ORIGINAL ARTICLE

Genetic heterogeneity in the leader and P1-coding regions of foot-and-mouth disease virus serotypes A and O in Africa

M. Chitray · T. A. P. de Beer · W. Vosloo ·
F. F. Maree

Abstract Genetic information regarding the leader (L) and complete capsid-coding (P1) region of FMD serotype A and O viruses prevalent on the African continent is lacking. Here, we present the complete L-P1 sequences for eight serotype A and nine serotype O viruses recovered from FMDV outbreaks in East and West Africa over the last 33 years. Phylogenetic analysis of the P1 and capsid-coding regions revealed that the African isolates grouped according to serotype, and certain clusters were indicative of transboundary as well as intra-regional spread of the virus. However, similar analysis of the L region revealed random groupings of isolates from serotypes O and A. Comparisons between the phylogenetic trees derived from the structural coding regions and the L region pointed to a possibility of genetic recombination. The intertypic nucleotide and amino acid variation of all the isolates in this study supported results from previous studies where the externally located 1D was the most variable whilst the internally located 1A was the most conserved, which likely reflects the selective pressures on these proteins. Amino acids identified previously as important for FMDV structure and functioning were found to be highly conserved. The information gained from this study will contribute to the construction of structurally designed FMDV vaccines in Africa.

M. Chitray (✉) · W. Vosloo · F. F. Maree
Agricultural Research Council, Onderstepoort Veterinary Institute, Transboundary Animal Diseases, Private Bag X05, Onderstepoort, Pretoria, South Africa
e-mail: chitraym@arc.agric.za

M. Chitray · W. Vosloo · F. F. Maree
Department of Veterinary Tropical Diseases, Faculty of Veterinary Sciences, University of Pretoria, Onderstepoort 0110, South Africa

T. A. P. de Beer
Bioinformatics and Computational Biology Unit, University of Pretoria, Pretoria 0002, South Africa

Present Address:
T. A. P. de Beer
EMBL, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK

Present Address:
W. Vosloo
Australian Animal Health Laboratory, Geelong, Victoria, Australia

Introduction

Foot-and-mouth disease (FMD) is a highly contagious disease that affects domestic and wild cloven-hoofed animals [2, 77]. Despite all the information accumulated over the years on many aspects of FMD basic biology, there is still a lack of information regarding FMD virus transmission, maintenance, virulence and host range. Although FMD is referred to as a single disease [18], the causative agent of the disease, FMD virus (FMDV), consists of seven immunologically distinct serotypes [23, 24]. The FMDV serotypes, i.e., A, O, C, Asia 1 and the South African Territories (SAT) types 1, 2 and 3, have different global geographical distribution patterns [8–10, 18, 44, 73, 88] and are endemic in many countries. Even on the African continent, the distribution of serotypes is variable, with the SAT serotypes occurring in most regions of sub-Saharan Africa but A and O confined mostly to the central and northern parts of the region [88]. Mortality is usually low,

but morbidity can reach 100 % and therefore remains a major economic concern for livestock health in many developing countries and a continued threat to disease-free countries [44]. The eradication and control of FMDV in Africa is complex and difficult due to the role of wildlife in virus spread and maintenance [82] and the presence of six of the seven serotypes, i.e., A, O, C, SAT1, SAT2 and SAT3. Serotype C has not been reported since 2004 [22].

FMDV is a non-enveloped virus containing a single-stranded RNA genome of positive polarity in the genus *Aphthovirus* of the family *Picornaviridae* [1, 2, 27]. The large open reading frame (ORF) of $\sim$6,996 nt, which differs in length between the different serotypes [20], encodes a single polypeptide, which is co- and posttranslationally cleaved by viral proteases to give rise to the structural and non-structural proteins [3, 13, 55, 67]. Ten of the 13 cleavage events are catalysed by the virally encoded 3C protease [15, 58, 67, 78]. Translation takes place from a single open reading frame by a cap-independent mechanism at the internal ribosome entry site (IRES) [49], located in the 5' untranslated region (UTR). There are two different sites on the RNA at which the initiation of protein synthesis occurs, resulting in the generation of two forms of L proteinase (L^{pro}), Lb and the less abundant Lab, where Lb is the truncated version, which arises after the initiation of translation at the second AUG start codon [13]. Lab and Lb can cleave the L/P1 junction and ensure the proteolytic degradation of the cellular cap-binding protein complex (eIF4G), which results in the shutoff of host translation [22]. The P1 region is the viral capsid precursor and consists of the proteins 1A (VP4), 1B (VP2), 1C (VP3) and 1D (VP1). The antigenicity of the viral particles is dependent on the amino acid (aa) residues that are exposed on the surface of the capsid [56, 85]. Furthermore, it has been shown that the external capsid proteins play a role in binding to the FMDV cell-surface receptors, i.e., the RGD-dependant integrins [14, 25, 37–39, 59, 60] and heparan sulphate proteoglycans (HSPGs) [4, 36, 68].

The genetic heterogeneity of the virus, which is due to the lack of a proofreading mechanism during virus replication, has resulted in the occurrence of extensive variability as well as different lineages and antigenic variants within a serotype that have established themselves in different geographical regions [reviewed in refs. 8–10, 44, 70, 71, 75, 76, 88]. This has resulted in the need for multiple vaccine strains required for each serotype to cover the antigenic diversity when using vaccination as a control option [26]. In Africa and countries bordering Europe, the disease is mainly controlled using vaccination and restriction of animal movement. Thus, it is imperative to obtain as much information as possible regarding the FMDV prevalent on the African continent to further our knowledge on FMD epidemiology, define genetic relationships of

viruses causing outbreaks [45, 47] and to enable better control strategies by successful vaccine development.

Genetic information regarding the leader (L) and complete capsid-coding (P1) region of serotype A and O viruses prevalent on the African continent is lacking, although the SAT isolates have been broadly studied in the past [8–10, 86]. For this study, the L and P1 coding regions for eight FMDV A and nine FMDV O viruses isolated between 1975 and 2003 were successfully sequenced and analysed using phylogenetic analysis, examination of sequence variability, and identification of highly conserved genomic regions relating to previously identified FMDV functional and structural biological capabilities. Non-conservative substitutions were mapped to the available O (O1BFS) [53] and A (A$_{10}$/HOL/61) [29] capsid structures, and amino acid substitutions that may be involved in antigenic divergence were identified.

Materials and methods

Viruses included in this study

The sub-Saharan African isolates included in this study belong to different topotypes of FMDV serotypes A and O as defined by 1D sequencing and represent a broad geographical distribution of viruses within East and West Africa. The nine FMDV serotype O isolates and eight serotype A isolates were obtained from the Institute for Animal Health, Pirbright Laboratory, Pirbright, United Kingdom (Table 1). For the purpose of analysis, a select few complete L and P1 FMDV sequences currently available in GenBank were included (Table 1).

Cell culture propagation of viruses

The FMDV type O viruses were passaged for a previous study and were used directly in this study for processing, whereas the FMDV A isolates were first propagated on IB-RS-2 cells (Instituto Biologico renal suino cell line, a pig kidney cell line) to obtain a high viral titer. The IB-RS-2 cells were maintained in RPMI medium (Sigma) supplemented with 10 % foetal calf serum (FCS; Delta Bio-products) and 1x Antibiotic-Antimycotic (100×, Gibco®), Invitrogen). Virus was added to prepared cells containing RPMI supplemented with 1 % (v/v) FCS and 1× Antibiotic-Antimycotic mixture and incubated at 37 °C until complete CPE was attained (after 48 h). Clarified cell culture supernatant containing virus was stored at −80 °C until further use.

Chinese hamster ovary (CHO) cells strain K1 (ATCC CCL-61) were maintained in Ham's F-12 medium (Invitrogen) supplemented with 10 % FCS. Plaque assays were

Table 1 Description of viruses used for genetic analysis of the coding sequences of the L and capsid proteins

Virus strain[a,d]	Country of origin	References	GenBank accession no.	Passage history[b]	Topotypes[c]
O/ETH/3/96*	Ethiopia	This study	EU919240	RS2	East Africa (EA)
O/UGA/5/96*	Uganda	This study	EU919247	RS2	East Africa (EA)
O/KEN/10/95*	Kenya	This study	EU919242	RS3	East Africa (EA)
O/SUD/4/80*	Sudan	This study	EU919239	RS2	East Africa (EA)
O/UGA/17/98*,▲	Uganda	This study	EU919245	RS2	East Africa (EA)
O/UGA/1/75*,▲	Uganda	This study	EU919244	RS2	East Africa (EA)
O/UGA/6/76*,▲	Uganda	This study	EU919246	RS2	East Africa (EA)
O/TAN/3/96*	Tanzania	This study	EU919241	RS2	East Africa (EA)
O/UGA/7/03*	Uganda	This study	EU919243	PK1 RS1	East Africa (EA)
O/UKG/35/2001	United Kingdom	Carrillo et al. [29]	AJ539141	-	Middle East-South Asia (ME-SA)
OFRA/1/2001	France	Nobiron et al. [61]	AJ633821	-	Middle East-South Asia (ME-SA)
O/SAR/19/2000	South Africa	Carrillo et al. [29]	AJ539140	-	Middle East-South Asia (ME-SA)
O/TAW/2/99	Taiwan	Carrillo et al. [29]	AJ539137	-	Middle East-South Asia (ME-SA)
O/TIBET/CHA/99	China	Carrillo et al. [29]	AJ539138	-	Middle East-South Asia (ME-SA)
O/CHINA/1/99	China	Zhang et al. [89]	AF506822	-	Middle East-South Asia (ME-SA)
O/SKR/2000	South Korea	Carrillo et al. [29]	AJ539139	-	Middle East-South Asia (ME-SA)
O/O10PHIL76	Philippines	Carrillo et al. [29]	AY593812	-	South East Asia (SEA)
O/O10PHIL54	Philippines	Carrillo et al. [29]	AY593811	-	South East Asia (SEA)
O/O1MANISA87	Turkey	Carrillo et al. [29]	AY593823	-	Middle East-South Asia (ME-SA)
O/AKESU/58	China	Li et al. [52]	AF511039	-	Middle East-South Asia (ME-SA)
O/11INDONESIA52	Indonesia	Carrillo et al. [29]	AY593813	-	South-East Asia
O/O1BRUGGE79	Belgium	Carrillo et al. [29]	AY593817	-	Europe-South America (Euro-SA)
O/O1ARGENTINA65	Argentina	Carrillo et al. [29]	AY593814	-	Europe-South America (Euro-SA)
O/O1CAMPOS94	Argentina	Carrillo et al. [29]	AY593819	-	Europe-South America (Euro-SA)
O/O1CAMPOS96	Brazil	Carrillo et al. [29]	AY593818	-	Europe-South America (Euro-SA)
O/O1BFS46	United Kingdom	Carrillo et al. [29]	AY593816	-	Europe-South America (Euro-SA)
O/O1BFS18	United Kingdom	Carrillo et al. [29]	AY593815	-	Europe-South America (Euro-SA)
A/CIV/4/95*,▲	Cote d'Ivoire	This study	EU919236	BTY1 RS2	Africa
A/ERI/3/98*,▲	Eritrea	This study	EU919238	BTY1 RS2	Africa
A/ETH/2/79*	Ethiopia	This study	EU919233	BTY5 RS2	Africa
A/ETH/7/92*,▲	Ethiopia	This study	EU919235	BTY1 RS2	Africa
A/NIG/4/79*,▲	Nigeria	This study	EU919234	BTY2 BHK4 RS2	Africa
A/SEN/10/97*,▲	Senegal	This study	EU919237	BTY2 RS2	Africa
A/SOM/1/78*	Somalia	This study	EU919231	BTY2 RS2	Africa
A/TAN/4/80*	Tanzania	This study	EU919232	BTY2 RS2	Africa
A/A18ZULIA40	Venezuela	Carrillo et al. [29]	AY593758	-	Europe-South America (Euro-SA)
A/A1BRAZIL75	Brazil	Carrillo et al. [29]	AY593753	-	Europe-South America (Euro-SA)
A/A17AGUARULBOS83	Brazil	Carrillo et al. [29]	AY593757	-	Europe-South America (Euro-SA)
A/APHILIPPINES50	Philippines	Carrillo et al. [29]	AY593793	-	Asia
A/A29PERU37	Peru	Carrillo et al. [29]	AY593773	-	Europe-South America (Euro-SA)
A/BRAZIL67	Brazil	Carrillo et al. [29]	AY593788	-	Europe-South America (Euro-SA)
A/A24CRUZEIRO71	Brazil	Carrillo et al. [29]	AY593768	-	Europe-South America (Euro-SA)
A/A4SPAIN62	Spain	Carrillo et al. [29]	AY593778	-	Europe-South America (Euro-SA)
A/A14SPAIN39	Spain	Carrillo et al. [29]	AY593754	-	Europe-South America (Euro-SA)
A/A5WESTERWALD73	West Germany	Carrillo et al. [29]	AY593781	-	Europe-South America (Euro-SA)
A/A5ALLIER45	France	Carrillo et al. [29]	AY593780	-	Europe-South America (Euro-SA)
A/A12VALLE119/20	Great Britain	Carrillo et al. [29]	AY593752	-	Europe-South America (Euro-SA)
A/A10HOLLAND82	Holland	Carrillo et al. [29]	AY593751	-	Europe-South America (Euro-SA)

Table 1 continued

Virus strain[a,d]	Country of origin	References	GenBank accession no.	Passage history[b]	Topotypes[c]
A/A1BAYERN41	Germany	Carrillo et al. [29]	AY593759	-	Europe-South America (Euro-SA)
A/A3MECKLENBURG81	Germany	Carrillo et al. [29]	AY593776	-	Europe-South America (Euro-SA)
A/A2SPAIN7	Spain	Carrillo et al. [29]	AY593774	-	Europe-South America (Euro-SA)
A/A4WGERMANY72	West Germany	Carrillo et al. [29]	AY593779	-	Europe-South America (Euro-SA)
A/A4WGERMANY42	West Germany	Carrillo et al. [29]	AY593777	-	Europe-South America (Euro-SA)

[a] The viruses labeled "*" represent isolates from Africa

[b] RS is the number of passages on IB-RS-2 (Instituto Biologico renal suino) porcine kidney cells; BTY, on primary bovine thyroid cells; PK, on primary porcine kidney cells; and BHK, on baby hamster kidney cells. The number following the cell line indicates the number of times a virus was passaged in that particular cell line. (-) information is not available

[c] The topotypes are as described by Knowles and Samuel [18]

[d] The viruses labeled "▲" represent the African O isolates that have a codon insertion between nt 77 and 79 and the African A isolates that have a codon insertion or deletion between nt 54 and 61

performed by infecting monolayer cells with the virus for 1 h, followed by the addition of a 2-ml tragacanth overlay [66] and staining with 1 % (w/v) methylene blue [54].

RNA extraction, RT-PCR and sequencing

Total viral RNA was extracted using a modified guanidinium thiocyanate (GuSCN)-silica method [17]. The viral RNA template was reverse transcribed at 42 °C for 1 h using 10 U of AMV reverse transcriptase (Promega) and the antisense P1 primer (WDA; 5'-GAAGGGCCCAGGG TTGGACTC-3') [12] as described previously [7]. Amplification of the L-P1 region was undertaken using the antisense P1 (WDA) primer and the sense NCR1 primer (5'-TACCAAGCGACACTCGGGATCT-3') followed by PCR reactions using long-template Taq DNA polymerase (Roche) and thermal cycling conditions described by van Rensburg et al. [86].

PCR products of *ca.* 2,820 bp were excised from a 1 % agarose gel and purified using a Nucleospin® Extract Kit (Macherey-Nagel). Purified PCR products were sequenced using a genome-walking approach with genome-specific oligonucleotides and an ABI PRISM™ BigDye® Terminator Cycle Ready Reaction Kit v3.1 (Applied Biosystems). Sequences were analysed using an ABI Prism 3100 Genetic Analyser (Applied Biosystems).

Data analysis

Ambiguous nucleotides (nt) of the L-P1 sequences were resolved manually and assembled into a contig using the SEQUENCHER™ 4.7 DNA sequence analysis software (Gene Codes Corporation, Ann Arbor, MI, USA). A consensus sequence representing the most probable nt for each position of the sequence was obtained for each isolate. Consensus sequences were translated in BioEdit 5.0.9 DNA sequence analysis software [32], and the complete L-P1 nt and aa sequences were aligned using ClustalX 1.8.1 [83]. Hypervariable regions in the complete aa alignment were defined as a linear 10-aa region containing more than 50 % variable residues. The phylogenetic analysis included the newly determined sequences as well as sequences of non-African serotype A and O isolates obtained from GenBank (Table 1). Maximum-likelihood analysis of the aligned sequences was carried out in PAUP [79] under the Aikake Information Criterion. Phylogenetic trees were constructed using the neighbour-joining (NJ), minimum-evolution (ME) and maximum-parsimony (MP) methods included in the MEGA 4.0 program [50] for the L, 1A, 1B, 1C, 1D-coding regions separately as well as the full P1-coding region. Node reliability was estimated by 1000 bootstrap replications for NJ, ME and MP trees, whilst the nucleotide substitution model of Kimura 2-parameter was employed for the NJ and ME trees and close-neighbour-interchange (CNI) with search level 1 in effect for the MP and ME trees. MEGA 4.0 [50] was utilised to determine the nt and aa variation.

Plots representing the aa variation, hydrophobicity and secondary structures for each protein were drawn using Python (http://python.org) and the matplotlib package (http://matplotlib.sourceforge.net). The number of different amino acids occurring at a specific position was used as a measure of variation, and the hydrophobicity scale of Kyte and Doolittle [51] was used to measure relative aa hydrophobicity.

The crystallographic protomers of the capsid proteins of O_1BFS (1FOD) [53] and A_{10}/HOL/61 ((1ZBE) [29] were visualized and the surface-exposed residues identified with PyMol v1.1rc2pre (DeLano Scientific LLC).

Fig. 1 Neighbour-joining tree depicting genetic relationships for the P1 region of FMDV A and O type viruses. The Kimura 2-parameter model and bootstrap analysis (1000 replications) were applied
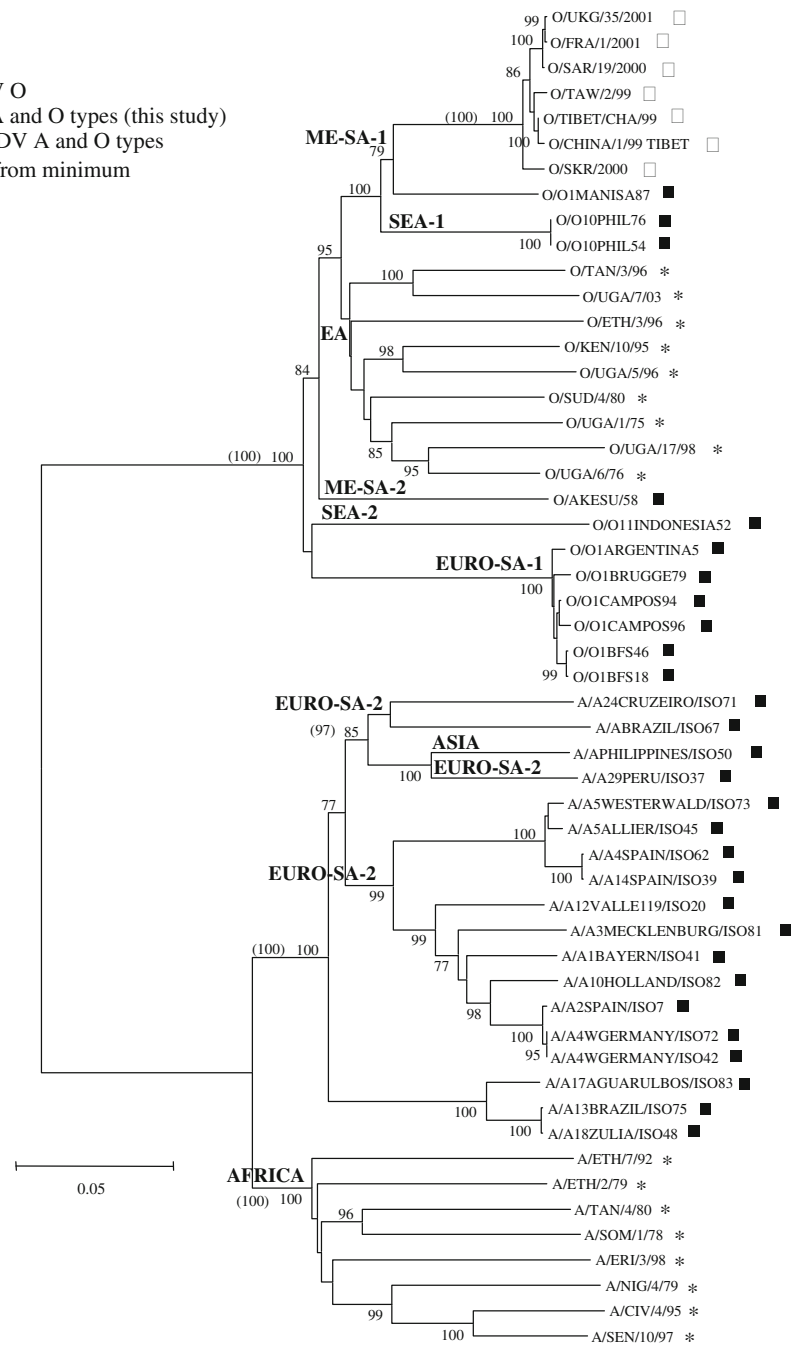
Results

Phylogenetic relationships and genetic heterogeneity of the serotype A and O isolates in Africa

Phylogenetic trees based on the P1 (Fig. 1), 1B, 1C and 1D regions that included all of the A and O isolates used in this study (Table 1) revealed groupings strictly according to serotype, irrespective of the phylogenetic methodology applied. In general, analysis of the entire structural protein-coding region improved bootstrap

values compared to 1D analysis alone. Phylogenetic clusters of the A and O isolates were described using the names published previously by Knowles and Samuel [44] (Table 1).

The P1 phylogeny for serotypes O viruses revealed that the African isolates clustered separately from the Pan-Asian O and non-African O isolates (Fig. 1), the latter belonging to the Middle East-South Asia (ME-SA) topotype based on 1D phylogeny [44]. The exception is O/SAR/19/2000, which was isolated in South Africa in 2000 during an outbreak caused by illegal feeding of swill to pigs [74].

This outbreak was controlled, and the virus no longer exists in southern Africa.

For the African isolates, significant bootstrap support was obtained for a group consisting of O/TAN/3/96 and O/UGA/7/03 (East African; EA-2) as well as for a group with O/KEN/10/95 and O/UGA/5/96 (East African; EA-1) and lastly for O/SUD/4/80, O/UGA/1/75, O/UGA/17/98 and O/UGA/6/76 (East African; EA-4) (Fig. 1). Furthermore, these P1 groupings were also observed when ME and MP phylogenetic models were utilised (not shown).

Clustering similar to that of the P1 region was observed for the separate gene regions, but with low bootstrap support except for 1B (O/UGA/7/03 and O/TAN/3/93, O/UGA/6/76 and O/UGA/17/98), 1C (O/UGA/17/98 and O/UGA/6/76) and 1D (OKEN/10/95 and O/UGA/5/96) groupings, which had high bootstrap support (Supplementary data, S1-S3). O/ETH/3/96 is the only representative of the EA-3 topotype in this study; thus, it did not cluster with the other isolates (Fig. 1). The nt sequence differences in the P1-coding region between members of each topotype were typically more than 15 %, similar to the cutoff defined for a topotype [47].

Globally, FMDV serotype A exists in three geographically distinct topotypes, Asia, Africa and Europe-South America (Euro-SA), based on the genetic relationships of 1D sequences [44]. Using the sequence information of the African A isolates together with P1 sequences of serotype A viruses available in the GenBank database, at least two separate clusters were observed for the type A viruses, *i.e.*, non-African and African A isolates, supported by 100 % bootstrap values for all phylogenetic methods used for the P1 (Fig. 1), 1B, 1C and 1D gene regions (Supplementary data, S1-S3). Two East African isolates, A/TAN/4/80 and A/SOM/1/78 formed a well-supported subgroup for the P1 (Fig. 1) and 1D NJ trees (Supplementary data, S3). In addition, there was a consistently strong grouping for three

West African isolates, A/NIG/4/79, A/CIV/4/95 and A/SEN/1/97, in the P1 (Fig. 1), 1B, 1C and 1D NJ analyses (Supplementary data, S1-S3).

The serotype A non-African and African viruses displayed similar genetic variability when compared to serotype O. The intratypic nt sequence variation in an alignment of the 2222-nt P1-coding region for type A was calculated to be 40.4 %, whilst the corresponding region (2202-2205 nt) of type O only revealed 38.5 % variable nucleotides.

Analysis of the 1A gene region resulted in phylogenetic groupings that differed from those of the P1, 1B, 1C and 1D analyses. When performing phylogenetic analysis on the combined O and A dataset, the FMDV A and O isolates did not group strictly according to serotype (Supplementary data, S4). For example, three non-African FMDV A strains, isolated from Brazil and Venezuela (A17/AGUA-RULBOS/ISO83, A18/ZULIA/ISO48 and A13/BRAZIL/ISO75), grouped with O viruses from the ME-SA, SEA and EA topotypes, but with low bootstrap support. As expected, the region encoding 1A was the most conserved, exhibiting 37.9 % variant nucleotides and was the only capsid-coding region with the highest average %Ts/Tv rate of 1.0 % (Table 2). In contrast, 1D had the highest variability of 58.7 % and lowest average %Ts/Tv rate of 0.28 % (Table 2).
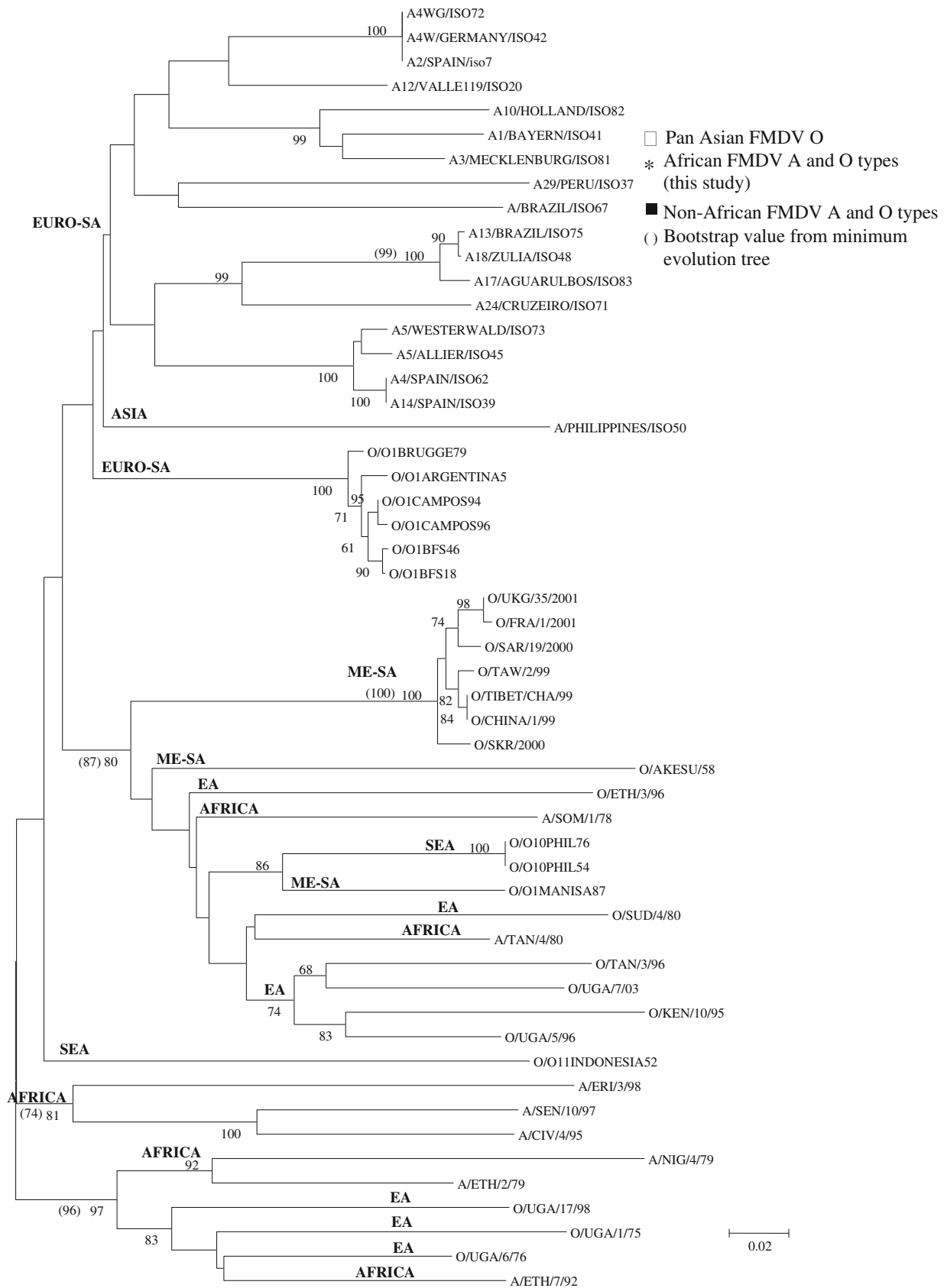
The phylogenetic trees based on the L^{pro}-coding region for the combined serotype O and A dataset had similar tree topologies for the A and O isolates, independent of the phylogenetic methods employed. The NJ tree of the L^{pro}-coding region (Fig. 2) showed that the viruses did not group strictly according to serotype, in contrast to those

Table 2 Variation within the nucleotide and amino acid sequences of the L and P1 polyprotein in a complete alignment of the non-African and African A and O sequences

Genome region	No. of nt positions aligned[a]	No. of variant nt			% of variant nt[b]			Av. %Ts/Tv rate	No. of aa positions aligned[a]	No. of variant aa			% of variant aa[b]		
		All	A	O	All	A	O	All		All	A	O	All	A	O
L	618	318	274	267	51.5	44.3	43.0	0.470	206	111	96	61	53.9	46.6	29.6
1A	256	97	94	76	37.9	36.7	29.7	1.045	85	13	13	4	15.3	15.3	4.7
1B	657	309	253	227	47.0	38.5	34.6	0.314	218	72	53	25	33.0	24.3	11.5
1C	667	318	254	275	47.7	38.1	41.2	0.345	222	89	63	36	40.1	28.4	16.2
1D	642	377	293	277	58.7	45.6	43.0	0.283	214	116	92	64	54.2	43.0	29.9
P1	2222	1162	894	855	52.3	40.4	38.5	0.104	739	296	224	128	40.1	30.3	17.2

[a] The number of nucleotides and amino acids were based on Clustal X alignments of the complete P1-coding region of the A and O serotypes

[b] The number of variant nucleotides or amino acids for each genomic region or capsid protein relative to the total number of positions was used to estimate the percentage (%) variability
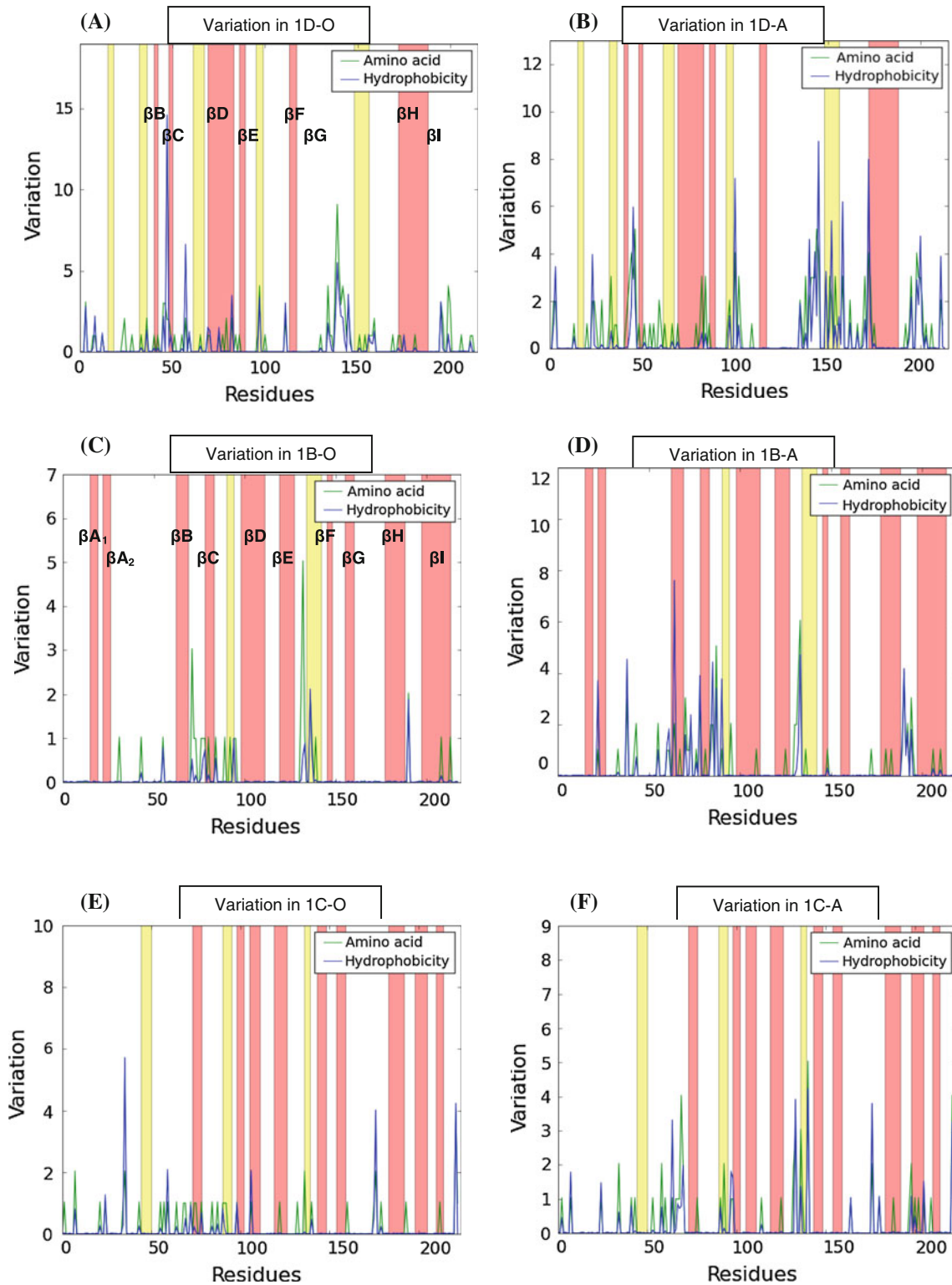
□ Pan Asian FMDV O

∗ African FMDV A and O types (this study)

■ Non-African FMDV A and O types

( ) Bootstrap value from minimum evolution tree

Fig. 3 Plots representing the aa variation and hydrophobicity for 1D (**a**, **b**), 1B (**c**, **d**) and 1C (**e**, **d**) for the African type A (**b**, **d**, **f**) and O (**a**, **c**, **e**) viruses from this study. The pink bars represent the beta strands of the FMDV structure, whilst the yellow represent the alpha helices. Areas of FMDV hydrophobicity and aa variation are represented by blue and green lines, respectively. Regions of variability or hypervariable sites were defined as sites on the P1 that had five or more variable aa residues within a window of 10 residues

based on the structural proteins. The non-African A and O isolates that form a part of the Euro-SA lineage [44] formed separate subgroupings in the L^{pro}-coding sequence NJ tree (Fig. 2). The Pan-Asian isolates formed a separate grouping with high bootstrap support (100 %). The five non-African O isolates that do not form a part of the Euro-SA lineage (isolates from the Philippines, Indonesia and China as well as the vaccine strain O1 Manisa originally isolated in Turkey; Table 1) grouped within the cluster of the African isolates, but as separate groupings. The majority of the African A and O viruses were found within two separate clusters, *i.e.*, one containing western Africa type A isolates and the other, eastern Africa type O and A (Fig. 2). Interestingly, one West African isolate, A/NIG/4/79, grouped with significantly high support with A/ETH2/79, an East African isolate. Also, a similarly strong grouping was observed for one East African (A/ERI/3/98) and two West African viruses (A/SEN/10/97 and A/CIV/4/95) (Fig. 2). There were also close relationships between African O and A viruses (O/ETH/3/96 and A/SOM/1/789; O/SUD/4/80 and A/TAN/4/80; O/UGA/6/76 and A/ETH/7/92), but these were not supported by high bootstrap values.

The nt variation for the L^{pro}-coding sequence was 51.5 % compared to 52.3 % for the P1 region (Table 2). The variation in the L^{pro}-coding sequence of the African isolates was further highlighted by a codon insertion between nt 77 and 79 for the Ugandan isolates O/UGA/17/98, O/UGA/1/75 and O/UGA/6/76 as well as by a codon insertion or deletion in A/CIV/4/95, A/ERI/3/98, A/ETH/2/79, A/ETH/7/92, A/NIG/4/79 and A/SEN/10/97 between nt 54 and 61 (Table 1).

Distribution of aa variation and hypervariability of the L and P1 polypeptides

The L^{pro} aa sequence displayed significant variation for a functional protein: 46.6 % for the serotype A alignment and 29.6 % for the serotype O isolates (Table 2). At least 30.3 % (224 of 739 aa) of the aa residues were variable in the alignment of the structural proteins (translated from the P1 region) of the 26 serotype A isolates, whilst the
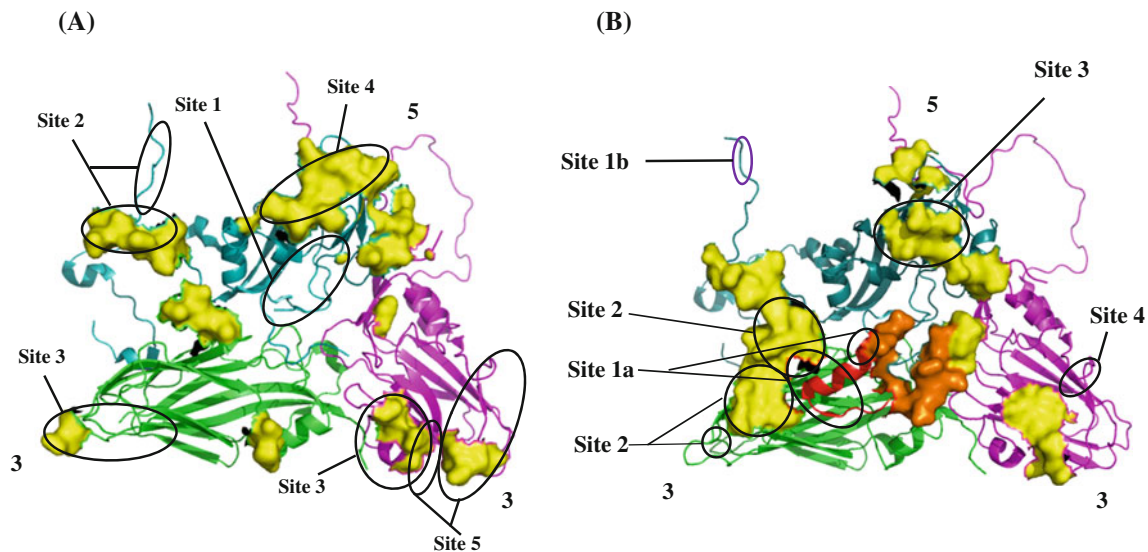
Table 3 Amino acid variation at amino acid position 56 of 1C and the relative infectivity of African O viruses in CHO-K1 cells

African O viruses	1C residue 56[a]	Titer in CHO-K1 cells (pfu/ml)[2]	Plaque size (in mm)[b]
O/UGA/5/96	His	-	-
O/SUD/4/80	His	-	-
O/ETH/3/96	His	-	-
O/TAN/3/96	His	-	-
O/KEN/10/95	Arg	3.63×10^4	<2
O/UGA/7/03	His	-	-
O/UGA/1/75	His	-	-
O/UGA/17/98	His	-	-
O/UGA/6/76	His	-	-

[a] His=histidine; Arg=arginine

[b] (-) No growth was observed in CHO-K1 cells

(A)   **(B)**

Fig. 4 The location of surface-exposed amino acid differences in the capsid proteins of African serotype A (**A**) and O (**B**) viruses on the crystallographic protomers of O1BFS (1FOD; Logan et al. [53]) and $A_{10}/61$ (1QQP; Fry et al. [28]). The protein subunits are colour-coded: 1D (teal), 1B (green), 1C (magenta) and the G-H loop of 1D (red). The G-H loop of $A_{10}/61$ (residues 138-158) is not resolved in the available structure and is not indicated here (**A**). The hypervariable amino acid positions in the G-H loop of O1BFS are shown in orange (**B**). 1A has been hidden from the structure. The fivefold and threefold axes of the capsid are indicated. The positions of surface-exposed residues with high variability are indicated in yellow. The surface view was made in Pymol

Table 4 Comparison of hypervariable regions identified in this study and previously identified neutralizing sites of type A and O FMDV

FMDV	Axis	Capsid	B-sheet structure	Hypervariable region indentified in this study[a,c]	Previously identified neutralisation sites[b,c]
Type O	5×	1D	B-C	1034-1060	Site 3: 1043-1048
					T-help: 1024-1042
	2×	1D	G-H	1135-1147	Site 1a: 1144-149;1154
		1B	E-F	2131-2141	Site 2: 2131-2134; 2188
	3×	1B	B-C	2072-2085	Site 2: 2070-2078
		1C	B-C		Site 4: 3056-3058
		1D	C$_T$	1196-1213	Site 1b: 1206-1208
Type A	5×	1D	H-I	1167-1176	Site 4: 1169; 1175-1178
	2×	1D	G-H	1135-1146 & 1150-1163	Site 1: 1142-1157;
					1138-1144
		1B	H-I	2188-2198	Site 3: 2196
	3×	1B			Site 3: 2079; 2082-2088
		1C	B-B	3058-3072	Site 5: 3058-3061
		1C	B-C	3058-3072	Site 5: 3069-3070
		1D	C$_T$	1193-1207	Site 2: 1198; 1200-1212
		1C	E-F	3132-3142	Site 3: 3136-3139

[a] The hypervariable regions were derived from the alignment of serotype A and O capsid sequences

[b] The antigenic sites are a summary of those described by Thomas et al. [80]; Baxt et al. [11]; Bolwell et al. [16]; Saiz et al. [72] (type A); Kitson et al. [43]; Crowther et al. [21]; Barnett et al. [5]; Barnett et al. [6] (type O)

[c] The amino acid residues have been numbered independently for each FMDV protein. For each residue, the first digit indicates the protein (1D, 1B or 1C), and the last three digits indicate the amino acid position

corresponding region of the 25 serotype O isolates displayed 17.2 % (128 of 739 aa) variable residues (Table 2). As expected, the internally located 1A (15.3 % for the FMDV A and 4.7 % for FMDV O) remained the most conserved, with 1D (43.0 % for FMDV A and 29.9 % for FMDV O) the most variable FMDV capsid protein. The variability for 1C was 28.4 % for FMDV A and 16.2 % for FMDV O, compared to 24.3 % for FMDV A and 11.5 % for FMDV O when looking at 1B (Table 2).

A systematic analysis of the capsid proteins revealed the variation not to be random but focused in local regions of hypervariability. The most variable capsid region, 1D, displayed the most regions of hypervariability. Figure 3A shows the hypervariable regions of type O at aa positions 34-60, 76-87, 135-147, 152-160, 196-213. At least seven discrete hypervariable regions (21-63, 80-87, 97-104, 135-146, 150-163, 167-176, 193-207) were identified in 1D of type A (Fig. 3B).

The conserved N-terminal motif of 1B, DKKTEETTL-LEDRIL-TTRNGHTTSTTQSSVG, described by Carrillo et al. [20], was present in the African A and O sequences (results not shown). Two hypervariable sites, residues 72-85 within the βB-βC loop and 131-141 in the βE-βF loop, were mapped within 1B of type O (Fig. 3C). 1B of type A displayed the same two hypervariable regions, residues 61-92 and 129-139, and a third hypervariable region, 188-198 (βH-βI loop; Fig. 3D).

Most of the 1C aa substitutions for type O were concentrated in one hypervariable region, i.e. 68-80. A second region with significant variability worth mentioning was residues 175-181, where three residue positions displayed high entropy and were located within a surface-exposed loop of 1C (Fig. 3E). The latter was situated in the β-β 'knob' of 1C and included the epitope site 4 for serotype O [43]. At least three hypervariable regions were identified in the type A alignment, i.e. residues 58-72, 132-142 and 197-211 (Fig. 3F).

The 1A protein of serotype O was most conserved, with only four variable residues and hypervariable regions that were not common for 1A (not shown).

The amino acids that have previously been identified as critical for FMDV were compared to the complete aa sequence alignment of the African and non-African A and O isolates from this study and are summarized in Supplementary data S5, showing that the aa residues important for FMDV function are conserved.

Investigation of possible heparan sulphate usage for O/KEN/10/95

Amino acid sequence alignments of all the African FMDV A and O viruses investigated in this study revealed an Arg at position 56 of 1C for only one African virus, O/KEN/10/95 (results not shown). The studies of Sa-Carvalho et al.

[68] and Fry et al. [28] confirmed the importance of the R56 residue of 1C for HS binding and cell culture adaptation. FMDV plaque assays in CHO-K1 cells (Table 3) confirmed that O/KEN/10/95 was the only virus that was able to infect and replicate in this cell line. Taking all of the serotype O capsid-sequence data together, 25 of the 27 O isolates had a His residue at position 56 of 1C, and they might therefore require integrins to replicate in cell culture.

Amino acid sequence variation in relation to structure

Vaccines based on A22/Iraq/64, A/ERI/98 and O1Manisa are recommended for the control of FMD in Africa [33]. We examined the variation within the deduced amino acid sequences of the capsid proteins of the African O and A isolates and compared the surface-exposed regions with those of the three recommended vaccine strains. Regions with high aa variability in an alignment of the capsid proteins were mapped onto the X-ray crystallographic structures of type A (A_{10}/HOL/61; 1QQP) [29] and O (O_1BFS; 1FOD) [53] viruses. Figure 4 shows that the regions of variability were mostly located on surface-exposed regions of the virion. Not all of the aa side chains within a variable region were exposed on the surface. Closer inspection of each aa position within a region of hypervariability indicated that positions with high variability had side chains exposed to the microenvironment of the virion.

For serotype A viruses, most of the hypervariable regions outside the 1D βG-βH loop were concentrated around the 5-fold and 3-fold axes of the virion and the C-terminus of 1D (Fig. 4) and correlated to a large extent with residues previously found to be involved in escape from neutralization by monoclonal antibodies (Table 4). Furthermore, many of the putative epitopes were probably discontinuous. For example, there was close proximity of 1B residue 2191 and 1C residues 3068-3071 and 3197-3198 around the 3-fold pore of the virion (Fig. 4). Similarly the regions of variability for type O correlated strongly with epitopes previously identified with distribution around the 5-fold and 3-fold axes of the virion (Fig. 4; Table 4).

Discussion

The data from the analysis of the complete capsid-coding region, P1, as well as the individual capsid-coding regions indicated that very similar tree topologies existed for the different genomic regions when comparing the African A and O viruses with those from other regions of the world. In general, analysis of the entire structural protein-coding region improved bootstrap values relative to 1D analysis alone. The longer the capsid-coding region included in the analysis, the more accurate the relationship conclusion. This supports the view that sequencing of the entire capsid-coding region, rather than 1D alone, is desirable in molecular evolution studies.

Phylogeny based on the NJ trees of the P1, 1B, 1C and 1D sequences resulted in the grouping of viruses according to serotype. In addition, the A and O virus clusters could be further divided into separate groupings of the African and non-African A and O isolates, which were observed for the P1, 1B, 1C and 1D NJ, ME and MP trees.

The separate groupings of the African and non-African A viruses support previous findings for type A viruses. These could be grouped into three major restricted genotypes, i.e., Euro-South America, Asia and Africa, based on 1D phylogeny (this study only included FMDV A viruses from Euro-South America and Africa) [44, 46, 57].

Similarly, based on 1D phylogeny, type O viruses were divided into three groups: those originating from Asia, Europe-South America and the Far East [44, 69, 73, 74]. The P1 phylogeny therefore supports the three major virus groups within serotype O. The eastern and western African O viruses were grouped together with the SEA and ME-SA lineages, together with the Pan-Asia strain [44, 45, 73], albeit as lineages restricted to geographic regions (East Africa-1, 2, 3, 4 and West Africa). Furthermore, the phylogeny is indicative of the transboundary spread of FMDV in Africa among the East African countries, Uganda, Kenya, Somalia and Tanzania, that are in close proximity to each other, which is also true for the West African countries, i.e. Nigeria, Ivory Coast and Senegal. The groupings also indicated that the East African and West African viruses fall into separate large groups. Another well-supported grouping was observed for the P1, 1B and 1C trees (all methodologies) for O/UGA/1/75, O/UGA/6/76 and O/UGA/17/98, with a maximum of 15.1 % nt and 6.5 % aa substitutions in any pairwise alignment. This grouping most likely signifies that the 1998 outbreak strains re-emerged from older strains that have been maintained in the endemic area since the early 1970s, i.e. from 1975 to 1998 (23 years).

There was a difference in the groupings for the 1A trees when compared to the P1 and other capsid-coding gene regions where three non-African A isolates clustered with the non-African O viruses (for all phylogenetic methodologies). The phylogenetic tree representing the region encoding the L protein differed from that of the structural proteins where sub-grouping according to serotype was much less apparent, which was consistent with previous findings for this region [81, 86].

Interestingly, certain A and O African viruses clustered together and also did not separate into geographical regions such as East and West Africa as observed for the structural

coding regions. For example, bootstrap support of 73 % for the L-region NJ tree was observed for the grouping of O/UGA/17/98, O/UGA/1/75, O/UGA/6/76 & A/ETH/7/92, which was not observed with the 1A, 1B, 1C, 1D and P1 phylogenetic analysis. This suggests that the African viruses share similarities or are closely related when comparing the L sequences, irrespective of serotype. Taking into account the extensive, uncontrolled movement of animals across the borders and the ease of virus spread and infection of multiple serotypes in one animal, the role of recombination events in the genetic diversification of FMDV cannot be excluded. Although we did not perform a study on the occurrence of recombination, the similarities present between FMDV A and O L sequences could be due to the occurrence of intertypic recombination events [30, 40–42, 86].

Due to the high mutation rates of FMDV, it is likely that even brief epidemics might result in the generation of substantial antigenic variability [35]. However, the adaptive significance of this variation remains unclear [34]. The antigenicity of FMDV is attributed to the aa residues that are exposed on the surface of the capsid [56]. An important immunogenic determinant, the 1D G-H loop [3], exhibited a high degree of variation for the A and O isolates included in this study. Consequently, aa changes in this region are most likely involved in the appearance of novel antigenic types. Analyses of antigenic sites of picornaviruses have been carried out using neutralising monoclonal antibodies (Mabs) to select and screen Mab-resistant mutants. Sequence analysis of these mutants resulted in the identification of five antigenic sites of serotype O virus, i.e., O_1 Kaufbeuren [21, 44], and six sites for the FMDV A viruses [44]. Alignments of the aa sequences of the African A and O viruses indicated that the regions of variability identified corresponded to the known antigenic sites, which points to the fact that the location of antigenic sites are structurally conserved for the African A and O viruses. In addition to these sites, other regions of variability were identified for both the FMDV O and A African isolates from the aa variability plots. These regions could potentially be antigenic determinants, which may be difficult to map by the classical methodology of MAb-resistant escape mutants. We have recently shown that an approach combining sequence variation with structural data and antigenic variation results in the reasonably accurate identification of novel antigenic determinants on the virion surface [65].

The aligned L^{pro} aa sequences displayed marked variation in both the Lab and Lb regions (not shown); however, despite this variation, the aa residues identified as being critical for the L^{pro} function were highly conserved, i.e., the residues C53, H153 and D168 required for L^{pro} catalytic activity, the E81 residue required for L^{pro} autocatalysis, and two His residues (H114 and H143) important for

cleavage of the translation initiation factor, eIF4G, [31, 48, 62, 63]. A comparison of the L/P1 cleavage sequence at the C-terminus of the L protein and N-terminus of the 1A protein of the FMDV non-African A types revealed a sequence of R(Q/W)KLK*GAGQ (* indicates cleaved peptide bond), whereas the African A types included in this study had the sequence K(R)R(K)LK*GAGQ (results not shown). Both the FMDV non-African and African O types revealed a sequence of (K/R)(K/R)L(K/R)*GAGQ (* indicates cleaved peptide bond) (results not shown). These observations compared well with the L^{pro}/1A junction previously described for serotypes A, O and C [76], where the residues K(R)R(K)LK(R) at the L^{pro} C terminus and the GAGQ at the 1A N terminus were observed. These results suggest that for all the A and O types included in this study, the conserved sequence XXLK(R)*GAGQ (where X is either K or R) is sufficient for L/P1 cleavage by L^{pro}.

The degree of hydrophobicity/hydrophilicity of the loops connecting the β chains varied between the African A and O surface proteins. Hydrophilic β-β loops tend to be exposed on the protein surface, sometimes protruding from the protein core, and are candidates for antibody binding [87]. Overall, the aa sequence variation observed for the FMDV A and O viruses included in this study showed that the A viruses exhibited more variation, possibly indicating that the A viruses evolved rapidly, which supports studies by Bachrach [2] and Brooksby [19]. Additionally, Tully and Fares [84] showed that among all of the FMDV serotypes, serotype A is the most divergent and that adaptive evolution has occurred in the 3C protease (involved in RNA replication and processing of the polyprotein) and 2B (involved in membrane rearrangements), which supports the hypothesis of selection for faster replication in serotype A.

Neff et al. [59] showed that a variant of the type O1 virus containing an Arg at residue 56 of 1C required only HS binding to replicate in CHO-K1 cells but that another variant with a His residue at this position required integrins to replicate in cell culture. Interestingly, in this study, it was shown that O/KEN/10/95 was the only African virus to have this Arg residue at residue 56 of 1C, and it was indeed able to replicate in CHO-K1 cells. However this virus has been passaged three times on IB-RS-2 cells, and it is possible that the mutation arose during cell culture passage. Additionally, various aa residues that were previously identified as important for playing a role in various functions for FMDV were found to be conserved for the A and O isolates (see "Results").

It is clear from the outbreaks of FMD during the last two decades that there is a continuing threat to the livestock industry. The results presented here show distinct geographical grouping of serotype A and O viruses in Africa, although common ancestry with the Euro-South American-

Asian topotypes is clear. The natural diversification of FMDV occurs during replication in infected animals and results in the rapid generation of mutants and the ability to persist and to spread amongst livestock. Thus, continuous surveillance and an active molecular epidemiology program increases our knowledge with regard to FMDV phylogenetic relationships, virus antigenicity, and the ability of existing vaccine strains to provide protection against emerging and re-emerging viruses.
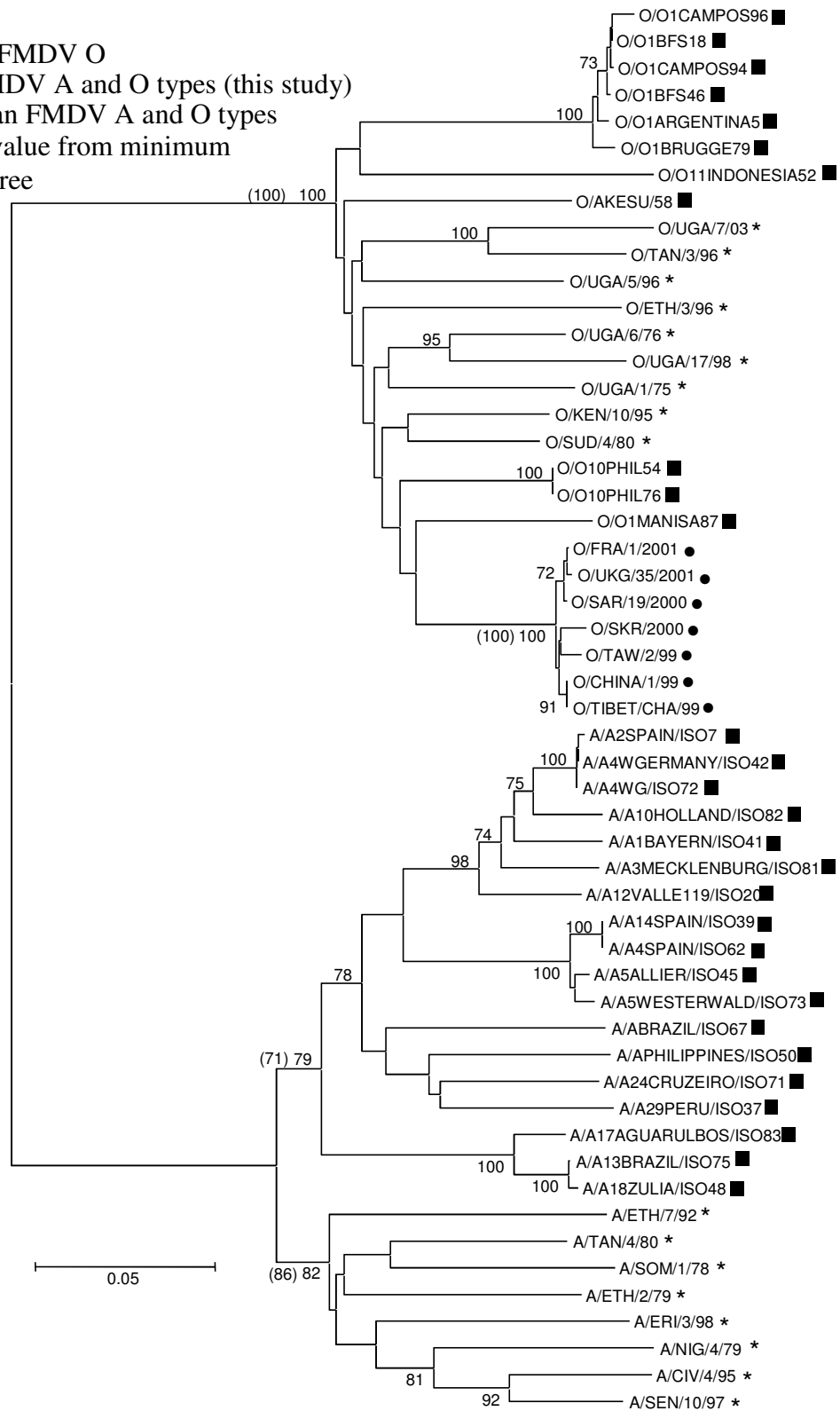
References

1. Alexandersen S, Zhang Z, Donaldson AI, Garland AJM (2003) The pathogenesis and diagnosis of foot-and-mouth disease. J Comp Path 129:1–36
2. Bachrach HL (1968) Foot-and-mouth disease. Annu Rev Microbiol 22:201–244
3. Bachrach HL, Morgan DO, McKercher PD, Moore DM, Robertson BH (1982) Foot-and-mouth disease virus: immunogenicity and structure of fragments derived from capsid protein VP and of virus containing cleaved VP. Vet Microbiol 7:85–96
4. Baranowski E, Sevilla N, Verdaguer N, Ruiz-Jarabo CM, Beck E, Domingo E (1998) Multiple virulence determinants of foot-and-mouth disease virus in cell culture. J Virol 72:6362–6372
5. Barnett PV, Ouldridge EJ, Rowlands DJ, Brown F, Parry NR (1989) Neutralization epitopes of type O foot-and-mouth disease virus. I. Identification and characterization of three functionally independent, conformational sites. J Gen Virol 70:1483–1491
6. Barnett PV, Samuel AR, Pullen L, Ansell D, Butcher RN, Parkhouse RME (1998) Monoclonal antibodies, against O″ serotype foot-and-mouth disease virus, from a natural bovine host, recognize similar antigenic features to those defined by the mouse. J Gen Virol 79:1687–1697
7. Bastos AD (1998) Detection and characterisation of foot-and-mouth disease in sub-Saharan Africa. Onderstepoort J Vet Res 65:37–47
8. Bastos AD, Haydon DT, Forsberg R, Knowles NJ, Anderson EC, Bengis RG, Nel LH, Thomson GR (2001) Genetic heterogeneity of SAT-1 type foot-and-mouth disease viruses in southern Africa. Arch Virol 146:1537–1551
9. Bastos AD, Haydon DT, Sangaré O, Boshoff CI, Edrich JL, Thomson GR (2003) The implications of virus diversity within the SAT 2 serotype for control of foot-and-mouth disease in sub-Saharan Africa. J Gen Virol 84:1595–1606
10. Bastos AD, Anderson EC, Bengis RG, Keet DF, Winterbach HK, Thomson GR (2003) Molecular epidemiology of SAT3-type foot-and-mouth disease. Virus Genes 27:283–290
11. Baxt B, VakhariaV, Moore DM, Franke AJ, Morgan DO (1989) Analysis of neutralising antigenic sites on the surface of type A12 foot-and-mouth disease virus. J Virol 63:2143–2151
12. Beck E, Strohmaier K (1987) Subtyping of European foot-and-mouth disease virus strains by nucleotide sequence determination. J Virol 61:1621–1629
13. Belsham GJ (1993) Distinctive features of foot-and-mouth disease virus, a member of the picornavirus family; aspects of virus protein synthesis, protein processing and structure. Biophys Molec Biol 60:241–260
14. Berinstein A, Rovainen M, Hovi T, Mason PW, Baxt B (1995) Antibodies to the vitronectin receptor (integrin αvβ3) inhibit binding and the infection of foot-and-mouth disease virus to cultured cells. J Virol 69:2664–2666
15. Birtley JR, Knox SR, Jaulent AM, Brick P, Leatherbarrow RJ, Curry S (2005) Crystal structure of foot-and-mouth disease virus 3C protease: New insights into catalytic mechanism and cleavage specificity. J Biol Chem 280:11520–11527
16. Bolwell C, Clarke BE, Parry NR, Oulridge EJ, Brown F, Rowlands DJ (1989) Epitope mapping of foot-and-mouth disease virus with neutralizing monoclonal antibodies. J Gen Virol 70:59–68
17. Boom R, Sol CJ, Salimans MMM, Jansen CL, Wertheim-van Dillen PME, Van Der Noordaa J (1990) Rapid and simple method for purification of nucleic acids. J Clin Microbiol 28:495–503
18. Bronsvoort BM, Sørensen KJ, Anderson J, Corteyn A, Tanya VN, Kitching RP, Morgan KL (2004) Comparison of two 3ABC enzyme-linked immunosorbent assays for diagnosis of multiple-serotype foot-and-mouth disease in a cattle population in an area of endemicity. J Clin Microbiol 42:2108–2114
19. Brooksby JB (1982) Portraits of viruses: foot-and-mouth disease virus. Intervirology 18:1–23
20. Carrillo C, Tulman ER, Delhon G, Lu Z, Carreno A, Vagnozzi A, Kutish GF, Rock DL (2005) Comparative genomics of Foot and mouth disease virus. J Virol 79:6487–6504
21. Crowther JR, Farias S, Carpenter WC, Samuel AR (1993) Identification of a fifth neutralisable site on type O foot-and-mouth disease virus following characterization of single and quintuple monoclonal antibody escape mutants. J Gen Virol 74:1547–1553
22. Di Nardo A, Knowles NJ, Paton DJ (2011) Combining livestock trade patterns with phylogenetics to help understand the spread of foot-and-mouth disease in sub-Saharan Africa, the Middle East and Southeast Asia. Rev Sci Tech Off Int Epiz 30(1):63–85
23. Domingo E, Martinez-Salas E, Sobrino F, de la Torre CJ, Portela A, Ortin J, Lopez-Galindez C, Perez-Brena P, Villanueva N, Najera R, VandePol S, Steinhauer D, DePolo N, Holland J (1985) The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevance—a review. Gene 40:1–8
24. Domingo E, Baranowski E, Escarmis C, Sobrino F (2002) Foot-and-mouth disease virus. Comp Imm Micro Infect Dis 25:297–308
25. Duque H, Baxt B (2003) Foot-and-mouth disease virus receptors: Comparison of bovine αv integrin utilization by type A & O viruses. J Virol 77:2500–2511
26. European Commission (2004) Report of the session of the research group of the standing technical Committee of the European commission for the control of foot-and-mouth disease (EUFMD) held at Chania, Crete (Greece)
27. Forss S, Strebel K, Beck E, Schaller H (1984) Nucleotide sequence and genome organization of foot-and-mouth disease virus. Nucleic Acids Res 12:6587–6601
28. Fry E, Lea SM, Jackson T, Newman JWI, Ellard FM, Blakemore WE, Ghazaleh RA, Samuel A, King AMQ, Stuart DI (1999) The structure and function of a foot and mouth disease virus-oligosaccharide receptor complex. EMBO J 18:543–554

29. Fry EE, Stuart DI, Rowlands DJ (2005) The structure of foot-and-mouth disease virus. Foot-and-mouth Disease Virus. Springer, Berlin, pp 71–102

30. Giraudo AT, Sagedahl A, Bergmann IE, La Torre JL, Scodeller EA (1987) Isolation and characterization of recombinants between attenuated and virulent aphthovirus strains. J Virol 61:419–425

31. Gorbalenya AE, Koonin EV, Lai MM (1991) Putative papain-related thiol proteases of positive strand RNA viruses. Identification of rubi- and aphthovirus proteases and delineation of a novel conserved domain associated with proteases of rubi-, alpha- and coronaviruses. FEBS Lett 288:201–205

32. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Sym Ser 41:95–98

33. Hammond J (2011) OIE/FAO FMD reference laboratory network annual report

34. Haydon D, Lea S, Fry L, Knowles N, Samuel AR, Stuart D, Woolhouse EJ (1998) Characterising sequence variation in the VP1 capsid proteins of foot-and-mouth disease virus (serotype O) with the respect to virion structure. J Mol Evol 46:465–475

35. Haydon DT, Bastos AD, Knowles NJ, Samuel AR (2001) Evidence for positive selection in foot-and-mouth disease virus capsid genes from field isolates. Genetics 157:7–15

36. Jackson T, Ellard FM, Ghazaleh RB, Brookes SM, Blakemore WE, Corteyn AH, Stuart DI, Newman JWI, King AM (1996) Efficient infection of cells in culture by type O foot-and-mouth disease virus requires binding to cell surface heparan sulfate. J Virol 70:5282–5287

37. Jackson T, Sheppard D, Denyer M, Blakemore W, King AMQ (2000) The epithelial integrin αvβ6 is a receptor for foot-and-mouth disease virus. J Virol 74:4949–4956

38. Jackson T, Mould AP, Sheppard D, King AMQ (2002) Integrin αvβ1 is a receptor for foot-and-mouth disease virus. J Virol 76:935–941

39. Jackson T, Clark S, Berryman S, Burman A, Cambier S, Mu D, Nishimura S, King AMQ (2004) Integrin αvβ8 functions as a receptor of foot-and-mouth disease virus: Role of the β-chain cytodomain in integrin-mediated infection. J Virol 78:4533–4540

40. Jackson AL, O'Neill H, Maree FF, Blignaut B, Carrillo C, Rodriguez L, Haydon DT (2007) Mosaic structure of foot-and-mouth disease virus genomes. J Gen Virol 88:487–492

41. King AMQ, McCahon D, Slade WR, Newman JWI (1982) Recombination in RNA. Cell 29:921–928

42. King AMQ, McCahon D, Saunders K, Newman JWI, Slade WR (1985) Multiple sites of recombination within the RNA genome of foot-and-mouth disease virus. Virus Res 3:373–384

43. Kitson JDA, McCahon D, Belsham GJ (1990) Sequence analysis of monoclonal antibody resistant mutants of type O foot and mouth disease virus: evidence for the involvement of the three surface exposed capsid proteins in four antigenic sites. Virology 179:26–34

44. Knowles NJ, Samuel AR (2003) Molecular epidemiology of foot-and-mouth disease virus. Virus Res 91:65–80

45. Knowles NJ, Samuel AR, Davies PR, Midgley RJ, Valarcher JF (2005) Pandemic strain of foot-and-mouth disease virus serotype O. Emerg Infect Dis 11:1887–1893

46. Knowles NJ, Wadsworth J, Reid SM, Swabey KG, El-Kholy AA, El-Rahman AOA, Soliman HM, Ebert K, Ferris NP, Hutchings GH, Statham RJ, King DP, Paton DJ (2007) Foot-and-mouth disease virus serotype A in Egypt. Emerg Infect Dis 13:1593–1596

47. Knowles NJ, Nazem Shirazi MH, Wadsworth J, Swabey KG, Stirling JM, Statham RJ, Li Y, Hutchings GH, Ferris NP, Parlak U, Ozyörük F, Sumption KJ, King DP, Paton DJ (2009) Recent spread of a new strain (A-Iran-05) of foot-and-mouth disease virus type A in the Middle East. Transbound Emerg Dis 56:157–169

48. Kronovetr J, Skern T (2002) Foot-and-mouth disease virus leader proteinase: a papain-like enzyme requiring and acidic environment in the active site. FEBS Lett 528:58–62

49. Kühn R, Luz N, Beck E (1990) Functional analysis of the internal translation initiation site of foot-and-mouth disease virus. J Virol 64:4625–4631

50. Kumar S, Tamura K, Nei M (1993) MEGA: Molecular evolutionary genetics analysis version 3.0. Pennsylvania State University, USA

51. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132

52. Li D, Liu ZX, Bao HF, Lu ZJ, Guo JH, Cao YM, Li PH, Bai XW, Chen YL, Xie BX, Cai XP, Xie QG (2007) The complete genome sequence of foot-and-mouth disease virus O/Akesu/58 strain and its some molecular characteristics. Arch Virol 152:2079–2085

53. Logan D, Abu-Ghazaleh R, Blakemore W, Curry S, Jackson T, King A, Lea S, Lewis R, Newman J, Parry N, Rowlands D, Stuart D, Fry E (1993) Structure of a major immunogenic site on foot-and-mouth disease virus. Nature 362:566–568

54. Maree FF, Blignaut B, de Beer TA, Visser N, Rieder EA (2010) Mapping of amino acid residues responsible for adhesion of cell culture-adapted foot-and-mouth disease SAT type viruses. Virus Res 153:82–91

55. Mason PW, Pacheco JM, Zhao QZ, Knowles NJ (2003) Comparisons of the complete genomes of Asian, African and European isolates of a recent foot-and-mouth disease virus type O pandemic strain (PanAsia). J Gen Virol 84:1583–1593

56. Mateu MG, Camarero JA, Giralt E, Andreu D, Domingo E (1995) Direct evaluation of the immunodominance of a major antigenic site of foot-and-mouth disease virus in a natural host. Virology 206:298–306

57. Nayak B, Pattnaik B, Tosh C, Sanyal A, Hemadri D, Patil SS, Venkataramanan R (2001) Genetic and antigenic analysis of type A foot-and-mouth disease viruses isolated in India during 1987-1996. Acta Virol 45:13–21

58. Nayak A, Goodfellow IG, Woolaway KE, Birtley J, Curry S, Belsham GJ (2006) Role of RNA structure and RNA binding activity of foot-and-mouth disease virus 3C Protein in VPg uridylylation and virus replication. J Virol 80:9865–9875

59. Neff S, Sa-Carvalho D, Rieder E, Mason PW, Blystone SD, Brown EJ, Baxt B (1998) Foot-and-mouth disease virus virulent for cattle utilizes the integrin αvβ3 as its receptor. J Virol 72:3587–3594

60. Neff S, Mason PW, Baxt B (2000) High efficiency utilization of the bovine integrin αvβ3 as a receptor for foot and mouth disease virus is dependant on the bovine β3 subunit. J Virol 74:7298–7306

61. Nobiron I, Remond M, Kaiser C, Lebreton F, Zientara S, Delmas B (2005) The nucleotide sequence of foot-and-mouth disease virus O/FRA/1/2001 and comparison with its British parental strain O/UKG/35/2001. Virus Res 108:225–229

62. Piccone ME, Rieder E, Mason PW, Grubman MJ (1995) The foot-and-mouth disease virus leader proteinase gene is not required for viral replication. J Virol 69:5376–5382

63. Piccone ME, Zellner M, Kumosinski TF, Mason PW, Grubman MJ (1995) Identification of the active-site residues of the L-proteinase of foot-and-mouth disease virus. J Virol 69:4950–4956

64. Pringle CR (1965) Evidence of genetic recombination in foot-and-mouth disease virus. Virology 25:48–54

65. Reeve R, Blignaut B, Esterhuysen JJ, Opperman P, Matthews L, Fry EE, de Beer TA, Theron J, Rieder E, Vosloo W, O'Neill HG, Haydon DT, Maree FF (2010) Sequence-based prediction for vaccine strain selection and identification of antigenic variability in foot-and-mouth disease virus. PLoS Comput Biol 6:e1001027

66. Rieder E, Bunch T, Brown F, Mason PW (1993) Genetically engineered foot-and-mouth disease viruses with poly(C) tracts of two nucleotides are virulent in mice. J Virol 67:5139–5145

67. Rueckert RR (1996) Picornaviridae: the viruses and their replication. In: Fields BN, Knipe BN, Howley PM (eds) Virology. Lippincott-Raven Publishers, Philadelphia, pp 609–654

68. Sa-Carvalho D, Reider E, Baxt B, Rodarte R, Tanuri A, Mason PW (1997) Tissue culture adaptation of foot-and-mouth disease virus selects viruses that bind to heparin and are attenuated in cattle. J Virol 71:5115–5123

69. Sahle M, Venter EH, Dwarka RM, Vosloo W (2004) Molecular epidemiology of serotype O foot-and-mouth disease virus isolated from cattle in Ethiopia between 1979–2001. Onderstepoort J Vet Res 71:129–138

70. Sahle M, Dwarka RM, Venter EH, Vosloo W (2007) Comparison of SAT-1 foot-and-mouth disease virus isolates obtained from East Africa between 1971 and 2000 with viruses from the rest of sub-Saharan Africa. Arch Virol 152:797–804

71. Sahle M, Dwarka RM, Venter EH, Vosloo W (2007) Study of the genetic heterogeneity of SAT-2 foot-and-mouth disease virus in sub-Saharan Africa with specific focus on East Africa. Onderstepoort J Vet Res 74:289–299

72. Saiz JC, Gonzalez MJ, Borca MV, Sobrino F, Moore DM (1991) Identification of neutralising antigenic sites on VP1 and VP2 of type A5 foot-and-mouth disease virus, defined by neutralisationresistant variants. J Virol 65:2518–2524

73. Samuel AR, Knowles NJ (2001) Foot-and-mouth disease type O viruses exhibit genetically and geographically distinct evolutionary lineages (topotypes). J Gen Virol 82:609–621

74. Sangare O, Bastos AD, Marquardt O, Venter EH, Vosloo W, Thomson GR (2001) Molecular epidemiology of serotype O foot-and-mouth disease virus with emphasis on West and South Africa. Virus Genes 22:345–351

75. Sangare O, Bastos ADS, Venter EH, Vosloo W (2003) Retrospective genetic analysis of SAT-1 type foot-and-mouth disease outbreaks in West Africa (1975–1981). Vet Microbiol 93:279–289

76. Seipelt J, Guarne A, Bergmann E, James M, Sommergruber W, Fita I, Skern T (1999) The structures of picornaviral proteinases. Virus Res 62:159–168

77. Sutmoller P, Barteling SS, Olascoaga RC, Sumption KJ (2003) Control and eradication of foot-and-mouth disease. Virus Res 91:101–144

78. Sweeney TR, Roque-Rosell N, Birtley JR, Leatherbarrow RJ, Curry S (2007) Structural and mutagenic analysis of foot-and-mouth disease virus 3C protease reveals the role of the β-Ribbon in proteolysis. J Virol 81:115–124

79. Swofford DL (1998) PAUP: phylogenetic analysis using parsimony (and other methods). Sinauer Associates, Sunderland

80. Thomas AA, Woortmeijer RJ, Puijk W, Barteling SJ (1988) Antigenic sites on foot-and-mouth disease virus type A10. J Virol 62:2782–2789

81. Tosh C, Mittal M, Sanyal A, Hemadri D, Bandyopadhyay SK (2004) Molecular phylogeny of leader proteinase gene of type A of foot-and-mouth disease virus from India. Arch Virol 149:523–536

82. Thomson GR (1994) Foot-and-mouth disease. In: Coetzer JAW, Thomson GR, Tustin RC, Kriek NPJ (eds) Infectious diseases of livestock with special reference to Southern Africa. Oxford University Press, Cape Town, pp 825–852

83. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nuc Acids Res 24:4876–4882

84. Tully DC, Fares MA (2009) Shifts in the Selection-Drift Balance Drive the Evolution and Epidemiology of Foot-and-Mouth Disease Virus. J Virol 83(2):781–790

85. Usherwood EJ, Johnston ICD, Lovelidge LJ, Tonks P, Nash AA (1995) Lymphocyte recognition elements on the VP1 protein of Theiler's virus. Immunology 85:190–197

86. Van Rensburg H, Haydon D, Joubert F, Bastos A, Heath L, Nel L (2002) Genetic heterogeneity in the foot-and-mouth disease virus Leader and 3C proteinase. Gene 289:19–29

87. Verdaguer N, Sevilla N, Valero ML, Stuart D, Brocchi E, Andreu D, Giralt E, Domingo E, Mateu MG, Fita I (1998) A similar pattern of interaction for different antibodies with a major antigenic site of foot-and-mouth disease virus: implications for intratypic antigenic variation. J Virol 72:739–748

88. Vosloo W, Bastos ADS, Sangare O, Hargreaves SK, Thomson GR (2002) Review of the status and control of foot and mouth disease in sub-Saharan Africa. Rev sci tech Off Int Epiz 21:437–449

89. Zhang X, Liu Z, Zhao Q, Chang H, Xie Q (2004) Sequencing and analysis for the full-length genome RNA of foot-and-mouth disease virus China/99. Sci China C Life Sci 47:74–81

S1: Neighbour-joining tree depicting gene relationships for the 1B region of A and O type viruses. Kimura-2 parameter and bootstrap (1000 replications) were applied.

● Pan Asian FMDV O
∗ African FMDV A and O types (this study)
■ Non-African FMDV A and O types
( ) Bootstrap value from minimum Evolution tree

O/O1CAMPOS96 ■
O/O1BFS18 ■
O/O1CAMPOS94 ■
O/O1BFS46 ■
O/O1ARGENTINA5 ■
O/O1BRUGGE79 ■
O/O11INDONESIA52 ■
O/AKESU/58 ■
O/UGA/7/03 ∗
O/TAN/3/96 ∗
O/UGA/5/96 ∗
O/ETH/3/96 ∗
O/UGA/6/76 ∗
O/UGA/17/98 ∗
O/UGA/1/75 ∗
O/KEN/10/95 ∗
O/SUD/4/80 ∗
O/O10PHIL54 ■
O/O10PHIL76 ■
O/O1MANISA87 ■
O/FRA/1/2001 ●
O/UKG/35/2001 ●
O/SAR/19/2000 ●
O/SKR/2000 ●
O/TAW/2/99 ●
O/CHINA/1/99 ●
O/TIBET/CHA/99 ●
A/A2SPAIN/ISO7 ■
A/A4WGERMANY/ISO42 ■
A/A4WG/ISO72 ■
A/A10HOLLAND/ISO82 ■
A/A1BAYERN/ISO41 ■
A/A3MECKLENBURG/ISO81 ■
A/A12VALLE119/ISO20 ■
A/A14SPAIN/ISO39 ■
A/A4SPAIN/ISO62 ■
A/A5ALLIER/ISO45 ■
A/A5WESTERWALD/ISO73 ■
A/ABRAZIL/ISO67 ■
A/APHILIPPINES/ISO50 ■
A/A24CRUZEIRO/ISO71 ■
A/A29PERU/ISO37 ■
A/A17AGUARULBOS/ISO83 ■
A/A13BRAZIL/ISO75 ■
A/A18ZULIA/ISO48 ■
A/ETH/7/92 ∗
A/TAN/4/80 ∗
A/SOM/1/78 ∗
A/ETH/2/79 ∗
A/ERI/3/98 ∗
A/NIG/4/79 ∗
A/CIV/4/95 ∗
A/SEN/10/97 ∗

73
100
(100) 100
100
95
100
72
(100) 100
91
100
75
74
98
100
100
78
(71) 79
100
100
(86) 82
81
92

0.05

S2: Neighbour-joining tree depicting gene relationships for the 1C region of A and O type viruses. Kimura-2 parameter and bootstrap (1000 replications) were applied.

- ● Pan Asian FMDV O
- * African FMDV A and O types (this study)
- ■ Non-African FMDV A and O types
- ( ) Bootstrap value from minimum Evolution tree

: Neighbour-joining tree depicting gene relationships for the 1D region of A and O type viruses. Kimura-2 parameter and bootstrap (1000 replications) were applied.

- Pan Asian FMDV O
* African FMDV A and O types (this study)
■ Non-African FMDV A and O types
( ) Bootstrap value from minimum
   Evolution tree

76 O/UKG/35/2001 •
96 O/FRA/1/2001 •
O/SAR/19/2000 •
92 O/TAW/2/99 •
O/TIBET/CHA/99 •
(100) 100 O/CHINA/1/99 •
O/SKR/2000 •
93 O/O10PHIL76 ■
100 O/O10PHIL54 ■
O/O1MANISA87 ■
O/TAN/3/96 *
O/UGA/7/03 *
73 O/KEN/10/95 *
O/UGA/5/96 *
85 O/UGA/17/98 *
O/SUD/4/80 *
O/UGA/1/75 *
O/UGA/6/76 *
76 O/ETH/3/96 *
O/AKESU/58 ■
100 O/O11INDONESIA52 ■
O/O1BRUGGE79 ■
O/O1ARGENTINA5 ■
100 O/O1CAMPOS94 ■
O/O1CAMPOS96 ■
O/O1BFS46 ■
99 O/O1BFS18 ■
A/A24CRUZEIRO/ISO71 ■
77 A/ABRAZIL/ISO67 ■
A/A29PERU/ISO37 ■
99 A/APHILIPPINES/ISO50 ■
99 A/A4SPAIN/ISO62 ■
A/A14SPAIN/ISO39 ■
100 A/A5WESTERWALD/ISO73 ■
A/A5ALLIER/ISO45 ■
91 A/A12VALLE119/ISO20 ■
79 A/A10HOLLAND/ISO82 ■
A/A1BAYERN/ISO41 ■
69 A/A3MECKLENBURG/ISO81 ■
A/A2SPAIN/ISO7 ■
99 A/A4WG/ISO72 ■
94 A/A4WGERMANY/ISO42 ■
(98) 99 A/A17AGUARULBOS/ISO83 ■
A/A13BRAZIL/ISO75 ■
100 A/A18ZULIA/ISO48 ■
100 A/ETH/2/79 *
A/SOM/1/78 *
83 A/TAN/4/80 *
A/ETH/7/92 *
A/ERI/3/98 *
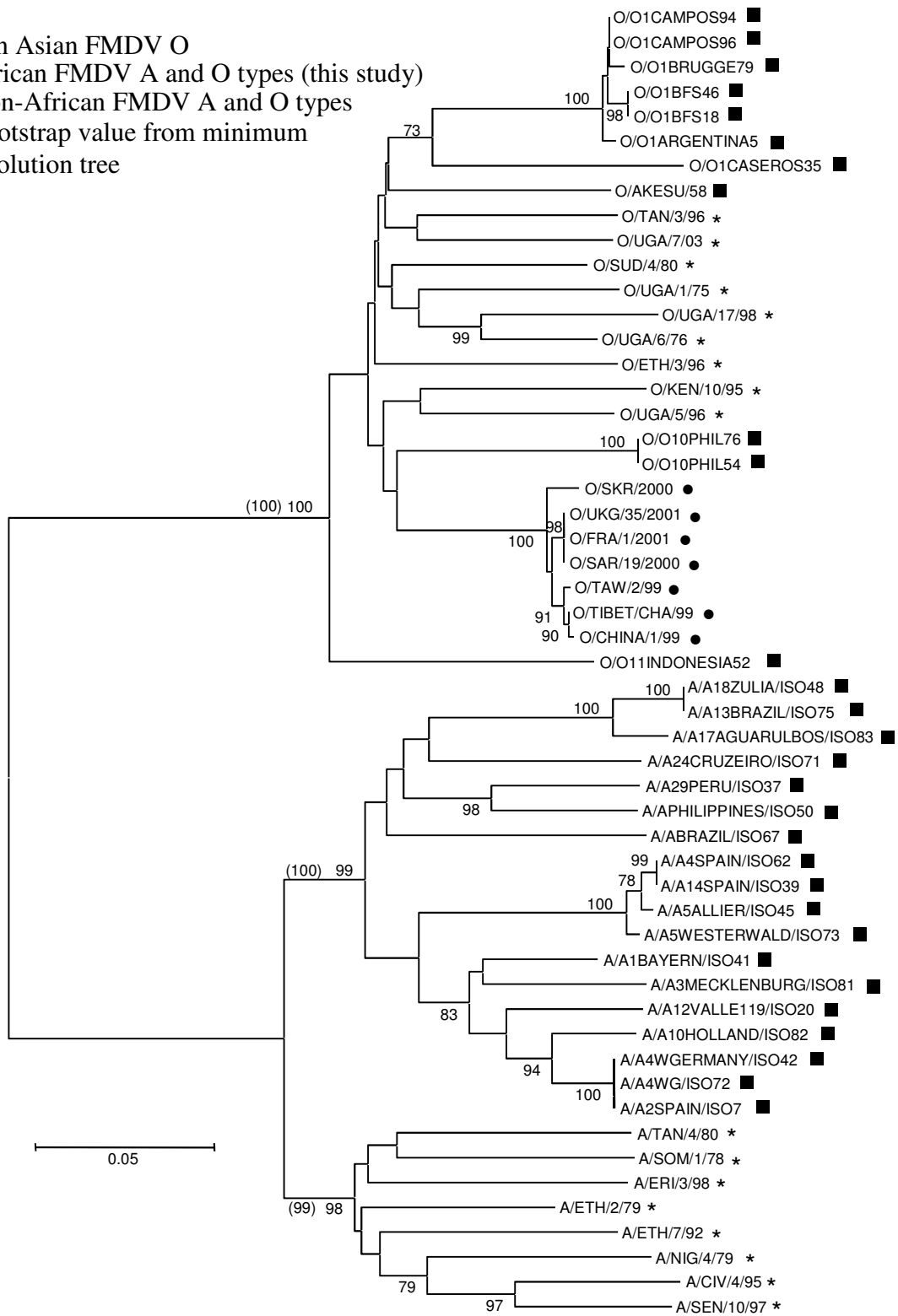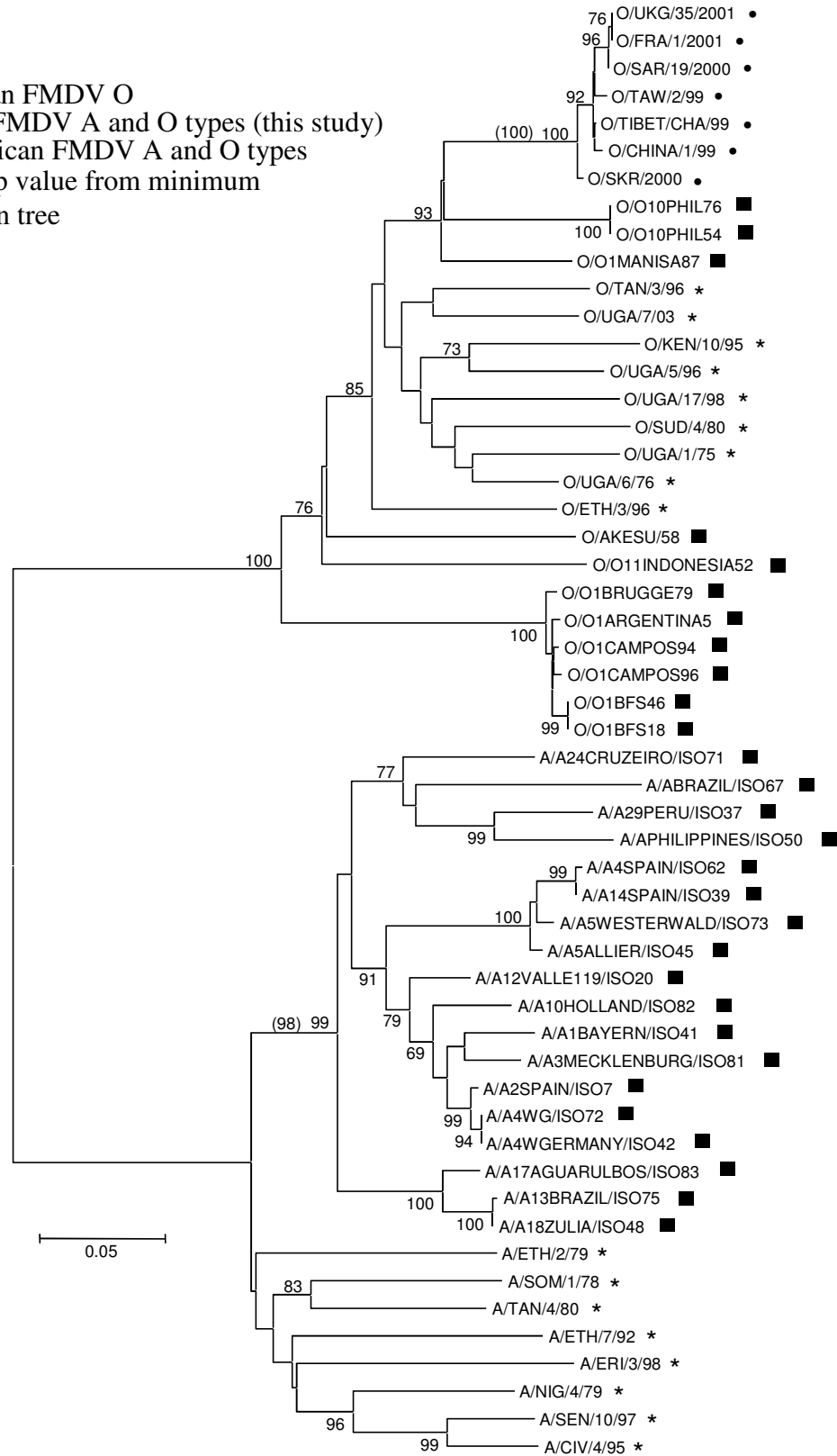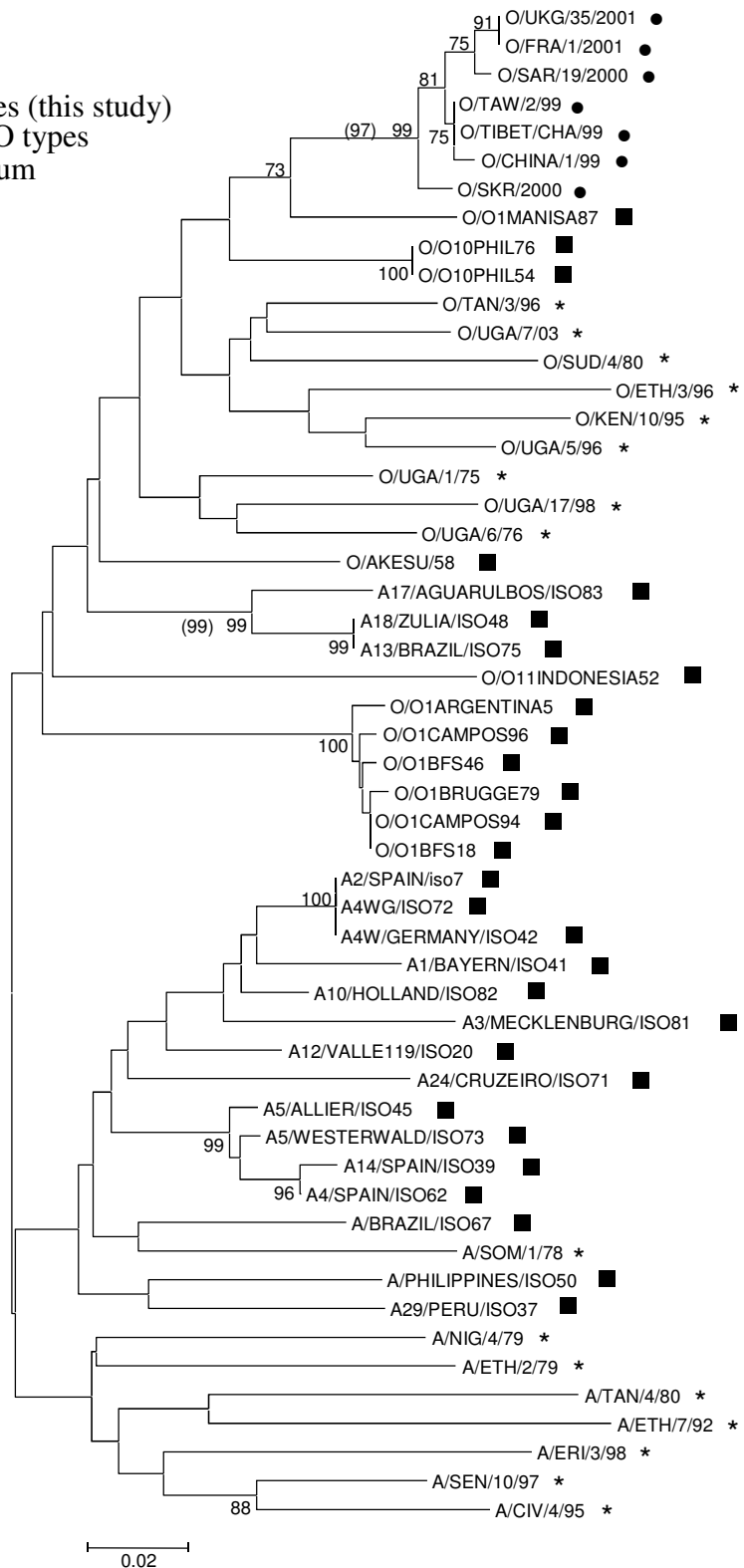A/NIG/4/79 *
96 A/SEN/10/97 *
99 A/CIV/4/95 *

0.05

S4: Neighbour-joining tree depicting gene relationships for the 1A region of A and O type viruses. Kimura-2 parameter and bootstrap (1000 replications) were applied.

● Pan Asian FMDV O
* African FMDV A and O types (this study)
■ Non-African FMDV A and O types
( ) Bootstrap value from minimum
   Evolution tree

91 O/UKG/35/2001 ●
75 O/FRA/1/2001 ●
81 O/SAR/19/2000 ●
(97) 99 O/TAW/2/99 ●
75 O/TIBET/CHA/99 ●
O/CHINA/1/99 ●
73 O/SKR/2000 ●
O/O1MANISA87 ■
O/O10PHIL76 ■
100 O/O10PHIL54 ■
O/TAN/3/96 *
O/UGA/7/03 *
O/SUD/4/80 *
O/ETH/3/96 *
O/KEN/10/95 *
O/UGA/5/96 *
O/UGA/1/75 *
O/UGA/17/98 *
O/UGA/6/76 *
O/AKESU/58 ■
A17/AGUARULBOS/ISO83 ■
(99) 99 A18/ZULIA/ISO48 ■
99 A13/BRAZIL/ISO75 ■
O/O11INDONESIA52 ■
O/O1ARGENTINA5 ■
100 O/O1CAMPOS96 ■
O/O1BFS46 ■
O/O1BRUGGE79 ■
O/O1CAMPOS94 ■
O/O1BFS18 ■
A2/SPAIN/iso7 ■
100 A4WG/ISO72 ■
A4W/GERMANY/ISO42 ■
A1/BAYERN/ISO41 ■
A10/HOLLAND/ISO82 ■
A3/MECKLENBURG/ISO81 ■
A12/VALLE119/ISO20 ■
A24/CRUZEIRO/ISO71 ■
A5/ALLIER/ISO45 ■
99 A5/WESTERWALD/ISO73 ■
A14/SPAIN/ISO39 ■
96 A4/SPAIN/ISO62 ■
A/BRAZIL/ISO67 ■
A/SOM/1/78 *
A/PHILIPPINES/ISO50 ■
A29/PERU/ISO37 ■
A/NIG/4/79 *
A/ETH/2/79 *
A/TAN/4/80 *
A/ETH/7/92 *
A/ERI/3/98 *
A/SEN/10/97 *
88 A/CIV/4/95 *

0.02

S5: A comparison of previously identified critical residues for FMDV to the FMDV A and O types from this study

aa Regions identified as important	Protein: aa residue/s *	A (African and non-African viruses): N (number of isolates aligned) = 26	A (African viruses): N=8	O (African and non-African viruses): N=27	O (African viruses): N=9
N-terminal myristylation site & swine and bovine T-cell epitope (Blanco et al., 2001)	**1A**: 19 - 35	c	c	Y31H (1)	Y31H (1)
A Q residue distinguishes the SAT types from the Euro-Asiatic lineages (Carrillio et al., 2005).	**1A**: 73	S (25) N(1)	S (8)	S (27)	S (9)
The I residue is potentially specific for the SAT2 and SAT3 viruses (Carrillo et al., 2005)	**1A**: 76	F (26)	F (8)	F (27)	F (9)
A valine is potentially specific for SAT1 (Carrillo et al., 2005).	**1A**: 80	F (26)	F (8)	F (27)	F (9)
Identified 1B T-cell epitopes (Perez et al., 2000).	**1B**: 133 – 153	I at position 141 (2) T at position 141 (1) Y at position 146 (3) R at position 149 (1) F at position 150 (17) H at position 150 (7) E at position 153 (3)	R at position 149 (1) H at position 150 (7)	A at position 141 (3)	A at position 141 (2)
	198 – 215	I at position 211 (1)	c	c	c
	263 - 271	I at position 265 (1) L at position 268 (1)	c	c	c
Side chains involved in non-covalent interactions between pentamer subunits in the FMDV capsid (Mateo et al., 2003)	**1B**: E96, I99, R103, H106, T108, T111, Q112, L136, T138, Q142, R145, K148, Y183, E193, V197, N199, Q200, F201,	c	c	c	c

	M239, K283, Y285, N287, E298, T107				
	K173	H (12), R (5), S (1), A (2), T (5)	A (2), T (5)	*c*	*c*
	K181	D (19), E (5), N (2)	D (5), E (1), N (2)	D (26), E (1)	D (9)
	T195	S(25)	S (7)	*c*	*c*
	1C: K422, R424, M426, H446, H449, E451, D453, L456, N457, I494, T495, K498	*c*	*c*	*c*	*c*
	D372	G (2)	*c*	E (1)	E (1)
	Q374	T (16), N (3)	T (2)	D (25), F (1)	D (8), F (1)
H residues identified in the 1B/1C regions and are proposed to mediate 1B/1C hydrogen bonding (Acharya *et al.*, 1989).	**1B:** 106 172 230 242 259	*c* *c* *c* *c* *c*	*c* *c* *c* *c* *c*	*c* *c* *c* *c* *c*	*c* *c* *c* *c* *c*
	1C: 388 412 446 449 496	*c* L (1) *c* *c* *c*	*c* L (1) *c* *c* *c*	T (1) *c* *c* *c* D (1)	T (1) *c* *c* *c* D (1)
C residues at these positions are important for the formation a disulphide bond at the base of the G-H loop	**1B:** 215 **1D:** 660	K (26) S (15) N (9) D (2)	K (8) S (7) N (1)	C (27) S (1)	C (9) *c*
A R residue is important for HS binding (Fry *et al.*, 1999).	**1C:** 359	C (26)	C (8)	H (25) R (2)	H (8) R (1)

The H residue plays a significant role on FMDV capsid destabilization (v. Vlijmen *et al.*, 1998; Ellard *et al.*, 1999).	**1C**: 449	*c*	*c*	*c*	*c*
The P residue contributes to the 1A/1B cleavage pocket (Carrillo *et al.*, 2005).	**1D**: 715	*c*	*c*	*c*	*c*
The RGD residue is important for receptor binding (Fox *et al.,* 1989; Mason *et al.,* 1994; Berinstein *et al.,* 1995; Jackson *et al.*, 1997).	**1D**: 671 - 673	*c*	*c*	SGD (1)	*c*
A L residue is most common and important for virus receptor (αvβ6) recognition and plays a role in the stability of the virus/integrin complex (DiCara *et al.*, 2008)	**1D**: 674

677 | M (6)
S (2)
T (3)

H (2)
P(2) | M (1)

c | *c*

c | *c*

c |

* The aa numbering is according to the P1 alignment in Appendix A8. <u>Note</u>: Not all the sequences referred to in this table is included in Appendix A8. This table is based on the complete alignment of all the isolates included in this study (results not shown).

c Refers to the aa residues that are conserved in the P1 alignment.
The numbers in brackets ( ) indicate the number of isolates where there is a difference in aa and the letter next to the brackets refers to the aa change for those isolates.

N The number of isolates aligned