Infection, Genetics and Evolution 20 (2013) 276–283

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

# Consensus sequence determination and elucidation of the evolutionary history of a rotavirus Wa variant reveal a close relationship to various Wa variants derived from the original Wa strain

Johannes F. Wentzel [a], Lijuan Yuan [b], Shujing Rao [c], Alberdina A. van Dijk [a], Hester G. O'Neill [d,*]

[a] Biochemistry Division, North-West University, Potchefstroom, South Africa
[b] Department of Biomedical Sciences and Pathobiology, Virginia-Maryland Regional College of Veterinary Medicine, Virginia Tech, Blacksburg, VA, USA
[c] Department of Entomology, Soils, and Plant Sciences, Clemson University, Clemson, SC, USA
[d] Department of Microbial, Biochemical and Food Biotechnology, University of the Free State, Bloemfontein, South Africa

ARTICLE INFO

*Article history:*
Received 11 June 2013
Received in revised form 14 August 2013
Accepted 3 September 2013
Available online 20 September 2013

*Keywords:*
Rotavirus Wa
Sequence-independent genome amplification
Consensus sequence
Phylogenetic analysis
Molecular clock analysis
Genome segment rearrangement

ABSTRACT

The consensus nucleotide sequence of a human rotavirus Wa strain, with only a partially known passage history, was determined with sequence-independent amplification and next generation 454® pyrosequencing. This rotavirus Wa strain had the expected genome constellation of G1-P[8]-I1-R1-C1-M1-A1-N1-T1-E1-H1 and was designated RVA/Human-tc/USA/WaCS/1974/G1P[8]. Phylogenetic analyses revealed a close relationship to four human rotavirus Wa variants (Wag5re, Wag7/8re, ParWa and VirWa) derived from the original 1974 human isolate. There were rearrangements in the Wag5re- and Wag7/8re variants in genome segments 5 (Wag5re) and 7 and 8 (Wag7/8re), which were not present in WaCS. Pairwise comparisons and a combined molecular clock for the Wa rotavirus genome indicated a close relationship between WaCS and ParWa and VirWa. These results suggest that WaCS is most probably an early cell culture adapted variant from the initial gnotobiotic pig passaged Wa isolate. Evolutionary pressure analysis identified a possible negative selected amino acid site in VP1 (genome segment 1) and a likely positive selected site in VP4 (genome segment 4). The WaCS may be more appropriate as a rotavirus Wa reference sequence than the current composite Wa reference genome.



## 1. Introduction

Whole-genome analyses of human rotavirus strains are fundamental in studying evolutionary patterns and genetic affiliations to other strains (Ghosh and Kobayashi, 2011). Matthijnssens and co-workers suggested a novel classification system based on the whole genome sequence of all 11 rotavirus genome segments in order to obtain a more complete picture of rotavirus strain diversity (Matthijnssens et al., 2008b). Nowadays, whole genome characterization has become the sought after procedure for viral strain characterization as next generation sequencing technology becomes more widely available and affordable. The easily accessible public sequence databases contain massive amounts of sequencing data, facilitating complex analysis and strain comparisons.

Since a link between rotavirus and gastroenteritis was established by Ruth Bishop and her colleagues 40 years ago (Bishop et al., 1973), this virus is now recognized to be the leading cause of severe dehydrating gastroenteritis among children under the age of five. In excess of 80% of the nearly half a million annual rotavirus related deaths occur in the developing regions, Asia and Africa, where access to proper medical care is limited (Madhi et al., 2010; Mwenda et al., 2010; Sanchez-Padilla et al., 2009). As part of the *Reoviridae* family, rotaviruses are non-enveloped particles, consisting of 11 double-stranded RNA (dsRNA) genome segments with a total size of about 18,600 base pairs. Ten of these genome segments encode a single viral protein, namely 6 structural proteins (VP1-4, VP6 and VP7) and 4 non-structural proteins (NSP1-4). Genome segment 11 encodes two non-structural proteins, NSP5 and NSP6. The genome is encapsulated in a triple-layered particle. The inner capsid particle consists of a RNA-dependent RNA polymerase (VP1), the RNA capping enzyme (VP3) and genomic double-stranded RNA (dsRNA), all encapsulated in the VP2 protein lattice. The inner capsid particle is surrounded by a middle layer protein (VP6), and is collectively known as the double-layered particle (DLP). The outer layer consists of VP7 and the spike protein VP4 to form the triple-layered particle (TLP) which is infectious (Estes and Kapikian, 2007).

Rotaviruses are conventionally classified into several serogroups (A–E, and possibly F and G). Type A rotaviruses are further classified using a binary system, G- and P-types, which refer to

* Corresponding author. Address: Department of Microbial, Biochemical and Food Biotechnology, University of the Free State, PO Box 339, Bloemfontein 9301, South Africa. Tel.: +27 51 401 2122.
*E-mail address:* oneillhg@ufs.ac.za (H.G. O'Neill).

1567-1348/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.meegid.2013.09.007

glycoprotein VP7 encoded by genome segment 9 and the protease sensitive VP4, encoded by genome segment 4. Thus far, 37 P and 27 G genotypes have been identified (Matthijnssens and van Ranst, 2012; Trojnar et al., 2013). The most prevalent rotavirus A strains found in humans are the genotypes G1–4, G9 and G12 in combination with P[4], P[6] and P[8] (Heiman et al., 2008; Matthijnssens et al., 2010). Group A rotaviruses include the AU-1 (G3P[9]), DS-1 (G2P[4]) and Wa (G1P[8]) genogroups. The AU-1 genogroup is fairly uncommon globally, in contrast to the DS-1-like and Wa-like genogroups that occur widespread among humans and various animal species (Matthijnssens and van Ranst, 2012). Phylogenetic analysis of rotavirus whole genome sequences indicated that DS-1-like strains are descendant from bovine rotaviruses, while Wa-like strains share a common ancestor with porcine rotaviruses (Matthijnssens et al., 2008a).

The Wa strain (Rotavirus A strain Human-tc/USA/Wa/1974/G1P[8]) was originally isolated in the United States in 1974 from an infant with severe diarrhoea. It was also one of the first rotaviruses to be successfully adapted to cultured cells (Wyatt et al., 1980), making the Wa strain one of the best-studied human rotaviruses to date.

In this paper we describe the consensus sequence, obtained by sequence-independent genome amplification and next generation 454® pyrosequencing, of a rotavirus Wa strain (obtained from Murdoch Children's Research Institute) derived from the original 1974 isolate. The evolutionary history of this strain was investigated through phylogenetic and molecular clock analyses combined with nucleotide substitution rate and evolutionary pressures analyses.

## 2. Materials and methods

### 2.1. Rotavirus and cell culture propagation

A cell culture adapted rotavirus Wa sample was obtained from Dr. Carl Kirkwood at the Murdoch Children's Research Institute (MCRI), Melbourne, Australia. This strain was originally obtained by Dr. Ruth Bishop from Dr. Richard Wyatt (National Institutes of Health, USA) in 1983. This particular Wa strain is a cell culture adapted variant from the original 1974 isolate but the exact passaging history is unknown (Dr. Carl Kirkwood and Dr. Ruth Bishop, personal communication). At MCRI the strain was passaged 9 times in MA104 cells.

Following activation with 10 μg/ml porcine trypsin IX (Sigma) at 37 °C for 30 min, the virus was passaged a further 7 times in African green monkey cells (MA104) at the North-West University (NWU), South Africa. The cells were cultured in serum free Dulbecco's modified essential medium (D-MEM; Hyclone) containing 1 μg/ml porcine trypsin (1x), 1% penicillin/streptomycin/amphotericin B (Gibco) and 1% non-essential amino acids (Lonza). Cells were cultured at 37 °C in a humidified atmosphere of 5% $CO_2$.

### 2.2. Sequence-independent cDNA synthesis and genome amplification

Rotavirus double-stranded RNA (dsRNA) was isolated as described by Potgieter et al. (2009). Infected cells were harvested when about 70% cytopathic effect was reached by freeze-thawing the cell/virus suspension twice. A phenol–chloroform extraction was performed using the Trizol reagent (Invitrogen) and the single-stranded RNA was removed by precipitation with 2 M LiCl (Sigma) at 4 °C for 14 h. Subsequently, the solution was centrifuged at 16,000*g* for 30 min at 4 °C and the supernatant was purified using the MinElute kit (Qiagen) as described by the manufacturer. A PC3-T7 loop primer (5'p-GGATCCCGGGAATTCGGTAATACGACTCACTATATTTTTATAGTGAGTCGTATTA-OH3') (TibMolBiol) was ligated to the purified RNA and the genome was subsequently amplified as cDNA using a sequence-independent genome amplification technique (Potgieter et al., 2009 with slight modifications. The purified ligated dsRNA was denatured using 300 mM methyl mercury hydroxide (Alfa Aesar). The cDNA was synthesised using AMV reverse transcriptase (Fermentas) followed by amplification of the genome with Phusion High Fidelity DNA polymerase (Finnzymes). The QIAquick (Qiagen) PCR purification kit was employed in order to purify the amplified cDNA according to the manufacturer's instructions. This rotavirus Wa-amplicon cocktail was sequenced using 454® pyrosequencing technology (GS FLX Titanium, Roche) at Inqaba Biotec (South Africa) as described before (Jere et al., 2011).

### 2.3. Sequence and phylogenetic analyses

The Lasergene™ 8.1.2 suite (DNASTAR®) was used for sequence assembly. The consensus sequence of all 11 genome segments was determined using the SeqMan module of this software suite. The GenBank accession numbers are listed in Table 1. The nucleotide and deduced protein sequences were analysed with the *Basic Local Alignment Search Tool* (*BLAST*) and compared with Wa sequences available in GenBank. Sequences of the 11 genome segments of all rotavirus strains (Supplementary Table 1) that closest resembled the sequenced Wa strain were retrieved from GenBank and aligned using MEGA 5.1. The evolutionary history was determined using the Neighbour-Joining method (Saitou and Nei, 1987) conducted in MEGA 5.1 (Tamura et al., 2011) with a bootstrap value of 10,000. In order to obtain a more comprehensive phylogenetic overview, the rotavirus DS-1, AU-1 and D reference strains were also included. The evolutionary distances were computed using the Maximum Composite Likelihood method (MEGA 5.1) and are in the units of the number of base substitutions per site (Tamura et al., 2004). Codon positions included were 1st + 2nd + 3rd + noncoding. All positions containing gaps and missing data were eliminated.

### 2.4. Molecular clock analyses and evolutionary rate estimations

Bayesian Evolutionary Analysis Sampling Trees (BEAST) is a multifaceted evolutionary package for phylogenetic and population genetics analysis. Bayesian phylogenetic reconstructions were performed using the Markov chain Monte Carlo (MCMC) analysis contained in the BEAST software suite (1.6.2) (Drummond and Rambaut, 2007). Aligned rotavirus sequences were converted to the NEXUS format using Data Analysis in Molecular Biology Evolution (DAMBE) software 5.2.76 (http://dambe.bio.uottawa.ca/dambe.asp). JModelTest (http://darwin.uvigo.es/software/software.html) was used to determine the most suitable nucleotide substitution model and subsequently all strains were analysed using a HKY model with gamma distributed rate variation and a relaxed clock lognormal model with a flexible Bayesian skyline tree prior. One hundred million MCMC simulations were performed (Matthijnssens et al., 2010). Tree files of all 11 genome segments were generated and annotated with TreeAnnotator. Additionally, all 11 tree files were combined with LogCombiner 1.6.2, in order to produce a tree representing the entire genome of the rotaviruses examined. Trees were visualized by FigTree 1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

Evolutionary rates were estimated for all 11 genome segments of the closely related rotavirus variants in the PAML 4.5 software package (Yang, 1997) using codon substitution models with a single non-synonymous/synonymous substitution rate (dN/dS). To elucidate general evolutionary pressures acting on protein-coding regions, non-synonymous–synonymous substitution ratios (ω) were also employed (Yang and Nielsen, 2002; Yang et al., 2000)