# Intragenic Recombination Influences Rotavirus Diversity and Evolution

Irene Hoxie[a,b] and John J. Dennehy[a,b,#]

[a]Biology Department, Queens College of The City University of New York, Queens, NY

[b]The Graduate Center of The City University of New York, New York, NY

Running head: Rotavirus Recombination and Evolution

#Address correspondence to John J. Dennehy, john.dennehy@qc.cuny.edu

Word counts

Abstract: 212

Text: excl. refs, table and figure legends

*Abstract*

Because of their replication mode and segmented dsRNA genome, homologous recombination is assumed to be rare in the rotaviruses. We analyzed 23,627 complete rotavirus genome sequences available in the NCBI Virus Variation database and found numerous instances of homologous recombination, some of which prevailed across multiple sequenced isolates. In one case, recombination may have generated a novel rotavirus VP1 lineage. We also found strong evidence for intergenotypic recombination in which more than one sequence strongly supported the same event, particularly between different genotypes of the serotype protein, VP7. In addition, we found evidence for recombination in nine of eleven rotavirus A segments. Only two segments, 7 (NSP3) and 11 (NSP5/NSP6), did not show strongly supported evidence of prior recombination. The abundance of putative recombinants with substantial amino acid replacements occurring between different G types suggests a fitness advantage for the resulting recombinant strains. Protein structural predictions indicated that despite the sometimes substantial, amino acid replacements resulting from recombination, the overall protein structures remained relatively unaffected. Notably, recombination junctions appear to occur non-randomly with hot spots corresponding to secondary RNA structures, a pattern that is seen consistently across segments. Collectively, the results of our computational analyses suggest that, contrary to the prevailing sentiment, recombination may be a significant driver rotavirus evolution and influence circulating strain diversity.

*Keywords*: dsRNA, genetic diversity, *Reoviridae*, segmented genome, RDP4

## Introduction

The non-enveloped, dsRNA rotaviruses of the family *Reoviridae* are a common cause of acute gastroenteritis in young individuals of many bird and mammal species (1). The rotavirus genome consists of 11 segments, each coding for a single protein with the exception of segment 11, which encodes two proteins, NSP5 and NSP6 (1). Six of the proteins are structural proteins (VP1-4, VP6 and VP7), and the remainder are non-structural proteins (NSP1-6). The infectious virion is a triple-layered particle consisting of two outer-layer proteins, VP4 and VP7, a middle layer protein, VP6, and an inner capsid protein, VP2. The RNA polymerase (VP1) and the capping-enzyme (VP3) are attached to the inner capsid protein. For the virus to be infectious (at least when not infecting as an extracellular vesicle), the VP4 spike protein's head (VP8*) must be cleaved by a protease, which results in the protein VP5* (2). Because they comprise the outer layer of the virion, VP7 and VP4 are capable of eliciting neutralizing antibodies, and are used to define G (glycoprotein) and P (protease sensitive) serotypes respectively (3, 4). Consequently, VP7 and VP4 are likely to be under strong selection for diversification to mediate cell entry or escape host immune responses (5-7).

Based on sequence identity and antigenic properties of VP6, 10 different rotavirus groups (A–J) have been identified, with group A rotaviruses being the most common cause of human infections (8-10). A genome classification system based on established nucleotide percent cutoff values has been developed for group A rotaviruses (3, 11). In the classification system, the segments VP7-VP4-VP6-VP1-VP2-VP3-NSP1-NSP2-NSP3-NSP4-NSP5/6 are represented by the indicators Gx-P[x]-Ix-Rx-Cx-Mx-Ax-Nx-Tx-Ex-Hx, (x = Arabic numbers starting from 1), respectively (3, 11). To date, between 20 to 51 different genotypes have been identified for each segment, including 51 different VP4 genotypes (P[1]–P[51]) and 36 different VP7 genotypes (G1–G36), both at 80% nucleotide identity cutoff values (12).

The propensity of rotavirus for coinfection and outcrossing with other rotavirus strains makes it a difficult pathogen to control and surveil, even with current vaccines (3, 7, 13-16). Understanding rotaviral diversity expansion, genetic exchange between strains (specifically between the clinically significant type I and type II genogroups), and evolutionary dynamics resulting from coinfections has important implications for disease control (3, 7, 13-16). Rotavirus A genomes have high mutation rates (16-18), undergo frequent reassortment (15, 19-21), and the perception is that these two processes are the primary drivers of rotavirus evolution (16, 22). Genome rearrangements may also contribute to rotavirus diversity, but are not believed to be a major factor in rotavirus evolution (23). Homologous recombination, however, is thought to be especially rare in rotaviruses due to their dsRNA genomes (20, 21, 24). Unlike +ssRNA (25) viruses and DNA viruses (26), dsRNA viruses are not easily able to undergo intragenic recombination because their genomes are not transcribed in the cytoplasm by host polymerases, but rather within nucleocapsids by viral RNA-dependent RNA polymerase (20, 27). Genome encapsidation should significantly reduce the opportunities for template switching, the presumptive main mechanism of intramolecular recombination (26, 28).

Despite the expectation that recombination should be rare in rotavirus A, there are nevertheless numerous reports of recombination among rotaviruses in the literature (29-37). However, a comprehensive

survey of 797 rotavirus A genomes failed to find any instances of the same recombination event in multiple samples (38). There are two possible explanations for this outcome. Either the putative recombinants were spurious results stemming from poor analytical technique or were poorly fit recombinants that failed to increase in frequency in the population such that they would be resampled (38). The main implications of this study are that recombination among rotavirus A is rare and usually disadvantageous.

Since the number of publicly available rotavirus A whole segment genomes is now over 23,600, it is worth revisiting these conclusions to see if they are still valid. To this end, we used bioinformatics tools to identify possible instances of recombination among all available complete rotavirus A genome sequences available in the NCBI Virus Variation database as of May, 2019. We found strong evidence for recombination events among all rotavirus A segments with the exception of NSP3 and NSP5. In several cases, the recombinants were fixed in the population such that several hundred sampled strains showed remnants of this same event. These reports suggest that rotavirus recombination occurs more frequently than is generally appreciated, and can significantly influence rotavirus A evolution.

## Methods

### *Sequence Acquisition and Metadata Curation*

We downloaded all complete rotavirus genomes from the Virus Variation Resource as of May 2019 (39). Laboratory strains were removed from the dataset. Genomes that appeared to contain substantial insertions were excluded as well. Avian and mammalian strains were analyzed separately as recombination analyses between more divergent genomes can sometimes confound the results. No well-supported events were identified among the avian strains. For each rotavirus genome, separate fasta files were downloaded for each of the 11 segments. Metadata including host, country of isolation, collection date, and genotype (genotype cutoff values as defined by (40)) were recorded. The genotype cutoff values allowed for the differentiation of events that occurred between two different genotypes—intergenotypic—from those which occurred between the same genotype—intragenotypic.

### *Recombination Detection and Phylogenetic Analysis*

All 11 segments of each complete genome were separately aligned using MUSCLE v3.8.31 (41) after removing any low quality sequence (e.g., 'Ns'). Putative recombinants were identified using RDP4 Beta 4.97, a program which employs multiple recombination detection methods to minimize the possibility of false positives (42). All genomes were analyzed, with a p-value cut off $< 10^{-9}$, using 3Seq, Chimæra, SiScan, MaxChi, Bootscan, Geneconv, and RDP as implemented in RDP4 (42). We eliminated any strains that did not show a putative recombination event predicted by at least 6 of the above listed programs. We then ran separate phylogenic analyses on the 'major parent' and 'minor parent' sequences of putative recombinants in BEAST v1.10.4 using the general time reversible (GTR) + Γ + I substitution model (43). 'Parent' in this case does not refer to the actual progenitors of the recombinant strain, but rather those members of the populations whose genome sequences most closely resemble that of the recombinant. Significant phylogenetic incongruities with

high posterior probabilities between the ‘major parent’ and ‘minor parent’ sequences were interpreted as convincing evidence for recombination.

*Phylogenetic Analysis of VP1 and VP3*

Segment 1 (VP1) and segment 3 (VP3) each showed evidence for recombination events resulting in a novel lineage, wherein many isolates reflected the same recombination event. To analyze these events more thoroughly, for each segment we split alignments of a subset of environmental isolates to include flagged clades, minor parent clades, and major parent clades along with outgroup clades to improve accuracy of tip dating. We generated minor and major parent phylogenies using BEAST v1.10.4 (43). We used tip dating to calibrate molecular clocks and generate time-scaled phylogenies. The analyses were run under an uncorrelated relaxed clock model using a time-aware Gaussian Markov random field Bayesian skyride tree prior (44). The alignments were run using a GTR + Γ + I substitution model and partitioned by codon position. Log files in Tracer v1.7.1 (45) were analyzed to confirm sufficient effective sample size (ESS) values, and trees were annotated using a 20% burn in. The alignments were run for three chains with a 200,000,000 Markov chain Monte Carlo (MCMC) chain length, analyzed on Tracer v1.7.1 (45), and combined using LogCombiner v1.10.4 (46). Trees were annotated with a 10% burn, and run in TreeAnnotator v1.10.4 (43). The best tree was visualized using FigTree v1.4.4 (47) with the nodes labeled with posterior probabilities.

*RNA Secondary Structure Analysis*

To test the hypothesis that recombination junctions were associated with RNA secondary structure, we generated consensus secondary RNA structures of different segments and genotypes using RNAalifold in the ViennaRNA package, version 2.4.11 (48, 49). The consensus structures were visualized and mountain plots were generated to identify conserved structures that may correspond to breakpoint locations. We made separate consensus alignments both by using the same genotype and by combining genotypes. Due to the high sequence variability of the observed recombinants, particularly in segments 4 (VP4) and 9 (VP7), we note that the consensus structure may vary substantially depending on the sequences used in the alignments. Segments 7 (NSP3), 10 (NSP4), and 11 (NSP5) were not analyzed due to only having one or no recombination events. Consensus structures could not reliably be made for segment 2 (VP2). While the VP2 protein is relatively conserved across genotypes, it contains insertions particularly in the C1 genotype, yet shows recombination across C1 and C2 genotypes.

*Protein Structure and Antigenic Epitope Predictions*

To verify that events did not result in substantial (deleterious) protein structure changes, we employed LOMETS2 (Local Meta-Threading-Server) I-TASSER (Iterative Threading Assembly Refinement) (50-52) to predict secondary and tertiary protein structures. I-TASSER generates a confidence (C) score for estimating the quality of the protein models.

To determine if any of the putative recombinants possessed recombined regions containing epitopes, we analyzed the amino acid sequences of all VP4, VP7, and VP6 recombinants and their major and minor parents.

We used the Immune Epitope Database (IEDB) (53) and SVMTriP (54) to predict conserved epitopes (Supplementary Table 2).

## Results

*Strong Evidence for Homologous Recombination in Rotavirus A*

We identified 117 putative recombination events (identified by 6/7 RDP4 programs; Table 1, Supplementary Table 1). Of these, 79 recombination events were strongly supported, meaning they were detected by 7/7 recombination programs as implemented in RDP4 (Table 1). Most recombination events detected were observed in sequences uploaded to the Virus Variation database since the last large scale analysis (38) so differences between these results and prior studies may simply reflect the recent increase in available genome sequences.

| | **VP1** | **VP2** | **VP3** | **VP4** | **VP6** | **VP7** | **NSP1** | **NSP2** | **NSP3** | **NSP4** | **NSP5** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Number of Sequences Analyzed** | 1710 | 1600 | 1905 | 1990 | 2176 | 3887 | 1962 | 2186 | 1881 | 2430 | 1900 |
| **Putative Recombination Events[a]** | 17 | 13 | 17 | 16 | 12 | 24 | 11 | 3 | 0 | 3 | 1 |
| **Strongly Supported Events[b]** | 11 | 8 | 12 | 12 | 10 | 17 | 7 | 1 | 0 | 1 | 0 |
| **Recombination Frequency[c]** | 6.4E-3 | 5.0E-3 | 6.3E-3 | 8.0E-3 | 4.6E-3 | 4.4E-3 | 4.4E-3 | 4.6E-4 | 3.2E-3 | 4.1E-4 | 0 |
| **Intergenotypic Recombination ratio** | 2/17 | 6/13 | 4/17 | 7/16 | 7/12 | 22/24 | 3/9 | 1/3 | -- | 2/3 | 1/1 |

***Table 1.*** *Recombination events identified among all mammalian rotavirus A genome sequences downloaded from NCBI Virus Variation database in May 2019 (39). a6/7 programs implemented in RDP4 identified event (see Methods). b7/7 programs implemented in RDP4 identified event or a putative recombination event where more than one downloaded sequence showed the same event (see Methods). cStrongly supported events (Row 3) divided by number of sequences analyzed (Row 1).*

*Putative Recombination Events Observed in Multiple Environmental Isolates*

Eleven of the recombination events identified in Table 1 were observed in more than one environmental isolate (Fig. 1; Table 2). The observation of multiple sequenced strains with the same recombination event is strong evidence that the observed event was not spurious, and was not a consequence of improper analytical technique or experimental error. There are only two ways that multiple sequenced isolates will show the same recombinant genotype. Either multiple recombination events with the same exact breakpoints occurred at approximately the same time, or the event happened once and descendants of the recombined genotype were

subsequently isolated from additional infected hosts. The latter scenario is more parsimonious, and suggests that the new genotype was not reproductively impaired. Indeed, some recombined genotypes may be more fit than their predecessors, but this outcome would need to be experimentally validated with infectivity assays.

| Recombined Segment | Genotype(s) Involved | Number of Independent Isolates | Corrected and Uncorrected Multiple Comparisons | Breakpoints[a] | Accession Numbers |
|---|---|---|---|---|---|
| Segment 1 (VP1) | R1 | 2 | 4.787 E-18 and 1.349 E-10 | 35 – 1448 (1290-1498) | KC579647<br>KC580176 |
| Segment 1 (VP1) | R2 | flagged in 285 | 1.305E-16 and 3.332E-09 | (541-711) - (1327-1445) | KU199270<br>*see Fig. 2 |
| Segment 1 (VP1) | R2 | 2 | 1.756 E-20 and 4.951 E-13 | 656 (541-711)- 2661 (2600-2735) | KU356662<br>KU356640 |
| Segment 3 (VP3) | M2 | flagged in 551 | 1.78E-12 and 4.982 E-05 | 1258 (872-1389)- 1798 (1662-1917) | KX655453 |
| Segment 3 (VP3) | M1 | flagged in 719 | 2.491E-20 and 6.946E-13 | 2158 (2129-2174) – 2531 (undetermined) | KJ919553<br>KJ919517<br>KJ919551<br>KJ753665<br>JQ069727 |
| Segment 7 (NSP1) | A8 (porcine) | 3 | 3.01 E -37 and 8.622 E-30 | (1434-25)- (573-596) | KP753174<br>KJ753184<br>KP752951 |
| Segment 8 (NSP2) | N1 | 3 | 3.689 E-13 and 4.054 E-06 | 485 (455-499)- 889 (868-914) | KJ753657<br>KM026663<br>KM026664 |
| Segment 9 (VP7) | G6 (bovine) | 2 | 6.240 E-12 and 2.780 E-05 | 1048 (1032-128)-481 (448-506) | HM591496<br>KF170899 |
| Segment 9 (VP7) | G1 & G2 | 2 | 5.507 E-38 and 1.664 E-30 | 55 (994-60) – 291 (262-296) | KC443034<br>MG181727 |
| Segment 9 (VP7) | G1 & G3 | 2 | 1.4 E-23 and 4.230 E-16 | 857 (833-859)- 1019 (991-51) | KJ751729<br>KP752817 |
| Segment 9 (VP7) | G1 & G9 | 2 | 5.292 E-19 and 1.599 E -11 | 362 (346-366) – 589 (558-605) | AF281044<br>GQ433992 |

***Table 2**. Recombination events observed in multiple independent environmental isolates (i.e., isolates from different patients and/or sequenced by different laboratories). a99% Confidence Intervals.*

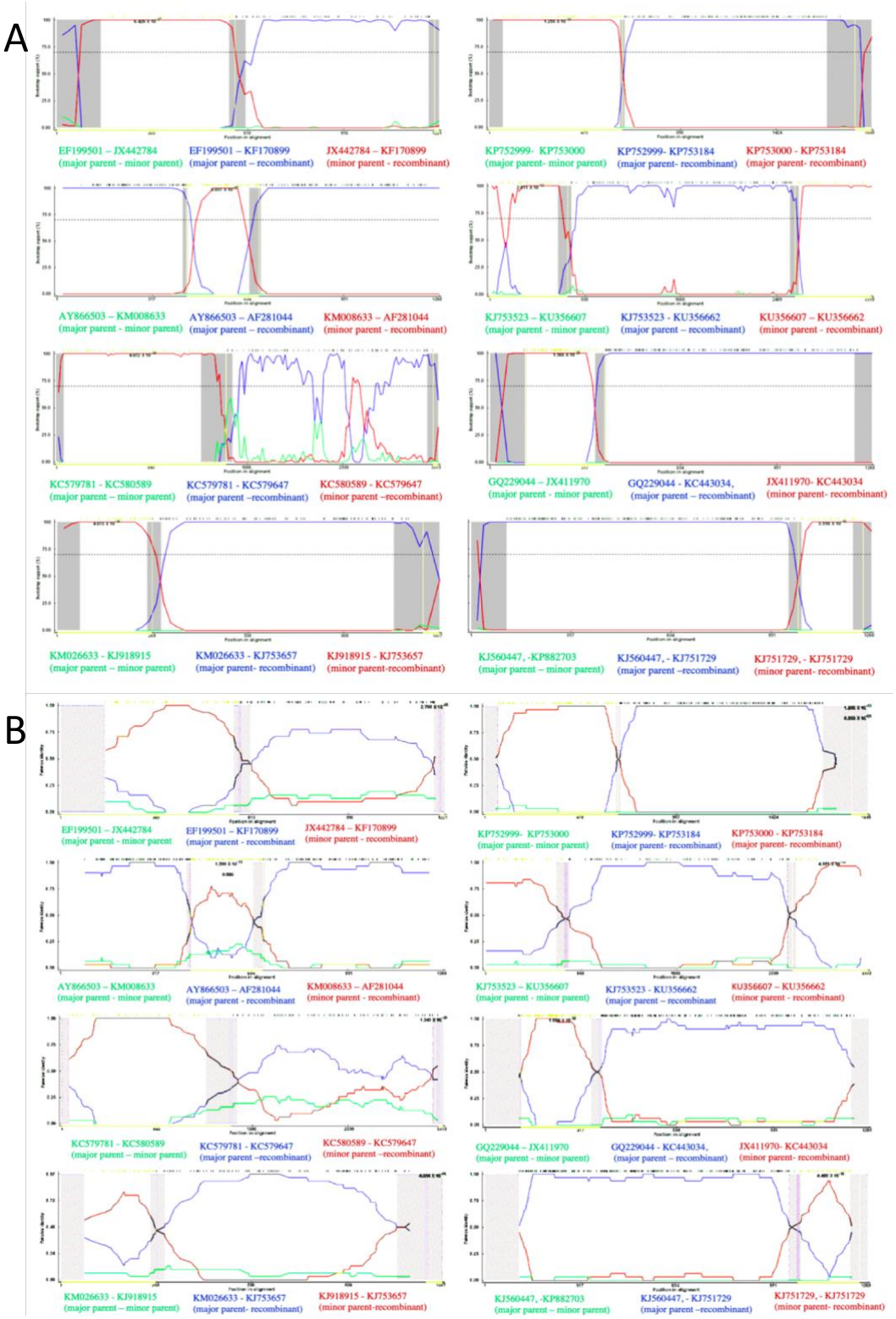



***Figure 1.*** *Putative recombinants where more than one sequence supported the event. a) Bootscan b) RDP. Top left showing a G6-G6 event in VP7, second from the top left showing a G9-G1 event, third from the top left showing a R1-R1 event in VP1, bottom left showing a N1-N1 event in NSP2. Top right is showing an A8-A8 event in NSP1, second from the top right is showing an R2-R2 event in VP1, third from top right is showing a G2-G1 event in VP7 and bottom right is showing a G3-G1 event.*

*Segment 1 Intragenotypic Recombination Resulting in a New Lineage*

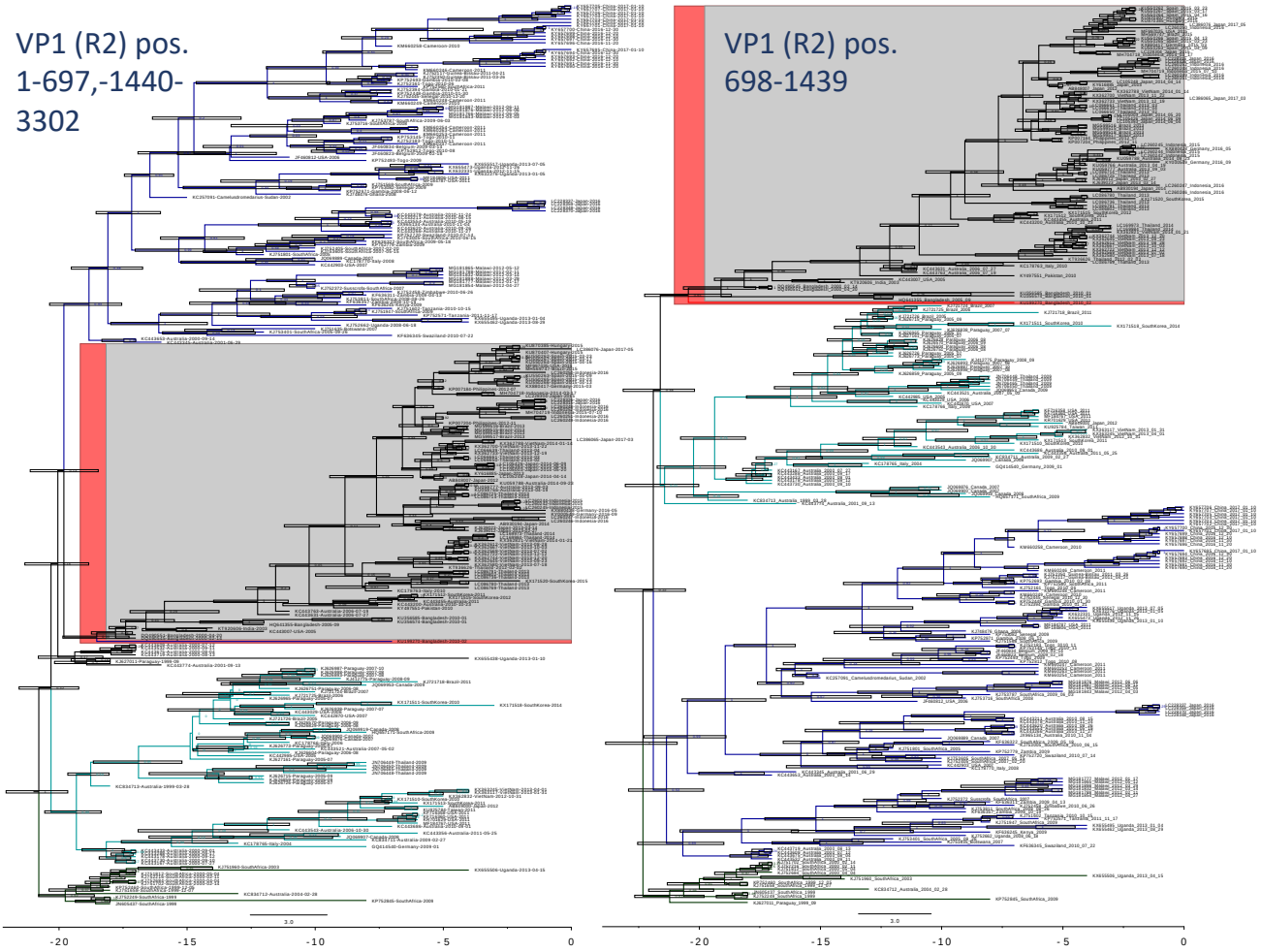



***Figure 2.*** *Bayesian phylogenetic analysis of a possible recombination event in VP1 that became fixed in the population. The strain closest to the recombination event is highlighted in red. Major clades are colored to show the phylogenetic incongruity. Nodes are labeled with posterior probabilities with 95% node height intervals shown. Time axis at the bottom is in years before 2017.*

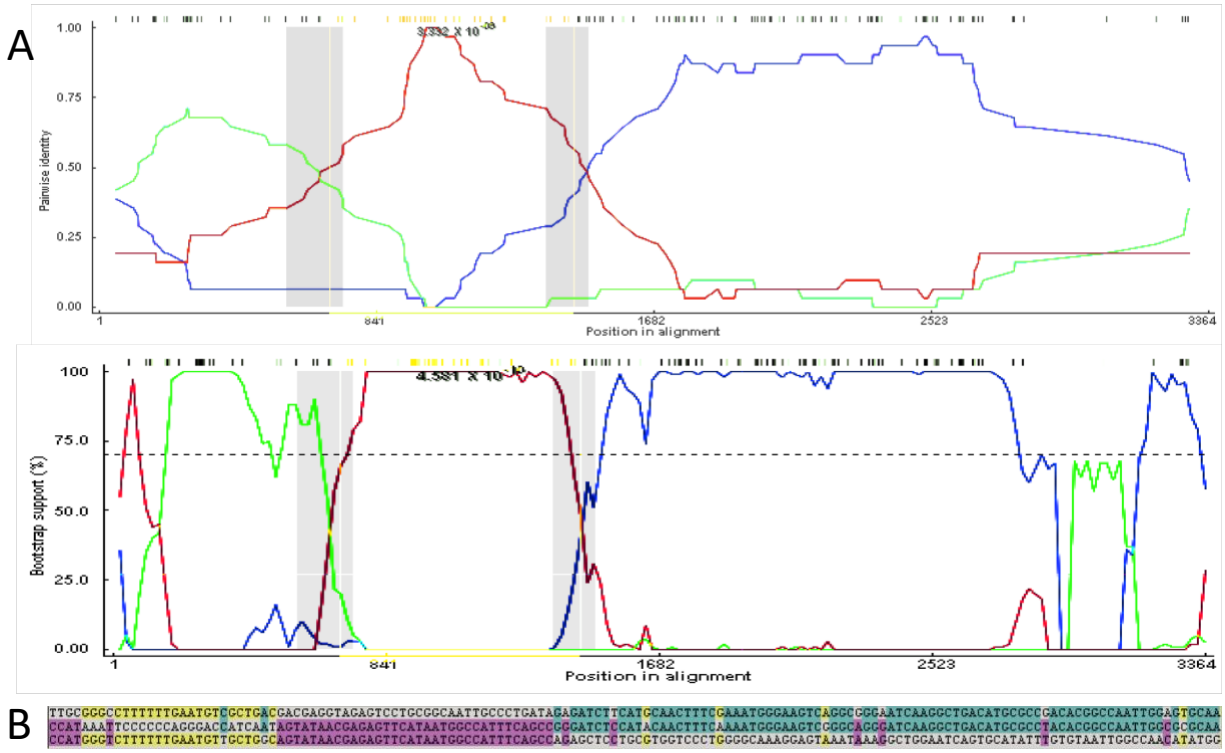



***Figure 3***. *A) RDP analysis (top) and BootScan analysis (bottom) shown for a putative recombinant between two R2 segment 1 genotypes. The red line is comparing the minor parent to the recombinant, blue line is comparing the major parent to the recombinant, and the green line is comparing the major parent to the minor parent. The Y axis for RDP (top) pairwise identity, while the Y axis for BootScan (bottom) is Bootstrap support. The X axis is the sequence along segment 1. B) The relevant sites shown below color-coded by which strain the recombinant matches. The Recombinant is the middle sequence, the minor parent is the lowest sequence, and the major parent is the top sequence. Mutations matching the major sequence are shown in blue, while mutations matching the minor parent are shown in purple. Yellow mutations are showing mutations not present in the recombinant sequence but which match between the major and minor parent, possibly suggesting a second recombination event.*

Segment 1 (VP1; ~3,302 base pairs) showed evidence of a recombination event within the R2 clade that was fixed in the population, and resulted in a new lineage (highlighted clade in Fig. 2; Fig. 3; Table 2). The multiple comparison (MC) uncorrected and corrected probabilities were 1.305E-16 and 3.332E-09 respectively. The sequence most closely related to the recombined sequence was KU199270, a human isolate from Bangladesh in 2010 (55). Phylogenetic analysis using tip calibration suggests that the recombination event occurred no later than 2000-2005 (node Cis), so if this is a true recombination event, the 2010 sequence is not the original recombinant. The recombinant region is 100% similar to an isolate also from Bangladesh in 2010 (KU248372) (55), which also has a putative recombinant sequence in another region of its genome. The breakpoint regions (99% CI: (541-711) - (1327-1445)) may represent a potential hotspot for segment 1 recombination as multiple recombination events show breakpoints in this area (Table 2). Phylogenetic analysis of the alignment of 250 representative sequences containing only the putative recombinant region compared with the rest of the segment 1 sequence showed a consistent subclade shift within the R2 clade. The

recombination events resulted in the incorporation of the following amino acid substitutions in the recombinant strains when compared to the major parent strain: 227 (K→E), 293 (D→N), 297 (K→R), 305 (N→K), and 350 (K→E).

*Segment 3 Intragenic Recombination*

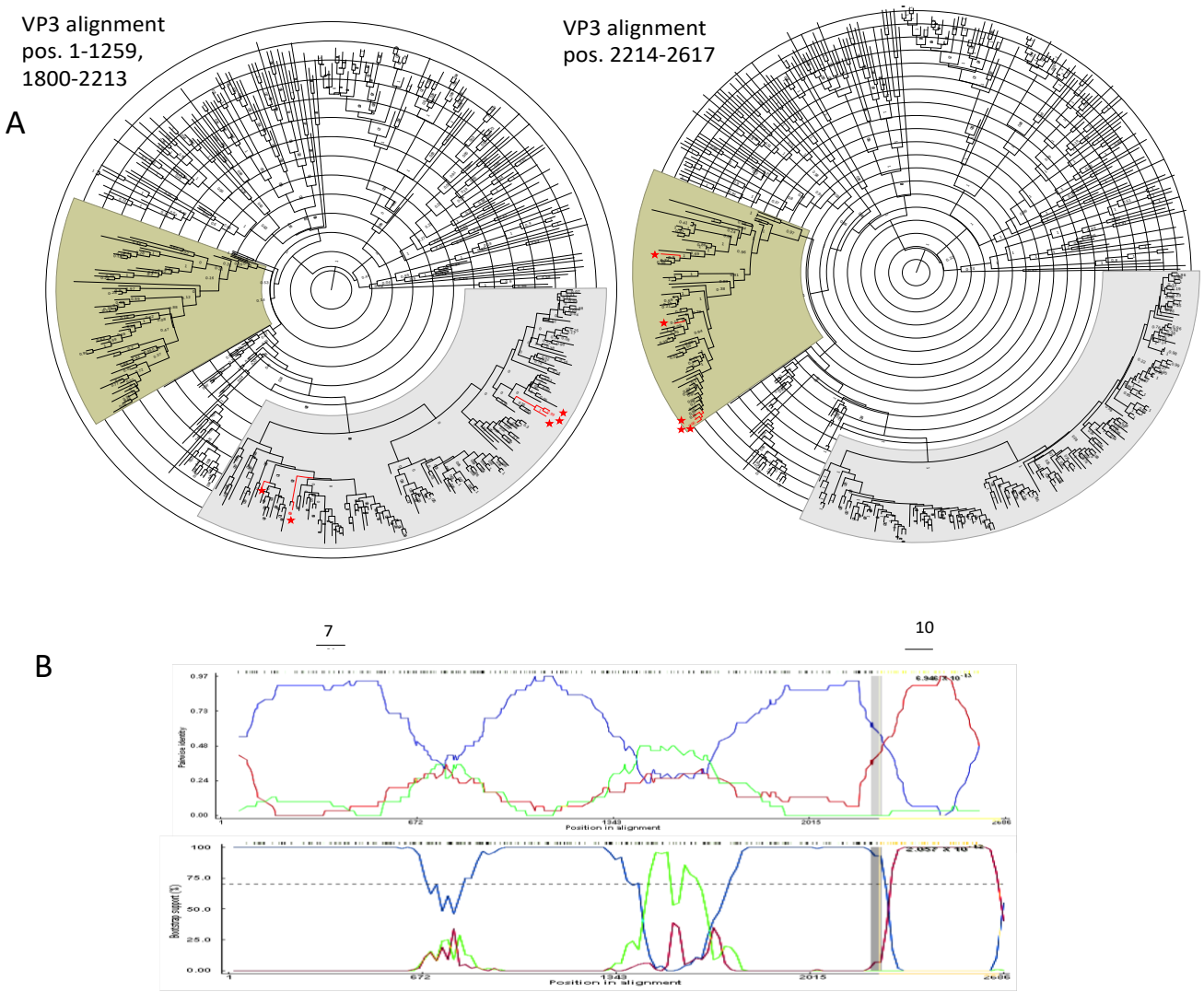



***Figure 1.*** *Segment 3 Recombination Event With. Multiple Isolates Supporting. a) Phylogenetic trees made using alignments from (left) nucleotide position 1-1259 and 1800-2213 representing the major parent region excluding second recombination event (Fig. 5), and (right) nucleotide position 2214-2617 representing the minor parent sequence. B) BootSccan (top) and RDP (bottom). The red line is comparing the minor parent to the recombinant, blue line is comparing the major parent to the recombinant, and the green line is comparing the major parent to the minor parent. The Y axis for RDP (bottom) pairwise identity, while the Y axis for BootScan (top) is Bootstrap support. The X axis is the sequence along segment 3.*

Two recombination events occurring in segment 3 (VP3) appear to have fixed in the population, and are seen in many descendant sequences (Table 2). The first strain identified as a putative recombinant is KJ753665, an isolate from South Africa in 2004. The recombinant region occurs between positions (2129-2174) - 2531.

This region is 98.7% identical to a porcine strain isolated in Uganda in 2016 (KY055418)(56), while the rest of segment 3 is 95.8% similar to a 2009 human isolate from Ethiopia (KJ752028). MC uncorrected and corrected probabilities were 2.491E-20 and 6.946E-13, respectively, with 107 isolates flagged as possibly derived from the recombination event. Phylogenetic analysis using 450 randomly selected VP3 sequences (excluding sequences lacking collection dates) within the putative recombinant region, along with an analysis of the genome excluding the two major recombination events, resulted in five sequences showing a significant phylogenetic incongruity (Fig. 4). The incongruity appeared between sub-lineages within the larger M1 lineage of VP3. Amino acid substitutions in the recombinant region included positions 748 (M→T) and 780 (T→M).

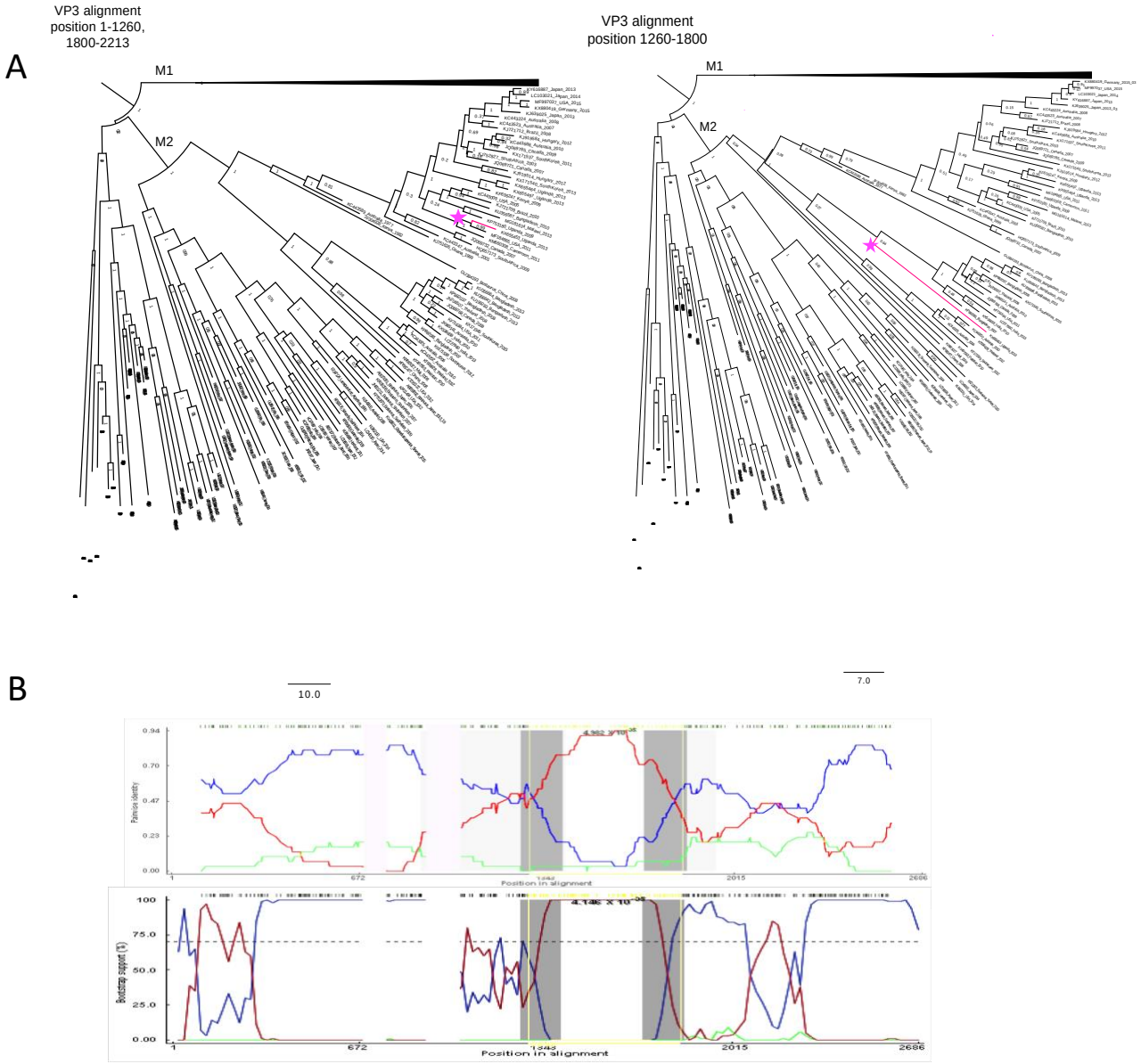



***Figure 5.*** *A) Phylogenetic analysis of segment 3. Left tree (major parent) was made from an anlignment of nucelotide positions 1-1260 and 1900-2213 and right tree (minor parent) was made from nucleotide positions 1260-1800. M2 strains have been collapsed, and recombinant is colored in pink. B) BootSccan (top) and RDP*

*(bottom of VP3 M2 putative recombinant.. The red line is comparing the minor parent to the recombinant, blue line is comparing the major parent to the recombinant, and the green line is comparing the major parent to the minor parent. The Y axis for RDP (bottom) pairwise identity, while the Y axis for BootScan (top) is Bootstrap support. The X axis is the sequence along segment 3.*

A second potential recombination event was identified in segment 3 between sublineages within the M2 lineage (Table 2; Fig. 5). However, when we created split alignments and ran a phylogenetic analysis in BEAST, only one sequence showed a phylogenetic incongruity supporting this event (KX655453). Amino acid substitutions as a result of this event included positions 405 (I→V), 412 (V→M), 414 (N→D), 441 (N→D), 458 (I→V), 459 (I→T), 468 (L→F), 473 (N→D), 486 (M→I), 518 (N→S), and 519 (E→G).

One possible complication with relying on phylogenetic incongruity as evidence of major recombination events is that the recombinant virus may evolve faster than the parental strains, and, even if possessing an initial advantage such as immune avoidance, may eventually converge back toward the major parent's sequence. Thus, phylogenetic analysis likely underestimates the frequency of transiently stable recombination events in the population. This may account for the fact that the number of VP3 isolates flagged as deriving from a recombination event based on sequence is greater than the prevalence of recombination predicted by the phylogenetic analysis.

*Intergenotypic Recombination*

We found evidence for intergenotype recombination in all segments except segment 7 (NSP3), segment 10 (NSP4), and segment 11 (NSP5) (Table 1; Supplementary Table 1). Instances of intergenotypic recombination in segment 1 (VP1) (KU714444, JQ988899) only occurred in regions where the amino acid sequence was highly conserved across genotypes, so these events resulted in few, if any, nonsynonymous mutations**.** Of the putative events observed in more than two environmental isolates with strong support from detection methods, only events in segment 9 (VP7) occurred between different genotypes (Table 2). The segment 9 recombinant region amino acid substitutions that match the minor parent and differ from the major parent are shown in Supplementary Table 2.

*Structural Prediction of Recombinant Proteins*

The protein models generated by I-TASSER for the intergenotypic recombinant G proteins showed that, although the amino acid changes for the G1-G2 recombinant were substantial (25 changes) (Supplementary Table 2), the secondary and tertiary structures seemed largely intact (Fig. 6). The G3-G1 and G6-G6 recombinants each showed four amino acid changes. The G9-G1 recombinant showed the most secondary structure disruption, including a loss of beta sheets, a loss of antiparallel beta sheets, and slightly shorter beta sheets/helices, which could indicate lower stability. However, based on the structural modeling, the tertiary structure appeared to be maintained, suggesting that the putative recombinant glycoproteins were able to form properly folded G-proteins (Fig. 6).

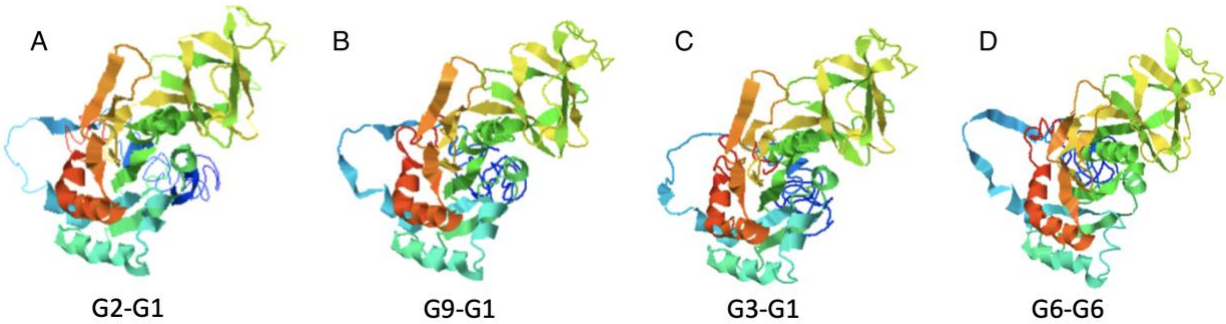



***Figure 6.*** *VP7 protein models predicted from I-TASSER. Proteins were predicted from amino acid alignments of the four strongly supported G recombinants a) G2-G1, C=-1.38 KC443034 b) G9-G1:AF281044, C=-1.29 c) G3-G1, C=-1.27 KJ751729 d) G6-G6: KF170899, C=-1.42.*

Antigenic epitope predictions generated by IEDB (53) and SVMTriP (54), as well as a large study done on mammalian G-types (57), showed that VP7 recombination occasionally results in amino acid substitutions in conserved epitopes (Supplementary Table 2). For example, the amino acid sequence RVNWKKWWQV is usually flagged as, or part of, an epitope in most G types including G2, G3, G4, G6, G9, but not in G1. In the G2-G1 recombinant flagged in multiple sequences (Table 2; Figure 1), this region is altered (sometimes a KR substitution and sometimes multiple amino acid substitutions) (Supplementary Table 2). Moreover, the region where the recombination occurred, although containing highly variable regions, had low solvent accessibility. Thus, although the amino acid sequences may diverge, the protein folding and three-dimensional structure remains relatively conserved (Fig. 6). The G3-G1 recombinant had amino acid substitutions in two of four conserved epitope regions due to the recombination event. There was also a conserved epitope sequence around amino acids 297-316 in G9 proteins that was altered in the G9-G1 recombinant so it no longer appeared as an epitope. In other words, changes in the amino acid sequence that do not result in significant changes to the protein configuration may nevertheless reduce binding by antibodies or T-cell receptors, and may provide a selective advantage allowing the virus to avoid immune surveillance.

*Recombination junctions often correspond to RNA secondary structural elements*

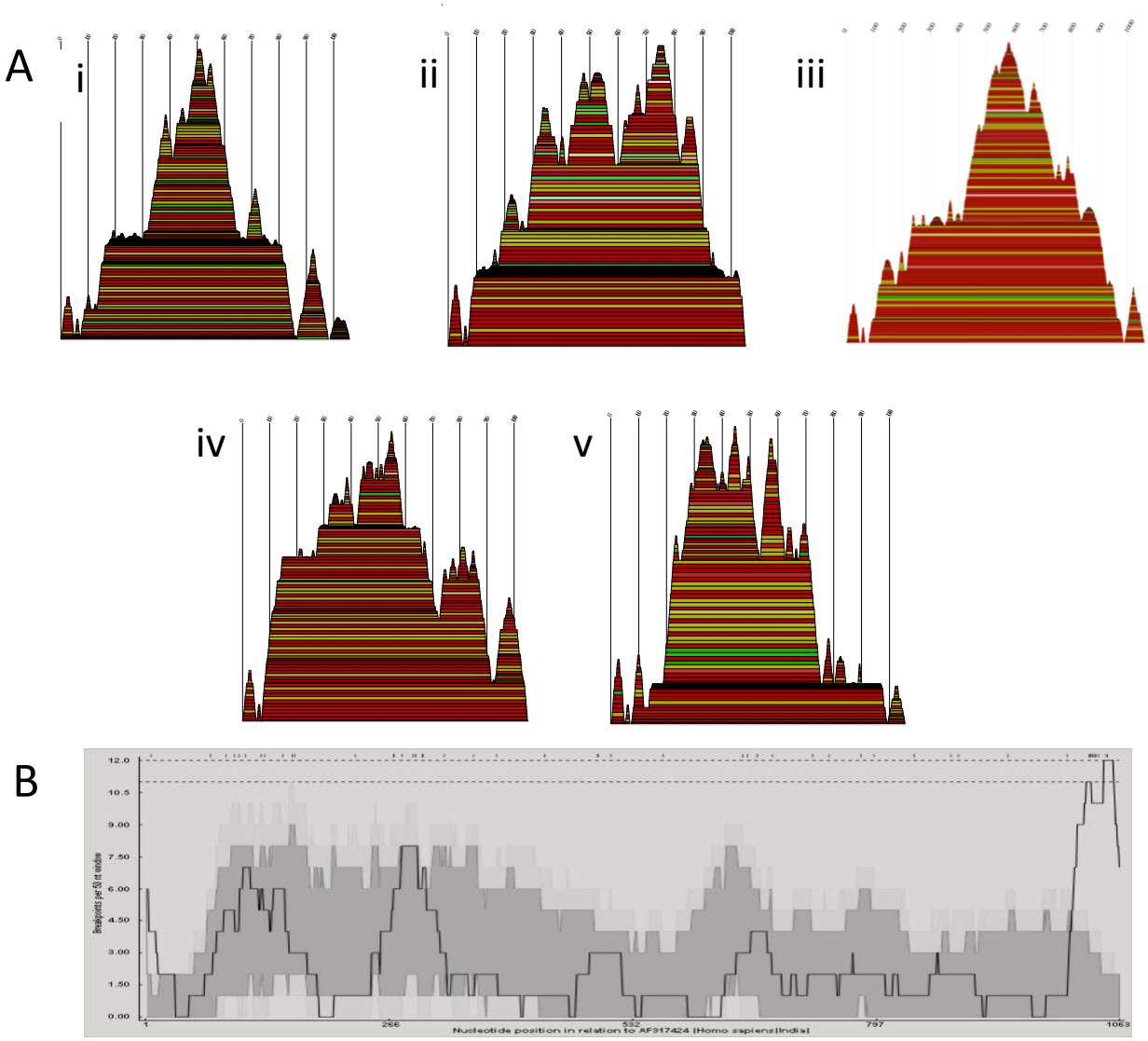



***Figure 7.*** *Consensus mountain plots made in RNAalifold using the ViennaRNA package, of the predicted RNA secondary structures based on alignments made using full sequences for each of the five G types involved in the intergenotypic recombination events: (i) G9 (ii) G6 (iii) G3 (iv) G1 (v) G2. Peaks represent hairpin loops, slopes correspond to helices, and plateaus correspond to loops. X axis corresponds to the sequence of the segment. Each base-pairing is represented by a horizontal box where the height of the box corresponds with the thermodynamic likeliness of the pairing. The colors correspond to the variation of base pairings at that position. Red indicates the base pairs are highly conserved across all the sequences, black indicates the least conservation of those base pairings, B) Breakpoint distribution plot made in RDP4, of putative recombinants in segment 9. X axis is showing the position in the sequence, and Y axis is showing the number of breakpoints per a 50-nucleotide window. Highest peaks are around x=115, 285, 1050.*

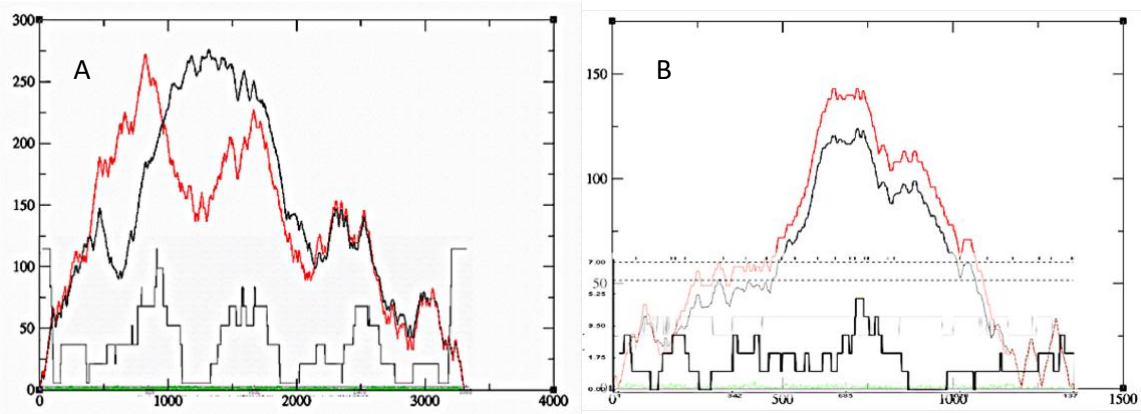



***Figure 8.*** *Consensus Mountain Plots overlaid with breakpoint distribution plots of recombination events for a) VP6, and b) VP1. Position of nucleotide sequence is on the X axis. Entropy curve represented in green. The black curve represents the pairing probabilities, and the red curve represents the MFE structure with well-defined regions having low entropy.*

Breakpoint distribution plots showed the sequence regions with the most breakpoints. These breakpoints often corresponded to hairpins predicted by RNAalifold. Secondary RNA structure predictions for segment 9 (VP7) genotypes G1, G2, G6 and G9 are shown as mountain plots (Fig. 7). The breakpoints of the segment 9 recombination events correspond to areas leading to the peaks in the mountain plots (Fig. 7). The peaks indicate a conserved hairpin loop, with the sequences leading up to the peak being the double-stranded portion of the hairpin. Breakpoint distributions also appeared to correspond to secondary structure predictions made from alignments of segment 1 (VP1) and segment (VP6) (Fig. 8). Segment 4 (VP4) RNA secondary structure predictions (Supplementary Fig. 1) showed greater variation across genotypes, so while breakpoints did coincide with secondary RNA structures as predicted by the models, there were not enough events in each genotype to provide strong support that the breakpoints were correlated with secondary structure. The breakpoint distributions for NSP1 and VP3 were also non-randomly distributed across the sequence (Supplementary Fig. 1), however secondary structure predictions from these alignments were not consistent, so they are not shown (Supplementary Fig. 2).

## Discussion

### *Recognizing Type I and Type II Error*

Apparent instances of recombination may actually be the result of convergent evolution, lineage-specific rate variation, sequencing error, poor sequence alignments, laboratory contamination or improper bioinformatics analysis (58-61). For example, the rates of type I (falsely identifying a recombination event when none exists) and type II (failure to detect a true recombination event) error incurred during the use of the recombination detection software, RDP4, were measured in a +ssRNA flavivirus, which has low rates of recombination (62, 63). The results of the analyses indicated that recombination was overestimated in these viruses, and certain detection methods were more prone to type I error. 3Seq and GENECONV display the lowest false positive rates as well as the lowest detection power of true events, while MaxChi, Chimaera, and

SiScan show higher false positive rates, but also had higher detection power for true positives. False positives using RDP4 are especially common among closely related strains. That being said, recombination events between sequences with low substitution rates and divergence percentages would be undetectable, yet most likely occur given their close spatial/temporal proximity. Furthermore, in RDP4, the *P*-value of .05 does not correspond to a 5% rate of false positives, which is why the cut-off value was set to 10E-04 and focused only on events where at least six RDP4 programs detected the putative recombinant. The length of the recombination region, if too short, also contributes to both type I and type II error (62-64). We ignored any putative recombination events of < 100nt, with the exception of one isolate in segment 4 (NSP4) and one isolate in segment 5 (NSP5), which we noted in Supplementary Table 1, as the putative recombinant regions were both at conserved regions at the ends of the segments.

Several steps can be taken to minimize incorrect attribution of viral recombination. Ideally this process should begin at the time of sequencing. If a putative recombination event is observed, then it should be confirmed that the originating sample did not come from a host infected with multiple genotypes by purifying, amplifying, and sequencing multiple plaques (58). Similarly, care must be taken when sequencing multiple samples of the same virus to minimize the possibility of cross-contamination.

For sequences obtained from online repositories, such precautions are rarely possible. Instead careful bioinformatics procedures can help minimize possible errors. Sequence alignments should be visually inspected to exclude misaligned sequences (58). Splitting alignments by major and minor event, followed by carefully parameterized BEAST runs, may help distinguish genuine phylogenetic incongruity signals from spurious false positives. Moreover, false positive recombinants are likely to be present as single isolates in phylogenetic trees (58).

The more isolates showing the same event, the greater the probability that it represents a true recombination event, especially if isolates were acquired and sequenced by different laboratories. Events that showed strong support, but were only isolated in one sequence are noted (Supplementary Table 1), but not discussed, as it is difficult to rule out the possibility of type I error due to PCR or mosaic contig assembly (24, 58). For example, pseudo-recombinants may be generated from template switching during PCR amplification, and can cause significant confusion when performing any kind of phylogenetic analysis or divergence dating.

Similarly, recombination facility likely depends on genetic similarity, so caution should be used in inferring events between highly dissimilar genotypes. When a positive recombination signal has been detected, it is essential to assess its statistical significance. Programs such as 3SEQ (65) and Simplot (66) can be used to identify recombination breakpoints and assign p-values assessing the likelihood that the recombination signal was non-random (58). In addition, the presence of unique polymorphisms differing from the parent strains within the suspected recombination region may provide evidence that recombination events are not laboratory artifacts.

Sequence metadata can also identify unlikely recombination events. For an event to be plausible, the major and minor parents should have had opportunity to coinfect the same host, which is only possible if they

are congruent in time and space (58). For example, one study identified influenza A virus strain A/Taiwan/4845/99 as a recombinant of A/Wellington/24/2000 and A/WSN/33 (67). Given that the two parents were isolated 77 years apart in different parts of the world, it is exceedingly unlikely that is a natural recombination event. Any putative recombination events should be carefully screed to determine if the parental strains could have plausibly interacted.

*Naturally High Coinfection in Rotavirus*

Some features of rotavirus biology make it plausible that recombination is not only possible, but also relatively common. Rotaviruses are often released from cells as aggregates of approximately 5-15 particles contained within extracellular vesicles (68). Collective infection of host cells by rotavirus-containing extracellular vesicles may increase the likelihood of recombination between genetically different viruses in the wild, an effect that may be missed in laboratory coinfection assays. Moreover, although the transcription of dsRNA in the viral core reduces the opportunity for recombination, it does not prevent it entirely or exclude adaptive recombination events. The physical barriers to recombination in dsRNA viruses may be offset by the high rates of coinfection resulting from vesicle transmission of rotaviruses.

Furthermore, infection of hosts by multiple rotavirus strains appears to be relatively common. In a study of 100 children in the Detroit area, G and P typing, which identifies the serotype of the VP7 and VP4 proteins respectively, revealed that ~10% of patients were infected with multiple rotavirus A strains (69). Similarly high frequencies of G and P mixed genotype infections were observed in children sampled in India (three studies showing multiple G types in 11.3 %, 12% and 21% of samples) (70-72), Spain (>11.4% of samples) (73), Kenya (5.9%) (74), Africa (12%) (75), and Mexico (5.6% in 2010, 33.5% in 2012) (76). Even higher frequencies of mixed genotype infections were observed in whole genome studies. For example, among 39 Peruvian fecal samples genotyped using multiplexed PCR, 33 (84.6%) showed evidence of multiple rotavirus genotypes (77). In another study, whole genome deep sequencing revealed that 15/61 (25%) samples obtained in Kenya contained multiple rotavirus genotypes (78). Given the high genetic diversity of rotavirus populations (7, 15, 16), and their proficiency in infecting a broad range of mammalian hosts, including many domesticated animal species (22, 79), the high frequencies of hosts infected with multiple genotypes is not entirely surprising. These coinfections present abundant opportunities for rotavirus recombination.

*Recombination Between Divergent Genotypes*

Rotavirus genotype classification varies by segment, with NSP5 having the highest % nucleotide identity cutoff (91% nucleotide identity) and NSP1 having the lowest % nucleotide identity cutoff (~79% nucleotide identity)(3). The serotype proteins, VP4 and VP7, have the most different genotypes, with 51 and 36 respectively (12). In particular, VP4 has especially variable secondary RNA structure (80). Since VP4 forms the host receptor-binding spike protein and VP7 is a glycoprotein that makes up the virus outer surface, it is likely that antibody neutralization drives the evolution of VP4 and VP7 diversity (6). Given this genetic diversity, the chances of two viruses with different G or P types coinfecting a cell is substantially higher than other segments. In addition, both VP4 and VP7 seem to be more prone to reassortment, and to tolerate more divergent genetic

backgrounds or genome constellations (5, 6, 81, 82). This diversity and tolerance of many different genetic backgrounds suggests that VP4 and VP7 may be more tolerant of recombination between divergent strains than the other segments. Consistent with this expectation, we found substantial evidence of intergenotypic recombination in segment 4 (VP4) and in segment 9 (VP7) (Table 2; SI Table 1). A previous study reported intergenotypic recombination events in segment 6 VP6, segment 8 (NSP2), and segment 10 (NSP4) (35), so our results add to the number of segments able to tolerate intergenotypic recombination.

Specifically, we observed several instances of intergenotypic recombination among G6 and G8 genotypes of segment 9 (SI Table 1). In addition, we identified intergenotypic recombination between P4-P6, P4-P8, P6-P8, and P8-P14 genotypes of segment 4 (SI Table 1). P4, P6 and P8 are all relatively closely related being in the P[II] genogroup, while P14 is in the P[III] genogroup. These putative recombination events suggest that relevant serotype diversity in human hosts is expanding. However, as P4, P6 and P8 are also the dominant P types in human infections, there is a sampling bias towards this genogroup as the genotypes within this group are more likely to coinfect humans, so caution should be taken with this conclusion.

*Generation of Escape Mutants*

Recombination involving regions encoding conserved epitopes, especially in segments that encode proteins involved in host cell attachment and entry, may provide selective advantages to rotaviruses by allowing them to evade inactivation by host-produced antibodies. These escape mutants may be generally less fit than wildtype viruses, but competitively advantaged in hosts because of a lack of host recognition. Subsequent intrahost adaptation may then select for compensatory fitness-increasing mutations allowing these strains to be competitive with circulating rotavirus strains. Many of the recombination events observed in our study appeared to generate such escape mutants. For example, we found two instances of the same segment 9 (VP7) G1-G3 recombination event (Table 2; KJ751729 and KP752817) with amino acid changes in two of four conserved epitope regions due to the recombination event, which suggests that this strain prevailed in the population because it was better able to evade antibody neutralization.

Most of the VP4 recombination events involved the spike head of the VP8* protein or the spike body/stalk region of VP5* (antigen domain). Escape mutant studies (4, 83-85) for VP4 show the VP8* spike head recognizes histo-blood group antigens, which is one of rotavirus's main host range expansion barriers (86-88). VP4 recombinants therefore may help the virus expand host range and potentially aid in immune evasion.

While VP4 is immunogenic, antibodies must bind to many VP8* protein fragments to inhibit host binding and viral entry. Because of this, VP7 specific monoclonal antibodies are more efficient at neutralizing rotavirus infection (89). Not only is there more VP7 protein on the virus surface, it is neutralized by inhibiting decapsidation (84). Our study found that a G3-G1 recombinant (KJ751729) had amino acid changes in two out of four total conserved epitope regions due to the recombination event, which could indicate this strain was able to prevail in the population by being better able to evade neutralization.

VP6 also showed recombination events resulting in substantial amino acid changes. VP6 is a more conserved protein, but also an antigenic protein that interacts with naïve B cells (90). This feature suggests that

there may be selection for VP6 escape mutants to evade host immune responses. A secondary structure analysis of VP6 indicated that it was relatively conserved across genotypes, which may explain why VP6 seems to have more frequent intergenotypic recombination. Conserved epitopes in VP6 exist around amino acid positions 197 to 214, and 308 to 316 (91).

NSP1 also showed recombination between highly divergent strains. It is expected that NSP1 could tolerate more mutations or even recombination, as NSP1 is not strictly required for viral replication (92). NSP1 is also the least conserved of the proteins, including the serotype proteins VP4 and VP7 (93), indicating intergenotypic recombination causing more disruption of the amino acid sequence may be less likely to be deleterious.

Recombination has the potential to decouple RNA-RNA interactions from protein-protein interactions. For reassortment to occur, segments packaging requires not only compatible RNA-RNA and RNA-protein interactions, but also the newly generated reassortant must be able to outcompete its progenitor strains. This feature implies that a reassortant virus with a segment originating from a more genetically divergent virus may be infectious, but less likely to be generated during a coinfection due to competing RNA interactions. Intragenic recombination could allow for novel reassortant proteins if breakpoints are within open reading frames and do not disrupt packaging signals. This could allow a virus to replicate and infect efficiently in a new genetic background.

Recombination around conserved epitope regions, especially in the VP4 and VP7 proteins involved in host cell attachment and entry, may provide a selective advantage to rotaviruses by allowing them to evade inactivation by host-produced antibodies. These escape mutants may be generally less fit than wildtype viruses, but competitively advantaged in the host because of a lack of host recognition. Subsequent intrahost adaptation may then select for compensatory fitness-increasing mutations allowing these strains to be competitive with circulating rotavirus strains.

Based on the I-TASSER structural predictions, the putative recombinant VP7 proteins detected in our survey appear able to fold properly and form functional proteins despite containing amino acid sequence from different ‘parental’ genotypes (Fig. 6). While the secondary structure appeared slightly altered (e.g., shorter beta sheets), the recombinant VP7 proteins generally maintained their 3-dimensional shape. The selective advantage from swapping epitopes may outweigh any potential decrease in protein stability resulting from recombination.

Segment 7 (NSP2), 8 (NSP3), 10 (NSP4), and 11 (NSP5) all had low rates of recombination. These segments are short thus recombination is expected to be less likely, however segment 9 (VP7), one of the smallest segments, defies this pattern, having a high number of events observed. The smaller segments may also be less able to tolerate recombination events due to the important roles they play during the formation and stabilization of the supramolecular RNA complex (94) during rotavirus packaging and assembly (80, 95-97).

*Drivers of Recombination*

Homologous recombination has not previously been considered a significant driver in rotavirus evolution (20, 38). While recombination is common in positive-sense RNA viruses (98), and even required for

retroviral genome replication and maintaining genome integrity (99), it is rarely reported in dsRNA viruses. This discrepancy may stem from divergent selective pressures among different types of RNA viruses. While recombination may be an important mechanism for purging deleterious mutations and lowering clonal interference in some non-segmented viruses (100), segmented virus can accomplish the same outcome through reassortment without breaking up any genes (21). Segmented viruses may be producing recombinant viruses that are less fit than reassortant viruses, and therefore do not dominate populations. While some segmented positive-sense RNA viruses can also reassort, such as the *Bromoviridae*, they have fewer genome segments, increasing clonal interference. Moreover, double-stranded RNA is a more stable molecule than single-stranded RNA so dsRNA viruses are more resistant to damage (101). Recombination as a mechanism for genetic repair, therefore, may be more important in segmented +ssRNA viruses than segmented dsRNA viruses.

Recombination is usually expected to be deleterious as the breakage of open reading frames disrupts RNA secondary structure and may alter protein functionality (28, 100). However, recombination, as with reassortment (15, 19, 102-104), may further increase rotavirus A genetic diversity due to epistatic interactions resulting in reassortant-specific or recombinant-specific mutations (105). Formerly deleterious mutations may become beneficial when the genetic background changes, resulting in an increase in circulating pathogenically relevant viral strains.

In our study, recombination events usually occurred between strains of the same genotype. Intertypic recombination may be more likely to be selected against as it may disrupt proteins or secondary RNA structure. Divergent genotypes may also have a lower likelihood of coinfect the same cell due to regional dispersal. However, the isolation of several strains resulting from recombination between different genotypes of segment 9, the VP7 coding segment, indicates recombination may be a mechanism for increasing the diversity of circulating VP7 genotypes (Table 2).

We also note that undetectable recombination would be expected to be more frequent as it is less likely to be deleterious, however this study shows recombination between strains different enough to result in significant phylogenetic incongruities occurs in rotaviruses. These events seem to be beneficial as evidenced by the fact that the recombinant genotypes persist in populations long enough for multiple samples showing the same event to be sampled. For example, the same mosaic VP7 G6 gene (KF170899) (Figure 8) was sequenced in multiple bovine strains, one isolated in 2009, and another in 2012, by two separate research groups. While still detectable, the parents are closely related. However, in some instances, events between highly divergent genotypes seem to have been able to persist in populations. Both a 2014 Malawi isolate and a 2006 isolate from the United States showed similar G1-G2 VP7 mosaic gene (accession number KU356662) (Figure 8). Given that these two G genotypes are highly divergent from one another, were identified in different years, in different locations, and by different research groups, supports the contention that it is a true recombination event. This example also indicates mosaic genes formed from two divergent genotypes can persist in the population, affecting the diversity of prevailing strains.

The results of this study also suggest the possibility of a recombinant sub-lineage within the R2 clade of segment 1, the polymerase-encoding segment. This result would indicate that homologous recombination has had a long-term effect on rotavirus diversity as the same event is represented in strains isolated years apart from geographically distant locations. Strains isolated in different countries showing the same recombination pattern indicates at least some recombination events were sufficiently favorable to be maintained in the population.

Recombination has had a significant impact on the diversity of other *Reoviridae* members such as bluetongue virus (106). For example, He et al., found that homologous recombination in bluetongue virus played a role in shaping the virus's evolution and epidemiology (106). Multiple bluetongue virus sequences isolated in different countries at different times showed evidence of the same recombination event, indicating that a mosaic gene successfully spread within the population.

*Recombination in Other dsRNA Viruses*

While recombination may result lead to fitness benefits, rates of recombination across viruses are more likely determined by RNA polymerase processivity (average number of nucleotides added by the polymerase per association with the template strand) (107). This supposition is evidenced by the fact that -ssRNA viruses are the least likely to undergo homologous recombination due to their genome architecture and replication mode. Since the expectation is that dsRNA viruses seldom recombine, researchers are rarely motivated to investigate recombination in dsRNA viruses, further exacerbating the lack of reported recombination events in these viruses.

Template switching, the presumptive main mechanism of RNA virus recombination, requires two different genomes to be in close proximity (26, 100, 108). Because of this requirement, recombination is expected to be rare in the dsRNA viruses because their genomes are transcribed within nucleocapsids, which limits opportunities for the genomes of different virus genotypes to interact (109, 110). Despite these limitations, recombination has been reported in several dsRNA virus species. For example, a study of 692 complete bluetongue virus segments found evidence for at least 11 unique recombinant genotypes (1.6%) (106). The case for recombination among bluetongue viruses is strengthened by the fact that viruses containing the same (or similar) recombinant segments were isolated by different research groups in different countries at different times, indicating that the recombinant viruses persisted and spread following the recombination event (106, 111). In another study, 1,881 sequences from members of the family *Birnaviridae* were analyzed for evidence of recombination (112). While no interspecies recombination was observed, at least eight putative instances of intraspecies recombination were observed among the infectious bursal disease viruses and the aquabirnaviruses (112). We note that Birnaviruses' genetic material is in a complex with ribonucleoprotein, while the genetic material of *Reoviridae* members is free in the virion, which could be a factor in differing rates of recombination across these dsRNA viruses. Li et al. reported several instances of recombination among the dsRNA Southern rice black-streaked dwarf virus which is also a member of the *Reoviridae* family, but infects plants (113). Putative recombinants were identified in six of the ten segments in both structural and nonstructural proteins. No instances of intrasegment/homologous) recombination were observed in a survey of the dsRNA cystoviruses

(114), but the medium segment of phage Φ13 is believed to have recombined with the medium segment of Φ6 (110).

*Possible Mechanism of Recombination in dsRNA Viruses*

Genetic recombination is a significant driver of evolution in many viruses, including human immunodeficiency virus (115), Western equine encephalitis virus (116), and poliovirus (25), however, recombination rates vary dramatically across different RNA viruses due to differences in genome replication and transcription, and due to differing propensities for coinfection. Negative-sense RNA viruses have lower rates of recombination because, inside host cells, their RNAs are encased by ribonucelocapsids (117) thus less available for recombination to occur. DsRNA viruses also undergo RNA synthesis after being packaged, and most do not have free double-stranded RNA in the cell. Prior to packaging, rotavirus ssRNA can form temporary associations even at low homology, and become tangled, which could result in an extra RNA being occasionally mispackaged. The packaging of more than 11 segments may provide opportunities for recombination among between the surplus segment and its homolog.

The precise mechanism for how rotaviruses package RNA is unknown, but there are leading hypotheses, which share similarities to the well-established systems Φ6 and influenza. The most-accepted hypothesis is that the eleven +ssRNA segments are bound to a protein complex consisting of the VP1 polymerase and the VP3 capping enzyme. RNA structures in the non-translated terminal regions (NTRs) aid in the formation of the supramolecular RNA complex (94, 97) and determine whether the segments are packaged (80, 96). Subsequently the supramolecular RNA complex is encapsidated by the core shell protein, VP2. Once the particle is assembled, RNA synthesis occurs resulting in the dsRNA genome. The packaging mechanism of bacteriophage Φ6 and other Cystoviruses is well understood, and distinct from the *Reoviridae* family. In the Cystoviruses, each RNA segment is packaged separately in a precise order (118). Because of this, the packaging of an extra segment would be more difficult, which may explain the lack of homologous recombination reported in Cystoviruses (110).

Genomic rearrangement events (non-homologous recombination) in immunocompromised patients with acute infection were shown to be generated non-randomly and exhibited recombination hotspots (119). Secondary structure has been linked as a mechanism for recombination (120). If secondary structures form around breakpoints, this may cause recombination to occur, as is the case with poliovirus (121). This may be because regions, especially near the base of the hairpin would act as a stall for the polymerase. If a sequence forms a hairpin with itself, it could also hybridize with a second RNA having a similar but not identical sequence, which would allow for possible crossing over.

It is difficult to quantify the true frequency of recombination, as presumably most events would go unnoticed if they were between near-identical sequences or be too deleterious to result in viable viruses that could be isolated and sequenced. However, there is strong evidence showing that recombination does indeed occur between distinct strains of rotavirus A, and at least some of these events are able to persist well enough to be detected in sequences from multiple subsequent isolates.

## References


1. Desselberger U. 2014. Rotaviruses. Virus Research 190:75-96.
2. Arias CF, Romero P, Alvarez V, Lopez S. 1996. Trypsin activation pathway of rotavirus infectivity. J Virol 70:5832-9.
3. Matthijnssens J, Ciarlet M, Heiman E, Arijs I, Delbeke T, McDonald SM, Palombo EA, Iturriza-Gomara M, Maes P, Patton JT, Rahman M, Van Ranst M. 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 82:3204-19.
4. Nair N, Feng N, Blum LK, Sanyal M, Ding S, Jiang B, Sen A, Morton JM, He XS, Robinson WH, Greenberg HB. 2017. VP4- and VP7-specific antibodies mediate heterotypic immunity to rotavirus in humans. Sci Transl Med 9.
5. McDonald SM, Matthijnssens J, McAllen JK, Hine E, Overton L, Wang S, Lemey P, Zeller M, Van Ranst M, Spiro DJ, Patton JT. 2009. Evolutionary Dynamics of Human Rotaviruses: Balancing Reassortment with Preferred Genome Constellations. PLOS Pathogens 5:e1000634.
6. Patton JT. 2012. Rotavirus Diversity and Evolution in the Post-Vaccine World. Discovery Medicine 13:85-97.
7. Kirkwood CD. 2010. Genetic and Antigenic Diversity of Human Rotaviruses: Potential Impact on Vaccination Programs. Journal of Infectious Diseases 202:S43-S48.
8. Matthijnssens J, Otto PH, Ciarlet M, Desselberger U, Van Ranst M, Johne R. 2012. VP6-sequence-based cutoff values as a criterion for rotavirus species demarcation. Archives of Virology 157:1177-1182.
9. Mihalov-Kovacs E, Gellert A, Marton S, Farkas SL, Feher E, Oldal M, Jakab F, Martella V, Banyai K. 2015. Candidate new rotavirus species in sheltered dogs, Hungary. Emerg Infect Dis 21:660-3.
10. Banyai K, Kemenesi G, Budinski I, Foldes F, Zana B, Marton S, Varga-Kugler R, Oldal M, Kurucz K, Jakab F. 2017. Candidate new rotavirus species in Schreiber's bats, Serbia. Infect Genet Evol 48:19-26.
11. Matthijnssens J, Ciarlet M, McDonald SM, Attoui H, Bányai K, Brister JR, Buesa J, Esona MD, Estes MK, Gentsch JR, Iturriza-Gómara M, Johne R, Kirkwood CD, Martella V, Mertens PPC, Nakagomi O, Parreño V, Rahman M, Ruggeri FM, Saif LJ, Santos N, Steyer A, Taniguchi K,

Patton JT, Desselberger U, Van Ranst M. 2011. Uniformity of rotavirus strain nomenclature proposed by the Rotavirus Classification Working Group (RCWG). Archives of virology 156:1397-1413.

12. Steger CL, Boudreaux CE, LaConte LE, Pease JB, McDonald SM. 2019. Group A Rotavirus VP1 Polymerase and VP2 Core Shell Proteins: Intergenotypic Sequence Variation and In Vitro Functional Compatibility. Journal of Virology 93:e01642-18.

13. Rahman M, Matthijnssens J, Yang X, Delbeke T, Arijs I, Taniguchi K, Iturriza-Gomara M, Iftekharuddin N, Azim T, Van Ranst M. 2007. Evolutionary history and global spread of the emerging G12 human rotaviruses. Journal of Virology 81:2382-2390.

14. Matthijnssens J, Bilcke J, Ciarlet M, Martella V, Banyai K, Rahman M, Zeller M, Beutels P, Van Damme P, Van Rans M. 2009. Rotavirus disease and vaccination: impact on genotype diversity. Future Microbiology 4:1303-1316.

15. Ghosh S, Kobayashi N. 2011. Whole-genomic analysis of rotavirus strains: current status and future prospects. Future Microbiology 6:1049-1065.

16. Sadiq A, Bostan N, Yinda KC, Naseem S, Sattar S. 2018. Rotavirus: Genetics, pathogenesis and vaccine advances. Reviews in Medical Virology 28.

17. Donker NC, Kirkwood CD. 2012. Selection and evolutionary analysis in the nonstructural protein NSP2 of rotavirus A. Infection Genetics and Evolution 12:1355-1361.

18. Matthijnssens J, Heylen E, Zeller M, Rahman M, Lemey P, Van Ranst M. 2010. Phylodynamic Analyses of Rotavirus Genotypes G9 and G12 Underscore Their Potential for Swift Global Spread. Molecular Biology and Evolution 27:2431-2436.

19. Ramig RF, Ward RL. 1991. GENOMIC SEGMENT REASSORTMENT IN ROTAVIRUSES AND OTHER REOVIRIDAE. Advances in Virus Research 39:163-207.

20. Ramig RF. 1997. Genetics of the rotaviruses. Annual Review of Microbiology 51:225-255.

21. McDonald SM, Nelson MI, Turner PE, Patton JT. 2016. Reassortment in segmented RNA viruses: mechanisms and outcomes. Nature Reviews Microbiology 14:448-460.

22. Doro R, Farkas SL, Martella V, Banyai K. 2015. Zoonotic transmission of rotavirus: surveillance and control. Expert Review of Anti-Infective Therapy 13:1337-1350.

23. Desselberger U. 1996. Genome rearrangements of rotaviruses. Advances in Virus Research, Vol 46 46:69-95.

24. Varsani A, Lefeuvre P, Roumagnac P, Martin D. 2018. Notes on recombination and reassortment in multipartite/segmented viruses. Curr Opin Virol 33:156-166.
25. Lukashev AN. 2005. Role of recombination in evolution of enteroviruses. Rev Med Virol 15:157-67.
26. Pérez-Losada M, Arenas M, Galán JC, Palero F, González-Candelas F. 2015. Recombination in viruses: Mechanisms, methods of study, and evolutionary consequences. Infection, Genetics and Evolution 30:296-307.
27. Patton JT, Vasquez-Del Carpio R, Tortorici MA, Taraporewala ZF. 2007. Coupling of rotavirus genome replication and capsid assembly. Adv Virus Res 69:167-201.
28. Lai MM. 1992. RNA recombination in animal and plant viruses. Microbiol Rev 56:61-79.
29. Suzuki Y, Gojobori T, Nakagomi O. 1998. Intragenic recombinations in rotaviruses. FEBS Lett 427:183-7.
30. Parra GI, Bok K, Martinez M, Gomez JA. 2004. Evidence of rotavirus intragenic recombination between two sublineages of the same genotype. J Gen Virol 85:1713-6.
31. Phan TG, Okitsu S, Maneekarn N, Ushijima H. 2007. Evidence of intragenic recombination in G1 rotavirus VP7 genes. J Virol 81:10188-94.
32. Phan TG, Okitsu S, Maneekarn N, Ushijima H. 2007. Genetic heterogeneity, evolution and recombination in emerging G9 rotaviruses. Infection Genetics and Evolution 7:656-663.
33. Martinez-Laso J, Roman A, Rodriguez M, Cervera I, Head J, Rodriguez-Avial I, Picazo JJ. 2009. Diversity of the G3 genes of human rotaviruses in isolates from Spain from 2004 to 2006: cross-species transmission and inter-genotype recombination generates alleles. J Gen Virol 90:935-43.
34. Donker NC, Boniface K, Kirkwood CD. 2011. Phylogenetic analysis of rotavirus A NSP2 gene sequences and evidence of intragenic recombination. Infect Genet Evol 11:1602-7.
35. Jere KC, Mlera L, Page NA, van Dijk AA, O'Neill HG. 2011. Whole genome analysis of multiple rotavirus strains from a single stool specimen using sequence-independent amplification and 454(R) pyrosequencing reveals evidence of intergenotype genome segment recombination. Infect Genet Evol 11:2072-82.
36. Jing Z, Zhang X, Shi H, Chen J, Shi D, Dong H, Feng L. 2018. A G3P 13 porcine group A rotavirus emerging in China is a reassortant and a natural recombinant in the VP4 gene. Transboundary and Emerging Diseases 65:e317-e328.

37. Esona MD, Roy S, Rungsrisuriyachai K, Sanchez J, Vasquez L, Gomez V, Rios LA, Bowen MD, Vazquez M. 2017. Characterization of a triple-recombinant, reassortant rotavirus strain from the Dominican Republic. Journal of General Virology 98:134-142.
38. Woods RJ. 2015. Intrasegmental recombination does not contribute to the long-term evolution of group A rotavirus. Infection Genetics and Evolution 32:354-360.
39. Hatcher EL, Zhdanov SA, Bao Y, Blinkova O, Nawrocki EP, Ostapchuck Y, Schaffer AA, Brister JR. 2017. Virus Variation Resource - improved response to emergent viral outbreaks. Nucleic Acids Res 45:D482-d490.
40. Matthijnssens J, Ciarlet M, Rahman M, Attoui H, Bányai K, Estes MK, Gentsch JR, Iturriza-Gómara M, Kirkwood CD, Martella V, Mertens PPC, Nakagomi O, Patton JT, Ruggeri FM, Saif LJ, Santos N, Steyer A, Taniguchi K, Desselberger U, Van Ranst M. 2008. Recommendations for the classification of group A rotaviruses using all 11 genomic RNA segments. Archives of virology 153:1621-1629.
41. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792-7.
42. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol 1:vev003.
43. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol 4:vey016.
44. Minin VN, Bloomquist EW, Suchard MA. 2008. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. Mol Biol Evol 25:1459-71.
45. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Systematic Biology 67:901-904.
46. Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7:214.
47. Rambaut A. 2019. Figtree v1.4.4. Accessed
48. Lorenz R, Bernhart SH, Honer Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. Algorithms Mol Biol 6:26.
49. Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF. 2008. RNAalifold: improved consensus structure prediction for RNA alignments. BMC Bioinformatics 9:474.

50. Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5:725-38.
51. Yang J, Zhang Y. 2015. Protein Structure and Function Prediction Using I-TASSER. Curr Protoc Bioinformatics 52:5 8 1-15.
52. Zhang Y. 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9:40.
53. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, Wheeler DK, Sette A, Peters B. 2019. The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Res 47:D339-d343.
54. Yao B, Zhang L, Liang S, Zhang C. 2012. SVMTriP: a method to predict antigenic epitopes using support vector machine to integrate tri-peptide similarity and propensity. PLoS One 7:e45152.
55. Aida S, Nahar S, Paul SK, Hossain MA, Kabir MR, Sarkar SR, Ahmed S, Ghosh S, Urushibara N, Kawaguchiya M, Aung MS, Sumi A, Kobayashi N. 2016. Whole genomic analysis of G2P[4] human Rotaviruses in Mymensingh, north-central Bangladesh. Heliyon 2:e00168-e00168.
56. Bwogi J, Jere KC, Karamagi C, Byarugaba DK, Namuwulya P, Baliraine FN, Desselberger U, Iturriza-Gomara M. 2017. Whole genome analysis of selected human and animal rotaviruses identified in Uganda from 2012 to 2014 reveals complex genome reassortment events between human, bovine, caprine and porcine strains. PLoS One 12:e0178855.
57. Ghosh A, Chattopadhyay S, Chawla-Sarkar M, Nandy P, Nandy A. 2012. In silico study of rotavirus VP7 surface accessible conserved regions for antiviral drug/vaccine design. PLoS One 7:e40749.
58. Boni MF, de Jong MD, van Doorn HR, Holmes EC. 2010. Guidelines for identifying homologous recombination events in influenza A virus. PLoS One 5:e10434.
59. Bertrand Y, Töpel M, Elväng A, Melik W, Johansson M. 2012. First Dating of a Recombination Event in Mammalian Tick-Borne Flaviviruses. PLOS ONE 7:e31981.
60. Boni MF, Smith GJD, Holmes EC, Vijaykrishna D. 2012. No evidence for intra-segment recombination of 2009 H1N1 influenza virus in swine. Gene 494:242-245.
61. Worobey M, Rambaut A, Pybus OG, Robertson DL. 2002. Questioning the evidence for genetic recombination in the 1918 "Spanish flu" virus. Science 296:211 discussion 211.
62. Bertrand YJK, Johansson M, Norberg P. 2016. Revisiting Recombination Signal in the Tick-Borne Encephalitis Virus: A Simulation Approach. PLOS ONE 11:e0164435.
63. Norberg P, Roth A, Bergstrom T. 2013. Genetic recombination of tick-borne flaviviruses among wild-type strains. Virology 440:105-16.

64. Boni MF, Zhou Y, Taubenberger JK, Holmes EC. 2008. Homologous Recombination Is Very Rare or Absent in Human Influenza A Virus. Journal of Virology 82:4807.

65. Boni MF, Posada D, Feldman MW. 2007. An Exact Nonparametric Method for Inferring Mosaic Structure in Sequence Triplets. Genetics 176:1035.

66. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, Ray SC. 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol 73:152-60.

67. He CQ, Han GZ, Wang D, Liu W, Li GR, Liu XP, Ding NZ. 2008. Homologous recombination evidence in human and swine influenza A viruses. Virology 380:12-20.

68. Santiana M, Ghosh S, Ho BA, Rajasekaran V, Du WL, Mutsafi Y, De Jesus-Diaz DA, Sosnovtsev SV, Levenson EA, Parra GI, Takvorian PM, Cali A, Bleck C, Vlasova AN, Saif LJ, Patton JT, Lopalco P, Corcelli A, Green KY, Altan-Bonnet N. 2018. Vesicle-Cloaked Virus Clusters Are Optimal Units for Inter-organismal Viral Transmission. Cell Host Microbe 24:208-220.e8.

69. Abdel-Haq NM, Thomas RA, Asmar BI, Zacharova V, Lyman WD. 2003. Increased prevalence of G1P[4] genotype among children with rotavirus-associated gastroenteritis in metropolitan Detroit. J Clin Microbiol 41:2680-2.

70. Husain M, Seth P, Dar L, Broor S. 1996. Classification of rotavirus into G and P types with specimens from children with acute diarrhea in New Delhi, India. Journal of Clinical Microbiology 34:1592.

71. Khetawat D, Dutta P, Bhattacharya SK, Chakrabarti S. 2002. Distribution of rotavirus VP7 genotypes among children suffering from watery diarrhea in Kolkata, India. Virus Res 87:31-40.

72. Jain V, Das BK, Bhan MK, Glass RI, Gentsch JR. 2001. Great diversity of group A rotavirus strains and high prevalence of mixed rotavirus infections in India. J Clin Microbiol 39:3524-9.

73. Sánchez-Fauquier A, Montero V, Moreno S, Solé M, Colomina J, Iturriza-Gomara M, Revilla A, Wilhelmi I, Gray J, Gegavi V-NG. 2006. Human rotavirus G9 and G3 as major cause of diarrhea in hospitalized children, Spain. Emerging infectious diseases 12:1536-1541.

74. Kiulia NM, Peenze I, Dewar J, Nyachieo A, Galo M, Omolo E, Steele AD, Mwenda JM. 2006. Molecular characterisation of the rotavirus strains prevalent in Maua, Meru North, Kenya. East Afr Med J 83:360-5.

75. Mwenda JM, Ntoto KM, Abebe A, Enweronu-Laryea C, Amina I, McHomvu J, Kisakye A, Mpabalwani EM, Pazvakavambwa I, Armah GE, Seheri LM, Kiulia NM, Page N, Widdowson M-A, Duncan Steele A. 2010. Burden and Epidemiology of Rotavirus Diarrhea in Selected African Countries: Preliminary Results from the African Rotavirus Surveillance Network. The Journal of Infectious Diseases 202:S5-S11.

76. Anaya-Molina Y, De La Cruz Hernández SI, Andrés-Dionicio AE, Terán-Vega HL, Méndez-Pérez H, Castro-Escarpulli G, García-Lozano H. 2018. A one-step real-time RT-PCR helps to identify mixed rotavirus infections in Mexico. Diagnostic Microbiology and Infectious Disease 92:288-293.

77. Rojas M, Dias HG, Gonçalves JLS, Manchego A, Rosadio R, Pezo D, Santos N. 2019. Genetic diversity and zoonotic potential of rotavirus A strains in the southern Andean highlands, Peru. Transboundary and Emerging Diseases 0.

78. Mwanga M, Nyaigoti CA, Njeru R, Nokes DJ, Matthew C, Phan VTM, Holmes EC. 2018. Unbiased whole genome deep sequencing of Rotavirus Group A positive samples from rural Kenya, 2012-14 reveals high frequency of coinfection and genetic reassortment. International Journal of Infectious Diseases 73:203.

79. Martella V, Banyai K, Matthijnssens J, Buonavoglia C, Ciarlet M. 2010. Zoonotic aspects of rotaviruses. Veterinary Microbiology 140:246-255.

80. Li W, Manktelow E, von Kirchbach JC, Gog JR, Desselberger U, Lever AM. 2010. Genomic analysis of codon, sequence and structural conservation with selective biochemical-structure mapping reveals highly conserved and dynamic structures in rotavirus RNAs with potential cis-acting functions. Nucleic Acids Res 38:7718-35.

81. Gentsch JR, Laird AR, Bielfelt B, Griffin DD, Banyai K, Ramachandran M, Jain V, Cunliffe NA, Nakagomi O, Kirkwood CD, Fischer TK, Parashar UD, Bresee JS, Jiang B, Glass RI. 2005. Serotype diversity and reassortment between human and animal rotavirus strains: implications for rotavirus vaccine programs. J Infect Dis 192 Suppl 1:S146-59.

82. Martella V, Ciarlet M, Pratelli A, Arista S, Terio V, Elia G, Cavalli A, Gentile M, Decaro N, Greco G, Cafiero MA, Tempesta M, Buonavoglia C. 2003. Molecular analysis of the VP7, VP4, VP6, NSP4, and NSP5/6 genes of a buffalo rotavirus strain: identification of the rare P[3] rhesus rotavirus-like VP4 gene allele. Journal of clinical microbiology 41:5665-5675.

83. Aoki ST, Settembre EC, Trask SD, Greenberg HB, Harrison SC, Dormitzer PR. 2009. Structure of rotavirus outer-layer protein VP7 bound with a neutralizing Fab. Science 324:1444-7.

84. Ludert JE, Ruiz MC, Hidalgo C, Liprandi F. 2002. Antibodies to rotavirus outer capsid glycoprotein VP7 neutralize infectivity by inhibiting virion decapsidation. J Virol 76:6643-51.

85. Zhou YJ, Burns JW, Morita Y, Tanaka T, Estes MK. 1994. Localization of rotavirus VP4 neutralization epitopes involved in antibody-induced conformational changes of virus structure. J Virol 68:3955-64.

86. Hu L, Sankaran B, Laucirica DR, Patil K, Salmen W, Ferreon ACM, Tsoi PS, Lasanajak Y, Smith DF, Ramani S, Atmar RL, Estes MK, Ferreon JC, Prasad BVV. 2018. Glycan recognition in globally dominant human rotaviruses. Nat Commun 9:2631.

87. Huang P, Xia M, Tan M, Zhong W, Wei C, Wang L, Morrow A, Jiang X. 2012. Spike protein VP8* of human rotavirus recognizes histo-blood group antigens in a type-specific manner. J Virol 86:4833-43.

88. Lee B, Dickson DM, deCamp AC, Ross Colgate E, Diehl SA, Uddin MI, Sharmin S, Islam S, Bhuiyan TR, Alam M, Nayak U, Mychaleckyj JC, Taniuchi M, Petri WA, Jr., Haque R, Qadri F, Kirkpatrick BD. 2018. Histo-Blood Group Antigen Phenotype Determines Susceptibility to Genotype-Specific Rotavirus Infections and Impacts Measures of Rotavirus Vaccine Efficacy. J Infect Dis 217:1399-1407.

89. Ruggeri FM, Greenberg HB. 1991. Antibodies to the trypsin cleavage peptide VP8 neutralize rotavirus by inhibiting binding of virions to target cells in culture. J Virol 65:2211-9.

90. Parez N, Garbarg-Chenon A, Fourgeux C, Le Deist F, Servant-Delmas A, Charpilienne A, Cohen J, Schwartz-Cornil I. 2004. The VP6 protein of rotavirus interacts with a large fraction of human naive B cells via surface immunoglobulins. J Virol 78:12489-96.

91. Aiyegbo MS, Eli IM, Spiller BW, Williams DR, Kim R, Lee DE, Liu T, Li S, Stewart PL, Crowe JE, Jr. 2014. Differential accessibility of a rotavirus VP6 epitope in trimers comprising type I, II, or III channels as revealed by binding of a human rotavirus VP6-specific antibody. J Virol 88:469-76.

92. Hua J, Chen X, Patton JT. 1994. Deletion mapping of the rotavirus metalloprotein NS53 (NSP1): the conserved cysteine-rich region is essential for virus-specific RNA binding. J Virol 68:3990-4000.

93. Arnold MM, Patton JT. 2011. Diversity of interferon antagonist activities mediated by NSP1 proteins of different rotavirus strains. J Virol 85:1970-9.

94. Fajardo T, Jr., Sung PY, Roy P. 2015. Disruption of Specific RNA-RNA Interactions in a Double-Stranded RNA Virus Inhibits Genome Packaging and Virus Infectivity. PLoS Pathog 11:e1005321.

95. Fajardo T, Sung PY, Celma CC, Roy P. 2017. Rotavirus Genomic RNA Complex Forms via Specific RNA-RNA Interactions: Disruption of RNA Complex Inhibits Virus Infectivity. Viruses 9.

96. Suzuki Y. 2015. A candidate packaging signal of human rotavirus differentiating Wa-like and DS-1-like genomic constellations. Microbiol Immunol 59:567-71.

97. Borodavka A, Dykeman EC, Schrimpf W, Lamb DC. 2017. Protein-mediated RNA folding governs sequence-specific interactions between rotavirus genome segments. Elife 6.

98. Chare ER, Holmes EC. 2006. A phylogenetic survey of recombination frequency in plant RNA viruses. Arch Virol 151:933-46.

99. Rawson JMO, Nikolaitchik OA, Keele BF, Pathak VK, Hu W-S. 2018. Recombination is required for efficient HIV-1 replication and the maintenance of viral genome integrity. Nucleic Acids Research 46:10535-10545.

100. Decrey L, Kazama S, Kohn T. 2016. Ammonia as an <em>In Situ</em> Sanitizer: Influence of Virus Genome Type on Inactivation. Applied and Environmental Microbiology 82:4909.

101. Simon-Loriere E, Holmes EC. 2011. Why do RNA viruses recombine? Nat Rev Microbiol 9:617-26.

102. Iturriza-Gomara M, Isherwood B, Desselberger U, Gray J. 2001. Reassortment in vivo: driving force for diversity of human rotavirus strains isolated in the United Kingdom between 1995 and 1999. J Virol 75:3696-705.

103. Jere KC, Chaguza C, Bar-Zeev N, Lowe J, Peno C, Kumwenda B, Nakagomi O, Tate JE, Parashar UD, Heyderman RS, French N, Cunliffe NA, Iturriza-Gomara M, Consortium V. 2018. Emergence of Double- and Triple-Gene Reassortant G1P 8 Rotaviruses Possessing a DS-1-Like Backbone after Rotavirus Vaccine Introduction in Malawi. Journal of Virology 92.

104. Schumann T, Hotzel H, Otto P, Johne R. 2009. Evidence of interspecies transmission and reassortment among avian group A rotaviruses. Virology 386:334-343.

105. Zeldovich KB, Liu P, Renzette N, Foll M, Pham ST, Venev SV, Gallagher GR, Bolon DN, Kurt-Jones EA, Jensen JD, Caffrey DR, Schiffer CA, Kowalik TF, Wang JP, Finberg RW. 2015. Positive

Selection Drives Preferred Segment Combinations during Influenza Virus Reassortment. Mol Biol Evol 32:1519-32.

106. He CQ, Ding NZ, He M, Li SN, Wang XM, He HB, Liu XF, Guo HS. 2010. Intragenic recombination as a mechanism of genetic diversity in bluetongue virus. J Virol 84:11487-95.

107. Simon-Loriere E, Holmes EC. 2011. Why do RNA viruses recombine? Nature Reviews Microbiology 9:617-626.

108. Bentley K, Evans DJ. 2018. Mechanisms and consequences of positive-strand RNA virus recombination. J Gen Virol 99:1345-1356.

109. McDonald SM, Patton JT. 2011. Assortment and packaging of the segmented rotavirus genome. Trends Microbiol 19:136-44.

110. Mindich L. 2004. Packaging, replication and recombination of the segmented genomes of bacteriophage Phi 6 and its relatives. Virus Research 101:83-92.

111. Carpi G, Holmes EC, Kitchen A. 2010. The Evolutionary Dynamics of Bluetongue Virus. Journal of Molecular Evolution 70:583-592.

112. Hon CC, Lam TT, Yip CW, Wong RT, Shi M, Jiang J, Zeng F, Leung FC. 2008. Phylogenetic evidence for homologous recombination within the family Birnaviridae. J Gen Virol 89:3156-64.

113. Li Y, Xia Z, Peng J, Zhou T, Fan Z. 2013. Evidence of recombination and genetic diversity in southern rice black-streaked dwarf virus. Arch Virol 158:2147-51.

114. Silander OK, Weinreich DM, Wright KM, O'Keefe KJ, Rang CU, Turner PE, Chao L. 2005. Widespread genetic exchange among terrestrial bacteriophages. Proceedings of the National Academy of Sciences of the United States of America 102:19009-19014.

115. Robertson DL, Hahn BH, Sharp PM. 1995. Recombination in AIDS viruses. J Mol Evol 40:249-59.

116. Hahn CS, Lustig S, Strauss EG, Strauss JH. 1988. Western equine encephalitis virus is a recombinant virus. Proc Natl Acad Sci U S A 85:5997-6001.

117. Chare ER, Gould EA, Holmes EC. 2003. Phylogenetic analysis reveals a low rate of homologous recombination in negative-sense RNA viruses. J Gen Virol 84:2691-703.

118. Patton JT, Jones MT, Kalbach AN, He YW, Xiaobo J. 1997. Rotavirus RNA polymerase requires the core shell protein to synthesize the double-stranded RNA genome. J Virol 71:9618-26.

119. Schnepf N, Deback C, Dehee A, Gault E, Parez N, Garbarg-Chenon A. 2008. Rearrangements of rotavirus genomic segment 11 are generated during acute infection of immunocompetent children and do not occur at random. J Virol 82:3689-96.

120. Dedepsidis E, Kyriakopoulou Z, Pliaka V, Markoulatos P. 2010. Correlation between recombination junctions and RNA secondary structure elements in poliovirus Sabin strains. Virus Genes 41:181-91.

121. Tolskaya EA, Romanova LI, Blinov VM, Viktorova EG, Sinyakov AN, Kolesnikova MS, Agol VI. 1987. Studies on the recombination between RNA genomes of poliovirus: the primary structure and nonrandom distribution of crossover regions in the genomes of intertypic poliovirus recombinants. Virology 161:54-61.

## *SUPPLEMENTARY INFORMATION*

**Supplementary Table 1**. Supported putative recombinants identified by RDP4 for whole genome Rotavirus A sequences. 117 events were identified where at least 6/7 methods listed supported the event. Shaded rows indicate strongly supported events (identified by 7/7 RDP4 programs). Events that were flagged by RDP4 but showed less support are not listed here.

| Segment | Accession Number | Breakpoint Cis (Excluding Gaps) | p-Values | Representative Major Parent Sequence, Source Species, Year of Isolation, % ANI, Genogroup | Representative Minor Parent Sequence, Source Species, Year of Isolation, % ANI, Genogroup |
|---|---|---|---|---|---|
| VP1 | JQ988899<br>*Homo sapiens*, Croatia, 2006 | (972), (353-509)-(1702-1760) | RDP 1.00E-43<br>GENECONV. 9.08E-47<br>BootScan 4.28E-27<br>MaxChi. 5.44E-18<br>Chimera 3.97E-18<br>SiScan 3.46E-27<br>3Seq 2.83E-09 | KJ751768<br>*Homo sapiens*, South Africa, 94.7% ,R1 | GQ414540<br>*Homo sapiens*, Germany, 2009, 98.7%, R2 |
| VP1 | KU739903<br>*Sus scrofa*, Taiwan, 2015 | 1-(747-805) | RDP 1.19E-24<br>GENECONV 1.41E-21<br>BootScan. 5.14E-21<br>MaxChi. 8.41E-16<br>Chimera 3.57E-16<br>SiScan. 9.64E-16<br>3Seq 2.83E-09 | KU739900<br>*Sus scrofa*, Taiwan, 2015, 95.8%, R1 | KU739904<br>*Sus scrofa*, Taiwan, 2015, 99.1%, R1 |
| VP1 | KF812721<br>*Homo sapiens*, South Korea, 2010 | (2374-2537)-(3268) | RDP 7.11E-23<br>GENECONV 1.75E-16<br>BootScan 2.44E-21<br>MaxChi 9.35E-16<br>Chimera 5.18E-15<br>SiScan 2.49E-21<br>3Seq 5.15E-29 | JX027681<br>*Homo sapiens*, Australia, 2007, 98.1%, R1 | Unknown, closest = JQ069926, R1 |
| VP1 | JQ069926<br>*Homo sapiens*, Canada, 2008 | 2498 (2356-2507) – 3268 | RDP 9.14E-22<br>GENECONV 5.71E-19<br>BootScan 1.00E-18<br>MaxChi 1.30E-15<br>Chimera 2.51E-15<br>SiScan. 7.42E-23<br>3Seq 2.78E-25 | JX195074<br>*Homo sapiens,* Italy, 2010, 98.1%<br>R1 | JX027681<br>*Homo sapiens*, Australia, 2007, 99.6%<br>R1 |
| VP1 | KM454503<br>*Equus callabus*, USA, 1981 | 840 (802-951)-3281 | RDP 2.77E-15<br>GENECONV 1.36E-19<br>BootScan 4.79 E-21<br>MaxChi 1.55E-17<br>SiScan 1.27E-40 | Unknown, closest = DQ838638, R2 | KM454492<br>*Equus callabus*, United Kingdom, 1976, 99.1%, R2 |

| | | | | | |
|---|---|---|---|---|---|
| | | | 3Seq 8.22E-05 | | |
| VP1 | KU714444<br>*Homo sapiens*, Malawi, 2000 | 1656 (1636-1673)-1752 (1740-1766) | RDP 1.35E-18<br>GENECONV 1.25E-17<br>BootScan 3.76E-11<br>MaxChi 3.81E-02<br>Chimera 4.33E-02<br>3Seq 5.67E-09 | JN706454, *Homo sapiens,* Thailand, 2009, 97.4%, R1 | KX655451, *Homo sapiens*, Uganda, 2013, 97.9%, R2 |
| VP1 | JQ069936, *Homo sapiens*, Canada, 2009 | 1670 (1531, 1703)-2419 (2308-2436) | RDP 1.83E-14<br>GENECONV 7.68E-13<br>MaxChi 2.73E-08<br>Chimera 2.48E-08<br>SiScan 2.65E-08<br>3Seq 5.15E-19 | KU248405, *Homo sapiens*, Bangladesh, 2010 (98.7%), R2 | HQ657171, *Homo sapiens*, South Africa, 2009, 99.7%, R2 |
| VP1 | KU739903, *Sus scrofa*, Taiwan, 2015 | 1543 (1383-1588)-2220 (2160-2260) | RDP 3.06E-13<br>GENECONV 6.19E-04<br>BootScan 9.40E-11<br>MaxChi 2.98E-08<br>Chimera 5.67E-08<br>SiScan 1.19 E-06<br>3Seq 2.83E-09 | KU739900<br>*Sus scrofa*, Taiwan, 2015, 97.8%, R1 | KU739904<br>*Sus scrofa*, Taiwan, 2015, 96.5%, R1 |
| VP1 | JQ988899<br>*Homo sapiens*, Croatia, 2006 | 390 (241-428)-971 (894-1847) | RDP 3.76E-12<br>GENECONV 4.88E-10<br>MaxChi 7.74E-09<br>Chimera 4.5E-07<br>SiScan 6.32E-14<br>3Seq 5.64E-09 | JX195074<br>*Homo sapiens*, Italy, 2010, 93.5%<br>R1 | Unknown, JQ069918, *Homo sapiens*, Canada, 2008<br>R1 |
| VP1 | KC580176, *Homo sapiens*, USA, 1988 | 1471 (1328-1521)-3270 | RDP 9.77E-12<br>GENECONV 3.81E-08<br>MaxChi 6.16E-12<br>Chimera 2.36E-04<br>SiScan. 7.42E-23<br>3Seq 1.91E-20 | KC580526, *Homo sapiens,* USA, 1979, 99%, R1 | KC580601, *Homo sapiens,* USA, 1988, 100%, R1 |
| VP1 | JQ069918, *Homo sapiens*, Canada, 2008 | 604 (541-662)-956 (889-1068) | RDP 2.42E-10<br>GENECONV 3.25E-08<br>BootScan 2.38E-09<br>MaxChi 5.62E-03<br>Chimera 5.62E-03<br>3.28E-03<br>SiSeq 1.10E-04<br>3Seq 1.13E-08 | KJ751834, *Homo sapiens*, South Africa, 2009, 99.5%, R1 | JQ069924, *Homo sapiens*, Canada, 2008, 99.7%, R1 |
| VP1 | KC579647, *Homo sapiens*, USA, 1988 | 35 – 1448 (1290-1498) | GENECONV 4.30E-04<br>BootScan 2.63E-05<br>MaxChi 2.34E-09<br>Chimera 1.41E-08<br>SiScan 6.55E-11<br>3Seq 4.34E-19 | KC580601, *Homo sapiens,* USA, 1988, 100%, R1 | KC579509, *Homo sapiens,* USA, 1989, 99.4%, R1 |

| | | | | | |
|---|---|---|---|---|---|
| VP1 | KU199270, *Homo sapiens*, Bangladesh, 2010 | 670 (541-711)-1403 (1327-1445) | Flagged for 285 seqs, p-value= 3.33E-09 | KX536654, *Homo sapiens*, India, 2011, 98.6%<br>R2 | KU356640, *Homo sapiens,* Bangladesh, 2013, 100%<br>R2 |
| VP1 | KU356662, KU356640, *Homo sapiens*, 2013 | 656 (541-711)-2661 (2600-2735) | RDP 5.87E-08<br>GENECONV 1.43E-04<br>MaxCh 9.02E-10<br>SiScan 6.71E-20<br>3Seq 6.88 E-23 | KU356607, *Homo sapiens*, Bangladesh, 2013, 99.5%, R2 | KC178768, *Homo sapiens*, Italy, 2007, 99.6%, R2 |
| VP1 | KU199270, *Homo sapiens*, Bangladesh, 2010 | 7-315 | RDP 4.80E-03<br>GENECONV 3.11E-03<br>BootScan 1.68E-02<br>MaxChi 4.91E-02<br>Chimera 4.65E-02<br>3Seq 2.43E-03 | MG181755, *Homo sapiens*, Malawi, 2014, 98.8%, R2 | KU248372, *Homo sapiens*, Bangladesh, 2010, 100%, R2 |
| VP2 | KJ753425, *Homo sapiens*, Uganda, 2011 | 599 (561-623) – 991 (967-1019) | RDP 6.28E-47<br>GENECONV 5.13E-45<br>BootScan 2.53E-44<br>MaxChi 5.39E-12<br>Chimera 6.10E-12<br>SiScan 8.38E-15<br>3Seq 1.96E-09 | JF490148, *Homo sapiens*, Australia, 2004, 98.3% ,C1 | KP752972, *Homo sapiens*, Gambia, 2008, 99.7%, R2-C2 |
| VP2 | KU714445, *Homo sapiens*, Malawi, 2000 | 1155(1139-1170) – 1887 (1862-1908) | RDP 2.37E-41<br>GENECONV 1.62E-33<br>BootScan 5.28E-39<br>MaxChi 1.26E-16<br>Chimera 1.42E-16<br>SiScan 8.61E-25<br>3Seq 1.96E-09 | KP752983, *Homo sapiens*, South Africa, 2009, 94.3%, C1 | KU714456, *Homo sapiens*, Malawi, 2000, 99.9%, R2-C2 |
| VP2 | KX655439, *Homo sapiens*, Uganda, 2013 | 2713 (2672-2717)-158 (143-171) | RDP 1.34E-25<br>GENECONV 9.67E-24<br>MaxChi 1.54E-03<br>Chimera 8.56E-04<br>SiScan 1.61E-05<br>3Seq 2.15E--08 | Unknown closest = KX632343, *Homo sapiens*, Uganda, 2013), C2 | MG181349, *Homo sapiens*, Malawi, 2002, 99.4%, C1 |
| VP2 | KJ753617, *Homo sapiens*, South Africa, 2004 | 548 (464-560)-784 (754-842) | RDP 5.31E-27<br>GENECONV 4.03E-24<br>BootScan 7.13E-23<br>MaxChi 3.85E-05<br>Chimera 4.51E-05<br>3Seq 1.96E-09 | KJ752329, *Homo sapiens*, South Africa, 2003, 99.4%, R1-C1 | KX655439, *Homo sapiens*, Uganda, 2013, 98.4%, R2-C2 |
| VP2 | HM988959, *Bos taurus*, South Korea, Unknown | 558 (524-586)-2685 | RDP 1.14E-21<br>GENECONV 2.68E-25<br>BootScan 2.00 E-21<br>MaxChi 1.29E-05<br>Chimera 1.11E-03<br>SiScan 3.88E-15<br>3Seq 5.94E-21 | JQ069842, *Homo sapiens,* Canada, 2008, 95.2%, C1 | JX971570, *Bos taurus*, South Korea, 2004, 100%, C1 |

| | | | | | |
|---|---|---|---|---|---|
| Vp2 | MG181591, *Homo sapiens*, Malawi, 2013 | 1129 (1106-1140)-1299 (1292-1318) | RDP 3.47 E-21<br>GENECONV 2.01E-19<br>BootScan 4.15E-15<br>MaxChi 4.04E-03<br>Chimera 3.95E-03<br>3Seq 1.96E-09 | KJ753828, *Homo sapiens*, Zimbabwe, 99.1%, C2 | KJ753425, *Homo sapiens*, Uganda, 2011, 99.4% ,C1 |
| VP2 | KX632343, *Homo sapiens*, Uganda, 2013 | 2666 (2614-39)-593 (560-612) | RDP 9.89E-21<br>GENECONV 2.26 E-17<br>BootScan 5.20E-19<br>MaxChi 4.65E-13<br>Chimera 3.18 E-13<br>SiSeq 2.53E-24<br>3Seq 1.37E-05 | JN258812, *Homo sapiens*, Belgium, 2000, 91.7%, C1 | KJ53402, *Homo sapiens*, South Africa, 2005, 99.5% |
| VP2 | KX632343, *Homo sapiens*, Uganda, 2013 | 1357 (1322-1368)-1808 (1790-1830) | RDP 1.14E-21<br>GENECONV 3.44E-17<br>MaxChi 1.69E-11<br>Chimera 3.57E-10<br>SiScan 2.84E-14<br>3Seq 1.96E-09 | MG181371, *Homo sapiens*, Malawi, 2002, 95.3%, C1 | KJ870879, *Homo sapiens*, Democratic Republic of the Congo, 99.3%, C2 |
| VP2 | KY055428, *Capra aegagrus*, Uganda, 2014 | 983 (960-1009) – 1238 (1194-1243) | RDP 1.64E-18<br>GENECONV 1.12E-16<br>BootScan 2.13E-10<br>MaxChi 1.67E-05<br>Chimera 4.73E-04<br>SiScan 4.35E-05<br>3Seq 1.96E-09 | JN831232, *Bos taurus*, South Africa, 2007, 94.7%, C2 | KF63257, *Bos taurus*, South Africa, 2007, 99.6% |
| VP2 | KJ870901, *Homo sapiens*, Democratic Republic of the Congo, Unknown | 2330 (2304-2359)-2648 (2589-1) | RDP 1.96E-18<br>GENECONV 2.99E-16<br>BootScan 5.51E-16<br>MaxChi 1.71E-03<br>Chimera 7.67E-04<br>SiScan 6.29E-06<br>3Seq 1.96E-09 | KY055417, *Sus scrofa*, Uganda, 2016, 95%, C1 | LC367260, *Homo sapiens*, Nepal, 2008, 99.4%, C1 |
| VP2 | KC257092, *Camelus dromedaries*, Sudan, 2002 | 2047 (2017-2102)-733 (661-754) | RDP 2.53E-16<br>GENECONV 4.72E-06<br>BootScan 7.95E-23<br>MaxChi 2.68E-13<br>Chimera 8.40E-12<br>SiScan 8.51E-11<br>3Seq 9.20E-08 | Unknown, closest = DQ480724, *Homo sapiens*, Iran, C2 | FJ347123, *Lama guanicoe*, Argentina, 1998, 99.9%, C2 |
| VP2 | KX632343, *Homo sapiens*, Uganda, 2013 | 1962 (1932-1978) – 2133 (2099-2136) | RDP 1.44E-13<br>GENECONV 7.46E-10<br>MaxChi 3.26E-06<br>Chimera 4.13E-05<br>SiScan 1.49E-07<br>3Seq1.96E-09 | JN706477, *Homo sapiens*, Thailand, 2010, 93%<br>C1 | KX655439, *Homo sapiens*, Uganda, 2013, 97.7%<br>C2 |

| | | | | | |
|---|---|---|---|---|---|
| VP2 | KJ627064, *Homo sapiens*, Paraguay, 2002 | 2675 (2615-95)-409 (370-449) | RDP 7.01E-06<br>GENECONV 2.53E-03<br>BootScan 1.26E-04<br>MaxChi 4.49E-02<br>Chimera 3.89E-02<br>SiScan 1.20E-02<br>3Seq 9.21E-04 | KJ752074, *Homo sapiens*, South Africa, 2005, 98.5% , C1 | KJ752859, *Homo sapiens*, Zimbabwe, 2011, 100%, C1 |
| VP3 | KJ412679, *Homo sapiens,* Paraguay, 2009 | 1 (2497-3) – 1068 (1047-1080) | RDP 1.04E-72<br>GENECONV 1.11E-59<br>BootScan 1.41E-73<br>MaxChi 3.58E-31<br>Chimera 5.31E-31<br>SiScan 6.77E-38<br>3Seq 2.83E-09 | JX185760, *Homo sapiens*, Italy, 2007, 97.8%, M1 | KJ412622, *Homo sapiens*, Paraguay, 2009, 99.9%, M3 |
| VP3 | KP007164, *Homo sapiens*, Philippines, 2012 | 968 (968-996)-1753 (1727-1777) | RDP 1.97E-38<br>GENECONV 2.93E-35<br>BootScan 2.16E-37<br>MaxChi 2.45E-16<br>Chimera 1.76E-15<br>SiScan 2.85E-19<br>3Seq 2.83E-19 | KY000551, *Homo sapiens*, Germany, 2016, 99.4%, M2 | KP882081, *Homo sapiens*, Bangladesh, 2007, 99.7%, M2 |
| VP3 | KX171542, *Homo sapiens*, South Korea, 2014 | 1465 (1432-1538)-2482 (2461-59) | RDP 1.40E-37<br>GENECONV 7.17E-32<br>BootScan 2.84E-35<br>MaxChi 2.03E-19<br>Chimera 3.57E-19<br>SiScan 3.27E-23<br>3Seq 2.83E-09 | KJ752827, *Homo sapiens,* South Africa, 2003, 98.8%, M2 | KX171538, *Homo sapiens,* South Korea, 2012, 99.7%, M2 |
| VP3 | KX171543, *Homo sapiens*, South Korea, 2014 | 1650 (1607-1688)-2472 (2455-59) | RDP 5.19E-37<br>GENECONV 2.46E-33<br>BootScan 2.81E-34<br>MaxChi 1.95E-17<br>Chimera 1.05E-17<br>SiScan 1.55E-19<br>3Seq 2.83E-09 | KU248374, *Homo sapiens*, Bangladesh, 2010, 99.1%, M2 | KC443523, *Homo sapiens*, Australia, 2007, 99.6%, M2 |
| VP3 | AB779630, *Sus scrofa,* Thailand, 2008 | 1590 (1578-1645)-90 (2490-96) | RDP 3.19E-33<br>GENECONV 1.69E-25<br>BootScan 2.39E-31<br>MaxChi 2.05E-16<br>Chimera 3.42E-15<br>SiScan 9.38E-25<br>3Seq 1.42E-24 | MG781054, *Sus scrofa*, Thailand, 92.7%, M1 | AB779632, *Sus scrofa*, Thailand, 2009, 99.9%, M1 |
| VP3 | MG181592, *Homo sapiens*, Malawi, 2013 | 2334 (2318-2349)-196 (150-204) | RDP 1.09E-32<br>GENECONV 6.85E-18<br>BootScan 5.72E-27<br>MaxChi 6.16E-12<br>Chimera 9.20E-12<br>SiScan 3.92E-10 | KC442894, *Homo sapiens*, USA, 2008, 96.4%, M2 | KT919912, *Homo sapiens*, USA, 2013, 99%, M1 |

| | | | | | |
|---|---|---|---|---|---|
| | | | 3Seq 3.54E-27 | | |
| VP3 | JN013991, *Homo sapiens*, South Africa, 2008 | 82 (2550-99)- 260 (253-268) | RDP 1.69E-29<br>GENECONV 8.98 E-27<br>BootScan 1.24E-17<br>MaxChi 9.47E-06<br>Chimera 1.58E-05<br>SiScan 1.76E-02<br>3Seq 5.67E-09 | KJ752816, *Homo sapiens*, South Africa, 2011, 97.4%, M1 | Kp883027, *Homo sapiens*, Mali, 2008, 99.4%, M2 |
| VP3 | KU356576, *Homo sapiens*, Bangladesh, 2010 | 844 (815-855)- 1194 (1142-1219) | RDP 4.55E-24<br>GENECONV 1.69E-21<br>BootScan 3.82E-20<br>MaxChi 1.89E-09<br>Chimera 1.77E-09<br>SiSeq 1.18E-07<br>3Seq 2.83E-09 | KC178787, *Homo sapiens*, Italy, 2008, 97.4%, M2 | KP882081, *Homo sapiens*, Bangladesh, 2007, 100%, M2 |
| VP3 | JQ0692727, *Homo sapiens*, Canada, 2007 | 51 (2482-59)- 845 (817-932) | RDP 4.27E-23<br>GENECONV 9.70E-19<br>BootScan 6.96E-18<br>MaxChi 8.76E-14<br>Chimera 4.71E-14<br>SiScan 2.58E-17<br>3Seq 3.98E-35 | JQ069748, *Homo sapiens*, Canada, 2008, 99.1%, M1 | JQ069762, *Homo sapiens*, Canada, 2008, 99.7%, M1 |
| Vp3 | JQ069754, *Homo sapiens*, Canada, 2008 | 835 (811-879)- 1621 (1591-1631) | RDP 6.24E-20<br>GENECONV 4.13E-17<br>BootScan 4.05E-17<br>MaxChi 1.69E-09<br>Chimera 1.31E-09<br>SiScan 1.91E-10<br>3Seq 1.71E-27 | KJ627117, *Homo sapiens,* Paraguay, 1999, 99%, M1 | KP645258, *Homo sapiens*, Australia, 2010, 99.7%, M1 |
| VP3 | MG181592, *Homo sapiens*, Malawi, 2013 | 1736 (1712-1755)- 1941 (1889-1953) | RDP 2.97E-19<br>GENECONV 7.80E-18<br>BootScan 5.65E-15<br>MaxChi 5.47E-04<br>Chimera 5.27E-04<br>SiSeq 9.43E-07<br>3Seq 2.83E-09 | KC178787, *Homo sapiens,* Italy, 2008, 97%, M2 | LC374087, *Homo sapiens,* Nepal, 2009, 100%, M1 |
| VP3 | KJ753665, *Homo sapiens,* South Africa, 2004 | 2158 (2127-2201)- 2517 (2447-64) | Flagged in ~307, KA p-value= 2.23E-18, Global p-value 5.69E-11 | KJ753481, *Homo sapiens*, Ethiopia, 2012, 94.9%<br>M1 | KY055418, *Sus scrofa,* Uganda, 2016, 98.6%<br>M1 |
| VP3 | AB374145, *Bos taurus*, Japan, Unknown | 810 (765-875)- 1544 (1486-1639) | RDP 1.85E-08<br>GENECONV 1.83E-04<br>BootScan 3.64E-05<br>MaxChi 1.68E-06<br>Chimera 3.72E-06<br>SiScan 2.14E-07<br>3Seq 4.41E-10 | LC340015, *Homo sapiens*, Japan, 2014, 95.5%, M2 | KJ411434, *Homo sapiens,* USA, 2012, 97.4%, M2 |

| | | | | | |
|---|---|---|---|---|---|
| VP3 | KJ753665, *Homo sapiens,* South Africa, 2004 | 1212 (1140-1312)-1732 (1632-1801) | RDP 3.84E-07<br>GENECONV 3.50E-05<br>BootScan 1.92E-05<br>MaxChi 2.67E-05<br>Chimera 1.59E-05<br>SiScan 3.24E-09<br>3Seq 2.08E-04 | HQ392255, *Homo sapiens*, Belgium, 2008, 96.2%, M1 | KX778614, *Homo sapiens*, Dominican Republic of the Congo, 2013, 99.4%, M1 |
| VP3 | KJ753665, *Homo sapiens*, South Africa, 2004 | 580 (498-612)- 922 (831-964) | GENECONV 7.32E-11<br>BootScan 1.64E-09<br>MaxChi 3.90E-05<br>Chimera 3.85E-06<br>SiScan 9.81E-09<br>3Seq 7.82E-20 | KP752825, *Homo sapiens,* South Africa, 2002, 99.6%, M1 | JQ069760, *Homo sapiens*, Canada, 2008, 99.7%, M1 |
| VP3 | KX655453, *Homo sapiens*, Uganda, 2013 | 1258 (872-1389)-1798 (1662-1917) | Flagged as many sequences, Global p-value= 1.78E-12 | KP882576, *Homo sapiens*, Ghana, 2009, 94.9%, M2 | KU714446, *Homo sapiens*, Malawi, 2000, 96.7%, M2 |
| VP4 | KX655509, *Homo sapiens*, Uganda, 2013 | 321 (308-323)- 683 (616-634)<br>VP8* domain | RDP 2.16E-54<br>GENECONV 1.17E052<br>BootScan 1.59E-51<br>MaxChi 7.47E-19<br>Chimera 1.44E-18<br>SiScan 8.23E-23<br>3Seq 2.07E-09 | KP902549, *Homo sapiens,* South Africa, 2003, 99.3%, P8 | KJ52687, *Homo sapiens*, South Africa, 2000, 99.3%, P8 |
| VP4 | KX655487, *Homo sapiens*, Uganda, 2013 | 334(326-347)-682(670-698)<br>VP8* domain | RDP 1.06E-55<br>GENECONV 5.94E-51<br>BootScan 1.77E-50<br>MaxChi 4.55E-19<br>Chimera 5.96E-19<br>SiScan 1.60E-17<br>3Seq 2.07E-09 | KX646644, H*omo sapiens*, India, 2012, 96.6%, P6 | MG181340, *Homo sapiens*, Malawi, 2002, 99.7%, P8 |
| VP4 | MG181725, *Homo sapiens*, Malawi, 2014 | 374 (367-382)- 542 (526-553)<br>VP8* domain | GENECONV 5.45E-46<br>BootScan 2.45E-41<br>MaxChi 4.02E-11<br>Chimera 3.97E-11<br>SiScan 2.83E-13<br>3Seq 2.07E-09 | MG181758, *Homo sapiens*, Malawi, 2014, 99.2%, P8 | MG181890, *Homo sapiens*, Malawi, 2012, 100%, P4 |
| VP4 | KU714447, *Homo sapiens*, Malawi, 2000 | 1380 (1378-1381)-1562 (1543-1571) | RDP 1.47E-39<br>GENECONV 4.07E-38<br>BootScan 104E-29<br>MaxChi 1.24E-07<br>Chimera 3.25E-07<br>SiScan 7.45E-12<br>3Seq 4.14E-09 | GQ869840, *Homo sapiens*, Malawi, 2000, P8 | KU714458, *Homo sapiens,* Malawi, 2000, 99.5%, P14 |
| VP4 | KX655441, *Homo sapiens*, Uganda, 2013 | 374 (357-382)-484 (475-488)<br>VP8* domain | RDP 5.41E-34<br>GENECONV 1.32E-31<br>BootScan2.38E-30<br>MaxChi 1.11E-05<br>Chimera 6.90E-06 | LC260226, *Homo sapiens*, Indonesia, 99.2%, P6 | JQ069692, *Homo sapiens*, Canada, 2009, 99.1%, P8 |

| | | | | | |
|---|---|---|---|---|---|
| | | | SiScan 4.86E-10<br>3Seq 2.07E-09 | | |
| VP4 | MG181637, *Homo sapiens,* Malawi, 2013 | 350 (325-368)-459 (464-466)<br>VP8* domain | GENECONV 5.13E-28<br>BootScan 5.75E-15<br>MaxChi 2.13E-04<br>Chimera 1.00E-04<br>SiScan 4.71E-05<br>3Seq 4.14E-09 | MG181758, *Homo sapiens*, Malawi, 2014, 99.5%, P8 | KJ870892, *Homo sapiens*, Democratic Republic of the Congo, 95.5%, P6 |
| VP4 | JQ069674, *Homo sapiens,* Canada, 2008 | 1568 (1541-1574)-2329 (2251-101) | RDP 1.59E-23<br>GENECONV 2.53E-21<br>BootScan 1.59E-23<br>MaxChi 2.96E-12<br>Chimera 2.13E-12<br>SiScan 7.86E-14<br>3Seq 1.44E-35 | HQ392253, *Homo sapiens*, Belgium, 2008, 99.5%, P8 | HM773714, *Homo sapiens*, USA, 2008, 99.5%, P8 |
| VP4 | KC579762, *Homo sapiens*, USA, 1991 | 1073 (1039-1124)-2332 (2269-8) | RDP 4.67E-23<br>GENECONV 4.72E-17<br>BootScan 8.46E-23<br>MaxChi 1.99E-14<br>Chimera 1.76E-11<br>SiScan 7.07E-25<br>3Seq 4.04E-40 | HQ392253, *Homo sapiens*, Belgium, 2008, 98.7%, P8 | FJ947895, *Homo sapiens*, USA, 1991, 99.9%, P8 |
| VP4 | MG181659, *Homo sapiens*, Malawi, 2013 | 1116 (1064-1146)-1462 (1402-1501) | RDP 6.53E-11<br>GENECONV 2.20E-09<br>BootScan 1.08E-09<br>MaxChi 1.781E-05<br>Chimera 1.74E-05<br>SiScan 5.94E-06 | MG181725, *Homo sapiens*, Malawi, 2013, 98.1%, P8 | KJ752208, *Homo sapiens*, South Africa, 2012, 98.3%, P4 |
| VP4 | JF490554, *Homo sapiens*, Australia, 2005 | 462 (397-482)- 683 (637-722) | RDP 2.17E-10<br>GENECONV 8.75E-09<br>BootScan 1.89E-10<br>MaxChi 3.26E-04<br>Chimera 1.04E-03<br>SiScan 1.14E-04<br>3Seq 4.14E-09 | KJ751716, *Homo sapiens,* Gambia, 2010, 98.4%, P8 | JQ069674, *Homo sapiens*, Canada, 2008, 99.5%, P8 |
| VP4 | AB008291, *Homo sapiens,* 1995 | 2291 (2223-71)-807 (762-849) | RDP 1.02E-03<br>BootScan 4.42E-02<br>MaxChi 5.00E-05<br>Chimera 1.32E-04<br>SiScan 1.51E-05<br>3Seq 1.66E-08 | AB039939, *Homo sapiens*, Japan, 1990, 98.6%, P8 | AB008277, *Homo sapiens*, China, 1992, 98.9%, P8 |
| VP6 | KJ482497, *Sus scrofa*, Brazil, 2013 | 589 (584-605) – 801 (790 – 809) | RDP 1.81E-25<br>GENECOV 7.17E-23<br>Bootscan 7.09E018<br>Maxchi 2.98E-10<br>SiScan 2.41E-11<br>3Seq 1.41E-09 | DQ119822, *Sus scrofa,* China, 98.1%, I2 | KJ482491, *Sus scrofa*, Brazil, 2013, 100%, I5 |

| | | | | | |
|---|---|---|---|---|---|
| VP6 | KX632346, *Homo sapiens*, Uganda, 2013 | 1269 (1250-1272)-56 (1333-103) | RDP 7.77E-19<br>GENECONV 1.33E-17<br>Bootscan 7.27E-09<br>Maxchi 2.1E-02<br>Chimera 3.24E-03<br>SiScan 1.02E-07<br>3Seq 3.06E-11 | KU714448, *Homo sapiens*, Malawi, 2000, 92.7%, I1 | KX655455, *Homo sapiens*, Uganda, 2013, 96.5%, I2 |
| VP6 | JN014005, *Homo sapiens*, South Africa, 2008 | 160 (33-220_- 384 (379-451) | RDP 1.98E-18<br>GENECONV 6.96E-18<br>BootScan 4.63E-16<br>MaxChi 1.00E-06<br>Chimera 2.77E-07<br>3Seq 1.41E-09 | JN706549, *Homo sapiens*, Thailand, 2010, 97.8%, I1 | KJ752622, *Homo sapiens, Senegal*, 2009, 96.4%, I2 |
| VP6 | JX040423, *Homo sapiens*, India, 2009 | 1177(1143-1181)-160(140-176) | RDP 1.30E-15<br>GENECONV 4.97E-09<br>BootScan 1.95E-11<br>MaxChi 4.38E-11<br>Chimera 9.18E-12<br>SiScan 7.20E-15<br>3Seq 2.82E-09 | KF740531, *Homo sapiens*, India, 2009, I12 | Unknown, closest =KC579918, *Homo sapiens*, USA, 1976), I1 |
| VP6 | JN014003, *Homo sapiens,* South Africa, 2008 | 202 (1333-208)-384(354-421) | RDP 6.30E-15<br>GENECONV 2.39E-15<br>BootScan 1.04E-12<br>MaxChi 3.76E-05<br>Chimera 3.02E-05<br>SiScan 4.51E-05<br>3Seq 1.41E-09 | KP752725, *Homo sapiens*, Swaziland, 2010, 95.5%, I2 | JX027952, *Homo sapiens*, Australia, 2010, 96.7%, I1 |
| VP6 | KU714448, *Homo sapiens*, Malawi, 2000 | 653(635- 664)-831(818-846) | RDP 1.24E-12<br>GENECONV 7.55E-10<br>BootScan 5.50E-07<br>MaxChi 1.15E-05<br>Chimera 3.38E-05<br>SiScan 1.38E-04<br>3Seq 1.41E-09 | JN014005, *Homo sapiens*, South Africa, 95.9%, I1 | KX655455, *Homo sapiens*, Uganda, 2013, 99.4%, I2 |
| VP6 | KU714448, *Homo sapiens*, Malawi, 2000 | 317 (290-333)- 491 (474-507) | RDP 3.94E-15<br>GENECONV 9.37E-14<br>BootScan 2.42E-13<br>MaxChi 2.80E<br>Chimera 2.58E-05<br>SiScan 1.73E-07<br>3Seq 1.41E-09 | KX638567, *Homo sapiens*, India, 2010, 96.6%, I1 | GU984758, *Bos taurus*, India, 2007, 98.9%, I2 |
| VP6 | HQ11009, *Homo sapiens*, Belgium, 2007, 98.5% | 754 (697-775)-1357 (1329-32) | GENECONV 2.23E-05<br>BootScan 1.61E-03<br>MaxChi 3.32E-11<br>Chimera 1.08E-10<br>SiScan 5.16E-14<br>3Seq 2.54E-19 | MF469405, *Homo sapiens*, USA, 2015, 99.3%, I1 | HQ392161, *Homo sapiens*, Belgium, 2007, 98.5%, I1 |

| | | | | | |
|---|---|---|---|---|---|
| VP6 | FJ685614, *Homo sapiens*, India, 1992 | 698(672-712)-1357 (1231-70) | RDP 1.72E-08<br>GENECONV 1.61E-07<br>BootScan 2.85E-08<br>MaxChi 1.40E-09<br>Chimera 2.14E-09<br>SiScan 4.04E-11<br>3Seq 1.54E-23 | KX638599, *Homo sapiens*, India, 2010, 99.4% ,I1 | KX632291, *Homo sapiens*, Uganda, 2013, 99.1%, I1 |
| VP6 | KJ753780, *Homo sapiens*, South Africa, 2004 | 513 (463-560)-1008 (952-1031) | GENECONV 1.89E-05<br>BootScan 2.60E-05<br>MaxChi 9.01E-09<br>Chimera 1.46E-08<br>SiScan 1.38E-09<br>3Seq 1.30E-19 | JN258839, *Homo sapiens*, Belgium, 2001, 99.6%, I1 | LC056797, *Homo sapiens*, Viet Nam, 2007, 98.8%, I1 |
| VP6 | HM988972, *Bos taurus*, South Korea, Unknown | 751 (340-811)-1113 (1080-1132) | RDP 2.08E-08<br>GENECONV 1.92E-08<br>BootScan 2.91E-08<br>MaxChi 1.45E-05<br>Chimera 2.41E-05<br>SiScan 3.56E-05<br>3Seq 5.55E-05 | Unknown, closest = MH238302, *Sus scrofa*, Spain, 2017, I5 | MH423866, *Sus scrofa*, China, 2016, 99.4% ,I5 |
| VP7 | AB158431, *Bos taurus,* Japan, 2000 | 901 (868-907)-1051 (1037-80) | RDP 3.71E-16<br>GENECONV 1.25E-13<br>MaxChi 6.99E-07<br>Chimera 3.69E-05<br>SiScan 8.07E-10<br>3Seq 4.95E-10 | AB077053, *Bos taurus*, Japan, 1995-1996, 95.9%, G8 | U50332, *Bos taurus,* USA, 1991, 97.2%, G6 |
| VP7 | HM591496, *Bos taurus*, India, 2009<br>KF170899, *Bos taurus*, India, 2012 | 1048 (1032-128)-481 (448-506) | RDP 2.78E-05<br>GENECONV 4.77 E-04<br>BootScan 4.86E-04<br>MaxChi 1.02E-03<br>Chimera 3.78E-05<br>SiSeq 7.21E-09<br>3Seq 3.90E-08 | EF199501, *Bos taurus,* India, 96.6%, G6 | JX442784, *Bos taurus*, India, 2010, 100%, G6 |
| VP7 | D86274, *Homo sapiens,* 1990 | 353 (338-368)- 745 (730-756) | RDP 1.73E-18<br>GENECONV 1.43E-16<br>BootScan 2.50E-14<br>MaxChi 3.26E-13<br>Chimera 6.54E-13<br>SiScan 2.47E-10<br>3Seq 3.55E-33 | D86282, *Homo sapiens,* 1981, 98.7%, G3 | Kf636217, *Homo sapiens*, South Africa, 2010, 94.9%, G1 |
| VP7 | MG181727, *Homo sapiens*, Malawi, 2014 | 316 (1010-314)-470 (437-476) | RDP 3.69E-17<br>GENECONV 1.84E-16<br>BootScan 8.96E-12<br>MaxChi 2.23E-08<br>Chimera 3.80E-08<br>SiScan 2.65E-12<br>3Seq 2.69E-23 | JQ069508, *Homo sapiens*, Canada, 2008, 98%, G1 | KJ753362, *Homo sapiens*, South Africa, 2003, 98.7%, G2 |

| | | | | | |
|---|---|---|---|---|---|
| VP7 | KX655511, *Homo sapiens*, Uganda, 2013 | 840 (823-851)-1052 (1037-79) | RDP 1.20E-16<br>GENECONV 3.11E-15<br>BootScan 2.21E-16<br>MaxChi 6.12E-05<br>Chimera 6.98E-06<br>SiScan 2.52E-04<br>3Seq 3.36E-09 | EU839936, *Homo sapiens*, Bangladesh, 2005, 97.4%, G9 | KX655456, *Homo sapiens*, Uganda, 2013, 100%, G8 |
| VP7 | KJ753024, *Homo sapiens*, South Africa, 2009 | 1013 (985-59)-130(121-143) | RDP 1.53E-16<br>GENECONV 2.32E-14<br>BootScan 1.27E-10<br>MaxChi 9.65E-04<br>Chimera 5.96E-04<br>SiScan 8.42E-13<br>3Seq 2.38E-16 | KM008672, *Homo sapiens*, India, 2010, 97.8%, G12 | D16344, *Homo sapiens*, 96.5%, G1 |
| VP7 | KJ751729, *Homo sapiens*, Ethiopia, 2010<br>KP752817, *Homo sapiens*, Togo, 2010 | 857 (833-859)-1019 (991-51) | RDP 4.0E-15<br>GENECONV 1.60E-10<br>BootScan 1.21EE-11<br>MaxChi 1.12E-03<br>Chimera 6.78E-04<br>SiScan 1.73E-03<br>3Seq 1.34E-05 | KJ560447, *Homo sapiens*, USA, 2011, 99%, G3 | KP882703, *Homo sapiens*, Kenya, 99.3%, G1 |
| VP7 | MG181661, *Homo sapiens,* Malawi, 2013 | 590 (582-616)- 175 (170-181) | RDP 7.82E-30<br>GENECONV 7.33E-23<br>BootScan 5.06E-31<br>MaxChi 3.49E-21<br>Chimera 1.06E-25<br>3Seq 2.28E-47 | JQ069502, *Homo sapiens,* Canada, 2008, 96.6%, G1 | MG181859, *Homo sapiens*, Malawi, 2012, 99.5%, G2 |
| VP7 | MG181661, *Homo sapiens*, Malawi, 2013 | 590 (582-616)- 175 (170-181) | RDP 7.82E-30<br>GENECONV 7.33E-23<br>BootScan 5.06E-31<br>MaxChi 3.49E-21<br>Chimera 1.06E-25<br>3Seq 2.28E-47 | JQ069502, *Homo sapiens*, Canada, 2008, 96.6%, G1 | MG181859, *Homo sapiens*, Malawi, 2012, 99.5%, G2 |
| VP7 | KF170899, *Bos taurus,* India, 2010 | (71) – (448-506) | RDP 2.78 E-05<br>GENEECONV 4.77 E-04<br>BootScan 4.86 E-04<br>MaxChi 1.02 E-03<br>Chimaera 3.78 E-05<br>SiScan 7.21 E-09<br>3Seq 3.90 E-08 | EF199501, *Bos taurus,* India, 97.3%, G6 | JX442784, *Bos taurus*, India, 99.5%, G6 |
| VP7 | KJ412857, *Homo sapiens,* Paraguay, 2009 | 772 (754-779)-1014(988-51) | GENECONV 1.23E-19<br>BootScan 6.33E-20<br>MaxChi 2.56E-07<br>Chimera 4.38E-07<br>SiScan 2.01E-09<br>3Seq 3.36E-09 | KJ412858, *Homo sapiens,* Paraguay, 2009, 100%<br><br>**G3** | KJ412802, *Homo sapiens*, Paraguay, 2009, 99.2%<br><br>G1 |

| | | | | | |
|---|---|---|---|---|---|
| VP7 | KC443034, *Homo sapiens,* USA, 2006<br>MG181727, *Homo sapiens,* Malawi, 2014 | 55 (994-60)-291 (262-296) | RDP 1.66 E-30<br>GENECONV 5.07E-26<br>BootScan 2.32E-24<br>MaxChi 1.03E32<br>Chimaera 1.10E-17<br>3Seq 3.36E-09 | GQ229044, *Homo sapiens*, India, 97.4%, G2 | JX411970, *Homo sapiens*, India, 2011, 98.7%, G1 |
| VP7 | MG181771, Homo sapiens, Malawi, 2012 | 715 (707-722) – 950 (921-971) | RDP 2.65E-29<br>GENECONV 6.45E-26<br>BootScan 5.84E-17<br>MaxChi 1.70E-11<br>Chimaera 2.36E-11<br>SiScan 4.78E-11<br>3Seq 3.55E-09 | KU356667, *Homo sapiens*, Bangladesh, 2013, 94.9%<br>G2 | KP222809, *Homo sapiens*, Mozambique, 2011, 99.6%<br>G1 |
| VP7 | KP752498, Homo sapiens, Togo, 2009 | 618 (605-641)-1018 (1000-1020) | RDP 1.47E-28<br>GENECONV 4.49E-23<br>BootScan 2.78E-14<br>MaxChi 2.05E-14<br>Chimaera 7.10E-19<br>SiScan 6.78E-39 | D86266, Homo sapiens, 95.7%<br>G3 | MG181661, Homo sapiens, Malawi, 2013, 98.8%<br>G1 |
| VP7 | AY261339, *Homo sapiens*, South Africa | 783 (774-791)-14(1044-18) | RDP 3.82E-28<br>GENECONV 2.35E-25<br>BootScan 2.41E-27<br>MaxChi 5.44E-10<br>Chimera 2.33E-10<br>SiScan 3.15E-14<br>3Seq 3.36E-09 | GU567778, *Homo sapiens*, Russia, 96.9%<br>G2 | JX458964, *Homo sapiens*, Argentina, 2001, 100%<br>G1 |
| VP7 | KX655489, *Homo sapiens*, Uganda, 2013 | 151 (132-155)- 316 (311-325) | RDP 2.14E-27<br>GENECONV 1.27E024<br>BootScan 3.15E-19<br>MaxChi 7.75E-09<br>Chimera 7.42E-09<br>SiScan 8.44E-20<br>3Seq 4.13E-03 | DQ146676, *Homo sapiens*, Bangladesh, 2003, 98.1%, G12 | JN232074, *Homo sapiens,* Brazil, 2004, 99.4%, G1 |
| VP7 | D86273, *Homo sapiens*, 1986 | 347 (339-353)- 726 (697-763) | RDP 1.70E-24<br>GENECONV 1.57E-22<br>BootScan 1.26E-17<br>MaxChi 7.23E-14<br>Chimera 1.37E-15<br>SiScan 1.73E-13<br>3Seq 6.15E-47 | AF450293, *Homo sapiens*, China, Unknown, 97.4%, G3 | D50121, *Homo sapiens*, 96.3%, Japan, 1992-1993, G2 |
| VP7 | Kj4122858, *Homo sapiens*, Paraguay, 2009 | 881 (859-914)-1016 (991-54) | RDP 4.80E-14<br>GENECONV 3.04E-13<br>BootScan 3.81E-13<br>MaxChi 5.36E-03<br>Chimera 1.05E-02 | AB585923, *Homo sapiens*, Hong Kong, 2006, 98.5%, G3 | HQ425288, *Homo sapiens*, South Korea, 2004, 99.2%, G4 |

| | | | | | |
|---|---|---|---|---|---|
| | | | SiScan 1.36E-04<br>3Seq 3.36E-09 | | |
| VP7 | KU714449, *Homo sapiens*, Malawi, 2000 | 825 (795-835)- 93 (85-93) | RDP 1.04E-13<br>GENECONV 4.64E-09<br>BootScan 1.71E-09<br>MaxChi 1.20E-10<br>Chimera 9.96E-11<br>SiScan 3.89E-10<br>3Seq 7.72E-08 | KM008635, *Homo sapiens*, India, 2009, 92.5%, G1 | KU714460, *Homo sapiens,* Malawi, 2000, 98.2%, G6 |
| VP7 | KJ412888, *Homo sapiens*, Paraguay, 2009 | 873 (823-890)- 1023 (1000-81) | RDP 2.17E-12<br>GENECONV 7.16E-11<br>BootScan 3.63E-12<br>MaxChi 5.48E-04<br>Chimera 4.30E-04<br>SiScan 8.10E-03<br>3Seq 1.12E-05 | D86275, *Homo sapiens*, 88%, G3 | JF490143, *Homo sapiens*, Australia, 2004, 99.3%, G1 |
| VP7 | KX655445, *Homo sapiens*, Uganda, 2013 | 1042 (1006-110)- 274 (269-287) | RDP 6.32E-12<br>GENECONV 5.98E-10<br>BootScan 8.41E-08<br>MaxChi 2.20E-06<br>Chimera 3.68E-08<br>SiScan 3.20E-04<br>3Seq 8.29E-11 | KJ560447, *Homo sapiens,* USA, 2011, 95.8%, G3 | KJ753326, *Homo sapiens*, Kenya, 94.9%, G8 |
| VP7 | D86277, *Homo sapiens,* China, 1989 | 335 (323-353)- 754 (710-762) | RDP 8.20E-20<br>GENECONV 2.21E-15<br>BootScan 1.81E-10<br>MaxChi 1.91E-12<br>Chimera 1.95E-13<br>SiScan 3.36E-13<br>3Seq 7.22E-27 | D86279, *Homo sapiens,* China, 1992, 89.6%, G3 | KP883055, *Homo sapiens,* Mali, 2009, G2 |
| VP7 | AF281044, *Homo sapiens,* Ireland, 1997-1999<br>GQ433992, *Homo sapiens,* Ireland, Unknown | 362 (346-366)- 589 (558-605) | RDP 1.32E-09<br>GENECONV 1.25E-02<br>BootScan 4.96E-05<br>MaxChi 2.75E-05<br>Chimera 1.50E-04<br>3Seq 3.36E-09 | AY866503, *Homo sapiens*, Thailand, 1995-1997, 98.9%, G9 | KM008633, *Homo sapiens,* India, 2009, 89%, G1 |
| VP7 | KU714449, *Homo sapiens*, Malawi, 2000 | 440 (421-446)- 604(641-704) | RDP 3.12E-10<br>GENECONV 7.36E-11<br>BootScan 1.56E-07<br>MaxChi 1.83E-03<br>Chimera 1.69E-03<br>3Seq 1.08E-12 | JQ069503, *Homo sapiens*, Malawi, 2000, 100%, G1 | KU714460, *Homo sapiens*, Malawi, 2000, 100%, G6 |
| VP7 | KX655444, *Homo sapiens*, Uganda, 2013 | 110 (88-114)- 191(183-197) | RDP 2.84E-08<br>GENECONV 3.00E-05<br>MaxChi3.11E-02<br>Chimera 3.11E-02<br>SiScan 1.01E-03<br>3Seq 8.25E-05 | KC579965, *Homo sapiens*, USA, 96.2%, G3 | AF254137, *Homo sapiens*, 93.9% ,G1 |

| | | | | | |
|---|---|---|---|---|---|
| NSP1 | KP753174, *Sus scrofa,* South Africa, 2008, KJ753184, *Sus scrofa,* South Africa, 2008, KP752951, *Sus scrofa*, South Africa, 2008 | 1498 (1435-25)-585(573-596) | RDP 7.45E-28<br>GENECONV 5.26E-27<br>BootScan 7.87E-25<br>MaxChi 8.49E-16<br>Chimera 5.73E-16<br>SiScan 9.95E-18<br>3Seq 1.09E-48 | KP752999, *Sus scrofa*, South Africa, 2009, 99.9%, A8 | KP753000, *Sus scrofa*, South Africa, 2009, 99.7%, A8 |
| NSP1 | KX655446, *Homo sapiens*, Uganda, 2013 | 485 (455-499)- 889 (868-914) | RDP 3.98E-25<br>GENECONV 2.55E-22<br>BootScan 2.53E-21<br>MaxChi 1.96E-14<br>Chimera 8.97E-15<br>SiScan 1.22E-19<br>3Seq 3.18E-09 | LC406825, *Homo sapiens*, Kenya, 2012, 95.9%, A2 | MG181277, *Homo sapiens*, Malawi, 2000, 96.5%, A1 |
| NSP1 | DQ199658, *Homo sapiens*, Australia, 2001 | 485 (440-517)- 942 (931-956) | RDP 2.25E-22<br>GENECONV 7.04E-20<br>BootScan 1.52E-21<br>MaxChi 2.09E-12<br>Chimera 1.10E-12<br>SiScan 6.14E-12<br>3Seq 2.61E-35 | KP883160, *Homo sapiens*, Malawi, 2008, 98.1%, A1 | DQ199657, *Homo sapiens,* Australia, 1997, 98.7%, A1 |
| NSP1 | KJ482257, *Sus scrofa*, Brazil, 2013 | 1534 (1490-44)-888 (853-909) | RDP 3.24E-20<br>GENECONV 2.11E-19<br>BootScan 5.21E-10<br>MaxChi 9.52E-18<br>SiScan 2.81E-21<br>3Seq 1.81E-29 | KJ482252, *Sus scrofa*, Brazil, 2013, 92.9%, A8 | KJ482255, *Sus scrofa*, Brazil, 2013, 99.9%, A8 |
| NSP1 | KX778616, *Homo sapiens*, Dominican Republic, 2013 | 1147 (1127-1155)-1462 (1433-5) | RDP 1.14E-19<br>GENECONV 4.51E-17<br>BootScan 8.47E-16<br>MaxChi 1.54E-11<br>Chimera 1.93E-12<br>SiScan 9.28E-13<br>3Seq 3.18E-09 | KJ752290, *Homo sapiens*, Zambia, 2009, 96.3%, A1 | KX363351, *Sus scrofa*, Vietnam, 2012, 95.9%, A8 |
| NSP1 | KJ482254, *Sus scrofa*, Brazil, 2013 | 534 (499-559)- 886 (849-911) | RDP 7.72E-18<br>GENECONV 1.84E-16<br>BootScan 3.89E-14<br>MaxChi 2.53E-10<br>Chimera 2.27E-10<br>SiScan 1.55E-08<br>3Seq 8.01E-16 | KJ482247*, Sus scrofa*, Brazil, 2013, 100%, A8 | KJ482256, *Sus scrofa*, Brazil, 2013, 100%, A8 |
| NSP1 | KX632348, *Homo sapiens*, Uganda, 2013 | 580 (508-605)- 844 (829-874) | RDP 1.91E-10<br>GENECONV 3.85E-10<br>BootScan 1.02E-06<br>MaxChi 2.32E-04<br>Chimera 2.29E-04<br>SiScan 6.75E-08<br>3Seq 3.18E-09 | KP752785, *Homo sapiens*, Ethiopia, 2010, 94.9%, A1 | KP882434, *Homo sapiens,* Ghana, 2009, 92.8%, A2 |

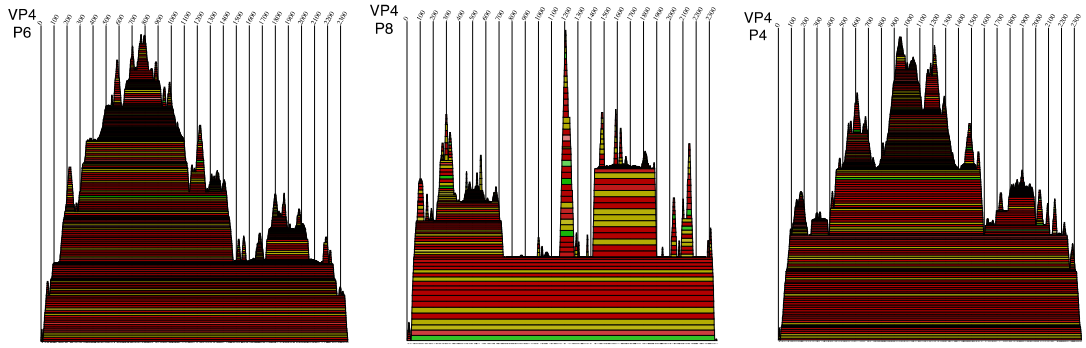
| | | | | | |
|---|---|---|---|---|---|
| NSP1 | KP752999, *Sus scrofa*, South Africa, 2009 | 1394 (1368-31)-601 (552-631) | RDP 1.38E-06<br>BootScan 5.20E-05<br>MaxChi 2.04E-08<br>Chimera 2.74E-08<br>SiScan 3.04E-06<br>3Seq 2.78E-12 | KF835942, *Homo sapiens,* Hungary, 2005, 93.9%, A8 | Unknown (close to KP753000 *Sus scrofa*, South Africa, 2009), A8 |
| NSP1 | JX040426, *Homo sapiens*, India, 2009 | 662 (559-677)- 842 (783-884) | RDP 8.72E-03<br>GENECONV 1.35E-05<br>BootScan 4.19E-02<br>MaxChi 1.57E-04<br>Chimera 5.50E-05<br>SiScan 4.50E-03<br>3Seq 6.62E-07 | KU292525, *Homo sapiens,* India, 2014, 98.4%, A11 | KP882478, *Homo sapiens*, Ghana, 2008, 95%, A11 |
| NSP1 | KP753000, *Sus scrofa*, South Africa, 2009 | 1389 (1359-41)-602 (504-631) | RDP 4.27E-03<br>BootScan 4.30-03<br>MaxChi 1.12E-04<br>Chimera 1.52E-05<br>SiScan 3.90E-04<br>3Seq 5.58E-05 | JQ309141, *Equus callabus,* United Kingdom, 1975, 87.2%, A8 | KF835942, *Homo sapiens*, Hungary, 2005, 92.8%, A8 |
| NSP1 | KJ482254, *Sus scrofa,* Brazil, 2013 | 1534 (1461-44) – 533 (499-909) | R ---<br>GENECONV: 1.50E-13<br>BootScan: 5.96E-10<br>MaxChi: 3.77E-08<br>Chimaera: 6.43E-08<br>SiScan: 1.88E-11<br>3Seq: 1.46 E-24 | KJ482247, *Sus scrofa,* Brazil, 2013, 99.1%, A8 | KJ482255, *Sus scrofa,* Brazil, 2013, 100%, A8 |
| NSP2 | KJ753657, *Homo sapiens,* South Africa, 2004<br>KM026663, *Homo sapiens,* Brazil, 2009<br>KM026664, *Homo sapiens,* Brazil, 2009 | (0)- (219-278) | RDP 2.01E-03<br>GENECONV 1.6E-04<br>BootScan 4.07E-06<br>MaxChi 3.51E-03<br>Chimera 3.30E-03<br>SiScan 1.90E-07<br>3Seq 1.57E-04 | KM026633, *Homo sapiens*, Brazil, 2003, 98.6%, N1 | KJ918915, *Homo sapiens*, Hungary, 2012, 99.5%, N1 |
| NSP2 | MF18783, *Homo sapiens*, USA, 2011 | 1015 (1002-57)-286 (273-297) | GENECONV 1.60E-14<br>BootScan 1.55E-14<br>MaxChi 6.46E-04<br>Chimera 3.74E-06<br>SiScan 4.37E-08<br>3Seq 5.75E-22 | KJ752146, *Homo sapiens*, South Africa, 2011, 99.7%, N1 | KX655480, *Homo sapiens,* Uganda, 2012, 99.3%, N2 |
| NSP2 | MF18783, *Homo sapiens*, USA, 2011 | 1015 (1002-57)-286 (273-297) | GENECONV 1.60E-14<br>BootScan 1.55E-14<br>MaxChi 6.46E-04<br>Chimera 3.74E-06<br>SiScan 4.37E-08<br>3Seq 5.75E-22 | KJ752146, *Homo sapiens*, South Africa, 2011, 99.7%<br>N1 | KX655480, *Homo sapiens*, Uganda, 2012, 99.3%<br>N2 |

| | | | | | |
|---|---|---|---|---|---|
| NSP2 | AB779641, *Sus scrofa*, Thailand, 2008 | 870 (833-921)- 216 (150-396) | RDP 6.30E-05<br>GENECONV 4.67 E-03<br>BootScan 1.24 E-03<br>MaxChi 1.28 E-03<br>Chimaera 6.20 E-03<br>3Seq 1.48 E-05 | KX363374, *Sus scrofa*, Vietnam, 2012, 96.5%<br>N1 | JF796719, *Sus scrofa*, South Korea, 2006, 100%<br>N1 |
| NSP4 | MF58092, *Homo sapiens,* China, 2013 | 599 (546-616)- (745-94) | RDP 1.64 E-11<br>GENECONV 2.38 E-08<br>BootScan 1.14 E-09<br>MaxChi 1.01 E-09<br>Chimaera 4.58 E-10<br>SiScan 4.77 E-18<br>3Seq 6.43 E-21 | JF813107, *Homo sapiens*, China, 2008, 97.9%<br>E1 | MG781050, *Homo sapiens*, Thailand, 2011, 73.1%<br>E1 |
| NSP4 | AB361290, *Homo sapiens*, India, 2006 | 624 (548-635)- 726 (712-50) | RDP 1.66 E-09<br>GENECONV 4.38 E-08<br>BootScan 6.96 E-07<br>MaxChi 1.96 E-03<br>Chimaera 2.49 E-05<br>3Seq 1.15 E-12 | AB326338, *Homo sapiens,* India, 1989, 95.8%<br>E2 | Unknown (FJ492834, *Sus scrofa*, Ireland)<br>E9 |
| NSP4 | AB326966, *Homo sapiens*, India, 2001 | 605 (564-614)- 675 (670-46) | RDP 2.77 E-09<br>GENECONV 1.10 E-10<br>BootScan 2.26 E -02<br>MaxChi 2.54 E -05<br>Chimaera 1.63 E -03<br>3Seq 8.89 E-09 | FJ685615, *Homo sapiens*, India, 1992, 98.2%<br>E1 | Unknown (MG181929, *Homo sapiens*, Malawi, 2012)<br>E2 |
| NSP5 | M33608, *Homo sapiens, 1993* | 619 (505-624)- 668 (665-182) | RDP 1.44 E -05<br>GENECONV 2.64 E-06<br>BootScan 4.1 E-05<br>MaxChi 4.54 E-02<br>Chimaera 3.70 E-02<br>SiScan 2.49 E-05<br>3Seq 6.61 E-04 | MG996138, *Homo sapiens*, Singapore, 2016, 95%<br>H2 | JQ358774, *Homo sapiens*, India, 2011, 100%<br>H3 |

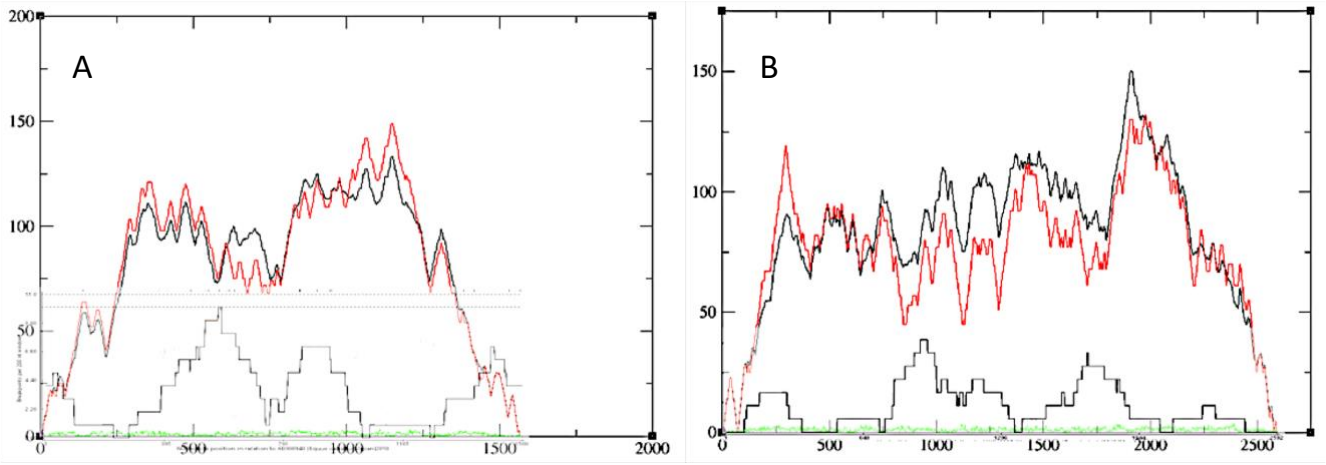
**Supplementary Table 2.** All amino acid changes observed in the segment 9 (VP7) recombinants which had more than one isolate supported the event. If changes were shown to occur in a predicted epitope region, it is noted.

| Gene recombinant | Accession Number | Position | Major | Minor | Epitope region? |
|---|---|---|---|---|---|
| G6-G6 | KF170899 | 40 | M | T | - |
| | | 46 | I | T | - |
| | | 48 | V | T | - |
| | | 116 | V | I | - |
| G3-G1 | KJ751729 | 281 | V | I | Yes |
| | | 303 | V | I | Yes |
| | | 309 | A | V | Yes |
| | | 318 | N | D | |
| G9-G1 | AF281044 | 135 | S | C | - |
| | | 162 | D | E | - |
| | | 163 | R | W | - |
| | | 169 | V | D | |
| | | 170 | N | I | |
| | | 176 | L | Q | |
| | | 180 | G | E | Yes |
| G2-G1 | KC443034 | 9 | I | T | - |
| | | 10 | F | I | - |
| | | 21 | S | Y | - |
| | | 25 | T | S | - |
| | | 26 | I | V | - |
| | | 28 | N | R | - |
| | | 29 | P | I | - |
| | | 31 | A | D | - |
| | | 32 | P | Y | - |
| | | 35 | F | Y | - |
| | | 41 | I | F | - |
| | | 42 | A | V | - |
| | | 43 | L | A | - |
| | | 44 | M | L | - |
| | | 45 | S | F | - |
| | | 47 | F | L | - |
| | | 48 | G | T | - |
| | | 49 | R | K | - |
| | | 50 | T | A | - |
| | | 56 | Y | N | - |
| | | 65 | A | T | - |
| | | 68 | T | S | - |
| | | 72 | S | R | - |

| | | | | | |
|---|---|---|---|---|---|
| | | 73 | G | E | - |
| | | 75 | S | V | - |




**Supplementary Figure 1**. P8, P6, and P4 consensus RNA secondary structure mountain plots. Mountain plot showing consensus secondary structure of VP4 genotypes P8, P6, and P4. Lighter colors indicate lower probabilities of base-pairing. Peaks correspond to hairpin-loops, plateus to loops, and slopes to helices. Plot was generated using RNAalifold in ViennaRNA package.




**Supplementary Figure 2.** Consensus mountain plot with recombination breakpoint distribution plot in grey. A) NSP1 B) VP3