Determination of common genetic variants within the non-structural proteins of foot-and-mouth disease viruses isolated in sub-Saharan Africa

P. Nsamba [a,b], T.A.P. de Beer [c,1], M. Chitray [a], K. Scott [a], W. Vosloo [b,d], F.F. Maree [a,e,*]

[a] *Agricultural Research Council, Onderstepoort Veterinary Institute, Transboundary Animal Diseases Programme, Private Bag X05, Onderstepoort 0110, South Africa*

[b] *Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, Private Bag X04, Onderstepoort 0110, South Africa*

[c] *European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, United Kingdom*

[d] *CSIRO-Australian Animal Health Laboratory, Private Bag 24, Geelong 3220, VIC, Australia*

[e] *Department of Microbiology and Plant Pathology, Faculty of Agricultural and Natural Sciences, University of Pretoria, Pretoria 0002, South Africa*

Keywords:
Foot-and-mouth disease virus
Non-structural proteins
Leader-protease, 2C/AAA+ ATPase
3A-protein
3B/Vpg-protein
3C-protease, RNA-dependent RNA polymerase

A B S T R A C T

The non-structural proteins of foot-and-mouth disease virus (FMDV) are responsible for RNA replication, proteolytic processing of the viral polyprotein precursor, folding and ssembly of the structural proteins and modification of the cellular translation apparatus. nvestigation of the amino acid heterogeneity of the non-structural proteins of seventy-nine FMDV isolates of SAT1, SAT2, SAT3, A and O serotypes revealed between 29 and 62% amino acid variability. The Leader protease (L^{pro}) and 3A proteins were the most variable whilst the RNA-dependent RNA polymerase ($3D^{pol}$) the most conserved. Phylogeny based on the non-structural protein-coding regions showed separate clusters for southern African viruses for both the L^{pro} and 3C protease ($3C^{pro}$) and sequences unique to this group of viruses, e.g. in the 2C and $3C^{pro}$ proteins. These groupings were unlike serotype groupings based on structural protein-coding regions. The amino acid substitutions and the nature of the naturally occurring substitutions provide insight into the functional domains and regions of the non-structural proteins that are critical for structure–function. The L^{pro} of southern African SAT type isolates differed from A, O and SAT isolates in northern Africa, particularly in the auto-processing region. Three-dimensional structures of the 3C protease ($3C^{pro}$) and $3D^{pol}$ showed that the observed variation does not affect the enzymatic active sites or substrate binding sites. Variation in the $3C^{pro}$ cleavage sites demonstrates broad substrate specificity.

1. Introduction

Foot-and-mouth disease (FMD) is widely considered the most economically important disease of livestock, and is a notifiable disease in many countries due to its highly contagious nature and associated productivity losses among even-toed ungulates (*Artiodactyla*). The disease is

* Corresponding author at: Transboundary Animal Diseases Programme, Onderstepoort Veterinary Institute, Private Bag X05, Onderstepoort 0110, South Africa. Tel.: +27 12 529 9584/93; fax: +27 12 529 9595/05.

E-mail address: mareef@arc.agric.za (F.F. Maree).

[1] Current address: Biozentrum & SIB Swiss Institute of Bioinformatics, University of Basel, Klingelbergstrasse 50/70, 4056 Basel, Switzerland.

endemic to large parts of the African continent and an impediment to lucrative export markets for animal products (Vosloo et al., 2002). The different serotypes of FMD virus (FMDV) cause a clinically indistinguishable vesicular disease in cloven-hoofed animals and display different geographical distributions and epidemiology (Bastos et al., 2001, 2003a,b; Bronsvoort et al., 2004; Knowles and Samuel, 2003; Samuel and Knowles, 2001). Of the seven serotypes, the South African Territories (SAT) types 1, 2 and 3 are confined to sub-Saharan Africa, although incursions into the Middle East by SAT1 (1961–1965 and 1970) and SAT2 (1990, 2000 and currently in North Africa) viruses have been recorded (Bastos et al., 2001; Ferris and Donaldson, 1992; Valdazo-González et al., 2012; records of the Office International des Épizooties or OIE). In contrast, serotypes A and O occur globally (Samuel and Knowles, 2001) with the exception of southern Africa (Vosloo et al., 2002).

FMDV is a small non-enveloped virus, a member of the *Aphthovirus* genus within the family *Picornaviridae*. The icosahedral capsid consists of 60 copies of four structural proteins, VP1 to 4, arranged in a pseudo $T = 3$ composition. The three surface-exposed proteins, VP1 (1D), VP2 (1B) and VP3 (1C), assemble into a protomeric subunit, with the smaller VP4 (1A) located internally (Acharya et al., 1989; Curry et al., 1995; Sobrino et al., 2001). Despite the high levels of genetic and antigenic variation (Vosloo et al., 1995; Reeve et al., 2010; Maree et al., 2011), a consequence of the high mutation rate of the virus, the structural arrangement of the capsid is remarkably conserved, indicating plasticity within the three-dimensional structure of the capsid proteins (Acharya et al., 1989; Curry et al., 1995; Fry et al., 1999; Lea et al., 1994). The capsid encloses a ca. 8.5 kilobase, positive-sense, single-stranded RNA genome with a single open reading frame and two in-frame translation-initiation codons. Covalently linked to the 5′ end of the genome is the small viral protein 3B (or VPg), while the 3′ end is poly-adenylated (Carrillo et al., 2005). Upon virus infection, the interactions between VP2 and VP3 at the pentameric interfaces are disrupted by acidification within cellular endosomes, thereby releasing the viral RNA (Ellard et al., 1999; Knipe et al., 1997). The viral genome is rapidly translated into a polyprotein which is co- and post-translationally cleaved by viral proteinases into several partially cleaved intermediates and ultimately into 12 mature proteins (Pereira, 1981; Rueckert, 1996).

In addition to the capsid proteins, the ORF of the viral RNA genome encodes eight non-structural proteins, each with its unique function within the replication cycle of FMDV (Belsham, 1993, 2005). The non-structural proteins include three proteases, i.e. L^{pro}, 2A and $3C^{pro}$, responsible for cleavage of the viral polyprotein and shut-down of the host cap-dependent translational system (Bablanian and Grubman, 1993; Martinez-Salas et al., 1996). Although several of the picornavirus proteins involved in RNA replication (2B, 2C and 3A) have membrane binding properties and disrupt protein trafficking in the cell (Moffat et al., 2005, 2007), their particular functions during viral replication are still unknown. The 2B protein has been implicated in virus-induced cytopathic effect (CPE) (van Kuppeveld et al., 1997), while the 2C protein has recently been classified as an AAA+ ATPase enzyme that may act as an RNA helicase (Sweeney et al., 2010). The 3D gene encodes the viral RNA-dependent RNA polymerase (RDRP), and together with the 3A co-localizes with ER membrane-associated replication complexes (Lama et al., 1994).

Based on the genetic variability of the VP1-coding region, the FMDV strains that exist among the serotypes, group into topotypes that are geographically specific (Knowles and Samuel, 2003; Samuel and Knowles, 2001). Serotype A has three topotypes, of which the Africa topotype is endemic to sub-Saharan Africa. Of the eleven topotypes defined for serotype O, five are endemic in Africa; the East Africa (EA1–EA4) and West Africa (WA) topotypes (Di Nardo et al., 2011). On the other hand, the SAT serotypes are genetically more diverse, and nine, fourteen and five topotypes have been defined for SAT1, SAT2 and SAT3 respectively (Bastos et al., 2001, 2003a,b; Knowles et al., 2010).

A few studies have looked at genome comparisons mainly focusing on serotypes A, O, C and Asia-1 with a geographic distribution in Euro-Asia and South America (Carrillo et al., 2005; Mason et al., 2003; Pereda et al., 2002). However, a limited number of complete non-structural protein analyses for viruses belonging to the SAT serotypes have been described (Carrillo et al., 2005). Here we have compared the non-structural proteins and their coding regions for the three SAT serotypes and viruses from serotype A and O found in sub-Saharan Africa from 1974 to 2006. Additionally, the natural variation found within the non-structural protein sequences of 79 viruses to identify structurally and possibly functionally conserved regions were mapped. Variation in the $3C^{pro}$ and $3D^{pol}$ was also mapped onto the protein structures to improve understanding of the plasticity of these enzymes. The deduced amino acid sequences of the non-structural proteins of two closely related SAT2 viruses, causing numerous of outbreaks in North Africa and the Middle East in 2012, were also included (Valdazo-González et al., 2012).

2. Materials and methods

2.1. Virus isolates

The viruses included in this study were either supplied by the World Reference Laboratory (WRL) for FMD at the Pirbright Institute (United Kingdom) or form part of the virus collection at the Transboundary Animal Diseases Programme, Onderstepoort (South Africa). The SAT1 ($n = 30$), SAT2 ($n = 26$), SAT3 ($n = 7$), serotype A ($n = 7$) and serotype O ($n = 10$) FMDV isolates from 17 countries in sub-Saharan Africa were selected for genomic characterization. The isolates span a 32 year time period and represent various geographic locations and animal species. The viruses were propagated in IB-RS-2 cells prior to RNA extraction, cDNA synthesis and amplification of the relevant genome regions by PCR (Maree et al., 2011). A description of the passage histories, host species and representative topotypes can be found in Table 1.

Table 1
Viruses used in this study.

	Virus strain	Topo-type[b]	Species[c]	Passage history[d]	Country of origin	GenBank accession number
1	SAT1/KNP/196/91	1	Buffalo	PK1 RS1	South Africa	KF647852
2	SAT1/KNP/148/91	1	Buffalo	PK1 RS5	South Africa	KF647853
3	SAT1/KNP/41/95	1	Buffalo	PK1 RS4	South Africa	KF647856
4	SAT1/SAR/9/81	1	Impala	Epithelium	South Africa	
5	SAT1/ZIM/HV/03/90	1	Buffalo	BTY1 RS3	Zimbabwe	KF647854
6	SAT1/ZIM/GN/13/90	1	Buffalo	BTY1 PK1 RS3	Zimbabwe	KF647855
7	[a]SAT1/BOT/1/68	1	N/A	BTY3 BHK5	Botswana	AY593845
8	[a]SAT1/RHOD/5/66	1	N/A	BTT1	Rhodesia	AY593846
9	[a]SAT1/RV/11/37	1	N/A	IBRS3	Unknown	AY593839
10	SAT1/NAM/307/98	2	Buffalo	PK1 RS4	Namibia	KF647863
11	SAT1/ZIM/6/94	2	Bovine	PK1 RS3	Zimbabwe	KF647864
12	[a]SAT1/BEC/1/70	2	N/A	LK2	Botswana	AY593838
13	[a]SAT1/SR/2/58	2	N/A	IBRS3	Rhodesia	AY593841
14	[a]SAT1/SWA/1/49	2	N/A	IBRS4	Unknown	AY593840
15	[a]SAT1/SA/5/61	2	N/A	IBRS3	Unknown	AY593842
16	[a]SAT1/SWA/40/61	2	N/A	IBRS3	Unknown	AY593843
17	SAT1/ZIM/25/90	3	Buffalo	BTY2 RS4	Zimbabwe	KF647857
18	SAT1/ZAM/2/93	3	Buffalo	PK1 RS3	Zambia	KF647862
19	SAT1/MOZ/3/02	3	Bovine	PK1 RS5	Mozambique	KF647858
20	SAT1/TAN/1/99	3	Bovine	PK2 RS4	Tanzania	KF647860
21	SAT1/KEN/05/98	3	Bovine	BTY1 RS3	Kenya	KF647861
22	SAT1/UGA/1/97	5	Buffalo	PK1 RS4	Uganda	KF647865
23	SAT1/UGA/3/99	4	Bovine	BTY1 RS4	Uganda	KF647866
24	SAT1/SUD/3/76	7	Bovine	BTY1 RS3	Sudan	KF647868
25	SAT1/NIG/5/81	7	Bovine	BTY2 RS2	Nigeria	KF647867
26	SAT1/NIG/15/75	8	Bovine	BTY1 RS3	Nigeria	KF647869
27	SAT1/NIG/8/76	8	Bovine	BTY1 RS5	Nigeria	KF647871
28	SAT1/NIG/6/76	8	Bovine	BTY1 RS5	Nigeria	KF647870
29	SAT2/KNP/19/89	I	Buffalo	BHK4	South Africa	KJ144902
30	SAT2/KNP/2/89	I	Impala	CFK2 RS2 BHK4	South Africa	KJ144903
31	SAT2/KNP/51/93	I	Impala	PK1 RS6	South Africa	KJ144904
32	SAT2/SAR/16/83	I	Impala	B1 BHK8	South Africa	KJ144905
33	SAT2/ZIM/1/88	I	Buffalo	CFK1 RS4	Zimbabwe	KJ144908
34	[a]SAT2/106/67	I	N/A	IBRS3	Unknown	AY593848
35	[a]SAT2/ZIM/7/83	II	Bovine	B1 BHK5 B1	Zimbabwe	AF540910
36	SAT2/ZIM/14/90	II	Buffalo	BTY1 RS3	Zimbabwe	KJ144910
37	SAT2/ZIM/17/91	II	Buffalo	BTY2 RS4	Zimbabwe	KJ144911
38	SAT2/ZIM/GN/10/91	II	Buffalo	BTY2 PK1 RS3	Zimbabwe	KJ144907
39	SAT2/ZIM/34/90	II	Buffalo	BTY3 RS4	Zimbabwe	KJ144909
40	SAT2/ZIM/8/94	II	Buffalo	BTY1 RS3	Zimbabwe	KJ144906
41	[a]SAT2/RHO/1/48	II	Bovine	BTY2 RS2	Zimbabwe	AY593847
42	SAT2/ZAM/07/96	III	Bovine	BTY1 RS2	Zambia	KJ144912
43	SAT2/KEN/8/99	IV	Bovine	BTY2 RS4	Kenya	KJ144913
44	SAT2/SEN/5/75	V	Bovine	BTY1 RS1 BHK5	Senegal	KJ676543
45	SAT2/GHA/8/91	V	Bovine	BTY1 RS3	Ghana	KJ144917
46	SAT2/SEN/7/83	VI	Bovine	CFK1 RS1	Senegal	KJ144916
47	SAT2/ERI/12/98	VII	Bovine	BTY2 PK1 RS5	Eritrea	KJ144921
48	SAT2/SAU/6/00	VII	Bovine	BTY1 RS1	Saudi Arabia	KJ144920
49	SAT2/RWA/2/01	VIII	Bovine	PK1 RS1	Rwanda	KJ144919
50	[a]SAT2/KEN/3/57	IX	Bovine	N/A	Kenya	KJ144915
51	[a]SAT2/KEN/11/60	IX	N/A	IBRS3	Kenya	AY593849
52	SAT2/ANG/4/74SAT2/	XI	Bovine	BTY3 RS3	Angola	KJ144914
53	UGA/2/02 [a]SAT2/	XII	Bovine	PK1 RS1	Uganda	KJ144918
54	UGA/MBF/4/02[a]SAT2/	XII	Buffalo	N/A	Uganda	FJ461346
55	EGY/9/12[a]SAT2/	VII	Bovine	N/A	Egypt	JX014255
56	PAT/1/12	VII	Bovine	N/A	Palestinian Autonomous Territories	JX014256
57	SAT3/KNP/10/90	I	Buffalo	PK1 RS1	South Africa	KF647849
58	[a]SAT3/SA/57/59	I	N/A	LK1	Unknown	AY593850
59	[a]SAT3/BEC/1/65	II	Bovine	LK1	Botswana	AY593853
60	[a]SAT3/KEN/11/60	IV	N/A	IBRS4	Kenya	AY593852
61	[a]SAT3/BEC/20/61	II	N/A	LK1	Botswana	AY593851
62	SAT3/ZIM/5/91	III	Buffalo	BTY1 RS4	Zimbabwe	KF647851
63	SAT3/ZAM/4/96	IV	Buffalo	BTY1 RS1	Zambia	KF647850
64	A/CIV/4/95	I	Bovine	BTY1	Coˆted'Ivoire	KJ144932
65	A/SEN/10/97	I	Bovine	BTY2	Senegal	KJ144933

	Virus strain	Topo-type[b]	Species[c]	Passage history[d]	Country of origin	GenBank accession number
66	A/SOM/1/78	III	N/A	BTY2	Somalia	KJ144935
67	A/TAN/4/80	III	N/A	BTY2	Tanzania	KJ144937
68	A/ETH/2/79	IV	N/A	BTY5	Ethiopia	KJ144936
69	A/ETH/7/92	IV	N/A	BTY1	Ethiopia	KJ144934
70	A/ERI/3/98	V	N/A	BTY1	Eritrea	KJ144931
71	O/UGA/5/96	II	N/A	IBRS2	Uganda	KJ144924
72	O/UGA/6/76	II	N/A	IBRS2	Uganda	KJ144926
73	O/UGA/17/98	II	N/A	IBRS2	Uganda	KJ144927
74	O/UGA/1/75	II	N/A	IBRS2	Uganda	KJ144925
75	O/UGA/7/03	II	Bovine	PK1	Uganda	KJ144922
76	O/KEN/10/95	II	N/A	IBRS2 IBRS3	Kenya	KJ144928
77	O/SUD/4/80	III	N/A	IBRS2	Sudan	KJ144930
78	O/ETH/3/96	IV	N/A	IBRS2	Ethiopia	KJ144929
79	O/TAN/3/96	VI	N/A	IBRS2	Tanzania	KJ144923

[a] Genome sequences obtained from GenBank.
[b] Reference for topotypes: Bastos et al. (2001, 2003a,b), Maree et al. (2011) and Knowles and Samuel (2003).
[c] The species the virus was isolated from. N/A: not available.
[d] The passage history of the virus is indicated as follows: the cell type, followed by the number of times the virus has been passaged in that particular cell type, in numerals (1, 2, 3 etc.). BHK – baby hamster kidney cells; B – bovine; BTT – bovine tongue tissue; BTY – bovine thyroid; CFK – calf foetal kidney; LK – lamb kidney cells; PK – pig kidney; RS/IBRS2 – Instituto Biologico Renal Suino-2 Cells; N/A – information not available.

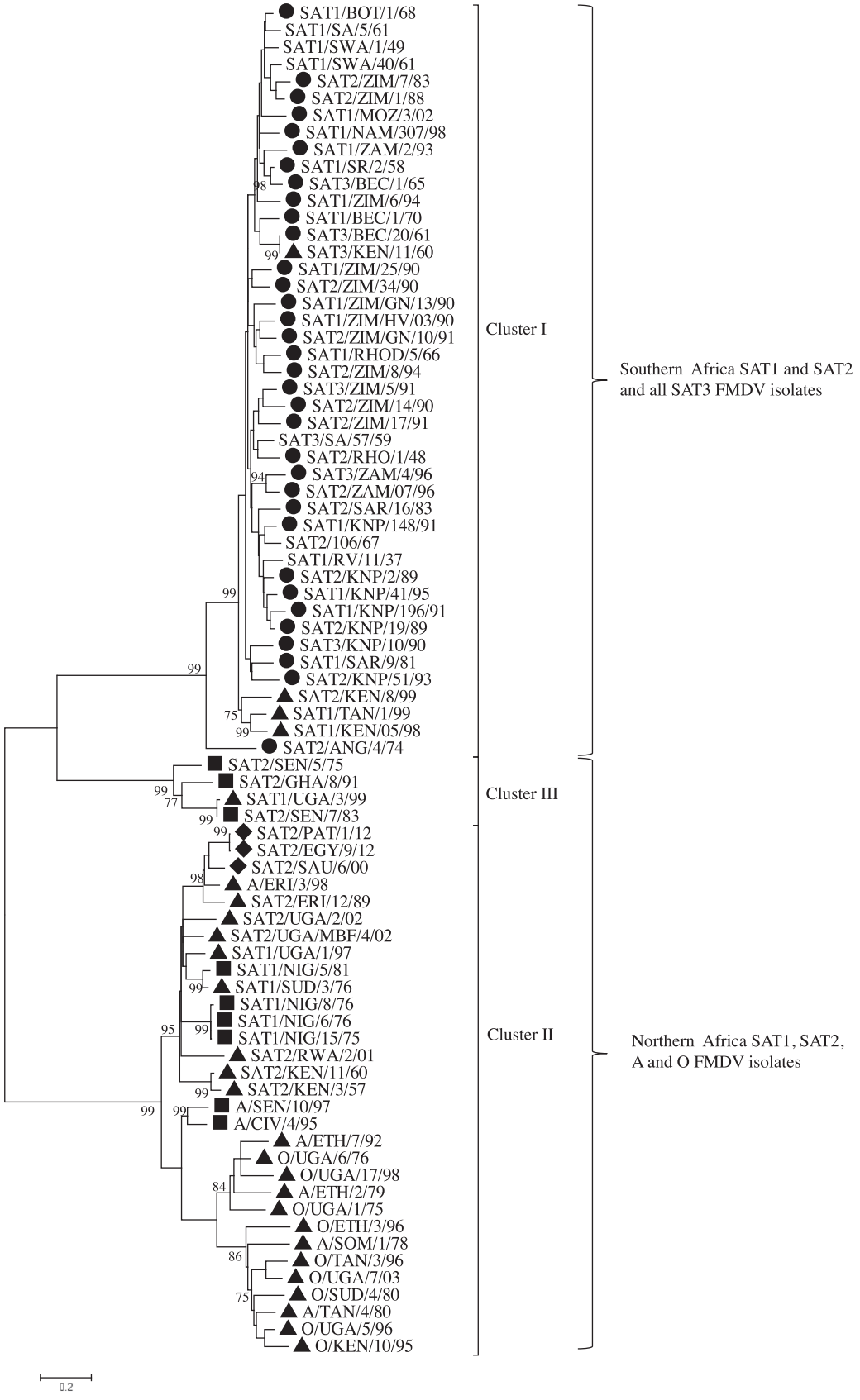
2.2. RT-PCR, sequencing and analysis

Viral RNA was extracted from infected cell culture supernatant using a modification (Bastos, 1998) of the guanidinium-silica based method described by Boom et al. (1990). To facilitate amplification of the untranslated regions (UTR's), Leader-P1-2A and P2-P3-coding regions, the viral genomic RNA was reverse transcribed with SuperScript III™ (Life Technologies) using either the oligonucleotide 2B-208R (Bastos et al., 2001) for the 5′UTR, Leader-P1-2A coding region or a modified oligo-dT, CCATGGCGGCCGCTTTTTTTTTTTTTTT, for the P2-P3 coding region and 3′UTR. The non-structural protein coding region and 3′UTR were subsequently amplified by three or four separate PCR reactions using the oligonucleotides detailed in Table 2 and Expand Long template Taq DNA polymerase™ (Roche). The cycling conditions were 95 °C for 20 s, 56 °C for 20 s, 68 °C for 2–3 min (30 cycles). Direct DNA sequencing of amplicons, using the ABI PRISM™ BigDye Terminator Cycle Sequencing Ready Reaction Kit v3.0 (Perkin Elmer Applied Biosystems) and genome-specific primers, yielded a consensus sequence

Table 2
A summary of the oligonucleotides that was used for PCR amplification of the L^{pro}, P2- and P3-coding regions. The orientation, their location on the FMDV genome and the serotype specificity are indicated.

PCR fragment amplified in the different serotypes	Primer oligonucleotide				
	Name	Sequence (5′–3′)	Orientation	Location	Serotype specificity
L-Frag 5′UTR	P664	CAGCGTCTCCCCCCCCCCCCCCC (this study)	Forwards	Poly C	SATs, A and O
	SEQ 130	CCGTTGTAVAGTGTGAATTCC (this study)	Reverse	Leader	SATs
Leader-P1-2A (SATs)	NCR1	TACCAAGCGACACTCGGGATCT (Maree et al., 2011)	Forward	5′UTR	SATs, A and O
Leader-P1 (A and O)	NCR2	GCTTCTATGCCTGAATAGG (Maree et al., 2011)	Forward	5′UTR	SATs
	2B-208R	ACAGCGGCCATGCACGACAG (Bastos et al., 2001)	Reverse	2B	SATs
	2BR-UGA	GCTACAGCGGCCATRCAYGACA (this study)	Reverse	2B	SATs East and West Africa
	W-DA	GAAGGGCCCAGGGTTGGACTC (Beck and Strohmaier, 1987)	Reverse	VP1/2B	A and O
2A (A and O)	VP_1 O	GATTTGTGAAGGTGACACC (Rodriguez et al., 1994)	Forward	VP1	O-type
	A1C-562F	TACCAAATTACACACGGGAA (Reid et al., 2000)	Forward	VP3	A-type
	2B-208R	ACAGCGGCCATGCACGACAG (Bastos et al., 2001)	Reverse	2B	A and O
	2BR-UGA	GCTACAGCGGCCATRCAYGACA (this study)	Reverse	2B	A and O
2BC, $3AB_{123}$, 3C (SATs)	SEQ 91	GAGTCCAACCCTGGGCCCTTCTTCTTC (this study)	Forward	VP1/2A	All
2BC, $3AB_{123}$ (A and O)	SEQ 92	CGCTCYTCAACWTCTCTGGTGTC (this study)	Reverse	3D	SATs
	SEQ 126	TCCATRCACACTACAATGTC (this study)	Reverse	3C	A and O
3C–3′UTR (all serotypes)	NC3+	CCKGTGAAGAAGCCTGTCGC (this study)	Forward	3B	All
	Poly SAT DT	CCATGGCGGCCGCTTTTTTTTTTTTTTT (this study)	Reverse	Poly A	All

Y = CT; M = AC; K = AG; W = AT; S = CG; K = TG; D = AGT; V = AGC; H = ACT; B = GCT.


SAT1/BOT/1/68
SAT1/SA/5/61
SAT1/SWA/1/49
SAT1/SWA/40/61
SAT2/ZIM/7/83
SAT2/ZIM/1/88
SAT1/MOZ/3/02
SAT1/NAM/307/98
SAT1/ZAM/2/93
SAT1/SR/2/58
SAT3/BEC/1/65
SAT1/ZIM/6/94
SAT1/BEC/1/70
SAT3/BEC/20/61
SAT3/KEN/11/60
SAT1/ZIM/25/90
SAT2/ZIM/34/90
SAT1/ZIM/GN/13/90
SAT1/ZIM/HV/03/90
SAT2/ZIM/GN/10/91
SAT1/RHOD/5/66
SAT2/ZIM/8/94
SAT3/ZIM/5/91
SAT2/ZIM/14/90
SAT2/ZIM/17/91
SAT3/SA/57/59
SAT2/RHO/1/48
SAT3/ZAM/4/96
SAT2/ZAM/07/96
SAT2/SAR/16/83
SAT1/KNP/148/91
SAT2/106/67
SAT1/RV/11/37
SAT2/KNP/2/89
SAT1/KNP/41/95
SAT1/KNP/196/91
SAT2/KNP/19/89
SAT3/KNP/10/90
SAT1/SAR/9/81
SAT2/KNP/51/93
SAT2/KEN/8/99
SAT1/TAN/1/99
SAT1/KEN/05/98
SAT2/ANG/4/74
SAT2/SEN/5/75
SAT2/GHA/8/91
SAT1/UGA/3/99
SAT2/SEN/7/83
SAT2/PAT/1/12
SAT2/EGY/9/12
SAT2/SAU/6/00
A/ERI/3/98
SAT2/ERI/12/89
SAT2/UGA/2/02
SAT2/UGA/MBF/4/02
SAT1/UGA/1/97
SAT1/NIG/5/81
SAT1/SUD/3/76
SAT1/NIG/8/76
SAT1/NIG/6/76
SAT1/NIG/15/75
SAT2/RWA/2/01
SAT2/KEN/11/60
SAT2/KEN/3/57
A/SEN/10/97
A/CIV/4/95
A/ETH/7/92
O/UGA/6/76
O/UGA/17/98
A/ETH/2/79
O/UGA/1/75
O/ETH/3/96
A/SOM/1/78
O/TAN/3/96
O/UGA/7/03
O/SUD/4/80
A/TAN/4/80
O/UGA/5/96
O/KEN/10/95
Cluster I
Cluster III
Cluster II
Southern Africa SAT1 and SAT2 and all SAT3 FMDV isolates
Northern Africa SAT1, SAT2, A and O FMDV isolates
0.2

representing the most probable nucleotide for each position. Each nucleotide was represented at least twice in the sequencing data. Sequences of the ca. 3.6 kb 5′UTR-Leader-P1-2A and 4.3 kb P2-P3-coding regions and 3′UTR were compiled and edited using BioEdit 5.0.9 software (Hall, 1999). The nucleotide data for eight non-structural protein coding regions (Leader, 2A, 2B, 2C, 3A, 3B, 3C, 3D), were determined in this study, while the nucleotide data for the P1 region have previously been described (Maree et al., 2011).

The nucleotide and deduced amino acid sequences were aligned using ClustalX (Thompson et al., 1997). Phylogenetic relationships were determined using Maximum-Likelihood methods conducted in MEGA5 (Tamura et al., 2011). The jModelTest 2.1.3 (Darriba et al., 2012) was used to predict the most appropriate model of evolution. The General Time Reversible model with gamma distribution and invariable rates (GIR + I + G) best described the pattern of nucleotide substitution. MEGA5 software was also used to identify hypervariable amino acid regions in a total alignment of the deduced amino acid sequences, defined as regions with more than 60% variable residues within a window of 10 residues. Entropy plots were drawn from the deduced amino acid alignments using BioEdit 5.0.9 software (Hall, 1999) and were defined as the uncertainty at each amino acid position with high values being an indication of high variation (Schneider and Stephens, 1990). The relative hydrophobicity of the individual proteins were predicted using the Kyte and Doolittle (1982) method operated in the BioEdit 5.0.9 software (Hall, 1999). Potential recombination within the aligned open reading frame sequences of the African FMDV were examined using the GENECONV (Sawyer, 1989) recombination detection method of the Recombination Detection Program v4.43 (Martin et al., 2005).

2.3. Structural modelling

Homology models of the 3C protease ($3C^{pro}$) and 3D RNA-polymerase ($3D^{pol}$) of SAT1 and SAT2 were built using Modeller 9v3 (Sali and Blundell, 1993) with FMDV A10 $3C^{pro}$ (pdb id: 2j92) or type C RNA-polymerase (1UO9) as templates. Alignments were performed with ClustalX. Structures were visualized and the surface-exposed residues identified with PyMol v1.1 (Schrödinger, LLC, New York, NY).

2.4. Nucleotide sequence accession numbers

All nucleotide sequences determined in this study have been submitted to GenBank under the accession numbers indicated in Table 1.

3. Results

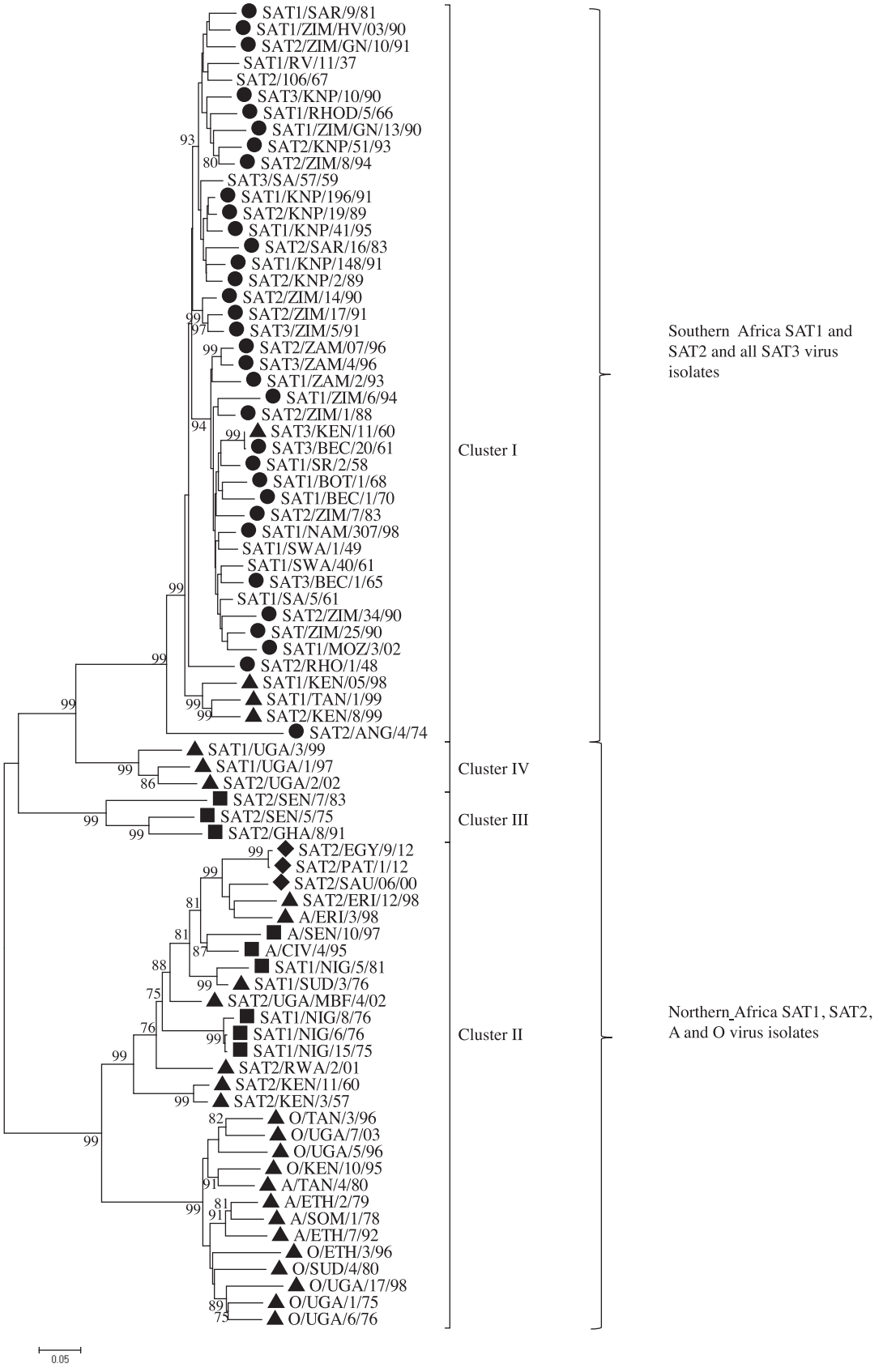
3.1. Phylogenetic trees of the various non-structural protein coding regions

For the 79 sub-Saharan African FMDV from 5 serotypes (SAT1, 2, 3, A and O) studied, 606, 1416 and 2721 nucleotide positions were aligned for the L^{pro}-, P2- and P3-coding regions, respectively. The L^{pro}-coding region displayed greater nucleotide variability i.e. 66.2% as compared to 46.7% and 47.9%, observed for the P2- and P3-coding regions, respectively.

Midpoint rooted neighbour joining trees were used to determine the phylogenetic relationships of the L^{pro}-coding region and the P3-coding regions for the sub-Saharan African FMDV. Also included in the analysis were sequences of two recent SAT2 viruses from the 2012 FMD outbreak in Egypt and the Middle East (Valdazo-González et al., 2012) obtained from GenBank (JX014255 and JX014256).

Based on a cut-off value of less than 16% nucleotide difference, three non-serotype specific clusters were observed for the L^{pro}-coding region (Fig. 1) and four clusters for the P3-coding regions (Fig. 2), supported by strong bootstrap values. Based on phylogeny of the L^{pro}- and P3-coding regions, the SAT1 and SAT2 viruses from southern Africa i.e. Angola, Botswana, Malawi, Mozambique, Namibia, Zambia, Zimbabwe and South Africa and the SAT3 viruses (from southern and East Africa), as well as two isolates from Kenya (SAT1/KEN/5/98 and SAT2/KEN/8/99) and one from Tanzania (SAT1/TAN/1/99), grouped together (i.e. cluster I; Figs. 1 and 2). The A and O types of viruses that originated from Côte d'Ivoire, Senegal (western Africa), Eritrea, Ethiopia, Kenya, Somalia, Sudan, Tanzania and Uganda (eastern Africa), grouped together with SAT1 and SAT2 viruses from Eritrea, Nigeria, Rwanda, Sudan and Saudi Arabia (i.e. cluster II; Figs. 1 and 2). Three SAT2 viruses, two from Kenya (KEN/3/57 and KEN/11/60), and one from Uganda (UGA/MBF/4/02) were also found in cluster II, based on L^{pro} and P3 phylogeny (Figs. 1 and 2). It is interesting to note that three SAT2 viruses from Senegal and Ghana in West Africa (SAT2/SEN/5/75, SAT2/SEN/7/83, and SAT2/GHA/8/91) and a SAT1 virus from Uganda (SAT1/UGA/3/99) formed a strongly supported cluster based on L^{pro}-sequence phylogeny with 20.4% nucleotide differences within the group (cluster III; Fig. 1). Based on the P3-coding region, these West African SAT2 viruses shared a separate cluster III (Fig. 2), while three Ugandan isolates grouped together (SAT1/UGA/1/97, SAT1/UGA/3/99 and SAT2/UGA/2/02) (cluster IV; Fig. 2). Two virus strains from the recent SAT2 FMD outbreak in Egypt and the Palestinian Autonomous Territories (PAT) (EGY/9/12 and SAT2/PAT/1/12) grouped in cluster II, demonstrating a close genetic

Fig. 1. Maximum-Likelihood tree depicting nucleotide relationships between the sub-Saharan African FMDV (SAT1, SAT2, SAT3, A and O) for the Leader-coding region rooted against the mid-point was constructed using MEGA5.2 (Tamura et al., 2011). The sub-Saharan southern African viruses are indicated by (●) eastern Africa viruses (▲), western Africa viruses (■), and the North African and Middle East viruses (◆). Model assumptions predicted by jModel 2.1.3 (Darriba et al., 2012) are based on the GIR + I + G (General Time Reversible) model of nucleotide substitution. There was non-uniform evolutionary rates among sites modelled using both discrete Gamma distribution (+G) with 4 rate categories and assuming that a certain fraction of sites are evolutionarily invariable (+I). Proportion of invariable sites is 0.2830, gamma shape is 0.795. The scale bar indicates nucleotide substitutions per site. The robustness of the tree topology was assessed using 1000 bootstrap replications.


SAT1/SAR/9/81
SAT1/ZIM/HV/03/90
SAT2/ZIM/GN/10/91
SAT1/RV/11/37
SAT2/106/67
SAT3/KNP/10/90
SAT1/RHOD/5/66
SAT1/ZIM/GN/13/90
SAT2/KNP/51/93
SAT2/ZIM/8/94
SAT3/SA/57/59
SAT1/KNP/196/91
SAT2/KNP/19/89
SAT1/KNP/41/95
SAT2/SAR/16/83
SAT1/KNP/148/91
SAT2/KNP/2/89
SAT2/ZIM/14/90
SAT2/ZIM/17/91
SAT3/ZIM/5/91
SAT2/ZAM/07/96
SAT3/ZAM/4/96
SAT1/ZAM/2/93
SAT1/ZIM/6/94
SAT2/ZIM/1/88
SAT3/KEN/11/60
SAT3/BEC/20/61
SAT1/SR/2/58
SAT1/BOT/1/68
SAT1/BEC/1/70
SAT2/ZIM/7/83
SAT1/NAM/307/98
SAT1/SWA/1/49
SAT1/SWA/40/61
SAT3/BEC/1/65
SAT1/SA/5/61
SAT2/ZIM/34/90
SAT/ZIM/25/90
SAT1/MOZ/3/02
SAT2/RHO/1/48
SAT1/KEN/05/98
SAT1/TAN/1/99
SAT2/KEN/8/99
SAT2/ANG/4/74
SAT1/UGA/3/99
SAT1/UGA/1/97
SAT2/UGA/2/02
SAT2/SEN/7/83
SAT2/SEN/5/75
SAT2/GHA/8/91
SAT2/EGY/9/12
SAT2/PAT/1/12
SAT2/SAU/06/00
SAT2/ERI/12/98
A/ERI/3/98
A/SEN/10/97
A/CIV/4/95
SAT1/NIG/5/81
SAT1/SUD/3/76
SAT2/UGA/MBF/4/02
SAT1/NIG/8/76
SAT1/NIG/6/76
SAT1/NIG/15/75
SAT2/RWA/2/01
SAT2/KEN/11/60
SAT2/KEN/3/57
O/TAN/3/96
O/UGA/7/03
O/UGA/5/96
O/KEN/10/95
A/TAN/4/80
A/ETH/2/79
A/SOM/1/78
A/ETH/7/92
O/ETH/3/96
O/SUD/4/80
O/UGA/17/98
O/UGA/1/75
O/UGA/6/76
Cluster I
Cluster IV
Cluster III
Cluster II
Southern Africa SAT1 and SAT2 and all SAT3 virus isolates
Northern Africa SAT1, SAT2, A and O virus isolates
0.05

relationship with SAT2 viruses from Saudi Arabia and Eritrea (Figs. 1 and 2).

When using the L^{pro} and P3 phylogenies to investigate whether or not intertypic recombination was prevalent between the African FMDV, potential recombination events were noted in the non-structural protein coding sequences at nucleotide positions 770–818 and 3675–3944. Interestingly, when only the SAT viruses were analyzed, a high rate of intertypic recombination events were also observed at nucleotide positions 2856–2962.

3.2. The untranslated regions (UTR's)

The analysis of the 5′UTR included an alignment of the L-fragment from the 79 African isolates in this study. The 5′UTR L-fragment ranged from 672 to 760 nucleotides in length and showed 31% invariant nucleotides. Insertions/deletions were observed, including an 83 nucleotide insertion which was 11 nucleotides downstream of the poly (C) tract in SAT1/ZIM/25/90. However, this was an exception where mostly an insertion of 40 nucleotides was present at this position for the southern African isolates ($n = 25$). When the 5′UTR L-fragment of only the SAT viruses was compared, 47% of the nucleotide positions was found to be invariable. In pairwise comparisons, SAT viruses display higher than 86% nucleotide identity between any two isolates, with most variation observed between isolates from southern and West Africa. This suggests that although the FMDV 5′UTR may tolerate nucleotide changes the overall conservation is retained, likely for the maintenance of functional secondary structures. SAT-specific conserved motifs were observed in the IRES, especially nucleotides 615–686 and 714–762 downstream of the poly (C) tract.

The African FMDV 3′UTR was variable in length, ranging from 91 to 101 nucleotides, with only 33% invariant nucleotide positions. SAT-specific conserved motifs were present between residues 15–29 and 35–66, based on geography, being incongruent for viruses originated from southern, West and East Africa.

3.3. Amino acid variation in the non-structural proteins of sub-Saharan FMDV

The nucleotide sequences for the non-structural protein-coding regions were translated and the deduced amino acid sequences aligned.

3.3.1. Leader protease

The L^{pro}, 199–202 residues in length, was the most variable of the non-structural proteins with 61.4% variable amino acid positions in a complete alignment of the 79 sub-Saharan FMDV sequences (Table 3). All the SAT sequences and one serotype A virus (ERI/3/98) contained a three amino acid deletion between residues 24 and 28. At least 35.6% of the positions in viruses from southern Africa [Angola, Botswana, Mozambique, Namibia, Tanzania, Zambia, Zimbabwe, and South Africa (i.e. cluster I; Fig. 1)], were variable (Fig. 3). Similarly 38.6% of the positions in the SAT viruses from western and eastern Africa i.e. Egypt, Eritrea, Ghana, Nigeria, Senegal, Sudan, Rwanda and Uganda (clusters II and III combined), demonstrated variability (Table 3). The amino acid residues between the methionine's of the Lab and Lb isoforms of L^{pro} was highly variable (Fig. 3) and contained a two or three amino acid insertion in the SAT viruses. The residues required for L^{pro} catalytic activity, namely C52, H149 and D165 (Guarné et al., 1998), were conserved in all the isolates. Another residue, E77, suggested to be involved in autocatalysis, showed a substitution of E77K for SAT2/RWA/2/01. Among the non-structural proteins, entropy was highest for the L^{pro}, where 29 residue positions had entropy values of 1.0–1.9; especially towards the N-terminus (residues 4–24).

3.3.2. Peptides encoded by the P2-coding region

The 2A of 18 amino acids in length (Table 4) had 56% (10 of 18 residues) conserved residues for SAT viruses, irrespective of the geographic location. At least 83% of residues (15 of 18 residues) in a complete alignment of the 2A sequences of African A and O viruses were identical. The C-terminal 13 amino acids of 2A, with the sequence LLKLAGDVESNPG, were highly conserved for all the 79 African viruses. However, a residue substitution, D12N, in the conserved ^{12}DVEXNPG^{18} motif (Carrillo et al., 2005; Ryan et al., 2004) was observed in two SAT1 viruses from Uganda and Nigeria (SAT1/UGA/1/97 and SAT1/NIG/6/76). The N-terminal residue position 2 of the 2A protein showed most variation with an entropy value of 1.1 (Table 4).

The 2B protein (154 amino acids) contained no insertions or deletions and showed 69.5% conserved residues in a complete alignment of all the sub-Saharan FMDV sequences included in this study (Table 3; Fig. 3). Most of the variation was found in two hypervariable domains in the N-terminal half of the protein, i.e. residues 5–29 and 44–53 (Fig. 3), while the sequence ^{64}IKLLSRLSC-MAAVAAR(S/A)KDPVLVAIMLADTGLEILDSTFVVKKI^{107} was highly conserved among all the sub-Saharan sequences. Another conserved motif, located between residues 115 and 137 (FHVPAPVFSFGAPILLAGLVKVA) contained a hydrophobic domain (Fig. 3). Entropy values of 1.0–1.5 were observed at four residue positions (20, 23, 45 and 52) at the N-terminus of the protein.

The FMDV 2C protein is an AAA+ ATPase with RNA binding activity (Sweeney et al., 2010), and was found to be

Fig. 2. A Maximum-Likelihood tree showing nucleotide relationships between the sub-Saharan African FMDV (SAT1, SAT2, SAT3, A and O serotypes) for the P3-coding region. The phylogenetic tree was rooted against the mid-point and was constructed in MEGA5.2 software (Tamura et al., 2011). The sub-Saharan southern African viruses are marked (●) eastern Africa viruses (▲), western Africa viruses (■), and the North African and Middle East viruses (◆). The most suitable model for nucleotide pattern substitution was determined by jModel 2.1.3 (Darriba et al., 2012). It was predicted the GIR + I + G (General Time Reversible) model of nucleotide substitution. There was non-uniform evolutionary rates among sites modelled using both discrete Gamma distribution (+G) with 4 rate categories and assuming that a certain fraction of sites are evolutionarily invariable (+I). Proportion of invariable sites is 0.428, gamma shape is 0.706. The scale bar indicates nucleotide substitutions per site. Confidence levels of the tree branches were tested using 1000 bootstrap replications.

Table 3
Variability in the deduced amino acid alignments of the non-structural proteins for sub-Saharan FMDV.

Region in Genome	No. of residue positions aligned	Complete alignment		Southern Africa SAT NSP alignment		A, O, western and eastern Africa SAT1 and SAT2 NSP alignment	
		No. of variant residues	% of variant residues	No. of variant residues	% of variant residues	No. of variant residues	% of variant residues
L^{pro}	202	124	61.4	67	33.2	77	38.1
2A	18	8	44.4	5	27.8	7	38.9
2B	154	47	30.5	29	18.8	28	18.2
2C	318	110	34.6	75	23.6	67	21.0
3A	153	86	56.0	51	33.4	62	40.5
$3B_{123}$	71	38	53.5	22	31.0	28	39.4
$3C^{pro}$	213	89	41.7	42	19.7	51	23.9
$3D^{pol}$	470	136	29.0	67	14.3	95	20.2

318 amino acids in length; mostly hydrophilic towards the C-terminus and contained 65.4% invariant residue positions when sequences of isolates in this study were compared (Table 3; Fig. 3). Most variation was focussed in two hypervariable regions, one in the N-terminus (residue 35–101) and one in the C-terminus (258–306) of the protein (Fig. 3). Entropy values of 1.0–1.2 occurred at three residue positions (83, 92, and 291). A highly conserved, hydrophobic motif was present between residues 17–34, i.e. ^{17}EWLVKLILAIRDWIKAWI^{34}. This amphipathic helix is most probably involved in the attachment of 2C to the membrane (Echeverri and Dasgupta, 1995). Residues 136–152, i.e. ^{136}DSVWYCPPDPDHFDGYN^{152} were 100% conserved in all African isolates. Conserved residues at positions 110–117 (GKSGQGKS), 156–161 (VVVMDD) and 201–207 (VIIATTN) included the characteristic Walker A, Walker B and C motifs of an AAA+ ATPase (Sweeney et al., 2010). A D160N substitution in the Walker B motif has been observed once in SAT3/ZAM/4/96. However, the C motif for southern Africa SAT viruses displayed a conservative T206S ($n = 68$) substitution while I203 V ($n = 2$) and I202L ($n = 1$) substitutions were also observed (Fig. 3). There were residues unique to the SAT1, SAT2 and SAT3 viruses originating from the larger southern African region (Angola, Botswana, Mozambique Namibia, South Africa, Tanzania, Zambia and Zimbabwe). These sequences were located within the N-terminal 100 amino acids of 2C, i.e. residue positions 40–42 (YIS) in the southern African SATs ($n = 34$) as compared to FVT in SATs of the larger northern Africa [Sudan, Ethiopia, Eritrea, Nigeria, Senegal and Ghana ($n = 11$)]. Other positions of unique substitutions for the southern SATs are G49R, K53C, D56N, A66S, D71E, Q75E and V92N (Fig. C; supplementary data).

3.3.3. *Peptides encoded by the P3-coding region*

The 3A protein is 152–153 amino acids in length. Viral genomes of the all southern and the Ugandan SAT serotypes (clusters I and III; Fig. 2) generally encoded a 3A protein one residue shorter compared to the SAT viruses from West Africa and all the African A and O serotypes in cluster II (Fig. 2) with a deletion at amino acid position 148 (Fig. D; supplementary data). An exception is the 3A protein of the SAT2 virus, SEN/7/83 that contained a deletion of 11 amino acids in the C-terminus, between residues 134 and 146. In the total alignment, at least 56% of residue positions were variable, while 34.6% were variable for the SATs from the southern part of Africa (Angola, Botswana, Mozambique Namibia, South Africa, Tanzania, Zambia and Zimbabwe) and 45.7% for the West and East Africa SAT1, SAT2, A and O viruses (Ethiopia, Eritrea, Ghana, Kenya, Nigeria, Senegal, Tanzania, Rwanda and Uganda). The sequence ^{115}EVVDKP^{120} in southern African SAT viruses (cluster I; Fig. 2) distinguished them from the northern African lineages (cluster II; Fig. 2). The N-terminal 41 amino acids were relatively conserved (Fig. 3) while a hydrophobic domain, with mean hydrophobicity >3.0 (Kyte and Doolittle, 1982), was located between residues 60–74 (^{60}EIVALVVVLLANIII^{74}). The C-terminus was highly variable (Fig. 3), with nine residue positions that showed entropy values of 1.0–1.3.

The three copies of 3B varied in length between 23 ($3B_1$) and 24 ($3B_2$ and $3B_3$) residues demonstrating 53.5% overall variability, whilst each copy varied by 56.5%, 50% and 54.2%, respectively (Table 5). The 3B proteins of the southern African SAT1, SAT2 and SAT3 viruses ($n = 59$) were more similar with 31% variation, while SAT1 and SAT2 viruses ($n = 9$) from Nigeria, Senegal, Ghana, Ethiopia and Eritrea showed 39.4% variability (Table 3). Nine of the eleven SAT1 and SAT2 viruses in cluster II (Fig. 2) from western (Nigeria) and eastern Africa (Ethiopia, Eritrea, Kenya and Uganda) had residues in common with serotype A and O viruses (Table 5). The N-terminal motif, GPYXGP (where X is any amino acid), was conserved for all the sub-Saharan viruses.

The $3C^{pro}$-coding region translated into 213 amino acids with 41.7% variable positions in an overall alignment (Fig. E; supplementary data). Forty-two variable residues (19.7%) were observed in the southern African SAT viruses (Table 3), 35 (16.4%) in the East Africa (Kenya, Uganda, Tanzania and Rwanda) SAT1, SAT2, A and O viruses, 5 variable residues (2.3%) in the West Africa (Ghana and Senegal) SAT2 isolates and only 2 variable residues (<1%) was found among the four Ugandan SAT1 and SAT2 isolates. The variation was not random, but defined in hyper-variable regions separated by highly conserved residues. The conserved residues point towards the significant contribution of these residues to structural and/or functional constraints. The N-terminal 60 amino acids of $3C^{pro}$, especially 17–48 [^{17}(K/R)PVELILDGK(T/I)VA(L/I)CCATGVFGTAYLVPRHLF^{48}], were highly conserved

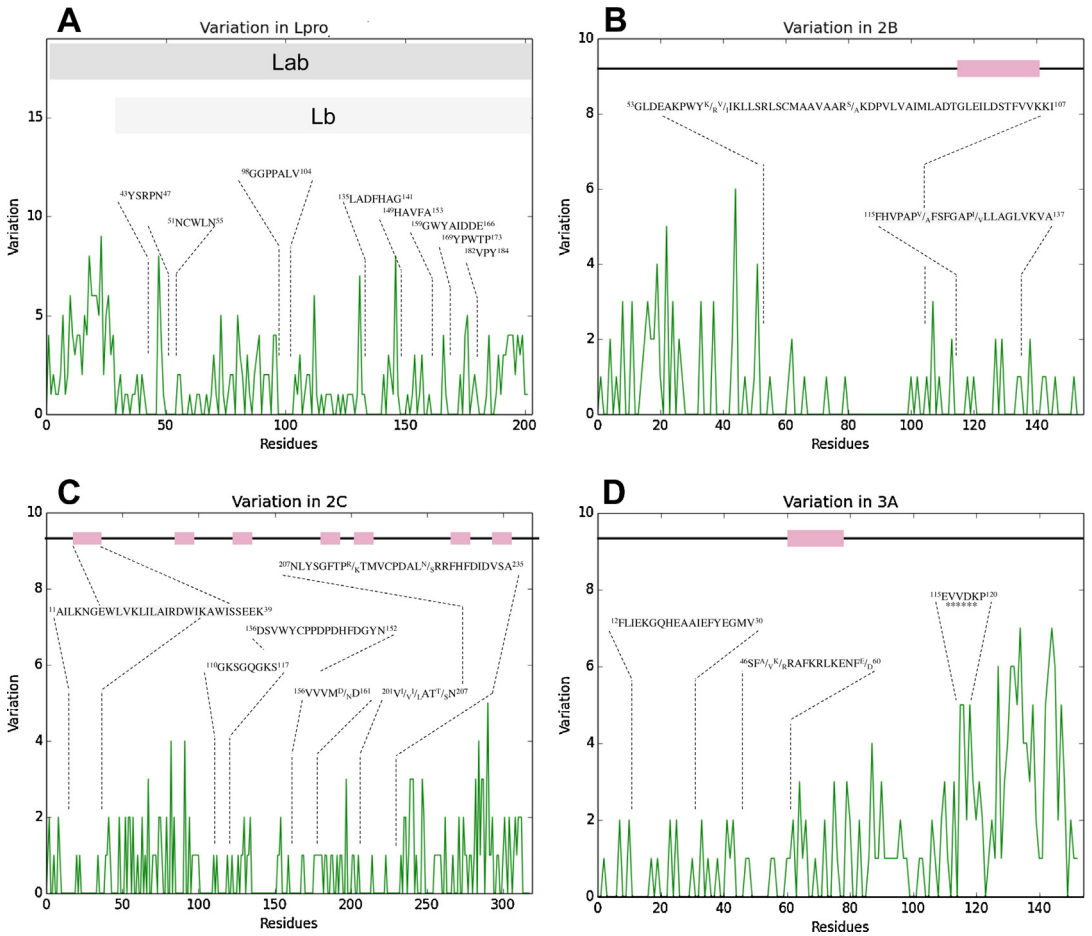


Fig. 3. Graphs representing the amino acid variation of the two most variable non-structural proteins, L^{pro} (A) and 3A (D) and two of the more conserved non-structural proteins, 2B (B) and 2C (C). Regions of hypervariability in a complete alignment were defined as sites which had 5 or more variable aa residues within a window of 10 residues. The pink bars at the top of the graph represent hydrophobic domains and is fitted in the correct position based on the amino acid numbering on the horizontal axis. The amino acid sequence of conserved domains are indicated.

Table 4
Variation in 2A in a complete alignment of sub-Saharan African FMDV isolates.

2A	Deduced amino acid sequence Residue position[a]																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
SAT types	L_{54} M_4 T_3	C_{47} A_6 L_5 F_1 G_1 S_1	N_{60} S_1	F_{59} Y_1 C_1	D_{57} E_2 A_2	L_{61}	L_{61}	K_{60} M_1	L_{60} Q_1	A_{61}	G_{61}	D_{59} N_2	V_{61}	E_{61}	S_{61}	**N**$_{61}$	P_{61}	G_{61}
O and A- types	*L_{14}*	*L_{16}*	*N_{16}*	*F_{14}*	*D_{16}*	*L_{16}*	*L_{16}*	*K_{16}*	*L_{16}*	*A_{16}*	*G_{16}*	*D_{16}*	*V_{16}*	*E_{16}*	*S_{14}*	*N_{16}*	*P_{16}*	*G_{16}*
	S_1 *T_1*			*Y_2*											*L_1* *P_1*			

[a] The individual columns represent the deduced amino acid alignment for the 2A for 79 sub-Saharan virus isolates. The consensus sequence for SAT types is given in the 1st row (in normal text) while the consensus for the O and A-types is indicated in the bottom row (in italics). Alternative amino acids occurring in the same position are indicated in the subsequent rows below each consensus. The numerals in the subscripts indicate the number of viruses with that particular amino acid.

Table 5
A complete alignment of the $3B_{123}$ (VPg) protein for the sub-Saharan African FMDV isolates.

	Residue position Deduced amino acid sequence[a]																						
$3B_1$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
SAT types	**G_{59}**	**P_{59}**	**Y_{59}**	A_{57} V_1 T_1	**G_{59}**	**P_{59}**	L_{48} M_8 V_2 F_1	E_{58} D_1	R_{59}	Q_{58} H_1	Q_{48} K_9 R_1 T_1	P_{59}	L_{59}	K_{59}	L_{49} V_{10}	K_{56} R_2 Q_1	A_{48} T_{10} V_1	K_{57} R_2	L_{58} P_1	P_{59}	Q_{27} K_{22} R_5 L_2	A_{48} Q_{11}	E_{59}
O and A types	**G_{16}**	**P_{16}**	**Y_{16}**	*A_{10}* *S_2* *T_2* *V_2*	**G_{16}**	**P_{16}**	*L_{12}* *M_3* *F_1*	*E_{16}*	*R_{16}*	*Q_{16}*	*K_{15}* *Q_1*	*P_{16}*	*L_{16}*	*K_{15}* *R_1*	*V_{16}*	*K_{11}* *R_5*	*A_{14}*	*K_{15}* *R_1*	*L_{15}* *P_1*	*P_{16}*	*Q_{14}* *L_2*	*Q_{16}*	*E_{16}*
$3B_2$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
SAT types	**G_{59}**	**P_{59}**	**Y_{59}**	A_{57} V_2	**G_{58}** R_1	**P_{59}**	L_{49} M_8 T_1 V_1	E_{59}	K_{37} R_{22}	Q_{59}	Q_{58} K_1	P_{59}	L_{59}	K_{57} R_1 E_1	L_{49} V_9 F_1	K_{55} R_4	A_{45} T_{10} V_2 S_1 E_1	K_{46} R_{14}	L_{49} A_5 P_4 G_1	P_{59}	V_{59}	A_{49} V_{10}	K_{59}	E_{59}
O and A types	**G_{16}**	**P_{16}**	**Y_{16}**	*A_{16}*	**G_{16}**	**P_{16}**	*M_{16}*	*E_{16}*	*R_{15}* *K_1*	*Q_{16}*	*K_{12}* *Q_4*	*P_{16}*	*L_{16}*	*K_{16}* *R_2*	*V_{16}*	*K_{13}* *R_2* *T_1*	*A_9* *V_7*	*K_{15}* *R_1*	*A_{14}* *T_1* *V_1*	*P_{16}*	*V_{16}*	*V_{14}* *A_1* *T_1*	*K_{16}*	*E_{16}*
$3B_3$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	24	24
SAT types	**G_{59}**	**P_{59}**	**Y_{59}**	E_{58} D_1	**G_{59}**	**P_{59}**	V_{58} L_1	K_{59}	K_{59}	P_{59}	V_{56} A_3	A_{59}	L_{59}	K_{59}	V_{57} E_1 L_1	K_{58} R_1	A_{55} T_4	K_{53} R_6	A_{41} N_{10} S_4 P_3 T_1	P_{43} M_9 A_3 L_2 V_2	I_{59}	V_{57} I_2	T_{59}	E_{59}
O and A types	**G_{16}**	**P_{16}**	**Y_{16}**	*E_{16}*	**G_{16}**	**P_{16}**	*V_{16}* *L_1*	*K_{16}*	*K_{16}*	*P_{16}*	*V_{15}* *A_1*	*A_{16}*	*L_{16}*	*K_{16}*	*V_{16}*	*K_{16}*	*A_{16}*	*K_{16}*	*N_{16}*	*L_{13}* *M_3*	*I_{16}*	*V_{16}* *I_2* *L_1*	*T_{16}*	*E_{16}*

[a] The individual columns represent the deduced amino acid alignment. For each copy of the 3B peptide, the consensus amino acids in the alignment are given in the top row for the SAT types (in normal print) and the complete alignments of A and O 3B sequences (in Italics). Amino acids substitutions that occurred are indicated in the subsequent rows, and the numerals in subscripts indicating the frequency of occurrence of a particular amino acid. The conserved GPYXGP motif (where X is any amino acid) is indicated in bold script.

and constituted a hydrophobic domain ^{27}(T/I)VA(L/I)CCATGVFGTAYLVP^{44}. Other conserved domains included ^{66}D(F/Y)RVFEFE(V/I)KVKGQ(D/E)M(L/M)SDAAL(M/R)(V/I)L(H/N)^{91}, ^{129}FSG(D/E)ALTYKD(L/V/I)VVCMDGDTMPGLFAY(R/K)A^{156} and ^{161}GYCG^{164}. Residue substitutions found in these regions were mostly conservative mutations. The active triad of $3C^{pro}$, consisting of residues H46, D84 and C163 (Birtley et al., 2005), showed complete conservation except in SAT3/ZIM/5/91 where D84Y substitution was observed. Interestingly, the sequence ^{107}MKLSKGS^{113} in southern African SAT viruses (cluster I; Fig. 2) distinguished them from the northern African lineages (cluster II; Fig. 2) that had the sequence ^{107}ARM(R/K)KGT^{113}. Notably, the four viruses from the Ugandan lineage (cluster IV; Fig. 2) displayed ^{107}VRVAKGN^{113} and ^{169}TKSGSQ^{174} potentially specific to this lineage (Fig. E; supplementary data).

The $3D^{pol}$ was found to be the most conserved of all the non-structural proteins. The 470 amino acid peptide, the longest of the non-structural proteins, demonstrated 29% variable residues (Table 3). The variation was 25% when only the three SAT serotypes ($n = 67$) were considered. The results presented here suggest an average of 15%, 18.5%, 3.2%, 9.8% and 10.2% of variable residues for $3D^{pol}$ of SAT1, SAT2, SAT3 and serotypes A and O, respectively. The $3D^{pol}$ variation was not limited to certain areas as seen for $3C^{pro}$. Previously five conserved motifs were described for $3D^{pol}$ of FMDV (Doherty et al., 1999; Ferrer-Orta et al., 2004). The ^{240}DYSAFD^{245}, ^{297}PSG^{299}, ^{336}YGDD^{339} and ^{385}FLKR^{388} motifs were conserved in all the sub-Saharan FMDV sequences included here (Fig. F; supplementary data). However, the ^{164}KDELR^{168} motif was present in the A, O and SAT sequences either as KDEIR or KDEVR. Although the sequences analyzed here showed some variability in the three hypervariable regions (residue positions 1–12, 64–76, and 143–153) that were previously described for $3D^{pol}$ (George et al., 2001), it was mostly one or two possible residue substitutions occurring at 14 out of the 23 variable residue positions among the sequences. The $3D^{pol}$ hypervariable region, between residues 143–154, showed the highest entropy of 1.15 as a result of 4 possible residues at position 145. All other entropy was less than 0.64.

3.4. Structural implications of the non-structural protein variation

Crystallography studies suggest FMDV 3 C^{pro} typically consists of two β-barrel domains, each composed by eight anti-parallel β-sheets. Upon alignment of the 79 sequences in this study, considerable variation was observed in the amino acid sequences of the β-sheets. However, these

were mostly conservative mutations (Fig. 4). The substrate (S1) pocket include residues from the βC_2-βD_2 loop (residues 154–160) and the C-terminal end of the βE_2 strand (residues 181–186) and were conserved in this study. The sequence of the S1 pocket, that interact with the P1 amino acid side chain of the substrate, are ^{154}Y(R/K)A(G/A)TK(V/A)^{160} and ^{181}HSAGGN^{186}. The amino acid variation in the African FMDV 3C^{pro} was almost entirely surface

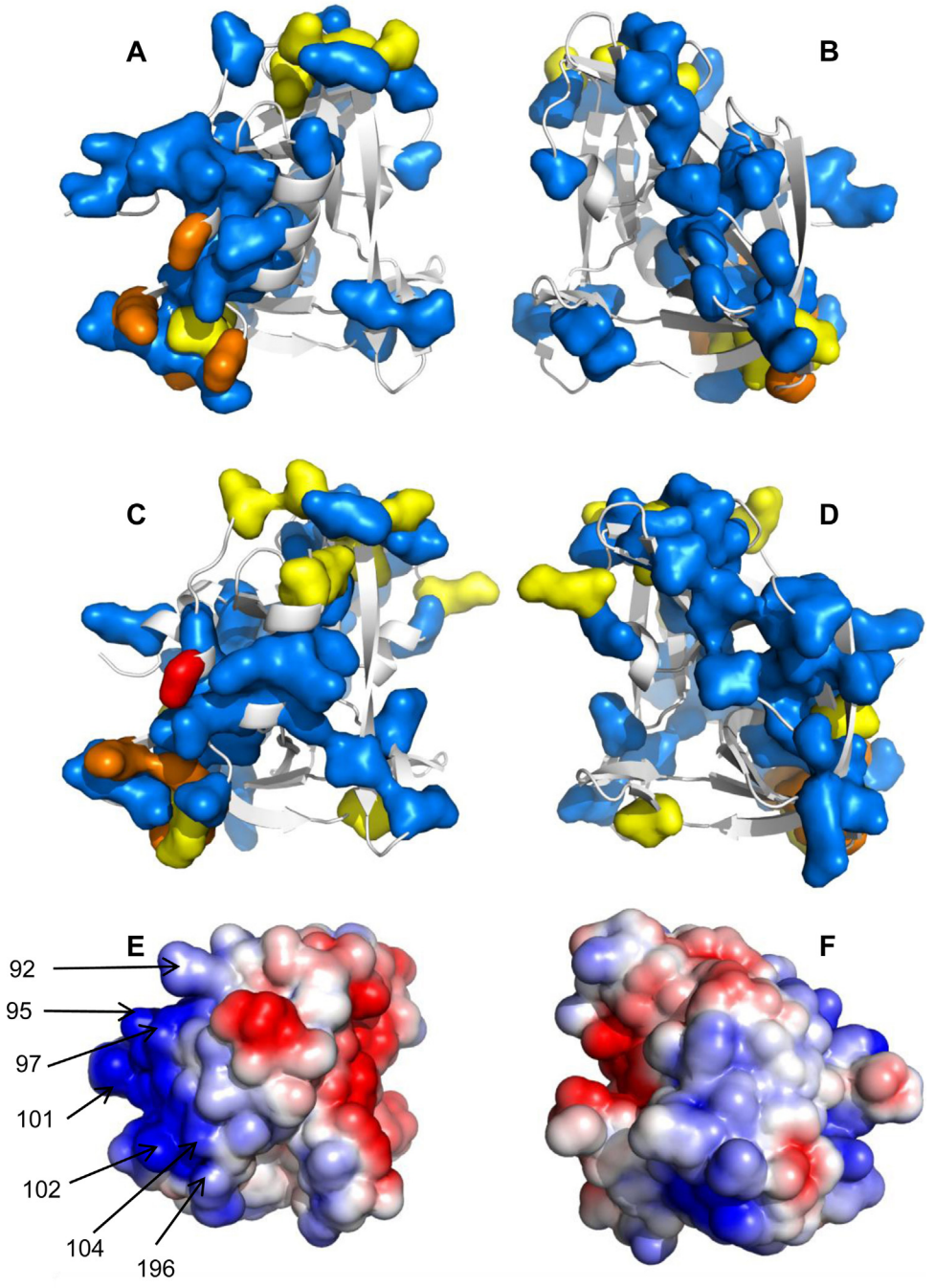


Fig. 4. Structural representation of FMDV 3C^{pro}. The position of amino acid variation observed in a complete alignment of all African SAT1 (A and B) and SAT2 (C and D) viruses is mapped on the modelled structure. Two possible residues at a position are indicated in blue, three residues in yellow, four in orange and five in red. Finally, the surface electrostatic potential is presented for SAT1/SAR/9/81 3C^{pro} (E and F). The electrostatic potential was coloured with positive charge as blue and negative in red and the scale of colouring was kept constant. The electrostatic potential is conserved in viruses across the five serotypes from Africa. The residues involved in the positively charged region on the 3C^{pro} are indicated.

located and directed away from the substrate binding site (Fig. 4A–D).

Opposite the active site of the protease, exposed on the surface of the $3C^{pro}$ was a local region of basic amino acids of which conserved residues R95 and R97 were exposed on the surface (Fig. 4E and F). This positively charged motif was highly conserved for the African A, O and northern African SAT viruses as ^{95}RVRDI^{99}. The sequence for the southern Africa SAT viruses displayed a characteristic I99L substitution. One SAT1 virus, NIG/5/81, had a characteristic ^{95}RVSVI^{99} sequence (Fig. E; supplementary data). An interesting observation was that the conserved R92 (Nayak et al., 2006), although present in African A, O and northern African SAT viruses, was replaced by either an S or T in all the southern African SAT types. Similarly residue K101 was substituted with a G or A in the southern African SAT viruses (Fig. E; supplementary data).

The FMDV $3D^{pol}$ has a structure similar to that of other RDRP's. Analysis of the variation in relation to the structure indicated that, although conserved, $3D^{pol}$ was more tolerant to substitutions than previously found (Carrillo et al., 2005; Koonin, 1991). The variation seemed to be limited to the outer edges of the protein, was mostly conservative substitutions and seldom involved more than three possible substitutions in an amino acid position (Fig. 5).

4. Discussion

The FMDV genome exhibits a quasispecies nature like many RNA viruses as a result genetic drift due to error prone replication and recombination occurs (Domingo et al., 1992; Eigen, 1996; Heath et al., 2006; Holland and Domingo, 1998). Only limited numbers of non-deleterious mutations can occur in regions of functional conservation in the non-structural proteins like the 2A, 2B, 2C, 3A, $3C^{pro}$ and $3D^{pol}$ of FMDV (Tully and Fares, 2009). This preserves both the integrity of structure and function especially of the major enzymes, like the RDRP (Domingo et al., 1990, 1992). In this study we present the coding regions for the non-structural proteins of 79 FMDV isolates from sub-Saharan Africa (SAT1, 2, 3, A and O types).

The phylogenetic comparisons of nucleotides coding the Leader- and P3-regions both generated a separate cluster for the southern Africa SAT viruses and the northern Africa SAT, A and O viruses. Three SAT2 viruses from Senegal and Ghana and three isolates from Uganda did not conform to the phylogenetic pattern, forming two separate clusters based on the P3-coding region. The incongruent tree topologies observed for the Leader- and P3-coding regions are consistent with previous reports (King et al., 1985; Krebs and Marquardt, 1992; van Rensburg et al., 2002), implicating intertypic recombination in the non-structural coding regions while maintaining serotype-specific phylogeny when looking at the structural protein coding regions. Recombination is one of the putative mechanisms for rapid viral evolution and diversification (Simmonds and Welch, 2006; Jackson et al., 2007) and analysis of our extensive African FMDV dataset confirmed previous studies of largely non-recombining structural protein coding regions, but with frequent and complex recombination events in the non-structural protein coding regions.

We observed variation in the deduced amino acid sequence alignments for the eight FMDV non-structural proteins (L^{pro}, 2A, 2B, 2C, 3A, $3B_{123}$, $3C^{pro}$, $3D^{pol}$), that ranged from 61.4% for L^{pro}, 56% for 3A, 53.5% for $3B_{123}$, 50% for 2A, 41.7% for $3C^{pro}$, 34.6% for 2C, 30.5% for 2B and 29% for $3D^{pol}$.

The L^{pro}, which is the first protein to be synthesized during virus replication, cleaves itself from the rest of the growing polypeptide (Strebel and Beck, 1986; Belsham and Brangwyn, 1990) before cleaving the eukaryotic translation initiation factor eIF-4G (Piccone et al., 1995; Guarné et al., 1998). Despite the high variability observed (61.4%), residue conservation among the African FMDV was maintained amongst the essential auto-catalytic residues (C52, H149 and D165).

The 2A protein induces a modification of the cellular translation apparatus resulting in 2A release (Ryan et al., 1991; Donnelly et al., 2001). This is achieved by modifying the activity of the ribosome to promote hydrolysis of the peptidyl (2A)-tRNAGly ester linkage and the release of the P1-2A precursor in the translational complex (Donnelly et al., 2001). We observed only 9 out of the 18 amino acids as invariable in a complete alignment; however the functional domain of ^{12}DVEXNPG^{18} was conserved indicating structural and functional constraints associated with this domain. 2A is cleaved from the P1 polypeptide by the $3C^{pro}$ in the later stage of processing and its function as an independent protein is unknown.

The small, hydrophobic 2B protein of FMDV associates with the ER and may cause rearrangement of the ER membrane (O'Donnell et al., 2011; Moffat et al., 2005). The hydrophobic motif at residue positions 115–137 is likely to be responsible for positioning 2BC complexes to allow its membrane bound activities at sites of FMDV replication in the ER-derived vesicles in the host cytoplasm (Grubman and Baxt, 2004; Moffat et al., 2005, 2007). The hydrophobic character of this domain was highly conserved in the African viruses.

The FMDV 2C protein is an AAA+ ATPase that affects initiation of minus strand RNA synthesis (Sweeney et al., 2010), and localizes with Golgi-derived membrane structures. The three ATPase binding motifs were highly conserved among the African viruses. The interaction of 2C, 3A and a cellular poly (A)-binding protein with the RNA helicase A (RHA) leads to a ribonucleoprotein complex formation at the 5′ end of the genome and has been shown to play an important role in FMDV replication (Lawrence and Rieder, 2009). The 2C protein and its precursor, 2BC, induce vesicle formation in the cytoplasm (Moffat et al., 2005). Unique sequences in the 2C peptide for the southern Africa SAT viruses included residues 40–42 (FVT → YIS) and in the Walker C motif, T206 → S, I203 → V, I202 → L which are all conservative substitutions.

The 3A protein is proposed to be the membrane anchor for the picornavirus replication complex (Weber et al., 1996; Xiang et al., 1997). It is associated with viral-induced membrane vesicles and contributes to the cytopathic effect and the inhibition of protein secretion (Doedens and Kirkegaard, 1995; Wessels et al., 2006). The 3A peptide has

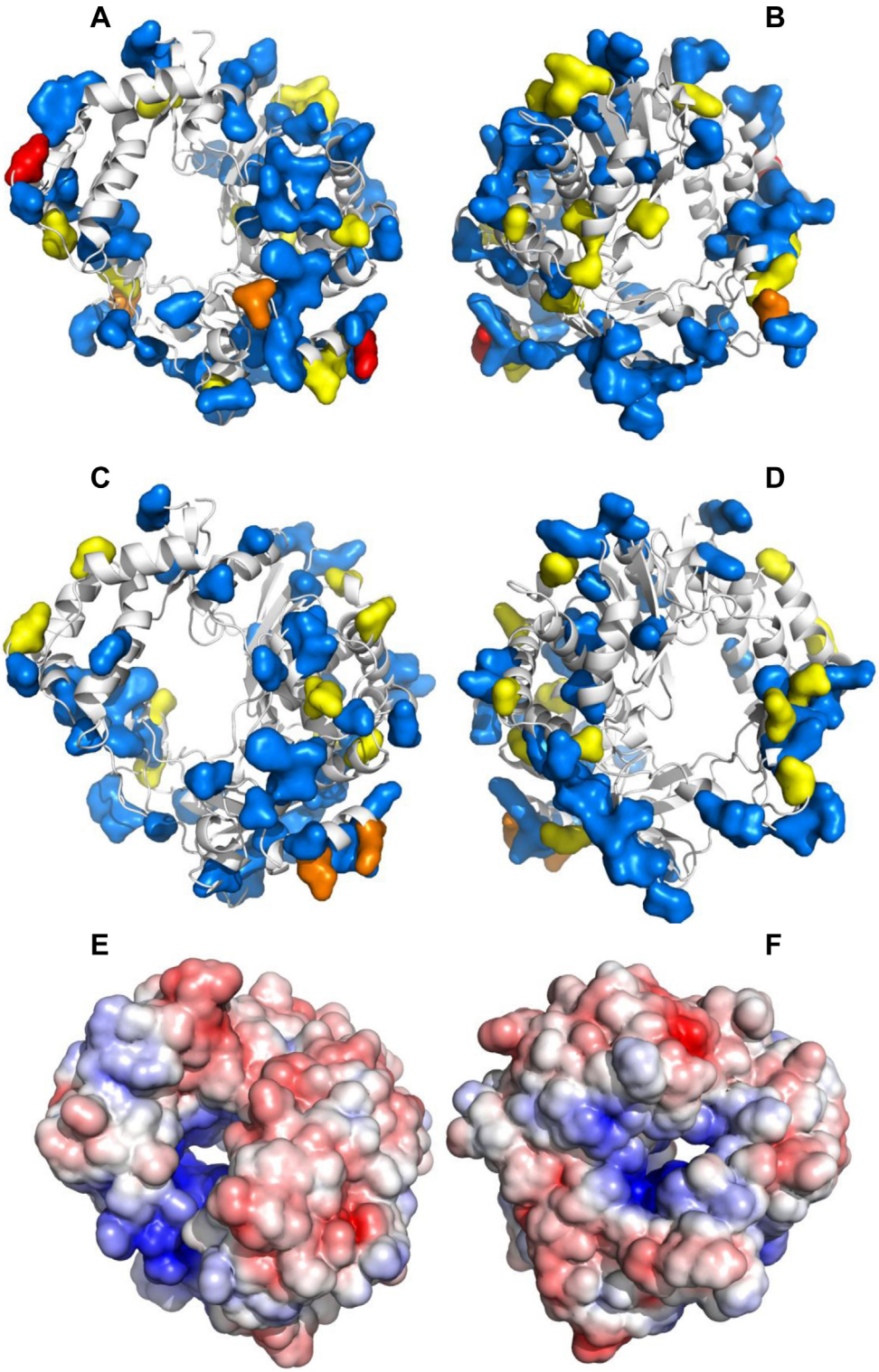


Fig. 5. Variation in the $3D^{pol}$ protein, observed in a complete alignment of African FMDV sequences, has been mapped to the three-dimensional structure of the protein. The electrostatic surface potential is shown with positive charge as blue and negative in red (A). The electrostatic potential is conserved in viruses across the five serotypes from Africa. The variable amino acid positions observed for SAT1 (B) and SAT2 (C) viruses were indicated with two possible residues at a position as blue, three residues as yellow, four as orange and five as red. The orientation of the protein was kept the same.

been previously associated with virulence in picornaviruses (Heinz and Vance, 1996; Lama et al., 1998; Pacheco et al., 2003), and deletions in this peptide at the C-terminus (133–143), collectively with other genome changes in Euro-South American and Asiatic FMDV, correlated to altered host range and slow growth in bovine derived cells but accelerated growth in porcine derived cells (Giraudo et al., 1990; Knowles et al., 2001; Pacheco et al., 2003). Interestingly, an 11 amino acid deletion was observed in the 3A C-terminal region (136–146) of only one virus, SAT2/SEN/7/83, which also did not reach titres of more than 3.0×10^5 pfu/ml after 10 passages in BHK-21 cells (results not shown). The 3A protein is, after the L^{pro} and VP1, one of the most variable proteins in the FMDV proteome.

Although all three copies of the 3B/VPg protein were present in the African isolates, they were highly variable in the complete alignments. However, the N-terminal motif ^{1}GPYXGP^6 was conserved in all the viruses. The VPg protein participates in the initiation of RNA replication and plays a role in the encapsidation of viral RNA (Hogle et al., 1985; Xiang et al., 1997; Barclay et al., 1998). Each of the VPg protein contains a 3Y, which is known to be involved in phosphodiester linkage to the viral RNA (Forss and Schaller, 1982) and was conserved in the African viruses in this investigation.

The $3C^{pro}$ is a cysteine protease (Birtley et al., 2005) responsible for catalyzing 10 of the 13 proteolytic cleavage events necessary for polyprotein processing (Vakharia et al., 1987; Clarke and Sangar, 1988). In the $3C^{pro}$, there were notably two changes to neutral residue substitutions in a conserved motif of a SAT1 virus NIG/5/81, i.e. residues R97 → S and D98 → V. However, there is conservation of the active triad for the $3C^{pro}$, residues H46, D84 and C163 as well as the substrate pocket. The pocket contains a H181 hydrogen bonded with Y154 and T158, which donate hydrogen bonds to the P1 peptide substrate (Birtley et al., 2005), all of which were invariable in the sub-Saharan African viruses. Studies further showed residues 138–150 of FMDV $3C^{pro}$ in a β-ribbon structure that overlies the substrate binding cleft (Sweeney et al., 2007). The C142 at the apex of this loop plays an important role in substrate binding and showed complete conservation in the $3C^{pro}$ amino acid comparison of the African isolates. Residues located on the surface of the $3C^{pro}$, opposite from the catalytic site of the protease, have been shown to be essential for VPg uridylation, which is the first stage in the replication of viral RNA, by binding RNA (Nayak et al., 2006). Although SAT-specific substitutions were observed within this positively charged region, this region was conserved in the 3D structure.

As in other picornaviruses, protein $3D^{pol}$ is the RDRP responsible for the replication of the RNA genome via negative strand intermediates (Doherty et al., 1999; Ferrer-Orta et al., 2004). The $3D^{pol}$ demonstrated the least amount of variation indicating the importance of conserving the structural and functional integrity of the RDRP. When variation was observed, the variable residues were mostly conservative in nature and were pointing away from the active site.

This comparative study of the FMDV non-structural proteins provides an outline into their evolution, sequence variability and common elements among the representative topotypes circulating in the different geographical regions of sub-Saharan Africa. Genetic diversity of the non-structural proteins may be of consequence to control of FMD, as the non-structural proteins are targets for antiviral therapeutics (Curry et al., 2007) or in diagnostic assays for example to differentiate infected from vaccinated animals, or DIVA tests (Clavijo et al., 2004; Mackay, 1998; van Rensburg et al., 2002).

Acknowledgements

This work was supported by funding from THRIP of the National Research Foundation of South Africa. Peninah Nsamba was a recipient of an Organization for Women in Science for the Developing World (OWSD) from 2007 to 2010. The SAT1 and SAT2 viruses were received from the regional FMD reference laboratory at Transboundary Animal Diseases of the ARC-OVI. The authors would also like to thank Drs. Lia Rotherham and Junita Liebenberg for critical reading of the manuscript.

References

Acharya, R., Fry, E., Stuart, D., Fox, G., Rowlands, D., Brown, F., 1989. The three-dimensional structure of foot-and-mouth disease virus at 2.9 Å resolution. Nature 337, 709–716.

Bablanian, G.M., Grubman, M.J., 1993. Characterization of the foot-and-mouth disease virus 3C protease expressed in *Escherichia coli*. Virology (New York) 197, 320–327.

Barclay, W., Li, Q.Y., Hutchinson, G., Moon, D., Richardson, A., Percy, N., Almond, J.W., Evans, D.J., 1998. Encapsidation studies of poliovirus subgenomic replicons. J. Gen. Virol. 79, 1725–1734.

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

Bastos, A.D.S., Anderson, E.C., Bengis, R.G., Keet, D.F., Winterbach, H.K., Thomson, G.R., 2003a. Molecular epidemiology of SAT3-type foot-and-mouth disease. Virus Genes 27, 283–290.

Bastos, A.D.S., Haydon, D.T., Forsberg, R., Knowles, N.J., Anderson, E.C., Bengis, R.G., Nel, L.H., Thomson, G.R., 2001. Genetic heterogenicity of SAT-1 type foot-and-mouth disease viruses in southern Africa. Arch. Virol.146,1537–1551.

Bastos, A.D.S., Haydon, D.T., Sangaré, O., Boshoff, C.I., Edrich, J.L., Thomson, G.R.,2003b.Theimplicationsofvirusdiversitywithin theSAT2sero-type forcontroloffoot-and-mouthdiseaseinsub-SaharanAfrica.J. Gen.Virol. 84,1595–1606.

Beck, E., Strohmaier, K., 1987. Subtyping of European foot-and-mouth disease virus strains by nucleotide sequence determination. J. Virol. 61, 1621–1629.

Belsham, G.J., Brangwyn, J.K., 1990. A region of the 5′noncoding region of foot-and-mouthdiseasevirusRNAdirectsefficientinternalinitiation of proteinsynthesiswithincells:involvementwiththeroleofL proteasein translationalcontrol.J.Virol.64,5389–5395.

Belsham, G.J., 1993. Distinctive features of foot-and-mouth disease virus, a member of the picornavirus family; aspects of virus protein synthesis, protein processing and structure. Prog. Biophys. Mol. Biol. 60, 241–260.

Belsham, G.J., 2005. Translation and replication of FMDV RNA. Curr. Top. Microbiol. Immunol. 288, 43–70.

Birtley, J.R., Knox, S.R., Jaulent, A.M., Brick, P., Leatherbarrow, R.J., Curry, S., 2005. Crystal structure of foot-and-mouth disease virus 3C protease. J. Biol. Chem. 280, 11520–11527.

Boom, R., Sol, C.J., Salimans, M.M., Jansen, C.L., Wertheim-van Dillen, P.M., VanderNoordaa,J.,1990.Rapidandsimplemethodforpurificationof nucleicacids.J.Clin.Microbiol.28,495–503.

Bronsvoort B.M.d.C., Radford, A.D., Tanya, V.N., Nfon, C., Kitching, R.P., Morgan,K.L.,2004.Molecularepidemiologyoffoot-and-mouthdis-ease viruses in the Adamawa Province of Cameroon. J. Clin. Microbiol. 42, 2186–2196.

Carrillo, C., Tulman, E.R., Delhon, G., Lu, Z., Carreno, A., Vagnozzi, A., Kutish, G.F., Rock, D.L., 2005. Comparative genomics of foot-and-mouth disease virus. J. Virol. 79, 6487–6504.

Clarke, B.E., Sangar, D.V., 1988. Processing and assembly of foot-and-mouth-disease virus proteins using subgenomic RNA. J. Gen. Virol. 69, 2313–2325.

Clavijo, A., Wright, P., Kitching, P., 2004. Developments in diagnostic techniques for differentiating infection from vaccination in foot-and-mouth disease. Vet. J. 167, 9–22.

Curry, S., Abrams, C.C., Fry, E., Crowther, J.C., Belsham, G.J., Stuart, D.I., King, A.M.Q., 1995. Viral RNA modulates the acid sensitivity of foot-and-mouth disease virus capsids. J. Virol. 69, 430–438.

Curry, S., Roque-Rosell, N., Zunszain, P.A., Leatherbarrow, R.J., 2007. Foot-and-mouth disease virus 3C protease: recent structural and function-al insights into an antiviral target. Int. J. Biochem. Cell Biol. 39, 1–6.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9, 772.

Di Nardo, A., Knowles, N.J., Paton, D.J., 2011. Combining livestock trade patterns with phylogenetics to help understand the spread of foot and mouthdiseaseinsub-SaharanAfrica,theMiddleEastandSoutheast Asia. Rev.Sci.Tech.Off.Int.Epiz.30,63–85.

Doedens, J.R., Kirkegaard, K., 1995. Inhibition of cellular protein secretion by poliovirus proteins 2B and 3A. EMBO J. 14, 894.

Doherty, M., Todd, D., McFerran, N., Hoey, E.M., 1999. Sequence analysis of a porcine enterovirus serotype 1 isolate: relationships with other picornaviruses. J. Gen. Virol. 80 (Pt 8), 1929–1941.

Domingo, E., Escarmis, C., Martinez, M., Martinez-Salas, E., Mateu, M., 1992. Foot-and-mouth disease virus populations are quasispecies. Curr. Top. Microbiol. Immunol. 176, 33.

Domingo, E., Mateu, M.G., Martinez, M.A., Dopazo, J., Moya, A., Sobrino, F., 1990. Genetic Variability and Antigenic Diversity of Foot-and-Mouth Disease Virus, vol. 2. Plenum Publishing Corporation, NY, , pp. 233–266.

Donnelly, M.L.L., Luke, G., Mehrotra, A., Li, X.J., Hughes, L.E., Gani, D., Ryan, M.D., 2001. Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanismindicatesnotaproteolyticreaction,butanoveltransla-tional effect:aputativeribosomal'skip'.J.Gen.Virol.82,1013–1025.

Echeverri, A.C., Dasgupta, A., 1995. Amino terminal regions of poliovirus 2C protein mediate membrane binding. Virology 208, 540–553.

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

Ellard, F.M., Drew, J., Blakemore, W.E., Stuart, D.I., King, A.M.Q., 1999. Evidence for the role of His-142 of protein 1C in the acid-induced disassembly of foot-and-mouth disease virus capsids. J. Gen. Virol. 80, 1911–1918.

Ferrer-Orta, C., Arias, A., Perez-Luque, R., Escarmis, C., Domingo, E., Verdaguer, N., 2004. Structure of foot-and-mouth disease virus RNA-dependent RNA polymerase and its complex with a template-primer RNA.J.Biol.Chem.279,47212–47221.

Ferris, N.P., Donaldson, A.I., 1992. The World Reference Laboratory for Foot and Mouth Disease: a review of thirty-three years of activity (1958–1991). Rev. Sci. Tech. 11, 657.

Forss, S., Schaller, H., 1982. A tandem repeat gene in a picornavirus. Nucleic Acids Res. 10, 6441–6450.

Fry, E.E., Lea, S.M., Jackson, T., Newman, J.W.I., Ellard, F.M., Blakemore, W.E., Abu-Ghazaleh, R., Samuel, A., King, A.M.Q., Stuart, D.I., 1999. The structure and function of a foot-and-mouth disease virus-oligosac-charidereceptorcomplex.EMBOJ.18,543–554.

George, M., Venkataramanan, R., Pattnaik, B., Sanyal, A., Gurumurthy, C.B., Hemadri, D., Tosh, C., 2001. Sequence analysis of the RNA polymerase geneoffoot-and-mouthdiseasevirusserotypeasia1.VirusGenes22, 21–26.

Giraudo, A.T., Beck, E., Strebel, K., de Mello, P.A., La Torre, J., Scodeller, E.A., Bergmann, I.E., 1990. Identification of a nucleotide deletion in parts of polypeptide 3A in two independent attenuated aphthovirus strains. Virology177,780–783.

Grubman, M.J., Baxt, B., 2004. Foot-and-mouth disease. Clin. Microbiol. Rev. 17, 465–493.

Guarne´ ,A.,Tormo,J.,Kirchweger,R.,Pfistermueller,D.,Fita,I.,Skern,T., 1998. Structure of the foot-and-mouth disease virus leader protease: a papain-like fold adapted for self-processing and eIF4G recognition. EMBO J. 17, 7469–7479.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NTIn: Nucleic Acids Symposium Series, pp. 95–98.

Heath, L., van Der Walt, E., Varsani, A., Martin, D.P., 2006. Recombination patterns in aphthoviruses mirror those found in other picornaviruses. J. Virol. 80, 11827–11832.

Heinz, B.A., Vance, L.M., 1996. Sequence determinants of 3A-mediated resistance to enviroxime in rhinoviruses and enteroviruses. J. Virol. 70, 4854–4857.

Hogle, J., Chow, M., Filman, D., 1985. Three-dimensional structure of poliovirus at 2.9 A resolution. Science 229, 1358–1365.

Holland, J., Domingo, E., 1998. Origin and evolution of viruses. Virus Genes 16, 13–21.

Jackson, A.L., O'Neill, H., Maree, F., Blignaut, B., Carrillo, C., Rodriguez, L., Haydon, D.T., 2007. Mosaic structure of foot-and-mouth disease virus genomes. J. Gen. Virol. 88, 487–492.

King, A.M.Q., McCahon, D., Saunders, K., Newman, J.W.I., Slade, W.R., 1985. Multiple sites of recombination within the RNA genome of foot-and-mouth disease virus. Virus Res. 3, 373–384.

Knipe, T., Rieder, E., Baxt, B., Ward, G., Mason, P.W., 1997. Characterization of synthetic foot-and-mouth disease virus provirions separates acid-mediated disassembly from infectivity. J. Virol. 71, 2851–2856.

Knowles, N.J., Davies, P.R., Henry, T., O'Donnell, V., Pacheco, J.M., Mason, P.W., 2001. Emergence in Asia of foot-and-mouth disease viruses with altered host range: characterization of alterations in the 3A protein. J. Virol.75,1551–1556.

Knowles, N.J., Samuel, A.R., 2003. Molecular epidemiology of foot-and-mouth disease virus. Virus Res. 91, 65–80.

Knowles, N.J., Wadsworth, J., Hammond, J.M., King, D.P., 2010. Foot-and-mouth disease virus genotype definitions and nomenclature. In: Open Session of the European Commission for the Control of Foot-and-Mouth Disease Standing Technical Committee, 28 September–1 October, Vienna.

Koonin, E.V., 1991. The phylogeny of RNA-dependent RNA-polymerases of positive-strand RNA viruses. J. Gen. Virol. 72, 2197–2206.

Krebs,O.,Marquardt,O.,1992.Identificationandcharacterizationof foot-and-mouth disease virus O1 Burgwedel/1987 as an intertypic recombinant. J. Gen. Virol. 73, 613–619.

Kyte, J., Doolittle, R.F., 1982. A simple method for displaying the hydro-pathic character of a protein. J. Mol. Biol. 157, 105–132.

Lama,J.,Paul,A.V.,Harris,K.S.,Wimmer,E.,1994.Propertiesofpurified recombinant poliovirus protein 3AB as substrate for viral proteinases and as cofactor for RNA-polymerase 3D(pol). J. Biol. Chem. 269, 66–70.

Lama, J., Sanz, M.A., Carrasco, L., 1998. Genetic analysis of poliovirus protein 3A: characterization of a non-cytopathic mutant virus defec-tive in killing Vero cells. J. Gen. Virol. 79, 1911–1921.

Lawrence,P.,Rieder,E.,2009.IdentificationofRNAhelicaseAasanew host factor in the replication cycle of foot-and-mouth disease virus. J. Virol. 83, 11356–11366.

Lea, S., Hernandez, J., Blakemore, W., Brocchi, E., Curry, S., Domingo, E., Fry, E.,Abughazaleh,R.,King,A.,Newman,J.,Stuart,D.,Mateu,M.G.,1994. The structure and antigenicity of a type-C foot-and-mouth-disease virus. Structure2,123–139.

Mackay, D.K.J., 1998. Differentiating infection from vaccination in foot-and-mouth disease. Vet. Q. 20, 2–5.

Maree, F.F., Blignaut, B., Esterhuysen, J.J., de Beer, T.A.P., Theron, J., O'Neill, H.G., Rieder, E., 2011. Predicting antigenic sites on the foot-and-mouth disease virus capsid of the South African Territories types using virus neutralizationdata.J.Gen.Virol.92,2297–2309.

Martin, D.P., Williamson, C., Posada, D., 2005. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics 21, 260–262.

Martinez-Salas,E.,Regalado,M.P.,Domingo,E.,1996.Identificationofan essential region for internal initiation of translation in the aphtho-virus internal ribosome entry site and implications for viral evolution. J. Virol. 70,992–998.

Mason, P.W., Pacheco, J.M., Zhao, Q.Z., Knowles, N.J., 2003. Comparisons of thecompletegenomesofAsian,AfricanandEuropeanisolatesofa recent foot-and-mouth disease virus type O pandemic strain (PanA-sia). J. Gen. Virol.84,1583–1593.

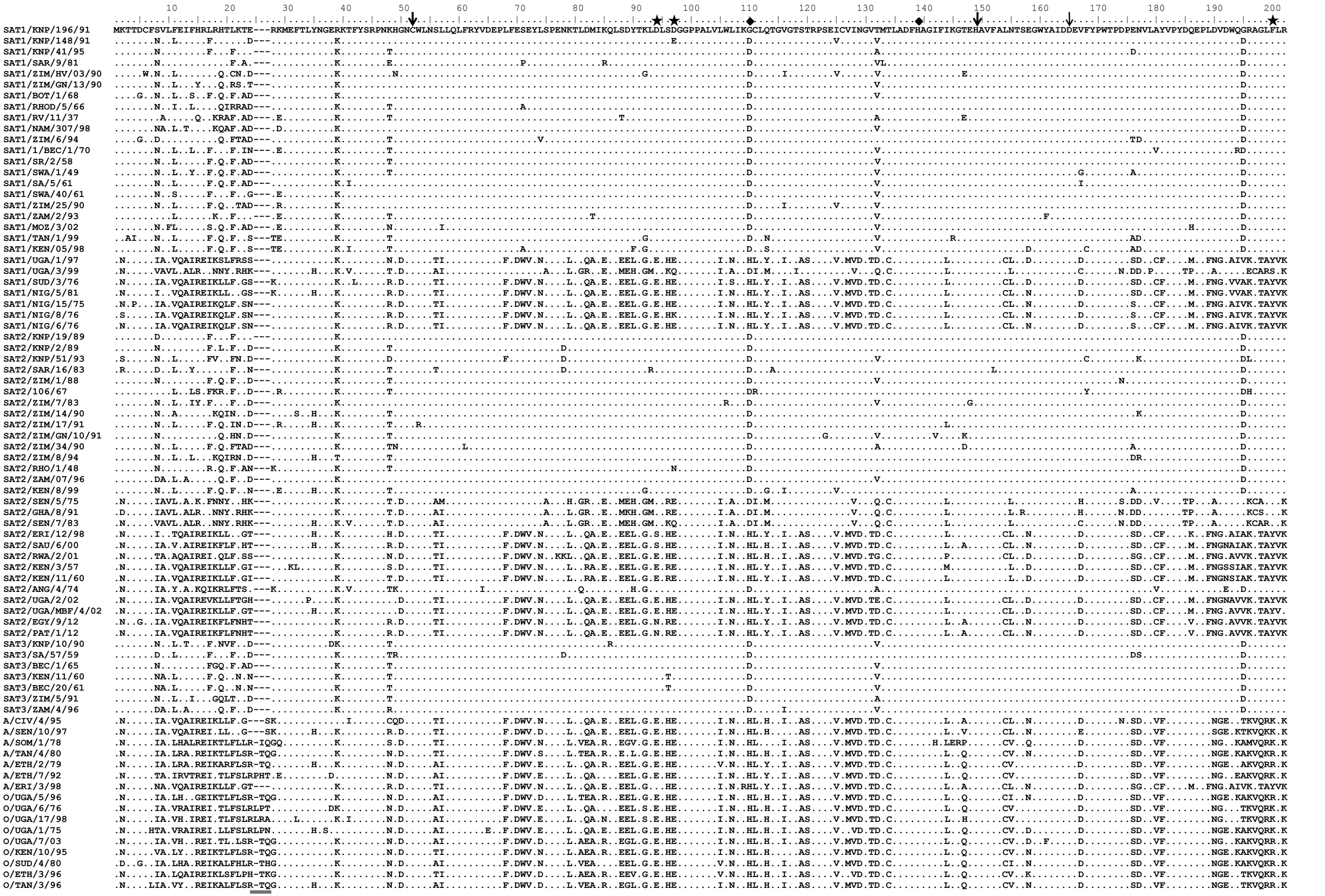
Moffat, K., Howell, G., Knox, C., Belsham, G.J., Monaghan, P., Ryan, M.D., Wileman, T., 2005. Effects of foot-and-mouth disease virus nonstruc-tural proteins on the structure and function of the early secretory pathway: 2BC but not 3A blocks endoplasmic reticulum-to-Golgi transport.J.Virol.79,4382–4395.

Moffat, K., Knox, C., Howell, G., Clark, S.J., Yang, H., Belsham, G.J., Ryan, M., Wileman, T., 2007. Inhibition of the secretory pathway by foot-and-mouth disease virus 2BC protein is reproduced by coexpression of 2B

with 2C, and the site of inhibition is determined by the subcellular location of 2C. J. Virol. 81, 1129–1139.
Nayak, A., Goodfellow, I.G., Woolaway, K.E., Birtley, J., Curry, S., Belsham, G.J., 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.
O'Donnell, V., Pacheco, J.M., LaRocco, M., Burrage, T., Jackson, W., Rodriguez, L.L., Borca, M.V., Baxt, B., 2011. Foot-and-mouth disease virus utilizes an autophagic pathway during viral replication. Virology 410, 142–150.
Pacheco, J.M., Henry, T.M., O'Donnell, V.K., Gregory, J.B., Mason, P.W., 2003. Role of nonstructural proteins 3A and 3B in host range and pathogenicity of foot-and-mouth disease virus. J. Virol. 77, 13017–13027. Pereda, A.J., Ko¨nig, G.A., Chimeno Zoth, S.A., Borca, M., Palma, E.L., Piccone, M.E., 2002. Full length nucleotide sequence of foot-and-mouth disease virus strain O 1 Campos/Bra/58. Arch. Virol. 147, 2225–2230.
Pereira, H.G., 1981. Foot-and-mouth Disease, vol. 2. Academic Press Inc., London, , pp. 333–363.
Piccone, M.E., Zellner, M., Kumosinski, T.F., Mason, P.W., Grubman, M.J., 1995.Identificationoftheactive-siteresiduesoftheLproteinaseof foot-and-mouth disease virus. J. Virol. 69, 4950–4956.
Reeve, R., Blignaut, B., Esterhuysen, J.J., Opperman, P., Matthews, L., Fry, E.E.,deBeer,T.A.P.,Theron,J.,Rieder,E.,Vosloo,W.,O'Neill,H.G., Haydon, D.T., Maree, F.F., 2010. Sequence-based prediction for vac-cine strain selection and identification of antigenic variability in foot-and-mouth diseasevirus.PLoSComput.Biol.6,e1001027.
Reid, S.M., Ferris, N.P., Hutchings, G.H., Samuel, A.R., Knowles, N.J., 2000. Primary diagnosis of foot-and-mouth disease by reverse transcription polymerase chain reaction. J. Virol. Methods 89, 167–176.
Rodriguez, A., Nunez, J.I., Nolasco, G., Ponz, F., Sobrino, F., Deblas, C., 1994. Direct PCR detection of foot-and-mouth-disease virus. J. Virol. Methods 47, 345–349.
Rueckert, R.R., 1996. Picornaviridae: The Viruses and their Replication, vol. 1. , pp. 609–654.
Ryan, M.D., Donnelly, M.L.L., Flint, M., Cowton, V.M., Luke, G., Hughes, L.E., Knox, C., Felipe, P., Sobrino, F., Domingo, E., 2004. Foot-and-mouth disease virus proteinases. Foot Mouth Dis.: Curr. Perspect. 53–76.
Ryan, M.D., King, A.M.Q., Thomas, G.P., 1991. Cleavage of foot-and-mouth-disease virus polyprotein is mediated by residues located within a 19 amino-acid-sequence. J. Gen. Virol. 72, 2727–2732.
Sali, A., Blundell, T.L., 1993. Comparative protein modeling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779–815.
Samuel, A.R., Knowles, N.J., 2001. Foot-and-mouth disease type O viruses exhibit genetically and geographically distinct evolutionary lineages (topotypes). J. Gen. Virol. 82, 609–621.
Sawyer, S.A., 1989. Statistical tests for detecting gene conversion. Mol. Biol. Evol. 6, 526–538.
Schneider, T.D., Stephens, R.M., 1990. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 18, 6097–6100.
Simmonds, P., Welch, J., 2006. Frequency and dynamics of recombination withindifferentspeciesofhumanenteroviruses.J.Virol.80,483–493.
Sobrino,F.,Sa´ iz,M.,Jime´ nez-Clavero,M.A.,Nu´ n˜ez,J.I.,Rosas,M.F.,Baranowski, E., Ley, V., 2001. Foot-and-mouth disease virus: a long known virus: but a current threat. Vet. Res. 32, 1–30.
Strebel, K., Beck, E., 1986. A second protease of foot-and-mouth disease virus. J. Virol. 58, 893–899.
Sweeney, T.R., Cisnetto, V., Bose, D., Bailey, M., Wilson, J.R., Zhang, X., Belsham, G.J., Curry, S., 2010. Foot-and-mouth disease virus 2C is a hexamericAAA+proteinwithacoordinatedATPhydrolysismecha-nism. J.Biol.Chem.285,24347–24359.
Sweeney, T.R., Roque-Rosell, N., Birtley, J.R., Leatherbarrow, R.J., Curry, S., 2007. Structural and mutagenic analysis of foot-and-mouth disease virus3Cproteaserevealstheroleofthebeta-ribboninproteolysis.J. Virol. 81,115–124.
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.
Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997.TheCLUSTAL_Xwindowsinterface:flexiblestrategiesformul-tiple sequencealignmentaidedbyqualityanalysistools.Nucleic AcidsRes.25, 4876–4882.
Tully, D.C., Fares, M.A., 2009. Shifts in the selection-drift balance drive the evolution and epidemiology of foot-and-mouth disease virus. J. Virol. 83, 781–790.
Vakharia, V.N., Devaney, M.A., Moore, D.M., Dunn, J.J., Grubman, M.J., 1987. Proteolytic processing of foot-and-mouth disease virus polyproteinsexpressedinacell-freesystemfromclone-derivedtran-scripts. J.Virol.61,3199–3207.
Valdazo-Gonza´lez, B., Knowles, N.J., Hammond, J., King, D.P., 2012. GenomesequencesofSAT2foot-and-mouthdiseasevirusesfromEgypt and Palestinianautonomousterritories(GazaStrip).J.Virol.86, 8901–8902.

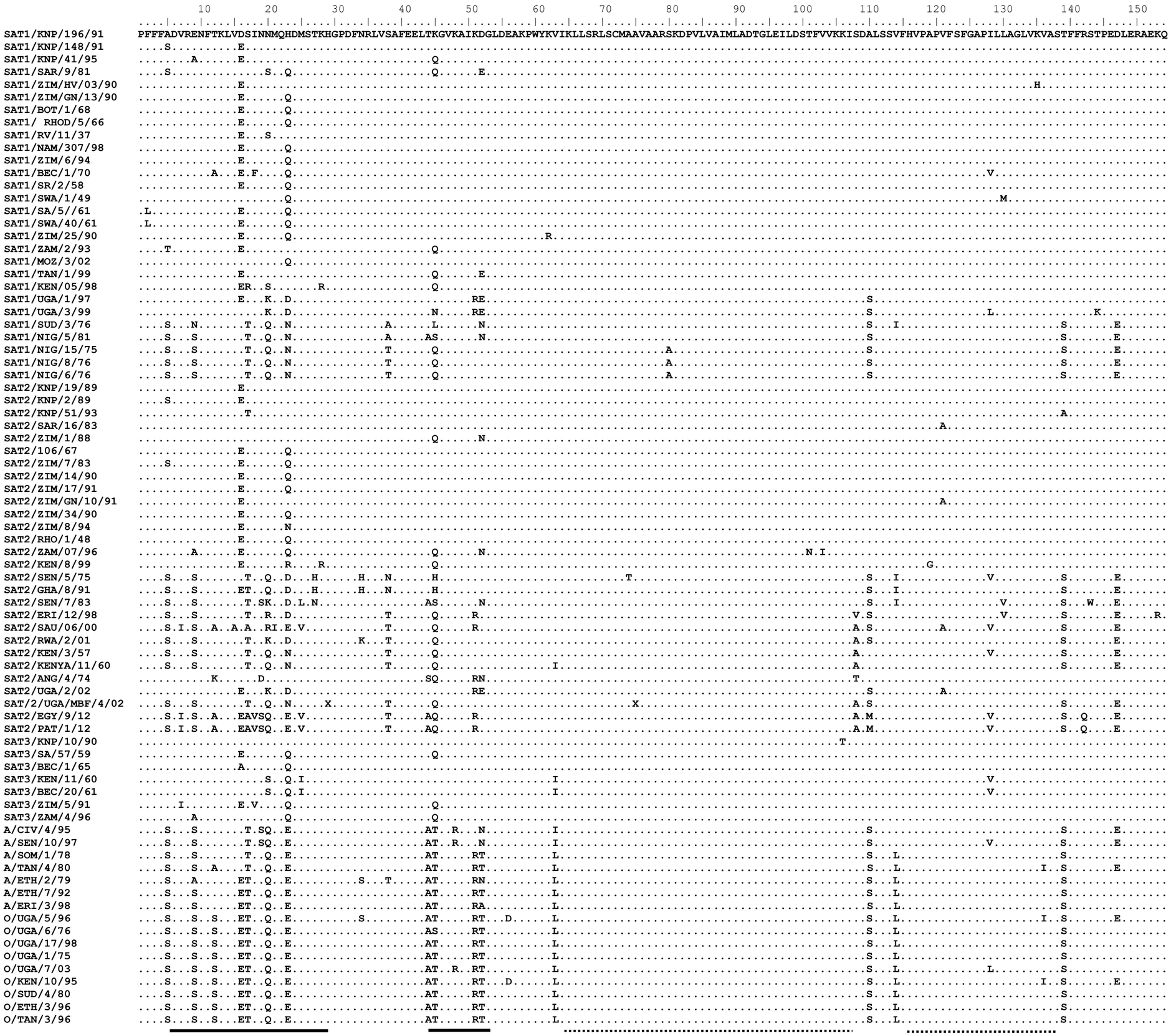
van Kuppeveld, F.J.M., Hoenderop, J.G.J., Smeets, R.L.L., Willems, P.H.G.M., Dijkman, H.B.P.M., Galama, J.M.D., Melchers, W.J.G., 1997. Coxsackievirus protein 2B modifies endoplasmic reticulum membrane and plasmamembranepermeabilityandfacilitatesvirusrelease.EMBOJ. 16, 3519–3532.
van Rensburg, H., Haydon, D., Joubert, F., Bastos, A., Heath, L., Nel, L., 2002. Genetic heterogeneity in the foot-and-mouth disease virus Leader and 3C proteinases. Gene 289, 19–29.
Vosloo, W., Bastos, A.D.S., Sangare, O., Hargreaves, S.K., Thomson, G.R., 2002. Review of the status and control of foot and mouth disease in Sub-Saharan Africa. Rev. Sci. Tech. Off. Int. Epiz. 21, 437–445.
Vosloo, W., Kirkbride, E., Bengis, R.G., Keet, D.F., Thomson, G.R., 1995. Genome variation in the SAT types of foot-and-mouth disease viruses prevalentinbuffalo(*Synceruscaffer*)intheKrugerNational Parkandother regionsofsouthernAfrica,1986–93.Epidemiol.Infect. 114,203.

Weber, S., Granzow, H., Weiland, F., Marquardt, O., 1996. Intracellular membrane proliferation in *E. coli* induced by foot-and-mouth disease virus 3A gene products. Virus Genes 12, 5–14.
Wessels, E., Duijsings, D., Lanke, K.H.W., van Dooren, S.H.J., Jackson, C.L., Melchers, W.J.G., van Kuppeveld, F.J.M., 2006. Effects of picornavirus 3A proteinsonproteintransportandGBF1-dependentCOP-Irecruit-ment.J. Virol.80,11852–11860.
Xiang, W., Paul, A.V., Wimmer, E., 1997. RNA signals in entero-and rhinovirus genome replication. Semin. Virol. 256–273.



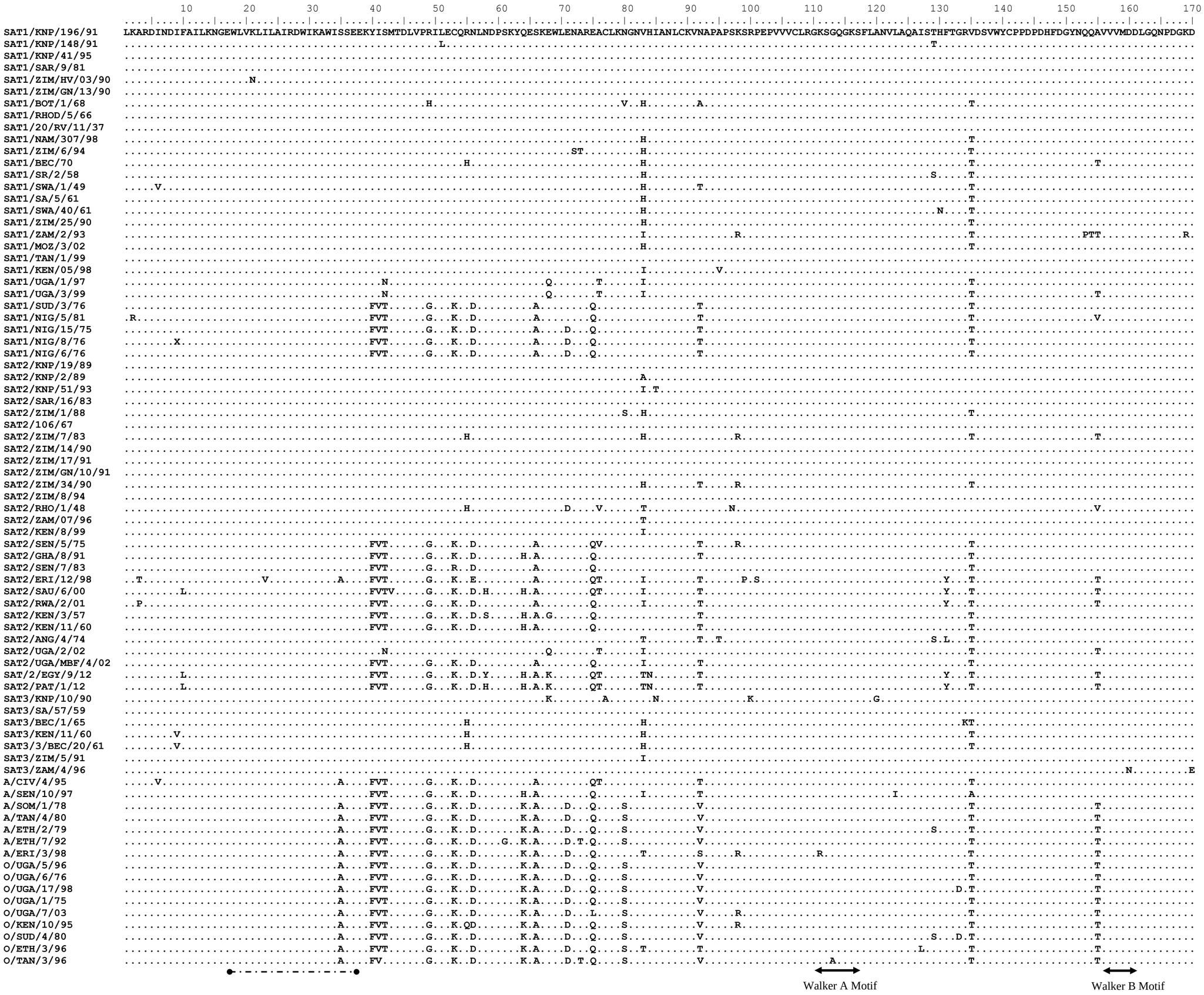
Supplementary Fig. A:

An alignment of 202 amino acid deduced positions for the Leader protease for 79 African FMDV isolates. The dot (.) indicate same amino acid residue as the reference sequence (SAT1/KNP/196/96) and a dash (-) depicts a gap generated during alignment of a particular sequence. Illustrated above the alignment are the critical residues involved in catalysis of L-VP4: C52, H149 and D165 (Guarné et al., 1998), indicated by the downward arrow (↓). Three other residuesE94, E97 and K200 associated with catalysis of L-VP4 (Guarné et al., 2000; Guarné et al., 1998) are indicated by the star (★). The position of the residues involved in eIF4G cleavage: H110 and H139 (Piccone et al., 1995) are shown by the kites (♦). The position of three deletions (amino acid positions 25-27) in the alignments for the SAT virus isolates is indicated below the alignment with a double underline (═══) at the bottom of the alignment.



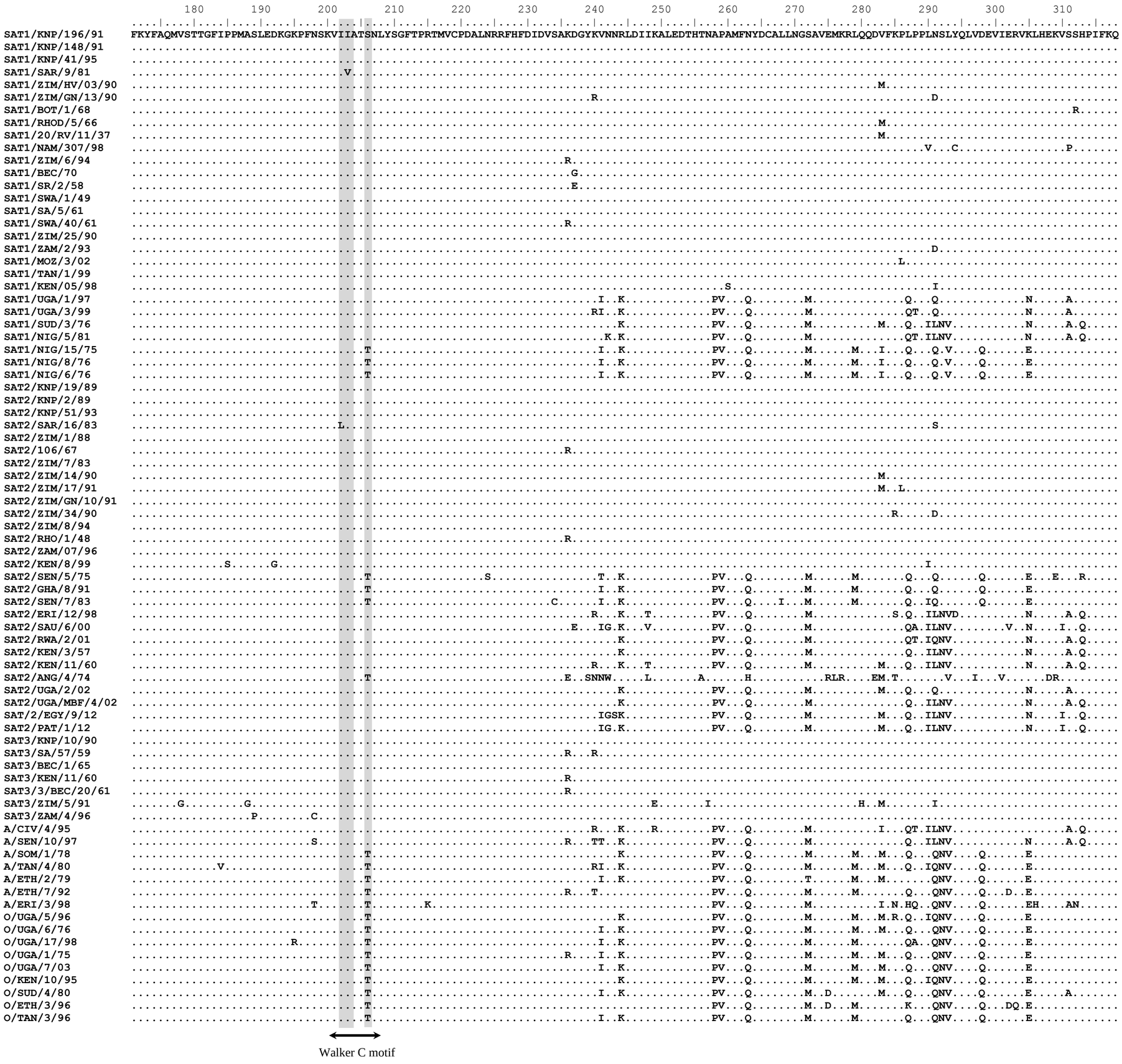
Supplementary Fig. B:

An alignment of 154 amino acid positions deduced for the 2B peptide for 79 African FMDV isolates. The dots (**.**) indicate similarity to the reference sequence (SAT1/KNP/196/96). The hypervariable domains (residues positions 5-29 and 44-53) in the alignment are shown by the horizontal block bars (_____) at the bottom of the alignment. Highly conserved motifs (residue positions 64-107 and 115-137) are depicted using horizontal dotted bars (.....). The latter conserved motif is a hydrophobic domain (Carrillo et al., 2005).

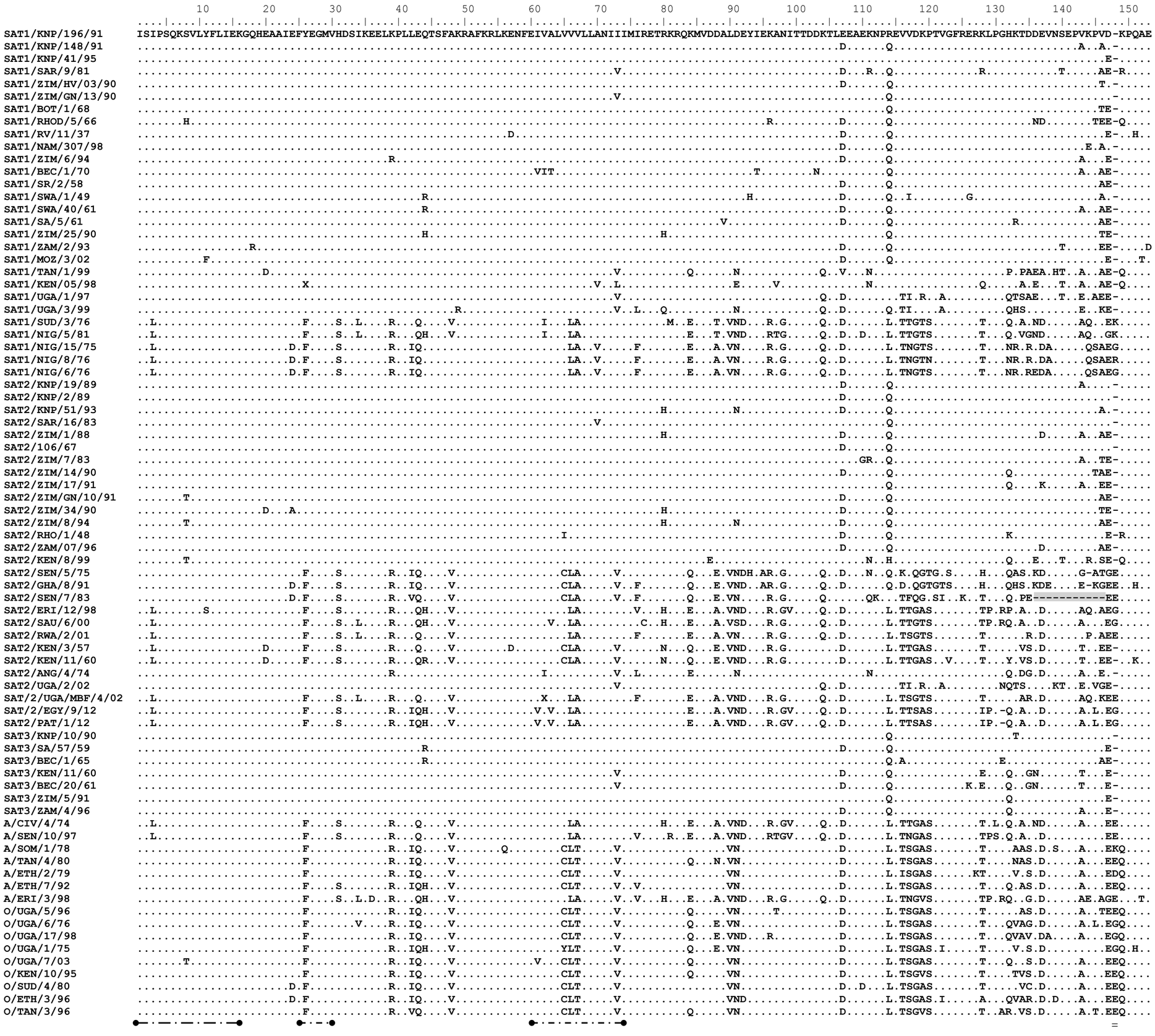


Supplementary Fig. C:

An alignment of 318 amino acid positions deduced for the 2C peptide for a total of 79 African FMDV isolates. The dots (**.**) indicate similarity to the reference sequence (SAT1/KNP/196/96). The hydrophobic domain in the alignment (residue positions 17-34) is indicated by horizontal dash-dot bar ( ) at the bottom. The horizontal double arrows (↔) define the residue positions in the Walker A motif (110-117), Walker B motif (156-161) and the Walker C motif (201-207) (Sweeney et al., 2010). The vertical grey bars highlight residue substitutions I202→L, I203→V and T206→S observed in the southern SAT virus isolates that occurred in the conserved Walker C motif.

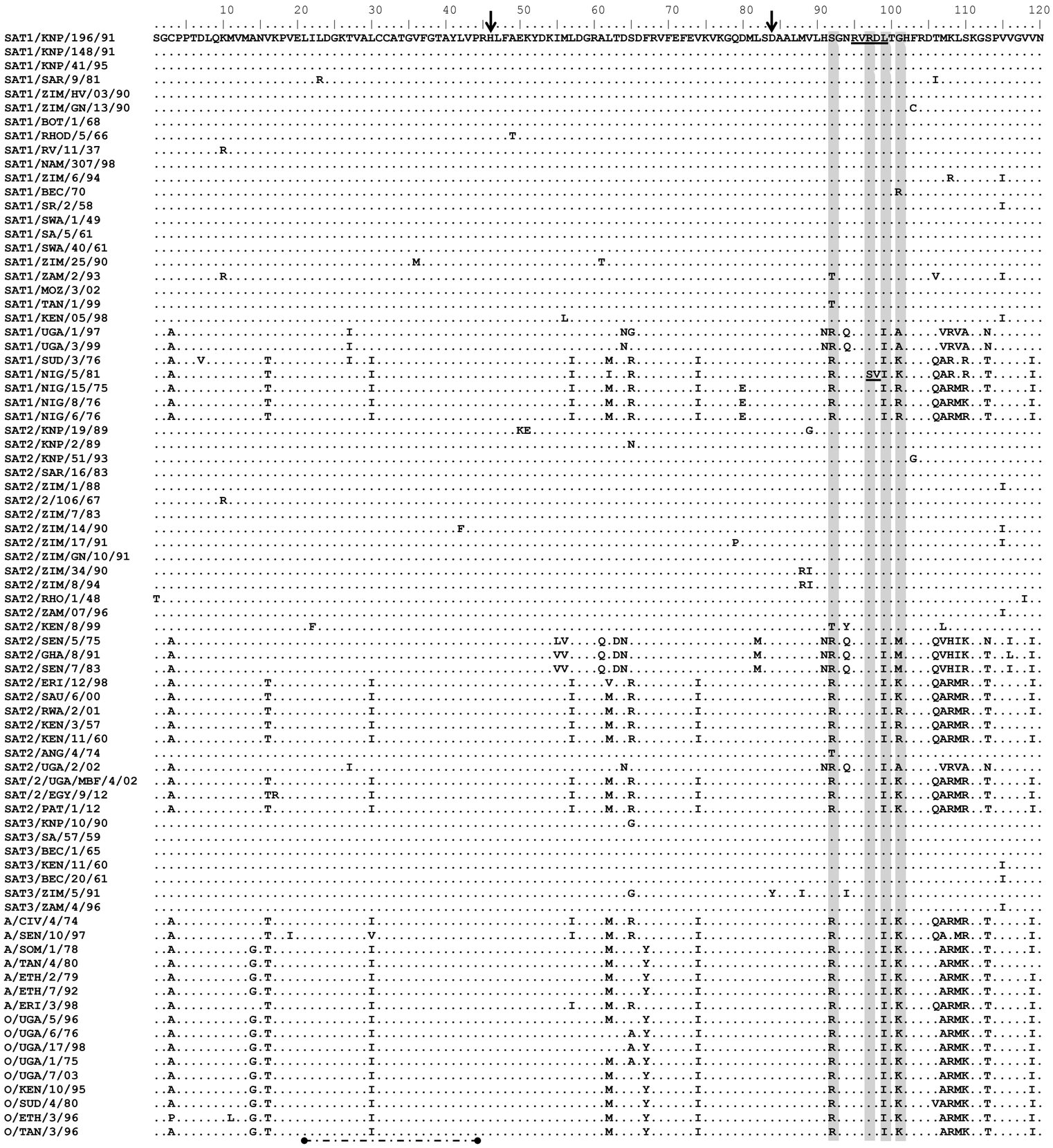


Supplementary Fig. C (continued)



Supplementary Fig. D:

An alignment of 153 deduced amino acid positions for the 3A peptide for a total of 79 African FMDV isolates. The dots (.) indicate similarity with the reference sequence (SAT1/KNP/196/96) and a dash (-) depicts a gap generated during the alignment. Illustrated among other things at the bottom of the alignment are the hydrophobic domains (residue positions 1-16: ISIPSQKSVLYFLIEK, 25-30: FYEGMV and 60-74: EIVALVVVLLANIII) depicted using a dash-dot line (_._._._). A double underline (═══) at amino acid position 148 shows the deletion observed in southern SAT virus isolates that grouped cluster I. The eleven residue deletions observed in SAT2/SEN/7/83 are highlighted in the grey horizontal bar.



Supplementary Fig. E:

An alignment of 213 deduced amino acid positions for the 3C protease for a total of 79 African FMDV isolates. The dots (**.**) indicate identity with the reference sequence (SAT1/KNP/196/96). At the bottom of the alignment are horizontal dotted bars (.....) which indicate highly conserved areas in the sequence alignment (residue positions 17-48, 68-90, 129-156, 161-164). The horizontal dash-dot bars (-.-.-.) depict a hydrophobic domain observed at residue positions 27- 44. At the top of the alignment, the residue positions of the active triad (H46, D84 and C163) (Birtley et al., 2005) are shown using the downwards arrow (↓). The residues (154-160, 181-186) that comprise substrate pocket (S1) (Birtley et al., 2005) are indicated by the wavy underline (~~~~~). The three conserved residues Y154, T158 and H181 that donate hydrogen bonds to the P1 substrate (Birtley et al., 2005) are indicated using the kites (♦). Underlined (___) at the top of the sequence alignment is the conserved motif 95-RVRDI-99 necessary for VPg uridylation (Nayak et al., 2006). Highlighted with grey vertical bars are residues R92, R97, I99 and K101 that contribute towards the uridylation process (Nayak et al., 2006). The following substitutions were observed in the southern SAT virus isolates R92→ S/T, R97→S, I99→L and K101→ G/A. The residue substitutions of 97R→S and 98D→V in this conserved motif for SAT1/NIG/5/81 are underlined.

```
                                130       140       150       160       170       180       190       200       210
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|...
SAT1/KNP/196/91     NADVGRLIFSGDALTYKDLVVCMDGDTMPGLFAYRAGTKVGYCGAAVLAKDGAKTVIVGTHSAGGNGVGYCSCVSRSMLLQMKAHIDPPPHTE
SAT1/KNP/148/91     ............X.............X...X...R.....X.................Y...S...C....F...................
SAT1/KNP/41/95      ....................................................................................................
SAT1/SAR/9/81       ....................................................................................................
SAT1/ZIM/HV/03/90   ....................................................................................................A....
SAT1/ZIM/GN/13/90   ...........S........I.....................................................................A....
SAT1/BOT/1/68       ....................................................................................................
SAT1/RHOD/5/66      ....................................................................................................A....
SAT1/RV/11/37       ....................................................................................................A....
SAT1/NAM/307/98     ....................................................................................................
SAT1/ZIM/6/94       ........................X...X.....................R.........................X.....................
SAT1/BEC/70         ....................................................................................................
SAT1/SR/2/58        ....................................................................................................
SAT1/SWA/1/49       ....................................................................................................
SAT1/SA/5/61        ....................................................................................................
SAT1/SWA/40/61      ....................................................................................................
SAT1/ZIM/25/90      ....................................................................................................
SAT1/ZAM/2/93       ....................................................................................................
SAT1/MOZ/3/02       ....................................................................................................
SAT1/TAN/1/99       ....................................................................................................
SAT1/KEN/05/98      ..........................................................................................K.....
SAT1/UGA/1/97       ................I.........................T.S.SQ...I.................K...D........A....
SAT1/UGA/3/99       ................I.........................T.S.SQ...I.................K...D........A....
SAT1/SUD/3/76       ...........E.....I.............K.A..A....G........E.F.................K.........V..E..H.
SAT1/NIG/5/81       ...........E.....I.............K.A..A....G........E.F.....................................E..H.
SAT1/NIG/15/75      ...........E.....I.............K.A..A....G........E.F.................K.........V..E..H.
SAT1/NIG/8/76       ...........E.....I.............K.A..A....G........E.F.................K.........V..E..H.
SAT1/NIG/6/76       ...........E.....I.............K.A..A....G........E.F.................K.........V..E..H.
SAT2/KNP/19/89      ....................................................................................................
SAT2/KNP/2/89       ....................................................................................................
SAT2/KNP/51/93      ....................................................................................................A....
SAT2/SAR/16/83      ....................................................................................................
SAT2/ZIM/1/88       ....................................................................................................
SAT2/2/106/67       ....................................................................................................A..N.
SAT2/ZIM/7/83       ....................................................................................................
SAT2/ZIM/14/90      ....................................................................................................
SAT2/ZIM/17/91      ....................................................................................................
SAT2/ZIM/GN/10/91   ....................................................................................................A....
SAT2/ZIM/34/90      ....................................................................................................
SAT2/ZIM/8/94       ....................................................................................................A....
SAT2/RHO/1/48       .......V............................................................................................
SAT2/ZAM/07/96      ....................................................................................................
SAT2/KEN/8/99       ..........................................................................................K.....
SAT2/SEN/5/75       ................V.....................G..MT....H...I.................S...L......E.R..
SAT2/GHA/8/91       ................V.....................G..MT....H...I.................S...L......E.R..
SAT2/SEN/7/83       ................V.....................G.IMT....H...I.................S...L......E.R..
SAT2/ERI/12/98      ...........E.....I.............K.A..A....G........E.F.....................................V..E..H.
SAT2/SAU/6/00       ...........E.....I.............K.A..A....G........E.F.....................................V..E..H.
SAT2/RWA/2/01       ...........E.....I.............K.A..A....G........E.F.................K.....................E..H.
SAT2/KEN/3/57       ...RW......E.....I.............K.A..A....G........E.F.................K.........V..E..H.
SAT2/KEN/11/60      ...........E.....I.............K.A..A....G........E.F.................K.........V..E..H.
SAT2/ANG/4/74       ..........................................I.....................................L.......
SAT2/UGA/2/02       ................I.........................T.S.SQ...I.................K...D........A....
SAT/2/UGA/MBF/4/02  ...........E.....I.............K.A..A....G........E.F.................K.........V..E..H.
SAT/2/EGY/9/12      ...........E.....I.............K.A..A....G........E.F.................K.....................E..H.
SAT2/PAT/1/12       ...........E.....I.............K.A..A....G........E.F.................K.....................E..H.
SAT3/KNP/10/90      ....................................................................................................
SAT3/SA/57/59       ....................................................................................................
SAT3/BEC/1/65       ....................................................................................................
SAT3/KEN/11/60      ....................................................................................................
SAT3/BEC/20/61      ....................................................................................................
SAT3/ZIM/5/91       ....................................................................................................
SAT3/ZAM/4/96       ........................P...........................................................................
A/CIV/4/74          ...........E.....I.................A..A....G........E.F.....................................V..E..H.
A/SEN/10/97         .....K.....E.....I.............K.A..A....G........E.F.....................................V..E..H.
A/SOM/1/78          ...........E.....I.................A..A....G........D.F.....................K.......E..H.
A/TAN/4/80          ...........E.....I.................A..A....G........D.F.................K....K....V..E..H.
A/ETH/2/79          ...........E.....I.................A..A....G........D.F.....................K.......E..H.
A/ETH/7/92          ...........E.....I.................A..A....G........D.F.....................K.......E..H.
A/ERI/3/98          ...........E.....I.............K.A..A....G........E.F.....................................V..E.YH.
O/UGA/5/96          ...........E.....I.................A..A....G........E.F.....................K.......E..H.
O/UGA/6/76          ...........E.....I.................A..A....G........D.F.....................K.......E..H.
O/UGA/17/98         ...........E.....I.................A..A....G........D.F.....................K.......E..H.
O/UGA/1/75          ...........E.....I.................A..A....G........D.F.....................K.......E..H.
O/UGA/7/03          ...........E.....I.................A..A....G........D.F.....................K.......E..H.
O/KEN/10/95         ...........E.....I.................A..A....G........N.F.....................K....V..E..H.
O/SUD/4/80          ...........E.....I......E..........A..A....G........D.F.....................K.......E..H.
O/ETH/3/96          ................I.................A..A....G........D.F.....................K.......E..H.
O/TAN/3/96          ...........E.....I.................A..A....G........S.F.....................K....V..E..H.
                            ....................................     .....
```

Supplementary Fig. E (continued)

```
                                10        20        30        40        50        60        70        80        90       100       110       120       130       140       150       160       170       180
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SAT1/KNP/196/91     GLVVDTREVEERVHVMRKTKLAPTVAHGVFQPEFGPAALSNNDKRLNEGVVLDEVIFSKHKGDAKMSEADKRLFRLCAADYASHLHNVLGTANSPLSVFEAIKGVDGLDAMEPDTAPGLPWALQGKRRGALIDFENGTVGPEIEQALKLMEKKEYKFTCQTFLKDEIRPLEKVKAGKTRI
SAT1/KNP/148/91     ....................................................................................................................................................................................
SAT1/KNP/41/95      ....................................................................................................................................................................................
SAT1/SAR/9/81       ........................Y.....................................K.....................................................................................................................
SAT1/ZIM/HV/03/90  ....................................................................................................................................................................................
SAT1/ZIM/GN/13/90  ....................................S.....................K.............D.....................................................................................M.....
SAT1/BOT/1/68      ....................................................................................................................A...F.....................................................
SAT1/RHOD/5/66     .....................................K.....................................................................................................................
SAT1/RV/11/37      ....................................................................................................................................................................................
SAT1/NAM/307/98    ...................................T.....................E.....................................A.....................................................
SAT1/ZIM/6/94      ....................D.....................K.....................................N.....A.....R.................G.....
SAT1/BEC/1/70      ....................T.....................A.....................I.....
SAT1/SR/2/58       ....................................................................A.....T.....
SAT1/SWA/1/49      ...................T.....................DA.....
SAT1/SA/5/61       ....................V.....................A.....
SAT1/SWA/40/61     ...L.....D.....................A.....R.....
SAT1/ZIM/25/90     ....................................................................A.....
SAT1/ZAM/2/93      ..............L.....................A.....
SAT1/MOZ/3/02      ....................................................................D.....A.....R.....
SAT1/TAN/1/99      ....................X.....K.....R.....M.....
SAT1/KEN/05/98     ....................................................................NRN.....M.....
SAT1/UGA/1/97      ..I.V.....GY.....D.....K...M.....T.....N.....V.....M...R.....
SAT1/UGA/3/99      ..I.....Y.....D.....K...M.....T.....T.....S.....V.....M...R.....
SAT1/SUD/3/76      ..I..S.D.....N.D.....R.....I..D.....T...E..K.....S.....Y.....T.....E.....R.....M...R.....
SAT1/NIG/5/81      ..I....D.....N.D.....S.....I..D.....T...E.....S.....Y.....T.....E.....R.....M...R.....
SAT1/NIG/15/75     ..I....D.....N.D.....I..D.....T...E..K...M.....S.....A...IY.....T.....A.....R.....M...R.....
SAT1/NIG/8/76      ..I....D.....N.D.....I..D.....T...E..K...M.....S.....A...IY.....T.....A.....RR.....M...R.....
SAT1/NIG/6/76      ..I....D.....N.D..R.....I..D.....T...E..K...M.....S.....A...IY.....T.....A.....R.....M...R.....
SAT2/KNP/19/89     ....................................................................................................................................................................................
SAT2/KNP/2/89      ....G.....K.....R.....
SAT2/KNP/51/93     ....G.....T.....K.....
SAT2/SAR/16/83     ....G.....Y.....K.....E.....
SAT2/ZIM/1/88      ....................................................................A.....
SAT2/106/67        ....................V.....
SAT2/ZIM/7/83      ....................S.....R.....S...A.....
SAT2/ZIM/14/90     ....................K.....E.....T.....
SAT2/ZIM/17/91     ....................E...R.....
SAT2/ZIM/GN/10/91  .....R.....K.....
SAT2/ZIM/34/90     ....................A.....
SAT2/ZIM/8/94      ....G.....K.....
SAT2/RHO/1/48      ....................K.....T.....E.....
SAT2/ZAM/07/96     ....................A.....
SAT2/KEN/8/99      ....................R.....M.....
SAT2/SEN/5/75      ..I.....D.....R..H.....D.....T.....S.....E.....E.....R.....R.....
SAT2/GHA/8/91      ..I.....D.....R.....D.....T.....S.....E.....E.....RR....S.....C.....
SAT2/SEN/7/83      ..I.....D.....R.....D.....S.....G.....E.....R.....R.....
SAT2/ERI/12/98     ..I....D.....N.D.....D.....D.....G..K...A.....S.....A....Y.....IE.....T.....V.E.....R.....M...R.....
SAT2/SAU/6/00      ..I....D.....N.....DQ.....D.....T...E..K...A.....S.....A...IY.....I.....T.....A...V.E.....R.....M...R.....
SAT2/RWA/2/01      ..I....D.....N..Y.....D.....T...E..K.....S.....T...IY.....T.....E.....R.....R.....
SAT2/KEN/3/57      ..I....D.....R..L...Y.....T...E..K.....S.....A...IY.....T.....A.....NR.....R.....
SAT2/KEN/11/60     ..I....D.....N..Y.....T...E..K.....R..S.....A...IY.....T.....A.....R.....V.....R.....
SAT2/ANG/4/74      ..........I.....S.....T.....P.....Q.DERN.....R.....
SAT2/UGA/2/02      ..I..S.....Y.....D.....T.....K.....T.....T.....S.....V.....M...R.....
SAT/2/UGA/MBF/4/02 ..I....D.....X.....N.D.....I..D.....T...E..K...M.....S.....A...IY.....T.....E.....R.....M...R.....
SAT2/EGY/9/12      ..I....D.....N.D.....D...I..D.....K...E..K...T.....S.....A...IY.....I.....T.....V.E.....R.....M...R.....
SAT2/PAT/1/12      ..I....D.....N.D.....D.....D.....K...E..K...T.....S.....A...IY.....I.....T.....V.E.....R.....M...R.....
SAT3/KNP/10/90     .W.....G...T.....R.....M.....
SAT3/SA/57/59      .....K.....R.....
SAT3/BEC/1/65      .....T.....A.....
SAT3/KEN/11/60     ....................A.....
SAT3/BEC/20/61     ....................A.....
SAT3/ZIM/5/91      .....K.....
SAT3/ZAM/4/96      .W.....I.....A.....
A/CIV/4/74         ..I....D.....N.D.....R..D...I..D.....T...E..K.....S.....A...IY.....T.....E.....R.....M...R.....
A/SEN/10/97        ..I....D.....N.D.....T..V.V.....T...E..K...I.....S.....A...IY.....T.....E.....R.....M...RT....
A/SOM/1/78         ..I....D.....N.....K.P.....T..T.E..A...R.....R..S.....A...IY.....V.....VAA..E....R....A.....M...R.....
A/TAN/4/80         ..I....D.....N.....K.P.....T..T.E..A...R.....R..S.....A...IY.....VAAP.E....R....A.....M...R.....
A/ETH/2/79         ..I....D.....N.....K.P.....T..T.E..A...R.....R..S.....A...IY.....VAA..E....R....A.....V..M...R.....
A/ETH/7/92         ..I....D.....N.....K.P.....T..T.E..A...R.....R..S.....A..NIY.....VAA..E....R....A.....M...R.....
A/ERI/3/98         ..IL...D.....Y...N.D.....R..D.....D.....T...E..K...T.....SI.....A....Y.....I.....T.....V.E.....R.....M...R.....
O/UGA/5/96         ..I....D.....K.....K.P.....T.....A...R.....S.....A...IY.....VAA..E....QQ...A.....M...R.....
O/UGA/6/76         ..I....D.....N.....V...R.P.....T....E..A...R.V.....R..S.....A...IY.....VAA..E....R....A.....M...R.....
O/UGA/17/98        ..II...D.....R.....N.....K.P.....T....E..A...R.....R..S.....A...IY.....VAT..E....RN...A.....M...R.....
O/UGA/1/75         ..I....D.....N.D.....K.P.....T..N.E..A...R.....R..S.....A...IY.....VAA..E....R....A.....R.....
O/UGA/7/03         ..I....D.....N.....S...K.P.....T....E..A...R.....R..SE.....A....Y.....VVA..E....R....A.....M...R.....
O/KEN/10/95        ..I..S.D.....N.....K.P.....T.E..P...R.....S.....A...IY.....VAA..E....R....A.....M...R.....
O/SUD/4/80         ..I....D.....K.....K.P.....V..T.E..A...R.....S.....A...IY.....E.....VAA..E....R....A.....M...R.....
O/ETH/3/96         ..II...D.....N.....R.P.....T....E..A...R.....S..M..A....Y.....VTA..E....R....A.....M...R.....
O/TAN/3/96         ..I....D.....N.....K.P.....A.....K....E..A...R.....R..G.....A...IY.....VTA..E....RQ...A.....M...R.....
```

Supplementary Fig. F:

An alignment of 470 deduced amino acid positions for the 3D RNA Dependent RNA Polymerase for a total of 79 African FMDV isolates. The dots (.) indicate similarity with reference sequence (SAT1/KNP/196/96). Underlined at the top of the alignment are five highly conserved motifs among the RNA polymerases (Doherty et al., 1999; Ferrer-Orta et al., 2004) i.e. KDEIR (positions164-168), DYSAFD (positions 240-245), PSG (positions 297-299), YGDD (positions 336-339) and FKLR (positions 385-388). The KDEVR motif observed in A/ETH/2/79 and SAT2/KEN/11/60 is highlighted in horizontal grey bars. The block horizontal bar (_____) at the bottom of the alignment indicates a hypervariable region in the sequence alignment between residues 143-154.

```
                                190       200       210       220       230       240       250       260       270       280       290       300       310       320       330       340       350       360
                                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SAT1/KNP/196/91                 VDVLPVEHIIYTRMMIGRFCAQMHSNNGPQIGSAVGCNPDVDWQRFGCHFAQYKNVWDIDYSAFDANHCSDAMNIMFEEVFREEFGFHPNAVWILKTLINTEHAYENKRITVEGGMPSGCSATSIINTILNNIYVLYALRRHYEGVELSHYTMISYGDDIVVASDYDLDFEALKPHFKSL
SAT1/KNP/148/91                 ....................................................................................................................................................................
SAT1/KNP/41/95                  ....................................................................................................................................................................
SAT1/SAR/9/81                   ..........................................................R.........................................................................................................
SAT1/ZIM/HV/03/90               ....................................................................................................................................................................
SAT1/ZIM/GN/13/90               ..................................................................S.................................................................................................
SAT1/BOT/1/68                   ..........................................T.........................................................N.........................................................
SAT1/RHOD/5/66                  ....................................................................................................................................................................
SAT1/RV/11/37                   ....................................................................................................................................................................
SAT1/NAM/307/98                 ..........................................T.........................................................................................................
SAT1/ZIM/6/94                   ..........................................T.........................................................................................................
SAT1/BEC/1/70                   ..........................................T.........................................................................................................
SAT1/SR/2/58                    ......................S...................T.........................................................................................................
SAT1/SWA/1/49                   ..........................................T.........................................................................................................
SAT1/SA/5/61                    ..........................................T.....................................I...................................................................
SAT1/SWA/40/61                  ............................T.........I.....T.........................................................................................................
SAT1/ZIM/25/90                  ..........................................T.........................................................................................................
SAT1/ZAM/2/93                   ..........................................T.........................................................................................................
SAT1/MOZ/3/02                   ..........................................T.....................................................H...................................................
SAT1/TAN/1/99                   ..........................................T.........................................................................................................
SAT1/KEN/05/98                  ..........................................T.........................................................................................................
SAT1/UGA/1/97                   .........L...........N..................T.........V..................P........E......V.....................................................V
SAT1/UGA/3/99                   .........L...........N..................T.........V..................P........E......V.....................................................V
SAT1/SUD/3/76                   .........L.......N......................T.........V..................P........E......V.....................T.S.............F.............V
SAT1/NIG/5/81                   .........L.......N......................T.........V..................P........E......V.....M...............T.S.............F.............V
SAT1/NIG/15/75                  .........L.......K......................T.....R....V..................P........E......VD............V.....................S.S.....................V
SAT1/NIG/8/76                   .........L.......K......................T.....R....V..................P........E......V.....................S.S.....................V
SAT1/NIG/6/76                   .........L.......K......................T.....R....V..................P........E......VD....................S.S.....................V
SAT2/KNP/19/89                  ....................................................................................................................................................................
SAT2/KNP/2/89                   ....................................................................................................................................................................
SAT2/KNP/51/93                  ....................................................................................................................................................................
SAT2/SAR/16/83                  .........L..........................................................................................................................................................
SAT2/ZIM/1/88                   ..........................................T.....R.....................................V.....................................................
SAT2/106/67                     ....................................................................................................................................................................
SAT2/ZIM/7/83                   ..........................................T.........................................................................................................
SAT2/ZIM/14/90                  ..........................................T.........................................................................................................
SAT2/ZIM/17/91                  ..........................................T.........................................................................................................
SAT2/ZIM/GN/10/91               ....................................................................................................................................................................
SAT2/ZIM/34/90                  ..........................................T.........................................................................................................
SAT2/ZIM/8/94                   ....................................................................................................................................................................
SAT2/RHO/1/48                   ..........................................T.....................................T.....G.....................................A.....................
SAT2/ZAM/07/96                  ..........................................T.........................................................................................................
SAT2/KEN/8/99                   ..........................................T.........................................................................................................
SAT2/SEN/5/75                   .........L.................................T.........V.........G..........P........E......V........R.....................................V
SAT2/GHA/8/91                   .........L.................................T.........V.........G..........P........E......V.....................................................V
SAT2/SEN/7/83                   .........L.................................T.........V..................P........E......V.....................................................V
SAT2/ERI/12/98                  .........L.......K..........R.............T.....R....V..................P........E......V...........S.....................T.S.............F.............V
SAT2/SAU/6/00                   .........L.......K.....................Y.T.....R....V..................P........E......V.....................T.S.............H.............I
SAT2/RWA/2/01                   .........L.................................T.....R....V..................P........E......V.....................TS.S.............H.............V
SAT2/KEN/3/57                   .........L.......K.........................T.........V..................P........E......V.....................C.....T.DS.S.....................V
SAT2/KEN/11/60                  .........L.......K.........................T.........V..................P........E......V.....................S......T.DS.S.....................V
SAT2/ANG/4/74                   ......................A....................T..SH.....................................R.....................................................
SAT2/UGA/2/02                   .........L...........N.....................T.........V..................P........E......V.....................................................V
SAT2/UGA/MBF/4/02               .........L....X..K.........................T.........V..................XP.......E......V.....................S.S..F.....................V
SAT2/EGY/9/12                   .........L.......K..........R.............T.....R....V..................P........E......V.....................T.S.............F.....S.......V
SAT2/PAT/1/12                   .........L.......K..........R.............T.....R....V..................P........E......V.....................T.S.............F.............V
SAT3/KNP/10/90                  ....................................................................................................................................................................
SAT3/SA/57/59                   ..................................................................................V.................................................................
SAT3/BEC/1/65                   ..........................................T.........................................................................................................
SAT3/KEN/11/60                  ..........................................T.........................................................................................................
SAT3/BEC/20/61                  ..........................................T.........................................................................................................
SAT3/ZIM/5/91                   ..........................................T.........................................................................................................
SAT3/ZAM/4/96                   ..........................................T.........................................................................................................
A/CIV/4/74                      .........L.......K.........................T.........V..................P........E......V.....................T.S.............F.............I
A/SEN/10/97                     .........VL......K.........................T.....R....L..................P........E......V.....................L.....T.S.............F.............V
A/SOM/1/78                      .........L.................................T.....R....V..................T........E......V.....................DT....................................
A/TAN/4/80                      .........L.................................T.....R....V..................T........E......V.....................DT....................................
A/ETH/2/79                      .........L.................................T.....R....V..................T........E......V.....................DT....................................
A/ETH/7/92                      .........L.................................T.....R....V..................T........E......V.....................DT....................................
A/ERI/3/98                      .........L.......K..........R.............Y.T.........V..................P........E......V.....................T.S.V...........F.............V
O/UGA/5/96                      .........L............L....H.............T.....R....V..................NT.......V........R....................DT....................................
O/UGA/6/76                      .........L.................................V.....R....V..................T........E......V.....................DT....................................
O/UGA/17/98                     .........L.................................T.....R....V..................T........E......V........R....................DT....................................
O/UGA/1/75                      .........L.................................T.....R....V..............L.....T........E......V.....................DT....................................
O/UGA/7/03                      .........L............T....R.............T.....R....V..................T........E......V.....................DT....................................
O/KEN/10/95                     .........L............L....................T.....R....V..................TD.......E......VY....................DT....................................
O/SUD/4/80                      .........L............L....................T.........V..................TD.......E.V....V...........V.........DT....................................
O/ETH/3/96                      .........L.................................T..S..R....V..................TD.......E......V.....................DT....................................
O/TAN/3/96                      .........L.................................T.....R....V..................NT.......E......V.....................DT.....................D..........
```

Supplementary Fig. F (continued)

```
                          370       380       390       400       410       420       430       440       450       460       470
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SAT1/KNP/196/91           GQTITPADKSDKGFVLGQSITDVTFLKRHFHLDYETGFYKPVMASKTLEAILSFARRGTIQEKLISVAGLAVHSGQDEYRRLFEPFQGTFEIPSYRSLYLRWVNAVCGDA
SAT1/KNP/148/91           ..........................................................................................................
SAT1/KNP/41/95            ..........................................................................................................
SAT1/SAR/9/81             ......................X.........G.....................................P.....................................
SAT1/ZIM/HV/03/90         ..........................................................................P.....................................
SAT1/ZIM/GN/13/90         ..........................................................................P.....................................
SAT1/BOT/1/68             ..........................................................................................................
SAT1/RHOD/5/66            ..........................................................................P.....................................
SAT1/RV/11/37             ..........................................................................H.....................................
SAT1/NAM/307/98           ..........................................................................................................
SAT1/ZIM/6/94             ..........................................................................................................
SAT1/BEC/1/70             ..........................................................................................................
SAT1/SR/2/58              ..........................................................................................................
SAT1/SWA/1/49             ..................................................T.......................................................
SAT1/SA/5/61              ..........................................................................................................
SAT1/SWA/40/61            ................................G.........................................................................
SAT1/ZIM/25/90            ..........................................................................................................
SAT1/ZAM/2/93             ..........................................................................................................
SAT1/MOZ/3/02             ..........................................................................................................
SAT1/TAN/1/99             ..........................................................................................................
SAT1/KEN/05/98            ............................Y.............................................................................
SAT1/UGA/1/97             ................................G.....................................P.........T.L.....................
SAT1/UGA/3/99             ................................G.....................................P.........T.L.....................
SAT1/SUD/3/76             ................................FG....................T................P.........T.L.....................
SAT1/NIG/5/81             ................................FG....................................P.........T.L.....................
SAT1/NIG/15/75            ................................G.....................................P.........T.L.....................
SAT1/NIG/8/76             ................................G.....................................P.........T.L.....................
SAT1/NIG/6/76             ................................G.....................................P.........T.L.....................
SAT2/KNP/19/89            ..........................................................................................................
SAT2/KNP/2/89             ..........................................................................................................
SAT2/KNP/51/93            ............................Y.............................................................................
SAT2/SAR/16/83            ..........................................................................G.....................................
SAT2/ZIM/1/88             ................................G.....................................................................
SAT2/106/67               ..........................................................................P.....................................
SAT2/ZIM/7/83             ..........................................................................................................
SAT2/ZIM/14/90            ..........................................................................................................
SAT2/ZIM/17/91            ..................................................Y.......................................................
SAT2/ZIM/GN/10/91         ..........................................................................P.....................................
SAT2/ZIM/34/90            ..A.......................................................................................L.....................
SAT2/ZIM/8/94             ..S.......................................................................P.....................................
SAT2/RHO/1/48             ............................Y.............................................................................
SAT2/ZAM/07/96            ..........................................................................................................
SAT2/KEN/8/99             ..........................................................................................................
SAT2/SEN/5/75             ................................FG....................................P.........A.L.....................
SAT2/GHA/8/91             ................................FG....................................P.........T.L.....................
SAT2/SEN/7/83             ................................FG....................................P.........N.L.....................
SAT2/ERI/12/98            ............................M..G.....................................P.........A.L.....................
SAT2/SAU/6/00             ...........R..................M.....................................P...Q.....A.L.....................
SAT2/RWA/2/01             ................................G.....................................P.........T.L.....................
SAT2/KEN/3/57             ............................M..G.....................................P.............L.....................
SAT2/KEN/11/60            ............................M..G.....................................P.............L.....................
SAT2/ANG/4/74             ............................K...G.....................................P.....................................
SAT2/UGA/2/02             ................................G.....................................P.........T.L.....................
SAT/2/UGA/MBF/4/02        ................................FG..................................X.P.....X......LXX.....................
SAT/2/EGY/9/12            ............................M..G.....................................P...Q.....T.L.....................G.
SAT2/PAT/1/12             ............................M..G.....................................P...Q.....T.L.....................G.
SAT3/KNP/10/90            ..........................................................................................................
SAT3/SA/57/59             ..........................................................................................................
SAT3/BEC/1/65             ..................................................................G.....................................
SAT3/KEN/11/60            ......................S...................................................................................
SAT3/BEC/20/61            ......................S...................................................................................
SAT3/ZIM/5/91             ..................................................Y.......................................................
SAT3/ZAM/4/96             ..........................................................................................................
A/CIV/4/74                ............................M.FG....................................PG.........T.L.....................
A/SEN/10/97               ............................M..G.....................................P.........T.L.....................
A/SOM/1/78                ..............H.............M..G.....................................P.............L.....................
A/TAN/4/80                ..............H.............M..G.....................................P.............L.....................
A/ETH/2/79                ..............H.............M..G.....................................P.............L.....................
A/ETH/7/92                ..............H..A..........M..G.....................................P.............L.....................
A/ERI/3/98                ............................M..G.....................................P...Q.....T.L.....................
O/UGA/5/96                ..A.........................M..G.....................................P.............L.....................
O/UGA/6/76                ..............H.............M..G.....................................P.............L.....................
O/UGA/17/98               ..............H.....P.......V..G.....................................PE............L.....................
O/UGA/1/75                ..............HA............M..G.....................................P.............L.....................
O/UGA/7/03                ..............H.............M..G...S.................................P.............L.....................
O/KEN/10/95               ..............H.............M..G.....................................P.............L.....................
O/SUD/4/80                ..............H.............M..G...........I.............A..........P.............L.....................
O/ETH/3/96                ..............H.............M..G.....................................P.............L.....................
O/TAN/3/96                ..............H.............M..G.....................T...............P.............L.....................
```

Supplementary Fig. F (continued)