

# Complete Genome Analysis of 33 Ecologically and Biologically Diverse Rift Valley Fever Virus Strains Reveals Widespread Virus Movement and Low Genetic Diversity due to Recent Common Ancestry<sup>∇</sup>

Brian H. Bird,<sup>1,3</sup> Marina L. Khristova,<sup>2</sup> Pierre E. Rollin,<sup>1</sup> Thomas G. Ksiazek,<sup>1</sup> and Stuart T. Nichol<sup>1\*</sup>

*Special Pathogens Branch, Division of Viral and Rickettsial Diseases, National Center for Infectious Diseases,<sup>1</sup> and Biotechnology Core Facility Branch,<sup>2</sup> Centers for Disease Control and Prevention, 1600 Clifton Road MS G-14, Atlanta, Georgia 30329, and University of California, Davis, School of Veterinary Medicine, Davis, California 95616<sup>3</sup>*

Received 25 September 2006/Accepted 18 December 2006

Rift Valley fever (RVF) virus is a mosquito-borne RNA virus responsible for large explosive outbreaks of acute febrile disease in humans and livestock in Africa with significant mortality and economic impact. The successful high-throughput generation of the complete genome sequence was achieved for 33 diverse RVF virus strains collected from throughout Africa and Saudi Arabia from 1944 to 2000, including strains differing in pathogenicity in disease models. While several distinct virus genetic lineages were determined, which approximately correlate with geographic origin, multiple exceptions indicative of long-distance virus movement have been found. Virus strains isolated within an epidemic (e.g., Mauritania, 1987, or Egypt, 1977 to 1978) exhibit little diversity, while those in enzootic settings (e.g., 1970s Zimbabwe) can be highly diverse. In addition, the large Saudi Arabian RVF outbreak in 2000 appears to have involved virus introduction from East Africa, based on the close ancestral relationship of a 1998 East African virus. Virus genetic diversity was low (~5%) and primarily involved accumulation of mutations at an average of  $2.9 \times 10^{-4}$  substitutions/site/year, although some evidence of RNA segment reassortment was found. Bayesian analysis of current RVF virus genetic diversity places the most recent common ancestor of these viruses in the late 1800s, the colonial period in Africa, a time of dramatic changes in agricultural practices and introduction of nonindigenous livestock breeds. In addition to insights into the evolution and ecology of RVF virus, these genomic data also provide a foundation for the design of molecular detection assays and prototype vaccines useful in combating this important disease.

Rift Valley fever (RVF) virus causes large explosive epidemics of animal and human illness throughout Africa and, more recently, the Arabian Peninsula. Case reports of an illness of sheep consistent with RVF disease were first reported in 1910 in western Kenya (30, 41, 56). Transmitted primarily by mosquitoes, RVF virus was first isolated in 1931 in Kenya as the agent of enzootic hepatitis of sheep (20). RVF virus epizootics are characterized by large sweeping “abortion storms” and mortality in livestock (primarily sheep and cattle), with newborn mortality approaching 100% (10, 15, 50). Human infections typically occur due to the bites of infected mosquitoes or percutaneous or aerosol exposure during the handling of aborted fetal materials or the slaughtering of diseased animals. In most human cases, the disease is characterized by a self-limiting febrile illness (2 to 5 days), which progresses to more serious complications in 1 to 2% of infected individuals. These include hepatitis, encephalitis, retinitis, blindness, or a hemorrhagic syndrome, with a hospitalized case fatality of 10 to 20% (37, 38, 40). RVF virus epidemics can greatly strain the ability

of public health infrastructures to provide adequate medical care, given the large numbers of infected individuals.

The ability of RVF virus to cross geographic or national boundaries increases the importance of understanding the basic RVF ecology and genomic diversity. RVF virus was first isolated outside of continental Africa in Madagascar in 1979 and has since become endemic there (43). RVF virus has at least twice been responsible for large “virgin-soil” epidemics. In 1977, RVF virus was recorded for the first time north of the Sahara desert in Egypt and resulted in a massive epizootic/epidemic, during which more than 200,000 people were estimated to have been infected (39). Later, in 2000, the virus was isolated for the first time outside of Africa across the Red Sea in Saudi Arabia and Yemen (9). The potential for further introductions of RVF virus into previously unaffected countries via importation of infected livestock or mosquito translocation, or through intentional release, illustrates the need for safe and effective veterinary and human vaccines and broadly based pan-RVF virus real-time molecular diagnostic assays.

The RVF virus genome is comprised of three negative-sense, single-stranded RNA genomic segments with a total combined length of approximately 11.9 kb (45). Like other members of the genus *Phlebovirus*, the small (S) ambisense segment possesses one open reading frame (ORF) coding for the nucleoprotein (NP; 27 kDa) in the antigenomic strand and one coding for the nonstructural (NSs; 31-kDa) protein in the genomic strand (45). These two regions are separated by a

\* Corresponding author. Mailing address: Special Pathogens Branch, Division of Viral and Rickettsial Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, 1600 Clifton Road MS G-14, Atlanta, GA 30329. Phone: (404) 639-1115. Fax: (404) 639-1118. E-mail: snichol@cdc.gov.

<sup>∇</sup> Published ahead of print on 27 December 2006.

poly(C)-rich intergenic region (in the genomic sense) of approximately 81 nucleotides (nt). The NSs protein has been demonstrated to function in the down regulation of RNA polymerase II activity, resulting in host cell transcription shutoff and, via this mechanism, to cause antagonism of host cell interferon responses (6, 33). Recent evidence suggests that both genomic and antigenomic sense templates of the S segment can be packaged within viral particles, allowing the direct transcription/translation of this major *in vivo* virulence factor without requiring polymerase-dependent virus replication (27). The medium (M) segment encodes at least four viral proteins in a single ORF: the 14-kDa NSm of unknown function, two major envelope surface glycoproteins (the 55-kDa Gn and 58-kDa Gc), and a 78-kDa fusion of the NSm and Gn proteins (57). We and others recently demonstrated that the RVF virus NSm proteins were dispensable for virus growth in cell culture (22, 60). The large (L) segment encodes the 237-kDa virus polymerase in a single 6.4-kb ORF (44).

Here, we report an extensive analysis of the complete genomes of 33 diverse RVF viruses. These data are invaluable for the identification of highly conserved regulatory or catalytic domains that can be targeted by a reverse-genetics approach for the development of safe and effective veterinary and human vaccines and can provide a basis for further investigations into the molecular determinants of RVF virus pathogenesis in animal models.

#### MATERIALS AND METHODS

**RNA isolation and RT-PCR.** VeroE6 infected cell supernatants were combined with TriPure (Roche, Indianapolis, IN) at 1:10 and incubated at room temperature for 10 min. Tubes containing RNA preparations were surface decontaminated and transferred to a biosafety level 3 laboratory, and RNA was extracted using the RNaid Kit (Obiogene, Carlsbad, CA). The RNA was resuspended in 50  $\mu$ l H<sub>2</sub>O. The RNA (2  $\mu$ l) was reverse transcriptase (RT)-PCR amplified using SuperScript III with Platinum *Taq* High Fidelity (Invitrogen, San Diego, CA). The optimized cycle parameters for the S segment were 1 cycle of 51°C for 30 min and 94°C for 2 min, followed by 40 cycles of 94°C for 15 seconds, 56°C for 30 seconds, and 68°C for 2 min, followed by a final extension at 68°C for 5 min. For the M and L segments, the parameters were 1 cycle of 51°C for 30 min and 94°C for 2 min, followed by 40 cycles of 94°C 15 seconds, 56°C 30 seconds, and 68°C for 4 min, followed by a final extension at 68°C for 10 min. The entire S and M segments were amplified in one piece, utilizing the following RT-PCR primers: RVFS-AFwd, 5'-ACACAAAGTCCCTAGAGATAC-3', and RVFS-ARev, 5'-ACACAAAGACCCCTAGTG-3'; RVFM-AFwd, 5'-ACACAAAGACGGTGC-3', and RVFM-ARev, 5'-ACACAAAGACCGTGC-3'. Amplification of the L segment required two overlapping sections (LA and LB); these were amplified using RVFL-AFwd, 5'-ACACAAAGGCCCAATC-3', and RVFL-3482Rev, 5'-GGAAGCATATAGCTGCGG-3' (LA region), and RVFL-2845Fwd, 5'-GAGACAATAGCCAGTGC-3', and RVFL-ARev, 5'-ACACAAAGACCCCAATATTG-3' (LB region).

**DNA purification and sequencing.** RT-PCR products were purified using the QiaEX II gel purification kit (QIAGEN, Valencia, CA) and sequenced by primer walking both strands, using ABI Big-Dye 3.1 dye chemistry and ABI 3730XL automated DNA sequencers (Applied Biosystems, Foster City, CA). For each segment, the nucleotide sequences were aligned, and sequencing primers representing the most conserved regions were designed for coverage of the entire segment. For sequencing, totals of 21, 52, and 50 sequencing primers were used for the S, M, and L segments, respectively. Some additional strain-specific sequencing primers were used to cover difficult-to-sequence regions. Approximately 140 to 150 reads were obtained for each genome, resulting in an average six-fold redundancy at each base position. Chromatogram data were assembled and analyzed using Seqmerge (GCG Wisconsin Package [10.3]; Accelrys, San Diego, CA), Phred/Phrap, and Consed software (18, 19, 25).

**Sequence analyses.** Datasets were analyzed using GCG, BIOEDIT version 5.0.6 (North Carolina State University, Raleigh, NC), CLUSTAL-W, and the PAUP\* program, version 4.0b10 (Sinauer Associates, Inc., Sunderland, MA).

For phylogenetic analysis, appropriate nucleotide substitution models were selected using Modeltest/Modelscore (49). Further analyses utilized the Bayesian analysis software package, BEAST, BEAUTi, and Tracer (14). Evidence of intrasegmental RNA recombination was analyzed using either individual protein ORF datasets to examine ORF-specific changes in tree topology within PAUP\* or with the SIMPLOT v.3.5.1 software package (RaySoft, Baltimore, MD) (35). BootScan and similarity plot analyses, which allow genomewide screening of multiple virus genomes for recombination based on a sliding window of defined length, were carried out utilizing a 200-nt window and a 10-nt step interval. Mean nonsynonymous ( $d_N$ ) and synonymous ( $d_S$ ) substitutions per site (ratio,  $d_N/d_S$ ) were calculated for each gene data set using the SLAC method within the HYPHY software package (32).

#### RESULTS

We report here the successful development of a high-throughput technique for rapidly generating complete S, M, and L segment RT-PCR fragments and the complete genome sequence of the entire genetic spectrum of RVF virus represented by 33 RVF virus strains. Virus diversity was found to be relatively low, with identity differences of only approximately 5% and 2% at the nucleotide and amino acid levels, respectively. Comparison of the single previously available complete genome sequence (GenBank entries NC\_002045, NC\_002044, and NC\_002043) for RVF virus strain MP12 (derived from ZH548) revealed a number of unique nonsynonymous substitutions and insertions and deletions at various positions across the genome relative to the parental ZH548 virus and all others sequenced here. In addition, the L segment of MP12 (GenBank entry NC\_002045) is 202 nt longer than that of the parental ZH548 reported here; the ORF is also longer, and extended regions are out of frame relative to all of the other RVF viruses (44). Correction of the MP12 L segment sequence and length to 6,404 nt was reported in a 1994 paper (44), which suggests that many of these sequence differences likely reflect earlier sequencing inaccuracies rather than authentic mutations accumulated during passage of the MP12 virus.

**S segment nucleotide and deduced NP and NSs protein sequences.** All 33 RVF virus S segments were found to be highly conserved. The maximum pairwise identity differences were 4% and 1% at the nucleotide and the deduced amino acid levels, respectively. A remarkable lack of significant length variation was observed, with virus S segments differing by only 2 nucleotides (1690 to 1692). These insertions and deletions were in the intergenic region between the N and NSs ORFs, a region that exhibited higher sequence divergence (11%) than was observed in the N and NSs ORFs (~4%). While no variation was found in the deduced amino acid length of the NP (245 aa) or NSs (264 aa) protein, the NSs gene was observed to be slightly more variable than the N gene (4.5% versus 3.5%), presumably reflecting less constraint on the evolution of the nonstructural protein.

**M segment nucleotide and deduced envelope glycoprotein (NSm, Gn, and Gc) sequences.** Our expectation was that considerable diversity would be found among virus M RNA segments, which encode the surface glycoproteins. Positive selective pressure might be expected to mold the evolutionary pattern observed, given that these glycoproteins are the targets for neutralizing antibodies and influence virus cell attachment and tissue tropism. However, a remarkably high level of conservation at the nucleotide and deduced amino acid levels was found among the 33 strains included in this study. No variation

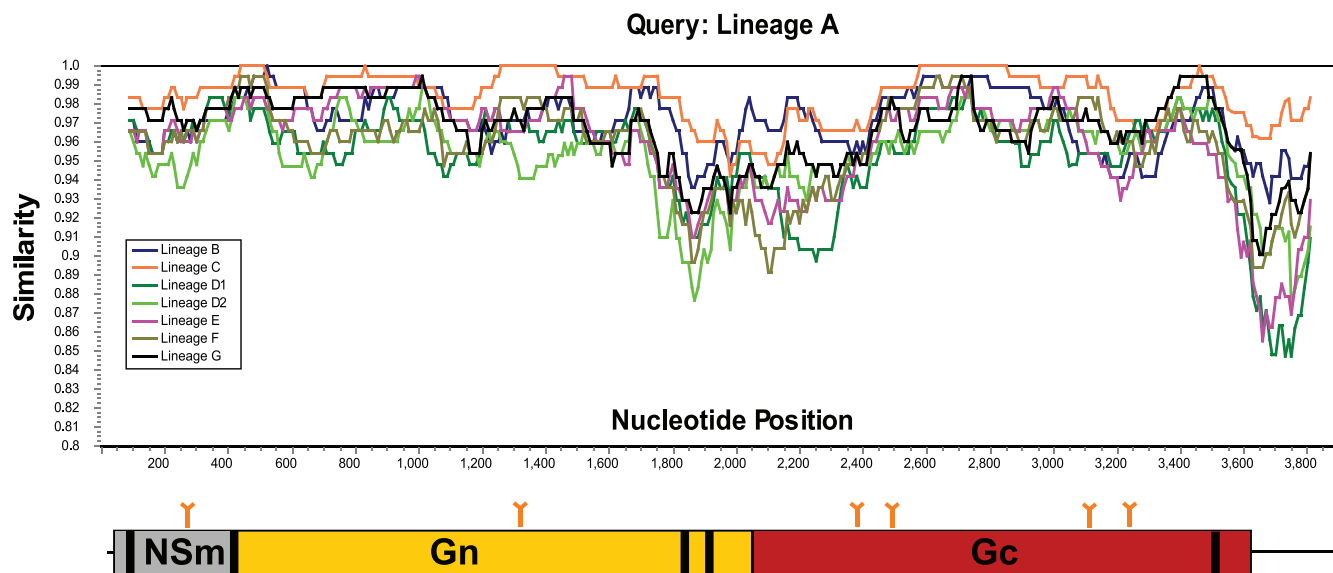


FIG. 1. Simplot analysis of 33 RVF virus complete M segment nucleotide alignments grouped by virus lineage (A to G), indicating nucleotide positions (antigenomic sense) along the *x* axis and, on the *y* axis, percent nucleotide similarity. Depicted beneath the *x* axis is a cartoon representation of the 5' UTR; the NSm, Gn, and Gc coding regions; and the 3' UTR encoded by the RVF virus M segment. Potential glycosylation sites are indicated above by arrows. Predicted hydrophobic transmembrane domains are depicted by heavy black lines. Simplot analyses (sliding window, 180 nt; step, 20 nt) revealed regions of greater nucleotide diversity, approaching 20% at the carboxy terminus of Gn and the amino terminus of Gc and at the 3' UTR.

in length was detected for the virus M RNA segments (3,885 nt) or encoded polypeptides (1,197 aa). In addition, overall pairwise sequence identity differences were only approximately 5% and 2% at the nucleotide and deduced amino acid levels, respectively. The greatest nucleotide variation was found in the 290-nt untranslated region (UTR) located at the 3' end of the antigenomic RNA. The nucleotide divergence in this 3' UTR approached 15%, as determined by a sliding 200-nt window analysis using Simplot (Fig. 1). In addition, the central ORF region encoding the Gn carboxy terminus and the Gc amino terminus represents another variable region within the M segment (Fig. 1). Putative polyprotein glycosylation sites exist in the NSm protein region prior to Gn (aa 88) and within Gn (aa 438) and Gc (aa794, 829, 1035, and 1077). Experimental data have shown glycosylation at amino acids 88 and 438 and at least three of the positions in Gc (31). Consistent with their usage and a functional role, all of these positions are completely conserved.

A notable feature of the RVF virus M segment is the presence of five in-frame AUG-methionine start codons within the NSm protein coding region at antigenomic sense positions 21, 135, 174, 411, and 426 (11, 23, 57). These multiple in-frame AUGs are responsible for at least two forms of glycoprotein (a 14-kDa NSm protein and a 78-kDa NSm and Gn fusion polypeptide), in addition to the expected Gn and Gc envelope glycoproteins. The formation of these proteins is dependent on the in-frame AUG that is utilized. All 33 RVF virus M segments analyzed retained all five in-frame AUG codons. Interestingly, despite the finding of the dispensability of the NSm proteins for productive infection of tissue culture cells, this conservation suggests a functional role *in vivo* for the 14-kDa NSm and 78-kDa NSm-Gn fusion glycoproteins (22, 60). In addition, the conservation of the third in-frame AUG codon in

all 33 natural RVF virus strains is surprising, given the earlier data using the RVF virus prototype strain ZH-501, which indicated that no protein was found to initiate from the third AUG (57). This suggests that although previously undetected, there may be an evolutionarily conserved role for an as-yet-unknown protein arising from this third translation start codon or, perhaps more likely, a functional requirement for a methionine at that position.

**L segment nucleotide and deduced polymerase protein sequences.** As expected, the L segments of the 33 RVF virus strains in this study demonstrated the highest levels of conservation at both the nucleotide and deduced amino acid levels compared with the S and M segments. No variation in L segment nucleotide (6,404-nt) or amino acid (2,092-aa) length was found. The maximum overall pairwise differences were 4% and 1% at the nucleotide and amino acid levels, respectively. The deduced amino acid sequences for several proposed functional motifs conferring RNA polymerase activity were found to be completely conserved in all 33 RVF virus strains analyzed (4, 44).

**Insights into RVF ecology from analyses of S, M, and L segment sequences.** A diverse array of RVF virus isolates were included in this analysis, including strains from throughout the known geographic range of the virus and spanning 56 years from 1944 to 2000. They also included virus isolates collected from mosquitoes, cattle, sheep, bats, and humans and from both endemic and epizootic/epidemic settings. In addition, some isolates were known to exhibit different pathogenic features in animal models. Extensive phylogenetic analyses using unconstrained maximum likelihood (ML) and Bayesian statistical (BEAST) methods were completed. For each segment, the ML tree and the maximum a posteriori (MAP) tree topologies were robust and congruent. ML bootstrap values (500

replicates) (data not shown) and Bayesian posterior support values were found to be comparable at each major tree node. The MAP trees for the S, M, and L segments are depicted in Fig. 2, 3, and 4. These analyses show that the RVF viruses can be separated into seven distinct genetic lineages (A to G).

RVF virus strains with diverse geographic origins can be found in each lineage, which is indicative of widespread dispersal and movement of RVF virus genotypes throughout Africa. Lineage A, which is composed primarily of isolates collected from the 1977 to 1979 Egyptian RVF virus epizootic/epidemic, also contains an isolate obtained from Zimbabwe in 1974 and an isolate from Madagascar from 1979. Lineage B contains viruses from Kenya, the Central African Republic (CAR), and Saudi Arabia. A clear genetic link can be seen with all three complete segments between the RVF virus introduced into the Arabian Peninsula in 2000 and a representative virus isolate collected in Kenya during an extremely large RVF outbreak that occurred throughout eastern Africa in 1997 and 1998. This reinforces the earlier conclusion that the RVF outbreak in Saudi Arabia and Yemen involved introduction of the virus from an area of eastern Africa involved in the massive outbreak 2 years earlier (55). Also within lineage B is an earlier Kenyan 1983 isolate and a strain from the CAR collected in 1973. Lineage C is a large heterogeneous collection of strains from Guinea, CAR, and Zimbabwe spanning the years 1969 to 1984. Similarly heterogeneous, lineage D contains RVF virus strains from Mauritania, Burkina Faso, Kenya, Zimbabwe, and South Africa spanning 1965 to 1987. Lineage E contains the classic Ugandan Entebbe strain 1944 and a strain from Zimbabwe isolated in 1974. Lineage F is comprised of a single Zimbabwe strain collected in 1974, and lineage G is composed of a single strain collected in South Africa in 1951.

No distinct mutually exclusive correlation of virus genotype and geographic origin was seen. While representatives from one geographic area do tend to cluster together within each RVF virus lineage, strains from geographically distant areas can be found within each lineage. This indicates that widespread geographic dispersal of genotypes has occurred across continental Africa. This movement suggests that the viral ecology of RVF virus is dynamic and that multiple introductions of virus genotypes have occurred as these viruses have evolved from their most recent common ancestor.

**Evidence of RNA segment reassortment.** The overall topologies of the S, M, and L trees were remarkably congruent, suggesting that virus RNA segment reassortment is not a common occurrence. However, detailed analysis revealed a limited number of incongruencies, indicative of reassortment events. The S and M segments of strain 73HB1230 are firmly placed (with 98% posterior support) in lineage B; however, the L segment moves (with 100% support) into lineage A, suggesting that this virus represents an L segment reassortant. In addition, lineage B is monophyletic with lineage A for the M and L segments (62% and 94% support) but monophyletic with lineage C (96% support) for the S segment. This is suggestive of a more ancestral S segment reassortment event. Strong support was also found for a reassortment event occurring with the M segments of lineages D and E. Lineage D contains two subgroups: a West African group (D1) and a group comprised of strains collected in southern Africa and Kenya (D2). While these two subgroups are monophyletic based on S and L seg-

ment tree topology, the M segment of the southern Africa/Kenya group (D2) is monophyletic with lineage E viruses. Together, these findings indicate that coinfections with differing RVF viral genotypes do occur and have resulted in the generation of reassortant viruses multiple times over the evolutionary history of these virus strains.

**Overall lack of evidence for RNA recombination.** Various detailed tests of recombination were completed on all 33 S, M, and L full genome segments, including similarity plotting, bootscanning, and informative site analyses. For all 33 M and L segments, no statistically significant evidence of recombination was found (data not shown). There was suggestive evidence of a possible S segment recombination event involving a 100-nt region of the NSs coding region between RVF virus strains 2269/74 and ZH-501 (data not shown); however, this evidence did not alter the overall S segment phylogeny.

**Molecular evolutionary rates and time to most common recent ancestor (TMRCA).** Molecular evolutionary rates (nucleotide substitutions per site per year) were estimated by Bayesian analysis of sequence differences among the S, M, and L segments of the 33 RVF virus strains. Independent estimation of the molecular evolutionary rates of the RVF virus S, M, and L segments revealed rates very similar to one another and to those observed for other negative-sense single-stranded RNA viruses (13, 29). For each segment, a mean rate and a 95% high posterior distribution (shown in parentheses) were calculated, and values of  $2.35 \times 10^{-4}$  ( $1.28 \times 10^{-4}$  to  $3.4 \times 10^{-4}$ ),  $2.42 \times 10^{-4}$  ( $1.8 \times 10^{-4}$  to  $3.0 \times 10^{-4}$ ), and  $2.78 \times 10^{-4}$  ( $2.0 \times 10^{-4}$  to  $3.5 \times 10^{-4}$ ) nt substitutions per site per year were obtained for the S, M, and L segments, respectively. Not surprisingly, the highest nucleotide substitution rates in the genomic segments were found when the data were limited to the third position of each amino acid codon, with mean evolutionary rates of  $6.5 \times 10^{-5}$ ,  $2.7 \times 10^{-5}$ , and  $6.3 \times 10^{-4}$  nt substitutions per site per year for the first, second, and third codon positions, respectively, or the 3' and 5' UTRs flanking each ORF (data not shown).

In a broad sense, the most recent common ancestor of any collection of genomes is defined as the most recently occurring ancestral genotype that is the progenitor of that group. Utilizing a Bayesian statistical approach, the TMRCA of the 33 RVF virus genotypes was estimated. These calculations rely on the date of collection of each viral-strain genotype, the estimated molecular evolutionary rate, and the overall genomic diversity to provide an estimate of the TMRCA, or "root height," measured in years prior to the most contemporary virus in the data set. Surprisingly, Bayesian analyses revealed that the TMRCA for these 33 RVF virus strains suggests a relatively recent ancestry for this sample of the viruses identifiable today as RVF virus. For each segment, the values of the overall root height coalesced toward mean values of 110 to 120 years prior to the year 2000, i.e., approximately 1880 to 1890. The mean TMRCA and (in parentheses) 95% high posterior distribution values for the S, M, and L segments were as follows: 108.6 (77.5 to 149.9), 117.3 (95.0 to 143.0), and 112.5 (93.2 to 134.3) years before the year 2000, respectively. These findings suggest that the most recent common ancestor of this collection of RVF viruses existed sometime in the 1880s, with an earliest calculated date of approximately 1850.

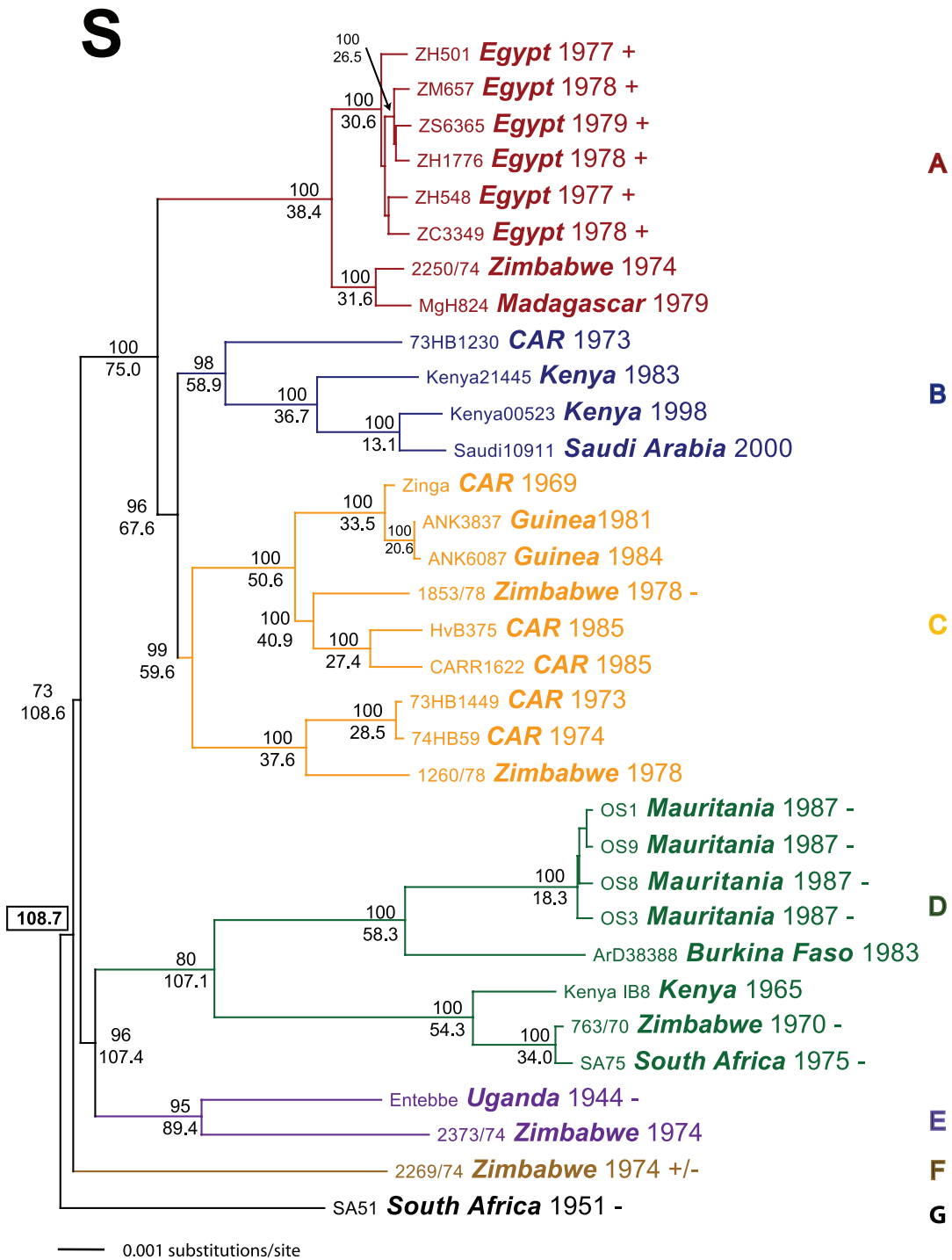


FIG. 2. Thirty-three complete RVF S segments were analyzed by either unconstrained ML techniques with 500 replicate bootstrap values (PAUP\*b10) or the Bayesian statistical program (BEAST) with an MCMC chain length of  $3.0 \times 10^7$ , a  $3.0 \times 10^6$  burn in, a GTR+ $\Gamma$ +I nucleotide rate substitution model, a strict molecular clock, and sampling of every 1,000 states. A strict molecular clock was chosen after multiple analyses; using both relaxed exponential and relaxed logarithmic clock models revealed no significant deviation in the coalescence of the overall evolutionary rate or tree topologies. Trees and support values from the ML and BEAST analyses were similar. The Bayesian MAP log likelihood value tree was chosen from the posterior distribution and is depicted here. Posterior support values were calculated from consensus analysis of all Bayesian posterior trees, with values over 0.5 indicated above each node. The estimated TMRCA of the tree nodes was calculated by Bayesian analyses and is reported below each node as years prior to the year 2000. The overall TMRCA of the entire S segment analysis is indicated in a text box located adjacent to the root node of SA51. Each taxon name indicates the strain, country of origin, and date of isolation. The GenBank accession numbers for the virus S segments are DQ380143 to -6, DQ380149, DQ380151 to -3, DQ380156, and DQ380158 to -81. Strains used in previous studies of virulence in WF rats are indicated with either + (lethal; LD<sub>50</sub>,  $\sim 1.0$  PFU), +/-, (lethal; LD<sub>50</sub>,  $\sim 2 \times 10^3$  PFU), or - (nonlethal).

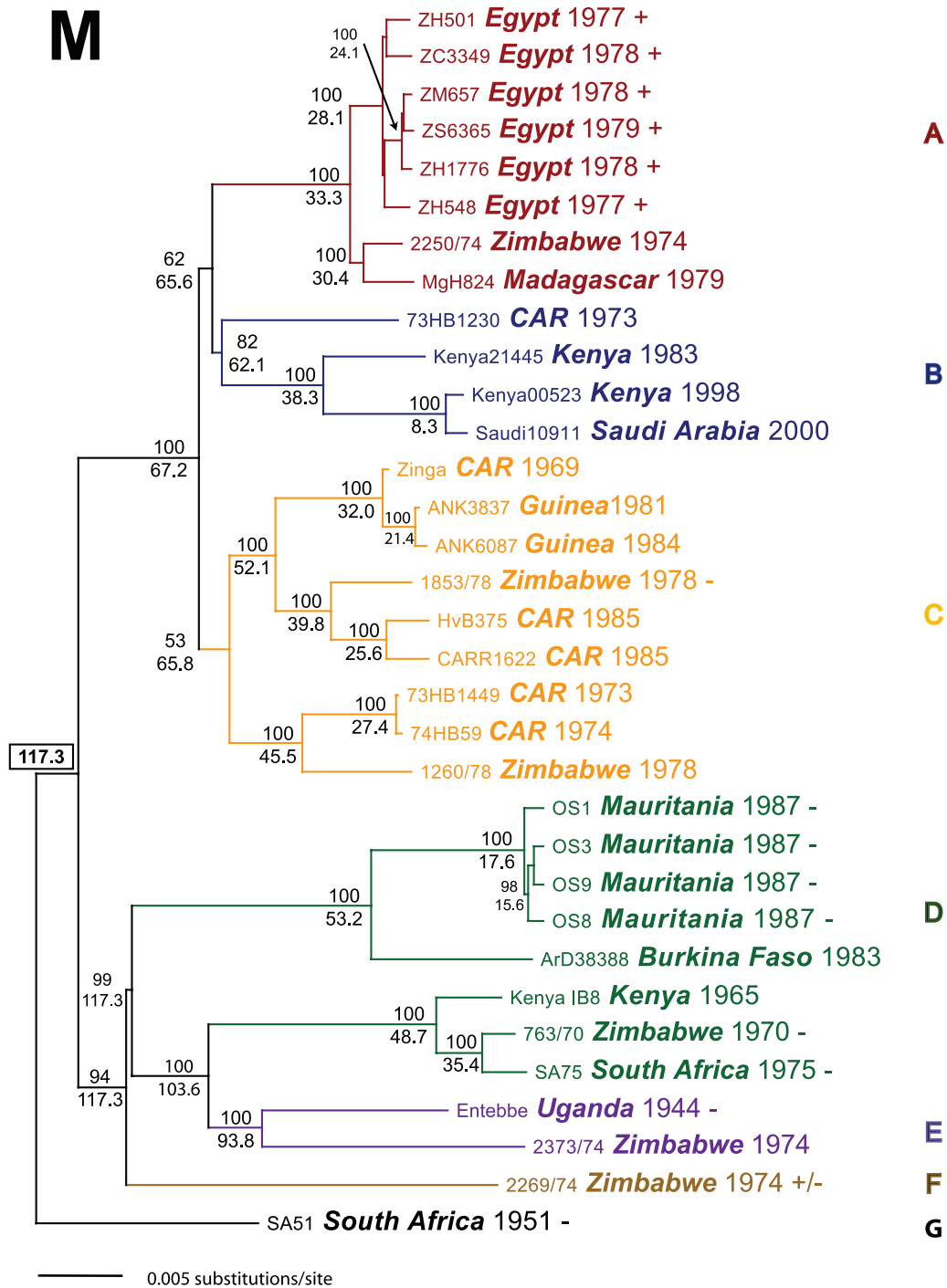


FIG. 3. Thirty-three complete RVF M segments were analyzed by either unconstrained ML techniques with 500 replicate bootstrap values (PAUP\*b10) or the Bayesian statistical program (BEAST) with an MCMC chain length of  $3.0 \times 10^7$ , a  $3.0 \times 10^6$  burn in, a GTR+ $\Gamma$ +I nucleotide rate substitution model, a strict molecular clock, and sampling of every 1,000 states. Trees and support values from the ML and BEAST analyses were similar. The Bayesian MAP log likelihood value tree was chosen from the posterior distribution and is depicted here. Posterior support values were calculated from consensus analysis of all Bayesian posterior trees, with values over 0.5 indicated above each node. The estimated TMRCAs of tree nodes was calculated by Bayesian analyses and is reported below each node as years prior to the year 2000. The overall TMRCAs of the entire M segment analysis is indicated in a text box located adjacent to the root node of SA51. Each taxon name indicates the strain, country of origin, and date of isolation. GenBank accession numbers for the virus M segments are DQ380183 to -91, DQ380194 to -8, DQ380200, DQ380203 to -7, DQ380209 to -12, and DQ380214 to -22. Strains used in previous studies of virulence in WF rats are indicated with either + (lethal; LD<sub>50</sub>, ~1.0 PFU), +/- (lethal; LD<sub>50</sub>, ~2 × 10<sup>3</sup> PFU), or - (nonlethal).

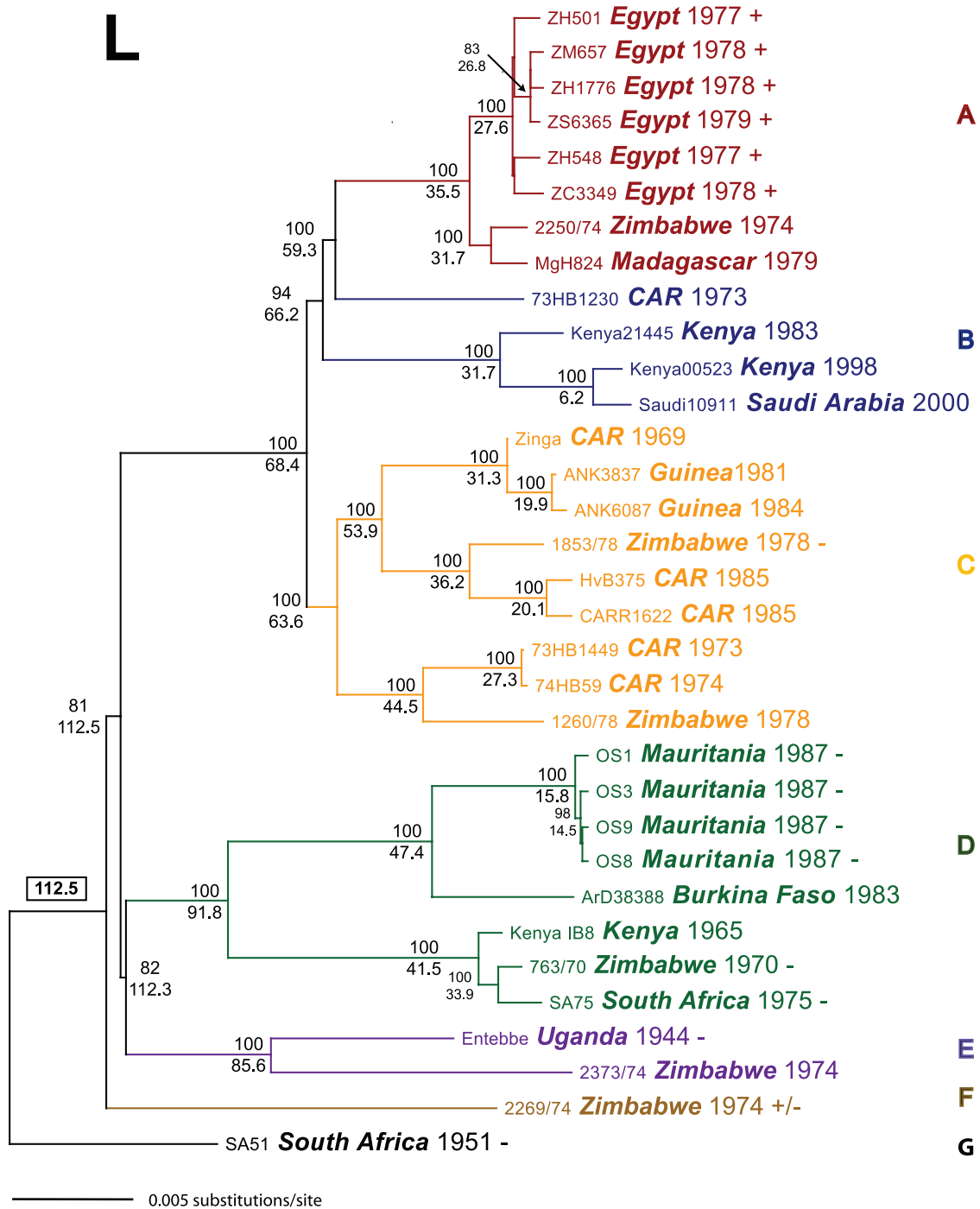


FIG. 4. Thirty-three complete RVF L segments were analyzed by either unconstrained ML techniques with 500 replicate bootstrap values (PAUP\*b10) or the Bayesian statistical program (BEAST) with an MCMC chain length of  $3.0 \times 10^7$ , a  $3.0 \times 10^6$  burn in, a GTR+ $\Gamma$ +I nucleotide rate substitution model, a strict molecular clock, and sampling of every 1,000 states. Trees and support values from the ML and BEAST analyses were similar. The Bayesian MAP log likelihood value tree was chosen from the posterior distribution and is depicted here. Posterior support values were calculated from consensus analysis of all Bayesian posterior trees, with values over 0.5 indicated above each node. The estimated TMRCA of tree nodes was calculated by Bayesian analyses and is reported below each node as years prior to the year 2000. The overall TMRCA of the entire L segment analysis is indicated in a text box located adjacent to the root node of SA51. Each taxon name indicates the strain, country of origin, and date of isolation. GenBank accession numbers for the virus L segments are DQ375395 to -403, DQ375406, DQ375409 to -16, DQ375418 to -29, and DQ375432 to -4. Strains used in previous studies of virulence in WF rats are indicated with either + (lethal; LD<sub>50</sub>, ~1.0 PFU), +/- (lethal; LD<sub>50</sub>, ~2 × 10<sup>3</sup> PFU), or - (nonlethal).

TABLE 1. RVF virus strains analyzed

Virus strain	Isolated from:	Country of origin	Yr isolated	Passage history <sup>a</sup>	SPB log ID no. <sup>b</sup>
763/70	Bovine	Zimbabwe	1970	SMB-1, AM-2, FRhL-1, E6-1	808766
1260/78	Bovine	Zimbabwe	1978	SMB-1, AM-2, FRhL-1, E6-1	808785
1853/78	Bovine	Zimbabwe	1978	SMB-1, AM-2, FRhL-1, E6-1	808768
2250/74	Bovine	Zimbabwe	1974	SMB-1, AM-2, FRhL-1, E6-1	808780
2269/74	Bovine	Zimbabwe	1974	SMB-1, AM-2, FRhL-1, E6-1	808767
2373/74	Bovine	Zimbabwe	1974	SMB-1, AM-2, FRhL-1, E6-1	808784
73HB1230	Human	CAR	1973	SMB-6, E6-4	808818
73HB1449	Human	CAR	1973	SMB-5, E6-1	808822
74HB59	Human	CAR	1974	SMB-6, E6-1	808812
ArD38388	<i>Aedes cuminsi</i>	Burkina Faso	1983	AP61-1, E6-4	808807
AnK3837	<i>Hipposideros caffer</i> (bat)	Guinea	1981	SMB-7, E6-1	808802
AnK6087	<i>Micropterus pusillus</i> (bat)	Guinea	1984	Unknown	808808
CAR R1622	Human	CAR	1985	SMB-4, E6-4	808811
Entebbe	<i>Eretmapodites</i> mosquito (pool)	Uganda	1944	AM-184IP, FRhL-1, E6-1	808754
HvB375	Human	CAR	1985	SMB-4, E6-4	808806
Kenya 56 (IB8)	Bovine	Kenya	1965	HA-1, LTC-25, SMB-8, FRhL-1, E6-1	808773
Kenya 83 (21445)	<i>Aedes macintoshi</i>	Kenya	1983	HA-1, FRhL-1, E6-1	808816
Kenya 98 (00523)	Human	Kenya	1998	E6-2	808171
MgH824	Human	Madagascar	1979	SMB-6, E6-4	808801
OS-1	Human	Mauritania	1987	FRhL-1, E6-1	808791
OS-3	Human	Mauritania	1987	FRhL-1, E6-1	808792
OS-8	Human	Mauritania	1987	FRhL-1, E6-1	808796
OS-9	Human	Mauritania	1987	FRhL-1, E6-1	808797
SA-51	Ovine	South Africa	1951	AS-3, SMB-1, FRhL-1, E6-1	808759
SA-75	Human	South Africa	1975	FRhL-2, E6-1	808787
Saudi-2000 (10911)	Human	Saudi Arabia	2000	E6-1	808860
ZC-3349	Bovine	Egypt	1978	SMB-2, FRhL-1, E6-1	808764
ZH-501	Human	Egypt	1977	SMB-2, FRhL-1, E6-2	810544
ZH-548	Human	Egypt	1977	SMB-2, FRhL-1, E6-1	808783
ZH-1776	Human	Egypt	1978	SMB-2, FRhL-1, E6-1	808813
ZM-657	Mosquito	Egypt	1978	SMB-2, FRhL-1, E6-1	808765
ZS-6365	Ovine	Egypt	1979	SMB-2, FRhL-1, E6-1	808763
Zinga	Human	CAR	1969	SMB-8, FRhL-1, E6-1	808782

<sup>a</sup> AM, adult mouse; AP-61, AP-61 cell line; AS, adult sheep; E6, Vero E6 cell line; FRhL, fetal rhesus lung; HA, hamster; IP, intraperitoneal; LTC, lamb testicular cell line; SMB, suckling mouse brain.

<sup>b</sup> SPB log ID no. refers to the unique identifier number assigned within our laboratory. Virus stocks were prepared in VERO E6 cells and stored in liquid nitrogen. All infectious virus work was conducted in the biosafety level 4 laboratory.

#### RVF virus epidemic/epizootic versus endemic viral ecology.

This study included representative samples of multiple RVF virus isolates from two temporally, phylogenetically, and geographically distinct epizootics/epidemics. They included the largest known “virgin-soil” RVF virus outbreak, which occurred in Egypt in 1977 to 1979, and a significant epizootic/epidemic that occurred 10 years later in Mauritania and Senegal, during 1987. Our data show that a unique viral genotype predominated within each outbreak. The genomic-sequence data for the RVF virus strains from within each of these outbreaks form monophyletic groupings with little nucleotide diversity (Fig. 2, 3, and 4; see Table 2). It is clear that during the 1977-to-1979 Egyptian outbreak, a single viral genotype was involved, which was virtually identical regardless of the source of the virus isolate (mosquito, human, or livestock), with a maximum pairwise difference at the nucleotide level of 0.33% (Tables 1 and 2). Similarly, RVF virus strains collected during the 1987 Mauritanian epizootic/epidemic form a very distinct monophyletic group with maximum pairwise nucleotide difference of 0.26% (Table 2).

These findings contrast sharply with the high genomic diversity observed for isolates collected during the 1970s from Zimbabwe and the CAR, at which time these regions experienced

either low-level enzootic activity or geographically localized epizootics/epidemics (24, 58). The virus strains in this analysis included six RVF virus strains from Zimbabwe that were collected from aborted bovine fetuses over a period of 8 years (1970 to 1978) within a 170-km radius of the capital city, Harare. Remarkably, these Zimbabwean RVF virus strains can be found within five of the seven distinct RVF virus phylogenetic lineages (Fig. 2, 3, and 4). Phylogenetic groupings of the three CAR strains that were collected within a 2-month period at a time of low-level RVF virus activity near Bangui, CAR, indicate that two of the strains (73HB1449 and 74HB59) lie within lineage C and have almost identical sequences for their S, M, and L segments. In contrast, the remaining strain, 73HB1230, was phylogenetically distant. As was shown above in the evidence of reassortment, the S and M segments of strain 73HB1230 are located within lineage B (ancestral to the other members of this Kenya/Saudi Arabian lineage), whereas the L segment is within lineage A. Taken together, these findings illustrate the potential for multiple distinct virus genotypes to coexist over relatively small geographic areas.

**Evidence for RVF virus introduction across natural geographic boundaries.** This complete genome data set also establishes a firm phylogenetic linkage between RVF virus



TABLE 2. Comparison of M segment nucleotide identity differences among endemic versus epizootic/epidemic RVF virus strains

Strain no.	Strain	Nucleotide identity differences with strain no.:																	
		Endemic									Epidemic								
		Zimbabwe 1970–1978						CAR 1973–74			Mauritania 1987				Egypt 1977–1979				
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	763/70	<b>4.7</b>	<b>4.5</b>	<b>3.5</b>	<b>4.5</b>	<b>4.4</b>	4.3	4.3	4.3	4.8	4.7	4.8	4.7	4.7	4.6	4.7	4.7	4.7	4.8
2	2250/74		<b>4.1</b>	<b>4.3</b>	<b>2.4</b>	<b>2.2</b>	1.9	2.3	2.3	4.7	4.6	4.7	4.7	0.4	0.4	0.4	0.4	0.4	0.5
3	2269/74			<b>3.8</b>	<b>4.1</b>	<b>3.7</b>	3.5	3.9	3.9	4.3	4.2	4.3	4.3	4.1	4.0	4.0	4.1	4.1	4.1
4	2373/74				<b>4.3</b>	<b>4.0</b>	3.9	4.0	4.0	4.2	4.1	4.2	4.1	4.3	4.2	4.2	4.3	4.2	4.3
5	1260/78					<b>2.1</b>	2.1	1.2	1.2	4.6	4.6	4.6	4.6	2.4	2.4	2.4	2.4	2.4	2.5
6	1853/78						1.6	1.9	1.9	4.3	4.2	4.3	4.3	2.2	2.1	2.2	2.2	2.2	2.3
7	73HB1230							<b>1.8</b>	<b>1.8</b>	4.2	4.2	4.2	4.2	1.9	1.8	1.9	1.9	1.9	2.0
8	73HB1449								<b>0.0</b>	4.3	4.2	4.3	4.3	2.3	2.3	2.3	2.3	2.3	2.4
9	74HB59									4.3	4.2	4.3	4.3	2.3	2.3	2.3	2.3	2.3	2.4
10	OS1										<b>0.2</b>	<b>0.3</b>	<b>0.2</b>	4.5	4.6	4.5	4.6	4.6	4.7
11	OS3												<b>0.1</b>	<b>0.0</b>	4.5	4.5	4.5	4.5	4.7
12	OS8													<b>0.1</b>	4.6	4.6	4.6	4.6	4.8
13	OS9														4.6	4.6	4.6	4.6	4.7
14	ZH501															<b>0.2</b>	<b>0.2</b>	<b>0.2</b>	<b>0.3</b>
15	ZH548																<b>0.1</b>	<b>0.1</b>	<b>0.3</b>
16	ZH1776																	<b>0.1</b>	<b>0.3</b>
17	ZM657																	<b>0.1</b>	<b>0.3</b>
18	ZS6365																		<b>0.3</b>
19	ZC3349																		<b>0.3</b>

<sup>a</sup> The numbering of strains along the top of the table corresponds to the RVF strain numbers in the first column. Boldface numbers indicate differences among viruses that are within a given enzootic area or epizootic.

strains collected outside of continental Africa (Saudi Arabia and Madagascar), across physical geographic barriers, and virus strains circulating earlier within continental Africa. The first RVF virus isolation outside of continental Africa occurred in 1979 from a mosquito pool collected in a forested area of Madagascar. On subsequent handling in the laboratory, this virus resulted in a human infection and the resulting strain, MgH824. This virus was found to be closely linked phylogenetically with a strain (2250/74) that was collected in 1974 from an aborted bovine fetus in Zimbabwe. Bayesian TMRCA analyses indicated that movement of RVF virus to Madagascar may have involved a common ancestor that was circulating in Zimbabwe in the early 1970s. Further evidence of the ability of RVF virus to cross physical boundaries was found with the phylogenetic linkage of a virus (strain Kenya98-00523) collected during the large epizootic/epidemic that occurred in eastern Africa from Kenya south to the Kruger Park in South Africa in 1997 and 1998 with the subsequent introduction of RVF virus (strain Saudi2000-10911) across the Red Sea onto the Arabian Peninsula.

DISCUSSION

The high conservation of the RVF genome sequence suggests either that the overall tolerance for mutation within the RVF virus genome is very low (an inherently slow clock) or that the viruses in the group (as represented by the current isolate collection) have a relatively recent common ancestor. The concept of a “double filter” operating to constrain arbovirus genome evolution to a sequence space where the virus can operate equally effectively within the insect and mammalian host cell environments may provide a plausible explanation for the low diversity seen with RVF virus (59). With a

“double filter” in operation, to maintain the numbers of persistently infected mosquitoes above a critical population size requires that RVF virus retain the ability to infect suitable mammalian amplification hosts easily and generate the high-titer viremia necessary to ensure the infection of large numbers of naive mosquitoes. This ecological constraint would be reflected at the genome level by increased purifying selection, limiting evolution at amino acid-changing (nonsynonymous) sites. Consistent with purifying selection, mean values for nonsynonymous ( $d_N$ ) and synonymous ( $d_S$ ) substitutions per site (ratio,  $d_N/d_S$ ) of 0.02861, 0.07327, 0.03786, 0.02971, and 0.15075 were obtained for the RVF L, Gn, Gc, N, and NSs ORF data sets, respectively. In addition, mean RVF virus evolutionary rates of  $6.5 \times 10^{-5}$ ,  $2.7 \times 10^{-5}$ , and  $6.3 \times 10^{-4}$  nt substitutions per site per year for the first, second, and third codon positions, respectively, were found, consistent with constraint at the protein-coding level. However, recent comparative work has shown that while RVF and some other vectored RNA viruses, such as western and eastern equine encephalitis viruses, have low overall molecular evolutionary rates, several mammalian-restricted viruses, such as measles virus, human parainfluenza virus, and rabies virus, were also found to have nucleotide substitution rates similar to those of these arboviruses (29). Additionally, in vitro evidence contrary to the “double-filter” hypothesis was found on genetic analysis of vesicular stomatitis viruses during consecutive insect-to-insect or mammal-to-mammal cell culture passages compared to serial insect-mammal-insect cell passages (46). The detailed Bayesian analyses of RVF virus sequence differences presented here also suggest that the low overall nucleotide diversity may have more to do with a relatively recent common ancestor than a “double filter” in operation.

The Bayesian analyses of the 33 complete genomes rather

TABLE 3. Groupwise amino acid mutations found between lethal, intermediate, and nonlethal RVF virus strains in WF rats\*

Virus strain	Lethality <sup>a</sup>	Amino acid at position <sup>b</sup> :														
		Gn				Gc	NSs				Pol			NP		
		595	605	631	659	1059	23	167	217	242	23	278	302	407	663	159
ZH501	+	I	R	I	V	S	F	A	V	I	F	S	V	G	A	G
ZH548	+	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
ZH1776	+	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
ZM657	+	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
ZS6365	+	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
ZC3349	+	*	*	*	*	*	*	*	*	T	*	*	*	*	*	*
2269/74	+/-	V	K	V	A	T	I	V	A	V	Y	N	I	D	T	G
763/70	-	V	K	V	A	T	I	V	A	V	Y	N	I	D	T	E
1853/78	-	I	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Entebbe	-	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
OS1	-	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
OS3	-	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
OS8	-	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
OS9	-	*	R	*	*	*	*	*	*	*	*	*	*	*	*	*
SA51	-	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
SA75	-	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

<sup>a</sup> Lethality criteria were defined as follows: +, lethal (LD<sub>50</sub>, ~1.0 PFU); +/-, intermediate (LD<sub>50</sub>, ~3.0 × 10<sup>3</sup> PFU); and -, nonlethal (LD<sub>50</sub> > 1.0 × 10<sup>6</sup> PFU).  
<sup>b</sup> The numbering corresponds to the amino acid position in each individual protein ORF (NSs, Pol, and NP) or in the glycopolyprotein precursor molecule (Gn and Gc). The asterisks indicate amino acid identity with either the ZH501 or 763/70 strain within each lethality group.

surprisingly estimated the time to a recent common ancestor at only 108 to 117 years prior to the year 2000 (the date of isolation of the most contemporary virus sampled). This recent ancestry for the currently defined RVF virus diversity was unexpected. The first case reports of an illness similar to RVF occurred in the early 1900s as an enzootic hepatitis of sheep in the western regions of Kenya (30). During that period, large numbers of highly susceptible European breed cattle and sheep were being imported into eastern and southern Africa, replacing the traditional zebu cattle and sheep of indigenous tribal populations (26, 36). It is interesting to speculate that these changes in agricultural practices may be linked to the estimated TMRCA, which places the most recent common ancestor of the currently defined RVF virus diversity in the late 1800s. The TMRCA and actual veterinary case reports suggest that at some point between 1850 and 1910 an unknown arbovirus ancestor of what we now know as RVF virus exploited a newly formed ecological niche created by the presence of large concentrations of highly susceptible exotic cattle and sheep breeds and rapidly established itself endemically throughout much of eastern and southern Africa. Later, the movement of infected animal herds or mosquitoes would have potentially allowed the spread of RVF virus throughout sub-Saharan Africa. Evidence supportive of the last point is found in the current phylogenetic analysis, where it has been shown that although there is a rough geographic correlation of most RVF virus lineages, indicating that within a specific region there may be a predominant overall viral genotype, the finding of RVF virus strains from Zimbabwe, CAR, or South Africa within multiple virus lineages suggests that the movement of RVF virus genotypes is an ongoing process and can occur over large geographic distances.

From the data set of 33 complete genomes, the dynamics of virus introduction and maintenance can be compared and contrasted between endemic and epidemic periods. This study contains multiple samples from two large epizootics/epidemics

that affected tens of thousands of animals and people. The genotypes sampled from these outbreaks, Egypt in 1977 and Mauritania in 1987, point to a single virus genotype introduction and to that genotype being responsible for the majority of human and animal illness during these periods. That contrasts sharply with data from six RVF virus isolates collected in Zimbabwe during the 1970s, which was a period of low-level endemicity and localized epizootics. The multiple virus genotypes collected from Zimbabwe during this period suggest that in areas of endemicity, such as southern Africa, multiple virus genotypes can circulate concurrently.

Questions regarding the timing and locations of epizootics and why a particular virus genotype is predominant during an outbreak likely have complex answers and involve the confluence of a number of factors. These may include the existence of transovarially virus-infected mosquito eggs, favorable conditions for the hatching of such eggs, large numbers of susceptible vertebrate amplifying hosts (e.g., naive livestock), and environmental conditions for efficient transmission and spread before acquired immunity or death in the local susceptible populations reduces the reproductive threshold (*R*<sub>0</sub>) below the critical threshold. These conditions are often met by periods of sporadic, unusually heavy rains in semiarid regions that allow the buildup of naive animal populations during the intervening dry years (34).

Evidence of RNA segment reassortment has been found both in vitro and in vivo for several viruses throughout the family *Bunyaviridae* (5, 7, 21, 52). The demonstration here of incongruities among the RVF virus S, M, and L segment tree topologies, together with earlier such findings based on partial genome sequences, indicates that RVF virus RNA segment reassortment does occur in nature (53, 54). The dominance of a single viral genotype during epizootic/epidemic periods, the short duration of infected-host viremia, and the strongly cross-protective and sterilizing immunity elicited by RVF virus infection in livestock and humans all point to a relatively small

window of opportunity for coinfection of hosts with multiple viral genotypes. However, the existence of multiple viral genetic lineages in relatively small geographic areas during periods of endemic or localized epizootic activity suggests that there is ample opportunity for reassortment to occur in these settings.

Although RVF virus genetic diversity is low, remarkable differences in mammalian pathogenesis among different virus isolates have been well documented (17, 47). It is clear from historic epidemiologic data that a wide spectrum of human disease outcomes result from RVF virus infection. RVF disease in humans is typically a self-limiting febrile illness that, in 1 to 2% of infected individuals, can progress to more serious complications, including hepatitis, encephalitis, retinitis, blindness, or a hemorrhagic syndrome that results in fatal outcomes in 10 to 20% of hospitalized cases. Animal models of infection have been used to mimic these disparate syndromes with various degrees of success (1, 2, 8, 12, 16, 42, 48). Unfortunately, even though moderate hematologic derangement has been detected in nonhuman primate models of RVF virus infection, to date no effective rodent model of hemorrhagic RVF disease has been identified (12). While RVF virus infection in adult mice and neonatal animals (rodents and livestock) is almost uniformly fatal, more variable results have been obtained with adult inbred Wistar-Furth (WF) strain *Rattus norvegicus* (48, 51). These studies demonstrated that RVF virus from the lineage A 1977-to-1979 Egyptian lineage, such as strain ZH501, induced fatal hepatic disease in WF rats, with a 50% lethal dose ( $LD_{50}$ ) of 1 to 5 PFU. However, when representative RVF virus strains from sub-Saharan Africa, found in lineage C, D, or E, that were lethal for humans or livestock were inoculated, these viruses failed to kill, and often failed to infect, WF rats (47, 3). These results are summarized in Fig. 2, 3, and 4 as lethal or nonlethal in the WF rat model. Pairwise comparisons of the deduced amino acid sequences revealed several striking patterns of conservative and nonconservative changes between Egyptian- and sub-Saharan-lineage viruses that have been assayed in WF rats (Table 3). These lethal versus nonlethal groupwise mutations were found in the NSs, NP, Gn, Gc, and polymerase coding regions. Interestingly, several of these mutations cluster together in a 64-aa region at the C terminus of Gn, suggesting that these mutations may play roles in virus receptor binding in the WF rat (Table 3). These and other mutations that are found exclusively in the lineage A virus strains may contribute either individually or synergistically to their increased virulence in the WF rat model. It is likely that in human or livestock populations other host-related factors, such as genetic polymorphisms, overall immune function, exposure history, and nutritional status, play significant roles in the outcome of RVF virus infection in nature. Detailed insights into the molecular determinants of RVF virus pathogenesis will be greatly assisted by the complete genome data set presented here for use as a road map in reverse-genetics approaches (22, 28).

With the elucidation of the molecular mechanisms of pathogenesis, targeted therapeutic interventions or reverse-genetics-derived live-attenuated vaccine constructs may be developed to reduce the burden of RVF disease. The highly conserved nature of the RVF virus genome suggests that potential live-attenuated or recombinant vaccine constructs would confer

protective immunity against challenge by nonhomologous RVF strains. This could allow one efficacious vaccine construct to be employed throughout Africa, thereby conferring protection against all RVF virus lineages.

The potential risk of RVF virus spread is high. Our RVF genomic analysis documents the previous long-distance movement of RVF lineages and demonstrates clearly that a single introduction of RVF virus, given the proper environmental conditions, could spread rapidly, resulting in catastrophic economic losses and significant human disease. These genomic data will be invaluable in assessing such threats, as they will allow the development of pan-RVF virus molecular detection assays and bioforensic capability to provide rapid identification of potential RVF virus introductions. Finally, control of any potential RVF virus introduction via natural or intentional means will require the close cooperation of medical and veterinary authorities and should be included in any governmental emergency planning.

#### ACKNOWLEDGMENTS

Many of the RVF virus isolates included in this study were obtained with the generous assistance of John Morrill of the United States Army Medical Research Institute of Infectious Diseases (USAMRIID), Fort Detrick, MD, or Robert Swanepoel of the Special Pathogens Unit, the National Institute for Communicable Diseases, Sandringham, South Africa. We thank Roman Biek and Andrew Rambaut for their insightful comments regarding the Bayesian statistical analyses presented here. B.H.B. thanks N. James MacLachlan of the University of California, Davis, School of Veterinary Medicine (SVM) and Frederick Murphy of the University of Texas Medical Branch for their steadfast support and mentoring throughout this study.

B.H.B. was supported during this work by the Students Training in Advanced Research (STAR) fellowship and the Veterinary Scientist Training Program (VSTP) of the University of California, Davis, SVM.

#### REFERENCES

- Anderson, G. W., Jr., W. Lawrence, J. O. Lee, and William C. Hall. Ocular sequelae associated with Rift Valley fever virus (RVFV) infection in inbred rat strains. *Abstr. VIIIth Int. Congr. Virol.*, abstr. P70-002.
- Anderson, G. W., Jr., and C. J. Peters. 1988. Viral determinants of virulence for Rift Valley fever (RVF) in rats. *Microb. Pathog.* 5:241-250.
- Anderson, G. W., Jr., J.-F. Saluzzo, T. G. Ksiazek, J. F. Smith, W. Ennis, D. Thureen, C. J. Peters, and J.-P. Digoutte. 1989. Comparison of in vitro and in vivo systems for propagation of Rift Valley fever virus from clinical specimens. *Res. Virol.* 140:129-138.
- Aquino, V. H., M. L. Moreli, and L. T. Moraes Figueiredo. 2003. Analysis of oropouche virus L protein amino acid sequence showed the presence of an additional conserved region that could harbour an important role for the polymerase activity. *Arch. Virol.* 148:19-28.
- Beatty, B. J., D. R. Sundin, L. J. Chandler, and D. H. Bishop. 1985. Evolution of bunyaviruses by genome reassortment in dually infected mosquitoes (*Aedes triseriatus*). *Science* 230:548-550.
- Billecocq, A., M. Spiegel, P. Vialat, A. Kohl, F. Weber, M. Bouloy, and O. Haller. 2004. NSs protein of Rift Valley fever virus blocks interferon production by inhibiting host gene transcription. *J. Virol.* 78:9798-9806.
- Borucki, M. K., L. J. Chandler, B. M. Parker, C. D. Blair, and B. J. Beatty. 1999. Bunyavirus superinfection and segment reassortment in transovarially infected mosquitoes. *J. Gen. Virol.* 80:3173-3179.
- Bucci, T. J., M. I. Moussa, and O. L. Wood. 1981. Experimental Rift Valley fever encephalitis in ACI rats. *Contrib. Epidemiol. Biostat.* 3:60-67.
- Centers for Disease Control and Prevention. 2000. Update: outbreak of Rift Valley fever—Saudi Arabia, August–November 2000. *Morb. Mortal. Wkly. Rep.* 49:982-985.
- Coetzer, J. A. 1982. The pathology of Rift Valley fever. II. Lesions occurring in field cases in adult cattle, calves and aborted foetuses. *Onderstepoort J. Vet. Res.* 49:11-17.
- Collett, M. S., L. T. Kakach, J. A. Suzich, and T. L. Wasmoe. 1989. Gene products and expression strategy of the M segment of the phlebovirus Rift Valley fever virus, p. 49-57. *In* D. Kolakofsky and B. W. J. Mahy (ed.), *Genetics and pathogenicity of negative strand viruses*. Elsevier, Amsterdam, The Netherlands.
- Cosgriff, T. M., J. C. Morrill, G. B. Jennings, L. A. Hodgson, M. V. Slayter,

- P. H. Gibbs, and C. J. Peters. 1989. Hemostatic derangement produced by Rift Valley fever virus in rhesus monkeys. *Rev. Infect. Dis.* **11**:S807–S814.
13. Drake, J. W., and J. J. Holland. 1999. Mutation rates among RNA viruses. *Proc. Natl. Acad. Sci. USA* **96**:13910–13913.
  14. Drummond, A. J., S. Y. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**:e88.
  15. Easterday, B. C., M. H. McGavran, J. R. Rooney, and L. C. Murphy. 1962. The pathogenesis of Rift Valley fever in lambs. *Am. J. Vet. Res.* **23**:470–479.
  16. Easterday, B. C., and L. C. Murphy. 1963. Studies on Rift Valley fever in laboratory animals. *Cornell Vet.* **53**:423–433.
  17. Erasmus, B. J., and J. A. W. Coetzer. The symptomatology and pathology of Rift Valley fever in domestic animals. *Contrib. Epidemiol. Biostat.* **3**:77–82.
  18. Ewing, B., and P. Green. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**:186–194.
  19. Ewing, B., L. Hillier, M. C. Wendl, and P. Green. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**:175–185.
  20. Findlay, G. M., and R. Daubney. 1931. The virus of Rift Valley fever or enzootic hepatitis. *Lancet* **ii**:1350–1351.
  21. Gerrard, S. R., L. Li, A. D. Barrett, and S. T. Nichol. 2004. Ngari virus is a Bunyamwera virus reassortant that can be associated with large outbreaks of hemorrhagic fever in Africa. *J. Virol.* **78**:8922–8926.
  22. Gerrard, S. R., B. H. Bird, C. G. Albarino, and S. T. Nichol. The NSm proteins of Rift Valley fever virus are dispensable for maturation, replication and infection. *Virology*, in press.
  23. Gerrard, S. R., and S. T. Nichol. 2007. Synthesis, proteolytic processing and complex formation of N-terminally nested precursor proteins of the Rift Valley fever virus glycoproteins. *Virology* **357**:124–133.
  24. Gonzalez, J.-P., J. C. Bouquety, J.-L. Lesbordes, M. C. Madelon, C. C. Mathiot, D. M. Y. Meunier, and A. J. Georges. 1987. Rift Valley fever virus and haemorrhagic fever in the Central African Republic. *Ann. Inst. Pasteur Virol.* **138**:385–390.
  25. Gordon, D., C. Abajian, and P. Green. 1998. Consed: a graphical tool for sequence finishing. *Genome Res.* **8**:195–202.
  26. Hanotte, O., D. G. Bradley, J. W. Ochieng, Y. Verjee, E. W. Hill, and J. E. Rege. 2002. African pastoralism: genetic imprints of origins and migrations. *Science* **296**:336–339.
  27. Ikegami, T., S. Won, C. J. Peters, and S. Makino. 2005. Rift Valley fever virus NSs mRNA is transcribed from an incoming anti-viral-sense S RNA segment. *J. Virol.* **79**:12106–12111.
  28. Ikegami, T., S. Won, C. J. Peters, and S. Makino. 2006. Rescue of infectious Rift Valley fever virus entirely from cDNA, analysis of virus lacking the NSs gene, and expression of a foreign gene. *J. Virol.* **80**:2933–2940.
  29. Jenkins, G. M., A. Rambaut, O. G. Pybus, and E. C. Holmes. 2002. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *J. Mol. Evol.* **54**:156–165.
  30. Kabete Veterinary Laboratories. Diseases of sheep. Kenya Veterinary Department annual report—1910. Kenya Department of Veterinary Services, Nairobi, Kenya.
  31. Kakach, L. T., J. A. Suzich, and M. S. Collett. 1989. Rift Valley fever virus M segment: phlebovirus expression strategy and protein glycosylation. *Virology* **170**:505–510.
  32. Kosakovsky Pond, S. L., S. D. W. Frost, and S. V. Muse. 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**:2531–2533.
  33. Le May, N., S. Dubaele, L. Proietti De Santis, A. Billecocq, M. Bouloy, and J. M. Egly. 2004. TFIIH transcription factor, a target for the Rift Valley hemorrhagic fever virus. *Cell* **116**:541–550.
  34. Linthicum, K. J., A. Anyamba, C. J. Tucker, P. W. Kelley, M. F. Myers, and C. J. Peters. 1999. Climate and satellite indicators to forecast Rift Valley fever epidemics in Kenya. *Science* **285**:397–400.
  35. Lole, K. S., R. C. Bollinger, R. S. Paranjape, D. Gadkari, S. S. Kulkarni, N. G. Novak, R. Ingersoll, H. W. Sheppard, and S. C. Ray. 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J. Virol.* **73**:152–160.
  36. MacOwan, K. D. 1994. The development of a livestock industry in Kenya. *Vet. Hist.* **8**:29–37.
  37. Madani, T. A., Y. Y. Al-Mazrou, M. H. Al-Jeffri, A. A. Mishkhas, A. M. Al-Rabeah, A. M. Turkistani, M. O. Al-Sayed, A. A. Abodahish, A. S. Khan, T. G. Ksiazek, and O. Shobokshi. 2003. Rift Valley fever epidemic in Saudi Arabia: epidemiological, clinical, and laboratory characteristics. *Clin. Infect. Dis.* **37**:1084–1092.
  38. McIntosh, B. M., D. Russell, I. dos Santos, and J. H. Gear. 1980. Rift Valley fever in humans in South Africa. *S. Afr. Med. J.* **58**:803–806.
  39. Meegan, J. M. 1981. Rift Valley fever in Egypt: an overview of the epizootics in 1977 and 1978. *Contrib. Epidemiol. Biostat.* **3**:100–113.
  40. Meegan, J. M., R. H. Watten, and L. W. Laughlin. 1981. Clinical experience with Rift Valley fever in humans during the 1977 Egyptian epizootic. *Contrib. Epidemiol. Biostat.* **3**:114–123.
  41. Montgomery, R. E., and R. J. Stordy. Report of Veterinary Department for 1912–1913. Annual Report of the Department of Agriculture, Kenya Colony.
  42. Morrill, J. C., F. K. Knauert, T. G. Ksiazek, J. M. Meegan, and C. J. Peters. 1989. Rift Valley fever infection of rhesus monkeys: implications for rapid diagnosis of human disease. *Res. Virol.* **140**:139–146.
  43. Morvan, J., P. E. Rollin, S. Laventure, I. Rakotoarivony, and J. Roux. 1992. Rift Valley fever epizootic in the central highlands of Madagascar. *Res. Virol.* **143**:407–415.
  44. Muller, R., O. Poch, M. Delarue, D. H. L. Bishop, and M. Bouloy. 1994. Rift Valley fever virus L segment—correction of the sequence and possible functional role of newly identified regions conserved in RNA-dependent polymerases. *J. Gen. Virol.* **75**:1345–1352.
  45. Nichol, S. T. 2001. Bunyaviruses, p. 1603–1633. *In* B. N. Fields and D. M. Knipe (ed.), *Fields virology*, 4th ed. Lippincott Williams & Wilkins, Philadelphia, PA.
  46. Novella, I. S., C. L. Hershey, C. Escarmis, E. Domingo, and J. J. Holland. 1999. Lack of evolutionary stasis during alternating replication of an arbovirus in insect and mammalian cells. *J. Mol. Biol.* **287**:459–465.
  47. Peters, C. J., and G. W. Anderson, Jr. 1981. Pathogenesis of Rift Valley fever. *Contrib. Epidemiol. Biostat.* **3**:21–41.
  48. Peters, C. J., and T. W. Slone. 1982. Inbred rat strains mimic the disparate human response to Rift Valley fever virus infection. *J. Med. Virol.* **10**:45–54.
  49. Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**:817–818.
  50. Rippy, M. K., M. J. Topper, C. A. Mebus, and J. C. Morrill. 1992. Rift Valley fever virus-induced encephalomyelitis and hepatitis in calves. *Vet. Pathol.* **29**:495–502.
  51. Ritter, M., M. Bouloy, P. Vialat, C. Janzen, O. Haller, and M. Frese. 2000. Resistance to rift valley fever virus in *Rattus norvegicus*: genetic variability within certain 'inbred' strains. *J. Gen. Virol.* **81**:2683–2688.
  52. Saeed, M. F., H. Wang, M. Suderman, D. W. Beasley, A. Travassos da Rosa, L. Li, R. E. Shope, R. B. Tesh, and A. D. Barrett. 2001. Jabobal virus is a reassortant containing the small RNA of Orpouche virus. *Virus Res.* **77**:25–30.
  53. Sall, A. A., P. M. de Zanolto, H. G. Zeller, J.-P. Digoutte, Y. Thiongane, and M. Bouloy. 1997. Variability of the NSs protein among Rift Valley fever virus isolates. *J. Gen. Virol.* **78**:2853–2858.
  54. Sall, A. A., P. M. de Zanolto, O. K. Sene, H. G. Zeller, J.-P. Digoutte, Y. Thiongane, and M. Bouloy. 1999. Genetic reassortment of Rift Valley fever virus in nature. *J. Virol.* **73**:8196–8200.
  55. Shoemaker, T., C. Boulianne, M. J. Vincent, L. Pezzanite, M. M. Al-Qahtani, Y. Al-Mazrou, A. S. Khan, P. E. Rollin, R. Swanepoel, T. G. Ksiazek, and S. T. Nichol. 2002. Genetic analysis of viruses associated with emergence of Rift Valley fever in Saudi Arabia and Yemen, 2000–01. *Emerg. Infect. Dis.* **8**:1415–1420.
  56. Stordy, R. J. 1913. Mortality among lambs. Annual Report Department of Agriculture, British East Africa, 1912–1913.
  57. Suzich, J. A., L. T. Kakach, and M. S. Collett. 1990. Expression strategy of a phlebovirus: biogenesis of proteins from the Rift Valley fever virus M segment. *J. Virol.* **64**:1549–1555.
  58. Swanepoel, Robert. 1981. Observations on Rift Valley fever in Zimbabwe. *Contrib. Epidemiol. Biostat.* **3**:83–91.
  59. Weaver, S. C. 2006. Evolutionary influences in arboviral disease. *Curr. Top. Microbiol. Immunol.* **299**:285–314.
  60. Won, S., T. Ikegami, C. J. Peters, and S. Makino. 2006. NSm and 78-kilodalton proteins of Rift Valley fever virus are nonessential for viral replication in cell culture. *J. Virol.* **80**:8274–8278.