scientific reports

OPEN

Novel insights of waterborne human rotavirus A in Riyadh (Saudi Arabia) involving G2 predominance and emergence of a thermotolerant sequence

Islam Nour[1], Atif Hanif[1], Ibrahim O. Alanazi[2], Ibrahim Al-Ashkar[3,4], Abdulkarim Alhetheel[5] & Saleh Eifan[1✉]

The routine evaluation of water environments is necessary to manage enteric virus-mediated fecal contamination and the possible emergence of novel variants. Here, we detected human rotavirus A (HRVA) circulating in two wastewater treatment plants, two lakes, irrigation water and a wastewater landfill located in Riyadh. VP7-derived surface protein sequences were assessed by phylogenetic analyses and inspection of thermotolerance-mediated secondary structure and seasonal variation. HRVA was most prevalent at An-Nazim wastewater landfill (AN-WWLF; 63.89%). Phylogenetic analyzes revealed the predominance of HRVA G2 lineage for the first time in Saudi Arabia. Moreover, a single HRVA sequence (2B64I-ANLF3/2018) was recovered at 45 °C from AN-WWLF; secondary structure prediction indicated that this sequence was thermotolerant with a high hydrophobicity, an absence of Ramachandran outliers, and a higher content of proline patches on the protein surface. Varied relationships were significantly observed between sampling areas influenced by temperature ranges ($p < 0.05$). HRVA prevalence was influenced by seasonal variations, favoring moderate temperatures in late autumn and early winter in all locations. However, a significant temperature impact was detected in Wadi-Hanifah Lake ($p = 0.01$). Our study extends the knowledge of currently circulating HRVA genotypes, and indicates the probable emergence of thermotolerant strains and seasonally mediated HRVA prevalence.

Wastewater (WW) is usually discharged to various surface water resources[1,2]. In addition to the negative impact of WW-mediated fecal contamination on aquatic environments, several public health issues are raised. Enteric viruses are a common cause of waterborne diseases associated with recreational water that were contaminated by the accidental release of feces or body fluids[3]. Moreover, their resistance to various environmental conditions has imposed a greater concern in terms of its durability and infectivity. Of a particular concern, rotavirus particles showed a significant tolerance to different pH, temperatures, electrolytes, etc.[4–6]. For instance, rotavirus can survive at a wide pH range (3–11) and for at least 7, 14, 22 and 32 days at 37 °C, 20 °C, 4 °C and − 20 °C, respectively with a slight loss of infectivity at 37 °C after 4 days[7,8].

Rotaviruses are major causes of acute diarrhea in infants, leading to 9% mortality in the under-5 age group globally[9]. The high rotavirus incidence rate imposes a significant burden on both low- and high-income countries[10]. Rotavirus infection results in approximately 39% of diarrheal-mediated mortality, with a larger proportion occurring in developing countries[11]; most children will have experienced at least a single episode of rotavirus infection before they are 5 years old[12]. However, mortality rate due to rotavirus

[1]Botany and Microbiology Department, College of Science, King Saud University, Riyadh, Saudi Arabia. [2]National Center for Biotechnology, King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia. [3]Biotechnology Laboratory, Plant Production Department, College of Food and Agriculture Sciences, King Saud University, Riyadh, Saudi Arabia. [4]Agronomy Department, Faculty of Agriculture, Al-Azhar University, Cairo, Egypt. [5]Department of Pathology and Laboratory Medicine, College of Medicine, King Saud University, Riyadh, Saudi Arabia. ✉email: seifan@ksu.edu.sa

infection in < 5 year-old children has declined worldwide[13]. This could be owing to natural rotavirus infection mediated acquired immunity[14] or vaccine introduction[15].

Rotavirus is a non-enveloped, double-stranded RNA virus with a triple layered icosahedral capsid[16]. Six structural viral proteins (VP1-VP4, VP6, and VP7) and six functional proteins (NSP1-NSP6) which are associated with rotavirus virulence[17]. Moreover, two surface proteins, including the glycosylated VP7 (G type) and the protease-sensitive VP4 proteins (P type), constitute the basis for dual classification-dependent rotavirus genotyping[18]. In addition, several G and P type combinations have been detected as their corresponding genes were discretely passed to progeny viruses via reassortment. So far, 36 G and 51 P types demonstrate intratypic variation[19]. The G1, G2, and G9 and the P^1, P^4, and P^8 genotypes have been frequently recorded at high prevalence in Saudi Arabia and its surrounding countries[20–23]. Moreover, fluctuations in rotavirus G-P type combinations are influenced by seasonal variations and geographical locations[19]. Interestingly, VP7 protein molecules exists on virion surface in larger concentrations of about 6.5 times than that of VP4[24]. Consequently, VP7-based monoclonal antibodies (mAbs) possess higher viral neutralizing capacity than VP4-mAbs[25]. Moreover, the genetic variation and mutation rate of VP7 gene is significantly high which even influences primers used causing genotyping failure[26]. Therefore, continuous monitoring of the circulating G lineage could enhance the vaccination program efficacy. However, few studies were specifically performed on the circulating waterborne rotaviruses in Saudi Arabia.

The current study aimed to detect the circulating human rotavirus A (HRVA) in six distinct locations in Riyadh, the capital of Saudi Arabia, including two WW treatment plants, two major lakes, irrigation water and a WW landfill. Moreover, the VP7-derived sequences from the HRVA detected were subjected to phylogenetic analysis. As Saudi Arabia is characterized by extreme heat[23], the secondary structure of the detected sequences was investigated to justify probable differences in viral tolerance to higher temperatures. Additionally, the seasonal variation was inspected in the context of HRVA incidence in these sampling areas.

Methods

Sample collection. To monitor the HRVA in different water environments in Riyadh, samples were collected from two wastewater treatment plants (WWTPs): the KSU-WWTP and Manfouhah (MAN)-WWTP (also called the Southern Plant of Riyadh Wastewater Treatment Plant), two lakes (Wadi-Hanifah and Wadi-Namar), KSU irrigation treated water (IRTW) and a WW landfill of Riyadh eastern region (An-Nazim landfill [AN-WWLF]) (Fig. S1). Moreover, KSU-WWTP-dependent irrigation water was monitored for enteric viruses. The KSU-WWTP lies to the north of Riyadh and is designed with an average capacity of 9100 m^3 day^{-1} and an average flow rate of 9000 m^3 day^{-1}. Primary treatment, involving pre-aeration, bar screen, grit chamber, comminutors, and primary sedimentation, are applied together with secondary treatment, comprising trickling filters and final sedimentation[27]. The KSU-WWTP is used for the irrigation of landscapes and for cooling purposes in the KSU power plant. The MAN-WWTP is south of Riyadh and was established with an average capacity of 200,000 m^3 day^{-1} and a maximum flow rate of 250–320 m^3 day^{-1}[27,28]. Mechanical screens, pre-aeration, and primary sedimentation are implemented for primary treatment, followed by high-rate trickling filters and humus tanks for secondary treatment, and eventually, 52 sand filters for advanced treatment. The MAN-WWTP is utilized for industrial needs, restricted irrigation, and sewer flushing. Furthermore, Wadi Hanifah lake (WHL) includes the mainstream that extends for 120 km along Riyadh from the Al Uyaynah at the north of to Al-Hair city at the south[29]. Wadi-Namar lake (WNL) is one of the secondary valleys that branch from Wadi-Hanifah and includes the drainage of the Riyadh western area. This was used previously as a water source, but is currently used for partial WW treatment[30]. Water samples (1L each) were collected using 2-L low density polyethylene bottle (BRAND, Merck KGaA, Darmstadt, Germany) from WWTPs effluents, mid-stream of lakes (at a depth of about 30 cm), landscape flushing system endpoints (IRTW), WW landfill drainage endpoints (AN-WWLF). Water bottles were stored in a cooler box and transferred to the lab for processing within 2 h to avoid samples distortion and incidence of false negatives. Water samples were collected thrice every month for 1 year from April 2018 to March 2019 from each sampling area to check the seasonal influence on the HRVA prevalence. Moreover, both high and low temperatures were recorded at each collection day, obtained from accuweather website (https://www.accuweather.com/en/sa/riyadh). Water pH was measured by the electrode probe method using Orion Star A111 pH meter (Thermo Scientific Co., USA) and adjusted to neutral pH prior to viral concentration.

Viral concentration. Rotavirus A from water samples was recovered using the electronegative-charged membrane-based adsorption/elution concentration method with a recovery efficiency of 56%[31]. To begin with, a HA filter of 0.45 μm pore size (Sigma-Aldrich Co., St. Louis, MO, USA) complying with U.S. EPA Standard methods was used to filter 5-mL aliquots of 0.25 M $AlCl_3$. Each water sample (1 L) was then filtered through the Al^{3+}-coated filter. The filter membrane was then flushed with 200 mL H_2SO_4 (0.5 mM) at pH 3. Subsequently, rotavirus A was eluted with 10 mL NaOH (1 mM). The eluent was neutralized with 100 μL equal volumes of H_2SO_4 (50 mM) and 100 × Tris–EDTA buffer (pH 8.0). The final

volume reduction to 700 μL was performed using Amicon Ultra-15 (Merck Millipore Ltd., Carrigtwohill Co., Cork, Ireland) for achieving > 10^3 ×, followed by storage at − 80 °C.

Nucleic acid extraction and RT-PCR. HRVA RNA was extracted from 200 μL of the viral concentrates using a ZymoBIOMICS RNA Mini Kit (Zymo Research, Irvine, CA 92614, USA) following the manufacturer's instructions. HRVA RNA was reverse transcribed using Sensiscript Reverse Transcription (RT) Kit (SensRT; Qiagen GmbH, Hilden, Germany). RT was performed in a 20-μL reaction mixture consisting of 2 μL template RNA and 18 μL RT mixture containing 2 μL 10 × RT buffer, 2 μL dNTPs, 2 μL random hexamers (final concentration of 10 μM), 1 μL Sensiscript RT, 1 μL RNase inhibitor (40 U μL^{-1}), and 10 μL nuclease-free water. The RT mixture was incubated at 37 °C for 1 h. RT was followed by polymerase chain reaction (PCR) using the 2 × Phusion Master Mix with GC buffer (Thermo Scientific, New England Biolabs, Ipswich, MA 01,938, USA). PCR was performed in a 20-μL reaction mixture consisting of 2 μL cDNA template, RComb-F: 5′-CCACAAYTDTTGTGATTA-3′, RComb-R 5′-CCCATYGTATCC AYTTATT-3′ (500 nM each) (designed for targeting 173 bp, located at 313–485 nucleotides in VP7, Reference strain: BLU5, Accession number: LC500701), and the 1 × Phusion Master Mix under the following reaction conditions: 98 °C for 30 s, followed by 40 cycles of 98 °C for 10 s, 50 °C for 30 s, and 72 °C for 30 s each, and a final extension at 72 °C for 5 min. A HRVA-positive control was obtained from the Virology Unit, King Khalid University Hospital, Riyadh, Saudi Arabia.

Amplicon purification and sequencing. PCR products were visualized by 2 × agarose gel electrophoresis and 173-bp amplicons were excised and purified using the Wizard SV Gel and PCR Clean-Up System (Promega Co., Madison, WI, USA), following the manufacturer's instructions. The purified amplicons were sequenced in both forward and reverse directions (3X) using the BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems, Foster City, California, USA) with individual PCR primers yielding the second round PCR product. Sequences were obtained using an ABI genetic analyzer 3130Xl (Applied Biosystems, Carlsbad, California 92008. USA).

Phylogenetic analysis. The sequences obtained were cleaned by removing the overlapping sequences after comparison with the sequences obtained from paired-end sequencing using Bioedit v7.2 (Nucleics Co., Sydney, Australia). To compare nucleotide sequences obtained from water samples, phylogenetic analyses were performed for the resulting HRVA sequences against the HRVA G1 lineage represented by G1P[8], G9 lineage, including G9P[8], G2 lineage, including G2P[4] and G2P[8], as the later genotypes were reported as highly prevalent in Saudi Arabia and nearby countries[20–23]. The G12 lineage was designated as the outgroup. The sequence alignments were generated using the multiple sequence alignment tool ClustalW in MEGA X software[32], with an opening penalty of 15 and an extension penalty of 6.66. Subsequently, the aligned sequences were assessed for the best fitting substitution model to construct the phylogenetic tree on analysis. The models with the lowest Bayesian information criterion were considered as the best depiction of the substitution pattern. Moreover, the Akaike information criterion, corrected value, maximum likelihood value (*lnL*), and the count of parameters used (including branch lengths) were considered for each model[33]. The non-uniformity of evolutionary rates among sites was modeled by applying a discrete Gamma distribution (+ *G*) with five rate categories and assuming that a certain fraction of sites is evolutionarily invariable (+ *I*). Furthermore, tree topology was automatically computed to estimate ML values. A phylogenetic tree was constructed using aligned sequences by the neighbor-joining method implementing 1000 bootstrap replicates using MEGA X software[32]. This analysis involved 94 nucleotide sequences. The genetic distances were estimated by the Kimura three-parameter method.

Secondary structure investigation. The influence of the secondary structure of the detected HRVA sequence (located at amino acids 105–161 of VP7 of the reference strain BLU5, Accession number: BBN91702) on thermotolerance, including any possible deviation from mesophilic-detected HRVA isolates, was assessed using SWISS-MODEL (https://swissmodel.expasy.org/)[34]. Moreover, the detection and orientation of distinctive residues and the abundance of possible residual outliers were investigated using Ramachandran plotting and MolProbity[35]. The sequences of 2B64I-ANLF3/2018 and 2B64I-ANLF5/2018 (target) were modeled against the structure of the VP7 outer layer protein using an atomic model of an infectious rotavirus particle (protein databank [PDB]:4V7Q.1.U) as a template. Furthermore, the models obtained for both 2B64I-ANLF3/2018 and 2B64I-ANLF5/2018 were used for alignment and comparison via the CLC Main Workbench V20.0 (Qiagen).

Statistical analyses. Pearson's correlation coefficient matrix was implemented to define the probable relationship between the sampling areas over the 1-year period. Moreover, the influence of seasonal variation, denoted by a wide range of high and low temperatures, on HRV detection was examined. One-way analysis of variance was performed to test the influence and significance of high- and low-temperature ranges on HRVA incidence, regardless of the sampling area. The relationships between different sampling areas (as dependent variables) and high and low temperatures (as independent variables) were fitted using

Sampling area	HRVA + ve 2018	HRVA + ve 2019	HRVA prevalence (%)
KSU-WWTP	3	2	13.89
MAN-WWTP	2	0	5.56
WHL	7	3	27.78
WNL	8	3	30.56
AN-WWLF	16	7	63.89
IRTW	1	1	5.56

Table 1. HRVA prevalence in different sample locations.

linear curve-fitting. All statistical analyses were conducted using the XL-STAT statistical package software (Ver. 2019, Excel Add-ins soft SARL, New York, NY, USA).

Results

AN-WWLF prevalence over other sampling areas.   Out of 216 water samples tested for HRVA, 53 (24.5%) were positive as denoted by the detection of the 173-bp amplicon. The highest HRVA prevalence (approximately 64%) was observed in the AN-WWLF, whereas the lowest prevalence occurred in the MAN-WWTP and IRTW (approximately 6%) (Table 1). The molecular characterization of amplicon sequences by Sanger sequencing showed the presence of 35 sequences; however, 18 sequences were not characterized owing to overlapped electropherograms, indicating the presence of more than one HRVA type.

Predominance of lineage G2 of HRVA.   The phylogenetic tree displayed a clear relationship between these HRVA sequences and the G2 lineage rather than the G1 and G9 lineages, as the current study sequences were entirely in the G2 lineage (Fig. 1). Pairwise distancing uncovered the presence of eight HRVA isolates of null distance to other isolates and/or strains (Tables 2 and S1). Initially, IRTW-derived HRVA sequences were related closely to HRVA sequences isolated from the KSU-WWTP obtained in 2019 (2B64-KSU1). However, another KSU-WWTP-derived HRVA sequence detected in 2018 (2B64-KSU1) was related to an HRVA Canadian strain (RT125-07). Moreover, sequences recovered from the MAN-WWTP were more linked to each other than to isolates from other sources. Remarkably, the HRVA sequence obtained from the KSU-WWTP (2B64I-KSU2/2018) was typically related to several genotypes G2P[4] and G2P[8] from Asia, Australia, and Europe. However, another AN-WWLF isolate (2B64-ANLF1/2019) was more related to isolates obtained from WHL, WNL, and AN-WWLF rather than to strains from other countries. Moreover, the constructed tree depicted two closely related AN-WWLF sequences (2B64-ANLF5/2018 and 2B64-ANLF6/2018), which were relatively distant to another cluster of AN-WWLF sequences (2B64-ANLF1–4 in 2018). Interestingly, the latter cluster comprised a sequence of an HRVA isolate (2B64-ANLF3/2018) recovered at a relatively high temperature (45 °C), unlike the HRVA sequences (2B64-ANLF5/2018 and 2B64-ANLF6/2018) in the former branch obtained at lower temperatures (22 °C and 23 °C, respectively). Therefore, the secondary structure of 2B64-ANLF3/2018 was compared with those obtained at the lowest temperature because of the absence of any sequence obtained from other sampling areas at 45 °C.

Thermotolerance indicated by predicted secondary structure hydrophobicity and fewer Ramachandran outliers of HRVA 2B64I-ANLF3/2018 isolate.   The alignment of amino acid sequences identified a shared identity for both 2B64I-ANLF5/2018 and 2B64I-ANLF6/2018; therefore, 2B64I-ANLF5/2018 was used as a representative for further analyses. Initially, the alignment of the predicted amino acids sequences showed the substitution of valine (a non-polar amino acid) in 2B64I-ANLF3/2018 by lysine (a polar amino acid) in 2B64I-ANLF5/2018 at codon 34 (Fig. 2). Moreover, the proline content of 2B64I-ANLF3/2018 (Movie 1) was higher than that of 2B64I-ANLF5/2018 (Movie 2) that patched over the external surface, mediating structural stability. Furthermore, outliers were absent in the Ramachandran plot for 2B64I-ANLF3/2018, and 94.44% of the constructed structure was in the favorable region of the plot, whereas 2B64I-ANLF5/2018 had 1.85% outliers and 88.89% of the structure in the favored region (Fig. 3). In addition, a three-dimensional secondary structure alignment showed that the substitution of valine and lysine in 2B64I-ANLF3/2018 with leucine and serine residues in 2B64I-ANLF5/2018 (Fig. 2) resulted in an initial shift in the β-sheet at codon 9, which then affected the structural alignment between codons 29 and 33 and a looped distance at codon 49 (Fig. S2). However, the substitution of glycine (2B64I-ANLF3/2018) by glutamic acid (2B64I-ANLF5/2018) at the base of an α-helix had no effect on the structural alignment. In addition, both sequences contained an N-glycosylation site (NTS**G** for 2B64I-ANLF3/2018 and NTS**E** for 2B64I-ANLF3/2018; however, 2B64I-ANLF5/2018 was shown to possess an extra protein kinase C phosphorylation (PKCP) site (SNK) (Fig. S3). This combina-

tion of structural features indicates that this isolate can tolerate higher temperatures, unlike other HRVA isolates. Moreover, this sequence was not detected in any other sampling area. Therefore, the relationship between the sampling area and the impact of seasonal variation on HRVA incidence was further investigated.

Relationship differences between sampling areas. The relationships detected between the studied sampling areas at high temperatures (the numbers above the gray highlight, Table 3) differed from those obtained at low temperatures (the numbers below the gray highlight, Table 3). At high-temperature range, IRTW was of the highest significant correlation ($p = 0.012$) to the KSU-WWTP, followed by a slightly lower correlation of AN-WWLF to WHL and WNL ($p = 0.02$ and 0.021, respectively) and WNL to WHL ($p = 0.037$) in terms of the % HRVA detection. On the contrary, at the low-temperature range, IRTW was of the highest correlation to AN-WWLF ($p = 0.002$) followed by a significant correlation of WHL to AN-WWLF, IRTW, and WNL ($p = 0.003$, 0.005, and 0.014, respectively) and KSU-WWTP to WNL and WHL ($p = 0.006$ and 0.013, respectively).

Seasonal variation influenced HRVA incidence. Regardless of the sampling area, the incidence of HRVA was significantly affected by temperature variations ($p = 0.03$), with a higher prevalence at lower temperatures (Fig. 3). However, when the sampling areas were segregated, HRVA was detected mostly at moderate temperature ranges (26 °C–29 °C) at five sampling areas in the high-temperature range, unlike WNL that favored a lower temperature range (22 °C–25 °C) (Fig. 4). HRVA detection followed a similar pattern at the low-temperature range of 14 °C–17 °C, which roughly corresponded to that at 26 °C–29 °C in the high-temperature range, thus favoring late autumn and early winter seasons (Fig. S3). Strikingly, HRVA was detected in the AN-WWLF at both high and low temperatures, unlike the other sampling areas. Despite the existence of clustered HRVA at relatively low temperatures in all sampling areas, a significant influence of the high-temperature range on HRVA existence was detected only at WHL ($p = 0.01$) (Table 4).

Discussion

Enteric viruses, particularly the non-enveloped viruses, are highly stable in water environments and fomite[36], leading to potential viral dissemination[37]. Likewise, the current study addressed the detection of HRVA in WWTPs, surface waters, and a landfill. HRVA prevalence in AN-WWLP was similar to the previously reported rotavirus prevalence of 65% at the Al-Misk Lake in Jeddah, Saudi Arabia[38]. However, we detected a higher prevalence of HRVA in both lakes (around 27–31%) than that recovered from the Biljurashi dams and Mikhwah sites (~ 13%) in the Albaha region, Saudi Arabia[39]. The current higher prevalence of HRVA might be owing to the 1-year sampling plan in comparison with the 3-month period of the earlier study[39]. HRVA prevalence in the MAN-WWTP was less than that observed in the KSU-WWTP, which may be because of the additional advanced treatment phase following the secondary water treatment that is absent in the KSU-WWTP[27].

Phylogenetic analyses of the HRVA detected in this study lay entirely within the G2 lineage. However, as reported by several studies in Saudi Arabia[22,23,40], Lebanon[21], Bahrain[41], Kuwait[42], and Oman[43], the predominant lineage of HRVA was G1, whereas G3 was the most prevalent genotype in Qatar[19]. A recent study highlighted a surge in G2 incidence in Saudi Arabia, particularly the G2P[4] genotype, occurring after vaccination, which may partially justify our results[39]. Thus, vaccination-mediated selection pressure may have favored the prevalence of G2 rather than G1 strains of the HRVA. However, the discrepancy with the Qatar findings could be because of the contribution of expats (~ 80%) of the community structure, leading to the dominance of the G3 genotype[44,45]. Moreover, the close relationship between sequences isolated from IRTW (2B64I-IRTW1/2018 and 2B64I-IRTW1/2019) and KSU-WWTP (2B64I-KSU1/2019) was found due to the fact that the source of irrigation water belongs to the KSU-WWTP effluent. Strikingly, two closely related AN-WWLF sequences (2B64-ANLF5/2018 and 2B64-ANLF6/2018) were found distantly related to the 2B64-ANLF3/2018 sequence, which belonged to the same source; however, the latter sequence was recovered at a relatively high temperature, thus postulating the existence of possible differences between secondary structure of these sequences.

The current study detected higher hydrophobicity in the 2B64-ANLF3/2018 sequence in comparison to low-temperature-prevailing sequences and this might have enhanced high-temperature resistance, as described elsewhere[46]. In addition, the molecular thermodynamics of thermophilic proteins is enhanced by the increased use of small non-polar amino acids (e.g., valine) in addition to greater hydrophobicity, which is in agreement with our study findings[47]. Moreover, a higher proline content was detected in the higher-temperature-prevailing sequence (Movie 1) and led to improved structural tolerance to such harsh conditions. Proline introduction resulted in reduction of steric hindrance and the unfolded state entropy besides its side chain cyclic structure that supported conformational rigidity required for thermotolerance[48,49]. Furthermore, thermophilic proteins were reported to have fewer average residues in Ramachandran plot with more condensed residues in the β-sheet and right- and left- handed helix core areas than those of a mesophilic counterpart, which is in line with our results[50]. Moreover, the N-glycosylation motif was detected in both sequences and was implicated in conferring protein stability and thermotolerance for

Figure 1. Phylogenetic tree for the HRVA VP7-derived sequences constructed by the maximum likelihood method and Tamura three-parameter model. The highest log likelihood tree is displayed (− 898.28). The percentage of associated taxa that are clustered together are provided at each branch. Heuristic search-dependent initial trees were produced automatically via Neighbor-Join and BioNJ algorithms applied to the pairwise distances matrix and assessed by the maximum composite likelihood (MCL) approach, followed by the highest log likelihood-resulted topology selection. The rate variation model was permitted to be evolutionarily invariable for several sites ([+ I], 44.78% sites), according to the best fitting substitution model validation (Table S2). The horizontal distance connecting two HRVA sequences is proportional to the genetic distance between these two HRVA sequences. The distance is expressed as the number of the nucleotide substitutions per site. Strain CNMC122 (G12P[8]) was used as an outgroup. Accession numbers of sequences used for phylogenetic analysis are displayed in Table S3.

many viruses, including the severe acute respiratory syndrome coronavirus 2[51,52], influenza A virus[53], and West Nile virus[54]. However, 2B64-ANLF3/2018 contains a glycine in the N-glycosylation motif (NTSG) instead of the glutamic acid usually favored in thermophilic proteins in the helix structure (Fig. S2)[50,55]. Moreover, phosphorylation at PKCP motif declined along with temperature elevation, as recorded for the brome mosaic virus[56] indicating sensitivity of PKCP motif to high temperature which approves PKCP motif presence in the low-temperature favoring sequence.

In contrast, this thermotolerant sequence was absent in sequences isolated from other locations; thus, varied relationships between sampling areas were found. KSU-WWTP is the treated water source of IRTW and this imposed the greatest relationship between them. However, at the low-temperature range, IRTW was highly correlated to AN-WWLF, which could be because of the discharge of agricultural wastes to the AN-WWLF, although this reason is still doubtful because of the approximately 25-km distance between these sites that would make HRVA vulnerable and less likely to persist. Moreover, the significant relationship between the lakes is because of their geographical nature, as WNL is a subsidiary branch of WHL[30].

Seasonal variation was found to influence the prevalence of HRVA, indicating that there was a higher prevalence at lower temperatures, which agrees with results of various studies that favored its prevalence in winter months[22,57,58]. Interestingly, significant HRVA prevalence was noted at the lower temperature range at WNL, which could be attributed to uncovered surface water exposed to various limiting conditions like direct sunlight, UV light, and high temperatures[59]. Moreover, the stable pattern of HRVA prevalence in AN-WWLF might be owing to HRVA sequence nature, fecal-mediated contamination from livestock, or contamination from vaccine strains[60]. Feces from vaccinated children younger than 6-month old were shown to boost HRVA prevalence in surface water[61].

The current study provides knowledge regarding HRVA prevalence and seasonal influences in different water environments in Riyadh. The dominance of the G2 lineage was reported for the first time in Saudi Arabia. The emergence of an unexpectedly thermotolerant HRVA sequence at 45 °C demands the extensive monitoring of circulating strains in the future.
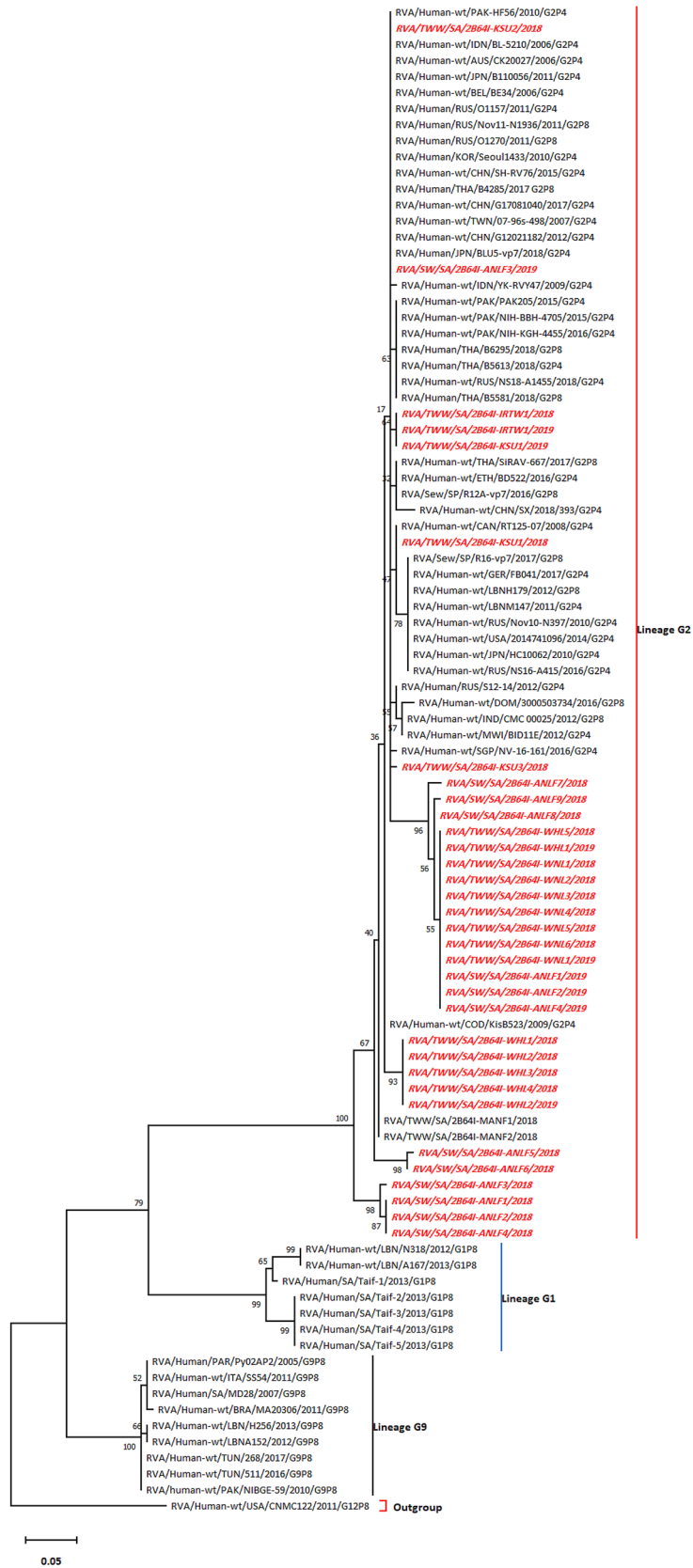
Studied sequence	Similar sequence(s)	Country of origin*
2B64I-KSU1/2018	RT125-07/2008/G2P4	Canada
2B64I-MANF1/2018	2B64I-MANF2/2018	Saudi Arabia
2B64I-WHL1/2018	2B64I-WHL2/2018	Saudi Arabia
	2B64I-WHL3/2018	
	2B64I-WHL4/2018	
2B64I-WHL1/2019	2B64I-WHL2/2018	
	2B64I-WHL3/2018	
	2B64I-WHL4/2018	
2B64I-ANLF1/2019	2B64I-WHL1/2019	Saudi Arabia
	2B64I-WHL5/2018	
	2B64I-WNL1/2018	
	2B64I-WNL2/2018	
	2B64I-WNL3/2018	
	2B64I-WNL4/2018	
	2B64I-WNL5/2018	
	2B64I-WNL6/2018	
	2B64I-WNL1/2019	
	2B64I-ANLF2/2019	
	2B64I-ANLF4/2019	
2B64I-ANLF1/2018	2B64I-ANLF2/2018	Saudi Arabia
	2B64I-ANLF4/2018	
2B64I-KSU2/2018	BLU5-vp7/2018/G2P4	Japan
	G12021182/2012/G2P4	China
	07-96s-498/2007/G2P4	Taiwan
	G17081040/2017/G2P4	China
	B4285/2017/G2P8	Thailand
	SH-RV76/2015/G2P4	China
	Seoul1433/2010/G2P4	South Korea
	O1270/2011/G2P8	Russia
	Nov11-N1936/2011/G2P8	Russia
	O1157/2011/G2P4	Russia
	BE34/2006/G2P4	Belgium
	B110056/2011/G2P4	Japan
	CK20027/2006/G2P4	Australia
	BL-5210/2006/G2P4	Indonesia
	PAK-HF56/2010/G2P4	Pakistan
	2B64I-ANLF3/2019	Saudi Arabia
2B64I-IRTW1/2018	2B64I-KSU1/2019	Saudi Arabia
	2B64I-IRTW1/2019	

Table 2. HRVA sequences grouping (Grouping is based on zero-distance between studied sequences as well as extraneously with other strains used for phylogenetic analysis implemented by pairwise distancing using Kimura three-parameter method). *Country of origin of similar sequences.
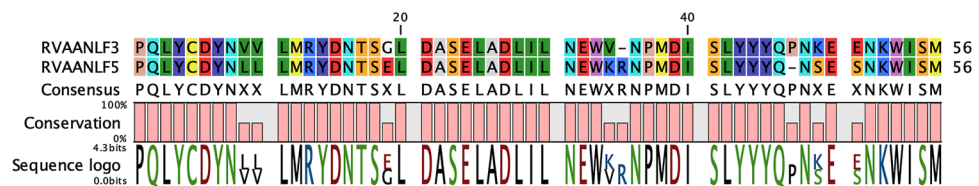

Figure 2. Amino acid sequence alignment of 2B64I-ANLF3/2018 (designated RVAANLF3) and 2B64I-ANLF5/2018 (designated RVAANLF5).
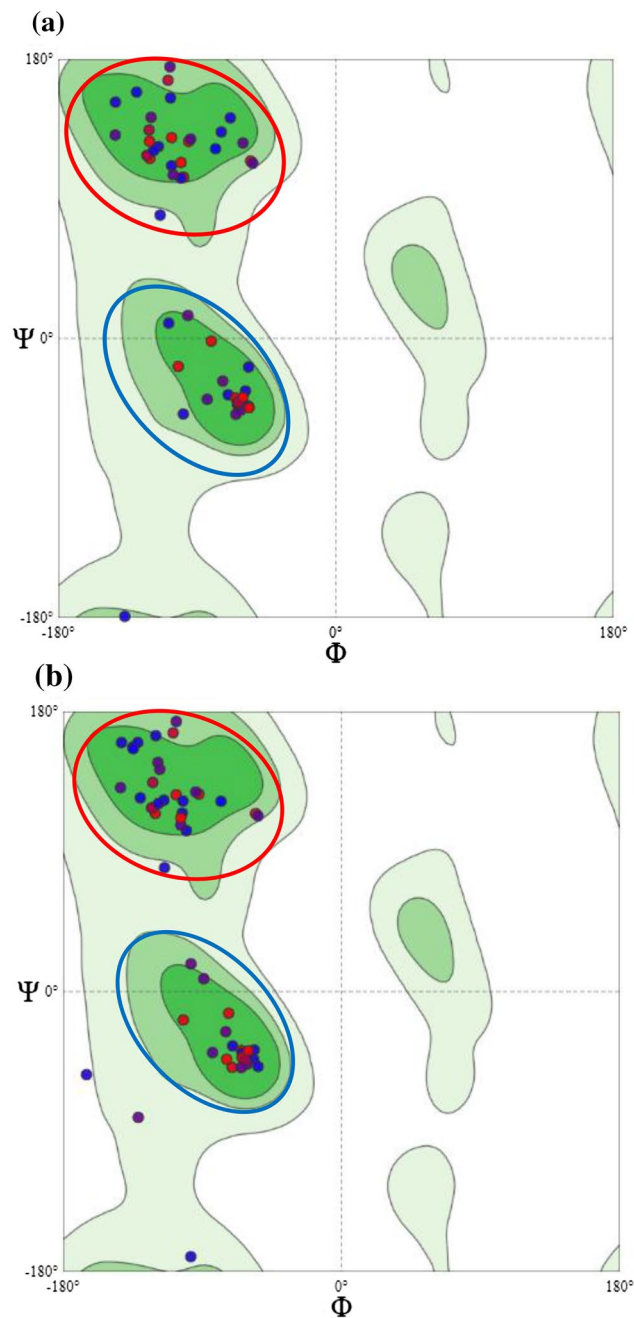
8

(a)

(b)

Figure 3. Ramachandran plot showing hydrophobic residues for (**a**) 2B64I-ANLF3/2018 and (**b**) 2B64I-ANLF5/2018. Ramachandran plot was obtained by using SWISS-MODEL (https://swissmodel.expasy.org/). The red circle refers to hydrophobic residues in the β-sheet. The blue circle denotes the hydrophobic residues in right-handed helix. Residues outside the circles represent residual outliers.

	% Det$_{KSU-WWTP}$	% Det$_{MAN-WWTP}$	% Det$_{WHL}$	% Det$_{WNL}$	% Det$_{AN-WWLF}$	%Det$_{IRTW}$
% Det$_{KSU-WWTP}$		0.773	0.796	0.627	0.637	**0.908***
% Det$_{MAN-WWTP}$	0.686		0.347	0.423	0.196	0.632
% Det$_{WHL}$	**0.905**	0.631		**0.838**	**0.882**	0.614
% Det$_{WNL}$	**0.935**	0.466	**0.900**		**0.881**	0.267
% Det$_{AN-WWLF}$	0.751	0.537	**0.957**	0.786		0.402
% Det$_{IRTW}$	0.759	0.632	**0.943**	0.799	**0.967**	

Table 3. Pearson's correlation matrix of HRVA detection percentage at the various sampling areas. The numbers above the grey-highlighted diagonal refers to correlation values at high-temperature range, whereas correlation values at low-temperature range are below the diagonal. % Det denotes HRVA detectability percentage at different sampling areas, for instance % Det$_{KSU-WWTP}$ refers to HRVA detection percentage at KSU-WWTP. *Significant correlation values are displayed as bold numbers.
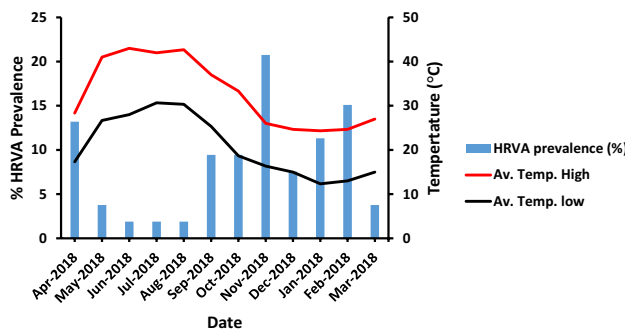

Figure 4. Temperature variation influence on the HRVA prevalence in water samples. "Av. Temp. High" refers to the average high temperature and "Av. Temp. Low" refers to the average low temperature.

Sampling area	Temperature range	R²	RMSE	Equation
KSU-WWTP	High	0.353	21.022	% Prev$_{KSU-WWTP}$ = 74.24−1.86* T$_H$‡
	Low	0.61	10.511	% Prev$_{KSU-WWTP}$ = 46.52−1.57* T$_L$
MAN-WWTP	High	0.043	28.242	% Prev$_{MAN-WWTP}$ = 38.81−0.71* T$_H$
	Low	0.043	28.242	% Prev$_{MAN-WWTP}$ = 30.24−0.71* T$_L$
WHL	High	**0.844***	8.223	% Prev$_{WHL}$ = 87.52−2.29* T$_H$
	Low	0.442	18.797	% Prev$_{WHL}$ = 54.67−2* T$_L$
WNL	High	0.601	13.715	% Prev$_{WNL}$ = 79.07−2.01* T$_H$
	Low	0.553	13.189	% Prev$_{WNL}$ = 49.98−1.75* T$_L$
AN-WWLF	High	0.502	9.051	% Prev$_{AN-WWLF}$ = 50.36−1.09* T$_H$
	Low	0.281	12.88	% Prev$_{AN-WWLF}$ = 34.96−0.96* T$_L$
IRTW	High	0.154	41.975	% Prev$_{IRTW}$ = 83.096−2.14* T$_H$
	Low	0.154	41.975	% Prev$_{IRTW}$ = 57.38−2.14* T$_L$

Table 4. The significance of impact of high or low-temperature ranges on HRVA prevalence in various sampling areas. % Prev refers to HAdV prevalence percentage at different sampling areas. RMSE denotes the root mean squared error which is an absolute measure of fit. *Significant at $p < 0.05$. ‡T$_H$ represents the highest temperature detected whereas T$_L$ denoted lowest temperature detected.

References

1. Braeye, T., De Schrijver, K., Wollants, E., Van Ranst, M. & Verhaegen, J. A large community outbreak of gastroenteritis associated with consumption of drinking water contaminated by river water, Belgium, 2010. *Epidemiol. Infect.* **143**, 711–719 (2015).
2. Mellou, K. *et al.* A large waterborne gastroenteritis outbreak in central Greece, March 2012: Challenges for the investigation and management. *Epidemiol. Infect.* **142**, 40–50 (2014).

3. Bonadonna, L. & La Rosa, G. A review and update on waterborne viral diseases associated with swimming pools. *Int. J. Environ. Res. Public Health.* **16**, 166 (2019).
4. Alidjinou, E. K., Sane, F., Firquet, S., Lobert, P. E. & Hober, D. Resistance of enteric viruses on fomites. *Intervirology* **61**, 205–213 (2018).
5. Desselberger, U. Rotaviruses. *Virus Res.* **22**, 75–96 (2014).
6. Estes, M. K., Graham, D. Y., Smith, E. M. & Gerba, C. P. Rotavirus stability and inactivation. *J. Gen. Virol.* **43**, 403–409 (1979).
7. Pinon, A. & Vialette, M. Survival of viruses in water. *Intervirology* **61**, 214–222 (2018).
8. Meng, Z. D., Birch, C., Heath, R. & Gust, I. Physicochemical stability and inactivation of human and simian rotaviruses. *Appl. Environ. Microbiol.* **53**, 727–730 (1987).
9. Goel, A. K., Chawla, S., Dhingra, A., Thiyagarajan, V. & Nair, N. P. Rotavirus diarrhea and its determinants among under-five children admitted in a tertiary care hospital of Southern Haryana, India. *Indian J. Pediatr.* **88**, 16–21 (2021).
10. Dbaibo, G. *et al.* Hospital-based surveillance study of rotavirus gastroenteritis in children under 5 years of age in Lebanon. *Trials Vaccinol.* **2**, 25–30 (2013).
11. World Health Organization. World health statistics 2018: monitoring health for the SDGs. Geneva: WHO. http://apps.who.int/iris/bitstream/handle/10665/272596/9789241565585-eng.pdf?ua=1 (2018).
12. Grimwood, K., Lambert, S. B. & Milne, R. J. Rotavirus infections and vaccines. *Paediatr. Drugs* **12**, 235–256 (2010).
13. Tate, J. E. *et al.* Global, regional, and national estimates of rotavirus mortality in children < 5 years of age. *Clin. Infect. Dis.* **62**, 96–105 (2016).
14. Velázquez, F. R. *et al.* Rotavirus infection in infants as protection against subsequent infections. *N. Engl. J. Med.* **335**, 1022–1028 (1996).
15. Troeger, C. *et al.* Rotavirus vaccination and the global burden of rotavirus diarrhea among children younger than 5 years. *JAMA Pediatr.* **172**, 958–965 (2018).
16. Desselberger, U. Rotaviruses. *Virus Res.* **190**, 75–96 (2014).
17. Estes, M. K. & Greenberg, H. B. Rotaviruses. In *Fields Virology (6th ed)* (eds Knipe, D. M., Howley, P. M. *et al.*) 1347–1401 (Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia, 2013).
18. O'Ryan, M. The ever-changing landscape of rotavirus serotypes. *Pediatr. Infect. Dis. J.* **28**, S60–S62 (2009).
19. Mathew, S., Al Ansari, K., Al Thani, A. A., Zaraket, H. & Yassine, H. M. Epidemiological, molecular, and clinical features of rotavirus infections among pediatrics in Qatar. *Eur. J. Clin. Microbiol. Infect. Dis.* **1**, 1–14 (2021).
20. Almalki, S. S. Circulating rotavirus G and P strains post rotavirus vaccination in Eastern Mediterranean Region. *Saudi Med. J.* **39**, 755 (2018).
21. Ali, Z. *et al.* Rotavirus genotypes and vaccine effectiveness from a sentinel, hospital-based, surveillance study for three consecutive rotavirus seasons in Lebanon. *PLoS ONE* **11**, e0161345 (2016).
22. Aly, M., Al Khairy, A., Al Johani, S. & Balkhy, H. Unusual rotavirus genotypes among children with acute diarrhea in Saudi Arabia. *BMC Infect. Dis.* **15**, 1–7 (2015).
23. Kheyami, A. M. *et al.* Molecular epidemiology of rotavirus diarrhea among children in Saudi Arabia: First detection of G9 and G12 strains. *J. Clin. Microbiol.* **46**, 1185–1191 (2008).
24. Ghosh, A., Chattopadhyay, S., Chawla-Sarkar, M., Nandy, P. & Nandy, A. In silico study of rotavirus VP7 surface accessible conserved regions for antiviral drug/vaccine design. *PLoS ONE* **7**, e40749 (2012).
25. Ruggeri, F. M. & Greenberg, H. B. Antibodies to the trypsin cleavage peptide VP8 neutralize rotavirus by inhibiting binding of virions to target cells in culture. *J. Virol.* **65**, 2211–2219 (1991).
26. Rahman, M. *et al.* Typing of human rotaviruses: Nucleotide mismatches between the VP7 gene and primer are associated with genotyping failure. *Virol. J.* **2**, 1–5 (2005).
27. Al-Jasser, A. O. Saudi wastewater reuse standards for agricultural irrigation: Riyadh treatment plants effluent compliance. *J. King Saud Univ. Eng. Sci.* **23**, 1–8 (2011).
28. Salman, A. S., Al Dosari, A., Zaidi, F. K. & Loni, O. A. Environmental assessment of soil, groundwater, and surface water quality in the south of the Riyadh, Saudi Arabia. *Arab. J. Geosci.* **10**, 490 (2017).
29. Al-Othman, A. A. Evaluation of the suitability of surface water from Riyadh Mainstream Saudi Arabia for a variety of uses. *Arab. J. Chem.* **12**, 2104–2110 (2019).
30. Mahboob, S. *et al.* Biomarkers of oxidative stress as indicators of water pollution in Nile tilapia (*Oreochromis niloticus*) from a water reservoir in Riyadh, Saudi Arabia. *Toxicol. Environ. Chem.* **96**, 624–632 (2014).
31. Sibanda, T. & Okoh, A. I. Real-time PCR quantitative assessment of hepatitis A virus, rotaviruses and enteroviruses in the Tyume River located in the Eastern Cape Province, South Africa. *Water SA* **39**, 295–304 (2013).
32. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
33. Nei, M. & Kumar, S. *Molecular Evolution and Phylogenetics* (Oxford University Press, 2000).
34. Waterhouse, A. *et al.* SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
35. Chen, V. B. *et al.* MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
36. Seitz, S. R. *et al.* Norovirus infectivity in humans and persistence in water. *Appl. Environ. Microbiol.* **77**, 6884–6888 (2011).
37. Geoghegan, J. L., Senior, A. M., Di Giallonardo, F. & Holmes, E. C. Virological factors that increase the transmissibility of emerging human viruses. *Proc. Natl. Acad. Sci.* **113**, 4170–4175 (2016).
38. Redwan, N. & Attar, R. Rotavirus infection and its monitoring in waste water using RT-PCR in Jeddah, Saudi Arabia. *Int. Res. J. Microbiol.* **3**, 94–100 (2012).
39. Almalki, S. S. Molecular detection of hepatitis a virus and rotavirus in water samples collected from Albaha, Saudi Arabia, Egypt. *Acad. J. Biol. Sci. C.* **10**, 59–68 (2018). https://doi.org/10.21608/EAJBSC.2018.13656 (2018).
40. Abdel-Moneim, A. S. *et al.* Sequence diversity of VP4 and VP7 genes of human rotavirus strains in Saudi Arabia. *Foodborne Pathog. Dis.* **12**, 937–944 (2015).
41. Al Musawi, M., Zainaldeen, H., Shafi, F., Anis, S. & DeAntonio, R. Rotavirus gastroenteritis in children under 5 years in the Kingdom of Bahrain: Hospital-based surveillance. *Clin. Epidemiol.* **5**, 269 (2013).
42. Marmash, R. W. *et al.* Genotypic characterization of rotaviruses and prevalence of serotype-specific serum antibodies in children in Kuwait. *Epidemiol. Infect.* **135**, 1331–1337 (2007).
43. Al-Awaidy, S. A. *et al.* Considerations for introduction of a rotavirus vaccine in Oman: Rotavirus disease and economic burden. *J. Infect. Dis.* **200**, S248–S253 (2009).
44. Mazid, R. *et al.* Resurgence and predominance of G3P [8] human rotaviruses in north-central Bangladesh, 2018–2019. *New Microb. New Infect.* **33**, 100621. https://doi.org/10.1016/j.nmni.2019.100621 (2020).
45. Goodman, A. The development of the Qatar healthcare system: A review of the literature. *Int. J. Clin. Med.* **6**, 177 (2015).
46. Zhou, X. X., Wang, Y. B., Pan, Y. J. & Li, W. F. Differences in amino acids composition and coupling patterns between mesophilic and thermophilic proteins. *Amino Acids* **34**, 25–33 (2008).
47. Sawle, L. & Ghosh, K. How do thermophilic proteins and proteomes withstand high temperature?. *Biophys. J.* **101**, 217–227 (2011).

48. Hait, S., Mallik, S., Basu, S. & Kundu, S. Finding the generalized molecular principles of protein thermal stability. *Proteins* **88**, 788–808 (2020).
49. Barik, S. Evolution of protein structure and stability in global warming. *Int. J. Mol. Sci.* **21**, 9662 (2020).
50. Panja, A. S., Bandopadhyay, B. & Maiti, S. Protein thermostability is owing to their preferences to non-polar smaller volume amino acids, variations in residual physico-chemical properties and more salt-bridges. *PLoS ONE* **10**, e0131495 (2015).
51. Rath, S. L. & Kumar, K. Investigation of the effect of temperature on the structure of SARS-Cov-2 spike protein by molecular dynamics simulations. *Front. Mol. Biosci.* **7**, 523 (2020).
52. Nour, I., Alanazi, I. O., Hanif, A. & Eifan, S. Molecular adaptive evolution of SARS-COV-2 spike protein in Saudi Arabia. *Saudi J. Biol. Sci.* **28**(6), 3325–3332 (2021).
53. Wu, C. Y. *et al.* Influenza A surface glycosylation and vaccine design. *Proc. Natl. Acad. Sci.* **114**, 280–285 (2017).
54. Maharaj, P. D. *et al.* N-linked glycosylation of the West Nile virus envelope protein is not a requisite for avian virulence or vector competence. *PLoS Negl. Trop. Dis.* **13**, e0007473 (2019).
55. Warren, G. L. & Petsko, G. A. Composition analysis of α-helices in thermophilic organisms. *Protein Eng. Des. Sel.* **8**, 905–913 (1995).
56. Hoover, H. S. *et al.* Phosphorylation of the brome mosaic virus capsid regulates the timing of viral infection. *J. Virol.* **90**, 7748–7760 (2016).
57. Howidi, M. *et al.* Burden and genotyping of rotavirus disease in the United Arab Emirates: A multicenter hospital-based surveillance. *Hum. Vaccines Immunother.* **10**, 2284–2289 (2014).
58. Bortagaray, V. *et al.* Detection, quantification, and microbial risk assessment of group A rotavirus in rivers from Uruguay. *Food Environ. Virol.* **12**, 89–98 (2020).
59. Osuolale, O. & Okoh, A. Human enteric bacteria and viruses in five wastewater treatment plants in the Eastern Cape, South Africa. *J. Infect Public Health* **10**, 541–547 (2017).
60. Miura, T., Gima, A. & Akiba, M. Detection of norovirus and rotavirus present in suspended and dissolved forms in drinking water sources. *Food Environ. Virol.* **11**, 9–19 (2019).
61. Moresco, V., Damazo, N. A. & Barardi, C. R. M. Rotavirus vaccine stability in the aquatic environment. *J. Appl. Microbiol.* **120**, 321–328 (2016).

Acknowledgements

Author contributions

Conceptualization: S.E., I.N., A.H., I.O.A., A.A.; data curation: S.E., I.N., A.H.; formal analysis: I.N., I.O.A., I.A.; funding acquisition: A.H.; investigation: I.N., I.O.A., I.A., A.A.; methodology: I.N., I.O.A., I.A.; supervision: S.E., A.H., A.A.; validation: I.N., I.O.A.; writing—original draft: I.N., A.H., I.A.; writing—review & editing: I.N., A.H., I.O.A., I.A., A.A., S.E.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-91607-3.

Correspondence and requests for materials should be addressed to S.E.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.