**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 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 genes (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.

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). 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.

*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.

*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.

*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 identify 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).