Retrospective genetic characterisation of ***Encephalomyocarditis viruses*** from African elephant and swine recovers two distinct lineages in South Africa

James H. d. T. van Sandwyk[1], Nigel C. Bennett[1], Robert Swanepoel[2,3] & Armanda D. S. Bastos[1,*]

[1]Mammal Research Institute, Department of Zoology and Entomology, University of Pretoria, Private Bag X20, Hatfield, Pretoria, 0028, South Africa

[2]National Institute for Communicable Diseases, National Health Laboratory Service, Private Bag X4,
Sandringham, Johannesburg, 2131, South Africa

[3]Department of Medical Virology, University of Pretoria, South Africa

Abstract

Encephalomyocarditis virus (EMCV) outbreaks are rare in southern Africa. Only two have been reported to date from South Africa, both coinciding with rodent irruptions. The first outbreak manifested as acute myocarditis in pigs in 1979, whilst the second, occurring from 1993 to 1994, was linked to the deaths of 64 free-ranging adult African elephants (*Loxodonta africana*). The P1 genome region, inclusive of the flanking leader (L) and 2A genes, of three South African isolates, one from swine and two from elephants, was characterised by PCR amplification and sequencing of up to 11 overlapping fragments. In addition to the resulting 3329 nucleotide dataset, the 3D region that is widely used in molecular epidemiology studies, was characterised, and three datasets (P1, VP1/3 and 3D), complemented with available homologous EMCV data, were compiled for analyses. Phylogenetic inferences revealed the near-identical elephant outbreak strains to be most closely related to a mengovirus from rhesus macaques (*Macaca mulatta*) in Uganda, differing from the latter by between 11 % (3D) and 15 % (VP3/1). The South African pig isolate differed by 4 % (3D) and 11 % (VP3/1) from available European and Asian pig virus sequences. This study confirms the presence of two genetically distinct EMCV lineages from wild and domestic hosts in southern African, and provides valuable baseline data for future outbreak eventualities in the sub-region.

Keywords: *Encephalomyocarditis virus*, South Africa, *Sus scrofa*, *Loxodonta africana,* Mengovirus, epidemiology, phylogeny

Introduction

Encephalomyocarditis virus (EMCV) and *Theilovirus* (ThV), are members of the *Cardiovirus* genus (family Picornaviridae), comprising one and 12 serotypes, respectively (Minor et al. 1995). Initially, the degree of relatedness of diverse poliomyelitis and mouse encephalomyelitis viruses was not clear (Dick 1949; Bautista et al. 1957), but the original Columbia-SK, EMC and Mengo isolates, as well as viruses from diverse hosts, sampled across a broad geographical range, are now all recognised as members of a single EMCV species and serotype (Craighead 1965, Zimmerman 1994).

The virus has an extremely broad host range, occurring in murid rodents, domestic and wild pigs (Maurice et al. 2005), various captive species (Thomson et al., 2001) and even humans (Dick et al. 1948, Tesh 1978, Obertse et al. 2009). Clinical outcomes range from inapparent, to acute and highly lethal infections (Thomson et al. 2001). Aside from outbreaks in pigs (Knowles et al. 1998; Maurice et al. 2005; Denis et al. 2006), sporadic outbreaks are recorded in zoos and invariably affect a wide range of species (Reddacliff et al. 1997, Canelli et al. 2010), however, primates and elephants appear to be particularly susceptible (Simpson et al. 1977, Tesh & Wallace 1978, Seaman & Finnie 1987, Jones et al. 2011). Murids are generally considered to be the virus reservoir hosts (Zimmerman 1994, Thomson et al., 2001), but a possible role for wild boars in the epidemiology of the disease in domestic pigs in Europe, has been suggested (Billinis 2009). Although EMC has an essentially worldwide distribution, the molecular characterisation of viruses primarily corresponds to pig isolates from Europe, Asia and America. Data from Africa is limited to isolates from rhesus macaques, sampled in Uganda in the 1940s and to an isolate from a semi-wild bonobo in the Democratic Republic of Congo, sampled in 2009 (Jones et al. 2011).

EMC outbreaks are rarely recognised in southern Africa, however, two outbreaks affecting domestic and wildlife species have been documented in South Africa (Williams, 1981, Grobler et al. 1995). The first in 1979 was restricted to pigs in Kwa-Zulu Natal Province (Williams, 1981), and the second occurred in a free-ranging population of African elephants (*Loxodonta africana*), between December 1993 and November 1994 in the Kruger National Park (KNP). The latter outbreak was notable for its marked gender-bias; of the 64 animals that succumbed, 83 % were adult bulls (Grobler et al. 1995).

In common with other picornaviruses, the *Cardiovirus* genome consists of a single open reading frame 7800-8200 bases in length, encoding a single polyprotein comprising three regions,P1, P2 and P3, that each encode five (L, 1A-1D), three (2A-2C) and four (3A-3D) non-overlapping genes (Hughes, 2004; Simmonds, 2006). The P1 region, encoding the VP4 (1A), VP2 (1B), VP3 (1C) and VP1 (1D) structural proteins, contains the major antigenic sites of the virus (Scraba & Palmenberg, 1999), and the VP3/1 junction has proved valuable for determining relatedness of outbreak strains (Knowles et al. 1998). Similarly, the 3D gene, encoding the RNA-dependent RNA polymerase (RNA replicase), is often used for molecular epidemiology studies (Vanderhallen & Koenen 1998; Koenen *et al.,* 1999; Denis et al. 2006; Jones et al. 2011), and is the best-represented gene in the GenBank database, with sequences for more than 60 strains, predominantly from swine, presently available.

As replication in lytic viruses takes place several times in a single infectious cycle, and as the high mutation rate of picornavirus RNA replicase (3D gene) results in approximately one mutation per replication cycle (Drake *et al.,* 1998), the resulting viral genome copies are not identical. Instead a ‘master sequence’ predominates in the quasispecies or ‘swarm of mutants’ (Eigen, 1996), that can be determined by PCR amplification and sequencing of all variants present. This was the approach taken here to characterise one pig and two

elephant isolates from South Africa, with the aim of determining relatedness and possible origins of these viruses, and to establish a regional reference sequence database.

Materials & Methods

RNA extraction and cDNA synthesis

Three isolates, one designated SAR/1979 obtained from the outbreak in pigs in 1979, and two from the 1993/4 outbreak in African elephants, designated KNP/17/94 and KNP/19/94 were obtained from the South African National Institute for Communicable Diseases (NICD) virus bank. Total RNA was extracted using the Zymo Research Mini RNA Isolation II kit (Zymo Research) and cDNA was synthesised using a random hexanucleotide approach which has previously been used for detection and characterisation of picornavirus genes (Meyer *et al.,* 1991; Bastos, 1998). Briefly, RNA was reverse-transcribed in the presence of 1 x AMV Reverse Transcriptase buffer (Fermentas), 5 µM random hexamer primers (IDT), 0.2 µM dNTPs (Fermentas) and 10 U Ribolock^{TM} RNase inhibitor (Fermentas). Reactions were performed at 42°C in a final reaction volume of 10 µl, containing 5% DMSO and 10U of AMV Reverse Transcriptase (Fermentas). Following incubation at 42°C for one hour, samples were heat-inactivated at 80°C for 1 minute and stored at -20°C.

PCR amplification

Primers, designed specifically for this study (summarised in Table 1), were used to generate up to 11 overlapping genome fragments, ranging from 371 to 1500 nucleotides (nt) in length, and spanning the entire P1 region. In addition, the epidemiologically informative 3D region was targeted with primers previously used to screen for *Cardiovirus* genome presence in invasive *Mus* from sub-Antarctic Marion Island (de

Bruyn et al. 2008). All genomic amplifications followed the same general protocol, in which a master mix was prepared containing 1 x Dream*Taq*™ Buffer (Fermentas), 0.2 µM dNTPs (Fermentas), and 0.4 µM of each primer (synthesised by Inqaba Biotech and IDT). Each reaction contained 1.25 U of Dream*Taq*™ DNA polymerase (Fermentas) and 100-200ng of cDNA template and was performed in a final reaction volume of 50µl. Annealing temperature and extension times varied between reactions and were guided by the primer with the lowest calculated Tm of each primer pair (Table 1), and by the length of the fragment being targeted, respectively. The size of amplified products was assessed by 1.5 % agarose gel electrophoresis against the GeneRuler™ 100 nt Plus DNA Ladder (Fermentas).

Nucleotide sequencing and analysis

PCR products of the expected size were purified using a High Pure PCR Product Purification Kit (Roche Applied Science). Dye-terminator cycle sequencing was performed using the ABI PRISM Big Dye™ Terminator version 3.1 Cycle Sequencing Ready Reaction Kit (Applied Biosystems), with each of the PCR primers, at supplier-prescribed reaction conditions. Reactions were run on an ABI 3130 automated sequencer (Applied Biosystems), viewed with Chromas v. 1.43 and aligned to reference sequence DQ294633, a cloned Mengovirus, termed Rz-pMwt derived from the 'M' strain (A. Palmenberg, pers. comm.), in Mega v.5 (Tamura *et al.* 2011). Alignment of the overlapping fragments resulted in a contiguous full-length P1 gene sequence inclusive of the flanking leader (L) protein and 2A coding regions. Each of the resulting South African EMCV sequences, ranging from 3371 nucleotides (for the historical swine virus) to 3688 nucleotides (for the two elephant strains) in length, were deposited in the GenBank database under accession numbers, JN800421-3 and used in nucleotide Blast searches (www.ncbi.nlm.nih.gov/blast) to identify all available closely related sequences.

Similarly, the 3D sequences generated for the South African isolates, were submitted to GenBank under accession numbers JX102656-8 and were used to identify homologous sequences in the database.

Three datasets (P1, VP3/1 and 3D) were ultimately compiled for phylogenetic analysis and each contained all available, homologous EMCV data currently in GenBank. The final, aligned P1 dataset used for phylogenetic analysis and comprising the three South African isolates and 21 GenBank entries, was 3329 nucleotides in length, and had an average empirically-determined nucleotide base frequency of A = 25.2 %, C = 26.1 %, G = 22.6 % and T = 26.2 %. The transition:transversion ratio (R) was 2.0 and 993 of the 1143 variable sites were parsimony informative (Pi). The VP3/1 junction dataset, which was 715 nucleotides in length, comprised 42 taxa, had an average base frequency of A = 25.2 %, C = 26.0 %, G = 21.4 % and T = 27.4 %, an R of 1.71 and of the 307 variable sites, 274 were parsimony informative. The 248 nucleotide long 3D dataset consisted of 70 taxa, had an average empirically-determined nucleotide frequency of A = 27.4 %, C = 24.9 %, G = 20.6 % and T = 27.1 %, an R of 2.3 and 82 of the 103 variable sites were parsimony informative. The best-fit model selected under the Akaike Information Criterion (AIC) in Mega5, for each of the datasets was: the General Time Reversible (GTR) model, with a Gamma distribution shape parameter (G) of 1.78 and proportion of invariant sites (I) of 0.58 for the P1 dataset; the Tamura 3-parameter (T92) model, with a G of 3.08 and I of 0.526 for the VP3/1 dataset; and the T92 model with a G of 0.29 for the 3D dataset. The best-fit model identified for each of the datasets guided parameter selection for the Minimum evolution (ME) and Maximum Likelihood (ML) analyses performed in MEGA v. 5 (Tamura *et al.,* 2011) and PhyML v 3.0 (Guindon & Gascuel, 2003), respectively, and for the selection of priors for the Bayesian inference (BI) performed with Mr Bayes v 3.1 (Huelsenbeck & Ronquist, 2001). Nodal support for ME

and ML was assessed by 10,000 and 5,000 nonparametric bootstrap replications, respectively. For the BI analyses four chains were run for 10×10^6 generations using random starting trees with default heating and swap parameters. Resulting trees and parameters were recorded and split frequencies were compared every 1000^{th} generation to ensure convergence. Tracer plots were inspected to ensure that trees were sampled from the region of stationarity, and two independent runs were performed for each dataset.

Results

Phylogenetic inferences of all three genome regions confirmed the sister relationship between the 1994 KNP elephant outbreak strains and the original 'Mengo' virus isolates from Uganda (Figures 1-3). High levels of support were obtained for this grouping with all three gene regions and across all methods of phylogenetic inference. In contrast, although the P1 analysis (Fig. 1) confirmed the sister relationship of the 1979 swine outbreak strain from South Africa (SAR/1979) to a lineage that is comprised primarily of isolates derived from global (China, Korea, Belgium and Panama) swine EMCV outbreaks and of rodent strains from the USA (100% for node J), nodal support for this relationship was ≤ 75 % and ≤ 57 % in the VP3/1 (Fig. 2) and 3D (Fig. 3) phylogenies, respectively.

The P1 and VP3/1 phylogenies recovered three monophyletic EMCV lineages, corresponding to previously defined groups (Koenen et al. 1999). In keeping with the original description of these major groups, these lineages are denoted A to C (Figs 1-3), and were consistently recovered for all three gene regions analysed. A fourth lineage, tentatively denoted 'Lineage D', is only known from two unpublished 3D gene records that list Orang Utang as the host species (Fig. 3). Based on available 3D data, the lineages / groups vary in their geographical distribution, with lineage A occurring on five

continents, lineage B on three continents and lineages C and D each occurring on a single continent, Africa and Asia, respectively. Lineages A and B both have wide host ranges that include domestic pigs, murid rodents and primates, whereas lineage C (the mengovirus lineage), has only been found in monkeys from Uganda and in African elephants from South Africa. Average within-lineage nucleotide distances for lineages A-C ranged from 5.5 % (within lineage B) to 9.8% (within lineage A) for the VP3/1 junction region, and from 2.2 % (within lineage B) to 6.6 % (within lineage C), for the 3D region. Between-lineage nucleotide distances ranged from 20.6% (between lineage A and B) to 22.8 % (between lineages B and C) for VP3/1, and from 12.4% (between lineage A and B) to 15.5 % (between lineages B and C) for the 3D region.

Comparison of the two elephant viruses, designated KNP17/94 and KNP19/94, revealed five variable nucleotide sites across the P1 and flanking genome region characterised, corresponding to an overall nucleotide sequence identity of 99.84 % across the ~3,300 nucleotide region. The five variable sites all occurred at third base positions and resulting in silent mutations at amino acid position 55 of the Leader (L) protein; amino acid positions 54, 157 and 256 of the VP2 (1B) gene, and at amino acid position 158 of the VP1 (1D) gene. For the VP3/1 and 3D regions analysed, the pairwise nucleotide sequence identity was 99.86 % and 100 %, respectively for the two elephant isolates.

The elephant isolates, KNP/17/94 and KNP/19/94, which form a sister clade to mengovirus isolates from Rhesus macaques (*Macaca mulatta*) in Uganda, differed by 13.3 % from the 'M' strain (L22089), across the P1 genome region characterised. Pairwise nucleotide sequence differences between each of the elephant strains and EMC viruses falling outside of lineage C (the elephant-mengovirus lineage) was ≥ 19 % across the P1 region. Nucleotide sequence differences between elephant and mengovirus strains were 15 % and 11 % across the VP3/1 and 3D regions, respectively. For the historical pig

virus, SAR/1979, pairwise nucleotide sequence similarity was highest (>89.9%) across the ~3.3 kbp (P1/flanking gene) region, to three lineage A EMCV strains (DQ288856, DQ717424, AF356822) of diverse geographic origin (USA, Korea and Belgium). For the VP1/3 and 3D genome regions for which substantially more data were available, SAR/1979 was most closely related (89 % sequence identity) to four isolates from China and Korea (HM641897, FJ604852, DQ517424, DQ464063) across the VP3/1 genome region, and had the highest nucleotide sequence identity (96 %) across the 3D gene region, to a domestic pig virus from an outbreak that occurred in Greece in 1997 (AJ235726).

Discussion

Reports of clinical EMC and of genomic characterisation of cardioviruses from Africa and southern Africa in particular (Dick 1949, Williams 1981, Grobler et al. 1995, Jones et al. 2011), are rare. Whereas numerous strains have been isolated and characterised globally from outbreaks in domestic pigs (Maurice et al. 2005) and from putative murid rodent hosts in close proximity (Billinis 2009), reports of the disease in wild and exotic animals are sporadic and generally limited to captive populations in North America, Australia and Europe and their associated rodent populations (Wells et al. 1989; Redacliff et al. 1997; Canelli et al. 2010). In contrast, Africa is the only continent from which outbreaks of the disease have been reported from a population of semi-wild primates (Jones et al. 2011) or from a natural population of elephants (Grobler et al. 1995). It is also the origin of one of the first EMC virus isolates, *viz.* the Mengo virus 'M' strain from Rhesus macaques in Uganda (Dick et al. 1949). Despite a long presence on the continent, isolated reports of the disease in pigs are confined to Cote d'Ivoire and to South Africa. In the present study, three EMCV isolates from outbreaks in domestic and wild animals in

southern Africa were genetically characterised for the first time. The 1979 outbreak which occurred in pigs was found to cluster within lineage A which is comprised of viruses of a wide geographical and host range, whereas the elephant isolates grouped with mengoviruses within lineage C. The 3D data revealed the South African pig virus to be most closely related to a 1997 domestic pig virus from Greece. At the time of the outbreak swine imports into South Africa was restricted from the USA due to an outbreak of Aujeszky's Disease and importation from European countries was suspended due to possible outbreaks of Classical Swine fever and Swine Vesicular Disease (Records of the South African Department of Agriculture). The geographic origin of the EMC strain recorded in South Africa at a time when importation of pigs and pig products was under scrutiny, therefore remains unclear. However, the availability of genomic data for a suid strain from Africa, together with the continued generation of homologous data is, likely to provide insights into probable origin and route of importation of this virus.

The origin of the two nearly identical elephant viruses is likewise unclear, but as the outbreak coincided with a rodent irruption and was confirmed by PCR in endemic *Mastomys* rodents from the same game park (Grobler et al. 1995), it is likely that the virus, which was most closely related to an isolate of East African origin (Fig 1-3), represents a regionally endemic lineage. Serological screening of sera from a range of rodents captured at the same time as the elephant outbreak revealed that *Mastomys* rodents had the highest seroprevalence (38 %), implicating members of this rodent genus as likely reservoirs of infection (Grobler et al. 1995). In addition, unpublished records at NICD indicate that 26 EMCV isolates were made from rodents, mainly *Mastomys*, collected in surveys at four widely separated locations in South Africa, between 1961 to 1968. Although the rodent isolates were not available for inclusion in the present study, the implication is clear that there is endemic circulation of virus in the country. Future

studies should therefore be directed at determining the rodent reservoir host species, using a combination of serology, virus isolation and nucleotide sequencing in order to determine the prevalence and diversity of EMC viruses in endemic South African rodents.

Acknowledgements

We would like to thank Dr. Gerdes of the ARC-OVI for performing the original virus isolations and for providing an EMCV RNA extract for positive control purposes. The project was funded by a DST-NRF South African Research Chair of Behavioural Ecology and Physiology awarded to NCB.

References

Bastos, A.D.S., 1998. Detection and characterisation of foot-and-mouth disease virus in sub-Saharan Africa. Onderstepoort J. Vet. Res. 65, 37-47.

Bautista, G., Jungeblut, C.W., Kodza, H., 1957. Nonspecific cross protection in hampsters between Columbia SK virus and poliomyelitis virus type II (MEF) virus infection. J. Infect. Dis. 100, 268-272.

Billinis, C., 2009. Encepahlomyocarditis virus infection in wildlife speices in Greece. J. Wildlife Dis. 45, 522-526.

Canelli, E., Luppi, A., Lavazza, A., Lelli, D., Sozzi, E., Moreno Martin, A. M., Gelmetti, D., Pascotto, E., Sandri, C., Magnone, W., Cordioli, P., 2010. Encephalomyocarditis virus infection in an Italian zoo. Virol. J. **7,** 64.

Craighead, J.E., 1965. Some properties of the encepahlomyocarditis, Columbia SK and Mengo viruses. Exp. Biol. Med. 119, 408-412.

De Bruyn, P.J.N., Bastos, A.D.S., Eadie, C., Tosh, C., Bester, M.N., 2008. Mass mortality of adult male subantarctic fur seals: are alien mice the culprits? PLoS ONE 3, e3757.

Denis, P., Liebig, H.D., Nowotny, N., Billinis, C., Papadopoulos, O., O'Hara, R.S., Knowles, N.J., Koenen, F., 2006. Genetic variability of *encephalomyocarditis virus* (EMCV) isolates. Vet. Microbiol. 113, 1-12.

Dick, G.W., Best, A.M., Haddown, A.J., Smithburn, K.C., 1948. Mengo encephalomyocaridits; a hirtheto unknown virus affecting man. Lancet 6251, 286-289.

Dick, G.W.A., 1949. The relationship of mengo encephalomyelitis, encephalomyocarditis Columbia-SK and M.M. viruses. J. Immunol. 62, 375-386.

Drake, J. W., Charlesworth, B., Charlesworth, D., Crow, J. F., 1998. Rates of spontaneous mutation. Genetics 148, 1667–1686.

Eigen, M., 1996. On the nature of virus quasispecies. Trends Microbiol. 4, 216-218.

Grobler, D.G, Raath, J.P., Braack, L.E.O, Keet, D.F., Gerdes, G.H., Barnard, B.J.H., Kriek, N.P.J., Jardine, J., Swanepoel, R., 1995. An outbreak of encephalomyocarditis-virus infection in free ranging African elephants in the Kruger National Park. Onderstepoort J. Vet. Res. 62, 97-108.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biol. 52, 696-704.

Huelsenbeck, J.P., Ronquist, F., 2001. MrBayes: Bayesian inferenc of phylogenetic trees. Bioinformatics 17, 754–755.

Hughes, A.L., 2004. Phylogeny of the Picornaviridae and differential evolutionary divergence of picornavirus proteins. Infection, Genet. Evol. 4, 143-152.

Jones, P., Cordonnier, N., Mahamba, C., Burt, F.J., Rakotovao, F., Swanepoel, R., André, C., Dauger, S. Bakkali Kassimi, L., 2011. Encephalomyocarditis virus mortality in semi-wild bonobos (*Pan panicus*). J. Med. Primatol. 40, 157-163.

Knowles, N.J., Dickinson, N.D., Wilsden, G., Carra, E., Brocchi, E., De Simone, F., 1998. Molecular analysis of encephalomyocaridits viruses isolated from pigs and rodents in italy. Virus Res. 57, 53-62.

Koenen, F., Vanderhallen, H., Dickinson, N.D., Knowles, N.J., 1999. Phylogenetic analysis of European encephalomyocarditis viruses: comparison of two genomic regions. Arch. Virol. 144, 893-903.

Maurice, H., Nielen, M., Brocchi, E., Nowotny, N., Bakkali-Kassimi, L., Billinis, C., Loukaides, P., O`Hara, R.S., Koenen, F., 2005. The occurence of encephalomyocarditis virus (EMCV) in European pigs from 1990 to 2001. Epidemiol. Infect. 133, 547-557.

Meyer, R.F., Brown, C.C., House, C., House, J.A., Molitor, T.W., 1991. Rapid and sensitive detection of foot-and-mouth disease virus in tissues by enzymatic RNA amplification of the polymerase gene. J. Virol. Methods, 34, 161-172.

Minor, P.D., Brown, F., Domingo, E., Holy, E., King, A., Knowles, N., Lemon, S., Palmenberg, A., Rueckert, R., Stanway, G., Wimmer, E., Yin-Murphy, M., 1995. Picornaviridae. In: F.A. Murphy, C.M. Fauguet, D.H.L. Bishop, S.A. Chabrial, A.V. Jarvis (Eds), Virus Taxonomy. Classification and Nomenclature of Viruses. Sixth Report of the International Committee on Taxonomy of Viruses, Springer, Vienna, pp. 329–336.

Oberste, M.S., Gotuzzo, E., Blair, P., Nix, W.A., Ksiazek, T.G., Comer, J.A., Rollin, P., Goldsmith, C.S., Olson, J. and Kochel, T.J., 2009. Human febrile illness caused by encephalomyocarditis virus infection, Peru. Emerg Infect Dis, 15, 640-646.

Reddacliff, L.A, Kirkland, P.D., Hartley, W.H., Reece, R.L., 1997. Encephalomyocarditis virus infections in an Australian Zoo. J. Zoo. Wildl. Med. 28, 153-157.

Scraba, D.G., Palmenberg, A.C. 1999. Cardioviruses (Picornaviridae). In: R.G. Webster, A. Granoff (Eds), Encyclopedia of Virology; Academic Press, London, pp 1–10.

Seaman, J.T., Finnie, E.P., 1987. Acute myocarditis in a captive African elephant *(Loxodonta africana*). J. Widl. Dis. 23, 170-171.

Simmonds, P., 2006. Recombination and selection in the evolution of picornaviruses and other mammalian positive-stranded RNA viruses. J. Virol. 80, 11124–11140.

Simpson, C.F., Lewis, A.L., Gaskin, J.M., 1977. Encephalomyocaarditis virus infection in captive elephants. J Am Vet Med Assoc 17, 902-905.

Tamura, K. Peterson, D. Peterson, N. Stecher, G. Nei, M., Kumar, S., 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731-2739.

Tesh, R.B., 1978. The prevalence of encephalomyocarditis virus neutralising antibodies among various human populations. Am. J. Trop. Med. Hyg. 27, 144-149.

Tesh, R.B., Wallace, G.D., 1978. Observations on the natural history of encephalomyocarditis virus. Am. J. Trop. Med. Hyg. 27, 133-143.

Thomson, G.R., Bengis, R.G., Brown, C.C., 2001. Picornavirus Infections. In: E.S. Williams, I.K. Barker (Eds), Infectious diseases of wild mammals, 3rd edn, Iowa State University Press, Iowa, pp. 119-130.

Vanderhallen, H., Koenen, F., 1998. Identification of encephalomyocarditis virus in clinical samples by reverse transcription-PCR followed by genetic typing using sequence analysis. J. Clin. Microbiol. 36, 3463-3467.

Wells, S.K., Gutter, A.E., Soike, K.F., Baskin, G.B., 1989. Encephalomyocarditis virus: Epizootic in a zoological collection. J. Zoo. Wildl. Med. 20, 291-296.

Williams, M.C., 1981. Encephalomyocarditis virus infection. J. S. Afr. Vet. Assoc. 52, 76.

Zimmerman, J.J., 1994. Encephalomyocarditis. In: G.W. Beran & J.H. Steele (Eds), Handbook of Zoonoses, 2nd edition, CPR Press, Florida, pp. 423-436.

Table 1: Summary of the primers designed to amplify overlapping fragments of the P1 and flanking genome regions, and the partial 3D gene region of the three historical South African isolates characterised in this study.

PCR primer pairs: Primer name and sequence (5' – 3') of the forward (F) and reverse (R) primers	**Primer binding site***	**Virus strain/s amplified**	**Expected size of amplicon**	**Sequence span***	**Genome region targeted**
EMC-AB1: GGCCGAAGCCGCTTGGAATA (F) EMC-AB2: ACGTGGCTTTTGGCCGCAGA (R)	213-232 478-497	SAR/1979 KNP/17/94 KNP/19/94	285nt	233-477	5'UTR
EMC- AB1: GGCCGAAGCCGCTTGGAATA (F) 1356R-AB: TGGGTGTTTGTGACCGTGTT (R)	213-232 1224-1243	SAR/1979 KNP/17/94 KNP/19/94	1031nt	233-1223	5'UTR-1B
386F-JvS: GGTCTGTTGAATGTCGTGAA (F) 884R-AB: GTTCCATGGTTGTAGCCAT (R)	366-385 753-771	SAR/1979 KNP/17/94 KNP/19/94	406nt	386-752	5'UTR-L
687F-AB: GCTCTCCTCAAGCGTATTCA (F) 1356R-AB: TGGGTGTTTGTGACCGTGTT (R)	575-594 1224-1243	SAR/1979 KNP/17/94 KNP/19/94	669nt	595-1223	L-1B
687F-AB: GCTCTCCTCAAGCGTATTCA (F) 2868R-AB: CCAATGGGACTAGACCTATCATA (R)	575-594 2733-2755	SAR/1979	2181nt	595-2732	L-1C
1156EPF-JvS: GGAGAACTTGTCTGATCGAGT (F) MasRAB-2392R: AGGCACAGTGAAGGAGTAAG (R)	1184-1204 2401-2420	SAR/1979	1237nt	800-900	1B-1D
1307F-AB: TCAGACCGAGTGTCTCAAGA (F) 1720R-AB: CTTGGACCATCTGTTGTCCAT (R)	1194-1213 1587-1607	KNP/17/94 KNP/19/94	414nt	1214-1586	1B
1307F-AB: TCAGACCGAGTGTCTCAAGA (F) 2113R-AB: CAGTTCCCACCATGCGGAAGTG (R)	1194-1213 2436-2457	KNP/17/94 KNP/19/94	1264nt	1214-2435	1B-1C
1347F-AB: CGGTCACAAACACCCARTCAAC (F) 2113R-AB: CAGTTCCCACCATGCGGAAGTG (R)	1228-1249 2436-2457	KNP/17/94 KNP/19/94	1230nt	1250-2435	1B-1C
1904F-AB: GCTTCCTGGACTTTGGTGAT (F) 2868R-AB: CCAATGGGACTAGACCTATCATA (R)	1791-1810 2733-2755	SAR/1979	965nt	1811-2732	1B-1D
2246MF-JvS: GTACCGTGGATCACTAGTCTA (F) 3204MR-JvS: GACAGCAGGTAGGACAGACAA (R)	2249-2269 3213-3233	KNP/17/94 KNP/19/94	985nt	2270-3212	1C-1D
VP3-FM: CACTTCCGCATGGTGGGAACTG (F) 2B-RM: CGGCAGTAGGGTTTGAGCCATT (R)	2736-2457 3947-3968	SAR/1979 KNP/17/94 KNP/19/94	1233nt	2458-3946	1D-2B
EMC-3DfAB: TCAGGTTGTGCAGCGACCTC (F) EMC-3DrAB: CTTACCGGGTAACGCGTTGT (R)	7128-7147 7651-7670	SAR/1979 KNP/17/94 KNP/19/94	318nt	7313-7630	3D

*Positions based on Mengovirus isolate Rz-pMwt, GenBank accession number DQ294633.

Figure 1: Neighbour Joining (NJ) tree indicating EMC virus relationships based on complete P1 and flanking L and 2A nucleotide data. Nodal support values ≥ 50 % from Minimum Evolution (ME), Maximum likelihood (ML) and Bayesian inference (BI) analyses are summarized for each of the nodes (labeled A-S) as follows: NJ / ML / ME / BI. Sequences highlighted in grey indicate isolates that are common to all three genome region phylogenies, whilst those generated in this study are indicated in bold. Taxon labels comprise the GenBank accession number, followed in brackets by the isolate name, host species, country of origin and year of isolation. * Indicates a laboratory strain (Rz-pMwt, derived from the 'M' mengo virus isolate).

Figure 2: Neighbour Joining (NJ) tree inferred using VP3/1 junction data of 42 EMC viruses. Nodal support values ≥ 50 % from Minimum Evolution (ME), Maximum likelihood (ML) and Bayesian inference (BI) analyses are summarized for each of the nodes (labeled A-X) as follows: NJ / ML / ME / BI. Sequences highlighted in grey indicate isolates that are common to all three genome region phylogenies, whilst those generated in this study are indicated in bold. Taxon labels comprise the GenBank accession number, followed in brackets by the isolate name, host species, country of origin and year of isolation. * Indicates a laboratory strain (Rz-pMwt, derived from the 'M' mengo virus isolate).

Figure 3: Neighbour Joining (NJ) tree inferred using partial 3D gene data. Nodal support values ≥ 50 % from Minimum Evolution (ME), Maximum likelihood (ML) and Bayesian inference (BI) analyses are summarized for each of the nodes (labeled A-T) as follows: NJ / ML / ME / BI. Sequences highlighted in grey indicate isolates that are common to all three genome region phylogenies, whilst those generated in this study are indicated in bold. Taxon labels comprise the GenBank accession number, followed in brackets by the isolate name, host species, country of origin and year of isolation. * Indicates a laboratory strain (Rz-pMwt, derived from the 'M' mengo virus isolate).

Fig. 1

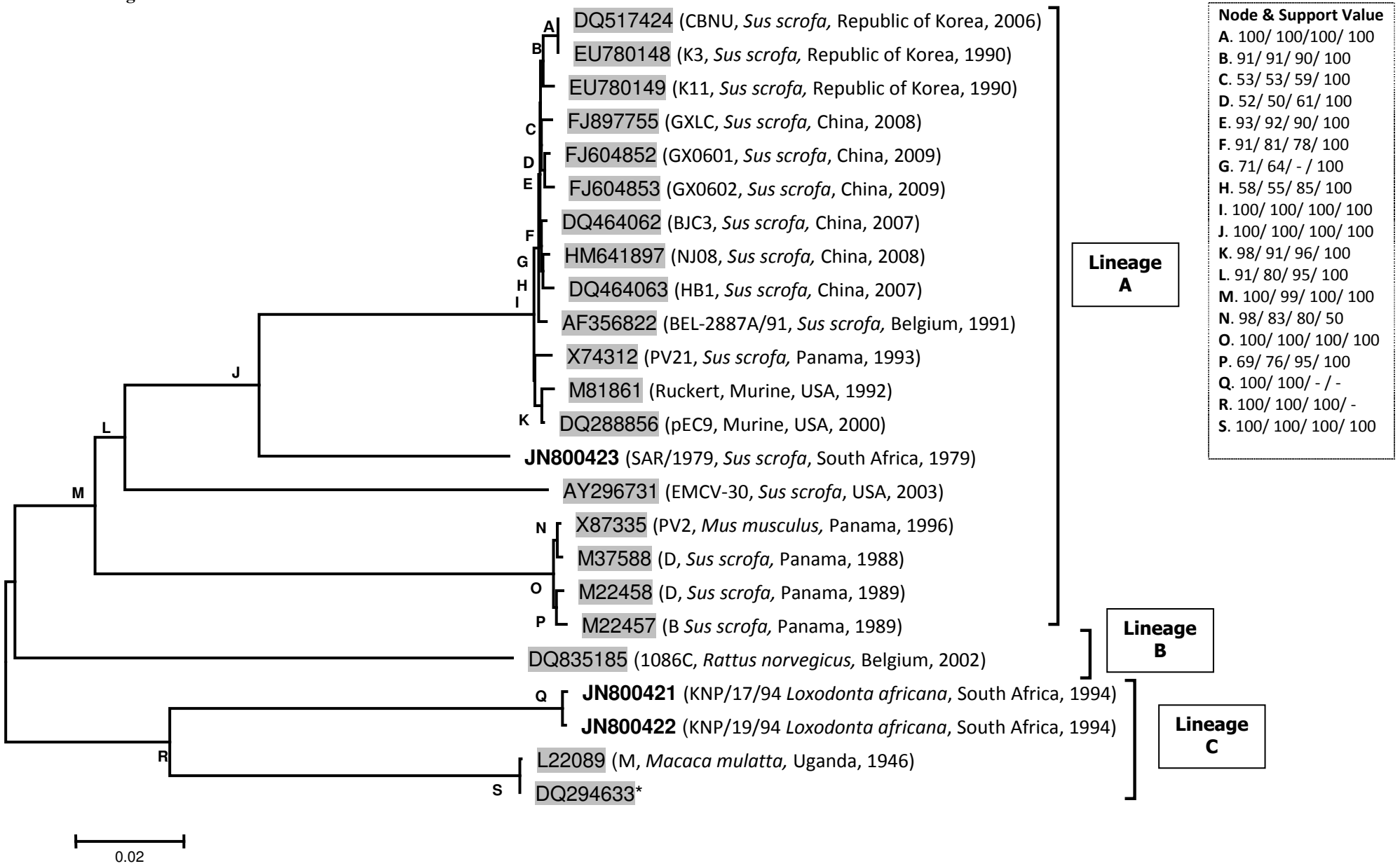
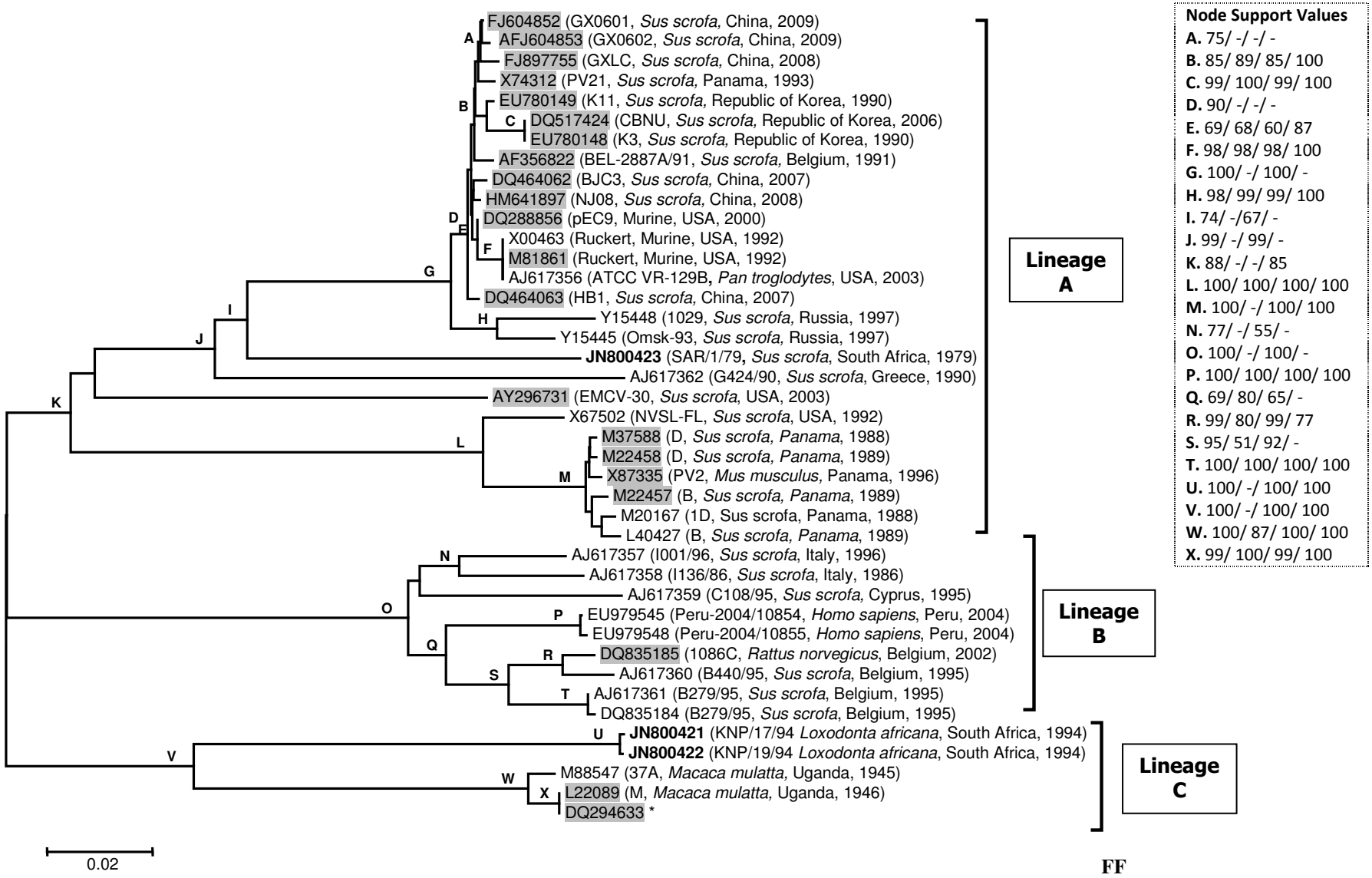
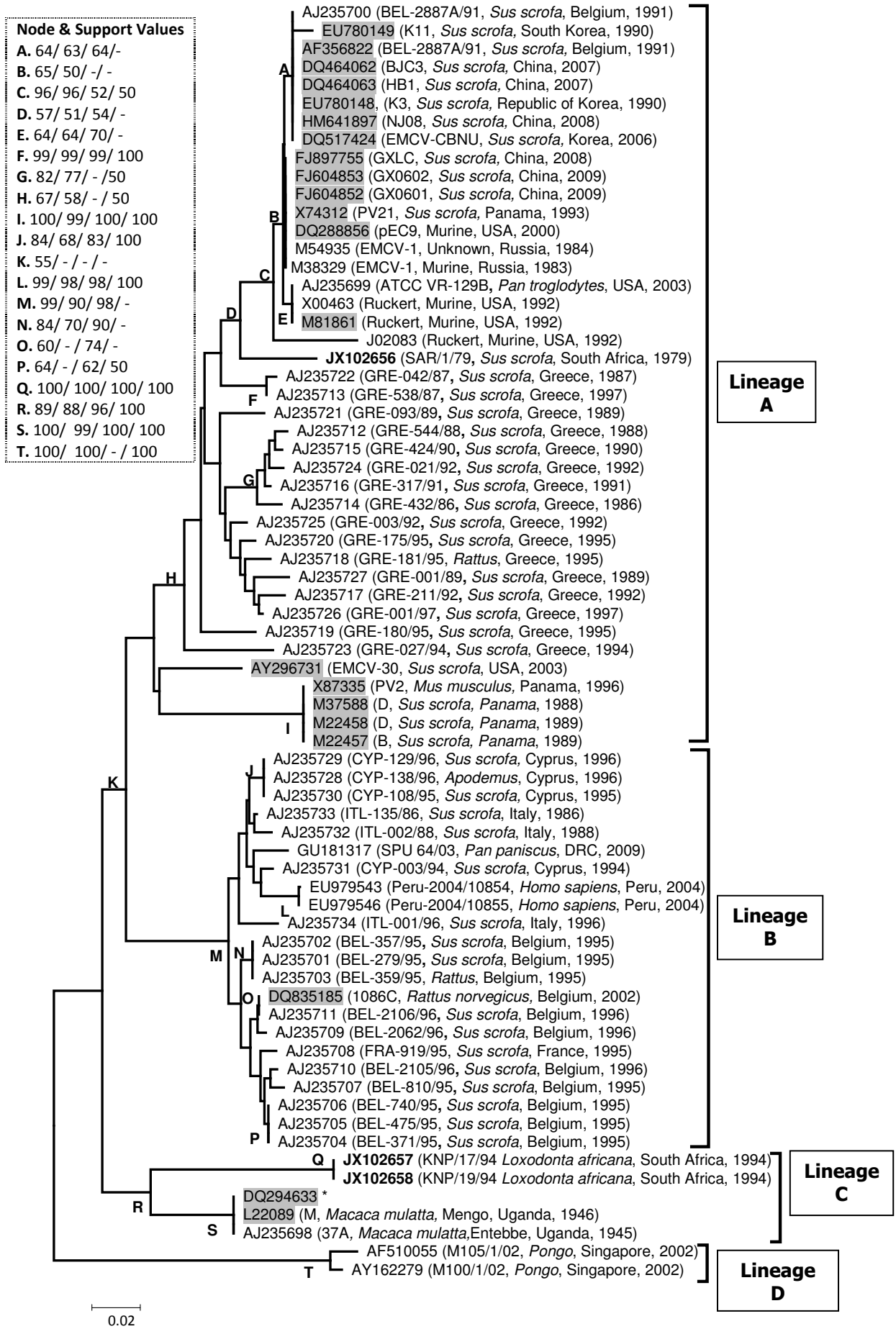

A DQ517424 (CBNU, Sus scrofa, Republic of Korea, 2006)
B EU780148 (K3, Sus scrofa, Republic of Korea, 1990)
EU780149 (K11, Sus scrofa, Republic of Korea, 1990)
C FJ897755 (GXLC, Sus scrofa, China, 2008)
D FJ604852 (GX0601, Sus scrofa, China, 2009)
E FJ604853 (GX0602, Sus scrofa, China, 2009)
DQ464062 (BJC3, Sus scrofa, China, 2007)
F HM641897 (NJ08, Sus scrofa, China, 2008)
G DQ464063 (HB1, Sus scrofa, China, 2007)
H
I AF356822 (BEL-2887A/91, Sus scrofa, Belgium, 1991)
X74312 (PV21, Sus scrofa, Panama, 1993)
J M81861 (Ruckert, Murine, USA, 1992)
K DQ288856 (pEC9, Murine, USA, 2000)
L JN800423 (SAR/1979, Sus scrofa, South Africa, 1979)
M AY296731 (EMCV-30, Sus scrofa, USA, 2003)
N X87335 (PV2, Mus musculus, Panama, 1996)
M37588 (D, Sus scrofa, Panama, 1988)
O M22458 (D, Sus scrofa, Panama, 1989)
P M22457 (B Sus scrofa, Panama, 1989)
DQ835185 (1086C, Rattus norvegicus, Belgium, 2002)
Q JN800421 (KNP/17/94 Loxodonta africana, South Africa, 1994)
JN800422 (KNP/19/94 Loxodonta africana, South Africa, 1994)
R L22089 (M, Macaca mulatta, Uganda, 1946)
S DQ294633*
0.02
Lineage A
Lineage B
Lineage C
Node & Support Value
A. 100/ 100/100/ 100
B. 91/ 91/ 90/ 100
C. 53/ 53/ 59/ 100
D. 52/ 50/ 61/ 100
E. 93/ 92/ 90/ 100
F. 91/ 81/ 78/ 100
G. 71/ 64/ - / 100
H. 58/ 55/ 85/ 100
I. 100/ 100/ 100/ 100
J. 100/ 100/ 100/ 100
K. 98/ 91/ 96/ 100
L. 91/ 80/ 95/ 100
M. 100/ 99/ 100/ 100
N. 98/ 83/ 80/ 50
O. 100/ 100/ 100/ 100
P. 69/ 76/ 95/ 100
Q. 100/ 100/ - / -
R. 100/ 100/ 100/ -
S. 100/ 100/ 100/ 100

Fig. 2


FJ604852 (GX0601, Sus scrofa, China, 2009)
AFJ604853 (GX0602, Sus scrofa, China, 2009)
FJ897755 (GXLC, Sus scrofa, China, 2008)
X74312 (PV21, Sus scrofa, Panama, 1993)
EU780149 (K11, Sus scrofa, Republic of Korea, 1990)
DQ517424 (CBNU, Sus scrofa, Republic of Korea, 2006)
EU780148 (K3, Sus scrofa, Republic of Korea, 1990)
AF356822 (BEL-2887A/91, Sus scrofa, Belgium, 1991)
DQ464062 (BJC3, Sus scrofa, China, 2007)
HM641897 (NJ08, Sus scrofa, China, 2008)
DQ288856 (pEC9, Murine, USA, 2000)
X00463 (Ruckert, Murine, USA, 1992)
M81861 (Ruckert, Murine, USA, 1992)
AJ617356 (ATCC VR-129B, Pan troglodytes, USA, 2003)
DQ464063 (HB1, Sus scrofa, China, 2007)
Y15448 (1029, Sus scrofa, Russia, 1997)
Y15445 (Omsk-93, Sus scrofa, Russia, 1997)
JN800423 (SAR/1/79, Sus scrofa, South Africa, 1979)
AJ617362 (G424/90, Sus scrofa, Greece, 1990)
AY296731 (EMCV-30, Sus scrofa, USA, 2003)
X67502 (NVSL-FL, Sus scrofa, USA, 1992)
M37588 (D, Sus scrofa, Panama, 1988)
M22458 (D, Sus scrofa, Panama, 1989)
X87335 (PV2, Mus musculus, Panama, 1996)
M22457 (B, Sus scrofa, Panama, 1989)
M20167 (1D, Sus scrofa, Panama, 1988)
L40427 (B, Sus scrofa, Panama, 1989)
AJ617357 (I001/96, Sus scrofa, Italy, 1996)
AJ617358 (I136/86, Sus scrofa, Italy, 1986)
AJ617359 (C108/95, Sus scrofa, Cyprus, 1995)
EU979545 (Peru-2004/10854, Homo sapiens, Peru, 2004)
EU979548 (Peru-2004/10855, Homo sapiens, Peru, 2004)
DQ835185 (1086C, Rattus norvegicus, Belgium, 2002)
AJ617360 (B440/95, Sus scrofa, Belgium, 1995)
AJ617361 (B279/95, Sus scrofa, Belgium, 1995)
DQ835184 (B279/95, Sus scrofa, Belgium, 1995)
JN800421 (KNP/17/94 Loxodonta africana, South Africa, 1994)
JN800422 (KNP/19/94 Loxodonta africana, South Africa, 1994)
M88547 (37A, Macaca mulatta, Uganda, 1945)
L22089 (M, Macaca mulatta, Uganda, 1946)
DQ294633 *
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Lineage A
Lineage B
Lineage C
0.02
Node Support Values
A. 75/ -/ -/ -
B. 85/ 89/ 85/ 100
C. 99/ 100/ 99/ 100
D. 90/ -/ -/ -
E. 69/ 68/ 60/ 87
F. 98/ 98/ 98/ 100
G. 100/ -/ 100/ -
H. 98/ 99/ 99/ 100
I. 74/ -/67/ -
J. 99/ -/ 99/ -
K. 88/ -/ -/ 85
L. 100/ 100/ 100/ 100
M. 100/ -/ 100/ 100
N. 77/ -/ 55/ -
O. 100/ -/ 100/ -
P. 100/ 100/ 100/ 100
Q. 69/ 80/ 65/ -
R. 99/ 80/ 99/ 77
S. 95/ 51/ 92/ -
T. 100/ 100/ 100/ 100
U. 100/ -/ 100/ 100
V. 100/ -/ 100/ 100
W. 100/ 87/ 100/ 100
X. 99/ 100/ 99/ 100

FF

Fig. 3


Node & Support Values
A. 64/ 63/ 64/-
B. 65/ 50/ -/ -
C. 96/ 96/ 52/ 50
D. 57/ 51/ 54/ -
E. 64/ 64/ 70/ -
F. 99/ 99/ 99/ 100
G. 82/ 77/ - /50
H. 67/ 58/ - / 50
I. 100/ 99/ 100/ 100
J. 84/ 68/ 83/ 100
K. 55/ - / - / -
L. 99/ 98/ 98/ 100
M. 99/ 90/ 98/ -
N. 84/ 70/ 90/ -
O. 60/ - / 74/ -
P. 64/ - / 62/ 50
Q. 100/ 100/ 100/ 100
R. 89/ 88/ 96/ 100
S. 100/ 99/ 100/ 100
T. 100/ 100/ - / 100
AJ235700 (BEL-2887A/91, Sus scrofa, Belgium, 1991)
EU780149 (K11, Sus scrofa, South Korea, 1990)
AF356822 (BEL-2887A/91, Sus scrofa, Belgium, 1991)
DQ464062 (BJC3, Sus scrofa, China, 2007)
DQ464063 (HB1, Sus scrofa, China, 2007)
EU780148, (K3, Sus scrofa, Republic of Korea, 1990)
HM641897 (NJ08, Sus scrofa, China, 2008)
DQ517424 (EMCV-CBNU, Sus scrofa, Korea, 2006)
FJ897755 (GXLC, Sus scrofa, China, 2008)
FJ604853 (GX0602, Sus scrofa, China, 2009)
FJ604852 (GX0601, Sus scrofa, China, 2009)
X74312 (PV21, Sus scrofa, Panama, 1993)
DQ288856 (pEC9, Murine, USA, 2000)
M54935 (EMCV-1, Unknown, Russia, 1984)
M38329 (EMCV-1, Murine, Russia, 1983)
AJ235699 (ATCC VR-129B, Pan troglodytes, USA, 2003)
X00463 (Ruckert, Murine, USA, 1992)
M81861 (Ruckert, Murine, USA, 1992)
J02083 (Ruckert, Murine, USA, 1992)
JX102656 (SAR/1/79, Sus scrofa, South Africa, 1979)
AJ235722 (GRE-042/87, Sus scrofa, Greece, 1987)
AJ235713 (GRE-538/87, Sus scrofa, Greece, 1997)
AJ235721 (GRE-093/89, Sus scrofa, Greece, 1989)
AJ235712 (GRE-544/88, Sus scrofa, Greece, 1988)
AJ235715 (GRE-424/90, Sus scrofa, Greece, 1990)
AJ235724 (GRE-021/92, Sus scrofa, Greece, 1992)
AJ235716 (GRE-317/91, Sus scrofa, Greece, 1991)
AJ235714 (GRE-432/86, Sus scrofa, Greece, 1986)
AJ235725 (GRE-003/92, Sus scrofa, Greece, 1992)
AJ235720 (GRE-175/95, Sus scrofa, Greece, 1995)
AJ235718 (GRE-181/95, Rattus, Greece, 1995)
AJ235727 (GRE-001/89, Sus scrofa, Greece, 1989)
AJ235717 (GRE-211/92, Sus scrofa, Greece, 1992)
AJ235726 (GRE-001/97, Sus scrofa, Greece, 1997)
AJ235719 (GRE-180/95, Sus scrofa, Greece, 1995)
AJ235723 (GRE-027/94, Sus scrofa, Greece, 1994)
AY296731 (EMCV-30, Sus scrofa, USA, 2003)
X87335 (PV2, Mus musculus, Panama, 1996)
M37588 (D, Sus scrofa, Panama, 1988)
M22458 (D, Sus scrofa, Panama, 1989)
M22457 (B, Sus scrofa, Panama, 1989)
AJ235729 (CYP-129/96, Sus scrofa, Cyprus, 1996)
AJ235728 (CYP-138/96, Apodemus, Cyprus, 1996)
AJ235730 (CYP-108/95, Sus scrofa, Cyprus, 1995)
AJ235733 (ITL-135/86, Sus scrofa, Italy, 1986)
AJ235732 (ITL-002/88, Sus scrofa, Italy, 1988)
GU181317 (SPU 64/03, Pan paniscus, DRC, 2009)
AJ235731 (CYP-003/94, Sus scrofa, Cyprus, 1994)
EU979543 (Peru-2004/10854, Homo sapiens, Peru, 2004)
EU979546 (Peru-2004/10855, Homo sapiens, Peru, 2004)
AJ235734 (ITL-001/96, Sus scrofa, Italy, 1996)
AJ235702 (BEL-357/95, Sus scrofa, Belgium, 1995)
AJ235701 (BEL-279/95, Sus scrofa, Belgium, 1995)
AJ235703 (BEL-359/95, Rattus, Belgium, 1995)
DQ835185 (1086C, Rattus norvegicus, Belgium, 2002)
AJ235711 (BEL-2106/96, Sus scrofa, Belgium, 1996)
AJ235709 (BEL-2062/96, Sus scrofa, Belgium, 1996)
AJ235708 (FRA-919/95, Sus scrofa, France, 1995)
AJ235710 (BEL-2105/96, Sus scrofa, Belgium, 1996)
AJ235707 (BEL-810/95, Sus scrofa, Belgium, 1995)
AJ235706 (BEL-740/95, Sus scrofa, Belgium, 1995)
AJ235705 (BEL-475/95, Sus scrofa, Belgium, 1995)
AJ235704 (BEL-371/95, Sus scrofa, Belgium, 1995)
JX102657 (KNP/17/94 Loxodonta africana, South Africa, 1994)
JX102658 (KNP/19/94 Loxodonta africana, South Africa, 1994)
DQ294633 *
L22089 (M, Macaca mulatta, Mengo, Uganda, 1946)
AJ235698 (37A, Macaca mulatta,Entebbe, Uganda, 1945)
AF510055 (M105/1/02, Pongo, Singapore, 2002)
AY162279 (M100/1/02, Pongo, Singapore, 2002)
Lineage A
Lineage B
Lineage C
Lineage D
0.02