Unbiased whole-genome deep sequencing of human and porcine stool samples reveals circulation of multiple groups of rotaviruses and a putative zoonotic infection

My V. T. Phan[1], Pham Hong Anh[2], Nguyen Van Cuong[2], Bas B. Oude Munnink[1], Lia van der Hoek[3,†], Phuc Tran My[2], Tue Ngo Tri[2], Juliet E. Bryant[2,4], Stephen Baker[2,4,5], Guy Thwaites[2,4], Mark Woolhouse[6], Paul Kellam[7,8,*,‡], Maia A. Rabaa[2,4], Matthew Cotten[1,9,*], on Behalf of the VIZIONS Consortium

[1]Virus Genomics, Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK, [2]Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam, [3]Laboratory of Experimental Virology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands, [4]Centre for Tropical Medicine, Nuffield Department of Medicine, University of Oxford, Oxford, UK, [5]London School of Tropical Medicine and Hygiene, London, UK, [6]Centre for Immunity, Infection & Evolution, University of Edinburgh, Edinburgh, UK, [7]Kymab Inc., Cambridge, UK, [8]Imperial College, London, UK and [9]Department of Viroscience, Erasmus Medical Center, Rotterdam, The Netherlands

*Corresponding authors: E-mail: mc13@sanger.ac.uk
†http://orcid.org/0000-0003-2803-642X
‡http://orcid.org/0000-0003-3166-4734

Abstract

Coordinated and synchronous surveillance for zoonotic viruses in both human clinical cases and animal reservoirs provides an opportunity to identify interspecies virus movement. Rotavirus (RV) is an important cause of viral gastroenteritis in humans and animals. In this study, we document the RV diversity within co-located humans and animals sampled from the Mekong delta region of Vietnam using a primer-independent, agnostic, deep sequencing approach. A total of 296 stool samples (146 from diarrhoeal human patients and 150 from pigs living in the same geographical region) were directly sequenced, generating the genomic sequences of sixty human rotaviruses (all group A) and thirty-one porcine rotaviruses (thirteen group A, seven group B, six group C, and five group H). Phylogenetic analyses showed the co-circulation of multiple distinct RV group A (RVA) genotypes/strains, many of which were divergent from the strain components of licensed RVA vaccines, as well as considerable virus diversity in pigs including full genomes of rotaviruses in groups B, C, and H, none of which have been previously reported in Vietnam. Furthermore, the detection of an atypical RVA genotype constellation (G4-P[6]-I1-R1-C1-M1-A8-N1-T7-E1-H1) in a human patient and a pig from the same region provides some evidence for a zoonotic event.

Key words: rotavirus; deep sequencing; whole genomes; virus surveillance; zoonotic infection.

1. Introduction

Rotavirus (RV) infections are the leading cause of acute gastro-enteritis globally, with a disproportionally greater morbidity and mortality in developing countries of Asia and sub-Saharan Africa (Tate et al. 2012, 2016). RV can infect humans and different animal species and is considered, in part, a zoonotic disease in humans (Estes et al. 2007). RV zoonotic infections and transmissions have been shown with animal strains moving into humans via direct contact with animals or exposure to environmental contamination (Cook et al. 2004; Martella et al. 2010; Ghosh and Kobayashi 2014), and present a challenge to infection control and management. RV is a non-enveloped double-stranded RNA virus forming the single genus *Rotavirus* in the Reoviridae family, with a 18.5-kb genome of eleven segments encoding six structural (VP1–4, VP6, and VP7) and five or six non-structural proteins (NSP1–NSP5/6) (Ramig et al. 2005; Estes et al. 2007). RVs are classified into eight established groups (A–H) and a new tentative group (I) based on the genetic and antigenic differences of VP6, with viruses of group A, B, C, and H known to infect both humans and other animals (Matthijnssens et al. 2012; Mihalov-kovács et al. 2015).

Within each RV group, strains are distinguished and classified into G and P genotypes based on VP7 surface glycoprotein and VP4 spike protein, respectively (Estes et al. 2007). For RV group A (RVA), at least 27 G and 37 P types have been detected in human and animals, with typical combinations of G-types (G1–G4, G9, G12) and P-types (P[4], P[6], P[8]) found in human infections globally, and different combinations of G- and P-types (G3–G5, G9, G11, P[6], P[7], P[13]) commonly found in pigs (Santos and Hoshino 2005; Matthijnssens et al. 2011; Trojnar et al. 2013; Ghosh and Kobayashi 2014; Dóró et al. 2015). A genotype classification system based on eleven genomic segments is recommended for RVA, with Gx-P[x]-Ix-Rx-Cx-Mx-Ax-Nx-Tx-Ex-Hx representing the genotypes of VP7-VP4-VP6-VP1-VP2-VP3-NSP1-NSP2-NSP3-NSP4-NSP5 segments. The most prevalent human strains belong to constellations of Wa-like (I1-R1-C1-M1-A1-N1-T1-E1-H1, commonly associated with G1P[8], G3P[8], G4P[8], G9P[8], G12P[8]) and DS-1-like (I2-R2-C2-M2-A2-N2-T2-E2-H2, commonly combined with G2P[4]) (Matthijnssens et al. 2008; Matthijnssens and Van Ranst 2012). Such an elegant whole genome genotyping system has not yet been established for non-RVA groups.

Determining the sequence of the 18.5-kb segmented genome for rotaviruses by standard methods can be biased and cumbersome, requiring an initial PCR step to identify and select primers specific for the RV genogroup and/or strain, with the 11-segment genome providing an additional complication for primer design. Such primer-based sequencing strategies can be further complicated by reassortment possibilities not predicted by the initial PCR typing, leading to sequencing failure of atypical and un-typeable RV strains whose frequency can vary by location, season, and environment (Cook et al. 1990; Gentsch et al. 2005; Santos and Hoshino 2005; Pitzer et al. 2009, 2011; Atchison et al. 2010). Next-generation sequencing has been recently employed for whole-genome sequencing of RVA with initial 11 PCR amplifications (Jere et al. 2011; Magagula et al. 2015; Nyaga et al. 2015); however, a single robust platform for whole-genome deep sequencing of multiple RV genogroups without prior genotype information would be useful. Routine identification of circulating RVA can be performed using commercial enzyme immunoassay kits (based on inner capsid protein VP6) and RT-PCR diagnostic and genotyping assays (based on outer capsid proteins VP4 and VP7) (Estes et al. 2007; Desselberger 2014). In addition, specific, rapid, and cost-effective assays are lacking for the detection of less common rotaviruses such as viruses in group B, C, and H (RVB, RVC, RVH), hindering our understanding of molecular epidemiology of these viruses and challenging efforts of genomic sequencing, particularly in resource-limited countries (Estes et al. 2007; Desselberger 2014).

The RV vaccines (Rotarix and RotaTeq) have been available since 2006 (Ruiz-Palacios et al. 2006; Vesikari et al. 2006), and offer a variable degree of protective immunity against human RVA infections. Reduced RVA vaccine efficacy has been observed in resource-limited countries in comparison to developed countries (Armah et al. 2010; Zaman et al. 2010; Breiman et al. 2012; Glass et al. 2014). The mechanism responsible for reduced vaccine efficacy in these settings is unclear, but may in part be due to local circulation of genetically and antigenically divergent RVA or zoonotic strains in developing countries (Santos and Hoshino 2005). Similar to other segmented viruses (Nomikou et al. 2015), genetic reassortment has been observed in RVs yielding significant genetic diversity, including a number of cross-species reassortants (Cook et al. 2004; Estes et al. 2007; Desselberger 2014). Hence, assessment of all eleven genome segments through full virus genome sequencing is essential for monitoring the overall RV genomic diversity, complex evolutionary dynamics, and the emergence of novel and zoonotic reassortants that may compromise vaccine protection (Matthijnssens et al. 2008, 2011; Matthijnssens and Van Ranst 2012).

Vietnam is a low-to-middle income country located in Southeast Asia and is considered one of the global hot spots of emerging infectious diseases (Hay et al. 2013). Diarrhoea is the fourth most common cause of mortality in children <5 years of age, accounting for 12% of deaths in this age group in 2013 (World Health Organization 2015). Among all diarrhoeal pathogens, RVA is responsible for 44–67.4 percent of all childhood diarrhoea cases requiring hospitalisation (Nishio et al. 2000; Nguyen et al. 2001, 2004, 2007; Doan et al. 2003; Landaeta et al. 2003; Van Man et al. 2005; Anh et al. 2006; Ngo et al. 2009; Tamura et al. 2010; Tra My et al. 2011; Trang et al. 2012; Vu Tra My et al. 2014; Thompson et al. 2015). Contrary to the clinical and public health importance, vaccination against RVA is currently not part of the Extended Program on Immunisation for Vietnamese infants. In addition, diagnosis for rotaviruses in diarrhoeal cases in humans and animals is not routinely performed and systematic genomic surveillance of circulating human and animal rotaviruses is limited. This leads to relatively little data on the overall RV prevalence and diversity in human and animal populations and their contribution to human infections and their potential to compromise RVA vaccine protection. Given the tropical climate of Vietnam prone to flooding, the close human–animals living proximity and high prevalence of infectious diseases, we hypothesized that RV zoonosis may occur in the region but is under-investigated and under-characterized. There is no report on the overall prevalence of other RVs (non-RVA) in both humans and animals in this region. To address this knowledge gap, we used focused sampling within human healthcare and animal farming populations, combined with high-throughput primer-independent direct genome sequencing from clinical materials (Cotten et al. 2014) to document RV diversity and transmission within and between humans and animals in a region of Vietnam.

2. Materials and Methods

2.1 Study setting and design

Human and porcine faecal samples were collected as part of the Vietnam Initiative on Zoonotic Infections (VIZIONS) project

from Dong Thap, a peri-urban province located in the south of Vietnam in the Mekong Delta region (see map, Supplementary Fig. S1). The human subjects were diarrhoeal patients (N = 146), regardless of age and gender, admitted to Dong Thap Provincial Hospital in the period from October 2012 to January 2014. Diarrhoea was defined as "at least three loose stools or one bloody stool within 24 hours ", according to the WHO guidelines (World Health Organization 2005). A stool specimen was collected from each individual within 24 h of hospital admission to avoid confounding by nosocomial infections. A total of 150 porcine faecal samples were randomly selected from a collection of porcine stool samples from pig farm baseline surveillance samples collected across the same province from January 2012 to April 2013. For four pigs in farms where no faecal specimens were obtained, a boot swab was collected (pig ID 12087_38, 14152_6, 14150_53, and 14250_12). The collection dates and ages of human enrollees and pigs in this study are provided in Supplementary Table 1. All collected faecal samples were stored in aliquots at −80°C until further processing. Ethical approval for the study was obtained from the Oxford Tropical Research Ethics Committee (OxTREC Approval No. 15-12) (Oxford, UK), the institutional ethical review board of Dong Thap Provincial Hospital (DTPH), and the Sub-Department of Animal Health Dong Thap province (Dong Thap, Vietnam).

2.2 Mapping of the patient residential and pig farm addresses

The residential district centroid was recorded for enrolled human patients to maintain participant anonymity, while the exact geographical location was recorded for the pig farms using an eTrex Legend GPS device (Garmin, UK). The decimal degrees of latitude and longitude were entered in a confidential database and kept separate from patient metadata so that patient identities could not be revealed based on the residence locations. These addresses were then validated in Google Earth Pro (https://www.google.com/earth/) and finally visualized in QGIS v2.2.0 (http://www.qgis.org/en/site/) overlaid with province-specific geographic data.

2.3 Sample preparation and nucleic acid extraction

Total nucleic acid extraction was performed as previously described (de Vries et al. 2011, 2012; Cotten et al. 2014). In brief, 110 μl of a 50 percent stool suspension in PBS was centrifuged for 10 min at 10,000 × g. Non-encapsidated DNA in the samples was degraded by addition of 20 U TURBO DNase (Ambion). Virion-protected nucleic acid was subsequently extracted using the Boom method (Boom et al. 1990). Reverse transcription was performed using non-ribosomal random hexamers (Endoh et al. 2005) that avoid transcription of rRNA, and second strand DNA synthesis was performed using 5 U of Klenow fragment 3′–5′ exo⁻ (New England Biolabs). Final purification of extracted nucleic acids was performed with phenol/chloroform and ethanol precipitation.

2.4 Library preparation and sequencing

Standard Illumina libraries were prepared for each sample. In short, nucleic acids in each sample were sheared to 400–500 nt in length, each sample's nucleic acid was separately indexed and samples were multiplexed at either seven samples per MiSeq run or ninety-six samples per HiSeq 2500 run, generating 2–3 million 149 nt (MiSeq) or 250 nt (HiSeq) paired-end reads per sample.

2.5 *De novo* assembly and identification of viral genomes

Raw sequencing reads were filtered to remove low-quality reads (Phred score >35) and trimmed to remove residual sequencing adapters using QUASR (Watson et al. 2013). The reads were assembled into contigs using *de novo* assembly with SPAdes (Bankevich et al. 2012) combined with sSpace (Boetzer et al. 2011). RV-encoding contigs and other mammalian virus contigs were identified with a modified SLIM algorithm (Cotten et al. 2014) combined with ublast (Edgar 2010). Coverage was determined for all contigs harvested to filter any process contamination sequences in each run, followed by additional filtering for minimum contig size cutoff (300 nt). Partial but overlapping contigs were joined into full-length sequences using Sequencher (Gene Codes Corporation, USA), and any ambiguities were resolved by consulting the original short reads. Final quality control of genomes included a comparison of the sequences, open reading frames (ORFs), and the encoded proteins with reference sequences retrieved from GenBank.

2.6 Genotyping and phylogenetic reconstruction

Assembled RVA sequences were genotyped using the online genotyping tool, RotaC v2.0 (http://rotac.regatools.be) (Maes et al. 2009), according to the guidelines for precise RVA classification using all eleven genomic segments (Matthijnssens et al. 2011). The resulting RVA, RVB, RVC, and RVH sequences were combined with additional full-length or nearly full-length sequences from previous Vietnamese studies (if available) and global representatives retrieved from GenBank. The complete genomes from the vaccine components of the monovalent vaccine Rotarix (Gautam et al. 2013) and the pentavalent vaccine RotaTeq (Matthijnssens et al. 2010) were retrieved from GenBank for phylogenetic reconstructions of all eleven RVA segments. Sequences were aligned using MUSCLE v3.8.31 (Edgar 2004) and manually checked in AliView (Larsson 2014); aligned sequences were trimmed to complete ORFs for subsequent analyses. Evolutionary model testing was implemented in IQ-TREE v3.10 (Nguyen et al. 2015) using the Akaike Information Criterion (AIC) to determine the best-fit models of nucleotide substitution for all genomic segments. Maximum likelihood (ML) phylogenetic trees were then inferred in IQ-TREE v3.10 with 500 bootstrap replicates under the best-fit model of evolution according to AIC (Supplementary Table 2 summarized the models determined for all segments). Resulting trees were visualized and edited using FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/). Genetic distances (p-uncorrected) were estimated using Geneious v9.0.4 (Biomatters Ltd).

2.7 Bayesian analysis for RVA NSP3 T7 genotype and for RVH VP6 gene sequences

Available sequences were retrieved from GenBank and new sequences obtained in this study were aligned using MUSCLE v3.8.31 (Edgar 2004), manually checked in AliView (Larsson 2014), and trimmed to complete ORF. An ML phylogenetic tree was constructed under the GTR + Γ₄ model of substitution in IQ-TREE v3.10 (Nguyen et al. 2015). For RVH VP6 sequences, highly similar sequences were removed before running Bayesian analyses (strains BR59, BR60, BR61, BR62, BR63, NC7_64_3, OK5_68_10). The molecular clock model was assessed in TempEst v1.5 (Rambaut et al. 2016), assessing the linear regression between root-to-tip divergence and the date of sampling. For RVA NSP3 T7 analysis, year of strain collection was used as tip dates since data on day and month were not available for

GenBank sequences; for RVH VP6 sequences, tip dates were defined as year, month, day of strain collection. A Bayesian Markov Chain Monte Carlo (MCMC) approach was then performed in BEAST v1.8.0 (Drummond and Rambaut 2007), using relaxed lognormal molecular clock under HKY85+Γ_4 substitution model. For RVA NSP3 T7 sequences, a Bayesian SkyGrid population process was employed and run in triplicate for 100 million generations chain with sampling performed every 10,000 runs. For RVH VP6 sequences, a non-parametric Guassian Markov Random Fields (GMRF) Bayesian Skyride population analyses were run in triplicate for 50 million generations with sampling performed every 5,000 generations. These triplicate runs were then combined using LogCombiner v1.8.0 (available within the BEAST package) with a removal of 10% burn-in, and analysed in Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer/) to ensure all parameters had converged with effective sample size (ESS) values >200 and to estimate the mean evolutionary rates across branches. Maximum clade credibility trees were annotated using TreeAnnotator v1.8.0 (BEAST) and visualised in FigTree v1.4.2.

2.8 GenBank accession numbers

All sequences generated in this study were deposited into GenBank under accession numbers KX362367–KX363442. Illumina raw read sets are available at the European Nucleotide Archive under submission ERR471259–ERR477293, ERR689707–ERR767572, ERR775471–ERR780002, ERR780013–ERR780019, ERR956666, ERR956667, ERR962074, ERR1300950–ERR1301100.

3. Results

3.1. Overall diversity of rotaviruses in human and pigs

Sequencing of human enteric samples from acute diarrhoeal patients admitted to Dong Thap Provincial Hospital from 2012 to 2014 yielded sixty *de novo* assembled RVA genome sequences from 146 samples (41.1 percent). No other RV genogroups were found in these human stool samples (Table 1). The same methods applied to 150 porcine faecal samples collected within the same geographic region (Supplementary Fig. 1) identified thirty-one rotaviruses from four different RV groups (A, B, C, and H) in a total of 150 samples (20.7 percent). These *de novo* assembled sequences included thirteen RVA (41.9 percent), seven RVB (22.6 percent), six RVC (19.4 percent), and five RVH (16.1 percent) (Table 1). The length of each assembled sequence was determined and expressed as percentage length coverage (length of assembled sequence divided by expected full length of that segment) for the corresponding segment (Fig. 1) . In samples where two distinct contigs were assembled for a segment (e.g. mixed infections), only the longer assembled contig was reported in the heatmap of segment coverage for the purpose of clarity (Fig. 1). The overall length coverage in human RVA sequences was higher than porcine RVA, RVB, RVC, and RVH, possibly be due to differential viral load or sample quality.

3.2. Human and porcine RVA genotype constellations

The genotype constellations were determined for all RVA strains according to established guidelines from the Rotavirus Classification Working Group (Maes et al. 2009; Matthijnssens et al. 2011). Among the human RVA, the two most common constellations were G1-P[8]-I1-R1-C1-M1-A1-N1-T1-E1-H1 (N = 33; Wa-like constellation) and G2-P[4]-I2-R2-C2-M2-A2-N2-T2-E2-H2 (N = 12; DS-1-like constellation) (Fig. 2). Reassorted RVA

strains between Wa-like and DS-1-like constellations were found in four human diarrhoeal patients, including G1-P[8]-I2-R2-C2-M2-A2-N2-T2-E2-H2 ($N = 3$) and G2-P[8]-I2-R2-C2-M2-A2-N2-T2-E2-H2 ($N = 1$). Interestingly, the genotype constellation G4-P[6]-I1-R1-C1-M1-A8-N1-T7-E1-H1 was found in one human and one pig sample. The closely related genotype constellation G4-P[6]-I1-R1-C1-M1-A8-N1-T1-E1-H1, which differs only by the NSP3 segment, predominated among the porcine RVA mono-infections ($N = 6$) (Fig. 2). Among all porcine RVA strains, the internal core gene cassette of R1-C1-M1-A8-N1-T1-E1-H1 (representing genotype of VP1-VP2-VP3-NSP1-NSP2-NSP3-NSP4-NSP5) was relatively conserved, with the exception of co-circulation of T1 and T7 genotypes in the NSP3 segment. The genotypes of capsid proteins (VP7-VP4-VP6) were more diverse in pigs, including G4-P[6]-I1 (six strains combined with NSP3 T1 and one strain with NSP3 T7), G5-P[13]-I5 ($N = 2$), G9-P[23]-I5 ($N = 1$), and G11-P[23]-I5 ($N = 1$).

Mixed infections were identified in nine samples, seven in humans and two in pigs (Fig. 2), with mixed infection being defined as the detection of two assembled but genetically distinct contigs in at least one segment with sufficient contig coverage to exclude potential process contamination among samples in the same run. The two homologous contig segments identified in mixed infections can have different or the same genotype; for example a mixed infection reported in an individual pig (sample ID 12070_4) contained two homologous VP7 segments, NSP1, NSP2, NSP3, NSP4, and NSP5 bearing the constellation of G1/G4-P[8]-I1-R1-C1-M1-A1/A8-N1/N1-T1/T1-E1/E1-H1/H1 (Fig. 2 and Supplementary Fig. S2). Another porcine sample was found with a mixture of G9/G11-P[13]/P[23]-I5/I5-R1/R1-C1-M1-A8/A8-N1/N1-T1/T1-E1/E1-H1 (sample ID 14150_53); however, it is important to note that this particular sample was a boot swab of faecal material in a cage-type pigsty, thus there is the possibility that the sample represents mixed environmental virus from more than one pig. Mixed human RVA infections typically contained genotypes 1 and 2 (Wa-like and DS-1 like, respectively) viruses.

3.3. Phylogenetic diversity of local human and porcine RVA

Phylogenetic trees were inferred for each RVA segment (Fig. 3A and B and Supplementary Fig. S3A and B) from assembled sequences in this study along with full-length sequences from previous studies in Vietnam, reference sequences in GenBank and sequences from the RVA vaccine formulations (RotaRix and RotaTeq). The local sequences clustered primarily by genotype as expected (Fig. 3A and B and Supplementary Fig. S3A and B); for example, VP7 G1 sequences in this study clustered with other G1 sequences from other regions and our G2 sequences clustered with other G2 (Fig. 3A). Sequences within the G4 genotype fell into two sub-lineages, with the human strain (16020_7) and porcine strains from this study clustering into one common sub-lineage (Fig. 3A). The mixed infection in the pig (12070_4) described above comprised two distinct contigs for the VP7 segment (belonging to the G1 and G4 genotypes), with the G1 sequence clustering within the human G1 lineage and the G4 sequence falling into a lineage with other G4 porcine sequences from this study (Fig. 3A). Similar observations were seen in the phylogenetic tree for VP4 sequences (Fig. 3B) and other gene segments (Supplementary Fig. S3A and B). It is also noteworthy that sequences from the vaccine strains (Rotarix and RotaTeq) were relatively distinct from the Vietnamese RVA sequences reported here, particularly for genotypes G5, G9, G11, P[6], P[13],

Rotavirus segment coverage
Rotavirus A

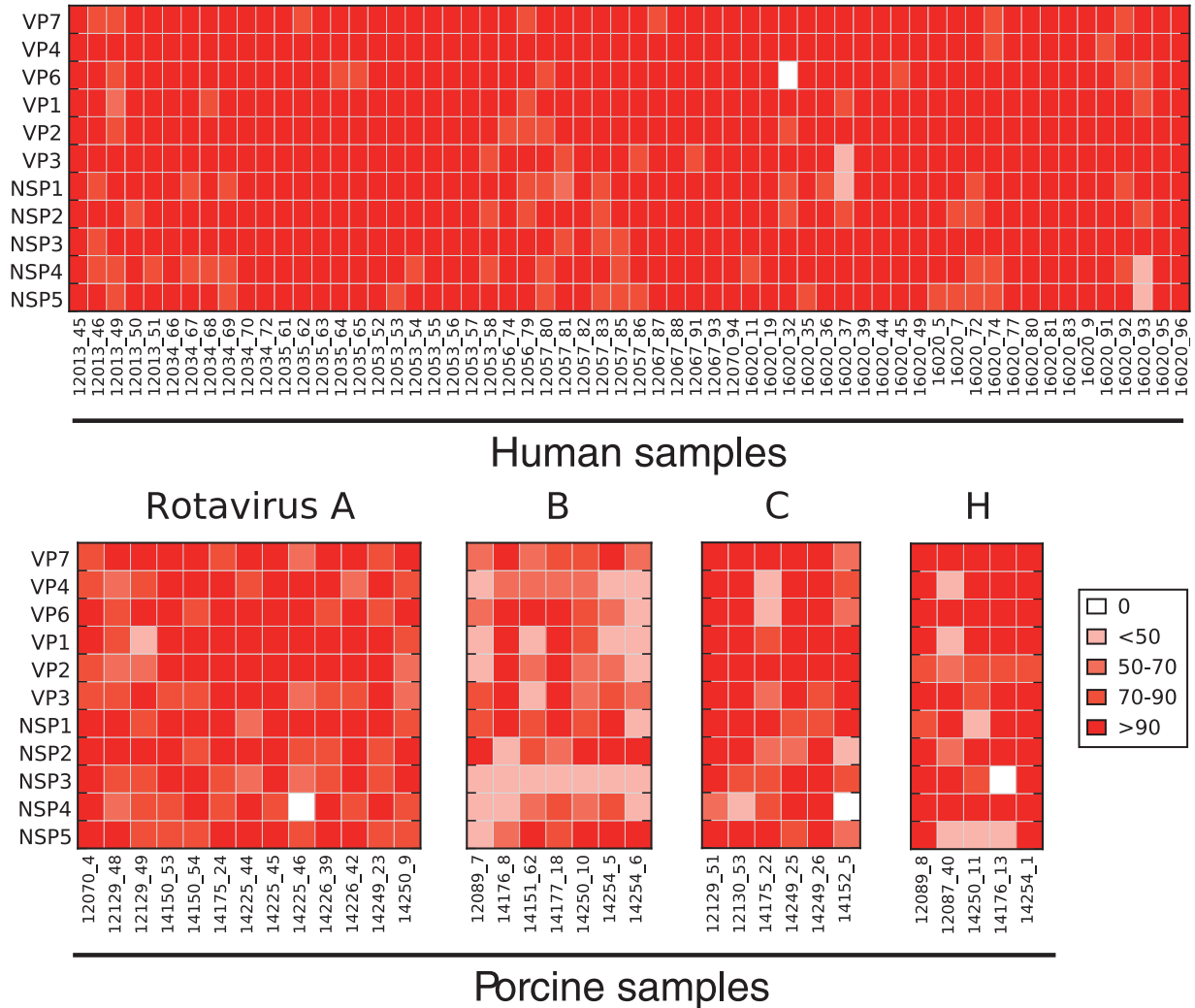## Human samples

Porcine samples

Figure 1. Heat map of RV sequence length coverage by segment detected in all samples. The sequence length coverage for each segment of all assembled rotaviruses by deep sequencing was calculated and expressed as [(length of assembled contig in nt)/(full-length of that segment in nt)]×100. Colour code for %genome coverage is indicated in figure key ranging from low (pale orange) to high (dark red). The value of 0 indicates that contig sequence for that segment was not identified or did not pass the stringent quality control criteria including % reads mapped and contig length (see Materials and Methods). All RVA sequences detected in human samples were shown in the first panel, with each column representing a sample and each row showing each RV segment. Similarly, porcine RV samples were shown, each panel representing RVA, RVB, RVC, and RVH sequences with the segment names given vertically and sample IDs horizontally.

Table 1. Number of samples tested and number of samples yielded up to full or nearly full RV genomic sequences.

Host	RVA	RVB	RVC	RVH	Total samples tested
Human	60 (41.1%)	0	0	0	146
Pig	13 (8.7%)	7 (4.7%)	6 (4%)	5 (3.3%)	150

Host indicates human or pig host from which the faecal samples were collected. Percentages of full or nearly full RV genomes were given in brackets. RVA, rotavirus genogroup A; RVB, group B; RVC, group C, RVH, group H.

P[23] of the two neutralizing antigens, VP7 and VP4 (Fig. 3A and B). Comparison of the amino acid sequences of VP7 and VP4 of the local strains to the Rotarix and RotaTeq vaccine sequences indicated a number of amino acid differences observed across the length of the proteins and particularly in the antigenic epitopes of VP7 and VP4 (data not shown). Taken together, multiple RVA genotypes co-circulate in human and pigs in this location; many of these genotypes are genetically dissimilar to currently used vaccine components.

3.4. Putative zoonotic infection of human with a porcine-human RVA virus

An atypical RVA genotype constellation of G4-P[6]-I1-R1-C1-M1-A8-N1-T7-E1-H1 was found in both a human patient (16020_7) and a weaning pig (14250_9), whose geographical distance (between the residence and the farm) was ~35 km apart (Fig. 4A and B). The genotype constellation of the core gene cassette (R1-C1-M1-A8-N1-T7-E1-H1) was also

|	Genotype constellation											Host count	
	VP7	VP4	VP6	VP1	VP2	VP3	NSP1	NSP2	NSP3	NSP4	NSP5	Human	Pig
Mono-infection	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	33	0
	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2	12	0
	G3	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2	0
	G4	P[6]	I1	R1	C1	M1	A8	N1	T7	E1	H1	1	1
	G4	P[6]	I1	R1	C1	M1	A8	N1	T1	E1	H1	0	6
	G5	P[13]	I5	R1	C1	M1	A8	N1	T7	E1	H1	0	1
	G11	P[13]	I5	R1	C1	M1	A8	N1	T7	E1	H1	0	1
	G9	P[23]	I5	R1	C1	M1	A8	N1	T1	E1	H1	0	1
	G5	P[13]	I5	R1	C1	M1	A8	N1	T7	x	H1	0	1
	G1	P[8]	I2	R2	C2	M2	A2	N2	T2	E2	H2	3	0
	G2	P[8]	I2	R2	C2	M2	A2	N2	T2	E2	H2	1	0
	G1	P[8]	x	R1	C1	M1	A1	N1	T1	E1	H1	1	0
Mixed infection	G1/G2	P[8]/P[4]	I1/I2	R1	C1/C2	M1/M2	A1/A2	N1/N2	T1/T2	E1/E2	H1/H2	1	0
	G1/G2	P[8]/P[4]	I1/I2	R1/R2	C1	M1/M2	A1/A2	N1/N2	T1/T2	E1/E2	H1	1	0
	G1/G2	P[8]/P[4]	I1/I2	R1/R2	C1/C2	M1/M2	A1/A2	N1/N2	T1/T2	E2	H2	1	0
	G1/G2	P[8]/P[4]	I1/I2	R1/R2	C1/C2	M1/M2	A1/A2	N1/N2	T1/T2	E1/E2	H1	1	0
	G1/G2	P[8]/P[4]	I1/I2	R1/R2	C1/C2	M1/M2	A1/A2	N1/N2	T1/T2	E1/E2	H1/H2	1	0
	G1/G2	P[8]/P[4]	I1/I2	R2	C1/C2	M2	A1/A2	N1/N2	T1/T2	E1/E2	H1/H2	1	0
	G1	P[8]	I1	R1	C1	M1	A1	N1/N1	T1	E1	H1	1	0
	G1/G4	P[8]	I1	R1	C1	M1	A1/A8	N1/N1	T1/T1	E1/E1	H1/H1	0	1
	G9/G11	P[13]/P[23]	I5/I5	R1/R1	C1	M1	A8/A8	N1/N1	T1/T7	E1/E1	H1	0	1

Legend:

Colour	Description
Wa-like genotype (G1/G3/G4/G9-P[8]-I1-R1-C1-M1-A1-N1-T1-E1-H1)	
DS-1-like genotype (G2/G8-P[4]-I2-R2-C2-M2-A2-N2-T2-E2-H2)	
G4-P[6]	
A8-T7	
G9-P[23]	
G5	
G11	
I5	
P[13]	
x	Sequence not determined

Figure 2. Genotype constellations of assembled RVA genomes. The genotypes of all assembled sequences were determined according to the guidelines of RV Classification Working Group, representing the genotype of VP7-VP4-VP6-VP1-VP2-VP3-NSP1-NSP2-NSP3-NSP4-NSP5. For mono-infection, each row represents one genotype constellation with the colour block used to illustrate different genotype patterns, such as common human types of Wa-like (orange) and DS-1-like (purple), and other less common genotypes shown in other colour blocks as indicated. The number of each of the genotype constellation identified in samples from human and pigs were given in the column Host count for Human and Pigs, respectively. For mixed infection where two distinct contigs were assembled for at least one segment, the genotypes of both contigs were given and indicated the host where strains were identified.

identified in four pig samples collected from two other farms (Fig. 4A); the farms that raised these pigs are also about 35 km away from the residential location of the 16020_7 case (Fig. 4B). RVA strains with the aforementioned genome constellation have been identified in paediatric diarrhoeal patients in Hungary (Papp et al. 2013), Argentina (Degiuseppe et al. 2013), Paraguay (Martinez et al. 2014), and Nicaragua (Bucardo et al. 2012), and were similarly thought to be of zoonotic origin (porcine-like). Given the diversity of geographic locations of reported zoonotic cases over a 15-year period, it is difficult to determine if this strain is sustained as a rare variant in human-to-human infections or has undergone multiple cross-species jumps from a porcine reservoir or an intermediate host.
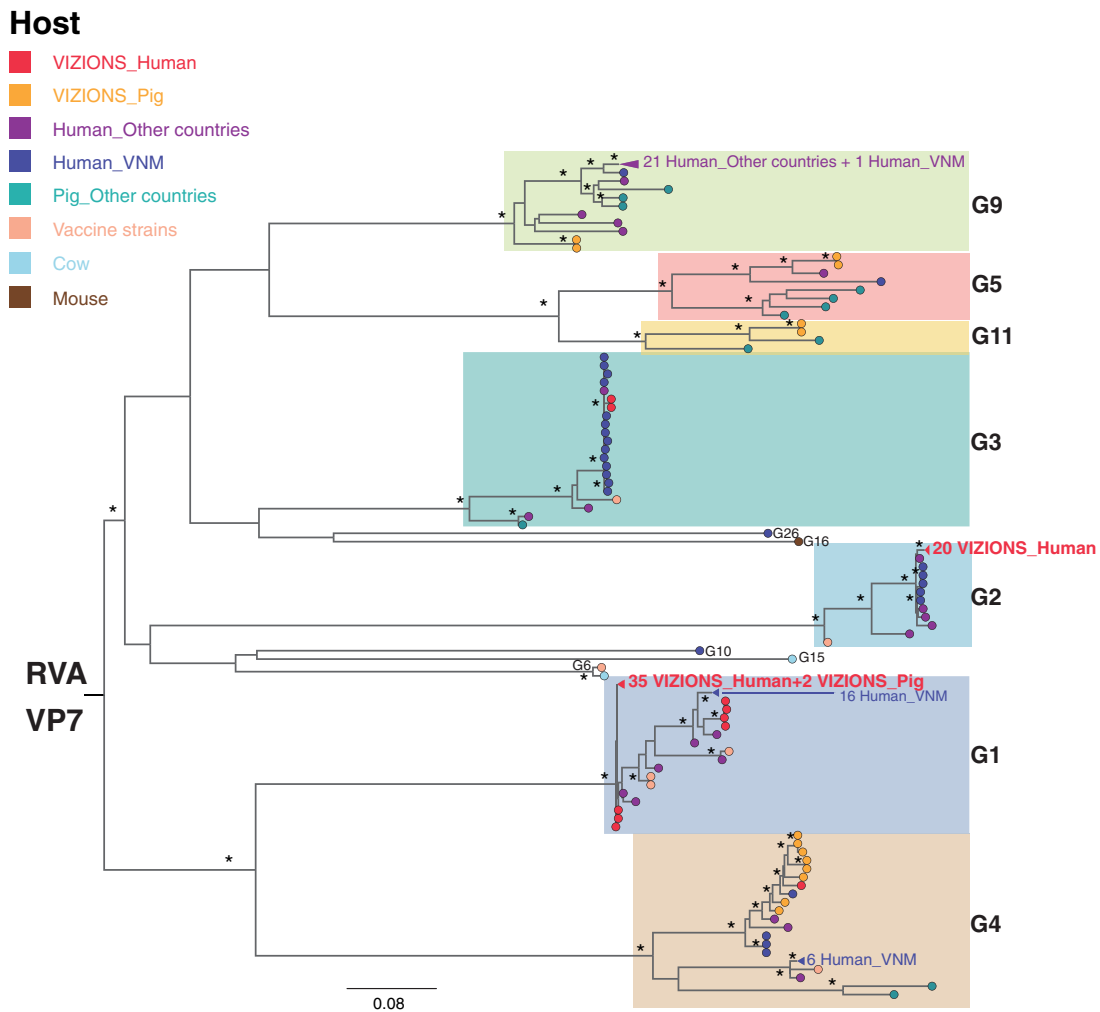
Figure 3. ML phylogenetic trees inferred from the assembled nucleotide sequences for RVA VP7 and VP4 genes. (A) ML tree of VP7 gene showed genetic relationships between sequences from this study and additional sequences of corresponding segments from GenBank. Strains were coloured according to the host species from which the strain was identified, and "VIZIONS" in the annotation refers to strains identified from this study. Tree is mid-point rooted for the purpose of clarity and bootstrap values of ≥75 percent are shown as asterisks. All horizontal branch lengths are drawn to the scale of nucleotide substitutions per site. (B) ML tree of VP4 gene showed genetic relationships between sequences from this study and additional sequences of corresponding segments from GenBank. The pattern of tree visualization is consistent with VP7 tree, see description of Fig. 3A for more information.

Phylogenetically, all eleven segments of the human 16020_7 and porcine 14250_9 viruses belonged to lineages comprising porcine and/or porcine-origin human sequences (Fig. 3A and B and Supplementary Fig. S3A and B). Genetic distance suggested that the strain 16020_7 was most similar to porcine strains: TM-a (for VP1; 96 percent nt similarity), CMP45 (NSP3; 93 percent), and porcine-origin human strain 30378 (NSP2; 99 percent) (Fig. 5). The remaining segments were most similar to porcine RVA strains obtained from this study, including the porcine 14150_53 (NSP1; 98.7 percent and NSP4; 99.1 percent) and 14225_44 (VP2; 99.2 percent and NSP5; 99.1 percent) strains. The capsid proteins of 16020_7 were most similar to the VP7 sequences of porcine samples 12129_48, 12129_49 and 12070_4 (G4 type; 97.2 percent), to the VP4 of pig 14226_39 (98.7 percent) and VP6 of pig 14226_42 (99.5 percent) (Fig. 5). The porcine sample 14250_9, despite possessing the same genotype constellation as 16020_7, shared the highest nucleotide similarity in only two internal genes, VP3 (97.5 percent) and NSP5 (99.1 percent), to the corresponding segments of strain 16020_7. Compared with the RVA vaccine

Rotarix, 16020_7 was relatively dissimilar, sharing as low as 75.1 and 75.6 percent nt similarity for the VP4 and VP7 segments, respectively (Fig. 5).

In this unusual genotype constellation reported, the NSP3 T7 type is a rare genotype that was first identified in a cow in Great Britain in 1973 (Ward et al. 1984), then in a bovine-like human strain (Mukherjee et al. 2011), and later in pigs (Martel-Paradis et al. 2013), porcine–bovine human reassortant (Wang et al. 2010) and porcine-like human strains (Bucardo et al. 2012; Zeller et al. 2012; Degiuseppe et al. 2013; Papp et al. 2013; Martinez et al. 2014) in various geographical locations. The inferred evolutionary rate of RVA NSP3 sequences bearing T7 genotype was 1.3261×10^{-3} substitutions per site per year (95 percent highest posterior density (HPD): 8.624×10^{-4}–1.793×10^{-3}), which is slightly lower than the estimated evolutionary rates for RVA VP7 capsid gene of 1.66×10^{-3} and 1.87×10^{-3} substitutions/site/year for G12 and G9 genotypes, respectively (Matthijnssens et al. 2010). The time-stamped MCC tree also indicate an interconnection among different host species indicating several host
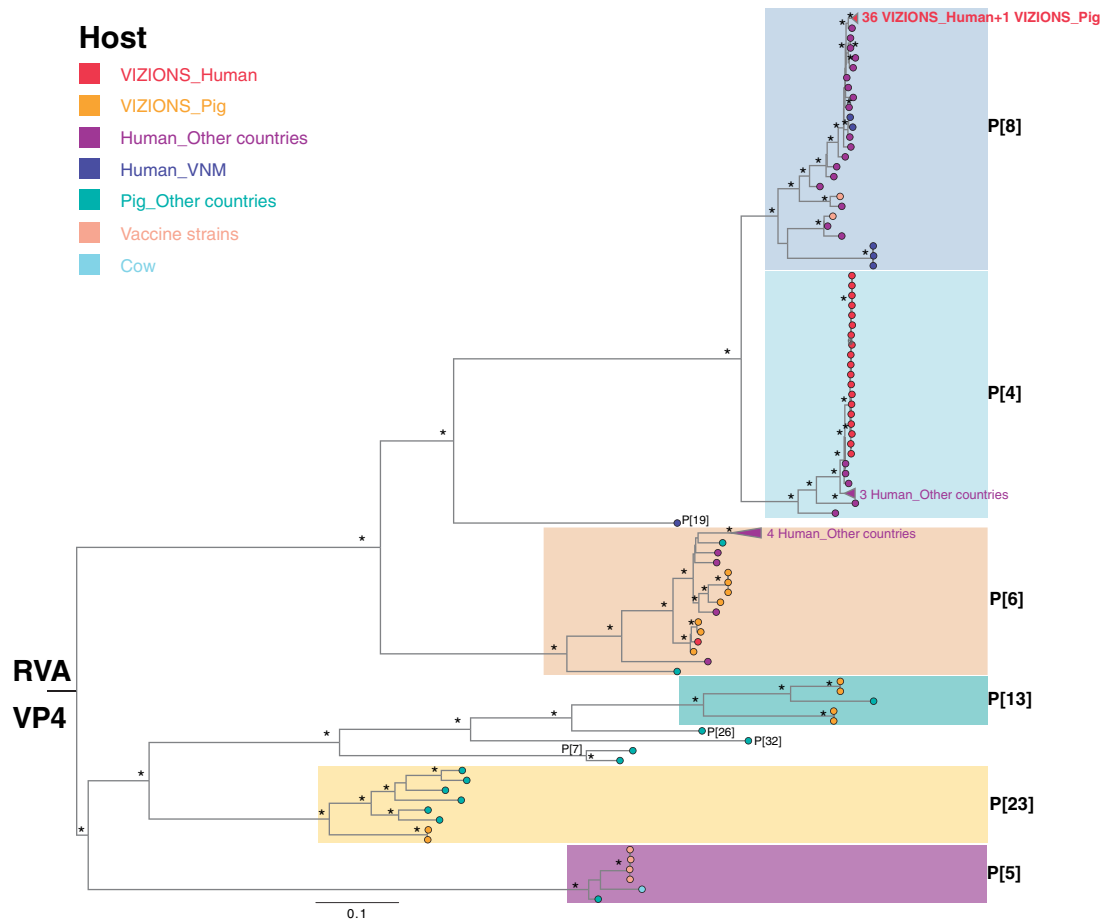
Figure 3. Continued

jump events particularly between pig and human hosts, suggesting that viral zoonotic chatter may occur more frequently than hitherto reported (Fig. 4C).

3.5. RV group H (RVH), group B (RVB), and C (RVC)

RVH was identified in five Vietnamese pigs (3.33 percent; 5/150) at several time points and locations with no temporal or geographical associations (Fig. 6), suggesting that these infections were sporadic and not linked to a single local outbreak. Furthermore, phylogenetic trees were inferred for all RVH segments to investigate the genetic diversity, comparing the RVH strains identified in this study with RVH sequences retrieved from GenBank (Fig. 6 and Supplementary Fig. S6). In general, RVH sequences typically clustered according to the host species, that is, all porcine RVH sequences belonged to a lineage that is separated from human or cow RVH lineages. Within the porcine clade of the VP6 gene (Fig. 6), sequences fell into two lineages: one lineage comprising sequences from USA and Japan, and the other lineage of Brazilian and Vietnamese sequences. The evolutionary rate was estimated to be 5.195×10^{-3} substitutions/site/year for sequences in the porcine lineage of RVH VP6 (95 percent HPD: 1.865×10^{-3}–8.976×10^{-3}).

RVB was found in seven pigs (4.67 percent; 7/150) and RVC was identified in six pigs (4 percent; 6/150). Phylogenetic trees of all segments of RVB and RVC showed that the local porcine sequences belonged to lineages comprising porcine sequences from other geographical locations for RVB (Fig. 7 and Supplementary Fig. S4) and RVC (Fig. 7 and Supplementary Fig. S5). In both RVB and RVC groups, the porcine lineages were relatively distant from lineages comprising of human sequences (Fig. 7 and Supplementary Figs. S4 and S5).

4. Discussion

This study represents the first unbiased genome-wide surveillance, targeting simultaneously multiple groups of rotaviruses infecting humans and animals in the same geographical location. Prior to this study, there were only three subgenomic RVC sequences (<300 nt) and nine complete or nearly complete RVA genomes reported from Vietnam in GenBank with no data on RVB and RVH. Data from this study document genomic sequences from sixty human RVA and thirty-one porcine RV (groups A, B, C, and H), providing the largest available collection of genome sequences from human and pigs from a single location in general and from Vietnam in particular. This is also the first report and the first genome characterisation of RVB, RVC and RVH from Vietnamese pigs.

Among the RVA, we identified a human (sample 16020_7) and a porcine sample (sample 14250_9) with atypical RVA genotype constellation G4-P[6]-I1-R1-C1-M1-A8-N1-T7-E1-H1, detected for the first time in Vietnam and in Asia. This variant may have originated from a direct zoonotic transmission or from reassortment event(s) involving porcine and
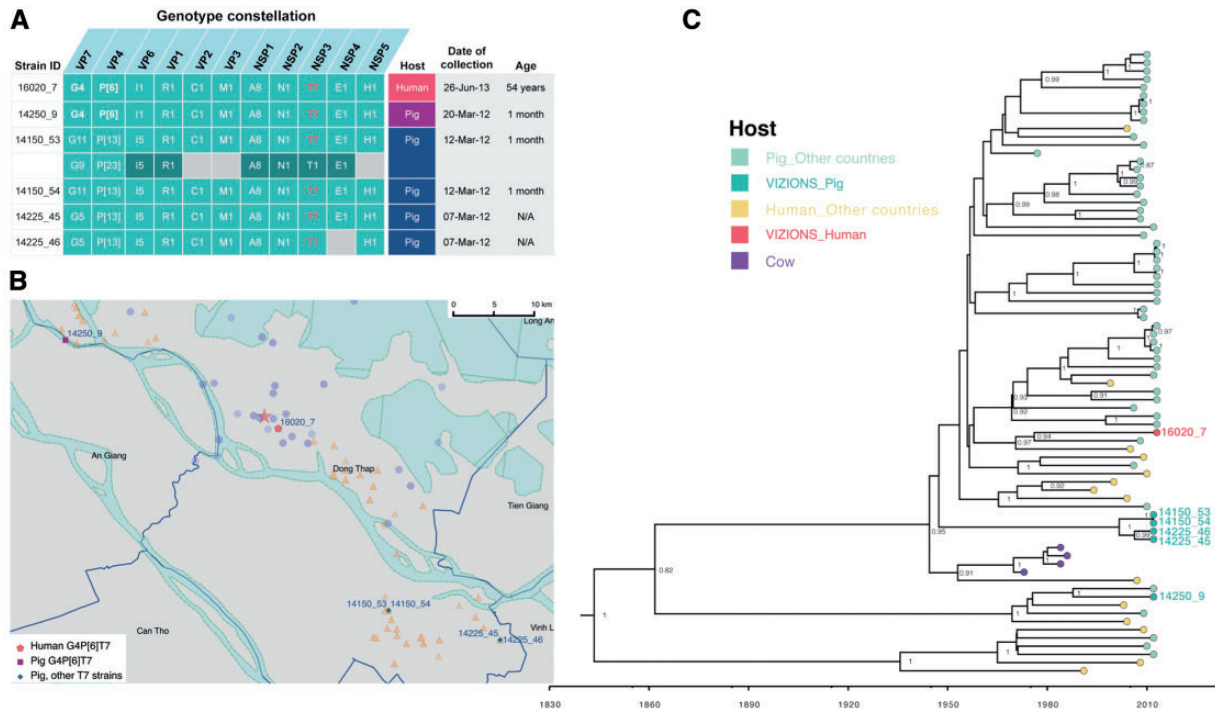
Figure 4. The analysis of RVA zoonotic strain G4P[6]T7 in the study. (**A**) The genotype constellation of the case in investigation 16020_7 and other porcine RVA strains with the NSP3 T7 genotype. Details on dates of collection and age of host are given. The colour code in the host column is consistent with colour illustration of corresponding case in the map (panel B) in this figure. (**B**) The geographical location of the human case's residency and pig farms that raised the pigs infected with RVA NSP3 T7 strains overlaid on the total sampling area (as shown in Supplementary Fig. S1). The colouring of the human case and pig farms is consistent with colour code presented in the column "Host" in panel of this figure. The red star indicates the Dong Thap Provincial Hospital where diarrhoeal patients were admitted. The map scale bar is shown in the units of geometric km. See Supplementary Fig. S1 for more information. (**C**) The time-resolved phylogenetic tree of RVA NSP3 T7 genotype sequences comparing local versus global sequences. Reference sequences were retrieved from GenBank [N = 69, excluded the duplicate sequence for TM-a strain (JX290174) and BP1901 (KF835960) which is 15 aa shorter than the complete ORF]. Strains were coloured according to the host species from which the strain was identified, and "VIZIONS" in the annotation refers to strains identified from this study. Sequences identified from this study were highlighted in red (human case 16020_7) or in turquoise (porcine sequences). The first T7 sequence identified (in a cow in 1973) was indicated in purple. Posterior probabilities of internal nodes with values ≥0.75 are shown and the scale axis indicates time in year of strain identification.

porcine-origin human strains. The human RVA strain (16020_7) was identified in a sample from a 54-year-old patient, admitted to the hospital due to acute diarrhoea. RV was the sole enteric pathogen identified from the stool sample and no other common viral and bacterial diarrhoeal pathogens were found by diagnostic testing (Dung et al. 2012; Thompson et al. 2015) for norovirus, astrovirus, sapovirus, adenovirus F, and aichi virus, *Shigella* spp., *Salmonella* spp., and *Campylobacter* spp. (data not shown). Although adults can be infected with RVA, such infections in immuno-competent individuals are typically asymptomatic, self-limiting, or cause mild disease (Anderson and Weber 2004). The RV infection in this particular case required hospitalization suggesting a moderate–severe end of the clinical spectrum of diarrhoeal disease. Although further studies are required to determine their significance and relevance in human and animal diseases, it is tempting to suggest that this atypical strain may be the cause of the moderate–severe diarrhoeal disease. The close proximity between humans and pigs and common use of river water (Mekong Delta River, Fig. 4B) for daily activities and farming might present an enhanced risk of transmitting water-borne infectious pathogens, in this case providing a plausible zoonotic route of atypical RV transmission.

Compared with RVA, human and animal infections with RVB, RVC, and RVH are not well understood and the detection rates for these groups of viruses are relatively low (Ghosh and Kobayashi 2011). This is probably because the majority of RV investigations have been focused on RVA given its clinical and public health relevance and the large genetic distance among these groups of rotaviruses as compared with RVA, which would likely be missed by commonly used diagnostic assays. Recently, there have been increasing numbers of reports on RV groups B, C, and H in animals (Marthaler et al. 2012; Amimo, Vlasova, and Saif 2013; Marthaler et al. 2013; Lahon et al. 2014; Marthaler et al. 2014; Molinari, Alfieri, and Alfieri 2014; Molinari et al. 2014; Marton et al. 2015; Otto et al. 2015) and humans (Alam et al. 2013; Lahon et al. 2013; Zhirakovskaia et al. 2016), which possibly reflect improved molecular methods to detect these viruses rather than an actual increase in their prevalence. In Vietnam, the frequencies and relative role in human and animal disease of RVB, RVC, and RVH viruses are not yet known.

Our study does have limitations. First, the sample size of 146 human and 150 porcine samples is relatively small. Despite this size, we were able to identify a potential zoonotic infection. This provides a baseline frequency for zoonotic infections and suggests that this may be occurring at a higher rate than previously considered. Second, the disease status of the sampled pigs was not well defined and there was no follow-up beyond the sampling time point so the clinical spectrum of diarrhoeal disease (e.g. mild, moderate, or severe) in the pigs is unknown. However, the primary objective of this study was characterization of RV pathogenesis or causation of the diarrhoeal disease. The presence of RV material at sufficiently high titres to allow full genome sequencing is consistent with these animals being a common source of the virus for movement to other species.

Strain ID	Country	VP7	VP4	VP6	VP1	VP2	VP3	NSP1	NSP2	NSP3	NSP4	NSP5
16020_7	VNM	G4	P[6]	I1	R1	C1	M1	A8	N1	T7	E1	H1
14250_9	VNM	G4	P[6]	I1	R1	C1	M1	A8	N1	T7	E1	H1
		96.9%	92.5%	97.3%	85.4%	97.9%	97.5%	81.6%	97.5%	87.5%	96.5%	99.1%
14150_53	VNM	G11/G9	P[13]/P[23]	I5/I5	R1/R1	C1	M1	A8/A8	N1N1	T1/T7	E1/E1	H1
		75.7%/72%	68.2%/72.2%	83.1%/83.5%	87.3%/85.7%	97.2%	86.5%	98.7%/81.9%	86.5%/93.4%	83.1%/91.6%	97.4%/99.1%	94.4%
14150_54	VNM	G11	P[13]	I5	R1	C1	M1	A8	N1	T7	E1	H1
		72.0%	68.4%	83.2%	85.6%	97.3%	86.4%	98.1%	86.4%	91.6%	97.5%	95.3%
14225_45	VNM	G5	P[13]	I5	R1	C1	M1	A8	N1	T7	E1	H1
		73.8%	67.7%	82.5%	85.6%	98.8%	98.3%	95.7%	97.1%	92.1%	88.7%	98.6%
14225_46	VNM	G5	P[13]	I5	R1	C1	M1	A8	N1	T7	X	H1
		73.9%	67.6%	82.3%	85.5%	98.7%	96.0%	95.7%	94.7%	91.9%	X	98.4%
14225_44	VNM	G4	P[6]	I1	R1	C1	M1	A8	N1	T1	E1	H1
		96.7%	98.3%	99.4%	87.3%	99.2%	86.6%	79.2%	93.8%	83.3%	90.9%	99.1%
14226_42	VNM	G4	P[6]	I1	R1	C1	M1	A8	N1	T1	E1	H1
		96.7%	97.9%	99.5%	87.3%	99.1%	86.2%	82.1%	95.5%	83.3%	90.9%	99.1%
14226_39	VNM	G4	P[6]	I1	R1	C1	M1	A8	N1	T1	E1	H1
		97.1%	98.7%	83.5%	85.8%	90.0%	87.1%	82.5%	95.5%	82.5%	90.1%	98.1%
14249_23	VNM	G4	P[6]	I1	R1	C1	M1	A8	N1	T1	E1	H1
		97.0%	92.4%	98.1%	87.7%	86.7%	86.5%	95.9%	96.3%	82.3%	96.6%	96.0%
12070_4	VNM	G4/G1	P[8]	I1	R1	C1	M1	A1/A8	N1/N1	T1/T1	E1/E1	H1/H1
		97.2%/75.6%	75.4%	99.0%	92.5%	87.5%	87.6%	76.9%/80.9%	97.7%/87.6%	84%/82.6%	91.4%/90.3%	98.3%/94%
12129_48	VNM	G4	P[6]	I1	R1	C1	M1	A8	N1	T1	E1	H1
		97.2%	91.3%	98.1%	86.8%	86.6%	86.1%	95.9%	96.6%	83.1%	95.6%	96.1%
12129_49	VNM	G4	P[6]	I1	R1	C1	M1	A8	N1	T1	E1	H1
		97.2%	92.4%	98.1%	66.6%	82.0%	86.4%	95.9%	96.5%	83.1%	96.3%	95.9%
30378§	VNM	G26	P[19]	I5	R1	C1	M1	A8	N1	T1	E1	H1
		74.8%	76.2%	83.4%	87.3%	94.1%	96.9%	82.2%	98.6%	83.8%	96.4%	98.8%
OL§	NCA	G4	P[6]	I1	R1	C1	M1	A8	N1	T7	E1	H1
		86.0%	86.2%	89.8%	86.1%	87.0%	85.3%	82.4%	92.7%	91.6%	88.0%	95.7%
E931§	CHN	G4	P[6]	I1	R1	C1	M1	A8	N1	T1	E1	H1
		97.1%	92.7%	97.7%	94.6%	91.0%	85.1%	82.2%	93.1%	84.1%	96.8%	95.4%
R946§	CHN	G3	P[6]	I1	R1	C1	M1	A1	N1	T1	E1	H1
		73.8%	94.5%	96.5%	95.4%	90.8%	84.8%	78.0%	88.9%	84.2%	92.8%	94.7%
CMP45/08	THA	G9	P[23]	I5	R1	C1	M1	A8	N1	T7	E1	H1
		75.5%	71.2%	83.1%	86.9%	92.5%	88.1%	92.4%	93.5%	93.1%	89.1%	93.3%
TM-a	CHN	G9	P[23]	I5	R1	C1	M1	A8	N1	T7	E1	H1
		75.5%	71.1%	82.9%	95.9%	90.5%	86.6%	80.0%	89.0%	88.2%	91.8%	94.9%
Wa	USA	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1
		74.8%	75.4%	92.9%	88.8%	88.1%	90.3%	76.5%	88.0%	83.1%	93.9%	93.3%
Rotarix	USA	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1
		75.6%	75.1%	92.5%	88.6%	88.1%	90.2%	76.2%	88.1%	83.1%	93.3%	92.8%

Figure 5. Genotype constellations of RVA 16020_7 compared with representative human and animal RVA of known genotypes. Segments are bold in red and shaded in green to indicate the segments with highest nucleotide sequence similarities to that of strain 16020_7. The strain was coloured according the host from which strain was identified, blue indicates human host, pink for pigs and green for vaccine component. All porcine samples from this study were shaded in light blue and the human case of interest (16020_7) was shaded in grey. §Human strains were previously shown to have porcine origin. Country of isolation abbreviation, VNM, Vietnam; NCA, Nicaragua; CHN, China; THA, Thailand; USA, United States of America.

Our findings indicate that porcine faecal material is a source of novel and possibly zoonotic viruses.

It is likely that with the ubiquity and falling costs of sequencing, the unbiased virus sequencing described here will become an important component of infectious disease surveillance and rapid responses to outbreaks (Woolhouse, Rambaut, and Kellam 2015). The ideal sampling rate, sample numbers, and geographical relationship between humans and animals for genetic surveillance are still being defined but this study provides a good starting point for future efforts. Even within the relatively modest sample set of 296 human and animal enteric samples, a considerable RV genetic diversity was observed including a potential zoonosis. The integration of targeted sampling, sequencing, and phylogeography or phylogenetics in different places in the world, perhaps informed by other risk mapping (Hay et al. 2013; Rabaa et al. 2015) has the ability to inform surveillance and to monitor zoonotic pathogens in human and animals.

Supplementary data

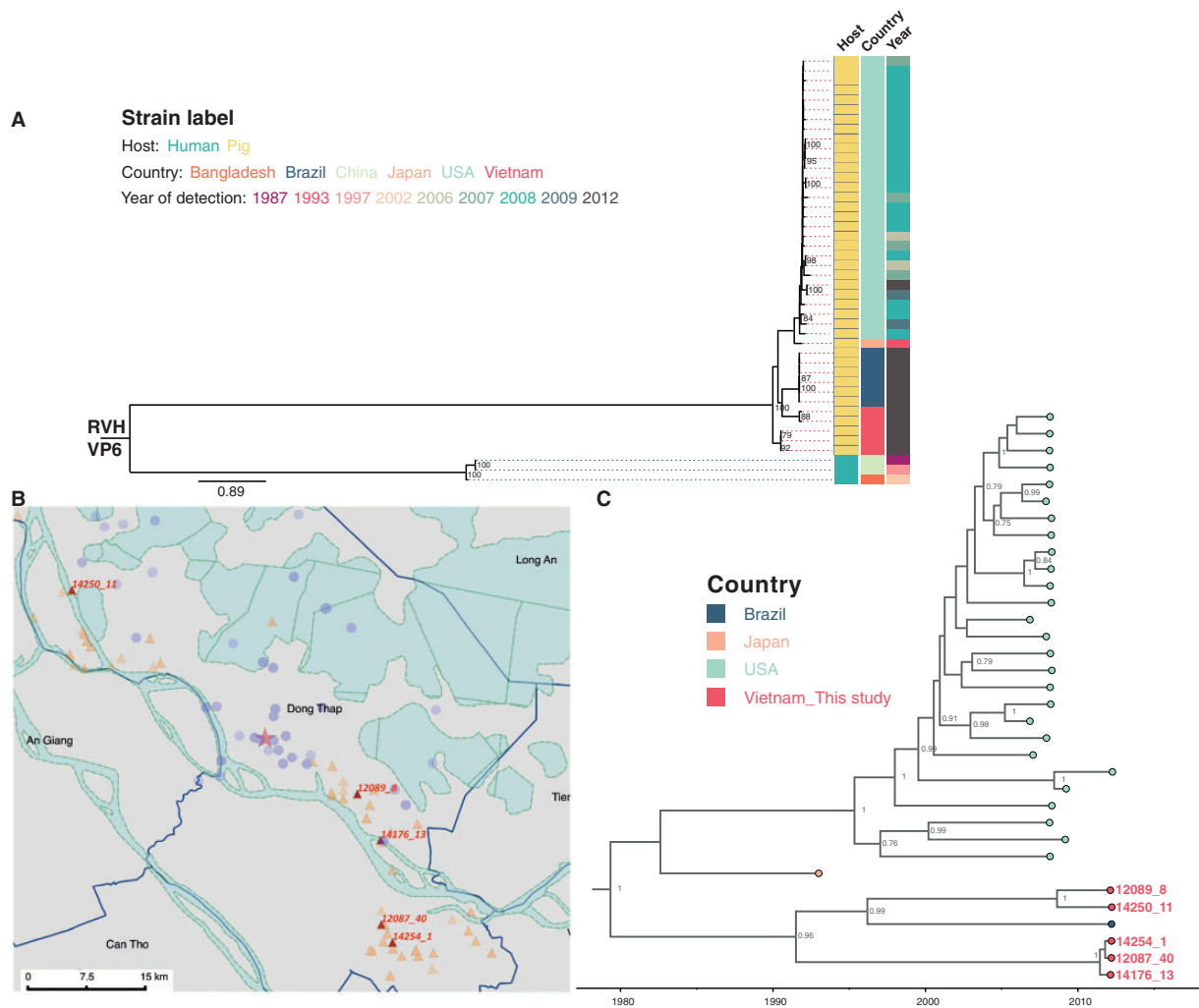Supplementary data are available at *Virus Evolution* online.

Figure 6. The analysis of RVH VP6 gene segment. (**A**) ML phylogenetic tree of RVH VP6 sequences. The RVH VP6 sequences in this study (N = 5) were compared with available full-length RVH VP6 sequences retrieved from GenBank (N = 39). Tree is mid-point rooted for the purpose of clarity and only bootstrap values of ≥ 75 percent are shown. All horizontal branch lengths are drawn to the scale of nucleotide substitutions per site in the tree. Strains were colour coded according to the host associated with the strain, the country where the strains were identified and the year of strain detection. (**B**) The geographical locations of pig farms that raised the pigs infected with RVH identified in this study. The farms were illustrated as red triangles with strain ID given, overlaid on the overall sampling area as shown in Supplementary Fig. S1. The red star indicates the Dong Thap Provincial Hospital. The map scale bar is shown in the units of geometric km. Refer to Supplementary Fig. S1 for more information on the background and provincial features colouring. (**C**) Time-resolved phylogenetic tree of porcine RVH VP6 sequences comparing local versus global sequences. Strains were coloured by the country where strains were identified. Porcine sequences identified from this study were highlighted in red, with strain ID given to link with geographical locations on map in panel B. Strains from Brazil were coloured in dark blue; orange indicates the Japanese porcine strain, and light green refers to the porcine strain from the USA. Posterior probabilities of internal nodes with values ≥0.75 are shown and the scale axis indicates time in year of strain identification.

Conflict of interest

The opinions expressed by the authors contributing to this article do not necessarily reflect the opinions of the Wellcome Trust, the British Council or the institutions with which the authors are affiliated or the grantee bodies. The authors declare no competing interests.
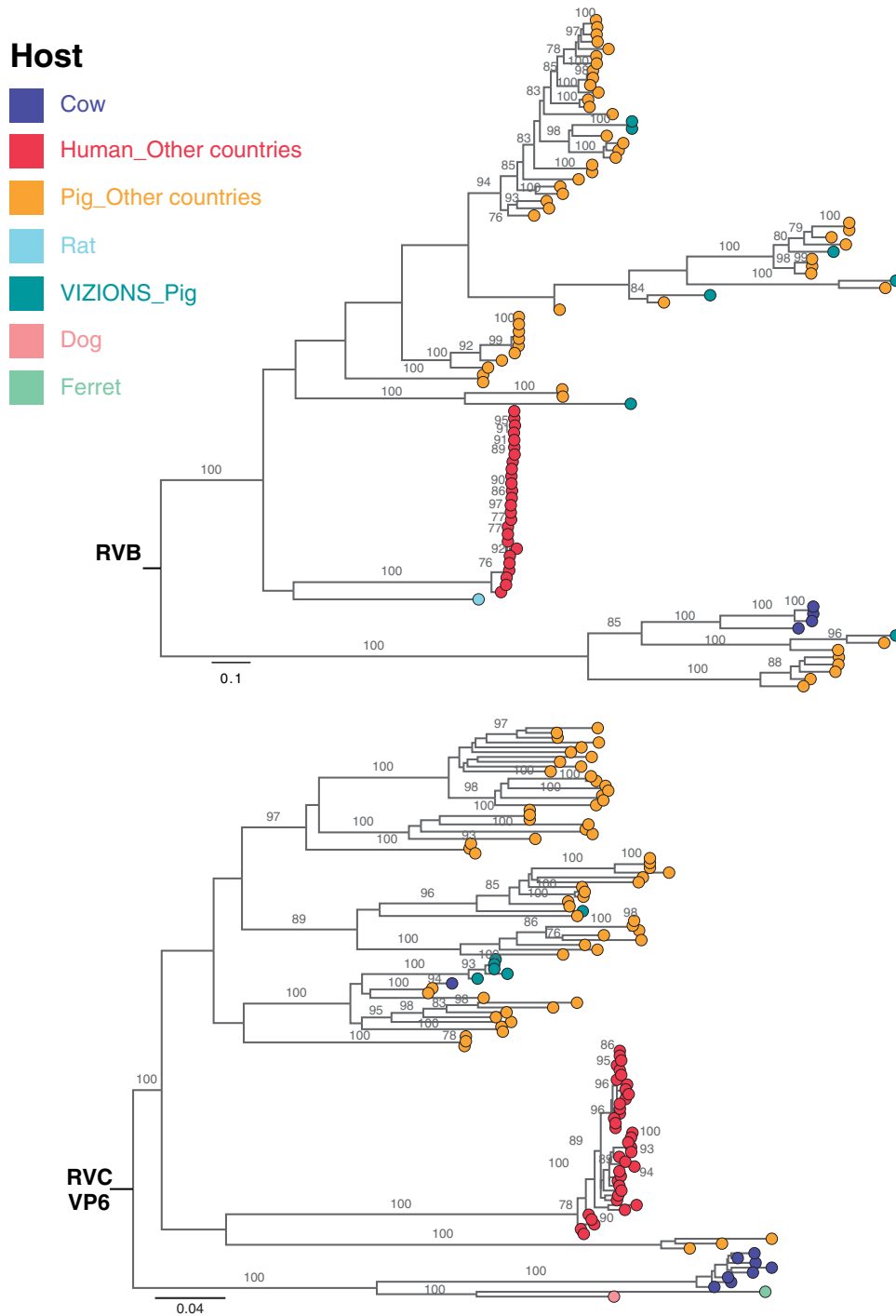
Figure 7. ML phylogenetic trees inferred from the assembled nucleotide sequences for VP6 gene of RVB and RVC. ML trees of RVB and RVC VP6 segments showed genetic relationships between assembled sequences from this study and full-length reference sequences of corresponding segments retrieved from GenBank. Trees are mid-point rooted for the purpose of clarity and only bootstrap values of ≥75 percent are shown. Scale bars are in the unit of nucleotide substitutions per site. Strains were coloured according to the host species that the sequences were identified from, and "VIZIONS" in the annotation refers to strains identified from this study.

References

Alam, M. M., et al. (2013) 'The First Identification of Rotavirus B from Children and Adults with Acute Diarrhoea in Kathmandu, Nepal', *Tropical Medicine and Health*, 41: 129–34.

Amimo, J. O., Vlasova, A. N., and Saif, L. J. (2013) 'Prevalence and Genetic Heterogeneity of Porcine Group C Rotaviruses in Nursing and Weaned Piglets in Ohio, USA and Identification of a Potential New VP4 Genotype', *Veterinary Microbiology*, 164: 27–38.

Anderson, E. J., and Weber, S. G. (2004) 'Rotavirus Infection in Adults', *Lancet Infectious Diseases*, 4: 91–9.

Anh, D. D., et al. (2006) 'The Burden of Rotavirus Diarrhea in Khanh Hoa Province, Vietnam: Baseline Assessment for a Rotavirus Vaccine Trial', *Pediatrics Infectious Disease Journal*, 25: 37–40.

Armah, G. E., et al. (2010) 'Efficacy of Pentavalent Rotavirus Vaccine Against Severe Rotavirus Gastroenteritis in Infants in Developing Countries in Sub-Saharan Africa: A Randomised, Double-Blind, Placebo-Controlled Trial', *Lancet*, 376: 606–14.

Atchison, C., et al. (2010) 'Spatiotemporal Dynamics of Rotavirus Disease in Europe: Can Climate or Demographic Variability Explain the Patterns Observed', *The Pediatric Infectious Disease Journal*, 29: 566–8.

Bankevich, A., et al. (2012) 'SPAdes: a New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing', *Journal of Computational Biology*, 19: 455–77.

Boetzer, M., et al. (2011) 'Scaffolding Pre-Assembled Contigs Using SSPACE', *Bioinformatics*, 27: 578–9.

Boom, R., et al. (1990) 'Rapid and Simple Method for Purification of Nucleic Acids', *Journal of Clinical Microbiology*, 28: 495–503.

Breiman, R. F., et al. (2012) 'Analyses of Health Outcomes From the 5 Sites Participating in the Africa and Asia Clinical Efficacy Trials of the Oral Pentavalent Rotavirus Vaccine', *Vaccine*, 30/1: A24–9.

Bucardo, F., et al. (2012) 'Vaccine-Derived NSP2 Segment in Rotaviruses From Vaccinated Children With Gastroenteritis in Nicaragua', *Infection, Genetics and Evolution*, 12: 1282–94.

Cook, N., et al. (2004) 'The Zoonotic Potential of Rotavirus', *Journal of Infectious*, 48: 289–302.

——, et al. (1990) 'Global Seasonality of Rotavirus Infections', *Bulletin of the World Health Organization*, 66: 171–7.

Cotten, M., et al. (2014) 'Full Genome Virus Detection in Fecal Samples Using Sensitive Nucleic Acid Preparation, Deep Sequencing, and a Novel Iterative Sequence Classification Algorithm', *PLoS One*, 9: e93269.

de Vries, M., et al. (2011) 'A Sensitive Assay for Virus Discovery in Respiratory Clinical Samples', *PLoS One*, 6: e16118.

——, et al. (2012) 'Performance of VIDISCA-454 in Feces-Suspensions and Serum', *Viruses*, 4: 1328–34.

Degiuseppe, J. I., et al. (2013) 'Complete Genome Analyses of G4P[6] Rotavirus Detected in Argentinean Children With Diarrhoea Provides Evidence Of Interspecies Transmission From Swine', *Clinical Microbiology and Infection*, 19: E367–71.

Desselberger, U. (2014) 'Rotaviruses', *Virus Research*, 190C: 75–96.

Doan, L. T. P., et al. (2003) 'Epidemiological Features of Rotavirus Infection Among Hospitalized Children with Gastroenteristis in Ho Chi Minh City, Vietnam', *Journal of Medical Virology*, 69: 588–94.

Dóró, R., et al. (2015) 'Zoonotic Transmission of Rotavirus: Surveillance and Control', *Expert Review of Anti-Infective Therapy*, 13: 1337–50.

Drummond, A. J., and Rambaut, A. (2007) 'BEAST: Bayesian Evolutionary Analysis by Sampling Trees', *BMC Evolutionary Biology*, 7: 214.

Dung, T. T. N., et al. (2012) 'The Validation and Utility of a Quantitative One-Step Multiplex RT Real-Time PCR Targeting Rotavirus A and Norovirus', *Journal of Virological Methods*, 187: 1–6.

Edgar, R. C. (2004) 'MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput', *Nucleic Acids Research*, 32: 1792–7.

—— (2010) 'Search and Clustering Orders of Magnitude Faster Than BLAST', *Bioinformatics*, 26: 2460–1.

Endoh, D., et al. (2005) 'Species-Independent Detection of RNA Virus by Representational Difference Analysis Using Non-Ribosomal Hexanucleotides for Reverse Transcription', *Nucleic Acids Research*, 33: e65.

Estes, M. K., et al. (2007) 'Rotaviruses', *in* D. M., Knipe, and P. M., Howley (eds.), *Fields Virology*, 5th ed., pp. 1917–1974. Philadelphia, PA: Lippincott Williams & Wilkins.

Gautam, R., et al. (2013) 'Real-Time RT-PCR Assays to Differentiate Wild-Type Group A Rotavirus Strains From Rotarix(®) and RotaTeq(®) Vaccine Strains in Stool Samples', *Human Vaccines and Immunotherapeutics*, 10: 767–77.

Gentsch, J. R., et al. (2005) 'Serotype Diversity and Reassortment Between Human and Animal Rotavirus Strains: Implications for Rotavirus Vaccine Programs', *Journal of Infectious Diseases*, 192: S146–59.

Ghosh, S., and Kobayashi, N. (2011) 'Whole-Genomic Analysis of Rotavirus Strains: Current Status and Future Prospects', *Future Microbiology*, 6: 1049–65.

——. (2014) 'Exotic Rotaviruses in Animals and Rotaviruses in Exotic Animals', *Virus Disease*, 25: 158–72.

Glass, R. I., et al. (2014) 'Rotavirus Vaccines: Successes and Challenges', *Journal of Infection*, 68/1: S9–18.

Hay, S. I., et al. (2013) 'Global Mapping of Infectious Disease', *Philosophical Transactions of the Royal Society of London Series B, Biological Science*, 368: 20120250.

Jere, K. C., et al. (2011) 'Whole Genome Analyses of African G2, G8, G9, and G12 Rotavirus Strains Using Sequence-Independent Amplification and 454-Pyrosequencing', *Journal of Medical Virology*, 83: 2018–42.

Lahon, A., et al. (2013) 'Group B Rotavirus Infection in Patients with Acute Gastroenteritis from India: 1994–1995 and 2004–2010', *Epidemiology and Infection*, 141: 969–75.

—— (2014) 'Molecular Characterization of Group B Rotavirus Circulating in Pigs From India: Identification of a Strain Bearing a Novel VP7 Genotype, G21', *Veterinary Microbiology*, 174: 342–52.

Landaeta, M. E., et al. (2003) 'Characterization of Rotaviruses Causing Diarrhoea in Vietnamese Children', *Annals of Tropical Medicine and Parasitology*, 97: 53–9.

Larsson, A. (2014) 'AliView: a Fast and Lightweight Alignment Viewer and Editor for Large Datasets', *Bioinformatics*, 30: 3276–8.

Maes, P., et al. (2009) 'RotaC: a Web-Based Tool for the Complete Genome Classification of Group A Rotaviruses', *BMC Microbiology*, 9: 238.

Magagula, N. B., et al. (2015) 'Whole Genome Analyses of G1P[8] Rotavirus Strains From Vaccinated and Non-Vaccinated South African Children Presenting With Diarrhea', *Journal of Medical Virology*, 87: 79–101.

Martella, V., et al. (2010) 'Zoonotic Aspects of Rotaviruses', *Veterinary Microbiology*, 140: 246–55.

Martel-Paradis, O., et al. (2013) 'Full-Length Genome Analysis of G2, G9 and G11 Porcine Group A Rotaviruses', *Veternary Microbiology*, 162: 94–102.

Marthaler, D., et al. (2012) 'Detection of Substantial Porcine Group B Rotavirus Genetic Diversity in the United States, Resulting in a Modified Classification Proposal for G Genotypes', *Virology*, 433: 85–96.

—— (2013) 'Identification, Phylogenetic Analysis and Classification of Porcine Group C Rotavirus VP7 Sequences From the United States and Canada', *Virology*, 446: 189–98.

—— (2014) 'Widespread Rotavirus H in Domesticated Pigs, United States', *Emerging Infectious Diseases Journal*, 20: 1195–8.

—— (2014) 'Whole-Genome Analyses Reveals the Animal Origin of a Rotavirus G4P[6] Detected in a Child With Severe Diarrhea', *Infection, Genetics and Evolution*, 27C: 156–62.

Marton, S., et al. (2015) 'Canine Rotavirus C Strain Detected in Hungary Shows Marked Genotype Diversity', *Journal of General Virology*, 96: 3059–71.

Matthijnssens, J., and Van Ranst, M. (2012) 'Genotype Constellation and Evolution of Group A Rotaviruses Infecting Humans', *Current Opinion in Virology*, 2: 426–33.

————, et al. (2008) 'Recommendations for the Classification of Group A Rotaviruses Using all 11 Genomic RNA Segments', *Archives of Virology*, 153: 1621–9.

————, et al. (2010) 'Molecular and Biological Characterization of the 5 Human-Bovine Rotavirus (WC3)-Based Reassortant Strains of the Pentavalent Rotavirus Vaccine, RotaTeq', *Virology*, 403: 111–27.

————, et al. (2010) 'Phylodynamic Analyses of Rotavirus Genotypes G9 and G12 Underscore Their Potential for Swift Global Spread', *Molecular Biology and Evolution*, 27: 2431–6.

————, et al. (2011) 'Uniformity of Rotavirus Strain Nomenclature Proposed by the Rotavirus Classification Working Group (RCWG)', *Archives of Virology*, 156: 1397–413.

———— (2012) 'VP6-Sequence-Based Cutoff Values as a Criterion for Rotavirus Species Demarcation', *Archives of Virology*, 157: 1177–82.

Mihalov-kovács, E., et al. (2015) 'Candidate New Rotavirus Species in Sheltered Dogs, Hungary', *Emerging Infectious Diseases*, 21: 660–3.

Molinari, B. L. D., Alfieri, A. F., and Alfieri, A. A. (2014) 'Genetic Variability of VP6, VP7, VP4, and NSP4 Genes of Porcine Rotavirus Group H Detected in Brazil', *Virus Research*, 197: 48–53.

————, et al. (2014) 'Species H Rotavirus Detected in Piglets with Diarrhea, Brazil, 2012', *Emerging Infectious Diseases Journal*, 20: 2012–5.

Mukherjee, A., et al. (2011) 'Full Genomic Analyses of Human Rotavirus G4P[4], G4P[6], G9P[19] and G10P[6] Strains from North-Eastern India: Evidence for Interspecies Transmission and Complex Reassortment Events', *Clinical Microbiology and Infection*, 17: 1343–6.

Munnink, B. B. O., et al. (2016) 'Complete Genome Characterization of Two Wild-Type Measles Viruses from Vietnamese Infants During the 2014 Outbreak', *Genome*, 4: 1–2.

Ngo, T. C., et al. (2009) 'Molecular Epidemiology of Rotavirus Diarrhoea Among Children in Haiphong, Vietnam: the Emergence of G3 Rotavirus', *Vaccine*, 27/5: F75–80.

Nguyen, L.-T., et al. (2015) 'IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies', *Molecular Biology and Evolution*, 32: 268–74.

————, et al. (2007) 'Diversity of Viruses Associated With Acute Gastroenteritis in Children Hospitalized with Diarrhea in Ho Chi Minh City, Vietnam', *Journal of Medical Virology*, 79: 582–90.

————, et al. (2004) 'Diarrhea Caused by Rotavirus in Children Less than 5 Years of Age in Hanoi, Vietnam', *Journal of Clinical Microbiology*, 42: 5745–50.

————, et al. (2001) 'The Epidemiology and Disease Burden of Rotavirus in Vietnam: Sentinel Surveillance at 6 Hospitals', *Journal of Infectious Diseases*, 183: 1707–12.

Nishio, O., et al. (2000) 'Rotavirus Infection Among Infants with Diarrhea in Vietnam', *Pediatrics International*, 42: 422–4.

Nomikou, K., et al. (2015) 'Widespread Reassortment Shapes the Evolution and Epidemiology of Bluetongue Virus Following European Invasion', *PLOS Pathogens*, 11: e1005056.

Nyaga, M. M., et al. (2015) 'Whole Genome Detection of Rotavirus Mixed Infections in Human, Porcine and Bovine Samples Co-infected with Various Rotavirus Strains Collected From Sub-Saharan Africa', *Infection, Genetics and Evolution*, 31: 321–34.

Otto, P. H., et al. (2015) 'Detection of Rotavirus Species A, B and C in Domestic Mammalian Animals with Diarrhoea and Genotyping of Bovine Species A Rotavirus Strains', *Veterinary Microbiology*, 179: 168–76.

Papp, H., et al. (2013) 'Zoonotic Transmission of Reassortant Porcine G4P[6] Rotaviruses in Hungarian Pediatric Patients Identified Sporadically Over a 15 Year Period', *Infection, Genetics and Evolution*, 19: 71–80.

Pitzer, V. E., et al. (2009) 'Demographic Variability, Vaccination, and the Spatiotemporal Dynamics of Rotavirus Epidemics', *Science*, 325: 290–4.

————, et al. (2011) 'Influence of Birth Rates and Transmission Rates on the Global Seasonality of Rotavirus Incidence', *Journal of the Royal Society Interface*, 8: 1584–93.

Rabaa, M. A., et al. (2015) 'The Vietnam Initiative on Zoonotic Infections (VIZIONS): A Strategic Approach to Studying Emerging Zoonotic Infectious Diseases', *Ecohealth*, 12: 726–35.

Rambaut, A., et al. (2016) 'Exploring the Temporal Structure of Heterochronous Sequences Using TempEst (Formerly Path-O-Gen)', *Virus Evolution*, 2: vew007.

Ramig, R. F., et al. (2005) 'Rotavirus', *in* C. M., Fauquet, M. A., Mayo, J., Maniloff, U., Desselberger, and L. A., Ball (eds.), *Virus Taxonomy: Eight Report of the International Committee on Taxonomy of Viruses*, pp. 484–496. Amsterdam: Elsevier Academic Press.

Ruiz-Palacios, G. M., et al. (2006) 'Safety and Efficacy of an Attenuated Vaccine Against Severe Rotavirus Gastroenteritis', *New England Journal of Medicine*, 354: 11–22.

Santos, N., and Hoshino, Y. (2005) 'Global Distribution of Rotavirus Serotypes/Genotypes and Its Implication for the Development and Implementation of an Effective Rotavirus Vaccine', *Reviews in Medical Virology*, 15: 29–56

Tamura, T., et al. (2010) 'Molecular Epidemiological Study of Rotavirus and Norovirus Infections Among Children with Acute Gastroenteritis in Nha Trang, Vietnam, December 2005–June 2006', *Japan Journal of Infectious Diseases*, 63: 405–11.

Tate, J. E., et al. (2012) '2008 Estimate of Worldwide Rotavirus-Associated Mortality in Children Younger Than 5 Years Before the Introduction of Universal Rotavirus Vaccination Programmes: A Systematic Review and Meta-Analysis', *Lancet Infectious Diseases*, 12: 136–41.

———— (2016) 'Global, Regional, and National Estimates of Rotavirus Mortality in Children <5 Years of Age, 2000–2013', *Clinical Infectious Diseases*, 62: S96–105.

Thompson, C. N., et al. (2015) 'A Prospective Multi-Center Observational Study of Children Hospitalized with Diarrhea in Ho Chi Minh City, Vietnam', *The American Journal of Tropical Medicine and Hygiene*, 92: 1045–52.

Tra My, P. V., et al. (2011) 'The Emergence of Rotavirus G12 and the Prevalence of Enteric Viruses in Hospitalized Pediatric Diarrheal Patients in Southern Vietnam', *The American Journal of Tropical Medicine and Hygiene*, 85: 768–75.

Trang, N. V., et al. (2012) 'Detection and Molecular Characterization of Noroviruses and Sapoviruses in Children Admitted to Hospital With Acute Gastroenteritis in Vietnam', *Journal of Medical Virology*, 297: 290–7.

Trojnar, E., et al. (2013) 'Identification of an Avian Group A Rotavirus Containing a Novel VP4 Gene with a Close Relationship to Those of Mammalian Rotaviruses', *Journal of General Virology*, 94: 136–42.

Van Man, N., et al. (2005) 'Epidemiological Profile and Burden of Rotavirus Diarrhea in Vietnam: 5 Years of Sentinel Hospital Surveillance, 1998–2003', *Journal of Infectious Diseases*, 192: S127–32.

Vesikari, T., et al. (2006) 'Safety and Efficacy of a Pentavalent Human-Bovine (WC3) Reassortant Rotavirus Vaccine', *New England Journal of Medicine*, 354: 23–33.

Vu Tra My, P., et al. (2014) 'Novel Porcine-Like Human G26P[19] Rotavirus Identified in Hospitalized Pediatric Diarrhea Patients in Ho Chi Minh City', *Journal of General Virology*, 95: 2727–33.

Wang, Y., et al. (2010) 'Full Genomic Analysis of a Porcine—Bovine Reassortant G4P[6] Rotavirus Strain R479 Isolated From an Infant in China', *Journal of Medical Virology*, 1102: 1094–102.

Ward, C. W., et al. (1984) 'Nucleotide Sequence of Gene Segment 9 Encoding a Nonstructural Protein of UK Bovine Rotavirus', *Virology*, 134: 249–53.

Watson, S. J., et al. (2013) 'Viral Population Analysis and Minority-Variant Detection Using Short Read Next-Generation Sequencing', *Philosophical Transactions of the Royal Society of London B: Biological Science*, 368: 20120205.

Woolhouse, M. E. J., Rambaut, A., and Kellam, P. (2015) 'Lessons from Ebola: Improving Infectious Disease Surveillance to Inform Outbreak Management', *Science Translational Medicine*, 7: 1–9.

World Health Organization. 2005. The treatment of diarrhoea: A manual for physicians and other senior health workers.

——. 2015. VietNam: WHO statistical profile.

Zaman, K., et al. (2010) 'Efficacy of Pentavalent Rotavirus Vaccine Against Severe Rotavirus Gastroenteritis in Infants in Developing Countries in Asia: a Randomised, Double-Blind, Placebo-Controlled Trial', *Lancet*, 376: 615–23.

Zeller, M., et al. (2012) 'Genetic Analyses Reveal Differences in the VP7 and VP4 Antigenic Epitopes Between Human Rotaviruses Circulating in Belgium and Rotaviruses in Rotarix and RotaTeq', *Journal of Clinical Microbiology*, 50: 966–76.

Zhirakovskaia, E., et al. (2016) 'First Genetic Characterization of Rotavirus C in Russia', *Infection, Genetics and Evolution*, 39: 1–8.