

Diversity and Relationships of Cocirculating Modern Human Rotaviruses Revealed Using Large-Scale Comparative Genomics

Sarah M. McDonald,^a Allison O. McKell,^a Christine M. Rippinger,^b John K. McAllen,^c Asmik Akopov,^c Ewen F. Kirkness,^c Daniel C. Payne,^d Kathryn M. Edwards,^e James D. Chappell,^{e,f} and John T. Patton^b

Virginia Tech Carilion Research Institute, Roanoke, Virginia, USA, and the Department of Biomedical Sciences and Pathobiology, Virginia-Maryland Regional College of Veterinary Medicine, Blacksburg, Virginia, USA^a; Rotavirus Molecular Biology Section, Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, USA^b; The J. Craig Venter Institute, Rockville, Maryland, USA^c; Centers for Disease Control and Prevention, Atlanta, Georgia, USA^d; and Vanderbilt University School of Medicine, Departments of Pediatrics^e and Pathology, Microbiology, and Immunology,^f Nashville, Tennessee, USA

Group A rotaviruses (RVs) are 11-segmented, double-stranded RNA viruses and are primary causes of gastroenteritis in young children. Despite their medical relevance, the genetic diversity of modern human RVs is poorly understood, and the impact of vaccine use on circulating strains remains unknown. In this study, we report the complete genome sequence analysis of 58 RVs isolated from children with severe diarrhea and/or vomiting at Vanderbilt University Medical Center (VUMC) in Nashville, TN, during the years spanning community vaccine implementation (2005 to 2009). The RVs analyzed include 36 G1P[8], 18 G3P[8], and 4 G12P[8] Wa-like genogroup 1 strains with VP6-VP1-VP2-VP3-NSP1-NSP2-NSP3-NSP4-NSP5/6 genotype constellations of I1-R1-C1-M1-A1-N1-T1-E1-H1. By constructing phylogenetic trees, we identified 2 to 5 subgenotype alleles for each gene. The results show evidence of intragenogroup gene reassortment among the cocirculating strains. However, several isolates from different seasons maintained identical allele constellations, consistent with the notion that certain RV clades persisted in the community. By comparing the genes of VUMC RVs to those of other archival and contemporary RV strains for which sequences are available, we defined phylogenetic lineages and verified that the diversity of the strains analyzed in this study reflects that seen in other regions of the world. Importantly, the VP4 and VP7 proteins encoded by VUMC RVs and other contemporary strains show amino acid changes in or near neutralization domains, which might reflect antigenic drift of the virus. Thus, this large-scale, comparative genomic study of modern human RVs provides significant insight into how this pathogen evolves during its spread in the community.

Group A rotaviruses (RVs) are important pathogens that cause acute gastroenteritis in infants and young children (11, 27). In developing regions of the world with reduced access to medical care, RV infections lead to the deaths of ~450,000 children each year (50). In industrialized countries, the burden of RV disease is mainly associated with the financial costs of treatment. Specifically, prior to the recent introduction of vaccines in the United States, it was estimated that RV-induced gastroenteritis caused more than 55,000 hospitalizations and 500,000 physician visits each year at a societal cost of ~\$1 billion (40, 54). In 2006 and 2008, respectively, the U.S. Advisory Committee on Immunization Practices recommended the live-attenuated vaccines RotaTeq (Merck) and Rotarix (Glaxo-Smith Kline) for the routine immunization of infants (8, 37). RotaTeq is a pentavalent vaccine consisting of five human-bovine RV reassortants, each of which carries a separate human RV VP7 gene (G1, G2, G3, or G4) or a human RV P[8] VP4 gene in the background of the bovine WC3 strain (G6P[5]) (31). In contrast, Rotarix is a monovalent vaccine derived from a human strain (89-12) with G1P[8] specificity (53). Postlicensure studies indicate that both RotaTeq and Rotarix prevent 85 to 100% of severe RV gastroenteritis in developed countries (15). For reasons that remain unclear, the efficacy of the vaccines was found to be lower in developing regions of the world (2, 25, 43, 55). Surveillance networks have been established at various geographical locations (i) to obtain information on the prevalence and types of circulating RV strains, (ii) to determine whether human RVs are changing in the face of vaccine pressures, and (iii) to continue monitoring for the safety and efficacy of

RotaTeq and Rotarix (5, 12, 13, 17, 19, 20, 24, 39). Importantly, the viral gene/genome sequences deduced via these epidemiological studies are illuminating the diversity and complex evolutionary dynamics of this common childhood pathogen.

RVs maintain their double-stranded RNA (dsRNA) genome as 11 separate segments, which can reassort when a host cell is infected with more than one strain (11, 34). These exchanges, described as genetic shift, have long been hypothesized to play an important role in generating viral diversity, allowing the virus to evolve rapidly in response to selection pressures (14). In addition, the error-prone nature of the viral polymerase leads to the accumulation of point mutations in the viral genome, causing RV strains to drift antigenically (11). However, because only limited numbers of complete genome sequences have been determined for naturally circulating human RVs, the frequency and significance of shift and drift are poorly understood. To aid in studies of RV diversity, a comprehensive classification system was developed that designates a genotype for each of the 11 viral genes (i.e., segments) based on established nucleotide identity cutoff values (28).

Received 3 May 2012 Accepted 7 June 2012

Published ahead of print 13 June 2012

Address correspondence to Sarah M. McDonald, mcdonaldsa@vtc.vt.edu.

Supplemental material for this article may be found at <http://jvi.asm.org/>.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JVI.01105-12

TABLE 1 RV G/P types collected and sequenced from sick children at Vanderbilt University Medical Center in Nashville, TN (2005 to 2009)

G/P type	Vaccination season and analysis status							
	2005-2006	2006-2007	Collected	Sequenced	Collected	Sequenced	Collected	Sequenced
G1P[8]	64	26	30	10	12	0	4	0
G2P[4]	5	0	4	0	0	0	5	0
G3P[8]	0	0	2	1	3	0	25	17
G9P[8]	0	0	0	0	0	0	2	0
G12P[8]	1	0	0	0	0	0	3	2
G12P[6]	4	2 ^a	0	0	0	0	0	0
Mixed	1	0	0	0	0	0	0	0
Total	75	28	36	11	15	0	39	19

^a Typed as P[6] by PCR but determined to be P[8] by sequence analysis.

The acronym Gx-P[x]-Ix-Rx-Cx-Mx-Ax-Nx-Tx-Ex-Hx is used to describe a virus based on segments encoding each viral protein(s) (i.e., VP7-VP4-VP6-VP1-VP2-VP3-NSP1-NSP2-NSP3-NSP4-NSP5/6). This system extends the well-known binomial classification of RVs that is based on the genes encoding outer capsid serotype antigens VP7 (G types) and VP4 (P types) to the other nine internal protein genes (encoding VP1 to VP3, VP6, and NSP1 to NSP5/6) (29).

The complete genome classification approach provided data to support the existence of two dominant human RV genogroups (Wa-like and DS-1-like), originally identified using differential RNA-RNA hybridization (35). Wa-like RVs almost invariably exhibit genotype 1 internal protein genes (i.e., I1-R1-C1-M1-A1-N1-T1-E1-H1) and tend to have G1P[8], G3P[8], G4P[8], and G9P[8] specificities (29). In contrast, DS-1-like RVs that have been fully sequenced are usually G2P[4] strains with genotype 2 internal protein genes (i.e., I2-R2-C2-M2-A2-N2-T2-E2-H2) (29). For the purposes of this paper, we will refer to the Wa-like genogroup as genogroup 1 (GG-1) and the DS-1-like genogroup as genogroup 2 (GG-2). Gene reassortment between GG-1 and GG-2 strains is possible, as evidenced by the isolation of human RVs containing both genotype 1 and genotype 2 genes from children suffering acute gastroenteritis (16, 47, 48). However, based on the available sequence data, intergenogroup reassortants seem to be less prevalent in the human population than pure GG-1 or GG-2 strains (29). It is possible that the genes (or the encoded proteins) of RVs belonging to the same genogroup have coevolved and operate best when kept together (3, 6, 16, 18). In this manner, it hypothesized that human RVs belonging to different genogroups create less fit reassortants which may not emerge in the human population. In contrast, segment exchange between RVs within a genogroup would be expected to occur more readily. Nevertheless, large-scale, complete genome sequence analyses of 62 archival GG-1 human RVs (51 G3P[8] and 11 G4P[8] strains) from Washington, DC, revealed that intragenogroup reassortants were less common than had been anticipated (32, 33). Comparative genomic studies of contemporary strains are needed to ascertain whether these previous results reflect the current status of RV diversity.

In this study, we report the complete genome sequence analysis of 58 GG-1 human RVs (36 G1P[8], 18 G3P[8], and 4 G12P[8] strains) from children seeking medical attention for severe, acute gastroenteritis at Vanderbilt University Medical Center (VUMC)

in Nashville, TN, from one prevaccine season (2005 to 2006) and two postvaccine seasons (2006 to 2007 and 2008 to 2009). Complete genome sequence analyses of modern human RVs are expected to enhance our understanding of the diversity and evolution of this important pediatric pathogen.

MATERIALS AND METHODS

Sample collection, RNA extraction, and G/P typing. Fecal specimens were collected from infants and children presenting with acute gastroenteritis at VUMC in Nashville, TN, during the years of 2005 to 2009 as part of the Centers for Disease Control and Prevention's (CDC) New Vaccine Surveillance Network (NVSN), which was approved by the CDC and VUMC institutional review boards. Details of patient enrollment and specimen collection have been previously published (40, 41). Briefly, children <36 months of age and hospitalized with diarrhea (≥ 3 episodes in 24 h) and/or vomiting (≥ 1 episode in 24 h), who were residents of Davidson County and who had informed consent from a parent or guardian, were enrolled in the study. Children were ineligible if they had a reported history of noninfectious diarrhea or clinical immunodeficiency. Bulk fecal samples were collected from the children within 7 days of enrollment (usually within 7 days of symptom onset) and tested for evidence of viral antigen using the commercial enzyme immunoassay (EIA) Premier Rotac隆 (Meridian Bioscience). Total RNA was extracted from deidentified RV-positive fecal specimens using TRIzol (Invitrogen), and samples were classified into G/P types using reverse transcription-PCR (RT-PCR) and/or nucleotide sequencing (Table 1). Clinical and demographic data relating to the 58 samples analyzed in the current study are shown in Table S1 in the supplemental material.

RT-PCR, nucleotide sequencing, and genotyping internal protein genes. Total RNA extracted from a representative subset of RV-positive fecal specimens (84 samples in total) was sent to the J. Craig Venter Institute (Rockville, MD) for high-throughput, RT-PCR, and Sanger sequencing. As described previously by McDonald et al. (33), oligonucleotide primers were designed every 600 bp along both sense and antisense strands of the viral genome, and M13-forward or M13-reverse tags were added to the 5' ends of the primers for use in sequencing. Primers for the G1 and G12 VP7 genes were designed based on the consensus of G-type-matched sequences available in public databases (see Table S2 in the supplemental material). Primers for the G3 VP7 gene, as well as the VP1 to VP4, VP6, and NSP1 to NSP5/6 genes, were described in McDonald et al. (33). RT-PCR and sequencing reactions were performed as described previously (32, 33). The genotype of each VUMC RV gene sequence was determined by RotaC assignment (<http://rotac.regatools.be>) (26).

Alignments, phylogenetic analyses, and structural predictions. Nucleotide sequence alignments and phylogenies were generated with Ge-

neious Pro v5.5.2 using the ClustalW and PhyML plugins. The Hasegawa-Kishino-Yano substitution model with gamma-distributed rate variation among sites (HKY+G) was chosen by following Bayesian information criterion ranking of each alignment as implemented in MEGA5.05 (45, 49). Trees were constructed with the open reading frame (ORF) nucleotide sequences, and 1,000 pseudoreplicates were generated for bootstrapping analyses. Subgenotype alleles were defined as tight phylogenetic clusters (containing 2 or more sequences) with strong bootstrap support (>75%) at separating nodes and were confirmed by visual inspection of nucleotide alignments. GenBank accession numbers and genotypes of the previously sequenced RV genes used in the phylogenetic analyses are provided in Tables S3 and S4 in the supplemental material. Amino acid alignments were constructed with Geneious Pro v5.5.2 using the ClustalW plugin with the BLOSUM cost matrix. Structural analysis of NSP2 (PDB number 2R7C; strain SA11), VP7 (PBD number 3FMG; strain RRV), and VP8* (PDB number 2DWR; strain Wa) was performed using the UCSF Chimera molecular modeling system (1, 4, 21, 42).

Nucleotide sequence accession numbers. The nucleotide sequences determined for the 11 genes of each of the 58 VUMC RVs (638 total sequences) were deposited in GenBank and are listed in Table 2.

RESULTS

Complete genome sequencing of VUMC RVs. Fecal specimens were collected from children <36 months of age presenting symptoms of gastroenteritis at VUMC in Nashville, TN (39–41). A total of 669 specimens were collected over four winter seasons (2005–2006, 2006–2007, 2007–2008, and 2008–2009), spanning the time of RotaTeq introduction in the community (beginning in spring 2006). Of these samples, 165 tested positive for the presence of RV antigen by EIA and then were sent to the CDC for G/P typing using RT-PCR (39). The G/P type distribution of RVs in the fecal specimens was typical of that seen in other regions of the United States during this time, with G1P[8], G2P[4], and G3P[8] being detected most years (Table 1) (17). The globally emerging G9P[8] and G12 strains were also detected, albeit in fewer numbers than the more traditional G/P-type strains (Table 1) (17, 30). Similar to what has been seen in many previous epidemiological studies, the dominant G/P type fluctuated from G1P[8] in 2005 to 2008 to G3P[8] in 2008 to 2009 (7, 17, 44). No correlation was found between G/P type and age, sex, and ethnicity of the child or between G/P type and disease severity (see Table S1 in the supplemental material).

We next determined the complete genome sequences of representative VUMC RVs using an established, semiautomated, RT-PCR and dideoxy-nucleotide Sanger sequencing pipeline at the J. Craig Venter Institute (32, 33). Unfortunately, material was unavailable for the 15 samples collected during the 2007–2008 season; therefore, we were not able to sequence RVs from this season (Table 1). Moreover, the pipeline was designed to preferentially amplify and sequence genotype 1 internal protein genes (32, 33). Consequently, no complete genome sequence information was obtained for the 14 G2P[4] samples from this collection (Table 1). Nonetheless, we were able to deduce the complete genome sequences of RVs in 36 G1P[8]-, 18 G3P[8]-, 2 G12P[8]-, and 2 G12P[6]-typed strains (58 RVs in total) that were collected during the 2005–2006, 2006–2007, and 2008–2009 seasons (Table 1). The entire ORFs and, in some cases, the 5' and 3' untranslated regions for all 11 genes of these 58 VUMC RVs were sequenced. Multiple, overlapping reads were generated, and the chromatographs showed little evidence of heterogeneity, indicating that a single RV isolate was dominant in each fecal specimen.

Using the web-based genotyping tool RotaC, we assigned genotypes for each of the 11 genes of the 58 VUMC RVs (26). The

G/P types of the viruses were verified, except for those of two strains that were initially typed as P[6] by RT-PCR analysis (Table 1). These two G12P[6] isolates (VU05-06-72 and VU05-06-74) from the 2005–2006 season were actually found to be G12P[8] strains by sequencing and RotaC. The nine other genes of each VUMC RV were classified as genotype 1, confirming that the strains sequenced in this study all belong to GG-1. Aside from their different G-type specificities, the 58 RVs share identical genotype constellations (i.e., Gx-P[8]-I1-R1-C1-M1-A1-N1-T1-E1-H1).

Subgenotype allele designation for VUMC RV genes. Given that the 58 VUMC RVs belong to GG-1 and have genes with the same genotypes (except VP7), we sought to examine their diversity at the subgenotype level. To do this, we created maximum likelihood phylogenetic trees for each gene of the 58 RVs using the ORF nucleotide sequences (Fig. 1). Similar to our previous analyses of archival GG-1 human RVs from Washington, DC, we defined subgenotype alleles as tight phylogenetic clusters (containing 2 or more sequences) that are separated by nodes with strong bootstrap values (>75%) (32, 33). Because the 58 RVs differed in their G-type specificities, we did not divide the VP7 genes into alleles in this study (Fig. 1A). Instead, the VUMC RV VP7 genes were compared phylogenetically to those of archival, contemporary, and vaccine strains and placed into established subgenotypic lineages (discussed below).

The phylogenetic analyses identified two to five subgenotype alleles for the VP1 to VP4, VP6, and NSP1 to NSP5/6 genes of the 58 VUMC RVs (Fig. 1B to K). For ease of discussion and visualization, the alleles were each assigned a letter (A, B, C, D, and E) and a corresponding color (red, green, cyan, purple, and orange, respectively). Consistent with our previous study of archival strains, we saw that isolates from different seasons could share nearly identical alleles, while those from the same infectious season could have genetically divergent alleles. Together, the phylogenetic data are most consistent with the notion that the VUMC alleles diverged prior to 2005, and that they remained genetically stable, accumulating minimal point mutations over the collection years (2005 to 2009).

Persistent clades (PCs) of VUMC RVs despite frequent gene reassortment. The summary of the color-coded allele constellations for each of the VUMC RVs reveals the complex genetic diversity of these GG-1 strains (Fig. 2). RVs with several different allele constellations cocirculated each epidemic season (Fig. 2A). In the 2005–2006 season, nine different clades with distinct allele constellations (8 G1P[8] and 1 G12P[8]) were identified, and in the 2006–2007 season, seven different clades (6 G1P[8] and 1 G3P[8]) were found. The 2008–2009 season was less diverse, with a majority of the viruses belonging to a single, homogeneous G3P[8] major clade. Two G12P[8] allele constellations, represented by VU08-09-6 and VU08-09-39, were also found in the 2008–2009 season. Two isolates (VU08-09-9 and VU08-09-28) of the major G3P[8] clade were from children who received two doses of RotaTeq, and another two isolates (VU08-09-17 and VU08-09-22) were from children vaccinated with all three doses (Fig. 2; also see Table S1 in the supplemental material). Overall, we found no allele-level constellation differences between G3P[8] viruses from vaccinated versus unvaccinated children. However, one G3P[8] isolate (VU09-08-22) from a child who was fully vaccinated with RotaTeq seemed related to the major 2008–2009 clade but contained different NSP1 and NSP3 gene alleles (Fig. 2).

This analysis showed that many VUMC RV isolates exhibit

TABLE 2 GenBank accession numbers for VUMC RVs sequenced in the current study

Full strain name	RV protein										
	VP7	VP4	VP6	VP1	VP2	VP3	NSP1	NSP2	NSP3	NSP4	NSP5/6
RVA/human-wt/USA/VU05-06-2/2005/G1P[8]	JF490561	JF490564	JF490563	JF490556	JF490566	JF490565	JF490558	JF490557	JF490562	JF490560	JF490559
RVA/human-wt/USA/VU05-06-9/2005/G1P[8]	JF490577	JF490573	JF490572	JF490567	JF490576	JF490575	JF490574	JF490568	JF490571	JF490570	JF490569
RVA/human-wt/USA/VU05-06-13/2005/G1P[8]	JF490582	JF490585	JF490584	JF490578	JF490588	JF490587	JF490586	JF490579	JF490583	JF490581	JF490580
RVA/human-wt/USA/VU05-06-15/2005/G1P[8]	JF490599	JF490595	JF490594	JF490589	JF490598	JF490597	JF490596	JF490590	JF490593	JF490592	JF490591
RVA/human-wt/USA/VU05-06-16/2005/G1P[8]	JF490608	JF490606	JF490605	JF490600	JF490610	JF490609	JF490607	JF490601	JF490604	JF490603	JF490602
RVA/human-wt/USA/VU05-06-17/2005/G1P[8]	JF490619	JF490616	JF490615	JF490611	JF490620	JF490621	JF490618	JF490617	JF490614	JF490613	JF490612
RVA/human-wt/USA/VU05-06-19/2005/G1P[8]	JF490627	JF490630	JF490629	JF490622	JF490632	JF490631	JF490624	JF490623	JF490628	JF490626	JF490625
RVA/human-wt/USA/VU05-06-20/2005/G1P[8]	JF490638	JF490641	JF490640	JF490633	JF490643	JF490642	JF490635	JF490634	JF490639	JF490637	JF490636
RVA/human-wt/USA/VU05-06-23/2005/G1P[8]	JF490648	JF490651	JF490650	JF490644	JF490653	JF490652	JF490654	JF490645	JF490649	JF490647	JF490646
RVA/human-wt/USA/VU05-06-26/2005/G1P[8]	JF490659	JF490663	JF490661	JF490655	JF490665	JF490664	JF490662	JF490656	JF490660	JF490658	JF490657
RVA/human-wt/USA/VU05-06-27/2005/G1P[8]	JF490671	JF490674	JF490673	JF490666	JF490676	JF490675	JF490668	JF490667	JF490672	JF490670	JF490669
RVA/human-wt/USA/VU05-06-28/2005/G1P[8]	JF490682	JF490685	JF490684	JF490677	JF490687	JF490684	JF490679	JF490678	JF490683	JF490681	JF490680
RVA/human-wt/USA/VU05-06-33/2005/G1P[8]	JF490691	JF490694	JF490693	JF490697	JF490696	JF490695	JF490695	JF490698	JF490688	JF490692	JF490690
RVA/human-wt/USA/VU05-06-40/2005/G1P[8]	JF490707	JF490702	JF490704	JF490699	JF490700	JF490701	JF490703	JF490706	JF490705	JF490708	JF490709
RVA/human-wt/USA/VU05-06-43/2005/G1P[8]	JF490717	JF490718	JF490714	JF490715	JF490720	JF490719	JF490716	JF490710	JF490713	JF490712	JF490711
RVA/human-wt/USA/VU05-06-46/2005/G1P[8]	JF490726	JF490729	JF490728	JF490721	JF490731	JF490730	JF490723	JF490722	JF490727	JF490725	JF490724
RVA/human-wt/USA/VU05-06-47/2005/G1P[8]	JF490737	JF490740	JF490739	JF490732	JF490742	JF490741	JF490734	JF490733	JF490738	JF490736	JF490735
RVA/human-wt/USA/VU05-06-48/2005/G1P[8]	JF490747	JF490750	JF490749	JF490743	JF490751	JF490753	JF490752	JF490744	JF490748	JF490745	JF490745
RVA/human-wt/USA/VU05-06-57/2005/G1P[8]	JF490762	JF490762	JF490764	JF490754	JF490758	JF490757	JF490760	JF490759	JF490755	JF490763	JF490761
RVA/human-wt/USA/VU05-06-59/2005/G1P[8]	JF490768	JF490770	JF490774	JF490772	JF490771	JF490775	JF490773	JF490765	JF490769	JF490767	JF490766
RVA/human-wt/USA/VU05-06-62/2005/G1P[8]	JF490784	JF490782	JF490781	JF490776	JF490786	JF490785	JF490783	JF490777	JF490780	JF490779	JF490778
RVA/human-wt/USA/VU05-06-67/2005/G1P[8]	JF490792	JF490797	JF490794	JF490787	JF490796	JF490795	JF490789	JF490788	JF490793	JF490791	JF490790
RVA/human-wt/USA/VU05-06-69/2005/G1P[8]	JF490803	JF490806	JF490805	JF490798	JF490808	JF490807	JF490800	JF490799	JF490804	JF490802	JF490801
RVA/human-wt/USA/VU05-06-70/2005/G1P[8]	JF490812	JF490815	JF490814	JF490817	JF490816	JF490819	JF490818	JF490809	JF490813	JF490811	JF490810
RVA/human-wt/USA/VU05-06-72/2005/G12P[8]	JF790296	JF790298	JF790293	JF790288	JF790295	JF790294	JF790297	JF790289	JF790292	JF790291	JF790290
RVA/human-wt/USA/VU05-06-73/2005/G1P[8]	JF490824	JF490827	JF490826	JF490820	JF490829	JF490828	JF490830	JF490821	JF490823	JF490822	JF490822
RVA/human-wt/USA/VU05-06-74/2005/G12P[8]	JF790307	JF790309	JF790304	JF790299	JF790306	JF790307	JF790305	JF790308	JF790300	JF790303	JF790301
RVA/human-wt/USA/VU05-06-76/2005/G1P[8]	JF490836	JF490839	JF490838	JF490831	JF490841	JF490840	JF490833	JF490832	JF490837	JF490835	JF490834
RVA/human-wt/USA/VU06-07-1/2006/G1P[8]	JF490883	JF490882	JF490881	JF490875	JF490885	JF490884	JF490877	JF490876	JF490880	JF490879	JF490878
RVA/human-wt/USA/VU06-07-7/2006/G1P[8]	JF490896	JF490894	JF490893	JF490886	JF490896	JF490895	JF490888	JF490887	JF490891	JF490890	JF490889
RVA/human-wt/USA/VU06-07-10/2006/G1P[8]	JF490905	JF490904	JF490903	JF490897	JF490906	JF490907	JF490899	JF490898	JF490902	JF490901	JF490900
RVA/human-wt/USA/VU06-07-21/2006/G3P[8]	JF490913	JF490916	JF490915	JF490908	JF490918	JF490917	JF490910	JF490909	JF490914	JF490912	JF490911
RVA/human-wt/USA/VU06-07-27/2006/G1P[8]	JF490923	JF490926	JF490925	JF490919	JF490928	JF490927	JF490929	JF490920	JF490924	JF490922	JF490921
RVA/human-wt/USA/VU06-07-29/2006/G1P[8]	JF490947	JF490850	JF490849	JF490842	JF490852	JF490851	JF490844	JF490843	JF490846	JF490845	JF490845
RVA/human-wt/USA/VU06-07-30/2006/G1P[8]	JF490857	JF490861	JF490859	JF490863	JF490860	JF490863	JF490862	JF490854	JF490853	JF490856	JF490855
RVA/human-wt/USA/VU06-07-31/2006/G1P[8]	JF490869	JF490872	JF490871	JF490864	JF490874	JF490873	JF490866	JF490865	JF490870	JF490868	JF490867
RVA/human-wt/USA/VU06-07-32/2006/G1P[8]	JF490935	JF490938	JF490937	JF490930	JF490940	JF490939	JF490932	JF490931	JF490936	JF490934	JF490933
RVA/human-wt/USA/VU06-07-33/2006/G1P[8]	JF490946	JF490950	JF490948	JF490941	JF490949	JF490951	JF490943	JF490942	JF490947	JF490945	JF490944
RVA/human-wt/USA/VU06-07-35/2006/G1P[8]	JF490960	JF490957	JF490956	JF490958	JF490962	JF490961	JF490959	JF490952	JF490955	JF490954	JF490953
RVA/human-wt/USA/VU08-09-5/2008/G3P[8]	JF490968	JF490971	JF490970	JF490963	JF490973	JF490972	JF490965	JF490964	JF490969	JF490967	JF490966
RVA/human-wt/USA/VU08-09-6/2008/G12P[8]	JF490979	JF490982	JF490981	JF490974	JF490983	JF490982	JF490976	JF490975	JF490980	JF490978	JF490977
RVA/human-wt/USA/VU08-09-7/2008/G3P[8]	JF490990	JF490993	JF490992	JF490985	JF490995	JF490994	JF490987	JF490986	JF490991	JF490989	JF490988
RVA/human-wt/USA/VU08-09-9/2008/G3P[8]	JF491001	JF491004	JF491003	JF490996	JF491006	JF491005	JF490998	JF490997	JF491002	JF491000	JF490999
RVA/human-wt/USA/VU08-09-12/2008/G3P[8]	JF491012	JF491015	JF491014	JF491007	JF491017	JF491016	JF491009	JF491008	JF491013	JF491011	JF491010
RVA/human-wt/USA/VU08-09-16/2008/G3P[8]	JF790316	JF790317	JF790314	JF790315	JF790319	JF790318	JF790320	JF790310	JF790313	JF790312	JF790311
RVA/human-wt/USA/VU08-09-17/2008/G3P[8]	JF491025	JF491023	JF491022	JF491024	JF491028	JF491026	JF491027	JF491018	JF491021	JF491020	JF491019
RVA/human-wt/USA/VU08-09-19/2008/G3P[8]	JF491039	JF491035	JF491034	JF491038	JF491037	JF491036	JF491030	JF491029	JF491033	JF491031	JF491031
RVA/human-wt/USA/VU08-09-20/2008/G3P[8]	JF790331	JF790327	JF790325	JF790326	JF790330	JF790328	JF790329	JF790321	JF790324	JF790323	JF790322
RVA/human-wt/USA/VU08-09-21/2008/G3P[8]	JF491047	JF491048	JF491044	JF491046	JF491045	JF491049	JF491050	JF491040	JF491043	JF491042	JF491041
RVA/human-wt/USA/VU08-09-22/2008/G3P[8]	JF491055	JF491058	JF491057	JF491060	JF491061	JF491059	JF491052	JF491051	JF491056	JF491054	JF491053
RVA/human-wt/USA/VU08-09-24/2008/G3P[8]	JF491066	JF491069	JF491068	JF491071	JF491072	JF491070	JF491063	JF491062	JF491067	JF491065	JF491064
RVA/human-wt/USA/VU08-09-25/2008/G3P[8]	JF491077	JF491080	JF491079	JF491083	JF491082	JF491081	JF491074	JF491073	JF491078	JF491076	JF491075
RVA/human-wt/USA/VU08-09-26/2008/G3P[8]	JF491089	JF491092	JF491091	JF491084	JF491094	JF491093	JF491086	JF491085	JF491090	JF491088	JF491087
RVA/human-wt/USA/VU08-09-27/2008/G3P[8]	JF790342	JF790345	JF790334	JF790332	JF790337	JF790336	JF790339	JF790338	JF790341	JF790340	JF790333
RVA/human-wt/USA/VU08-09-28/2008/G3P[8]	JF491105	JF491101	JF491100	JF491095	JF491103	JF491102	JF491104	JF491096	JF491094	JF491098	JF491097
RVA/human-wt/USA/VU08-09-29/2008/G3P[8]	JF491115	JF491111	JF491116	JF491114	JF491113	JF491112	JF491107	JF491106	JF491110	JF491111	JF491108
RVA/human-wt/USA/VU08-09-30/2008/G3P[8]	JF491122	JF491125	JF491124	JF491117	JF491127	JF491126	JF491119	JF491118	JF491123	JF491121	JF491120
RVA/human-wt/USA/VU08-09-39/2008/G12P[8]	JF790346	JF790352	JF790348	JF790349	JF790350	JF790351	JF790353	JF790343	JF790347	JF790345	JF790344

Downloaded from <http://jvi.asm.org/> on December 5, 2014 by University Libraries, Virginia Tech

genes with the same allele designations. This result is unlike what we described for the archival GG-1 DC RVs and is indicative of gene reassortment among the contemporary VUMC strains (32, 33). However, it is not possible to determine whether these genetic exchanges occurred during or prior to the years of study. Nonetheless, several allele constellations were seen repeatedly, even for

viruses isolated in different seasons. For instance, the G3P[8] isolate VU06-07-21 from the 2006-2007 season and the major clade of G3P[8] RVs from the 2008-2009 season contain identical allele constellations (Fig. 2A). These PCs of VUMC RVs can be identified when the color-coded allele constellations of VUMC RVs are ordered according to visual genome similarities rather than date of isolation

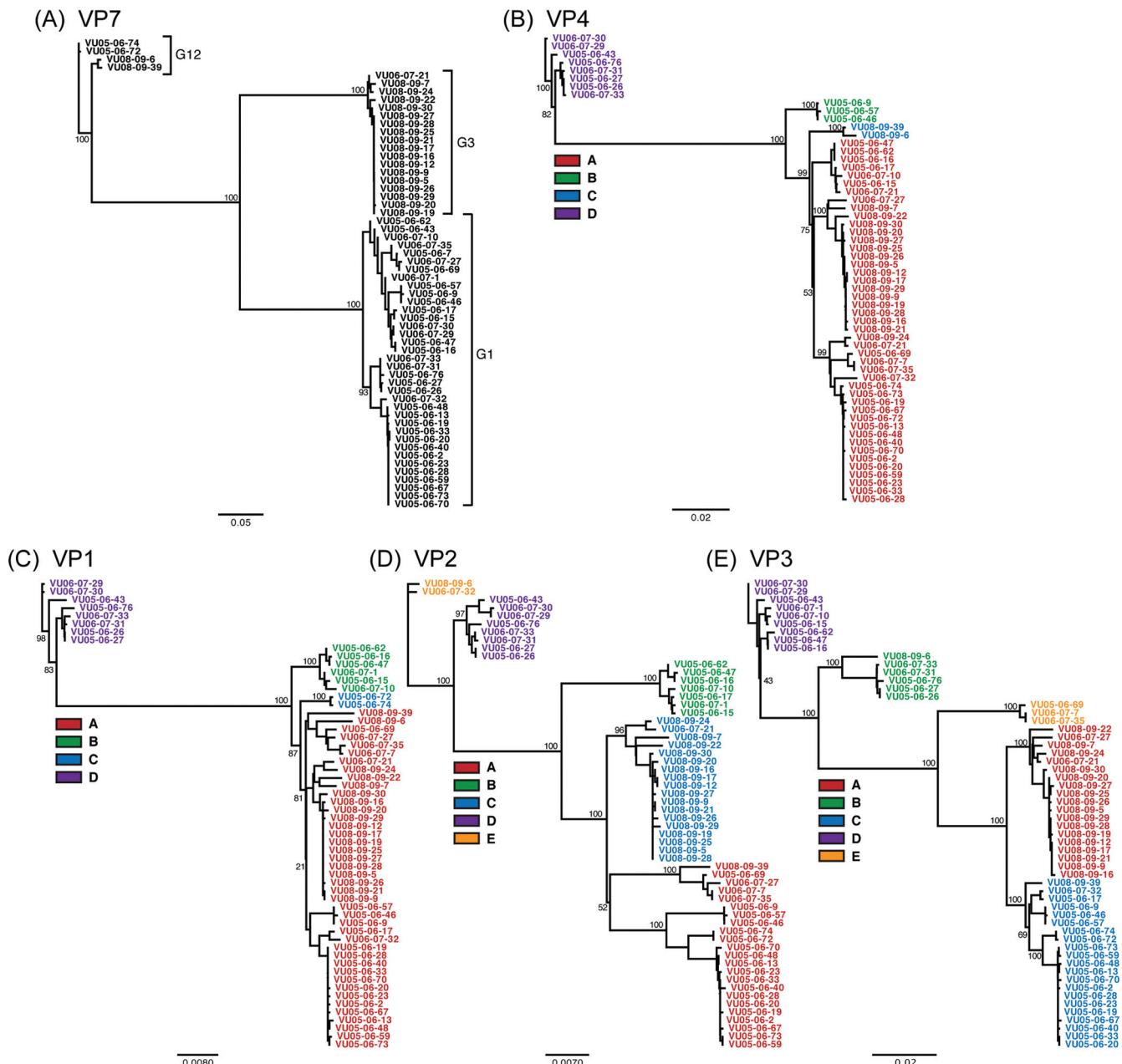


FIG 1 Genetic relationships among the individual genes of the VUMC RVs. The unrooted maximum likelihood trees were constructed using the ORF nucleotide sequences for each gene of each isolate. All horizontal branch lengths are drawn to scale (nucleotide substitutions per base), and bootstrap values are shown as percentages for key nodes. Sequences that cluster tightly in the phylogenetic trees and are supported by strong nodal bootstrap values (>75%) were assigned to the same color-coded allele group. Thus, separate colors indicate divergent gene alleles: allele A (red), allele B (green), allele C (cyan), allele D (purple), and allele E (orange). For the VP7 tree, no alleles were defined, and instead the G-type groupings (G1, G3, and G12) are indicated with brackets.

(Fig. 2B). PC-1, PC-3, PC-4, and PC-5 are each comprised of G1P[8] RVs found during the 2005–2006 and 2006–2007 infectious seasons. PC-2 is represented by the aforementioned single G3P[8] isolate from 2006 to 2007 (VU06-07-21) along with 16 G3P[8] isolates from 2 years later (2008 to 2009). Like G/P-type specificity, we found no correlation between allele constellation and age, sex, or ethnicity of the child or between allele constellation and disease severity (Fig. 2; also see Table S1 in the supplemental material).

Subgenotype 1 alleles of VUMC RVs encode different proteins. Having observed the intragenotypic diversity of the VUMC

RV genes at the nucleotide level, we next performed amino acid alignments to determine whether the alleles (A to E) encode different VP1 to VP4, VP6, and NSP1 to NSP5/6 proteins (data not shown). The results indicate that, in general, the proteins are different, with 1 to 78 allele-specific amino acid changes. The intermediate capsid protein of the virion (VP6) was the most highly conserved for the VUMC RVs, showing less than five allele-specific amino acid changes in all pairwise comparisons (Fig. 3A). In contrast, the innate immune antagonist (NSP1) was the most variable protein, with 19 to 78 amino acid differences between pro-

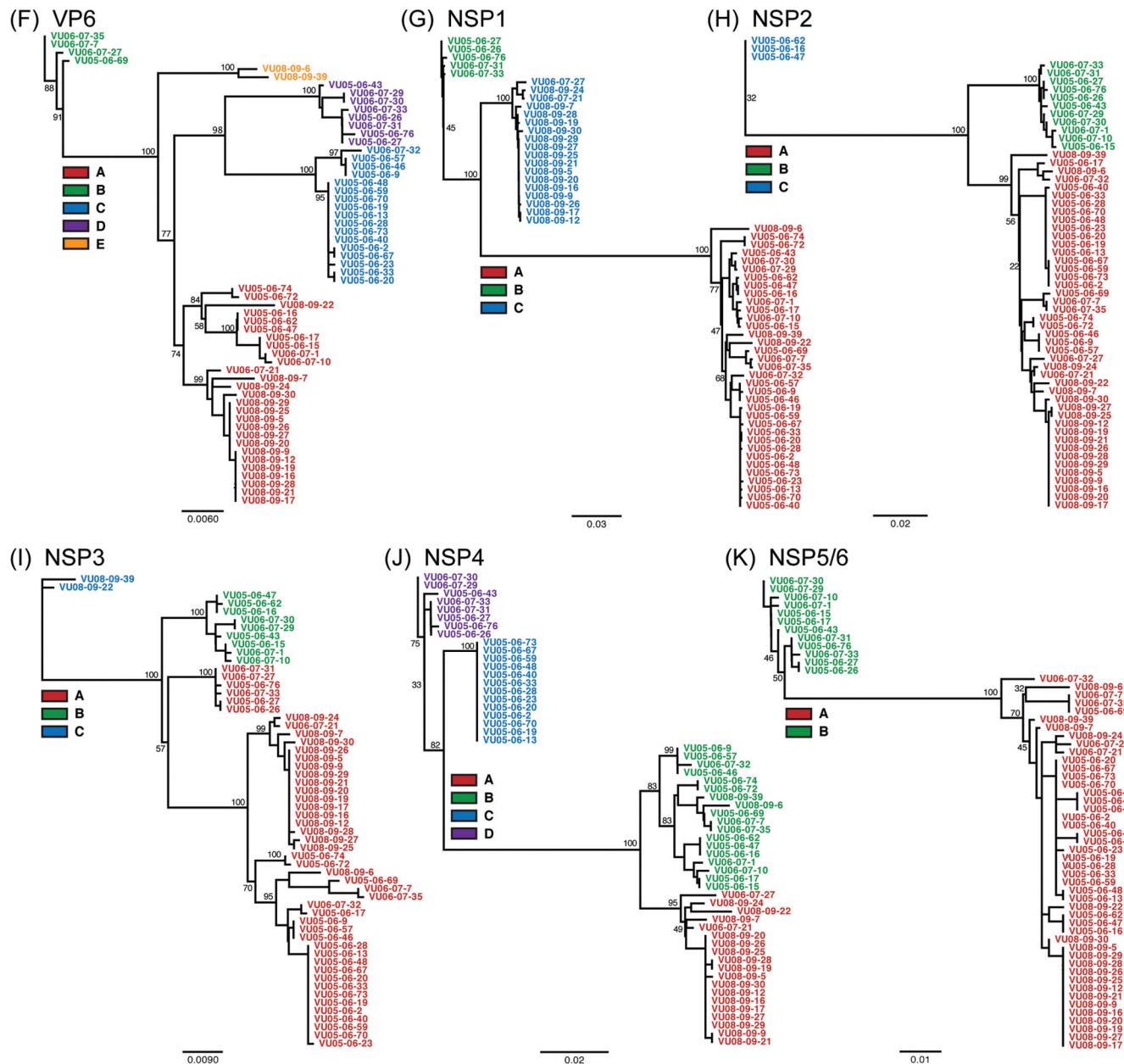


FIG 1 continued

teins encoded by different alleles (Fig. 3A). From this amino acid analysis, it became apparent that some gene alleles encode more similar proteins than do others. For instance, the viral RNA-dependent RNA polymerase encoded by the VP1 A allele (red) differs by only 2 amino acids from those encoded by either the B allele (green) or the C allele (cyan) (Fig. 3A). However, VP1 differs by 10 to 12 amino acids in alleles A and D (red versus purple), B and D (green versus purple), or C and D (cyan versus purple). For the VP4 spike attachment protein and P-type antigen, we found that proteins encoded by the D allele (purple) were much more divergent than the others, showing 35 unique differences. Thirteen of the 35 allele D (purple)-specific changes lie within VP8*, the distal cleavage fragment of spike protein (Fig. 3B) (4, 10). Seven of these VP8* changes (T78N, D113T, N120M, N125S, R131S, G146S,

D150E, and N190S) are located within antibody neutralization domains (8-1, 8-2, 8-3, and 8-4) (Fig. 3B) (23). Thus, the exchange of a VP4 D allele (purple) for a VP4 A, B, or C allele (red, green, or cyan, respectively) via gene reassortment could result in an antigenically distinct virus. We hypothesize that gene constellations at both the genotype level and allele level are influenced, at least in part, by the coevolution of viral proteins that must interact during the replication cycle (16, 33). In this manner, exchange of genes encoding nearly identical proteins would be evolutionarily neutral in terms of viral fitness. In contrast, nonconservative protein exchanges might result in suboptimal protein interactions during replication for which the virus could compensate by also exchanging the interacting protein via gene reassortment.

To gain a better understanding of whether the VUMC RVs

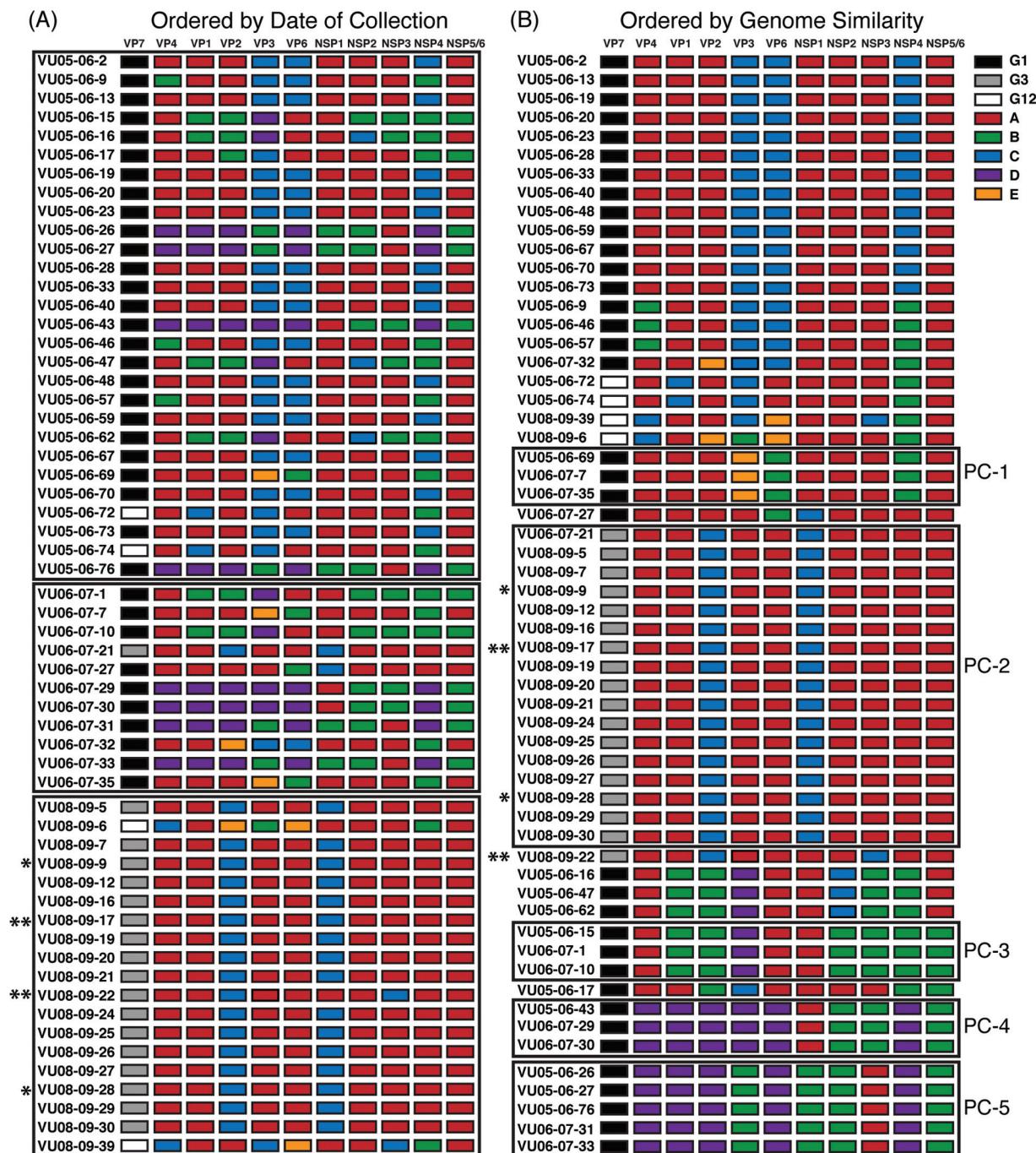


FIG 2 Allele-based gene constellations of the VUMC RVs. The schematic illustrates the color coding of each gene for the VUMC RVs based on the phylogenetic analysis shown in Fig. 1. The VUMC RV isolates are listed to the left of their corresponding allele constellations, and the protein(s) encoded by each gene is shown at the top. VP7 genes are colored according to their G types, and alleles are colored as described in the legend to Fig. 1. The allele constellations are ordered vertically by either date of collection (A) or visual allele-level genome similarity (B). Persistent clades (PCs) are comprised of isolates with identical allele constellations found in different seasons. One or two asterisks indicate viruses from children who received two or three doses of RotaTeq, respectively.

exchanged genes encoding relatively conservative versus nonconservative proteins, we analyzed three sets of putative multiallele G1P[8] reassortants from the 2005–2006 and 2006–2007 seasons (Fig. 3C). The viruses within each set showed identical allele constellations, with the exception of 2 to 3 genes, which we predict were reassorted. Set 1 RVs likely made two allele exchanges (VP4

and NSP4), each of which resulted in few amino acid changes in the encoded proteins. Set 2 viruses exchanged fairly conservative VP3 and NSP3 proteins (changing 9 or 3 amino acids, respectively) but also exchanged relatively nonconservative NSP1 proteins (78 amino acid changes). NSP1 is not expected to interact with other viral proteins, so such a dramatic swap might also be

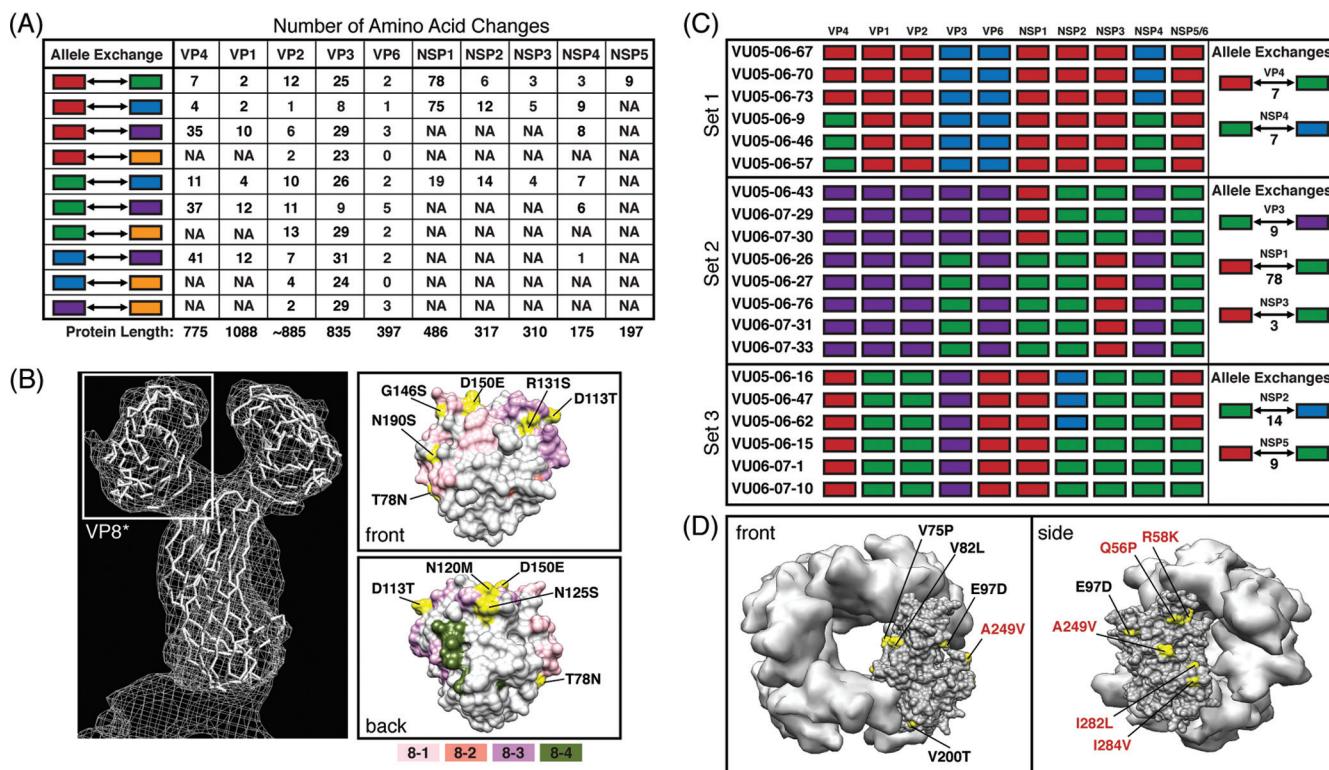


FIG 3 VUMC RV gene alleles encode divergent proteins. (A) Allele-specific amino acid changes. The table illustrates the number of amino acid changes between different alleles in pairwise combinations. This analysis is meant to determine how many amino acids would change for each protein assuming allele exchange as a result of gene reassortment. Alleles are colored as described in the legend to Fig. 1. The protein(s) encoded by each gene is shown at the top, and the length of each protein, in amino acids, is shown at the bottom. (B) Location of surface-exposed, allele D (purple)-specific VP8* amino acid changes. The left image shows a mesh-ribbon representation of the VP4 crystal structure (PDB number 1KQR). A white box defines the position of VP8*. The images on the right show the structure of VP8* (PDB number 2DWR) from the front (top) or back (bottom). Residues comprising the neutralization domains are colored as 8-1 (light pink), 8-2 (salmon), 8-3 (lavender), and 8-4 (olive). Amino acid changes seen for proteins encoded by allele D VP4 versus all other VUMC RV VP4 alleles are shown in yellow and are labeled. (C) Putative gene reassortment events and consequences at the protein level. Three different sets of RV reassortants are shown. The VUMC RV isolates are listed to the left of their corresponding allele constellations, and the protein(s) encoded by each gene is shown at the top. To the right of each grouping, a cartoon represents the allele exchanges and the number of amino acid changes (below the arrow) expected to occur in the encoded protein as a result of the gene reassortment event. (D) Location of surface-exposed, amino acid differences between the NSP2 proteins encoded by allele B (green) versus allele C (cyan). The left image shows a surface representation of an NSP2 monomer (PDB number 2R7C) in the context of an octamer (modeled), while the right image is rotated 90° to the left to show the NSP5-binding groove on the side of the protein. NSP2 amino acid changes for proteins encoded by allele B (green) versus allele C (cyan) are shown in yellow and are labeled. Those in red are at the NSP5-binding interface of NSP2.

considered neutral as related to viral protein coevolution. Set 3 viruses exchanged alleles encoding NSP2 and NSP5, resulting in a significant number of changes for these proteins considering their small sizes (15/317 and 9/197 amino acids, respectively). NSP2-NSP5 interactions play important roles during RV genome replication and core assembly (38). Nine of 15 amino acid changes between NSP2 proteins encoded by alleles B (green) and C (cyan) are surface exposed on the high-resolution structure of the octamer, and five of them (Q56P, R58K, A249V, I282L, and I284V) map to the tetramer-tetramer groove, which is a site of NSP5 binding (Fig. 3C) (21, 23). The nine changes in NSP5 (S37N, V41I, S43P, L108M, N121S, I126V, D131N, and R187Q) are generally spread throughout the linear amino acid sequence (data not shown). While no atomic structure of NSP5 exists, a region of the protein encompassing residues 66 to 188 has been shown to bind within the NSP2 groove (23). It is speculated that viruses in set 3 kept specific NSP2-NSP5 gene sets due to the important interactions among these proteins.

Comparison of VUMC RV VP1 to VP3, VP6, and NSP1 to NSP5/6 genes and proteins to those of other strains. We next

sought to determine how similar the protein gene alleles A to E of VUMC RVs are to those of other human and animal strains, particularly to those of archival and contemporary GG-1 viruses. Therefore, we constructed maximum likelihood phylogenetic trees for VP1 to VP3, VP6, and NSP1 to NSP5/6 using the ORF sequences of representative GG-1 strains for which complete genome sequences exist in GenBank (Fig. 4; also see Table S3 in the supplemental material). The results show that, in general, the VUMC RV alleles A to E group with those of other contemporary strains (isolated during 2002 to 2010) from various regions of the world, including 6361 and 061060 (India), BE00036 and BE00029 (Belgium), Dhaka16-03 and Matlab36-02 (Bangladesh), GER126-08 and GER172-08 (Germany), 2008747332 and 2008747336 (United States), and CK00005 and CK00034 (Australia) (Fig. 4; also see Table S3). The VP1 to VP3, VP6, NSP1, and NSP5/6 genes of contemporary human RVs are found in one or two putative clusters and are distinct from the genotype 1 genes of animal strains (e.g., YM, OSU, A253, etc.) (Fig. 4). One cluster seems to share evolutionary relationships with the genes of the GG-1 neonatal strain ST3 (G4P[6]), whereas another cluster con-

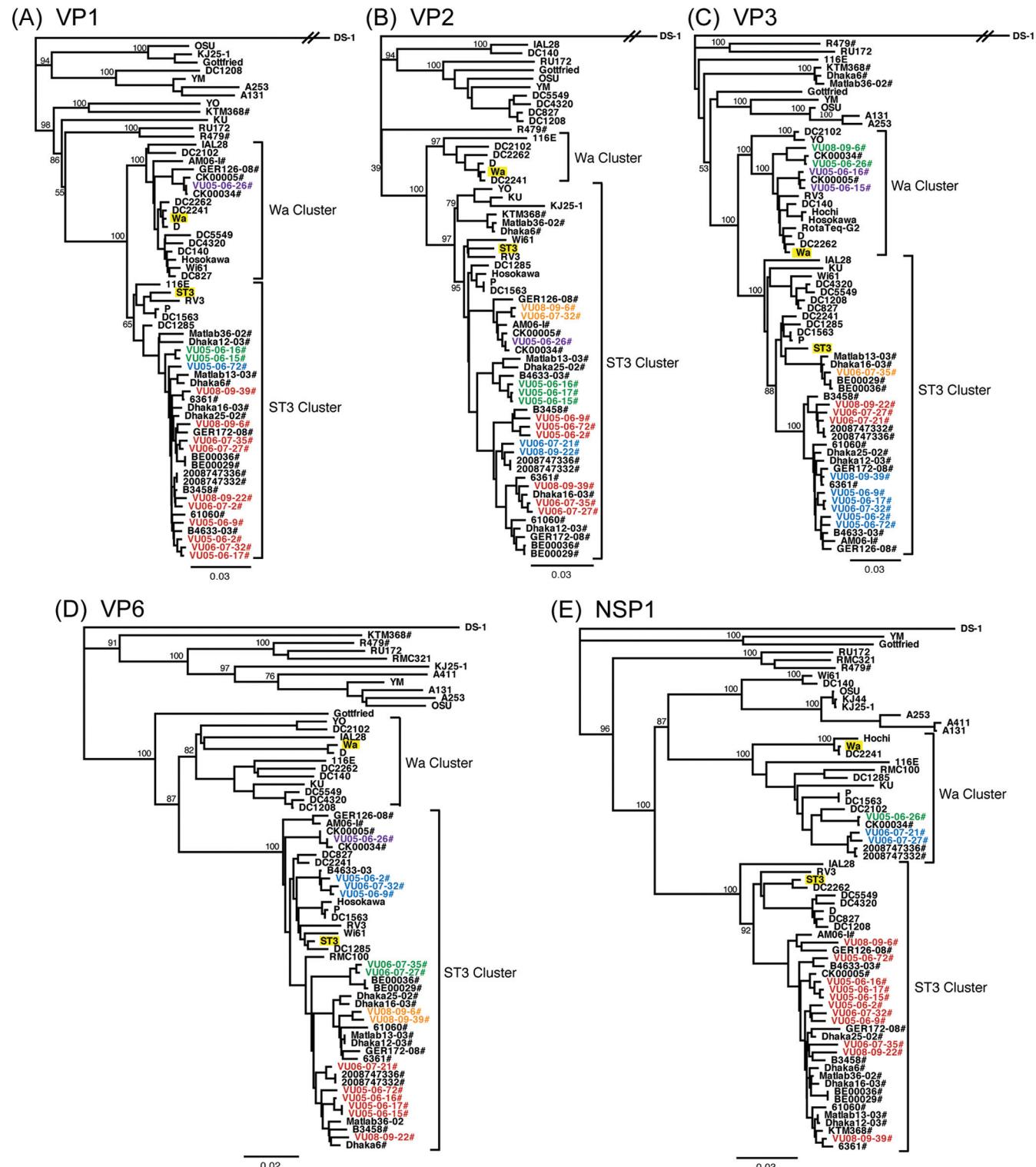


FIG 4 Comparison of VUMC RV VP1 to VP3, VP6, and NSP1 to NSP5/6 genes and proteins to those of other strains. The maximum likelihood trees were constructed using the ORF nucleotide sequences for representative human and animal strains (listed by common names). Horizontal branch lengths are drawn to scale (nucleotide substitutions per base), and bootstrap values are shown as percentages for key nodes. All trees are outgroup rooted to the GG-2 prototype strain DS-1 (G2-P[4]-I2-R2-C2-M2-A2-N2-T2-E2-H2). Representative VUMC RV alleles are indicated with a pound sign (#) following the strain name. Prototype strains ST3 and Wa are highlighted in yellow, and putative phylogenetic clusters related to these strains are indicated using brackets and are labeled.

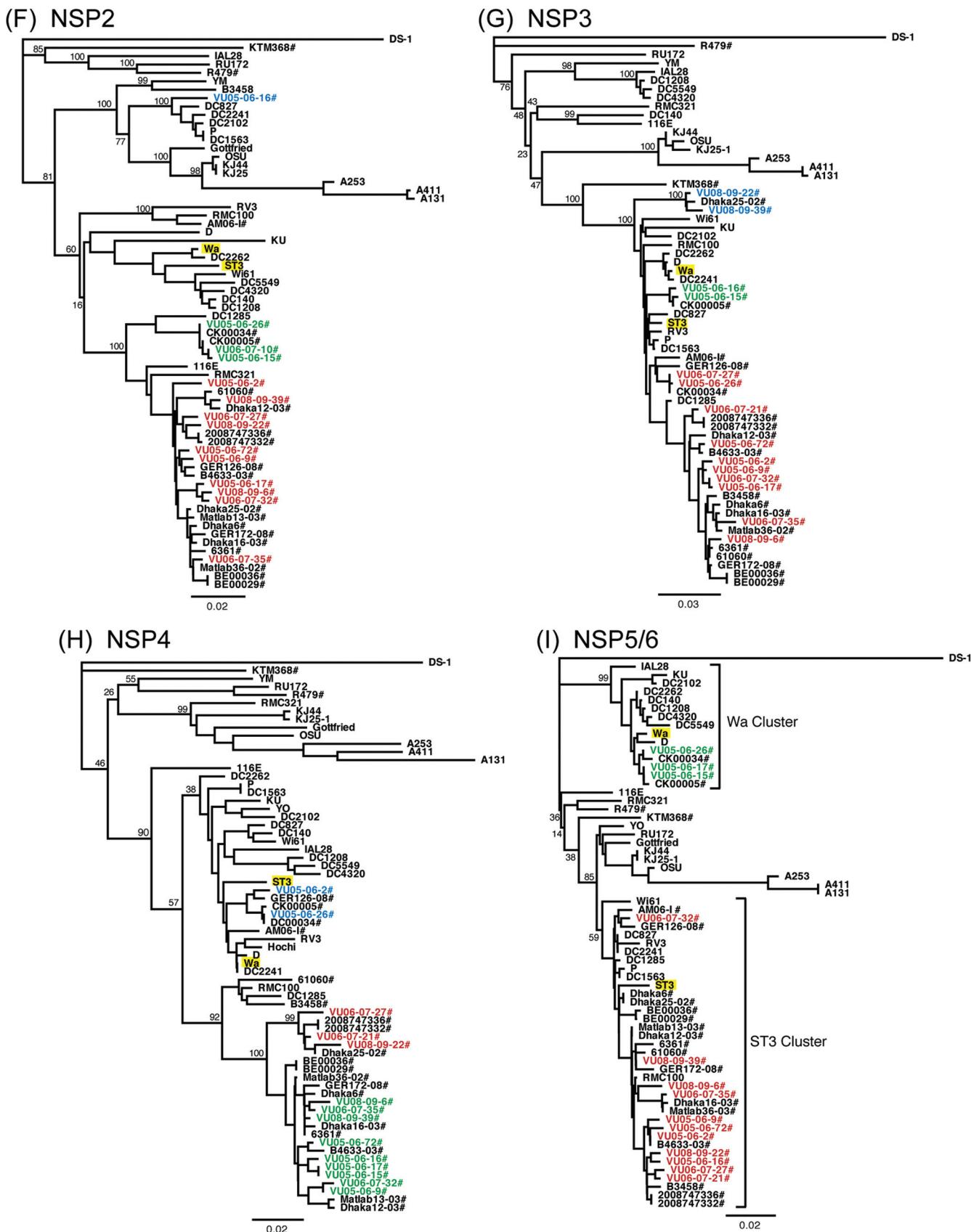


FIG 4 Continued

tains genes more closely related to those of the classic GG-1 prototype strain Wa (G1P[8]) (Fig. 4). To date, there is no evidence of modern VP2 or VP6 genes in the Wa cluster. For NSP2, NSP3, and NSP4, the bootstrap values were too low to clearly resolve any specific clusters of human RV genotype 1 genes (Fig. 4F to H). However, the NSP2 C allele (cyan), which is represented by VU05-06-16, seems more divergent than any of the modern RV genes and grouped closely with the NSP2 genes of several archival DC RVs (DC827, DC2241, and DC2102) in the phylogenetic tree (Fig. 4F).

Together, the analyses shown in Fig. 3 and 4 reveal that the VUMC RV alleles found to encode divergent proteins may actually belong to different phylogenetic clusters (i.e., Wa cluster versus ST3 cluster). For example, VP3 alleles B and D (green and purple, respectively) both belong to the Wa cluster and show only 9 amino acid changes between them (Fig. 3A and 4). However, VP3 allele B (green; Wa cluster) versus A, C, or D (red, cyan, or orange; ST3 cluster) exhibits 25 to 29 amino acid changes. This result expands our allele-based analysis of gene reassortment and suggests that the exchange of gene alleles belonging to the same evolutionary cluster occurs more readily than the exchange of gene alleles belonging to different clusters because they alter similar proteins.

Comparison of VUMC RV VP7 and VP4 genes and proteins to those of other strains. Since the 58 VUMC RVs were of various G-type specificities, we did not define subgenotype alleles for the VP7 genes. Nonetheless, we did seek to determine how similar the VP7 genes and proteins of these contemporary Nashville strains were to each other and to those of other modern, archival, and vaccine strains. This analysis is important, as anti-VP7 neutralizing antibodies are thought to play a critical role in vaccine-mediated immunological protection against RV disease. To determine the genetic relationships of VUMC RV VP7 genes to other strains, we first constructed maximum likelihood phylogenetic trees using the ORF sequences of representative G1, G3, or G12 RVs (Fig. 5A to C) (see Table S4). For the 36 VUMC strains with G1 specificity, the VP7 sequences cluster along with those of other contemporary strains in well-characterized lineage 1 or 2, and all of them are distinct from VP7 genes in lineages 3 (strains D and Wa) and 4 (strains K54, 421, and Kor-64) (Fig. 5A) (31). The VP7 gene of the Rotarix vaccine clusters phylogenetically with those of many modern G1 strains of lineage 1. The VP7 gene of the G1 component strain of the RotaTeq vaccine, on the other hand, is more divergent from those of contemporary strains and is found in lineage 3, consistent with what has been described previously (31).

This analysis revealed that the G1 VP7 genes belonging to both lineage 1 (19 VUMC RVs) and lineage 2 (17 VUMC RVs) were found to cocirculate in Nashville in 2005 to 2006. These lineages correspond to two phylogenetic clusters observed in the trees created using only VUMC RV VP7 gene sequences and could represent distinct subgenotype G1 alleles (Fig. 1A). By creating amino acid sequence alignments, we found that proteins encoded by G1 VP7 lineage 1 and 2 genes differ at nine positions (S37F, R49K, I55L, L57I, A66V, A68S, N94S, S123N, M217T, T281I, and K291R) (data not shown). When mapped onto the structure of the VP7 trimer, four amino acid changes (N94S, S123N, M217T, and K291R) are found in or near VP7 neutralization domains 7-1A, 7-1B, and 7-2 (Fig. 5A, inset) (1, 33). The G1 VP7 proteins of both RotaTeq and Rotarix match those encoded by lineage 1 genes at each of these four sites. A recent analysis of the G1 VP7 proteins of

contemporary Belgium strains versus those in RotaTeq and Rotarix also identified these four residues (56). This result suggests that G1 lineage 2 VP7 proteins are antigenically distinct from those of the G1 lineage 1 proteins as well as those of the current vaccine strains.

For the VUMC RVs with G3 specificity, the VP7 sequences were found to cluster with those of other contemporary strains within lineage 1 (Fig. 5B) (31). The VP7 gene of the G3 component strain of the RotaTeq vaccine is genetically more divergent from those of modern strains and groups within lineage 2, in agreement with other reports (31, 56). By performing amino acid alignments, we found six residue changes (P66S, E104G, A212T, K238N, D242N, and A278M) between the VP7 of RotaTeq and those of modern strains, including the 18 G3P[8] VUMC RVs (data not shown). Three of the changes (A212T, K238N, and D242N) are located in or near neutralization domains on the trimer structure (7-1A, 7-1B, and 7-2) and might change VP7 antigenicity (Fig. 5B, inset). Changes at positions 212, 238, and 242 were also found in the analysis of G3 VP7 proteins from contemporary Belgium strains versus RotaTeq G3 VP7 (56).

The VP7 sequences of the four VUMC RVs with G12 specificity were found within lineage 3, separate from those of lineage 1 (strain L26) and 2 (strains 10941 and K12) (Fig. 5C) (56). We found that lineage 3 could be further defined into 2 sublineages (3-A and 3-B), both of which contain contemporary human RVs. This observation suggests that lineages 3-A and 3-B diverged prior to the estimated time of G12 introduction to the United States in 1999 to 2000 (56). All four G12 strains sequenced in this study have lineage 3-B VP7 genes. However, there are no amino acid changes that distinguished lineage 3-A from 3-B G12 VP7 proteins (data not shown). Thus, while these genes are divergent at the nucleotide level, they are expected to encode antigenically identical VP7 proteins.

In addition to VP7, the spike attachment protein VP4 also induces neutralizing antibodies during infection and is important for immunological protection against RV disease. Thus, we sought to determine how similar the VP4 genes and proteins of the contemporary VUMC RV strains were to each other and to those of other modern, archival, and vaccine strains. To do this, we constructed maximum likelihood phylogenetic trees using the VP4 ORF sequences of representative P[8] RVs (Fig. 5D) (see Table S4). The results show that the VP4 genes of contemporary RVs cluster into two known phylogenetic lineages, 1 and 3; RVs with both lineage 1 (8 VUMC RVs) and lineage 3 (50 VUMC RVs) genes cocirculated in Nashville in 2005–2006 and 2006–2007 seasons (31). Lineage 1 is comprised mainly of the VP4 genes from archival DC RVs DC2262, DC2241, and DC2102 and the prototypic laboratory strains Hochi, Odelia, D, and Wa. However, the VP4 genes of VUMC RVs designated allele D (purple), and from other contemporary strains for which sequences are available in GenBank, form a cluster within lineage 1. These other allele D-like sequences are from human RV isolated after 2006 in the United States (strains 20007744509 and 2008747288), Belgium (strain BE00010), and Australia (strains CK00005 and CK00034). The VP4 gene of Rotarix is also lineage 1 and is closely related to this contemporary allele D-like cluster (56).

The VP4 gene of the P[8] component strain of the RotaTeq vaccine, considered lineage 2, is genetically more divergent from modern strains, in agreement with a previously published report (31). At the protein level, the RotaTeq P[8] VP4 protein is differ-

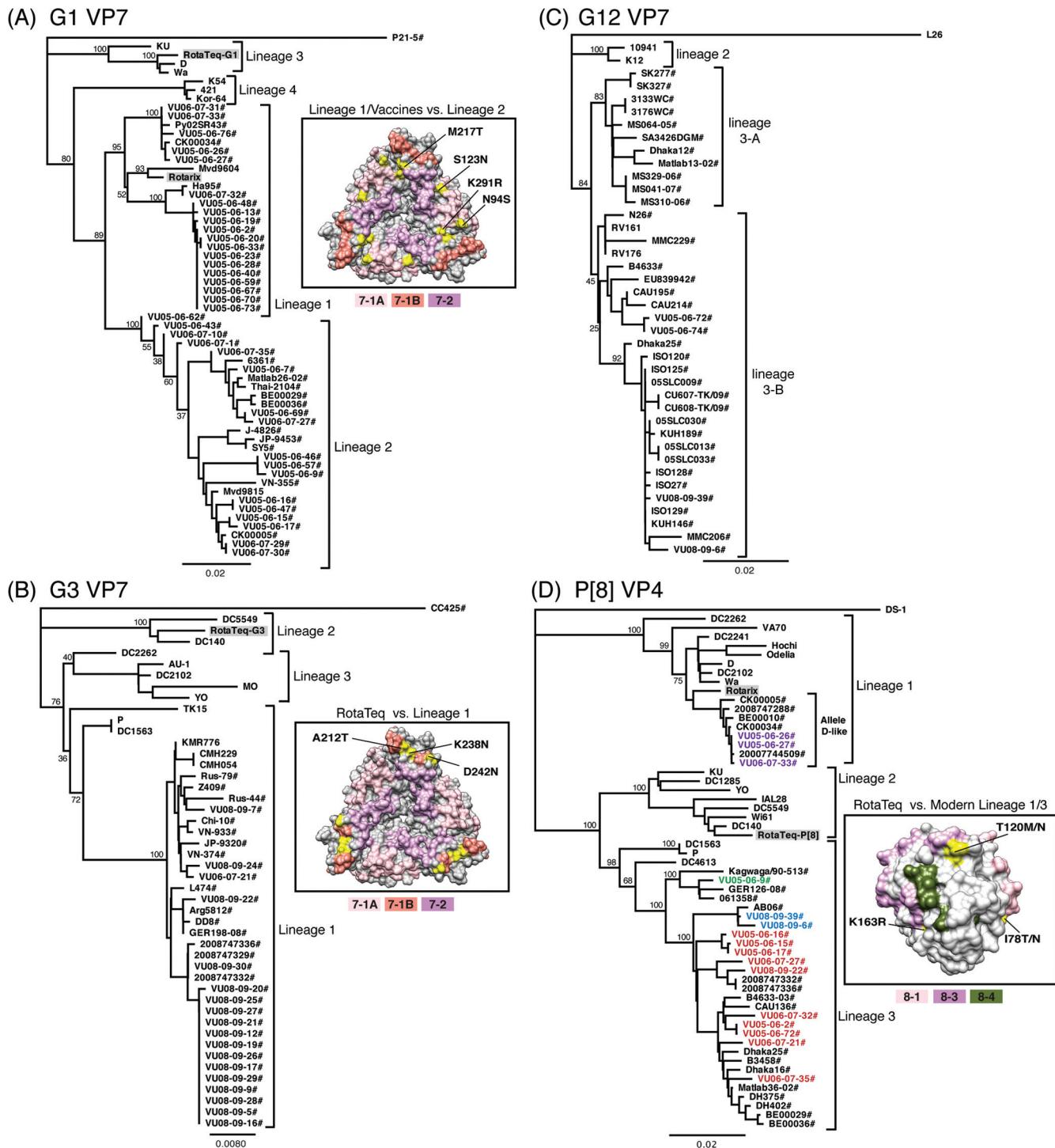


FIG 5 Comparison of VUMC RV VP7 and VP4 genes and proteins to those of other strains. The maximum likelihood trees were constructed using the VP7 ORF nucleotide sequences and are outgroup rooted to the most recent common ancestor. All horizontal branch lengths are drawn to scale (nucleotide substitutions per base), and bootstrap values are shown as percentages for key nodes. Lineages are indicated using brackets and are labeled. Human RV strains isolated during 2002 to 2010 are indicated with a pound sign (#) following the common strain name. Vaccine strains are highlighted with a gray box. (A) Phylogenetic analysis of G1 VP7 sequences. The inset image is a surface representation of the VP7 trimer (PDB number 3FMG). Residues comprising neutralization domains of VP7 are colored as 7-1A (light pink), 7-1B (salmon), and 7-2 (lavender). Amino acids that differ between VP7 proteins encoded by lineage 1 versus 2 genes are shown in yellow and are labeled. The G1 VP7 protein of the RotaTeq and Rotarix vaccines match lineage 1 proteins at these four sites. (B) Phylogenetic analysis of G3 VP7 sequences. The inset image is a surface representation of the VP7 trimer with neutralization domains colored as described for panel A. Amino acid differences between the G3 VP7 of RotaTeq and those of lineage 1 modern human RV strains are shown in yellow and are labeled. (C) Phylogenetic analysis of G12 VP7 sequences. (D) Phylogenetic analysis of P[8] VP4 sequences. Representative VUMC RV VP4 alleles are colored as described in the legend to Fig. 1. The inset image is a surface representation of VP8* (PDB number 2DWR) as shown from the back, similar to the bottom right image of Fig. 3B. Residues comprising the neutralization domains are colored as 8-1 (light pink), 8-3 (lavender), and 8-4 (olive). Amino acid changes seen for RotaTeq VP8* versus those of most contemporary strains are shown in yellow and are labeled.

ent from those encoded by the contemporary lineage 1 and 3 genes. Specifically, there are 40 amino acid differences between the P[8] VP4 protein of RotaTeq and the P[8] VP4 protein of contemporary lineage 1 strains (i.e., those encoded by allele D-like genes) (data not shown). The VP4 proteins of RotaTeq and contemporary lineage 3 strains are more similar but still show 28 amino acid changes when aligned to each other (data not shown). We found three amino acid residues (I78, T120, and K163) of RotaTeq VP4 that (i) differ from nearly every modern P[8] strain (i.e., isolated since 2000) for which sequences are available and (ii) are located in or near neutralization domain 8-1, 8-3, or 8-4 of VP8*, the distal cleavage product of trypsin-activated VP4 (Fig. 5D, inset). The location of these three residues suggests that they influence the antigenicity of VP4.

DISCUSSION

RV is a ubiquitous pediatric pathogen, infecting nearly every unvaccinated child by age 5 (27). Despite their medical significance, the genetic diversity of human RVs is not yet fully known, mainly due to the lack of complete genome sequences. In this study, we deduced the sequences of 58 human RVs isolated as part of active, prospective surveillance for community-acquired gastroenteritis at VUMC in Nashville, TN. The viruses are contemporary strains from the years 2005 to 2009, and they represent G/P types that are dominant in many regions of the world today. The years of sample collection spanned the widespread implementation of RotaTeq vaccination in the Nashville community (40, 41). We found that VUMC RVs circulating in the prevaccine season (i.e., the 2005–2006 season) were different from those circulating 2 to 3 years following community RotaTeq implementation (i.e., the 2008–2009 season). Specifically, in the 2005–2006 season, a heterogeneous mix of G1P[8] RVs, with various allele constellations and antigenically distinct VP7 and VP4 proteins, cocirculated in Nashville, TN. In contrast, the 2008–2009 season was marked by a single genetically and antigenically homogeneous G3P[8] RV clade. The 2006–2007 season, while technically considered postvaccine, had only an ~36% community vaccine compliance rate (40). The RV strains appearing in the 2006–2007 season are more similar to those found in the 2005–2006 season. While it is tempting to speculate that the differences in VUMC RVs from 2005–2006 and 2006–2007 seasons versus the 2008–2009 season is attributable to RotaTeq introduction, they might instead just reflect the well-documented, albeit poorly understood, seasonal fluctuation of RV G/P types (7, 17, 44). It is important to note that the majority of VUMC RVs were isolated from children who themselves did not receive any RV vaccine (Fig. 2; also see Table S1 in the supplemental material). However, two G3P[8] RV isolates (VU08-09-9 and VU8-09-28) were from children who received two doses of RotaTeq, and two G3P[8] RV isolates (VU08-09-17 and VU08-09-22) were from children vaccinated with all three doses (see Table S1). Overall, we found no VP7 or VP4 amino acid changes between G3 viruses from vaccinated and unvaccinated children. In fact, full-genome inspection of these isolates did not reveal any appreciable differences from RVs found in vaccinated versus unvaccinated children. The reason for the lack of protection against RV disease for these vaccinated children is not known, but immunologically driven evasion mechanisms do not appear to be at work.

The complete genome sequence analysis of the 58 VUMC RVs revealed that they belong in GG-1, and none contained genotype 2

or animal RV-like genes. Still, a major limitation of this study is in the design of primers to preferentially amplify genotype 1 genes during RT-PCR and sequencing via the established, semiautomated pipeline at the J. Craig Venter Institute. Future studies will rely on sequence-independent amplification and sequencing approaches to address the frequency of gene reassortment between GG-1 and GG-2 strains (22, 46). Nonetheless, the fact that all 58 VUMC RVs were GG-1, with genotype 1 internal protein genes, afforded us the unique opportunity to analyze modern RV diversity at the subgenotype level. As was done in our previous analyses of archival GG-1 RVs from Washington, DC, we used phylogenetic trees to designate each gene of each VUMC RV (except VP7) into alleles (32, 33). By plotting the allele constellations of the VUMC isolates, it became clear that genetic exchange occurred among these strains. However, we note that, given our level of analysis, it is not possible to determine whether such exchanges occurred during (rather than prior to) 2005 to 2009. Interestingly, we found that some allele constellations persisted over a 1- to 3-year time span, despite the observation that gene reassortment was possible for other cocirculating strains. Why these RV PCs continued to circulate in Nashville, nearly unchanged, over time is a fascinating question. One hypothesis is that specific combinations of alleles were advantageous to the viruses and provided selective pressure(s) against gene reassortment. The basis for preferred gene sets might be related to the interactions of the encoded viral proteins during replication. In this manner, we expect more closely related genes, such as those belonging to the same phylogenetic lineage, to be reassorted more freely. Future experiments in our laboratories will use reverse-genetic and biochemical approaches to test the hypothesis that certain alleles encode proteins that interact preferentially (51). We cannot exclude the alternative explanation that the PCs were maintained simply due to lack of gene reassortment opportunities. Specifically, instead of children being inoculated with a heterogeneous mix of GG-1 viruses (i.e., with several different allele constellations), they might have been infected with a single strain. As such, the lack of intrahost RV diversity would abrogate the chance for gene reassortment as analyzed at the allele level. Future deep-sequencing studies will elucidate the quasi-species dynamics of RVs within a single host, in addition to revealing reassortment frequencies.

In this study, we also analyzed the genetic and antigenic similarities among these 58 modern VUMC RVs and compared their VP7 and VP4 genes/proteins to those of other strains. We found that G1, G3, and P[8] RVs can encode different types of VP7 and/or VP4, which show changes in regions involved in neutralizing antibody binding. The VP7 and VP4 proteins of VUMC RVs and other modern strains also differ from those of the currently available vaccines, RotaTeq and Rotarix, a result consistent with recent findings by Zeller et al. (56). It is conceivable that amino acid differences between vaccine strains and currently circulating RVs could set the stage for the emergence of vaccine-resistant variants. However, immunological protection against RV is thought to be both homotypic (the same G/P type) as well as heterotypic (different G/P types) (9, 52). For example, Rotarix, which only contains the G1 and P[8] antigens, has proven effective at preventing moderate to severe gastroenteritis against non-G1P[8] strains (36). Still, given the evolutionary potential for RVs to rapidly adapt through gene reassortment and the accumulation of point mutations, ongoing genetic surveillance of strains is imperative. The complete genome sequences of human RVs circu-

lating in a defined region during or just prior to widespread vaccination will provide a baseline for monitoring changes in the viral landscape that could affect vaccine performance over time.

ACKNOWLEDGMENTS

We thank Elizabeth Teel, Mike Bowen, and Jon Gentsch (Centers for Disease Control and Prevention) for scientific support in the collection and typing of fecal specimens and for editorial suggestions on the manuscript. We also acknowledge David Spiro for assistance in the initial development of the sequencing pipeline at the J. Craig Venter Institute.

S.M.M. and A.O.M. were supported by the Virginia Tech Carilion Research Institute. J.T.P. and C.M.R. were supported by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases, National Institutes of Health. This project was also supported in part by a contract from the Centers for Disease Control and Prevention (contract no. 1U01IP00022) and by federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services (contract no. HHSN272200900007C).

REFERENCES

- Aoki ST, et al. 2009. Structure of rotavirus outer-layer protein VP7 bound with a neutralizing Fab. *Science* 324:1444–1447.
- Armah GE, et al. 2010. Efficacy of pentavalent rotavirus vaccine against severe rotavirus gastroenteritis in infants in developing countries in sub-Saharan Africa: a randomised, double-blind, placebo-controlled trial. *Lancet* 376:606–614.
- Benati FJ, Maranhao AG, Lima RS, da Silva RC, Santos N. 2010. Multiple-gene characterization of rotavirus strains: evidence of genetic linkage among the VP7-, VP4-, VP6-, and NSP4-encoding genes. *J. Med. Virol.* 82:1797–1802.
- Blanchard H, Yu X, Coulson BS, von Itzstein M. 2007. Insight into host cell carbohydrate-recognition by human and porcine rotavirus from crystal structures of the virion spike associated carbohydrate-binding domain (VP8*). *J. Mol. Biol.* 367:1215–1226.
- Carvalho-Costa FA, et al. 2011. Laboratory-based rotavirus surveillance during the introduction of a vaccination program, Brazil, 2005–2009. *Pediatr. Infect. Dis. J.* 30:S35–S41.
- Chaimongkol N, et al. 2012. Genotypic linkages of gene segments of rotaviruses circulating in pediatric patients with acute gastroenteritis in Thailand. *Infect. Genet. Evol.* 12:1381–1391.
- Clark HF, Lawley D, DiStefano D, Matthijssens J, Dinubile MJ. 2011. Distribution of rotavirus genotypes causing nosocomial and community-acquired acute gastroenteritis at the Children's Hospital of Philadelphia in the new rotavirus vaccine era. *Hum. Vaccine* 7:1118–1123.
- Cortese MM, Parashar UD. 2009. Prevention of rotavirus gastroenteritis among infants and children: recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Recomm. Rep.* 58:1–25.
- Desselberger U, Huppertz HI. 2011. Immune responses to rotavirus infection and vaccination and associated correlates of protection. *J. Infect. Dis.* 203:188–195.
- Dormitzer PR, Sun ZY, Wagner G, Harrison SC. 2002. The rhesus rotavirus VP4 sialic acid binding domain has a galectin fold with a novel carbohydrate binding site. *EMBO J.* 21:885–897.
- Estes MK, Kapikian AZ. 2007. Rotaviruses and their replication, p 1917–1974. In *Fields virology* (5th ed), Knipe DM, et al, Lippincott Williams & Wilkins, Philadelphia, PA.
- Gentsch JR, et al. 2009. G and P types of circulating rotavirus strains in the United States during 1996–2005: nine years of prevaccine data. *J. Infect. Dis.* 200(Suppl. 1):S99–S105.
- Gentsch JR, Parashar UD, Glass RI. 2009. Impact of rotavirus vaccination: the importance of monitoring strains. *Future Microbiol.* 4:1231–1234.
- Gouvea V, Brantly M. 1995. Is rotavirus a population of reassortants? *Trends Microbiol.* 3:159–162.
- Gray J. 2011. Rotavirus vaccines: safety, efficacy and public health impact. *J. Intern. Med.* 270:206–214.
- Heiman EM, et al. 2008. Group A human rotavirus genomics: evidence that gene constellations are influenced by viral protein interactions. *J. Virol.* 82:11106–11116.
- Hull JJ, et al. 2011. United States rotavirus strain surveillance from 2005 to 2008: genotype prevalence before and after vaccine introduction. *Pediatr. Infect. Dis. J.* 30:S42–S47.
- Iturriaga-Gomara M, et al. 2003. Evidence for genetic linkage between the gene segments encoding NSP4 and VP6 proteins in common and reassortant human rotavirus strains. *J. Clin. Microbiol.* 41:3566–3573.
- Iturriaga-Gomara M, et al. 2011. Rotavirus genotypes co-circulating in Europe between 2006 and 2009 as determined by EuroRotaNet, a pan-European collaborative strain surveillance network. *Epidemiol. Infect.* 139:895–909.
- Iturriaga-Gomara M, et al. 2009. Rotavirus surveillance in europe, 2005–2008: web-enabled reporting and real-time analysis of genotyping and epidemiological data. *J. Infect. Dis.* 200(Suppl. 1):S215–S221.
- Jayaram H, Taraporewala Z, Patton JT, Prasad BV. 2002. Rotavirus protein involved in genome replication and packaging exhibits a HIT-like fold. *Nature* 417:311–315.
- Jere KC, et al. 2011. Whole genome analyses of African G2, G8, G9, and G12 rotavirus strains using sequence-independent amplification and 454(R) pyrosequencing. *J. Med. Virol.* 83:2018–2042.
- Jiang X, et al. 2006. Cryoelectron microscopy structures of rotavirus NSP2-NSP5 and NSP2-RNA complexes: implications for genome replication. *J. Virol.* 80:10829–10835.
- Kirkwood CD, Boniface K, Bishop RF, Barnes GL. 2010. Australian Rotavirus Surveillance Program: annual report, 2009/2010. *Commun. Dis. Intell.* 34:427–434.
- Madhi SA, et al. 2010. Effect of human rotavirus vaccine on severe diarrhea in African infants. *N. Engl. J. Med.* 362:289–298.
- Maes P, Matthijssens J, Rahman M, Van Ranst M. 2009. RotaC: a web-based tool for the complete genome classification of group A rotaviruses. *BMC Microbiol.* 9:238. doi:10.1186/1471-2180-9-238.
- Malek MA, et al. 2006. Diarrhea- and rotavirus-associated hospitalizations among children less than 5 years of age: United States, 1997 and 2000. *Pediatrics* 117:1887–1892.
- Matthijssens J, et al. 2008. Full genome-based classification of rotaviruses reveals a common origin between human Wa-Like and porcine rotavirus strains and human DS-1-like and bovine rotavirus strains. *J. Virol.* 82:3204–3219.
- Matthijssens J, et al. 2011. Uniformity of rotavirus strain nomenclature proposed by the Rotavirus Classification Working Group (RCWG). *Arch. Virol.* 156:1397–1413.
- Matthijssens J, et al. 2010. Phylodynamic analyses of rotavirus genotypes G9 and G12 underscore their potential for swift global spread. *Mol. Biol. Evol.* 27:2431–2436.
- Matthijssens J, et al. 2010. Molecular and biological characterization of the 5 human-bovine rotavirus (WC3)-based reassortant strains of the pentavalent rotavirus vaccine, RotaTeq. *Virology* 403:111–127.
- McDonald SM, Davis K, McAllen JK, Spiro DJ, Patton JT. 2011. Intra-genotypic diversity of archival G4P[8] human rotaviruses from Washington, DC. *Infect. Genet. Evol.* 11:1586–1594.
- McDonald SM, et al. 2009. Evolutionary dynamics of human rotaviruses: balancing reassortment with preferred genome constellations. *PLoS Pathog.* 5:e1000634. doi:10.1371/journal.ppat.1000634.
- McDonald SM, Patton JT. 2011. Assortment and packaging of the segmented rotavirus genome. *Trends Microbiol.* 19:136–144.
- Nakagomi O, Nakagomi T, Akatani K, Ikegami N. 1989. Identification of rotavirus genogroups by RNA-RNA hybridization. *Mol. Cell. Probes* 3:251–261.
- O'Ryan M, Linhares AC. 2009. Update on Rotarix: an oral human rotavirus vaccine. *Expert Rev. Vaccines* 8:1627–1641.
- Parashar UD, Gibson CJ, Bresse JS, Glass RI. 2006. Rotavirus and severe childhood diarrhea. *Emerg. Infect. Dis.* 12:304–306.
- Patton JT, Silvestri LS, Tortorici MA, Vasquez-Del Carpio R, Taraporewala ZF. 2006. Rotavirus genome replication and morphogenesis: role of the viroplasm. *Curr. Top. Microbiol. Immunol.* 309:169–187.
- Payne DC, et al. 2008. Active, population-based surveillance for severe rotavirus gastroenteritis in children in the United States. *Pediatrics* 122:1235–1243.
- Payne DC, et al. 2011. Direct and indirect effects of rotavirus vaccination upon childhood hospitalizations in 3 US Counties, 2006–2009. *Clin. Infect. Dis.* 53:245–253.
- Payne DC, et al. 2009. Secular variation in United States rotavirus disease rates and serotypes: implications for assessing the rotavirus vaccination program. *Pediatr. Infect. Dis. J.* 28:948–953.

42. Pettersen EF, et al. 2004. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25:1605–1612.
43. Phua KB, et al. 2009. Safety and efficacy of human rotavirus vaccine during the first 2 years of life in Asian infants: randomised, double-blind, controlled study. *Vaccine* 27:5936–5941.
44. Pitzer VE, et al. 2011. Modeling rotavirus strain dynamics in developed countries to understand the potential impact of vaccination on genotype distributions. *Proc. Natl. Acad. Sci. U. S. A.* 108:19353–19358.
45. Posada D, Buckley TR. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Syst. Biol.* 53:793–808.
46. Potgieter AC, et al. 2009. Improved strategies for sequence-independent amplification and sequencing of viral double-stranded RNA genomes. *J. Gen. Virol.* 90:1423–1432.
47. Rahman M, et al. 2007. Evolutionary history and global spread of the emerging g12 human rotaviruses. *J. Virol.* 81:2382–2390.
48. Ramani S, et al. 2007. Geographic information systems and genotyping in identification of rotavirus G12 infections in residents of an urban slum with subsequent detection in hospitalized children: emergence of G12 genotype in south India. *J. Clin. Microbiol.* 45:432–437.
49. Takahashi K, Nei M. 2000. Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used. *Mol. Biol. Evol.* 17:1251–1258.
50. Tate JE, et al. 2012. 2008 estimate of worldwide rotavirus-associated mortality in children younger than 5 years before the introduction of universal rotavirus vaccination programmes: a systematic review and meta-analysis. *Lancet Infect. Dis.* 12:136–141.
51. Trask SD, Taraporewala ZF, Boehme KW, Dermody TS, Patton JT. 2010. Dual selection mechanisms drive efficient single-gene reverse genetics for rotavirus. *Proc. Natl. Acad. Sci. U. S. A.* 107:18652–18657.
52. Vesikari T, et al. 2007. Efficacy of human rotavirus vaccine against rotavirus gastroenteritis during the first 2 years of life in European infants: randomised, double-blind controlled study. *Lancet* 370:1757–1763.
53. Ward RL, Bernstein DI. 2009. Rotarix: a rotavirus vaccine for the world. *Clin. Infect. Dis.* 48:222–228.
54. Widdowson MA, et al. 2007. Cost-effectiveness and potential impact of rotavirus vaccination in the United States. *Pediatrics* 119:684–697.
55. Zaman K, et al. 2010. Efficacy of pentavalent rotavirus vaccine against severe rotavirus gastroenteritis in infants in developing countries in Asia: a randomised, double-blind, placebo-controlled trial. *Lancet* 376:615–623.
56. Zeller M, et al. 2012. Genetic analyses reveal differences in the VP7 and VP4 antigenic epitopes between human rotaviruses circulating in Belgium and rotaviruses in Rotarix and RotaTeq. *J. Clin. Microbiol.* 50:966–976.