JVI Journals.ASM.org

Covariation of Major and Minor Viral Capsid Proteins in Norovirus Genogroup II Genotype 4 Strains

Martin Chi-Wai Chan,[a] Nelson Lee,[b] Wing-Shan Ho,[a,b] Carmen Oi-Kwan Law,[a] Terrence Chi-Kong Lau,[c,d] Stephen Kwok-Wing Tsui,[c] and Joseph Jao-Yiu Sung[a,b]

Institute of Digestive Disease, The Chinese University of Hong Kong,[a] Department of Medicine and Therapeutics, The Chinese University of Hong Kong,[b] School of Biomedical Sciences, The Chinese University of Hong Kong,[c] and Department of Biology and Chemistry, City University of Hong Kong,[d] Hong Kong Special Administrative Region, People's Republic of China

We report sequence hypervariability in the viral protein 1 (VP1) interaction domain of VP2 in the norovirus (NoV) genogroup II genotype 4 (GII.4) lineage on 3 levels: (i) the global evolution of pandemic/epidemic strains from the mid-1970s through post-2006, (ii) the local emergence of an epidemic strain, and (iii) an immunocompromised patient chronically shedding NoV. When a quantitative yeast two-hybrid assay was used, VP2 was found to interact with VP1 in a time-ordered, strain-dependent manner among 3 NoV GII.4 strains. Our findings suggest that VP1 and VP2 may covary in virus evolution and that sequence hypervariability of VP2 may be functionally driven. Further investigations are warranted.

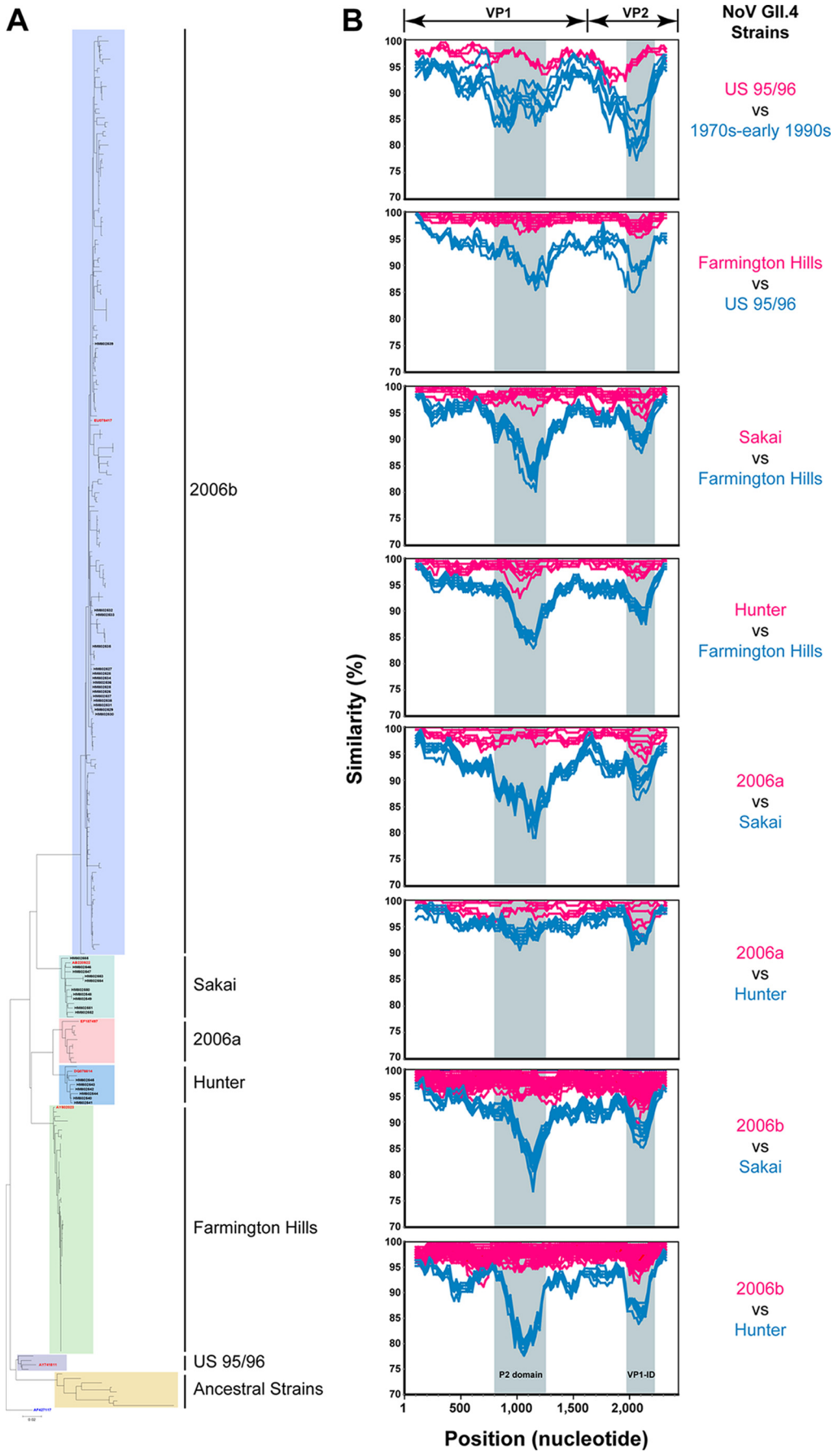
Human noroviruses (NoVs) are major causative agents of acute viral gastroenteritis and affect all age groups worldwide. Periodic emergence of widely circulating novel NoVs that belong to the genogroup II genotype 4 (GII.4) lineage has been demonstrated (2, 3, 15, 18). The capsid of NoVs is composed of two structural viral proteins: major viral protein 1 (VP1) and minor VP2 (13). Extensive epidemiological, phylogenetic, and clinical studies have shown that the protruding domain 2 (P2 domain) of VP1 is rapidly mutating and reflects continuous viral evolution of NoV GII.4 strains (reviewed in reference 7). In contrast, there are only a paucity of studies on VP2, and the question of whether this minor structural protein is associated with the evolution of the NoV GII.4 lineage remains largely unexplored. In this study, we performed sequence analysis of NoV GII.4 strains on global, local, and individual levels for evidence of covariation of VP1 and VP2 genes. We also evaluated whether VP1 and VP2 interact in a strain-dependent manner.

First, we carried out *in silico* sequence analysis of NoV GII.4 VP1 and VP2 genes based on data publicly available (as of October 2010) in GenBank. In brief, only entries containing nucleotide sequences of complete VP1 and VP2 genes were retrieved and studied. Novel NoV GII.4 strains that have emerged since 2008 were not included since their epidemiology in local and global perspectives remains uncertain. Phylogenetic analysis was performed using the software program MEGA, version 5.05. Global nucleotide identity was computed using the software program MatGat, version 2.02. Nucleotide similarity was scanned by the SimPlot program, version 3.5.1. Computing parameters are shown in figure legends, where applicable. A total of 314 sequences were studied (see Table S1 in the supplemental material). A maximum-likelihood tree of the complete VP1 genes showed that they represented a panel of known NoV GII.4 strains, including ancestral strains from the mid-1970s to 1992 and other previous widely circulating strains (US 95/96, Farmington Hills, Hunter, Sakai, 2006a, and 2006b) (Fig. 1A). While global comparison of the complete VP1 and VP2 genes revealed an overall high level of nucleotide identity (>95%), scanning nucleotide similarity analysis exclusively identified 2 local hypervariable regions that were consistently observed between time-ordered NoV GII.4 strains: the P2 domain of VP1 and nucleotide positions 300 to 600 (approximate) of VP2 (Fig. 1B). Specifically, pairwise nucleotide similarity of the 2 hypervariable regions decreased to a local minimum of 77 to 90% for both the VP1 and VP2 genes. Notably, homology alignment against the VP2 amino acid sequence of Norwalk virus (GenBank accession number M87661) suggest that this VP2 hypervariable region may map to the VP1 interaction domain of the virus (9). Molecular docking simulation also suggested that in the NoV GII.4 lineage, the VP2 hypervariable region may interact with VP1 (see Fig. S1).

Second, we studied the 3′ half of NoV GII.4 genomes (including the VP1 and VP2 genes) obtained locally from 33 adult patients diagnosed with NoV GII.4 infections during a 2-year period (November 2004 to November 2006). One stool specimen was collected from each patient. Complete nucleotide sequences of the VP1 and VP2 genes were determined (see Methods in the supplemental material). A maximum-likelihood tree of the complete VP1 gene showed that our viruses clustered into 3 distinct strains within the NoV GII.4 lineage according to year of collection (Fig. 2A): for those collected in 2004 and 2005, 6 belonged to the Hunter and 10 to the Sakai strains; of the remaining 15 viruses collected in 2006, all belonged to the 2006b strain. This suggests the Hunter and Sakai strains cocirculated in 2004/2005, before the emergence of the 2006b strain in 2006, and agrees well with findings of other local epidemiological studies (10, 11). Despite the overall high level of sequence homology, scanning nucleotide similarity analysis identified 2 local hypervariable regions in the 2006b strain, similar to observations of the global evolution of NoV GII.4 strains: the P2 domain of VP1 and the VP1 interaction domain of VP2, in which pairwise nucleotide similarity decreased to a local

Received 1 February 2011 **Accepted** 18 October 2011
Published ahead of print 9 November 2011
Address correspondence to Martin Chan, martin.chan@cuhk.edu.hk.
Supplemental material for this article may be found at http://jvi.asm.org/.

doi:10.1128/JVI.00228-11


A
2006b
Sakai
2006a
Hunter
Farmington Hills
US 95/96
Ancestral Strains
B
VP1
VP2
NoV GII.4 Strains
US 95/96 vs 1970s-early 1990s
Farmington Hills vs US 95/96
Sakai vs Farmington Hills
Hunter vs Farmington Hills
2006a vs Sakai
2006a vs Hunter
2006b vs Sakai
2006b vs Hunter
Similarity (%)
100
95
90
85
80
75
70
P2 domain
VP1-ID
1
500
1,000
1,500
2,000
Position (nucleotide)

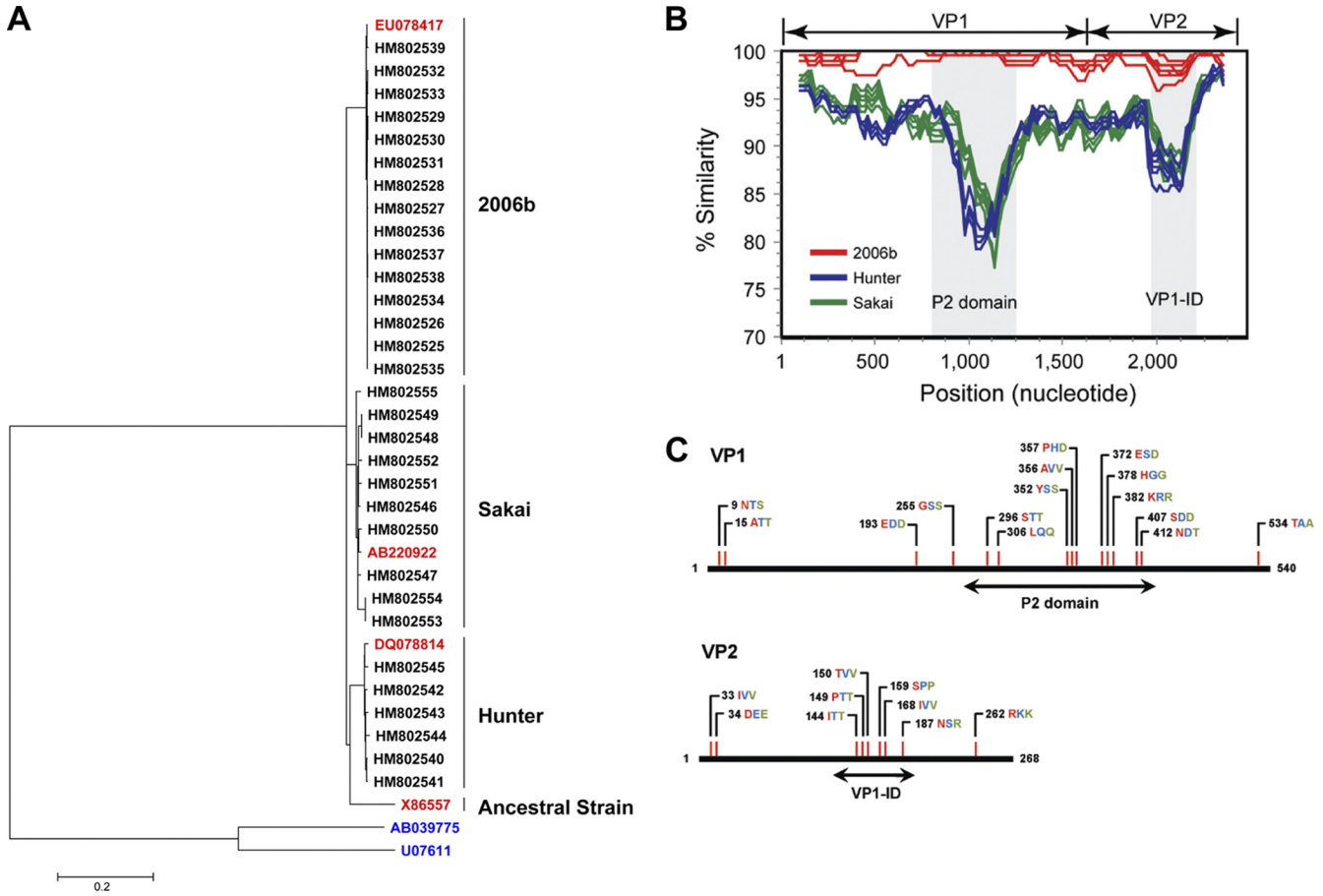


FIG 2 Hypervariable regions in protruding domain 2 (P2 domain) of viral protein 1 (VP1) and the VP1 interaction domain (VP1-ID) of VP2 in the local emergence of the epidemic norovirus genogroup II genotype 4 (NoV GII.4) 2006b strain. (A) Maximum-likelihood tree of the complete VP1 gene, showing major clusters of local sequences within the NoV GII.4 lineage. The phylogenetic tree was built as described in the Fig. 1 legend. Shown are GenBank accession numbers; all sequences obtained in this study are shown in black; reference strains and outgroup genotypes (GII.1 and GII.12) are shown in red and blue, respectively. (B) Scanning nucleotide similarity plot of the 3′ genome half of the 2006b strain against those of previously widely circulating local Hunter and Sakai strains. Analyzed regions include the complete VP1 and VP2 genes and the 3′ untranslated region (3′-UTR). Lines are color coded to denote different NoV GII.4 strains: 2006b (red), Hunter (blue), and Sakai (green). The P2 domain and VP1-ID are boxed in gray. Refer to the Fig. 1 legend for computing parameters. (C) Frequent unique amino acid substitutions in the P2 domain of VP1 and VP1-ID of VP2. Unique substitutions were defined as residues that were observed in the 2006b strain but not in the majority of both local GII.4 Hunter and Sakai strains. Residues are color coded to represent different NoV GII.4 strains as in panel B. Residue positions are numbered with reference to the 2006b strain. Panel C is not drawn to scale.

minimum of ~80% and ~85%, respectively (Fig. 2B). Localized nucleotide hypervariability translated into frequent unique amino acid substitutions in the corresponding regions (Fig. 2C). In VP1, 10 (67%) of the 15 amino acid changes were located in P2 domain; in VP2, 6 (67%) of the 9 residue changes were located in the hypervariable region.

Third, we examined *in vivo* virus evolution in an immunocompromised patient chronically shedding high levels of NoV (Fig. 3A). This patient was a 69-year old man with underlying agammaglobulinemia and a medical history of thymoma. Serial stool specimens were collected monthly from the patient over a 4-month period (July 2010 to November 2010). The complete nucleotide sequences of the noroviral VP1 and VP2 genes were determined (see Methods in the supplemental material). A

FIG 1 Hypervariable regions in protruding domain 2 (P2 domain) of VP1 and the VP1 interaction domain (VP1-ID) of VP2 in the global evolution of pandemic/epidemic norovirus genogroup II genotype 4 (NoV GII.4) strains from the mid-1970s through post-2006. (A) Maximum-likelihood tree of complete VP1 gene. Tree inference was performed using the software program MEGA, version 5.05. A general time-reversible model was used for nucleotide substitution, and substitution rates were assumed to be gamma distributed (G+I). The final tree was optimized using the heuristic nearest-neighbor-interchange method. The scale bar indicates the estimated number of nucleotide substitutions per site. GenBank accession numbers of reference NoV GII.4 strains, sequences deduced in the current study, and outgroups are shown in red, black, and blue, respectively. Sequences used are listed in Table S1 in the supplemental material. (B) Time-ordered scanning nucleotide similarity plots of the complete VP1 and VP2 genes. The P2 domain and VP1-ID are boxed in gray. The P2 domain was defined from amino acid residues 267 through 418 as previously described (3). VP1-ID was defined from residues 119 through 197 based on pairwise alignment against the VP2 amino acid sequence of Norwalk Virus (GenBank accession number M87661) (9). Residue positions are numbered with reference to the 2006b strain. The plots were generated using the program SimPlot, version 3.5.1, with the following computing parameters: window size = 200; step size = 20; distance model = Kimura 2 parameter.

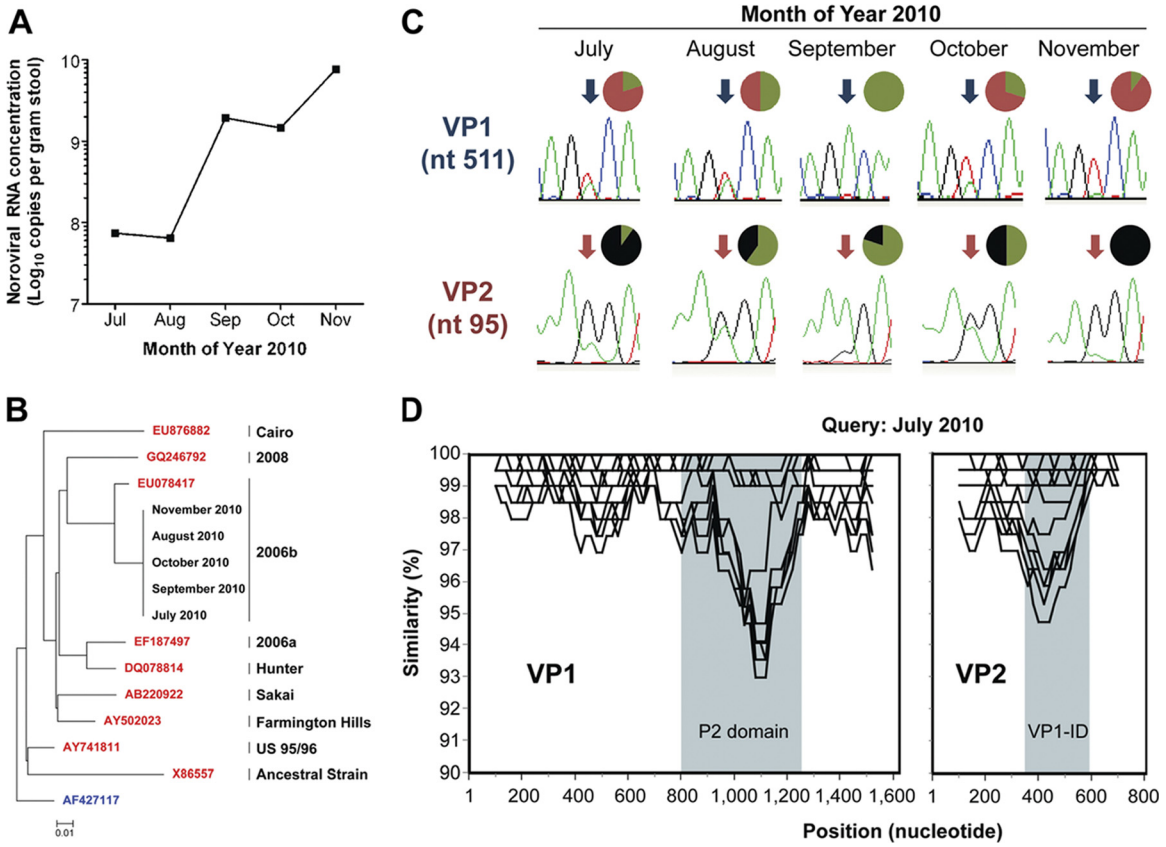


FIG 3 *In vivo* virus evolution in an immunocompromised patient chronically shedding a norovirus of genogroup II genotype 4 (NoV GII.4). (A) Fecal virus shedding profile during the study period from July 2010 to November 2010. Quantification of noroviral RNA from serial stool specimens was performed using a real-time reverse transcription-PCR assay targeting the polymerase/capsid gene junction as we have previously described (5). (B) Maximum-likelihood tree of the complete viral protein 1 (VP1) gene showing clustering of sequences within the NoV GII.4 2006b strain. Sequences were obtained from directly sequenced PCR products. The phylogenetic tree was built as described in the Fig. 1 legend. Sequences derived from the patient are shown in black with the month and year of specimen collection. GenBank accession numbers of reference strains and outgroups are shown in red and blue, respectively. (C) Presence and changing proportion of virus quasispecies in VP1 and VP2 genes. Data on one representative nucleotide position for each gene are shown. Chromatograms were obtained from directly sequenced PCR products. Insets (pie charts), quantitative measurement of virus quasispecies by counting 10 clones of completely sequenced cloned PCR products from each time point. nt, nucleotide. (D) Scanning nucleotide similarity plots of complete VP1 and VP2 genes. Lines represent individual clones ($n = 50$). Protruding domain 2 (P2 domain) of VP1 and the VP1 interaction domain (VP1-ID) of VP2 are boxed in gray. Refer to the Fig. 1 legend for computing parameters.

maximum-likelihood tree of the complete VP1 gene suggested that this patient was a chronic shedder of the NoV GII.4 2006b strain (Fig. 3B). Over a time period of 4 months, both the VP1 and VP2 genes of the virus were rapidly mutating, as shown by the presence and changing proportions of virus quasispecies (Fig. 3C). PCR products were also cloned into the pCR2.1-TOPO vector (Invitrogen) for further analysis. Ten clones were picked and sequenced for each PCR product. Clonal nucleotide sequences are available upon request. Importantly, scanning nucleotide similarity analysis of individual clones identified the same set of 2 local hypervariable regions observed in virus global evolution and local emergence: the P2 domain of VP1 and the VP1 interaction domain of VP2, in which pairwise nucleotide similarity decreased to a local minimum of ~93% and ~95%, respectively (Fig. 3D). Furthermore, analysis of these clones showed that 38 and 15 nucleotides were continuously changing in the VP1 and VP2 genes, respectively (see Table S2 in the supplemental material), and 9 and 5 of them were missense, nonsynonymous changes, resulting in different amino acids (see Table S2). Notably, three codons were found to be under positive selection in each of the VP1 and VP2 genes as determined by the ratio of the number of nonsynonymous to synonymous substitutions (see Table S3). However, preliminary analysis using the CAPS server (coevolution analysis using protein sequences; available at http://bioinf.gen.tcd.ie/caps) found no coevolving residues between VP1 and VP2. This may be attributed to the short time period (spanning 5 months only) covered. Cross-comparison of nucleotide variations among global, local, and individual data sets revealed few overlapping variations (data not shown).

Finally, since sequence comparisons consistently revealed a hypervariable region (VP1 interaction domain) in VP2, we sought to know whether the observed hypervariability could be functionally driven. Using qualitative yeast two-hybrid approach, we confirmed VP1/VP2 interactions among 3 NoV GII.4 strains (Hunter, Sakai, and 2006b) (Fig. 4A; see also Methods in the supplemental material). Further quantitative β-galactosidase liquid culture assay demonstrated time-ordered, strain-dependent VP1/VP2 interactions (Fig. 4B; see also Methods in the supplemental mate-

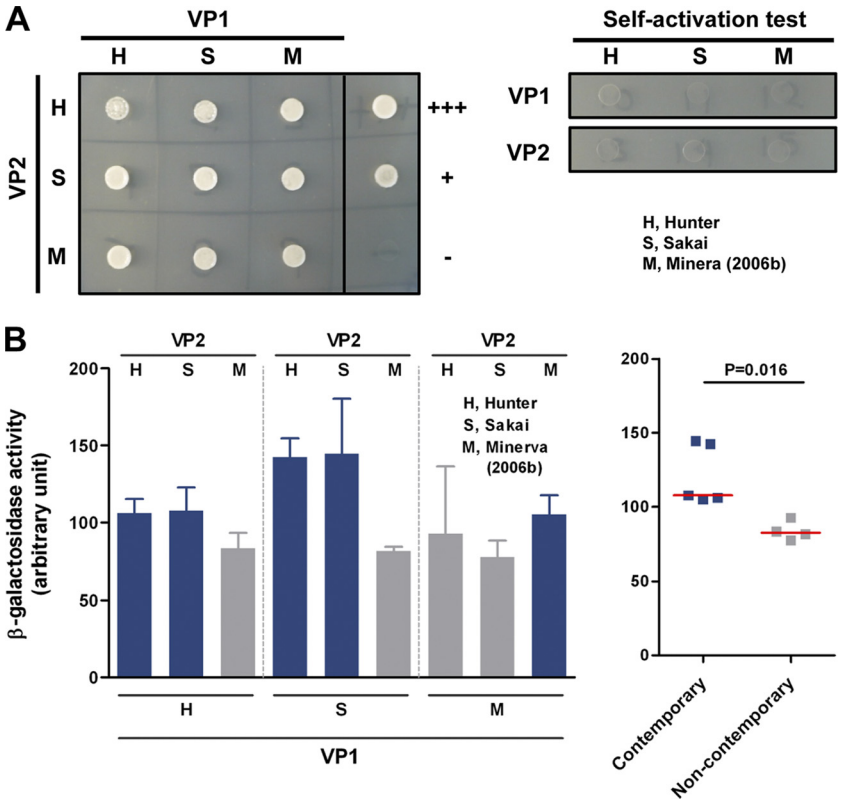


FIG 4 Physical protein-protein interaction between viral protein 1 (VP1) and VP2 among 3 human NoV GII.4 strains. (A) VP1 interacts with VP2 in qualitative yeast two-hybrid assay (left). Self-activation tests of individual yeast two-hybrid constructs are shown on the right. Shown are yeast colonies on selection agar plates (without leucine, tryptophan, and histidine) supplemented with 100 mM 3-amino-1,2,4-triazole 5 days posttransformation. The last column in the left panel denotes controls of yeast two-hybrid assay (+++, strong interaction control; +, weak interaction control; −, negative interaction control). (B) Time-ordered, strain-dependent VP1/VP2 interactions in quantitative yeast two-hybrid β-galactosidase liquid culture assay. Contemporary and noncontemporary VP1/VP2 combinations are in blue and light gray, respectively. Left and right panels contain the same set of data plotted in different formats. Error bars denote standard deviations. Horizontal red lines indicate median levels of β-galactosidase activity. H, Hunter strain; S, Sakai strain; M, Minerva strain.

rial). Specifically, VP1/VP2 interactions between contemporary strains (e.g., Sakai VP1 plus Hunter VP2 or Sakai VP1 plus Sakai VP2) were found to be stronger than those between noncontemporary strains (e.g., Sakai VP1 plus 2006b VP2) (Mann-Whitney *U* test, $P = 0.016$).

In this study, we found sequence hypervariability in the P2 domain of VP1 and, more importantly, in the VP1 interaction domain of VP2 in NoV GII.4 lineage on 3 levels: (i) in the global evolution of pandemic/epidemic strains from the mid-1970s through post-2006, (ii) in the local emergence of an epidemic strain, and (iii) in an immunocompromised patient chronically shedding NoV. Our findings agree with observations in earlier work that the VP2 gene is highly variable (based on multiple sequence alignment) (2, 4, 14). It is noteworthy that our findings from scanning nucleotide analysis establish the link between VP2 sequence hypervariability and the putative VP1 interaction domain, hinting at coevolution of VP1 and VP2. A similar observation of VP1/VP2 covariation has been implicated based on comparable topologies of their phylogenetic trees (12). We acknowledge that the presence of the VP1 interaction domain was documented only in VP2 of Norwalk virus (GI.1 strain) by means of coimmunoprecipitation and virus-like-particle incorporation assays (9); whether a similar domain is present in VP2 of the NoV GII.4 strains remain elusive. However, our preliminary findings that VP1 interacts with VP2 among 3 NoV GII.4 strains and that such interactions were mediated *in silico* through the VP2 hypervariable region as demonstrated by molecular docking support that the hypervariable region in VP2 which we observed in the NoV GII.4 lineage could be regarded as a VP1 interaction domain. Confirmatory crystallographic studies determining three-dimensional structures of VP2 and the VP1/VP2 complex are necessary.

It remains unknown whether VP1 and VP2 covary independently or through a mechanism driven by one another. Accumulating evidence suggests that VP2 has important virologic functions accessory to those of VP1. First, this minor structural protein is associated with recombinant virus-like particles (rVLPs) of Norwalk virus (8). Second, VP2 was found to bind to VP1 through its VP1 interaction domain and is important in stablizing Norwalk Virus rVLPs (1). In feline calicivirus, VP2 also interacts with VP1 and is necessary to form rVLPs with morphology similar to that of native virions (6) and to produce infectious virus (17). Following along this line of thought, one possible explanation of VP2 hypervariability could be that VP1 interaction of VP2 covaries as driven by VP1 in an attempt to adapt to the changing conformation of VP1 (presumably as a result of changing the P2 domain). In this way, VP2 continues to work in concert with altered virus capsid. Our observation of time-ordered, strain-dependent VP1/VP2 in-

teractions demonstrated that the selection of the contemporary VP1/VP2 combination is important for maintaining the strength of VP1/VP2 interactions. Considering that only one hypervariable region was present in VP2, such hypervariability is likely to be functionally driven by covariation with VP1. Further experiments to study whether VP1/VP2 interaction may affect VP1 dimerization and self-assembly of virions are needed. Notably, earlier study has shown that an average of 1.5 molecules of VP2 were present in Norwalk virus rVLPs (8). How VP2 functions in such a trace quantity remains elusive. Alternatively, it has been shown that the 3′ ends of NoV GII RNA genomes are highly structured and may affect infectivity of the closely related murine NoV (16); whether sequence hypervariability of VP2 may have a similar virologic influence on the viral fitness of NoV GII.4 strains remains to be determined. Further laboratory investigations are warranted but would be difficult due to the absence of an *in vitro* culture systems for human NoVs.

In summary, we report sequence hypervariability in the VP1 interaction domain of VP2 in the NoV GII.4 lineage, hinting at covariation of VP1 and VP2 in virus evolution. Time-ordered, strain-dependent interaction of VP1 and VP2 implicates that sequence hypervariability of VP2 may be functionally driven and deserves further investigation.

Nucleotide sequence accession numbers. GenBank accession numbers for sequences determined in this work are listed in Fig. 1 and 2.

ACKNOWLEDGMENTS

M.C.-W.C. designed and coordinated the study, performed the sequencing work and the yeast two-hybrid assay, analyzed data, and drafted the manuscript; N.L. coordinated patient recruitment and specimen collection; W.-S.H. performed the sequencing work and the yeast two-hybrid assay; C.O.-K.L. performed the sequencing work; T.C.-K.L. and S.K.-W.T. performed the molecular docking simulation; and J.J.-Y.S. designed and supervised the study. All authors read and commented on the manuscript. M.C.-W.C. has access to all study materials and data and is responsible for the integrity of the study.

This project forms part of a series of studies commissioned by the Food and Health Bureau of the Hong Kong SAR Government and was funded by the Research Fund for the Control of Infectious Diseases (CU-09-02-04; to J.J.-Y.S.). This project was also partly supported by the Institute of Digestive Disease, The Chinese University of Hong Kong (to J.J.-Y.S).

REFERENCES

1. **Bertolotti-Ciarlet A, Crawford SE, Hutson AM, Estes MK.** 2003. The 3′ end of Norwalk virus mRNA contains determinants that regulate the expression and stability of the viral capsid protein VP1: a novel function for the VP2 protein. J. Virol. **77**:11603–11615.
2. **Bok K, et al.** 2009. Evolutionary dynamics of GII.4 noroviruses over a 34-year period. J. Virol. **83**:11890–11901.
3. **Bull RA, Eden JS, Rawlinson WD, White PA.** 2010. Rapid evolution of pandemic noroviruses of the GII.4 lineage. PLoS Pathog. **6**:e1000831.
4. **Cauchi MR, Doultree JC, Marshall JA, Wright PJ.** 1996. Molecular characterization of Camberwell virus and sequence variation in ORF3 of small round-structured (Norwalk-like) viruses. J. Med. Virol. **49**:70–76.
5. **Chan MC, et al.** 2006. Fecal viral load and norovirus-associated gastroenteritis. Emerg. Infect. Dis. **12**:1278–1280.
6. **Di Martino B, Marsilio F.** 2010. Feline calicivirus VP2 is involved in the self-assembly of the capsid protein into virus-like particles. Res. Vet. Sci. **89**:279–281.
7. **Donaldson EF, Lindesmith LC, Lobue AD, Baric RS.** 2010. Viral shape-shifting: norovirus evasion of the human immune system. Nat. Rev. Microbiol. **8**:231–241.
8. **Glass PJ, et al.** 2000. Norwalk virus open reading frame 3 encodes a minor structural protein. J. Virol. **74**:6581–6591.
9. **Glass PJ, Zeng CQ, Estes MK.** 2003. Two nonoverlapping domains on the Norwalk virus open reading frame 3 (ORF3) protein are involved in the formation of the phosphorylated 35K protein and in ORF3-capsid protein interactions. J. Virol. **77**:3569–3577.
10. **Ho EC, Cheng PK, Lau AW, Wong AH, Lim WW.** 2007. Atypical norovirus epidemic in Hong Kong during summer of 2006 was caused by a new genogroup II/4 variant. J. Clin. Microbiol. **45**:2205–2211.
11. **Ho EC, Cheng PK, Wong DA, Lau AW, Lim WW.** 2006. Correlation of norovirus variants with epidemics of acute viral gastroenteritis in Hong Kong. J. Med. Virol. **78**:1473–1479.
12. **Kamel AH, et al.** 2009. Predominance and circulation of enteric viruses in the region of Greater Cairo, Egypt. J. Clin. Microbiol. **47**:1037–1045.
13. **Prasad BV, et al.** 1999. X-ray crystallographic structure of the Norwalk virus capsid. Science **286**:287–290.
14. **Seah EL, Gunesekere IC, Marshall JA, Wright PJ.** 1999. Variation in ORF3 of genogroup 2 Norwalk-like viruses. Arch. Virol. **144**:1007–1014.
15. **Siebenga JJ, et al.** 2010. Phylodynamic reconstruction reveals norovirus GII.4 epidemic expansions and their molecular determinants. PLoS Pathog. **6**:e1000884.
16. **Simmonds P, et al.** 2008. Bioinformatic and functional analysis of RNA secondary structure elements among different genera of human and animal caliciviruses. Nucleic Acids Res. **36**:2530–2546.
17. **Sosnovtsev SV, Belliot G, Chang KO, Onwudiwe O, Green KY.** 2005. Feline calicivirus VP2 is essential for the production of infectious virions. J. Virol. **79**:4012–4024.
18. **Zheng DP, Widdowson MA, Glass RI, Vinje J.** 2010. Molecular epidemiology of genogroup II-genotype 4 noroviruses in the United States between 1994 and 2006. J. Clin. Microbiol. **48**:168–177.