Virology **299,** 225–239 (2002)

Phylogenetic Analysis of the Complete Genome of 18 Norwalk-like Viruses

Kazuhiko Katayama,*[,1] Haruko Shirato-Horikoshi,* Shigeyuki Kojima,*[,†] Tsutomu Kageyama,† Tomoichiro Oka,† Fuminori B. Hoshino,† Shuetsu Fukushi,† Michiyo Shinohara,‡ Kazue Uchida,‡ Yoshiyuki Suzuki,§ Takashi Gojobori,§ and Naokazu Takeda*

**Department of Virology II, National Institute of Infectious Diseases, 4-7-1 Gakuen, Musashi-murayama, Tokyo, 208-0011, Japan; †Section of Infectious Disease, R&D Center, BML, Matoba 1361-1, Kawagoe, Saitama 350-1101, Japan; ‡Saitama Institute of Public Health, Kamiokubo 639-1, Urawa, Saitama 338-0824, Japan; and §Center for Information Biology, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan*

Received November 28, 2001; returned to author for revision January 10, 2002; accepted May 20, 2002

"Norwalk-like viruses" (NLV), a member of the family *Caliciviridae*, are the major causative agents of acute gastroenteritis and are genetically divided into two groups, genogroup I (GI) and genogroup II (GII). We have determined the complete nucleotide sequences of 10 new NLV strains. Using this information together with eight known NLV sequences, the criteria to further classify genotypes of NLV were investigated. Validation of the topological error based on the bootstrap value and the branch length (distance) allowed us to identify two potential subgenomic regions suitable for the genotyping. They were the putative 3D-like RNA-dependent RNA polymerase (polymerase) and the capsid N-terminal/Shell domains (capsid N/S domain). When the distance distribution analysis was performed, the polymerase-based classification did not separate the strains into internal clusters within the genogroup. Furthermore, a diversity plot analysis of the complete nucleotide sequences of WUG1, a NLV GI strain, and Saitama U1, a NLV GII strain, indicated that the genotype was different between the polymerase and capsid N/S domain, suggesting that these strains are the genetic recombinants. Therefore, polymerase is not suitable for genotyping. On the other hand, the clustering based on the capsid N/S domain successfully distinguished the NLV as well as the grouping based on the antigenicity, as determined by both antigen and antibody ELISAs with recombinant virus-like particles. As the nucleotide sequences of the primers for the capsid N/S domain are highly conserved among the NLV, the amplification of the unknown genotype can be easily performed. This method will facilitate global surveying as well as epidemiologic study on NLV.

Key Words: Norwalk-like viruses; complete genome sequence; phylogenetic analysis; genotyping; recombination.

INTRODUCTION

"Norwalk-like viruses (NLV)" in the family *Caliciviridae* are important causative agents of acute nonbacterial gastroenteritis in humans. Molecular cloning and the complete nucleotide sequence analyses of Norwalk virus and Southampton virus revealed that these viruses contain a positive-sense single-stranded RNA genome of ~7.7 kb with a poly(A) tail of ~20 nucleotides (Jiang *et al.,* 1990, 1993; Lambden *et al.,* 1993). Three open reading frames (ORFs) were identified in the genome. The ORF1 encodes a large polyprotein that is cleaved into several viral proteins, which contain amino acid sequence motifs conserved in NTPase, 3C-like protease, and a 3D-like RNA-dependent RNA polymerase (polymerase), as found in other ssRNA viruses (Lee *et al.,* 1977; Rueckert and Wimmer, 1984). The ORF2 encodes a capsid protein with an apparent molecular weight of 58,000. The ORF3 encodes a small protein rich in basic amino acid. Although this protein's role in the virus replication is unknown, the protein is probably a minor structural protein that interacts with the RNA genome during the virion formation (Glass *et al.,* 2000). As NLVs are uncultivable gastroenteritis viruses, a conventional neutralization test cannot be used for the serotyping. However, several antigenically distinct viruses have been identified by immune electron microscopy, solid-phase immune electron microscopy, and cross-challenge volunteers studies (Lewis, 1990; Lewis *et al.,* 1988; Madore *et al.,* 1990).

Amplification of a part of the genome that encodes the polymerase, capsid protein, and ORF3 protein followed by the sequencing analyses revealed that the strains of NLV are genetically divided into two distinct groups, genogroup I (GI) and genogroup II (GII) (Green *et al.,* 1994; Wang *et al.,* 1994). Each genogroup is further grouped into at least three genotypes in GI and four genotypes in GII. Although the partial nucleotide sequences were accumulated rapidly (Green *et al.,* 2000), most were from the polymerase and capsid regions, and the studies have been done independently. Therefore, the strains were often assigned to different clusters depending on whether the polymerase or capsid sequence was used in the analysis (Vinje *et al.,* 2000). This sug-

[1] To whom correspondence and reprint requests should be addressed at Department of Virus Diseases and Vaccine Control, National Institute of Infectious Diseases. 4-7-1 Gakuen, Musashi-murayama, Tokyo, 208-0011, Japan. Fax: +81-42-565-3315. E-mail: katayama@nih.go.jp.

0042-6822/02 $35.00

TABLE 1

Percent Nucleotide Identities of the Full Length Genome among NLV and SLV

	Reported GI strains				Reported GII strains				SLV strain
	NV68	SO	BS5	Chiba	LD	CW	Hawaii	MD145	Manchester (MC)
Norwalk (NV68)	[a]	**69.1**	**71.1**	**71.4**	51.2	52.0	55.8	55.0	44.5
Southampton (SO)		[a]	**71.2**	**70.6**	51.6	52.0	55.8	55.3	44.2
BS5			[a]	**72.9**	52.1	52.8	56.3	56.1	44.5
Chiba				[a]	55.2	55.3	55.9	55.0	45.1
Lordsdale (LD)					[a]	**93.2**	**78.9**	**94.4**	43.5
Camberwell (CW)						[a]	**78.9**	**96.7**	43.5
Hawaii							[a]	**79.3**	44.4
MD145-12 (MD145)								[a]	45.4
SzUG1	**69.8**	**70.8**	**73.9**	**75.6**	53.1	53.5	56.2	55.3	44.6
WUG1	**72.4**	**79.8**	**74.1**	**69.9**	56.0	55.6	56.1	55.6	45.6
U1	52.2	52.4	52.4	55.2	**80.7**	**80.6**	**86.4**	**82.5**	44.4
U3	52.2	52.1	52.6	55.4	**66.4**	**66.7**	**67.5**	**66.4**	44.9
U4	52.2	52.1	52.6	55.4	**66.4**	**66.7**	**67.5**	**66.4**	44.9
U16	51.5	52.5	51.6	55.5	**66.9**	**66.9**	**68.1**	**66.5**	43.9
U17	51.5	52.5	51.6	55.4	**66.9**	**66.9**	**68.1**	**66.7**	43.9
U18	51.7	52.2	51.1	55.2	**72.4**	**72.7**	**73.1**	**72.6**	43.7
U201	51.8	52.2	51.5	55.1	**72.5**	**72.9**	**73.1**	**72.6**	44.0
U25	52.9	52.3	52.5	55.0	**65.1**	**65.0**	**68.1**	**66.3**	44.4

Note. Bold letters indicate nucleotide identities of the intergroup.
[a] One hundred percent.

gests that genetic recombinations may have occurred. In fact, Jiang *et al.* recently reported a possible recombination within NLV (Jiang *et al.,* 1999). Therefore, it is unclear whether the genotype determined by one region correlates with that determined by an other. A genetic classification based on proper statistical analyses employing a suitable genome region is urgently needed.

We have determined the complete nucleotide sequences of 10 new NLV strains. Using this information together with eight known NLV sequences, phylogenetic trees were constructed, and the criteria to classify the genotypes of NLV were investigated. The topological errors were evaluated from the statistical significance of the bootstrap value, and the branch lengths (distance) were examined by histogram and distribution analyses. These analyses allowed us to identify two potential subgenomic regions for the genotyping. Furthermore, the suitability of these regions was evaluated by two phylogenetic methods. The phylogenetic analysis and the nucleotide identity window analysis showed evidence of genomic recombination.

RESULTS

Comparison of full-length genome sequences

We have determined the complete nucleotide sequences of 10 strains and compared them with those of 8 known strains (Table 1). When the nucleotide identity was examined with the entire genome, SzUG1 and WUG1 nucleotide sequences, respectively, showed 69.8–75.6 and 69.9–79.8% identities among four GI strains: Norwalk virus (NV68), Southampton virus (SO), BS5, and Chiba virus (Chiba), whereas they only showed 53.1–56.2 and 55.6–56.1% identities with GII strains, Lordsdale virus (LD), Camberwell virus (CW), Hawaii virus (HW), and MD145-12 virus (MD145). These findings indicate that SzUG1 and WUG1 are NLV GI. On the other hand, the remaining eight strains of nucleotide sequences showed 51.1–55.5% identities between the GI strains and 65.0–86.4% identities between the GII strains, indicating that these strains are NLV GII.

Saitama U3 (U3), Saitama U4 (U4), Saitama U16 (U16), and Saitama U17 (U17) were obtained from different individuals in the same outbreak. Two strains, U3 and U4, contained 12 different nucleotides (99.84% nucleotide identity). Similarly, U16 and U17 contained four different nucleotides (99.95% nucleotide identity). Sequences obtained from different outbreaks, e.g., Saitama U1 (U1) and U3, U4, and U16, and Saitama U25 (U25) and Saitama U201 (U201), showed 67.3–84.4% nucleotide identities to each other. However, in one combination, U201 and Saitama U18 (U18) showed 99.5% nucleotide identity (39 different nucleotides), although they were from two independent outbreaks.

Ten strains appeared to encode three ORFs as observed in eight known NLV (Table 2) (Dingle *et al.,* 1995; Jiang *et al.,* 1993; Lambden *et al.,* 1993; Someya *et al.,* 2000). ORF1 and ORF3 were encoded in frame 2, whereas ORF2 was encoded in frame 1. The C-terminal ORF1 overlapped with the N-terminal ORF2 by 14 nucle-

TABLE 2

Putative ORFs Location

Strain	Full length (nt[a])	ORF1 (nt)	ORF2 (nt)	ORF3 (nt)
Norwalk	7654	5–5375	5358–6950	6950–7588
Southampton	7708	5–5371	5355–6995	6995–7630
BS5	7598	5–5359	5343–6965	6965–7594
Chiba	7697	5–5362	5346–6980	5980–7606
SzUG1[b]	7700	5–5365	5352–6980	6983–7618
WUG1[b]	7688	5–5371	5355–6974	6974–7564
Lordsdale	7555	5–5104	5085–6704	6704–7510
Camberwell	7555	5–5104	5085–6704	6704–7510
Hawaii	7513	5–5104	5085–6692	6692–7471
MD145-12	7556	5–5104	5085–6704	6704–7510
Saitama U1[b]	7521	5–5101	5085–6689	6692–7468
Saitama U3[b]	7564	5–5095	5079–6728	6731–7507
Saitama U4[b]	7564	5–5095	5079–6728	6731–7507
Saitama U16[b]	7546	5–5095	5079–6719	6722–7495
Saitama U17[b]	7546	5–5095	5079–6719	6722–7495
Saitama U18[b]	7541	5–5110	5094–6737	6740–7501
Saitama U201[b]	7541	5–5110	5094–6737	6740–7501
Saitama U25[b]	7524	5–5101	5085–6695	6698–7468

[a] Nucleotide.
[b] Newly sequenced strain.

otides in NLV GI and by 17 nucleotides in NLV GII. The amino acid sequences of the predicted cleavage sites of the ORF1 proteins, N-terminal protein, NTPase, 3A-like protein, VPg, 3C-like protease, and polymerase were conserved in both NLV GI and NLV GII as described (Liu *et al.,* 1996; Seah *et al.,* 1999; Someya *et al.,* 2000).

To examine the nucleotide and amino acid sequence similarities, similarity plot analyses were performed. The plots of the nucleotide and putative amino acid sequences of NLV GI and NLV GII are depicted in Fig. 1. The highest nucleotide similarity score was found in the region from the C-terminal polymerase to the N-terminal capsid in both NLV GI and NLV GII, which correspond to the nucleotide position 5100–5380 in NV68 and 4980–5260 in LD. These regions included an 18 nucleotides consensus motif, 5′-GTR AAT GAW GAT GGC GTC-3′, which was previously suggested to be a packaging signal for the NLV genome or a transcription start signal (Lambden *et al.,* 1995). The nucleotide similarity score was above the average found in most parts of the ORF1 with the exceptions of the N-terminal and the 3A-like proteins. However, we could not find a region where a minimum 20 nucleotides were conserved. This is because many synonymous substitutions are scattered throughout ORF1. The lowest similarity score was observed in the middle portion of ORF2. This region corresponds to a highly variable P2 domain that protrudes outside of the NLV recombinant virus-like particles (VLPs), as examined by a three-dimensional structural analysis with VLPs (Prasad *et al.,* 1999).

The average amino acids similarity score, 4.6 for GI and 4.5 for GII, were highest in ORF1, followed by 4.1 for GI, and 3.9 for GII in ORF2. Those of ORF3, 3.8 for GI, and 3.1 for GII, were the lowest. The N- and C-terminal regions of the ORF2 and ORF3 proteins had higher similarity scores, and their middle region had lower similarity scores. Much lower scores were observed in ORF2 in both NLV GI (amino acid residue 330–400 in the NV68 genome) and NLV GII (amino acid residue 250–450 in the LD genome). These regions also correspond to the P2 domain of the capsid protein, forming the outer protrusions of the virion surface, possibly reflecting the antigenic diversity of NLV.

Phylogenetic analysis of the complete genome sequences

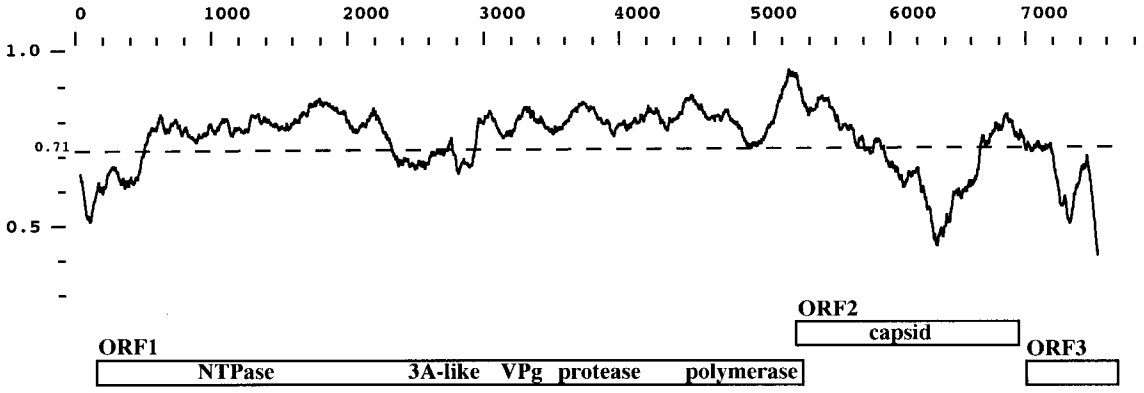
To identify the region in the NLV genome suitable for the genotyping, phylogenetic trees were constructed by the neighbor-joining method with the Manchester strain (MC) of "Sapporo-like viruses" (SLV) (Liu *et al.,* 1995) as an outgroup (Fig. 2, top). The bootstrap values of the operational taxonomic units (OTUs) of NLV GI and NLV GII, which represent the branches on the phylogenetic tree, were calculated, and the topological errors of the tree were evaluated. The bootstrap values of the OTUs were over 950 in the full-length genome and ORF1, and their topology was statistically valid (Efron *et al.,* 1996; Nei and Jin, 1989). Thus, the analysis employing the full-length genome and ORF1 were capable of segregating the strains into not only the proper genogroup but also the suitable genotype. On the other hand, some of the bootstrap values of the ORF2 and ORF3 were significantly low and the topology was not valid. However, determination of the full-length genome is time-consuming, and the nucleotide length is known to affect the bootstrap values of each ORF; therefore, ORF1 and ORF2 were divided into smaller segments according to their functional domains, and the bootstrap values were calculated (Fig. 2, bottom). The length of ORF3 is short enough to be compared without segregation. The bootstrap values of the OTUs between NLV GI and NLV GII were over 950 in the nine trees depicted in the lower panel of the figure as was that of ORF3, with the exception of the P domain. Less than three points with a bootstrap value under 950 were found in NTPase, 3A-like, and polymerase, and NLV GI and NLV GII were topologically separated. The bootstrap value of the OTUs of BS5 and WUG1 was 841, and the bootstrap value of the other strains was 845. This indicated that the P domain of ORF 2 could not divide NLV GI and NLV GII. Therefore, the P domain of ORF2 was eliminated from further analysis.

Statistical analyses of genetic distances

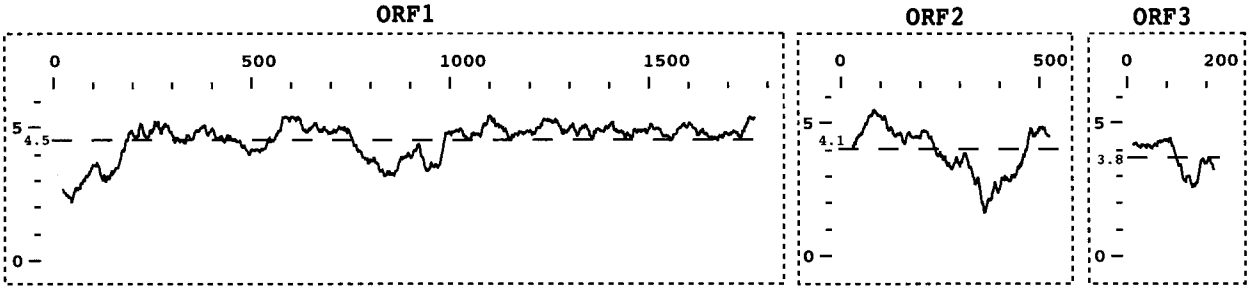
To determine whether the genogroups that distinguish NLV GI and NLV GII in the phylogenetic tree are stable,

a NLV GI (NV68, SO, BS5, Chiba, SzUG1, WUG1)

i) Nucleotide

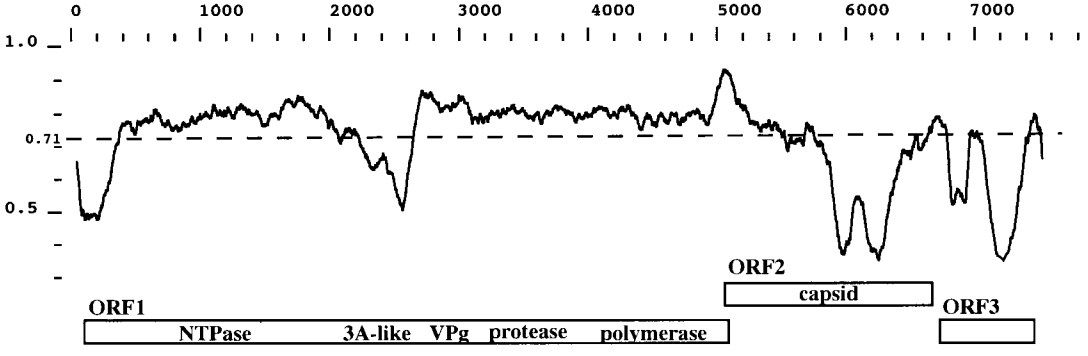


ii) Amino acid

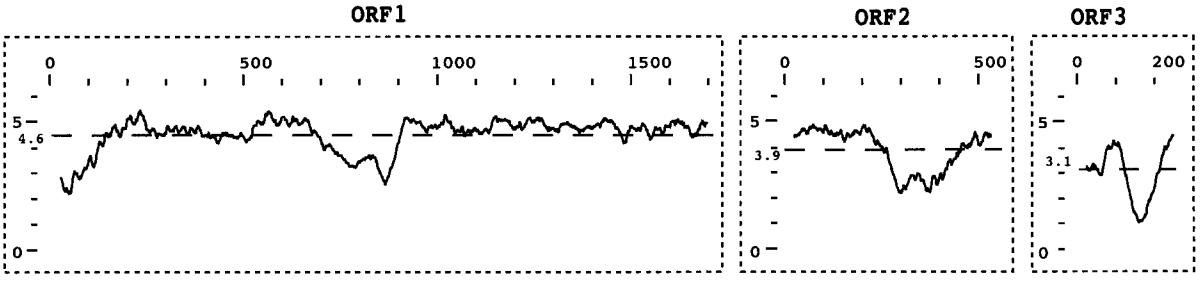


b NLV GII (LD, CW, Hawaii, MD145, U1, U3, U4, U16, U17, U18, U201, U25)

i) Nucleotide



ii) Amino acid



the pairwise distances were compared (Fig. 3). From the topology of the tree, the OTUs of NLV GI and NLV GII were distinguished by a bootstrap value of over 950 when the MC was used as the outgroup. Frequency distributions of the pairwise distances with bootstrap values over 950 were plotted. When the full-length genome was used in the analysis, the distributions of the pairwise distances between two genera, SLV and NLV, were grouped into C (the ninetieth percentile of 0.58–0.64) (Fig. 3a). Similarly, the pairwise distances between NLV GI and NLV GII were distributed in B (the ninetieth percentile of 0.46–0.49), indicating that independent clusters exist within the genus NLV. The pairwise distances represented within NLV GI or NLV GII were distributed in A (the ninetieth percentile of 0.15–0.35).

When the ORF1 and ORF2 were used for the analyses, the distributions of pairwise distances were similar to that of the full-length distance, showing the three distinct areas, A, B, and C (Figs. 3b and 3c). However, the distribution of ORF3 was different, and an overlap area between the genogroup (B) and the genus (C) was observed (Fig. 3d). Thus, the genogroup and genus were indistinguishable, and the ORF3 was eliminated from further analysis. When the comparison of the pairwise distances was carried out with each functional motif, the distributions of A and B were clearly distinguished in the polymerase and capsid N/S domain. Therefore, the polymerase and capsid N/S domains appeared to be suitable for the genotyping based both on topology and on pairwise distance analyses.

Identification of a recombination by nucleotide identity window search

When the phylogenetic trees of ORF1 and ORF2 were compared, WUG1 appeared to be located in a different manner (Fig. 2, bottom). WUG1 formed a cluster with the OTU containing SO in six ORF1 trees; however, this strain was more closely related to BS5 in the ORF2 and ORF3. Similarly, U1 was close to LD, CW, and MD145 in six ORF1 trees; however, this strain was closely related to the Hawaii strain and distantly related with LD, CW, and MD145 in ORF2 and ORF3. The topological difference between WUG1 and U1 in these trees suggested that a genetic recombination had occurred between ORF1 and ORF2.

To confirm this event, the nucleotide identity window search of WUG1 was performed with the five other NLV GI, U1 with LD, CW, Hawaii, U3, and U16 by using the DIVERT program (Simon *et al.,* 1998). The nucleotide identities among the strains were plotted along the entire nucleotide sequence (Fig. 4). WUG1 showed a close association with SO (over 80% identity) and was distantly related to NV68, BS5, Chiba, and SzUG1 in the nucleotide residue 1–5300. However, this strain showed a closer association with BS5 than these four strains in the nucleotide residue from 5300 to the 3′ end (Fig. 4a). U1 showed a close association with LD, CW, and Hawaii and was distantly related to U3 and U16 (under 70% identity) in the nucleotide residue 1–5100. However, this strain showed close identity with LD, CW, U3, and U16 in the nucleotide residue from 5100 to the 3′ end. These observations suggested that a recombination occurred at the junction of ORF1 and ORF2, where a highly conserved consensus sequence, 5′-GTRAATGAW GATGGCGTC-3′ corresponding to the nucleotide residue 5351–5368 in SO and the nucleotide residue 5081–5098 in LD were identified. Amplification of a nearly entire sequence of WUG1 and U1 followed by the sequencing analysis eliminated the possibility of a chimeric genome derived from two distinct fragments produced during the RT-PCR (data not shown).

Comparison of the distribution of the pairwise distances between the polymerase and capsid N/S domains

A large number of the polymerase and capsid N/S domain sequences have been reported previously (Ando *et al.,* 1995; Noel *et al.,* 1999; Vinje *et al.,* 1997, 2000; Wang *et al.,* 1994). To determine the genotype, 117 polymerase and 67 capsid N/S domain sequences larger than 200 nucleotides were selected from the database and phylogenetically analyzed along with those of the 10 strains newly sequenced in this study.

The frequency distribution of the pairwise distances formed two separate, nonoverlapping areas for the capsid N/S domains, A1 and A2 (Figs. 5a and 5b). The distances of area A1 ranged from 0 to 0.115 (mean 0.036) and those of area A2 ranged from 0.161 to 0.284 (mean 0.221) for NLV GI (Fig. 5a). Similarly, area A1 ranged from 0 to 0.111 (mean 0.033), and A2 ranged from 0.156 to 0.366 (mean 0.261) for NLV GII (Fig. 5b). These figures indicated that at least 99.7% of the pairwise distances were within ±3 SD. Area A1 was defined by the pairwise distance within the genotypes, namely the strains, whereas area A2 showed the distance between the genotypes. Student's *t* test indicated that two peaks of

FIG. 1. PlotSimilarity analyses of the 18 NLV. Sequence alignments were performed by the Pileup program, and the identities were calculated by the PlotSimilarity program of GCG (version 9). (a) Similarity scores of the six NLV GI strains (NV68, SO, BS5, Chiba, SzUG1, and WUG1) and (b) those of the 12 NLV GII strains (LD, CW, Hawaii, MD145, U1, U3, U4, U16, U17, U18, U201, and U25) were depicted with schematic representations of the NLV genome and three ORFs. (i) The nucleotide sequence similarities were calculated using 150-nucleotides sliding windows. (ii) The amino acid (aa) sequence similarities of ORF1, ORF2, and ORF3 were plotted by using 50-aa sliding windows. The dashed line represents the mean similarity scores across the genome or across each ORF.

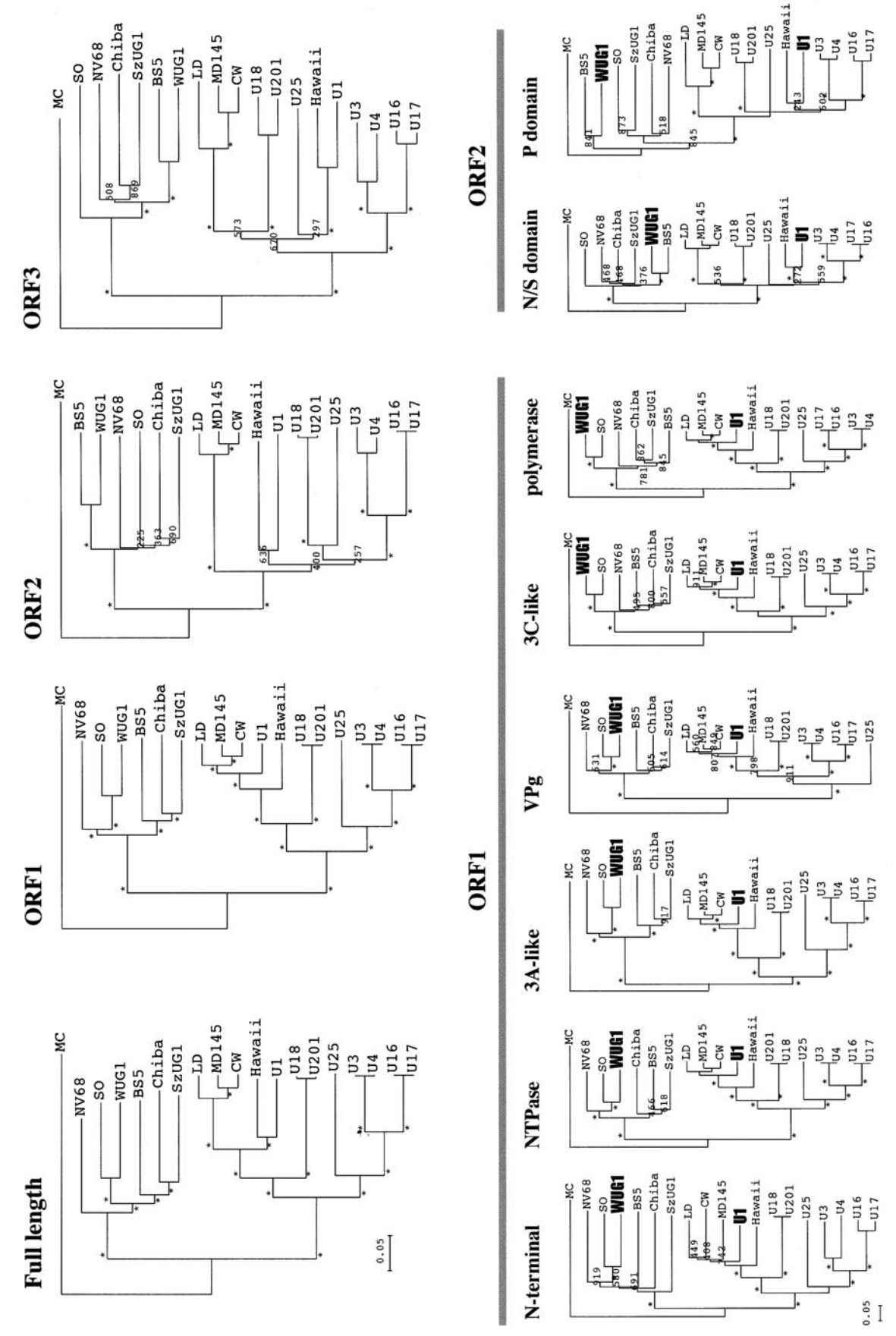


FIG. 2. The phylogenetic trees of 18 NLV reconstructed with the MC strain of the SLV as an outgroup. The nucleotide sequences of the full-length genome or the putative functional subgenomic regions were analyzed by the neighbor-joining method. The bootstrap values correspond to 1000 replications. The numbers on each branch indicate that the bootstrap value was less than 950. The asterisks show bootstrap values over 950.

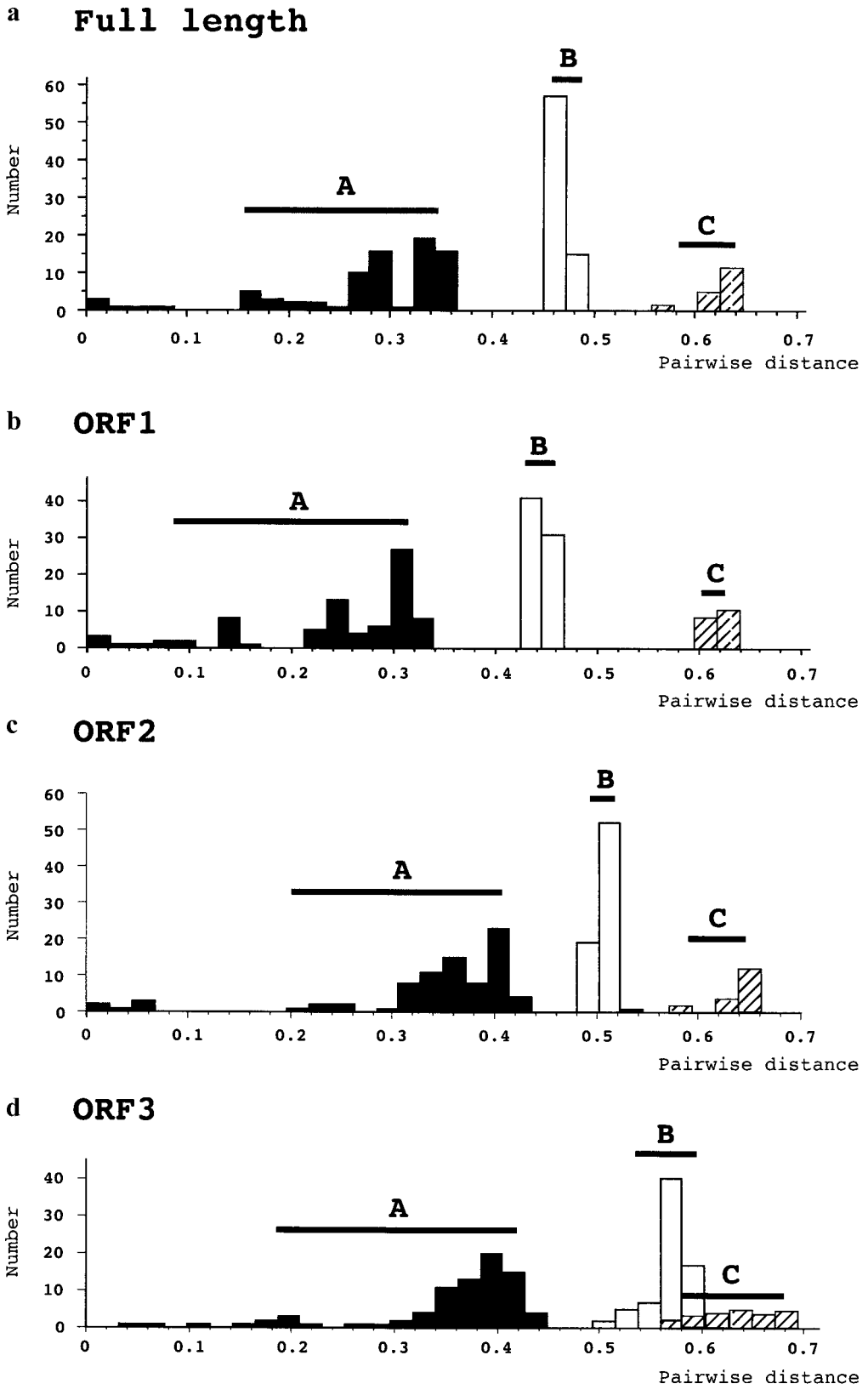


FIG. 3. The distribution of the pairwise distances. The numbers of the full-length genome (a), ORF1 (b), ORF2 (c), and ORF3 (d) are recorded on the *x*-axis in increments of 0.021 and range from 0 to 0.7 on the *y*-axis. The ninetieth percentile of three different nucleotide pairwise distances are indicated by intragenotype distances (A), intergenotype pairwise distances (B), and interout group pairwise distances (C). The dark bars indicate the number of intragenotype distances. The blank bars and striped bars indicate the number of intergenotype distances and intergenus distances, respectively.

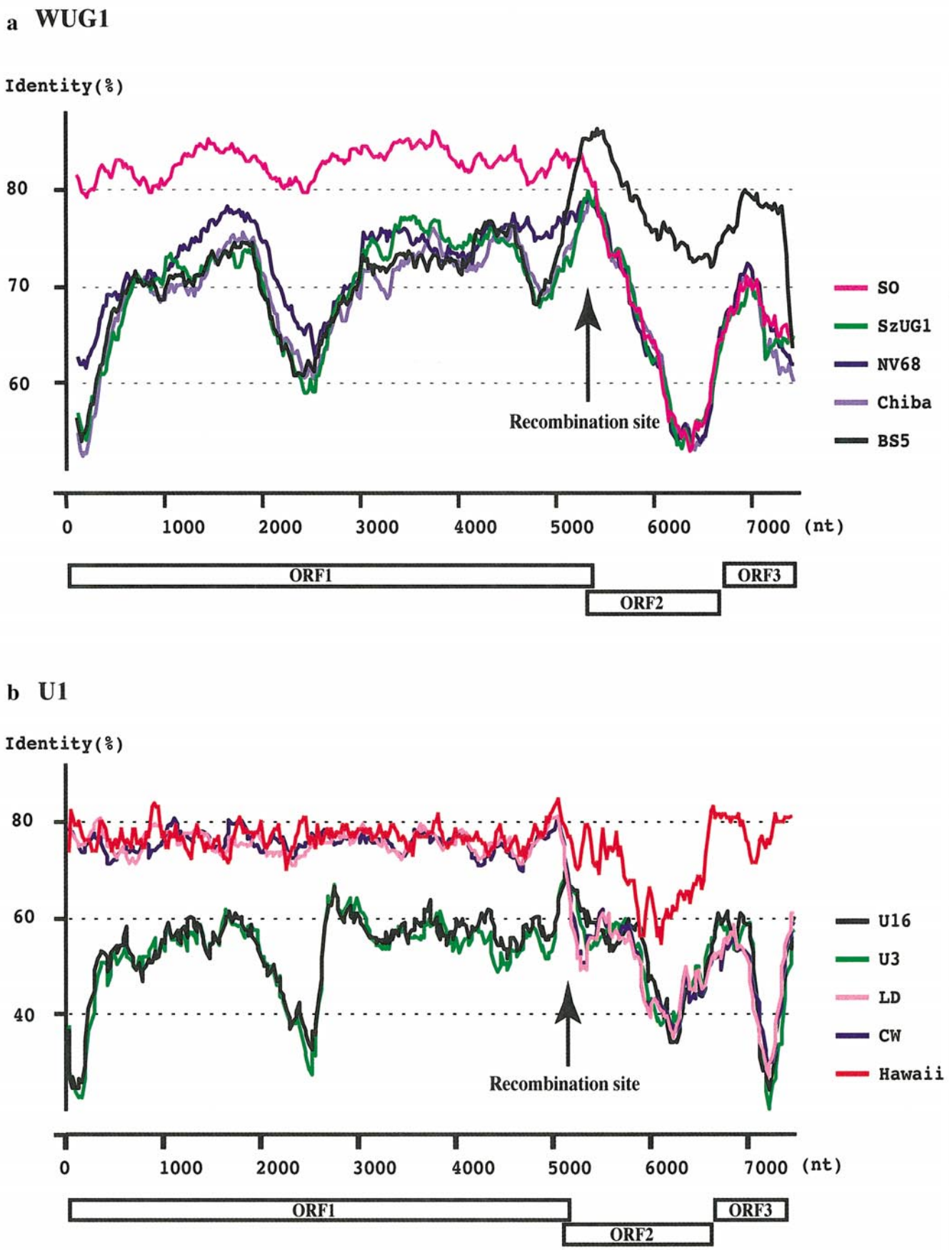


FIG. 4. Nucleotide identity plotting between the query sequence and the other five strains. The DIVERT program was used for pairwise comparisons in an incremented window. We used a window size of 90 nucleotides with an increment of 20. The vertical axis indicates the nucleotide identities between the query strain and the other reference strains expressed as percentages. The horizontal axis indicates nucleotide positions (nt). Colored lines correspond to the individual strains as indicated on the right side of the legend. The query sequence of (a) was WUG1 and that of (b) was U1.

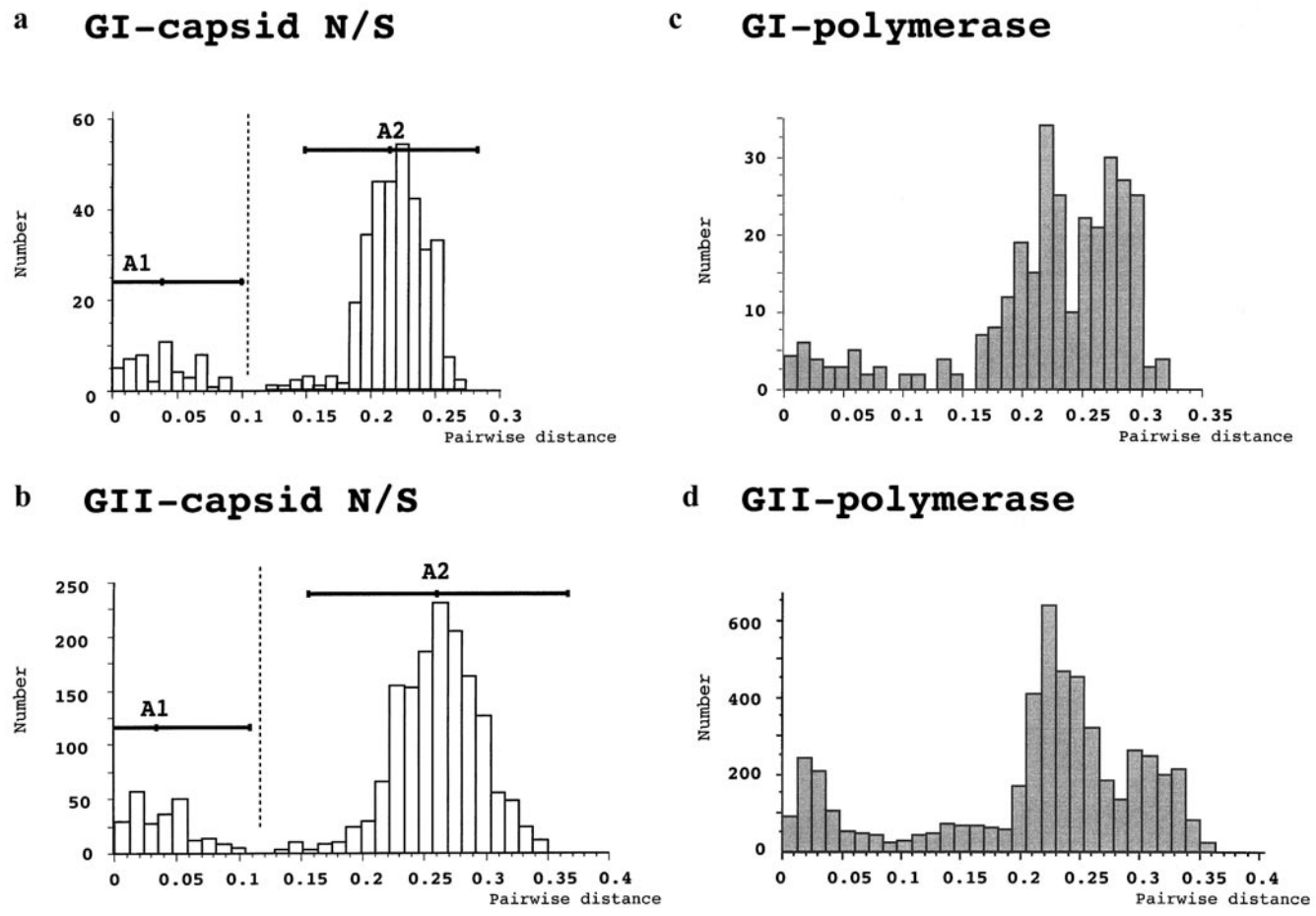


FIG. 5. The distribution of GI or GII intracluster evolutionary distances by pairwise comparison of NLV strains in which more than 200 nucleotide partial genome sequences were available in the genome database. The calculated evolutionary distance is indicated on the horizontal axis. Two distinct peaks are identified in the capsid N/S domain (c), (d). The horizontal bars indicate the mean value ± 3 SD for each distribution.

the capsid N/S domains were significant ($P < 0.001$). In contrast, distinct peaks were not found in the polymerase region of either NLV GI or NLV GII (Figs. 5c and 5d). Therefore, we concluded that the capsid N/S domain is the most suitable region for defining the genotypes.

Genotyping using the pairwise distances of the capsid N/S domain

The phylogenetic tree of the capsid N/S domain was constructed by the NJ method (Fig. 6). The OTUs were identified as the genotype clusters that showed a pairwise distance value of 0.221 ± 0.063 for NLV GI, and a value of 0.261 ± 0.105 for NLV GII. Therefore, 19 genotypes denoted GI/1 to GI/9 and GII/1 to GII/10 were statistically supported. In addition, the bootstrap values of each characterized genotype cluster were also statistically supported, with a few exceptions (e.g., BS5, AJ277608, AF195847). Although these three strains' OTUs were under 950, the distances between the other strains of the same genogroup indicated that their pairwise distances were within 0.221 ± 0.063 for NLV GI and 0.261 ± 0.105 for NLV GII. Therefore, each OTU formed distinct genotype clusters.

In the pairwise distance analysis, the genotype and strain were not separated into each cluster in the polymerase since no distinct peak was found. A phylogenetic analysis explored two genogroups in the polymerase (Fig. 7). Each genogroup contained at least six clusters in NLV GI, and nine or more clusters in NLV GII. However, the lack of a stable bootstrap value prevented further subgrouping within each genogroup. For example, the NLV GI cluster contains OTUs-1 (DS strain, AB026481, and U07612) with a pairwise distance of 0.05–0.15. However, it is difficult to separate the genotype and strain (Fig. 7, OTUs-1). Similarly, NLV GII contains OTUs-2 and OTUs-3 with pairwise distances of 0.05–0.15, but it was impossible to distinguish the genotypes from the strains due to a failure to meet the criteria (Fig. 7). Furthermore, the recombination added complications to the genotyping. Similar to WUG1 and U1, Arg320 could be classified into different genotypes depending on the region used for the analysis. This strain was clustered with NOR89JD,

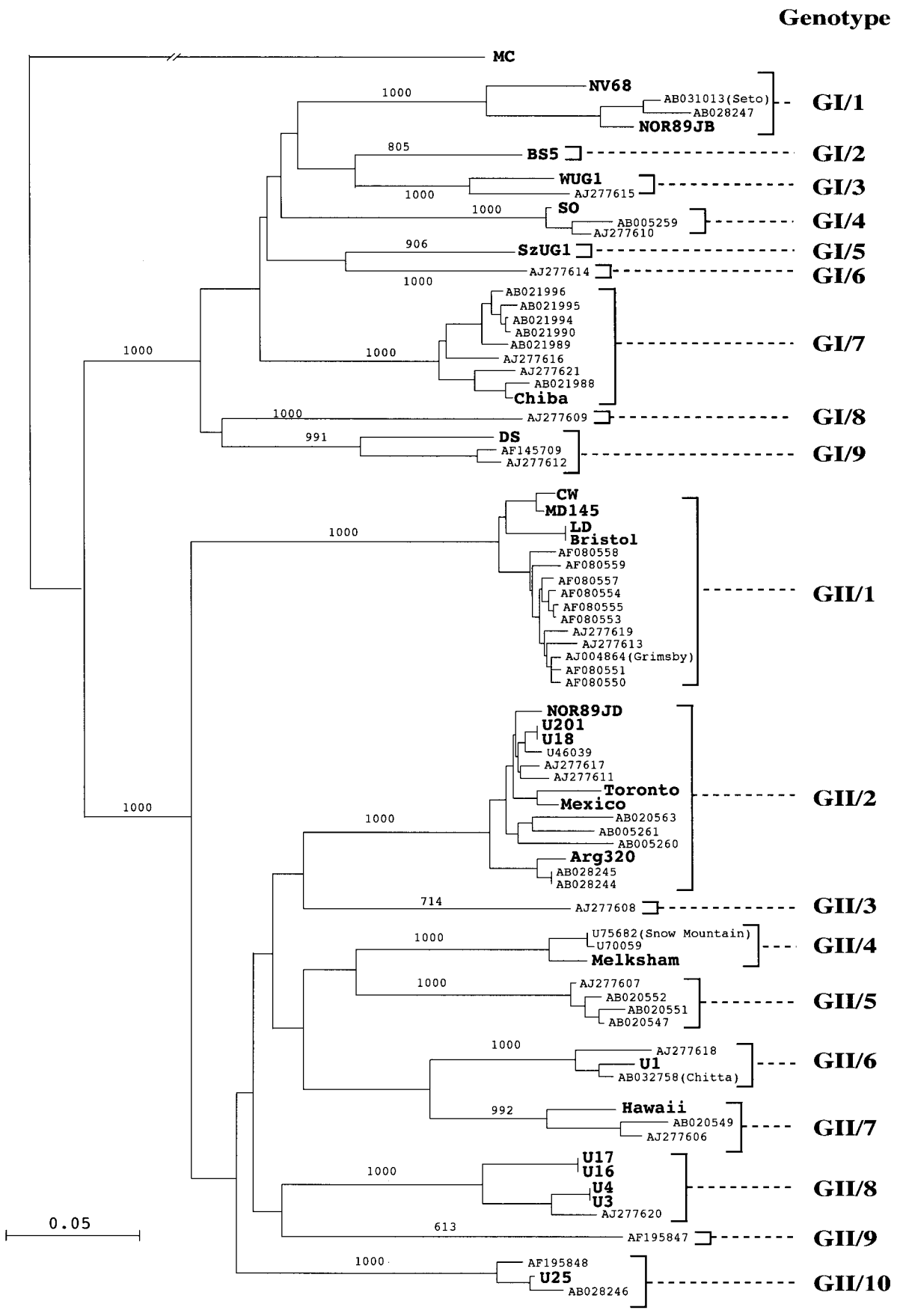


FIG. 6. The phylogenetic trees reconstructed for partial sequences of the capsid N/S domain region of NLV using MC as the outgroup. The numbers on each branch indicate the bootstrap values for the clusters supported by that branch. The putative genotypes are named and indicated on each cluster. The bold letters indicate the reference strains and the 10 newly sequenced strains with the characterized sequence from polymerase through the capsid N/S domain.

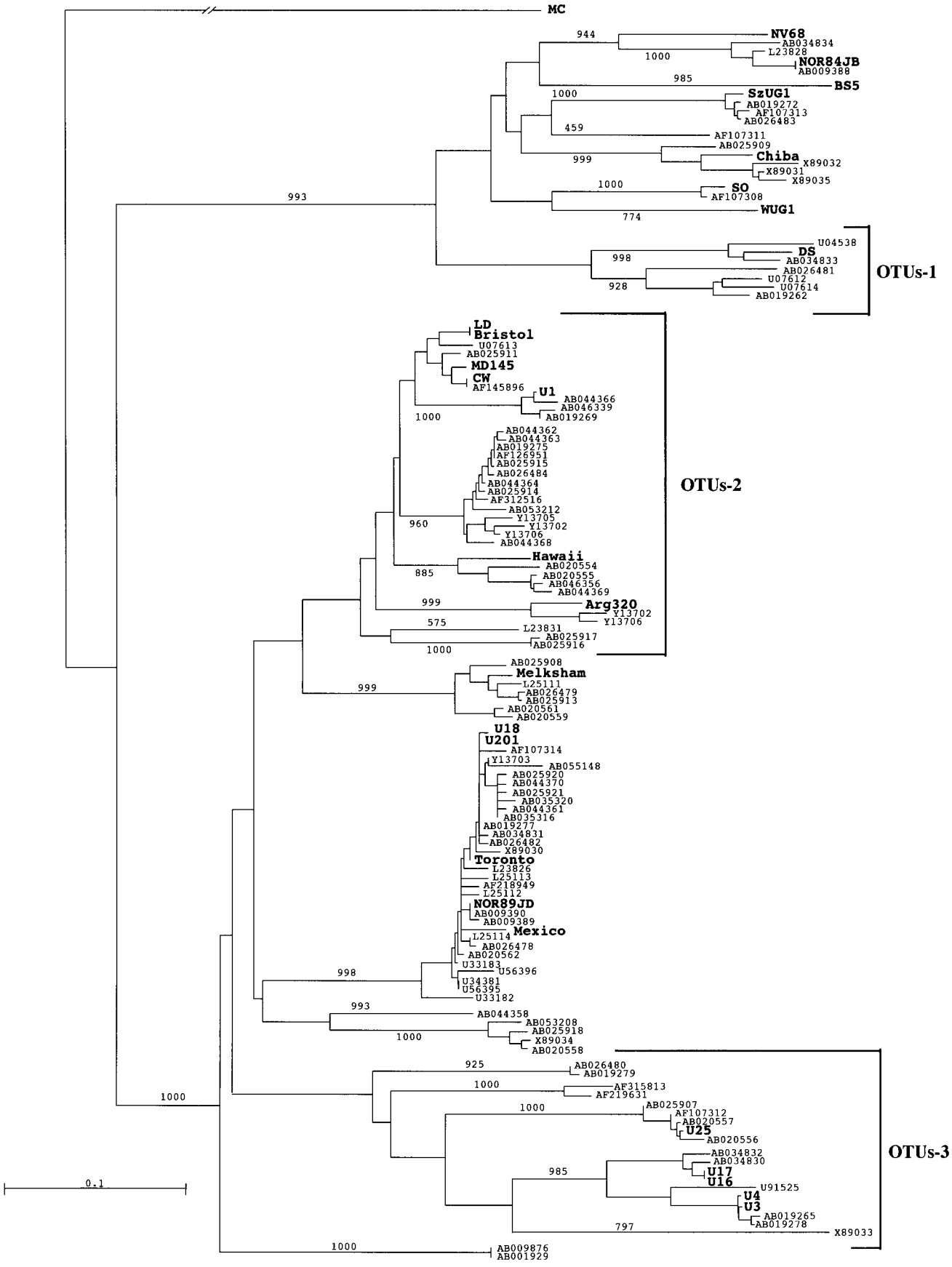


FIG. 7. The phylogenetic trees reconstructed for partial sequences of the RNA-dependent RNA polymerase region of NLV using MC as the outgroup. Bold letters signify that the strain contains the characterized sequence from polymerase through the capsid N/S domain. The unidentified clusters are indicated on the right side of OTUs.

U201, U18, Toronto, and Mexico when the polymerase was compared, but was characterized as an independent cluster in the capsid N/S domain (Fig. 7). This discrepancy was obviously due to the genetic cross within the ORF1 and ORF2 junction. Therefore, neither a distinct peak in the pairwise distance analysis nor a distinct genotype was obtained by the NJ tree based on the polymerase.

DISCUSSION

The polymerase region has been favored in gene amplification of NLV. The primers used for the RT-PCR contain degenerated sequences based on the conserved sequence motifs, GLPSG and YGDD (Ando *et al.,* 1995; Saito *et al.,* 1998). Although this region displayed a high average similarity score (Figs. 1a and 1b), the nucleotide sequence alignment of 18 NLV, including 10 strains newly determined in this study and the previously determined eight strains, indicated a large number of nonsynonymous substitutions in the third codons, suggesting that there may be problems when an efficient amplification of all NLV genotypes is examined (Kojima, 2002). Multiple alignment of the full-length genome showed numerous conserved segments of 15–27 nucleotides in the region from 100 nucleotides upstream of the initiation codon through the first 160 nucleotides of the ORF2. This region yielded the highest similarity scores in both NLV GI and NLV GII genomes. No other regions in the genome contain a conserved sequence of a successive 18 nucleotides. We have shown previously that RT-PCR targeting this region provided better results than that targeting the polymerase region (Kojima, 2002).

The average similarity score of ORF1 was higher than those of ORF2 and ORF3 in both nucleotide and amino acid sequences. The amino acid sequences required for the viral replication (e.g., NTPase, protease, and polymerase) might have lower evolution rates compared with those of ORF2 and ORF3. Although NLV capsid protein contains a highly conserved N/S domain, the average similarity score of nucleotide and amino acid sequences was lower than that of ORF1. This is largely due to the capsid P2 domain that showed the lowest score throughout the genome. The V3 loop of HIV (Holmes *et al.,* 1992) and the E2/hypervariable region of HCV (Kato *et al.,* 1992) contain highly variable sequences and show high amino acid mutation rates. These regions are located on the outer surfaces of HIV and HCV particles, conferring antigenic variability and the ability to evade host immunological surveillance (Kato *et al.,* 1993; Shioda *et al.,* 1992). Although NLV is not an enveloped virus, the P2 domain exposed on the surface of the capsid may have a similar interaction with the host immune system, and the antigenic variation may facilitate the reinfection.

In many RNA viruses, recombination and reassortment are known to be highly associated with evolution (Worobey and Holmes, 1999). For example, viruses included in the *Togaviridae* and *Coronaviridae* share homologous but differently arranged genes in the genome. The arrangement is probably due to a recombination, which is important in their evolutionary history (Hertz and Huang, 1992; Keck *et al.,* 1987; Weiss and Schlesinger, 1991). In enteroviruses, the exchange of a part of the genome occurs by a recombination, allowing the alterations to occur in a host and tissue tropism (Santti *et al.,* 1999). In Dengue virus, the recombination might have important implications for virus pathogenicity, vaccine safety and efficiency, and diagnosis (Tolou *et al.,* 2001).

In NLV, it was suggested that a recombination may have occurred between ORF1 and ORF2 (Jiang *et al.,* 1999; Vinje *et al.,* 2000). We have amplified the entire genome through a single long RT-PCR, and the full-length nucleotide sequence was determined using 10 NLV. The nucleotide identity window search and phylogenetic analysis demonstrated that the recombination might have occurred between not only the NLV GII strains but also the NLV GI strains (Fig. 4). This is the first report of the recombination of NLV GI. In both WUG1 and U1, the change of the genotype occurred in the 18-nucleotides consensus sequence, 5′-GTR AAT GAW GAT GGC GTC-3′, encoded between the 3′ end of ORF1 and the 5′ terminus of ORF1. Coinfection by NLV GI and NLV GII and infection by several genotypes of NLV are reported in NLV-related food poisoning (Gray *et al.,* 1997), which seems common in NLV (Ando *et al.,* 1995; Maunula *et al.,* 1999). Recent studies have indicated that NLV is found in calves and pigs (Liu *et al.,* 1999; Sugieda *et al.,* 1998; van Der Poel *et al.,* 2000). Therefore, the recombination may occur between NLV from humans and NLV from other species.

The capsid protein of the feline calicivirus is translated through a subgenomic RNA (Herbert *et al.,* 1997). Therefore, the capsid of NLV is likely to be translated through the subgenomic RNA. The 18 nucleotides consensus sequence seems to be involved in the production of the genomic and subgenomic RNA. Identification of the recombination emphasizes the importance of the full-length genomic sequence data from various genotypes. The data will improve further understanding of the genetic diversity of NLV.

In addition to the two genogroups, NLV GI and NLV GII, we identified nine genotypes (GI/1–GI/9) within the NLV GI cluster with the mean pairwise distances, 0.221 $\pm$ 0.021, and 10 genotypes (GII/1–GII/10) within the NLV GII cluster with the mean 0.261 $\pm$ 0.105 in an analysis of the capsid N/S domain. Four antigenically distinct strains (Norwalk, Hawaii, W-Ditchling, and Marin) were initially identified by Dolin *et al.* with an IEM (Dolin *et al.,* 1982). Jiang *et al.* produced hyperimmune antisera and confirmed that they were distinguished from each other by ELISA. Two distinct genogroups were identified when the nucleotide sequences of these strains became

available (Jiang *et al.,* 1995). Recently, Kobayashi *et al.* identified at least two antigenically different strains in NLV GI, namely Seto virus (AB031013) and Chiba virus, and several strains in NLV GII by an antigen ELISA; these antigenic classifications appeared to be correlated well with that determined by a phylogenetic analysis of the capsid N/S domain (Kobayashi *et al.,* 2000a,b). Therefore, it may be possible to predict antigenic types by phylogenetic analysis of the capsid N/S domain. In this study, Grimsby virus, Mexico virus, Snow Mountain virus, and Chitta virus of NLV GII were classified as the independent genotypes GII/1, GII/2, GII/4, and GII/7, respectively. Further study using the recombinant virus-like particles and their antisera are needed to antigenically classify these viruses. A global survey with the phylogenetic analysis described in this study will facilitate the epidemiologic study of NLV and further discovery of the genotypes and new antigenic types.

MATERIALS AND METHODS

Stool specimens and extraction of viral RNA

Ten stool specimens, SzUG1, WUG1, U1, U3, U4, U16, U17, U18, U201, and U25, were collected from adults and children involved in outbreaks and sporadic cases of gastroenteritis that occurred between 1997 and 2001 in Saitama and Wakayama Prefecture, Japan. The U3, U4, U16, and U17 were from different patients in the same outbreak in 1997. SzUG1, U1, U18, and U25, were collected from patients with sporadic cases of acute gastroenteritis associated with oysters between 1997 and 1999. U201 was from a child with acute gastroenteritis in a 1998 outbreak at an elementary school. WUG1 was collected from a child with acute gastroenteritis in a 2000 outbreak at a swimming school in Wakayama Prefecture, Japan. All stool samples were found to be positive for NLV by electron microscopy.

Viral RNA was extracted from 140 μl of the 10% stool suspension with the QIA viral RNA extraction kit (Qiagen) according to the manufacturer's instructions. The RNA was eluted with 60 μl of diethyl pyrocarbonate-treated water and kept at -80°C until use.

Amplification of the NLV sequence between the polymerase and the capsid coding region

The RNA solution (2 μl) was mixed with 1 μl (75 pmol) of random hexamers (pdN6, Pharmacia Biotech), heated at 94°C for 2 min, and chilled on ice. This solution was mixed with 17 μl of reverse transcription (RT) reaction mix containing 25 mM Tris–HCl, pH 8.3, 125 mM KCl, 12.5 mM $MgCl_2$, 1.25 mM dNTPs, 2.5 units RNasin (Promega), and 200 units SuperScript version II reverse transcriptase (Life Technologies). RT was performed at 37°C for 4 h, and the reaction was terminated by incubation at 94°C for 5 min and followed by chilling on ice. The cDNA (1 μl) was then mixed with 49 μl of PCR reaction mixture containing 20 pmol of the sense primer NLV4594S 5′-YTC YTT CTA TGG YGA TGA TGA-3′ and the same amount of the antisense primer G1R1 5′-CCA ACC CAR CCA TTR TAC ATT T-3′ or G2R11 5′-CCA CCW CAT AAC CAT TRT ACA T-3′, 2.5 units ExTaq DNA polymerase (Takara, Tokyo, Japan), 200 μM each of dNTP, 1× PCR buffer provided by the manufacturer. After denaturation at 94°C for 2 min, 45 cycles of amplification were performed at 94°C for 1 min, 55°C for 1 min, and 72°C for 2 min. The products were separated by electrophoresis on a 1% agarose gel, and the DNA fragments of the proper size were purified with QIA-quick gel extraction kits. The DNA sequence was determined with the Big-Dye terminator cycle sequence kit and an ABI 377A sequencer (Applied Biosystems Inc.).

Determination of the complete nucleotide sequences

The complete nucleotide sequences of the 10 strains were performed with two separate RT-PCR products of the genome. Using SuperScript version II reverse transcriptase, RT was performed with a tagged primer complementary to the poly(A) tail of the genome (TX30SXN 5′-GAC TAG TTC TAG ATC GCG AGC GGC CGC CT T TTT TTT TTT TTT TTT TTT TTT TTT TTT TT-3′, with the tag sequence underlined). Approximately 2300 nucleotides of the 3′ end of the genome was amplified with a forward primer designed in the junction between polymerase and capsid, NLV4594S (5′-YTC YTT CTA TGG YGA TGA TGA-3′), and a reverse primer containing the tag sequence, SXN (5′-GAC TAG TTC TAG ATC GCG AGC GGC CGC CC-3′). The 35 cycles of the amplification was performed with TbrEXT polymerase (FINZYME) according to the manufacturer's instructions. Similarly, approximately 5000 nucleotides of the 5′ fragment was obtained with a forward primer designed in the conserved sequence at the 5′ end of the genome, NLV1S25 (5′-GTG AAT GAA GAT GGC GTC TAA CGA C-3′) or NLV1S 31(5′-GTG AAT GAT GAT GGC GTC RAA AGA CGT CGT T-3′) and the strain-specific reverse primers designed in the junction between the polymerase and capsid protein, as follows: SzUG1/5077A (5′-CCT GTC CAA CCT GCC TTG AA-3′) for SzUG1, NLV4671A (5′-ATT CCT TAA GTT TTG CTG TC-3′) for U1, NLV4646A (5′-GTC AAG GTT TAT GTC AGT GC-3′) for U3 and U4, NLV4803A (5′-TGA TCC AGC TTC CCA AAC CA-3′) for U16 and U17, NLV4798A (5′-CAG CTT TCC AAA CCA ACC TG-3′) for U18 and U201, 4708A (5′-GCC CTC AGT TTT GTC TGG CC-3′) for U25, and WUG4898A (5′-ACT CTA TGT CAG TGG AAA CT-3′). The amplification of the two fragments resulted in a 100- to 200-nucleotides overlapping region.

The amplified fragments were sequenced directly with the Big-Dye terminator cycle sequence kit and ABI 377A sequencer (ABI), according to the manufacturer's instructions. Strain-specific primers were prepared and used for

the primer walking to determine both orientations. The nucleotide sequence of the 5′ end of the genome was determined by three different methods: 5′ rapid amplification of cDNA end (5′-RACE) with RNA–RNA ligation (Katayama *et al.,* 1995) with purified genomic RNA, DNA–DNA ligation (Katayama *et al.,* 1998) with cDNA, and TdT-tailing with cDNA (Hardy and Estes, 1996). A long RT-PCR was performed with a strain-specific forward primer designed from the 5′ terminal sequences and the reverse tag primer, SXN, to obtain the nearly full-length DNA (approximately 7600 nucleotides), and the nucleotide sequences were verified in their entirely.

Phylogenetic analyses

The nucleic acid sequence alignment was performed with Clustal X (http://www-igbmc.u-strasbg.fr/BioInfo/) with parameters provided in Clustal W1.6, and the column positions containing the gaps were removed. The genetic distance was calculated by Kimura's two-parameter method (Kimura, 1980), and a distance matrix file was created. A phylogenetic tree with 1000 bootstrap replicates was generated by the neighbor-joining method with Clustal X. Manchester virus in the genus Sapporo-like viruses was used as an outgroup. Bootstrap values of 950 or higher were considered statistically significant for the grouping(Efron *et al.,* 1996). Amino acid sequences were aligned with BLOSUM 30 in Clustal X, and the distances were calculated according to the Kimura method and plotted by STATVIEW. The nucleotide sequence identities were analyzed with GENETYX-MAC program. The nucleotide and amino acid sequence similarity scores were analyzed with the Plot-Similarity program in the GCG.

Nucleotide sequence accession numbers

All sequences obtained in this study were deposited in DDBJ and/or GenBank with the following accession numbers: AB039774, SzUG1; AB039775, U1; AB039776, U3; AB039777, U4; AB039778, U16; AB039779, U17; AB039781, U18; AB039782, U201; AB039780, U25; and AB081723, WUG1. The accession numbers of the complete genome sequences from the databases are M87661, NV68; L07418, SO; AF093797, BS5; AB042808, Chiba, AF145896, CW; X86557, LD; AY032605, MD145-12, U07611, HI; and X86560, MC. The accession numbers of other partial sequences are presented in Figs. 6 and 7.

ACKNOWLEDGMENTS

This work was supported in part by a grant for Research on Emerging and Reemerging Infectious Diseases and Research on Health Sciences focusing on Drug Innovation from the Ministry of Health, Labour, and Welfare, Japan.

REFERENCES

Ando, T., Monroe, S. S., Gentsch, J. R., Jin, Q., Lewis, D. C., and Glass, R. I. (1995). Detection and differentiation of antigenically distinct small round-structured viruses (Norwalk-like viruses) by reverse transcription-PCR and southern hybridization. *J. Clin. Microbiol.* **33,** 64–71.

Dingle, K. E., Lambden, P. R., Caul, E. O., and Clarke, I. N. (1995). Human enteric Caliciviridae: The complete genome sequence and expression of virus-like particles from a genetic group II small round structured virus. *J. Gen. Virol.* **76**(Pt. 9), 2349–2355.

Dolin, R., Reichman, R. C., Roessner, K. D., Tralka, T. S., Schooley, R. T., Gary, W., and Morens, D. (1982). Detection by immune electron microscopy of the Snow Mountain agent of acute viral gastroenteritis. *J. Infect. Dis.* **146,** 184–189.

Efron, B., Halloran, E., and Holmes, S. (1996). Bootstrap confidence levels for phylogenetic trees. *Proc. Natl. Acad. Sci. USA* **93,** 7085–7090.

Glass, P. J., White, L. J., Ball, J. M., Leparc-Goffart, I., Hardy, M. E., and Estes, M. K. (2000). Norwalk virus open reading frame 3 encodes a minor structural protein. *J. Virol.* **74,** 6581–6591.

Gray, J. J., Green, J., Cunliffe, C., Gallimore, C., Lee, J. V., Neal, K., and Brown, D. W. (1997). Mixed genogroup SRSV infections among a party of canoeists exposed to contaminated recreational water. *J. Med. Virol.* **52,** 425–429.

Green, J., Vinje, J., Gallimore, C. I., Koopmans, M., Hale, A., and Brown, D. W. (2000). Capsid protein diversity among Norwalk-like viruses. *Virus Genes* **20,** 227–236.

Green, S. M., Dingle, K. E., Lambden, P. R., Caul, E. O., Ashley, C. R., and Clarke, I. N. (1994). Human enteric Caliciviridae: A new prevalent small round-structured virus group defined by RNA-dependent RNA polymerase and capsid diversity. *J. Gen. Virol.* **75**(Pt. 8), 1883–1888.

Hardy, M. E., and Estes, M. K. (1996). Completion of the Norwalk virus genome sequence. *Virus Genes* **12,** 287–290.

Herbert, T. P., Brierley, I., and Brown, T. D. (1997). Identification of a protein linked to the genomic and subgenomic mRNAs of feline calicivirus and its role in translation. *J. Gen. Virol.* **78**(Pt. 5), 1033–1040.

Hertz, J. M., and Huang, H. V. (1992). Utilization of heterologous alphavirus junction sequences as promoters by Sindbis virus. *J. Virol.* **66,** 857–864.

Holmes, E. C., Zhang, L. Q., Simmonds, P., Ludlam, C. A., and Brown, A. J. (1992). Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient. *Proc. Natl. Acad. Sci. USA* **89,** 4835–4839.

Jiang, X., Espul, C., Zhong, W. M., Cuello, H., and Matson, D. O. (1999). Characterization of a novel human calicivirus that may be a naturally occurring recombinant. *Arch. Virol.* **144,** 2377–2387.

Jiang, X., Graham, D. Y., Wang, K. N., and Estes, M. K. (1990). Norwalk virus genome cloning and characterization. *Science* **250,** 1580–1583.

Jiang, X., Matson, D. O., Ruiz-Palacios, G. M., Hu, J., Treanor, J., and Pickering, L. K. (1995). Expression, self-assembly, and antigenicity of a snow mountain agent-like calicivirus capsid protein. *J. Clin. Microbiol.* **33,** 1452–1455.

Jiang, X., Wang, M., Wang, K., and Estes, M. K. (1993). Sequence and genomic organization of Norwalk virus. *Virology* **195,** 51–61.

Katayama, K., Kageyama, T., Fukushi, S., Hoshino, F. B., Kurihara, C., Ishiyama, N., Okamura, H., and Oya, A. (1998). Full-length GBV-C/HGV genomes from nine Japanese isolates: Characterization by comparative analyses. *Arch. Virol.* **143,** 1063–1075.

Katayama, K., Kurihara, C., Fukushi, S., Hoshino, F. B., Ishikawa, K., Nagai, H., Ando, T., and Oya, A. (1995). Characterization of the hog cholera virus 5′ terminus. *Virus Genes* **10,** 185–187.

Kato, N., Ootsuyama, Y., Ohkoshi, S., Nakazawa, T., Sekiya, H., Hijikata, M., and Shimotohno, K. (1992). Characterization of hypervariable regions in the putative envelope protein of hepatitis C virus. *Biochem. Biophys. Res. Commun.* **189,** 119–127.

Kato, N., Sekiya, H., Ootsuyama, Y., Nakazawa, T., Hijikata, M., Ohkoshi, S., and Shimotohno, K. (1993). Humoral immune response to hyper-

variable region 1 of the putative envelope glycoprotein (gp70) of hepatitis C virus. *J. Virol.* **67,** 3923–3930.

Keck, J. G., Stohlman, S. A., Soe, L. H., Makino, S., and Lai, M. M. (1987). Multiple recombination sites at the 5′-end of murine coronavirus RNA. *Virology* **156,** 331–341.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16,** 111–120.

Kobayashi, S., Sakae, K., Suzuki, Y., Ishiko, H., Kamada, K., Suzuki, K., Natori, K., Miyamura, T., and Takeda, N. (2000a). Expression of recombinant capsid proteins of chitta virus, a genogroup II norwalk virus, and development of an ELISA to detect the viral antigen. *Microbiol. Immunol.* **44,** 687–693.

Kobayashi, S., Sakae, K., Suzuki, Y., Shinozaki, K., Okada, M., Ishiko, H., Kamata, K., Suzuki, K., Natori, K., Miyamura, T., and Takeda, N. (2000b). Molecular cloning, expression, and antigenicity of seto virus belonging to genogroup I norwalk-like viruses. *J. Clin. Microbiol.* **38,** 3492–3494.

Kojima, S., Kageyama, T., Fukushi, S., Hoshino, F. B., Shinohara, M., Uchida, K., Natori, K., Takeda, N., and Katayama, K. (2002). Genogroup-specific PCR primers for detection of Norwalk-like viruses. *J. Virol. Methods* **100,** 107–114.

Lambden, P. R., Caul, E. O., Ashley, C. R., and Clarke, I. N. (1993). Sequence and genome organization of a human small round-structured (Norwalk-like) virus. *Science* **259,** 516–519.

Lambden, P. R., Liu, B., and Clarke, I. N. (1995). A conserved sequence motif at the 5′ terminus of the Southampton virus genome is characteristic of the Caliciviridae. *Virus Genes* **10,** 149–152.

Lee, Y. F., Nomoto, A., Detjen, B. M., and Wimmer, E. (1977). A protein covalently linked to poliovirus genome RNA. *Proc. Natl. Acad. Sci. USA* **74,** 59–63.

Lewis, D. C. (1990). Three serotypes of Norwalk-like virus demonstrated by solid-phase immune electron microscopy. *J. Med. Virol.* **30,** 77–81.

Lewis, D. C., Lightfoot, N. F., and Pether, J. V. (1988). Solid-phase immune electron microscopy with human immunoglobulin M for serotyping of Norwalk-like viruses. *J. Clin. Microbiol.* **26,** 938–942.

Liu, B., Clarke, I. N., and Lambden, P. R. (1996). Polyprotein processing in Southampton virus: Identification of 3C-like protease cleavage sites by in vitro mutagenesis. *J. Virol.* **70,** 2605–2610.

Liu, B. L., Clarke, I. N., Caul, E. O., and Lambden, P. R. (1995). Human enteric caliciviruses have a unique genome structure and are distinct from the Norwalk-like viruses. *Arch. Virol.* **140,** 1345–1356.

Liu, B. L., Lambden, P. R., Gunther, H., Otto, P., Elschner, M., and Clarke, I. N. (1999). Molecular characterization of a bovine enteric calicivirus: Relationship to the Norwalk-like viruses. *J. Virol.* **73,** 819–825.

Madore, H. P., Treanor, J. J., Buja, R., and Dolin, R. (1990). Antigenic relatedness among the Norwalk-like agents by serum antibody rises. *J. Med. Virol.* **32,** 96–101.

Maunula, L., Piiparinen, H., and von Bonsdorff, C. H. (1999). Confirmation of Norwalk-like virus amplicons after RT-PCR by microplate hybridization and direct sequencing. *J. Virol. Methods* **83,** 125–134.

Nei, M., and Jin, L. (1989). Variances of the average numbers of nucleotide substitutions within and between populations. *Mol. Biol. Evol.* **6,** 290–300.

Noel, J. S., Fankhauser, R. L., Ando, T., Monroe, S. S., and Glass, R. I. (1999). Identification of a distinct common strain of "Norwalk-like viruses" having a global distribution. *J. Infect. Dis.* **179,** 1334–1344.

Prasad, B. V., Hardy, M. E., Dokland, T., Bella, J., Rossmann, M. G., and Estes, M. K. (1999). X-ray crystallographic structure of the Norwalk virus capsid. *Science* **286,** 287–290.

Rueckert, R. R., and Wimmer, E. (1984). Systematic nomenclature of picornavirus proteins. *J. Virol.* **50,** 957–959.

Saito, H., Saito, S., Kamada, K., Harata, S., Sato, H., Morita, M., and Miyajima, Y. (1998). Application of RT-PCR designed from the sequence of the local SRSV strain to the screening in viral gastroenteritis outbreaks. *Microbiol. Immunol.* **42,** 439–446.

Santti, J., Hyypia, T., Kinnunen, L., and Salminen, M. (1999). Evidence of recombination among enteroviruses. *J. Virol.* **73,** 8741–8749.

Seah, E. L., Marshall, J. A., and Wright, P. J. (1999). Open reading frame 1 of the Norwalk-like virus Camberwell: Completion of sequence and expression in mammalian cells. *J. Virol.* **73,** 10531–10535.

Shioda, T., Levy, J. A., and Cheng-Mayer, C. (1992). Small amino acid changes in the V3 hypervariable region of gp120 can affect the T-cell-line and macrophage tropism of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. USA* **89,** 9434–9438.

Simon, F., Mauclere, P., Roques, P., Loussert-Ajaka, I., Muller-Trutwin, M. C., Saragosti, S., Georges-Courbot, M. C., Barre-Sinoussi, F., and Brun-Vezinet, F. (1998). Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nat. Med.* **4,** 1032–1037.

Someya, Y., Takeda, N., and Miyamura, T. (2000). Complete nucleotide sequence of the chiba virus genome and functional expression of the 3C-like protease in Escherichia coli. *Virology* **278,** 490–500.

Sugieda, M., Nagaoka, H., Kakishima, Y., Ohshita, T., Nakamura, S., and Nakajima, S. (1998). Detection of Norwalk-like virus genes in the caecum contents of pigs. *Arch. Virol.* **143,** 1215–1221.

Tolou, H. J., Couissinier-Paris, P., Durand, J. P., Mercier, V., de Pina, J. J., de Micco, P., Billoir, F., Charrel, R. N., and de Lamballerie, X. (2001). Evidence for recombination in natural populations of dengue virus type 1 based on the analysis of complete genome sequences. *J. Gen. Virol.* **82,** 1283–1290.

van Der Poel, W. H., Vinje, J., van Der Heide, R., Herrera, M. I., Vivo, A., and Koopmans, M. P. (2000). Norwalk-like calicivirus genes in farm animals. *Emerg. Infect. Dis.* **6,** 36–41.

Vinje, J., Altena, S. A., and Koopmans, M. P. (1997). The incidence and genetic variability of small round-structured viruses in outbreaks of gastroenteritis in The Netherlands. *J. Infect. Dis.* **176,** 1374–1378.

Vinje, J., Green, J., Lewis, D. C., Gallimore, C. I., Brown, D. W., and Koopmans, M. P. (2000). Genetic polymorphism across regions of the three open reading frames of "Norwalk-like viruses." *Arch. Virol.* **145,** 223–241.

Wang, J., Jiang, X., Madore, H. P., Gray, J., Desselberger, U., Ando, T., Seto, Y., Oishi, I., Lew, J. F., Green, K. Y., and et al. (1994). Sequence diversity of small, round-structured viruses in the Norwalk virus group. *J. Virol.* **68,** 5982–5990.

Weiss, B. G., and Schlesinger, S. (1991). Recombination between Sindbis virus RNAs. *J. Virol.* **65,** 4017–4025.

Worobey, M., and Holmes, E. C. (1999). Evolutionary aspects of recombination in RNA viruses. *J. Gen. Virol.* **80**(Pt. 10), 2535–2543.