PLOS | ONE

RESEARCH ARTICLE

Sequencing and Analysis of Globally Obtained Human Respiratory Syncytial Virus A and B Genomes

Michael E. Bose[1], Jie He[1], Susmita Shrivastava[2], Martha I. Nelson[3], Jayati Bera[2¤a], Rebecca A. Halpin[2], Christopher D. Town[2], Hernan A. Lorenzi[2], Daniel E. Noyola[4], Valeria Falcone[5], Giuseppe Gerna[6], Hans De Beenhouwer[7], Cristina Videla[8], Tuckweng Kok[9], Marietjie Venter[10¤b], John V. Williams[11], Kelly J. Henrickson[1]*

1 Midwest Respiratory Virus Program, Medical College of Wisconsin, Milwaukee, WI, USA, **2** J. Craig Venter Institute, Rockville, MD, United States of America, **3** Fogarty International Center, National Institutes of Health, Bethesda, MD, United States of America, **4** Departamento de Microbiología, Facultad de Medicina, Universidad Autónoma de San Luis Potosí, San Luis Potosí, Mexico, **5** Department of Virology, Institute for Medical Microbiology and Hygiene, University Medical Center Freiburg, Freiburg, Germany, **6** Laboratori Sperimentali di Ricerca, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy, **7** Department of Microbiology, OLV Hospital Aalst, Aalst, Belgium, **8** Clinical Virology Laboratory, Centro de Educación Médica e Investigaciones Clínicas University Hospital, Buenos Aires, Argentina, **9** School of Molecular and Biomedical Science, University of Adelaide, Adelaide, Australia, **10** Zoonoses Research Unit, Department Medical Virology, University of Pretoria, Pretoria, South Africa, **11** Departments of Pediatrics and Pathology, Microbiology, and Immunology, Vanderbilt University School of Medicine, Nashville, TN, United States of America

¤a Current address: U.S. Army Medical Research Institute of Infectious Diseases, Frederick, MD, United States of America.
¤b Current address: Global Disease Detection, US Centers for Disease Control and Prevention, Pretoria, South Africa.
* khenrick@mcw.edu

CrossMark
click for updates

OPEN ACCESS

Citation: Bose ME, He J, Shrivastava S, Nelson MI, Bera J, Halpin RA, et al. (2015) Sequencing and Analysis of Globally Obtained Human Respiratory Syncytial Virus A and B Genomes. PLoS ONE 10(3): e0120098. doi:10.1371/journal.pone.0120098

Academic Editor: Steven M. Varga, University of Iowa, UNITED STATES

Received: October 27, 2014

Accepted: January 19, 2015

Published: March 20, 2015

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Data Availability Statement: All genome sequence files are available from the GenBank database (bioproject id PRJNA73049 and accession numbers KF530258 - KF530269 and KF826816 - KF826860). All other relevant data are within the paper and its Supporting Information files.

Funding: This project has been funded in whole or part with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services (www.niaid.nih.gov) under contract number HHSN272200900007C and grant numbers 5U01AI070428 and 5U01AI077988. This work was

Abstract

Background

Human respiratory syncytial virus (RSV) is the leading cause of respiratory tract infections in children globally, with nearly all children experiencing at least one infection by the age of two. Partial sequencing of the attachment glycoprotein gene is conducted routinely for genotyping, but relatively few whole genome sequences are available for RSV. The goal of our study was to sequence the genomes of RSV strains collected from multiple countries to further understand the global diversity of RSV at a whole-genome level.

Methods

We collected RSV samples and isolates from Mexico, Argentina, Belgium, Italy, Germany, Australia, South Africa, and the USA from the years 1998-2010. Both Sanger and next-generation sequencing with the Illumina and 454 platforms were used to sequence the whole genomes of RSV A and B. Phylogenetic analyses were performed using the Bayesian and maximum likelihood methods of phylogenetic inference.

supported in part by the Multinational Influenza Seasonal Mortality Study (MISMS), led by the Fogarty International Center, National Institutes of Health, with funding from the Department of Health and Human Services (www.fic.nih.gov). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: DEN has participated as an advisory board member for Novartis and GlaxoSmithKline Biologicals and as a speaker for Abbvie and Roche. This does not alter the authors' adherence to PLOS ONE policies on sharing data and materials.

Results

We sequenced the genomes of 34 RSVA and 23 RSVB viruses. Phylogenetic analysis showed that the RSVA genome evolves at an estimated rate of 6.72×10^{-4} substitutions/site/year (95% HPD 5.61×10^{-4} to 7.6×10^{-4}) and for RSVB the evolutionary rate was 7.69×10^{-4} substitutions/site/year (95% HPD 6.81×10^{-4} to 8.62×10^{-4}). We found multiple clades co-circulating globally for both RSV A and B. The predominant clades were GA2 and GA5 for RSVA and BA for RSVB.

Conclusions

Our analyses showed that RSV circulates on a global scale with the same predominant clades of viruses being found in countries around the world. However, the distribution of clades can change rapidly as new strains emerge. We did not observe a strong spatial structure in our trees, with the same three main clades of RSV co-circulating globally, suggesting that the evolution of RSV is not strongly regionalized.

Introduction

Human respiratory syncytial virus (RSV) is the leading cause of respiratory tract illness in children globally, with nearly all children experiencing at least one infection by the age of two [1]. RSV can also regularly infect adults and causes severe illness in the elderly [2,3]. As a member of the *Paramyxoviridae* family, RSV is a non-segmented negative strand RNA virus whose genome contains ten genes that encode 11 proteins. RSV is divided into two distinct subtypes (A and B) and each have been divided into multiple genotypes that have been found to co-circulate [4,5]. Most genetic studies of RSV are focused on the attachment glycoprotein (G) gene which is the most variable and has been commonly used for genotyping [6–8]. The same genotypes have generally been shown to circulate globally [9–12].

When we started this study in 2009, there were relatively few sequences for RSV in GenBank. The majority were partial coding sequences, and the limited number of whole genome sequences were mainly from viruses collected more than 20 years prior or experimentally derived mutants. The lack of whole-genome data made it particularly difficult to design diagnostic tests for RSV capable of detecting the broad diversity of modern strains. To fill this void, we collected RSV viruses from Mexico, Argentina, Belgium, Italy, Germany, Australia, South Africa, and the USA from 1998–2010. From these isolates we sequenced the complete genomes of 57 RSV viruses (34 RSVA and 23 RSVB viruses). Since we started this study there have been parallel efforts to sequence RSV genomes, so we have included these additional data in our analyses. A phylogenetic analysis based on the whole genome, full G gene coding sequence (CDS), and the second variable region of the G gene identified multiple clades circulating globally.

Materials and Methods

Ethics Statement

Samples from the Virology Laboratory of CEMIC, University of Adelaide, University Medical Center Freiburg, and Onze Lieve Vrouw Ziekenhuis were collected for routine viral diagnostic testing and de-identified prior to shipment to the Medical College of Wisconsin. Samples from the Autonomous University of San Luis Potosí were collected with written informed consent as

approved by the Research and Ethics Committee. Samples from the Fondazione IRCCS Policlinico San Matteo were collected with written informed consent as approved by the Bioethics Committee of the Fondazione IRCCS Policlinico San Matteo. Samples from the University of Pretoria were collected with written informed consent as approved by the Health Sciences Research Ethics Committee. Samples from the Vanderbilt Vaccine Clinic were collected with written informed consent as approved by the Committee for the Protection of Human Subjects of the Vanderbilt University Medical Center. At Children's Hospital of Wisconsin and Froedtert Hospital samples were collected with written informed consent as approved by the Children's Hospital of Wisconsin Human Research Review Board. For samples collected from minors written informed consent was obtained from their parent or guardian on their behalf. Some samples had been de-identified and where tested under a protocol approved by the Children's Hospital of Wisconsin Human Research Review Board.

Sample Collection and RSV Identification from Various Locations

The Autonomous University of San Luis Potosí (San Luis Potosí, Mexico). Respiratory samples were collected as part of research projects carried out to analyze the epidemiology of viral respiratory infections and as part of a hospital-based infection control program; the research projects were approved by the corresponding Research and Ethics Committees. Samples were obtained by nasal wash or pharyngeal swab. Viral testing was carried out directly on respiratory samples. RSV was identified with a direct immunofluorescence assay.

Virology Laboratory of CEMIC (Buenos Aires, Argentina). Nasopharyngeal aspirates or swabs were submitted for routine viral diagnostics in a transport media (Hanks plus 2% FCS, penicillin, streptomycin, and amphotericin B). Samples were processed for antigen detection on pelleted cells by indirect immunofluorescence with monoclonal antibodies against RSV, adenovirus, influenza A and B and parainfluenza (EMD Millipore, Billerica, MA, USA). A fluorescein labeled anti-mouse IgG was used (Sigma-Aldrich Corp., St. Louis, MO, USA). Readings were performed with a C. Zeiss microscope provided with epifluorescent equipment and a mercury lamp.

University of Adelaide, IMVS-SA Pathology (Adelaide, Australia). Respiratory secretion specimens were submitted for routine viral diagnostics for seven viruses (adenovirus, influenza A & B, parainfluenza 1, 2, & 3 and respiratory syncytial virus) and *Mycoplasma pneumoniae* [13]. The viruses and *M. pneumoniae* were identified directly from the specimens by specific antibodies using in-house developed enzyme immunoassay. Specimens were also inoculated into 96-well microwell cell cultures for virus isolation, spun at 1000×g for 1hr at 35°C and incubated for 5–6 days at 37°C [14].

University Medical Center Freiburg (Freiburg, Germany). Nasopharyngeal swab or bronchoalveolar lavage samples were collected as part of routine clinical testing. RSV was identified using the ID-Tag RVP test (Luminex, Austin, TX, USA). The ethics policy of the hospital allows for left-over specimens to be used for investigational purposes as long as they are de-identified.

Fondazione IRCCS Policlinico San Matteo (Pavia, Italy). Approval for the study was obtained from the local Ethics Committee, and informed consent was obtained from patients or their parents. Nasopharyngeal aspirates were tested by cell culture, direct fluorescent antibody (DFA) staining, and quantified by real time RT-PCR [15].

Department of Medical Virology, University of Pretoria (Pretoria, South Africa). Nasopharyngeal aspirates were submitted for routine diagnosis from patients with lower respiratory tract infection, hospitalized in the Kalafong and Steve Biko Academic hospitals to the Tshwane National Health Laboratory Service laboratory, Department Medical Virology

University of Pretoria by direct immunofluorescent assay followed by confirmation by RT-PCR, as described previously [11]. Viruses were subsequently amplified in tissue culture for the purpose of this study. Characterization of strains by sequencing was approved and monitored by the human ethics committee, University of Pretoria (25/2006).

Laboratory Microbiology, Onze Lieve Vrouw Ziekenhuis (Aalst, Belgium). Samples were collected as nasopharyngeal aspirates as part of routine clinical testing. RSV was detected using real-time RT-PCR [16]. The ethics policy of the hospital allows for left-over specimens to be used for investigational purposes as long as they are de-identified.

Vanderbilt Vaccine Clinic (Nashville, TN, USA). Virus sequences were derived from specimens collected prospectively over a 20-year period from 1982–2001 in the Vanderbilt Vaccine Clinic, as previously described [17]. Nasal wash specimens were collected from children <5 years of age with acute upper or lower respiratory illness. Nasal washes were tested by cell culture, DFA staining, and quantified by real-time RT-PCR.

The Midwest Respiratory Virus Program, Children's Hospital of Wisconsin, and Dynacare Laboratories (Milwaukee, WI, USA). Nasopharyngeal swab samples were collected with informed consent under an IRB protocol or through routine clinical diagnosis and de-identified. RSV was identified by an in-house real-time RT-PCR.

Samples from the above locations where all frozen at -20°C to -85°C and held frozen until shipped to the Midwest Respiratory Virus Program for further analysis. Some of the viruses were amplified in tissue culture by inoculation in Hep-2 cells for 3–5 days prior to use in this study.

Sanger Sequencing of RSV Genomes

Two sets of 96 pairs of degenerate RSV-specific primers tiled across the entire viral genome were designed from the consensus sequences of either three RSV-A or three RSV-B reference genomes (RSV-A accessions: RSVA/WI/629–4071/98 (JF920065), RSVA/WI/629–9/06–07 (JF920070) and RSVA/WI/629-Q0282/10 (JF920054); RSV-B accessions: RSVB/WV/14647/85 ATCC-VR1400 (unpublished), RSVB/WI/629–5B/06–07 (JN032115) and RSVB/WI/629-Q0306/10 (JN032121)) using a PCR primer design pipeline developed at the J. Craig Venter Institute (JCVI) (S1 Text and S2 Text)[18]. The primer configuration was such that produced tiled short-amplicons with an average length of 650 bp, 150 bp overlap, and with at least two-fold amplicon coverage at every base. Each primer contains a M13 sequence tag at the 5' end used for Sanger sequencing.

Viral RNA was extracted from 100–200 μl of the samples using the ZR-96 Quick RNA(TM) kit (Zymo Research Corp., Irvine, CA, USA), following manufacturers recommendations or from 100–400 μl of the samples using the NucliSENS easyMag (bioMerieux, Inc., Durham, NC, USA) with elution in 25 μl. RNA from each sample was subjected to RT-PCR using a QIAGEN One-step kit (QIAGEN, Hilden, Germany) with four pairs of primers specific to either RSVA or B to determine the RSV type. Then, RT-PCR was performed to produce 650 bp amplicons from each of the 96 primer pairs from one of the two RSV-specific sets, A or B. Excess primers and dNTPs were removed by treatment with Exonuclease I (New England Biolabs, Ipswich, MA, USA) and shrimp alkaline phosphatase (Affymetrix, Santa Clara, CA, USA): 37°C for 60 min, followed by incubation at 72°C for 15 min, and then subjected to Sanger sequencing with M13 primers.

Sequencing reads were trimmed to eliminate amplicon primer and low-quality sequences, and assembled with Minimus, a program from the AMOS project [19]. Draft assemblies were evaluated with an in-house software, CLOE (Closure Editor, http://cloe.sourceforge.net), and targeted PCR-based sequencing reactions were conducted to close gaps and improve sequence coverage. Curated assemblies were validated and annotated with the viral annotation software

VIGOR [20] and predicted genes were subjected to manual inspection and quality control before submission to GenBank.

Sequencing of RSV Genomes by NextGen Technology

RSV isolates that could not be completely sequenced by Sanger were processed using JCVI's Next Generation Sequencing Pipeline. Extracted genomic RNA was reverse transcribed with RSV-specific degenerate primers scattered along the RSV genome at intervals of approximately 4 kb using SuperScript III (Thermo Fisher Scientific, Waltham, MA, USA) (S1 Table). The resulting cDNA was used to generate a set of ~4 kb PCR-amplicons encompassing the entire RSV genome. PCR reactions were carried out with Accuprime (Thermo Fisher Scientific) using pairs of degenerate primers selected from the two pools of 96 primer pairs used for the Sanger sequencing pipeline. Each amplicon was gel purified and then simultaneously amplified and bar-coded using a modified sequence-independent single-primer amplification (SISPA) approach [21]. A pool of these and other viral samples were used to construct Illumina and 454 paired-end libraries and sequenced on their respective platforms. After sequencing, reads from each sample were sorted by barcode and trimmed to eliminate low quality regions as well as SISPA hexamer primer and barcode sequences. The Illumina and 454 reads were then assembled de novo using the clc_novo_assemble program (QIAGEN). The resulting contigs were then used to pick the best reference RSV genome, which was then used for a reference-based assembly using the clc_ref_assemble_long program. The assembled sequences were annotated using VIGOR as described above.

Recombinant Analysis

All RSV genome sequences available in GenBank as of 9/24/2013 were downloaded and combined with the genomes produced in this study. Genomes from mutant RSV strains were excluded (AF013255, AF035006, U39661, U50362, U50363, and U63644). The genomes were separated by RSV subtype and aligned using MAFFT [22]. Each alignment was checked for recombination using the RDP, GENECONV, Chimaera, MaxChi, BootScan, SiScan, and 3Seq algorithms in RDP4 [23]. In order for a recombination event to be considered real it had to be detected by at least three of the algorithms. For the RSVA alignment there was a group of 14 genome sequences (GU591758—GU591771) from a single study that produced an excessive number of recombination events, as observed in a previous study [24]. These genomes were removed and the analyses were repeated. Any recombination events found were considered to most likely be sequencing errors and not true recombination. Therefore, the minor sequence in each recombinant was removed from these sequences and excluded from subsequent analyses.

Positive Selection

All sequences available for each RSV gene were downloaded from GenBank and separated by RSV type. Sequences less than 200 base pairs or from a patent were excluded. Sequences were aligned using MAFFT and were trimmed to only the coding sequence. Coding sequences less than 50 base pairs were removed and for some genes longer sequences were removed because of non-overlapping partial sequences not aligning properly (<600bp for RSVA N gene, <500bp for RSVA F gene, and <650bp for RSVA and B G gene). Positive selection was determined for each alignment using the SLAC, FEL, and FUBAR algorithms available on the Datamonkey webserver [25–28]. Sites were only considered positive if they met the cutoff criteria for at least two of the algorithms, that is a p-value of less than 0.05 for SLAC and FEL and posterior probability of greater than 0.95 for FUBAR.

Phylogenetic Analyses

The set of whole genome sequence alignments resulting from the recombination analysis (RSVA: 105 sequences, RSVB: 70 sequences), alignments of the full G gene CDS that have collection dates available in GenBank (RSVA: 181 sequences, RSVB: 128 sequences), and alignments of the G gene second hypervariable region for both RSVA and RSVB (RSVA: 1117 sequences, RSVB: 750 sequences) were used for the phylogenetic analyses with the Bayesian method of phylogenetic inference in BEAST v1.8.0 and the maximum likelihood method in MEGA 6.05 [29,30].

For the partial CDS alignments of the second hypervariable region in the G gene, there were many duplicate sequences. Duplicate sequences were removed with only one instance of each sequence left for each country for each year. Alignments were imported into BEAUti v1.8.0, which was used to generate xml files for use in BEAST v1.8.0. Collections dates were used to assign tip dates. A general-time reversible (GTR) model of nucleotide substitution was used with a gamma-distributed (Γ) rate variation among sites, with a proportion of invariant sites. An uncorrelated lognormal (UCLN) relaxed molecular clock was used with a flexible Bayesian skyline tree prior.

For the full CDS analysis the same model was used without a proportion of invariant sites. BEAST was used to perform a Bayesian MCMC analysis for the genome and full CDS alignments. The MCMC chain length was 250 million for the RSVA genomes, 50 million for the RSVB genomes, 150 million for the RSVA full G CDS, and three runs of 150 million for RSVB full G CDS that were combined using LogCombiner v1.8.0. The Markov chain was sampled 10,000 times for each run. BEAST results were analyzed using Tracer v1.6 and summary trees were produced using TreeAnnotator v1.8.0. Maximum likelihood trees were inferred for the genome, full CDS, and partial CDS alignments with MEGA 6.05 using the GTR model with gamma-distributed rate variant among sites. Trees were visualized and annotated in FigTree v1.4.0.

To compare the evolutionary rates between genes we extracted the complete CDS sequences for each of the 12 genes from the whole genome data sets. Alignments for each CDS were made using MAFFT. Bayesian MCMC analysis was performed with BEAST as described above for the full G CDS. A chain length of 50 million was used for each CDS sampling a total of 10,000 times. Results were analyzed using Tracer v1.6.

Results and Discussion

Of the 100 RSV samples or isolates we attempted to sequence we were able to sequence genomes from 57, resulting in 34 RSVA and 23 RSVB genomes. Of these, 11 genomes were not completely closed and contained at least one gap. These sequences could not be closed due to insufficient sample for additional sequencing. We suspect that the samples that could not be completely sequenced had insufficient viral nucleic acids due to low viral load or possible sample degradation. It is unlikely that we were not able to sequence these viruses due to genetic variability because the number of primers used should have at least obtained partial genome sequences if there was sufficient nucleic acid. We did not sequence the 5' and 3' termini of the genomes, so the genomes contain incomplete 5' and 3' non-coding regions. Since many of these sequences (41/57) were derived directly from patient specimens they will not contain mutations that can be selected for during growth in tissue culture. For those that had been grown in tissue culture we sequenced the lowest passage available (three passages or less) for sequencing to minimize the number of culture-derived mutations. Information on the sequenced viruses can be found in Table 1 and all sequences can be retrieved from GenBank using the bioproject id PRJNA73049.

Table 1. Information on the RSV Strains Sequenced.

Accession Number	Strain Name	Base Pairs	Gaps (bp)
KF826827	RSVA/Homo sapiens/ARG/159/2004	15209	-
KF826828	RSVA/Homo sapiens/ARG/162/2004	15191	-
KF530260	RSVA/Homo sapiens/ARG/170/2005[a]	14980	-
KF826838	RSVA/Homo sapiens/ARG/177/2006	15189	-
KF826841	RSVA/Homo sapiens/ARG/190/2007	15194	-
KF826846	RSVA/Homo sapiens/ARG/202/2008	15142	-
KF826847	RSVA/Homo sapiens/AUS/248/2007[c]	15179	-
KF826848	RSVA/Homo sapiens/AUS/249/2007[c]	15190	-
KF530261	RSVA/Homo sapiens/DEU/106/2008[a]	15100	-
KF826830	RSVA/Homo sapiens/DEU/107/2009	15186	-
KF826831	RSVA/Homo sapiens/DEU/108/2009	15195	-
KF826854	RSVA/Homo sapiens/ITA/119/2009	15192	-
KF826832	RSVA/Homo sapiens/ITA/120/2009	15176	-
KF826833	RSVA/Homo sapiens/ITA/121/2009	15129	-
KF826855	RSVA/Homo sapiens/ITA/123/2009	15163	-
KF826856	RSVA/Homo sapiens/ITA/125/2009	15188	-
KF826826	RSVA/Homo sapiens/MEX/23/2004	15197	-
KF826816	RSVA/Homo sapiens/MEX/25/2005	14948	1215
KF826836	RSVA/Homo sapiens/MEX/26/2006	15165	-
KF826837	RSVA/Homo sapiens/MEX/27/2006	15194	-
KF826840	RSVA/Homo sapiens/MEX/29/2007	15106	-
KF826817	RSVA/Homo sapiens/MEX/43/2009	14744	117
KF530268	RSVA/Homo sapiens/MEX/59/2007[a]	15128	-
KF826852	RSVA/Homo sapiens/USA/629–1/2007[c]	15167	-
KF826850	RSVA/Homo sapiens/USA/629–11–1/2008[c]	15193	-
KF530263	RSVA/Homo sapiens/USA/629–4/2007[a, c]	15118	1279
KF826823	RSVA/Homo sapiens/USA/629–4360/1998	15197	-
KF826824	RSVA/Homo sapiens/USA/629–4392/1998	15200	-
KF826821	RSVA/Homo sapiens/USA/629–8–2/2007[c]	15177	-
KF530269	RSVA/Hep2_lab/USA/629-Q0030_RSV60/2009[a, c]	15049	-
KF826849	RSVA/Hep2_lab/USA/629-Q0115_RSV89/2010[c]	15182	-
KF530267	RSVA/Homo sapiens/ZAF/323/2007[a, b, c]	15078	1518, 530, 1797, 130, 1078
KF530258	RSVA/Homo sapiens/ZAF/324/2007[a, b, c]	15123	592
KF530264	RSVA/Homo sapiens/ZAF/332/2008[a, b, c]	15039	1002, 670, 290, 174
KF826839	RSVB/Homo sapiens/ARG/187/2006	15278	-
KF826842	RSVB/Homo sapiens/ARG/195/2007	15269	-
KF826845	RSVB/Homo sapiens/ARG/201/2008	15147	-
KF530265	RSVB/Homo sapiens/DEU/111/2009[a]	14673	258, 670
KF826853	RSVB/Homo sapiens/DEU/114/2008	15183	-
KF530266	RSVB/Homo sapiens/DEU/115/2008[a]	15064	-
KF826834	RSVB/Hep2_lab/ITA/126_RSV78/2009[c]	15218	-
KF826835	RSVB/Hep2_lab/ITA/127_RSV79/2009[c]	15238	-
KF530262	RSVB/Homo sapiens/ITA/128/2009[a]	15045	-
KF826857	RSVB/Homo sapiens/ITA/129/2009	15278	-
KF826858	RSVB/Homo sapiens/ITA/130/2009	15233	-
KF826859	RSVB/Homo sapiens/ITA/131/2009	15263	-

(*Continued*)

Table 1. (*Continued*)

Accession Number	Strain Name	Base Pairs	Gaps (bp)
KF826818	RSVB/Homo sapiens/ITA/132/2009	14730	207, 1197, 1383, 1456, 687
KF826819	RSVB/Homo sapiens/ITA/133/2009	15200	926, 737, 388
KF826820	RSVB/Homo sapiens/ITA/134/2009	15181	54
KF826860	RSVB/Homo sapiens/ITA/135/2009	15249	-
KF826825	RSVB/Homo sapiens/MEX/20/2004	15259	-
KF826829	RSVB/Homo sapiens/MEX/24/2005	15215	-
KF826843	RSVB/Homo sapiens/MEX/51/2008	15280	-
KF826844	RSVB/Homo sapiens/MEX/62/2008	15218	-
KF826851	RSVB/Homo sapiens/USA/629–24/2007[c]	15279	-
KF826822	RSVB/Homo sapiens/USA/629–5/2007[c]	15269	-
KF530259	RSVB/Homo sapiens/ZAF/319/2006[a, b, c]	15157	308

[a] These sequences were sequenced by Next-Gen sequencing.

[b] Partial G CDS sequences have previously been published for these samples with different strain names. Accession numbers for these sequences are HQ711732, HQ711709, HQ711688, and HQ711801 [11].

[c] These sequences were produced from viruses isolated in tissue culture.

doi:10.1371/journal.pone.0120098.t001

Recombination

We identified eight recombination events in six RSVA genomes and two events in two RSVB genomes. Five of the RSVA genomes (JF920059, JF920061, JF920064, JF920067, and JF920068) and one RSVB genome (JN032121) were from a previous study from our group [31], and were identified as potential recombinants in a previous study [24]. Upon reexamining the sequencing reads we found that the reads contributing to the minor components for these genomes appeared to be from a different RSV strain due to mismatches in overlapping reads near the predicted recombination breakpoints. In one genome (JF920068) we found additional reads in locations not detected by the recombination analysis that appeared to be from a different strain. These reads were removed and the genomes were reassembled. Since we suspected that all predicted recombination events were sequencing errors the predicted recombinant regions from the remaining two genomes (JX015495 and JX576753) were also removed. These results highlight the importance of running a recombination analysis when performing genome sequencing even when one does not expect true recombination events.

Entropy Plots of the RSVA and B Protein Sequences

From the whole genome alignments the CDS sequences for each RSV gene were concatenated into a single sequence and then translated into predicted protein sequences. The entropy values for each amino acid position were calculated using the Entropy (H(x)) plot function in BioEdit 7.0 (Fig. 1). From these plots it is clear that the G protein is the most variable in both RSVA and B. The high variability of the G gene/protein makes it a good target for evolutionary analyses, which has been the primary goal of most RSV sequencing studies. Therefore, there is a large amount of sequence data for this gene. For the remaining genes/proteins there is significantly less sequence data available. These more conserved regions of the genome are critical for the development of robust diagnostics that continue to detect currently circulating strains.

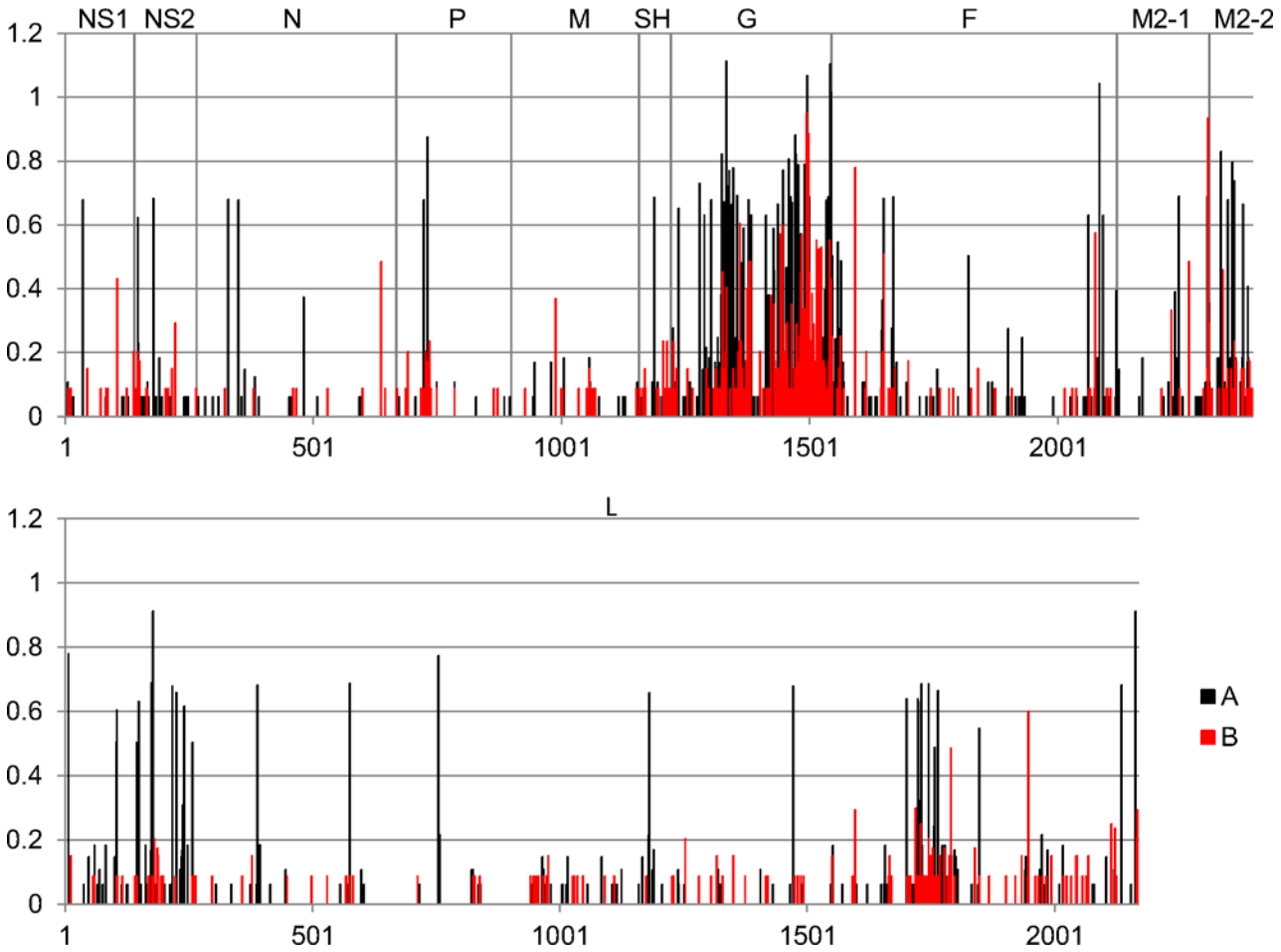


Fig 1. Protein Entropy Plots. This is an entropy plot of the concatenated predicted protein sequences of all of the viruses sequenced in this study. Entropy values were calculated using BioEdit 7.0 and the plot was generated using Microsoft Excel. Black bars are for RSVA sequences and red bars are for RSVB sequences. The higher the bar is the greater the variation at that position in the protein sequence. Across the top of the plot are listed the abbreviated protein sequence names in the order in which the CDS sequences for the proteins appear in the genome.

doi:10.1371/journal.pone.0120098.g001

Positively Selected Sites

We used the SLAC, FEL, and FUBAR algorithms on the Datamonkey webserver to identify potential positively selected sites for each of the coding regions in RSV [27,28]. These programs take alignments of coding sequences and use algorithms to predict if codon sites are under different types of selection. In general, positions under positive selection show a shift over time from one amino acid to another presumably due to a fitness benefit. Positions under negative selection tend to rapidly remove amino acid variants because they may be deleterious. There were multiple positively selected sites predicted in the G and F coding regions (Table 2). These results are not surprising considering both of these genes produce surface glycoproteins that likely face selective pressure from the host immune system. The predicted positively selected sites in the fusion protein were located in the N-terminal signal peptide and the C-terminal heptad repeat. Two additional sites (553 and 573) had initially been predicted in the RSVB F gene cytoplasmic tail region, but upon close inspection they were found to be artifacts from primer sequences used in previous studies [32,33]. In the attachment glycoprotein, predicted

Table 2. Sites under positive selection in RSVA and RSVB.

Type	Gene	Site	SLAC		FEL		FUBAR		Reference Position[a]
			dN-dS	p-value	dN-dS	p-value	dN-dS	Post. Pr.	
RSVA	G	4	3.237	0.042	1.091	0.002	0.779	0.994	4
		124	0.752	0.019	0.433	0.043	0.176	0.877	**124**
		161	0.338	0.233	0.346	0.009	0.202	0.957	**161**
		162	0.474	0.129	0.35	0.006	0.212	0.969	162
		247	1.333	0.002	1.162	0.005	0.847	0.998	**244**
		258	0.405	0.175	0.453	0.003	0.239	0.977	**255**
		301	1.686	0.001	0.807	0.062	0.697	0.98	**274**
		313	1.879	<0.001	0.776	0.002	0.637	0.994	**286**
		317	2.885	0.003	0.834	0.017	0.716	0.989	**290**
		324	2.009	0.15	1.751	0.004	2.034	0.998	**297**
	F	19	5.37	0.083	1.931	0.029	1.025	0.964	19
		518	6.54	0.026	0.787	0.008	0.324	0.956	518
RSVB	G	219	4.19	<0.001	1.318	0.001	0.81	1	**219**
		286	1.288	0.005	0.498	0.005	0.248	0.961	261
		292	1.571	0.011	1.043	0.003	0.806	0.996	**267**
		305	0.894	0.027	0.617	0.035	0.207	0.874	280

[a] Position in reference sequence. Accession M74568 is the reference for RSVA and AF013254 is the reference for RSVB. Sites in bold were also predicted to be under positive selection in previous publications.

doi:10.1371/journal.pone.0120098.t002

positively selected sites were located in the N-terminal cytoplasmic domain and the mucin-like regions. It is unclear what would be driving positive selection in the few sites that are located on the cytoplasmic regions of these proteins. None of the other RSV coding regions had sites that met the criteria of having two of the three algorithms predicting positive selection with the cutoffs used in this study. Since the criteria we used were fairly conservative our results do not eliminate the possibility that there are sites under positive selection in these coding regions.

Since these algorithms were designed for sequences from divergent populations [34], Kryazhimskiy and Plotkin 2008 recommend that caution be taken when interpreting results from using these algorithms with closely related sequences belonging to the same species. Generally these algorithms use different methods to estimate the ratio of nonsynonymous changes per nonsynonymous site (dN) to synonymous changes per synonymous site (dS). Sites with a dN/dS ratio > 1 are considered to be under positive selection and those < 1 under negative selection. They showed that these assumptions may not always hold true with sequences from the same population/species and sites under positive selection may be underestimated. Therefore, it is likely that there are additional sites under positive selection that were not identified in this study and other similar studies. Despite these limitations, other studies have also tried to identify positively selected sites in RSV.

We reviewed seven previously published studies to identify which sites have previously been predicted to be under positive selection [24,33,35–39]. Most of these studies focused on the G CDS, and the majority of the positively selected sites identified in this study were also identified in one or more of the previous studies (8/10 for RSVA and 2/4 for RSVB). Across all studies most of the sites predicted to be under positive selection were identified in only one study (60% for RSVA and 74% for RSVB). Differences between studies can be attributed to differences in

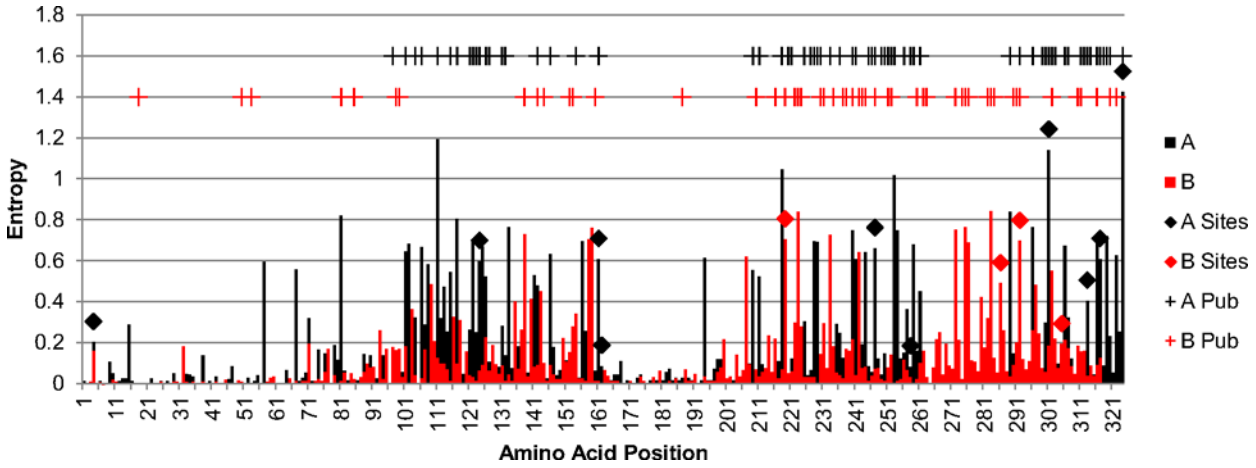


Fig 2. G Protein Entropy Plot and Positive Selected Sites. This is an entropy plot of the G protein sequences with positively selected sites also shown. Entropy values were calculated using BioEdit 7.0 from the alignments used for the positive selection analysis and the plot was generated using Microsoft Excel. Black bars are for RSVA sequences and red bars are for RSVB sequences. Sites predicted to be under positive selection in this study are shown with black diamonds for RSVA and red diamonds for RSVB. Near the top are shown sites predicted to be under positive selection or diversifying selection from previously published studies with black pluses for RSVA and red pluses for RSVB.

doi:10.1371/journal.pone.0120098.g002

data sets, prediction algorithms, and cutoff criteria used. However, like the previous studies most of the positively selected sites were identified in the highly variable mucin-like regions of the G gene (Fig. 2). It has previously been demonstrated that both humans and rabbits produce antibodies to these highly variable regions [40,41]. One of these studies used linear peptides representing natural amino acid variations in the carboxy-terminal mucin-like region of the G protein to demonstrate that human serum can contain antibodies to these regions of the protein and that even a single mutation can eliminate reactivity [41]. Additionally, reactivity of the serum with the peptides was dependent on the genotype of the RSV virus that infected the individual, and a mutation in one of the predicted positively selected sites (position 244 in M74568) was found to produce an antibody escape mutant [41]. Taken together these results support the concept that immunologic pressure is driving selection of mutations in these regions and that mutations in these regions play a role in the ability of RSV to re-infect individuals throughout their lives.

Gene Start (GS), Gene End (GE), and Stop Codon Sequences

Each gene in RSV has a specific nine nucleotide "gene start" sequence (3'-CCCCGUUUA-5') that directs transcription initiation that is almost completely conserved in all genes in both RSVA and RSVB, with only the polymerase gene (3'-CCCUGUUUU-5') in both RSVA and RSVB and the small hydrophobic gene (3'-CCCCAUUUA-5') in RSVB typically having a different sequence. In one of our RSVB genome sequences (KF826829) we found a mutation in the GS sequence of the SH gene resulting in 3'-CCCUAUUUA-5' as the GS. It is unknown how this mutation affects the viability of the virus. We found two other published RSVB sequences (JX576736, JX576757) that had a C->U mutation in the same position of the GS sequences in the M and F genes. Interestingly, this mutation actually matches the base at this position of the L gene GS. This suggests that this specific mutation may be tolerated better by the transcriptional machinery than other mutations in the GS sequences. To support this suggestion, a mutagenesis study of the GS showed that this mutation had no negative impact on expression levels [42]. The only other published sequences with mutations in a GS were a temperature sensitive mutant of the RSV A2 strain (U63644) with an A->G mutation at the last position of the

M2 gene GS, which was shown to be the cause of temperature sensitivity [43], and a sequence from China with two mutations in the GS of the F gene (AY198177), which appears to have been introduced by the amplification primers [44].

Each gene in RSV also ends with a less conserved sequence that directs transcription termination and polyadenylation that follows the motif (3'-UCAAUN$_{1-4}$U$_{4-8}$–5') (Table 3). Unlike the GS, these GE sequences are usually different for each gene, different between RSVA and RSVB, and exhibit more strain-to-strain variation. The first five bases are conserved in most of the GE sequences, however, some variation is observed in positions three and four. In our previous genome sequencing study we identified previously unseen variations of the GE sequences [31]. Again in this study we found additional variant GE sequences highlighting the relative flexibility of these sequences, which have been shown to affect the efficiency of transcription termination which in turn affects the expression levels of downstream gene sequences [45]. Most of the variations are the result of differences in the length of the poly U tract.

Stop codon locations are generally well conserved in both location and sequence for most of the genes in RSV. The gene with the most variable stop codons is the G gene for which variation in stop codons is common resulting in proteins of varying lengths [46]. The G gene stop codon positions of sequences produced in this study matched the positions of those in previously published sequences. These positions were codons 298 and 299 in RSVA and 289, 292, and 296 in RSVB using accessions M74568 and AF013254 as reference sequences. In RSVA both stop codons are utilized with about equal frequency, while in RSVB the stop codon at position 289 is utilized most frequently in the BA clade which represents the majority of recent isolates. In one of our RSVB sequences (KF826839) we identified a premature stop codon in the SH gene resulting in a reduction in the expected protein size from 65 aa to 56 aa. We were able to grow this virus in culture and we confirmed by sequencing the SH gene that the isolate also contained the mutation. The SH protein is a viroporin that has been shown to form a pentameric ring in the cell membrane that functions as a cation channel [47]. The missing 9 aa would be part of an extended loop structure of the extracellular C-terminal domain. It is not known if these amino acids are important for protein function.

Evolutionary Rates

The evolutionary rates of the whole genome sequences were 6.72×10^{-4} substitutions/site/year [95% HPD 5.61×10^{-4} to 7.91×10^{-4}] for RSVA and 7.69×10^{-4} substitutions/site/year [95% HPD 6.81×10^{-4} to 8.62×10^{-4}] for RSVB. The evolutionary rates based on the G CDS were 1.86×10^{-3} substitutions/site/year [95% HPD 1.64×10^{-3} to 2.08×10^{-3}] for RSVA and 2.43×10^{-3} substitutions/site/year [95% HPD 2.05×10^{-3} to 2.84×10^{-3}] for RSVB. Evolutionary rates varied among individual genes, but all remained within the range of 4.22×10^{-4}–3.73×10^{-3} substitutions/site/year (Fig. 3), similar to what has been observed previously [38].

Phylogenetic Trees

We inferred maximum likelihood (ML) and maximum clade credibility (MCC) trees for the whole genome and full G gene CDSs and only ML trees for the partial G gene CDSs. For RSVA, the partial G gene CDS covered nucleotides 677–891 (216 nucleotides) for sequence JN257702 and for RSVB, it covered nucleotides 640–853 (214 nucleotides) for sequence AY353550. The size of the CDS fragment varied due to various insertions. This region was reported most frequently in previous RSV phylogenetic studies and includes the 72 nucleotide insertion found in the recent RSVA ON1 genotype and 60 nucleotide insertion found in the RSVB BA genotype. We found that the trees produced by both methods for the whole genome and full G gene CDS sequences were similar, and that there was good agreement between the

Table 3. Gene End Sequences of RSVA and RSVB.

Gene	RSVA		RSVB	
	Sequence 3'→5'[a]	Occurrences[b]	Sequence 3'→5'	Occurrences
NS1	UCAAUUAUAUUUU	34	UCAAUUAUAUUUU	23
NS2	UCAUU-AAAUUUU	17	UCAUU-A-AUUUUU	23
	UCAAU-AAAUUUU	16		
N	UCAAUU—AUUUUUU	27	UCAAUU—GUUUUUU	22
	UCAAUU—AUUUUUUU	1	UCAAUU—GUUUUU	1
	UCAAUU—AUUUUU	3		
	UCAAUU—GUUUUU	2		
P	UCAAU—-GUUUUUUU	27	UCAUU—-GUUUUUUU	22
	UCAAU—-GUUUUUUUU	2	UCAUU—-GUUUUUUUU	1
	UCAAU—-GUUUUUU	2		
	UCAAU—-GUUUUU	1		
	UCAAU—GCUUUUUU	2		
M	UCAAUU—AUUUUUU	12	UCCAUUU-AUUUU	22
	UCAAUU—AUUUUU	19	UCUAUUU-AUUUU	1
	UCAAUU—AUUUUUUU	1		
SH	UCAAUU-AAUUUUU	26	UCAAU-AAAUUUUU	21
	UCAAUU-AAUUUUUU	6	UCAAU-AAAUUUU	2
G	UCAGU—AAUUUUU	29	UCAAU-AAAUUUUU	20
	UCAGU—AAUUUUUU	2	UCAAU-AAGUUUUU	3
	UCAAU—AAUUUUU	2		
F	UCAAU-AUAUUUU	29	UCAAU-GUAUUUUU	21
	UCAAU-AUAUUUUU	1	UCAAU-AUAUUUUU	2
	UCAAU-AUAUUUUUU	2		
	UCAAU-AUAUUUUUUU	1		
	UCAGU-AUAUUUU	1		
M2	UCAAU-AAAUUUU	30	UCAAU-AGAUUUU	20
L	UCAAU-AAAUUUU	31	UCAAU—AAUUUUUU	21
			UCAAU—AAUUUUU	1
			UCAGU—AAUUUUUU	1

[a] Underlined sequences were identified previously. Dashes were added so the U-tracts line up.

[b] These numbers only include sequences from this study.

whole genome and full G gene CDS trees. However, there were noticeable topological differences in the partial G gene CDS trees relative to the whole genome and full G gene CDS trees. These rearrangements were most evident in RSVB, for which sequences belonging to SAB3 genotype that lack the 60 nucleotide insertion were found to cluster within the BA genotype only in the partial CDS tree. These results suggest that both the whole genome and full G gene CDS are suitable for evolutionary analyses, but the partial G CDS alone may produce misleading results.

RSVA Phylogenetic Analysis

The MCC tree inferred for the RSVA genome (Fig. 4) includes viruses that have previously been characterized as belonging to one of the defined genotypes: GA1, GA2, GA5, GA7, and

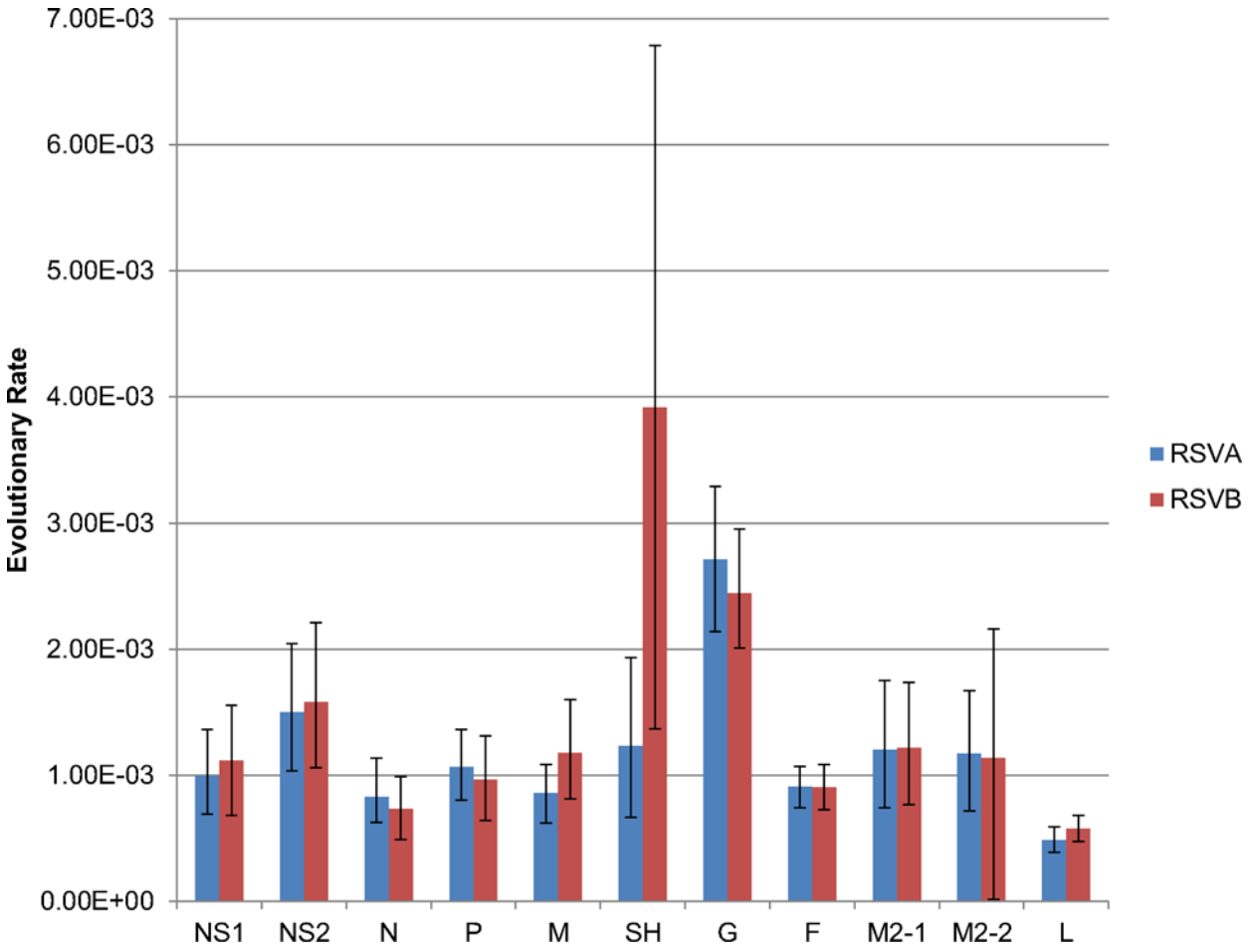


Fig 3. CDS Evolutionary Rates. This plot shows the estimated evolutionary rates for each CDS in RSVA and RSVB. Error bars represent the 95% HPD values. Rates were determined using BEAST v1.8.0 with the GTR model of nucleotide substitution, a gamma-distributed rate variation among sites, an uncorrelated lognormal relaxed molecular clock, and a flexible Bayesian skyline tree prior. Each CDS was ran with a chain length of 50 million and sampled 10,000 times.

doi:10.1371/journal.pone.0120098.g003

ON1. The full CDS tree (S1 Fig.) also included viruses from genotypes NA1 and CB-A, while the partial CDS tree (S2 Fig.) includes viruses from NA2 and a single virus from GA3 and SAA1. We found that the majority of sequences in our data set that were collected globally over the past 10 years belong to one of two clades of RSVA. The larger of these clades encompasses the GA2, NA1, NA2, CB-A, and ON1 genotypes, with the oldest virus dating back to 1994. Different studies have used different reference sequences for defining genotypes, making it difficult at times to distinguish between the genotypes. Therefore, we found genotypes not to be particularly useful to characterize viral diversity in our study, and instead will refer to this large clade simply as the GA2 clade based on the oldest genotype found within it. The ON1 genotype contains a 72 nucleotide insertion in the second variable region of the G gene and includes sequences from 2012–2013 from Canada, Italy, Germany, China, Japan, South Korea, South Africa, Croatia and India [6,9]. The RSVB BA genotype also has a large insertion in the same region (60 nt) that emerged in the late 1990s and then rapidly spread globally. It will be interesting to see if the ON1 genotype becomes the predominant RSVA genotype as the BA genotype did for RSVB.

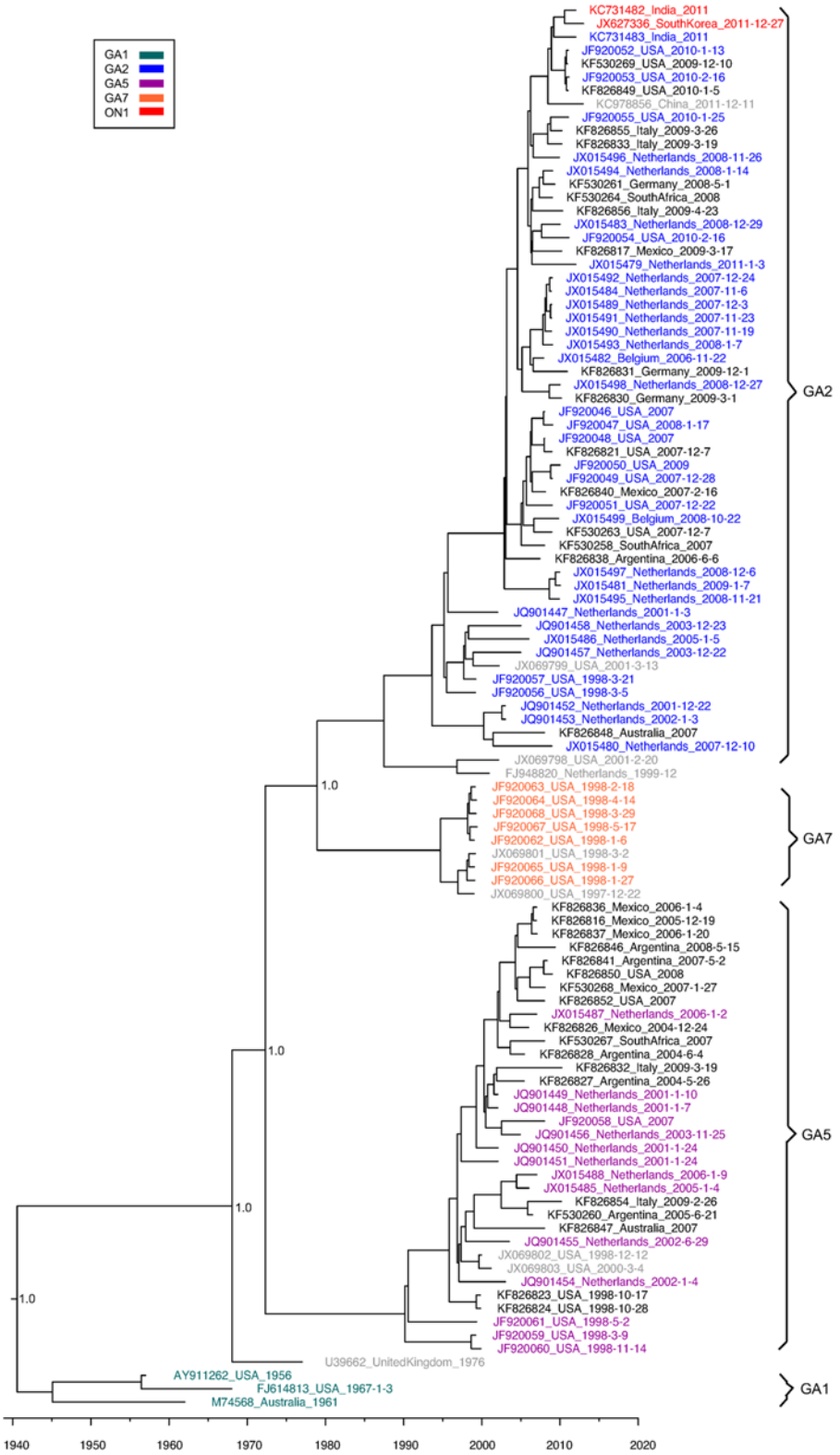


Fig 4. Tree of RSVA Genome Sequences. This is a maximum clade credibility tree of RSVA genome sequences generated in this study and retrieved from GenBank. Tip times correspond to date of collection with the scale axis across the bottom showing the years. Tip labels show the accession number, country of isolation, and collection date. The labels are color coded with black for sequences from this study (FTS), grey

for sequences with an undetermined genotype (UND), and the remaining colors corresponding to previous published genotypes as show in the key in the upper left corner. Brackets highlight the major clades. Bayesian posterior probabilities are shown for key nodes.

doi:10.1371/journal.pone.0120098.g004

The second predominant clade contains sequences from the GA5 genotype, with the oldest one from 1993. Of the 34 RSVA genomes sequenced in this study, 16 belonged to the GA2 clade and 18 belonged to the GA5 clade. The other two clades represent the GA1 and GA7 genotypes. The GA1 clade contains the oldest identified RSVA viruses (1956–1967), two viruses from our previous study (JF920069 and JF920070, not shown due to potential contamination), and a group of viruses from Iran from 2008–2009 (only in the partial CDS tree). One of the Iran viruses (GU339399) is a 100% match to the A2 strain from 1961 (M74568) and the remaining strains are much less divergent than one would expect given the time frame [48,49].

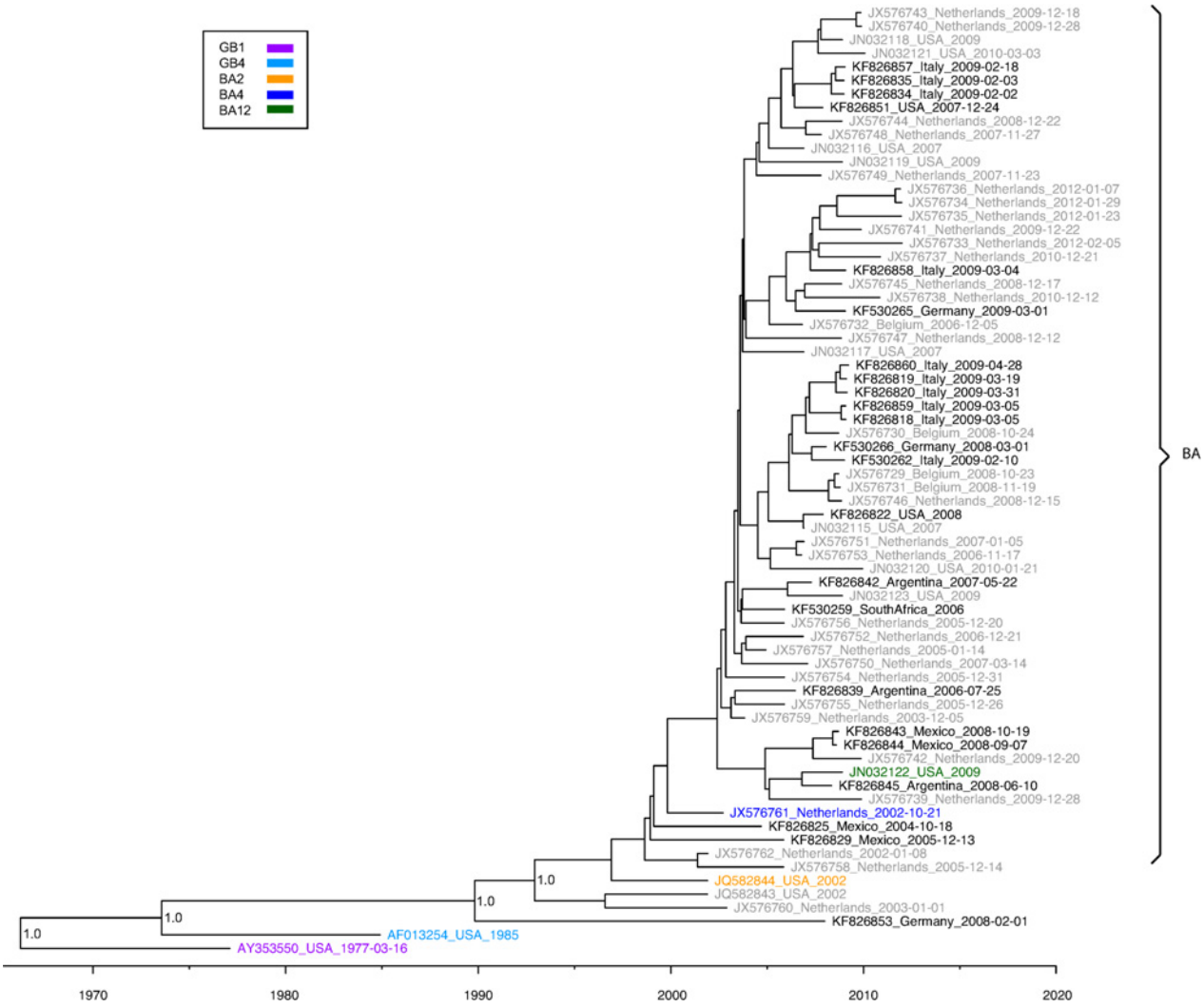


Fig 5. Tree of RSVB Genome Sequences. This is a maximum clade credibility tree of RSVB genome sequences generated in this study and retrieved from GenBank. Tip times correspond to date of collection with the scale axis across the bottom showing the years. Tip labels show the accession number, country of isolation, and collection date. The labels are color coded with black for sequences from this study (FTS), grey for sequences with an undetermined genotype (UND), and the remaining colors corresponding to previous published genotypes as show in the key in the upper left corner. Brackets highlight the major clades. Bayesian posterior probabilities are shown for key nodes.

doi:10.1371/journal.pone.0120098.g005

Therefore, we suspect that this could be a re-introduction of a laboratory strain into the population and is now circulating in Iran. Since we retrieved the sequences used in this study, another study has been published showing additional viruses from this clade isolated in 2012 and 2013 in Iran, providing further evidence that the this virus is indeed currently circulating and not simply a lab/sequencing error [50]. For the GA7 clade only one virus is from after 2002 (JX256946) and is a 100% match to a 1994 virus (JX256947) from the same study suggesting possible contamination [51]. The relatively low number of GA7 sequences suggests that this clade may no longer be circulating or present infrequently enough that it is rarely detected.

RSVB Phylogenetic Analysis

The RSVB genome tree (Fig. 5) includes viruses that had previously been described as belonging to genotypes GB1, GB3, GB4, SAB3, and BA. The full CDS tree (S3 Fig.) also includes viruses from genotypes SAB1 and SAB2 while the partial CDS tree (S4 Fig.) includes viruses from SAB4, GB2 and GB12. Since the RSVB BA genotype emerged in the late 1990s it spread globally and became the predominant genotype [52]. Matching this data, we found the BA clade to represent the largest proportion of sequences in this study, with 22 of the 23 viruses we sequenced belonging to this clade. Many studies have divided members in the BA clade into as many as 13 different genotypes. However, since different studies use different reference sequences for the genotypes and disagree on the total number of genotypes there is much overlap between genotypes, as evidenced by the partial CDS tree. For example, sequences classified as belonging to the BA4 genotype are intermixed with those reported as belonging to the BA7, BA8, BA9, and BA10 genotypes. Despite some fairly distinct clusters of sequences most of the BA viruses could not readily be divided into separate clades. Only one of the viruses (KF826853) sequenced in this study belonged to a genotype other than BA (GB3). Viruses in the GB3/SAB4 clade were identified from 2000 to 2012 in Africa, Asia, Europe, North America, and South America showing that this clade has continued to circulate globally at low levels [53]. Viruses have also been identified as recently as 2011 for the GB2 clade [54]. This shows that even though the viruses that belong to the BA clade have almost completely replaced the older clades some of the non-BA clades have continued to circulate either in distinct regions or globally at low levels.

Conclusions

Through this work we have contributed an additional 57 RSVA and RSVB whole virus genomes to GenBank, significantly increasing the total number of available genomes for RSV. In addition, these genomes have considerable geographic diversity. By incorporating additional publically available sequences in our analyses we were able to develop a better understanding of the evolution and global circulation of RSV viruses. We found that the GA2 and GA5 clades for RSVA and the BA clade for RSVB have been the predominant RSV clades circulating globally for at least the past 10 years. Importantly, we did not observe a strong spatial structure in the tree, with evidence that both main RSVA clades co-circulate globally. These findings suggest that the evolution of RSVA and RSVB is not strongly regionalized, although some small clades were only identified in one country (e.g., GA7 in the United States). Further surveillance and sequencing of RSV is required to determine if these minor clades are still circulating in under sampled regions, as well as to know whether certain RSV lineages dominate over sustained time periods within a defined region, or fluctuate year-to-year. Our phylogenies also indicate that previously used genotype classifications may not be monophyletic, and therefore highlight the importance of developing improved nomenclature for RSV.

The majority of RSV sequencing has been focused on RSV evolution and therefore has primarily targeted only a partial region of a single RSV gene. Until recently, this meant that there were only very few sequences for the conserved regions of RSV, which are critical for the development of robust diagnostics. Since viruses continuously evolve, the availability of more whole virus genome sequences from recent isolates representing the diversity of the circulating RSV population will aid in the development of reliable diagnostics and monitoring for changes in areas targeted by current and future RSV therapeutics.

Supporting Information

S1 Fig. Tree of RSVA Attachment Glycoprotein CDS Sequences. This is a maximum clade credibility tree of RSVA attachment glycoprotein CDS sequences generated in this study and retrieved from GenBank. Tip times correspond to date of collection with the scale axis across the bottom showing the years. Tip labels show the accession number, country of isolation, and collection date. The labels are color coded with black for sequences from this study (FTS), grey for sequences with an undetermined genotype (UND), and the remaining colors corresponding to previous published genotypes as show in the key in the upper left corner. Brackets highlight the major clades. Bayesian posterior probabilities are shown for key nodes.
(PDF)

S2 Fig. Tree of RSVA Attachment Glycoprotein Partial CDS Sequences. This is a maximum likelihood tree of RSVA attachment glycoprotein partial CDS sequences corresponding to the second hyper variable region generated in this study and retrieved from GenBank. Tip labels show the accession number, country of isolation, and collection date. The labels are color coded with black for sequences from this study (FTS), grey for sequences with an undetermined genotype (UND), and the remaining colors corresponding to previous published genotypes as show in the key in the upper left corner. Brackets highlight the major clades.
(PDF)

S3 Fig. Tree of RSVB Attachment Glycoprotein CDS Sequences. This is a maximum clade credibility tree of RSVB attachment glycoprotein CDS sequences generated in this study and retrieved from GenBank. Tip times correspond to date of collection with the scale axis across the bottom showing the years. Tip labels show the accession number, country of isolation, and collection date. The labels are color coded with black for sequences from this study (FTS), grey for sequences with an undetermined genotype (UND), and the remaining colors corresponding to previous published genotypes as show in the key in the upper left corner. Brackets highlight the major clades. Bayesian posterior probabilities are shown for key nodes.
(PDF)

S4 Fig. Tree of RSVB Attachment Glycoprotein Partial CDS Sequences. This is a maximum likelihood tree of RSVB attachment glycoprotein partial CDS sequences corresponding to the second hyper variable region generated in this study and retrieved from GenBank. Tip labels show the accession number, country of isolation, and collection date. The labels are color coded with black for sequences from this study (FTS), grey for sequences with an undetermined genotype (UND), and the remaining colors corresponding to previous published genotypes as show in the key in the upper left corner. Brackets highlight the major clades.
(PDF)

S1 Table. Primer Pairs Used for Next Gen Sequencing.
(XLS)

S1 Text. Primer Pairs Used for Sanger Sequencing of RSVA.
(TXT)

S2 Text. Primer Pairs Used for Sanger Sequencing of RSVB.
(TXT)

Acknowledgments

We would like to thank Cecilia Rebuffo-Scheer and Swati Kumar for coordinating the collection of the RSV positive samples and isolates from all the different sites. We would also like to thank Marthi Pretorius and Amelia Buys for propagating the RSV positive specimens from the Department Medical Virology at the University of Pretoria.

Disclaimer. The findings and conclusions in this publication are those of the author(s) and do not necessarily represent the views of the Centers for Disease Control and Prevention.

Author Contributions

Conceived and designed the experiments: MEB JH RAH CDT HAL JVW KJH. Performed the experiments: MEB JH SS JB. Analyzed the data: MEB JH SS MIN JB RAH CDT HAL. Contributed reagents/materials/analysis tools: DEN VF GG HDB CV TK MV JVW KJH. Wrote the paper: MEB JH SS MIN HAL KJH.

References

1. Glezen WP, Taber LH, Frank AL, Kasel JA. Risk of primary infection and reinfection with respiratory syncytial virus. Am J Dis Child. 1986; 140(6):543–6. PMID: 3706232
2. Falsey AR. Respiratory syncytial virus infection in elderly and high-risk adults. Exp Lung Res. 2005; 31 Suppl 1:77. PMID: 16395866
3. Sorvillo FJ, Huie SF, Strassburg MA, Butsumyo A, Shandera WX, Fannin SL. An outbreak of respiratory syncytial virus pneumonia in a nursing home for the elderly. J Infect. 1984; 9(3):252–6. PMID: 6527041
4. Mufson MA, Orvell C, Rafnar B, Norrby E. Two distinct subtypes of human respiratory syncytial virus. J Gen Virol. 1985; 66 (Pt 10):2111–24. PMID: 2413163
5. Cane PA, Matthews DA, Pringle CR. Identification of variable domains of the attachment (G) protein of subgroup A respiratory syncytial viruses. J Gen Virol. 1991; 72 (Pt 9):2091–6. PMID: 1895054
6. Eshaghi A, Duvvuri VR, Lai R, Nadarajah JT, Li A, Patel SN, et al. Genetic variability of human respiratory syncytial virus A strains circulating in Ontario: A novel genotype with a 72 nucleotide G gene duplication. PLoS One. 2012; 7(3):e32807. doi: 10.1371/journal.pone.0032807 PMID: 22470426
7. Parveen S, Sullender WM, Fowler K, Lefkowitz EJ, Kapoor SK, Broor S. Genetic variability in the G protein gene of group A and B respiratory syncytial viruses from India. J Clin Microbiol. 2006; 44(9):3055–64. PMID: 16954227
8. Choi EH, Lee HJ. Genetic diversity and molecular epidemiology of the G protein of subgroups A and B of respiratory syncytial viruses isolated over 9 consecutive epidemics in Korea. J Infect Dis. 2000; 181(5):1547–56. PMID: 10823752
9. Tsukagoshi H, Yokoi H, Kobayashi M, Kushibuchi I, Okamoto-Nakagawa R, Yoshida A, et al. Genetic analysis of attachment glycoprotein (G) gene in new genotype ON1 of human respiratory syncytial virus detected in Japan. Microbiol Immunol. 2013; 57(9):655–9. doi: 10.1111/1348-0421.12075 PMID: 23750702
10. Reiche J, Schweiger B. Genetic variability of group A human respiratory syncytial virus strains circulating in Germany from 1998 to 2007. J Clin Microbiol. 2009; 47(6):1800–10. doi: 10.1128/JCM.02286-08 PMID: 19386848
11. van Niekerk S, Venter M. Replacement of previously circulating respiratory syncytial virus subtype B strains with the BA genotype in South Africa. J Virol. 2011; 85(17):8789–97. doi: 10.1128/JVI.02623-10 PMID: 21715483

12. Peret TC, Hall CB, Schnabel KC, Golub JA, Anderson LJ. Circulation patterns of genetically distinct group A and B strains of human respiratory syncytial virus in a community. J Gen Virol. 1998; 79 (Pt 9):2221–9. PMID: 9747732

13. Kok T, Mickan LD, Burrell CJ. Routine diagnosis of seven respiratory viruses and Mycoplasma pneumoniae by enzyme immunoassay. J Virol Methods. 1994; 50(1–3):87–100. PMID: 7714062

14. Schepetiuk SK, Kok T. The use of MDCK, MEK and LLC-MK2 cell lines with enzyme immunoassay for the isolation of influenza and parainfluenza viruses from clinical specimens. J Virol Methods. 1993; 42(2–3):241–50. PMID: 8514846

15. Gerna G, Campanini G, Rognoni V, Marchi A, Rovida F, Piralla A, et al. Correlation of viral load as determined by real-time RT-PCR and clinical characteristics of respiratory syncytial virus lower respiratory tract infections in early infancy. J Clin Virol. 2008; 41(1):45–8. PMID: 18082444

16. Bonroy C, Vankeerberghen A, Boel A, De Beenhouwer H. Use of a multiplex real-time PCR to study the incidence of human metapneumovirus and human respiratory syncytial virus infections during two winter seasons in a Belgian paediatric hospital. Clin Microbiol Infect. 2007; 13(5):504–9. PMID: 17263836

17. Williams JV, Harris PA, Tollefson SJ, Halburnt-Rush LL, Pingsterhaus JM, Edwards KM, et al. Human metapneumovirus and lower respiratory tract disease in otherwise healthy infants and children. N Engl J Med. 2004; 350(5):443–50. PMID: 14749452

18. Li K, Shrivastava S, Brownley A, Katzel D, Bera J, Nguyen AT, et al. Automated degenerate PCR primer design for high-throughput sequencing improves efficiency of viral sequencing. Virol J. 2012; 9:261,422X-9–261. doi: 10.1186/1743-422X-9-261 PMID: 23131097

19. Treangen TJ, Sommer DD, Angly FE, Koren S, Pop M. Next generation sequence assembly with AMOS. Curr Protoc Bioinformatics. 2011;Chapter 11:Unit 11.8.

20. Wang S, Sundaram JP, Stockwell TB. VIGOR extended to annotate genomes for additional 12 different viruses. Nucleic Acids Res. 2012; 40(Web Server issue):W186–92. doi: 10.1093/nar/gks528 PMID: 22669909

21. Djikeng A, Halpin R, Kuzmickas R, Depasse J, Feldblyum J, Sengamalay N, et al. Viral genome sequencing by random priming methods. BMC Genomics. 2008; 9:5,2164–9–5. doi: 10.1186/1471-2164-9-5 PMID: 18179705

22. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. 2013; 30(4):772–80. doi: 10.1093/molbev/mst010 PMID: 23329690

23. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P. RDP3: A flexible and fast computer program for analyzing recombination. Bioinformatics. 2010; 26(19):2462–3. doi: 10.1093/bioinformatics/btq467 PMID: 20798170

24. Tan L, Lemey P, Houspie L, Viveen MC, Jansen NJ, van Loon AM, et al. Genetic variability among complete human respiratory syncytial virus subgroup A genomes: Bridging molecular evolutionary dynamics and epidemiology. PLoS One. 2012; 7(12):e51439. doi: 10.1371/journal.pone.0051439 PMID: 23236501

25. Delport W, Poon AF, Frost SD, Kosakovsky Pond SL. Datamonkey 2010: A suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics. 2010; 26(19):2455–7. doi: 10.1093/bioinformatics/btq429 PMID: 20671151

26. Pond SL, Frost SD, Muse SV. HyPhy: Hypothesis testing using phylogenies. Bioinformatics. 2005; 21(5):676–9. PMID: 15509596

27. Kosakovsky Pond SL, Frost SD. Not so different after all: A comparison of methods for detecting amino acid sites under selection. Mol Biol Evol. 2005; 22(5):1208–22. PMID: 15703242

28. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, et al. FUBAR: A fast, unconstrained bayesian approximation for inferring selection. Mol Biol Evol. 2013; 30(5):1196–205. doi: 10.1093/molbev/mst030 PMID: 23420840

29. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 2007; 7:214. PMID: 17996036

30. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013; 30(12):2725–9. doi: 10.1093/molbev/mst197 PMID: 24132122

31. Rebuffo-Scheer C, Bose M, He J, Khaja S, Ulatowski M, Beck ET, et al. Whole genome sequencing and evolutionary analysis of human respiratory syncytial virus A and B from Milwaukee, WI 1998–2010. PLoS One. 2011; 6(10):e25468. doi: 10.1371/journal.pone.0025468 PMID: 21998661

32. Agenbach E, Tiemessen CT, Venter M. Amino acid variation within the fusion protein of respiratory syncytial virus subtype A and B strains during annual epidemics in South Africa. Virus Genes. 2005; 30(2):267–78. PMID: 15744582

33. Gaunt ER, Jansen RR, Poovorawan Y, Templeton KE, Toms GL, Simmonds P. Molecular epidemiology and evolution of human respiratory syncytial virus and human metapneumovirus. PLoS One. 2011; 6(3):e17427. doi: 10.1371/journal.pone.0017427 PMID: 21390255

34. Kryazhimskiy S, Plotkin JB. The population genetics of dN/dS. PLoS Genet. 2008; 4(12):e1000304. doi: 10.1371/journal.pgen.1000304 PMID: 19081788

35. Woelk CH, Holmes EC. Variable immune-driven natural selection in the attachment (G) glycoprotein of respiratory syncytial virus (RSV). J Mol Evol. 2001; 52(2):182–92. PMID: 11231898

36. Zlateva KT, Lemey P, Moes E, Vandamme AM, Van Ranst M. Genetic variability and molecular evolution of the human respiratory syncytial virus subgroup B attachment G protein. J Virol. 2005; 79(14): 9157–67. PMID: 15994810

37. Zlateva KT, Lemey P, Vandamme AM, Van Ranst M. Molecular evolution and circulation patterns of human respiratory syncytial virus subgroup A: Positively selected sites in the attachment G glycoprotein. J Virol. 2004; 78(9):4675–83. PMID: 15078950

38. Tan L, Coenjaerts FE, Houspie L, Viveen MC, van Bleek GM, Wiertz EJ, et al. The comparative genomics of human respiratory syncytial virus subgroups A and B: Genetic variability and molecular evolutionary dynamics. J Virol. 2013; 87(14):8213–26. doi: 10.1128/JVI.03278-12 PMID: 23698290

39. Botosso VF, Zanotto PM, Ueda M, Arruda E, Gilio AE, Vieira SE, et al. Positive selection results in frequent reversible amino acid replacements in the G protein gene of human respiratory syncytial virus. PLoS Pathog. 2009; 5(1):e1000254. doi: 10.1371/journal.ppat.1000254 PMID: 19119418

40. Norrby E, Mufson MA, Alexander H, Houghten RA, Lerner RA. Site-directed serology with synthetic peptides representing the large glycoprotein G of respiratory syncytial virus. Proc Natl Acad Sci U S A. 1987; 84(18):6572–6. PMID: 3476962

41. Cane PA. Analysis of linear epitopes recognised by the primary human antibody response to a variable region of the attachment (G) protein of respiratory syncytial virus. J Med Virol. 1997; 51(4):297–304. PMID: 9093944

42. Kuo L, Fearns R, Collins PL. Analysis of the gene start and gene end signals of human respiratory syncytial virus: Quasi-templated initiation at position 1 of the encoded mRNA. J Virol. 1997; 71(7):4944–53. PMID: 9188557

43. Whitehead SS, Firestone CY, Collins PL, Murphy BR. A single nucleotide substitution in the transcription start signal of the M2 gene of respiratory syncytial virus vaccine candidate cpts248/404 is the major determinant of the temperature-sensitive and attenuation phenotypes. Virology. 1998; 247(2):232–9. PMID: 9705916

44. Deng J, Wang Z, Qian Y, Gen XH, Sun XO, Zhu RN, et al. A preliminary study on the characteristics of fusion proteins genes of respiratory syncytial virus field strains isolated from children with epidemic asthmalike pneumonia. Chinese Journal of Pediatrics. 1999; 37(5):287–90.

45. Sutherland KA, Collins PL, Peeples ME. Synergistic effects of gene-end signal mutations and the M2–1 protein on transcription termination by respiratory syncytial virus. Virology. 2001; 288(2):295–307. PMID: 11601901

46. Almajhdi FN, Farrag MA, Amer HM. Genetic diversity in the G protein gene of group A human respiratory syncytial viruses circulating in Riyadh, Saudi Arabia. Arch Virol. 2014; 159(1):73–81. doi: 10.1007/s00705-013-1792-6 PMID: 23884633

47. Gan SW, Tan E, Lin X, Yu D, Wang J, Tan GM, et al. The small hydrophobic protein of the human respiratory syncytial virus forms pentameric ion channels. J Biol Chem. 2012; 287(29):24671–89. doi: 10.1074/jbc.M111.332791 PMID: 22621926

48. Faghihloo E, Salimi V, Rezaei F, Naseri M, Mamishi S, Mahmoodi M, et al. Genetic diversity in the G protein gene of human respiratory syncytial virus among Iranian children with acute respiratory symptoms. Iran J Pediatr. 2011; 21(1):58–64. PMID: 23056765

49. Faghihloo E, Rezaie F, Salimi V, Naseri M, Mamishi S, Mahmoodi M, et al. Molecular epidemiology of human respiratory syncytial virus in Iran. Acta Virol. 2011; 55(1):81–3. PMID: 21434709

50. Faghihloo E, Yavarian J, Jandaghi NZ, Shadab A, Azad TM. Genotype circulation pattern of human respiratory syncytial virus in Iran. Infect Genet Evol. 2014; 22:130–3. doi: 10.1016/j.meegid.2014.01.009 PMID: 24462624

51. Khor CS, Sam IC, Hooi PS, Chan YF. Displacement of predominant respiratory syncytial virus genotypes in Malaysia between 1989 and 2011. Infect Genet Evol. 2013; 14:357–60. doi: 10.1016/j.meegid.2012.12.017 PMID: 23305888

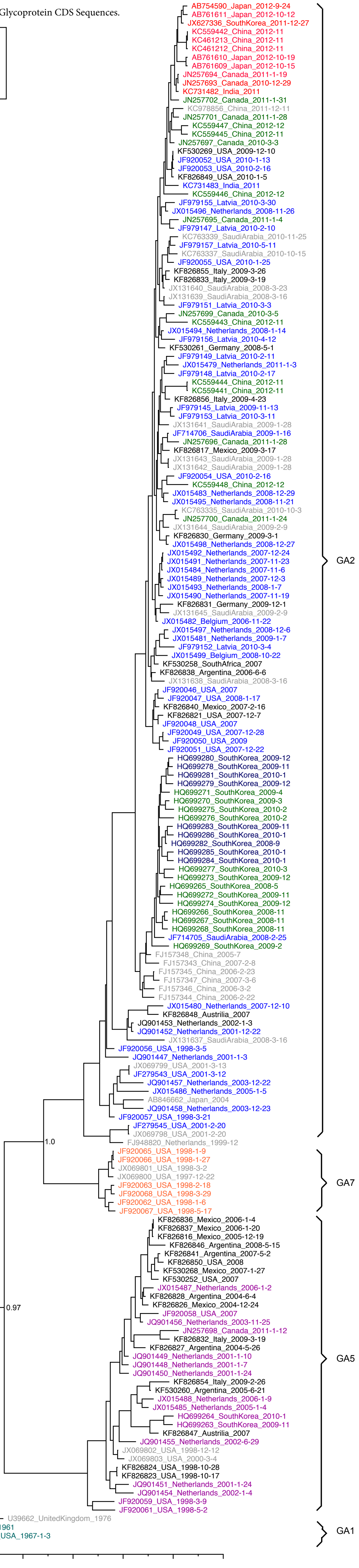
52. Trento A, Galiano M, Videla C, Carballal G, Garcia-Barreno B, Melero JA, et al. Major changes in the G protein of human respiratory syncytial virus isolates introduced by a duplication of 60 nucleotides. J Gen Virol. 2003; 84(Pt 11):3115–20. PMID: 14573817

53. Arnott A, Vong S, Mardy S, Chu S, Naughtin M, Sovann L, et al. A study of the genetic variability of human respiratory syncytial virus (HRSV) in Cambodia reveals the existence of a new HRSV group B genotype. J Clin Microbiol. 2011; 49(10):3504–13. doi: 10.1128/JCM.01131-11 PMID: 21865418

54. Zhang ZY, Du LN, Chen X, Zhao Y, Liu EM, Yang XQ, et al. Genetic variability of respiratory syncytial viruses (RSV) prevalent in southwestern china from 2006 to 2009: Emergence of subgroup B and A RSV as dominant strains. J Clin Microbiol. 2010; 48(4):1201–7. doi: 10.1128/JCM.02258-09 PMID: 20147636

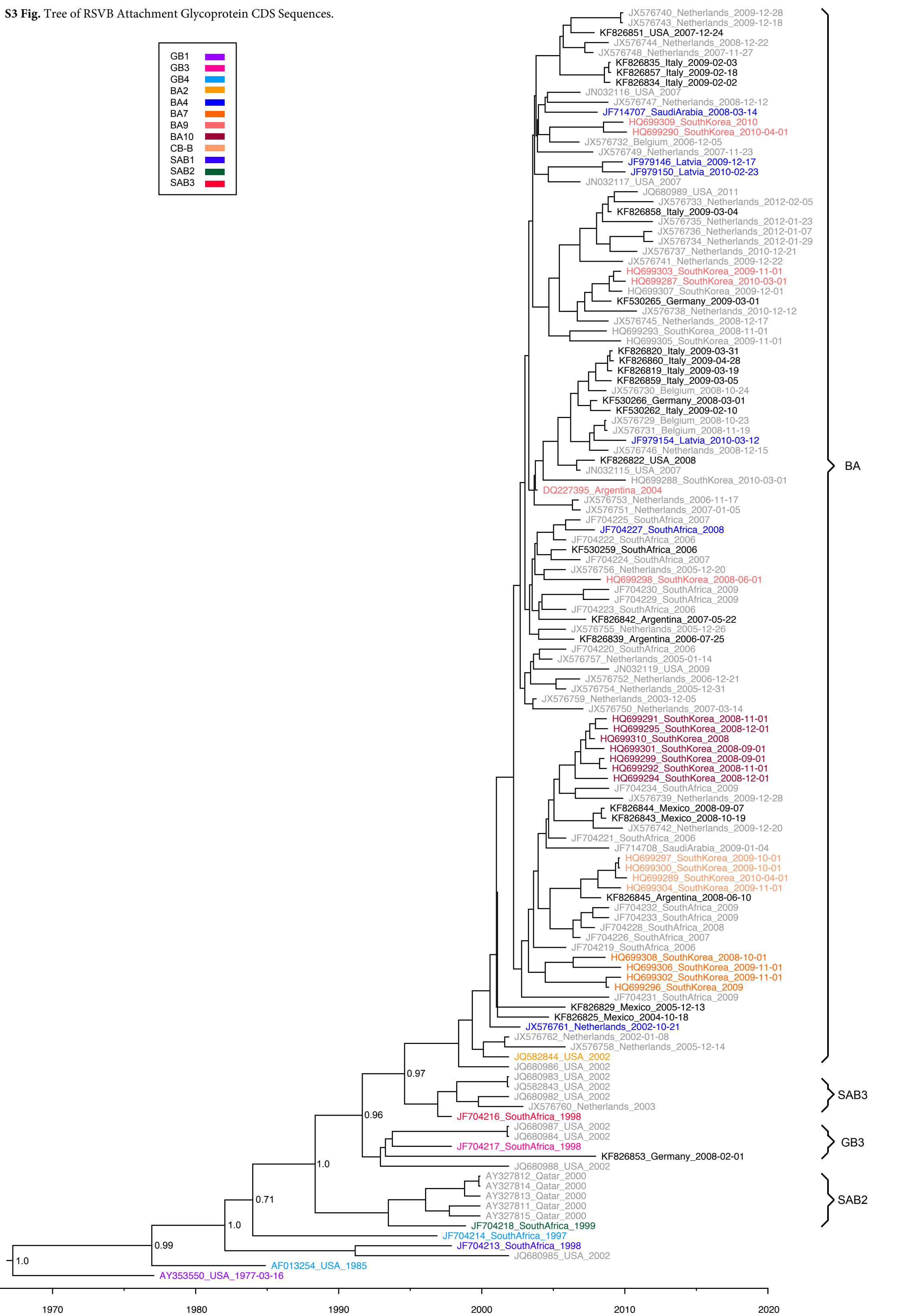
Supporting Information

S1 Fig. Tree of RSVA Attachment Glycoprotein CDS Sequences.



RdRp Tree of RTX-A Alveolate Bipartite Partial CDS Sequences
Alveolata
Stramenopiles
Rhizaria
Haptophyta
Cryptophyta
Archaeplastida
Amoebozoa
Opisthokonta
Excavata
Unclassified
Viruses
Bacteria
Archaea
Unknown
Unclassified
Viruses
Bacteria
Archaea
Unknown
Unclassified
Viruses
Bacteria
Archaea
Unknown

S3 Fig. Tree of RSVB Attachment Glycoprotein CDS Sequences.



S4 Fig. Tree of HXVB Attachment Glycoprotein Partial CDS Sequences.

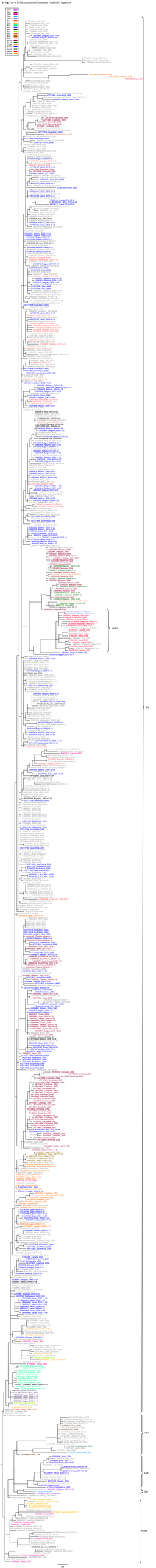


Table S1. Primer Pairs Used for Next Gen Sequencing.

Forward Primer Name	Forward Primer Sequence	Reverse Primer Name	Reverse Primer Sequence
RSV_Seg1_01F	ACGCGAAAAAATGCGTACTACAAAC	RSV_Seg1_4017R	CGTGTAGCTGTRTGYTTCCAA
RSV_Seg2_3905F	GGAGCATTCAAATAYATMAAGCC	RSV_Seg2_7432R	TRCTTARTGTRACTGGTGTG
RSV_Seg3_7215F	TGATGCATCAATATCTCAAGTCA	RSV_Seg3_11165R	GRCCTATDCCTGCATACTC
RSV_Seg4_10959F	TGGACCATWGAAGCYATATCA	RSV_Seg4_15333R	AGTGTCAAAAACTAATRTCTCGT
RSV_Seg1_6F	AAAAAATGCGTACTACAAACTTGC	RSV_Seg4_2_15344R	ACGAGAAAAAAAGTGTCAAAAACTAAT

S1 Text. Primer Pairs Used for Sanger Sequencing of RSVA.

>MAIN_RSVAT01_F_1M13 T01::A01 FORWARD
TGTAAAACGACGGCCAGTAAAAAATGCGTACTACAAACTTGC
>MAIN_RSVAT01_R_671M13 T01::A01 REVERSE
CAGGAAACAGCTATGACCCTGTGATCATCAGTCTTTGTG
>MAIN_RSVAT01_F_2M13 T01::A02 FORWARD
TGTAAAACGACGGCCAGTAAAAATGCGTACTACAAACTTGC
>MAIN_RSVAT01_R_671M13_v2 T01::A02 REVERSE
CAGGAAACAGCTATGACCCTGTGATCATCAGTCTTTGTGG
>MAIN_RSVAT01_F_2M13 T01::A03 FORWARD
TGTAAAACGACGGCCAGTAAAAATGCGTACTACAAACTTGC
>MAIN_RSVAT01_R_670M13 T01::A03 REVERSE
CAGGAAACAGCTATGACCTGTGATCATCAGTCTTTGTGG
>MAIN_RSVAT01_F_33M13 T01::A04 FORWARD
TGTAAAACGACGGCCAGTAAAAATGGGGCAAATAAGAA
>MAIN_RSVAT01_R_695M13 T01::A04 REVERSE
CAGGAAACAGCTATGACCTCTCAAGTGAYAAYGGTCTCAT
>MAIN_RSVAT01_F_315M13 T01::A05 FORWARD
TGTAAAACGACGGCCAGTATCCAATTTCACAACAATGC
>MAIN_RSVAT01_R_954M13 T01::A05 REVERSE
CAGGAAACAGCTATGACCGTAGGCTTAATGCCAATGC
>MAIN_RSVAT01_F_492M13 T01::A06 FORWARD
TGTAAAACGACGGCCAGTGGATTTGAYCTYAATCCATAAATYA
>MAIN_RSVAT01_R_1182M13 T01::A06 REVERSE
CAGGAAACAGCTATGACCTTGATCTTTGTTGAGTGTRTCA
>MAIN_RSVAT01_F_533M13 T01::A07 FORWARD
TGTAAAACGACGGCCAGTGCAAATCAATGTCACTAACACC
>MAIN_RSVAT01_R_1218M13 T01::A07 REVERSE
CAGGAAACAGCTATGACCGCTCCGTTGGATGGTRTATT
>MAIN_RSVAT01_F_790M13 T01::A08 FORWARD
TGTAAAACGACGGCCAGTGGCCACATTTACATTYCTGG
>MAIN_RSVAT01_R_1429M13 T01::A08 REVERSE
CAGGAAACAGCTATGACCTTGTTAYATCYACTCCATTTGC
>MAIN_RSVAT01_F_957M13 T01::A09 FORWARD
TGTAAAACGACGGCCAGTGCAYACWCCCATAATATACAAG
>MAIN_RSVAT01_R_1657M13 T01::A09 REVERSE
CAGGAAACAGCTATGACCCYCCTGCTGCTAATTTRGTT
>MAIN_RSVAT01_F_1054M13 T01::A10 FORWARD
TGTAAAACGACGGCCAGTCATAGTCYARATGGAGCCTG
>MAIN_RSVAT01_R_1756M13 T01::A10 REVERSE
CAGGAAACAGCTATGACCTGTTGGCTATRTCCTTGGGT
>MAIN_RSVAT01_F_1289M13 T01::A11 FORWARD
TGTAAAACGACGGCCAGTAATCACWGAAGATGCYAATCA
>MAIN_RSVAT01_R_1970M13 T01::A11 REVERSE
CAGGAAACAGCTATGACCGCTTGYACACTAGCRTGTCCT
>MAIN_RSVAT01_F_1553M13 T01::A12 FORWARD
TGTAAAACGACGGCCAGTAAGAAATGGGAGAGGTRGC
>MAIN_RSVAT01_R_2215M13 T01::A12 REVERSE
CAGGAAACAGCTATGACCCATTTTCTTTGAGTTGTTCAGC
>MAIN_RSVAT01_F_1736M13 T01::B01 FORWARD
TGTAAAACGACGGCCAGTACCCAAGGAYATAGCCAACA
>MAIN_RSVAT01_R_2433M13 T01::B01 REVERSE

CAGGAAACAGCTATGACCGGTGATGTGAATTTGCCCT
>MAIN_RSVAT01_F_1817M13 T01::B02 FORWARD
TGTAAAACGACGGCCAGTAGCACAATCTTCTACCAGAGG
>MAIN_RSVAT01_R_2506M13 T01::B02 REVERSE
CAGGAAACAGCTATGACCAGGGCTTTCTTTGGTTACTTC
>MAIN_RSVAT01_F_2020M13 T01::B03 FORWARD
TGTAAAACGACGGCCAGTGGWGAAGCAGGATTMTACC
>MAIN_RSVAT01_R_2635M13 T01::B03 REVERSE
CAGGAAACAGCTATGACCRGGATTATCACTTGGYGTAGG
>MAIN_RSVAT01_F_2263M13 T01::B04 FORWARD
TGTAAAACGACGGCCAGTGCTATCAAACATCAGCTTAATCC
>MAIN_RSVAT01_R_2946M13 T01::B04 REVERSE
CAGGAAACAGCTATGACCTCATTCCTGAGTCTTGCCAT
>MAIN_RSVAT01_F_2322M13 T01::B05 FORWARD
TGTAAAACGACGGCCAGTRTGGGGCAAATAAATCATCW
>MAIN_RSVAT01_R_3017M13 T01::B05 REVERSE
CAGGAAACAGCTATGACCTGTTCARTTTCTCTGATGTTGG
>MAIN_RSVAT01_F_2511M13 T01::B06 FORWARD
TGTAAAACGACGGCCAGTCATCAAATTCAACYATTATMAACC
>MAIN_RSVAT01_R_3132M13 T01::B06 REVERSE
CAGGAAACAGCTATGACCGRGTTGGTTWGTTTGTTGG
>MAIN_RSVAT01_F_2650M13 T01::B07 FORWARD
TGTAAAACGACGGCCAGTCAAAGAAACCATMRAAACATTTG
>MAIN_RSVAT01_R_3348M13 T01::B07 REVERSE
CAGGAAACAGCTATGACCGCAGGRTCATCGTCTTTTTC
>MAIN_RSVAT01_F_2674M13 T01::B08 FORWARD
TGTAAAACGACGGCCAGTAACAATGARGAAGAATCYMGC
>MAIN_RSVAT01_R_3374M13 T01::B08 REVERSE
CAGGAAACAGCTATGACCACATGGGCACCCATATTGTA
>MAIN_RSVAT01_F_2976M13 T01::B09 FORWARD
TGTAAAACGACGGCCAGTCAGATGAAGTGTCTCTCAATCC
>MAIN_RSVAT01_R_3684M13 T01::B09 REVERSE
CAGGAAACAGCTATGACCTCATGTGTWGGGTTGAGTGT
>MAIN_RSVAT01_F_3174M13 T01::B10 FORWARD
TGTAAAACGACGGCCAGTCAACAAAACAACCAGCCAAT
>MAIN_RSVAT01_R_3849M13 T01::B10 REVERSE
CAGGAAACAGCTATGACCAAGGKATGATTTTTGCATTTG
>MAIN_RSVAT01_F_3174M13_v2 T01::B11 FORWARD
TGTAAAACGACGGCCAGTCCAACAAAACAACCAGCCAATC
>MAIN_RSVAT01_R_3896M13 T01::B11 REVERSE
CAGGAAACAGCTATGACCTGCTCCTTTGTTGTCAGTCA
>MAIN_RSVAT01_F_3540M13 T01::B12 FORWARD
TGTAAAACGACGGCCAGTTGTCCTTGGATGAAAGAAGC
>MAIN_RSVAT01_R_4183M13 T01::B12 REVERSE
CAGGAAACAGCTATGACCGWTCTGATRAACAATGACTTGG
>MAIN_RSVAT01_F_3664M13 T01::C01 FORWARD
TGTAAAACGACGGCCAGTRACACTCAACCCWACACATG
>MAIN_RSVAT01_R_4351M13 T01::C01 REVERSE
CAGGAAACAGCTATGACCAAGGCCAGAATTTRCTTGAGA
>MAIN_RSVAT01_F_3739M13 T01::C02 FORWARD
TGTAAAACGACGGCCAGTACCAACATACCTAAGATCYATCAG
>MAIN_RSVAT01_R_4414M13 T01::C02 REVERSE
CAGGAAACAGCTATGACCTCATGATGGAGATTATGATTAKCAA
>MAIN_RSVAT01_F_4016M13 T01::C03 FORWARD

TGTAAAACGACGGCCAGTCCCATGGAAGATTAACCYYT
>MAIN_RSVAT01_R_4702M13 T01::C03 REVERSE
CAGGAAACAGCTATGACCTTGKTTTTGGACATGTTTGC
>MAIN_RSVAT01_F_4076M13 T01::C04 FORWARD
TGTAAAACGACGGCCAGTCYACATTCTTCACTTCACMATCA
>MAIN_RSVAT01_R_4746M13 T01::C04 REVERSE
CAGGAAACAGCTATGACCGAGTGTCCCAGGTCCTTTCT
>MAIN_RSVAT01_F_4297M13 T01::C05 FORWARD
TGTAAAACGACGGCCAGTCAACCAATGGAAAATACATCC
>MAIN_RSVAT01_R_4927M13 T01::C05 REVERSE
CAGGAAACAGCTATGACCGTTGCMTCTTGTRTGATTGC
>MAIN_RSVAT01_F_4515M13 T01::C06 FORWARD
TGTAAAACGACGGCCAGTCYCAAAACAGTWACCTTGCAT
>MAIN_RSVAT01_R_5215M13 T01::C06 REVERSE
CAGGAAACAGCTATGACCCATATGCTGCAGGGTACAAA
>MAIN_RSVAT01_F_4573M13 T01::C07 FORWARD
TGTAAAACGACGGCCAGTATYAACATCYCACCATGCAA
>MAIN_RSVAT01_R_5255M13 T01::C07 REVERSE
CAGGAAACAGCTATGACCTATTCTYTTGCAGATRGCCC
>MAIN_RSVAT01_F_4805M13 T01::C08 FORWARD
TGTAAAACGACGGCCAGTACAAATCACAYTATCYATTYTGG
>MAIN_RSVAT01_R_5493M13 T01::C08 REVERSE
CAGGAAACAGCTATGACCRGTTGAGTGGARGGTTTCC
>MAIN_RSVAT01_F_4953M13 T01::C09 FORWARD
TGTAAAACGACGGCCAGTRCATACCTCACCCAGAATCC
>MAIN_RSVAT01_R_5635M13 T01::C09 REVERSE
CAGGAAACAGCTATGACCTTCTGTTTRAKTTGGTCATGG
>MAIN_RSVAT01_F_5060M13 T01::C10 FORWARD
TGTAAAACGACGGCCAGTACCCYRCAATCCACAAYAGT
>MAIN_RSVAT01_R_5748M13 T01::C10 REVERSE
CAGGAAACAGCTATGACCGTGATGTTTTGACYRGAAGC
>MAIN_RSVAT01_F_5100M13 T01::C11 FORWARD
TGTAAAACGACGGCCAGTCAACYCAAAYACAACCYAGC
>MAIN_RSVAT01_R_5781M13 T01::C11 REVERSE
CAGGAAACAGCTATGACCACTGCACTGCATGTTGATTG
>MAIN_RSVAT01_F_5302M13 T01::C12 FORWARD
TGTAAAACGACGGCCAGTAACCAACYMTCAAKACAACC
>MAIN_RSVAT01_R_6001M13 T01::C12 REVERSE
CAGGAAACAGCTATGACCYCTTGGYAGTTCTCTTCTGG
>MAIN_RSVAT01_F_5435M13 T01::D01 FORWARD
TGTAAAACGACGGCCAGTCWCCAAYACCACAGGAAATC
>MAIN_RSVAT01_R_6103M13 T01::D01 REVERSE
CAGGAAACAGCTATGACCTGCAGATCCAACACCTAACA
>MAIN_RSVAT01_F_5466M13 T01::D02 FORWARD
TGTAAAACGACGGCCAGTAAGTCAARWGGAAACCYTCC
>MAIN_RSVAT01_R_6139M13 T01::D02 REVERSE
CAGGAAACAGCTATGACCGTGCAGGACCTTRGATACRG
>MAIN_RSVAT01_F_5686M13 T01::D03 FORWARD
TGTAAAACGACGGCCAGTAAATGCWATYACCACAATCC
>MAIN_RSVAT01_R_6384M13 T01::D03 REVERSE
CAGGAAACAGCTATGACCGCATTAACACTAAATTCYCTGG
>MAIN_RSVAT01_F_5934M13 T01::D04 FORWARD
TGTAAAACGACGGCCAGTAGAATTGCAGTTGCTCATGC
>MAIN_RSVAT01_R_6622M13 T01::D04 REVERSE

CAGGAAACAGCTATGACCAGGGGATGTRTGYAGTTTCC
>MAIN_RSVAT01_F_5974M13 T01::D05 FORWARD
TGTAAAACGACGGCCAGTARTCGRGCCAGAAGAGAACT
>MAIN_RSVAT01_R_6649M13 T01::D05 REVERSE
CAGGAAACAGCTATGACCYCCTTCYTTTGTGTTGGTYG
>MAIN_RSVAT01_F_6275M13 T01::D06 FORWARD
TGTAAAACGACGGCCAGTCCYATTGTKAACAAGCAAAGC
>MAIN_RSVAT01_R_7021M13 T01::D06 REVERSE
CAGGAAACAGCTATGACCACCTACAGACACAGTRTCMACC
>MAIN_RSVAT01_F_6452M13 T01::D07 FORWARD
TGTAAAACGACGGCCAGTTGCCTATAACAAATGATCAGAAAA
>MAIN_RSVAT01_R_7131M13 T01::D07 REVERSE
CAGGAAACAGCTATGACCGCATCAAATTCATCAGAGGG
>MAIN_RSVAT01_F_6482M13 T01::D08 FORWARD
TGTAAAACGACGGCCAGTTGTCCARCAATGTTCAAATAG
>MAIN_RSVAT01_R_7182M13 T01::D08 REVERSE
CAGGAAACAGCTATGACCCGRATAAATGCTAGRCTCTGRTT
>MAIN_RSVAT01_F_6787M13 T01::D09 FORWARD
TGTAAAACGACGGCCAGTCCAAGTGARGTAAATCTCTGC
>MAIN_RSVAT01_R_7459M13 T01::D09 REVERSE
CAGGAAACAGCTATGACCTCCAATGAYAGATGGGTTRTC
>MAIN_RSVAT01_F_6810M13 T01::D10 FORWARD
TGTAAAACGACGGCCAGTYRTTGACATATTCAACCCCAA
>MAIN_RSVAT01_R_7509M13 T01::D10 REVERSE
CAGGAAACAGCTATGACCTGAGATGGTTTATRKATGARAGTT
>MAIN_RSVAT01_F_7049M13 T01::D11 FORWARD
TGTAAAACGACGGCCAGTGAAGGYAAAAGTCTCTATGTAAAAGG
>MAIN_RSVAT01_R_7663M13 T01::D11 REVERSE
CAGGAAACAGCTATGACCTTMCCATTYAAGCAATGACC
>MAIN_RSVAT01_F_7283M13 T01::D12 FORWARD
TGTAAAACGACGGCCAGTTTAATTGCWGTTGGACTGCT
>MAIN_RSVAT01_R_7968M13 T01::D12 REVERSE
CAGGAAACAGCTATGACCTRGCTCTTCATKGTCCCTYA
>MAIN_RSVAT01_F_7304M13 T01::E01 FORWARD
TGTAAAACGACGGCCAGTYTATACTGYAAGGCCAGAAGC
>MAIN_RSVAT01_R_7934M13 T01::E01 REVERSE
CAGGAAACAGCTATGACCTCRCTRTTGAGTTCAGTGAGG
>MAIN_RSVAT01_F_7340M13 T01::E02 FORWARD
TGTAAAACGACGGCCAGTGYAARGATCAACTRAGTGGTATAAA
>MAIN_RSVAT01_R_8033M13 T01::E02 REVERSE
CAGGAAACAGCTATGACCTTGTTYTTCCTGTTGCTTTCA
>MAIN_RSVAT01_F_7796M13 T01::E03 FORWARD
TGTAAAACGACGGCCAGTCTGCAGAGWTGGACAGAAC
>MAIN_RSVAT01_R_8495M13 T01::E03 REVERSE
CAGGAAACAGCTATGACCCCCAYARCTTGAAYTGTTTGA
>MAIN_RSVAT01_F_7882M13 T01::E04 FORWARD
TGTAAAACGACGGCCAGTAAMAATCAGCATGTGTTGCC
>MAIN_RSVAT01_R_8537M13 T01::E04 REVERSE
CAGGAAACAGCTATGACCAAACATTAGCAGAATTTCCAYT
>MAIN_RSVAT01_F_7947M13 T01::E05 FORWARD
TGTAAAACGACGGCCAGTRCTRAGGGACMATGAAGAGC
>MAIN_RSVAT01_R_8639M13 T01::E05 REVERSE
CAGGAAACAGCTATGACCTGGTATAATCATTTTGGAGATAAGG
>MAIN_RSVAT01_F_8057M13 T01::E06 FORWARD

TGTAAAACGACGGCCAGTAAAGATTGCCAGCAGACG
>MAIN_RSVAT01_R_8754M13 T01::E06 REVERSE
CAGGAAACAGCTATGACCGTAGGTTCTTCTAWTTTTATTTCACC
>MAIN_RSVAT01_F_8168M13 T01::E07 FORWARD
TGTAAAACGACGGCCAGTTGCCAAAAATAATGATACTACCTGA
>MAIN_RSVAT01_R_8868M13 T01::E07 REVERSE
CAGGAAACAGCTATGACCTTGACATCACTDATTTCTATAGCTC
>MAIN_RSVAT01_F_8605M13 T01::E08 FORWARD
TGTAAAACGACGGCCAGTTTCAATGGTCCTTATCTCCAA
>MAIN_RSVAT01_R_9295M13 T01::E08 REVERSE
CAGGAAACAGCTATGACCRGTTGTCACAGTAATTCTTTTGAGT
>MAIN_RSVAT01_F_8650M13 T01::E09 FORWARD
TGTAAAACGACGGCCAGTAGACAAAATCCATTAATAGAACAYA
>MAIN_RSVAT01_R_9337M13 T01::E09 REVERSE
CAGGAAACAGCTATGACCTCTACTAAGGCTAATATCTTTCCA
>MAIN_RSVAT01_F_8726M13 T01::E10 FORWARD
TGTAAAACGACGGCCAGTAAAGGTGAAATAAAAWTAGAAGAACC
>MAIN_RSVAT01_R_9383M13 T01::E10 REVERSE
CAGGAAACAGCTATGACCATGTGTTCAARCAGTTACTAATCC
>MAIN_RSVAT01_F_9094M13 T01::E11 FORWARD
TGTAAAACGACGGCCAGTCATCCTCCATCATGGTTAATACA
>MAIN_RSVAT01_R_9796M13 T01::E11 REVERSE
CAGGAAACAGCTATGACCGGTGTCCAAATATTCTGAAC
>MAIN_RSVAT01_F_9153M13 T01::E12 FORWARD
TGTAAAACGACGGCCAGTAACACAGTATCGATCWARTGAGG
>MAIN_RSVAT01_R_9843M13 T01::E12 REVERSE
CAGGAAACAGCTATGACCGCAATTAAYTTTAACAGCATCCA
>MAIN_RSVAT01_F_9307M13 T01::F01 FORWARD
TGTAAAACGACGGCCAGTCTTGACATGGAAAGATATTAGCC
>MAIN_RSVAT01_R_9980M13 T01::F01 REVERSE
CAGGAAACAGCTATGACCCATCTTAAGGGTAAAACAATRGC
>MAIN_RSVAT01_F_9587M13 T01::F02 FORWARD
TGTAAAACGACGGCCAGTCAACATCACAGATGCTGCTAA
>MAIN_RSVAT01_R_10284M13 T01::F02 REVERSE
CAGGAAACAGCTATGACCTTCATTGAATTTGTTATCTCTTAAA
>MAIN_RSVAT01_F_9609M13 T01::F03 FORWARD
TGTAAAACGACGGCCAGTAAAGCTCAGAAAAATCTGCT
>MAIN_RSVAT01_R_10250M13 T01::F03 REVERSE
CAGGAAACAGCTATGACCAATACYCTTCTTGATTTRTCRCT
>MAIN_RSVAT01_F_9925M13 T01::F04 FORWARD
TGTAAAACGACGGCCAGTAAYAATTACAACAGATGGCCTAC
>MAIN_RSVAT01_R_10627M13 T01::F04 REVERSE
CAGGAAACAGCTATGACCGAAATGCTTGATTGAATTTGC
>MAIN_RSVAT01_F_10069M13 T01::F05 FORWARD
TGTAAAACGACGGCCAGTGWGARTTTCGGTTGCCTAA
>MAIN_RSVAT01_R_10753M13 T01::F05 REVERSE
CAGGAAACAGCTATGACCGGGGTGCATGCCTATATGT
>MAIN_RSVAT01_F_10119M13 T01::F06 FORWARD
TGTAAAACGACGGCCAGTAARGCTATATCACCTCCTAAAAA
>MAIN_RSVAT01_R_10806M13 T01::F06 REVERSE
CAGGAAACAGCTATGACCTCCACTYTGYTCATCTACATTGTT
>MAIN_RSVAT01_F_10467M13 T01::F07 FORWARD
TGTAAAACGACGGCCAGTYCCTGAAAGTCTTACAAGATATGG
>MAIN_RSVAT01_R_11073M13 T01::F07 REVERSE

CAGGAAACAGCTATGACCCTTTTAATTTGTGRCCTATKCC
>MAIN_RSVAT01_F_10603M13 T01::F08 FORWARD
TGTAAAACGACGGCCAGTTCAGCAAATTCAATCAAGCA
>MAIN_RSVAT01_R_11281M13 T01::F08 REVERSE
CAGGAAACAGCTATGACCAAACTGCATAATAGACTTTCACCTC
>MAIN_RSVAT01_F_10652M13 T01::F09 FORWARD
TGTAAAACGACGGCCAGTYGATGTRCTGGATGAAYTGC
>MAIN_RSVAT01_R_11344M13 T01::F09 REVERSE
CAGGAAACAGCTATGACCCATAATGCATGATTTTTWARTTG
>MAIN_RSVAT01_F_10947M13 T01::F10 FORWARD
TGTAAAACGACGGCCAGTTGATATAAGYAAACCARTCAGRCTC
>MAIN_RSVAT01_R_11643M13 T01::F10 REVERSE
CAGGAAACAGCTATGACCAAACGTGATTATGCATGTTAAGAA
>MAIN_RSVAT01_F_10974M13 T01::F11 FORWARD
TGTAAAACGACGGCCAGTGGAAGGTCAAWCTCATGCTC
>MAIN_RSVAT01_R_11664M13 T01::F11 REVERSE
CAGGAAACAGCTATGACCTTCAGCATTRGGGTTTTTGT
>MAIN_RSVAT01_F_11144M13 T01::F12 FORWARD
TGTAAAACGACGGCCAGTCCCAGCTAGTATAAAGAAAGTCC
>MAIN_RSVAT01_R_11792M13 T01::F12 REVERSE
CAGGAAACAGCTATGACCCTTTTGGAGAATATTTTGTTTGG
>MAIN_RSVAT01_F_11400M13 T01::G01 FORWARD
TGTAAAACGACGGCCAGTTTTAATCTTGATAAYATTGATACAGC
>MAIN_RSVAT01_R_12060M13 T01::G01 REVERSE
CAGGAAACAGCTATGACCGTTACAATCYAATGGWARTATCC
>MAIN_RSVAT01_F_11445M13 T01::G02 FORWARD
TGTAAAACGACGGCCAGTTRCCCATGTTATTTGGTGGT
>MAIN_RSVAT01_R_12147M13 T01::G02 REVERSE
CAGGAAACAGCTATGACCGTTAGRGAAAGATCTTGGTCTTTATCC
>MAIN_RSVAT01_F_11467M13 T01::G03 FORWARD
TGTAAAACGACGGCCAGTTGATCCCAACTTGTTATATCG
>MAIN_RSVAT01_R_12094M13 T01::G03 REVERSE
CAGGAAACAGCTATGACCGGTTTTCCATACTCAATATYTCTC
>MAIN_RSVAT01_F_11767M13 T01::G04 FORWARD
TGTAAAACGACGGCCAGTGCTCCAAACAAAATATTCTCC
>MAIN_RSVAT01_R_12462M13 T01::G04 REVERSE
CAGGAAACAGCTATGACCTAACCCAAGRRTTCCTATGC
>MAIN_RSVAT01_F_11928M13 T01::G05 FORWARD
TGTAAAACGACGGCCAGTTCMGGTACAAAATCTATAACTAACA
>MAIN_RSVAT01_R_12615M13 T01::G05 REVERSE
CAGGAAACAGCTATGACCGCGATTAATAGGGCTAGTRTCAA
>MAIN_RSVAT01_F_11989M13 T01::G06 FORWARD
TGTAAAACGACGGCCAGTTTGATAGAGCCRYTGAGATG
>MAIN_RSVAT01_R_12689M13 T01::G06 REVERSE
CAGGAAACAGCTATGACCAARCTAAGGCCAAAGCTTATACA
>MAIN_RSVAT01_F_12317M13 T01::G07 FORWARD
TGTAAAACGACGGCCAGTRACAATGCCAGTTTATAATAGACAA
>MAIN_RSVAT01_R_13010M13 T01::G07 REVERSE
CAGGAAACAGCTATGACCAGRATCCAATGTCCAGCYAA
>MAIN_RSVAT01_F_12372M13 T01::G08 FORWARD
TGTAAAACGACGGCCAGTGATCTRTTAGCAAAATTGGATTG
>MAIN_RSVAT01_R_13066M13 T01::G08 REVERSE
CAGGAAACAGCTATGACCCCTCTCCCCAATCTKTTTCA
>MAIN_RSVAT01_F_12549M13 T01::G09 FORWARD

TGTAAAACGACGGCCAGTGAATTCCCTGCATCWATACC
>MAIN_RSVAT01_R_13186M13 T01::G09 REVERSE
CAGGAAACAGCTATGACCAAGTRTTCATATCACACTCYAGC
>MAIN_RSVAT01_F_12857M13 T01::G10 FORWARD
TGTAAAACGACGGCCAGTAAGTTTGACTCAATATGTGGAA
>MAIN_RSVAT01_R_13508M13 T01::G10 REVERSE
CAGGAAACAGCTATGACCAATTCATCATTGAAYTTGTGTTT
>MAIN_RSVAT01_F_12898M13 T01::G11 FORWARD
TGTAAAACGACGGCCAGTCACTCAAATCTGGATCTMATGTTA
>MAIN_RSVAT01_R_13595M13 T01::G11 REVERSE
CAGGAAACAGCTATGACCTCAGAATTAGCAATCCTTATATGTT
>MAIN_RSVAT01_F_13047M13 T01::G12 FORWARD
TGTAAAACGACGGCCAGTTGAAAMAGATTGGGGAGAGG
>MAIN_RSVAT01_R_13648M13 T01::G12 REVERSE
CAGGAAACAGCTATGACCTYTCTAGGGTTTCTGGTGTAGG
>MAIN_RSVAT01_F_13274M13 T01::H01 FORWARD
TGTAAAACGACGGCCAGTCATTCTTAGCCARGATGCAA
>MAIN_RSVAT01_R_13953M13 T01::H01 REVERSE
CAGGAAACAGCTATGACCCCAAGGAAGCATGCARTARA
>MAIN_RSVAT01_F_13299M13 T01::H02 FORWARD
TGTAAAACGACGGCCAGTCATAGAGTAAAAGGATGTCATAGC
>MAIN_RSVAT01_R_13994M13 T01::H02 REVERSE
CAGGAAACAGCTATGACCCCTGTAGAACTAAATACAAAATTGA
>MAIN_RSVAT01_F_13433M13 T01::H03 FORWARD
TGTAAAACGACGGCCAGTTWGATCTTGTTAGAATGRGATTG
>MAIN_RSVAT01_R_14086M13 T01::H03 REVERSE
CAGGAAACAGCTATGACCATAAATTCCCTGCTCCTTCAC
>MAIN_RSVAT01_F_13657M13 T01::H04 FORWARD
TGTAAAACGACGGCCAGTCCAATCCRRTTAAAAGTAATGAYA
>MAIN_RSVAT01_R_14356M13 T01::H04 REVERSE
CAGGAAACAGCTATGACCTTATAATYTTRCTCCAGTTGACTRT
>MAIN_RSVAT01_F_13751M13 T01::H05 FORWARD
TGTAAAACGACGGCCAGTGAAGCTTATTAAATCGTCTRCAATG
>MAIN_RSVAT01_R_14444M13 T01::H05 REVERSE
CAGGAAACAGCTATGACCTCATCTTGAGCATGATATTTTAC
>MAIN_RSVAT01_F_13765M13 T01::H06 FORWARD
TGTAAAACGACGGCCAGTTCGTCTRCAATGATTAGAACC
>MAIN_RSVAT01_R_14385M13 T01::H06 REVERSE
CAGGAAACAGCTATGACCGCAYTTTCTTACATGCTTGC
>MAIN_RSVAT01_F_14182M13 T01::H07 FORWARD
TGTAAAACGACGGCCAGTGGYTGTACAATGGRCATATCAA
>MAIN_RSVAT01_R_14849M13 T01::H07 REVERSE
CAGGAAACAGCTATGACCAAACATGRTTGAACCAYTTTAAGA
>MAIN_RSVAT01_F_14248M13 T01::H08 FORWARD
TGTAAAACGACGGCCAGTCCAACAACATTCATTGGTCT
>MAIN_RSVAT01_R_14948M13 T01::H08 REVERSE
CAGGAAACAGCTATGACCCATTRGTTGTCAAGCTGTTT
>MAIN_RSVAT01_F_14359M13 T01::H09 FORWARD
TGTAAAACGACGGCCAGTGARTGGAGCAAGCATGTAAG
>MAIN_RSVAT01_R_15056M13 T01::H09 REVERSE
CAGGAAACAGCTATGACCCARTGTTAGTGTRTAGCTATGGGAA
>MAIN_RSVAT01_F_14517M13 T01::H10 FORWARD
TGTAAAACGACGGCCAGTGARGTTTACTTAGTCCTTACAATAGG
>MAIN_RSVAT01_R_15221M13 T01::H10 REVERSE

CAGGAAACAGCTATGACCTGTCAAAAACTAATATCTCGTAGTTTAGT
>MAIN_RSVAT01_F_14614M13 T01::H11 FORWARD
TGTAAAACGACGGCCAGTTTTCATCATGCCTAARAAAGC
>MAIN_RSVAT01_R_15222M13 T01::H11 REVERSE
CAGGAAACAGCTATGACCGTGTCAAAAACTAATATCTCGTAGTTTAGT
>MAIN_RSVAT01_F_14617M13 T01::H12 FORWARD
TGTAAAACGACGGCCAGTCATCATGCCTAARAAAGCTG
>MAIN_RSVAT01_R_15222M13_v2 T01::H12 REVERSE
CAGGAAACAGCTATGACCGTGTCAAAAACTAATATCTCGTAGTTTAG

S2 Text. Primer Pairs Used for Sanger Sequencing of RSVB.

>MAIN_RSVBT02_F_1M13 T02::A01 FORWARD
TGTAAAACGACGGCCAGTACGCGAAAAAATGCGTACTA
>MAIN_RSVBT02_R_691M13 T02::A01 REVERSE
CAGGAAACAGCTATGACCACAGRGGTCTCATGTCYGTG
>MAIN_RSVBT02_F_1M13_v2 T02::A02 FORWARD
TGTAAAACGACGGCCAGTACGCGAAAAAATGCGTACTAC
>MAIN_RSVBT02_R_676M13 T02::A02 REVERSE
CAGGAAACAGCTATGACCCYGTGATCATYAATCTTTTGC
>MAIN_RSVBT02_F_1M13_v3 T02::A03 FORWARD
TGTAAAACGACGGCCAGTACGCGAAAAAATGCGTACTACA
>MAIN_RSVBT02_R_678M13 T02::A03 REVERSE
CAGGAAACAGCTATGACCGTCYGTGATCATYAATCTTTTGC
>MAIN_RSVAT01_F_33M13 T02::A04 FORWARD
TGTAAAACGACGGCCAGTAAAAATGGGGCAAATAAGAA
>MAIN_RSVBT02_R_737M13 T02::A04 REVERSE
CAGGAAACAGCTATGACCTGTGTGATGATTTCYTTGGTG
>MAIN_RSVBT02_F_497M13 T02::A05 FORWARD
TGTAAAACGACGGCCAGTTGGGCTTGATCTCMATTCAT
>MAIN_RSVBT02_R_1197M13 T02::A05 REVERSE
CAGGAAACAGCTATGACCGACAGCAGCTGATCCTTATTT
>MAIN_RSVBT02_F_576M13 T02::A06 FORWARD
TGTAAAACGACGGCCAGTAMCTCATCAAAGGGAAATGG
>MAIN_RSVBT02_R_1265M13 T02::A06 REVERSE
CAGGAAACAGCTATGACCTTTGCACATCATAATTGGGAG
>MAIN_RSVBT02_F_679M13 T02::A07 FORWARD
TGTAAAACGACGGCCAGTATGAGACCYCTGTCGATGG
>MAIN_RSVBT02_R_1369M13 T02::A07 REVERSE
CAGGAAACAGCTATGACCCCTTCCTAACCTGGACATAGC
>MAIN_RSVBT02_F_1014M13 T02::A08 FORWARD
TGTAAAACGACGGCCAGTAAAACCAACYYAWCCAAACC
>MAIN_RSVBT02_R_1710M13 T02::A08 REVERSE
CAGGAAACAGCTATGACCAAGACATTGTTTGCCCTCCT
>MAIN_RSVBT02_F_1110M13 T02::A09 FORWARD
TGTAAAACGACGGCCAGTGCCARTAACATAAATTGGGG
>MAIN_RSVBT02_R_1768M13 T02::A09 REVERSE
CAGGAAACAGCTATGACCAAACTGTTRGCTATRTCCTTTGG
>MAIN_RSVBT02_F_1242M13 T02::A10 FORWARD
TGTAAAACGACGGCCAGTACACTCCCAATTATGATGTGC
>MAIN_RSVBT02_R_1914M13 T02::A10 REVERSE
CAGGAAACAGCTATGACCGCAWTAYTTGYCCTGAACCA
>MAIN_RSVBT02_F_1308M13 T02::A11 FORWARD
TGTAAAACGACGGCCAGTTGCAAATCATAAATTCACAGGA
>MAIN_RSVBT02_R_1986M13 T02::A11 REVERSE
CAGGAAACAGCTATGACCCCATTTMWGCYTGGACACTA
>MAIN_RSVBT02_F_1621M13 T02::A12 FORWARD
TGTAAAACGACGGCCAGTACTGTGTATAGCWGCMCTTG
>MAIN_RSVBT02_R_2307M13 T02::A12 REVERSE
CAGGAAACAGCTATGACCCRTCATCTTCTTTRGGGTTG
>MAIN_RSVBT02_F_1732M13 T02::B01 FORWARD
TGTAAAACGACGGCCAGTCAAGGGYCTMATACCAAAGG
>MAIN_RSVBT02_R_2431M13 T02::B01 REVERSE

CAGGAAACAGCTATGACCTTGCCYTTTATTGATTCTTAGG
>MAIN_RSVBT02_F_1832M13 T02::B02 FORWARD
TGTAAAACGACGGCCAGTTCWTCMACAAGAGGGGGTAG
>MAIN_RSVBT02_R_2523M13 T02::B02 REVERSE
CAGGAAACAGCTATGACCTGTTATCGGGCTCTCTTTRG
>MAIN_RSVBT02_F_1970M13 T02::B03 FORWARD
TGTAAAACGACGGCCAGTTGTCCARGCWKAAATGGA
>MAIN_RSVBT02_R_2635M13 T02::B03 REVERSE
CAGGAAACAGCTATGACCCACTTGGGGTGAGATCTTCT
>MAIN_RSVBT02_F_2253M13 T02::B04 FORWARD
TGTAAAACGACGGCCAGTCARCAGAAGAATTGGAAGCC
>MAIN_RSVBT02_R_2951M13 T02::B04 REVERSE
CAGGAAACAGCTATGACCTTCCTAAGTCTTGCCATAGCC
>MAIN_RSVBT02_F_2347M13 T02::B05 FORWARD
TGTAAAACGACGGCCAGTAWCATKGAGAAGTTTGCMCC
>MAIN_RSVBT02_R_3043M13 T02::B05 REVERSE
CAGGAAACAGCTATGACCCGTTGTCTTCCAACAAGTCA
>MAIN_RSVBT02_F_2517M13 T02::B06 FORWARD
TGTAAAACGACGGCCAGTGATAACATCTGGCACCAACA
>MAIN_RSVBT02_R_3214M13 T02::B06 REVERSE
CAGGAAACAGCTATGACCGCTGATKGATCARTTGGTTG
>MAIN_RSVBT02_F_2672M13 T02::B07 FORWARD
TGTAAAACGACGGCCAGTAACATTTGATAACMATGAAGAAGAA
>MAIN_RSVBT02_R_3372M13 T02::B07 REVERSE
CAGGAAACAGCTATGACCCCATATTGTTAGTGATGCRGG
>MAIN_RSVBT02_F_2707M13 T02::B08 FORWARD
TGTAAAACGACGGCCAGTCATATGARGAGATMAATGATCAAA
>MAIN_RSVBT02_R_3399M13 T02::B08 REVERSE
CAGGAAACAGCTATGACCTGGYACAGATGACTGGAACA
>MAIN_RSVBT02_F_3079M13 T02::B09 FORWARD
TGTAAAACGACGGCCAGTTCAGTGATCAACTCACTCAGC
>MAIN_RSVBT02_R_3871M13 T02::B09 REVERSE
CAGGAAACAGCTATGACCTACTAATCCWGCATAGGGAA
>MAIN_RSVBT02_F_3247M13 T02::B10 FORWARD
TGTAAAACGACGGCCAGTAAGAACAAGATGGGGCAAAT
>MAIN_RSVBT02_R_3934M13 T02::B10 REVERSE
CAGGAAACAGCTATGACCAAATTGACTCTGTGGCTTGA
>MAIN_RSVBT02_F_3257M13 T02::B11 FORWARD
TGTAAAACGACGGCCAGTATGGGGCAAATATGGAAAC
>MAIN_RSVBT02_R_3956M13 T02::B11 REVERSE
CAGGAAACAGCTATGACCGGTAGGCACCAAGATCYACT
>MAIN_RSVBT02_F_3670M13 T02::B12 FORWARD
TGTAAAACGACGGCCAGTAAGACATTCAACCCCACTCA
>MAIN_RSVBT02_R_4359M13 T02::B12 REVERSE
CAGGAAACAGCTATGACCATAGGGCCAAAATTTGCTTG
>MAIN_RSVBT02_F_3672M13 T02::C01 FORWARD
TGTAAAACGACGGCCAGTAGACATTCAACCCCACTCAT
>MAIN_RSVBT02_R_4360M13 T02::C01 REVERSE
CAGGAAACAGCTATGACCAATAGGGCCAAAATTTGCTT
>MAIN_RSVBT02_F_3771M13 T02::C02 FORWARD
TGTAAAACGACGGCCAGTTGTCAAAAACAAGGAYCTGAA
>MAIN_RSVBT02_R_4460M13 T02::C02 REVERSE
CAGGAAACAGCTATGACCCAGAATGTTTTATGTTCACTTAGC
>MAIN_RSVBT02_F_4059M13 T02::C03 FORWARD

TGTAAAACGACGGCCAGTTGACAGGTYCACATATATCCTCA
>MAIN_RSVBT02_R_4758M13 T02::C03 REVERSE
CAGGAAACAGCTATGACCTTRAGAGTATCCCAGGTCTTTTC
>MAIN_RSVBT02_F_4096M13 T02::C04 FORWARD
TGTAAAACGACGGCCAGTTCYAAACATCATRAACATCTACA
>MAIN_RSVBT02_R_4725M13 T02::C04 REVERSE
CAGGAAACAGCTATGACCGCAGTGCGTTGATTCTTGTK
>MAIN_RSVBT02_F_4308M13 T02::C05 FORWARD
TGTAAAACGACGGCCAGTAATGGGAAAYACATCCATCA
>MAIN_RSVBT02_R_4995M13 T02::C05 REVERSE
CAGGAAACAGCTATGACCCTAMCCCTTTCTGGTGRGA
>MAIN_RSVBT02_F_4568M13 T02::C06 FORWARD
TGTAAAACGACGGCCAGTCTTCTCACAGAGTCATGGTRTC
>MAIN_RSVBT02_R_5244M13 T02::C06 REVERSE
CAGGAAACAGCTATGACCTTGCARAGTYGATTGTTGCC
>MAIN_RSVBT02_F_4591M13 T02::C07 FORWARD
TGTAAAACGACGGCCAGTAAAACCAYGCCAACYATCAT
>MAIN_RSVBT02_R_5289M13 T02::C07 REVERSE
CAGGAAACAGCTATGACCGGTTTCTTCTTTKGTTTRTTGC
>MAIN_RSVBT02_F_4878M13 T02::C08 FORWARD
TGTAAAACGACGGCCAGTTCATCATCTCTGCCAATCAC
>MAIN_RSVBT02_R_5551M13 T02::C08 REVERSE
CAGGAAACAGCTATGACCGATTGCTGTTGGATTGTGTG
>MAIN_RSVBT02_F_4972M13 T02::C09 FORWARD
TGTAAAACGACGGCCAGTTCMAGTCYCACCAGAAAGG
>MAIN_RSVBT02_R_5665M13 T02::C09 REVERSE
CAGGAAACAGCTATGACCTTRAATAACTAAGCATGTGAYTGG
>MAIN_RSVBT02_F_5087M13 T02::C10 FORWARD
TGTAAAACGACGGCCAGTYACAWACCMAAGGCAGAACC
>MAIN_RSVBT02_R_5748M13 T02::C10 REVERSE
CAGGAAACAGCTATGACCGATCAGCARCTCCATGGTTA
>MAIN_RSVBT02_F_5134M13 T02::C11 FORWARD
TGTAAAACGACGGCCAGTAGCAYAAAACCACGYCYAA
>MAIN_RSVBT02_R_5810M13 T02::C11 REVERSE
CAGGAAACAGCTATGACCGACTTGAGGTGAGGTACAATGC
>MAIN_RSVBT02_F_5457M13 T02::C12 FORWARD
TGTAAAACGACGGCCAGTCAACCACATYARAACACACAGAA
>MAIN_RSVBT02_R_6157M13 T02::C12 REVERSE
CAGGAAACAGCTATGACCAAGAAGCCCAGAAATCTTCG
>MAIN_RSVBT02_F_5492M13 T02::D01 FORWARD
TGTAAAACGACGGCCAGTACCTTACAATCCATTGCGCT
>MAIN_RSVBT02_R_6187M13 T02::D01 REVERSE
CAGGAAACAGCTATGACCCCACTYGCTATTGCAGATCC
>MAIN_RSVBT02_F_5724M13 T02::D02 FORWARD
TGTAAAACGACGGCCAGTGCAAATAACCATGGAGYTGC
>MAIN_RSVBT02_R_6392M13 T02::D02 REVERSE
CAGGAAACAGCTATGACCTGTTTCAATGTTGGARATGC
>MAIN_RSVBT02_F_5932M13 T02::D03 FORWARD
TGTAAAACGACGGCCAGTAACCAAATGCAATGGAACTG
>MAIN_RSVBT02_R_6630M13 T02::D03 REVERSE
CAGGAAACAGCTATGACCATGCAAGGACTTCTTCCTTT
>MAIN_RSVBT02_F_6030M13 T02::D04 FORWARD
TGTAAAACGACGGCCAGTACACACCAGCTGYCAACAAC
>MAIN_RSVBT02_R_6713M13 T02::D04 REVERSE

CAGGAAACAGCTATGACCGATGTTGGTGGTGCAYAGAG
>MAIN_RSVBT02_F_6210M13 T02::D05 FORWARD
TGTAAAACGACGGCCAGTCACCTTGAAGGAGAAGTGAAC
>MAIN_RSVBT02_R_6878M13 T02::D05 REVERSE
CAGGAAACAGCTATGACCGGCTGACTTCACTTGGTAATG
>MAIN_RSVBT02_F_6362M13 T02::D06 FORWARD
TGTAAAACGACGGCCAGTACAGAGYTGTCGCATYTCC
>MAIN_RSVBT02_R_7031M13 T02::D06 REVERSE
CAGGAAACAGCTATGACCAATYCCACGATTTTTGTTGG
>MAIN_RSVBT02_F_6451M13 T02::D07 FORWARD
TGTAAAACGACGGCCAGTTCAATGCAGGTGTAACRACAC
>MAIN_RSVBT02_R_7136M13 T02::D07 REVERSE
CAGGAAACAGCTATGACCGTTYTTGCCTTCCAGCTTGT
>MAIN_RSVBT02_F_6560M13 T02::D08 FORWARD
TGTAAAACGACGGCCAGTAGCAMYGTTCARATAGTAAGGCA
>MAIN_RSVBT02_R_7208M13 T02::D08 REVERSE
CAGGAAACAGCTATGACCTGCATCAAACTCATCAGAAGG
>MAIN_RSVBT02_F_6862M13 T02::D09 FORWARD
TGTAAAACGACGGCCAGTCCAAGTGAAGTCAGCCTTTG
>MAIN_RSVBT02_R_7558M13 T02::D09 REVERSE
CAGGAAACAGCTATGACCCAGCCTGTRCTGTGATGTTG
>MAIN_RSVBT02_F_6883M13 T02::D10 FORWARD
TGTAAAACGACGGCCAGTAACACTGACATATTCAATTCCA
>MAIN_RSVBT02_R_7583M13 T02::D10 REVERSE
CAGGAAACAGCTATGACCGGTAGCATGATGYGARGAAA
>MAIN_RSVBT02_F_7109M13 T02::D11 FORWARD
TGTAAAACGACGGCCAGTCTATGTMAACAAGCTGGAAGG
>MAIN_RSVBT02_R_7803M13 T02::D11 REVERSE
CAGGAAACAGCTATGACCCTCACTAGYAAKGCATGAGG
>MAIN_RSVBT02_F_7387M13 T02::D12 FORWARD
TGTAAAACGACGGCCAGTCAAAGCCAAAAACACACCAG
>MAIN_RSVBT02_R_8058M13 T02::D12 REVERSE
CAGGAAACAGCTATGACCTGAATTGGGTTCTTCATTRTCTC
>MAIN_RSVBT02_F_7416M13 T02::E01 FORWARD
TGTAAAACGACGGCCAGTGCAAAGACCAACTAAGTGGAA
>MAIN_RSVBT02_R_8113M13 T02::E01 REVERSE
CAGGAAACAGCTATGACCTTTTTCTATTGCTCTCRATGTATGA
>MAIN_RSVBT02_F_7601M13 T02::E02 FORWARD
TGTAAAACGACGGCCAGTTCCTTAACTCATAGTTACATAAAAACC
>MAIN_RSVBT02_R_8280M13 T02::E02 REVERSE
CAGGAAACAGCTATGACCTTATCCGGTAATATCATTATTTTTGG
>MAIN_RSVBT02_F_7906M13 T02::E03 FORWARD
TGTAAAACGACGGCCAGTGCTCTTGGTATAGTTGGAGTGC
>MAIN_RSVBT02_R_8593M13 T02::E03 REVERSE
CAGGAAACAGCTATGACCGGGATCCATTTTGTCCCATA
>MAIN_RSVBT02_F_7954M13 T02::E04 FORWARD
TGTAAAACGACGGCCAGTMCAACATAACAAAACAATCAGCA
>MAIN_RSVBT02_R_8632M13 T02::E04 REVERSE
CAGGAAACAGCTATGACCCTCTGCTAATGTRTATCTAACTGAT
>MAIN_RSVBT02_F_8247M13 T02::E05 FORWARD
TGTAAAACGACGGCCAGTTGACCAAACCAAAAATAATGATATTACC
>MAIN_RSVBT02_R_8940M13 T02::E05 REVERSE
CAGGAAACAGCTATGACCTTTCTATAGCTCTTCGTATCYMY
>MAIN_RSVBT02_F_8271M13 T02::E06 FORWARD

TGTAAAACGACGGCCAGTACCGGATAAATATCCTTGTAGT
>MAIN_RSVBT02_R_8964M13 T02::E06 REVERSE
CAGGAAACAGCTATGACCTGGCGTACACCTTTACATCAC
>MAIN_RSVBT02_F_8527M13 T02::E07 FORWARD
TGTAAAACGACGGCCAGTYCATAAYGATTCTATATYATCCAACC
>MAIN_RSVBT02_R_9193M13 T02::E07 REVERSE
CAGGAAACAGCTATGACCCATGATGGAGGATGTTGC
>MAIN_RSVBT02_F_8734M13 T02::E08 FORWARD
TGTAAAACGACGGCCAGTAGACAAAGYCCACTACTAGAGCA
>MAIN_RSVBT02_R_9419M13 T02::E08 REVERSE
CAGGAAACAGCTATGACCGCTAAGGCTGATRTCTTTCCA
>MAIN_RSVBT02_F_8805M13 T02::E09 FORWARD
TGTAAAACGACGGCCAGTYCATAAAGGTGAACTGAAATTAGAA
>MAIN_RSVBT02_R_9500M13 T02::E09 REVERSE
CAGGAAACAGCTATGACCGAATCCRCATCTCARCCCTA
>MAIN_RSVBT02_F_9005M13 T02::E10 FORWARD
TGTAAAACGACGGCCAGTGCCCAACAAYAATTCAGG
>MAIN_RSVBT02_R_9686M13 T02::E10 REVERSE
CAGGAAACAGCTATGACCTGCATCTGTGATGTTATTTAGC
>MAIN_RSVBT02_F_9248M13 T02::E11 FORWARD
TGTAAAACGACGGCCAGTCGATCAAATGAGGTAAAAAGTCA
>MAIN_RSVBT02_R_9920M13 T02::E11 REVERSE
CAGGAAACAGCTATGACCAATTCTTACAGCATCCATTGC
>MAIN_RSVBT02_F_9325M13 T02::E12 FORWARD
TGTAAAACGACGGCCAGTTCAATATGGTTGYATYGTTTATCA
>MAIN_RSVBT02_R_10020M13 T02::E12 REVERSE
CAGGAAACAGCTATGACCTGTTGTAGGTATTTACAAACCCTTT
>MAIN_RSVBT02_F_9510M13 T02::F01 FORWARD
TGTAAAACGACGGCCAGTGTGYTATCACAATTATTYCTTTATGG
>MAIN_RSVBT02_R_10223M13 T02::F01 REVERSE
CAGGAAACAGCTATGACCTTTTGGAGGWGAAATRGC
>MAIN_RSVBT02_F_9626M13 T02::F02 FORWARD
TGTAAAACGACGGCCAGTGAAGAAGATCAATTTAGGAAWCG
>MAIN_RSVBT02_R_10304M13 T02::F02 REVERSE
CAGGAAACAGCTATGACCGAACTTCAACTTTTCATGTTCTAT
>MAIN_RSVBT02_F_9661M13 T02::F03 FORWARD
TGTAAAACGACGGCCAGTCATGCTAAATAACATCACAGATGC
>MAIN_RSVBT02_R_10328M13 T02::F03 REVERSE
CAGGAAACAGCTATGACCTCTTCTTGATCTGTCRCTTTCAG
>MAIN_RSVBT02_F_10128M13 T02::F04 FORWARD
TGTAAAACGACGGCCAGTTTTTATCAGGATTGCGGTTC
>MAIN_RSVBT02_R_10805M13 T02::F04 REVERSE
CAGGAAACAGCTATGACCGACAAGAGGTATTGTTAAATGCAA
>MAIN_RSVBT02_F_10150M13 T02::F05 FORWARD
TGTAAAACGACGGCCAGTTCGTGARTTTCATCTGCCTAA
>MAIN_RSVBT02_R_10843M13 T02::F05 REVERSE
CAGGAAACAGCTATGACCATRAAAGGAGGTGCATGTCT
>MAIN_RSVBT02_F_10235M13 T02::F06 FORWARD
TGTAAAACGACGGCCAGTGACTAGTTTTCCTAGAAATTACATGC
>MAIN_RSVBT02_R_10929M13 T02::F06 REVERSE
CAGGAAACAGCTATGACCGACACCAGCCCTCAATACC
>MAIN_RSVBT02_F_10607M13 T02::F07 FORWARD
TGTAAAACGACGGCCAGTGCAGGAATAAGCAACAAGTCA
>MAIN_RSVBT02_R_11274M13 T02::F07 REVERSE

CAGGAAACAGCTATGACCTTGTATTTATCCATGGACCTACT
>MAIN_RSVBT02_F_10682M13 T02::F08 FORWARD
TGTAAAACGACGGCCAGTGATCTTAGCAAATTYAATCAAGCA
>MAIN_RSVBT02_R_11358M13 T02::F08 REVERSE
CAGGAAACAGCTATGACCTGCATAATAAGCTTTCYCCTC
>MAIN_RSVBT02_F_10737M13 T02::F09 FORWARD
TGTAAAACGACGGCCAGTGATGTATTAGATGAACTGCATGG
>MAIN_RSVBT02_R_11419M13 T02::F09 REVERSE
CAGGAAACAGCTATGACCGCATGATTTCGGAGTTGC
>MAIN_RSVBT02_F_10995M13 T02::F10 FORWARD
TGTAAAACGACGGCCAGTTCACAGCTCTGATAAATGGTGA
>MAIN_RSVBT02_R_11690M13 T02::F10 REVERSE
CAGGAAACAGCTATGACCTCTATCATCTGGAAGATCCTGG
>MAIN_RSVBT02_F_11060M13 T02::F11 FORWARD
TGTAAAACGACGGCCAGTGGTCAGACCCATGCTCAA
>MAIN_RSVBT02_R_11743M13 T02::F11 REVERSE
CAGGAAACAGCTATGACCTCRGCATTKGGATTTTTATCR
>MAIN_RSVBT02_F_11318M13 T02::F12 FORWARD
TGTAAAACGACGGCCAGTAACACAGGAGTTAGAATACAGAGG
>MAIN_RSVBT02_R_11971M13 T02::F12 REVERSE
CAGGAAACAGCTATGACCARACTTTCATAAACAACTCTTAATCC
>MAIN_RSVBT02_F_11529M13 T02::G01 FORWARD
TGTAAAACGACGGCCAGTCTATGCTGTTTGGTGGTGGT
>MAIN_RSVBT02_R_12221M13 T02::G01 REVERSE
CAGGAAACAGCTATGACCTGACCAAGATCTTTCTCTTACA
>MAIN_RSVBT02_F_11580M13 T02::G02 FORWARD
TGTAAAACGACGGCCAGTGGAGAACTCCAGACTTCCTTACA
>MAIN_RSVBT02_R_12267M13 T02::G02 REVERSE
CAGGAAACAGCTATGACCTTGTGAACATAATACTKGGCGA
>MAIN_RSVBT02_F_11776M13 T02::G03 FORWARD
TGTAAAACGACGGCCAGTTAGGGTCTGAAAGGCAAGC
>MAIN_RSVBT02_R_12448M13 T02::G03 REVERSE
CAGGAAACAGCTATGACCTGGTCTCTTTGCTTTTTGG
>MAIN_RSVBT02_F_12040M13 T02::G04 FORWARD
TGTAAAACGACGGCCAGTAAAAACATCAGCRATAGATACAA
>MAIN_RSVBT02_R_12732M13 T02::G04 REVERSE
CAGGAAACAGCTATGACCTGTCGATATCTTCATCWCCA
>MAIN_RSVBT02_F_12075M13 T02::G05 FORWARD
TGTAAAACGACGGCCAGTGGGCTACTGATATGATGAGGAA
>MAIN_RSVBT02_R_12770M13 T02::G05 REVERSE
CAGGAAACAGCTATGACCCARGCTAAGACCAAAACTTATGC
>MAIN_RSVBT02_F_12363M13 T02::G06 FORWARD
TGTAAAACGACGGCCAGTAGCCATGGGTAGGTTCATCT
>MAIN_RSVBT02_R_13054M13 T02::G06 REVERSE
CAGGAAACAGCTATGACCMGCATTATGAAAATAATCAGACA
>MAIN_RSVBT02_F_12429M13 T02::G07 FORWARD
TGTAAAACGACGGCCAGTCCAAAAAGCAAAGAGACCAA
>MAIN_RSVBT02_R_13102M13 T02::G07 REVERSE
CAGGAAACAGCTATGACCTTGAATAATYAGAATCCAATGTC
>MAIN_RSVBT02_F_12659M13 T02::G08 FORWARD
TGTAAAACGACGGCCAGTAACAAATTATCATTTYGATACTAGTCC
>MAIN_RSVBT02_R_13330M13 T02::G08 REVERSE
CAGGAAACAGCTATGACCAACTTTAGACATAGATTTCCAGTAGC
>MAIN_RSVBT02_F_12774M13 T02::G09 FORWARD

TGTAAAACGACGGCCAGTCRGGTTGTGGAACAATTCAC
>MAIN_RSVBT02_R_13470M13 T02::G09 REVERSE
CAGGAAACAGCTATGACCTTAACAACCCAAGGGCATAC
>MAIN_RSVBT02_F_12883M13 T02::G10 FORWARD
TGTAAAACGACGGCCAGTTTCATCAAGTTGAAGCAAGTGA
>MAIN_RSVBT02_R_13544M13 T02::G10 REVERSE
CAGGAAACAGCTATGACCCATTTATTAACCCCATTCTWAC
>MAIN_RSVBT02_F_13061M13 T02::G11 FORWARD
TGTAAAACGACGGCCAGTTTAAGTACTAATTTAGCTGGACATTGG
>MAIN_RSVBT02_R_13728M13 T02::G11 REVERSE
CAGGAAACAGCTATGACCTCTAAAGTTTCTGGGGTTGG
>MAIN_RSVBT02_F_13274M13 T02::G12 FORWARD
TGTAAAACGACGGCCAGTTTCTTTGTGTTTTGGAGTTAATAGA
>MAIN_RSVBT02_R_13974M13 T02::G12 REVERSE
CAGGAAACAGCTATGACCTGGTGRTGTAAAGTTGGYTAGA
>MAIN_RSVBT02_F_13362M13 T02::H01 FORWARD
TGTAAAACGACGGCCAGTCAATCAAGACACAAGTTTGCAT
>MAIN_RSVBT02_R_14043M13 T02::H01 REVERSE
CAGGAAACAGCTATGACCTGACATGATGCCAAGGAAG
>MAIN_RSVBT02_F_13443M13 T02::H02 FORWARD
TGTAAAACGACGGCCAGTAATTTACCGTATGCCCTTGG
>MAIN_RSVBT02_R_14141M13 T02::H02 REVERSE
CAGGAAACAGCTATGACCGAATGCTATACAACTGGGRTC
>MAIN_RSVBT02_F_13788M13 T02::H03 FORWARD
TGTAAAACGACGGCCAGTTGGAAATACYGAATCTATRATGAYG
>MAIN_RSVBT02_R_14484M13 T02::H03 REVERSE
CAGGAAACAGCTATGACCGTGCAARTACTGTTCTTCTGTAA
>MAIN_RSVBT02_F_13796M13 T02::H04 FORWARD
TGTAAAACGACGGCCAGTCYGAATCTATRATGAYGTCAAC
>MAIN_RSVBT02_R_14496M13 T02::H04 REVERSE
CAGGAAACAGCTATGACCAAATGCATCTATTTACAGAAGAACA
>MAIN_RSVBT02_F_13828M13 T02::H05 FORWARD
TGTAAAACGACGGCCAGTAAAATGCATATTAAATCTTCCACTG
>MAIN_RSVBT02_R_14521M13 T02::H05 REVERSE
CAGGAAACAGCTATGACCTCTTGAGCATGATATTTTGCRAT
>MAIN_RSVBT02_F_14110M13 T02::H06 FORWARD
TGTAAAACGACGGCCAGTTTAAGATTAARGAYCCCAGTTG
>MAIN_RSVBT02_R_14769M13 T02::H06 REVERSE
CAGGAAACAGCTATGACCGGTAACAAAGGAAAGGTATTAAGC
>MAIN_RSVBT02_F_14306M13 T02::H07 FORWARD
TGTAAAACGACGGCCAGTACCATTCCTGCTACAGATGC
>MAIN_RSVBT02_R_15002M13 T02::H07 REVERSE
CAGGAAACAGCTATGACCTTCACTYAAGTAAGGATATGTGGA
>MAIN_RSVBT02_F_14359M13 T02::H08 FORWARD
TGTAAAACGACGGCCAGTAAAATTTGCAGAACCTATTAGYA
>MAIN_RSVBT02_R_15041M13 T02::H08 REVERSE
CAGGAAACAGCTATGACCCAGYTTCTTGAGCTCATTGG
>MAIN_RSVBT02_F_14566M13 T02::H09 FORWARD
TGTAAAACGACGGCCAGTYGTGTGCCTAGGTAGCAAGTT
>MAIN_RSVBT02_R_15305M13 T02::H09 REVERSE
CAGGAAACAGCTATGACCCGAGAAAAAAAGTGTCAAAAACTAATRTC
>MAIN_RSVBT02_F_14609M13 T02::H10 FORWARD
TGTAAAACGACGGCCAGTGTCCTTACAATAGGCCCTGC
>MAIN_RSVBT02_R_15277M13 T02::H10 REVERSE

CAGGAAACAGCTATGACCTCTCGTTGTGTTGTAAATGCAC
>MAIN_RSVBT02_F_14637M13 T02::H11 FORWARD
TGTAAAACGACGGCCAGTCTTCCTGTTTTTGATGTTGTGC
>MAIN_RSVBT02_R_15305M13_v2 T02::H11 REVERSE
CAGGAAACAGCTATGACCACGAGAAAAAAAGTGTCAAAAACTAATRT
>MAIN_RSVBT02_F_14639M13 T02::H12 FORWARD
TGTAAAACGACGGCCAGTTCCTGTTTTTGATGTTGTGC
>MAIN_RSVBT02_R_15305M13_v3 T02::H12 REVERSE
CAGGAAACAGCTATGACCACGAGAAAAAAAGTGTCAAAAACTAATRTC