



**HAL**  
open science

# Sequencing and phylogenetic analysis of the coding region of six common rotavirus strains: evidence for intragenogroup reassortment among co-circulating G1P[8] and G2P[4] strains from the United States

Krisztian Banyai

► **To cite this version:**

Krisztian Banyai. Sequencing and phylogenetic analysis of the coding region of six common rotavirus strains: evidence for intragenogroup reassortment among co-circulating G1P[8] and G2P[4] strains from the United States. *Journal of Medical Virology*, 2011, 83 (3), pp.532. 10.1002/jmv.21977 . hal-00610730

**HAL Id: hal-00610730**

**<https://hal.science/hal-00610730>**

Submitted on 24 Jul 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.




**Sequencing and phylogenetic analysis of the coding region of six common rotavirus strains: evidence for intragenogroup reassortment among co-circulating G1P[8] and G2P[4] strains from the United States**

|                               |   |
|-------------------------------|---|
| Journal:                      | <i>Journal of Medical Virology</i>  |
| Manuscript ID:                | JMV-10-2008.R1  |
| Wiley - Manuscript type:      | Research Article  |
| Date Submitted by the Author: | 13-Sep-2010   |
| Complete List of Authors:     | Banyai, Krisztian; Baranya County Institute of State Public Health Service, Regional Laboratory of Virology |
| Keywords:                     | evolution, genome sequencing, Wa-like rotaviruses, DS1-like rotaviruses, United States of America           |
|                               |   |

SCHOLARONE™  
Manuscripts

view


1  
2  
3  
4  
5  
6  
7  
8 Sequencing and phylogenetic analysis of the coding region of six  
9 common rotavirus strains: evidence for intragenogroup reassortment  
10 among co-circulating G1P[8] and G2P[4] strains from the United  
11 States  
12  
13  
14  
15  
16  
17  
18  
19

20 *Bányai K*<sup>1\*</sup>, *Mijatovic-Rustempasic S*<sup>2\*</sup>, *Hull JJ*<sup>2</sup>, *Esona MD*<sup>2</sup>, *Freeman MM*<sup>2</sup>,  
21 *Frace MA*<sup>2</sup>, *Bowen MD*<sup>2</sup>, *Gentsch JR*<sup>2</sup>

22  
23  
24  
25 <sup>1</sup> Veterinary Medical Research Institute, Budapest, Hungary

26 <sup>2</sup> Centers for Disease Control and Prevention, Atlanta, GA, USA  
27  
28  
29

30 \* K.B. and S.M.R. have equally contributed to this work  
31  
32

33  Division of Viral Diseases, National Center for Immunization and Respiratory Diseases,  
34 Centers for Disease Control and Prevention, Atlanta GA 30333; Phone: 404-639-2860, Fax: 404-  
35 639-3645, email: [jgentsch@cdc.gov](mailto:jgentsch@cdc.gov)  
36  
37  
38  
39  
40

41 Running head: Genome sequences of common US G1P[8] and G2P[4] rotavirus  
42 strains  
43  
44  
45

46 Disclaimer: The findings and conclusions in this report are those of the authors  
47 and do not necessarily represent the views of the CDC.  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Abstract

The segmented genome of rotaviruses provides an opportunity for rotavirus strains to generate a large genetic diversity through reassortment; however, this mechanism is considered to play little role in the generation of mosaic gene constellations between Wa-like and DS1-like strains in genes other than the neutralization antigens. A pilot study was undertaken to analyze these two epidemiologically important strains at the genomic level in order to (i) identify intergenogroup reassortment and (ii) to make available additional reference genome sequences of G1P[8] and G2P[4] for future genomics analyses. The full or nearly complete coding region of all 11 genes for 3 G1P[8] (LB2719, LB2758 and LB2771) and 3 G2P[4] (LB2744, LB2764 and LB2772) strains isolated from children hospitalized with severe diarrhea in Long Beach, California, where these strains were circulating at comparable rates during 2005-2006 are described in this study. Based on the full-genome classification system, all G1P[8] strains had a conserved genomic constellation: G1-P[8]-I1-R1-C1-M1-A1-N1-T1-E1-E1-H1 and were mostly identical to the few Wa-like strains whose genome sequences have already been determined. Similarly, the genome sequences of the 3 G2P[4] strains were highly conserved: G2-P[4]-I2-R2-C2-M2-A2-N2-T2-E2-E2-H2 and displayed an overall lesser genetic divergence with reference DS-1-like strains. While intergenogroup reassortment was not seen between the G1P[8] and G2P[4] strains studied here, evidence for intragenogroup reassortment events was identified. Similar studies in the post rotavirus genomic era will help uncover whether intergenogroup reassortment affecting the backbone genes could play a significant role in any potential vaccine breakthrough events by evading immunity of vaccinated children.

## Key words

evolution, genome sequencing, Wa-like rotaviruses, DS1-like rotaviruses, United States of America

## Introduction

*Rotavirus A* is the most important species within the *Rotavirus* genus (*Sedoreovirinae*, *Reoviridae*) and the main cause of acute dehydrating diarrhea in infants and young children worldwide [Estes and Kapikian, 2007]. The traditional classification scheme to designate individual strains based upon serotype specificities of the neutralizing antigens (G and P types) has been recently extended to the entire genome to allow specifying the genotype of all 11 genome segments. Use of this new classification system during sequencing studies confirmed the existence of the two main human genogroups (Wa and DS1) and permits more detailed studies of inter-strain variability and evolutionary history [Matthijnssens et al., 2008a, b].

The epidemiology of rotaviruses is complex. Typically, each year multiple strain variants co-circulate in a given area and strain prevalence changes year-to-year [Matthijnssens et al., 2009; Martella et al., 2010]. The existing genetic and antigenic diversity within many serotypes enables intratypic shifts to preserve the dominance of a particular antigenic type even in those epidemic years when serotype specificity of dominating strains remains the same [Bányai et al., 2009]. During the 1980s four serotypes (G1P[8], G2P[4], G3P[8], and G4P[8]) were predominant worldwide, mostly in temperate climates and an additional serotype, G9P[8], emerged to become epidemiologically important during the 1990s. While these 5 antigenic types remained important during the 2000s, additional globally emerging strains (G12P[6], G12P[8]) have been introduced and clonally expanded in many areas [Matthijnssens et al., 2009].

The first synchronized rotavirus strain surveillance network in the United States was launched in 1995-1996 by the Centers for Disease Control and Prevention (CDC; Atlanta, GA) with a group of participating laboratories located across the United States volunteering to provide rotavirus positive stool samples during each rotavirus season [Ramachandran et al., 1998; Griffin et al., 2000]. As summarized recently, during the 9-year pre rotavirus vaccine surveillance period

1  
2  
3 (1996-2005) G1P[8] was extremely predominant in the United States followed by  
4 G2P[4], and then by G9P[8], G9P[6], G3P[8], G4P[8], and several rare strains,  
5 even though temporal and geographical variation in strain distribution was  
6 marked [Gentsch et al., 2009].  
7  
8  
9

10  
11 Historic observations suggest that rotavirus strains belonging to different  
12 genogroups do not exchange readily their genome segments except for the outer  
13 capsid coding genes although a few exceptions have been noted [Nakagomi et  
14 al., 1989; Nakagomi and Nakagomi, 2002; Medici et al., 2007]. For example,  
15 Kaga and Nakagomi [1994] described an outbreak in Japan where a single  
16 segment reassortant between Wa and DS1 genogroup strains were identified.  
17 Also, Palombo et al. [1996] have described a multi-segment reassortant outbreak  
18 between Wa and DS1 genogroup strains in Australia. These findings contrast  
19 with the original perception on the restricted capability of strains belonging to  
20 different genogroups to reassort and thus it remains to be determined whether  
21 such reassortment events of non-serotype genes are truly rare or might be more  
22 common than expected based on the few studies published in this subject.  
23 Ironically, while the number of genome sequences of epidemiologically minor  
24 strains increases continuously [Bányai et al., 2009b, 2009c, 2009d; De Grazia et  
25 al., 2010; Esona et al., 2009a, 2009b; Matthijnssens et al., 2006a, 2006b, 2008a,  
26 2008b, 2008c, 2009b, 2010; Mukherjee et al., 2009; Pietsch et al., 2009;  
27 Tsugawa and Hoshino, 2008; Wang et al., 2010], very few genome sequences  
28 are available for representatives of the epidemiologically major Wa-like G1P[8]  
29 and DS1-like G2P[4] strains, an important shortcoming that does not allow  
30 evaluating comprehensively the significance of intergenogroup reassortment in  
31 their epidemiologic fitness and evolution. Because it seemed reasonable to  
32 assume that high frequency of co-circulating Wa-like G1P[8] and DS1-like G2P[4]  
33 strains could favor the generation of rotavirus strains with mixed gene  
34 configuration a pilot study was undertaken at CDC to analyze epidemiologically  
35 important strains at the genomic level. A single US surveillance area and season  
36 was sought in the records where high co-circulation of G1P[8] and G2P[4] strains  
37 was observed.  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Materials and Methods

### *Study design*

In this study the 2005-2006 season in Long Beach, California was chosen, where G1P[8] and G2P[4] strains co-circulated at comparable rates, 38% to 32%, respectively. A total of six strains (G1P[8], LB2719, LB2758, and LB2771; G2P[4], LB2744, LB2764, and LB2772) were selected for genetic analysis; to increase the chance to find any reassortants, these 6 strains were selected on the basis of distinctness in their RNA profiles in a polyacrylamide gel (not shown). Stool specimens were utilized in order to avoid any potential nucleotide changes that may be associated with cell culture adaptation.

### *Genome sequencing*

The strategy to obtain the nearly full length genomes of the selected strains included the application of type common primer sets designed to the ends of each genome segment in individual RT-PCR assays followed by the primer walking strategy. Briefly, genomic RNA was extracted from 10% stool suspension by TRIzol as described elsewhere [Jakab et al., 2004]. Then, 6  $\mu$ l RNA was mixed with the appropriate combination of gene specific primer pairs (2  $\mu$ l of 20  $\mu$ M of each primer) in a total volume of 10  $\mu$ l and then denatured at 97°C for 5 min. Then the RT-PCR mixture (One-step RT-PCR kit, QIAGEN) was added in a 30  $\mu$ l final reaction volume. The thermal profile included incubation at 42°C for 30 min for reverse transcription, followed by activation of the *Taq* DNA polymerase at 95°C for 15 min, and 40 cycles of amplification (94°C, 30 sec; 42°C, 30 sec; 72°C, 3 min) followed by final extension at 72°C for 7 min. Amplicons were run in low melting point agarose gels and stained with ethidium bromide, followed by excision and purification using the Qiaquick gel extraction kit (QIAGEN) according to the manufacturer's instructions. To avoid contamination of samples the genes of each strain were amplified and purified on separate days.

### *Computer analyses*

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Sequence contigs were generated by Sequencher 4.8 (Gene Codes Corporation, Inc., Ann Arbor, MI). The phylogenetic analysis was done by the neighbor-joining method in MEGA4.1 software [Tamura et al., 2007]. The significance of tree topologies was estimated by bootstrap analysis with 1,000 pseudoreplicates.

#### *GenBank accession numbers*

Sequences were deposited in GenBank under the following accession numbers: HM467916-HM467981.

## **Results**

In general, the complete or nearly complete ORFs of all 11 genes for 6 strains were determined. Short stretches remained undetermined for the VP1, VP2, VP4, and NSP1 genes, where PCR primers overlapped with few nucleotides at the 5' ends of the ORF. This resulted in the loss of 7 amino acids in the predicted coding region (~0.1% of the full length polypeptide sequence of Wa; Table 1). The missing sequences are highly conservative regions of 5' ends of the respective genes. As expected, the ORFs of Wa-like and DS1-like strains were similar in terms of sequence lengths; exceptions include the VP2 gene, where notable variance was seen in length at the N-terminal end of the protein. For example, the 3 Wa-like US G1P[8] strains had 4, 8 and 10 aa long inserts, respectively, when compared to the reference strain Wa, while DS1-like US G2P[4] strains did not display variation in this region. Concerning the segment lengths, only the NSP5/NSP6 encoding genome segments were remarkably different between Wa-like and DS1-like strains, due to a known 143 bp long insertion at the 3' end, downstream of the coding region [Matsui et al., 1990].

Phylogenetic analysis (Fig. 1) of the nt sequences of all 11 genes of 3 G1P[8] strains revealed a conserved genomic constellation (G1-P[8]-I1-R1-C1-M1-A1-N1-T1-E1-H1) with, but moderate sequence similarity to strains, Wa, KU, Dhaka 16-03, and D, the few G1P[8] strains whose genome sequences have already been determined [Matthijnssens, et al., 2008b]. Evidence for reassortment



1  
2  
3 among these 3 Californian strains was revealed by demonstrating an inverse  
4 phylogenetic relationship among the 3 strains in various genes. Thus, in the  
5 NSP1, NSP2, NSP3, NSP5, and VP3 gene phylogenies the strains LB2758 and  
6 LB2771 clustered together while the strain LB2719 was more divergent from  
7 these strains, whereas in the NSP4, VP1, VP2, VP4, and VP6 gene phylogenies  
8 the strains LB2719 and LB2771 clustered together and the strain LB2758 was  
9 more divergent from these 2 strains. Of interest, the VP7 gene of all 3 G1P[8]  
10 strains belonged to the same genetic cluster, designated lineage I (nt similarity,  
11 >96.4%; Table 2) .

12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
Phylogenetic analysis (Fig. 1) of the 3 G2P[4] strains identified a conserved  
genomic constellation (G2-P[4]-I2-R2-C2-M2-A2-N2-T2-E2-H2) with reference  
G2P[4] strains, DS1 and TB-Chen, and also displayed a lesser extent of genetic  
divergence to these strains and within the 3 selected strains in most gene  
phylogenies. In these comparisons, strain LB2744 and LB2772 always clustered  
together, although no absolute sequence identity was seen among them (range  
of sequence similarity, 98.7% for VP1 to 99.8% for NSP2 and NSP5), while strain  
LB2764 was slightly divergent from these strains (range of sequence similarity,  
95.3% for VP7 to 99.6% for NSP2). Despite the moderate sequence identity  
within the VP7 gene, all 3 Californian strains belonged to the same genetic  
cluster, designated lineage II, with up to 99.6% sequence identity to reference  
strains (Table 3). In the full genomic constellation, only NSP2 could have  
represented a shared gene among the 3 G2P[4] strains where the mean genetic  
similarity (99.6%, range, 99.4%-99.8%) was identical to the genetic similarity  
seen for the G1P[8] strains, LB2758 and LB2771, for which the common origin of  
NSP2 was anticipated. Thus, it is possible, that the strain LB2764 acquired its  
NSP2 gene through reassortment with a co-circulating LB2744-like strain.

Altogether, intragenogroup reassortment events within 3 co-circulating G1P[8]  
strains and likely among G2P[4] strains were identified, but no intergenogroup  
reassortment events were detected by the whole genome sequencing approach.  
The overall low detection rate of mixed infections (~5%) with involvement of  
strains belonging to different genogroups in hospitalized Californian children

1  
2  
3 might provide one explanation for this failure. It is planned that in the future  
4 additional strains from the CDC's strain collection will be analyzed in similar  
5 studies to better understand if intergenogroup reassortment of the non serotype  
6 genes is key in evolution and epidemiologic success of various strains detected  
7 in either US children treated in hospital due to rotavirus gastroenteritis or in other  
8 countries and settings, where mixed infection are more abundant.  
9  
10  
11  
12  
13

## 14 15 16 **Discussion**

17 Although these six strains represent the most important antigen combinations  
18 in the US and in many regions of the world [Gentsch et al., 2005, 2009; Santos  
19 and Hoshino, 2005; Matthijnssens et al., 2009a], their representation in terms of  
20 whole genome sequence data is limited [Matthijnssens et al., 2008a, 2008b].  
21 Thus, in addition to the primary objective to identify intergenogroup reassortment  
22 events among co-circulating strains, the present study was also conducted to  
23 provide additional reference genome sequences of G1P[8] and G2P[4]  
24 rotaviruses for future genomics and epidemiologic studies conducted on these  
25 medically important strains. As nucleotide sequencing becomes routine, it seems  
26 likely that surveillance studies in the post rotavirus vaccine era will utilize whole  
27 genome sequencing as part of the standard strain characterization algorithm.  
28 Such studies are needed to determine whether intergenogroup reassortment  
29 affecting the backbone genes could play a significant role in any potential  
30 vaccine breakthrough events by evading immunity of vaccinated children.  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45

46 **Acknowledgements** At the time the study was launched, K.B. was an  
47 international fellow of the Emerging Infectious Diseases Fellowship Program  
48 awarded by the Association of Public Health Laboratories.  
49  
50  
51

## 52 53 **References**

54 Ahmed K, Nakagomi T, Nakagomi O. 2005. Isolation and molecular  
55 characterization of a naturally occurring non-structural protein 5 (NSP5)  
56  
57  
58  
59  
60

- gene reassortant of group A rotavirus of serotype G2P[4] with a long RNA pattern. *J Med Virol* 77:323–330.
- Bányai K, Gentsch JR, Martella V, Bogdán Á, Havasi V, Kisfali P, Szabó A, Mihály I, Molnár P, Melegh B, Szücs G. 2009a. Trends in the epidemiology of human G1P[8] rotaviruses: a hungarian study. *J Infect Dis* 200 Suppl 1:S222–S227.
- Bányai K, Esona MD, Mijatovic S, Kerin TK, Pedreira C, Mercado J, Balmaseda A, Perez MC, Patel MM, Gentsch JR. 2009b. Zoonotic bovine rotavirus strain in a diarrheic child, Nicaragua. *J Clin Virol* 46:391–393.
- Bányai K, Esona MD, Kerin TK, Hull JJ, Mijatovic S, Váscónez N, Torres C, de Filippis AM, Foytich KR, Gentsch JR. 2009c. Molecular characterization of a rare, human-porcine reassortant rotavirus strain, G11P[6], from Ecuador. *Arch Virol* 154:1823–1829.
- Bányai K, Martella V, Molnár P, Mihály I, Van Ranst M, Matthijnssens J. 2009d. Genetic heterogeneity in human G6P[14] rotavirus strains detected in Hungary suggests independent zoonotic origin. *J Infect* 59:213–215.
- De Grazia S, Giammanco GM, Potgieter CA, Matthijnssens J, Banyai K, Platia MA, Colomba C, Martella V. 2010. Unusual assortment of segments in 2 rare human rotavirus genomes. *Emerg Infect Dis* 16:859–862.
- Esona MD, Geyer A, Banyai K, Page N, Aminu M, Armah GE, Hull J, Steele DA, Glass RI, Gentsch JR. 2009a. Novel human rotavirus genotype G5P[7] from child with diarrhea, Cameroon. *Emerg Infect Dis* 15:83–86.
- Esona MD, Geyer A, Page N, Trabelsi A, Fodha I, Aminu M, Agbaya VA, Tsion B, Kerin TK, Armah GE, Steele AD, Glass RI, Gentsch JR. 2009b. Genomic characterization of human rotavirus G8 strains from the African rotavirus network: relationship to animal rotaviruses. *J Med Virol* 81: 937–951.
- Estes MK, Kapikian AZ. 2007. Rotaviruses. In: Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA, Roizman B, Straus SE, editors. *Fields Virology*. Philadelphia: Lippincott Williams & Wilkins/Wolters Kluwer, pp. 1917–1974.

- 1  
2  
3 Gentsch JR, Laird AR, Bielfelt B, Griffin DD, Banyai K, Ramachandran M, Jain V,  
4 Cunliffe NA, Nakagomi O, Kirkwood CD, Fischer TK, Parashar UD,  
5 Bresee JS, Jiang B, Glass RI. 2005. Serotype diversity and reassortment  
6 between human and animal rotavirus strains: implications for rotavirus  
7 vaccine programs. *J Infect Dis* 192 Suppl 1:S146–S159.  
8  
9  
10  
11 Gentsch JR, Hull JJ, Teel EN, Kerin TK, Freeman MM, Esona MD, Griffin DD,  
12 Bielfelt-Krall BP, Banyai K, Jiang B, Cortese MM, Glass RI, Parashar UD,  
13 collaborating laboratories of the National Rotavirus Strain Surveillance  
14 System. 2009. G and P types of circulating rotavirus strains in the United  
15 States during 1996-2005: nine years of prevaccine data. *J Infect Dis* 200  
16 Suppl 1:S99–S105.  
17  
18  
19  
20  
21  
22  
23 Griffin DD, Kirkwood CD, Parashar UD, Woods PA, Bresee JS, Glass RI,  
24 Gentsch JR. 2000. Surveillance of rotavirus strains in the United States:  
25 identification of unusual strains. The National Rotavirus Strain  
26 Surveillance System collaborating laboratories. *J Clin Microbiol* 38:2784–  
27 2787.  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
- Jakab F, Meleg E, Bányai K, Meleg B, Tímár L, Péterfai J, Szucs G. 2004. One-year survey of astrovirus infection in children with gastroenteritis in a large hospital in Hungary: occurrence and genetic analysis of astroviruses. *J Med Virol* 74:71–77.
- Kaga E, Nakagomi O. 1994. Recurrent circulation of single nonstructural gene substitution reassortants among human rotaviruses with a short RNA pattern. *Arch Virol* 136:63-71.
- Martella V, Bányai K, Matthijnssens J, Buonavoglia C, Ciarlet M. 2010. Zoonotic aspects of rotaviruses. *Vet Microbiol* 140:246–255.
- Matsui SM, Mackow ER, Matsuno S, Paul PS, Greenberg HB. 1990. Sequence analysis of gene 11 equivalents from "short" and "super short" strains of rotavirus. *J Virol* 64:120-124.
- Matthijnssens J, Rahman M, Yang X, Delbeke T, Arijs I, Kabue JP, Muyembe JJ, Van Ranst M. 2006a. G8 rotavirus strains isolated in the Democratic Republic of Congo belong to the DS-1-like genogroup. *J Clin Microbiol*

- 1  
2  
3 44:1801–1809.  
4  
5 Matthijnssens J, Rahman M, Martella V, Xuelei Y, De Vos S, De Leener K,  
6 Ciarlet M, Buonavoglia C, Van Ranst M. 2006b. Full genomic analysis of  
7 human rotavirus strain B4106 and lapine rotavirus strain 30/96 provides  
8 evidence for interspecies transmission. *J Virol* 80:3801–3810.  
9  
10 Matthijnssens J, Ciarlet M, Heiman E, Arijs I, Delbeke T, McDonald SM, Palombo  
11 EA, Iturriza-Gomara M, Maes P, Patton JT, Rahman M, Van Ranst M.  
12 2008a. Full genome-based classification of rotaviruses reveals a common  
13 origin between human Wa-like and porcine rotavirus strains and human  
14 DS-1-like and bovine rotavirus strains. *J Virol* 82:3204–3219.  
15  
16 Matthijnssens J, Ciarlet M, Rahman M, Attoui H, Banyai K, Estes MK, Gentsch  
17 JR, Iturriza-Gomara M, Kirkwood CD, Martella V, Mertens PP, Nakagomi  
18 O, Patton JT, Ruggeri FM, Saif LJ, Santos N, Steyer A, Taniguchi K,  
19 Desselberger U, Van Ranst M. 2008b. Recommendations for the  
20 classification of group A rotaviruses using all 11 genomic RNA segments.  
21 *Arch Virol* 153:1621–1629.  
22  
23 Matthijnssens J, Rahman M, Van Ranst M. 2008c. Two out of the 11 genes of an  
24 unusual human G6P[6] rotavirus isolate are of bovine origin. *J Gen Virol*  
25 89:2630–2635.  
26  
27 Matthijnssens J, Bilcke J, Ciarlet M, Martella V, Bányai K, Rahman M, Zeller M,  
28 Beutels P, Van Damme P, Van Ranst M. 2009a. Rotavirus disease and  
29 vaccination: impact on genotype diversity. *Future Microbiol.* 4:1303–1316.  
30  
31 Matthijnssens J, Potgieter CA, Ciarlet M, Parreño V, Martella V, Bányai K,  
32 Garaicoechea L, Palombo EA, Novo L, Zeller M, Arista S, Gerna G,  
33 Rahman M, Van Ranst M. 2009b. Are human P[14] rotavirus strains the  
34 result of interspecies transmissions from sheep or other ungulates that  
35 belong to the mammalian order Artiodactyla? *J Virol* 83:2917–2929.  
36  
37 Matthijnssens J, Rahman M, Ciarlet M, Zeller M, Heylen E, Nakagomi T, Uchida  
38 R, Hassan Z, Azim T, Nakagomi O, Van Ranst M. 2010. Reassortment of  
39 human rotavirus gene segments into G11 rotavirus strains. *Emerg Infect*  
40 *Dis* 16:625–630.  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 McDonald SM, Matthijnssens J, McAllen JK, Hine E, Overton L, Wang S, Lemey  
4 P, Zeller M, Van Ranst M, Spiro DJ, Patton JT. 2009. Evolutionary  
5 dynamics of human rotaviruses: balancing reassortment with preferred  
6 genome constellations. *PLoS Pathog* 5:e1000634.  
7  
8  
9  
10 Medici MC, Abelli LA, Martella V, Martinelli M, Lorusso E, Buonavoglia C, Dettori  
11 G, Chezzi C. 2007. Characterization of inter-genogroup reassortant  
12 rotavirus strains detected in hospitalized children in Italy. *J Med Virol*  
13 79:1406–1412.  
14  
15  
16  
17 Mukherjee A, Dutta D, Ghosh S, Bagchi P, Chattopadhyay S, Nagashima S,  
18 Kobayashi N, Dutta P, Krishnan T, Naik TN, Chawla-Sarkar M. 2009. Full  
19 genomic analysis of a human group A rotavirus G9P[6] strain from Eastern  
20 India provides evidence for porcine-to-human interspecies transmission.  
21 *Arch Virol* 154:733–746.  
22  
23  
24  
25  
26 Nakagomi O, Nakagomi T, Akatani K, Ikegami N, 1989. Identification of rotavirus  
27 genogroups by RNA-RNA hybridization. *Mol Cell Probes* 3:251–261.  
28  
29  
30 Nakagomi O, Nakagomi T. 1991. Molecular evidence for naturally occurring  
31 single VP7 gene substitution reassortant between human rotaviruses  
32 belonging to two different genogroups. *Arch Virol* 119:67–81.  
33  
34  
35 Nakagomi O, Nakagomi T. 2002. Genomic relationships among rotaviruses  
36 recovered from various animal species as revealed by RNA-RNA  
37 hybridization assays. *Res Vet Sci* 73:207–214.  
38  
39  
40 Nakagomi T, Gentsch JR, Das BK, Kumar R, Bhan MK, Glass RI, Nakagomi O.  
41 2002. Molecular characterization of serotype G2 and G3 human rotavirus  
42 strains that have an apparently identical electropherotype of the short  
43 RNA pattern. *Arch Virol* 147:2187–2195.  
44  
45  
46  
47 Palombo EA, Bugg HC, Masendycz PJ, Coulson BS, Barnes GL, Bishop RF.  
48 1996. Multiple-gene rotavirus reassortants responsible for an outbreak of  
49 gastroenteritis in central and northern Australia. *J Gen Virol* 77:1223–  
50 1227.  
51  
52  
53  
54  
55 Phan TG, Khamrin P, Quang TD, Dey Sk, Takanashi S, Okitsu S, Maneekarn N,  
56 Ushijima H. 2007. Detection and genetic characterization of group A  
57  
58  
59  
60



- 1  
2  
3 rotavirus strains circulating among children with acute gastroenteritis in  
4 Japan. *J Virol* 81:4645-4653.  
5  
6  
7 Pietsch C, Petersen L, Patzer L, Liebert UG. 2009. Molecular characteristics of  
8 German G8P[4] rotavirus strain GER1H-09 suggest that a genotyping and  
9 subclassification update is required for G8. *J Clin Microbiol* 47:3569–3576.  
10  
11  
12 Rahman M, Matthijnsens J, Yang X, Delbeke T, Arijs I, Taniguchi K, Iturriza-  
13 Gómara M, Iftekharuddin N, Azim T, Van Ranst M. 2006. Evolutionary  
14 history and global spread of the emerging G12 human rotaviruses. *J Virol*  
15 81:2382–2390.  
16  
17  
18 Ramachandran M, Gentsch JR, Parashar UD, Jin S, Woods PA, Holmes JL,  
19 Kirkwood CD, Bishop RF, Greenberg HB, Urasawa S, Gerna G, Coulson  
20 BS, Taniguchi K, Bresee JS, Glass RI. 1998. Detection and  
21 characterization of novel rotavirus strains in the United States. *J Clin*  
22 *Microbiol* 36:3223–3229.  
23  
24  
25 Santos N, Hoshino Y. 2005. Global distribution of rotavirus serotypes/genotypes  
26 and its implication for the development and implementation of an effective  
27 rotavirus vaccine. *Rev Med Virol* 15:29–56.  
28  
29  
30 Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular evolutionary  
31 genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–  
32 1599.  
33  
34  
35 Trinh QD, Pham NTK, Nguyen TA, Phan TG, Yan H, Hoang LP, Khamrin P,  
36 Maneekarn N, Li Y, Okitsu S, Mizuguchi M, Ushijima H. 2010. Sequence  
37 analysis of the VP7 gene of human rotaviruses G2 and G4 isolated in  
38 Japan, China, Thailand, and Vietnam during 2001–2003. *J Med Virol*  
39 82:878–885.  
40  
41  
42 Tsugawa T, Hoshino Y. 2008. Whole genome sequence and phylogenetic  
43 analyses reveal human rotavirus G3P[3] strains Ro1845 and HCR3A are  
44 examples of direct virion transmission of canine/feline rotaviruses to  
45 humans. *Virology* 380:344–353.  
46  
47  
48 Wang Y-H, Kobayashi N, Nagashima S, Zhou X, Ghosh S, Peng J-S, Hu Q,  
49 Zhou D-J, Yang Z-Q. 2010. Full genomic analysis of a porcine-bovine  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 reassortant G4P[6] rotavirus strain R479 isolated from an infant in China.  
4 J Med Virol 82:1094–1102.  
5

6  
7 Ward RL, Nakagomi O, Knowlton DR, McNeal MM, Nakagomi T, Clemens JD,  
8 Sack DA, Schiff GM. 1990. Evidence for natural reassortants of human  
9 rotaviruses belonging to different genogroups. J Virol 64:3219–3225.  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review



Table 1. Overview of sequence completeness of 6 selected US strains.

| Strain | Length of the sequenced coding region based on the deduced amino acid sequences (%) in comparison to strain Wa * |                   |                |                |                |                |                 |                 |                 |                 |                     |
|--------|--|-------------------|----------------|----------------|----------------|----------------|-----------------|-----------------|-----------------|-----------------|---------------------|
|        | VP1,<br>1088 aa  | VP2,<br>890 aa ** | VP3,<br>835 aa | VP4,<br>775 aa | VP6,<br>397 aa | VP7,<br>327 aa | NSP1,<br>486 aa | NSP2,<br>317 aa | NSP3,<br>310 aa | NSP4,<br>175 aa | NSP5,<br>197 aa *** |
| LB2719 | 99.9   | 99.8              | 100            | 99.5           | 100            | 100            | 100             | 100             | 100             | 100             | 100                 |
| LB2758 | 99.9   | 99.8              | 100            | 99.5           | 100            | 100            | 100             | 100             | 100             | 100             | 100                 |
| LB2771 | 99.9   | 99.8              | 100            | 99.5           | 100            | 100            | 99.9            | 100             | 100             | 100             | 100                 |
| LB2744 | 99.9   | 99.8              | 100            | 99.5           | 100            | 100            | 100             | 100             | 100             | 100             | 100                 |
| LB2764 | 99.9   | 99.8              | 100            | 99.5           | 100            | 100            | 100             | 100             | 100             | 100             | 100                 |
| LB2772 | 99.9   | 99.8              | 100            | 99.5           | 100            | 100            | 100             | 100             | 100             | 100             | 100                 |

\* Missing data: VP1, missing 1 aa at N-terminal in all 6 US strains; VP2, missing 2 aa at N terminal in all 6 US strains; VP4, missing 4 aa at N terminal in all 6 US strains; NSP1, missing 1 aa at N terminal in strain LB2771

\*\* Predicted length of the deduced amino acid sequences of the VP2 of contemporary US G1P[8] and G2P[4] strains differed from the length of VP2 of Wa: LB2719, 894 aa; LB2758, 898 aa; LB2771, 900; all 3 DS1-like US strains, 877 aa.

\*\*\* 200 aa in the NSP5 of DS1-like strains

\*\* and \*\*\* Strain specific insertions in the VP2 gene of Wa-like strains and genogroup specific insertions in the NSP5 gene of DS1-like strains are not considered in the calculations of the percentile values.

Table 2. VP7 gene nucleotide identities of 3 Californian G1P[8] strains to 69 reference G1 strains.

| G1 lineage * | No. of reference strain compared | Range of nt similarities (%) to reference strains |           |           |
|--------------|----------------------------------|---|-----------|-----------|
|              |                                  | LB2719  | LB2758    | LB2771    |
| I            | 17                               | 96.4-99.4   | 96.9-99.3 | 96.7-99.3 |
| II           | 24                               | 92.4-94.3   | 93.2-94.6 | 93.2-94.6 |
| III          | 11                               | 90.1-91.7   | 91.0-92.2 | 91.2-92.5 |
| IV           | 2                                | 91.1-91.5   | 91.7-92.1 | 91.9-92.4 |
| V            | 1                                | 93.5  | 94.6      | 94.6      |
| VI           | 1                                | 84.9  | 85.3      | 85.6      |
| VII          | 1                                | 94.3  | 94.3      | 94.3      |
| VIII         | 1                                | 94.0  | 94.3      | 94.0      |
| IX           | 7                                | 94.7-95.4   | 95.0-95.7 | 94.4-95.4 |
| X            | 1                                | 85.6  | 85.4      | 85.7      |
| XI           | 3                                | 85.0-85.7   | 84.7-85.4 | 85.3-86.0 |

\* According to classification scheme proposed by Phan et al. [2007]

Note: A 720 bp long fragment was utilized for comparison due to sequence length differences of various reference strains.

Table 3. VP7 gene nucleotide identities of 3 Californian G2P[4] strains to 59 reference G2 strains

| G2 lineage * | No. of reference strain compared | Range of nt similarities (%) to reference strains |           |           |
|--------------|----------------------------------|---|-----------|-----------|
|              |                                  | LB2744  | LB2764    | LB2772    |
| I            | 9                                | 92.6-93.9   | 92.2-93.3 | 92.4-93.9 |
| II           | 45                               | 95.1-99.4   | 94.7-97.9 | 95.1-99.6 |
| III          | 3                                | 91.6-92.4   | 90.8-91.5 | 91.4-92.2 |
| IV           | 2                                | 82.1-82.4   | 81.7-81.9 | 82.2-82.3 |

\* According to classification scheme proposed by Trinh et al. [2010]

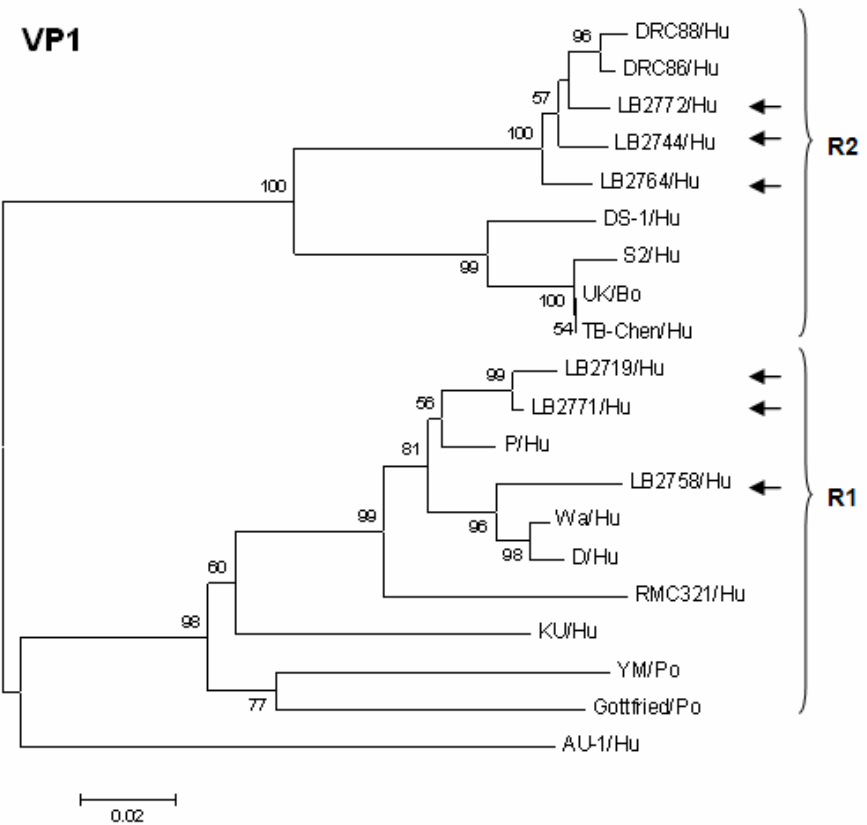
Note: A 994 bp long fragment was utilized for comparison due to sequence length differences of various reference strains.

**Figure legend.**

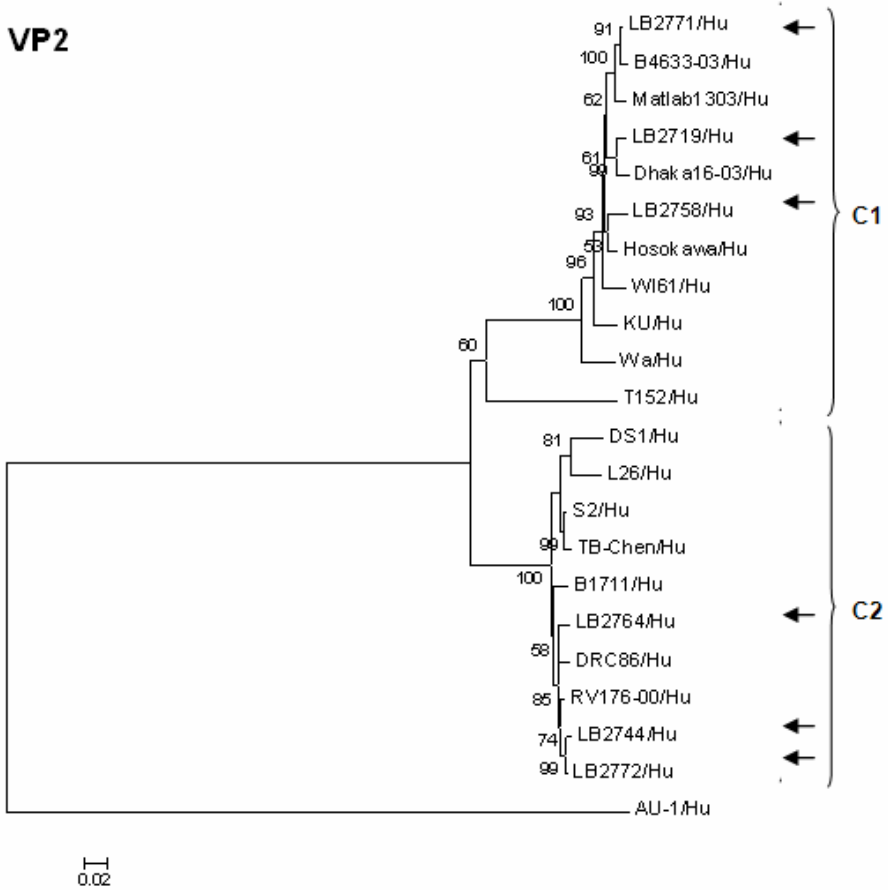
Fig. 1. Phylogenetic analysis of nucleotide sequences of 3 G1P[8] and 3 G2P[4] Long Beach strains (indicated by arrowheads) along with sequences obtained from GenBank. The trees were constructed using neighbor-joining method of the MEGA4 package with bootstrap of 1000 replicates and rooted with sequences of human rotavirus G3P[9] strain Au-1. Bootstrap values ( $\geq 50\%$ ) are indicated at the branch nodes. Abbreviations used are Po (Porcine), Bo (Bovine), Hu (Human) and Eq (Equine).

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

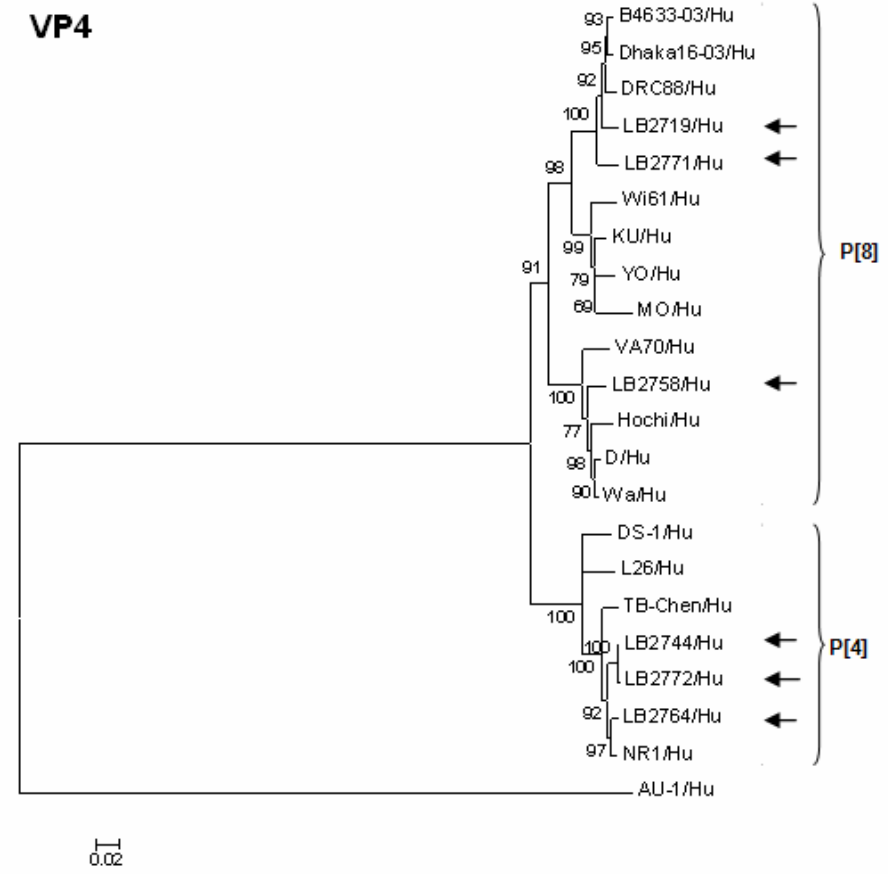
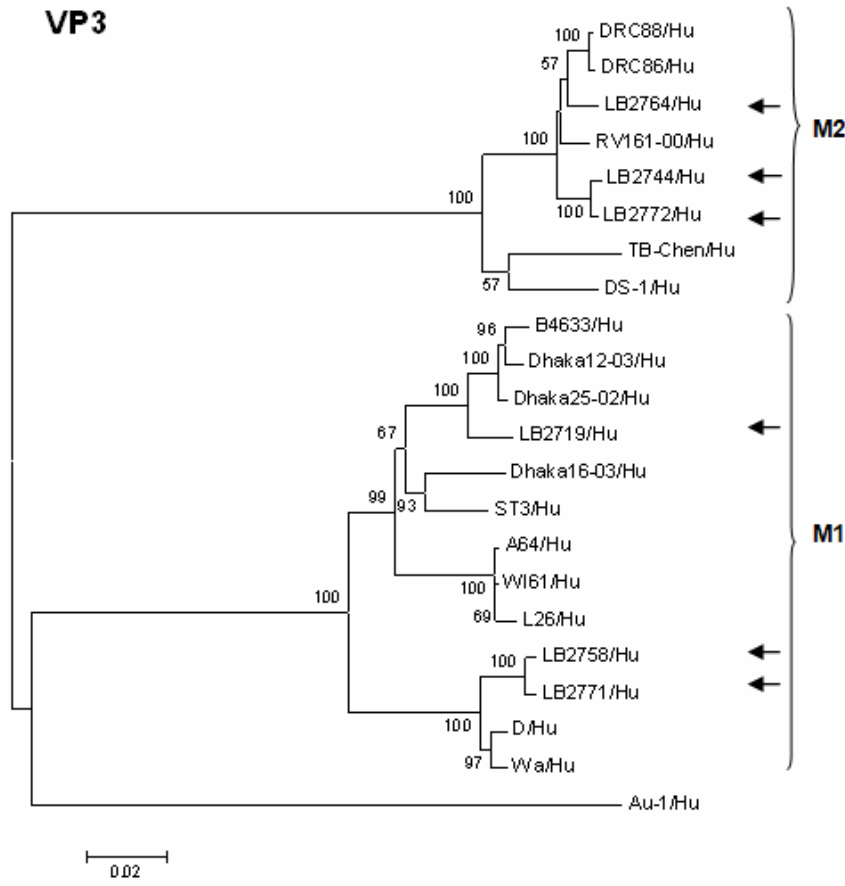
**VP1**



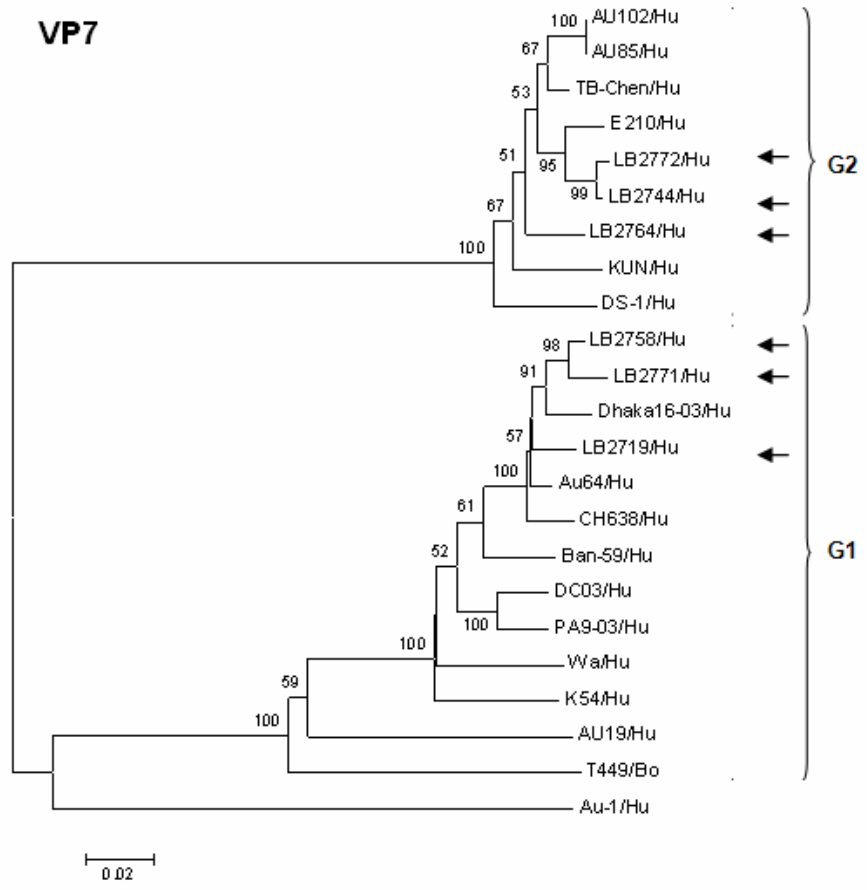
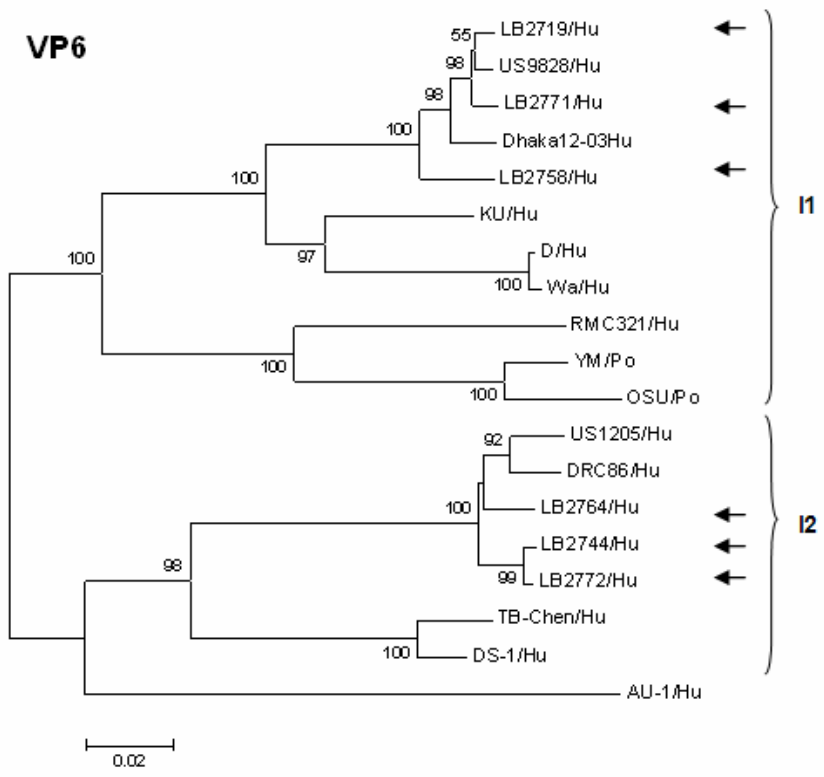
**VP2**

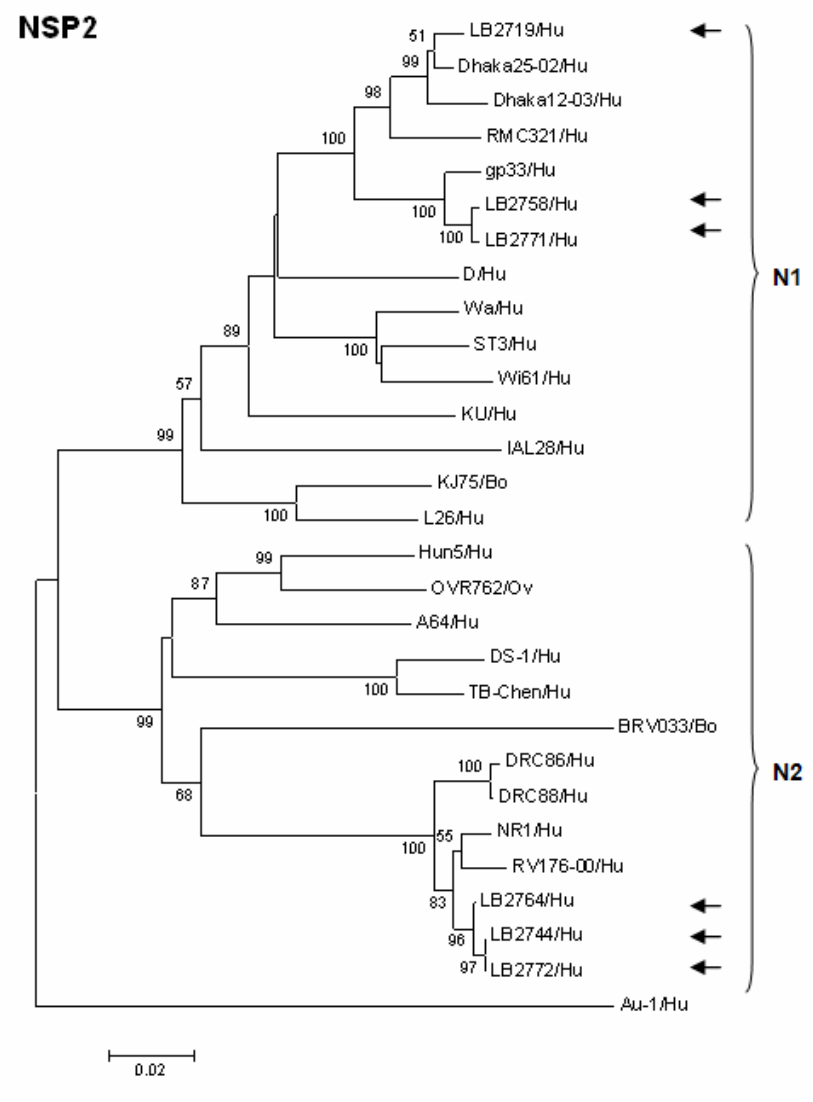
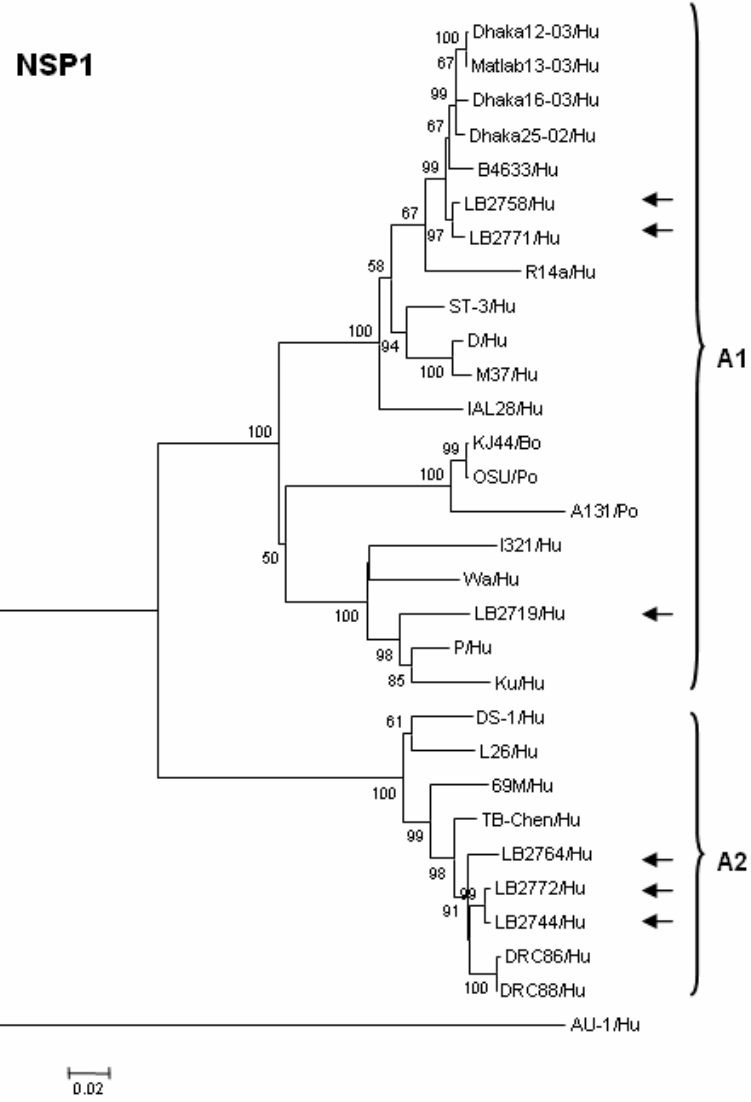


1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49



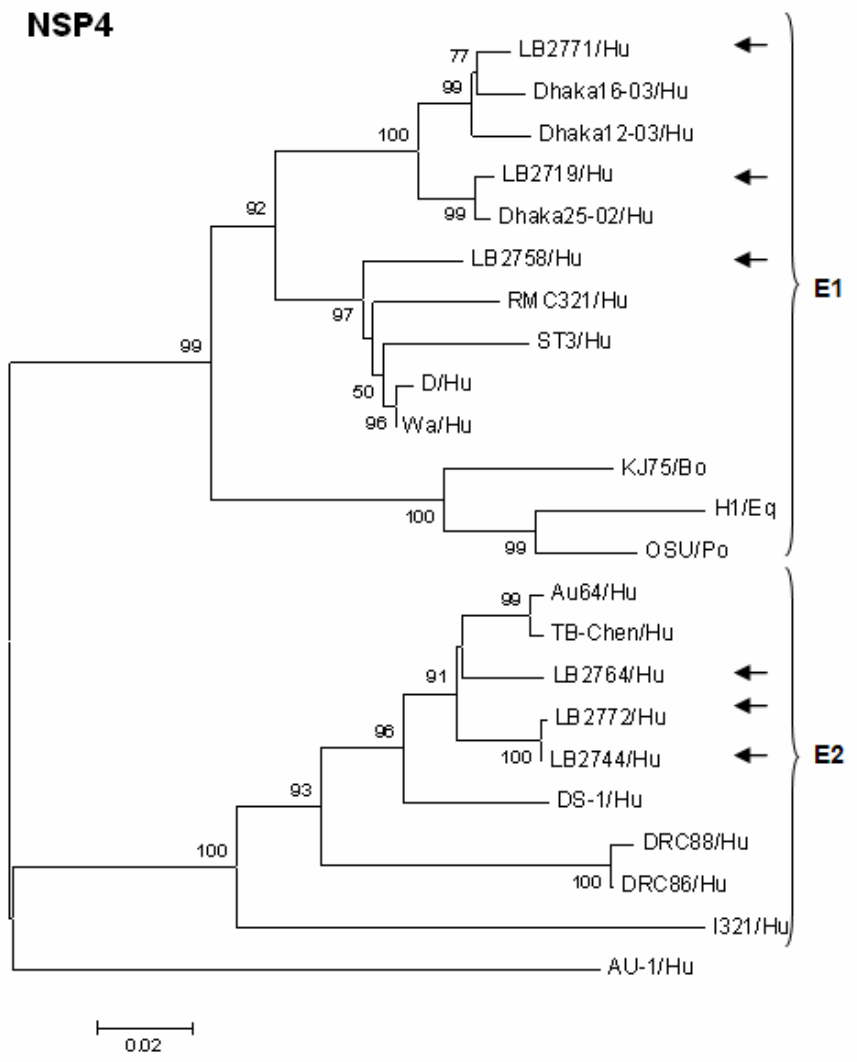
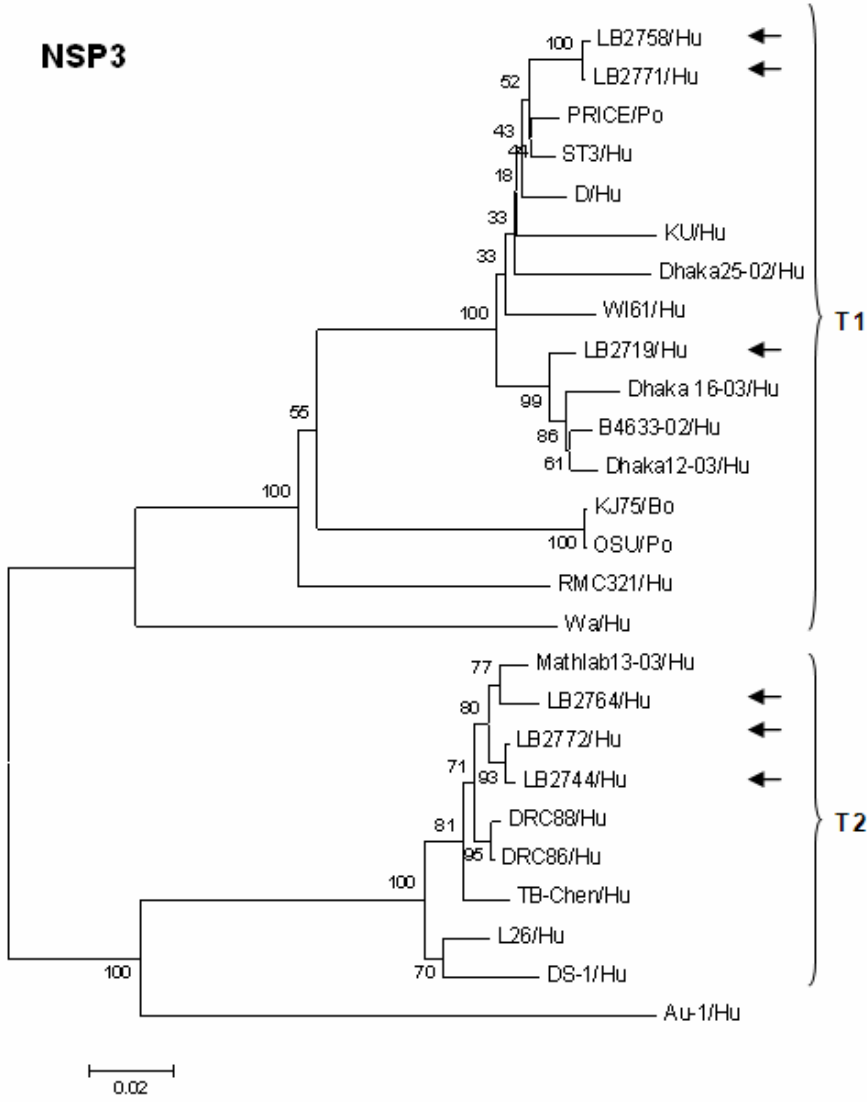
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

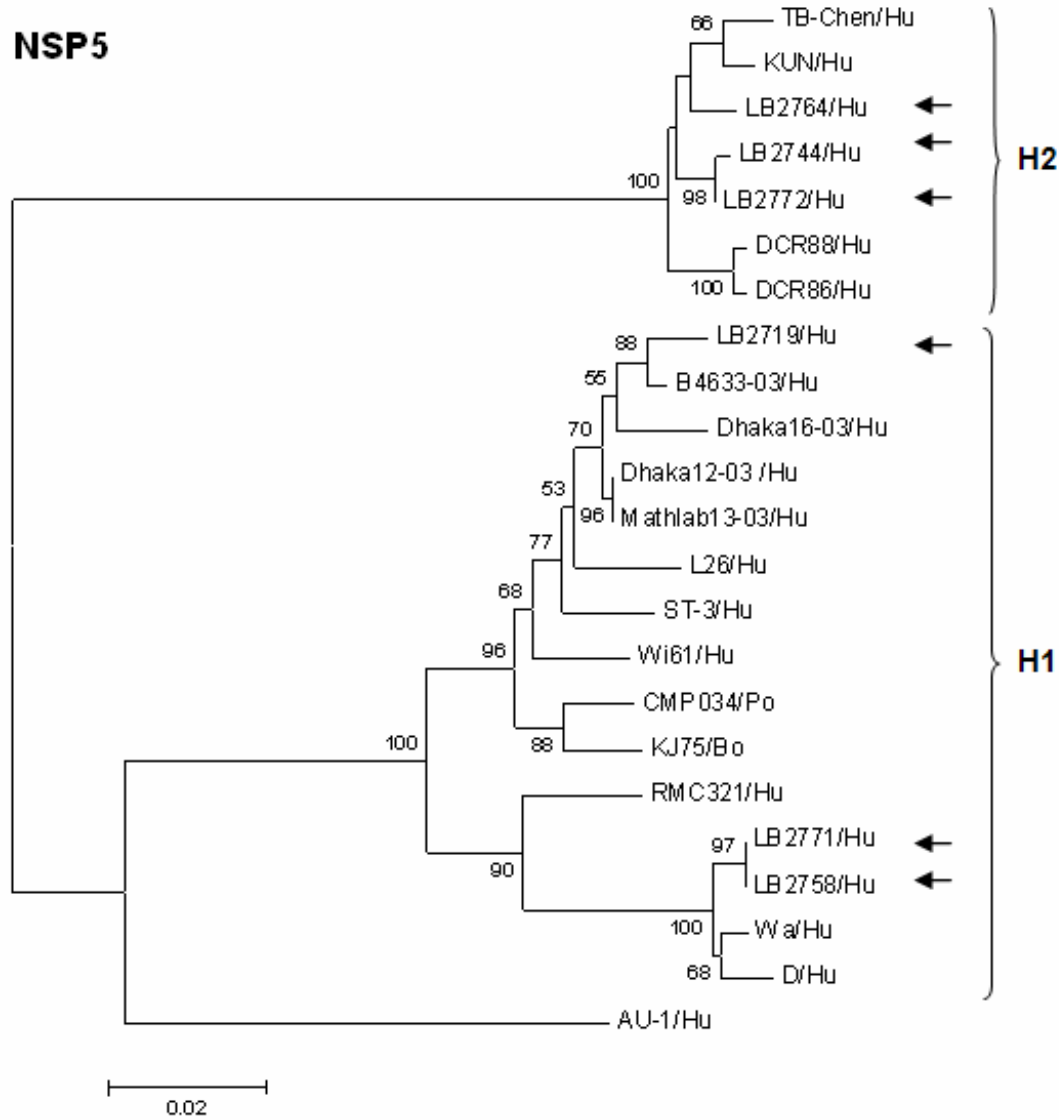






1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49





Me  
e  
e