****

**Part II Genetics Research Project:**

**Molecular Evolution of Rotavirus**

Qiancheng Xionga\*

aUniversity of Cambridge, Department of Genetics

\*Corresponding author: qx217@cam.ac.uk

Supervisors:

Dr Bethany L. Dearloveb

Dr Simon D. W. Frostb

bUniversity of Cambridge, Department of Veterinary Medicine

Abstract: 156 words. Main text: 3496 words.

**TABLE OF CONTENTS**

|  |  |
| --- | --- |
| **Section** | **Page** |
|  |  |
| *Acknowledgements* | *3* |
|  |  |
| *Abstract* | *4* |
|  |  |
| *Introduction* | *5* |
|  |  |
| *Methods* | *9* |
|  |  |
| *Results* | *12* |
|  |  |
| *Discussion* | *16* |
|  |  |
| *References* | *20* |
|  |  |
| *Data Tables* | *30* |
|  |  |
| *Figures* | *36* |
|  |  |
| *Appendix A - Software and Settings Used for Sequence Processing* | *45* |
|  |  |
| *Appendix B - Software and Settings Used for Recombination Checks* | *46* |
|  |  |
| *Appendix C - Summary Details of Datasets* | *47* |
|  |  |
| *Appendix D - BEAST Settings* | *48* |
|  |  |
|  |  |

**ACKNOWLEDGEMENTS**

I would to thank Bethany L. Dearlove and Simon D. W. Frost for their assistance and guidance on the project. I am also grateful to Mukarram Hossain for his help on the identification of segment numbers.

Lastly, I would like to thank the Department of Genetics and the Department of Veterinary Medicine for organising this project.

**ABSTRACT**

Rotavirus (RV) is a major cause of acute gastroenteritis and child mortality worldwide. Phylogenetic analyses may aid the development of RV vaccines or antivirals. Through estimating selection pressures and mutation rates at both the gene level and the site-by-site level, this project demonstrated evolutionary differences between RV species, proteins and hosts. In particular, higher mutation rate and diversifying selection estimates for species B may facilitate its expansion to geographical regions outside Asia. Capsid protein VP6 was also found to be a possible vaccine target due to its low mutation rate and diversifying selection estimates. Separately, reconstructed effective population size and estimated mutation rates for species A showed no significant change in genetic variation after the introduction of modern vaccines in 2005, but new sites under diversifying selection were identified on its gene for the vaccine targeted protein VP4. Finally, reconstructed phylogenies showed ongoing zoonotic transmission in species A but not in species B and C.

**INTRODUCTION**

Rotavirus (RV) is a major cause of acute gastroenteritis in human children (Bishop et al., 1973; Flewett et al. 1973), as well as in other mammals (Adams & Kraft, 1963; Malherbe & Harwin, 1963; Mebus et al., 1969) and birds (Otto et al., 2012; Trojnar et al., 2009). It still causes around 453,000 deaths annually among young children (Tate et al., 2012). More than half of the reported childhood mortality cases were reported from less-developed nations (Parashar et al., 2009; Tate et al., 2012). Despite lower mortality rates, RV is also an issue in developed nations. Before the wide use of vaccines, RV only caused an estimated 25-37 deaths annually in the United States. However, there were approximately 59,600 hospitalizations and the cost to the economy was around $893 million (Fischer et al., 2007; Widdowson et al., 2007; Esposito et al., 2011). RV-induced morbidity and mortality in farm animals also result in further economic losses (House, 1978; Martella et al., 2010).

*Viral Genome*

RV is a Group III double-stranded RNA virus in the *Reoviridae* family. Its genome is around 18,000 bp and consists of 11 segments numbered by decreasing size, ranging from 3,300-3,500 bp (segment 1) to 630-730 bp (segment 11) (Nagashima et al., 2008). These segments encode for six structural viral proteins (VPs: VP1-VP4, VP6 and VP7) and six non-structural proteins (NSPs: NSP1-NSP6) (Desselberger, 2014). NSP6 is only produced in some species of RV by using an out-of-phase open reading frame in the segment encoding for NSP5 (Mohan & Atreya, 2001). VP4 can also be cleaved by a protease such as trypsin to form VP5 and VP8 (Crawford et al., 2001). As the individual segments are numbered according to size, a segment with the same number can encode different proteins in different species or strains (Chen et al., 2002; Mlera et al., 2012; Desselberger, 2014).

*Classifications and Epidemiology*

Based on the antigenicity of VP6, RV is classified into eight different species (RVA-RVH) which differ in epidemiology (Matthijnssens et al., 2012). Of the eight, RVA, RVB and RVC are more widely studied. RVA is the most prevalent worldwide, accounting for >90% of all reported RV infections (Gentsch et al., 2005; Santos & Hoshino, 2004). RVC infections have also been reported globally, but in smaller numbers (Kuzuya et al., 2007; Schnagl et al., 2004; Araújo et al., 2011; Chang et al., 1999). In contrast, RVB infections typically occur in Asia (Tao et al., 1984; Kelkar and Zade, 2004; Saiada et al., 2010).

While RVA, RVB and RVC can be found in both humans and non-human animals, RVD, RVE, RVF and RVG are mainly restricted to animals such as birds or pigs (Estes & Kapikian, 2007). Reports on RVH infections are currently limited to pigs and humans (Jiang et al., 2008; Marthaler et al., 2014; Molinari et al., 2014). However, since most samples are obtained from patients and diseased animals, non-clinical or sub-clinical species such as RVC tend to be under-represented in prevalence estimates (Collins et al., 2008).

Additional G/P-genotyping (Glycoprotein/Protease-sensitive protein) using the antigenicities of VP7 and VP4, and genotyping based on all RV proteins, have been established for RVA (Matthijnssens et al., 2008; Matthijnssens et al., 2011) The same G-genotyping method has also been adapted for RVB and RVC (Marthaler et al., 2012; Marthaler et al., 2013).

*Immunity and Vaccines*

Children infected with RV will obtain immunity for subsequent infections (Bishop et al., 1983). For adults who have experienced repeated viral challenges, infections tend to be asymptomatic as memory B cells provide both long-term and heterotypic protection (Franco et al. 2006). This is achieved via humoral antibodies for VP4 and VP7, which are on the capsid (Offit, 1994; Feng et al., 1997; Jiang et al., 2002). These observations spurred the development of live-attenuated vaccines for RVA, including the monovalent Rotarix® (GlaxoSmithKline) in 2005 and the pentavalent RotaTeq® (Merck) in 2006 (Ruiz-Palacios et al., 2006; Vesikari et al., 2006). Both vaccines were generally successful, with vaccine efficacy of around 90% in the United States (Patel & Parashar, 2009; Tate et al., 2011a; Tate et al., 2011b; Cortese et al., 2013). They also have been found to provide herd immunity for unvaccinated children (Anderson et al., 2013; Yi and Anderson, 2013; Pollard et al., 2015).

*Significance of Study*

Despite successes with existing vaccines, a few issues affect their implementation, particularly in less-developed countries. First, RV has high mutation rates across its genome. This is due to frequent point mutations contributed by the error-prone RNA-dependent RNA polymerase (Blackhall et al., 1996), as well as re-assortments arising from co-infections of different strains (Iturriza-Gomara et al., 2001) and zoonotic transmissions (Todd et al., 2010; Steyer et al., 2008; Matthijnssens et al., 2011; Steyer et al., 2013). As a result, a rapidly evolving RV genome may render genotypic-specific vaccines less effective over time. Secondly, the efficacy of vaccines is found to be reduced in less developed countries, and plausibly requires further modifications to better adapt to different host populations and geographical locations (Lopman et al., 2012; Glass et al., 2014).

As such, phylogenetic analyses of RV may elucidate details on its future evolution, and the genotypic differences due to epidemiological factors. It will also aid the development of alternate vaccines or antivirals for RVA and other RV species. For instance, VP6 is also a potential vaccine target, as VP6-specific antibodies developed after infection were also shown to be protective (Burns et al., 1996; Jalilvand et al., 2015).

*Purpose*

Previous studies indicated that species of RV are separated into two major clades (Kindler et al., 2013). Separation between mammalian and avian RVA were also observed, along with host-specific evolution in RVB and RVC (Kindler et al., 2013). Additionally, genome constellations of RV were found to be stable, and a study on one strain of RVA only identified 6 codons under diversifying selection (Mcdonald et al., 2009; Zeller et al., 2015).

This project aims to expand the scope of current phylogenetic analysis on RV by using whole genome datasets from the most common species - RVA, RVB and RVC. By detecting diversifying selection and estimating mutation rates, differences in the evolution of RV between species, segments, and hosts shall be identified. Effects of modern vaccines on RV evolution after implementation in 2005 will also be studied through selection, mutation rate and effective population size estimates. Finally, the presence of zoonotic transmission will be analysed using reconstructed phylogenies.

**METHODS**

*Sequence Acquisition*

49,121 RV DNA sequences were downloaded from GenBank (Benson et al., 2013) using the keyword “rotavirus[Organism]” on January 18th, 2015, together with metadata on species, host, country, sampling date, genotype, and segment number. The temporal range of the sequences is 1958-2015.

*Processing*

Sequences were processed and converted into 33 separate datasets of multiple sequence alignments (MSAs) to facilitate analysis at both the gene level and the site-by-site level. Categorisation by proteins was chosen as the genes are numbered differently within each RV species. The software and settings used for sequence processing can be found in Appendix A.

Recombinants were also removed, as recombination would create mosaic sequences with sites of different evolutionary history. This would generate misleading phylogenetic analyses through inaccurate estimations of branch lengths, selection pressures and molecular clocks (Schierup & Hein, 2000; Posada, 2001; Posada & Crandall, 2002). The software and settings used for recombination checks can be found in Appendix B.

To compare the effects of modern vaccines on RVA evolution, sequences from 10 years post-vaccine introduction (2006-2015) and pre-vaccine introduction (1996-2005) were selected from datasets of genes for capsid VP4, VP6 and VP7 to form six separate datasets. VP4 and VP7 were chosen as they are the main vaccine targets, while the non-target VP6 served as a comparison.

The summary statistics of all processed datasets can be found in Appendix C.

*Selection Analyses*

Selection pressure differences between datasets were estimated by calculating non-synonymous substitution rates (β) and synonymous substitution rates (α) using the HyPhy package (version 2.2.4; Pond et al., 2005) on the Datamonkey (Pond & Frost, 2005; Delport et al., 2010) cluster. Datasets for RVA were randomly down-sampled to 500 sequences without replacement in order to satisfy upload limits of Datamonkey. The Fast Unconstrained Bayesian AppRoximation (FUBAR; Murrell et al., 2013) with default settings was used to identify sites which have experienced pervasive diversification, while MEME (Murrell et al., 2012) with default settings was used to identify sites which have experienced episodic diversification. In addition, gene level β and α were estimated for each dataset using FUBAR.

To compare selection differences between human and non-human hosts, phylogenetic trees for RVB datasets (NSP1, NSP2, NSP5, VP6 and VP7) and RVC datasets (NSP1, NSP2, NSP4, VP1, VP4 and VP6) were generated using IQ-TREE (version 1.4.0; Nguyen et al., 2014). These datasets were chosen as they had very good sequence alignments with minimal gaps. The trees were then partitioned in HyPhy using the Slatkin-Maddison (Slatkin & Maddison, 1989) test, and were subsequently subjected to analyses where different ratios of β to α were fitted to tree branches corresponding to human and non-human hosts.

*Molecular Clocks*

The BEAST (version 1.8.3; Drummond et al., 2012) software package was used to estimate the substitution rates within each dataset under the GMRF Bayesian Skyride coalescent model. This estimation utilised sampling dates in addition to the MSAs. BEAST was chosen over other programs because it allows substitution rates to vary between lineages, and allows uncertainty in the sampling dates to be incorporated.

All BEAST analyses were carried out on the CIPRES cluster (version 3.3; Miller et al., 2010). BEAST settings can be found in Appendix D. To complete analyses within the time limits of CIPRES (168 hours), every RVA dataset was randomly down-sampled without replacement to 100 sequences. Each BEAST analysis was repeated once to check for abnormalities, and to increase the effective sample size of Bayesian analysis. Log files from each repeat of 1 billion iterations were combined within BEAST’s LogCombiner (version 1.8.2), before being visualised in Tracer (version 1.6.0; Rambaut et al., 2014) with a 10% burn-in period.

*Effective Population Size and Zoonotic Transmission*

BEAST can also reconstruct phylogenies and generate maximum clade credibility (MCC) trees. Trees from two repeats of analyses were combined within LogCombiner. To study changes in effective population sizes due to the introduction of modern vaccines, GMRF Skyride reconstructions of effective population sizes based on RVA VP7 were visualised using the combined log files and combined tree files in Tracer for a duration of 20 years before 2015. As for identification of zoonotic transmissions, VP7 tree files for each species were processed using BEAST’s TreeAnnotator (version 1.8.2) and visualised in Figtree (version 1.4.2; Rambaut, 2014). VP7 was chosen for both population size and transmission analyses as its datasets are the largest within each species.

**RESULTS**

*Selection Analyses*

Selection pressures were compared across species and proteins at the gene level using FUBAR (Table 1; Figure 1). To enable comparison, a ratio of β to α for each dataset was computed at the gene level. A value closer to “0” suggests stronger purifying selection, while a value closer to “1” suggests stronger diversifying selection. Sometimes, higher β/α values could also imply relaxations of functional constraints, rather than more pressure to diversify.

Between species, it was observed that RVA had much stronger diversifying selection on genes for NSP4 and VP7 than other species. Likewise, RVB had much stronger diversifying selection on the gene for NSP3 than other species. On average, RVA had higher diversifying selection than RVB and RVC across both NSP genes (RVA: 0.219; RVB: 0.198; RVC: 0.174) and VP genes (RVA: 0.169; RVB: 0.127; RVC: 0.077). However, significance testing for these differences was unavailable as variance estimates were not provided by Datamonkey.

Across proteins, the average diversifying selection estimate for NSP genes (0.197) was much higher than that of VP genes (0.118). Within NSP genes, the gene for NSP2 had the weakest diversifying selection while the gene for NSP4 had the highest average diversifying selection estimate (0.317). The gene for VP6 had the weakest diversifying selection across VP genes while the gene for VP7 had the highest average diversifying selection estimate (0.191).

Comparisons of selection pressures between human and non-human hosts were achieved using RVB and RVC datasets in HyPhy (Table 2; Figure 2a-b). Ratios of β to α assigned to host branches in each dataset are measures of diversifying selection. The significance of host effect on selection pressure was indicated by a difference in AIC values between models with and without host effect: larger positive values indicate higher significance. Comparing within RVB datasets, host effect on selection was significant for genes of NSP1, NSP2 and NSP5. Within RVC datasets, host effect on selection was significant for genes of NSP1, NSP2, VP1 and VP4. However, within each species, there was no indication of human hosts consistently resulting in higher or lower diversifying selection pressures. Only the gene for NSP1 demonstrated weaker diversifying selection in humans for both species.

Gene level FUBAR was also used to study the effect of modern vaccines on selection pressures of genes for RVA VP4, VP7 and VP6 (Table 3; Figure 3). As only VP4 and VP7 were targeted by vaccines, no vaccine effect was expected for VP6. After the introduction of vaccines, diversifying selection weakened for genes of VP4 and VP7, while the gene for VP6 experienced a slight increase in diversifying selection.

Combining site-by-site results from MEME and FUBAR, eight codons across the three species were identified to be under both pervasive and episodic diversifying selection (Table 4). There was no common gene or codon among species that experienced both forms of diversifying selection.

The same site-by-site analysis was applied to study vaccine effects on selection pressures for RVA (Table 5). While the majority of sites on the gene of RVA VP4 remained under purifying selection, 2 new codons came under diversifying selection after the introduction of modern vaccines. While codon 161 was not detected by MEME, visualising the site-by-site changes using FUBAR confirmed its new diversifying selection pressure (Figure 4). Codon 311 was found to be an artefact due to gaps in the majority of sequences at this codon.

*Molecular Clocks*

The evolutionary rates over time for each dataset was estimated in BEAST using sampling dates in addition to MSAs (Table 6; Figure 5).

Between the three species, RVB had the highest average mutation rate. This could be attributed to high mutation rates in its genes for NSP1, NSP4, NSP5, VP2 and VP3.

Between proteins, the average mutation rate of NSP genes was higher than that of VP genes, but on the same order of magnitude. Across NSP genes, genes for NSP1 and NSP4 had the highest average mutations rates while the gene for NSP2 had the lowest average mutation rate. For the VP genes, the gene for VP3 had the highest average mutation rate, while genes for VP1 and VP6 had the lowest average mutation rates. The gene for RVA VP7 had a much larger 95% highest posterior density (HPD) interval than all other genes. This was not a result of inaccuracy as the BEAST MCMC chains converged on 2 separate attempts for the same estimation. Thus, this may be indicative of huge genetic diversities within the gene for RVA VP7.

BEAST analyses were also conducted for genes of VP4, VP7 and VP6, with 1996-2005 and 2006-2015 datasets of 500 randomly selected sequences. However, the MCMC probability distributions failed to converge after one week on CIPRES for two attempts.

*Changes in Effective Population Size Due to Modern Vaccines*

The effective population sizes of RVA were reconstructed using tree files of the largest RVA VP7 dataset for the temporal range of 1996-2015 (Figure 6). There was no significant change in the effective population sizes of RVA after the introduction of modern vaccines in 2005.

*Zoonotic Transmission*

MCC trees were reconstructed using BEAST analyses of VP7 datasets (Figures 7, 8 & 9). Phylogenies of RVB and RVC diverged between human and non-human hosts early in their evolutionary histories and showed no signs of zoonotic transmission. However, phylogeny of RVA suggests zoonotic transmission as there was no clear separation of clades for human hosts and non-human hosts. Furthermore, a few RVA branches in humans branched off from those in non-human hosts as recent as 5-6 years ago and vice versa.

**DISCUSSION**

*Comparison between Species*

RVA had a much higher diversifying selection estimate and variance of mutation rates for VP7, correlating to its huge variety of G genotypes based on VP7. As such, it may not be efficient to base future vaccines or drugs solely on a few genotypes of VP7.

RVB had higher mutation rates than RVA and RVC, especially in genes for NSP1 and NSP4. This observation was supported by high diversifying selection estimates. As both the NSP1 involved in suppressing host interferon response and the NSP4 enterotoxin are involved in host invasion, the high mutation rates and diversifying selection pressures may aid colonisation of new hosts and niches. Hence, although RVB is currently limited to Asia, we should be vigilant against its expansion to other geographical regions.

*Comparison between Proteins*

Between proteins, the structural VP proteins had lower diversifying selection and mutation rate estimates than non-structural NSP proteins. This suggests that the structural proteins are less tolerant to change or are less targeted by the immune response.

Proteins with core functions, such as the VP1 RNA polymerase and the NSP2 NTPase/helicase essential for dsRNA synthesis, also had lower diversifying selection and mutation rate estimates than those involved in host invasion, such as the NSP4 enterotoxin and the NSP1 involved in suppressing host interferon response. In particular, the high mutation rate of the gene for NSP1 may be responsible for evading innate immunity and correlates with the high incidence of the disease in young children.

Negatively-selected genes may be suitable targets for vaccines because changes to generate escape mutants are likely to be intolerable (Suzuki, 2004). Hence, this study also confirms the viability of utilising VP6 as a future vaccine and drug target across all species as its diversifying selection and mutation rate estimates were consistently the lowest.

*Comparison between Hosts*

Only the gene for NSP1 showed a lower diversifying selection for human hosts in both RVB and RVC. The host effect was insignificant in many other genes and human hosts did not always result in stronger or weaker diversifying selection. Hence, the host type is not a good predictor for the direction of selection in RVB and RVC. Similar analyses can be done for RVA in future but only on individual strains of RVA due to the complexity of RVA MSAs when taken as a whole.

*Effect of Modern Vaccines*

The selection pressures at the gene level became more negative for VP4 and VP7 after the introduction of vaccines in 2005. It is possible that the vaccine decreased the viability of diversifying selection due to the diverse but less fit genotypes it generated.

Although findings on effective population sizes of RVA showed no change after the introduction of modern vaccines and agree with previous works (Zeller et al., 2015), the new sites on the gene for VP4 under diversifying selection are of concern. In particular, codon 2 showed diversification from Ala (gct) to Val (gtt) while codon 161 was STOP (tag) changing to a variety of other amino acids, allowing read-throughs. Such changes arising *de novo* could potentially render vaccines less effective over time by altering the target VP4 proteins. Hence, we should continue to monitor diversifying selection on the genes of vaccine targets.

*Zoonotic Transmission*

The main RV hosts other than humans are domesticated animals including pigs and cows. Despite close human contact with these animals, only RVA displayed both direct and reverse zoonosis. As previous studies (Martella et al., 2010) also showed evidence of zoonotic transmission for RVA, it is crucial that future RVA vaccines continue to target sequences of both human and non-human RVA strains.

Even though RVB and RVC from animals cannot infect humans directly, we should still be wary of the possibility of genetic reassortments between these species and the zoonotic RVA during co-infection events in animals. It is more likely for RVB or RVC to gain zoonotic capabilities from RVA than to evolve the function *de novo*. Therefore, phylogenies of RVB and RVC should be reviewed frequently for any RVA sequences in their genomes.

*Future Directions*

BEAST analyses for molecular clock comparisons between time periods before and after vaccine introduction could not be completed after 2 attempts. Subsequent attempts could extend the run time to 5 weeks on other computing clusters without time limits, or employ a simpler model such as the Extend Bayesian Skyline Plot (Drummond, 2005) to decrease computing demands.

Furthermore, datasets were down-sampled to fit the time limitations of this project. Full size datasets can be utilised in future to generate more accurate estimates of mutation rates and selection pressures.

Lastly, other species RVD-RVH can also be studied. As they were omitted from this project due to insufficient sequence data, greater sampling efforts could yield useful insights, especially for human infecting species such as RVH.

**REFERENCES**

1. Adams, W. R., and Kraft, L. M. (1963). Epizootic Diarrhea of Infant Mice: Identification of the Etiologic Agent. Science *141*, 359–360.
2. Anderson, E. J., Shippee, D. B., Weinrobe, M. H., Davila, M. D., Katz, B. Z., Reddy, S., Cuyugan, M. G. K. P., Lee, S. Y., Simons, Y. M., Yogev, R., et al. (2013). Indirect Protection of Adults From Rotavirus by Pediatric Rotavirus Vaccination. Clinical Infectious Diseases *56*, 755–760.
3. Araújo, I. T., Heinemann, M. B., Fialho, A. M., and Leite, J. P. G. (2011). Detection and Molecular Characterization of Human Group C Rotavirus in Brazil. Intervirology *54*, 261–267.
4. Benson, D. A., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Sayers, E. W. (2013). GenBank. Nucleic Acids Research Nucl. Acids Res. *42*.
5. Bishop, R. F., Barnes, G. L., Cipriani, E., and Lund, J. S. (1983). Clinical Immunity after Neonatal Rotavirus Infection. New England Journal of Medicine N Engl J Med *309*, 72–76.
6. Bishop, R., Davidson, G., Holmes, I., and Ruck, B. (1973). Virus Particles In Epithelial Cells Of Duodenal Mucosa From Children With Acute Non-Bacterial Gastroenteritis. The Lancet *302*, 1281–1283.
7. Blackhall, J., Fuentes, A., and Magnusson, G. (1996). Genetic Stability of a Porcine Rotavirus RNA Segment during Repeated Plaque Isolation. Virology *225*, 181–190.
8. Boni, M. F., Posada, D., and Feldman, M. W. (2006). An Exact Nonparametric Method for Inferring Mosaic Structure in Sequence Triplets. Genetics *176*, 1035–1047.
9. Burns, J. W., Siadat-Pajouh, M., Krishnaney, A. A., and Greenberg, H. B. (1996). Protective Effect of Rotavirus VP6-Specific IgA Monoclonal Antibodies That Lack Neutralizing Activity. Science *272*, 104–107.
10. Chang, K. O., Nielsen, P. R., Ward, L. A., and Saif, L. J. (1999). Dual Infection of Gnotobiotic Calves with Bovine Strains of Group A and Porcine-Like Group C Rotaviruses Influences Pathogenesis of the Group C Rotavirus. Journal of Virology *73*, 9284–9293.
11. Chen, Z., Lambden, P. R., Lau, J., Caul, E., and Clarke, I. N. (2002). Human group C rotavirus: completion of the genome sequence and gene coding assignments of a non-cultivatable rotavirus. Virus Research *83*, 179–187.
12. Collins, P. J., Martella, V., and O'shea, H. (2008). Detection and Characterization of Group C Rotaviruses in Asymptomatic Piglets in Ireland. Journal of Clinical Microbiology *46*, 2973–2979.
13. Cortese, M. M., Immergluck, L. C., Held, M., Jain, S., Chan, T., Grizas, A. P., Khizer, S., Barrett, C., Quaye, O., Mijatovic-Rustempasic, S., et al. (2013). Effectiveness of Monovalent and Pentavalent Rotavirus Vaccine. Pediatrics *132*.
14. Crawford, S. E., Mukherjee, S. K., Estes, M. K., Lawton, J. A., Shaw, A. L., Ramig, R. F., and Prasad, B. V. V. (2001). Trypsin Cleavage Stabilizes the Rotavirus VP4 Spike. Journal of Virology *75*, 6052–6061.
15. Delport, W., Poon, A. F. Y., Frost, S. D. W., and Pond, S. L. K. (2010). Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics *26*, 2455–2457.
16. Desselberger, U. (2014). Rotaviruses. Virus Research *190*, 75–96.
17. Drummond, A. J. (2005). Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. Molecular Biology and Evolution *22*, 1185–1192.
18. Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian Phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution *29*, 1969–1973.
19. Drummond, A. J., Ho, S. Y. W., Phillips, M. J., and Rambaut, A. (2006). Relaxed Phylogenetics and Dating with Confidence. PLoS Biology PLoS Biol *4*.
20. Eddy, S. R. (1998). Profile hidden Markov models. Bioinformatics *14*, 755–763.
21. Esposito, D. H., Holman, R. C., Haberling, D. L., Tate, J. E., Podewils, L. J., Glass, R. I., and Parashar, U. (2011). Baseline Estimates of Diarrhea-associated Mortality Among United States Children Before Rotavirus Vaccine Introduction. The Pediatric Infectious Disease Journal *30*, 942–947.
22. Estes, M., and Kapikian, A. (2007). Rotavirus. In Fields' Virology, D. Knipe, P. Howley, D. Griffin, R. Lamb, M. Martin, B. Roizman, and S. Straus, eds. (Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins), pp. 1917–1974.
23. Feng, N., Franco, M. A., and Greenberg, H. B. (1997). Murine Model of Rotavirus Infection. Advances in Experimental Medicine and Biology Mechanisms in the Pathogenesis of Enteric Diseases, 233–240.
24. Fischer, T. K., Viboud, C., Parashar, U., Malek, M., Steiner, C., Glass, R., and Simonsen, L. (2007). Hospitalizations and Deaths from Diarrhea and Rotavirus among Children 195, 1117–1125.
25. Flewett, T., Bryden, A., and Davies, H. (1973). Virus Particles In Gastroenteritis. The Lancet *302*, 1497.
26. Franco, M. A., Angel, J., and Greenberg, H. B. (2006). Immunity and correlates of protection for rotavirus vaccines. Vaccine *24*, 2718–2731.
27. Gentsch, J. R., Laird, A. R., Bielfelt, B., Griffin, D. D., Bányai, K., Ramachandran, M., Jain, V., Cunliffe, N. A., Nakagomi, O., Kirkwood, C. D., et al. (2005). Serotype Diversity and Reassortment between Human and Animal Rotavirus Strains: Implications for Rotavirus Vaccine Programs. The Journal of Infectious Diseases J INFECT DIS *192*.
28. Gibbs, M. J., Armstrong, J. S., and Gibbs, A. J. (2000). Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics *16*, 573–582.
29. Glass, R. I., Parashar, U., Patel, M., Gentsch, J., and Jiang, B. (2014). Rotavirus vaccines: Successes and challenges. Journal of Infection *68*.
30. House, J. A. (1978). Economic impact of rotavirus and other neonatal disease agents of animals. Journal of the American Veterinary Medical Association *173*, 573–576.
31. Iturriza-Gomara, M., Isherwood, B., Desselberger, U., and Gray, J. (2001). Reassortment In Vivo: Driving Force for Diversity of Human Rotavirus Strains Isolated in the United Kingdom between 1995 and 1999. Journal of Virology *75*, 3696–3705.
32. Jalilvand, S., Marashi, S. M., and Shoja, Z. (2015). Rotavirus VP6 preparations as a non-replicating vaccine candidates. Vaccine *33*, 3281–3287.
33. Jiang, B., Gentsch, J. R., and Glass, R. I. (2002). The Role of Serum Antibodies in the Protection against Rotavirus Disease: An Overview. Clinical Infectious Diseases CLIN INFECT DIS *34*, 1351–1361.
34. Jiang, S., Ji, S., Tang, Q., Cui, X., Yang, H., Kan, B., and Gao, S. (2008). Molecular characterization of a novel adult diarrhoea rotavirus strain J19 isolated in China and its significance for the evolution and origin of group B rotaviruses. Journal of General Virology *89*, 2622–2629.
35. Katoh, K., and Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution *30*, 772–780.
36. Kelkar, S. D., and Zade, J. K. (2004). Group B rotaviruses similar to strain CAL-1, have been circulating in Western India since 1993. Epidemiol. Infect. Epidemiology and Infection *132*, 745–749.
37. Kindler, E., Trojnar, E., Heckel, G., Otto, P. H., and Johne, R. (2013). Analysis of rotavirus species diversity and evolution including the newly determined full-length genome sequences of rotavirus F and G. Infection, Genetics and Evolution *14*, 58–67.
38. Kuzuya, M., Fujii, R., Hamano, M., Nishijima, M., and Ogura, H. (2007). Detection and molecular characterization of human group C rotaviruses in Okayama Prefecture, Japan, between 1986 and 2005. J. Med. Virol. Journal of Medical Virology *79*, 1219–1228.
39. Lopman, B. A., Pitzer, V. E., Sarkar, R., Gladstone, B., Patel, M., Glasser, J., Gambhir, M., Atchison, C., Grenfell, B. T., Edmunds, W. J., et al. (2012). Understanding Reduced Rotavirus Vaccine Efficacy in Low Socio-Economic Settings. PLoS ONE *7*.
40. Malherbe, H., and Harwin, R. (1963). The cytopathic effects of vervet monkey viruses. South African Medical Journal *37*, 407–411.
41. Martella, V., Bányai, K., Matthijnssens, J., Buonavoglia, C., and Ciarlet, M. (2010). Zoonotic aspects of rotaviruses. Veterinary Microbiology *140*, 246–255.
42. Marthaler, D., Rossow, K., Gramer, M., Collins, J., Goyal, S., Tsunemitsu, H., Kuga, K., Suzuki, T., Ciarlet, M., and Matthijnssens, J. (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.
43. Marthaler, D., Rossow, K., Culhane, M., Collins, J., Goyal, S., Ciarlet, M., and Matthijnssens, J. (2013). Identification, phylogenetic analysis and classification of porcine group C rotavirus VP7 sequences from the United States and Canada. Virology *446*, 189–198.
44. Marthaler, D., Rossow, K., Culhane, M., Goyal, S., Collins, J., Matthijnssens, J., Nelson, M., and Ciarlet, M. (2014). Widespread Rotavirus H in Commercially Raised Pigs, United States. Emerg. Infect. Dis. Emerging Infectious Diseases *20*, 1203–1206.
45. Martin, D., and Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. Bioinformatics *16*, 562–563.
46. Martin, D. P., Murrell, B., Golden, M., Khoosal, A., and Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evolution *1*.
47. Martin, D., Posada, D., Crandall, K., and Williamson, C. (2005). A Modified Bootscan Algorithm for Automated Identification of Recombinant Sequences and Recombination Breakpoints. AIDS Research and Human Retroviruses *21*, 98–102.
48. Matthijnssens, J., Ciarlet, M., Heiman, E., Arijs, I., Delbeke, T., Mcdonald, S. M., Palombo, E. A., Iturriza-Gomara, M., Maes, P., Patton, J. T., et al. (2008). Full Genome-Based Classification of Rotaviruses Reveals a Common Origin between Human Wa-Like and Porcine Rotavirus Strains and Human DS-1-Like and Bovine Rotavirus Strains. Journal of Virology *82*, 3204–3219.
49. Matthijnssens, J., Ciarlet, M., Mcdonald, S. M., Attoui, H., Bányai, K., Brister, J. R., Buesa, J., Esona, M. D., Estes, M. K., Gentsch, J. R., et al. (2011). Uniformity of rotavirus strain nomenclature proposed by the Rotavirus Classification Working Group (RCWG). Arch Virol Archives of Virology *156*, 1397–1413.
50. Matthijnssens, J., Grazia, S. D., Piessens, J., Heylen, E., Zeller, M., Giammanco, G. M., Bányai, K., Buonavoglia, C., Ciarlet, M., Martella, V., et al. (2011). Multiple reassortment and interspecies transmission events contribute to the diversity of feline, canine and feline/canine-like human group A rotavirus strains. Infection, Genetics and Evolution *11*, 1396–1406.
51. Matthijnssens, J., Otto, P. H., Ciarlet, M., Desselberger, U., Ranst, M. V., and Johne, R. (2012). VP6-sequence-based cutoff values as a criterion for rotavirus species demarcation. Arch Virol Archives of Virology *157*, 1177–1182.
52. Mcdonald, S. M., Matthijnssens, J., Mcallen, J. K., Hine, E., Overton, L., Wang, S., Lemey, P., Zeller, M., Ranst, M. V., Spiro, D. J., et al. (2009). Evolutionary Dynamics of Human Rotaviruses: Balancing Reassortment with Preferred Genome Constellations. PLoS Pathog PLoS Pathogens *5*.
53. Mebus, C. A., Underdahl, N. R., Rhodes, M. B., and Twiehaus, M. J. (1969). Further studies on neonatal calf diarrhea virus. Proceedings, Annual Meeting of the United States Animal Health Association *73*, 97–99.
54. Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. 2010 Gateway Computing Environments Workshop (GCE).
55. Mlera, L., O’Neill, H. G., Jere, K. C., and Dijk, A. A. V. (2012). Whole-genome consensus sequence analysis of a South African rotavirus SA11 sample reveals a mixed infection with two close derivatives of the SA11-H96 strain. Arch Virol Archives of Virology *158*, 1021–1030.
56. Mohan, K. V. K., and Atreya, C. D. (2000). Comparative sequence analysis identified mutations outside the NSP4 cytotoxic domain of tissue culture-adapted ATCC-Wa strain of human rotavirus and a novel inter-species variable domain in its C-terminus. Arch. Virol. Archives of Virology *145*, 1789–1799.
57. Molinari, B. L., Lorenzetti, E., Otonel, R. A., Alfieri, A. F., and Alfieri, A. A. (2014). Species H Rotavirus Detected in Piglets with Diarrhea, Brazil, 2012. Emerg. Infect. Dis. Emerging Infectious Diseases *20*, 1019–1022.
58. Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Pond, S. L. K., and Scheffler, K. (2013). FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. Molecular Biology and Evolution *30*, 1196–1205.
59. Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., and Pond, S. L. K. (2012). Detecting Individual Sites Subject to Episodic Diversifying Selection. PLoS Genetics PLoS Genet *8*.
60. Nagashima, S., Kobayashi, N., Ishino, M., Alam, M. M., Ahmed, M. U., Paul, S. K., Ganesh, B., Chawla-Sarkar, M., Krishnan, T., Naik, T. N., et al. (2008). Whole genomic characterization of a human rotavirus strain B219 belonging to a novel group of the genus rotavirus. J. Med. Virol. Journal of Medical Virology *80*, 2023–2033.
61. Nguyen, L.-T., Schmidt, H. A., Haeseler, A. V., and Minh, B. Q. (2014). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology and Evolution *32*, 268–274.
62. Offit, P. A. (1994). Rotaviruses: Immunological Determinants of Protection Against Infection and Disease. Advances in Virus Research Advances in Virus Research Volume 44, 161–202.
63. Otto, P. H., Ahmed, M. U., Hotzel, H., Machnowska, P., Reetz, J., Roth, B., Trojnar, E., and Johne, R. (2012). Detection of avian rotaviruses of groups A, D, F and G in diseased chickens and turkeys from Europe and Bangladesh. Veterinary Microbiology *156*, 8–15.
64. Padidam, M., Sawyer, S., and Fauquet, C. M. (1999). Possible Emergence of New Geminiviruses by Frequent Recombination. Virology *265*, 218–225.
65. Paradis, E., Claude, J., and Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics *20*, 289–290.
66. Parashar, U. D., Burton, A., Lanata, C., Boschi‐Pinto, C., Shibuya, K., Steele, D., Birmingham, M., and Glass, R. I. (2009). Global Mortality Associated with Rotavirus Disease among Children in 2004. The Journal of Infectious Diseases J INFECT DIS *200*.
67. Patel, M. M., and Parashar, U. D. (2009). Assessing the Effectiveness and Public Health Impact of Rotavirus Vaccines after Introduction in Immunization Programs. The Journal of Infectious Diseases J INFECT DIS *200*.
68. Pollard, S. L., Malpica-Llanos, T., Friberg, I. K., Fischer-Walker, C., Ashraf, S., and Walker, N. (2015). Estimating the herd immunity effect of rotavirus vaccine. Vaccine *33*, 3795–3800.
69. Pond, S. L. K., Frost, S. D. W., and Muse, S. V. (2004). HyPhy: hypothesis testing using phylogenies. Bioinformatics *21*, 676–679.
70. Pond, S. L. K., and Frost, S. D. W. (2005). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics *21*, 2531–2533.
71. Posada, D., and Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. Proceedings of the National Academy of Sciences *98*, 13757–13762.
72. Posada, D. (2001). Unveiling the Molecular Clock in the Presence of Recombination. Molecular Biology and Evolution *18*, 1976–1978.
73. Posada, D., and Crandall, K. A. (2002). The Effect of Recombination on the Accuracy of Phylogeny Estimation. J Mol Evol Journal of Molecular Evolution *54*, 396–402.
74. R Core Team (2015). R: A Language and Environment for Statistical Computing. Available at: https://www.r-project.org/.
75. Rambaut, A., Suchard, M. A., Xie, D., and Drummond, A. J. (2014). Tracer v1.6. Tracer v1.6. Available at: http://beast.bio.ed.ac.uk/tracer [Accessed March 13, 2016].
76. Rambaut, A. (2014). FigTree. FigTree. Available at: http://tree.bio.ed.ac.uk/software/figtree/ [Accessed March 13, 2016].
77. Ruiz-Palacios, G. M., Pérez-Schael, I., Velázquez, F. R., Abate, H., Breuer, T., Clemens, S. C., Cheuvart, B., Espinoza, F., Gillard, P., Innis, B. L., et al. (2006). Safety and Efficacy of an Attenuated Vaccine against Severe Rotavirus Gastroenteritis. New England Journal of Medicine N Engl J Med *354*, 11–22.
78. Saiada, F., Rahman, H. N. A., Moni, S., Karim, M. M., Pourkarim, M. R., Azim, T., and Rahman, M. (2010). Clinical presentation and molecular characterization of group B rotaviruses in diarrhoea patients in Bangladesh. Journal of Medical Microbiology *60*, 529–536.
79. Santos, N., and Hoshino, Y. (2004). Global distribution of rotavirus serotypes/genotypes and its implication for the development and implementation of an effective rotavirus vaccine. Rev. Med. Virol. Reviews in Medical Virology *15*, 29–56.
80. Schierup, M. H., and Hein, J. (2000). Consequences of Recombination on Traditional Phylogenetic Analysis. Genetics *156*, 879–891.
81. Schnagl, R. D., Boniface, K., Cardwell, P., Mccarthy, D., Ondracek, C., Coulson, B., Erlich, J., and Morey, F. (2004). Incidence of Group C Human Rotavirus in Central Australia and Sequence Variation of the VP7 and VP4 Genes. Journal of Clinical Microbiology *42*, 2127–2133.
82. Slatkin, M., and Maddison, W. P. (1989). A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics *123*, 603–613.
83. Smith, J. (1992). Analyzing the mosaic structure of genes. J Mol Evol Journal of Molecular Evolution *34*.
84. Steyer, A., Poljsak-Prijatelj, M., Barlic-Maganja, D., and Marin, J. (2008). Human, porcine and bovine rotaviruses in Slovenia: evidence of interspecies transmission and genome reassortment. Journal of General Virology *89*, 1690–1698.
85. Steyer, A., Sagadin, M., Kolenc, M., and Poljšak-Prijatelj, M. (2013). Whole genome sequence analysis of bovine G6P[11] rotavirus strain found in a child with gastroenteritis. Infection, Genetics and Evolution *13*, 89–95.
86. Suzuki, Y. (2004). Negative selection on neutralization epitopes of poliovirus surface proteins: implications for prediction of candidate epitopes for immunization. Gene *328*, 127–133.
87. Tao, H., Changan, W., Zhaoying, F., Zinyi, C., Xuejian, C., Xiaoquang, L., Guangmu, C., Henli, Y., Tungxin, C., Weiwe, Y., et al. (1984). Waterborne Outbreak Of Rotavirus Diarrhoea In Adults In China Caused By A Novel Rotavirus. The Lancet *323*, 1139–1142.
88. Tate, J. E., Cortese, M. M., Payne, D. C., Curns, A. T., Yen, C., Esposito, D. H., Cortes, J. E., Lopman, B. A., Patel, M. M., Gentsch, J. R., et al. (2011). Uptake, Impact, and Effectiveness of Rotavirus Vaccination in the United States. The Pediatric Infectious Disease Journal *30*.
89. Tate, J. E., Mutuc, J. D., Panozzo, C. A., Payne, D. C., Cortese, M. M., Cortes, J. E., Yen, C., Esposito, D. H., Lopman, B. A., Patel, M. M., et al. (2011). Sustained Decline in Rotavirus Detections in the United States Following the Introduction of Rotavirus Vaccine in 2006. The Pediatric Infectious Disease Journal *30*.
90. Tate, J. E., Burton, A. H., Boschi-Pinto, C., Steele, A. D., Duque, J., and Parashar, U. D. (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. The Lancet Infectious Diseases *12*, 136–141.
91. Tavaré, S. (1986). Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. Lectures Math Life Sci *17*, 57–86.
92. Todd, S., Page, N. A., Steele, A. D., Peenze, I., and Cunliffe, N. A. (2010). Rotavirus Strain Types Circulating in Africa: Review of Studies Published during 1997–2006. The Journal of Infectious Diseases J INFECT DIS *202*.
93. Trojnar, E., Otto, P., and Johne, R. (2009). The first complete genome sequence of a chicken group A rotavirus indicates independent evolution of mammalian and avian strains. Virology *386*, 325–333.
94. Vesikari, T., Matson, D. O., Dennehy, P., Damme, P. V., Santosham, M., Rodriguez, Z., Dallas, M. J., Heyse, J. F., Goveia, M. G., Black, S. B., et al. (2006). Safety and Efficacy of a Pentavalent Human–Bovine (WC3) Reassortant Rotavirus Vaccine. New England Journal of Medicine N Engl J Med *354*, 23–33.
95. Widdowson, M.-A., Meltzer, M. I., Zhang, X., Bresee, J. S., Parashar, U. D., and Glass, R. I. (2007). Cost-effectiveness and Potential Impact of Rotavirus Vaccination in the United States. Pediatrics *119*, 684–697.
96. Yi, J., and Anderson, E. J. (2013). Rotavirus vaccination: short-term indirect herd protection, long-term uncertainty. Expert Review of Vaccines *12*, 585–587.
97. Zeller, M., Donato, C., Trovão, N. S., Cowley, D., Heylen, E., Donker, N. C., Mcallen, J. K., Akopov, A., Kirkness, E. F., Lemey, P., et al. (2015). Genome-Wide Evolutionary Analyses of G1P[8] Strains Isolated Before and After Rotavirus Vaccine Introduction. Genome Biol Evol Genome Biology and Evolution *7*, 2473–2483.

**DATA TABLES**













**FIGURES**

|  |
| --- |
|  |
| **Figure 1: FUBAR Analysis of Selection for Different RV Species and Proteins**  A value of "β/α" closer to “0” suggests stronger purifying selection, while a value closer to “1” suggests stronger diversifying selection. Datasets with notably higher β/α values were the genes of RVB NSP3, RVA NSP4 and RVA VP7. RVA had higher β/α values than RVB and RVC across both NSP genes (RVA: 0.219; RVB: 0.198; RVC: 0.174) and VP genes (RVA: 0.169; RVB: 0.127; RVC: 0.077). Across proteins, the average β/α value for NSP genes (0.197) was much higher than that of VP genes (0.118). Within NSP genes, the gene for NSP2 had the lowest β/α values while the gene for NSP4 had the highest average β/α value (0.317). The gene for VP6 had the lowest β/α values across VP genes while the gene for VP7 had the highest average β/α value (0.191). |

|  |
| --- |
|  |
|  |
| **Figure 2: Comparison of Selection Pressures in RV Genes Between Human and Non-Human Hosts**  A value of "β/α" closer to “0” suggests stronger purifying selection, while a value closer to “1” suggests stronger diversifying selection. a) Within the RVB datasets, host effect on selection was significant for genes of NSP1, NSP2 and NSP5. b) Within the RVC datasets, host effect on selection was significant for genes of NSP1, NSP2, VP1 and VP4. However, within each species and across all datasets, there was no indication of human hosts always resulting in higher or lower diversifying selection pressures. Only the gene for NSP1 demonstrated lowered diversifying selection pressures in humans for both RVC and RVB.  “X” denotes a significant host effect on the selection pressures as determined by AIC values. |

|  |
| --- |
|  |
| **Figure 3: FUBAR Analysis of Selection for RVA Proteins Before and After the Introduction of Modern Vaccines**  A value of "β/α" closer to “0” suggests stronger purifying selection, while a value closer to “1” suggests stronger diversifying selection. Higher diversifying selection pressures were observed for the genes of VP4 and VP7 before the introduction of vaccine, while the gene for VP6 experienced a slight increase in diversifying selection pressure after the introduction of vaccine. |

|  |
| --- |
|  |
| **Figure 4: Selection on RVA VP4 Before and After the Introduction of Modern Vaccines**  To minimize the number of artefact signals, only selection pressures with posterior probabilities of more than 0.80 were drawn. "β-α" is the difference between synonymous (α) and non-synonymous (β) substitution rates over sites: a positive value suggests diversifying selection while a negative value suggests purifying selection. Selection pressures on the gene for RVA VP4 before the introduction of modern vaccines (2006-2015; yellow) was superimposed onto the selection pressures on the gene for RVA VP4 after the introduction of modern vaccines (1996-2005; blue). Novel sites under diversifying selection after the introduction of vaccines were labelled in bold numbers. Codon 2 showed diversification from Ala (gct) to Val (gtt) while codon 161 was STOP (tag) changing to a variety of other amino acids, allowing read-throughs. The estimation at codon 311 was found to be an artefact due to gaps in the majority of sequences at this codon. |

|  |
| --- |
|  |
| **Figure 5: Mutation Rates of Different RV Genes**  Error bars denotes 95% highest posterior density (HPD) intervals. Between species, RVB (1.50E-02 substitutions/site/year) had the highest average mutation rate as compared to RVA (2.61E-03 substitutions/site/year) and RVC (3.15E-03 substitutions/site/year). This can be attributed to high mutation rates in the RVB genes for NSP1, NSP4, NSP5, VP2 and VP3. Between proteins, the average mutation rate of NSP genes (9.98E-03 substitutions/site/year) was higher than but on the same order of magnitude as that of VP genes (4.35E-03 substitutions/site/year). Across NSP genes, genes for NSP1 (1.79E-02 substitutions/site/year) and NSP4 (2.07E-02 substitutions/site/year) had the highest average mutations rates while the gene for NSP2 had the lowest average mutation rate (1.02E-03 substitutions/site/year). For the VP genes, the gene for VP3 had the highest average mutation rate (1.07E-02 substitutions/site/year), while genes for VP1 (1.57E-03 substitutions/site/year) and VP6 (1.94E-03 substitutions/site/year) had the lowest average mutation rates. The gene for RVA VP7 had a much larger 95% highest posterior density (HPD) interval than all other genes. |
|  |
| **Figure 6: Effective Population Size of RVA against Time**  Boundaries shaded in blue denote the 95% highest posterior density (HPD) intervals. There was no significant change in the effective population size of RVA after the introduction of modern vaccines in 2005. |

|  |
| --- |
| C:\rotavirus\20 Results\trees\VP7A_rad.mcc.jpg |
| **Figure 7: Maximum Credibility Tree of RVA Reconstructed Using VP7 Dataset**  Sequences from human isolates are marked in red. Sequences from non-human isolates are marked in blue. Sequences missing host data are greyed out. Tip labels indicate node ages. Phylogeny of RVA suggests zoonotic transmission as there is no clear separation of clades for human hosts and non-human hosts. Furthermore, a few RVA branches in humans branched off from those in non-human hosts as recent as 5-6 years ago and vice versa. |

|  |
| --- |
| C:\rotavirus\20 Results\trees\VP7B_rad.mcc.jpg |
| **Figure 8: Maximum Credibility Tree of RVB Reconstructed Using VP7 Dataset**  Sequences from human isolates are marked in red. Sequences from non-human isolates are marked in blue. Tip labels indicate node ages. Phylogeny of RVB diverged between human and non-human hosts early in its evolutionary histories and showed no signs of zoonotic transmission. |

|  |
| --- |
| C:\rotavirus\20 Results\trees\VP7C_rad.mcc.jpg |
| **Figure 9: Maximum Credibility Tree of RVC Reconstructed Using VP7 Dataset**  Sequences from human isolates are marked in red. Sequences from non-human isolates are marked in blue. Sequences missing host data are greyed out. Tip labels indicate node ages. Phylogeny of RVC diverged between human and non-human hosts early in its evolutionary histories and showed no signs of zoonotic transmission. |

**APPENDIX A – SOFTWARE AND SETTINGS USED FOR SEQUENCE PROCESSING**

The segment numbers of sequences without relevant host metadata were identified using homology to known sequences in HMMER (version 3.1b1; Eddy, 1998). Together with the segment numbers, metadata including the species of RV, host, country, date of sampling and genotype were also appended to the sequence names in R (version 3.2.3; R Core Team, 2015) using the APE package (version 3.4; Paradis et al., 2004). All sequences were then sorted into 33 separate datasets based on their species (RVA, RVB, RVC) and protein products (NSP1-5, VP1-4, VP6 and VP7). Multiple sequence alignments (MSAs) were then generated for each dataset using MAFFT (online version 7; Katoh & Standley, 2013) with default settings. Sequences with sequencing errors were then removed and MSAs were refined using reverse alignments of their respective translated sequences in MAFFT.  After duplicates were removed, the MSAs were trimmed using the first START codons and the last STOP codons. Finally, sequences with lengths less than 90% of the longest sequence in each dataset were removed and checked in HyPhy, with the exception of those for RVA NSP4, RVA VP4 and RVA VP7 which contain numerous sequences covering specific regions of the genes.

**APPENDIX B – SOFTWARE AND SETTINGS USED FOR RECOMBINATION CHECKS**

Recombination checks were carried out in each dataset using the RDP software (version 4.69; Martin et al., 2015) using a combination of RDP (Martin & Rybicki, 2000), GENECONV (Padidam et al., 1999), Chimaera (Posada & Crandall, 2001), MaxChi (Smith, 1992), and 3Seq (Boni et al., 2006) methods in the primary scan, followed by BootScan (Martin et al., 2005) and SiScan (Gibbs et al., 2000) methods in the secondary scans. Potential recombinants were identified when at least 3 methods reported detections. These were removed from the datasets upon manual inspection to verify that the detections were not a result of complex mutations.

**APPENDIX C – SUMMARY DETAILS OF DATASETS**



**APPENDIX D – BEAST SETTINGS**

The GTR model (Tavaré, 1986) of nucleotide substitution was used with an uncorrelated lognormal relaxed clock, as well as a gamma and invariant sites rate heterogeneity with 4 gamma categories (Drummond et al., 2006). For the clock priors, a uniform prior between 0.0 and 1.0 was assumed for the mean, and an exponential with mean 1/3 for the standard deviation. A uniform (Dirichlet) prior was used for the nucleotide frequencies. The Markov chain Monte Carlo (MCMC) chain length was setup for 1 billion iterations, with sample logging every 10,000 iterations.