Completely genomic and evolutionary characteristics of human-dominant G9P[8] group A rotavirus strains in Yunnan, China

Ziqin Dian,1,2† Binghui Wang,1† Mao Fan,3 Shuwei Dong,1 Yue Feng,1 A-Mei Zhang,1 Li Liu,1 Hua Niu,2 Yuanyue Li1 and Xueshan Xia1,*

Abstract

G9P[8] rotavirus A (RVA) has been identified as the predominant genotype circulating in Yunnan, China. To elucidate the molecular characteristics of its genetic composition at the whole-genome level, the genomes of 12 strains isolated from paediatric patients with diarrhoea were fully sequenced and characterized. Eleven of the 12 strains were genotyped as G9-P[8]-I1-R1-C1-M1-A1-N1-T1-E1-H1, which is closely related to the Wa-like genotype 1 constellation. In contrast, one strain was genotyped as G9-P[8]-I1-R1-C1-M1-A1-N2-T1-E1-H1, with the NSP2 gene characterized as a DS-1 like genotype. Bayesian phylogenetic analysis indicated that G9 strains had emerged in 1932 with an estimated average evolutionary rate of 1.63×10⁻³ substitutions/site/year. Considering the high prevalence and fast evolutionary rate of G9P[8] rotaviruses, our results suggest that G9P[8] RVA should be strictly monitored in China.

Rotavirus belongs to the family Reoviridae containing segmented, double-stranded RNA genomes [1]. Group A rotavirus (RVA) is an important pathogen that causes diarrhoea in human children [2]. The genome of RVA encodes six structural proteins (VP1–VP4, VP6 and VP7) and five or six non-structural proteins (NSP1–NSP6). Based on the phylogenies of capsid protein VP7 and spike protein VP4, at least 32 G types and 47 P types have been discriminated, and the number of genotypes is still expanding [3]. Of these, G1P[8], G2P[4], G3P[8], G4P[8] and G9P[8] have been documented as the most prevalent genotypes globally during 2009–2012 [4–8], but recently, G9 RVA has emerged as the predominant genotype in many countries, including China.

Analysis of G9 RVA at the whole-genomic level is essential to reveal its genetic evolution and viral reassortment characteristics [9]. In terms of the whole-genome classification system, the rotavirus genotype constellations are Gx-P[x]-Ix-Rx-Cx-Mx-Ax-Nx-Tx-Ex-Hx, based on the variation VP7-VP4-VP6-VP1-VP2-VP3-NSP1-NSP2-NSP3-NSP4-NSP5/6 genes, respectively. Human RVA can be classified into the Wa-like (G1/G3/G4-P[8]-I1-R1-C1-M1-A1-N1-T1-E1-H1), DS-1-like (G2-P[4]-I2-R2-C2-M2-A2-N2-T2-E2-H2) and AU-1-like (G3-P[9]-I3-R3-C3-M3-A3-N3-T3-E3-H3) [10]. Using this classification, multiple genotypes of rotavirus have been assigned, and their phylogenetic relationship can be further analysed.

In our previous study, G9P[8] RVA was found to be the predominant genotype in Yunnan, China [11]. To further elucidate the genotype constellations of these G9P[8] RVA, 12 strains were selected, and their RNA genome was extracted using the TIANamp Virus RNA Kit (Tiangen Biotech, Beijing, China). The 11 segments of the 12 strains were amplified using previously published primer sets [12] and the One Step RT-PCR Kit (TAKARA, Dalian, China) following the manufacturer’s protocol. After sequencing and assembly, nearly full-length sequences (except for the 5¢ terminal sequences) were obtained. The genotype for each genome segment was preliminarily determined using the RotaC v.2.0 webserver, and the obtained sequences were submitted to GenBank (Table S1, available in the online Supplementary Material). Eleven of the 12 strains shared the same genotype constellations, and were identified as G9-P[8]-I1-R1-C1-M1-A1-N1-T1-E1-H1, possessing a complete Wa-like genomic backbone. The remaining strain...
Fig. 1. Phylogenetic trees were constructed for the genes encoding (a) VP7, (b) VP4 and (c) NSP2. Maximum likelihood (ML) trees were constructed using MEGA 6.06 software and bootstrap values were calculated for 1000 repetitions. The nucleotide sequences for the 12 G9P[8] strains of this study are highlighted with black triangles. The sequences of the reference strains were derived from the GenBank database (www.ncbi.nlm.nih.gov). The genetic distance is indicated at the bottom. Bootstrap values are only shown for values of 70% and greater.
(km15118) was identified as G9-P[8]-I1-R1-C1-M1-A1-N2-T1-E1-H1, bearing a NSP2 genome segment 8. In order to elucidate the phylogeny profile of these G9-P[8] strains at the whole-genome level, MEGA 6.06 software was used to construct the maximum-likelihood (ML) trees for each genomic segment, and the nodal reliability of the ML trees was assessed by bootstrap (BS) analysis with 1000 pseudo-replicates. Tamura 3-parameter+G+I models were chosen for phylogenetic analysis, whereas the genetic distances were calculated using the Kimura-2 correction parameter [12]. In the VP7 phylogenetic tree, G9 strains could be separated into six lineages; our strains present two separate clusters in lineage III. Strains km15118, km15099, km15097, km15007 and km15093 were grouped into cluster 1; these strains grouped tightly with strains from Korea with a genetic distance of 0.013. Strains km15066, km15015, km15094, km15035, km15095, km15094, and km15097 were grouped into cluster 2. Strains km15105 and km15119 were clustered separately from the other strains.
km15119 and km15007 were grouped into cluster 2, together with Chinese and Japanese strains with genetic distances of 0.009 and 0.011 (Fig. 1a). For the VP4 gene, four lineages of P[8] have been described and our strains fell into lineage 3. Of these, 11 strains, with the exception of km15093, were phylogenetically clustered with strains from South Africa, Zimbabwe, Korea and the USA, and had a shorter genetic distance from the Korean strains than the South African strains (0.015 vs 0.016). Strain km15093 differs from the others but is closely related to Chinese and Australian strains with genetic distances of 0.030 and 0.029, respectively. Two rotavirus vaccines, Rotarix™ (G1P[8], Glaxo Smith Kline Biologicals, Belgium) and RotaTeq™ (G1-G4P[8], Merck & Co., Inc., USA), were clustered in P[8] lineages 2 and 1, respectively (Fig. 1b). Based on NSP2 nucleotide sequences, phylogenetic analysis showed that strain km15118 displayed a DS-1-like genotype in the NSP2 segment. In the ML tree, this strain has a close genetic distance (0.01) from the G2P [4] virus from Thailand and the G3P[4] virus from Hungary. However, the other 11 studied strains are more closely related to the G9P[8], G3P[8] and G1P[8] viruses from Zimbabwe (MRC-DPRU1102), China (L1621) and the USA (CNMC126), respectively (Fig. 1c). Meanwhile, phylogenetic analysis based on VP1, VP2, VP3, VP6, NSP1 and NSP3-5 nucleotide sequences demonstrated that the presented G9P[8] strains have high homology with cognate gene sequences of G9P[8], G1P[8], G3P[8], G12P[6] (NSP3) and G12P[8] (NSP5) strains from Korea, the USA, Australia, Canada, South Africa and China, respectively (data not shown).

In the phylogenetic trees, 11 of the 12 studied strains possessed a Wa-like genomic backbone and clustered closely with reference strains that had the same genetic backbone, mostly from the same geographical origin. Moreover, high homology with G1P[8] and G3P[8] rotaviruses indicates a close inter-genotype relationship between the G9 and those rotavirus genotypes. Interestingly, strain km15118 carried a DS-1-like genome segment 8 (NSP2) in the Wa-like genomic backbone, thus differing from the common G9P[8] strains. It has been thought that rotavirus evolution mainly occurs through the selection of point mutations or reassortment of the segmented genome [13]. Here, human rotavirus with a purely Wa-like or a DS-1-like genome constellation was more common, while reassorted viruses are rarely detected; even if they emerge, the reassorted viruses may be less fit for the environment than parental strains and unable to compete with them [14]. Genome segment exchanges in G9P[4], G9P[6] and G3P[9] RVAs have been reported [5, 7, 15], and now a Wa-like G9P[8] strain with a DS-1-like segment of NSP2 has been found.

Furthermore, VP7 was chosen for further evolutionary analyses in order to better understand the transmission of G9 strains. Together with the 12 studied strains, 110 nearly full-length G9 sequences from different geographical regions and lineages (obtained from the NCBI database) were used for evolutionary analyses. Bayesian coalescent analysis was performed in BEAST v 1.6.1, and the mean evolutionary rate and the time to the most recent common ancestor (TMRCA) was calculated by using the general time-reversible+G+I substitution model and the relaxed uncorrelated lognormal molecular clock [16]. Convergence of the Markov chain Monte Carlo (MCMC) sample on the posterior distribution was defined at an effective sample size (ESS) value of greater than 200, which was calculated with Tracer v 1.4 (http://beast.bio.ed.ac.uk/Tracer). The maximum clade credibility (MCC) tree was annotated with the TreeAnnotator [17] and then visualized using FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/). The mean value of the evolutionary rate for the VP7 segment was estimated to be $1.63 \times 10^{-3}$ substitutions/site/year (95 % HPD, 1.1704$\pm 2.1068 \times 10^{-3}$ substitutions/site/year). Using this estimated nucleotide substitution value, TMRCA analysis estimated that the G9 strains emerged in 1928. Previously, six lineages of the G9 rotavirus have been described [18]. Lineages I, II, IV and V were only found in humans; lineages III and VI were found in both humans and porcine sources [19]. Lineage I G9 viruses included those isolated during 1980s, with W161 as the typical strain; lineage II included those isolated during 1990s, with 116E as the typical strain; lineage IV and V included those isolated during 1997s, and the typical strains were 97’sz37 and om67, respectively [6]; lineage III and VI accounted for the most prevalent G9 viruses around the world [18]. In this study, the emergence time of lineage VI was calculated to be 33 years ago; this was estimated to be earlier than lineage III, which emerged 29 years ago. The other four lineages were estimated to be the earlier ancestors, emerging 45–48 years ago. From the MCC tree, lineages III and VI were estimated to originate from lineage I, which was identified in Hong Kong, the USA and Japan. Lineages IV and V were estimated to originate from lineage II, which was discovered in India and the USA (Fig. 2).

Of the 12 strains in this study, five clustered tightly in lineage III with strains from Asian countries, including China, Korea and Japan. This lineage was estimated to have emerged in 1988 and is thought to be the youngest one in the MCC tree. Furthermore, lineage III is prevalent worldwide, and most recently reported G9 strains belong to it [6, 7, 20]. Strains isolated from America and Africa were also reported, but the study strains are more closely related to strains circulating in Asia. Notably, seven studied strains that belonged to lineage III cluster 2 in the VP7 phylogenetic tree clustered in lineage VI in the MCC tree. Further analyses indicated that these strains were identical to either lineage III cluster 1 or lineage VI, because the genetic distance for each of them was 0.08. The topological discrepancies between the ML and Bayesian trees probably originated from the model selection for each analysis. The studied strains belong to lineage VI, which originated in 1984 and clustered tightly with the reference strains identified in 2012–2013, which were from China and neighbouring countries, including Japan and Russia (Fig. 2). In the MCC tree, the seven studied strains in lineage VI showed a very close evolutionary relationship with strains from different regions of China, such as Beijing, Shanghai...
and Jiangsu. Our strains had similar evolutionary features for the VP7 and VP4 genes, e.g. the time of origin and the major distribution area were close. When making a comparison with the previously reported G2P[4] [17], we found that the VP7 segment in G9 strains has a faster evolutionary rate than the segments of G2P[4], i.e. 1.63x10^{-3} versus 0.95x10^{-3} (for VP7), 1.11x10^{-3} (for VP3) and 1.05x10^{-3} (for NSP4) substitutions/site/year, respectively. Regarding the prevalence of G9 strains in China, it is reasonable to assume that the fast evolution of the VP7 gene may help the G9 strains to adapt to the environment and infect hosts more effectively. Moreover, the high variability of the VP7 genome may help the virus to escape from immune and drug pressure. Because there is no effective anti-rotavirus drug, vaccination is taken to be the most useful and readily available tool for the prevention of RVA infection. Two oral vaccines, Rotarix™ and RotaTeq™ are being used worldwide, but have not been approved for use in China. In the current study, the P[8] lineage of the study strains (lineage 3) is distant from the vaccines (lineages II and I). It has been suggested that sub-genotypic lineages of RVA strains with different antigenic properties may enable rotaviruses to escape from adaptive immunity [21]. So far, no significant vaccine-attributable change has been found. However, researchers worry about selective pressure on prevalent rotavirus strains and a possible change of their evolutionary rate [22]. Therefore, it is necessary to monitor the circulation of G9 strains before the introduction of RVA vaccines to this region.

Yunnan is located in southwestern China and borders various Southeast Asian countries, including Laos, Vietnam and Myanmar. Due to its special geographic location, imported infectious diseases, such as human immunodeficiency virus (HIV), hepatitis C virus (HCV) and Dengue virus (DENV) have been reported sporadically. As it is a hub region between mainland China and Southeast Asia, it would be worth studying further the molecular epidemiology and evolutionary mechanism of the RVA circulating in this region. In our study, the high detection frequency and high evolutionary rate of G9P[8] strains suggest that this RVA genotype is widespread, at least in Yunnan, and has the potential to be transmitted to other provinces of China. Our studies have provided the necessary basic data to control RVA transmission in Yunnan and effect future vaccine development in China.

**Funding information**

This study was supported by grants from the National Science and Technology Sport Program of China (2014BAI01B01), the National Natural Science Foundation of China (81260248) and the Innovation Talent Supporting Project of Yunnan Province (2015HC030).
Acknowledgements
The authors would like to thank the clinical laboratories of the First People’s Hospital of Yunnan Province and Kunming Children’s Hospital for provision of the RVA strains.

Conflicts of interest
The authors declare that they have no conflicts of interest.

Ethical statement
This study was reviewed and approved by the Institutional Ethical Committee of Kunming University of Science and Technology. The guardians of the recruited children in this study were informed of the aims of this investigation, and provided oral consent.

Reference