Reverse transcription–polymerase chain reaction has been used worldwide for the diagnosis of Norwalk-like virus (NLV) infection, yet a commonly accepted genetic classification scheme has not been established. Amino acid sequences from four regions of open-reading frame 2 (ORF2) were used to analyze 101 NLV strains, including 2 bovine strains. On the basis of this analysis, a genetic classification scheme is proposed that differentiates 99 human strains into 2 major genetic groups consisting of 5 and 10 genetic clusters, respectively. The 2 bovine strains constitute a newly defined third major genetic group composed of 2 putative clusters represented by each strain. This classification scheme is well supported by the analysis of the entire ORF2 sequences from 38 strains selected to represent the genetic diversity of the human strains used above. This scheme should provide a firm scientific basis for the unified classification of NLV strains detected around the world.

"Norwalk-like viruses" (NLVs) are a genetically and antigenically diverse group of viruses constituting 1 of the 4 genera of the family Caliciviridae [1]. The viruses in this group are the major cause of outbreaks of acute nonbacterial gastroenteritis worldwide [2–8]. Application of reverse transcription–polymerase chain reaction (RT-PCR) and DNA sequencing techniques to detect and characterize NLVs has markedly enhanced our understanding of the epidemiology of NLV infection [4, 7, 9–12]. Currently, RT-PCR is used around the world as a tool for the routine diagnosis of NLV infection; however, primers used in RT-PCR differ between laboratories both in product size and detection efficiency [13]. In addition, methods used in genetic characterization of detected NLV strains differ between laboratories both in criteria for strain classification and naming of classification units. This has produced confusion among researchers working on NLV infections and makes it difficult to compare strains detected in different laboratories. A commonly accepted scheme is needed to genetically classify NLV strains and standardize NLV classification. Such a scheme should be consistent with antigenic classification and properly represent the remarkable genetic and antigenic diversity of NLV strains.

For several years at the Centers for Disease Control and Prevention (CDC), we have been conducting studies on the genetic and antigenic characterization of NLV strains in an attempt to establish a unified scheme that classifies NLV strains both genetically and antigenically. In the course of this project, we recently reported a genetic classification of NLV strains based on a 277-nucleotide (nt) region in the second open-reading frame (ORF2); we also demonstrated good correlation between serum IgG responses in patients with gastroenteritis and genetic clusters of the strains infecting the patients [14]. This work was followed by amplification, cloning, and sequencing of a 3-kb region from the RNA polymerase region to the 3' poly(A) tail from 24 strains selected from a subset of strains genetically characterized by the preceding step of this project [15, 16]. With the progress toward the genetic and antigenic characterization of NLV strains, our understanding of the classification has evolved.

Herein we briefly review this progress at CDC. We also examine, in light of defined criteria for genetic classification of NLV strains, our current results and those of other laboratories for which sequence information is available in GenBank.

Definition of Genetic Cluster, Genogroup, and Prototype and Reference Strains

The term "genetic cluster" as used here represents a minimum classification unit consisting of strains that reproducibly group together on a distinct branch of a phylogenetic tree and are sufficiently close in either amino acid (aa) or nt sequence to be distinguished from strains falling outside the group. The term "genogroup" represents a minimum classification unit consisting of the genetic clusters that reproducibly group together on a distinct branch of a phylogenetic tree and are sufficiently close in both aa and nt sequences to be distinguished from genetic clusters falling outside the group. Two genogroups have been described, genogroup I (GI) and genogroup II (GII) [17, 18].
The clusters are named using the abbreviation of genogroup, followed by consecutive numbers in the order of the publication of the “prototype strains” for the clusters.

The term “genetic reference strain” refers to a strain for which sequence information of the entire ORF2 is available in GenBank. The prototype strain of a genetic cluster refers to the first reference strain that was described as being genetically distinct from other reference strains. At present, there are 14 genetic reference strains, of which 8 are referred to as the prototype strains (table 1). These include the following viruses: Norwalk (NV), Hawaii (HV), Snow Mountain (SMV), Jena (JV), Southampton (SOV), Toronto (TV), Desert Shield (DSV), and Bristol (BV). In cases where the entire ORF2 sequence of the prototype strain has not yet been published, the outbreak number of the candidate prototype strain used in this paper has been indicated.

The term “antigenic reference strain” refers to 9 strains previously classified into 4 antigenic types, designated UK1, UK2, UK3, and UK4, on the basis of solid-phase immune electron microscopy (SPIEM) [37]. The prototype strains for these SPIEM-based antigenic types had been reported as Taunton (UK1) [38], NV (UK2), HV (UK3), and SMV (UK4). The 9 antigenic reference strains used in this paper were coded by antigenic type and identification number: UK1-1, UK1-2, UK1-3, UK2-12, UK3-14, UK3-15, UK3-17, UK4-20, and UK4-22 [39]. Three of these 9 strains (UK1-1, UK2-12, UK4-22) originated in outbreaks in the United States [40], and the remaining 6 strains originated in outbreaks and sporadic cases in the United Kingdom [41]. These strains were used to examine the relationship between the SPIEM-based antigenic types and the genetic clusters described in this paper.

### Historic Background of the Reference Strains

In the 1970s and 1980s, detection of NLVs was largely based on direct electron microscopy [42], while typing of the detected NLVs relied solely on immunologic methods involving human clinical samples as the source of antigens and antibodies [21, 43–50]. Since neutralization tests are not applicable for classification of this virus group because of the lack of an in vitro cultivation system, researchers referred to the results of cross-challenge studies in volunteers conducted in the United States as the primary basis for antigenic typing of NLV strains. In the early 1970s, the results of two cross-challenge studies suggested the existence of at least 2 distinct antigenic types, represented by NV and HV [19, 51], which were confirmed shortly afterward by using immune electron microscopy [20]. Ten years later, SMV was found to represent a third antigenic type [21]. In subsequent studies on the antigenic typing of NLVs, these 3 strains have been used commonly as the antigenic reference strains.

In 1990, Okada et al. [52] classified strains detected in outbreaks and sporadic cases of gastroenteritis in Japan into 9 antigenic types by immune electron microscopy, which showed various degrees of antigenic relatedness to NV, HV, and Otofuke virus. One year later, in the United Kingdom, Lewis [37] defined the 4 SPIEM-based antigenic types described above. Although Okada’s classification scheme was the most compre.

### Table 1. Genetic reference strains used in this article, as first classified by electron microscopy (EM), immune EM (IEM), or molecular methods—by year of publication.

<table>
<thead>
<tr>
<th>Year</th>
<th>First publication based on</th>
<th>Reference</th>
<th>Molecular methods</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1972</td>
<td>Norwalk virus (NV)</td>
<td>[19]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1977</td>
<td>Hawaii virus (HV)</td>
<td>[20]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1982</td>
<td>Snow Mountain virus (SMV)</td>
<td>[21]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1987</td>
<td>Jena virus (JV)</td>
<td>[22]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td>SRSV-KY89/89/Jpn (KY89)</td>
<td>[24]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1993</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1994</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1994</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1994</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1994</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1994</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1994</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE. Bold indicates 8 strains referred to as the prototype strains that represent individual clusters. GenBank accession nos. follow: NV, M87661; HV, U07611; SMV, L23831; JV, AJ011099; KY89, L23828; OTH25, L23830; SOV, L07418; TV, U02030; DSV, U04469; BV, X76716; LV, X65557; MXV, U22498, MKV, X81879; CAV, AF145896. SRSV = small round structured virus.
hensive at the time, the strains characterized by Lewis’s typing scheme were more frequently used as reference strains for developing molecular diagnostics during the 1990s [13, 39, 53–56]. However, both of the proposed antigenic classification schemes had serious limitations in accuracy and reproducibility in other laboratories. These limitations were primarily associated with the cross-reactive antibody responses often observed in individuals [14, 52, 57–59]. Nevertheless, the results obtained in the early studies suggested two main characteristics of NLVs: the presence of considerable antigenic diversity among strains and a worldwide distribution of strains with similar antigenicity.

In addition to human enteric NLV strains, NLV strains have been detected from calves and pigs. Three viruses morphologically resembling NV and causing diarrheal disease have been isolated from calves by serial passage. Two of these, Newbury agents SVR-1 and SVR-2, were isolated in the United Kingdom in 1978 and shown to be antigenically distinct from each other by a cross-challenge study in gnotobiotic calves in 1984 [60, 61]. SVR-2 was later redesignated Newbury agent 2 (NA-2). The third virus, JV, was originally isolated in Germany in 1980 [22]. These 3 bovine enteric viruses had only been candidate NLV strains because of the absence of definitive sequence evidence linking them to the family Caliciviridae. In 1999, the entire genome of JV and a part of the genome of NA-2 were sequenced, and both of these bovine enteric viruses were reported as belonging to GI on the basis of the phylogenetic analysis of these sequences [36, 62]. The phylogenetic relationship between JV and NA-2 has not been reported.

For the swine enteric NLV strains, Sugieda et al. [63] in Japan detected NLV sequences in the caecum contents of healthy pigs [63]. These sequences were reported as belonging to GII on the basis of a 292-nt region in the RNA polymerase gene. However, the sequence in an ORF2 region of the swine strains has not been determined.

Molecular Diagnosis of NLV Infections

Although the early antigenic typing experiments had predicted a genetically diverse range of NLV strains in circulation, in reality this diversity was much greater than anticipated. This observation led many researchers to use an almost identical region of the genome for molecular diagnosis, the RNA polymerase region involving the GLPSG and YGDD motifs, where amino acids are highly conserved within the family Caliciviridae [18, 23, 25, 29, 64]. Figure 1 illustrates the relative location of the primer sets and RT-PCR products that have been used in...
most diagnostic laboratories worldwide [4, 12, 13, 17, 18, 53–56, 65–81]. Many amplified products derived from these primers have been sequenced, and the sequences have been deposited in GenBank. Unfortunately, they overlap in only a 63-base region near the YGDD motif.

The primers used at CDC for routine diagnosis of NLV infection amplify one of the smallest regions among those used around the world, a 123-nt region with 81 unique nt when primer sequences are excluded [39]. These primers were designed to broadly detect strains previously classified into the 4 SPIEM-based antigenic types. Subsequent application of this RT-PCR method to the outbreak investigations in the United States has allowed us to detect many genetically diverse NLV strains [6, 7, 9, 14]. On the basis of the analysis of the 81-nt sequences, we proposed, in 1995, an interim scheme to classify NLVs into 4 genetic groups, P1-A, P1-B, P2-A, and P2-B [39]. The criterion for this classification of strains was to estimate the genetic relatedness of strains to those of the SPIEM-based antigenic types rather than to accurately determine their phylogenetic relationship because the histogram showing frequency distribution of all possible sequence pairs of distances in this 81-nt region does not differentiate the distances within clusters (intracluster distances), between clusters (intercluster distances), and between genogroups (intergenogroup distances) (data not shown).

**Genetic Classification of NLVs Based on Four Regions in ORF2**

To establish a commonly acceptable genetic classification scheme requires that the scheme maintain its consistency throughout genetic and antigenic classification of NLV strains. This has been anticipated from phylogenetic analysis of sequences based on a region in ORF2. However, the sequences in ORF2 regions are more divergent than those in the RNA polymerase region and have proven to be difficult to amplify.

The regions in ORF2 have been routinely used in only a few laboratories for analysis of NLV strains in circulation [7, 14, 29, 82]. We used a region 277 bases in length and located between nt 312 and 588 from the 5’-end of ORF2 in the equivalent location of the Lordsdale virus (LV) genome. This region is composed of 2 conserved areas, to which primers were directed, spanning a variable region from which the differentiation of sequences was expected. Previously, we classified 41 NLV strains into the 2 genogroups, GI and GII, and further differentiated them into 10 genetic clusters on the basis of the phylogenetic analysis of this region; however, we did not define the criteria for their classification nor evaluate the results [14]. Similar results were reported by Green et al. [82], who used a 223-nt region in ORF2 from 20 NLV strains, including 9 detected in the United Kingdom.

In our current paper, using the same region and the newly defined criteria of genetic classification described above, we analyzed a total of 91 strains consisting of 68 strains from outbreaks that occurred in the United States between 1993 and 1999 and the 14 genetic and 9 antigenic reference strains described above. The 68 US outbreak strains were selected to represent the genetic diversity of NLV strains detected in ~300 outbreaks. Upon alignment of the 91 sequences, gaps of 6 and 3 nt were found between the 2 genogroups and between individual clusters, respectively. The resulting aligned sequences consisted of 283 nt corresponding to 94 deduced amino acids.

To evaluate the results of the analysis based on the 283-nt region, we compared the results with those obtained from the analysis of the entire ORF2 sequences (575-aa sequences after an alignment, corresponding to the aligned 1725-nt sequences) from 38 strains composing a subset of the 90 human strains used in the 283-nt sequence analysis. These 38 strains included 23 from US outbreaks, 1 from a UK outbreak, and the 14 genetic reference strains. Of these, 24 (excluding the genetic reference strains) were previously amplified, cloned, and sequenced in a 3-kb region from the RNA polymerase to the 3’-poly(A) tail [15, 16].

In addition to the sequences from the US outbreak strains, we analyzed a 349-nt sequence (corresponding to 116 aa; the equivalent location from the 5’-end of ORF2 in LV, 79–427 nt) from the bovine strain NA-2 and the 175-nt sequence (corresponding to 58 aa; the equivalent location from the 5’-end of ORF2 in LV, 300–474 nt) from the 9 UK outbreak strains reported by Green et al. [82], which were available in GenBank.

For analysis of the sequences, we constructed phylogenetic trees, using the neighbor-joining method for tree construction and minimum evolution as a distance correction method [83]. To evaluate the reproducibility of our genetic clusters and genogroups, bootstrap analysis was conducted to determine the statistical significance of a consensus tree based on 100 replicates.

To evaluate the consistency in estimating the genetic similarity of strains and clusters, a histogram showing the frequency distribution of pairwise distances was used to determine the differentiability of the intrACLuster, intercluster, and intergenogroup distances based on Kimura and Tamura distance methods for aa and nt sequences, respectively [83]. According to the definition described above, a single strain could represent a putative cluster or genogroup if it was clearly differentiable from other strains by both bootstrap values and distance scores.

**Analysis of the 91 NLV Strains on the Basis of the Deduced 94-aa and 283-nt Sequences**

The results of phylogenetic analysis of the 94-aa sequences provided the foundation for the genetic classification of NLV strains proposed in this paper. According to this classification scheme, the 91 strains segregated into 3 genogroups (GI, GII, and GIII) and further differentiated into a total of 15 clusters. These clusters were named by using consecutive numbers for
each genogroup as described above: GI/1–GI/5, GII/1–GII/9, and GIII/1 (figure 2, table 2). Although prototype names are not used in our naming system, for convenience we have indicated either the prototype strains or outbreak numbers of candidate prototype strains in parentheses after the cluster name: the 5 GI clusters, GI/1 (NV), GI/2 (SOV), GI/3 (DSV), GI/4 (Cruise Ship virus [CSV]), and GI/5 (318); the 9 GII clusters, GII/1 (HV), GII/2 (SMV), GII/3 (TV), GII/4 (BV), GII/5 (290), GII/6 (269), GII/7 (273), GII/8 (539), and GII/9 (378); and the 1 GIII cluster, GIII/1 (JV). Sequence in the 94-aa region was not available for strains that formed the putative clusters GII/10 (Yat/94/UK) and GIII/2 (NA-2). These 2 clusters were defined using sequences available from alternative regions, as described below.

Analysis of the 94-aa sequences. Since JV could not be classified on the basis of the analysis of the 94-aa and the 283-nt sequences alone, its classification was discussed in a separate section. The 90 strains of human origin differentiated into 2 major genogroups (GI and GII) and further segregated into 5 and 9 clusters, respectively (figure 2). Nine of the 14 clusters corresponded to those we previously called NV-, SOV-, CSV-, DSV-, TV-, LV-, SMV-, WR- (White River virus), and HV-clusters [14]. Of the remaining 5 clusters, 3 are newly defined in this study: GI/5 (318), GII/8 (539), and GII/9 (378). The 2 remaining clusters were composed of 6 strains detected in outbreak 273, 269, 274, 292, 330, and 363. Previously, we classified all 6 of these strains into a cluster named GV, which referred to the strain (Gwynedd virus) from outbreak 273 [7, 14]. However, in the current analysis the strain from outbreak 273 was segregated from 5 other strains forming a putative new cluster. This discrepancy was due to the fact that our previous classification was based on an analysis of the 81-nt sequence in

Figure 2. Phylogenetic tree based on 94-amino acid sequences from 91 strains. Consensus phylogenetic tree was constructed by PP Search software (Wisconsin Package, version 10; Genetics Computer Group, Madison, WI) with optimality criterion set to distance (minimum evolution). Bootstrap values are indicated as % of 100 replicates based on neighbor-joining and mean distance methods. Tree is unrooted. Genetic clusters are enclosed by circles. Prototype and candidate prototype strains for each cluster are highlighted in bold. Viruses: SMV = Snow Mountain, HV = Hawaii, JV = Jena, SOV = Southampton, NV = Norwalk, DSV = Desert Shield, BV = Bristol, LV = Lordsdale, TV = Toronto, MXV = Mexico. UK1–4 = antigenic types determined on basis of solid-phase immune electron microscopy. GI–III = genogroups I to III, respectively.
Table 2. Genetic classification scheme, by genogroup, as determined on the basis of analysis of open-reading frame 2 sequences.

<table>
<thead>
<tr>
<th>Genogroup</th>
<th>Genetic cluster based on Amino acid sequences</th>
<th>Nucleotide sequences</th>
<th>Strains used in establishment of scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>GI</td>
<td>1 1 NV, KY89</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 2 SOV</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 3A DSV</td>
<td>3B UK2-12,-,b 316*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4 4 184 (CSV),,- a 266, 277 *</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 5 318*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GII</td>
<td>1 1A HV, UK4-22,-,b 283,- a 295,- a 301,-,a 302,-,a 317,-,a 442,- a 448</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1B UK4-20,-,b 304,-,a 314,- a 315,-,a 323,-,a 393,-,a 396,-,a 534, 554</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 2 SMV, MKV, 403,-,a 416,- a 421,-,a 430,-,a 434*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 3 TV, OTH25, MXV, UK1-1,-,b UK1-2,-,b UK1-3,-,b 247,-,a 279,-,a 312,-,a 313,-,a 321,-,a 365,-,a 413,-,a 432,-,a 433,-,a 435,-,a 445,-,a 451,-,a 547, 564</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4 4 BV, LV, CAV, UK3-14,-,b UK3-15,-,b UK3-17,-,b 293,-,a 326,-,a 362,-,a 366,-,a 367,-,a 392,-,a 393,-,a 397,-,a 398,-,a 399,-,a 405,-,a 408,-,a 425,-,a 426,-,a 452,-,a 459,-,a 461,-,a 465,-,a 467,-,a 471,-,a</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 5 290 (WR),,- a 306*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6 6 269,-,a 274,-,a 292,-,a 330,-,a 363*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7 7 273 (GV),</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8 8 539, 263,-,a</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9 9 378,-,a 352*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 10 Yat/94/UK6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIII</td>
<td>1 1 JV</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 2 NA-2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE. NV = Norwalk virus; SOV = Southampton virus; DSV = Desert Shield virus; CSV = Cruise Ship virus; HV = Hawaii virus; SMV = Snow Mountain virus; MKV = Melksham virus; TV = Toronto virus; MXV = Mexico virus; BV = Bristol virus; LV = Lordsdale virus; CAV = Camberwell virus; WR = White River virus; GV = Gwynedd virus; JV = Jena virus; NA-2 = Newbury agent-2. US outbreak strains, except for antigenic reference strains (see text), are indicated by outbreak nos. Bold indicates prototype and candidate prototype strains for individual clusters.

a [7].
b [39].
c [14].
d [82].
e [62].

ORF1, as we previously had been unable to amplify this strain in the 277-nt region of ORF2 [14]. Accordingly, the 2 clusters previously grouped together in the GV-cluster were designated GII/6 (269) and GII/7 (273), respectively, in the new classification scheme.

Four of the 14 genetic clusters described here were suggested to correlate with the SPIEM-based antigenic types because the 9 antigenic reference strains segregated into the 4 clusters according to their antigenic types: 3 UK1 strains into GII/3 (TV), 1 UK2 strain into GII/3 (DSV), 3 UK3 strains into GII/4 (BV), and 2 UK4 strains into GII/1 (HV) (figure 2). However, the prototype strains of these clusters differed from those originally reported for these antigenic types. The precise relationship between genetic and antigenic characteristics of these strains remained to be determined.

The reproducibility of all 11 clusters comprised of >1 strain (clusters 1, 3, and 4 in GI and clusters 1–6, 8, and 9 in GII) was well supported with high bootstrap values, ranging from 94% to 100% on the basis of 100 replicates (figure 2). In the phylogenetic tree, GI/1 (NV), GI/2 (SOV), GI/4 (CSV), and GI/5 (318) were more closely related to each other than to GI/3 (DSV). Relatively high bootstrap values (87%) supported the reproducibility of the phylogenetic relatedness of these 4 clusters. Similarly, GII/1 (HV), GII/2 (SMV), and GII/5 (WR) were closely related to each other, and the reproducibility of their phylogenetic relatedness was supported by relatively high bootstrap values of 79%. While the majority of clusters were largely monophyletic, GII/1 (HV), consisting of 16 strains, formed 2 subdivisions. One subdivision comprised 9 strains, including HV (bootstrap values, 65%), and the other comprised the remaining 7 strains (bootstrap values, 83%).

The histogram showing the frequency distribution of pairwise distances demonstrated 3 major peaks corresponding to the intracluster, intercluster, and intergenogroup distances (figure 3A). The intracluster distances within all 11 clusters comprised of >1 strain (clusters 1, 3, and 4 in GI and clusters 1–6 and 8, 9 in GII) were distributed from 0 to 10 Kimura distance scores (Kimura distances). Of note, the intracluster distances of all
Figure 3. Histogram showing frequency distribution of pairwise distances between 91 sequences including Jena virus (JV). A, Histogram based on 94-amino acid (aa) sequences. Symbols a, b, and c highlight distances within clusters (light gray bars, 0–10 Kimura distances), between clusters (gray bars, 11–45 Kimura distances), and between genogroups (black bars, 45–85 Kimura distances), respectively. Distances calculated by Kimura protein distance correction method are indicated as estimated no. of substitutions per 100 aa. B, Histogram highlighting frequency distribution of pairwise distances of 94-aa sequences between JV and 90 human strains (black bars, d, 48–77 Kimura distances). Gray bars, representing intercluster distances illustrated in A, are re-indicated to show boundary of their distribution. C, Histogram highlighting frequency distribution of pairwise distances of 288-nucleotide sequences between JV and 90 human strains (black bars, d, 47–75 Tamura distances). Gray bars represent intercluster distances attributed to 90 human strains (b, 18–52 Tamura distances). Distances were calculated by Tamura distance correction method with 2:1 weighting of transversion to transitions.

possible pairwise combinations between strains in the subdivisions of cluster GII/1 (HV) ranged from 6 to 9 Kimura distances, supporting our classification of these 2 subdivisions into a single cluster.

The intercluster distances were distributed from 11 to 45 Kimura distances and formed a large peak between 18 and 45 Kimura distances accompanied by a small peak between 11 and 16 Kimura distances. The ability to differentiate between intracluster and intercluster distances strongly supports our estimation of the putative clusters described in this paper.

The small peak of the intercluster distances was attributed to the intercluster distances both between strains of GII/1 (HV) and GII/2 (SMV) and strains of GII/2 (SMV) and GII/5 (WR). These results indicated that clusters 1 (HV), 2 (SMV), and 5 (WR) in GII were closely related to each other in the 94-aa region, consistent with the relatively high bootstrap values for these clusters in the phylogenetic tree.

The intergenogroup distances were distributed from 43 to 85 Kimura distances overlapping the intercluster distances in the range from 43 to 45 Kimura distances. The intercluster distances in this overlapping range were mainly attributed to the distances between strains of GII/4 (BV) and GII/8 (539) and between strains of GII/8 (539) and GII/9 (378), while the distribution of intergenogroup distances in this range was solely attributed to the pairwise distances between strains of GI/3 (DSV) and those of GI/6 (269). These results indicated that
clusters 4 (BV), 8 (539), and 9 (378) in GII were most distantly related within this genogroup, while GI/3 (DSV) and GII/6 (269) were most closely related upon comparison of GI and G2 clusters.

**Analysis based on the 283-nt sequences.** The pattern of strain clustering based on the 283-nt sequences was similar to that based on the 94-aa sequences (data not shown). The 90 strains of human origin fell into GI and GII, which were further divided into 6 GI and 10 GII clusters. However, 3 GI strains (DSV, UK2-12, and 316) that formed a single cluster, GI/3, in the phylogenetic tree based on the aa sequences differentiated into 2 clusters in the tree based on the nt sequences. In a similar manner, the 2 subdivisions described above for GII/1 (HV) based on the aa analysis now clearly differentiated into 2 distinct clusters on the basis of the nt sequence analysis. To reconcile the differences in the cluster formation between the aa and nt sequences, the 4 clusters based on the nt sequence analysis were designated GI/3A, GI/3B, GII/1A, and GII/1B in accordance with their relatedness to the clusters based on the aa sequences (table 2).

The results of the distance analysis based on the nt sequences were also similar to those obtained from the analysis of the aa sequences. The histogram demonstrated three major peaks corresponding to intracluster, intercluster, and intergenogroup distances (data not shown).

**Analysis of the 38 Strains Based on the Entire ORF2 Sequences**

Similar analysis of the entire ORF2 sequences from the 38 strains described above demonstrated that both phylogenetic trees based on the aa and nt sequences displayed the same strain clustering as indicated by the analysis based on the 94-aa sequences with the exception of GII/8 (539) and GI/9 (352), for which the entire ORF2 sequences have not been determined (data not shown). The histogram based on both 574-aa and 1725-nt sequences demonstrated three clearly separated peaks corresponding to the intracluster, intercluster, and intergenogroup distances, as observed in the histogram based on the 94-aa sequences (figure 4A, B). These results indicate that the classification of NLV strains based on the 94-aa sequence is consistent with that in the entire ORF2 region.

**Analysis of JV Based on the 94-aa (283-nt) and Entire ORF2 Sequences**

In the phylogenetic trees based on both 94-aa (figure 2) and 283-nt (data not shown) sequences, JV was more closely related to human strains in GI than in GII. The pairwise distances of the 94-aa sequences between JV and the 90 strains of human origin were distributed between 48 and 77 Kimura distances and did not overlap the intercluster distances (figure 3B). The frequency of distribution of these distances formed a large peak between 66 and 77 and a small peak between 48 and 56 Kimura distances. The large peak corresponded to the frequency of distribution of the pairwise distances of sequences between JV and GII strains, while the small peak corresponded to the distances between JV and GI strains. These results confirm the observation based on the phylogenetic tree that JV was more closely related to GI strains than to GII strains (figure 2). More important, however, the results indicated that JV does not belong to either GI or GII and thus, on the basis of the 94-aa sequence analysis, may represent a putative cluster of a new genogroup, genogroup III (GIII).

In contrast to the aa sequences, the pairwise distances of the 283-nt sequences between JV and the 90 human strains overlapped the intercluster distances in a range from 47 to 52 Tamura distances (figure 3C). This overlapping region was attributed to the pairwise distances between JV and GI strains. These results, based on the analysis of the 283-nt sequences, indicated that JV could be classified into GI.

Since different results regarding the classification of JV were obtained from the analyses of the distances based on the 94-aa and the 283-nt sequence, we analyzed JV using the entire ORF2 sequences from the 38 strains described above. JV displayed the same clustering pattern in both phylogenetic trees based on the aa and nt sequences in the entire ORF2 region as that observed in the tree based on the 94-aa sequences (data not shown). The histogram based on the 574-aa sequences clearly demonstrated that the pairwise distances between JV and the 38 human strains fell into the intergenogroup distances of the human strains and did not overlap the intercluster distances at all (figure 4A, B). More important, the same results were obtained from the distance analysis based on the 1725-nt sequences. While the intracluster, intercluster, and intergenogroup distances attributed to the human strains were distributed between 1 and 15, 25 and 53, and 72 and 88 Tamura distances, respectively, the distances between JV and the human strains were distributed between 88 and 111 Tamura distances and did not at all overlap the intercluster distances of the human strains (figure 4B). These results indicate that JV should represent a putative cluster in GIII. We therefore designated this cluster GIII/1.

**Analysis of the Genetic Relationship between JV and NA-2**

In a similar manner, we used the entire ORF2 sequences from the 38 strains as the source of the reference sequences to determine the phylogenetic relationship between NA-2 and JV. We did this because the reported 349-nt sequence of NA-2 overlapped the 277-base sequence only in a 116-nt region in which the histogram did not clearly differentiate the intracluster, intercluster, and intergenogroup distances. The analysis was conducted on a 349-nt region (corresponding to 116-aa) located between nt 79 and 427 from the 5’ end of ORF2 in the equivalent
Figure 4. Results of distance analysis of sequences in 3 different regions of open-reading frame 2 (ORF2). Outlined, light, and dark gray bars indicate frequency distribution of pairwise distances solely attributed to 38 human strains, with symbols a, b, and c highlighting intracluster, intercluster, and intergenogroup distances, respectively. Amino acid (aa) distances calculated by Kimura distance correction method (for A and C) and nucleotide (nt) distances calculated by Tamura distance correction method with 2:1 weighting of transversion to transitions (for B) are indicated as estimated no. of substitutions per 100 aa and 100 nt. A. Histogram showing frequency distribution of pairwise distances of entire ORF2 aa sequences (575-aa) between Jena virus (JV) and 38 human strains. Symbols a, b, and c correspond to 1–17, 23–51, and 73–95 Kimura distances, respectively. Black bars (d) indicate pairwise distances between JV and 38 human strains (73–95 Kimura distances). B. Histogram showing frequency distribution of pairwise distances of entire ORF2 nt sequences (1725-nt) between JV and 38 human strains. Symbols a, b, and c correspond to 1–15, 25–53, and 72–88 Tamura distances, respectively. Black bars (d) indicate pairwise distances between JV and 38 human strains (88–111 Tamura distances). C. Histogram showing frequency distribution of pairwise distances of 116-aa sequences from 2 bovine strains, JV and Newbury agent-2 (NA-2), and 38 human strains. Black bars (d) indicate distances between all possible combinations of JV and NA-2 with 38 human strains (35–62 Kimura distances). Arrow indicates distance between JV and NA-2 (14 Kimura distance score).

location of LV. A phylogenetic tree based on the 116-aa sequences indicated that NA-2 and JV formed two closely related yet distinct branches in the tree with a bootstrap value of 100% (data not shown).

The histogram showing frequency distribution of the pairwise distances between the 38 human strains indicated that the intercluster distances were distributed between 10 and 31 Kimura distances and formed a small peak between 10 and 18 and a large peak between 19 and 31 Kimura distances (figure 4C). This small peak was mainly attributed to the distances between GII/1 (HV), GII/2 (SMV), and GII/5 (WR), as observed in the histogram based on the 94-aa sequences. The Kimura distance between JV and NA-2 was 14 and fell into this small peak. These results indicate that NA-2 is closely related to but distinguishable from JV and thus represents a putative cluster that is different from that represented by JV. We designated this cluster GIII/2 (table 2). Of note, the pairwise distances of the 166-aa sequences between these 2 bovine strains and the 38
human strains were distributed from 35 to 62 Kimura distances and did not overlap the intercluster distances of the human strains, as observed in the histogram based on the 94-aa sequences.

**Analysis of an Additional 9 ORF2 Sequences Available in GenBank**

Finally, analysis of the 58-aa sequences from 9 UK outbreak strains reported by Green et al. [82] for their relationship with the 38 human strains used above indicated that a single strain, Yat/94/UK, represented a putative new cluster in GII (data not shown). We designated this cluster GII/10 (table 2).

Taken together, the results described in this paper lead us to propose a genetic classification scheme that differentiates NLV strains into 3 genogroups with a total of 17 clusters on the basis of the aa sequence. Fifteen of these clusters are indicated in figure 2, and 2 of the clusters are represented by NA-2 and Yat/94/UK (table 2).

**Conclusion**

In the 1970s and 1980s, typing of NLV strains relied solely on immunologic methods involving human clinical samples as the source of antigens and antibodies. These methods had serious limitations in accuracy and reproducibility and never provided a totally reliable scheme for the antigenic classification of NLV strains. In the 1990s, the availability of molecular techniques to amplify, sequence, and express the genome of NLV strains provided the tools necessary to characterize NLV strains both genetically and antigenically. Currently, the NLV strains that prevail in communities are genetically characterized in many laboratories, largely on the basis of RNA polymerase regions. In addition, a growing number of laboratories are beginning to antigenically characterize NLV strains, using specific antibodies produced by the recombinant-expressed capsid proteins [14, 28, 30, 84–89]. However, neither genetic nor antigenic classification schemes for NLV strains have been commonly accepted. Such acceptance will require a scheme that maintains its consistency throughout the genetic and antigenic classification of NLV strains. It is also important that the scheme properly represent the remarkable genetic and antigenic diversity of NLV strains. Moreover, the scheme must be based on defined criteria for classification that can be practically evaluated.

In this brief review, we attempted to establish an interim scheme to genetically classify NLV strains by defining the minimum classification units based on the clustering pattern of strains in a phylogenetic tree and the pairwise sequence distances of the strains. These two criteria were applied to the analysis of a total of 101 strains in four regions of ORF2 (283-nt, 1725-nt, 344-nt, and 223-nt regions). The results were evaluated using the bootstrap values and a histogram showing frequency distribution of pairwise distances of the sequences.

Throughout these analyses, we obtained consistent results of strain clustering not only in the partial regions of ORF2 but also in the entire ORF2 region. This consistency allowed us to confirm the validity of our definition and evaluation of the genetic clusters as the criteria for typing of NLV strains. As a result, 99 human strains were classified into the 2 major genogroups, GI and GII, each of which consisted of 5 and 10 genetic clusters, respectively. Furthermore, we found that the 2 bovine strains could be classified into 2 putative clusters in a new genogroup, GIII.

The classification scheme of NLV strains proposed in this paper reflects the genetic diversity of the strains responsible for outbreaks in the United States. The 68 US outbreak strains analyzed here were selected to represent the genetic diversity of the strains detected in ~300 US outbreaks during 1993 to 1999 and screened on the basis of the 81-nt region in the RNA polymerase gene [7, 14, 39]. This classification scheme was further extended by including 13 of the 14 genetic reference strains selected to represent strains in circulation at different periods and in various geographic locations, 6 of the 9 antigenic reference strains screened by SPIEM to represent strains responsible for outbreaks and sporadic cases in the United Kingdom, an additional 9 UK outbreak strains detected between 1992 and 1995 [82], and 2 bovine strains that may represent outbreaks of gastroenteritis in calves [60].

This is an interim scheme, and the numbers of the genetic clusters will undoubtedly increase, but the main features for the genetic classification scheme proposed here may remain unchanged. This typing scheme should facilitate the designation and evaluation of new detection methods needed for standardized diagnosis of NLV infection. The development of such a standardized diagnostic method will play a crucial role in establishing an international surveillance network to rapidly identify international outbreaks of NLV-related gastroenteritis and to elucidate the mode(s) of transmission of the viruses [7]. The next step to establish a unified scheme for both genetic and antigenic classification of NLV strains will be to accurately determine the relationship between the genetic clusters and antigenic types of NLVs. For this effort, expressed capsid proteins from the strains representing the majority of the genetic clusters described here will be needed along with hyperimmune animal sera directed toward these proteins.

**Acknowledgments**

We gratefully acknowledge R. Glass and S. Monroe for fruitful discussion and C. Kirkwood for carefully reading the manuscript.

**References**

42. Caul EO. Viral gastroenteritis: small round structured viruses, caliciviruses.