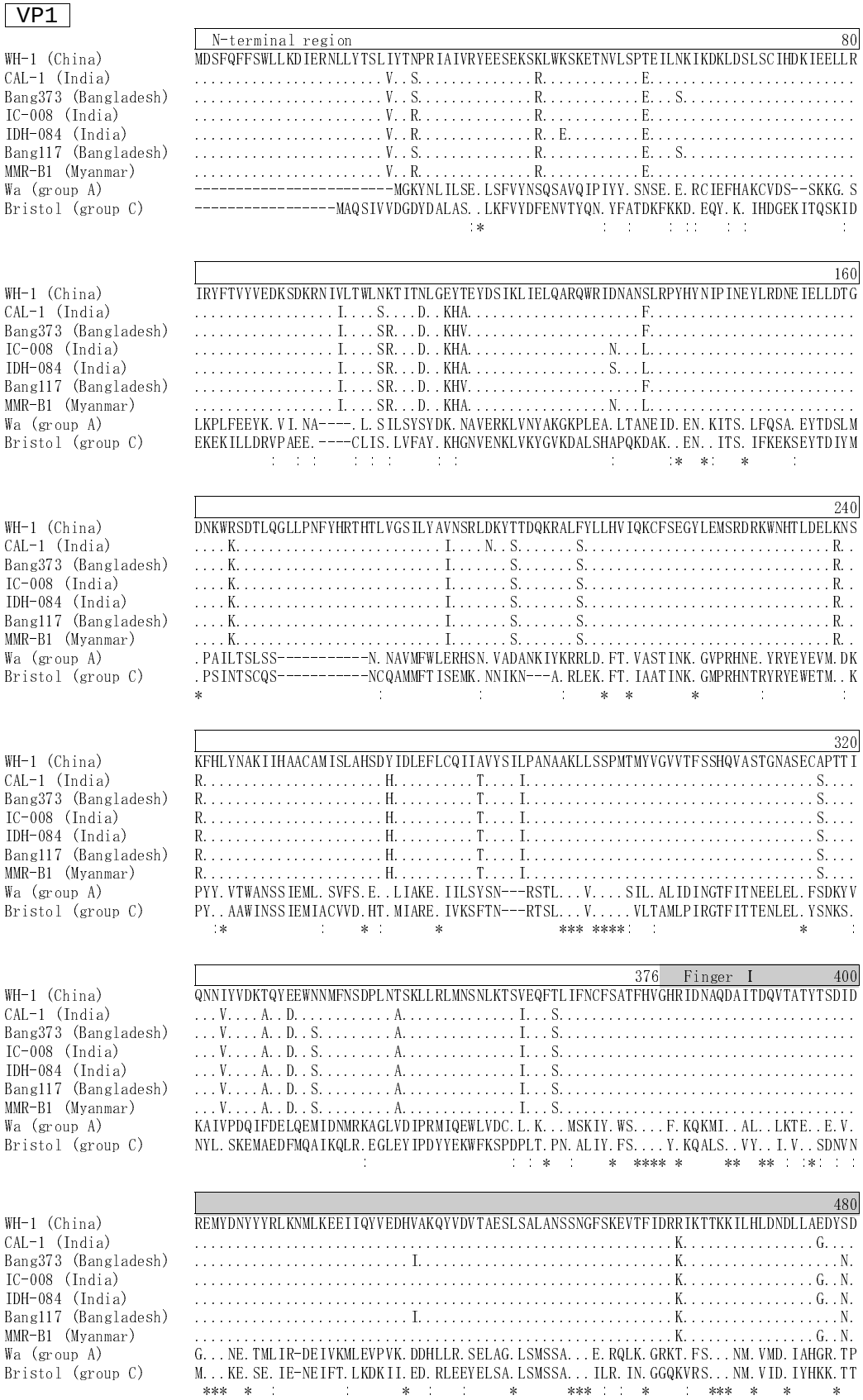

VP1
N-terminal region
80
160
240
320
376
Finger I
400
480
WH-1 (China)
CAL-1 (India)
Bang373 (Bangladesh)
IC-008 (India)
IDH-084 (India)
Bang117 (Bangladesh)
MMR-B1 (Myanmar)
Wa (group A)
Bristol (group C)

Fig. 2-continued

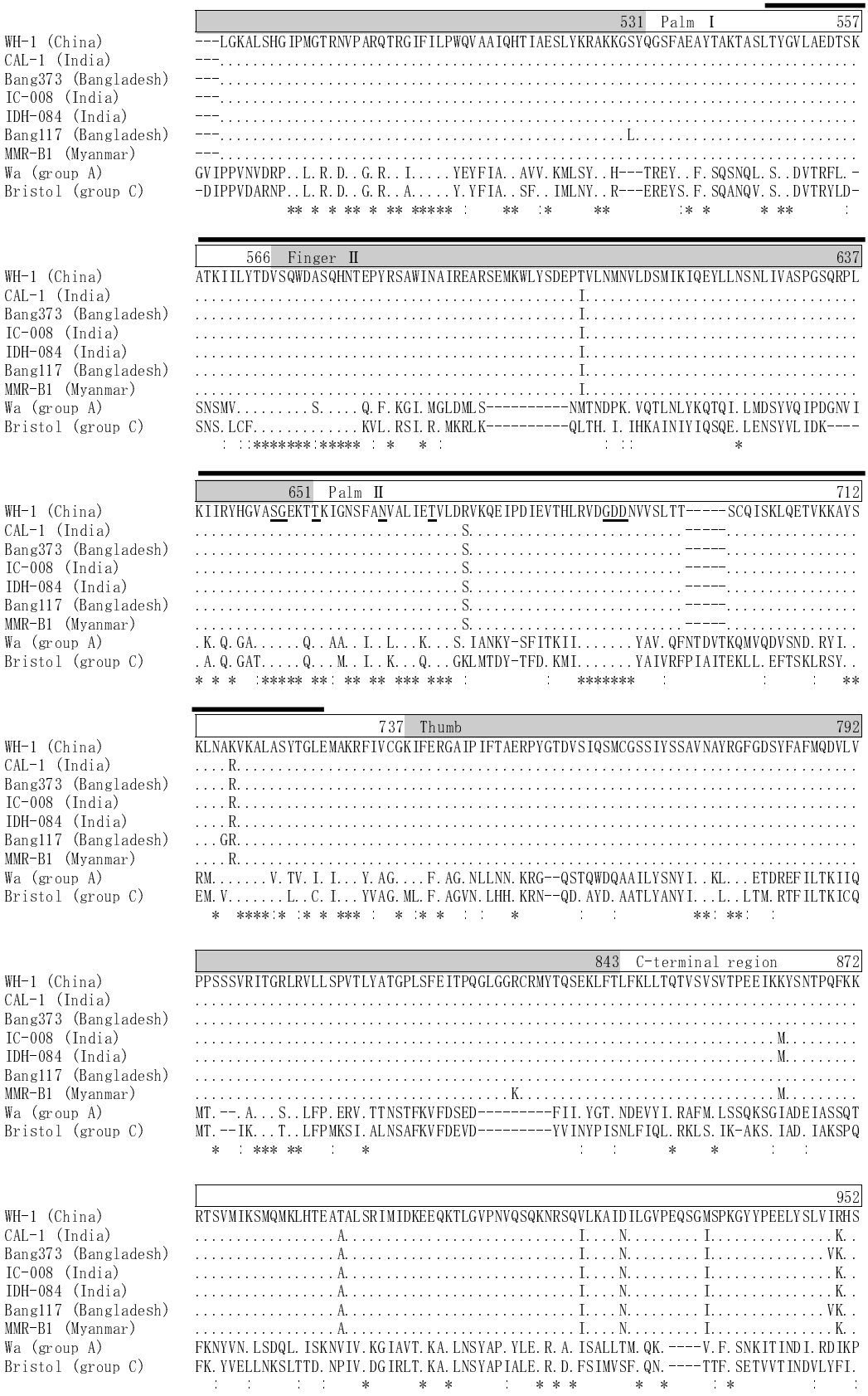


Fig. 2-continued

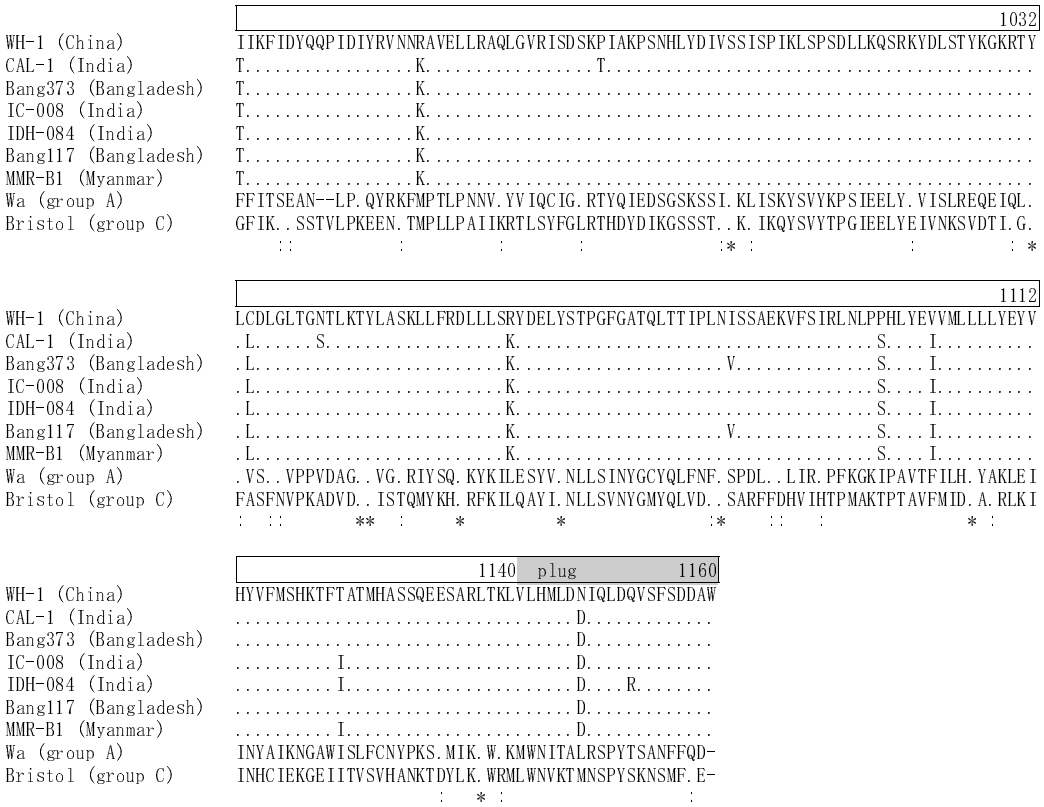


Fig.3

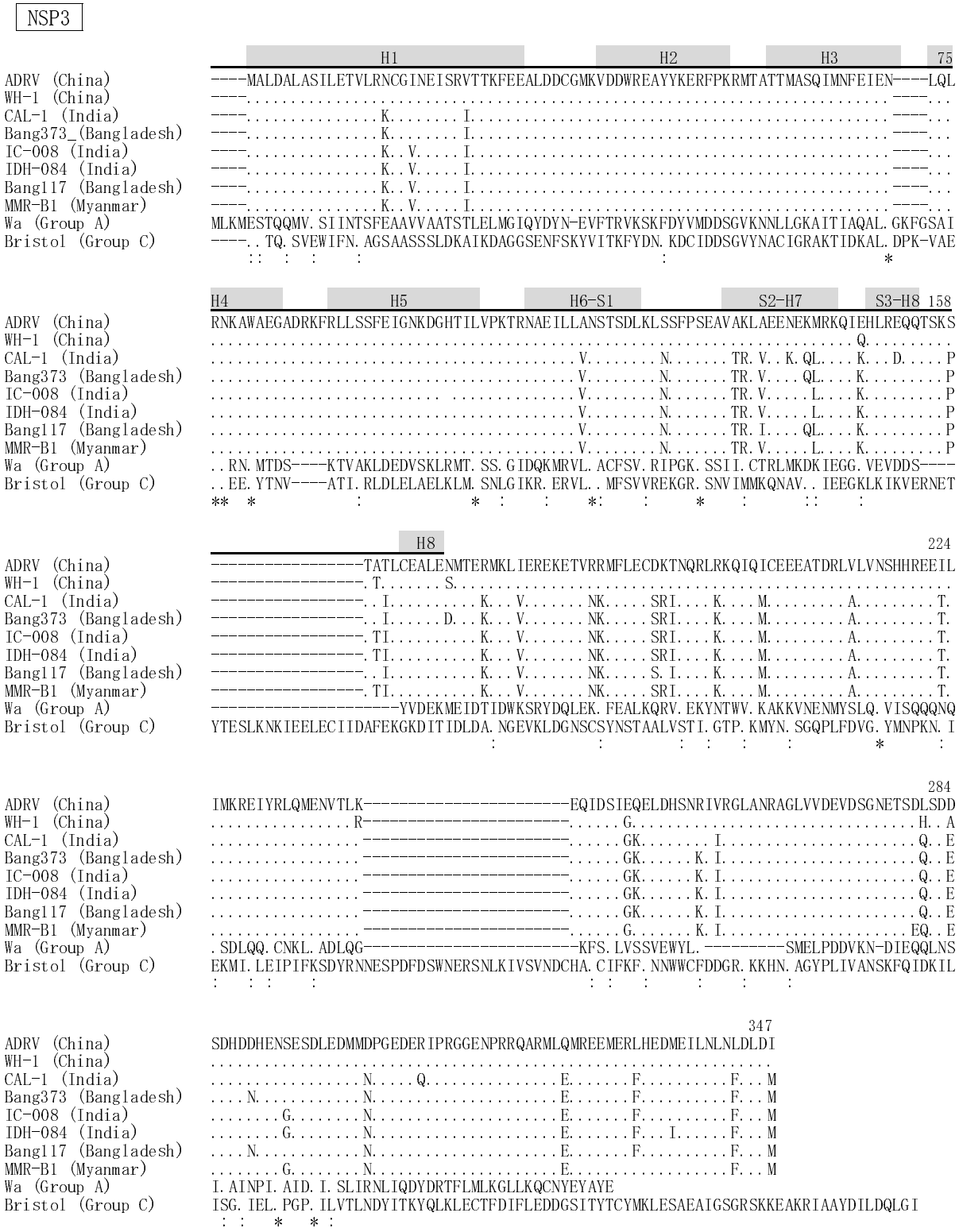
A. VP3

```
                            371                                                                           450
                                                                        __________
WH-1 (China)                MGDKLGTVYNEDDRIAAKANNLPNYVFGGVPFTAAALRFDYVNIALYSLSNSVNSPELIKATLSYDHIFTFPSYSKGDWR
CAL-1 (India)               ..N....T........V.T..................K...I......................................
Bang373 (Bangladesh)        ..N....T........V.T..................K...I......................................
IC-008 (India)              ..N....T........V.T..................K...I......................................
IDH-084 (India)             ..N....T........V.T..................K...I...................P..................
Bang117 (Bangladesh)        ..N....T........V.T..................K...I......................................
MMR-B1 (Myanmar)            ..N....T........V.T..................K...I......................................
Wa (group A)                WDIITIKRFIPKGVFY.FI..VTTENVFIQ.PFKLKTSPTDYIV...A...DL..RQDVINLINKQKQSLITVRINNTFK
Bristol (group C)           HII.RQKIKVTKSLMYNAI.TIYSDNVFISGKYSLRG-KTEGVL...C...TI.QK.KVIQYANSFSGTCMTVRLNNTYE
                                               *:         :             *** ***  *: :
```

B. NSP2

```
                            208                                                                           278
                                                       _________
ADRV (China)                EQPVNGMLALKAVAGN------QFFMYHGHGHIRTVPYHELADAIKSFARKDKETLEN---ISKSPLAAQCGSKFLDMLD
WH-1 (China)                ................------....................................---V..................
CAL-1 (India)               ................------.........................Y.........S---...................
Bang373 (Bangladesh)        ................------...................................S---...................
IC-008 (India)              ................------...................................S---...................
IDH-084 (India)             ................------...................................S---.................I.
Bang117 (Bangladesh)        ................------...................................S---...................
MMR-B1 (Myanmar)            ................------...................T...............S---.................I.
Wa (Group A)                DK.ISDVCIKEL..ELRWQYNRFAVIT..K..Y.V.K.SSV.NHADRVFATY.NSAKSGNV.DFNL.DQRIIWQNWYAFT
Bristol (Group C)           KTDIPDRNQTAFA.YIRYNFNKFAAIS..KR.W.L.LHSQ.MSHAERLD..I.SDKKHGRQF.YDDGDMAFVHPGWKTCI
                              :         :*              ** :* * * :  :        :: *         :
                                                               ▲
```

Fig. 4


NSP3
H1
H2
H3
H4
H5
H6-S1
S2-H7
S3-H8
H8
ADRV (China)
WH-1 (China)
CAL-1 (India)
Bang373_(Bangladesh)
IC-008 (India)
IDH-084 (India)
Bang117 (Bangladesh)
MMR-B1 (Myanmar)
Wa (Group A)
Bristol (Group C)