bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1

2

3 # A comparative genome analysis of Rift Valley Fever virus isolates from
4 foci of the disease outbreak in South Africa in 2008-2010

5

6

7

8 Moabi R. Maluleke[1,4], Maanda Phosiwa[1], Antoinette van Schalkwyk[1], George Michuki[2],
9 Baratang A. Lubisi[1], Phemelo S. Kegakilwe[3], Steve J. Kemp[2,] , Phelix A.O. Majiwa[1,4]*

10

11

12 [1] ARC-Onderstepoort Veterinary Research, Private Bag X05, Onderstepoort, 0110 Gauteng,
13 South Africa

14 [2]International Livestock Research Institute, Nairobi, 00100, Kenya

15 [3]Department of Agriculture, Land Reform and Rural Development, Veterinary Services,

16 Northern Cape Province, Private Bag X5018, Kimberley 8300, South Africa,

17 [4]Department of Veterinary Tropical Diseases, University of Pretoria, Private Bag X04,
18 Onderstepoort, 0110 Gauteng, South Africa

19

20

21 ***Corresponding author:** Phelix A.O. Majiwa, email: pmajiwa@gmail.com

22

23

bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

Rift Valley fever (RVF) is a re-emerging zoonotic disease responsible for major losses in livestock production, with negative impact on the livelihoods of both commercial and resource-poor farmers in sub-Sahara African countries. The disease remains a threat in countries where its mosquito vectors thrives. Outbreaks of RVF usually follow weather conditions which favour increase in mosquito populations. Such outbreaks are usually cyclical, occurring every 10-15 years.

Recent outbreaks of the disease in South Africa have occurred unpredictably and with increased frequency. In 2008 outbreaks were reported in Mpumalanga, Limpopo and Gauteng provinces, followed by a 2009 outbreak in KwaZulu-Natal, Mpumalanga and Northern Cape provinces and in 2010 in the Eastern Cape, Northern Cape, Western Cape, North West, Free State and Mpumalanga provinces. By August 2010, 232 confirmed infections had been reported in humans, with 26 confirmed deaths.

To investigate the evolutionary dynamics of RVF viruses (RVFVs) circulating in South Africa, we undertook complete genome sequence analysis of isolates from animals at discrete foci of the 2008-2010 outbreaks. The genome sequences of these viruses were compared with viruses from earlier outbreaks in South Africa and in other countries. The data indicates that one 2009 and all the 2008 isolates from South Africa and Madagascar (M49/08) cluster in Lineage C or Kenya-1. The remaining of the 2009 and 2010 isolates cluster within Lineage H, except isolate M259_RSA_09, a probable segment M reassortant.

bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

24

25 ## Author summary

26 A single RVF virus serotype exists, yet differences in virulence and pathogenicity of the virus
27 have been observed. This necessitates the need for detailed genetic characterization of
28 various isolates of the virus. The RVF virus isolates that caused the 2008-2010 disease
29 outbreaks in South Africa were most probably reassortants. Reassortment results from
30 exchange of portions of the genome, particularly those of segment M. Although clear
31 association between RVFV genotype and phenotype has not been established, various amino
32 acid substitutions have been implicated in the phenotype. Viruses with amino acid substitutions
33 from glycine to glutamic acid at position 277 of segment M have been shown to be more virulent
34 in mice in comparison to viruses with glycine at the same position. Phylogenetic analysis
35 indicated the viruses responsible for the 2008-2010 RVF outbreaks in South Africa were not
36 introduced from outside the country, but mutated in time and caused the outbreaks when
37 environmental conditions became favourable.

38

39 ## Introduction

40 Rift Valley fever (RVF), a mosquito-borne viral disease, affects humans and some species of
41 ruminants including sheep, cattle, goats and buffalos. The causative agent, Rift Valley Fever
42 virus (RVFV), belongs to the genus *Phlebovirus* in the family *Bunyaviridae.* The disease in
43 livestock is characterized by abortion storms and high mortality of young animals. In humans,
44 it manifests as febrile illness, resulting in retinal degeneration, severe encephalitis,
45 haemorrhage; fatal hepatitis occurs in less than 1% of patients [1]. Protection of animals from

bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

46 the disease can be conferred by vaccination; however, there is currently no approved vaccine
47 for use in humans. Transmission of RVF can be prevented by eliminating the mosquito vectors
48 and avoiding human contact with infected tissues of infected animals.

49 The RVF virus has been isolated from more than 30 species of mosquitoes, belonging to at
50 least six genera (*Aedes, Culex, Anopheles, Eretmapodites, Mansonia* and *Coquillettidia*). In a
51 number of mosquito species, the virus has been isolated from both the insects and their eggs,
52 suggesting different modes of transmission [2, 3]. Furthermore, the virus has been isolated from
53 unfed mosquitoes reared from eggs obtained during inter-epidemic periods in Kenya and South
54 Africa [3,4]. Eggs from floodwater mosquitoes can remain viable for several years, hatching
55 during conducive climatic conditions associated with periods of high rainfall. Changing climate
56 and farming systems can create conditions favorable for mosquito breeding, resulting in
57 unexpected outbreaks of the disease [5].

58 The disease is endemic in eastern and southern Africa, but outbreaks have been reported in
59 Egypt, Madagascar, Mauritania, Saudi Arabia, Sudan and Yemen [6,7,8]. No significant
60 antigenic differences have thus far been demonstrated among isolates of this virus from
61 different geographic locations, confirming the existence of a single RVFV serotype. Despite the
62 single serotype, differences in virulence and pathogenicity of the virus have been observed,
63 necessitating the need for detailed genetic characterization of the various isolates [9,10].

64 The genome of RVFV, like that of the other *Bunyaviridae*, consists of three-segmented, single-
65 stranded negative- and ambi-sense RNAs with a total size of 12kb. The L (Large) segment
66 codes for the viral RNA polymerase. The M (Medium) segment encodes a single precursor
67 protein which is cleaved to produce the envelope glycoproteins G1 and G2, and two non-

bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

68 structural proteins of 78kDa and 14 kDa. In contrast, the ambisense S (Small) segment codes
69 for the nonstructural protein NSs in the genomic sense and the nucleocapsid protein N in the
70 antigenomic sense [9,11,12,13].

71 Because of its segmented structure, the genome of RVFV is thought to undergo recombination
72 through reassortment, thereby contributing to its evolutionary dynamics [11,14]. In general, the
73 RVFV genome is characterized by low genetic diversity (~5%); consequently, it is difficult to
74 statistically detect intragenic recombination events [9,15]. Similar to other arboviruses, all the
75 genes of RVFV are under purifying selection and have evolved at distinct rates by accumulating
76 mutations at 1.9×10^{-4} to 2.5×10^{-4} substitutions per site per year [9,14]. The previously
77 estimated time to most recent common ancestor (TMRCA) is at around 124 to 133 years. This
78 coincides with the importation of highly susceptible European breeds of cattle and sheep [10]
79 into East Africa where the disease was first reported [15].

80 Despite the high sequence identity, nucleotide sequences of partial M segments from viruses
81 isolated over 60 years from various countries have been grouped into 15 lineages [16]. The
82 phylogenetic trees constructed from the nucleotide sequences of the three partial genome
83 segments suggested the existence of reassortment, specifically of a 2010 isolate from a patient
84 in South Africa. The individual was accidentally co-infected with live RVF animal vaccine and a
85 RVF virus in lineage H [16]. Around this time, a RVF outbreak with unusual clinical presentation
86 in animals was observed in South Africa. This outbreak had two distinguishing features: first, it
87 occurred atypically in the absence of abnormally high rainfall; secondly, in addition to causing
88 abortions storms, it had a high mortality among pregnant adult cattle [17].

bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

89 The first case of RVF in South Africa occurred in the summer of 1950-1951 in animals and
90 subsequently it was diagnosed in humans in 1951 [18, 19]. Three major outbreaks of the
91 disease occurred in South Africa in 1950-1951, 1974-1976 and, most recently, in 2008-2011.
92 There were minor incidents in the inter-episodic periods interspersing these outbreaks [20].

93 Confirmation of suspected cases of RVF in animals in South Africa is normally done at
94 Agricultural Research Council–Onderstepoort Veterinary Research (ARC-OVR). Over time, the
95 institute has accumulated a large collection of RVFV isolates from a majority of reported cases
96 of the disease in South Africa. In order to obtain comprehensive information on the genetic
97 composition of the RVF viruses (RVFVs) circulating in South Africa, we performed full genome
98 sequence analysis of some of the viruses isolated from animals at discrete foci of the outbreaks
99 which occurred during the 2008-2010 period. The genome sequences of these viruses were
100 compared with the genome sequences of other RVFVs from earlier outbreaks in South Africa
101 and other countries where the disease has occurred.

102

103 ## Methods

104 ### Growth, isolation and purification of RVF virus isolates

105

106 Presence of RVF virus nucleic acids in samples collected from animals suspected to be
107 affected by the disease was confirmed by real-time PCR, using a slight modification of an
108 established method [21].
109 Optimum conditions for efficient infection of Baby Hamster Kidney (BHK 21) cells (obtained
110 from AATC) with RVF virus were established empirically using isolate M35/74, the challenge

bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

111 strain of RVFV [22]. The BHK 21 cells were grown in DMEM-F12 supplemented with 5%FBS
112 (LONZA) and 1% pen/strep Amphotericin B (LONZA). These conditions were applied to infect
113 BHK cells at a MOI resulting in the highest viral load. The infected cells were pelleted by
114 centrifugation at 2 500 rpm for 5 minutes and the supernatant recovered. The supernatant
115 was committed to sequence independent single primer amplification (SISPA) [23]. Briefly, the
116 viral particles in the supernatant were treated with 100U DNase I and 4μg RNase at 37^0C for
117 2h to remove possible host nucleic acids contamination. Viral RNA was extracted using
118 TRIZOL LS kit (Invitrogen) according to the procedure provided by the supplier (Invitrogen).
119 The RNA was recovered and used as the template in the first strand cDNA synthesis primed
120 with FR26RV-N (5'GCC GGA GCT CTG CAG ATA TCN NNN NN3' [23]. The single-stranded
121 cDNA was the template for double-stranded cDNA synthesis using random 20mer primers
122 and Klenow fragment of *E. coli* DNA polymerase. These products were subjected to PCR
123 amplification using the 20-mer region of the above primer (FR26RV: 5'GCC GGA GCT CTG
124 CAG ATA TC3') in a reaction incubated in a thermocycler programmed to denature at 94 °C,
125 2 min then 35 cycles of 94 °C, 30 sec; 55 °C, 30 sec; 68 °C, 30s; with a final extension at 68
126 °C, for 10 min. The SISPA products were resolved by electrophoresis in 1% agarose gels.
127
128 Construction of cDNA libraries
129
130 The SISPA products ranging in size from 0.2kb to 1.5kb were recovered from agarose gels and
131 used in the preparation of library for sequencing reactions on the New Generation Sequencing
132 (NGS) platforms exactly as described by the manufacturers (Roche Applied Science or
133 Illumina). The sequencing was done on the Genome Sequencer 454 platform (GSFLX; 454 Life

bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

134 Sciences, Roche Applied Science; http://www.454.com); SISPA products from two random
135 isolates were sequenced also on Illumina platform (http://www.illumina.com).
136

137 Bioinformatics analyses of the sequence data

138 The sequence data obtained was processed and assembled into contigs using the appropriate
139 software set to default values (Roche/454 Newbler for 454 Life Sciences Corporation, Software
140 Release: 2.8 — 20120726_1306 or CLC Genomics Workbench, QIAGEN Bioinformatics).
141 The sequence data were subjected to further analyses using a combination of bioinformatics
142 software. The nucleotide sequences were aligned using Clustal W [24] within the Molecular
143 Evolutionary Genetics Analysis (MEGA) [25] set to optimum parameters for each sequence
144 type. The best fitting nucleotide substitution model was determined for each genome segment
145 using MEGA 6 and then applied in all the subsequent analyses. The aligned nucleotide
146 sequences were used in calculating the mean pairwise distances and to derive phylogenetic
147 trees using Maximum likelihood under 1000 bootstrap iterations [26].
148 Evidence for possible intragenic recombination events among the isolates was sought using
149 different methods available from RDP3 [27]. Rates of molecular evolution for individual genome
150 segments were estimated using Bayesian Markov Chain Monte Carlo implemented in the
151 BEAUTI v1.8.1, BEAST v1.8.1, Tracer and FigTree packages [28]. The substitutions rates were
152 estimated using both strict and relaxed uncorrelated lognormal molecular clock under General
153 Time Reversible (GTR) model with gamma distribution (T4). The general Bayesian skyline
154 coalescent prior was used and the MCMC allowed to run for sufficient number of generation
155 with sampling every 1000 states, to ensure convergence of all parameters [28].
156

bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

157

158

159 Rift Valley fever virus genome sequence accession numbers

160

161 The nucleotide sequences of all the segments of the RVF isolates analyzed in the current study
162 have been deposited in GenBank with accession numbers indicated in Table 1.

163

164 **Table I**. List of RVF virus isolates analyzed in the study and GenBank accession numbers assigned to
165 the nucleotide sequences of their respective L, M and S segments. With the exception of Madagascar,
166 origin imply the South African Province and nearest town from which the isolate originated.

167

bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

168

Isolate #/year	GenBank Accession number			Host	Tissue source	Origin
	L segment	M segment	S segment			
M48/08	KX944866	KX944843	KX944820	Bovine	Liver	Madagascar
M47/08	KX944865	KX944842	KX944819	Bovine	Liver	Warmbad, (LP)
M39/08	KX944864	KX944841	KX944818	Buffalo calf	Whole foetus	Neilspruit (MP)
M37/08	KX944863	KX944840	KX944817	Buffalo calf	Whole foetus	Hoedspruit (LP)
M85/08	KX944871	KX944848	KX944825	Bovine Calf	Carcase	Irene (GP)
M84/08	KX944870	KX944847	KX944824	Bovine	Blood	Irene (GP)
M80/08/2	KX944869	KX944846	KX944823	Bovine	Blood	Irene (GP)
M66/09	KX944868	KX944845	KX944822	bovine	Liver	Modderfontein (GP)
M260/09	KX944860	KX944837	KX944814	Bovine	Foetus organ	Upington (NC)
M259/09	KX944859	KX944836	KX944812	Bovine	Blood	Upington (NC)
M247/09	KX944857	KX944834	KX944811	Ovine	Liver, lung	Upington (NC)
M127/09	KX944851	KX944828	KX944804	Bovine	Blood, fresh liver	Cascade (KZN)
M33/10	KX944862	KX944839	KX944816	Ovine	Liver	Middleburg (EC)
M29/10	KX944861	KX944838	KX944815	Bovine	Blood	Pretoria (GP)
M25/10	KX944858	KX944835	KX944813	Ovine	Organs	Bultfontein (FS)
M12/10	KX944850	KX944827	KX944805	Ovine	Blood	Bultfontein (FS)
M19/10	KX944853	KX944830	KX944809	Ovine	Liver, Lung, Brain and spleen	Heldefontein (FS)
M16/10	KX944852	KX944829	KX944806	Ovine	Organs	Bultfontein (FS)
M06/10	KX944849	KX944826	KX944803	Bovine	Blood	Sterkfontein (EC)
M21/10	KX944856	KX944833	KX944810	Ovine	Organ pool	Bloemfontein (FS)
M57/74	KX944867	KX944844	KX944821	N/A	N/A	N/A
M1975Bov	KX944855	KX944832	KX944808	Bovine	N/A	N/A
M1955	KX944854	KX944831	KX944807	N/A	N/A	N/A

169 N/A: Not applicable (the information is unknown). LP: Limpopo Province; MP: Mpumalanga Province;
170 GP: Gauteng Province; NC: Northern Cape Province; KZN: KwaZulu Natal Province; EC: Eastern Cape
171 Province; FS: Free State Province.

bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

172

173
Results

174 Rift Valley fever virus SISPA products yield identical nucleotide sequence data irrespective of
175 NGS platform used. This study focused on RVF viruses isolated during the disease outbreaks
176 in South Africa in the period spanning 2008 to 2010, but also included viruses from the other
177 major outbreaks in 1955 (M1955) and 1974-1975 (M57/74 and M1975Bov). The complete
178 genome sequences of 23 isolates were determined using a combination of SISPA [23] and two
179 independent NGS technologies. A representative profile of SISPA products obtained from the
180 virus isolates is shown in Fig. 1. The 23 RVFV isolates whose genome sequences were
181 determined represent four of the 15 reported outbreaks in 2008, three of the 19 outbreaks
182 reported in 2009 and six of the 484 outbreaks in 2010 (Table 1) [29]. The complete genome
183 sequence of isolate M48/08 from Madagascar was also determined (Table 1). The genomes of
184 isolates M260/09 and M247/09 were sequenced on both GSFLX 454 and Illumina platforms,
185 for comparison of sequence data obtained from SISPA products on either of these NGS
186 technologies. The sequencing data obtained indicated that either of the technologies can be
187 utilized with SISPA to obtain accurate genome sequences.

188

189 **Fig1.** Photograph of a representative agarose gel in which SISPA products of RVF viruses were resolved. Lanes
190 contain products as follows: lane 1: M03/10, lane 2: M15/10, lane 3: M06/10, lane 4: M19/10, lane 5: M21/10, lane
191 6: M22/10, lane 7: M23/10, lane 8: M25/10, lane 9: M26/10, lane 10: M33/10 lane 11: no DNA. Lanes labeled M
192 contain DNA size markers, with corresponding sizes of some indicated in kilobasepairs (kb).

193

194

bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

195 The nucleotide sequence data of all the RVFVs were assembled for complete S, M and L
196 segments analyses and were deposited in GenBank with accession numbers as indicated in
197 Table 1.

198 Coding regions under selection pressure

199 Sequence alignments were generated for each of the three segments using all the available
200 RVFV sequence data in GenBank. The alignments, which included full genome sequences
201 from 120 – 140 isolates depending on the segment, were used in evaluating the evolutionary
202 dynamics acting on each of the three segments.

203 Generally, sequence diversity among the segments were <5% among S or L segments, and
204 <6% among M segments. Bayesian coalescent estimations of RVF genomes indicated that the
205 segments evolve at a mean rate between 3.9 $x10^{-4}$ and 4.17 $x10^{-4}$ substitutions per site per
206 year, regardless of the molecular model used. This is in agreement with previous Bayesian
207 estimations [14,28]. Similarly, the estimated Time to Most Recent Common Ancestor (TMRCA)
208 supports previous estimations of between 1880 and 1890 [14,28]. In order to determine the
209 influence of substitution rate on biological function, we estimated the effect of differential
210 selection pressures by calculating the rate of non-synonymous (d_N) to synonymous (d_S)
211 substitutions. All the coding regions were found to be under purifying selection pressure (d_N/d_S
212 <1).

213 Evidence for M segment reassortment but no intragenic recombination

214 Using Maximum likelihood trees, the phylogenetic relationship of the 23 RVFV isolates was
215 assessed in relation to those of 50 other isolates genomes sequences of which were already
216 in GenBank (S1 Table). Incongruences among the phylogenies of the individual genome

bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

217 segments were observed (Fig 2A-C), prompting the investigation into possible influence of
218 recombination and reassortment.

219 **Fig2A-C**. Phylogenetic trees derived from nucleotide sequence data of the three genomic segments of RVF virus
220 segments. Segment S (A), Segment M (B) and Segment L (C) follows the same Lineage names as described by
221 Grobbelaar et al., 2013. All Lineages containing RVF viruses from South Africa are presented in green. The two
222 segment M re-assortments are indicated in red.
223

224 Insignificant statistical support for intragenic recombination events were predicted in all three
225 segments (data not shown). This is explained by the low genetic diversity among the sequences
226 [30]. Reassortment has been described for RVFV [11,14] and was therefore investigated
227 utilizing the current data.

228 Fifteen lineages have been described using nucleotide sequences of parts of the three
229 segments individually [16]. Our data indicate that one 2009 and all the 2008 isolates from South
230 Africa and Madagascar (M49/08) clustered in Lineage C or Kenya-1[10,15] (Figs 2A, 2B and
231 2C). The remaining of the 2009 and 2010 isolates clustered within Lineage H [14], with the
232 exception of isolate M259_RSA_09. The latter originated from serum of a bovine in Upington,
233 Northern Cape Province, South Africa. Both its segments L and S cluster in Lineage K together
234 with that of isolate JQ068143 from Kakamas (also in the Northern Cape Province); however,
235 its segment M clustered in Lineage H along with the rest of the 2009 – 2010 isolates (Fig 2B).
236 This indicates that isolate M259_RSA_09 is probably a segment reassortant from a coinfection
237 with RVFVs in Lineages H and K. Whether this event occurred in an insect vector or an animal
238 host Is not clear.

239 Segments L and S of isolate M1955_RSA_55 cluster in Lineage I together with the 1951 South
240 African Van-Wyck isolate (DQ380158) (Figs 2A and C). However, segment M of this isolate

bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

241 (M1955_RSA_55) places it in Lineage L with the isolates of the 1974 and 1975 outbreaks in
242 South Africa (Fig 2B). Thus, isolate M1955_RSA_55 was the second RVFV in this study, which
243 had sequence features suggesting that it may be a segment reassortant.

244 Since both these putative reassortment events relate only to Segment M, which encodes two
245 glycoproteins (Gc and Gn), the segment was subjected to additional analysis. The amino acid
246 sequences of the glycoproteins encoded by the M segment of different RVF virus isolates from
247 the 2008-2010 outbreak were compared to those previously published (S1 Table). The
248 predicted amino acid residues are conserved with <3% sequence identity. Of the amino acid
249 changes, 55% are conservative, 9.8% result in loss of a charge, 17% in gain of a charge and
250 2.7% in change of a charge. The positions of amino acid substitutions relative to the proportion
251 of sequences with that change and those resulting in a change of charge are shown in Fig 3.
252 Even though the majority of the substitutions are at the C-terminal region of the glycoprotein
253 Gn, they are only observed in few of the sequences, the majority of them being conservative
254 substitutions. One exception found was a change from D (Aspartic acid) to N (Asparagine) at
255 the amino acid position 95, which is prominent in the 2008 – 2009 isolates in Lineage C, Fig
256 2B.

257

258 **Fig3.** Welling antigenicity plots of proteins encoded by the M segments of isolates M33_RSA_10 in blue, ZH501-
259 Egy-77 in black and M37_RSA_08 in green. Differences in amino acids residues among the three isolates are
260 indicated on top of the antigenicity plots, with each isolate represented in its assigned colour. A graphical
261 representation of the Non-structural protein (NSm), and glycoproteins (Gn) and (Gc) regions separate the
262 antigenicity plots from the graph depicting the proportion of substitutions per amino acid position. These are
263 representative of the 23 sequences of the M segments analyzed in this study and those of the previously published
264 RVF viruses listed in S1 Table.

265

bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

266 To investigate the possible influence of the amino acid changes on the antigenic properties of
267 the viruses, we performed antigenicity predictions using Welling [32], with a window size of 11.
268 Antigenicity plots for isolate M33_RSA_10, M37_RSA_08 and T1 ZH-501 isolated in Egypt in
269 1977 are shown in Fig 3. These isolates represent Lineages H, C and A, respectively. Virus
270 ZH501 from the 1977 outbreak in Egypt has been shown to be associated with increased
271 virulence in rats [33]. It was therefore included in this analysis for comparison with isolates from
272 Lineages H and C [16]. The major differences in antigenicity predictions between ZH501-77
273 and the South African isolates are at positions 60 and 631 (Fig 3). Isolate ZH501-77 has a
274 valine at both of these positions in contrast to the South African isolates which have isoleucine.
275 Association of virulence with amino acid substitutions at positions 595, 631, 659 and 1059 has
276 been shown in previous studies [10,33].

277

278 Discussion

279 South Africa has experienced three major periods of RVF outbreaks, the first in 1950-1951, the
280 largest in 1974-1976 and lastly in 2008-2011. In 2008 a total of 15 outbreaks were reported,
281 localized to the central provinces of Limpopo, Mpumalanga, North West and Gauteng [29].
282 Complete genome sequence analysis of viruses isolated from the 2008 outbreak, clusters them
283 in Lineage C together with isolates from a 2007 outbreak in Kenya, known as Lineage Kenya-
284 1, Fig 2A, Fig 2B and Fig 2C [10,16]. In contrast to the centralized outbreak of 2008, 19
285 outbreaks which were reported in 2009 were distributed, with single cases in Mpumalanga and
286 Gauteng, and the rest in KwaZulu-Natal, the Eastern Cape or along the Orange River in the
287 Northern Cape [29,34]. Similar to the isolates from the 2008 outbreaks, the 2009 isolate from

bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

288 Gauteng clustered in Lineage C, within Lineage Kenya-1 (Fig 2A and B). This 2009 Gauteng
289 outbreak appears to have been caused by the 2008 RVF viruses present in that region. Isolates
290 from both the geographically distinct KwaZulu-Natal and Northern Cape outbreaks of 2009,
291 clustered in Lineage H (Fig 2A). This lineage includes isolate M259_RSA_09, a segment M
292 reassortant, whose segments L and S cluster with the 2009 Kakamas isolate (JQ068142) in
293 Lineage K (Fig 2B). It is therefore possible that the RVF viruses associated with the majority of
294 the outbreaks in 2009 originated from a single source.

295 In 2010 a total of 484 outbreaks were reported in every province of South Africa, except
296 KwaZulu-Natal [20]. The initial report of an outbreak was from Bultfontein and Brandfort, both
297 in the Free State and subsequently cases were reported from across the country [29]. Similar
298 to the virus isolates from the 2009 outbreaks in KwaZulu-Natal and the Northern Cape, all the
299 isolates from the 2010 outbreaks clustered in Lineage H (Fig 2A and B). The eight isolates from
300 the 2010 outbreak analyzed in this study (Table 1 and Fig 2) are not necessarily statistically
301 representatives of the 14342 cases reported in that year, but analyses of their nucleotide
302 sequence data support speculations that the 2010 outbreak was a continuation of the 2009 in
303 KZN and Northern Cape outbreaks. The clustering of isolates in lineage H (Fig 2A) gives an
304 indication that new strains could evolve due to nucleotide substitution (Fig 3), albeit at slow/low
305 rate. It is possible that these viruses were not introduced from elsewhere outside South Africa,
306 but rather that they mutated over time and caused outbreaks when suitable conditions
307 prevailed.

308 This study has contributed full genome sequence of RVFVs M57_RSA_74 isolated during the
309 1974 outbreaks and M1955_RSA_55 isolated from one of the 28 outbreaks in 1955 [29]. The
310 largest RVF epidemic reported in South Africa were between 1973 and 1976, with mortality

bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

311 rates of 95% and cases reported from every province [29, 35]. Previous studies have clustered
312 the 1973-1975 isolates into Lineage L along with a 1970 isolate from Zimbabwe and a 1956
313 isolate from Kenya[16]; as expected isolate M57_RSA_74 clustered with these (Fig 2A, 2B
314 and C). In contrast, Segment S and Segment L of isolate M1955_RSA_55 clustered with a
315 1951 South African isolate known as van-Wyck in Lineage I (Fig 2A and C) [16], but Segment
316 M clustered with isolates from the 1973-1976 outbreaks in Lineage L (Fig 2B), making this
317 1955 isolate a segment M reassortant. The occurrence of segment M reassortment in
318 M1955_RSA_55 indicates that multiple RVF virus lineages can co-circulate, resulting in
319 reassortant viruses re-emerging decades later causing disease outbreaks.
320 The evolutionary dynamics of RVFVs are characterized by low substitutions rates (3.9 $x10^{-4}$
321 and 4.17 $x10^{-4}$ substitutions per site per year) under strong purifying or negative selection with
322 the major genomic diversity resulting from reassortment [11,14]. Similar evolutionary dynamics
323 have been described in other arboviruses such as bluetongue virus and Epizootic
324 haemorrhagic disease virus, due to the obligatory replication of the virus in both its insect vector
325 and mammalian host [36]. The majority of reasortment events described in RVFV involves the
326 exchange of segment M, resulting in antigenic shift due to the two glycoproteins Gn and Gc
327 encoded by this segment [10,14,16].
328 Although RVF virus is antigenically homogenous, various isolates of the virus exhibit
329 differences in virulence, evident upon infection of the mammalian host [33,36]. Whereas some
330 of these differences may be attributable to the individual host, others are inherent to the virus.
331 Differences in virulence and lethality of RVF virus isolates have been observed during the
332 experimental infection of BHK cells [37], mice [33], sheep [38] and cattle [37]. Significant
333 differences associated with the severity of RVF in humans have been observed [39, 40]. An

bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

increase in the severity of RVF since the 1977 outbreak in Egypt to the devastating outbreak during the 2006 -2008 in East Africa have been observed [41]. A clear association between RVFV genotype and lethal phenotype has not been established; however, various amino acid substitutions have been implicated in this phenotype [37]. Therefore, indicative amino acid changes in some of the RVF proteins were investigated. The most prominent substitution in the glycoproteins are 595 I>V, 605 R>K, 631 I>V, 659 V>A located in Gn and 1059 S>T within Gc [10]. Another variation was identified in ZH501, isolate from a human in Egypt during an outbreak in 1977, which resulted in the change of Glycine to Glutamic acid at position 277. The virus with the Glutamic acid displayed an increased virulence in mice, compared to the virus with Glycine in the same position [33].

The majority of RVF viruses analysed in this study had Glutamine at position 277, except wild type isolate 763/70 from a foetus aborted during an outbreak of the disease in Zimbabwe in 1970 [10]. This study identified additional substitutions between the lethal isolate ZH501-77 from Egypt and isolates belonging to Lineage H from 2010 in South Africa (Fig 3B). The substitutions included 602 V>I, 987 D>E and 1131 T>I. The impacts of each of these substitutions on the pathogenicity of RVFVs remain to be investigated.

The fact that the sequence data of two isolates (M260/09 and M247/09) generated by different NGS platforms clustered together (Fig 2A) demonstrates that the two platforms produce identical sequences. One caveat with the dataset analyzed in this study is that the isolates might not be representative of the RVF viruses circulating during the 2008-2010 outbreak. This is inherent in the way the study was done: samples brought for testing at the ARC-OVR are opportunistic and are not necessarily representative of cases of RVF in animals in South Africa. During this period; the RVFVs whose genomes could be analyzed are the viruses that infect

bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

357 BHK 21 cells growing in culture media; and finally, good quality sequence data could not be
358 obtained from all RVFVs, which were isolated in cell culture. A different picture of viruses and
359 their potential quasispecies might emerge when the analyses are performed on viruses
360 obtained directly from representative proven clinical cases. This is currently unattainable in our
361 system, but determining the entire RVF viral genome sequence directly from clinical samples
362 is being investigated.

363

bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

364 ## Acknowledgements

365 We thank Dr Marco Romito for performing PCR tests which identified the samples positive for
366 RVF virus, the technical staff in Virology Laboratory for isolating RVF viruses from the positive
367 samples; and Drs Keoagile Oliver Bezuidt and Oleg Reva, FABI, University of Pretoria, for
368 assistance with some parts of the bioinformatics work.

369 Preliminary parts of this work have previously been presented at: Interregional Conference on
370 Rift Valley Fever in the Middle East and Horn of Africa, 21 - 23 April 2015, Djibouti City, Djibouti
371 and 11th Annual Sequencing, Finishing and Analysis in the Future (SFAF) conference, 1st - 3rd
372 June 2016, Santa Fe, NM, USA.

373

374

375 ## References

376 1. Daubney R, Hudson J, Garnhann P. Enzootic hepatitis or Rift Valley fever. An undescribed virus
377 disease of sheep, cattle and man from East Africa. J Pathol Bacteriol 1931; 34: 545-579.

378 2. Linthicum KJ, Davies FG, Kairo A, Bailey CL. Rift Valley fever virus (family Bunyaviridae, genus
379 Phlebovirus). Isolations from Diptera collected during an inter-epizootic period in Kenya. J Hyg.
380 1985; 95:197-209.

381 3. Turell MJ, Presley SM, Gad AM, Cope SE, Dohm DJ, Morrill JC, et al. Vector competence of
382 Egyptian mosquitoes for Rift Valley fever virus. Am J Trop Med Hyg. 1996; 54:136-9.

bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

383 4. Alexander RA. Discussion on both Rift Valley Fever (R.V.F.) and Wesselsbron Virus Disease
384 (W.V.D.). In: Proceedings of the IVth Annual Meeting of the Inter African Advisory Committee on
385 Epizootic Disease. 1957 p. 29. Dakar: Interafrican Bureau for Epizootic Diseases.

386 5. Anyamba A, Chretien J-P, Small J, Tucker CJ, Formenty PB, Richardson JH, et al. Prediction of a
387 Rift Valley fever outbreak. Proc Natl Acad Sci U S A. 2009; 106:955-9.

388 6. Schoemaker T, Boulianne C, Vincent MJ, Pezzanite L, Al-Qahtani MM, Al-Mazrou Y, et al. Genetic
389 Analysis of viruses associated with emergence of Rift Valley fever in Saudi Arabia and Yemen,
390 2000-01. Emerg Infec Dis. 2002; 8: 1415-20.

391 7. Morvan J, Lesbordes JL, Rollin PE, Mouden JC, Roux J. First fatal human case of Rift Valley fever
392 in Madagascar. Trans R Soc Trop Med Hyg. 1992; 86:320.

393 8. Hassan OA, Ahlm C, Sang R, Evander M. The 2007 Rift Valley fever outbreak in Sudan. PLoS
394 Negl Trop Dis. 2011; 5(9):e1229. doi:10.1371/journal.pntd.0001229.

395 9. Muller R, Saluzzo JF, Lopez N, Dreier T, Turell M, Smith J, et al. Characterization of clone 13, a
396 naturally attenuated avirulent isolate of Rift Valley fever virus, which is altered in the small
397 segment. Am J Trop Med Hyg. 1995; 53(4):405-11.

398 10. Bird BH, Khristova ML, Rollin PE, Ksiazek TG, Nichol ST. Complete genome analysis of 33
399 ecologically and biologically diverse Rift Valley fever virus strains reveals widespread virus
400 movement and low genetic diversity due to recent common ancestry. J Virol. 2007; 81:2805-16.

401 11. Sall AA, Zanotto PM De A, Sene OK, Zeller HG, Digoutt JP, Thiongane Y, et al. Genetic
402 reassortment of Rift Valley fever virus in nature. J Virol. 1997; 3:8196-8200.

403 12. Liu L, Celma CCP, Roy P. Rift Valley fever virus structural proteins: expression, characterization
404 and assembly of recombinant proteins. Virol J. 2008; 5:82. DOI: 10.1186/1743-422X-5-82.

bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

405 13. Bouloy M, Weber F. Molecular biology of Rift Valley fever virus. Open Virol J.2010; 4:8-14.

406 14. Freire CCM, Iamarino A, Ly Soumare PO, Faye O, Sall AA, Zanotto PMA. Reassortment and
407 distinct evolutionary dynamics of Rift Valley fever virus genomic segments. Sci Rep. 2015; 5:11353
408 DOI:10.1038/srep11353.

409 15. Davies FG. Observations on the epidemiology of Rift Valley fever in Kenya. J. Hygiene 1975;
410 75:219-30.

411 16. Grobbelaar AA, Weyer J, Leman PA, Kemp A, Paweska JT, Swanepoel R. Molecular epidemiology
412 of Rift Valley fever virus. Emerg Infect Dis. 2011; 17:2270-6. Doi:0.3201/eid1712.111035.

413 17. Kegakilwe, P. S. An atypical outbreak of Rift Valley fever in the Northern Cape in October 2009.
414 In: Proceedings of the 9th annual congress of the Southern African Society for Veterinary
415 Epidemiology and Preventive Medicine. 18-20 August 2010. Farm Inn, Republic of South Africa

416 18. Alexander RA. Rift Velley fever in the Union. J S Afr Vet Med Assoc. 1951; 22:105.

417 19. Schulz KH. Rift valley fever in South Africa, Special Report No 5/51. Union Department of Health,
418 Plague Research Laboratory, Johannesburg, South Africa, May 1951.

419 20. Pienaar NJ. A retrospective analysis of the epidemiology of Rift Valley fever in South Africa
420 [dissertation]. Pretoria (South Africa): University of Pretoria; 2011.

421 21. Drosten C, Göttig S, Schilling S, Asper M, Panning M, Schmitz H, et al. Rapid detection and
422 quantification of RNA of Ebola and Marburg viruses, Lassa virus, Crimean-Congo hemorrhagic
423 fever virus, Rift Valley fever virus, dengue virus, and yellow fever virus by real-time reverse
424 transcription-PCR. J Clin Microbiol. 2002; 40:2323-30.

425 22. von Teichman B, Engelbrecht A, Zulu G, Dungu B, Pardini A, Bouloy M. Safety and efficacy of Rift
426 Valley fever Smithburn and Clone 13 vaccines in calves. Vaccine. 2011; 29:5771-7. doi:
427 10.1016/j.vaccine.2011.05.055.

bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

428 23. Djikeng A, Halpin R, Kuzmickas R, Depasse J, Feldblyum J, Sengamalay N, et al. Viral genome
429 sequencing by random priming methods. BMC Genomics. 2008; 9:5. doi: 10.1186/1471-2164-9-
430 5.

431 24. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive
432 multiple sequence alignment through sequence weighting, position-specific gap penalties and
433 weight matrix choice. Nucleic Acids Res. 1994; 22:4673-80.

434 25. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S. MEGA5: Molecular
435 Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum
436 Parsimony Methods. Mol Biol Evol. 2011; 28: 2731-2739. DOI: 10.1093/molbev/mrs121.

437 26. Hall BG. Building phylogenetic trees from molecular data with MEGA. Mol Biol Evol. 2013;
438 30:1229-35.

439 27. Heath L, van der Walt E, Varsani A, Martin DP. Recombination patterns in aphthoviruses mirror
440 those found in other picornaviruses. J Virol. 2006; 80:11827-32.

441 28. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUTi and the
442 BEAST 1.7. Mol Bio Evol. 2012; 29:1969-73.

443 29. Pienaar NJ, Thompson PN. Temporal and spatial history of Rift Valley Fever in South Africa: 1950
444 to 2011. Onderstepoort J Vet Res. 2013; 80:384. http:// dx.doi.org/10.4102/ojvr. v80i1.384
445

446 30. Prosada D, Crandall KA, Holmes EC. Recombination in evolutionary genomics. Ann Rev Genet.
447 2002: 36. 75-97. DOI: 10.1146/annurev.genet.36.040202.111115
448

bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

449 31. Pepin M, Bouloy M, Bird BH, Kemp A, Paweska J. Rift Valley fever virus (Bunyaviridae:
450 Phlebovirus): an update on pathogenesis, molecular epidemiology, vectors, diagnostics and
451 prevention. Vet Res. 2010; 41:61.

452 32. Welling GW, Weijer WJ, van der Zee R and Welling-Wester S. Prediction of sequential antigenic
453 regions in proteins. FEBS Letters, 1985; 188(2): 215-218.
454

455 33. Morrill JC, Ikegami T, Yoshikawa-Iwata N, Lokugamage N, Won S, Terasaki K, et al. Rapid
456 accumulation of virulent Rift Valley fever virus in mice from an attenuated virus carrying a single
457 nucleotide substitution in the mRNA. PLoS One. 2010; 5: e9986. Erratum in: PLoS One. 2010;
458 5(4). doi: 10.1371/annotation/326c5f5b-3da9-4b0b-b889-28d4f33d2543.

459 34. Williams R, Malherbe J, Weepener H, Majiwa P, Swanepoel R. Anomalous High Rainfall and Soil
460 Saturation as Combined Risk Indicator of Rift Valley Fever Outbreaks, South Africa, 2008-2011.
461 Emerg Infect Dis. 2016; 22:2054-62. doi: 10.3201/eid2212.151352.

462 35. Swanepoel R, Coetzer JAW. Rift Valley fever. In: Coetzer JAW,Tutsin RC, editors. Infectious
463 disease of livestock with special reference to Southern Africa. Cape Town: Oxford University
464 Press: 2004. 1037-70.

465 36. Wilson WC, Gaudreault NN, Jasperson DC, Johnson DJ, Ostlund EN, Chase CL, et al. Molecular
466 evolution of American field strains of Bluetongue and Epizootic haemorrhagic disease viruses. Vet
467 Ital. 2015; 51(4). 269-273. doi: 10.12834/VetIt.555.2627.1

468 37. Wilson WC, Davis AS, Gaudreault NN, Faburay B, Trujillo JD, Shivanna V, et al. Experimental
469 Infection of Calves by Two Genetically-Distinct Strains of Rift Valley Fever Virus. Viruses. 2016;
470 8(5). pii: E145. doi: 10.3390/v8050145.

bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

471 38. Moutailler S, Roche B, Thiberge JM, Caro V, Rougeon F, Failloux AB. Host alternation is
472 necessary to maintain the genome stability of Rift Valley fever virus. PLoS Negl Trop Dis. 2011; 5:
473 e1156. doi: 0.1371/journal.pntd.0001156.

474 39. Faburay B, Gaudreault NN, Liu Q, Davis AS, Shivanna V, Sunwoo SY, et al. Development of a
475 sheep challenge model for Rift Valley fever. Virol. 2016; 489:128-40. doi:
476 10.1016/j.virol.2015.12.003.

477

478 40. Al-Hazmi M, Ayoola EA, Abdurahman M, Banzal S, Ashraf J, El-Bushra A, et al. Epidemic Rift
479 Valley fever in Saudi Arabia: a clinical study of severe illness in humans. Clin Infect Dis. 2003;
480 36:245-52.

481 41. Baba M, Masiga DK, Sang R, Villinger J. Has Rift Valley fever virus evolved with increasing
482 severity in human populations in East Africa? Emerg Microbes Infect. 2016; 5: e58. doi:
483 10.1038/emi.2016.57.

484

485

486

487

488

489

490

bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

491 **S1 Table**. Previously published RVF viruses used in the study

Isolate Name	Country	Year	Segment L	Segment M	Segment S
T1: mosquito which fed on hamster infected with ZH-501	Egypt	1977	DQ375407 Egy Sha T1 77	DQ380201 Egy T1 ZH-501 77	DQ380150 Egy T1 ZH-501 77
ZH-1776	Egypt	1978	DQ375411 Egy Gha ZH-1776 78	DQ380203 Egy Gha ZH-1776 78	DQ380153 Egy Gha ZH-1776 78
ZM-657	Egypt	1978	DQ375409 Egy Sha ZM-657 78	DQ380204 Egy Sha ZM-657 78	DQ380146 Egy Sha ZM-657 78
ZS-6365	Egypt	1979	DQ375410 Egy Cha ZS-6365 79	DQ380205 Egy Gha ZS-6365 79	DQ380145 Egy Gha ZS-6365 79
ZH-548	Egypt	1977	DQ375403 Egy Sha ZH-548 77	DQ380206 Egy Sha ZH-548 77	DQ380151 Egy Sha ZH-548 77
ZC-3349	Egypt	1978	DQ375412 Egy Asy ZC-3349 78	DQ380207 Egy Asy ZC-3349 78	DQ380152 Egy Asy ZC-3349 78
763/70	Zimbabwe	1970	DQ375426 Zim Sal 763 70	DQ380188 Zim Sal 763 70	DQ380174 Zim Sal 763 70
2373/74	Zimbabwe	1974	DQ375432 Zim Sal 2373 74	DQ380194 Zim Sal 2373 74	DQ380159 Zim Sal 2373 74
2250/74	Zimbabwe	1974	DQ375413 Zim Bea 2250 74	DQ380209 Zim Bea 2250 74	DQ380143 Zim Bea 2250 74
1260/78	Zimbabwe	1978	DQ375418 Zim Sal 1260 78	DQ380214 Zim Sal 1260 78	DQ380164 Zim Sal 1260 78
1853/78	Zimbabwe	1978	DQ375424 Zim Sin 1853 78	DQ380220 Zim Sin 1853 78	DQ380168 Zim Sin 1853 78
2269/74	Zimbabwe	1974	DQ375434 Zim Sin 2269 74	DQ380222 Zim Sin 2269 74	DQ380173 Zim Sin 2269 74
MgH824	Madagascar	1979	DQ375414 Mad MgH824 79	DQ380210 Mad MgH824 79	DQ380144 Mad MgH824 79
200803166	Madagascar	1991	JF311372 Mad Ant 200803166 91	JF311381 Mad Ant 200803166 91	JF311390 Mad Ant 200803166 91
200803167	Madagascar	1991	JF311373 Mad Ant 200803167 91	JF311382 Mad Ant 200803167 91	JF311391 Mad Ant 200803167 91

bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

200803168	Madagascar	2008	JF311374 Mad Mia 200803168 08	JF311383 Mad Mia 200803168 08	JF311392 Mad Mia 200803168 08
200803169	Madagascar	2008	JF311375 Mad Ant 200803169 08	JF311384 Mad Ant 200803169 08	JF311393 Mad Ant 200803169 08
OS-9	Mauritania	1987	DQ375397 Mau OS-9 87	DQ380183 Mau OS-9 87	DQ380179 Mau OS-9 87
OS-8	Mauritania	1987	DQ375395 Mau OS-8 87	DQ380185 Mau OS-8 87	DQ380177 Mau OS-8 87
OS-1	Mauritania	1987	DQ375398 Mau OS-1 87	DQ380186 Mau OS-1 87	DQ380180 Mau OS-1 87
Hv-B375	Central African Republic	1985	DQ375422 CAR Mba Hv-B375 85	DQ380218 CAR Aba Hv-B375 85	DQ380161 CAR Hv-B375 85
CAR-R1622	Central African Republic	1985	DQ375423 CAR Ban R1622 85	DQ380219 CAR Ban R1622 85	DQ380160 CAR R1622 85
73HB1230	Central African Republic	1973	DQ375425 CAR 73HB1230 73	DQ380221 CAR 73HB1230 73	DQ380172 CAR 73HB1230 73
Zinga	Central African Republic	1969	DQ375419 CAR Zinga 69	DQ380217 CAR Zinga 69	DQ380167 CAR Zinga 69
Kenya 56 (IB8)	Kenya	1965	DQ375427 Ken-IB8 56	DQ380190 Ken-IB8 65	DQ380176 Ken 56-IB8 65
Kenya 57 (Rintoul)	Kenya	1951	DQ375431 Ken Rintoul 57	DQ380192 Ken Rintoul-57 51	DQ380155 Ken Rintoul-57 51
Kenya 9800523	Kenya	1998	DQ375400 Ken 9800523 98	DQ380196 Ken 9800523 98	DQ380169 Ken 9800523 98
Kenya 83 (21445)	Kenya	1983	DQ375402 Ken Rui 21445 83	DQ380198 Ken Rui 21445 83	DQ380171 Ken Rui 21445 83
2007004194	Kenya	2007	EU574004 Ken Kia 2007004194 07	EU574031 Ken Kia 2007004194 07	EU574057 Ken Kia 2007004194 07
2007004193	Kenya	2007	EU574005 Ken Nai 2007004193 07	EU574032 Ken Nai 2007004193 07	EU574058 Ken Nai 2007004193 07
2007003644	Kenya	2007	EU574006 Ken Bar 2007003644 07	EU574033 Ken Bar 2007003644 07	EU574059 Ken Bar 2007003644 07
2007002060	Kenya	2007	EU574013 Ken Nai	EU574039 Ken Nai	EU574066 Ken Nai 2007002060 07

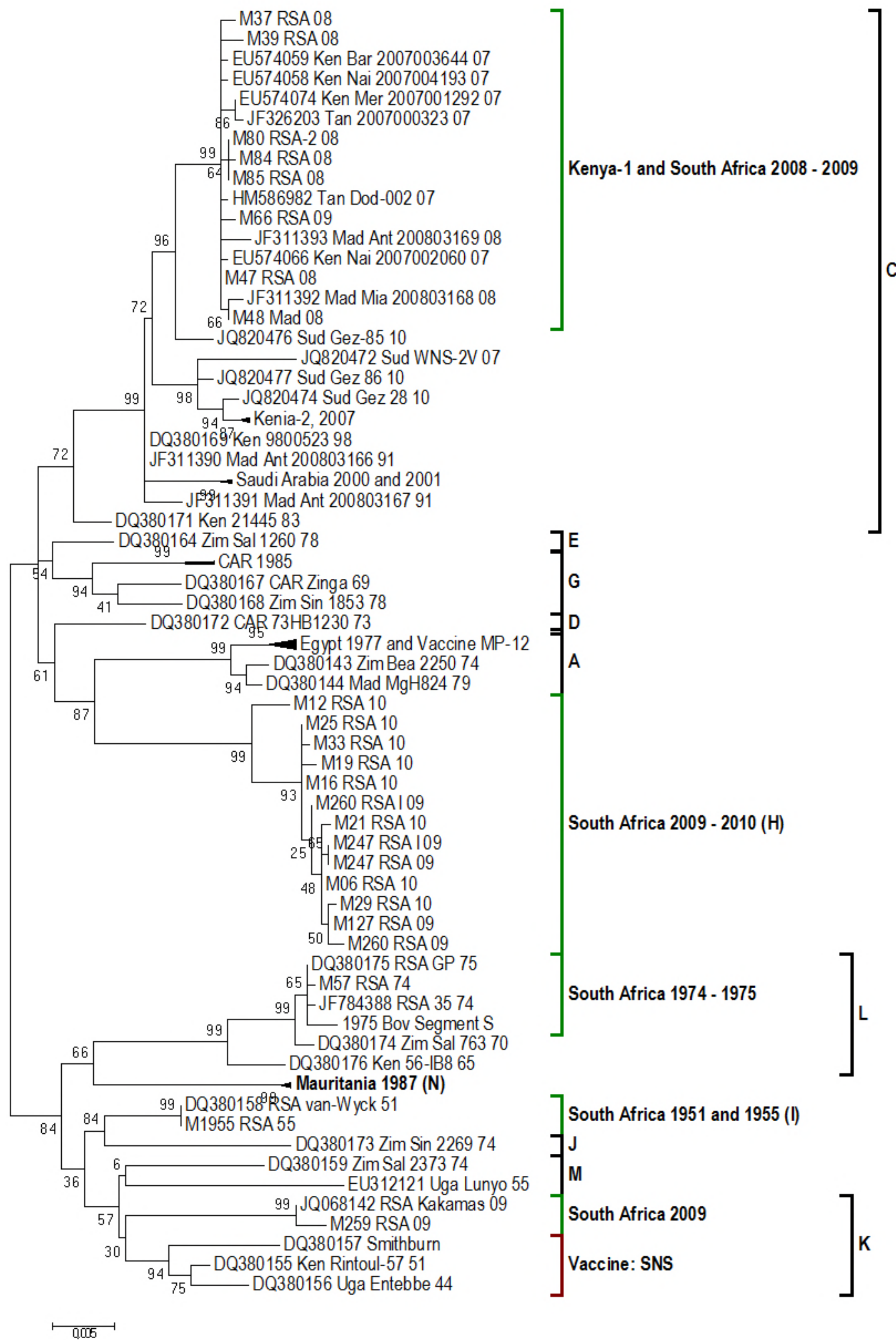
bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

			2007002060 07	2007002060 07	
2007001564	Kenya	2007	EU574017 Ken Mur 2007001564 07	EU574044 Ken Mur 2007001564 07	EU574072 Ken Mur 2007001564 07
2007001292	Kenya	2007	EU574019 Ken Mer 2007001292 07	EU574046 Ken Meru 2007001292 07	EU574074 Ken Mer 2007001292 07
Saudi 2000-10911	Saudi Arabia	2000	DQ375401 SA 10911 00	DQ380197 Saudi 10911 00	DQ380170 Saudi 10911 00
SA01-1322	Saudi Arabia	2001	KX096941 SA 1322 01	KX096942 Saudi 1322 01	KX096943 Saudi 1322 01
SA-75	South Africa	1975	DQ375428 RSA 75	DQ380189 RSA 75	DQ380175 RSA 75
SA-51 (Van Wyck)	South Africa	1951	DQ375433 RSA VanWyck 51	DQ380195 RSA VanWyck 51	DQ380158 RSA VanWyck 51
M35/74	South Africa	1974	JF784386 RSA M35 74	JF784387 RSA M35 74	JF784388 RSA M35 74
Kakamas	South Africa	2009	JQ068144 RSA Kakamas 09	JQ068143 RSA Kakamas 09	JQ068142 RSA Kakamas 09
Sudan 85-2010	Sudan	2010	JQ820485 Sud Gez-85 10	JQ820488 Sud Gez-85 10	JQ820476 Sud Gez-85 10
Sudan 86-2010	Sudan	2010	JQ820484 Sud Gez 86 10	JQ820489 Sud Gez 86 10	JQ820477 Sud Gez 86 10
Sudan 2V-2007	Sudan	2007	JQ820483 Sud WNS-2V 07	JQ820490 Sud WNS 2V 07	JQ820472 Sud WNS-2V 07
Sudan 28-2010	Sudan	2010	JQ820486 Sud Gez-28 10	JQ820491 Sud Gez 28 10	JQ820474 Sud Gez 28 10
TAN/Tan-001/07	Tanzania	2007	HM586959 TAN Tan-001 07	HM586970 Tan 001 07	HM586981 Tan Tan-001 07
TAN/Dod-002/07	Tanzania	2007	HM586960 Tan Dod-002 07	HM586971 Tan 002 07	HM586982 Tan Dod-002 07
Tan 2007000323	Tanzania	2007	JF326189 Tan 2007000323 07	JF326194 Tan 2007000323 07	JF326203 Tan 2007000323 07
Entebbe	Uganda	1944	DQ375429 Uga Entebbe 44	DQ380191 Uga Entebbe 44	DQ380156 Uga Entebbe 44
Smithburn	Uganda	1944	DQ375430 Smithburn	DQ380193 Smithburn	DQ380157 Smithburn
Lunyo	Uganda	1955	KU167027 Uga Lunyo 55	KU167026 Uga Lunyo 55	EU312121 Uga Lunyo 55

492

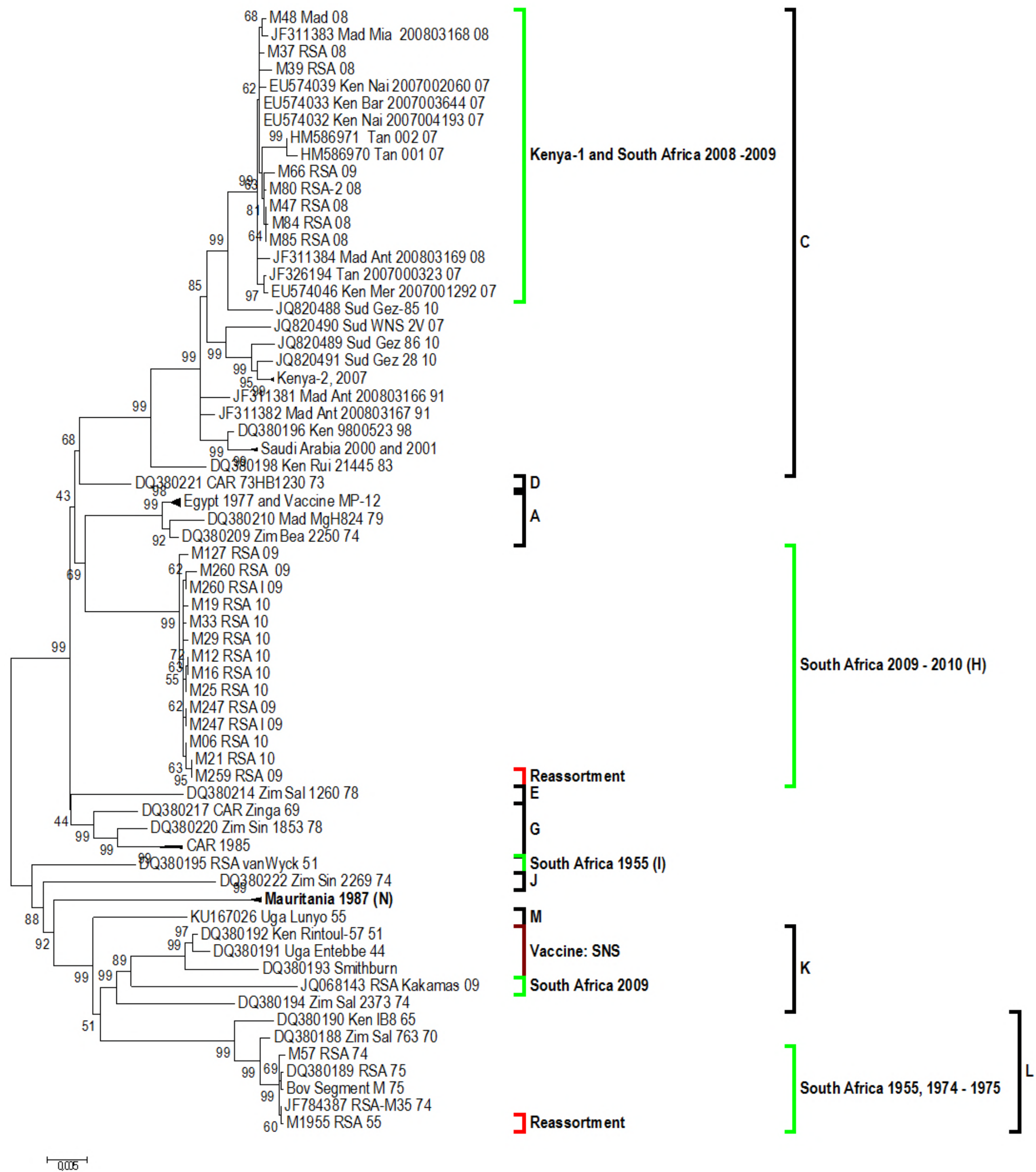
bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Fig2A.



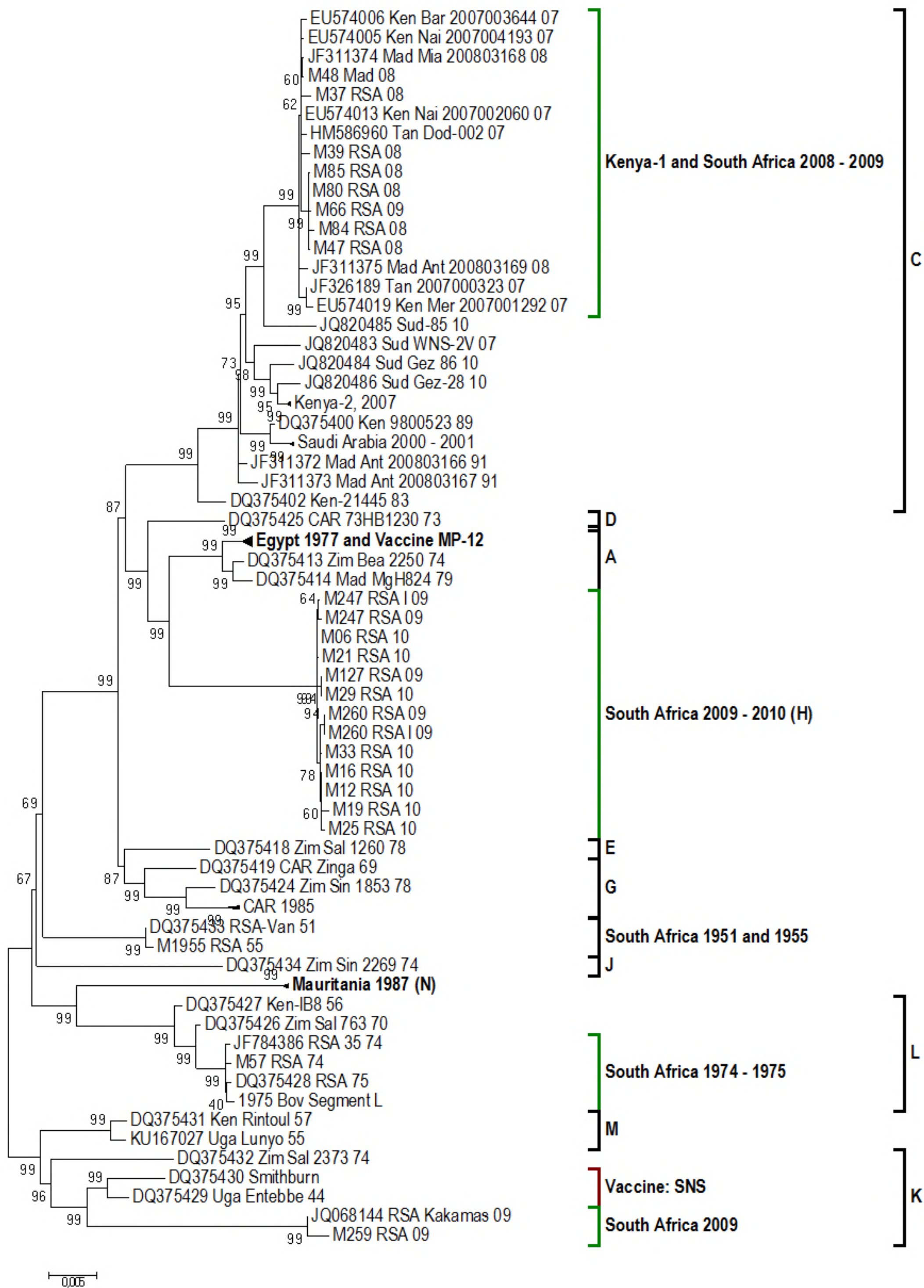
bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Fig2B.



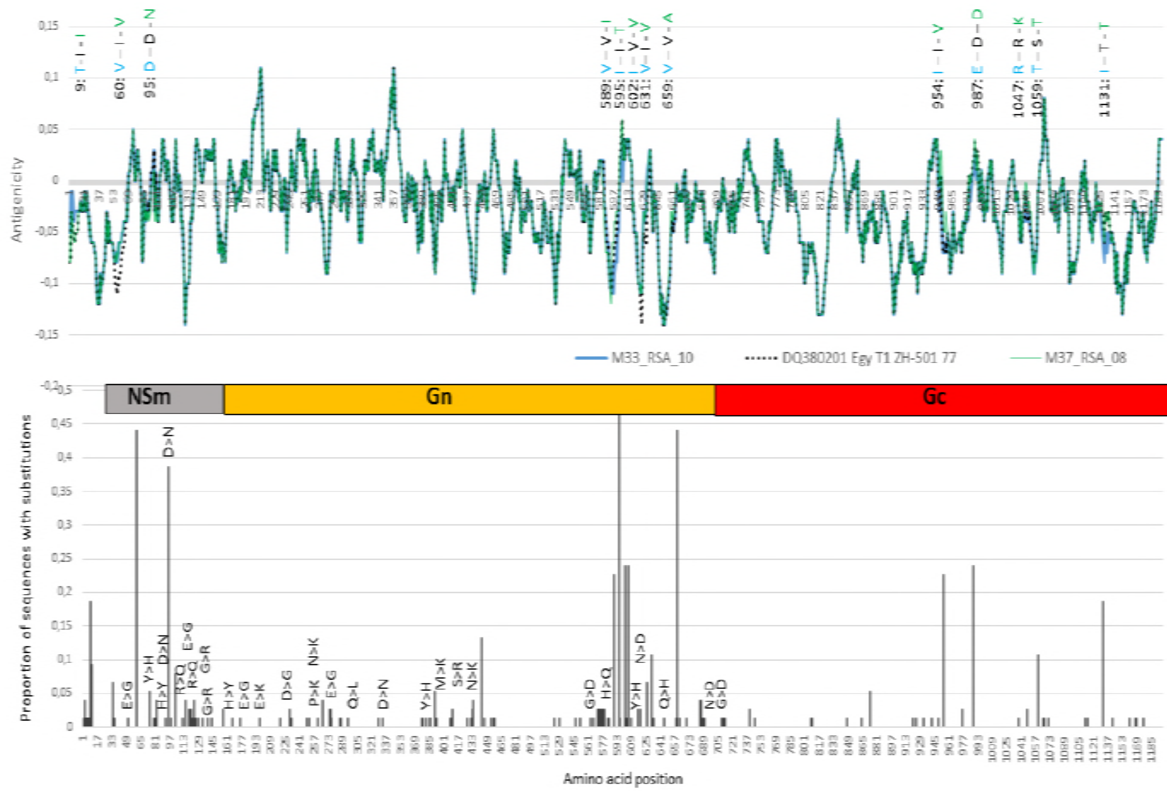
bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Fig2C.



bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Fig3.



bioRxiv preprint doi: https://doi.org/10.1101/338251; this version posted June 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

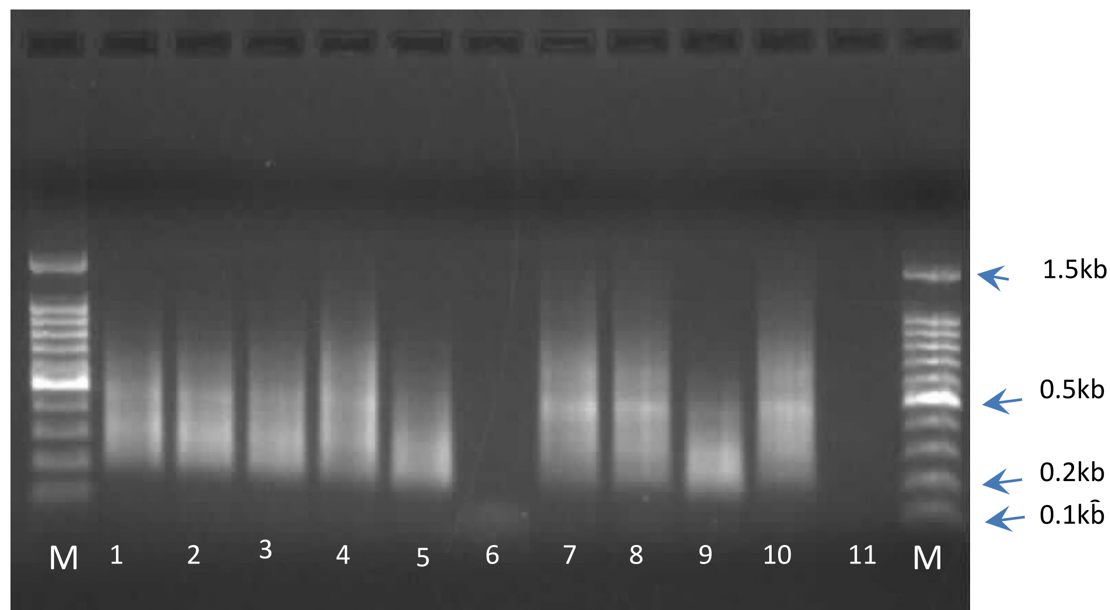


Figure 1.