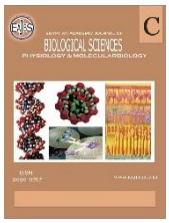
.

Citation: *Egypt.Acad.J.Biolog.Sci. (C.Physiology and Molecular biology) Vol. 16(1) pp175-189 (2024)*
DOI: 10.21608/EAJBSC.2024.340634

Egypt. Acad. J. Biolog. Sci., 16(1):175-189 (2024)

Egyptian Academic Journal of Biological Sciences
C. Physiology & Molecular Biology
ISSN 2090-0767
www.eajbsc.journals.ekb.eg



Exploration of the Pathogenic Potential of Mutations Incurred in VP4 and VP7 Proteins of the Rotavirus Strain from Saudi Arabia Based on A Molecular Evolutionary Model

Mohammad Asrar Izhari
Department of Laboratory Medicine, Faculty of Applied Medical Sciences, Al-Baha University,KSA.
***E-mail:** aazhari@bu.edu.sa

ARTICLE INFO
Article History
Received:29/12/2023
Accepted:10/2/2024
Available:14/2/2024

Keywords:
Rotavirus, mutation, conservation analysis, stability analysis.

ABSTRACT

Background/aims: Rotavirus-associated gastroenteritis is one of the main reasons for morbidity and mortality in Arab populations. The current investigation aimed to analyze distinctive structural mutations of VP4 and VP7 protein sequences from Saudi Arabia and the determination of the phylogenetic relationship using an evolutionary model. **Materials and methods:** Query and reference sequences were mined from a viral database of NCBI. Sequence variation analysis, phylogenetic analysis, intertaxon clade classification, the functional impact of mutation mutational stability analysis, and evolutionary conservation determination were executed by employing multiple sequence alignment version 3.8.31 (MUSCLE: v.3.8.31)/Clustal Omega (V2.1), iterative Tree of Life (iTOL: v5), Predict SNP & MutPred2, I-Mutant version 3.0/support vector machine (SVM)/MUpro, and ConSurf web-server respectively. **Results:** Among thirty-five mutational sites in total, with predicted conservation scores of two sites (I108T, S258A) of VP7 protein and five sites (G145N, I123V, Y102S, D110N, and T195K) of VP4 protein were identified as deleterious. Mutational sites (D110N and S258A) were found to decrease the stability of the VP4 and VP7 proteins. D110N and S258A were found to impact the protein function (by decreasing the stability) as variable mutation positions are subjected to variable protein functions. **Conclusion:** Understanding the pathogenesis relies on structural mutations in protein VP4 and VP7 sequences. Pathogenic potential and stability of 35 variations, and deleterious D110N and S258A variations were ascertained. This could be used to design a vaccine construct for rotavirus serving as a potential treatment. These discoveries are important because the mutation may be able to confer co-protection against all rotavirus.

INTRODUCTION

One of the leading causes of infant/child (age < 5 years) mortality is diarrhea with a global burden of 9% (Parashar *et al.*, 2006) and rotaviral diarrhea constitutes approximately 39% of the global mortality (Goel *et al.*, 2021). Rotavirus (contagious virus) causes dehydrating gastrointestinal abnormalities and takes the fecal-oral route of transmission of the infection (Bishop *et al.* 1973; Parashar *et al.* 2013). In addition to infant diarrhea, one of the major causes of dehydrating severe diarrhea in immunosuppressed individuals is a rotaviral infection (Sue E Crawford *et al.*, 2017).

Several efficacious vaccines are currently available in the market that prevent serious and prolonged rotavirus infection-associated hospitalization (Al-Ayed *et al.*, 2017; Carvalho & Gill 2018; Dennehy 2008; Flewett *et al.*, 1973). This is one of the vital available measures for the prevention of rotavirus infection, however, other preventative measures including routine hand washing, proper hygiene, and avoiding contact with infected people are effective too.

ds RNA genome of rotavirus encodes structural proteins (SP): VP1 through VP4, VP6 (n = 6), and VP7 and non-structural proteins (NSP): NSP1 through NSP6 (n = 6) (S. E. Crawford *et al.*, 2017). The VP4 protein, along with VP1, VP2, and VP3, is the most vital primary rotaviral structural (S) protein (S. E. Crawford *et al.*, 2017). VP4 (outer surface protein) of virion throughout the early phases of virus infection, notably during the process of cell attachment and entrance (Greenberg & Estes 2009). Viral spikes aid in the virus's adsorption and penetration of the host cell membrane which is well mediated by VP4 (Brunet *et al.* 2000; Sen *et al.* 2022). The VP4 protein is important in determining the rotavirus serotype, which describes the numerous strains or variants of the virus too. Each of the nine rotavirus serotypes is distinguished by the specificity of its VP4 and VP7 proteins (Michelangeli *et al.* 1995; Sen *et al.*, 2007). Cleaved VP4 protein makes the virus unstable, however, it renders the virus highly infectious, and neutralizing antibodies (NAbs) to a particular region of this protein can protect against recurring infections with the same serotype of the virus (Crawford *et al.*, 2001; Ludert *et al.*, 1996; Morris *et al.*, 1999; Sánchez-Tacuba *et al.*,2022).

VP7, an outer layer protein, is the key antigenic protein and the most crucial target for protective antibodies (Abs) (Aoki *et al.*, 2009). It is the vital structural protein (proteins of the outer layer) (Blutt *et al.*, 2004; Solberg *et al.*, 2009). The development of the glycoprotein spikes on the virus's outer surface is caused by VP7. These spikes aid the virus in adhering to and entering host cells, coupled with the spikes produced by the VP4 protein (Trask & Dormitzer 2006). The role of VP7 protein in determining the rotavirus serotype is paramount, which describes the various strains or viral variants that can exist (Glass *et al.* 1996). The specificity of each of the nine rotavirus serotypes' VP4 and VP7 proteins distinguishes it from the others. The principal target of neutralizing antibodies (NAbs) produced by the Ab-mediated immunity of the host following the infection is the VP7 protein (Green & Kapikian 1992). These antibodies can protect against future infections with the same viral serotype. The VP7 and VP4 proteins are also an important component of rotavirus vaccines, as it is included in vaccine formulations to generate virus protection (Usman *et al.*, 2023).

Mutation is an event that will be investigated and studied to comprehend the evolution and for proper understanding of disease surveillance (D'souza *et al.* 2008). RNA virus mutation rates alter how proteins function. These may disrupt transcribed proteins' ability to perform their intended functions which impacts their ability to survive and direct relationships with additional biological elements (Sue E Crawford *et al.*, 2017). Mutations impact the protein stability and related interactions that lead to the potential effect on disease severity, therefore, it is crucial for understanding a variety of processes at the biological level, for example, disease resistance and treatment resistance (Shen *et al.*, 1994). To facilitate the speedy and parallel data analysis at the genomic level, computational tools are applied to anticipate the result of mutations on protein's stability and pathogenicity (Ghosh *et al.*, 2012). Moreover, substitutions and deletions in the spike region are identified, and these mutation sites can be used to find new drugs and vaccines against specific targets.

The current study involves bioinformatics-based exploration of evolutionary and possible mutations among VP4 and VP7 protein sequences of rotavirus

extracted from Saudi Arabia. Additionally, Identification of patterns of mutation and evolutionary process characterization was performed in this study.

MATERIALS AND METHODS

Dataset Collection:

The reference sequences served as the cornerstone for the studies on growth, mutation, and disease. They are considered an essential databank for clinical/medical, functional, and other different research (Barrett *et al.*, 2012; Houldcroft *et al.*, 2017). Accession numbers were utilized to mine the target viral sequences from the virus database of NCBI. By April 23, 2023, the viral sequences (reference genome sequence and query protein segments from Saudi Arabia's rotavirus genomes) were mined from the viral database of NCBI. Concerned sequences of VP4 and VP7 for every isolate were iterated in the required format to estimate the evolutionary relatedness and measure optimal substitutions of nucleotide bases at the genome level. The methodological strategy and the computational tools employed in the current study have been well demonstrated by a flow chart (Fig. 1).

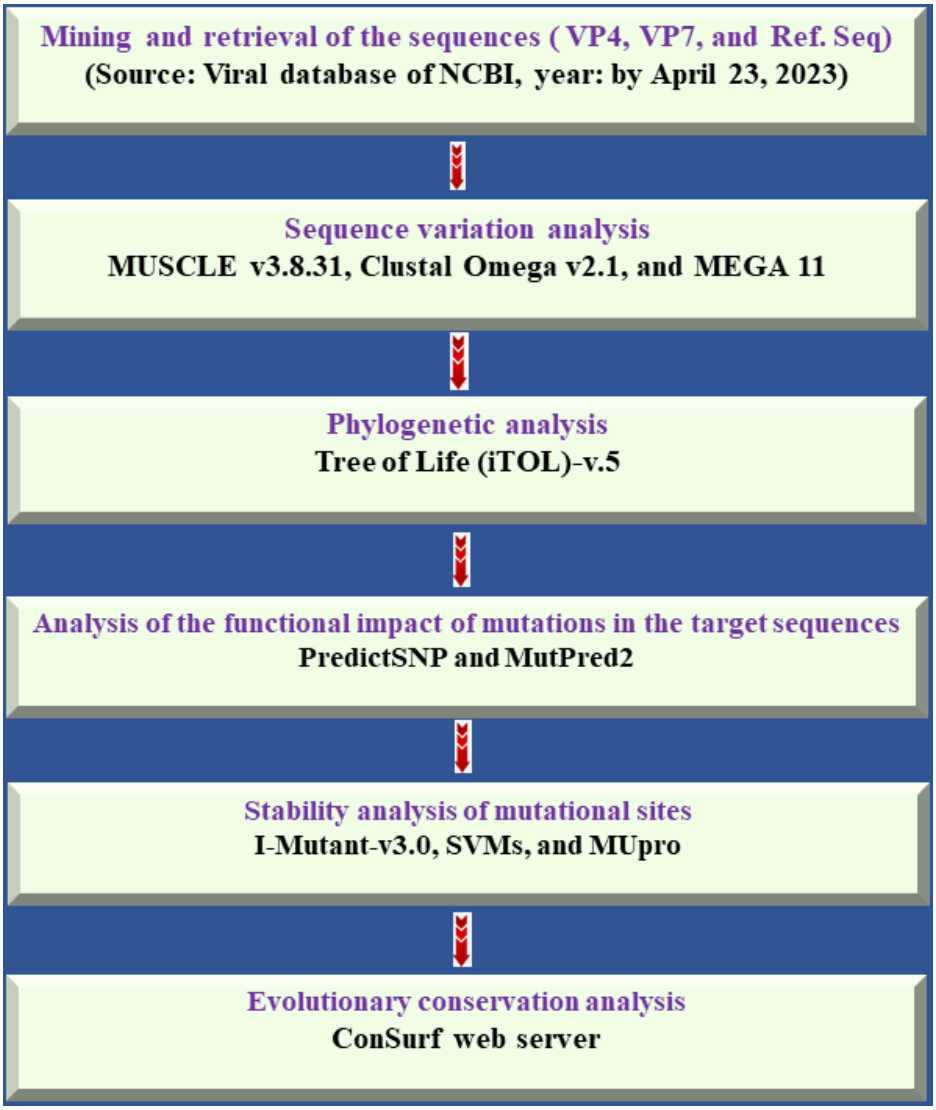


Fig. 1: Depiction of methodological approach and computational tools adopted for this study.

Sequence Variation Analysis:

Multiple sequence alignment was performed by employing MUSCLE (version-3.8.31) and the identity matrices were obtained by processing the data through Clustal Omega (version 2.1) (Edgar 2004;

Thompson *et al.*, 2003). Multiple sequence alignment rendered the identification of conserved sequence patterns and domains among query sequences. Clustal Omega computed the percent identity at each residue position as the number of matches in alignment in a row relative to alignment length. This results in showcasing the genome-level variations and point-mutation sites in protein sequences. As a result, the relationship among each polymorphism site, evolutionary rates, and phylogenetic relationship was assessed using MEGA 11 (Tamura *et al.*, 2021).

Phylogenetic Analysis:

Genomic sequences (VP7 and VP4 regions) of *Rotavirus* were subjected to phylogenetic analysis to generate a maximum likelihood-based best substitution model at default parameters (Junier & Zdobnov 2010; Tamura *et al.*, 2004). The visualization of the intertaxon classification of clades showcasing viral diversity was carried out by using Tree of Life (iTOL)-version-5 (Letunic & Bork 2019,2021). Tajima's relative rate test was performed to compare rates between lineages (Tajima 1993).

Assessment of Functional Impact of Mutations:

The functionality assessment was done by the PredictSNP tool (Bendl *et al.*, 2014) and MutPred2 software (Pejaver *et al.*, 2020). PredictSNP is a combinatorial classifier of several tools that exploit the physio-chemical properties and sequence similarity to assess the functional impact of mutation. Deleterious and neutral mutations are genetic changes that negatively impact an organism's fitness and survival, often affecting protein-coding regions. Deleterious mutations can lead to functional changes, disrupted cellular processes, or harmful traits. They are associated with diseases and are often linked to genetic disorders. Neutral mutations have little to no effect on fitness and are often found in non-coding regions. They are selectively neutral, meaning they do not significantly alter the phenotype or contribute to the overall fitness of the organism (Bao *et al.*, 2022). The probabilistic measure of variation in pathogenicity was estimated by processing genetic and molecular data using MutPred2 (http://mutpred.mutdb.org/).

Stability Analysis of Mutational Sites:

The biological significance of altered sequences from Rotavirus genomes of Saudi Arabian origin was assessed using a combination of methods to determine whether they should be considered neutral or deleterious. I-Mutant (v-3.0) (Capriotti *et al.*, 2008) was employed to predict the stability of the protein by incorporating the principles of linear regression provided by support vector machines (Cheng *et al.*, 2006; Jakkula 2006). Furthermore, the MUpro protein-stability analysis tool (Khan & Vihinen 2010) was used to facilitate the determination of the stability of the protein by combining neural networks and SVM.

Evolutionary Conservation Analysis:

ConSurf web server was used to facilitate computational evaluation and assessment of conservation-degree at every aligned point/location of VP4 and VP7 protein segments (Glaser *et al.*, 2003) of rotavirus. To detect and track substitutions, residues were clustered at default window length, and the evolutionary survival rate was calculated using Bayesian inference. Using a score system at default parameters, the influence of amino acids on protein conformations and related functions was determined. It also allows functional inference by highlighting conserved residues, guiding experimental studies and identifying critical regions for further investigation. The ConSurf web server offers a user-friendly interface, multiple sequence alignment integration, normalized conservation scores, downloadable results, flexible input options, and accessibility for researchers (Glaser *et al.*, 2003).

RESULTS

Analysis of Sequence Variations:

The alignment of sequences revealed shared structural and functional traits as well as common evolutionary descents. Alignment at default parameters of the MUSCLE program followed by visualization and

tranking of variation by MEGA11 among aligned sequences established the phylogenetic relationship between strains and finally various mutation sites at the genomic level were identified. All the mutation sites identified in VP4 and VP7 proteins of different strains have been tabulated in detail in Table 1.

Table 1: List of total number of mutation sites among proteins (VP4 and VP7) of rotavirus genome of Saudi Arabian origin.

VP4		
S.No.	**Accession Number**	**Mutation sites**
1.	ALD51917.1	N63D
2.	ALD51918.1	T195K, I123T,
3.	ALD51919.1	T204I, T195K
4.	ALD51920.1	T204I, T195K, L185S, I123T,
5.	ALD51921.1	T195K, I123T
6.	ALD51922.1	T195R, T149I, G145N, N142D, I123V, T25K, N39D, P64Q, V65T, D66N, N70V, R81E, D83S, Y102S, D110N,
7.	ALD51923.1	N63D
8.	ALD51924.1	T204I, T195K, H2L,3GN,34P, I5S
VP7		
9.	ALD51911.1	S41F, A46T, A65T, A68S, Q72R, G74E, I108T, M217T, S258A, T281I
10.	ALD51912.1	n/a
11.	ALD51913.1	Q92E
12.	ALD51914.1	n/a
13.	ALD51915.1	G74E, I108T
14.	ALD51916.1	A46P

Phylogenetic Analysis:

The phylogenetic analysis provides better insights into the evolution of species and shared similarities and differences among query genomic sequences. The output of protein sequence alignment was subjected to phylogenetic evaluation and phylogenetic trees developed and the evolutionary relationship deduced has been illustrated in Figure 2a (VP4) and Figure 2b (VP7). The relative maximum substitution matrix and evolutionary rate models computed by utilizing the Gamma distribution function implemented in the Jones-Taylor-Thornton model have been depicted (Figs. 3a, 3b, 4a, & 4b). Approximated value of 0.05 of the distribution's shape parameters was measured and maximum likelihood for variation rates was computed for VP4. The proportion of amino acid frequencies evaluated in the VP4 protein segment was 5.11%, 7.69%, 6.64%, 4.25%, 2.03%, 6.18%, 7.47%, 2.30, 9.11%, 2.34%, 4.05%, 5.05%, 5.85%, 1.43%, 5.13%, 3.23%, 4.11%, 5.26%, 5.95%, and 6.82%, for (R), (A), (V) (N), (C), (E), (G), (H), (L), (M), (F), (P), (T), (W), (D), (Y), (Q), (I), (K), and (S) respectively with a highest logarithmic value of likelihood ratio (-857.067) as illustrated in Figure 3a. The similarity of evolutionary rate between sequences (ID/name: ALD51917.1/VP4 partial Human rotavirus A, and (ID/name: ALD51918.1/VP4 partial Human rotavirus A, with sequence (ID/name: ALD51919.1/VP4 partial Human rotavirus A (Fig. 2a) was used in relative rate test of Tajima that is based on Chi-square (χ^2 test) statistical test considering statistical significance cut off level as *P*-value < 0.05. And $\chi^2 = 1.00$ (*P*-value = 0.31731) was measured (Fig. 3b) which was statistically insignificant.

Approximated value of 0.05 of the distribution's shape parameters was measured and maximum likelihood for variation rates was computed for VP7. The proportion of amino acid frequencies assessed in VP7 protein segment (at default consensus setting with r = 100) was 5.11%, 7.69%, 6.64%, 4.25%, 2.03%, 6.18%, 7.47%, 2.30, 9.11%, 2.34%, 4.05%, 5.05%, 5.85%, 1.43%, 5.13%, 3.23%, 4.11%, 5.26%, 5.95%, and 6.82%, for (R), (A), (V) (N), (C), (E), (G), (H), (L), (M),

(F), (P), (T), (W), (D), (Y), (Q), (I), (K), and (S) respectively with highest logarithmic value of likelihood ratio (-958.342) as depicted in Figure 4a. The heterogeneity of viral sequence prediction results was statistically significant with an evaluated transition/Transversion bias (R) = 2.90. The equality of evolutionary rate between sequences (ID/name: ALD51911.1/VP7 partial Human rotavirus A and (ID/name: ALD51912.1/VP7 partial Human rotavirus A with sequence (ID/name: ALD51913.1/VP7 partial Human rotavirus A (Fig. 2b) was used as an outgroup in relative rate test of Tajima based on Chi-square (χ^2 test) statistical test considering statistical significance cut off level as *P*-value < 0.05. And $\chi^2 = 23.00$ (*P*-value < 0.001) which was statistically significant (Fig 4b). Thus the phylogenetic analysis revealed the heterogeneity and maximum plausible mutations in ALD51911.1/VP7 partial human rotavirus A.

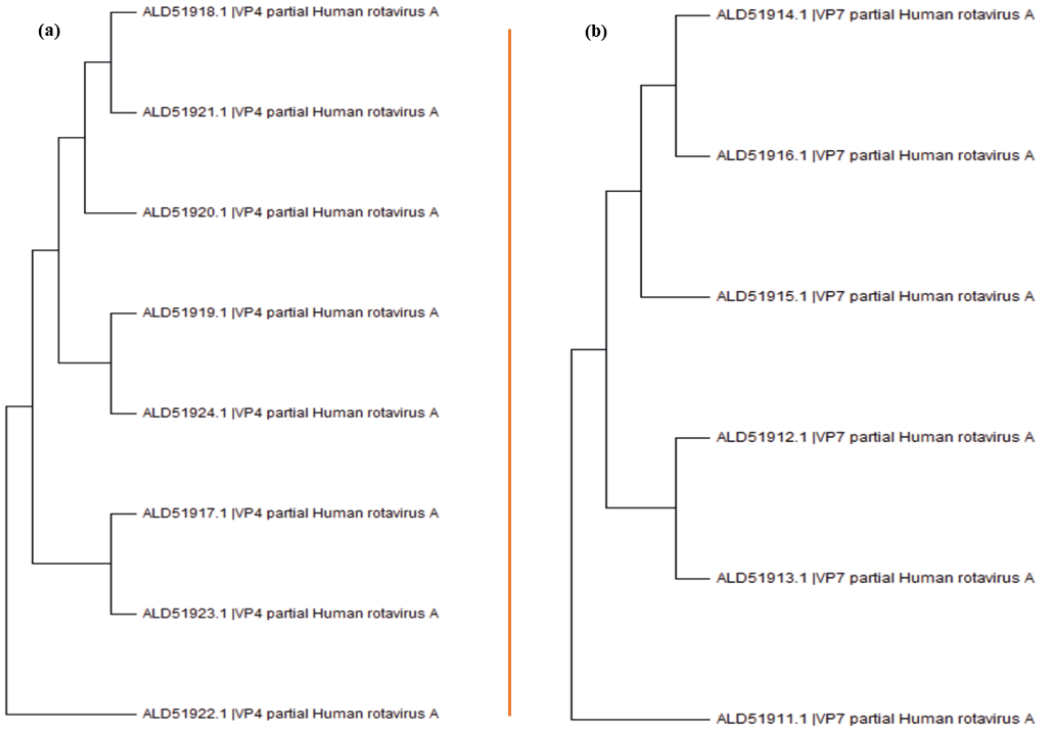


Fig. 2: Phylogenetic tree of protein sequences from (a) VP4 and (b) VP7 segment of Rotavirus genome from Saudi Arabia using MEGA11.

Maximum Likelihood Estimate of Substitution Matrix

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	-	0.14	0.12	0.22	0.06	0.12	0.34	0.67	0.03	0.10	0.15	0.11	0.06	0.03	0.51	1.37	1.39	0.01	0.02	1.01
R	0.21	-	0.10	0.04	0.11	0.64	0.10	0.53	0.38	0.07	0.18	2.01	0.05	0.01	0.19	0.35	0.20	0.09	0.04	0.06
N	0.22	0.12	-	1.48	0.03	0.16	0.19	0.30	0.48	0.13	0.06	0.78	0.04	0.02	0.03	1.79	0.71	0.00	0.12	0.06
D	0.33	0.04	1.22	-	0.01	0.11	2.49	0.49	0.12	0.03	0.03	0.09	0.02	0.01	0.03	0.21	0.13	0.00	0.08	0.11
C	0.23	0.27	0.07	0.03	-	0.02	0.02	0.21	0.09	0.04	0.08	0.02	0.05	0.14	0.03	0.76	0.14	0.08	0.35	0.21
Q	0.22	0.80	0.17	0.14	0.01	-	1.09	0.09	0.68	0.02	0.33	0.91	0.06	0.01	0.42	0.19	0.16	0.01	0.04	0.06
E	0.43	0.08	0.13	2.07	0.01	0.73	-	0.43	0.03	0.03	0.05	0.53	0.02	0.01	0.05	0.11	0.10	0.01	0.01	0.16
G	0.69	0.36	0.17	0.34	0.06	0.05	0.36	-	0.02	0.01	0.03	0.08	0.02	0.01	0.05	0.66	0.10	0.04	0.01	0.16
H	0.09	0.85	0.89	0.27	0.08	1.21	0.08	0.08	-	0.05	0.26	0.16	0.04	0.10	0.30	0.26	0.14	0.01	0.98	0.04
I	0.14	0.06	0.11	0.03	0.02	0.02	0.04	0.02	0.02	-	1.10	0.06	0.59	0.16	0.03	0.14	0.77	0.01	0.05	3.28
L	0.12	0.10	0.03	0.02	0.02	0.15	0.03	0.03	0.06	0.64	-	0.05	0.47	0.53	0.28	0.21	0.08	0.04	0.04	0.61
K	0.15	1.73	0.56	0.08	0.01	0.63	0.55	0.10	0.06	0.06	0.07	-	0.08	0.01	0.06	0.17	0.29	0.01	0.01	0.04
M	0.19	0.11	0.07	0.05	0.04	0.10	0.06	0.05	0.04	1.32	1.82	0.19	-	0.09	0.04	0.10	0.64	0.01	0.03	1.05
F	0.06	0.02	0.02	0.01	0.07	0.01	0.01	0.02	0.05	0.21	1.18	0.01	0.05	-	0.04	0.33	0.04	0.04	0.92	0.20
P	0.78	0.19	0.03	0.03	0.01	0.34	0.06	0.08	0.14	0.03	0.50	0.07	0.02	0.03	-	0.99	0.36	0.01	0.02	0.07
S	1.55	0.27	1.12	0.16	0.23	0.12	0.10	0.73	0.09	0.11	0.28	0.15	0.03	0.20	0.73	-	1.45	0.02	0.11	0.14
T	1.83	0.17	0.52	0.11	0.05	0.11	0.11	0.12	0.06	0.70	0.13	0.30	0.26	0.03	0.31	1.69	-	0.01	0.03	0.39
W	0.03	0.33	0.01	0.02	0.12	0.04	0.04	0.21	0.02	0.04	0.25	0.03	0.02	0.11	0.02	0.11	0.02	-	0.13	0.08
Y	0.06	0.06	0.15	0.12	0.22	0.05	0.02	0.02	0.70	0.08	0.11	0.03	0.02	1.15	0.03	0.22	0.06	0.06	-	0.06
V	1.16	0.05	0.04	0.08	0.07	0.04	0.15	0.18	0.01	2.60	0.83	0.04	0.37	0.12	0.06	0.14	0.35	0.02	0.03	-

(a)

Amino acid frequencies: 7.69% (A), 5.11% (R), 4.25% (N), 5.13% (D), 2.03% (C), 4.11% (Q), 6.18% (E), 7.47% (G), 2.30% (H), 5.26% (I), 9.11% (L), 5.95% (K), 2.34% (M), 4.05% (F), 5.05% (P), 6.82% (S), 5.85% (T), 1.43% (W), 3.23% (Y), and 6.64% (V)

Table. Results from the Tajima's test for 3 Sequences

Configuration	Count
Identical sites in all three sequences	203
Divergent sites in all three sequences	0
Unique differences in Sequence A	3
Unique differences in Sequence B	1
Unique differences in Sequence C	1

(b)

Fig. 3: (a) Maximum likelihood estimates of substitution matrix and **(b)** Tajima's relative rate test of VP7 protein segment.

Maximum Likelihood Estimate of Substitution Matrix

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	-	0.14	0.12	0.22	0.06	0.12	0.34	0.67	0.03	0.10	0.15	0.11	0.06	0.03	0.51	1.37	1.39	0.01	0.02	1.01
R	0.21	-	0.10	0.04	0.11	0.64	0.10	0.53	0.38	0.07	0.18	2.01	0.05	0.01	0.19	0.35	0.20	0.09	0.04	0.06
N	0.22	0.12	-	1.48	0.03	0.16	0.19	0.30	0.48	0.13	0.06	0.78	0.04	0.02	0.03	1.79	0.71	0.00	0.12	0.06
D	0.33	0.04	1.22	-	0.01	0.11	2.49	0.49	0.12	0.03	0.03	0.09	0.02	0.01	0.03	0.21	0.13	0.00	0.08	0.11
C	0.23	0.27	0.07	0.03	-	0.02	0.02	0.21	0.09	0.04	0.08	0.02	0.05	0.14	0.03	0.76	0.14	0.08	0.35	0.21
Q	0.22	0.80	0.17	0.14	0.01	-	1.09	0.09	0.68	0.02	0.33	0.91	0.06	0.01	0.42	0.19	0.16	0.01	0.04	0.06
E	0.43	0.08	0.13	2.07	0.01	0.73	-	0.43	0.03	0.03	0.05	0.53	0.02	0.01	0.05	0.11	0.10	0.01	0.01	0.16
G	0.69	0.36	0.17	0.34	0.06	0.05	0.36	-	0.02	0.01	0.03	0.08	0.02	0.01	0.05	0.66	0.10	0.04	0.01	0.16
H	0.09	0.85	0.89	0.27	0.08	1.21	0.08	0.08	-	0.05	0.26	0.16	0.04	0.10	0.30	0.26	0.14	0.01	0.98	0.04
I	0.14	0.06	0.11	0.03	0.02	0.02	0.04	0.02	0.02	-	1.10	0.06	0.59	0.16	0.03	0.14	0.77	0.01	0.05	3.28
L	0.12	0.10	0.03	0.02	0.02	0.15	0.03	0.03	0.06	0.64	-	0.05	0.47	0.53	0.28	0.21	0.08	0.04	0.04	0.61
K	0.15	1.73	0.56	0.08	0.01	0.63	0.55	0.10	0.06	0.06	0.07	-	0.08	0.01	0.06	0.17	0.29	0.01	0.01	0.04
M	0.19	0.11	0.07	0.05	0.04	0.10	0.06	0.05	0.04	1.32	1.82	0.19	-	0.09	0.04	0.10	0.64	0.01	0.03	1.05
F	0.06	0.02	0.02	0.01	0.07	0.01	0.01	0.02	0.05	0.21	1.18	0.01	0.05	-	0.04	0.33	0.04	0.04	0.92	0.20
P	0.78	0.19	0.03	0.03	0.01	0.34	0.06	0.08	0.14	0.03	0.50	0.07	0.02	0.03	-	0.99	0.36	0.01	0.02	0.07
S	1.55	0.27	1.12	0.16	0.23	0.12	0.10	0.73	0.09	0.11	0.28	0.15	0.03	0.20	0.73	-	1.45	0.02	0.11	0.14
T	1.83	0.17	0.52	0.11	0.05	0.11	0.11	0.12	0.06	0.70	0.13	0.30	0.26	0.03	0.31	1.69	-	0.01	0.03	0.39
W	0.03	0.33	0.01	0.02	0.12	0.04	0.04	0.21	0.02	0.04	0.25	0.03	0.02	0.11	0.02	0.11	0.02	-	0.13	0.08
Y	0.06	0.06	0.15	0.12	0.22	0.05	0.02	0.02	0.70	0.08	0.11	0.03	0.02	1.15	0.03	0.22	0.06	0.06	-	0.06
V	1.16	0.05	0.04	0.08	0.07	0.04	0.15	0.18	0.01	2.60	0.83	0.04	0.37	0.12	0.06	0.14	0.35	0.02	0.03	-

(a)

Amino acid frequencies: 7.69% (A), 5.11% (R), 4.25% (N), 5.13% (D), 2.03% (C), 4.11% (Q), 6.18% (E), 7.47% (G), 2.30% (H), 5.26% (I), 9.11% (L), 5.95% (K), 2.34% (M), 4.05% (F), 5.05% (P), 6.82% (S), 5.85% (T), 1.43% (W), 3.23% (Y), and 6.64% (V)

Configuration	Count
Identical sites in all three sequences	257
Divergent sites in all three sequences	0
Unique differences in Sequence A	23
Unique differences in Sequence B	0
Unique differences in Sequence C	0

(b)

Fig. 4: (a) Maximum likelihood estimate of substitution matrix and **(b)** Tajima's relative rate test of VP7 protein segment.

Pathogenicity Analysis:

The functional capacity of protein sequence is affected by possible amino acid substitutions. The classification of mutational sites (neutral or deleterious) was based on the underlying algorithm of the tools. PredictSNP displayed the consensus findings from a combination of tools which were summarized using an Excel sheet (Supplementary File: Sheet 1). Various deleterious and neutral mutations assessed were highlighted by using red and green color codes respectively (Figs. 5a and 5b). In totality, thirty-five mutational sites (n =35) were characterized. Out of n = 35, only two sites (n = 02) in VP7 protein (I108T, S258A) and five sites (n = 5) in VP4 protein (G145N, I123V, Y102S, D110N, and T195K) with their predicted conservation scores were characterized as deleterious mutations. Additionally, two harmful mutational sites (D110N and S258A) were further assessed to decrease the stability of the VP4 and VP7 proteins by employing a combination of prediction tools *viz.* I-Mutant (v-3.0) and MuPro (Supplementary File: Sheet 2).

Fig. 5: Demonstration of conservation analysis of the target protein sequences (a) VP4 and (b) VP7 segment of the Rotavirus genome from Saudi Arabia using ConSurf.

Analysis of Conservation of Protein:

The pattern of conserved domains demonstrates that an amino acid residue has remarkable potential in conserving the protein structure and function. The frequency of amino acid stability at each aligned position to highlight the localized succession is examined by ConSurf. The related alignment of multiple sequences was submitted to the ConSurf web server for the preservation of rate of development analysis including Bayesian empirical interface computation (Figure 5a and Figure 5b). HMMER algorithm dependent UNIREF90 was employed for homologs identification with an E-value of 0.0001 at 1 number of iterations, maximal percent ID between sequences = 95, minimal percent ID for homologs = 35) which indicated the variable mutation position. As the variable mutation positions are subjected to variable protein function, D110N and S258A positions may have a negative impact by making proteins less stable.

DISCUSSION

By calculating the expected number of nucleotide alterations at each position, the genetic distance between homologous sequences may be estimated, enabling lineage sharing and ancestor identification (Park *et al.* 2014). The foundation for genome evolution at each succession is laid by the ideas of speciation and horizontal gene transfer (Boto 2010; Domingo 2007). The process of creating molecular evolutionary models involves three stages: (1) searching for global similarity between query sequences, (2) creating a mathematical model based on query sequences' physio-chemical properties, and (3) using statistical techniques to estimate the phylogenetic relationship between query sequences (Boto 2010; Domingo 2007). Moreover, by computing the molecular evolutionary model, any variation and associated functional change at the genome level are estimated. The current research aimed to examine the molecular evolutionary rates of the rotavirus genome's VP4 and VP7 protein sequences. In support, conservation analysis enhances our understanding of viral infections by identifying common features critical for the virus's survival and interaction with the host. It also provides evolutionary insights into the virus's evolution, enabling the development of diagnostic markers for viral infections, and facilitating the development of targeted therapeutic strategies. Conservation analysis can also be used for epidemiological surveillance, allowing early detection of emerging variants and informing public health interventions (Pybus and Rambaut, 2009).

Genetic analysis revealed a high mutation rate of VP7 protein (1.45×10^{-3} mutations/site/year) highlighting the multiple amino acid differences among VP4 and dVP7 antigenic epitopes (Sue E Crawford *et al.* 2017) which corroborates with findings of this study. Results of the current research were indicative of the deleterious alteration in VP4 and VP7 proteins of Saudi Arabian origin which are consistent with the reports of several studies describing the capability of rotavirus of undergoing antigenic drift and drift-mediated zoonotic transmission (Flewett *et al.* 1973; Sen *et al.* 2022). Furthermore, the genetic alteration observed in VP4 and VP7 proteins signifies the evolution of the rotavirus in the region which could impact the efficacy of the vaccine in use because evolution includes the concepts of natural selection and genetic alteration (genetic shift and genetic drift) (McDonald *et al.* 2009; Satyam *et al.* 2021; Satyam *et al.* 2020; Satyam *et al.* 2018). Additionally, the phylogenetic analysis relies on the possibility of resolving discrepancies in genetic-based studies (Menet *et al.,* 2022).

Various genome-level mutation or variation sites or potential variant sites were identified in the retrieved complete sequences (VP4 and VP7) of rotavirus of Saudi Arabian origin (Table 1) which corroborates with the reports of other studies (Song & Hao 2009; Song *et al.* 2010). In addition to that, phylogenetic analysis of VP4 and VP7 sequences at the individual level identifies the evolutionary rates and maximum possible substitutions at a global level (Maurya *et al.*

2021; Said *et al.* 2023). In the case of VP4 segments of rotavirus, ALD51922.1 is a distant member of the complete clade which might be the result of maximum point mutations followed by ALD51920.1. The Jones-Taylor-Thornton model was used to analyze the molecular evolutionary models for 8 protein sequences of the VP4 segment of the rotavirus genome utilizing discrete Gamma distribution. This model showed that during transmission, amino acids are present at different rates and show varying degrees of transitions and transversions (Stecher *et al.* 2020). The proportion of amino acid frequencies evaluated in the VP4 and VP7 protein segments was 5.11%, 7.69%, 6.64%, 4.25%, 2.03%, 6.18%, 7.47%, 2.30, 9.11%, 2.34%, 4.05%, 5.05%, 5.85%, 1.43%, 5.13%, 3.23%, 4.11%, 5.26%, 5.95%, and 6.82%, for (R), (A), (V) (N), (C), (E), (G), (H), (L), (M), (F), (P), (T), (W), (D), (Y), (Q), (I), (K), and (S) respectively with a highest logarithmic value of likelihood ratio (-857.067) in VP4 and (-958.342) in VP7. Moreover, the heterogeneity of viral sequence prediction with an evaluated transition/Transversion bias (R) = 2.90 was significant. The equality of evolutionary rate between sequences (ID/ name: ALD51911.1/VP7 partial Human *rotavirus A* and (ID/name: ALD51912.1/VP7 partial Human *rotavirus A* with sequence (ID/name: ALD51913.1/VP7 partial Human *rotavirus A* (Figure 2b) was statistically significant ($\chi^2 = 23.00$; *P*-value < 0.001). Thus the phylogenetic analysis revealed the heterogeneity and maximum plausible mutations in ALD51911.1/VP7 partial human rotavirus A of Saudi origin. This finding was consistent with the results of other studies (Ianiro *et al.* 2016; Maunula & Von Bonsdorff 2002; Nyaga *et al.* 2018). Whereas the similarity of evolutionary rate between sequences of VP4 (ID/name: ALD51917.1/VP4 partial Human rotavirus *A*, and (ID/name: ALD51918.1/VP4 partial Human rotavirus *A*, with sequence **(**ID/name: ALD51919.1/VP4 partial Human *rotavirus A* (Figure 2a) was statistically not significant *(*$\chi^2 = 1.00$; *P*-value = 0.31731). The maximum-likelihood method offers advantages over heuristic algorithms, including statistical rigor, parameter uncertainty quantification, consistency and asymptotic properties, and model comparison by allowing quantification of parameter uncertainty, allowing for confidence intervals or likelihood ratio tests to assess precision and reliability (Zhou *et al.,* 2018). In addition to that, stability analysis identified D110N in the VP4 segment and S258A in the VP7 segment respectively as having decreased stability over evolution. In whole genome sequence (WGS) and phylogenetic analysis of *rotavirus A.* There is scanty information regarding the genetic variability of VP4 (outer capsid protein), however, genetic variability plays a vital role in impacting the vaccine program (Kirkwood 2010). D110 is a crucial position in *rotavirus A* strain (SLO/D110–15) identified in 2015 (Jamnikar-Ciglenecki *et al.* 2017) and substitution at this position with decreased stability (D110N in the VP4) could alter the transmissibility and pathogenicity of the rotavirus. S258A in the VP7 could impact the effectiveness of the currently running vaccine program because VP7 is one of the key components of subunit vaccines (Khodabandehloo *et al.* 2012).

Future researchers will utilize this study as a foundation to investigate the structural conformity in variations caused by sites of mutation. Such sites are subjected to additional research to discover novel drugs and vaccines. Provided the impact of epidemiological and genomic characteristics of rotavirus, it is critical to perform epidemiological surveillance on rotavirus to design disease control in endemic areas.

Conclusion

The rate of the evolutionary path is suggested by the increase in variations patterns in the rotavirus genome as compared to present studies. The possibility of creating a vaccine that can be useful for Saudi Arabia has been made possible by the identification of mutational signatures among VP4 and VP7 protein sequences from rotavirus. The findings of this research will serve as a model for rational drug and vaccine constructs that help in reducing the deaths caused by

rotavirus making it a global issue. Continued research in these areas will contribute to our understanding of viral pathogenesis, inform public health strategies, and guide the development of effective therapeutic and preventive measures in the face of evolving viral threats. The scarcity of knowledge related to the surveillance of the current outbreak necessitates additional research to serve the insights and platforms required to better predict the dynamics of disease and eradication.

Declarations:

Ethics Approval and Consent to Participate. Not applicable

Consent for Publication: Not applicable

Availability of data and material:All the data related to this manuscript is available from the corresponding author which can be obtained upon request.

Competing interests:No non-financial and non-financial competing interests

Funding: This investigation has not received any funding.

Acknowledgements: Not applicable

REFERENCES

Al-Ayed M S Z, Asaad A M, Qureshi M A, & Hawan A A (2017). Epidemiology of group A rotavirus infection after the introduction of monovalent vaccine in the National Immunization Program of Saudi Arabia. *Journal of Medical Virology,* 89(3): 429-434.

Aoki S T, Settembre E C, Trask S D, Greenberg H B, Harrison S C, & Dormitzer P R (2009). Structure of rotavirus outer-layer protein VP7 bound with a neutralizing Fab. *Science,* 324(5933): 1444-1447.

Bao K, Melde RH, Sharp NP (2022). Are mutations usually deleterious? A perspective on the fitness effects of mutation accumulation. *Evolutionary ecology,* 36(5):753-766.

Barrett T, Wilhite S E, Ledoux P, Evangelista C, Kim I F, Tomashevsky M, Marshall K A, Phillippy K H, Sherman P M, & Holko M (2012). NCBI GEO: archive for functional genomics data sets update. *Nucleic acids research,* 41(D1): D991-D995.

Bendl J, Stourac J, Salanda O, Pavelka A, Wieben E D, Zendulka J, Brezovsky J, & Damborsky J (2014). PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. PLoS computational biology, 10(1): e1003440.

Bishop R, Davidson G, Holmes I, & Ruck B (1973). Virus particles in epithelial cells of duodenal mucosa from children with acute non-bacterial gastroenteritis. *The Lancet,* 302 (7841): 1281-1283.

Blutt S E, Crawford S E, Warfield K L, Lewis D E, Estes M K, & Conner M E (2004). The VP7 outer capsid protein of rotavirus induces polyclonal B-cell activation. *Journal of virology,* 78(13): 6974-6981.

Boto L (2010). Horizontal gene transfer in evolution: facts and challenges. *Proceedings of the Royal Society B: Biological Sciences,* 277(1683): 819-827.

Brunet J-P, Cotte-Laffitte J, Linxe C, Quero A-M, Géniteau-Legendre M, & Servin A (2000). Rotavirus infection induces an increase in intracellular calcium concentration in human intestinal epithelial cells: role in microvillar actin alteration. *Journal of virology,* 74(5): 2323-2332.

Capriotti E, Fariselli P, Rossi I, & Casadio R (2008). A three-state prediction of single point mutations on protein stability changes. B*MC bioinformatics,* 9(2): 1-9.

Carvalho M F, & Gill D (2018). Rotavirus vaccine efficacy: current status and areas for improvement. *Human vaccines & immunotherapeutics,* 15(6):1237-1250.doi: 10. 1080/ 21645515.2018.1520583.

Cheng J, Randall A, & Baldi P (2006). Prediction of protein stability changes for single-site mutations using support vector machines.

Proteins: Structure, Function, and *Bioinformatics,* 62(4): 1125-1132.

Crawford S E, Mukherjee S K, Estes M K, Lawton J A, Shaw A L, Ramig R F, & Prasad B V (2001). Trypsin cleavage stabilizes the rotavirus VP4 spike. *Journal of virology,* 75(13): 6052-6061.

Crawford S E, Ramani S, Tate J E, Parashar U D, Svensson L, Hagbom M, Franco M A, Greenberg H B, O'Ryan M, & Kang G (2017). Rotavirus infection. *Nature Reviews Disease Primers,* 3(1): 1-16.

Crawford S E, Ramani S, Tate J E, Parashar U D, Svensson L, Hagbom M, Franco M A, Greenberg H B, O'Ryan M, Kang G, Desselberger U, & Estes M K (2017). Rotavirus infection. *Nature Reviews Disesease Primers,* 3: 17083. https://doi.org/10. 1038/ nrdp.2017.83

D'souza R, Hall G, & Becker N (2008). Climatic factors associated with hospitalizations for rotavirus diarrhoea in children under 5 years of age. *Epidemiology & Infection,* 136(1): 56-64.

Dennehy P H (2008). Rotavirus vaccines: an overview. *Clinical microbiology reviews,* 21(1): 198-208.

Domingo E (2007). Virus evolution. *Fields virology,* 5: 391-422.

Edgar R C (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research,* 32(5): 1792-1797.

Flewett T, Bryden A, & Davies H (1973). Virus particles in gastroenteritis. *Lancet;*2(7844):1497.doi: 10.1016/ s0140-6736(73)92760-8.

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: e40749

Glaser F, Pupko T, Paz I, Bell R E, Bechor-Shental D, Martz E, & Ben-Tal N (2003). ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics,* 19(1): 163-164.

Glass R I, Kilgore P E, Holman R C, Jin S, Smith J C, Woods P A, Clarke M J, Ho M S, & Gentsch J R (1996). The epidemiology of rotavirus diarrhea in the United States: surveillance and estimates of disease burden. *Journal of Infectious Diseases,* 174 (Supplement_1): S5-S11.

Goel A K, Chawla S, Dhingra A, Thiyagarajan V, & Nair N P (2021). Rotavirus Diarrhea and its Determinants Among Under-Five Children Admitted in a Tertiary Care Hospital of Southern Haryana, India. *Indian journal of Pediatrics,* 88(Suppl 1): 16-21. https://doi. org/ 10.1007/s12098-020-03616-1

Green K, & Kapikian A Z (1992). Identification of VP7 epitopes associated with protection against human rotavirus illness or shedding in volunteers. *Journal of virology,* 66(1): 548-553.

Greenberg H B, & Estes M K (2009). Rotaviruses: from pathogenesis to vaccination. *Gastroenterology,* 136 (6): 1939-1951.

Houldcroft C J, Beale M A, & Breuer J (2017). Clinical and biological insights from viral genome sequencing. *Nature Reviews Microbiology,* 15(3): 183-192.

Ianiro G, Delogu R, Fiore L, Ruggeri F M, & Group R-I S (2016). Genetic variability of VP7, VP4, VP6 and NSP4 genes of common human G1P [8] rotavirus strains circulating in Italy between 2010 and 2014. *Virus research,* 220: 117-128.

Jakkula V (2006). Tutorial on support vector machine (svm). *School of EECS, Washington State University,* 37 (2.5): 3.

Jamnikar-Ciglenecki U, Kuhar U, Steyer A, & Kirbis A (2017). Whole genome

sequence and a phylogenetic analysis of the G8P[14] group A rotavirus strain from roe deer. *BMC Veterinary Research,* 13(1): 353. https://doi.org/10.1186/s12917-017-1280-4

Junier T, & Zdobnov E M (2010). The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics,* 26(13): 1669-1670.

Khan S, & Vihinen M (2010). Performance of protein stability predictors. *Human mutation,* 31(6): 675-684.

Khodabandehloo M, Shahrabadi M S, Keyvani H, Bambai B, & Sadigh Z (2012). Recombinant outer capsid glycoprotein (VP7) of rotavirus expressed in insect cells induces neutralizing antibodies in rabbits. *Iranian Journal of Public Health,* 41(5): 73-84.

Kirkwood C D (2010). Genetic and antigenic diversity of human rotaviruses: potential impact on vaccination programs. *Journal of Infectious Diseases,* 202(Supplement_1): S43-S48.

Letunic I, & Bork P (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic acids research,* 47(W1): W256-W259.

Letunic I, & Bork P (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic acids research,* 49(W1): W293-W296.

Ludert J E, Krishnaney A A, Burns J W, Vo P T, & Greenberg H B (1996). Cleavage of rotavirus VP4 in vivo. *Journal of general virology,* 77(3): 391-395.

Maunula L, & Von Bonsdorff C-H (2002). Frequent reassortments may explain the genetic heterogeneity of rotaviruses: analysis of Finnish rotavirus strains. *Journal of virology,* 76(23): 11793-11800.

Maurya A P, Chikhale R V, & Pandey P (2021). Transmission of SARS-CoV-2 in South Asian countries: molecular evolutionary model based phylogenetic and mutation analysis. *Environmental Sustainability,* 4: 533-541.

McDonald S M, Matthijnssens J, McAllen J K, Hine E, Overton L, Wang S, Lemey P, Zeller M, Van Ranst M, & Spiro D J (2009). Evolutionary dynamics of human rotaviruses: balancing reassortment with preferred genome constellations. *PLoS pathogens,* 5(10): e1000634.

Menet H, Daubin V, Tannier E (2022). Phylogenetic reconciliation. *PLoS Computational Biology,* 2022 18(11):e1010621.

Michelangeli F, Liprandi F, Chemello M E, Ciarlet M, & Ruiz M-C (1995). Selective depletion of stored calcium by thapsigargin blocks rotavirus maturation but not the cytopathic effect. *Journal of virology,* 69(6): 3838-3847.

Morris A P, Scott J K, Ball J M, Zeng C Q-Y, O'Neal W K, & Estes M K (1999). NSP4 elicits age-dependent diarrhea and Ca2+ mediated I− influx into intestinal crypts of CF mice. *American Journal of Physiology-Gastrointestinal and Liver Physiology,* 277(2): G431-G444.

Nyaga M M, Tan Y, Seheri M L, Halpin R A, Akopov A, Stucker K M, Fedorova N B, Shrivastava S, Steele A D, & Mwenda J M (2018). Whole-genome sequencing and analyses identify high genetic heterogeneity, diversity and endemicity of rotavirus genotype P [6] strains circulating in Africa. *Infection, Genetics and Evolution,* 63: 79-88.

Parashar U D, Gibson C J, Bresee J S, & Glass R I (2006). Rotavirus and severe childhood diarrhea. *Emerging infectious diseases,* 12(2): 304.

Parashar U D, Nelson E A S, & Kang G (2013). Diagnosis, management, and

prevention of rotavirus gastroenteritis in children. *BMJ*, 2013;347:f7204. https://doi.org/ 10. 1136/bmj.f7204

Park J-G, Kim D-S, Matthijnssens J, Kwon H-J, Zeller M, Alfajaro M M, Son K-Y, Hosmillo M, Ryu E-H, & Kim J-Y (2014). Comparison of pathogenicities and nucleotide changes between porcine and bovine reassortant rotavirus strains possessing the same genotype constellation in piglets and calves. *Veterinary microbiology,* 172(1-2): 51-62.

Pejaver V, Urresti J, Lugo-Martinez J, Pagel K A, Lin G N, Nam H-J, Mort M, Cooper D N, Sebat J, & Iakoucheva L M (2020). Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nature communications,* 11(1): 5918.

Pybus OG, Rambaut A (2009). Evolutionary analysis of the dynamics of viral infectious disease. *Nature Reviews Genetics,1*0(8):540-50.

Said K B, Alsolami A, Alshammari F, Alshammari K F, Alazmi M, Bhardwaj T, Najm M Z, Singh R, & Kausar M A (2023). Molecular evolutionary model based on phylogenetic and mutation analysis of SARS-CoV-2 spike protein sequences from Asian countries: A phylogenomic approach. *Informatics in Medicine Unlocked,* 38: 101221.

Sánchez-Tacuba L, Kawagishi T, Feng N, Jiang B, Ding S, & Greenberg H B (2022). The role of the VP4 attachment protein in rotavirus host range restriction in an in vivo suckling mouse model. *Journal of virology,* 96(15): e00550-00522.

Satyam R, Bhardwaj T, Goel S, Jha N K, Jha S K, Nand P, Ruokolainen J, Kamal M A, & Kesari K K (2021). miRNAs in SARS-CoV 2: a spoke in the wheel of pathogenesis. *Current pharmaceutical design*, 27(13): 1628-1641.

Satyam R, Bhardwaj T, Jha N K, Jha S K, & Nand P (2020). Toward a chimeric vaccine against multiple isolates of Mycobacteroides-An integrative approach. *Life sciences,* 250: 117541.

Satyam R, Janahi E M, Bhardwaj T, Somvanshi P, Haque S, & Najm M Z (2018). In silico identification of immunodominant B-cell and T-cell epitopes of non-structural proteins of Usutu Virus. *Microbial pathogenesis,* 125: 129-143.

Sen A, Bansal R, Bhardwaj T, Rathi B, Almalki A H, Janahi E M, Alsulimani A, Tewari B N, Somvanshi P, & Haque S (2022). In-silico analysis of multiepitope based vaccine targeting respiratory viruses sars, Mers and sars-coV-2. *Minerva Biotechnology & Biomolecular,* 34(3).

Sen A, Sen N, & Mackow E R (2007). The formation of viroplasm-like structures by the rotavirus NSP5 protein is calcium regulated and directed by a C-terminal helical domain. *Journal of virology,* 81(21): 11758-11767.

Shen S, Burke B, & Desselberger U (1994). Rearrangement of the VP6 gene of a group A rotavirus in combination with a point mutation affecting trimer stability. *Journal of virology,* 68(3): 1682-1688.

Solberg O D, Hasing M E, Trueba G, & Eisenberg J N (2009). Characterization of novel VP7, VP4, and VP6 genotypes of a previously untypeable group A rotavirus. *Virology,* 385(1): 58-67.

Song X-f, & Hao Y (2009). Adaptive evolution of rotavirus VP7 and NSP4 genes in different species. *Computational biology and chemistry,* 33(4): 344-349.

Song X, Hao Y, & Han P. (2010). Amino Acid Substitution Positions on the VP7 and NSP4 Molecules of Rotavirus Circulating in China. International

Conference on Bio-Science and Bio-Technology, (pp. 265-272). Berlin, Heidelberg: Springer Berlin Heidelberg.

Stecher G, Tamura K, & Kumar S (2020). Molecular evolutionary genetics analysis (MEGA) for macOS. *Molecular biology and evolution,* 37(4): 1237-1239.

Tajima F (1993). Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics,* 135(2): 599-607.

Tamura K, Nei M, & Kumar S (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences,* 101(30): 11030-11035.

Tamura K, Stecher G, & Kumar S (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Molecular biology and evolution,* 38(7): 3022-3027.

Thompson J D, Gibson T J, & Higgins D G (2003). Multiple sequence alignment using ClustalW and ClustalX. *Current protocols in bioinformatics,* (1): 2.3. 1-2.3. 22.

Trask S D, & Dormitzer P R (2006). Assembly of highly infectious rotavirus particles recoated with recombinant outer capsid proteins. *Journal of virology,* 80(22): 11293-11304.

Usman M, Ayub A, Habib S, Rana M S, Rehman Z, Zohaib A, Jamal S B, Jaiswal A K, Andrade B S, & de Carvalho Azevedo V (2023). Vaccinomics Approach for Multi-Epitope Vaccine Design against Group A Rotavirus Using VP4 and VP7 Proteins. *Vaccines,* 11(4): 726.

Zhou X, Shen XX, Hittinger CT, Rokas A (2018). Evaluating Fast Maximum Likelihood-Based Phylogenetic Programs Using Empirical Phylogenomic Data Sets. *Molecular Biology Evolution,*35(2):486-503.

Supplementary Files

Sheet 1:Pathogenicity analysis.

Mutations	PredictSNP	MAPP	PhD-SNP	PolyPhen-1	PolyPhen-2	SIFT	SNAP	nsSNPAnalyzer
N63D	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
T195K	Deleterious	Neutral	Deleterious	Deleterious	Neutral	Neutral	Deleterious	Deleterious
I123T	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
T204I	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
L185S	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
H2L	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
G3N	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
E4P	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
I5S	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
D110N	Deleterious	Neutral	Neutral	Deleterious	Deleterious	Neutral	Deleterious	Neutral
Y102S	Deleterious	Deleterious	Deleterious	Neutral	Deleterious	Neutral	Neutral	Deleterious
R81E	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
N70V	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
D66N	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
V65T	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
P64Q	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
N39D	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
T25K	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
I123V	Deleterious	Neutral	Neutral	Neutral	Deleterious	Neutral	Deleterious	Deleterious
N142D	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
G145N	Deleterious	Deleterious	Deleterious	Deleterious	Neutral	Neutral	Neutral	Deleterious
T149I	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
T195R	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
For VP7								
S41F	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
A46T	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
A65T	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
A68S	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
Q72R	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
G74E	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
I108T	Deleterious	Neutral	Neutral	Deleterious	Neutral	Deleterious	Neutral	Deleterious
M217T	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
S258A	Deleterious	Deleterious	Deleterious	Deleterious	Neutral	Neutral	Neutral	Deleterious
T281I	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
Q92E	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
A46P	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral

Sheet 2:Stability analysis.

Mutations	PredictSNP	I-Mutant	MuPro
T195K	Deleterious	increase stability	increase stability
D110N	Deleterious	**decrease stability**	**decrease stability**
Y102S	Deleterious	increase stability	**decrease stability**
I123V	Deleterious	increase stability	increase stability
G145N	Deleterious	increase stability	increase stability
For VP7			
I108T	Deleterious	increase stability	increase stability
S258A	Deleterious	**decrease stability**	**decrease stability**