Genetic characterisation of six novel African swine fever viruses isolated from a pig, warthog, wild boar, and ticks

By

Sandy Sibusiso Ndlovu

Thesis submitted to the University of Cape Town in fulfilment of the degree Doctor of Philosophy (PhD) Medical Virology

University of Cape Town

Division of Medical Virology

Department of Pathology

Faculty of Health Sciences

University of Cape Town

University of Cape Town

December 2020

Supervisor: Professor Anna-Lise Williamson

Co-supervisor: Dr Olivia Carulei

Co-supervisor: Dr Livio Heath

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

Table of Contents

Declaration vi

Acknowledgements vii

Abstract viii

List of Abbreviations ix

List of Tables xii

List of Figures xiii

CHAPTER 1: LITERATURE REVIEW 1

1.1. AFRICAN SWINE FEVER 1

1.1.1. History and geographical distribution 1

1.1.2. Host and pathogenesis 2

1.1.3. Clinical signs 5

1.1.4. Disease control 5

1.1.5. Socio-economic impact 6

1.2. *ASFARVIRIDAE* (AFRICAN SWINE FEVER VIRUS (ASFV)) 7

1.2.1 Classification and taxonomy 7

1.2.2. Phylogenetic analysis of ASFV isolates 9

1.2.3. ASFV structure 11

1.2.4. ASFV replication and morphogenesis 13

1.2.5. Immunology 14

1.2.5.1. Host immune response to ASFV infection 14

1.2.5.2. Modulation of host defence response 16

1.2.6. Diagnostics 17

1.3. AFRICAN SWINE FEVER VIRUS GENOME 17

1.3.1. Characteristics of the ASFV genome 17

1.3.2. Virulence genes 19

1.4. Vaccine development 20
1.4.1. Inactivated virus vaccines 20
1.4.2. Subunit vaccines 21
1.4.3. Live attenuated vaccines 22
1.5. Challenges of working with ASFV 22
1.6. PROJECT RATIONALE 23
1.7. AIMS 24
CHAPTER 2: THE GENOME SEQUENCES OF AFRICAN SWINE FEVER VIRUS (ASFV) ISOLATES FROM SOUTH AFRICA, ZAMBIA, AND ZAIRE 25
2.1. Introduction 25
2.2. MATERIALS AND METHODS 32
2.2.1. Growth of novel ASFV study isolates in cell-culture and purification of virus DNA 32
2.2.2. Viral DNA extraction and amplification 34
2.2.3. DNA sequencing, analysis, and comparative genomics 34
2.3. RESULTS 35
2.3.1. Sequencing statistics 35
2.3.2. Basic genome structure and statistics 37
2.3.3. Comparative genomics 48
2.4. DISCUSSION 66
CHAPTER 3: PHYLOGENETIC ANALYSIS OF SIX NOVEL AFRICAN SWINE FEVER VIRUSES ISOLATED FROM A DOMESTIC PIG, WARTHOG, WILD BOAR, AND TICKS 70
3.1. Introduction 70
3.2. MATERIALS AND METHODS 74
3.2.1. Methods and sequence sources 74
3.2.2. Phylogenetic tree construction 75
3.3. RESULTS 75

4.4. DISCUSSION 90

CHAPTER 4: COMPARISON OF AFRICAN SWINE FEVER VIRUS (ASFV) MULTIGENE FAMILIES (MGFs) 94

4.1. Introduction 94

4.2. MATERIALS AND METHODS 98

4.3. RESULTS 101

4.4. DISCUSSION 113

5. Conclusions 117

6. Future directions 118

7. References 119

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without acknowledgement of the source. The thesis if to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

Declaration

I, **Sandy Sibusiso Ndlovu**, hereby declare that the work on which this thesis is based is my original work and have not otherwise been submitted in any form for any degree or diploma at any University. The use of information from others has been referenced.

Signature: Date: 28/12/2020

Acknowledgements

I would like to thank my supervisors, Prof Anna-Lise Williamson, Dr Olivia Carulei, and Dr Livio Heath. This project would not have been possible if it was not for their constant involvement, guidance, encouragement, and moral support throughout this project.

Thank you to the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation (NRF), South Africa (#64815) and the Department of the Defense Threat Reduction Agency of the United States Department of Defense (HDTRA1-16-1-500).

I would like to thank my colleagues at the Onderstepoort Veterinary Institute, Agricultural Research Council (ARC-OVI), in the Transboundary Animal disease (TAD) laboratory for the help they provided during my training: Juanita van Heerden, Rivalani Mthombeni and Refiloe Malesa. Lastly, a big thank you to my dearest partner Moriti Lekalakala for the moral support and encouragement.

Abstract

African swine fever (ASF) is a disease that affects domestic pigs and wild boars, resulting in up to 100% case fatality rate, and there is currently no effective treatment or vaccine. To date, there are 67 ASFV complete genome sequences available, but most of the sequences represent only genotypes I-V and VII-X of the 24 genotypes identified based on p72 sequencing, limiting inter and intra-genotype comparative studies. ASFVs encode several multigene families (MGFs) involved in virulence and host range which are found at the genomic termini and the majority of genomic differences between isolates are due to the composition of these MGFs. The comparison of the MGFs across ASFV isolates is of utmost importance in understanding genome variability and their contribution to virulence. The p72 gene has historically been used in phylogenetic analysis of ASFV. However, it lacks the capacity for higher resolution between isolates belonging to the same genotype. This study aimed to analyse and characterise six novel ASFV isolates of African origin from a domestic pig, warthog, wild boar and ticks in terms of genomic makeup, MGF composition and phylogenetic relationships, including identification of additional phylogenetic markers, specifically for use in discrimination between closely related isolates.

Genomes of six novel isolates were sequenced and annotated by identifying open reading frames (ORFs) with a methionine START codon and performing BLASTx searches of each ORF against the NCBI data base. Differences between the genomes were analysed by generating dotplots and using Base-By-Base which showed them to be mostly collinear, but regions of difference were observed at the termini and the CCR. MGF analysis using sorting and clustering in Morpheus software, based on genotype, serogroup, country, host, virulence, and year, showed that genotype and serogroup play a role in the MGF arrangement patterns. Loci corresponding to regions of difference in the CCR were used for phylogenetic comparison to the previously identified marker p72. The tree topology of all of the alternative phylogenies differed from the current p72 classification. B117L and B169L provided slightly better resolution of genotypes I and II, respectively, and viruses from East Africa that are classified as belonging to genotype IX based on p72 were separated when using EP364R. This data adds to the pool of diverse ASFV isolates available for comparative genomics studies, and to the knowledge of ASFV in Africa. The sequencing of more diverse ASFV isolates of each genotype will help characterise the MGFs arrangement patterns among isolates. The novel alternative phylogenetic markers should further be investigated using more ASFV isolates representing the 24 genotypes described to date.

List of Abbreviations

AEC	Animal Ethics Committee
ARC-OVI	Agricultural Research Council Onderstepoort Veterinary Institute
ASF	African swine fever
ASFV	African swine fever virus
ATP	Adenosine triphosphate
bp	Base pairs
BSL-3	Biosecurity level 3
CCR	Central conserved region
CTL	Cytotoxic T-lymphocyte
DNA	Deoxy-ribonucleic acid
dsDNA	Double-stranded DNA
ELISA	Enzyme-linked immunosorbent assay
ER	Endoplasmic reticulum
FAO	Food and Agricultural Organization
g	Gravity
HAD50	50% haemadsorbing dose
HAI	Hemadsorption inhibition assay
IFN	Interferon
IFN-α	Interferon-alpha
Ig	Immunoglobulin
IL	Interleukin
Kbp	Kilobase pairs

LVR	Left variable region
Mbp	Megabase pairs
MGFs	Multigene families
MHC	Major Histocompatibility complex
ml	Millilitre
mRNA	Messenger Ribonucleic acid
MSA	Multiple sequence alignment
NAFT	Nuclear factor activated T-cells
NBS	Normal bovine serum
NCLDVs	Nucleocytoplasmic large DNA viruses
NFκB	Nuclear factor Kappa B
NK	Natural killer cells
nt	Nucleotide
nm	Nanometre
OIE	World Organisation for Animal Health
ORF	Open reading frame
PCR	Polymerase Chain Reaction
RPMI	Roswell Park Memorial Institute medium
qPCR	Quantitative Polymerase Chain Reaction
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
rpm	Revolutions per minute
RSA	Republic of South Africa
RVR	Right variable region

SG	Serogroup
TAD	Transboundary animal diseases
TCID50	50% tissue culture infective dose
TGF	Transforming growth factor
TNF-α	Tumour necrosis factor alpha
Vero	African green monkey kidney cell line
WGS	Whole genome sequencing

List of Tables

Table 2. 1. Complete ASFV genome sequences that are publicly available 30
Table 2. 2. Sequencing and basic genome statistics of six novel ASFV genomes. 36
Table 2. 3. Genome sequence statistics of six ASFV study isolates and publicly available reference sequences of the same/similar genotypes.....44
Table 2. 4. Open reading frames of the six study isolates and the BA71V reference sequence.. 45
Table 2. 5. 109 conserved ORFs in the CCR of the 73 ASFV isolates used in this study.....60
Table 2. 6. Pairwise comparison of the alignment of 109 concatenated conserved ORFs (nt) of six study isolates and reference strains.. 61
Table 2. 7. Fragmented ORFs in the central coding region (CCR) of the LIV_5_40 isolate. .62
Table 2. 8. The tetrameric repeats of the six study isolates and the reference sequences of the same genotype. 65
Table 3. 1. ASFV isolates selected for inclusion in a study comparing ASFV serogroups.....74
Table 4. 1. ASF viruses used in this study, and the representative isolate for some genotypes. 99
Table 4. 2. Comparison of the MGF 100, 110, 300, 360, and 505 members of six novel ASFV isolates.....102

List of Figures

Figure 1. 1. The transmission cycle of the ASF virus.....3
Figure 1. 2. Distribution of African swine fever endemic and non-endemic areas and predominant cycles in endemic areas where known.....4
Figure 1. 3. Phylogenetic tree of five genes that are nearly universal in nucleocytoplasmic large DNA virus (NCLDV).. 8
Figure 1. 4. Geographical distribution of the ASFV p72 genotypes.. 11
Figure 1. 5. ASFV atlas.....12
Figure 1. 6. Genome organisation of the African swine fever virus genome. The organisation of open reading frames (ORFs) on the genome of the virulent ASFV isolate Georgia 2007/1 is shown.. 19

Figure 2. 1. Distribution of the p72 genotypes of ASFV in sub-Saharan Africa. The image was taken from (Mulumba-Mfumu *et al*., 2019). 26
Figure 2. 2. Genome organisation of the ASFV LIV_5_40 isolate. ORFs are shown as arrows to indicate their size and direction of transcription. Fragmented ORFs are highlighted in blue colour, truncated ORFs are highlighted in red colour, commonly used ORFs for phylogenetic analysis are highlighted in purple, and ORFs identified as novel phylogenetic markers are highlighted in green colour. 38
Figure 2. 3. Genome organisation of the ASFV RSA_W1_1999 isolate. ORFs are shown as arrows to indicate their size and direction of transcription. Fragmented ORFs are highlighted in blue colour, commonly used ORFs for phylogenetic analysis are highlighted in purple, and ORFs identified as novel phylogenetic markers are highlighted in green. 39
Figure 2. 4. Genome organisation of the ASFV RSA_2_2004 isolate. ORFs are shown as arrows to indicate their size and direction of transcription. Fragmented ORFs are highlighted in blue colour, commonly used ORFs for phylogenetic analysis are highlighted in purple, and ORFs identified as novel phylogenetic markers are highlighted in green colour.....40
Figure 2. 5. Genome organisation of the ASFV RSA_2_2008 isolate. ORFs are shown as arrows to indicate their size and direction of transcription. Fragmented ORFs are highlighted in blue colour, commonly used ORFs for phylogenetic analysis are highlighted in purple, and ORFs identified as novel phylogenetic markers are highlighted in green colour.....41
Figure 2. 6. Genome organisation of the ASFV SPEC_57 isolate. ORFs are shown as arrows to indicate their size and direction of transcription. Fragmented ORFs are highlighted in red

colour, commonly used ORFs for phylogenetic analysis are highlighted in purple, and ORFs identified as novel phylogenetic markers are highlighted in green colour. 42
Figure 2. 7. Genome organisation of the ASFV Zaire isolate. ORFs are shown as arrows to indicate their size and direction of transcription. Fragmented ORFs are highlighted in blue colour, commonly used ORFs for phylogenetic analysis are highlighted in purple, and ORFs identified as novel phylogenetic markers are highlighted in green colour. 43
Figure 2. 8. Dot plots of the study isolates LIV_5_40 and Zaire genomes (x-axis) vs. reference genome sequences (y-axis). A) Mkuzi_1979 vs. LIV_5_40 B) BA71V vs Zaire. Major and minor deletions can be seen as breaks in the diagonal line, highlighted with red arrows.....50
Figure 2. 9. Dot plots of the study isolates SPEC_57 and RSA_W1_1999 genomes (x-axis) vs. reference genome sequences (y-axis). C) Warmbaths vs. SPEC_57 D) Warthog vs. RSA_W1_1999. Major and minor deletions can be seen as breaks in the diagonal line, highlighted with red arrows. 51
Figure 2. 10. Dot plots of the study isolates RSA_2_2004 and Zaire genome (x-axis) vs. reference genome sequences (y-axis). E) Pretoriuskop_96_4 vs. RSA_2_2004 F) Pretoriuskop_96_4 vs. Zaire. Major and minor deletions can be seen as breaks in the diagonal line, highlighted with red arrows. 52
Figure 2. 11. Dot plots of the study isolate RSA_2_2008 genome (x-axis) vs. reference genome sequences (y-axis). E) Pretoriuskop_96_4 vs. RSA_2_2008 F) RSA_2_2008 vs. RSA_2_2004. Major and minor deletions can be seen as breaks in the diagonal line, highlighted with red arrows. 53
Figure 2. 12. Visual summary of the alignment of the 73 ASFV isolate used in this study. The top greyscale track indicates similarity (black-perfect or 100 match; white-low or 0% match), and gaps are shown as dashes. 54
Figure 2. 13. Visual summary of the alignment of the six novel study isolate genomes vs. reference sequences of the same genotype. The red arrows indicate the location of the study isolates. The top greyscale track indicates similarity (black-perfect or 100 match; white-low or 0% match), and gaps are shown as dashes. The track showing pink and purple boxes indicates top and bottom strand open reading frames, respectively. 55
Figure 2. 14. Visual summary of the alignment of the Zaire study isolate genome vs. Ba71V, NHV, and OURT_88_3. The top greyscale track indicates similarity (black-perfect or 100 match; white-low or 0% match), and gaps are shown as dashes. The track showing pink and purple boxes indicates top and bottom strand open reading frames, respectively. 56

Figure 2. 15. Visual summary of the amino acid alignment of the EP153R protein encoded by the LIV_5_40 study isolate vs. NHV and OURT_88_3. 59
Figure 2. 16. Amino acid alignment of the proline-rich repeat region of the EP402R gene. a) Scale representing the length of the EP402R (CD2V) protein with the aligned region highlighted, b) Amino acid alignment of the proline-rich repeat region of the EP402R (CD2v) gene of 73 ASFV isolates analysed in this study. Red arrows indicate the location of the study isolates, residues in grey are conserved across all sequences, and coloured residues are different in at least one sequence. 64
Figure 3. 1. Phylogenetic analysis of a causative virus strain (China 2018/1) of an African swine fever outbreak, China, 2018.. 72
Figure 3. 2. Neighbour-Joining tree depicting p72 gene relationships of 36 African swine fever virus isolates, including six novel sequences described in this study (blue text).. 77
Figure 3. 3. Neighbour-Joining tree depicting B117L genes relationship of 36 African swine fever virus isolates, including six novel sequences described in this study (blue text).. 79
Figure 3. 4. Neighbour-Joining tree depicting B169L gene relationships of 36 African swine fever virus isolates, including six novel sequences described in this study (blue text).. 80
Figure 3. 5. Neighbour-Joining tree depicting B407L gene relationships of 36 African swine fever virus isolates, including six novel sequences described in this study (blue text).. 81
Figure 3. 6. Neighbour-Joining tree depicting I177L gene relationships of 36 African swine fever virus isolates, including six novel sequences described in this study (blue text).. 82
Figure 3. 7. Neighbour-Joining tree depicting EP364R gene relationships of 36 African swine fever virus isolates, including six novel sequences described in this study (blue text).. 84
Figure 3. 8. Neighbour-Joining tree depicting fivel novel concatenated genes from 36 African swine fever virus isolates, including six novel sequences described in this study (blue text).. 85
Figure 3. 9. Unrooted maximum-likelihood phylogenetic tree constructed from the alignment of the 109 core concatenated conserved set of genes from 36 ASFV isolates, including six novel sequences described in this study (blue text).. 86
Figure 3. 10. Unrooted maximum-likelihood phylogenetic tree constructed from the alignment of the EP402R gene.. 88
Figure 3. 11. Unrooted maximum-likelihood phylogenetic tree constructed from the alignment of the EP153R gene.. 89
Figure 4. 1. Display of all MGF paralogs as they appear, left to right, in the 16 ASFV strains as described by Imbery and Upton 2017 96

Figure 4. 2. Amino acid alignment of the MGF 110-6L gene encoding the ER retention motif sequence (KEDL), of 28 ASFV isolates. 104
Figure 4. 3. Amino acid alignment of the MGF 110-4L gene encoding the ER retention motif sequence (KEDL), of 30 ASFV isolates. 105
Figure 4. 4. Sorting of the MGFs by genotype in Morpheus. 108
Figure 4. 5. Sorting of the MGFs by serogroup in Morpheus. 109
Figure 4. 6. Hierarchical clustering of the MGFs. 110
Figure 4. 7. Hierarchical clustering of the MGFs in (**A**) European genotype I strains, (**B**) European genotype II strains, and (**C**) Southern African strains 112

CHAPTER 1: LITERATURE REVIEW

1.1. AFRICAN SWINE FEVER

1.1.1. History and geographical distribution

Montgomery first described African swine fever (ASF) in Kenya in the 20th century (Montgomery, 1921). However, the virus is thought to have existed long before it was officially reported in Africa. Since there are no apparent clinical signs in its natural host, the disease was only observed when imported, and susceptible, domesticated pigs were introduced. Mortalities started to occur after missionaries from Portugal and colonists introduced domestic pigs in sub-Saharan Africa five centuries ago (Penrith *et al*., 2004). In 1926, the first ASF outbreak was recorded in the Limpopo Province, which is thought to have resulted from the contact of wild (warthog) and domestic suids. In 1933, infected animals that were moved from the Transvaal province into the Western Cape caused two outbreaks. Angola reported the first outbreak in 1932, and the disease was first observed in Mozambique in 1954. The disease reached west Africa in the 1950s; the origin of the virus in the Iberian Peninsula from west Africa was recorded in 1957 and 1960 and subsequently moved to several European and Latin countries (Dixon *et al*., 2019b; Penrith *et al*., 2004).

ASF reappeared in 1971, where it threatened Central America, and Cuba was the first country to be affected (Penrith *et al*., 2004). In 1978, the disease severely affected pork industries in Brazil, Haiti, and the Dominican Republic (Penrith *et al*., 2004) but these countries were able to eradicate the disease by 1984 through the implementation of strict control measures. ASFV was eradicated from European countries in 1990. However, in Sardinia, the disease continues to be endemic since its introduction in 1978. After a period of 30 years, during which no outbreaks were reported, the disease reappeared in Kenya in 1994, and in 1997/98, it occurred in Madagascar for the first time (Penrith *et al*., 2004).

In 2007, the Caucasus region was severely affected, and the disease is thought to have remained in Armenia, Georgia, and South Russia (Rowlands *et al*., 2008). In 2014, the disease occurred in Estonia, Latvia, Lithuania, and Poland. Outbreaks of ASF were observed in Bulgaria, Czech Republic, Hungary, and Romania in 2018 (Dixon *et al*., 2019b). The first outbreak report of ASF in China was in the Shenyang City, Liaoning region in August 2018. In 2019, subsequent spread to Mongolia, Vietnam, and North Korea was observed. The cause

of these outbreaks includes interprovincial transportation of infected live pigs and pork products (Dixon *et al*., 2019b; Sur, 2019). In the Republic of Korea, the first outbreak was reported in September 2019 on a commercial farm with 2450 pigs (Kim *et al*., 2020a). More outbreaks occurred in the same area; however, the highest number of outbreaks was reported in January 2020 (Jo and Gortázar, 2020). Currently, 1013 outbreaks are reported to be on-going in the country (Disease information-Immediate notification 2021). In July 2019 the Philippines reported their first outbreak in the Rozal Province. The outbreak resulted in the loss of 525 pigs, and a total of 7416 pigs were culled. Outbreaks that occurred in January 2020 and February 2020 resulted in the loss of 1000 pigs and 2261 pigs, respectively (Penrith, 2020). At the moment, a total of 171 new outbreaks in backyard farms have been reported (Disease information-Immediate notification 2021). The first outbreak in Malaysia occurred in February 2020 and they are planning to cull a total of 3000 pigs (Disease information-Immediate notification 2021). In Indonesia, the first outbreak was reported in September 2019, and a total of 38123 pigs died (Penrith 2020). Further outbreaks that occurred from February 2021 resulted in the loss of 35000 pigs (Disease information-Immediate notification 2021). Timor-Leste reported their first outbreak in September 2019, and 405 pigs died (Penrith, 2020). Currently, a total of 126 outbreaks are currently on-going in the country (Disease information-Immediate notification 2021).

1.1.2. Host and pathogenesis

The African swine fever virus (ASFV) is maintained in the sylvatic cycle in east and southern Africa, which involves warthogs (*Phacochoerus africanus*) and soft ticks of the *Ornithodoors moubata* complex (Jori and Bastos, 2009). The virus is transmitted through a domestic cycle involving soft ticks, and a domestic cycle without soft ticks playing a role (Figure 1.1). The virus is also transmitted trans-stadially amongst ticks (Jori *et al*., 2013; Service, 1980). Soft ticks are found in warthog burrows, and young warthogs spend the first 4-6 weeks of their life in burrows and are infected when ticks are taking blood meals during this period. Ticks can be transported from burrows to farmed areas by adult warthogs, and this pattern of movement results in the start of the tick and domestic pig cycle (Jori *et al*., 2013). Figure 1.2 shows the distribution of the three cycles according to predominance, geographical areas, and the year of the reported outbreak (Penrith *et al.*, 2019). However, the last reported outbreak was in South Africa, in the Eastern Cape Province, in October 2020.

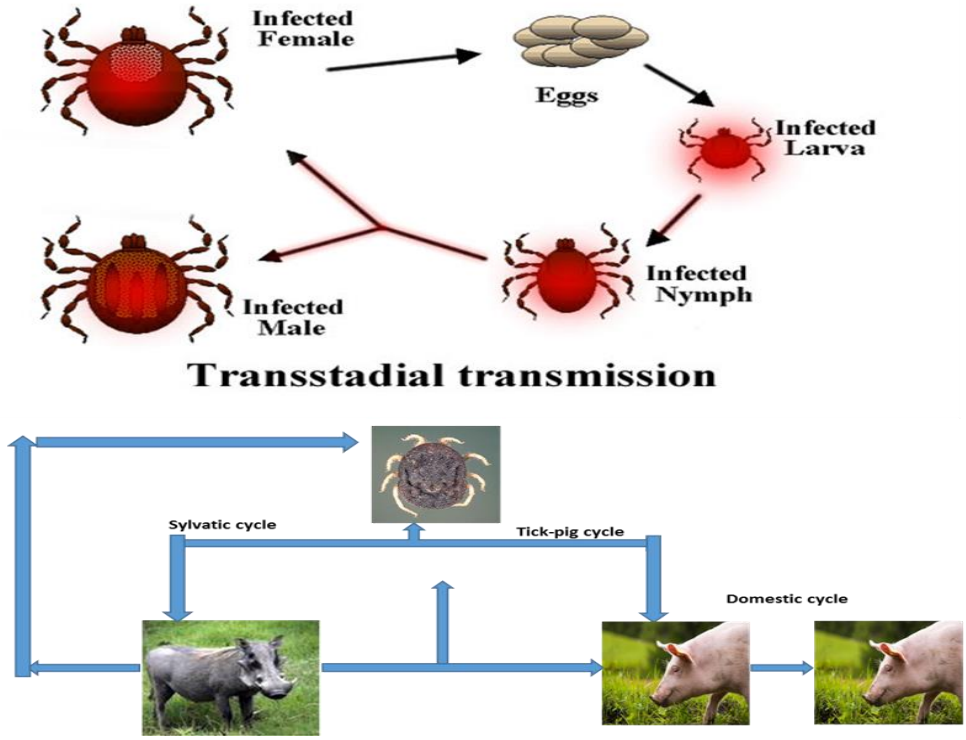


Figure 1. 1. The transmission cycle of the ASF virus. Soft ticks are reservoirs of ASFV, and the virus remains with the tick from one life stage to the next (trans-stadial). In sub-Saharan Africa, the virus is maintained in the sylvatic cycle, which involves ticks and warthogs. The virus is also transmitted in the domestic cycle with or without tick involvement. Adapted from (Service, 1980).

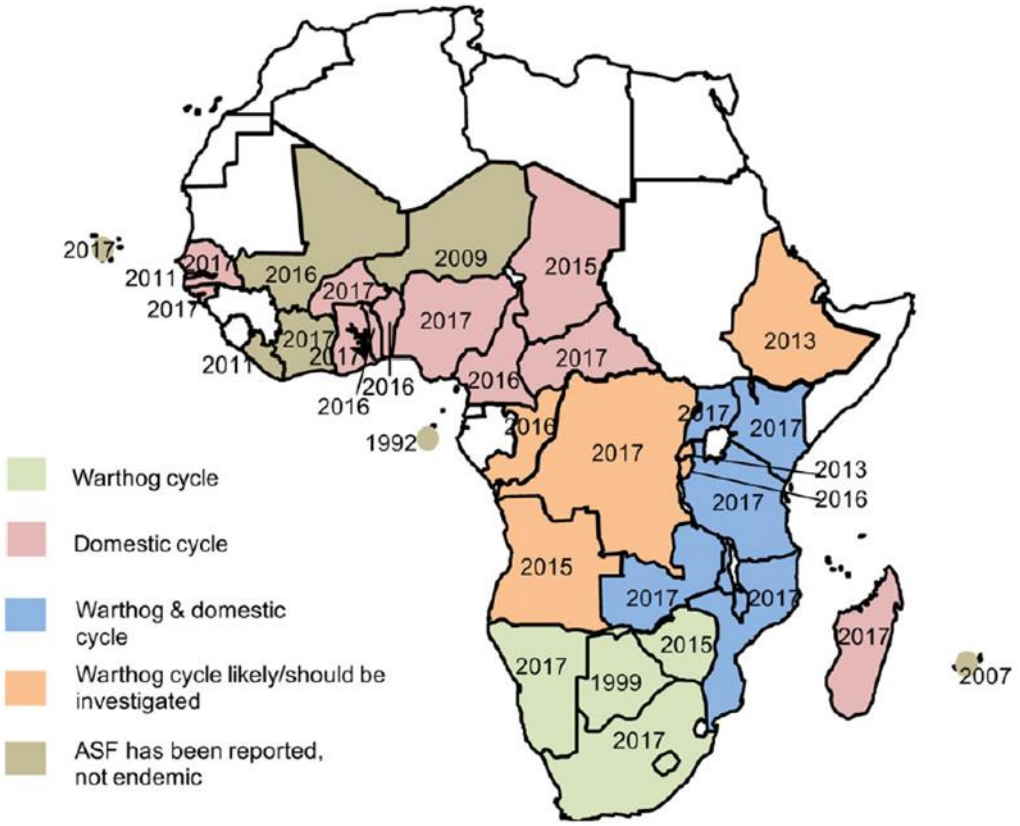


Figure 1. 2. Distribution of African swine fever endemic and non-endemic areas and predominant cycles in endemic areas where known. The date of the last reported outbreak for each country is provided. The image was taken from (Penrith *et al*., 2019).

Initial virus replication in soft ticks occurs in the midgut epithelium, and subsequent replication takes place in the salivary gland, coxal gland, and reproductive tissue. It takes a period of 2 to 3 weeks for the virus infection to move from the midgut to other tick tissues (Kleiboeker *et al*., 1998). Subclinical infection is observed in young warthogs, and the viraemic titres in young warthogs were found to range from 10^2 to 10^3 50% haemadsorbing dose per millilitre (HAD50)/ml. Adult warthogs are non-viraemic, but most seroconvert following infection with the virus restricted to the lymph nodes. The sylvatic cycle is mostly maintained by young warthogs between 3-4 weeks of age, which transmit the virus to naïve ticks (Jori and Bastos, 2009; Kleiboeker *et al*., 1998).

Previous studies have shown that during pig-to-pig transmission, the titres can reach up to10^9 HAD50/ml in blood, and up to 10^5 HAD50/ml in saliva and faeces following infection with high virulence strains circulating in Georgia (2007), Russia (2013), and Lithuania (2014) (Gallardo *et al*., 2017; Guinat *et al*., 2016). However, similar values were recorded following infections of pigs with moderately virulent strains which were circulating in the Netherlands, Portugal, and Malta in 1986, 1968, and 1978, respectively (de Carvalho Ferreira *et al*., 2012;

Gallardo *et al*., 2015; Wilkinson *et al*., 1981). This could be a result of other pathogenic factors. Experimental studies have shown that when wild boars (*Sus scrofa ferus*) are infected with high virulence ASFV isolates, they present with high viral titres and similar clinical symptoms to that observed in domestic pigs (*Sus scrofa domesticus*). Direct contact of infected wild boars with domestic pigs results in domestic pigs getting infected, and the mortality rate also approaches 100% (Gabriel *et al*., 2011; Pietschmann *et al*., 2015). The transmission of the ASFV from the sylvatic cycle to domestic pigs is achieved through ticks that drop off adult warthogs in close proximity to domestic pigs, thus, infecting domestic pigs when feeding and initiating a tick-domestic pig cycle (Jori *et al*., 2013). Convalescent animals can transmit the ASFV to naïve domestic pigs through excretions and secretions, but for a limited time (Bellini *et al*., 2016).

1.1.3. Clinical signs

ASFV in soft ticks and warthogs results in persistent infection, and clinical signs are not apparent. However, varying clinical signs are recorded in domestic pigs and wild boars, which present as hyperacute, acute, or chronic disease, depending on the virulence of the ASFV strain. Highly virulent isolates cause hyperacute clinical signs, and pigs die between 2-4 days post-infection, with no apparent clinical signs other than fever observed. Characteristics of acute infections include fever, leucopenia, and thrombocytopenia. Reddening of the skin is observed in white pigs on tips of the ear, tail, chest, and abdomen. Vomiting, diarrhoea, and eye discharge are observed, and death may occur within 6-13 or up to 20 days post-infection (Salguero, 2020).

ASFV isolates with moderate virulence cause subacute forms of the disease, which are characterised by intermittent fever for up to a month and followed by recovery in most cases. Anorexia, coughing, dyspnoea, and abortion in pregnant sows are recorded, and death can occur between 15-45 days after infection. Chronic infection is characterised by loss of weight, irregular peaks of temperature, and respiratory signs, but these signs may not be apparent or absent in some cases. However, these animals are vulnerable to secondary infection (Blome *et al*., 2012; Dixon *et al*., 2019a; Galindo-Cardiel *et al*., 2013; Salguero, 2020).

1.1.4. Disease control

ASF is endemic in sub-Saharan Africa, and there is currently no vaccine. Thus, the biosafety approach is the most effective option to reduce the risk of outbreaks. During outbreaks,

control measures include surveillance and monitoring programmes that are employed as a countermeasure to try and eliminate the disease from the source and to contain the disease. ASF is listed as a notifiable disease by the World Organisation for Animal Health (OIE) and the Food and Agricultural Organization (FAO) (Domenech *et al*., 2006). However, different approaches have to be applied to detect various clinical manifestations. Surveillance activities during acute infection can be based on clinical signs; however, during chronic infections, where clinical signs are not apparent, diagnostic testing should supplement surveillance measures. An epidemiological investigation based on sequencing the C-terminus of the p72 gene is performed to classify the ASFV isolate that caused the outbreak. These methods are combined with strict quarantine and biosafety measures that include movement restriction of pigs and culling of all infected herds. Carcasses must be appropriately disposed of through incineration, and all premises at the site of the infection must be thoroughly disinfected (Dixon *et al*., 2019b; Oura, 2013; Penrith, 2009; Penrith *et al*., 2013).

1.1.5. Socio-economic impact

ASF outbreaks result in a significant socio-economic impact in developing and developed countries. Small-scale farmers are significantly affected because they lack financial support to put in place proper control measures and face challenges to start production again after outbreaks. Initial outbreaks between 1933 and 1934 in South Africa affected 11, 000 animals, and only 862 survived (De Cock *et al*., 1940). The introduction of this disease in Cote d'Ivoire and Madagascar led to a loss of approximately 50% of the pig population (Roger *et al*., 2001). The emergency eradication measures that were implemented in Cuba in 1980 resulted in a total cost of US $9.4 million (Costard *et al*., 2009) and the estimated cost of eradication in Spain was US $92 million. In 1994, it was estimated that the implementation of strict control measures in preventing the introduction of ASF resulted in the USA spending US $450 million (Costard *et al*., 2009).

In 2011, in Russia, ASF outbreaks resulted in an estimated economic loss of US $267 million (Sanchez-Cordon *et al*., 2018b). Disease outbreaks that occurred in Estonia, Lithuania, Latvia, and Poland between 2014 and 2015 reduced the export value by US $961 million. Investigations show that should Denmark experience ASF outbreaks, it will have a US $349 million loss in exports (Sanchez-Cordon *et al*., 2018b). As a result of outbreaks since August 2018, China has culled about 5 million pigs, which is equivalent to 1% of their total pig population (Yun, 2020). The World Organisation for Animal Health (OIE) reported that 33 African countries experienced ASF outbreaks by the end of 2017 (Achenbach *et al*., 2017).

Recent ASF outbreaks have significantly affected the global pork market. It was estimated that by mid-2019 about 150-200 million pigs were infected by the ASFV, which is about 30% of Chinese pig population (Mason-D'Croze *et al*., 2020). The impact of the ASF epidemics on pork prices were assessed using two linked economic models and it was demonstrated that the price of pork increased by 17-85%. The rise in pork prices resulted in the decline in pork demand globally. However, much of the decline in pork demand happened outside of China, and greatly affecting Europe, while notable declines were observed in North America, Latin America, and the rest of East Asia and the Pacific (Mason-D'Croze *et al*., 2020).

1.2. *ASFARVIRIDAE* (AFRICAN SWINE FEVER VIRUS (ASFV))

1.2.1 Classification and taxonomy

ASFV is a member of the nucleo-cytoplasmic large DNA viruses (NCLDVs), which are comprised of a large double-stranded DNA (dsDNA) genome ranging in length from 100 kilobase pairs (kbp) in *Iridoviridae* up to 2.5 megabase pairs (Mbp) in Pandoraviridae (Abergel *et al*., 2015; Gallot-Lavallee and Blanc, 2017). The NCLDVs are classified as a group of monophyletic viruses (Figure 1.3). Phylogenetic analysis revealed that these viruses share a set of 40 core genes that play a role in virion structure, assembly, and replication, thus supporting the monophyletic origin of the NCLDVs from a common virus ancestor (Koonin and Yutin, 2018). The NCLDVs are classified into eight taxonomic families, *Ascoviridae*, *Asfarviridae*, *Iridoviridae*, *Marseilleviridae*, *Mimiviridae*, *Pithoviridae*, *Phycodnaviridae*, and *Poxviridae* (Freitas *et al*., 2019; Gallot-Lavallee and Blanc, 2017; Koonin and Yutin, 2018). Initially, ASFV was classified as a member of the *Iridoviridae* family until 1988 (Bastos *et al*., 2003). The ASFV is of the genus *Asfivirus* and classified into the *Asfarviridae* family (Alonso *et al*., 2018). Amendments to the 2019 taxonomy release of the International Committee on Taxonomy of Viruses (EC 51, Berlin, Germany, July 2019), included the *Asfarviridae* into the class *Pokkesviricetes*, and order *Asfuvirales*. ASFV is the sole member of the *Asfarviridae*, and is also the only known DNA arbovirus. Partial sequencing of the Heterocapsa circularis DNA virus (HcDNAV) revealed that it is closely related to the ASFV, and it was further shown that recently isolated viruses, faustovirus, kaumoebavirus, and pacmanvirus form a new group that is closely related to the *Asfarviridae* family. A phylogenetic tree generated using the DNA polymerase gene reveals that ASFV, faustoviruses, kaumoebavirus, and pacmanvirus have a common ancestor. However, the divergence of the kaumoebavirus suggests that it preceded the other three viral groups, and it

was further shown that the pacmanvirus would form a new viral group (Andreani *et al*., 2017; Bajrai *et al*., 2016; Ogata *et al*., 2009; Reteno *et al*., 2015).

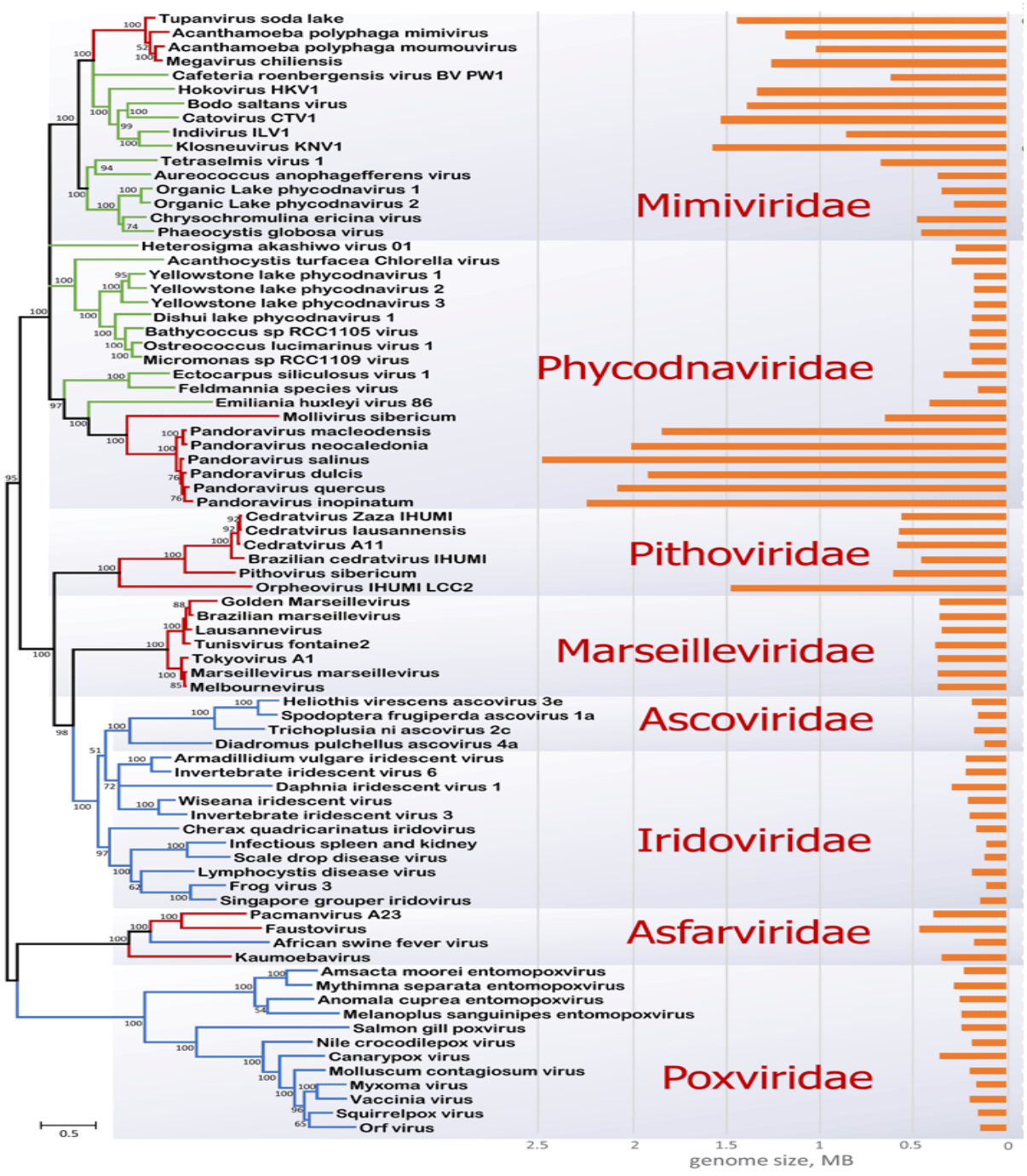


Figure 1. 3. Phylogenetic tree of five genes that are nearly universal in nucleocytoplasmic large DNA virus (NCLDV). The tree was constructed from a concatenated multiple sequence alignment of five (nearly) universally conserved NCLDV proteins: DNA polymerase, major capsid protein, packaging ATPase, A18-like helicase, and poxvirus late transcription factor VLTF3. The branch colour denotes confirmed or likely hosts: red, Amoebozoa; green, other protists; blue, Metazoa. The tree was constructed by using the FastTree software60 with default parameters. The image was adapted from (Koonin & Yutin, 2018).

1.2.2. Phylogenetic analysis of ASFV isolates

Early studies on the origin of ASFV outbreaks and the disease outbreak patterns at the molecular level were based on the restriction fragment length polymorphism (RFLP) approach. This method is based on the digestion of the viral genome into pieces of various sizes by a restriction enzyme, followed by electrophoresis to determine the size of the DNA fragments (Blasco *et al*., 1989b; Dixon and Wilkinson, 1988; Sumption *et al*., 1990). The RFLP method identified eight variable regions in ASF viruses isolated in Malawi and Zambia (Sumption *et al*., 1990). The *BamHI* restriction enzyme was the preferred enzyme because it gave better results when determining the relatedness of diverse ASFV isolates, and the results were also reproducible. This enzyme also produced more fragments, which assisted in establishing relationships among closely related strains (Sumption *et al*., 1990). However, this approach required considerable time, did not provide good discrimination and was not the best method to use to characterise ASFV isolates during outbreaks (Costard *et al*., 2009).

Polymerase chain reaction (PCR)-sequencing methods are the most suitable methods implemented for molecular epidemiology as they are faster and more accurate in identifying major genotypes. Subsequent studies amplified the viral DNA by PCR, the PCR products were sequenced, and phylogenetic trees were generated to classify the ASFV isolates. Genes within the highly conserved region of the genome that showed some sequence variability were identified and used to cluster ASFV isolates into different genotypes. Initial studies used the C-terminus of the B646L gene (which encodes the p72 protein) to genotype ASFV isolates. The p72 protein is the major component of the viral capsid, making up 32% of the total protein mass of the virion, and it surrounds the inner envelope of the virus (Cobbold and Wileman, 1998). Study by Sanz *et al*., (1985), demonstrated that other viral proteins change when the virus is passaged in macrophages, however, the p72 protein remains unchanged. During natural infections, the p72 protein is the major immunogen, and it has since been used in diagnostic tests (Tabares *et al*., 1980b). The antigenic stability of p72 was further confirmed by comparing the full length of the B646L genes from the Dominican Republic (DR-2) isolate, two Spanish Isolates (BA71V and E70), and one isolate from Uganda (UGA). It was found that at the nucleotide level, there is 95.5% to 100% sequence similarity, and 97.8% to 100% similarity at the amino acid level (Yu *et al*., 1996). In a study by Bastos *et al*., (2003), a nucleotide and amino variability plot of the central and C-terminus regions of the B646L gene showed the C-terminus was rich with synonymous and non-synonymous mutations. It was observed that, at a nucleotide level, 59 of the total 415 sites corresponded to

variable sites, and 12 of the 59 sites resulted in non-synonymous amino acid substitutions. The C-terminus was found to be surrounded by conserved nucleotide sequences that were used to design oligonucleotide primers that are complementary to the flanking sites. Thus, the C-terminus was chosen for the phylogenetic analysis of ASFV isolates instead of the central region and demonstrated that Eastern and Southern African strains segregated into 21 genotypes, while East European isolates clustered in genotype II (Bastos *et al*., 2003). It has been shown that ASFV strains from the Caribbean (Genotype I), Europe (Genotype I and II), South America (Genotype I), and Central and West Africa are closely related to each other.

Viruses in South and East Africa are more diverse (Figure 1.4) (Bastos *et al*., 2003; Boshoff *et al*., 2007; Lubisi *et al*., 2005). This observation could be due to the maintenance of the virus within the sylvatic cycle (Bastos *et al*., 2009; Penrith *et al*., 2013). The virus was first reported in East Africa (Kenya) in 1921 (Montgomery, 1921), and it was reported in North of South Africa around 1926 (Steyn, 1932). The virus is able to survive in ticks for many years, while the tick is still waiting for host feeding. The sylvatic cycle can indefinitely maintain the virus circulation and also the persistence of ancient and emergence of new variants. The spill over of viruses from the sylvatic cycle to the domestic pigs is very rare. Reflecting the long term evolution and selective constrains imposed by replication, in and transmission between these hosts (Dixon *et al*., 2013). One of the gaps is the lack of studies on homologous recombination in ASFV at a genetic level, and the role of recombination on the genetic diversity and evolution of the ASFV is still unclear (Zhu *et al*., 2019). To date, 24 genotypes of ASFV have been identified through sequencing of the C-terminus of the B646L gene (Achenbach *et al*., 2017; Bastos *et al*., 2003; Boshoff *et al*., 2007; Lubisi *et al*., 2005; Quembo *et al*., 2018). It was shown that the phylogenetic tree generated with the p72 gene formed four major lineages (Michaud *et al*., 2013). Lineage 1 consisted of previously described genotypes I, II, XVII and XVIII, while lineage 2 was comprised of isolates representing genotypes III, IV, V, VI, VII, XIX, XX, XXI, and XXII. Lineage 3 was formed by isolates corresponding to genotypes VIII, XI, XII, XIII, XV, and XVI, and lineage 4 included genotypes IX and X. However, B646L is too genetically conserved to discriminate ASFV isolates within the same genotype. Other genes were subsequently sequenced to get better phylogenetic resolution between closely related isolates. This includes the B602L, CP204L (p30), E183L (p54), EP402R (CD2v), and the intergenic region between I73R and I329L genes (Gallardo *et al*., 2009; Nix *et al*., 2006; Rowlands *et al*., 2008; Sanna *et al*., 2017).

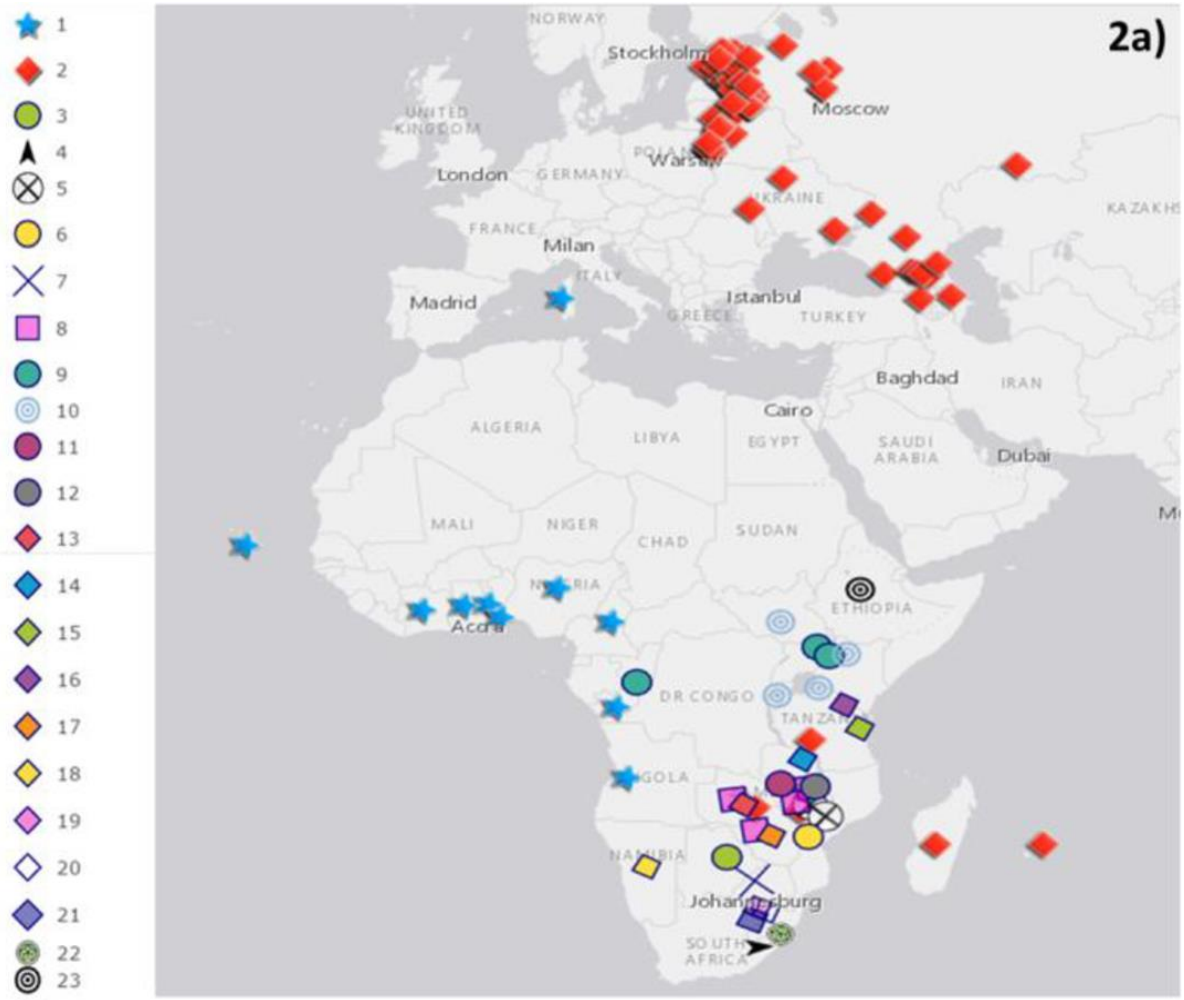


Figure 1. 4. Geographical distribution of the ASFV p72 genotypes. The image was taken from (Arias *et al.*, 2017).

1.2.3. ASFV structure

The virus particle is comprised of a multi-layered structure that includes the outer envelope, cap, inner envelope, core-shell, and the nucleoid (Figure 1.5). The virus particle on average is 260 to 300 nanometres (nm) in diameter, which is larger than what it was previously observed which was 200 nm (Wang *et al*., 2019). The outer envelope is acquired through budding (Rouiller *et al*., 1998), and the EP402R gene that encodes for the CD2v protein involved in haemadsorption localises at the outer envelope (Ruiz-Gonzalvo *et al*., 1996). The capsid is formed by capsomers, which appear as hexagonal prisms with a length of 13 nanometres (nm) (Rouiller *et al*., 1998). The capsomers were further shown to be formed of homotrimers of the p72 protein, possessing a double jelly-roll fold (Andrés *et al*., 2020). The three-dimensional reconstruction of the capsid structure revealed that, it is comprised of the major capsid protein (p72) and four minor proteins M1249L, p17, p49, and H240R, which are

organized into pentasymmetrons and trisymmetrons (Wang *et al*., 2019). The p72 protein encoded by the B646L gene, the C terminus region of which has been used extensively in ASFV genotyping. It was found that p72 and p17 provide 35% of the mass of the virion, and 25% by structural proteins derived from the polyprotein vp220 (vp150, p37, p34, and p14) (Alcami *et al*., 1992; Cobbold and Wileman, 1998). The inner envelope of the virus is thought to be acquired from the endoplasmic reticulum (ER), but little is known about the mechanism (Andres *et al*., 1998). The p54 protein is one of the capsomers forming the inner envelope, which is encoded by the E183L gene. This gene has been used in conjunction with other ASFV genes to give better resolution in phylogenetic analysis studies (Gallardo *et al*., 2009). The core-shell is a thick protein of about 30 nm and contains mature products encoded as polyprotein precursors (pp220 and pp62), which is not observed in other large DNA viruses (Andres *et al*., 1997; Carrascosa *et al*., 1984). The nucleoid is an 80 nm structure that contains the viral genome and also proteins involved in nucleotide metabolism, transcription, replication, and repair (Alonso *et al*., 2018; Salas and Andres, 2013).

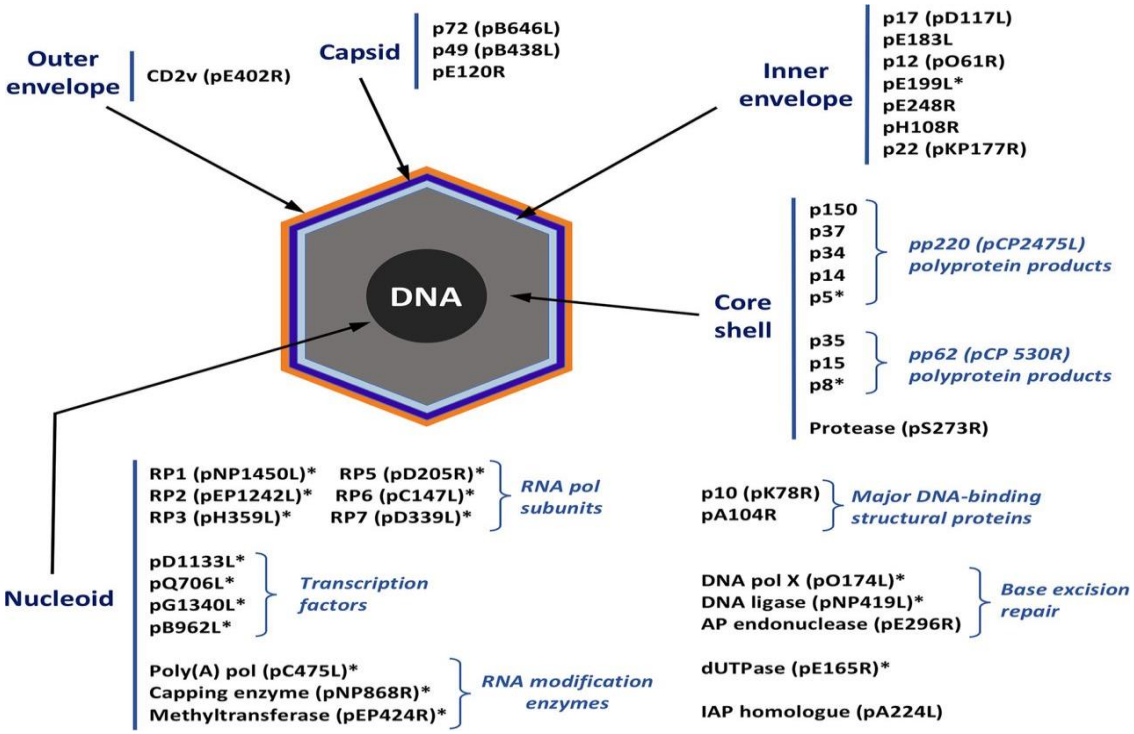


Figure 1. 5. ASFV atlas. The subviral localisation of 40 viral proteins among the five structural domains of the ASFV particle is shown. The distribution of proteins marked with an asterisk was inferred from the predicted or known role, while that of the remaining proteins was determined by immunoelectron microscopy. pol, polymerase. The image was taken from (Alejo *et al*., 2018).

1.2.4. ASFV replication and morphogenesis

ASFV has a preferential tropism for monocytes and macrophages, and these cells are components of the innate immunity. Monocytes are white blood cells that circulate in the blood and spleen. In response to infection, these cells defend the host from pathogens through phagocytosis, antigen presentation, and secretion of chemokines. Macrophages respond to infection by secreting chemokines and recruiting other immune cells, migrating to local lymph nodes, and presenting processed antigens (Dixon *et al*., 2004). Although macrophages remain the preferred cells for replication both *in vitro* and *in vivo*, ASFV infects dendritic cells, endothelial cells, fibroblasts, and reticular cells. However, these cells are infected in the late stages of infection, after the onset of apparent ASF symptoms, and thus, have a limited role in the disease pathology (Abrams *et al*., 2008). Micro-organisms, including viruses, have developed mechanisms to successfully replicate and maintain infection in these cells, regardless of their innate defence mechanisms (Basta *et al*., 2010).

The ASFV entry into the host cell involves a process known as receptor-mediated endocytosis and this mechanism is regulated by energy and temperature. Post-internalisation, the virus particle loses the outer envelope and the inner envelope fuses with the endosome under acidic conditions, releasing the viral core into the cell cytoplasm of the infected host (Alonso *et al.*, 2013). The ASFV can also gain entry into the host through a mechanism known as dynamin-dependent and clathrin-mediated endocytosis (Hernaez and Alonso, 2010). Immediately after gaining entry to the host cell, early genes are transcribed using the RNA-polymerase incorporated in the virion. Subsequently, the messenger ribonucleic acids (mRNA) are polyadenylated and capped by other enzymes packaged into the virus core. There are mRNAs that are transcribed at early or intermediate stages; however, they remain silent until the onset of replication (Rodriguez and Salas, 2013). Morphogenesis begins by the formation of the inner envelope, which results from the accumulation of viral membranes identified as small curved and open structures that become precursors for the inner envelope of the ASFV particle. The capsid is assembled through a mechanism that is adenosine triphosphate (ATP) and calcium-dependent on the convex face of the viral membrane (Cobbold *et al*., 2000), and the core-shell is assembled underneath the concave face of the viral envelope.

The nucleoid is thought to be formed in the last step of morphogenesis because icosahedral particles with or without core structures are observed in the viral factories (Andres *et al*., 2002; Epifano *et al*., 2006a; Epifano *et al*., 2006b; Garcia-Escudero *et al*., 1998). However,

the process of nucleoid formation is poorly understood as empty particles are also found at the plasma membrane to be released by budding (Andres *et al*., 1998; Suarez *et al*., 2015).

1.2.5. Immunology

1.2.5.1. Host immune response to ASFV infection

The tools needed to study the immune response in suids during infection with ASFV, primarily the T cell-mediated responses, are poorly developed (Zhu *et al*., 2019). In addition, the lack of available host (warthog) genome sequences and sequences of ASFVs isolated from warthogs add to the challenge of understanding why African wild suids do not present with apparent clinical signs, while wild boars show clinical symptoms. Warthogs and soft ticks can be persistently infected with the virus and not show any clinical signs of disease (Jori *et al*., 2013; Netherton *et al*., 2019). A serological survey in Eastern and Southern Africa revealed that the infection rate could be higher than 80% in warthogs where ASFV is endemic (Jori and Bastos, 2009; Thomson, 1985). There is a correlation between seroprevalence and age in Uganda (Queen Elizabeth National Park), which is reported to be 58% in young warthogs (4-12 months) and 8% after 24 months of age. Post-infection, young warthogs have been shown to have viraemia levels of 10^3-10^4 HAD50/ml sufficient to infect naïve ticks, but adult warthogs cannot transmit the virus to ticks taking a blood meal as their viral titre levels rarely exceed 10^2 HAD50/ml (Jori and Bastos, 2009).

Highly virulent strains of ASFV cause high mortality rates in domestic pigs; however, animals infected with a low virulence strain that manage to survive the infection are characterised by a short period of virus replication and excretion (Penrith *et al*., 2004). The chronically infected pigs become insusceptible to challenge with a homologous strain. Thus, suggesting animals can develop protective immunity against ASFV (Takamatsu *et al*., 2013). The release of the p36 viral protein by infected macrophages suppresses the immune response in domestic pigs, and also results in reduced expression of the major histocompatibility complex (MHC) class I and II antigens (Penrith *et al*., 2004). The use of sera from pigs that were infected with ASFV strains Malta or OURT_88_3/OURT_88_1, demonstrated that fourteen ASFV serological immunodeterminants that were characterised, induced antibody responses (Kollnberger *et al*., 2002). Six of the fourteen proteins (B602L, C44L, CP312R, E184L, K145R, and K205R) are encoded by previously uncharacterised open reading frames (ORFs), five (A104R, p10, p32, p54, and p72) are structural proteins, and three (F334L, NP419L, and K196R), are non-structural proteins. It was further shown that antibodies in the sera of bush pigs recognised these immunodeterminants after a secondary challenge with a

virulent Malawi ASFV isolate (Kollnberger *et al*., 2002). Twelve ASFV proteins known to be the main targets for serological immunity were previously investigated using sera collected from pigs inoculated experimentally with the ASFV/NH/P68 strain. Strong total immunoglobulin G (IgG) responses were observed against viral proteins p54, K205R, A104R, and B602L. The K205R viral protein induced an IgM response, not seen with other viral proteins, and an IgG1 response was detected against all 12 proteins suggesting an immune response dominated by T helper cell 2 (Th2) (Reis *et al*., 2007).

Previous studies have demonstrated that the humoral and cellular immune responses are essential for the survival of pigs infected with ASFV. Domestic pigs infected with a low virulence ASFV isolate develop a cytotoxic T lymphocyte (CTL) response and show increased activity of the natural killer (NK) cells, and these animals are protected against challenge with a high virulence ASFV strain (Gil *et al*., 2008; Kollnberger *et al*., 2002). A previous study evaluated the immunogenic peptides expressed by an ASFV genomic library (Jenson *et al*., 2000). A total of 72 recombinant plasmids were randomly selected and expressed after transfection into MHC I (+) and MHC II (-) matched pig skin cells, thus, serving as antigen-presenting cells. Each recombinant plasmid was screened with a proliferation assay for recognition by CD8+ T lymphocytes from convalescent pigs, and 14 plasmids were recognised with 10 recognised as non-overlapping and distinct nucleic acid sequences.

The essential role played by CD8+ lymphocytes was evaluated, whereby domestic pigs vaccinated with a non-virulent ASFV-OURT 88/3 isolate were protected against a second challenge with a virulent ASFV strain (Oura *et al*., 2005). However, when the $CD8^+$ lymphocytes were depleted, the protection of the animals from the disease was abolished. It has been shown that low virulence strains induce the production of essential regulatory cytokines as compared to virulent strains. In a previous study, the temporal expression of regulatory cytokines was assessed (Gil *et al*., 2008). It was demonstrated that at 2, 4, and 6 hours post-infection, there was an increase in interferon-alpha (IFN-alpha), tumour necrosis factor (TNF-alpha), and interleukin (IL)-12p40 expression with the low virulence NHV strain, and a lower expression of transforming growth factor-beta (TGF-beta) was observed with the virulent L60 strain.

1.2.5.2. Modulation of host defence response

Previous studies have shown that more than half of the genes encoded by ASFV do not play a role in virus replication, but they have an essential role in establishing and maintaining infection in its host (Dixon *et al*., 2004). Characterisation of these genes will give more insight into understanding how the virus evades the host immune response to infection and avoids elimination, and will also provide knowledge of host antiviral mechanisms and discovery of novel vaccines.

The ASFV has developed different mechanisms to avoid detection by the host immune system. ASFV encodes the A238L protein that has the ability to inhibit the activation of the nuclear factor activated T-cell (NFAT) transcription factor, and it is achieved by binding calcineurin. This mechanism directly inhibits the activity of the nuclear factor Kappa B (NFκB). The ASFV A238L protein inhibits the role of the NFκB transcriptional pathway because it induces a wide range of proinflammatory and antiviral mediators and cytokines. It has been found that the deletion of this ORF in a virulent ASFV strain did not affect the virus virulence in domestic pigs. Thus, suggesting that the pathway inhibited by the A238L gene does not affect the ASFV pathogenesis, or that other genes are encoded which play a supplementary role in the absence of the A238L gene (Abrams *et al*., 2008; Miskin *et al*., 2000; Neilan *et al*., 1997; Rodriguez *et al*., 2002; Salguero *et al*., 2008).

ASFV encodes at least three proteins that are known to have the ability to inhibit apoptosis. Apoptosis is an essential innate cellular mechanism used to prevent virus infection, however, ASFV has developed ways of inhibiting or delaying this mechanism (Brun *et al*., 1996). The A179Lgene encodes a Bcl-2 homologue that targets the Bcl-2 pro-apoptotic protein resulting in the inhibition of apoptosis (Banjara *et al*., 2019). The interaction of the A224L protein with caspase-3 prevents MHC I antigen presentation by using the ER153R (C-type lectin), thus, promoting cell survival (Hurtado *et al*., 2004; Nogal *et al*., 2001).

The EP402R gene encodes the CD2v protein which is responsible for attaching red blood cells to ASFV infected macrophages, facilitating the dissemination of the virus to target tissue, and in soft ticks, it enhances the virus uptake and replication (Rowlands *et al*., 2009). The expression of CD2v protein in infected macrophages prevents T-cell division (Goatley and Dixon, 2011), reducing the impact of these cells during infection. A delay in the onset of viraemia is observed when the ASFV EP402R gene is deleted from the genome. In ASFV infected ticks, the same deletion results in reduced viral replication (Rowlands *et al*., 2009).

It has been demonstrated that the proteins encoded by the multigene family (MGF) 360 and 505/530 genes in highly virulent ASFV isolates inhibit the induction of the host type I IFN response. The type I interferon family of proteins promotes the maturation of dendritic cells, induces viral antigen presentation to CD8+ cells, and increases natural killer cell activity. Deletion of members of the MGF 360 and MGF 505/530 from the genome of the virulent Benin_97_1 strain reduced its virulence, and inoculated animals were protected against the challenge strain (Afonso *et al*., 2004; Golding *et al*., 2016; Reis *et al*., 2016).

1.2.6. Diagnostics

ASF and classical swine fever display similar clinical signs of disease and laboratory confirmation is required for diagnosis using real-time PCR assays (Grau *et al*., 2015; Gallardo *et al*., 2019). Other challenges include a lack of infrastructure or expertise in some of the endemic countries. Laboratory tests should be performed to confirm surveillance measures. There is a wide variety of techniques used to detect ASFV and antibodies during infection, and they are used in combination to get better results. In ASF endemic countries, seroprevalence in the domestic pig is determined through serodiagnostic tests. Tests that are used can identify antibodies produced immediately after infection as well as antibodies in convalescent animals. In cases where clinical symptoms are not apparent, a combination of virus isolation and serodiagnosis should be used (Brown and Bevins, 2018). The most commonly used techniques include enzyme-linked immunosorbent assay (ELISA) and virus DNA detection by quantitative polymerase chain reaction (qPCR) (Hutchings and Ferris, 2006; King *et al*., 2003; Zsak *et al*., 2005). Virus isolation and genotyping techniques are of utmost importance in determining virus origin and genetic variation among ASFV isolates in cases of sporadic ASF outbreaks (Bastos *et al*., 2003; Bastos *et al*., 2004; Boshoff *et al*., 2007; Lubisi *et al*., 2005, 2007; Lubisi *et al*., 2009).

1.3. AFRICAN SWINE FEVER VIRUS GENOME

1.3.1. Characteristics of the ASFV genome

ASFV is comprised of a linear, double-stranded DNA (dsDNA) genome and contains identical inverted terminal repeats (ITR), and hairpin loops at both ends. The genome length ranges between 170 and 194 kilobase pairs (kbp) and encodes 150 to 167 ORFs (Chapman *et al*., 2011; Chapman *et al*., 2008; de Villiers *et al*., 2010; Forth *et al*., 2019). The central conserved region (CCR) of the genome is approximately 125 kbp and is flanked by the left

variable region (LVR) that ranges from 38 kbp to 47 kbp and the right variable region (RVR) that ranges from 13 kbp to 16 kbp in length (Woźniakowski *et al*., 2016).

The first ASF virus to have its genome fully sequenced was that of the cell-culture adapted BA71V isolate, which was annotated based on the EcoRI restriction enzyme fragmentation (Yanez *et al.*, 1995). The number of available ASF virus isolates publically available for comparative studies has increased with the development of sequencing technology. However, the challenge is that the vast majority of sequences represent genotypes I, II, and IX (Achenbach *et al*., 2017; Bastos *et al*., 2003; Boshoff *et al*., 2007; Lubisi *et al*., 2005), thus, limiting the understanding of inter- and intra-genotype genomic differences, geographical origins, and outbreak patterns.

The ASF virus encodes about 200 proteins, and most of them are not fully characterised. The characterised proteins include structural proteins, enzymes involved in genome replication, repair or transcription, and evasion of host defences (Figure 1.6) (Dixon *et al*., 2013). To date, the ASFV Kenya_1950 isolate remains the longest (193, 886 bp) and BA71V, the shortest isolate (170,101 bp). The genome size differences observed between ASFV isolates are mostly due to the gain or loss of the MGF members. BA71V is the only cell-culture adapted strain fully sequenced to date, including the terminal inverted repeats (Rodriguez *et al*., 2015). Pairwise comparison between ASFV genomes in previous studies has shown that the percentage identity of ASFV isolates ranges from 88% to 98% (Chapman *et al*., 2008; Portugal *et al*., 2015). Most of the variations between ASFV isolates are found in the left and right variable regions, which are mainly comprised of MGF members that are not always in tandem and have different numbers of paralogs (Imbery and Upton, 2017). Comparative studies of ASFV genome sequences has provided more knowledge about genetic variability and evolution, as well as the phylogenetic relationships of the isolates representing different genotypes. Comparative studies have helped in understanding outbreak patterns, routes of spread, and genetic variation among viruses from different geographical areas (Chapman *et al*., 2011; de Villiers *et al*., 2010; Giammarioli *et al*., 2011).

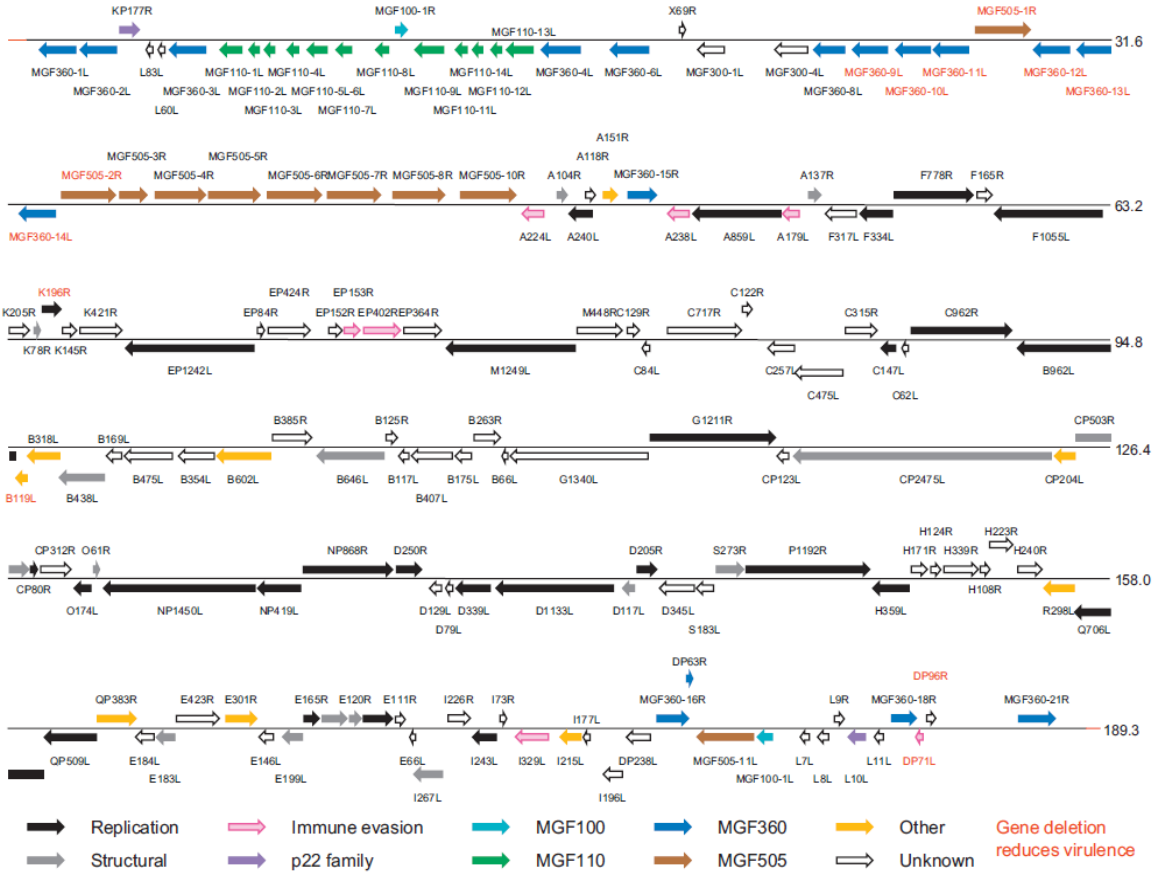


Figure 1. 6. Genome organisation of the African swine fever virus genome. The organisation of open reading frames (ORFs) on the genome of the virulent ASFV isolate Georgia 2007/1 is shown. ORFs are shown as arrows to indicate their size and direction they are read. The colours indicate ORFs with known functions. Black indicates ORFs encoding enzymes and factors involved in genome replication, repair or transcription. Grey indicates ORFs encoding structural proteins. Pink indicates ORFs encoding proteins involved in evading the host defences. Turquoise, blue, green, brown and mauve indicate members of multigene families. ORFs encoding proteins with other predicted functions are shown in yellow. ORFs encoding proteins of unknown function are shown in white. Red text indicates ORFs whose deletion reduces virus virulence. The image was taken from (Dixon *et al*., 2013).

1.3.2. Virulence genes

In domestic pigs, the severity of the disease ranges from lethal haemorrhagic to subclinical based on whether the ASFV isolate is virulent or non-virulent. The deletion of genes involved in virulence has resulted in the attenuation of the virus in domestic pigs (Zsak *et al*., 1998). Genes that play a role in virulence have been characterised, which include the dUTPase (E165R gene), thymidine kinase (K196R gene), UK (DP96R gene), NL (DP71L gene), and the DP148R gene (Oliveros *et al*., 1999; Reis *et al*., 2017a; Sanford *et al*., 2016).

The dUTPase gene is responsible for reducing the dUTP concentration during a natural infection, ensuring proper incorporation of the deoxyuridine into the viral DNA, and ensuring

the accuracy of genome replication. The thymidine kinase gene is also involved in nucleotide and nucleic acid metabolism. These genes have been shown to play a role in the host range, and the deletion of these two genes from the virus genome reduces its ability to replicate in macrophages (Ramirez-Medina *et al*., 2019).

The exact functions of the DP96R and the DP71L genes are not clearly understood, but it has been shown that they are involved in viral pathogenesis and virulence in domestic pigs but do not affect replication in macrophages *in vitro*. The DP96R gene is expressed early during infection and is highly conserved among ASFV isolates, with 92% to 96% amino acid identity. However, it is not encoded in the genome of the virulent Malawi_LiL-20_1 isolate. The deletion of this gene from the genome of virulent ASFV isolates results in attenuation, although replication in macrophages is not affected (Zsak *et al*., 1998). The NL protein encoded by the DP71L gene is similar to MyD116 and a neuro-virulence factor ICP34.5 found in the herpes simplex virus. It was shown that deletion of the recently characterised DP148R gene reduced the virulence of the Benin_97_1 isolate in pigs and gave protection against challenge with the homologous strain (Bacciu *et al*., 2016; Ramirez-Medina *et al*., 2019; Reis *et al*., 2017a; Zsak *et al*., 1996).

1.4. Vaccine development

It was previously demonstrated that convalescent pigs were resistant to a second challenge with a homologous virulent strain and that there was a high level of protection (Detray, 1957). This evidence played a role in initiating vaccine development studies; however, the progress has been slow, and to date, there is currently no vaccine. It was demonstrated that antibodies play a role in protection through the passive transfer of antibodies from convalescent pigs to naïve animals, which resulted in partial protection against challenge (Onisk *et al*., 1994; Wardley *et al*., 1985). It was shown that CD8+ cells also provide protection; however, when they were depleted, the protection induced by live attenuated isolates was abrogated (Oura *et al*., 2005). Thus, an effective vaccine should be able to induce both protective antibodies and cellular responses.

1.4.1. Inactivated virus vaccines

The approach of developing inactivated vaccines was tested by immunising animals with glutaraldehyde-fixed ASFV infected alveolar macrophages and detergent treated antigens (Coggins, 1974; Forman *et al*., 1982; Mebus, 1988), however, it was not successful. Subsequent studies showed that binaryethylenimine-inactivated ASFV with polygen or

emulsigen adjuvants did not induce a protective immune response (Blome *et al*., 2014; Forman *et al*., 1982; Mebus, 1988). Inactivated vaccines are known to induce antibodies that prevent virus infection of cells; however, these antibodies have failed to inhibit ASFV infection. This is thought to be a result of two infectious forms of the virus particle, which are extracellular and intracellular mature particles (Andres, 2017; Hernaez *et al*., 2016).

1.4.2. Subunit vaccines

Since antibody-mediated immunity was achieved through passive transfer of antibodies from recovered pigs to naïve animals (Onisk *et al*., 1994), several groups have investigated subunit vaccines as an approach for vaccine development. ASFV encodes about 167 proteins, which makes it challenging to choose candidate antigens that will induce a protective immune response. Proteins that have been described to be immunogenic include p54, p30, pp220, pp62, p72, and CD2v, which have been incorporated in subunit vaccines (Galindo and Alonso, 2017; Gomez-Puertas *et al*., 1998; Gomez-Puertas *et al*., 1996; Jia *et al*., 2017). Partial immunity has been achieved with baculovirus-expressed p54, p30, and p72, and it has also been demonstrated that baculovirus-expressed CD2v, p54, and p30 conferred protection against virulent ASFV challenge (Gomez-Puertas *et al*., 1998; Ruiz-Gonzalvo *et al*., 1996). It was shown that baculovirus-expressed p22, p30, p54, and p72 proteins, failed to induce sufficient protection (Neilan *et al*., 2004).

DNA vaccines have also been tested as an alternative, and they are known to provide better T cell immune responses because the antigens are expressed intracellularly and presented to MHC-I, thus, playing an essential role in activating the CD8+ lymphocyte cells. However, the use of DNA vaccines encoding p30 and p54 to immunise pigs did not induce a protective immune response (Argilaguet *et al*., 2012). To improve this approach, CD2v protein was fused to p30 and p54, and a strong humoral and cellular immune response was observed, however, there was no specific antibody response (Argilaguet *et al*., 2012; Lacasta *et al*., 2014). Another approach to try and induce both humoral and cell-mediated responses was with the use of the viral vector BacMan which is a baculovirus-based vector that can transduce mammalian cells to express the target genes. This vector was used to deliver the sHA/p54/p30 fusion construct to domestic pigs for protection against a sub-lethal challenge. It was demonstrated that 4 out of 6 pig were protected; however, no specific antibody response was detected and the protection was correlated to a strong virus-specific T cell response. To date, several vectors have been tested in pigs, and only a few have been tested against virulent ASFV challenge (Gaudreault and Richt, 2019).

1.4.3. Live attenuated vaccines

Live attenuated ASFV vaccines have been shown to provide up to 100% protection, but there are safety concerns which include adverse effects as a result of the vaccine strains and the potential transmission of the virus in the field (King *et al*., 2011; Leitao *et al*., 2001; O'Donnell *et al*., 2017; Reis *et al*., 2016; Reis *et al*., 2017a). Attenuation was achieved by either passaging strains in cell culture or deleting genes that are associated with virulence from genomes of virulent strains (Boinas *et al*., 2004; Leitao *et al*., 2001). Deleted genes include thymidine kinase (TK), DP71L, B119L, DP96R, CD2v, and members of the MGF 360 and 505/530 (Monteagudo *et al*., 2017; O'Donnell *et al*., 2017; Sanford *et al*., 2016). It has been demonstrated that the deletion of six members of the MGF 360 (MGF 360-9L, 10L, 11L, 12L, 13L, and 14L), and MGF 505 along with B119L resulted in attenuation of the Georgia_2007_1 strain with improved safety. However, it did not induce a protective immune response when vaccinated pigs were challenged with a virulent parent strain (O'Donnell *et al*., 2016). It was further demonstrated that safety and protection were improved when the B119L and DP96R were both deleted rather than deleting the B119L gene alone. It was shown that deletion of the EP402R (CD2v) gene from the genome of the cell-culture adapted strain BA71V provided protection against a homologous challenge with the parental BA71 strain, and further conferred protection against a heterologous challenge with a genotype I E75 strain, and genotype II Georgia_2007_1 strain (Monteagudo *et al*., 2017). It has been shown that through the deletion of virulent genes, different immune responses were achieved. The deletion of the CD2v protein in the BA71V strain resulted in virus attenuation and induced a protective immune response. However, when the CD2v protein was deleted in other virulent strains, attenuation was not observed (Borca *et al*., 1998; Monteagudo *et al*., 2017). To date, live attenuated ASFV isolates remain the only method that has effectively protected pigs against challenge with homologous virulent strains. However, it is going to take years to produce a licensed vaccine that can be used in the field.

1.5. Challenges of working with ASFV

ASFV is a difficult virus to research as a biosecurity level 3 (BSL-3) laboratory is required (Manual, 2008), and the virus is slow-growing with low titres. During the process of virus purification and viral DNA extraction, the recovered viral DNA is mostly contaminated by host DNA. Thus, purification and extraction protocols have to be optimised to reduce the amount of host DNA which is time-consuming (Carrascosa *et al*., 2011; Carrascosa *et al*.,

1985). The decontamination process, when sending samples out of the BSL3 facility for sequencing, also affects the DNA quality of the samples. A further challenge is the lack of stable cell lines to passage ASFV isolates. Other cell lines, such as Vero cells, have been used to adapt some ASFV isolates. However, it was shown that they cause genomic changes that resulted in reduced virus replication in pigs and lack of induction of a protective immune response (Krug *et al*., 2015).

1.6. PROJECT RATIONALE

To date, there are 67 complete genome sequences publically available, but most of the sequences represent genotypes I-V and VII-X of the 24 genotypes identified based on p72 sequencing, limiting comparative studies. Although ASFV is only a single species, there is a significant difference in genome size (170-194 kbp) and composition of MGFs in the available sequences. Little is known about the genes involved in virulence, host range, and viral-vector-host interactions. One important gap in the field is the lack of available diverse ASFV genome sequences. The limited sequence data of ASFV isolates from different origins and hosts, including domestic pigs, wild suids, and soft ticks, results in limited interpretation of the genetic diversity between ASFV isolates. This study aimed to use next-generation sequencing technologies to sequence and characterise six novel full-length genomes of ASFV isolates from different genotypes (I, III, IV, XX, and XXII) and hosts including a domestic pig, warthog, wild boar, and soft ticks, and to compare them to publicly available reference genomes. It has been shown that the concatenation of multiple genes can resolve phylogenetic ambiguities observed when using single genes. A good phylogenetic marker was previously defined as a gene that can give a similar phylogenetic tree to the one obtained with concatenated core genes (de Villiers *et al*., 2010). This study further aimed to identify novel loci that could be phylogenetically informative especially for isolates from the same genotype. The role of the MGFs is not adequately understood, and previous studies have shown that they are implicated in virulence and host range (Afonso *et al*., 2004). MGFs of sixteen ASFV isolates were recently reannotated, and gene maps were created that could be edited to incorporate newly sequenced ASFV genomes and MGF orthologs when discovered (Imbery and Upton, 2017). Thus, we have used this framework to conduct an updated MGF analysis of six novel ASFV isolates and 51 publicly available fully sequenced isolates from GenBank.

1.7. AIMS

1. To characterise six novel ASFV isolates of African origin from a domestic pig, warthog, wild boar, and soft ticks by complete genome sequencing and comparison to publicly available reference sequences.

2. To analyse a multiple sequence alignment of 73 complete ASFV genome sequences to identify and investigate novel loci in the CCR that could be phylogenetically informative.

3. To compare the MGFs across 73 ASFV isolates using sorting and clustering to understand their contribution to genome variation.

CHAPTER 2: THE GENOME SEQUENCES OF AFRICAN SWINE FEVER VIRUS (ASFV) ISOLATES FROM SOUTH AFRICA, ZAMBIA, AND ZAIRE

2.1. Introduction

One of the major gaps in the field of ASFV research is the lack of publicly available, complete genome sequences representing the 24 identified genotypes (Bastos *et al.*, 2003, Lubisi *et al.*, 2005, Boshoff *et al.*, 2007, Achenbach *et al.*, 2017, Quembo *et al.*, 2018). Currently, only 9 of 24 ASFV genotypes (I-V, VII, IX, X, and XX) are represented by complete genome data. More than two-thirds of the publicly available sequences represent ASFV isolates from genotypes I and II, which belong to the same lineage and are the only genotypes found outside of Africa. Some of the technical challenges of sequencing ASFV genomes include the need for the use of a BSL-3 laboratory, slow growth of the virus, a lack of available, stable cell lines, and minimisation of host DNA contamination, making whole-genome sequencing (WGS) progress difficult and time-consuming; however, methodological improvements are being made (Carrascosa *et al*., 2011; de Leon *et al*., 2013; Olasz *et al*., 2019).

Genotypes that are not represented by complete genome data have all been isolated from sub-Saharan Africa (Figure 2.1) (Mozambique (Genotype VI; isolate from ticks), Botswana and South Africa (Genotype VII; isolated from ticks and domestic pigs), Zambia (Genotype XI; isolated from a ticks), Malawi and Zambia (Genotype XII; isolated from domestic pigs and ticks), Zambia (Genotypes XIII and XIV; isolated from ticks), Tanzania (Genotypes XV and XVI; isolated from domestic pigs), Zimbabwe (Genotype XVII; isolated from ticks), Namibia (Genotype XVIII; isolated from domestic pigs), South Africa (Genotypes XIX, XXI, and XXII; isolated from ticks), Ethiopia (Genotype XXIII; isolated from domestic pigs), and Mozambique (Genotype XXIV; isolated from ticks) (Bastos *et al*., 2003, Lubisi *et al*., 2005, Boshoff *et al*., 2007). African swine fever is endemic in many countries and known to be a threat to small scale and commercial pig farming. The lack of genotype representation limits genome comparison studies and results in challenges inferring the evolutionary relationships between ASFV isolates. Therefore, more isolates from different geographical areas and hosts that include domestic pigs, warthogs, wild boars, and ticks should be sequenced.

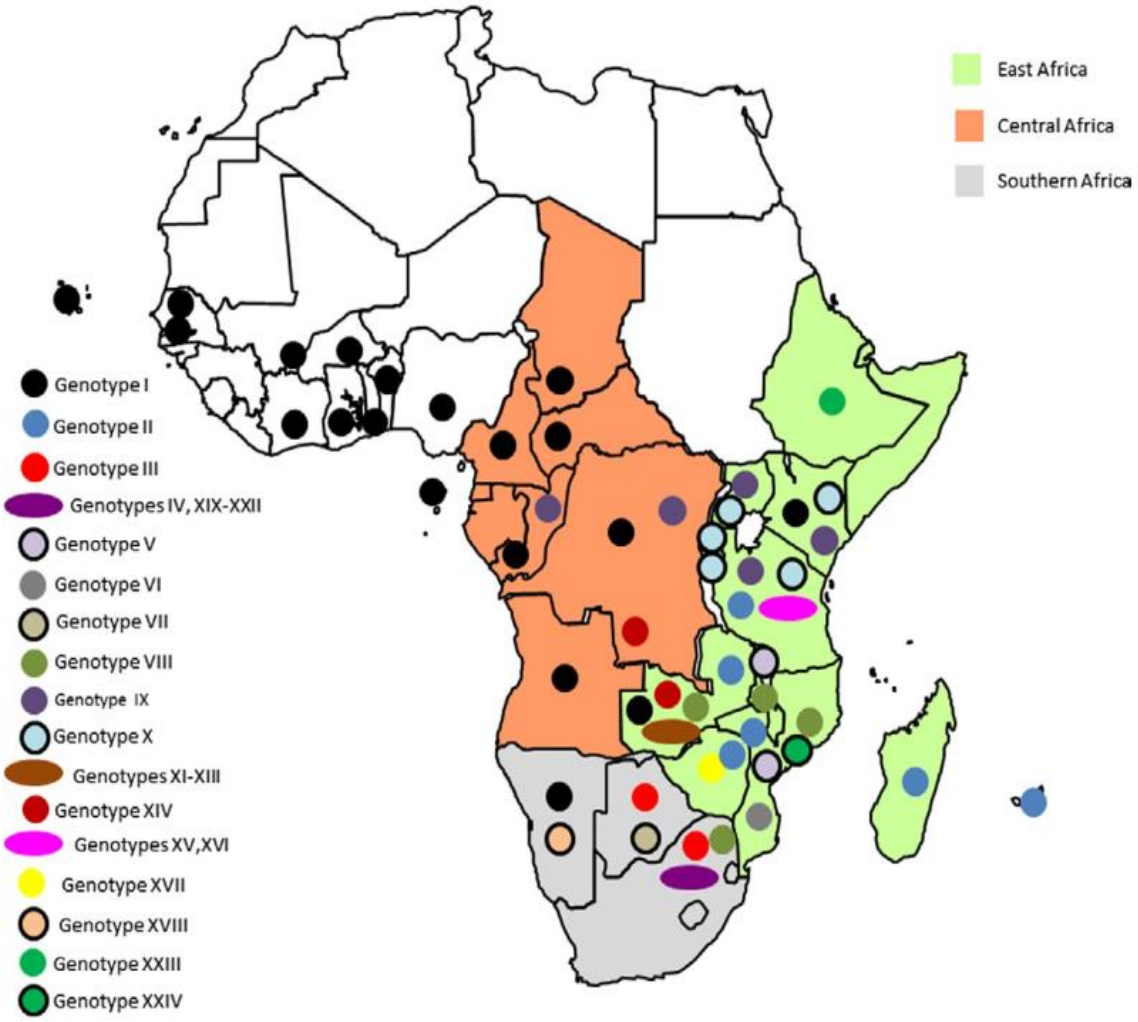


Figure 2. 1. Distribution of the p72 genotypes of ASFV in sub-Saharan Africa. The image was taken from (Mulumba-Mfumu *et al*., 2019).

An effective approach to understanding ASFV diversity is based on the partial sequencing of the B646L gene. However, increased discrimination between closely related ASFV isolates has been achieved by sequencing additional genes that include CP204L (p30), E183L (p54), and B602L genes (Nix *et al.*, 2006, Gallardo *et al.*, 2009). The combination of the B646L (p72) gene and the central variable region within the B602L gene are used for intra-genotyping of ASFV isolates and enhanced discrimination of phylogenetic data (Gallardo *et al*., 2009; Phologane *et al*., 2005). A comparison of the available genome sequences shows an overall high degree of conservation of the central conserved region (CCR) with the exception of several genes that are used for phylogenetic analyses (p72, p30, p54, and B602L). The terminal ends of the genome are the most variable regions, referred to as the left variable region (LVR) and the right variable region (RVR). These terminal ends are comprised mostly

of the multigene families (MGFs), which are made up of paralogous genes of various size and gene content (Chapman *et al*., 2008). Although their function is not fully understood, it has been shown that the MGF 360 and 505/530 genes play a role in virulence, host range specificity, and blocking of host innate response (O'Donnell *et al*., 2015a). Analysis of p72, p30, p54, and B602L does not capture the LVR and RVR diversity; however, WGS will allow the analysis of these differences.

The development of a vaccine is of utmost importance for African swine fever (ASF) control and eradication but is complicated by ASFV genome diversity and complexity. Approaches that have been attempted to develop a protective vaccine include inactivated viruses, recombinant proteins, and viral vectors to deliver antigens. However, they failed to induce a protective immune response (Blome *et al.*, 2014). Live attenuated ASFV isolates generated through serial passage in cell culture or through deletion of MGFs or other genes involved in virulence such as DP148R, EP402R, and B119L have induced protective immune responses against homologous challenge (Lewis *et al*., 2000; Monteagudo *et al*., 2017; Reis *et al*., 2017a). A genotype II, non-haemadsorbing, naturally attenuated strain (Lv17/WB/Rie1) resulting from an adenosine deletion that generates a truncated EP402R (CD2v) gene, conferred 92% protection in Eurasian wild boars against challenge with a virulent genotype II strain (Arm07) (Barasona *et al*., 2019). It was demonstrated that the deletion of the uncharacterised I177L from Georgia_2007_1 resulted in complete virus attenuation in swine and induced protection against challenge with the parental strain (Borca *et al*., 2020). This vaccine is reported as the first candidate to confer sterilising immunity against the circulating strain causing outbreaks in Eurasia. ASFV specific antibodies have been detected in serum collected from wild boars in Eastern Europe, suggesting that circulating ASFV strains are attenuated, since the acute form of disease causes animals to die before an antibody response develops (Nurmoja *et al*., 2017). A deletion at the 5'-end of the Estonia strain was associated with attenuation, and it resulted in deletion of 26 genes (Zani *et al*., 2018). Of the 26 genes, thirteen are members of the MGF 110 (1L-13L), while three are members of the MGF 360 (1L-3L). A duplication and an inverse binding of about 7kb occurred from the 3' to the 5' which resulted in duplication of two genes (DP71L and DP96R), which play a role in virulence (Oliveros *et al*., 1999; Reis *et al*., 2017a; Sanford *et al*., 2016). The Estonian variant was further characterised in three animal experiments (A, B, and C). The Estonia strain which was re-isolated from surviving pigs, demonstrated characteristics of attenuation in trials A

and B. However, all the animals developed acute clinical signs, but a survival rate between 75% and 100% was observed when the trials ended (Zani *et al*., 2018).

Genetically manipulated attenuated vaccine candidates have helped in identifying ASFV genes involved in virulence and immune evasion (O'Donnell *et al*., 2015b; Reis *et al*., 2016; Sanchez-Cordon *et al*., 2018a). Thus, the comparison of genomes of virulent and non-virulent ASFV strains will aid in identifying genes involved in virulence. Large genome deletions and sequence transpositions from one genome terminus to the other have been associated with reduced virulence (Dobson & Tscharke, 2015). Thus, the sequencing of ASFV isolates representing all 24 described genotypes is needed because the availability of genomic sequences helps in designing and developing safer vaccines. WGS will provide more information on the antigenic diversity among ASFV isolates and aid in identifying constant epitopes. This approach will provide substantial information on the genomic sequence of virulent and non-virulent ASFV strains and help in developing safer attenuated vaccines through the deletion of genes involved in virulence. Comparative genomics studies of different isolates of the same species of a pathogen, helps in understanding virulence associated genetic patterns, and identifying novel vaccines candidates. Complete genome sequences (Table 2.1) of haemadsorbing and non-haemadsorbing isolates of different genotypes from different hosts, and geographical areas, were selected for comparative genomics analysis.

The aim of this chapter is to characterise six novel ASFV isolates of African origin from a domestic pig, warthog, wild boar, and soft ticks by complete genome sequencing and comparison to publicly available reference sequences. In this study, six ASFVs (Table 2.2) were selected for WGS. The Zaire isolate is a non-haemadsorbing, cell-culture adapted virus that grows in Vero cells. The remaining isolates are haemadsorbing viruses that grow on primary swine macrophages. These isolates were selected to allow for whole genome comparison between cell-culture adapted and wildtype isolates, MGF comparison in Chapter 4 and testing of novel phylogenetic loci in Chapter 3. To date, the BA71V strain from Spain, isolated in 1971 from a domestic pig, and representing genotype I, remains the only publicly available cell-culture adapted reference sequence; the Zaire isolate will be the second to be reported. Multiple sequence alignment of the C-terminus of the B646L gene showed LIV_5_40 to belong to genotype I (Lubisi *et al*., 2005), RSA_W1_1999 to genotype IV (Bastos *et al*., 2003), RSA_2_2004 to genotype XX (ARC-OVI (TAD)), RSA_2_2008 to genotype XXII (ARC-OVI (TAD)), SPEC_57 to genotype III (Arnot *et al*., 2009), and Zaire

to genotype XX (Lubisi *et al*., 2005), with LIV 5_40 showing closest BLASTn identity to LIV_1217 from Zambia (genotype I; Genbank accession number AY351524), RSA_W1_1999 to Lillie from South Africa (genotype XX; Genbank accession number DQ250109), RSA_2_2004 to Warthog from Namibia (genotype IV; Genbank accession number AY578706), RSA_2_2008 to RSA_1_2002 from South Africa (genotype XXII; Genbank accession number JX403678), SPEC_57 to RSA_3_2003 from South Africa (genotype III; Genbank accession number JX403655), and Zaire to Benin_97_1 from Benin (genotype I; Genbank accession number EF121428) (Ndlovu *et al*., 2020a; Ndlovu *et al*., 2020b)

Table 2. 1. Complete ASFV genome sequences that are publicly available

Isolate	Year	Length (bp)	LVR (bp)	CVR (bp)	RVR (bp)	Country	Genotype	Host	Accession No.
BA71V	1971	170 101	29 795	122 033	11 927	Spain	I	Domestic pig	NC_001659
BA71	1971	180 365	37 944	129 623	14 064	Spain	I	Domestic pig	KP055815
Benin_97_1	1997	182 284	41 129	122 230	12 654	Benin	I	Domestic pig	AM712239
L60	1960	182 362	41 478	121 898	12 653	Portugal	I	Domestic pig	KM262844
E75	1975	181 187	40 433	119 968	12 422	Spain	I	Domestic pig	FN557520
OURT_88_3	1988	172 051	31 949	121 836	11 147	Portugal	I	Tick	AM712240
NHV	1968	171 719	32 296	121 856	11 146	Portugal	I	Domestic pig	KM262845
Mkuzi_1979	1979	192 714	48 676	121 297	15 785	South Africa	I	Tick	AY261362
26544_OG10	2010	182 906	41 202	121 900	13 539	Italy	I	Domestic pig	KM102979
ASFV-Patent	1971	180 365	47 966	129 623	14 064	Spain	I	Domestic pig	LP643842
56/Ca/78	1978	183 636	41 212	130 415	12 009	Italy	I	Domestic pig	MN270969
57/Ca/79	1979	183 639	41 214	130 415	12 010	Italy	I	Domestic pig	MN270970
139/Nu/81	1981	183 645	41 220	130 415	12 010	Italy	I	Domestic pig	MN270971
140/Or/85	1985	183 723	41 256	130 415	12 052	Italy	I	Domestic pig	MN270972
85/Ca/85	1985	181 816	40 310	130 259	11 247	Italy	I	Domestic pig	MN270973
141/Nu/90	1990	183 720	41 333	130 241	12 146	Italy	I	Domestic pig	MN270974
142/Nu/95	1995	183 724	41 336	130 241	12 147	Italy	I	Domestic pig	MN270975
60/Nu/97	1997	181 651	40 295	130 240	11 116	Italy	I	Domestic pig	MN270976
26/Ss/04	2004	181 869	41 762	130 240	9 867	Italy	I	Domestic pig	MN270977
72407/Ss/05	2005	181 699	40 303	130 241	11 155	Italy	I	Domestic pig	MN270978
97/Ot/12	2012	184 206	41 574	130 241	12 391	Italy	I	Domestic pig	MN270979
22653/Ca/14	2014	181 869	40 449	130 252	11 168	Italy	I	Domestic pig	MN270980
47_Ss_2008	2008	184 638	42 466	122 215	14 007	Italy	II	Domestic pig	KX354450
Georgia_2007_1	2007	189 344	46 983	121 525	14 349	Georgia	II	Domestic pig	FR682468
Estonia	2014	182 446	39 698	122 481	14 700	Estonia	II	Domestic pig	LS478113
Podlaskie	2015	189 394	47 003	121 605	14 354	Poland	II	Wild boar	MH681419
Odintsovo	2014	189 333	46 979	121 385	14 347	Russia	II	Domestic pig	KP843857
China_2018_AnhuiXCGQ	2018	189 393	46 990	129 332	14 355	China	II	Domestic pig	MK128995
Belgium_2018_1	2018	190 599	47 966	129 330	14 587	Belgium	II	Wild boar	LR536725
DB_LN_2018	2018	189 404	47 001	129 331	14 356	China	II	Domestic pig	MK333181
Pig_HLJ_2018	2018	189 404	47 001	129 331	14 356	China	II	Domestic pig	MK333180
Pol16_20186_o7	2016	189 401	46 998	129 331	14 356	Poland	II	Domestic pig	MG939583
Pol17_04461_C210	2017	189 401	46 998	129 331	14 356	Poland	II	Domestic pig	MG939588

Isolate	Year	Length (bp)	LVR (bp)	CVR (bp)	RVR (bp)	Country	Genotype	Host	Accession No
Pol17_03029_C201	2017	189 405	46 998	129 335	14 356	Poland	II	Domestic pig	MG939587
Pol16_20538_o9	2016	189 399	47 001	129 323	14 359	Poland	II	Domestic pig	MG939584
Pol16_29413_o23	2016	189 393	47 000	129 321	14 356	Poland	II	Domestic pig	MG939586
Pol17_05838_C220	2017	189 393	47 000	129 321	14 356	Poland	II	Domestic pig	MG939589
Pol16_20540_o10	2016	189 405	47 002	129 331	14 356	Poland	II	Domestic pig	MG939585
Belgium_Etalle_wb_ 2018	2018	190 202	47 046	129 330	15 110	Belgium	II	Wild boar	MK543947
Czech Republic_ 2017_1	2017	190 594	47 961	129 331	14 586	Czech Republic	II	Wild boar	LR722600
Moldova_2017_1	2017	190 598	47 964	129 331	14 587	Moldova	II	Domestic pig	LR722599
ASFV_Kyiv_2016_131	2016	191 911	48 653	129 327	15 215	Ukraine	II	Domestic pig	MN194591
ASFV-wbBS01	2018	189 394	47 000	129 322	14 356	China	II	Wild boar	MK645909
ASFV-SY18	2018	189 354	46 983	129 306	14 349	China	II	Domestic pig	MH766894
Georgia 2008/1	2008	189 465	46 971	129 329	14 449	Georgia	II	Domestic pig	MH910495
Georgia 2008/2	2008	189 315	46 911	129 334	14 354	Georgia	II	Domestic pig	MH910496
Liv13/33	2020	188 277	47 616	129 389	12 538	Zambia	II	tick	MN913970
Wuhan 2019-1	2019	190 576	48 165	129 338	14 357	China	II	Domestic pig	MN393476
Wuhan 2019-2	2019	190 576	48 165	129 338	14 357	China	II	Domestic pig	MN393477
Georgia 2007.1 (Reseq)	2017	190 584	47 961	129 321	14 586	Georgia	II	Domestic pig	LR743116
CHN_CAS19	2019	189 405	47 001	129 332	14 356	China	II	Domestic pig	MN172368
LT14.1490	2014	189 399	46 998	129 331	14 354	Lithuania	II	Domestic pig	MK628478
Kashino 04/13	2014	189 387	47 005	129 311	14 229	Russia	II	Wild boar	KJ747406
HU_2018	2018	190 601	47 955	129 332	14 598	Hungary	II	Domestic pig	MN715134
Warmbaths	1987	190 773	46 773	121 424	15 849	South Africa	III	Tick	AY261365
Warthog	1980	186 528	43 606	129 289	14 899	Namibia	IV	Warthog	AY261366
Tengani62	1962	185 689	42 061	121 446	15 383	Malawi	V	Domestic pig	AY261364
Malawi_Lil_20_1	1983	187 612	43 505	121 486	12 890	Malawi	VIII	Tick	AY261361
Ken05.Tk1	2005	191 058	47 392	121 162	14 636	Kenya	X	Domestic pig	KM111294
Kenya_1950	1950	193 886	49 916	121 215	14 361	Kenya	X	Domestic pig	AY261360
Ken06.Bus	2006	184 368	43 473	120 528	12 106	Kenya	IX	Domestic pig	KM111295
R7	2015	188 628	45 621	121 826	14 204	Uganda	IX	Domestic pig	MH025917
R8	2015	188 627	45 620	121 826	14 204	Uganda	IX	Domestic pig	MH025916
R25	2015	188 630	45 623	121 712	14 204	Uganda	IX	Domestic pig	MH025918
R35	2015	188 629	45 622	121 827	14 204	Uganda	IX	Domestic pig	MH025920
N10	2015	188 611	45 609	121 898	14 204	Uganda	IX	Domestic pig	MH025919
Pretoriuskop_96_4	2005	190 324	47 596	121 808	14 365	South Africa	XX	Tick	AY261363

2.2. MATERIALS AND METHODS

2.2.1. Growth of novel ASFV study isolates in cell-culture and purification of virus DNA

Passage of the ASFV Zaire isolate on Vero cells: The Zaire isolate was part of the collection stored in the virus bank at the Onderstepoort Veterinary Institute, Agricultural Research Council (ARC-OVI), in the Transboundary Animal disease (TAD) laboratory. This isolate was collected from domestic pigs during an outbreak in the province of Kinshasa in Zaire in 1977 and stored at -80°C. The Zaire isolate was passaged 20 times in Vero cells before its use in this study, followed by three passages in Vero cells grown in Roswell Park Memorial Institute growth medium (RPMI) with 10% normal bovine serum (NBS) under standard conditions in this study. Monolayers of Vero cells were infected with 1 ml of 10^4 50% tissue culture infective dose (TCID50)/ml of ASFV Zaire isolate and incubated at 37°C for 48 hours. Cultures were examined daily for plaque development, and once the plaques developed, the culture was stained with crystal violet, and the plaques were counted. The virus titre was estimated by the following equation:

$$\text{Viral titre} = \frac{D * P}{V}$$

Where:

D is the dilution of virus

P is the average number of plaques counted

V is the volume of diluted virus samples added to the monolayer of cells

Following infection of Vero cell monolayers, the virus was harvested by freezing the flask at -80°C for at least 30 min and thawing to lyse the cells. After thawing, the content was collected with a 10 ml pipette and transferred into 50 ml falcon tubes. The virus was purified following a method for purification of poxvirus DNA from lesion materials (Afonso *et al*., 2006). The virus was pelleted at 19000 x g for 60 min, through a 36% sucrose cushion, and further pelleted through a 20% sucrose cushion at 18750 x g for 40 min. The pellet was left to resuspend overnight in 0.5 ml Tris-EDTA (TE) buffer (pH 7.5), followed by DNA extraction.

Culture of the haemadsorbing ASFV isolates LIV_5_40, RSA_W1_1999, RSA_2_2004, RSA_2_2008, and SPEC_57 in bone marrow macrophages: In order to increase the amount of virus available to extract DNA they were amplified in bone marrow cultures.

Information on these haemadsorbing viruses is provided in Table 2.2. These viruses were passaged in primary swine macrophages acquired from bone marrow as described by (Malmquist and Hay, 1960). The bone marrow leukocytes were extracted from femoral bones collected from euthanised pigs. The animals were euthanised following the ARC/OVI animal ethics committee approval (AEC 7.18) and as part of the diagnostic department's routine procedure for collection of primary cells for use in virus isolation.

Bone marrow culture: The bone marrow cultures were prepared from the femoral bones collected from a euthanised pig. The bones were cleaned off the flesh and cartilage, put in 300 ml of wash buffer, and incubated for 20 minutes at 37°C. The bones were fragmented into small pieces using the bone-cutting forceps, poured into a one-litre conical flask containing 250 ml of wash buffer, then incubated at 37°C for 90 minutes while gently shaking the flask. A cell-rich solution was collected into 250 ml plastic centrifuge bottles (Beckman®) by placing butter muslin over funnels to filter the fragmented bone pieces. The cell-rich solution was centrifuged at 2500 g (Beckman® model J-6B) for five minutes at 25°C. The supernatant was removed and the sediment was resuspended in 0.82% ammonium chloride solution, and centrifuged at 2500 g (Beckman® model J-6B) for five minutes. The sediment was washed twice with wash buffer and then resuspended in the growth medium. One hundred microliters of the resuspended cells were poured into 96 well plates (NUNC^{TM}) and incubated at 37°C.

Haemadsorbption Assay: Haemadsorption is defined as the adherence of red blood cells to the surface of infected cells. Bone marrow cell culture were infected with 50 µl of 10^4 50% haemadsorbing dose (HAD50)/ml of ASFV LIV_5_40, RSA_W1_1999, RSA_2_2004, and RSA_2_2008, and SPEC_57, in a 96-well plate. The plates were incubated for 2 hours at 37°C to allow cells to settle, and then 25 µl of red blood cells were added. The plates were then sealed with plastic sealers followed by the lids and incubated at 37°C for 48 hours. The inoculated cells were monitored daily for haemadsorption. Positive wells were harvested by freezing the 96 well plates at –80°C for at least 30 minutes and thawing to lyse the infected cells. The contents of the positive wells were collected by piercing through the plastic sealers with a syringe and pooled into a 20 ml bottle. Centrifugation at 15000 revolutions per minute (rpm) for ten minutes followed. The virus rich supernatant was removed and transferred into 1 ml cryo-vial tubes, followed by DNA extraction.

2.2.2. Viral DNA extraction and amplification

Isolation of nucleic acid was performed using a High Pure PCR Template Preparation Kit (Roche Diagnostics, Germany) according to the manufacturer's protocol. The NEBNext Microbiome DNA Enrichment Kit (New England Biolabs, USA) was used to deplete the host methylated DNA according to the manufacturer's instructions. The DNA was amplified with the Genomiphi v3 DNA amplification kit (GE Healthcare, U.S.A) according to the manufacturer's instructions. A confirmatory PCR for the presence of ASF viral nucleic acid was performed to amplify the C-terminus region of the p72 gene using the following epidemiological primers P72-U (5'-GGCACAAGTTCGGACATGT-3') and P72-D (5'-GTACTGTAACGCAGCACAG-3') as described by (Bastos *et al*., 2003).

2.2.3. DNA sequencing, analysis, and comparative genomics

Viral DNA (1 ng total) was prepared for sequencing on an Illumina MiSeq sequencer using the Nextera XT DNA Sample Preparation kit (Illumina, U.S.A.). This sequencing method makes use of fluorescently labelled nucleotides to quantify input DNA. This technology yields highly accurate sequence data and eliminates sequences with specific errors. Fragmentation of input DNA, PCR amplification of tagged DNA, and PCR clean-up were performed as per the manufacturer's instructions. The libraries were pooled, denatured, and combined with the PhiX genome following Illumina protocols. Finally, library mixtures were loaded into the cartridge for simultaneous sequencing. Paired-end reads of approximately 300 bp were generated using Miseq v3 chemistry, for RSA_W1_1999, RSA_2_2004, and Zaire isolates. In contrast, the Illumina Hiseq sequencer was used to generate paired-end reads of approximately 2x250 bp using the v2 chemistry for LIV_5_40, RSA_2_2008, and SPEC_57 isolates. The sequencing was performed at the Biotech Platform (Onderstepoort Veterinary Institute, Agricultural Research Council (ARC-OVI)). Basic quality control was performed at the Biotech Platform (ARC-OVI). CLC Genomics Workbench version 9.0.1 was used to trim adaptor sequences and to remove poor-quality bases at the 3' end of long reads.

A *de novo* assembly of all reads was performed using SPAdes genome assembler (Version 3.13.0, https://www.liebertpub.com/doi/abs/10.1089/cmb.2012.0021). ORFs with a methionine start codon (ATG) were identified using CLC Genomics Workbench. The ORFs were annotated by performing a BLASTx search against the NCBI nr database. All ORFs intact at their 5' end but compromising less than 80% of orthologue length are annotated as truncated genes, while those ORFs disrupted at their 5' end are annotated as fragmented genes (Hendrickson *et al*., 2010). Fragmented ORFs are annotated with suffixes 1_2_3 etc,

ORFs in the reverse orientation are annotated as reading from START codon to STOP codon not in the order the fragments appear (Figure 2.2-2.7). Analysis and comparison of genome sequences were performed in CLC Genomics Workbench version 9.0.1 (https://qiagenbioinformatics.com), dot plots were generated using Gepard 1.4 with window size=10 (https://academic.oup.com/bioinformatics/article/23/8/1026/198110). Whole-genome multiple sequence alignment (MSA) was performed using MAFFT in Geneious Prime version 1.2 (https://www.geneious.com), and indels (insertions and deletions) were visualised using base-by-base (Tu *et al*., 2018).

2.3. RESULTS

2.3.1. Sequencing statistics

Table 2.2 outlines the sequencing statistics obtained from the Illumina Hiseq and Miseq platforms, and the contig lengths that resulted from *de novo* assembly with the SPAdes genome assembler. The RSA_2_2008 isolate is the longest compared to other isolates, and the highest average coverage was achieved with the RSA_2_2008 isolate with the Illumina Hiseq platform.

Table 2. 2. Sequencing and basic genome statistics of six novel ASFV genomes.

	ASFV Isolate Names					
	RSA_W1_1999	**RSA_2_2004**	**RSA_2_2008**	**LIV_5_40**	**SPEC_57**	**Zaire**
Year	1999	2004	2008	1983	1985	1977
Country	South Africa	South Africa	South Africa	Zambia	South Africa	Zaire
Genotype	IV	XX	XXII	I	III	XX
Host	Warthog	European Wild boar	Tick	Tick	Tick	Domestic pig
Cell line	Macrophages	Macrophages	Macrophages	Macrophages	Macrophages	Vero cells
# of passages in this study	4 times	4 times	2 times	4 times	4 times	3 times (Total # of passages: 23)
Seq. platform	Miseq	Miseq	Hiseq	Hiseq	Hiseq	Miseq
Total # of reads	3 726 962	3 749 576	9 914 962	22 993 178	36 764 148	8 525 098
# of virus specific reads	2 236 520	2 156 036	7 322 880	2 234 811	3 487 567	4 277 735
Read length (bp)	300	300	250	250	250	300
Coverage	3621 x	3431 x	9623 x	3048 x	4685 x	6924 x
GenBank accession #	MN641876	MN641877	MN336500	MN318203	MN394630	MN630494
SRA accession #	PRJNA 587581	PRJNA 587591	PRJNA 577445	PRJNA 577538	PRJNA 577546	PRJNA 587577

2.3.2. Basic genome structure and statistics

Table 2.3 outlines basic genome sequence statistics of the study isolates compared to publicly available reference sequences of the same genotypes. The longest novel genome was RSA_2_2008 (190, 065 bp), and the shortest was LIV_5_40 (183, 292 bp). RSA_W1_1999 and Zaire isolates have the same number of ORFs as the Warthog strain. Although the BA71V genome is shorter, it has a slightly higher coding density, and has a slightly higher GC content than the other sequences. Table 2.4 outlines ORFs that are present, shown with a (+) symbol and absent, with a (-), and highlighted in green and orange, respectively. Truncated ORFs are highlighted in red, and fragmented ORFs are highlighted in blue.

Overall genome structure: We have sequenced the coding regions of six novel ASFV isolates; however, the inverted terminal repeats (ITR) and the hairpin loop sequences of the terminal ends, were not sequenced. Cell-culture experiments were carried out at the ARC_OVI (TAD) laboratory before these isolates were used in this study, which classified the Zaire isolate as cell-culture adapted and non-haemadsorbing, while the other isolates were classified as haemadsorbing. A BLASTx search of the identified ORFs against the NCBI nr database revealed 191, 164, 166, 168, 165, and 164 predicted ORFs for LIV_5_40, RSA_W1_1999_RSA_2_2004, RSA_2_2008, SPEC_57, and Zaire, respectively (Figure 2.2-2.7). Overall, gene content and organisation were similar between all genome sequences, but differences were observed at the terminal ends, where most of the MGFs are found. LIV_5_40, RSA_W1_1999, RSA_2_2004, RSA_2_2008, SPEC_57, and Zaire isolates have a total of 42, 42, 41, 44, 43, and 43 MGF members, respectively. We further observed a truncation of the EP153R and EP402R ORFs in the LIV_5_40 isolate, highlighted in red, and fragmentation of the EP402R ORF in the Zaire isolate, highlighted in blue. Differences in the repetitive amino acid patterns were observed in the central variable region (CVR) of the B602L gene. Commonly used genes for phylogenetic analysis are highlighted in purple colour, the five novel genes (EP364R, B169L, B117L, B407L, and I177L), found in the CCR, which were less conserved and identified as potential novel phylogenetic markers (Chapter 3), are highlighted with a green colour. Study isolate LIV_5_40 encodes more fragmented ORFs highlighted in blue, compared to other study isolates. There are 109 conserved ORFs among these study isolates, which are involved in structural morphogenesis, nucleotide metabolism, transcription, replication and repair.

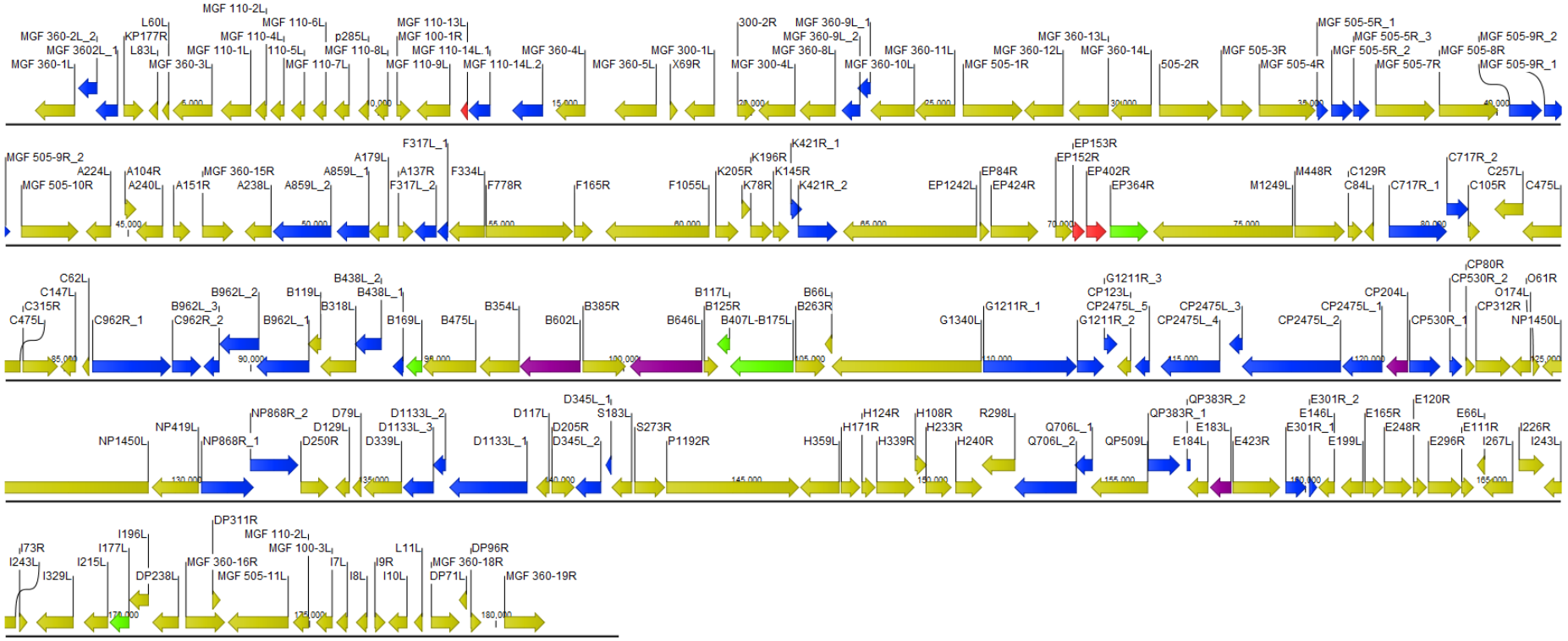


Figure 2. 2. Genome organisation of the ASFV LIV_5_40 isolate. ORFs are shown as arrows to indicate their size and direction of transcription. Fragmented ORFs are highlighted in blue colour, truncated ORFs are highlighted in red colour, commonly used ORFs for phylogenetic analysis are highlighted in purple, and ORFs identified as novel phylogenetic markers are highlighted in green colour.

Figure 2. 3. Genome organisation of the ASFV RSA_W1_1999 isolate. ORFs are shown as arrows to indicate their size and direction of transcription. Fragmented ORFs are highlighted in blue colour, commonly used ORFs for phylogenetic analysis are highlighted in purple, and ORFs identified as novel phylogenetic markers are highlighted in green.

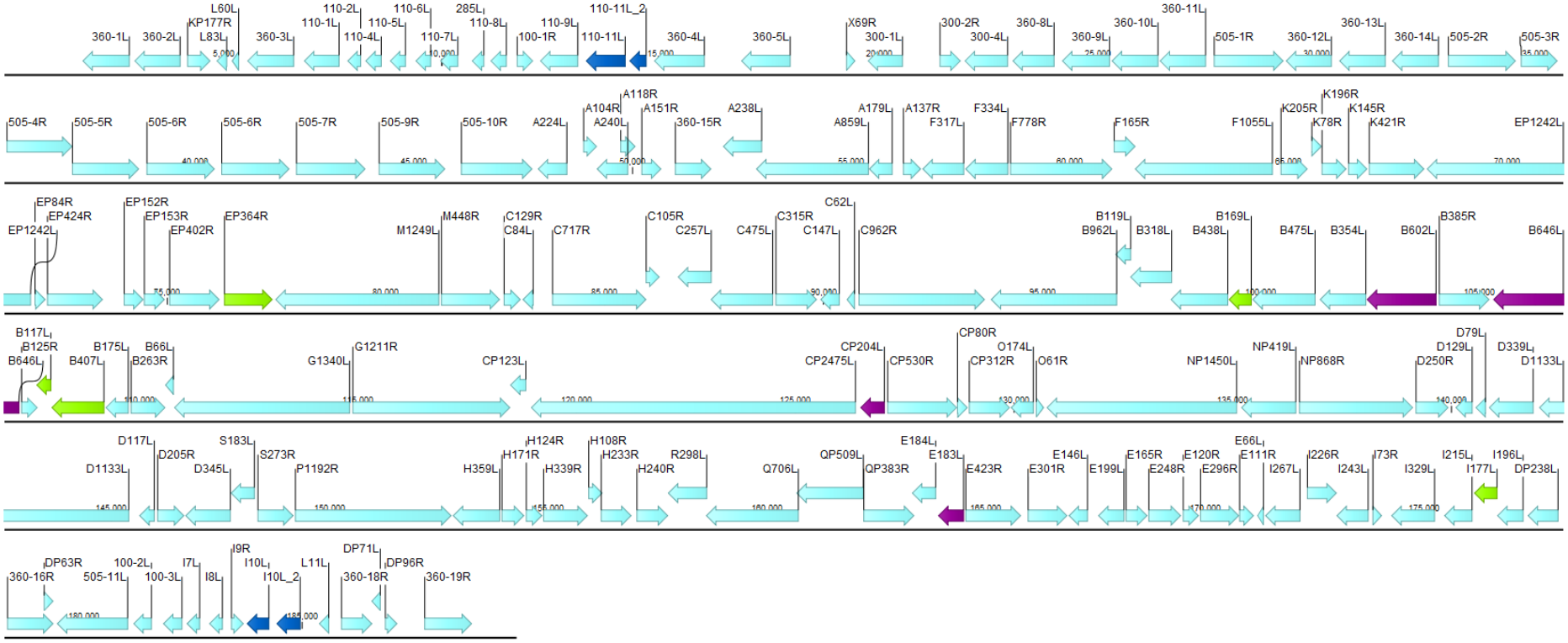


Figure 2. 4. Genome organisation of the ASFV RSA_2_2004 isolate. ORFs are shown as arrows to indicate their size and direction of transcription. Fragmented ORFs are highlighted in blue colour, commonly used ORFs for phylogenetic analysis are highlighted in purple, and ORFs identified as novel phylogenetic markers are highlighted in green colour.

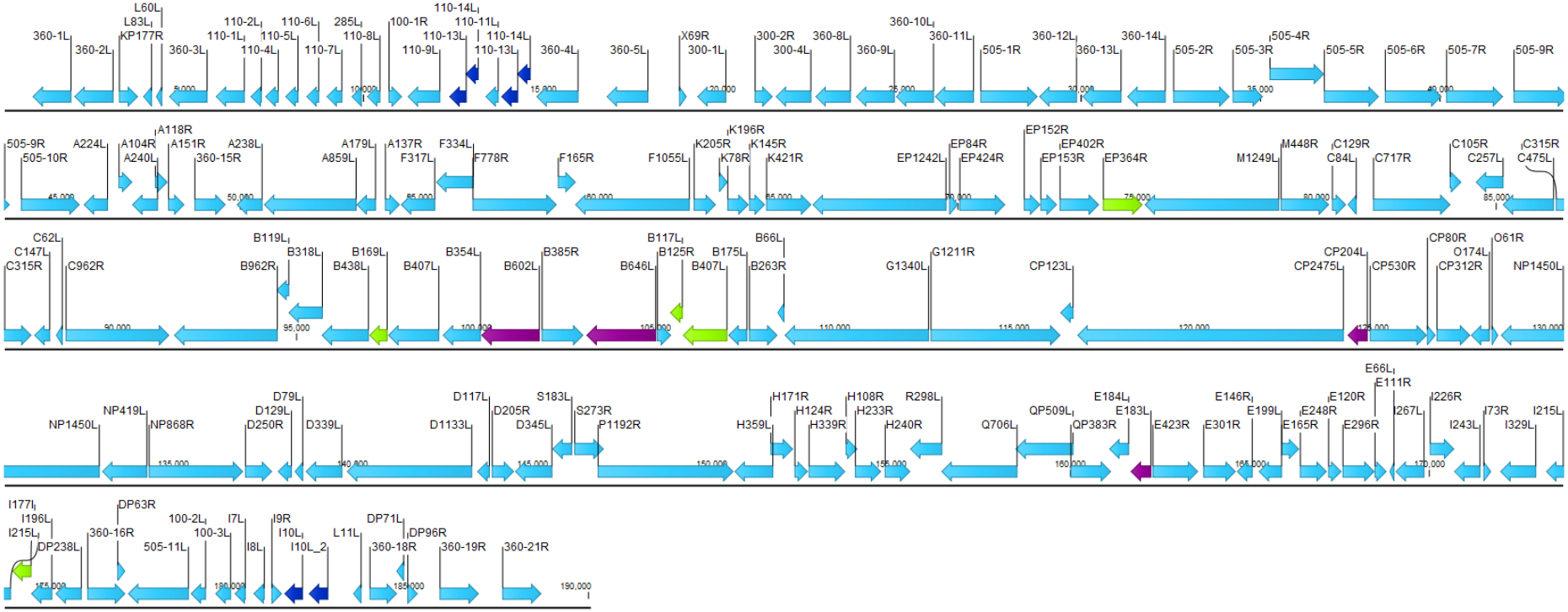


Figure 2. 5. Genome organisation of the ASFV RSA_2_2008 isolate. ORFs are shown as arrows to indicate their size and direction of transcription. Fragmented ORFs are highlighted in blue colour, commonly used ORFs for phylogenetic analysis are highlighted in purple, and ORFs identified as novel phylogenetic markers are highlighted in green colour.

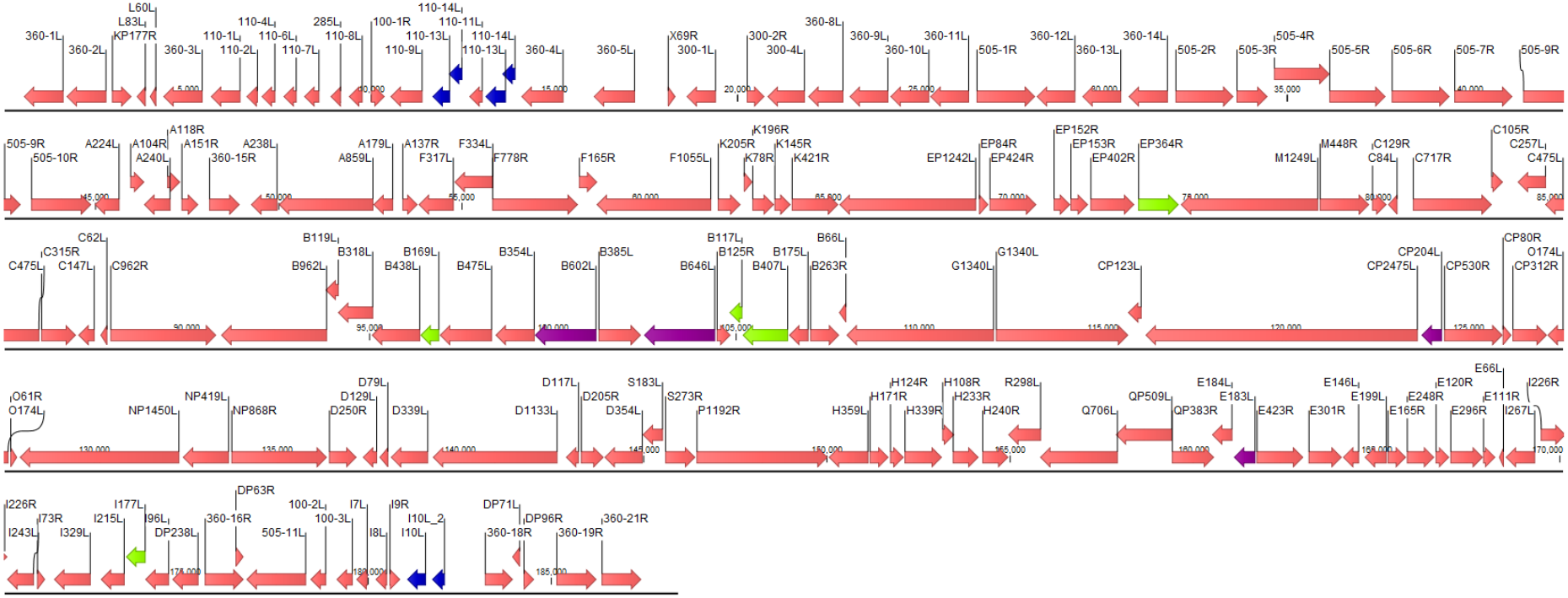


Figure 2. 6. Genome organisation of the ASFV SPEC_57 isolate. ORFs are shown as arrows to indicate their size and direction of transcription. Fragmented ORFs are highlighted in red colour, commonly used ORFs for phylogenetic analysis are highlighted in purple, and ORFs identified as novel phylogenetic markers are highlighted in green colour.

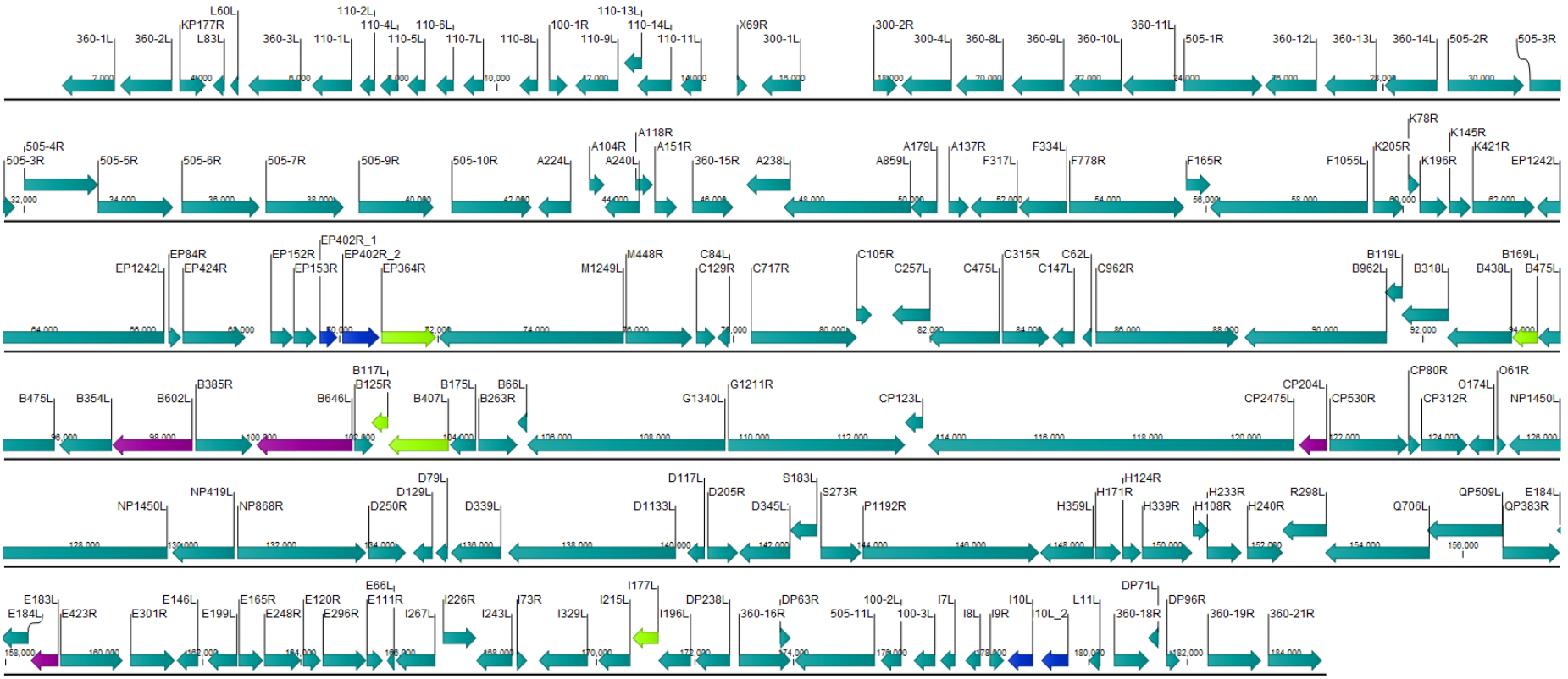


Figure 2. 7. Genome organisation of the ASFV Zaire isolate. ORFs are shown as arrows to indicate their size and direction of transcription. Fragmented ORFs are highlighted in blue colour, commonly used ORFs for phylogenetic analysis are highlighted in purple, and ORFs identified as novel phylogenetic markers are highlighted in green colour.

Table 2. 3. Genome sequence statistics of six ASFV study isolates and publicly available reference sequences of the same/similar genotypes

	RSA_W1_ 1999	**Warthog**	**RSA_ 2_2008**	**RSA_ 2_2004**	**Pretorius-kop**	**Zaire**	**BA71V**	**LIV_5_40**	**Mkuzi_ 1979**	**SPEC_57**	**Warm-baths**
Year	1999	1980	2008	2004	1996	1977	1971	1983	1979	1985	1987
Country	South Africa	Namibia	South Africa	South Africa	South Africa	Zaire	Spain	Zambia	South Africa	South Africa	South Africa
Genotype	IV	IV	XXII	XX	XX	XX	I	I	I	III	III
Host	Warthog	Warthog	Tick	European Wild boar	Tick	Domestic pig	Domestic Pig	Tick	Tick	Tick	Tick
Genome length (bp)	187 621	186 528	190 065	189 903	190 324	184 820	170 101	183 292	192 714	188 458	190 244
LVR (bp)	44 533	43 606	46 310	48 508	47 596	43 110	29 795	44 541	48 676	45 653	46 773
CCR (bp)	129 289	129 289	129 421	129 454	121 808	129 482	122 033	126 480	121 297	129 591	121 424
RVR (bp)	15 065	14899	15 597	13 219	14 365	13 569	11 927	13 131	15 785	14 537	15 849
# of ORFs	164	164	168	166	167	164	160	191	167	165	167
G + C content (%)	38,6	38,6	38,5	38,4	38,4	38,6	38,9	38,6	38,4	38.5	38,4

Table 2. 4. Open reading frames of the six study isolates and the BA71V reference sequence. ORF status: absent: - and orange, present: + and green, truncated: red, fragmented: blue. Isolates: **1**-LIV_5_40, **2**-RSA_W1_1999, **3**-RSA_2_2004, **4**-RSA_2_2008, **5**-SPEC_57, **6**-Zaire, and 7-BA71V.

ORFs	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Description/ putative function**
DP60R	-	-	-	-	-	-	+	Putative protein
DP86L	-	-	-	-	-	-	+	Transmembrane protein
KP93L	-	-	-	-	-	-	+	Transmembrane protein
360-1L	+	+	+	+	+	+	+	MGF 360
360-2L	+	-	+	+	+	+	+	MGF 360
KP177R	+	+	+	+	+	+	+	Structural protein p22
L83L	+	+	+	+	+	+	+	Putative protein
L60L	+	+	+	+	+	+	+	Putative protein
360-3L	+	+	+	+	+	+	+	MGF 360
110-1L	+	+	+	+	+	+	+	MGF 110
110-2L	+	+	+	+	+	+	+	MGF110
110-4L	+	+	+	+	+	+	+	MGF 110
110-5L	+	+	+	+	+	+	+	MGF 110
110-6L	+	+	+	+	+	+	+	MGF 110
110-7L	+	+	+	+	+	+	+	MGF 110
P285L	+	+	+	+	+	+	+	MGF 110
110-8L	+	+	+	+	+	+	-	MGF 110
100-1R	+	+	+	+	+	+	-	MGF 100
110-9L	+	+	+	+	+	+	-	MGF 110
110-11L	-	+	+	+	-	-	-	MGF 110
110-13L	+	-	-	+	+	+	-	MGF 110
110-14L	+	-	-	+	+	+	-	MGF 110
360-4L	+	+	+	+	+	-	-	MGF 360
360-5L	+	-	+	+	+	-	-	MGF 360
X69R	+	+	+	+	+	+	+	Transmembrane protein
300-1L	+	+	+	+	+	+	+	MGF 300
300-2R	+	+	+	+	+	+	+	MGF 300
300-4L	+	+	+	+	+	+	+	MGF 300
360-6L	-	-	-	-	-	-	+	MGF 360
360-8L	+	+	+	+	+	+	+	MGF 360
360-9L	+	+	+	+	+	+	+	MGF 360
360-10L	+	+	+	+	+	+	-	MGF 360
360-11L	+	+	+	+	+	+	-	MGF 360
505-1R	+	+	+	+	+	+	-	MGF 505
360-12L	+	+	+	+	+	+	-	MGF 360
360-13L	+	+	+	+	+	+	-	MGF 360
360-14L	+	+	+	+	+	+	-	MGF 360
505-2R	+	+	+	+	+	+	+	MGF 505/530
505-3R	+	+	+	+	+	+	+	MGF 505/530
505-4R	+	+	+	+	+	+	+	MGF 505/530
505-5R	+	+	+	+	+	+	+	MGF 505/530
505-6R	-	+	+	+	+	+	+	MGF 505/530
505-7R	+	+	+	+	+	+	+	MGF 505/530
505-8R	+	-	-	-	-	-	-	MGF 505/530
505-9R	+	+	+	+	+	+	+	MGF 505/530
505-10R	+	+	+	+	+	+	+	MGF 505/530
A224L	+	+	+	+	+	+	+	IAP apoptosis inhibitor
104R	+	+	+	+	+	+	+	Histone-like protein
A240L	+	+	+	+	+	+	+	Thymidylate kinase
A118R	-	+	+	+	+	+	+	Putative
A151R	+	+	+	+	+	+	+	Component of redox pathway

ORFs	1	2	3	4	5	6	7	Description/ putative function
360-15R	+	+	+	+	+	+	+	MGF 360
A238L	+	+	+	+	+	+	+	Inhibitor of host gene transcription
A859L	+	+	+	+	+	+	+	Member of helicase superfamily II
A179L	+	+	+	+	+	+	+	Bcl 2 apoptosis inhibitor
A137R	+	+	+	+	+	+	+	Structural protein p11.5
F317L	+	+	+	+	+	+	+	Putative protein
F334L	+	+	+	+	+	+	+	Ribonucleotide reductase small subunit
F778R	+	+	+	+	+	+	+	Ribonucleotide reductase large subunit
F165R	+	+	+	+	+	+	+	dUTPase
F1055L	+	+	+	+	+	+	+	Member of helicase superfamily II
K205R	+	+	+	+	+	+	+	Putative protein
K78R	+	+	+	+	+	+	+	Structural protein p10
K196R	+	+	+	+	+	+	+	Thymidine kinase
K145R	+	+	+	+	+	+	+	Putative protein
K421R	+	+	+	+	+	+	+	Putative protein
EP1242L	+	+	+	+	+	+	+	RNA polymerase subunit 2
EP84R	+	+	+	+	+	+	+	Transmembrane protein
EP424R	+	+	+	+	+	+	+	FTS J-like methyl transferase domain
EP152R	+	+	+	+	+	+	+	Transmembrane protein
EP153R	+	+	+	+	+	+	+	C-type lectin-like
EP402R	+	+	+	+	+	+	+	Causes Haemadsorption
EP364R	+	+	+	+	+	+	+	ERCC4 nuclease domain
M1249L	+	+	+	+	+	+	+	Putative protein
M448R	+	+	+	+	+	+	+	Putative protein
C129R	+	+	+	+	+	+	+	Putative protein
C84L	+	+	+	+	+	+	+	Putative protein
C717R	+	+	+	+	+	+	+	Putative protein
C105R	+	+	+	+	+	+	+	Putative protein
C257R	+	+	+	+	+	+	+	Transmembrane protein
C475L	+	+	+	+	+	+	+	PolyA polymerase large subunit
C315R	+	+	+	+	+	+	+	TFIIB like
C147L	+	+	+	+	+	+	+	RNA polymerase subunit 6
C62L	+	+	+	+	+	+	+	Putative protein
C962R	+	+	+	+	+	+	+	DNA primase
B962L	+	+	+	+	+	+	+	Member of helicase superfamily II
B119L	+	+	+	+	+	+	+	ERV 1-like. Involved in redox metabolism
B318L	+	+	+	+	+	+	+	Prenyltransferase
B438L	+	+	+	+	+	+	+	Formation of vertices in icosahedral capsid
B169L	+	+	+	+	+	+	+	Transmembrane protein
B475L	+	+	+	+	+	+	+	Putative protein
B354L	+	+	+	+	+	+	+	Putative protein
B602L	+	+	+	+	+	+	+	Chaperone. Involved in folding of capsid.
B385R	+	+	+	+	+	+	+	VV A2L-like transcription factor
B646L	+	+	+	+	+	+	+	Structural protein p72
B125R	+	+	+	+	+	+	+	Putative
B117L	+	+	+	+	+	+	+	Transmembrane protein
B407L	+	+	+	+	+	+	+	Putative protein
B175L	+	+	+	+	+	+	+	Putative transcription factor
B263R	+	+	+	+	+	+	+	Putative protein
B66L	+	+	+	+	+	+	+	Transmembrane protein
G1340L	+	+	+	+	+	+	+	VV A8L-like transcription factor
G1211R	+	+	+	+	+	+	+	DNA polymerase family B
CP123L	+	+	+	+	+	+	+	Transmembrane protein
CP2475L	+	+	+	+	+	+	+	Precursor of p150, p37, p14 and p34
CP204L	+	+	+	+	+	+	+	ASFV-induced protein p32
CP530R	+	+	+	+	+	+	+	Polyprotein precursor of p35 and p15
CP80R	+	+	+	+	+	+	+	RNA polymerase subunit 10

ORFs	1	2	3	4	5	6	7	Description/ putative function
CP312R	+	+	+	+	+	+	+	Putative protein
0174L	+	+	+	+	+	+	+	DNA polymerase X-like
061R	+	+	+	+	+	+	+	P12 attachment protein
NP1450L	+	+	+	+	+	+	+	RNA polymerase subunit 1
NP419L	+	+	+	+	+	+	+	DNA ligase
NP868R	+	+	+	+	+	+	+	Guanylyltransferase
D205R	+	+	+	+	+	+	+	8-hydroxy-dGTPase
D129L	+	+	+	+	+	+	+	Putative protein
D79L	+	+	+	+	+	+	+	Putative protein
D339L	+	+	+	+	+	+	+	Putative protein
D1133L	+	+	+	+	+	+	+	Helicase
D117L	+	+	+	+	+	+	+	Structural protein p17
D205R	+	+	+	+	+	+	+	RNA polymerase subunit 5
D345L	+	+	+	+	+	+	+	Lambda-like exonuclease
S183L	+	+	+	+	+	+	+	Putative protein
S273R	+	+	+	+	+	+	+	Ulp1 protease family
P1192R	+	+	+	+	+	+	+	Topoisomerase II
H359L	+	+	+	+	+	+	+	RNA polymerase subunit 3
H171R	+	+	+	+	+	+	+	Putative protein
H124R	+	+	+	+	+	+	+	Putative protein
H339R	+	+	+	+	+	+	+	Putative protein
H108R	+	+	+	+	+	+	+	J5R. Transmembrane domain
H233R	+	+	+	+	+	+	+	Transmembrane protein
H240R	+	+	+	+	+	+	+	Putative protein
R298L	+	+	+	+	+	+	+	Serine protein kinase
Q706L	+	+	+	+	+	+	+	Helicase
QP509L	+	+	+	+	+	+	+	Helicase
QP383R	+	+	+	+	+	+	+	NifS-like protein
E184L	+	+	+	+	+	+	+	Putative protein
E183L	+	+	+	+	+	+	+	Structural protein p54
E423R	+	+	+	+	+	+	+	Putative
E301R	+	+	+	+	+	+	+	Proliferating cell nuclear antigen protein.
E146L	+	+	+	+	+	+	+	Transmembrane protein
E199L	+	+	+	+	+	+	+	Transmembrane protein
E165R	+	+	+	+	+	+	+	dUTPase
E248R	+	+	+	+	+	+	+	Structural protein
E120R	+	+	+	+	+	+	+	Structural protein p14.5
E296R	+	+	+	+	+	+	+	AP endonuclease class II
E111R	+	+	+	+	+	+	+	Putative protein
E66L	+	+	+	+	+	+	+	Putative protein
I267L	+	+	+	+	+	+	+	Putative protein
I226R	+	+	+	+	+	+	+	Putative protein
I243L	+	+	+	+	+	+	+	Transcription factor SII homolog
I73R	+	+	+	+	+	+	+	Putative protein
I329L	+	+	+	+	+	+	+	Inhibitor of the TLR 3 signaling pathway
I215L	+	+	+	+	+	+	+	Ubiquitin-conjugating enzyme
I177L	+	+	+	+	+	+	+	Transmembrane protein
I196L	+	+	+	+	+	+	+	Putative protein
DP238L	+	+	+	+	+	+	+	Putative protein
360-16R	+	+	+	+	+	+	+	MGF 360
DP63R	+	+	+	+	+	+	+	MGF 360
505-11L	+	+	+	+	+	+	+	MGF 5050/530
100-2L	+	+	+	+	+	+	+	MGF 100
100-3L	+	+	+	+	+	+	+	MGF 100
I7L	+	+	+	+	+	+	-	Putative protein
I8L	+	+	+	+	+	+	-	Putative protein
I9R	+	+	+	+	+	+	-	Putative protein

ORFs	1	2	3	4	5	6	7	Description/ putative function
I10L	+	+	+	+	+	+	-	structural protein
LIIL	+	+	+	+	-	+	-	Putative protein
360-18R	+	+	+	+	+	+	+	MGF 360
DP71L	+	+	+	+	+	+	+	Similar to HSV ICP34.5
DP96R	+	+	+	+	+	+	+	Putative protein
360-19R	+	+	+	+	+	+	+	MGF 360
360-20R	-	-	-	-	-	-	+	MGF 360
360-21R	-	+	-	+	+	+	-	MGF 360

(+) for present genes, highlighted in green colour, (-) for absent genes, highlighted in orange colour. Red colour represents truncated genes, while blue colour represents fragmented ORFs.

2.3.3. Comparative genomics

The six novel study isolates were first compared to publicly available reference sequences of the same genotype (with the exception of the RSA_2_2008, which is a genotype XXII), to identify regions where the collinear structure is disrupted. The RSA_2_2008, which represents genotype XXII, was compared with the Pretoriuskop_96_4 isolate, which is a genotype XX isolate because they cluster closely together in the phylogenetic tree based on the C-terminus sequencing of the p72 gene (see Chapter 3), and further compared to study sequence RSA_2_2004, presenting genotype XX. Dot plot analysis of the six study isolates compared to reference sequences of the same genotype shows that the genomes are mostly collinear. However, a disruption of the collinearity at the LVR, CCR, and RVR was observed in LIV_5_40, and Zaire isolates, affecting the EP153R and EP402R ORFS in LIV_5_40, and disruption at the CCR was observed in the RSA_2_2004. Disruption at the LVR was observed in the RSA_W1_1999 and RSA_2_2008 isolates, and at the RVR was found in SPEC_57 isolate (Figures 2.8-2.11). The Base-by-base analysis showed that sequence insertions and sequence deletions (indels) are found at the terminal ends, and mostly at the LVR (Figure 2.12). We further observed that genotypes VIII (Malawi_Lil_20_1) and IX (N10, R7, R8, R25, R35, and Ken06.Bus) have a deletion between position 5300 and 5500. This deletion is also found in genotype I isolate BA71V, but not observed in other isolates. Comparing the study genome sequences with reference sequences of the same genotype showed that the LIV_5_40 is more similar to Mkuzi_1979 than to other genotype I isolates.

The SPEC_57 isolate is similar to Warmbaths, and RSA_W1_1999 is similar to Warthog, which represents genotype III and IV, respectively. The Zaire isolate contains a gapped region between position 20,000 and 30,000, which is not observed in other genotype XX isolates (Pretoriuskop_96_4 and RSA_2_2004). However, RSA_2_2008, which is a genotype XXII isolate, is similar to Pretoriuskop_96_4 and RSA_2_2004 (Figure 2.13). The cell-culture adapted Zaire isolate was further compared to attenuated strains BA71V, NHV, and

OURT 88_3. The comparison reveals a gapped region between position 38,000 and 50,000 in NHV and OURT 88_3 isolates not found in the Zaire isolate (Figure 2.14). These indels observed, affect members of the MGFs. Some of the affected genes are the same among the three isolates and will be further discussed in Chapter 4.

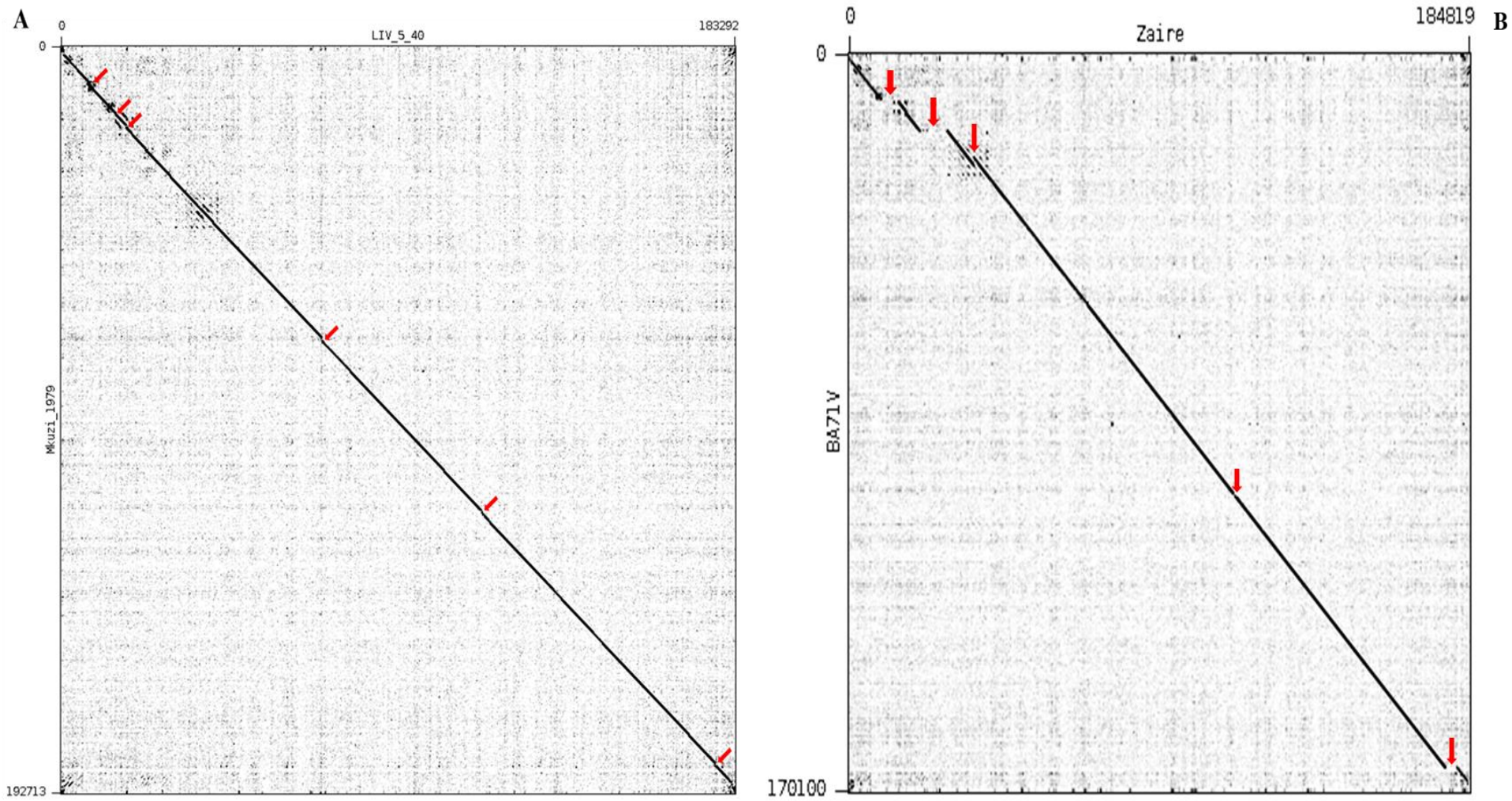


Figure 2. 8. Dot plots of the study isolates LIV_5_40 and Zaire genomes (x-axis) vs. reference genome sequences (y-axis). A) Mkuzi_1979 vs. LIV_5_40 B) BA71V vs Zaire. Major and minor deletions can be seen as breaks in the diagonal line, highlighted with red arrows.

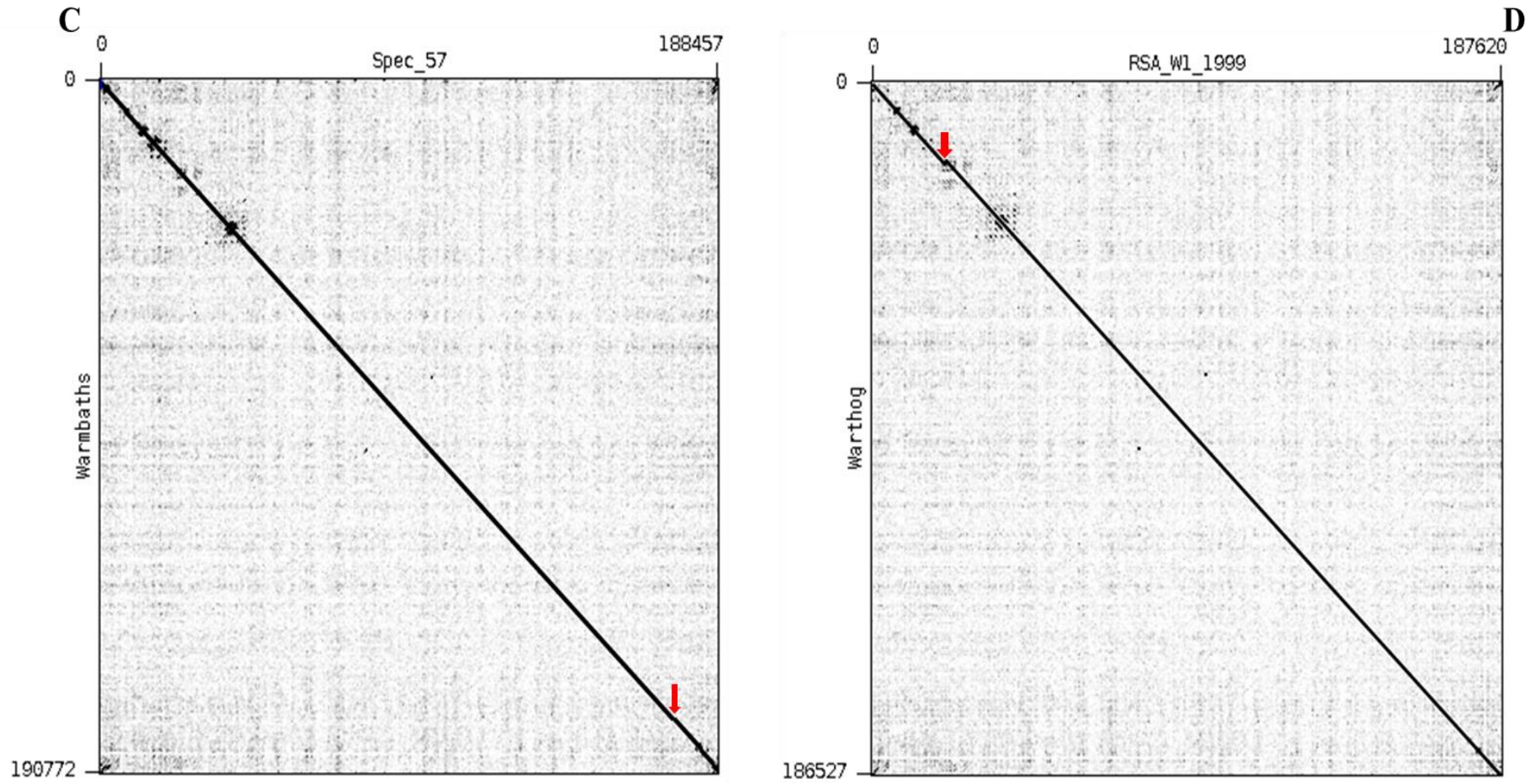


Figure 2. 9. Dot plots of the study isolates SPEC_57 and RSA_W1_1999 genomes (x-axis) vs. reference genome sequences (y-axis). C) Warmbaths vs. SPEC_57 D) Warthog vs. RSA_W1_1999. Major and minor deletions can be seen as breaks in the diagonal line, highlighted with red arrows.

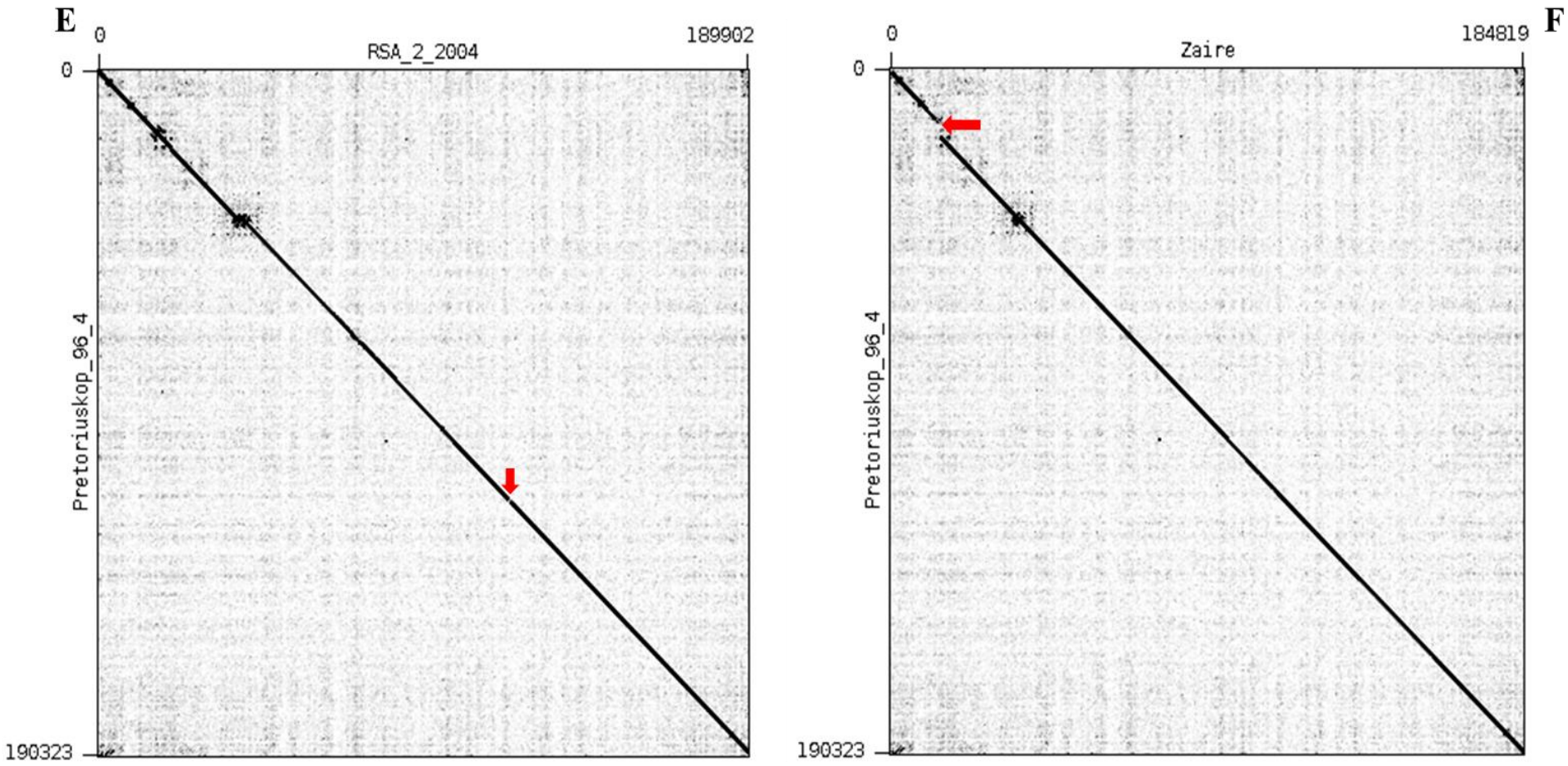


Figure 2. 10. Dot plots of the study isolates RSA_2_2004 and Zaire genome (x-axis) vs. reference genome sequences (y-axis). E) Pretoriuskop_96_4 vs. RSA_2_2004 F) Pretoriuskop_96_4 vs. Zaire. Major and minor deletions can be seen as breaks in the diagonal line, highlighted with red arrows.

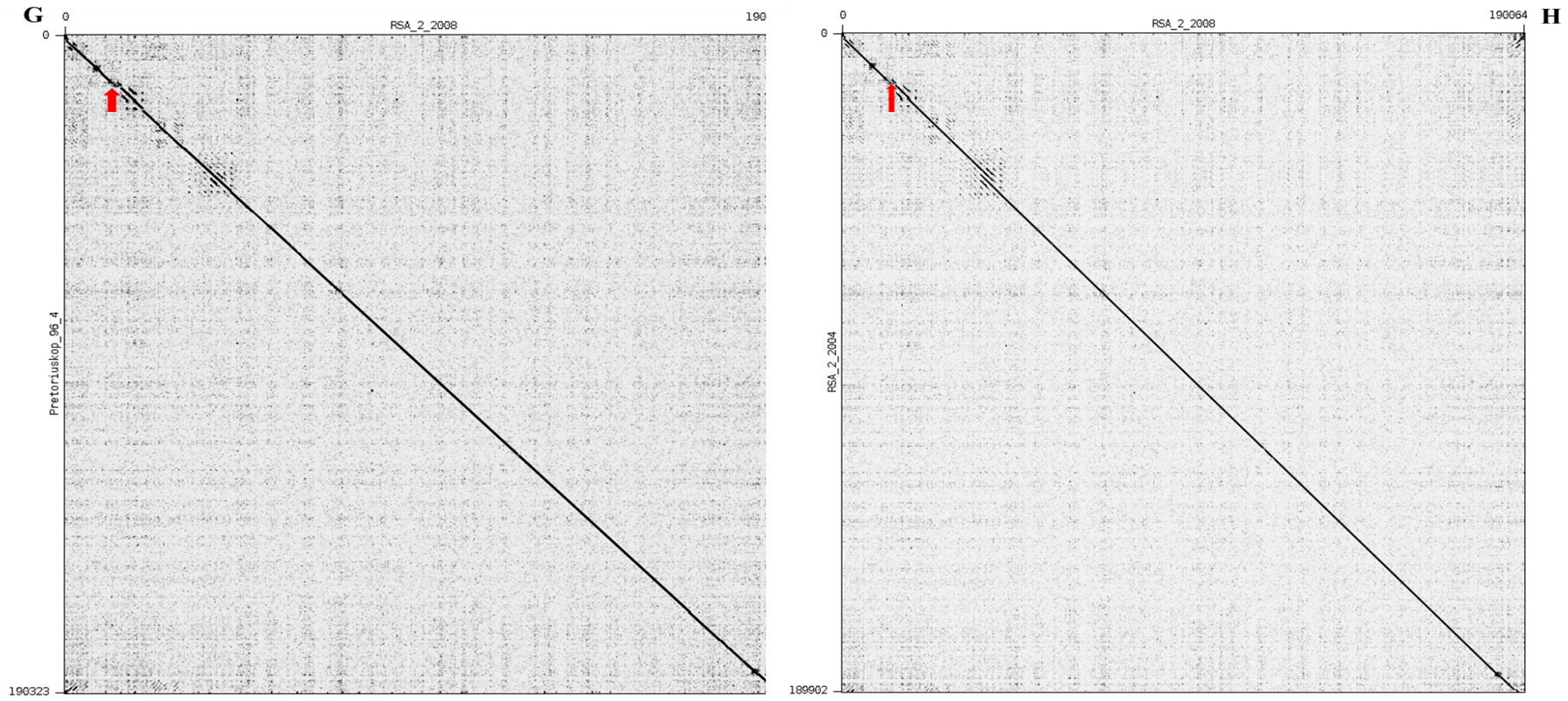


Figure 2. 11. Dot plots of the study isolate RSA_2_2008 genome (x-axis) vs. reference genome sequences (y-axis). E) Pretoriuskop_96_4 vs. RSA_2_2008 F) RSA_2_2008 vs. RSA_2_2004. Major and minor deletions can be seen as breaks in the diagonal line, highlighted with red arrows.

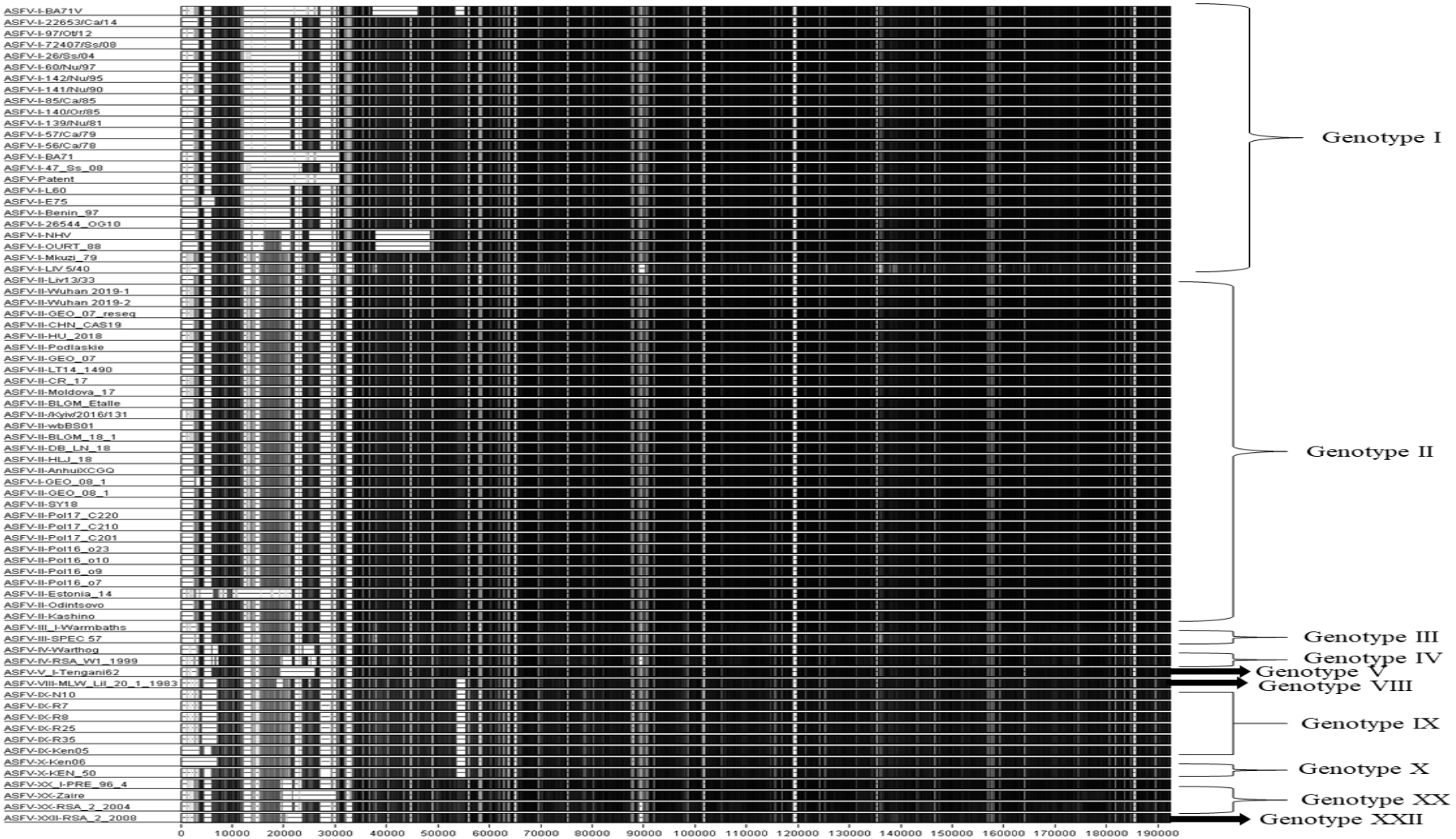


Figure 2. 12. Visual summary of the alignment of the 73 ASFV isolate used in this study. The top greyscale track indicates similarity (black-perfect or 100 match; white-low or 0% match), and gaps are shown as dashes.

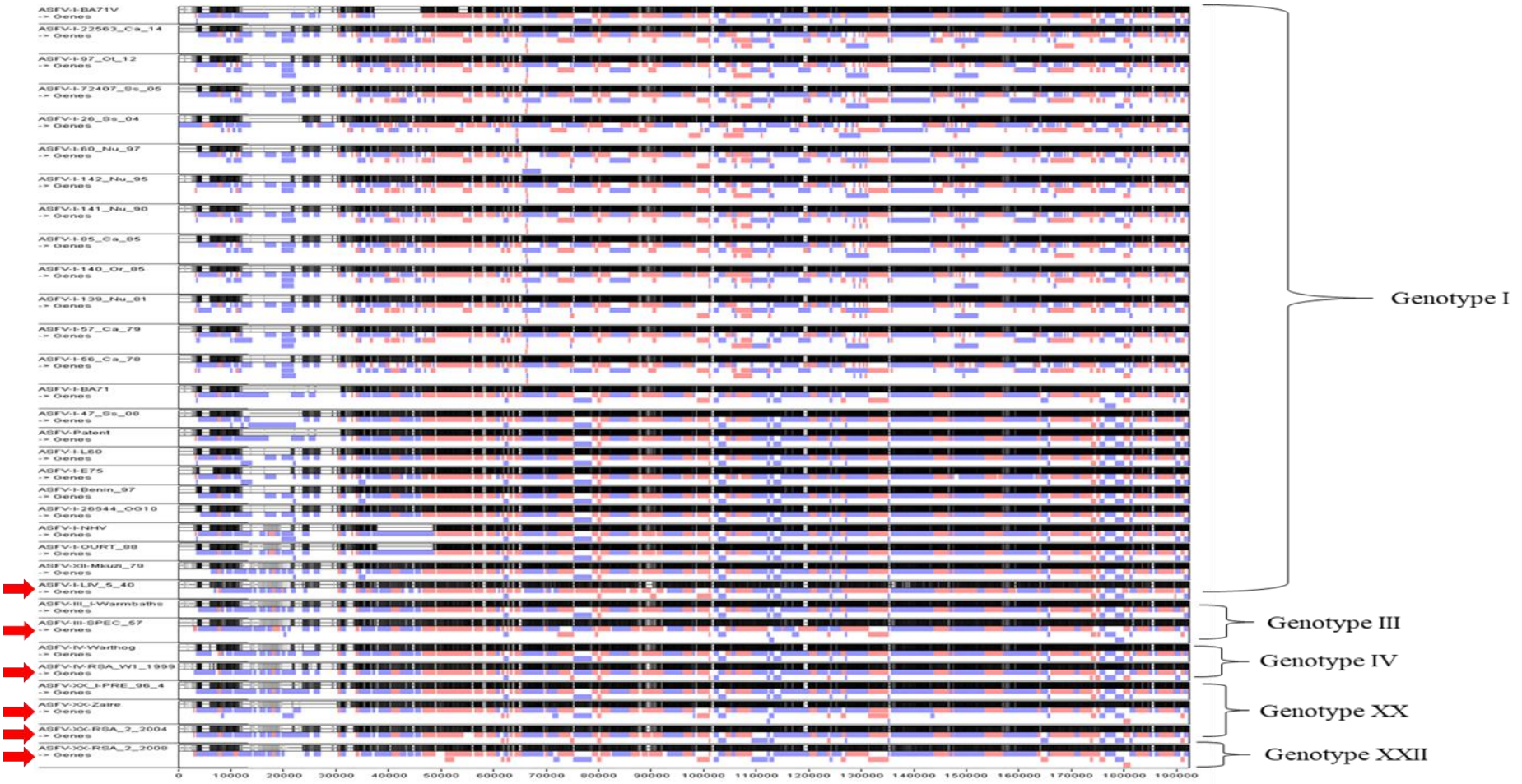


Figure 2. 13. Visual summary of the alignment of the six novel study isolate genomes vs. reference sequences of the same genotype. The red arrows indicate the location of the study isolates. The top greyscale track indicates similarity (black-perfect or 100 match; white-low or 0% match), and gaps are shown as dashes. The track showing pink and purple boxes indicates top and bottom strand open reading frames, respectively.

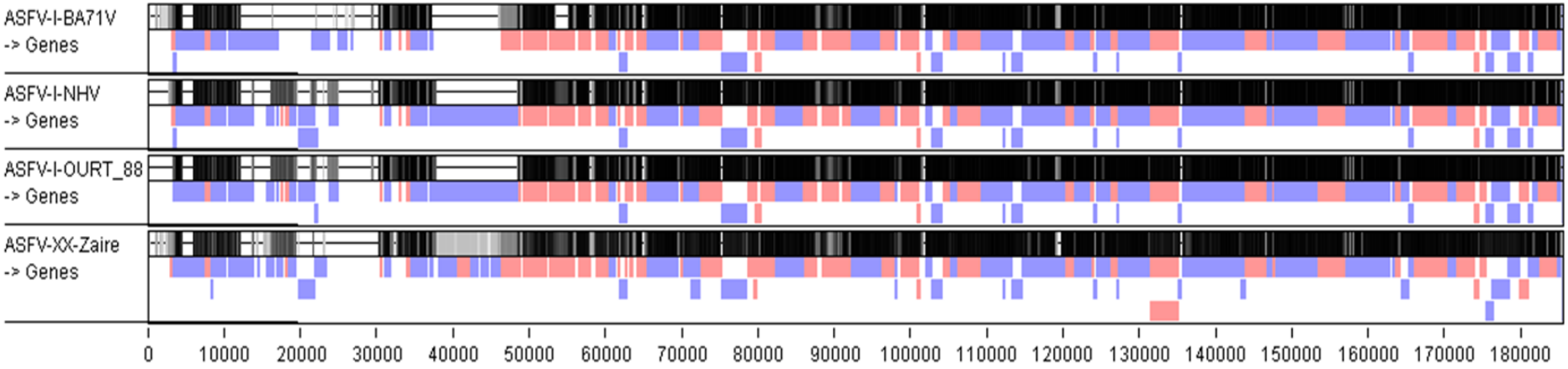


Figure 2. 14. Visual summary of the alignment of the Zaire study isolate genome vs. Ba71V, NHV, and OURT_88_3. The top greyscale track indicates similarity (black-perfect or 100 match; white-low or 0% match), and gaps are shown as dashes. The track showing pink and purple boxes indicates top and bottom strand open reading frames, respectively.

Left variable region (LVR) and the right variable region (RVR): Multiple sequence alignments were analysed to determine gene content. Dot plot analysis reveals that indels occurred mostly at the terminal ends of the genome. At the RVR of the genome, the RSA_W1_1999, RSA_2_2004, RSA_2_2008, SPEC_57, and Zaire isolates encode two fragments of the I10L ORF. This fragmented ORF is also found in Kenya_1950 (GenBank accession number: AY261360), Ken05/Tk1 (GenBank accession number: KM111294), Mkuzi_1979 (GenBank accession number: AY261362), Pretoriuskop_96_4 (GenBank accession number: AY261363), Tengani62 (GenBank accession number: AY261364), and Warthog (GenBank accession number: AY261366). A significant difference is observed in members of the MGFs encoded by the study isolates (See Chapter 4).

Central conserved region (CCR): The CCR starts at the A224L ORF, which encodes an apoptosis inhibitor (IAP) homologue and ends at the I196L ORF. A total of 109 ORFs (Table 2.5) are conserved in the 73 isolates analysed, and 44 of these ORFs are of unknown function. The identified ORFs encode proteins involved in nucleotide metabolism, transcription, replication and repair, structural proteins, and proteins involved in morphogenesis.

The pairwise comparison of the concatenated nucleotide alignment of the 109 conserved ORFs showed nucleotide identity varying between 88% and 100% among ASFV isolates used in this study (Table 2.6). The EP153R and the EP402R genes, which are found adjacent to each other in ASFV genomes, are disrupted in the genomes of natural non-haemadsorbing, low virulence ASFV isolates, and are interrupted by frameshift mutation close to the N-terminus in both the NHV and the OURT 88_3 strains (Leitao *et al.*, 2001, Chapman *et al*., 2008). The EP153R gene in NHV (GenBank accession number: KM262845), and the OURT 88_3 3 (GenBank accession number: AM712240) encode a polypeptide that is 113 amino acids long. The EP153R gene in LIV_5_40, which is a haemadsorbing and virulent isolate, encodes a polypeptide that is 108 amino acids long, resulting from a truncation (Figure 2.15). The LIV_5_40 further encodes fifteen fragmented ORFs (Table 2.7) found in the CCR. The LIV_5_40 encodes two fragments in the opposite direction of the F317L ORF, and the B407L and B175L ORFs are fused. Although the CCR contains highly conserved genes (e.g. A104R and B646L), there are regions of the CCR that contain genes which are less conserved (*i.e.* B602L, CP204L (p30), and E183L (p540), which are used for enhanced discrimination of ASFV isolation in phylogenetic analysis) (Gallardo *et al*., 2009). In this study, five novel genes (EP364R, B169L, B117L, B407L, and I177L) that are less conserved were identified.

They will be discussed in chapter 3 as novel phylogenetic marker loci for phylogenetic analysis.

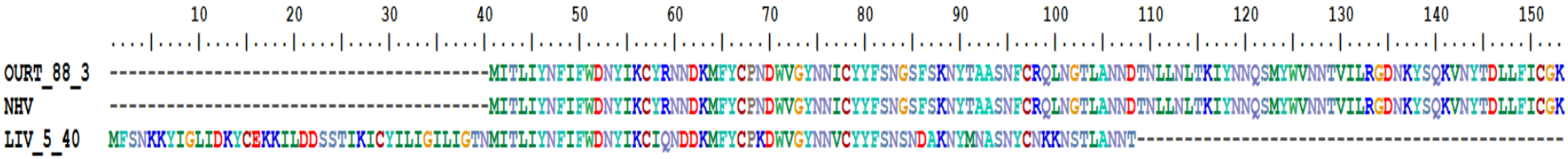


Figure 2. 15. Visual summary of the amino acid alignment of the EP153R protein encoded by the LIV_5_40 study isolate vs. NHV and OURT_88_3.

Table 2. 5. 109 conserved ORFs in the CCR of the 73 ASFV isolates used in this study.

ORFs	Description/functions	ORFs	Description/functions
A224L	IAP apoptosis inhibitor	B385R	VV A2L-like transcription factor
104R	Histone-like protein	B646L	Structural protein p72
A240L	Thymidylate kinase	B125R	Putative protein
A151R	pA151R. Component of redox pathway	B117L	Transmembrane protein
360-15R	Member of the MGF 360	B407L	Putative protein
A238L	Inhibitor of host gene transcription	B175L	Putative transcription factor
A859L	Helicase superfamily II	B263R	Putative protein
A179L	Bcl 2 apoptosis inhibitor	B66L	Transmembrane protein
A137R	Structural protein p11.5	G1340L	VV A8L-like transcription factor
F317L	Putative protein	G1211R	DNA polymerase family B
F334L	Ribonucleotide reductase small subunit	CP123L	Transmembrane protein
F778R	Ribonucleotide reductase large subunit	CP2475L	Precursor of p150, p37, p14 and p34.
F165R	dUTPase	CP204L	ASFV-induced protein p32
F1055L	Helicase superfamily II	CP530R	Polyprotein precursor of p35 and p15
K205R	Putative protein	CP80R	RNA polymerase subunit 10
K78R	Structural protein p10	CP312R	Putative protein
K196R	Thymidine kinase	0174L	DNA polymerase X-like
K145R	Putative protein	061R	P12 attachment protein
K421R	Putative protein	NP1450L	RNA polymerase subunit 1
EP1242L	RNA polymerase subunit 2	NP419L	DNA ligase
EP84R	Transmembrane protein	NP868R	Guanylyltransferase
EP424R	FTS J-like methyl transferase domain	D205R	8-hydroxy-dGTPase
EP152R	Transmembrane protein	D129L	Putative protein
EP153R	C-type lectin-like	D79L	Putative protein
EP402R	Causes haemadsorption	D339L	Putative protein
EP364R	ERCC4 nuclease domain	D1133L	Helicase protein
M1249L	Putative protein	D117L	Structural protein p17.
M448R	Putative protein	D205R	RNA polymerase subunit 5
C129R	Putative protein	D345L	Lambda-like exonuclease
C84L	Putative protein	S183L	Putative protein
C717R	Putative protein	S273R	Ulp1 protease family
C105R	Putative protein	P1192R	Topoisomerase II
C257R	Transmembrane protein	H359L	RNA polymerase subunit 3
C475L	PolyA polymerase large subunit	H171R	Putative protein
C315R	TFIIB like	H124R	Putative protein
C147L	RNA polymerase subunit 6	H339R	Putative protein
C62L	Putative	H108R	J5R. Transmembrane domain
C962R	DNA primase	H233R	Transmembrane protein
B962L	Helicase superfamily II	H240R	Putative protein
B119L	ERV 1-like. Involved in redox metabolism	R298L	Serine protein kinase
B318L	Prenyltransferase	Q706L	Helicase
B438L	Formation of vertices in icosahedral capsid	QP509L	Helicase
B169L	Transmembrane protein	QP383R	NifS-like protein
B475L	Putative protein	E184L	Putative protein
B354L	Putative protein	E183L	structural protein p54
B602L	Chaperone involved in folding of capsid.	E423R	Putative
E301R	Proliferating cell nuclear antigen-like protein.	I267L	Putative protein
E146L	Transmembrane protein	I226R	Putative protein
E199L	Transmembrane protein	I243L	transcription factor SII homolog
E165R	dUTPase	I73R	Putative protein
E248R	Structural protein	I329L	Inhibitor of TLR 3 signaling pathway.
E120R	Structural protein p14.5	I215L	Ubiquitin-conjugating enzyme
E296R	AP endonuclease class II	I177L	Transmembrane protein
E111R	Putative protein	I196L	Putative protein
E66L	Putative protein		

Table 2. 6. Pairwise comparison of the alignment of 109 concatenated conserved ORFs (nt) of six study isolates and reference strains. Nucleotide identity (%) is shown in the bottom left and the number of differences in the top right, with red and blue, indicating a higher and lower degree of identity, respectively.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
26544_OG10_2010_Italy_Sus_scrofa_Genotype_I	1		197	4008	310	3879	409	3878	255	3914	12385	12578	12867	283	2217	5213	11633	2350	12592	524	3917	536	3892	5510	12579	12579	12579	12579	5051	4885	5232	4903	5683	4765	5232	3885	5451
57_Ca_79_1979_Italy_Sus_scrofa_Genotype_I	2	99.85		4086	113	3957	266	3956	94	3992	12442	12568	12832	122	2196	5241	11538	2355	12557	363	3995	375	3970	5496	12544	12544	12544	12544	5114	4896	5230	4893	5780	4775	5230	3963	5442
ASFV_Kyiv_2016_131_Ukraine_Sus_scrofa_Genotype_II	3	96.92	96.86		4129	145	4250	144	4118	173	12411	12362	12836	4128	3491	6058	11261	3794	12357	4353	163	4368	165	5139	12342	12342	12342	12342	5146	4999	4886	4829	5408	5275	4886	151	5149
BA71V_1971_Spain_Sus_scrofa_Genotype_I	4	99.76	99.91	96.83		4002	313	4001	171	4035	12516	12641	12907	209	2277	5319	11607	2432	12630	342	4038	356	4015	5574	12617	12617	12617	12617	5194	4973	5309	4967	5855	4851	5309	4008	5518
Belgium_2018_1_Wild_boar_Genotype_II	5	97.01	96.96	99.89	96.92		4121	3	3999	46	12289	12237	12712	4001	3365	5929	11137	3665	12230	4226	50	4239	24	5004	12215	12215	12215	12215	5013	4870	4758	4702	5275	5148	4758	10	5014
Benin_97_1_1997_Sus_scrofa_Genotype_I	6	99.68	99.80	96.74	99.76	96.84		4120	290	4157	12606	12721	13001	342	2339	5406	11686	2494	12708	337	4160	349	4134	5669	12695	12695	12695	12695	5271	5057	5394	5056	5916	4936	5394	4127	5608
CzechRepublic_2017_1_Wild_boar_Genotype_II	7	97.02	96.96	99.89	96.92	100.00	96.84		3998	45	12288	12238	12711	4000	3366	5930	11136	3664	12231	4225	49	4238	23	5003	12216	12216	12216	12216	5012	4871	4757	4701	5274	5147	4757	9	5013
E75_1975_Spain_Sus_scrofa_Genotype_I	8	99.80	99.93	96.84	99.87	96.93	99.78	96.93		4027	12470	12596	12863	162	2200	5246	11567	2392	12585	379	4032	391	4010	5535	12572	12572	12572	12572	5153	4934	5267	4931	5807	4813	5267	4005	5481
Georgia_2007_1_Sus_scrofa_Genotype_II	9	96.99	96.93	99.87	96.90	99.96	96.81	99.97	96.90		12316	12264	12743	4034	3394	5966	11171	3700	12260	4259	45	4274	42	5038	12245	12245	12245	12245	5045	4907	4792	4738	5307	5183	4792	52	5048
Ken05_Tk1_2005_Kenya_Sus_scrofa_Genotype_X	10	90.59	90.56	90.56	90.51	90.65	90.44	90.65	90.54	90.63		4618	2790	12483	12310	14290	12195	12121	4641	12702	12325	12714	12297	12393	4634	4634	4634	4634	12174	12445	12183	12444	12703	12365	12183	12291	12375
Ken06.Bus_2006_Kenya_Sus_scrofa_Genotype_IX	11	90.43	90.45	90.59	90.39	90.68	90.34	90.68	90.43	90.66	96.47		5290	12610	12201	14180	12348	12203	220	12827	12273	12840	12246	12206	210	210	210	210	12373	12319	12119	12466	12801	12299	12119	12241	12334
Kenya_1950_Sus_scrofa_Genotype_X	12	90.26	90.29	90.28	90.24	90.37	90.18	90.37	90.27	90.34	97.88	95.98		12873	12691	14680	12733	12545	5291	13088	12751	13100	12718	12699	5281	5281	5281	5281	12541	12791	12521	12743	13062	12753	12521	12714	12651
L60_1960_Portugal_Sus_scrofa_Genotype_I	13	99.78	99.91	96.83	99.84	96.92	99.74	96.92	99.88	96.90	90.53	90.41	90.26		2254	5127	11577	2390	12599	421	4035	433	4012	5543	12586	12586	12586	12586	5160	4955	5269	4933	5820	4821	5269	4007	5489
Liv13_33_2020_Zambia_Tick_Genotype_I	14	98.29	98.31	97.31	98.25	97.41	98.20	97.41	98.31	97.39	90.64	90.71	90.39	98.26		4588	11242	2071	12195	2458	3398	2471	3379	5072	12182	12182	12182	12182	5060	4441	4746	4850	5564	4753	4746	3373	5010
LIV_5_40_1983_Zambia_Tick_Genotype_I	15	95.99	95.97	95.34	95.91	95.43	95.85	95.43	95.96	95.41	89.13	89.19	88.87	96.05	96.46		13380	4875	14172	5509	5967	5522	5944	7084	14162	14162	14162	14162	6973	6341	6812	6722	7549	6371	6812	5937	6986
Malawi_Lil_20_1_1983_Tick_Genotype_VII	16	91.23	91.30	91.50	91.25	91.59	91.20	91.59	91.28	91.57	90.84	90.72	90.47	91.27	91.51	89.89		11357	12318	11791	11176	11804	11145	11151	12301	12301	12301	12301	11456	11340	11279	11212	11898	11518	11279	11141	11189
Mkuzi_1979_RSA_Tick_Genotype_I	17	98.19	98.19	97.08	98.13	97.18	98.08	97.18	98.16	97.15	90.79	90.71	90.50	98.16	98.40	96.25	91.43		12170	2615	3703	2628	3677	4956	12157	12157	12157	12157	4559	4407	4699	4279	5387	4023	4699	3666	4897
N10_2015_Uganda_Sus_scrofa_Genotype_IX	18	90.42	90.45	90.59	90.40	90.68	90.35	90.68	90.43	90.66	96.46	99.83	95.98	90.42	90.71	89.20	90.74	90.73		12816	12269	12829	12239	12192	40	40	40	40	12361	12313	12110	12460	12834	12292	12110	12234	12342
NHV_1968_Portugal_Sus_scrofa_Genotype_I	19	99.60	99.72	96.66	99.74	96.76	99.74	96.76	99.71	96.73	90.38	90.27	90.12	99.68	98.11	95.77	91.13	97.99	90.28		4262	16	4237	5763	12803	12803	12803	12803	5381	5164	5498	5158	6037	5042	5498	4232	5709
Odintsovo_2014_Russia_Sus_scrofa_Genotype_II	20	96.98	96.93	99.87	96.90	99.96	96.81	99.96	96.90	99.97	90.62	90.65	90.34	96.90	97.38	95.40	91.56	97.15	90.65	96.73		4277	68	5042	12254	12254	12254	12254	5050	4911	4798	4741	5309	5187	4798	56	5052
OURT_88_3_1988_Portugal_Tick_Genotype_I	21	99.59	99.71	96.65	99.73	96.75	99.73	96.75	99.70	96.72	90.37	90.26	90.11	99.67	98.10	95.76	91.12	97.98	90.27	99.99	96.72		4250	5776	12816	12816	12816	12816	5394	5177	5511	5171	6050	5055	5511	4245	5722
Podlaskie_2015_Russia_Wild_boar_Genotype_II	22	97.00	96.95	99.87	96.91	99.98	96.83	99.98	96.92	99.97	90.65	90.67	90.36	96.91	97.40	95.42	91.59	97.17	90.67	96.75	99.95	96.74		5017	12224	12224	12224	12224	5024	4884	4767	4714	5284	5160	4767	30	5027
Pretoriuskop_96_4_1996_RSA_Tick_Genotype_XX	23	95.78	95.79	96.06	95.73	96.16	95.66	96.16	95.76	96.13	90.60	90.72	90.39	95.75	96.11	94.56	91.59	96.20	90.73	95.59	96.13	95.58	96.15		12177	12177	12177	12177	1731	3729	3012	3701	5077	3701	3012	5004	979
R25_2015_Uganda_Sus_scrofa_Genotype_IX	24	90.43	90.46	90.60	90.41	90.69	90.36	90.69	90.44	90.67	96.46	99.84	95.99	90.43	90.72	89.20	90.75	90.74	99.97	90.29	90.66	90.28	90.69	90.74		0	0	0	12346	12298	12095	12445	12819	12277	12095	12219	12325
R35_2015_Uganda_Sus_scrofa_Genotype_IX	25	90.43	90.46	90.60	90.41	90.69	90.36	90.69	90.44	90.67	96.46	99.84	95.99	90.43	90.72	89.20	90.75	90.74	99.97	90.29	90.66	90.28	90.69	90.74	100.00		0	0	12346	12298	12095	12445	12819	12277	12095	12219	12325
R7_2015_Uganda_Sus_scrofa_Genotype_IX	26	90.43	90.46	90.60	90.41	90.69	90.36	90.69	90.44	90.67	96.46	99.84	95.99	90.43	90.72	89.20	90.75	90.74	99.97	90.29	90.66	90.28	90.69	90.74	100.00	100.00		0	12346	12298	12095	12445	12819	12277	12095	12219	12325
R8_2015_Uganda_Sus_scrofa_Genotype_IX	27	90.43	90.46	90.60	90.41	90.69	90.36	90.69	90.44	90.67	96.46	99.84	95.99	90.43	90.72	89.20	90.75	90.74	99.97	90.29	90.66	90.28	90.69	90.74	100.00	100.00	100.00		12346	12298	12095	12445	12819	12277	12095	12219	12325
RSA_2_2004_Wild_boar_Genotype_XX	28	96.12	96.08	96.05	96.02	96.15	95.96	96.15	96.05	96.12	90.75	90.59	90.51	96.04	96.11	94.63	91.36	96.50	90.60	95.88	96.12	95.87	96.14	98.67	90.61	90.61	90.61	90.61		3680	2980	3271	4769	3287	2980	5013	1785
RSA_2_2008_Tick_Genotype_XXII	29	96.25	96.24	96.16	96.18	96.26	96.12	96.25	96.21	96.23	90.55	90.63	90.32	96.19	96.58	95.12	91.44	96.61	90.63	96.04	96.22	96.03	96.24	97.14	90.64	90.64	90.64	90.64	97.17		3289	2043	5132	1525	3289	4873	3629
RSA_W1_1999_Warthog_Genotype_IV	30	95.98	95.98	96.24	95.92	96.34	95.86	96.34	95.96	96.31	90.74	90.77	90.51	95.95	96.35	94.75	91.49	96.39	90.78	95.79	96.31	95.78	96.33	97.68	90.79	90.79	90.79	90.79	97.70	97.47		3201	4926	3315	0	4758	2887
SPEC_57_1985_RSA_Tick_Genotype_III	31	96.24	96.25	96.29	96.19	96.39	96.13	96.39	96.22	96.36	90.56	90.52	90.36	96.22	96.27	94.83	91.54	96.71	90.53	96.05	96.36	96.04	96.38	97.16	90.54	90.54	90.54	90.54	97.49	98.43	97.54		4767	1411	3201	4702	3558
Tengani62_1962_Malawi_Sus_scrofa_Genotype_V	32	95.64	95.57	95.85	95.52	95.95	95.48	95.95	95.55	95.93	90.37	90.28	90.13	95.54	95.73	94.20	91.04	95.87	90.25	95.39	95.93	95.38	95.95	96.11	90.27	90.27	90.27	90.27	96.34	96.06	96.22	96.34		4875	4926	5277	4940
Warmbaths_1987_RSA_Tick_Genotype_III	33	96.34	96.33	95.95	96.28	96.04	96.22	96.04	96.31	96.02	90.61	90.64	90.35	96.30	96.35	95.09	91.31	96.91	90.64	96.14	96.01	96.13	96.03	97.16	90.65	90.65	90.65	90.65	97.47	98.82	97.45	98.91	96.26		3315	5149	3625
Warthog_1980_Namibia_Genotype_IV	34	95.98	95.98	96.24	95.92	96.34	95.86	96.34	95.96	96.31	90.74	90.77	90.51	95.95	96.35	94.75	91.49	96.39	90.78	95.79	96.31	95.78	96.33	97.68	90.79	90.79	90.79	90.79	97.70	97.47	100.00	97.54	96.22	97.45		4758	2887
Wuhan_2019-1_China_Sus_scrofa_Genotype_II	35	97.01	96.95	99.88	96.92	99.99	96.83	99.99	96.92	99.96	90.65	90.67	90.37	96.92	97.40	95.43	91.59	97.18	90.68	96.75	99.96	96.74	99.98	96.16	90.69	90.69	90.69	90.69	96.15	96.25	96.34	96.39	95.95	96.04	96.34		5012
Zaire_1977_Sus_scrofa_Genotype_XX	36	95.82	95.83	96.05	95.77	96.15	95.70	96.15	95.80	96.12	90.60	90.62	90.42	95.79	96.15	94.63	91.56	96.24	90.61	95.63	96.12	95.62	96.14	99.25	90.63	90.63	90.63	90.63	98.62	97.21	97.78	97.27	96.21	97.21	97.78	96.15	

Table 2. 7. Fragmented ORFs in the central coding region (CCR) of the LIV_5_40 isolate.

ORFs	**No of fragments**
F317L	2 fragments
K421R	2 fragments
C717R	2 fragments
C962R	2 fragments
B962L	3 fragments
B438L	2 fragments
G1211R	3 fragments
CP2475L	5 fragments
CP530R	2 fragments
NP868R	2 fragments
D1133L	3 fragments
D345L	2 fragments
Q706L	2 fragments
QP383R	2 fragments
E301R	2 fragments

EP402R (CD2V): The CD2v protein encoded by the Zaire isolate is similar to that encoded by the Pretoriuskop_96_2 strain (AY261363), with 99.9% amino acid identity. In contrast, the CD2v protein encoded by the SPEC_57 isolate is most similar to that encoded by all genotype II reference sequences, with 88.7% amino acid identity. The CD2v protein encoded by LIV_5_40 is similar to that encoded by the study isolate Zaire, with a 53.6% amino acid identity. The one encoded by RSA_2_2004 is most similar to that encoded by the study isolate Zaire, with 79.6% similarity, while the one encoded by RSA_2_2008 isolate has highest similarity to that encoded by Liv13_33 (Genbank number: MN913970) with a 95.4 % amino acid identity. The CD2v protein encoded by the RSA_W1_1999 isolate is more similar to that encoded by Warthog strain (AY261366), with 99.9 % amino acid identity (Data not shown).

The C-terminus region of the EP402R gene is known to contain proline-rich tandem repeats. A comparison of the CD2v protein from 73 ASFV isolates (Figure 2.16) showed a variation in the number of tandem repeats between isolates. The CD2v protein encoded by the cell-culture adapted strain (Zaire) contains seven repeats of the proline-rich tandem repeat sequence (PPPKPC). The European cell-culture adapted isolate BA71V, also contains seven repeats of the tandem repeat sequence. The naturally attenuated isolates NHV and OURT_888_3 do not encode tandem repeat sequence, and all genotype II isolates have three tandem repeats. A genotype III tick isolate Warmbaths, contains five repeats of the tandem repeat sequence, while a genotype III study tick isolates SPEC_57 contains 8 repeats. The

study isolate LIV_5_40 contains five tandem repeats, and RSA_2_2008 contains 4 repeats. The RSA_W1_1999 isolate has six tandem repeats, and the RSA_2_2004 contains four repeats of the tandem repeat sequence. The RSA_2_2004 contains a substitution mutation of proline with serine at position 340, and RSA_2_2008 contains a similar substitution mutation at position 357. European genotype I isolates encode between seven and eight repeats. The SPEC_57 contains a substitution mutation of lysine with glutamic acid at position 354.

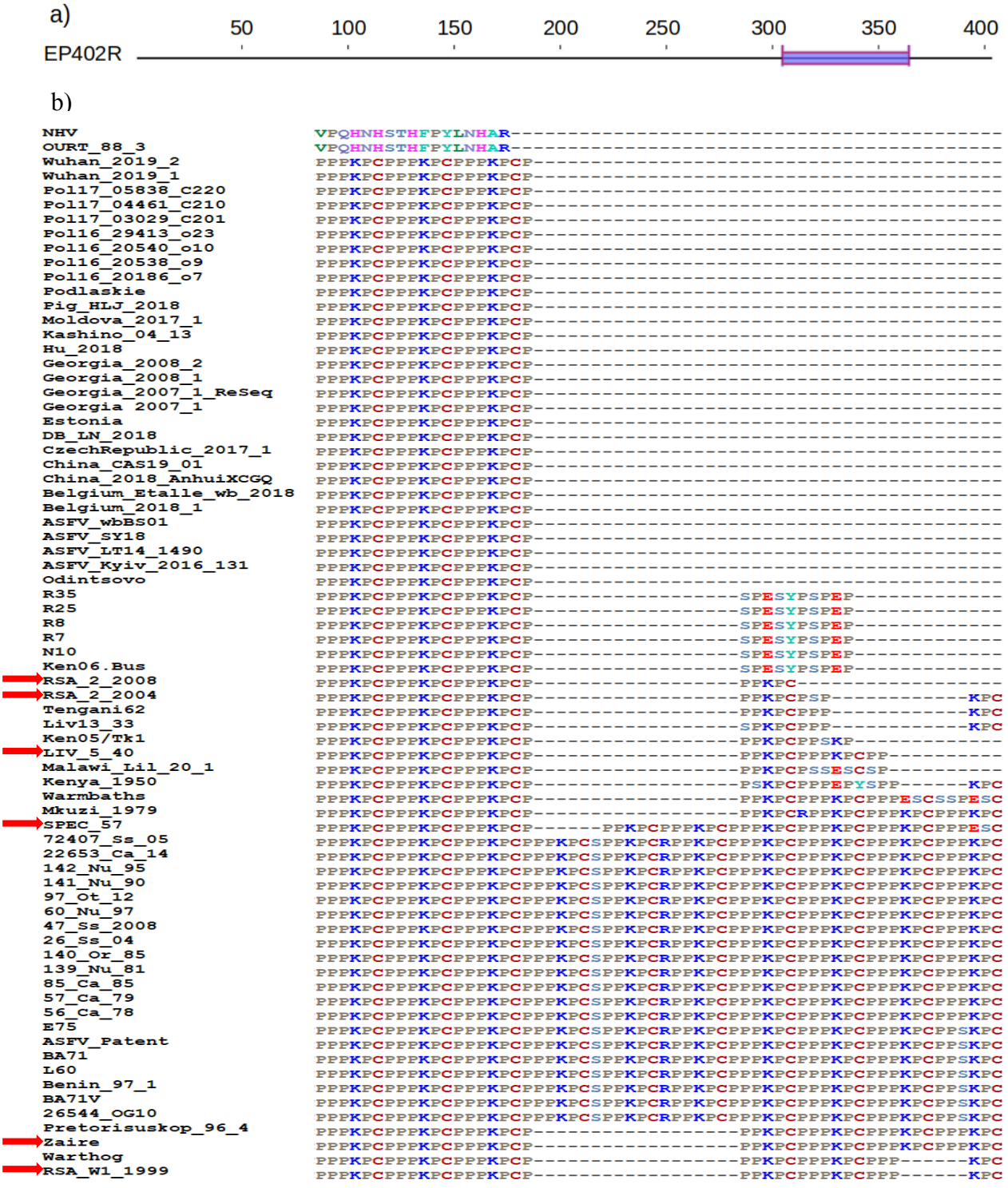


Figure 2. 16. Amino acid alignment of the proline-rich repeat region of the EP402R gene. a) Scale representing the length of the EP402R (CD2V) protein with the aligned region highlighted, b) Amino acid alignment of the proline-rich repeat region of the EP402R (CD2v) gene of 73 ASFV isolates analysed in this study. Red arrows indicate the location of the study isolates, residues in grey are conserved across all sequences, and coloured residues are different in at least one sequence.

B602L: The CVR of the B602L gene contains tetrameric repeat sequences that vary in number among genomes of different ASFV isolates. Differences in the number of tandem repetitive amino acid sequences TACD, TNVD, TCAS, and TGAS in the B602L (CVR region) gene were observed, and the number of differences is outlined in Table 2.8.

Table 2. 8. The tetrameric repeats of the six study isolates and the reference sequences of the same genotype. The repeats reflect tetrameric content only and not the order in which the repeats occur in each isolate.

Isolates	Genotype	# of tandem repeats (**CADT**)	# of tandem repeats (**NVDT**)	# of tandem repeats (**CAST**)	# of tandem repeats (**GAST**)
LIV_5_40	I	3	2	3	**A**
BA71	I	7	6	11	1
BA71V	I	6	5	13	1
ASFV-Patent	I	7	6	11	1
Benin_97_1	I	5	5	21	4
E75	I	6	5	11	1
L60	I	6	5	10	1
NHV	I	8	7	24	3
OURT 88_3	I	8	7	25	3
47_Ss_2008	I	2	2	6	**A**
26544_OG10	I	2	2	6	**A**
56/Ca/78	I	6	5	10	1
57/Ca/79	I	6	5	10	1
139/Nu/81	I	6	5	10	1
140/Or/85	I	6	5	10	1
85/Ca/85	I	2	2	6	**A**
141/Nu/90	I	2	2	6	**A**
142/Nu/95	I	2	2	6	**A**
60/Nu/97	I	2	2	6	**A**
26/Ss/04	I	2	2	6	**A**
72407/Ss/05	I	2	2	6	**A**
97/Ot/12	I	2	2	6	**A**
22653/Ca/14	I	2	2	6	**A**
Liv13_33	I	4	3	3	**A**
Mkuzi	I	2	2	4	**A**
Warmbaths	III	2	**A**	3	**A**
SPEC_57	III	1	**A**	3	**A**
Warthog	IV	3	1	4	**A**
RSA_W1_1999	IV	5	2	5	**A**
Pretoriuskop_96_4	XX	6	3	6	**A**
RSA_2_2004	XX	2	2	4	A
Zaire	XX	7	4	3	**A**
RSA_2_2008	XX	2	**A**	3	**A**

A-Absent (highlighted in green), study isolates in bold.

2.4. DISCUSSION

The genome sequences of six novel ASFV isolates have been determined, apart from the ITRs, which are incomplete. ASFV genomes have been shown to vary in length from 170-194 kbp (Chapman *et al.*, 2008, de Villiers *et al.*, 2010, Forth *et al.*, 2019), as well as the number of ORFs, ranging from 150-167 (Rodriguez *et al.*, 2015). The shortest genome from this study was from the tick isolate LIV_5_40 with a genome length of 183, 292 bp, and the longest was the RSA_2_2008 with a genome length of 190, 065 bp. The genome size differences observed are a result of the gain or loss of members of the five MGFs.

Nucleotide pairwise comparison of full genomes of the study isolates with reference sequences of the same genotype showed the LIV_5_40 to be more similar to Liv13_33 (94.1%) and Mkuzi_1979 (93.2%) than other genotype I reference sequences. SPEC_57 showed 97.6% similarity to Warmbaths, and RSA_W1_1999 had a 99.2% similarity to the Warthog isolate. Study isolates RSA_2_2004 and Zaire showed 97.2% and 96.8% similarity to Pretoriouskop_96_4, respectively. A genotype XXII study isolate, RSA_2_2008 showed 93.4% similarity to Pretoriuskop_96_4, which represents genotype XX. Of interest was high nucleotide similarity amongst study isolates, with 97.4% similarity observed between RSA_2_2008 and SPEC_57 and 95.0% similarity observed between RSA_2_2004 and Zaire.

The first analysis of the genomes was to use dot plots to determine the level of collinearity between the novel ASFV genomes and selected reference sequences. Dot plots showed interruption of the collinear structure at LVR, CCR, and RVR in the LIV_5_40 and Zaire isolates. At the same time, interruption of the collinear structure was observed mostly at the terminal ends in other isolates (Figure 2.8- 2.11), resulting in the truncation of some of the members of the MGFs relative to publicly available reference sequences (See Chapter 4).

The comparison of the cell-culture adapted isolate Zaire with the attenuated BA71V, NHV, and OURT_88_3 isolates, revealed that the Zaire isolate is significantly longer (184, 820 bp) than BA71V (170, 101 bp), NHV (171, 719 bp), and OURT_88_3 (172, 051 bp). The whole-genome nucleotide comparison of the Zaire, BA71V, NHV, and OURT_88_3 showed a sequence similarity varying between 86.0% and 99.9%. At the LVR region, the BA71V isolate has a deletion of 8.2 kbp that results in the deletion of five members of MGF 360 and one member of MGF 505. The NHV and OURT_88_3 isolates, which contain a 10 kbp deletion, resulting in loss of six members of the MGF 360 and three members of the MGF 505. The Zaire isolate contains an insertion at that region and contains all five members of

the MGF 360, and the one member of the MGF 505 deleted in BA71V, NVH, and OURT_88_3 strains. The BA71V further has a 3.2 kbp deletion at the RVR, resulting in the loss of five ORFs (I7L, I8L, I9L, I10L, and L11L), which are conserved in Zaire, NHV, and OURT_88_3 isolates. These ORFs are uncharacterised, except for the I10L ORF, which is similar to the KP177R ORF found at the LVR that encodes the p22 structural protein. It has been shown that at the amino acid level, the I10L proteins have a 40% amino acid sequence identity to the p22 structural protein (Chapman *et al*., 2008). However, the I10L protein found in the Benin_97_1 and OURT_88_3 strains is 100% conserved. The Kenya_1950, Mkuzi-1979, Pretoriuskop_96_4, and Tengani62 encode a two fragment protein (I10L_1 and I10L_2). This is also observed in five of the study isolates (RSA_W1_1999, RSA_2_2004, RSA_2_2008, SPEC_57 and Zaire). The amino acid alignment of the I10L_1 protein reveals that it shares between 53% to 100% identity and the I10L_2 protein shares between 76% to 100% identity among ASFV isolates (Chapman *et al*., 2008). The comparison of the I10L_1 protein encoded by the study isolates showed that they share between 72% to 99% identity and I10L_2 shares between 68% to 100% identity. The I10L has been shown to be expressed on the surface of ASFV infected cells, whether this gene has an effect on virulence needs further investigation (Vydelingum *et al*., 1993).

In the CCR, there are 109 conserved ORFs in all 73 isolates used in this study, with almost half of them having an unknown function (Chapman *et al*., 2011) and most of the remainder playing an essential role in the virus life cycle (Chapman *et al*., 2011; Yanez *et al*., 1995). Previous studies have demonstrated that the nucleocytoplasmic large DNA viruses (NCLDVs) replicate independently of the host replication or transcription system to complete their replication process. These viruses can achieve this independence by encoding several conserved proteins that include DNA polymerases, helicase, primase, RNA polymerase subunits, and transcriptional factors that function in transcription initiation and elongation (Yutin & Koonin, 2012).

The LIV_5_40 isolates encodes a few fragmented ORFs, however, similar fragmented ORFs are found in the in the Ukraine isolate ASFV_Kyiv_2016_131. These fragmented ORFs in the LIV_5_40 could be as a result of incorrect sequencing of homopolymer runs (data no shown).

Significant differences were observed in genes that are involved in haemadsorption (EP153R and EP402R). These ORFs are found adjacent to each other in the ASFV genomes (Figure

2.2-2.7), and are present in all ASFV isolates and are the most divergent in the ASFV genomes. The EP153R ORF encodes a C-type lection, and the EP402R ORF encodes the CD2v protein. The EP402R gene contains a proline-rich tandem repeat sequence (PPPKPC), which is required to bind the CD2v protein to the host actin-binding adaptor protein SH3P7 (Rowlands *et al*., 2009). The CD2v protein plays are role in binding red blood cells to ASFV infected macrophages (Borca *et al*., 1994; Rodriguez *et al*., 1993). More tandem repeats result in the early onset of viremia and enhance virus dissemination to the target tissue. The expression of the CD2v protein in ticks results in high virus titres through the enhancement of virus uptake across the gut wall (Reis *et al*., 2017b). It has been shown that the deletion of the EP153R gene reduces the number of red blood cells binding to infected cells, thus, suggesting that it plays a role in haemadsorption (Rowlands *et al*., 2009). These ORFs are disrupted in natural non-haemadsorbing isolates of low virulence (Leitao *et al*., 2001). However, the EP153R gene in the study isolate LIV_5_40, which is haemadsorbing was truncated (Figure 2.15), close to the C-terminus. Hence the weak heamadsorbing reaction of the LIV_5_40 isolate when it was passaged in swine macrophages. The NVH and OURT_88_3 strains have a deletion close to the N-terminus, while LIV_5_40 contains an insertion. The percentage identity of the EP153R and EP402R genes across all ASFV isolates is from 43% to 100% at the nucleotide level, and 55% to 100% at the amino acid level (Chapman *et al.*, 2008, Chapman *et al*., 2011).

The deletion of the EP402R gene from the genome of virulent isolates does not affect the mortality caused by an infection in pigs; however, it causes a delay in the onset of viraemia and virus dissemination to the target tissue (Borca *et al*., 1998). The European cell-culture adapted isolate BA71V isolate encodes the same number (seven) of proline-rich tandem repeat sequence (PPPKPC), as the southern African cell-culture adapted isolate Zaire. The complete deletion of the EP402R gene in the Zaire isolate might result in attenuation and may induce effective protective immunity, similar to the BA71ΔCD2 strain (Monteagudo *et al.*, 2017).

The EP153R and the EP402R genes have been used for serotyping ASFV isolates because the genetic locus that encodes the C-type lectin and CD2v proteins bring about serotype-specific haemadsorption inhibition (HAI) (Malogolovkin *et al*., 2015b). Eight ASFV serotype groups have been identified using this approach, and viruses within the same group have been shown to cross-protect against each other (Malogolovkin *et al*., 2015a; Sereda and Balyshev, 2011; Sereda *et al*., 1992). It has been demonstrated that the heterologous expression of the

C-type lectin and CD2v proteins in swine induced T-cell responses (Jancovich *et al*., 2018; Lopera-Madrid *et al*., 2017). A recent study further identified six T-cell epitopes on the C-type lectin and CD2v protein encoded by the ASFV Congo K-49 (Burmakina *et al*., 2019). Thus, serotyping of the study isolates is of utmost importance in identifying viruses that they can cross-protect against. It will provide more information that is necessary for vaccine design and development.

The B602L gene is a chaperone that plays a role in the assembly of the p72 protein into virions (Cobbold *et al*., 2001). In a study looking at the serological determinants that are detected post-infection of pigs with ASFV, the B602L gene was one of the 14 identified proteins that antibodies were generated against (Kollnberger *et al.*, 2002). Variations in the amino acid repeats in the B602L were observed, with SPEC_57 encoding the least with four tetramers, while the Zaire isolate encodes the most tetramers with 14. The CVR region of the B602L gene is a hyper-variable genetic marker that has been used extensively in phylogenetic analysis studies to enhance the discrimination of ASFV isolates. The difference in the number of tetramers could be essential in generating antigenic variations and aid the virus to evade the host antibody response.

The genome sequences of the LIV_5_40, RSA_W1_1999, RSA_2_2004, RSA_2_2008, SPEC_57, and Zaire have significantly added to the limited repository of genomic information available for viruses isolated from domestic pigs, ticks, warthogs, and wild boars. The overall comparison of the Zaire isolate to other attenuated strains (BA71V, NHV, and OURT_88_3) showed that the Zaire isolate could be a potential vaccine candidate. However, the potential of the Zaire isolate as a vaccine candidate could be further investigated with *in vivo* studies as host range and virulence are difficult to determine by analysis of gene content or sequence similarity alone.

CHAPTER 3: PHYLOGENETIC ANALYSIS OF SIX NOVEL AFRICAN SWINE FEVER VIRUSES ISOLATED FROM A DOMESTIC PIG, WARTHOG, WILD BOAR, AND TICKS

3.1. Introduction

The advent of phylogenetic analysis has enabled researchers to quantify biological diversity, classify organisms based on genetic difference and to understand the evolutionary relationships among and within species. In virology, phylogenetics is often used to trace the origins of disease outbreaks and to confirm the epidemiological links between disease events. Early studies on the origin of ASFV outbreaks were determined based on the restriction fragment length polymorphism (RFLP) (Blasco *et al*., 1989a; Dixon, 1988; Sumption *et al*., 1990). Enzymatic digestion of the viral genomes generates specific patterns that can be used to broadly classify viruses into empirically determined groups. However, the resolution of RFLP methods is low, and limiting the practical application at a geographically localised level. During the 1990's the use of sequence data replaced RFLP-based methods, which enabled reconstruction of phylogenetic trees by looking at the nucleotide or protein differences. However, interpretations of phylogenetic analyses are dependent on the selection of gene regions that represent the evolutionary processes operating at a species level. Selecting appropriate phylogenetic markers is therefore the bedrock of reliable and accurate phylogenetic analyses.

The B646L gene encoding the p72 protein has historically been used in phylogenetic analysis of ASFV. The p72 protein is the major component of the viral capsid, making up 32% of the total protein mass of the virion (Cobbold and Wileman, 1998). During natural infections, the p72 protein is the dominant immunogen, and it is often used in diagnostic tests (Tabares *et al*., 1980a). It was demonstrated that although some viral proteins undergo evolutionary changes when the virus is passaged multiple times in macrophages, p72 protein remains the unchanged (Sanz *et al*., 1985). The antigenic stability of p72 was further demonstrated by comparing the full length B646L gene of ASFV isolated from various part of the world, including the Dominican Republic, Spain and Uganda. It was found that at the nucleotide level, the sequence identity between the isolates ranged from 95.5% to 100%, translating to 97.8% to 100% similarity at amino acid level (Yu *et al*., 1996). However, the B646L gene of viruses from a broader geographical range showed that the C-terminus region of the gene is rich in synonymous and non-synonymous mutations and sufficiently diverse to enable the

differentiation of genetically disparate groups of viruses (Bastos *et al*., 2003). At nucleotide level, 59 of the 415 sites corresponded to variable sites, 12 of which resulted in non-synonymous substitutions. The C-terminus region of the gene was subsequently used for the phylogenetic analysis to define genotypes (Bastos *et al*., 2003). To date, 24 p72 genotypes have been described (Figure 3.1A). Sequencing of the partial B646L gene of isolates from Europe, South America, and West Africa confirmed that these isolates all clustered in genotype I. However, Eastern and Southern African isolates are more diverse and grouped into 23 genotypes. The higher sequence variation seen amongst African isolates may be due to the presence of the sylvatic cycle. Although the p72 gene is sufficiently genetically diverse to discriminate between genotypes, it lacks the capacity for higher resolution between isolates belonging to the same genotype.

Various other gene regions have been considered as alternatives to B646L, including genes encoding the minor structural proteins such as p54 and p30. Sequence diversity within these genes are similar to that of B646L with an average of 98% amino acid similarity among ASFV isolates. However, the B602L and EP402R genes, which have given better discrimination of closely related ASFV isolates, were less conserved. The B602L gene has an average of 88% amino acid identity among ASFV isolates (Bacciu *et al*., 2016; Gallardo *et al*., 2009; Sanna *et al*., 2017). Tetrameric amino acid repeats in the central variable region (CVR) of the protein encoded by B602L have proven useful to discriminate between closely related viruses or disaggregating viruses that are grouped in the same genotype based on B646L. The discriminatory power can be further enhanced by the combined analysis of B646L and CVR (Lubisi *et al*., 2007). Other genes used in the phylogenetic analysis of ASFV include K196R, A859L, and the intergenic region (IGR) between I73R and I329L genes. Similar to the CVR, the IGR tetrameric repeat sequences have provided valuable information in determining the origin and the spread of closely related ASFV isolates (de Villiers *et al*., 2010; Gallardo *et al*., 2014; Gallardo *et al*., 2009; Ge *et al*., 2018).

ASFV isolates have been serogrouped using the EP402R (CD2v) and the EP153R (C-type lectin) genes, which play a role in haemadsorption (Malmquist and Hay, 1960). The antigenic diversity of ASFV isolates was determined by assessing the ability of serum from convalescent animals to inhibit the haemadsorption reaction during the replication of the ASFV in swine macrophages. Eight serogroups (SG) have been identified using this haemadsorption inhibition assay (Malogolovkin *et al*., 2015a). It was further demonstrated that vaccination of animals with ASFV isolates within a serogroup are able to cross-protect

against challenge with viruses of the same serogroup (Sereda *et al*., 1992). It has been found that serogroup diversity is mainly found in Southern Africa, where ASFV is transmitted through multiple hosts. In contrast, low serogroup diversity is observed in non-endemic countries, where outbreaks result from a single genotype (Malogolovkin *et al*., 2015a). Phylogenetic assignment of isolates to genotypes based on the p72 gene does not consistently correspond to defined serogroups. Viruses belonging to serogroup 1, 2, and 4, formed a genotype I clade, and did not group with other genotypes (Malogolovkin *et al*., 2015b). The China_2018_1 strain was phylogenetically serotyped using the CD2v gene and the isolate clustered with viruses representing serogroup 8 (Figure 3.1B) (Ge *et al*., 2018).

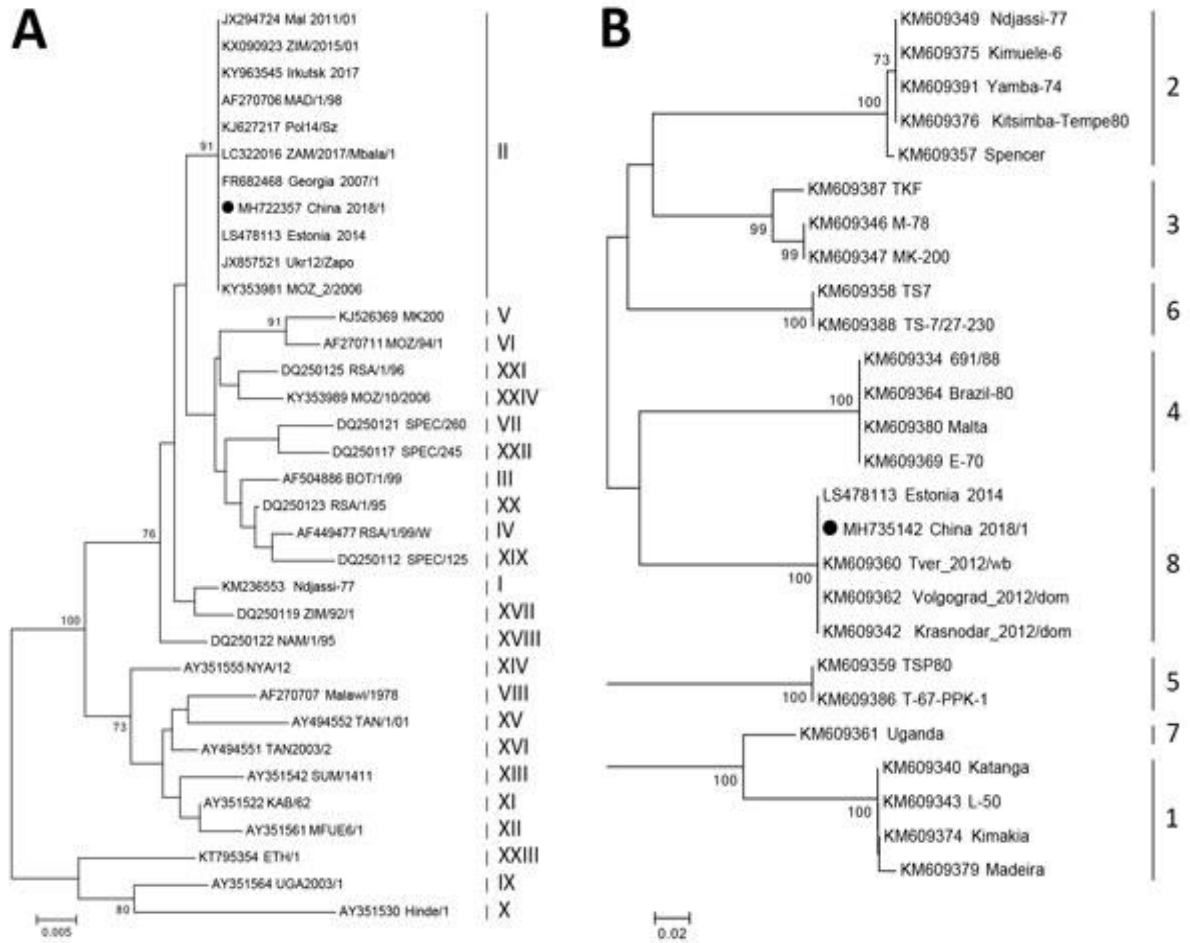


Figure 3. 1. Phylogenetic analysis of a causative virus strain (China 2018/1) of an African swine fever outbreak, China, 2018. A) p72 genotype; B) CD2v serogroup. The neighbour-joining method and Kimura 2-parameter model were used for construction of phylogenetic trees in MEGA 5.0 software. Numbers along branches indicate bootstrap values >70% (1,000 replicates). Black circles indicate causative virus from this study. Roman numerals to the right in panel A indicate p72 genotypes. Numbers to the right in panel B indicate CD2v serogroups. GenBank accession numbers are provided for all sequences. Scale bars indicate nucleotide substitutions per site. The image was taken from (Ge *et al*., 2018).

Phylogenetic analysis of the ASFV isolates collected from 1978 to 2014, in Sardinia, using the EP402R gene, revealed that these isolates could be divided into two subgroups. One group comprised of historical viruses isolated from 1978 to 1990, and the second group containing isolates recovered from 1990 to 2004 (Sanna *et al*., 2017). ASF viruses isolated in Sardinia from 1978 to 2009 showed no significant variations with the B646L (p72) and E183L (p54) genes. However, minor differences in the number of the tetramer amino acid repeats were observed in the B602L gene also separated the isolates into two subgroups (Giammarioli *et al*., 2011). The comparison of the B602L and KP86R genes, and the IGR, 43 closely related ASFV isolates from the Caribbean, Europe, West, and Central Africa were grouped into 17 different subgroups (Nix *et al*., 2006). They further grouped 81 different ASFV isolates into 31 subgroups based on the variation in the number of tandem repeats in the B602L gene.

Although phylogenetic analysis based on a single locus such as B646L is simple, but represents less than 1% of the complete genome and provides only a rough classification of the ASFV (Xiong *et al*., 2019). Other studies have used a genome-based approach, whereby 123 concatenated core conserved genes were used to study the phylogenetic relationship among ASFV isolates (de Villiers *et al*., 2010). It has also been shown that the concatenation of multiple genes can resolve ambiguities observed when using single genes (Gontcharov *et al*., 2004). Complete genome sequencing is considered the gold standard. However, it is not practical to sequence whole genomes for surveillance and would only be useful when several complete genome sequences are available to represent all genotypes described.

The aim of this chapter is to analyse a multiple sequence alignment of 73 complete ASFV genome sequences to identify and investigate novel loci in the CCR that could be phylogenetically informative. Identifying additional genetic markers that can discriminate between closely related ASFV isolates is needed. This study aimed to reassess the phylogenetic constructions of ASFV isolates by performing multiple sequence alignments of 73 complete ASFV genome sequences and identifying genomic regions with substantial genetic variability that can be used as novel phylogenetic markers to provide finer discrimination between closely related isolates. This study further aimed to phylogenetically serotype the six novel ASFV isolates, and other publicly available reference genome sequences in GenBank using the EP153R and EP402R genes.

3.2. MATERIALS AND METHODS

3.2.1. Methods and sequence sources

A total of 73 ASFV were used in this study with 67 of the 73 ASFV sequences (see Chapter 2, Table 2.1) downloaded from GenBank (http://www.ncbi.nlm.nih.gov). The additional six sequences were generated in this study from viruses isolated from domestic pigs, warthog, wild boar, and soft ticks. The six novel sequences (LIV_5_40, RSA_W1_1999, RSA_2_2004, RSA_2_2008, SPEC_57 and Zaire) as described in chapter 2. Reference sequences of the EP153R and EP402R genes representing the eight serogroups (Table 3.1) were downloaded from GenBank.

A multiple sequence alignment (MSA) of 73 ASFV genomes was generated using MAFFT in Geneious Prime, and the core conserved region of the genome was scrutinised for distinct regions that could be phylogenetically informative. Five novel loci (EP364R, B169L, B117L, B407L, and I177L), were identified as possible, alternative phylogenetic markers and compared to the standard classification based on the p72 gene.

Representative isolates were selected for groups of sequences that were genetically similar across the 7 loci used in the phylogenetic analysis. Seven isolates were selected to represent genotype II (ASFV_Kyiv_2016_131, Belgium_2018_1, CzechRepublic_2017_1, Georgia_2007_1, Odintsovo, Podlaskie, and Wuhan_2019-1). A reduced data set containing a total of 36 (see Table 3.1) isolates was used to generate phylogenetic trees based on p72, the five novel loci, five novel concatenated genes, and the 109 core concatenated conserved genes in order to increase the readability of the figures presented here.

Table 3. 1. ASFV isolates selected for inclusion in a study comparing ASFV serogroups.

Isolate	Country	Year	Serogroup	Accession no.
Katanga	Zaire	1957	SG1	KM609340
Kimakia	Kenya	N/A	SG1	KM609374
L-50	Portugal	1950	SG1	KM609343
L-57	Portugal	1957	SG1	KM609344
K-49	Zaire	1949	SG2	KM609339
K-73	Zaire	N/A	SG2	KM609370
Kk-262	Democratic Republic of Congo-	N/A	SG2	KM609371
Ndjassi-77	Democratic Republic of Congo	1977	SG2	KM609349

Yamba	Democratic Republic of Congo	1974	SG2	KM609391
M-78	Mozambique	1978	SG3	KM609346
MK-200	Mozambique	1978	SG3	KM609347
TKF	Tanzania	N/A	SG3	KM609387
Brazil-80	Brazil	1979	SG4	KM609364
Cuba-71	Cuba	1971	SG4	KM609366
F-32	France	1964	SG4	KM609337
O-77	USSR	1977	SG4	KM609350
P-60	Portugal	N/A	SG4	KM609351
PPA	Spain	N/A	SG4	KM609352
T-67_PPK	Tanzania	N.A	SG5	KM609386
TSP80	Tanzania	1980	SG5	KM609359
TS7	Tanzania	1984	SG6	KM609358
TS-7/27-230	Tanzania	1986	SG6	KM609388
Uganda	Uganda	N/A	SG7	KM609361
Tver_2012/wb	Russia	2012	SG8	KM609360
Volgograd_2012/dom	Russia	2012	SG8	KM609362
Zavidovo-2012	Russia	2012	SG8	KM609392

N/A –Not available

3.2.2. Phylogenetic tree construction

Multiple sequence alignments based on the EP402R and EP153R genes, C-terminus p72 gene, the 109 concatenated core conserved set of genes, the five novel concatenated genes, and each individual gene (B646L, EP364R, B169L, B117L, B407L, and I177), were performed using MAFFT in Geneious Prime. Phylogenetic analyses of the individual genes were conducted using a neighbour joining method (Saitou and Nei, 1987) implemented in Mega X (Kumar *et al.*, 2018) employing the p-distance nucleotide substitution model. Bootstrap confidence values were calculated on 10 000 replicates. For the concatenated ORFs maximum-likelihood trees were generated in MEGA X (Kumar *et al.*, 2018) using the Kimura 2-parameter model with 1000 bootstrap replicates. The phylogenetic trees were visualised in the Interactive tree of life (iTOL) v4 (Letunic and Bork, 2019), an online tool for display and annotation of phylogenetic trees.

3.3. RESULTS

The partial B646L gene sequences extracted from the 36 complete genome sequences, representing ASF viruses isolated between 1960 and 2020, were used to assign each isolate to

one of the 24 genotypes previously described (Achenbach *et al*., 2017; Bastos *et al*., 2003; Boshoff *et al*., 2007; Lubisi *et al*., 2005; Quembo *et al*., 2018). This gene is highly conserved and showed nucleotide similarity between 94% and 100% among the 36 isolates analysed in this study. The phylogenetic tree generated with the C-terminus portion of the p72 gene formed three clades (Figure 3.2) with ten of the 24 genotypes represented in the dataset. The phylogenetic groupings were consistent with published descriptions. Isolates from Europe and Asia were confined to two genotypes with viruses associated with outbreaks prior to 1990 falling within genotype I and viruses associated with outbreaks after 2007 all grouping within genotype II. Genotype I contained three viruses from Africa, namely Mkuzi_1997, LIV_5_40 and Liv_13_33. These viruses were isolated from ticks collected in South Africa and Zambia, respectively. The remainder of the viruses from Africa were assigned to eight different genotypes. This is representative of the higher genetic diversity of ASFV found in Africa. It is important to note that although viruses within genotype II originated from Africa after it was introduced into the Caucuses in 2007, a complete genome sequence of an African genotype II virus is not currently available.

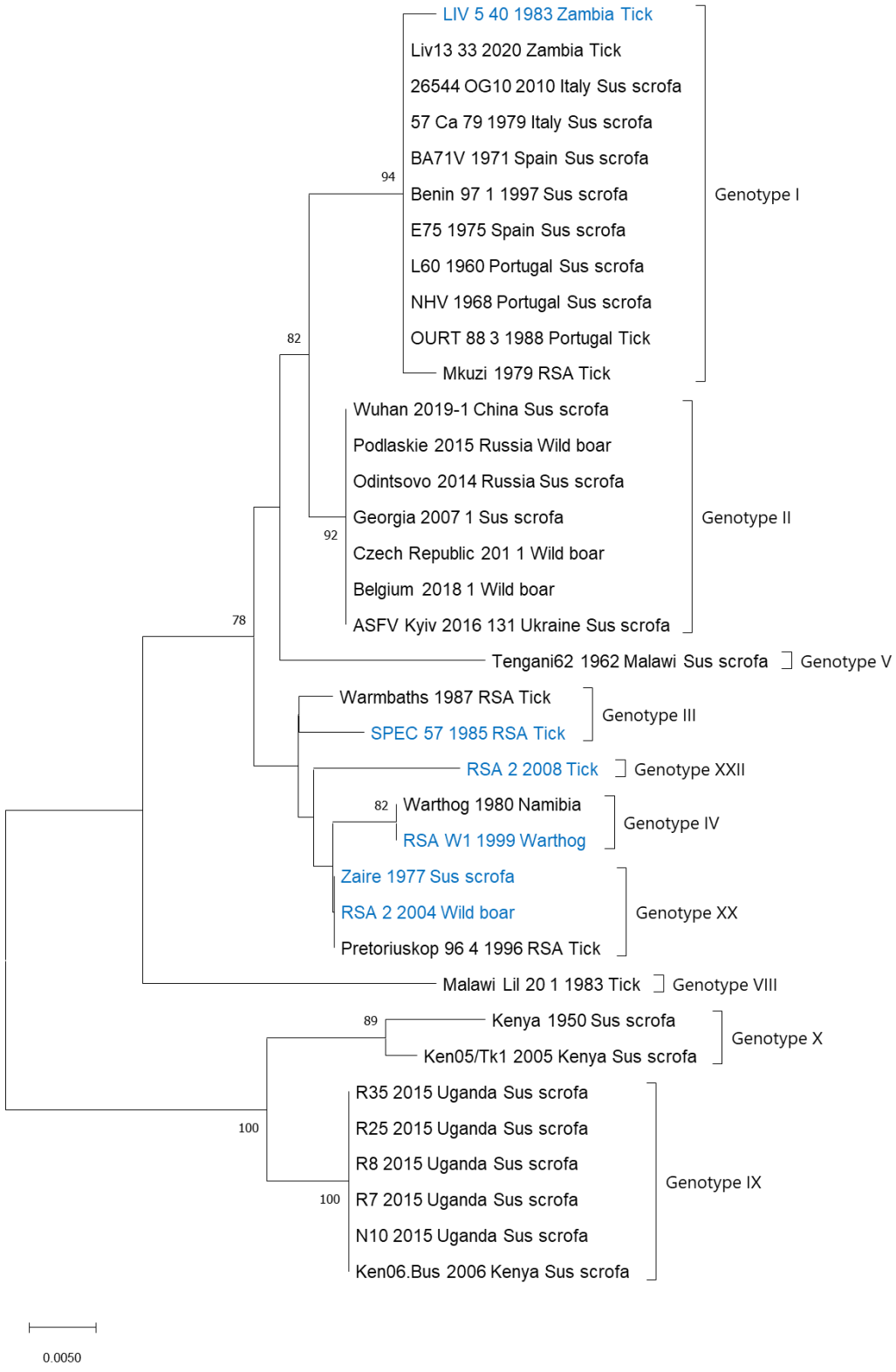


Figure 3. 2. Neighbour-Joining tree depicting p72 gene relationships of 36 African swine fever virus isolates, including six novel sequences described in this study (blue text). Isolate names are followed by the year of isolation, country, host, and genotype. Bootstrap values ≥ 70% are shown at branch nodes, and the scale bar (0.005) represents nt substitutions per site.

The genotypic assignment based on the p72 gene was used as a baseline to compare the ability of alternative gene regions to differentiate between the different ASFV isolates included in this study. The five novel genes, EP364R, B169L, B117L, B407L, and I177L, are found in the CCR of the genome (see Figure 2.2-2.7) and offer alternative to the use of the p72 gene in phylogenetic analyses. The pairwise nucleotide sequence identity was found to vary between 80% to 100% for B169L and B117L, 81% to 100% for I177L and 85% to 100% for EP364R and B407L.

The phylogenetic tree generated using the B117L ORF was not concordant with the p72 phylogeny. The phylogenetic grouping of genotypes II, V, VIII, IX and X were unchanged compared to the p72 phylogeny, clustering as separate monophyletic groups. However, the phylogenetic clustering of the remaining viruses differs significantly from the classical genotyping based on the p72 gene sequence, with the isolates grouping into 3 defined clusters (Figure 3.3). Genotypes I and II collapses into a single cluster with further sub-clustering within Cluster A not well supported. Interestingly, the isolates recovered from ticks collected in the Livingstone Game Park, Zambia, in 1983 appear to be genetically distinct from other viruses within Genotype I. Similarly, viruses collected from ticks in South Africa formed a supported cluster (Cluster B) distinct from viruses recovered from suids (Cluster C). The overall topology of the phylogenies based on the B169L (Figure 3.4) and B407L (Figure 3.5) were similar to that of B117L, but had lower bootstrap support for the various clusters. The clustering of southern African isolates based on host species was also not consistently retained in these phylogenies. However, the African genotype I viruses (Mkuzi_1997, Liv_5_40 and Liv 13_33) clustered as a monophyletic group distinct from the other genotype I isolates, suggesting that these viruses are evolutionarily distinct from the isolates that caused outbreak of the disease in Europe prior to 1990.

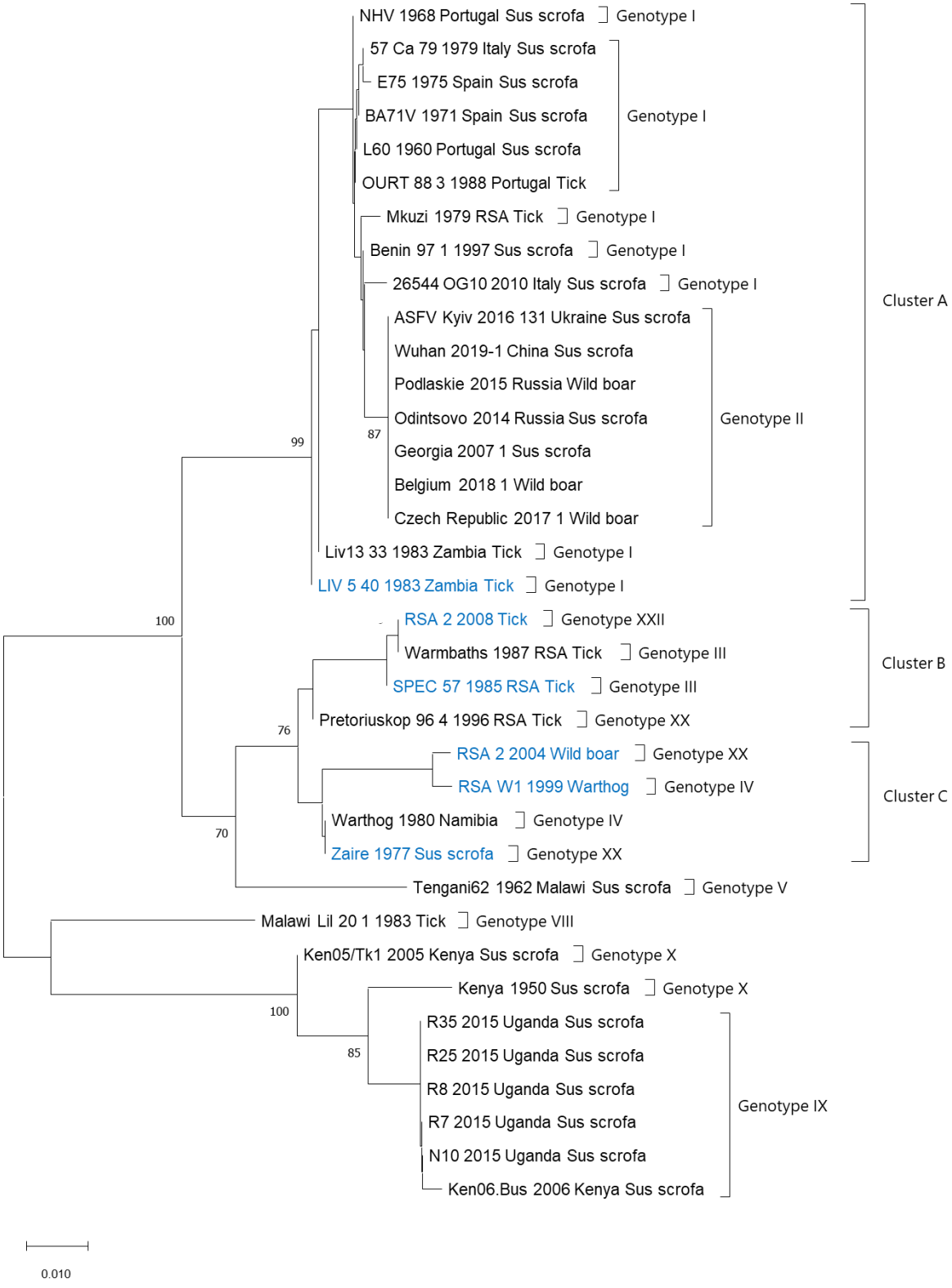


Figure 3. 3. Neighbour-Joining tree depicting B117L genes relationship of 36 African swine fever virus isolates, including six novel sequences described in this study (blue text). Isolate names are followed by the year of isolation, country, host, and genotype. Bootstrap values ≥ 70% are shown at branch nodes, and the scale bar (0.001) represents nt substitutions per site.

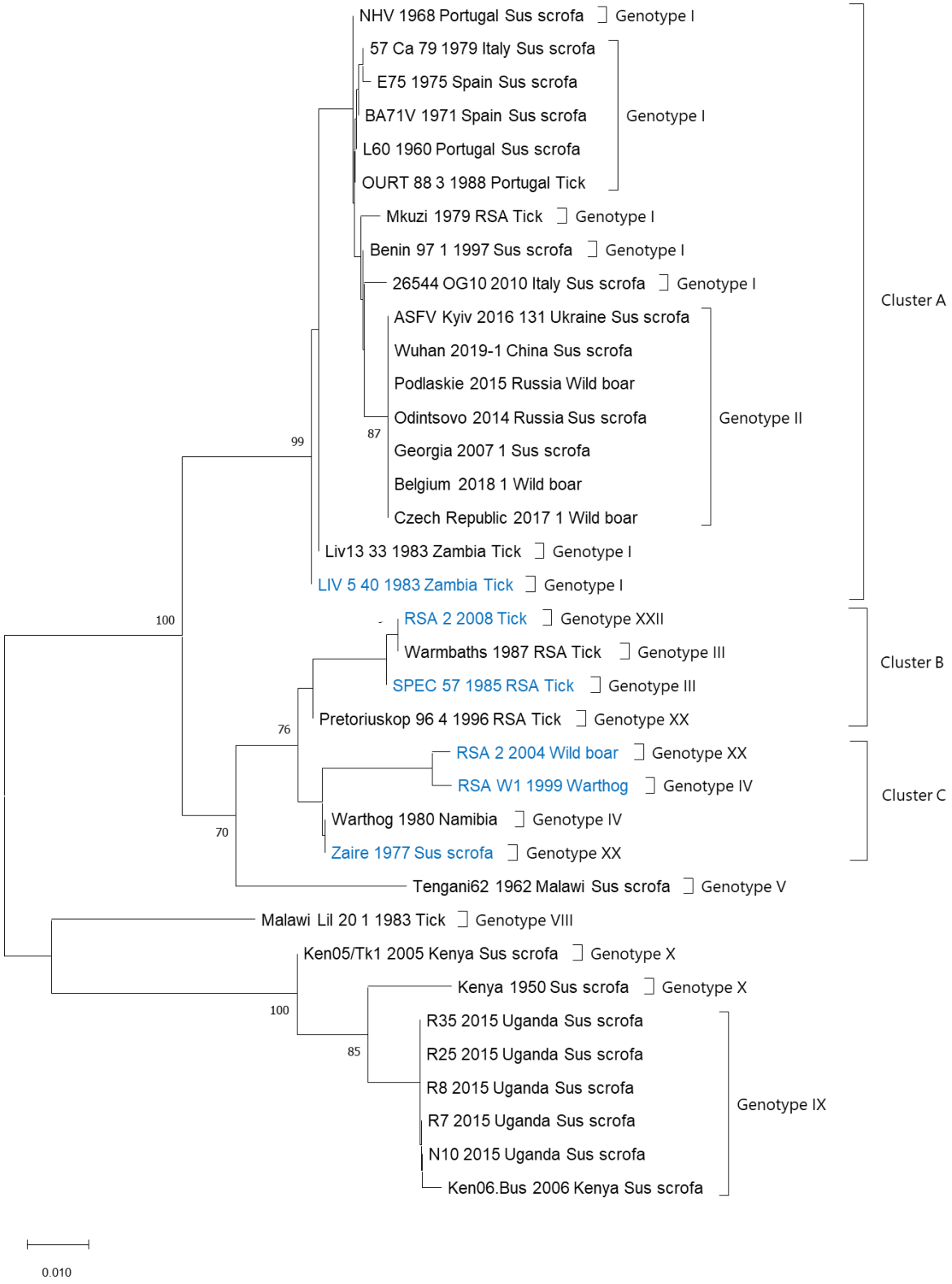


Figure 3. 4. Neighbour-Joining tree depicting B169L gene relationships of 36 African swine fever virus isolates, including six novel sequences described in this study (blue text). Isolate names are followed by the year of isolation, country, host, and genotype. Bootstrap values ≥ 70% are shown at branch nodes, and the scale bar (0.010) represents nt substitutions per site.

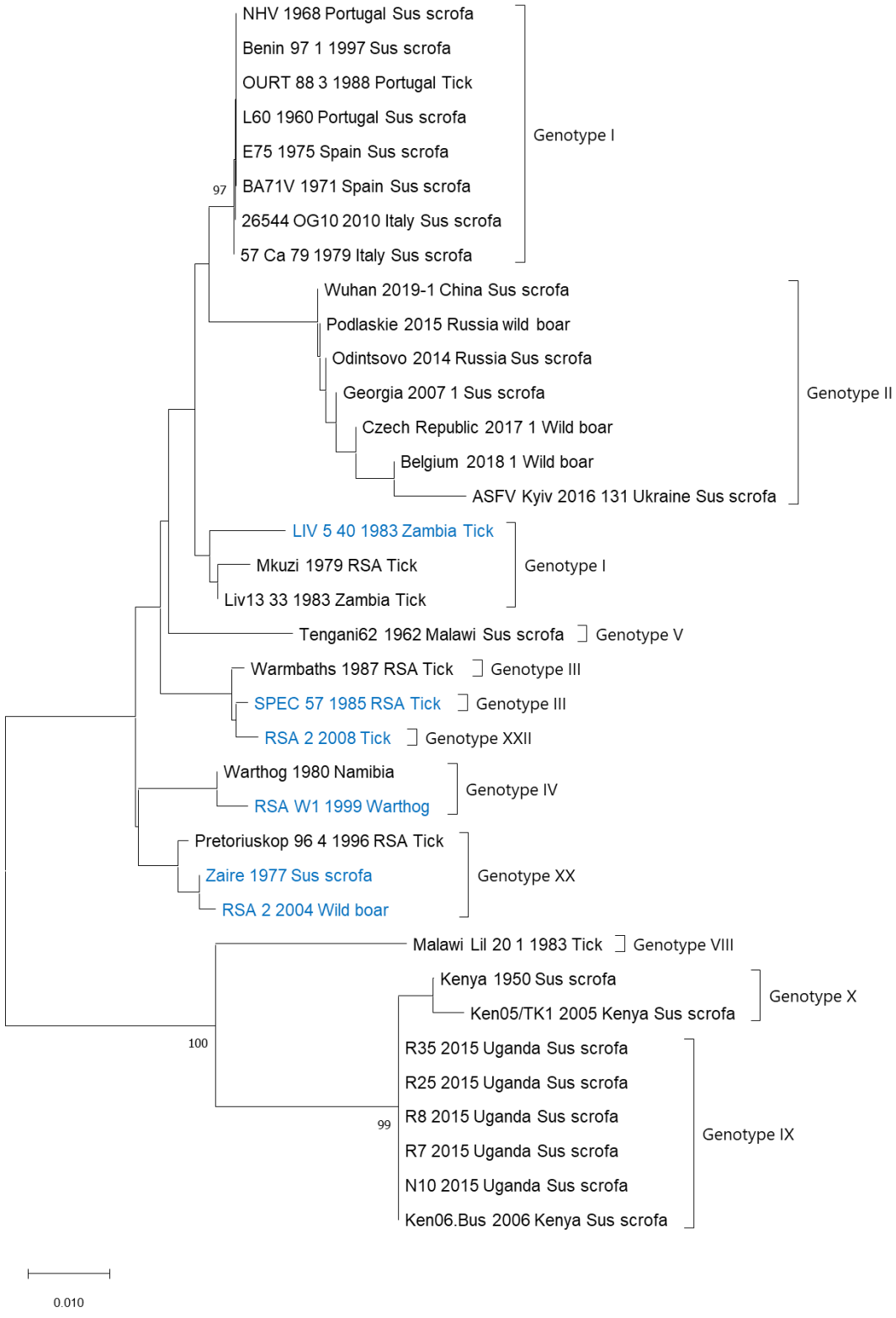


Figure 3. 5. Neighbour-Joining tree depicting B407L gene relationships of 36 African swine fever virus isolates, including six novel sequences described in this study (blue text). Isolate names are followed by the year of isolation, country, host, and genotype. Bootstrap values ≥ 70% are shown at branch nodes, and the scale bar (0.010) represents nt substitutions per site.

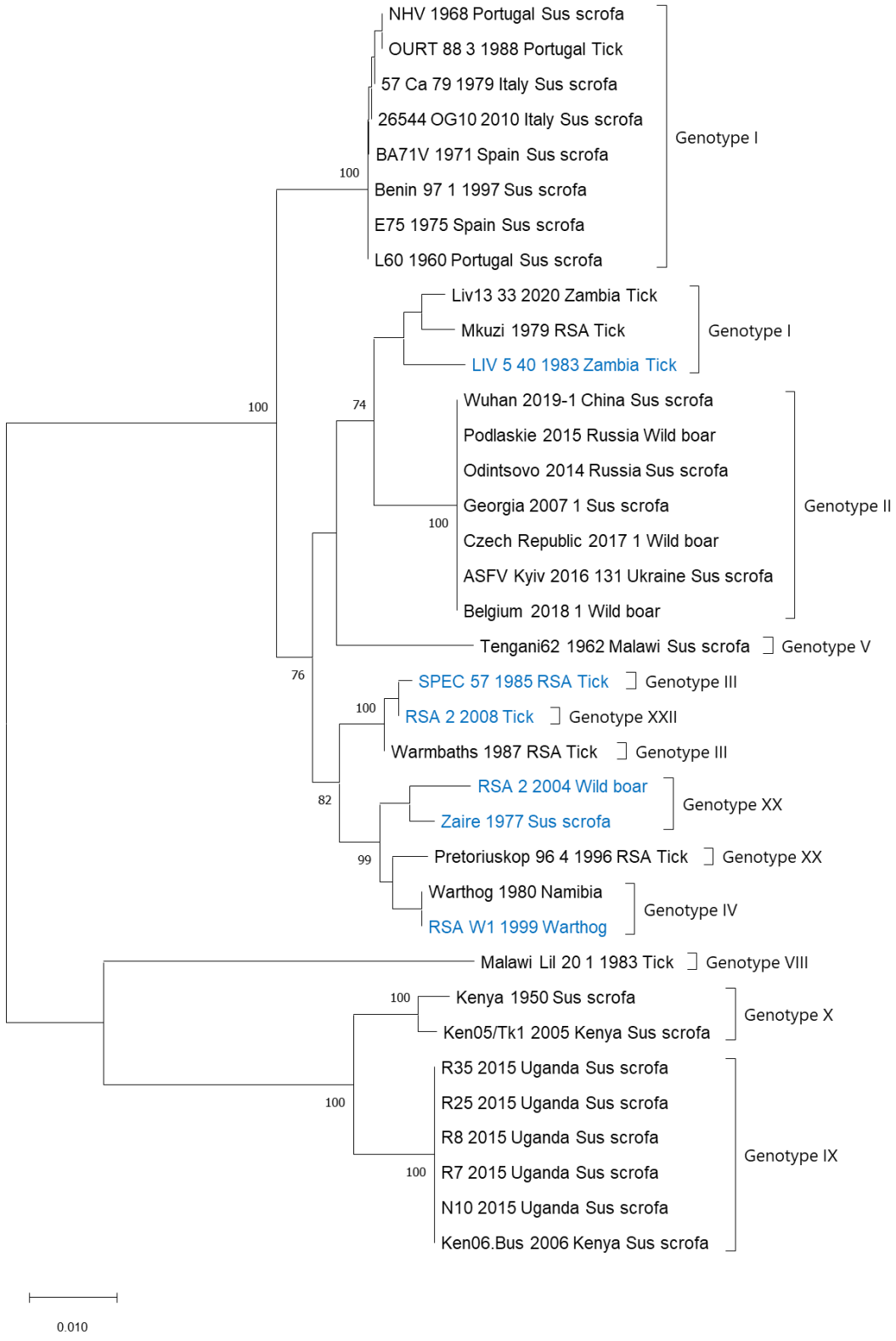


Figure 3. 6. Neighbour-Joining tree depicting I177L gene relationships of 36 African swine fever virus isolates, including six novel sequences described in this study (blue text). Isolate names are followed by the year of isolation, country, host, and genotype. Bootstrap values ≥ 70% are shown at branch nodes, and the scale bar (0.010) represents nt substitutions per site.

The overall topology of the phylogenies based on the I177L (Figure 3.6) and EP364R (Figure 3.7) differed significantly from that of the other genes. The monophyletic cluster of the African genotype 1 viruses was also present in the phylogeny based on the I177L gene, clustered as a monophyletic group distinct from the other genotype I isolates. The phylogenetic grouping of genotypes II, VIII, IX and X were unchanged compared to the p72 phylogeny. Genotypes IV and V formed a new sub-cluster, associated with viruses belonging to genotype 1, but the cluster separation of the sub-clusters was not supported by sufficiently high bootstrap values. The classification of the remaining African isolates was also not concordant with the p72 phylogeny. The biggest difference in tree topologies was seen when the EP364R (Figure 3.7) and p72 phylogenies were compared. The discordance between the two phylogenies was partially attributed to a number of large indels within the EP364R gene. However, the topologies remained incongruent even when all positions containing gaps and missing data were eliminated.

To assess whether larger portions of the genome would increase the resolution of the phylogenetic analysis, phylogenetic trees were generated using a concatenated dataset combining the five genes described in the preceding section (Figure 3.8.), as well as the core central conserved region (CCR) of the genome (see Chapter 2, Table 2.5). The CCR is defined as the 109 ORFs conserved in all ASFV isolates starting and ending at A224L and I196L ORFs. The ORFs situated within the CCR encode proteins involved in nucleotide metabolism, transcription, replication and repair, structural proteins, and proteins involved in morphogenesis. The phylogenetic arrangement of isolates based on the CCR (Figure 3.9) closely resembles that p72 phylogeny. The three African genotype I viruses (Mkuzi_1997, LIV_5_40 and Liv13_33) were distinct from the remaining genotype I viruses further supporting the contention that, these viruses are evolutionarily distinct from the isolates that caused outbreak of the disease in Europe prior to 1990. Interestingly, RSA_2_2008, isolated from ticks in South Africa appeared to be closely related to viruses within genotype III. This suggests that the phylogeny based on p72 reflect more recent evolutionary divergence that are only discernible when individual genes as analysed. The five novel concatenated genes formed a sub-cluster formed by a warthog isolate RSA_W1_1999 and a wild boar isolate RSA_2_2004, not observed in trees generated with the p72 and the 109 core concatenated conserved set of genes.

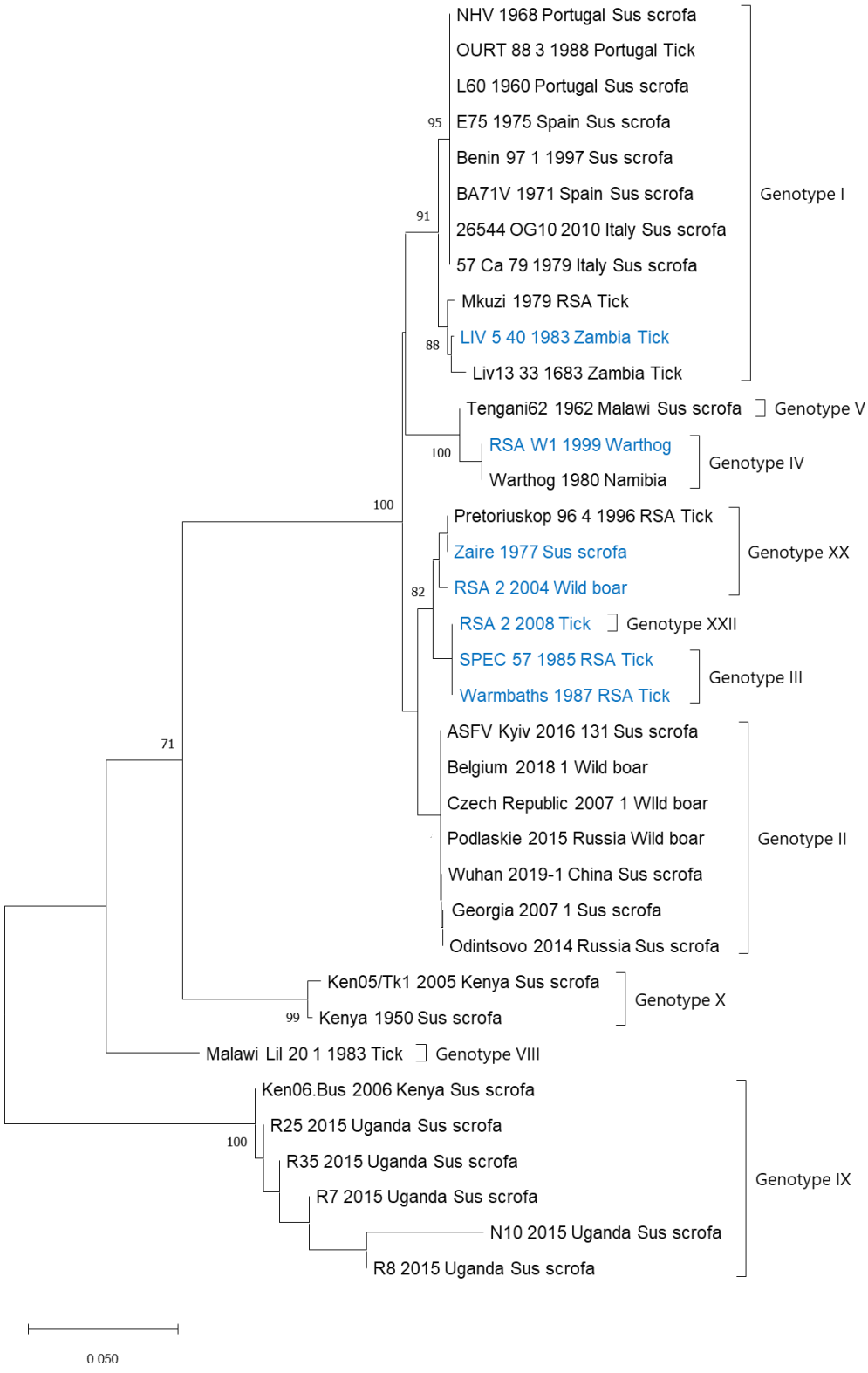


Figure 3. 7. Neighbour-Joining tree depicting EP364R gene relationships of 36 African swine fever virus isolates, including six novel sequences described in this study (blue text). Isolate names are followed by the year of isolation, country, host, and genotype. Bootstrap values ≥ 70% are shown at branch nodes, and the scale bar (0.050) represents nt substitutions per site.

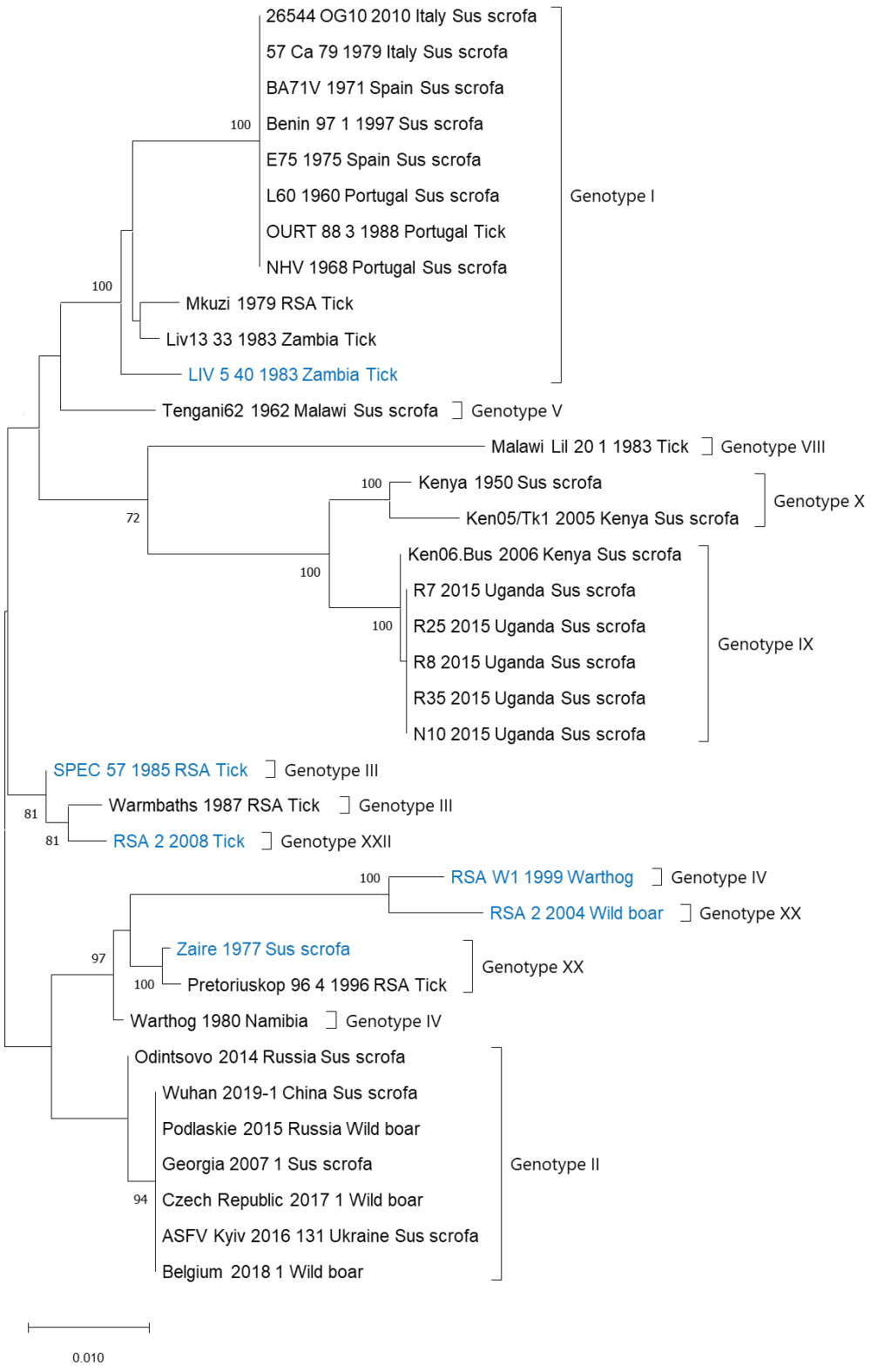


Figure 3. 8. Neighbour-Joining tree depicting five novel concatenated genes from 36 African swine fever virus isolates, including six novel sequences described in this study (blue text). Isolate names are followed by the year of isolation, country, host, and genotype. Bootstrap values ≥ 50% are shown at branch nodes, and the scale bar (0.010) represents nt substitutions per site.

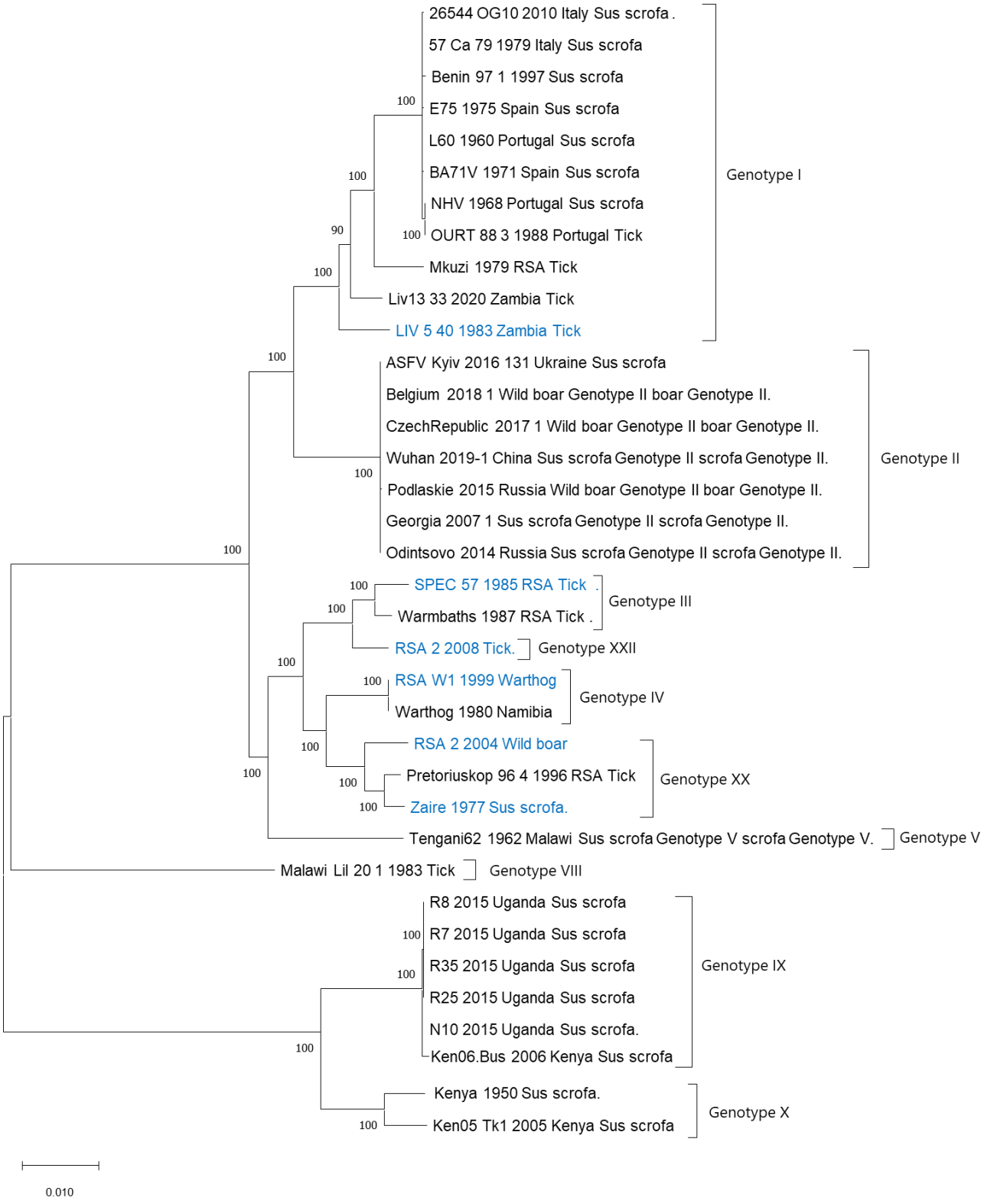


Figure 3. 9. Unrooted maximum-likelihood phylogenetic tree constructed from the alignment of the 109 core concatenated conserved set of genes from 36 ASFV isolates, including six novel sequences described in this study (blue text). Isolate names are followed by the year of isolation, country, host, and genotype. Bootstrap values ≥ 70% are shown at branch nodes, and the scale bar (0.010) represents nt substitutions per site.

To date, the publicly available whole-genome sequences that have been grouped within serotypes using phylogenetics include, BA71V, Benin_97_1, E75, L60, NHV, Georgia_2007_1, Kenya_1950, Malawi_Lil_20_1, Mkuzi_1979, Pretoriuskop_96_4 Tengani62, Warmbaths, and Warthog. European isolates and western African isolates representing genotype I clustered in SG4 (BA71V, E75, L60, NHV, and Benin_97_1). The European genotype II isolate (Georgia_2007_1) and the African isolate Malwi_Lil_20_1 clustered in SG 8, while Eastern and southern African isolates Warthog, Tengani62, and Kenya_1950, representing genotype IV, V, and X clustered in SG2, SG6, and SG7, respectively. Mkuzi_1979, Warmbaths, and Pretoriuskop_96_4 isolates from southern Africa representing genotype I, III, and XX, did not cluster with any of the known eight serogroups.

In this study, we classified the six ASFV for which the complete genome sequences were generated (highlighted in blue) into the previously defined serogroups based on the EP402R (Figure 3.10) and EP153R (Figure 3.11) gene sequences and compared it to serogrouping of all publicly available sequences used in this study (highlighted in green). Four of the virus sequences in this study could be assigned to serotypes based on gene sequences. LIV_5_40 (genotype I) clustered within SG1 whilst SPEC_57 (genotype III) and Zaire (genotype XX) clustered with viruses in SG8 and SG3, respectively. RSA_2_2008, representing genotype XXII, clustered in SG4. RSA_W1_1999 and RSA_2_2004 (genotype IV) grouped as outliers, and could not be assigned to a serotype. All European strains from genotype I and II analysed in this study clustered in SG4 and SG8, consistent with previous reports. Isolates from Uganda (R7, R8, R25, R35, and N10), and a Kenyan Isolates Ken06.Bus representing genotype IX, grouped with SG1 strains, whilst Ken05.Tk1 isolated in Kenya representing genotype X, clustered in SG6.

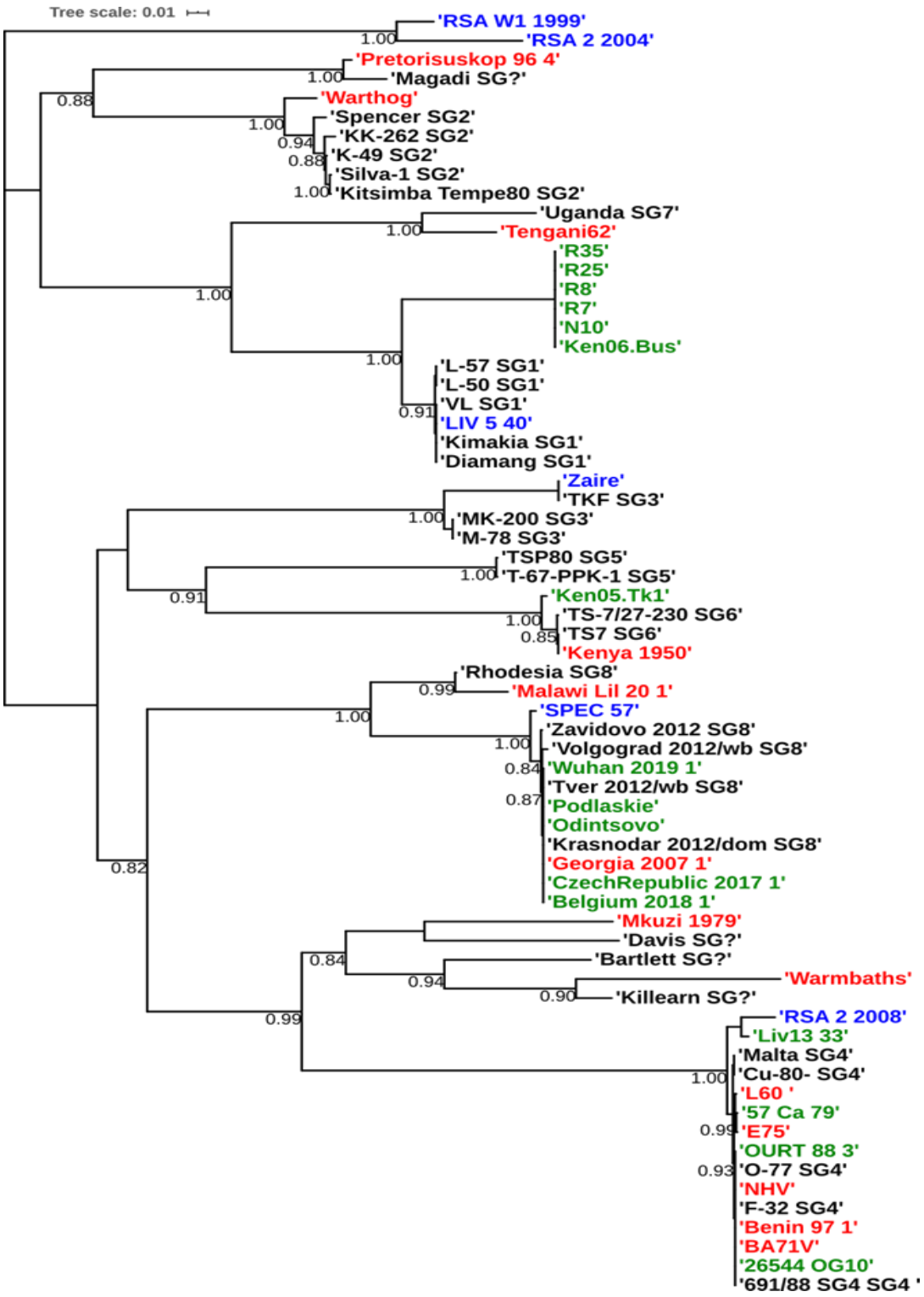


Figure 3. 10. Unrooted maximum-likelihood phylogenetic tree constructed from the alignment of the EP402R gene. The six novel sequences described in this study (blue text), reference sequences described in this study (green text), previously described references sequences (red text). Bootstrap values ≥70% are shown at branch nodes, and the scale bar (0.01) represents nt substitutions per side.

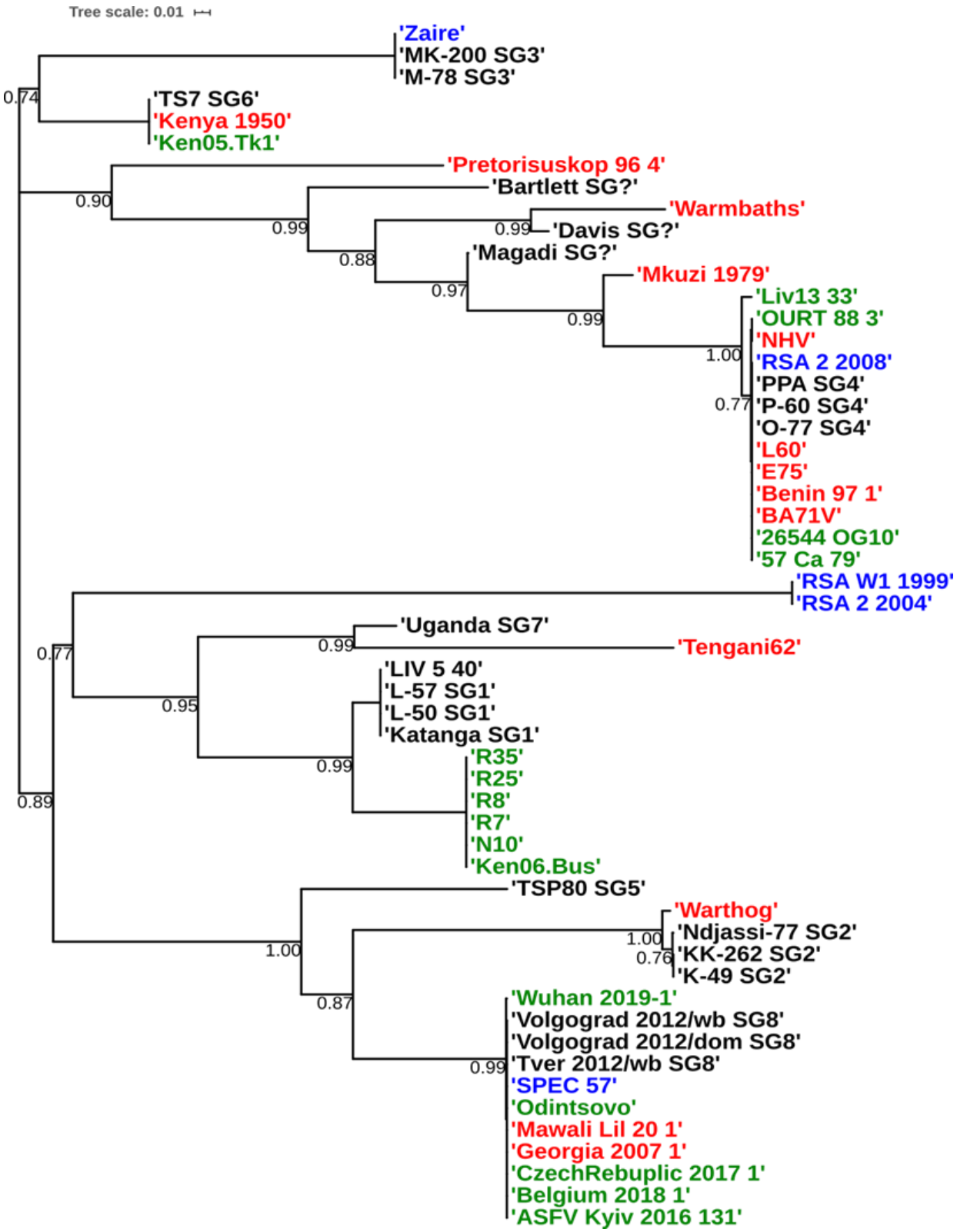


Figure 3. 10. Unrooted maximum-likelihood phylogenetic tree constructed from the alignment of the EP153R gene. The six novel sequences described in this study (blue text), reference sequences described in this study (green text), previously described references sequences (red text). Bootstrap values ≥70% are shown at branch nodes, and the scale bar (0.01) represents nt substitutions per side.

4.4. DISCUSSION

Phylogenetic analysis of ASFV based on individual gene sequences has proven useful in delineating the epidemiology of the disease. However, the inability of p72 phylogenies to discriminate between epidemiologically unrelated isolates prompted investigations into alternative targets that would enable researchers to unravel evolutionary relationships. The search for alternative targets focused predominantly on genes encoding proteins that make up part of the virus structure. This was predicated on the assumption that these genes would undergo positive selections driven by immune pressure. Despite this, the designation of isolates based on p72 remains the preferred option with regards to ASFV molecular epidemiology.

It was shown that ASFV isolates were clustered into four major lineages based on the partial sequence of the p72 gene (Michaud, Randriamparany, & Albina, 2013). Lineage 1 consisted of previously described genotypes I, II, and XVII, while lineage 2 was comprised of isolates representing genotypes III, IV, V, VI, VII, XIX, XX, XXI, and XXII. Lineage 3 was formed by isolates corresponding to genotypes VIII, XI, XII, XIII, XV, and XVI, and Lineage 4 included genotypes IX and X. The six novel ASFV isolates were classified into Lineage 1 (LIV_5_40; Genotype I), and Lineage 2 (SPEC_57; Genotype III, RSA_W1_1999; Genotype IV, RSA_2_2004; Genotype XX, Zaire; Genotype XX, and RSA_2_2008; Genotype XXII).

However, the p72 gene does not provide an adequate resolution between closely related ASFV isolates within genotypes. Sequencing of other genes that include B602L, KP86R, and E183L (p54), have provided better resolution between related ASFV isolates. It was demonstrated that the B602L and KP86R genes and the IGR discriminated 43 closely related ASFV isolates into 17 different sub-groups (Nix *et al*., 2006). It was shown that 67 ASFV isolates from Europe, America, and East and West Africa, which clustered as genotype I by the partial sequencing of the p72 gene, were further discriminated into four major sub-groups based on the E183 (p54) gene (Gallardo *et al*., 2009). Previous studies have suggested that more ASFV genotypes might exist, thus, more genetic markers should be identified for intra-genotype classification (Achenbach *et al*., 2017, Gallardo *et al.*, 2014).

The recent increase in the number of ASFV whole genome sequences that have been generated may provide further insight into the selection of alternative genes that could be used in phylogenetic analyses. The introduction of ASF into Asia resulted in a significant increase in the number of publically available whole genome sequences. Unfortunately, the

majority of recently published genome sequences are from virus isolates that group within genotype II. In addition, it is reasonable to assume that all of the reports of ASFV outbreaks in western Europe and Asia since 2007 can ultimately be traced back to a single introduction of the virus into Georgia (Chapman *et al*., 2011). Analysis of the genotype II whole genome sequences of ASF viruses from Western Europe and Asia revealed very low genetic diversity, with these viruses sharing more than 99.9% nucleotide sequence identity across the genome. This suggests that the ASF genome is genetically stable and that the virus evolves slowly when circulating within a domestic pig cycle.

It is therefore essential that the breadth of whole genome sequence data be expanded to include multiple genotypes and epidemiological scenarios. In this study, we compared the whole genome sequences from six novel viruses from warthogs, domestic pigs, European wild boar and ornithodoros spp. ticks with publically available whole genome sequences in an attempt to identify gene targets that would provide increased phylogenetic resolution when comparing closely related viruses.

In this study, five novel loci (EP364R, B169L, B117L, B407L, and I177L), were identified as possible alternative phylogenetic markers. The tree topology of all of the alternative phylogenies differed from the current p72 classification. B117L and B169L provided slightly better resolution of genotypes I and II, respectively. Viruses from East Africa that are classified as belonging to genotype IX based on p72 were separated when using EP364R.

Genotype I isolates clustered closely together. Previous studies have shown that European, Caribbean, South American, and Western Africa (ESAC-WA) isolates are closely related and cluster together, most likely due to a single introduction of the virus from the west-coast shipping route between Africa and Europe (Bastos *et al*., 2003). In this study, we observed eastern and southern African genotype I isolates (LIV_5_40 and Mkuzi_1979) forming part of the ESAC-WA genotype. However, this association was not consistently maintained for all of the genes analysed, suggesting that the p72 phylogeny does not accurately reflect that evolutionary relatedness of the genotype I viruses. It is essential that additional information on the epidemiology of outbreaks also be considered when interpreting the phylogenetic classification of ASFV. This is further exemplified when considering the phylogeny of the European genotype II viruses. The low rate of genetic changes observed in ASFV may preclude the use of classical phylogenetic approaches based on the use of individual genes to

discriminate between epidemiologically linked viruses. Thus, full genome sequencing could be used to discriminate between epidemiological linked viruses.

To assess the usefulness of using multiple genes in phylogenetic analyses, we investigated the use of a concatenation of the five alternative genes and the central conserved region of the genome. The 109 core concatenated conserved set of genes and the five novel concatenated genes also clustered isolates into three clades. Trees generated with the concatenated genes both have a clade formed by eastern African isolates representing genotypes VIII, IX, and X. A major drawback of the phylogenies described above is that they do not relate to biological significance. To address this concern, classification of ASFV into serogroups based on the ability of sera from convalescent animals to inhibit the haemadsorption in swine macrophages was proposed (Malogolovkin *et al*., 2015a). Eight serogroups (SG) have been identified based on this haemadsorption inhibition assay (Malogolovkin *et al*., 2015a). These serogroups correspond to the phylogenetic analyses of the EP402R (CD2v) and the EP153R (C-type lectin) genes, which play a role in haemadsorption (Malmquist and Hay, 1960). In this study, we used EP402R and EP153R genes to phylogenetically assign the six novel ASFV isolates to the appropriate serogroup. Four of the six novel ASFV isolates, LIV_50, Zaire, RSA_2_2008, and SPEC_57, were assigned to SG1, SG3, SG4, and SG8, respectively. The other two isolates RSA_W1_1999 and RSA_2_2004 clustered as outliers and could not be assigned to existing serogroups. Of the 67 reference genome sequences in the GenBank, only ten isolates were successfully serotyped by phylogenetics (Malogolovkin *et al*., 2015b). In this study, we further assigned all other reference genome sequences representing genotype I and II to SG4 and SG8, respectively. Six Eastern Africa isolates representing genotype IX clustered in SG1, while a Kenyan genotype X isolate, Ken05.Tk1, clustered in SG6. The three ASFV isolates Mkuzi_1979, Pretoriuskop_96_4, and Warmabths, all originating from South Africa could not be phylogenetically serotyped (Malogolovkin *et al*., 2015b). Study isolate RSA_W1_1999, and RSA_2_2004 that we could not serotype phylogenetically were also isolated in South Africa. ASFV isolates clustering as outliers to reference serogroups, could represent novel serogroups, but this will need to be empirically confirmed. The data generated in this study support previous studies that the EP402R and EP153R genes are sufficient in serotyping ASFV isolates by phylogenetics and also that strain diversity is observed in geographical areas where more than one mechanism of transmission exists. Future vaccine research could use this information when developing

country specific vaccines because ASFV isolates have the ability to cross-protect against challenge with viruses of the same serogroup.

The aim of this study was to use the p72 gene as baseline to compare the ability of the five novel genes to differentiate between closely related ASFV isolates included in this study. The presence of multiple genotypes within a country adds to the diversity of ASF epidemiology, specifically in Eastern and Southern Africa, where the sylvatic cycles exist. These novel genes were able to cluster the ASFV isolates by genotype, geographical, and by the host. We propose these novel genes as alternative phylogenetic markers to the p72 gene that could be used to analyse other ASFV isolates representing all 24 genotypes from different geographical areas, in order to investigate their ability in tracing virus spread. The data obtained in this study confirms the need to identify and investigate other genome regions to further characterise closely related ASFV isolates. Regions with reasonable number of mutations were assessed in the CCR, given the relative conservation of this region. Viral genomes can mutate at the nucleotide level and yet maintain their essential characteristics at the protein level and virion level, and this is the basis for molecular characterisation of viruses. However, it is not always possible to find genes that a conserved across all viral genomes, or sufficiently diverged to be used for phylogenetic analysis. Thus, phylogenetic analysis based of full genomes will be the gold standard to characterise different ASF viruses.

CHAPTER 4: COMPARISON OF AFRICAN SWINE FEVER VIRUS (ASFV) MULTIGENE FAMILIES (MGFs)

4.1. Introduction

A multigene family is defined as a set of similar genes that have descended as a result of duplication of common ancestor genes, which are referred to as paralogs (Nei and Rooney, 2005). Comparison of ASFV genome sequences has shown high variability at the terminal ends of the genome. The terminal ends are comprised mainly of the MGFs and are characterised as regions of insertions and deletions of up to 10 kbp (Portugal *et al.*, 2015). The variations observed at the left and right variable regions are mainly a result of gain or loss of the genes in the MGFs. The members of the MGFs were annotated based on the average number of amino acids encoded by each protein and the direction they are read in (Chapman *et al.*, 2008). There are five MGF members: MGF 100, MGF 110, MGF 300, MGF 360, and MGF 505/530 (Almendral *et al*., 1990; Chapman *et al*., 2008; Gonzalez *et al*., 1990; Rodriguez *et al*., 1994). The MGFs are found at the terminal ends of the genome, except for the MGF 360-15R, which is located in the central conserved region.

The MGFs are found adjacent to one another and reading in the same direction, showing that they have evolved by gene duplication (Rodriguez *et al*., 1994). A similar method of gene amplification under strong natural selection is observed in poxviruses (Elde *et al*., 2012). Thus, it is thought that ASFV isolates use gene duplication for altered host tropism. It has been shown that repeated passage of the poxviruses in cell culture results in attenuation through the loss of multigene families, specifically, those that encode for the ankyrin repeat proteins, which are thought to play an essential role in the poxvirus host-range (Haller *et al*., 2014; Laidlaw and Skinner, 2004). Genes in the MGF 360 and 505/530 are associated with host range and virulence (Afonso *et al*., 2004; Neilan *et al*., 2002; Tulman and Rock, 2001), thus, the loss in some of the MGF 360 and 505/530 families is associated with attenuation.

The difference in gene copy number between ASFV isolates has resulted in challenges when annotating the MGFs based on the BA71V nomenclature, specifically for unique genes and paralogs within each MGF. The truncation and fusion of genes in the ASFV genomes results in inconsistencies or mis-annotations of these regions. Thus, the MGFs of ASFVs were re-annotated using standard nomenclature, and gene maps were created that will be used in future genome comparison studies (Imbery & Upton, 2017). Their approach to these challenges in annotating MGF was to create whole-genome multiple sequence alignments

using MAFFT, to highlight the differences. Dot plots were created for each MGF gene and compared against full genome reference sequences. MGF maps were created, (Figure 4.1) which could be edited to incorporate newly sequenced ASFV genomes and new MGF orthologous genes when discovered, which will make annotations of the ASFV genomes much easier (Imbery & Upton, 2017).

The MGFs 100, 110, and 300 have not been studied in detail. The MGF 100 and MGF 300 have fewer paralogs and are less variable. The MGF 100 contains three paralogs, MGF 300 contains four paralogs, and the MGF 110 contains fourteen paralogs (Imbery & Upton, 2017). Previous studies have shown that adaptation of ASFV virus isolates on tissue-culture results in the loss of members of the MGF 110; the L60 isolate lost two MGF 110, while the BA71V lost seven MGF 110 genes (Almendral *et al*., 1990; de la Vega *et al*., 1990; Pires *et al*., 1997). Subsequent studies analysed the progressive adaptation of Georgia (ASFV-G) and Odintsovo strains on Vero cells and CV-1 Cells, respectively. The MGF 110 members were not deleted during the adaptation process in either isolates, however, point mutations and frameshifts were observed in a limited number of genes in the ASFV-G isolate. The Odintsovo isolate contains a similar deletion of 3 000 bp as the BA71V strain, which results in the deletion of the I7L, I8L, I9L, I10L, L11L, and MGF360-18R genes (Mazloum *et al*., 2018). It has been suggested that the potential function of the MGF 110-4L and MGF 110-6L might be preparing the endoplasmic reticulum for a role in morphogenesis because some members of MGF 110 contain an endoplasmic reticulum (ER) retention motif sequence at their C-terminus. The MGF 110-6L contains a retention sequence KDEL, while the MGF 110-4L contains a KEDL homologous sequence (Netherton *et al*., 2004).

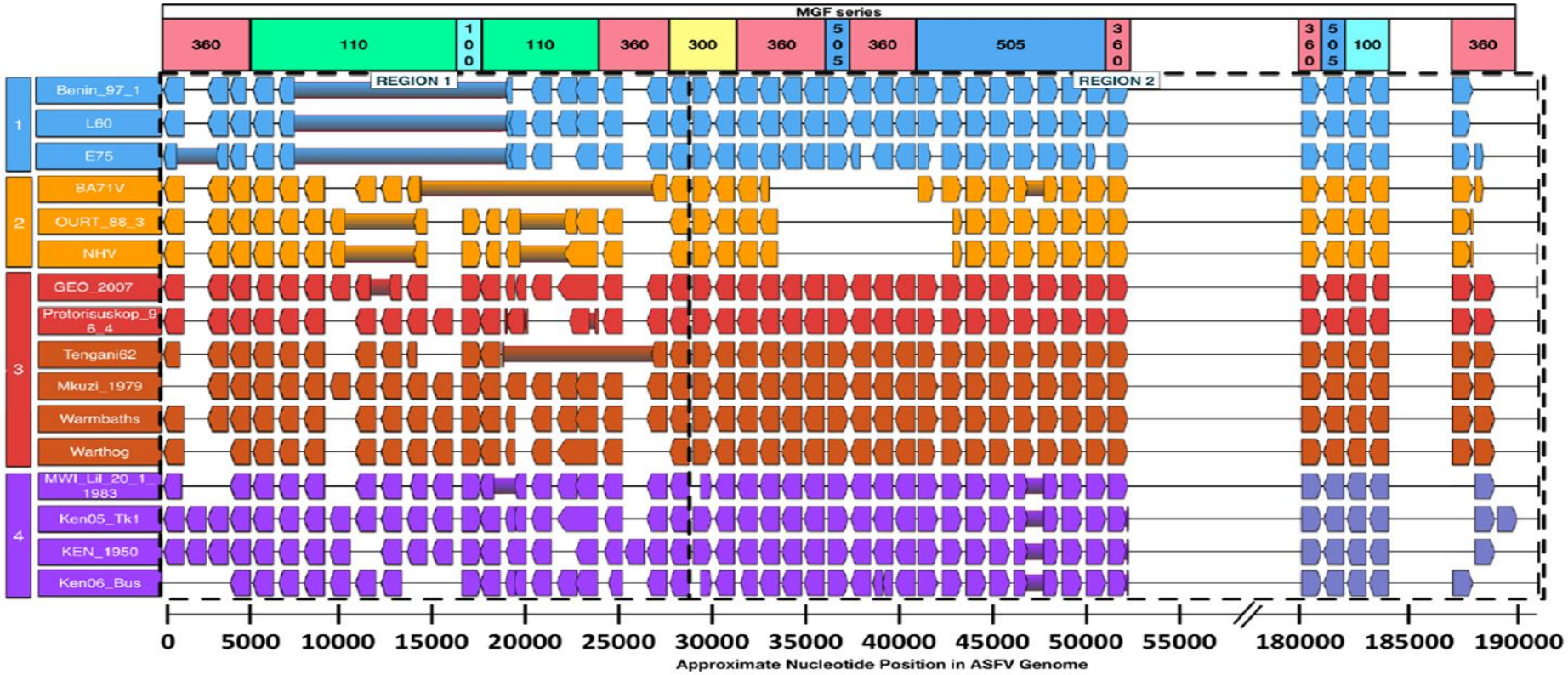


Figure 4. 1. Display of all MGF paralogs as they appear, left to right, in the 16 ASFV strains as described by Imbery and Upton 2017. Columns of boxes represent individual MGF orthologues. Location of the different MGF series is annotated above the diagram. The arrangement of MGF orthologues separates the genomes (coloured rows) into 4 groups that correlate well with the phylogenetic tree. It is notable that the MGF compilation can be split into 2 regions, between MGF 300-1L and MGF 300-2R, reflecting significantly more variation in region1 than 2. Nucleotide positions are given at the bottom of the figure. The image was taken from (Imbery & Upton, 2017).

The roles of the MGF 360 and MGF 505/530 are not fully understood, but several studies have shown that they have unspecified roles in host range and virulence by promoting the survival of infected cells (Neilan *et al*., 2002; Zsak *et al*., 2001b). The highest copy number and most variation are observed in the MGF 360, which contains 22 paralogs. The cell-culture adapted ASFV isolate BA71V, and the low virulence isolates, NHV and OURT_88_3, contain less of the MGF 360, both encoding 11 copies of the 22 paralogs. The MGF 505/530 contains 11 paralogs, and ten of them are found at the left variable region, while the MGF 505-11L is located at the right variable region of the genome (Chapman *et al*., 2008, Rodriguez *et al*., 2015, Imbery & Upton, 2017).

The deletion of the three MGF 360 genes (MGF 360-12L, 13L, and 14L) from the Pretoriuskop_96_4 isolate reduced its ability to replicate in soft ticks (Burrage *et al*., 2004). It was demonstrated that restoration of six copies of the MGF 360 genes (MGF 360-9L, 10L, 11L, 12L, 13L, and 14L) along with the MGF 505-1R, restored the ability of the cell-culture adapted BA71V isolate to replicate in macrophages. Deletion of five of the MGF 360 and three MGF 505/530 in the highly virulent Benin_97_1 and Georgia_2007_1 strains resulted in attenuation (Zsak *et al*., 2001a; Zsak *et al*., 2001b). These MGF families are also thought to play a role in virulence, based on the fact that, the OURT_88_3 which is a low virulence strain, does not contain five of the MGF 360 (MGF 360-10L, 11L, 12L, 13L, and 14L) and two of the MGF 505/530 (MGF 505-1R, and 2R) which are found in virulent strains (Zsak *et al*., 2001, Chapman *et al*., 2008). The virulent isolates can withstand the host interferon response as compared to low virulence strains (OURT 88_3), where the host IFN response results in limited virus replication (Golding *et al*., 2016).

To date, 24 genotypes have been described (Bastos *et al*., 2003, Lubisi *et al*., 2005, Boshoff *et al*., 2007, Achenbach *et al*., 2017), and 16 full genome sequences were used in the study by Imbery and Upton (2017), which represent genotypes I, II, III, IV, V, VIII, IX and XXII (Bishop *et al*., 2015; Chapman *et al*., 2008; de Villiers *et al*., 2010; Portugal *et al*., 2015; Rodriguez *et al*., 2015). It has been shown that European, Caribbean, South American, and West African (ESAC-WA) isolates, representing genotype I cluster together (Bastos *et al*., 2003). In this study, we have used the framework by Imbery and Upton (2017) to conduct an updated MGF analysis of six novel ASFV isolates representing genotypes I, III, IV, XX, and XXII, and 30 publicly available sequences from GenBank. Although there are similarities in the MGFs between ASFV, some differences are specific to ASFV isolates that cluster together (Imbery & Upton, 2017). The comparison of the MGFs across ASFV isolates is of

utmost importance in understanding genome variability and their contribution to virulence. The aim of this chapter is to compare the MGFs across 73 ASFV isolates using sorting and clustering to understand their contribution to genome variation.

4.2. MATERIALS AND METHODS

Data set: MGF paralogs from 73 genome sequences were analysed in this study, however, a total of 36 isolates were used for comparison (Table 4.1). Some ASFV isolates belonging to genotypes I, II, and IX have similar MGF arrangement patterns, with minor differences that include fusion, truncation, and fragmentation of ORFs, and one isolate was chosen to represent them. Five genotype IX isolates from Uganda have similar MGFs arrangement, and one strain (N10) was selected from the group to represent them. Genotype II isolates from Belgium, China, Georgia, and Poland are also similar to each other, and one isolate (Belgium_2018_1, China_CAS19-01, Georgia_2007_1, and Pol16_20538_o9) was chosen from each country. A total of eleven isolates (ASFV_Kyiv_2016_131, Belgium_2018_1, China_CAS19-01, CzechRepublic_2017_1, Georgia_2007_1, Hu_2018, LT14_1490, Moldova, Odintsovo, Podlaskie, Pol16_20538_o9), were chosen to represent genotype II isolates. Fourteen isolates from Italy representing genotype I are also similar to each other, and one isolate (57_Ca_79) was selected for comparison. Ten genotype I isolates (BA71V, Benin_97_1, E75, L60, LIV_5_40, Mkuzi_1979, NHV, OURT_88_3, and 57_Ca_79), were used for comparison (Table 4.1).

Eleven representatives of the European genotype II isolates, a representative (57_Ca_79) of the fourteen Italian isolates, a representative (N10) of the five Ugandan isolates, and all other isolates encoding the MGF 110-4L and 6L were selected. Thus, 28 ASFV isolates were used for comparison of the ER motif (KDEL) (Figure 4.2), and 30 ASFV isolates for (KEDL) (Figure 4.3). Multiple sequence alignments (MSA) were performed using MAFFT in Geneious Prime version 1.2 (https://www.geneious.com) and visualised in Bioedit (Hall, 1999). Morpheus software (Morpheus, https://software.broadinstitute.org/morpheus) was used to cluster and sort the MGFs. In the Morpheus software, the colours are numerically coded, 0 coded for absent genes with no colour. 0.49 coded for truncated genes with a red colour, 0.5 coded for fused genes with a yellow colour, 0.51 coded for fragmented genes with green colour, and 1 coded for present genes with a blue colour. This numerical colour coding was done to allow differentiation between gene states and to allow for hierarchical clustering based on similarities between gene status. The true cost of a truncation vs fusion vs

fragmentation could only be arbitrarily estimated and thus was kept in a tight range of 0.49 – 0.51. The hierarchical clustering was based on the Euclidean distance metric and average linkage methods.

Table 4. 1. ASF viruses used in this study, and the representative isolate for some genotypes.

Representative isolates	**Isolate**	**Country**	**Genotype**	**Accession no**
Not applicable	L60	Portugal	I	KM262844
	NHV	Portugal	I	KM262845
	BA71V	Spain	I	NC_001659
	E75	Spain	I	FN557520
	Mkuzi_1979	South Africa	I	AY261363
	LIV_5_40	Zambia	I	MN318203
	Liv13/33	Zambia	I	MN913970
	OURT_88_3	Portugal	I	AM712240
	Benin_97_1	Benin	I	AM712239
57/Ca/79	56/Ca/78	Italy	I	MN270969
	57/Ca/79	Italy	I	MN270970
	139/Nu/81	Italy	I	MN270971
	140/Or/85	Italy	I	MN270972
	85/Ca/85	Italy	I	MN270973
	141/Nu/90	Italy	I	MN270974
	142/Nu/95	Italy	I	MN270975
	60/Nu/95	Italy	I	MN270976
	26/Ss/04	Italy	I	MN270977
	72407/Ss/05	Italy	I	MN270978
	97/Ot/12	Italy	I	MN270979
	222653/Ca/14	Italy	I	MN270980
	47_Ss_2008	Italy	I	KX354450
	26544_OG10	Italy	I	KM102979
Not applicable	Estonia	Estonia	II	LS478113
Representative isolates	**Isolate**	**Country**	**Genotype**	**Accession no**
China_CAS19-01	China_2018_AnhuiXCGQ	China	II	MK128995

	Pig_HLJ_2018	China	II	MK333180
	China_CAS19-01	China	II	MK33181
	ASFV-wbBS01	China	II	MK645909
	ASFV-SY18	China	II	MH766894
	DB_LN_2018	China	II	MN172368
	Wuhan 2019-1	China	II	MN393476
	Wuhan 2019-2	China	II	MN393477
Pol16_20538_o9	Pol16_20186_o7	Poland	II	MG939583
	Pol16_20540_o19	Poland	II	MG939585
	Pol16_29413_o23	Poland	II	MG939586
	Pol16_20538_o9	Poland	II	MG939584
	Pol17_03029_C201	Poland	II	MG939587
	Pol17_04461_C210	Poland	II	MG939588
	Pol17_05838_C220	Poland	II	MG939589
Belgium_2018_1	Belgium_Etalle_wb_2018	Belgium	II	MK543947
	Belgium_2018_1	Belgium	II	LR536725
Georgia_2007_1	Georgia_2007_1	Georgia	II	FR682468
	Georgia_2008_1	Georgia	II	MH910495
	Georgia_2008_2	Georgia	II	MH910496
Not applicable	Kashino_04/13	Russia	II	KJ747406
	Odintsovo	Russia	II	KP843857
	LT14/1490	Luthuania	II	MK628478
	Podlaskie	Poland	II	MH681419
	ASFV_Kyiv_2016_131	Ukraine	II	MN194591
	Pol16_20538_o9	Poland	II	MG939584
	CzechRepublic_2017_1	Czech Republic	II	LR722600
	Moldova	Moldova	II	LR722599
	Hu_2018	Hungary	II	MN715134
	SPEC_57	South Africa	III	MN394630
	Warmbaths	South Africa	III	AY261365
	Warthog	Namibia	IV	AY261366
	RSA_W1_1999	South Africa	IV	MN641876
	Tengani62	Malawi	V	AY261364
Representative isolates	**Isolate**	**Country**	**Genotype**	**Accession no**
Not applicable	Malawi_Lil_20_1	Malawi	VIII	AY261361
N10	R7	Uganda	IX	MH025917

	R8	Uganda	IX	MH025916
	R25	Uganda	IX	MH025918
	R35	Uganda	IX	MH025920
	N10	Uganda	IX	MH025919
Not applicable	Ken06.Bus	Kenya	IX	KM111295
	Kenya_1950	Kenya	X	AY261360
	Ken05_Tk1	Kenya	X	KM111294
	Zaire	Zaire	XX	MN630494
	Pretoriuskop_96_4	South Africa	XX	AY261363
	RSA_2_2004	South Africa	XX	MN641877
	RSA_2_2008	South Africa	XXII	MN336500

Study isolates in bold.

4.3. RESULTS

Dot plot analysis (Figure 2.8-2.11) of the six novel ASFV isolates revealed deletions at the LVR, CCR, and RVR in LIV_5_40 and Zaire isolates. Deletions at the LVR were observed in RSA_W1_1999 and RSA_2008, while SPEC_57 showed deletions at the RVR. These deletions resulted in the truncation of the MGF 110-13L and MGF 360-4L ORFs in LIV_5_40. The LIV_5_40 isolate further encodes fragmented MGF 110-14L, 360-2L, 360-9L, 505-5R, and 505-9R ORFs. The RSA_W1_1999 and RSA_2_2004 isolates encode fragmented MGF 110-11L ORFs. The MGF 110-13L and 14L ORFs are fragmented in the RSA_2_2008 and SPEC_57 isolates. (Table 4.2). All study isolates are missing the MGF 110-3L, 110-10L, MGF 300-3L, and the MGF 360-6L, 7L, 17R, 20R, and 22R ORFs. The study isolates were further compared to other ASFV isolates from GenBank.

Table 4. 2. Comparison of the MGF 100, 110, 300, 360, and 505 members of six novel ASFV isolates.

						MGF 100			MGF 110													MGF 300				MGF360																							MGF 505										
Isolates	Year	Country	Host	Serotype	Genotype	1R	2L	3L	1L	2L	3L	4L	5L	6L	7L	8L	9L	10L	11L	12L	13L	14L	1L	2R	3L	4L	1L	2L	3L	4L	5L	6L	7L	8L	9L	10L	11L	12L	13L	14L	15R	16R	17R	18R	19R	20R	21R	22R	1R	2R	3R	4R	5R	6R	7R	8R	9R	10R	11L
LIV_5_40	1983	Zambia	Tick	SG1	I																																																										
SPEC_57	1985	RSA	Tick	SG8	III																																																									
RSA_W1_1999	1999	RSA	Warthog	N/A	IV																																																									
RSA_2_2004	2004	RSA	W. boar	N/A	XX																																																									
Zaire	1977	Zaire	Pig	SG3	XX																																																									
RSA_2_2008	2008	RSA	Tick	SG4	XXII																																																									

N/A-Not available

	Absent
	Present
	Fused
	Fragmented
	Truncated

MGF 100: The MGF 100 members are highly conserved among ASFVs, and all ASFV isolates used in this study encode three paralogs of the MGF 100, except for L60, BA71V, E75, Benin_97_1, Estonia, and ASFV_Kyiv_2016_131 strains. The MGF 100-2L ORF in the ASFV_Kyiv_2016_131 isolate encodes a protein with two fragments, not observed in other isolates (Figure 4.6).

MGF 110: A significant degree of difference is observed in members of the MGF 110 encoded by different ASFV strains. The MGF 110-11L ORF in isolates RSA_W1_1999 and RSA_2_2004 encode a protein with two fragments, also found in the LT14_1490, Ken06.Bus and Ken05_Tk1 strains. Cell-culture adapted isolates BA71V and Zaire, are both missing MGF 110-3L, 10L, and 12L ORFs. Genotype I strains L60, E75, Benin_97_1, and 57_Ca_79, isolated from domestic pigs, are missing MGF 110-3L, 4L, 5L, 6L, 7L, 8L, 9L, and 10L ORFs. All other ASF viruses isolated in Italy with the exception of 25644_OG10 strain encode two fragments of the MGF 110-13L ORFs, and a similar fragmented MGF 110-13L is found in Moldova_2017_1 and Belgium_2018_1 isolates, and study isolates RSA_2_2008 and SPEC_57. The MGF 110-13L ORF is truncated in the LIV_5_40 isolate. The MGF 110-14L ORF in isolates LIV_5_40, RSA_2_2008, and SPEC_57 encode a protein with two fragments, not observed in other isolates. Fused ORFs are encoded mostly by ASFV isolates representing genotype I and II isolates, compared to other ASFV isolates representing other genotypes. The MGF 110-13L and 14L ORFs are both missing in BA71V, Tengani62, and RSA_2_2004 strains, and present in other ASFV isolates. All genotype II isolates contain a fused ORF comprised of MGF 110-5L and 6L, except for Estonia. The MGF 110-5L is missing in the SPEC_57 isolate. The MGF 110-2L and 11L ORFs are fused in Benin_97_1, E75, and NHV strains, and not observed in other isolates. A fusion of the MGF 110-4L and 110-7L is observed in the NHV and OURT 88_3 strains, and not in other isolates. The MGF 110-3L and 4L, MGF 110-10L, and 110-14L are fused in the Czech Republic_2017_1, Moldova_2007_1, and Belgium_2018_1 strains, not observed in other genotype II isolates. The fusion of the MGF 110-7L and MGF 360-6L ORFs is only found in the BA71V strain. The MGF 110-8L ORF is encoded by all isolates from Uganda but only annotated in the R7 strain.

The NHV is the only strain that encodes a fused ORF comprised of the MGF 110-11L, 13L, and 14L. A genotype VIII isolate, Tengani62, encodes fused MGF 110-9L and 11L ORFs, not present in any other isolates used in this study. The MGF 110-7L is truncated only in the Tengani62 isolate. The MGF 110-13L and 14L ORFs are fused in the NHV, Italian isolates,

Georgia_2007_1, Warthog, Pretoriuskop_96_4, and Ken05_Tk1 strains, and not found in other isolates. The MGF 110-14L in the Warmbaths isolate is present but not annotated (Figure 4.6).

To identify whether the six novel study isolates contain the ER retention motif sequence (KDEL) and the homologous sequence (KEDL) at the C-termini of the MGF 110-6L gene and MGF 110-4L, a MSA of these genes from the six novel study isolates against reference sequences of these ORFs was performed and visualised in Bioedit. All six novel ASFV isolates contain the ER retention motif sequence (KDEL). However, no genotype II isolates contain the KDEL sequences at their C-terminus, due to substitution mutations. In genotype II strains, Lysine (K) is substituted by glutamic acid (E), aspartic acid (D) by asparagine (N), glutamic acid (E) by asparagine (N), and leucine (L) by serine (S) (Figure 4.2) giving ENNS. A deletion is observed at position 104-190 in ASFV isolates representing genotype I, III, IV, V, VIII, IX, X, XX, and XXII; in contrast, genotype II isolates contain an insertion in that position. At the amino acid level, the percentage similarity of the MGF 110-6L protein ranged from 45.4 % to 100% among isolates used in this study.

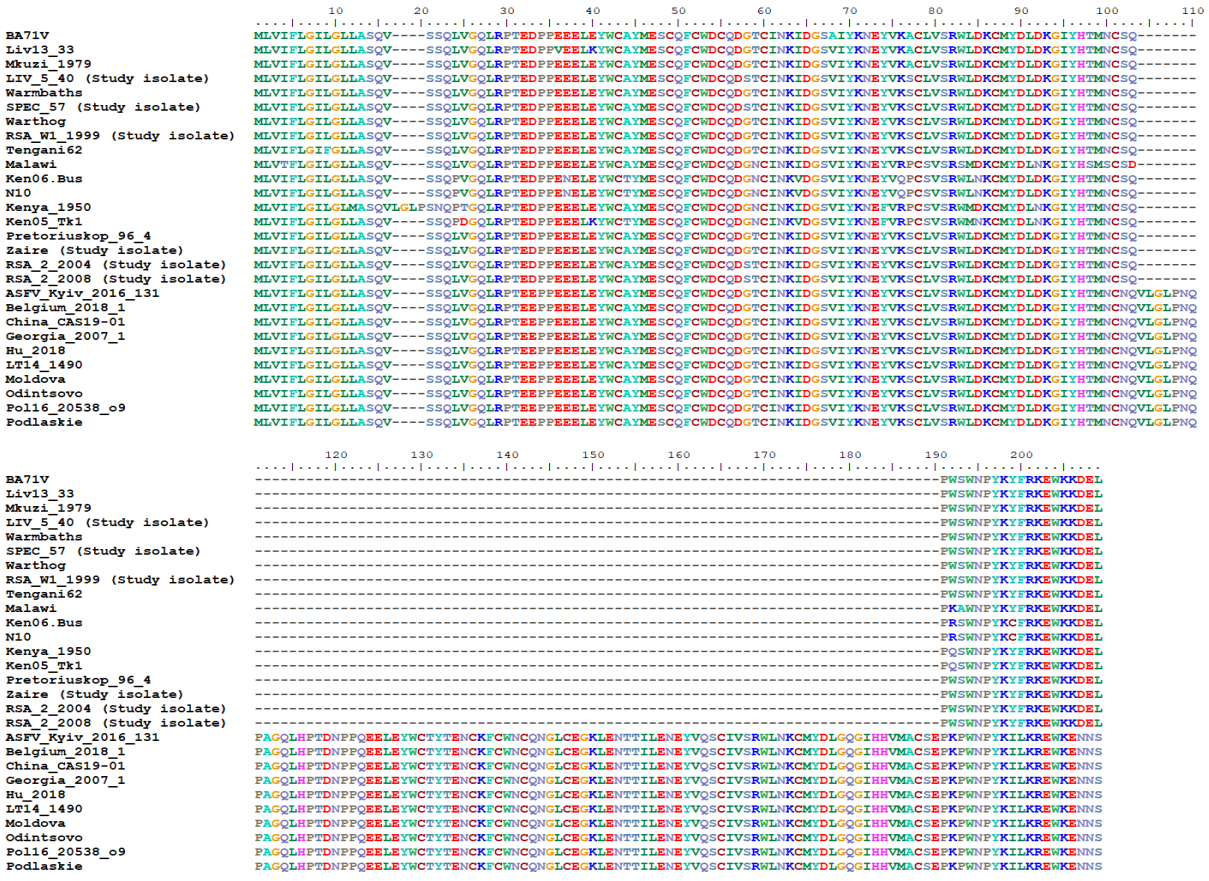


Figure 4. 2. Amino acid alignment of the MGF 110-6L gene encoding the ER retention motif sequence (KEDL), of 28 ASFV isolates.

The analysis of the homologous retention motif sequence (KEDL) found in the MGF 110-4L revealed that the glutamic acid (E) is substituted by lysine (K) in genotype II strains (Figure 4.3), except for Belgium_2018_1 and Moldova, where glutamic acid (E) is substituted by aspartic acid (D). At the amino acid level, the percentage similarity of the MGF 110-4L protein ranged from 84.7 % to 100% among isolates used in this study.

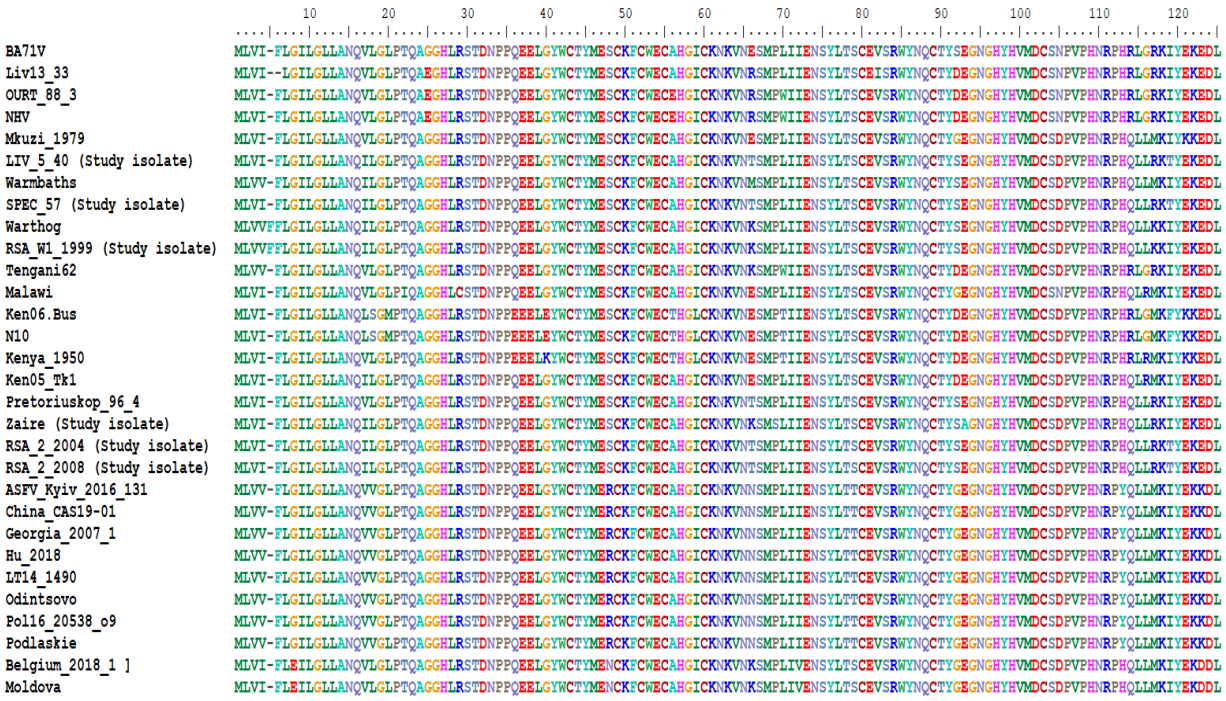


Figure 4. 3. Amino acid alignment of the MGF 110-4L gene encoding the ER retention motif sequence (KEDL), of 30 ASFV isolates.

MGF 300: Similar to the MGF 100, the MGF 300 paralogs are highly conserved among ASFV isolates. There are four paralogs of the MGF 300; however, the MGF 300-3L ORF is only found in E75, Benin_97_1, 57_Ca_79, (14 Italian isolates), Liv13_33, and N10 (5 isolates from Uganda). The MGF 300-3L and 4L ORFs are fused in E75, Benin_97_1, and 57_Ca_79 (Italian isolates). All isolates from Uganda are missing the MGF 300-4L ORF. The MGF 300-2R ORF is absent in the ASFV_Kyiv_2016_131 isolate, and it encodes two fragments of the MGF 300-4L ORF not found in other isolates (Figure 4.6).

MGF 360: The MGF 360 contains 22 members, and their presence or absence are closely associated with virus pathogenesis. The MGF 360-1L ORF in genotype II isolates from Belgium, China, Czech Republic, Hungary, Lithuania, Moldova, Poland, Russia, and Ukraine, encode a protein with two fragments. The MGF 360-1L ORF is found at the RVR in the Ken05.Tk1 isolate. A two fragment MGF 360-2L ORF is found in the LIV_5_40 not found in other isolates. The MGF 360-2L is missing in the RSA_W1_1999 isolate, truncated

in the N10 strains, and it is located at the RVR in the Warthog isolate. A fusion of the MGF 360-1L and 2L ORFs is only found in the E75 strain. The MGF 360-5L ORFs is missing in the RSA_W1_1999. BA71V and Zaire isolates, which are cell-culture adapted, are missing the MGF 360-4L, 5L, and 7L ORFs. A two fragment MGF 360-9L ORF is found in the LIV_5_40, and not observed in other isolates, while the RSA_W1_1999 encodes duplicate ORFs of the MGF 360-9L. The MGF 360-11L ORF is present in the N10 strain but not annotated. A truncated MGF 360-12L ORF is only found in the E75 and the MGF 360-13L ORF in the Ken06.Bus isolate encodes a protein with two fragments. The MGF 360-3L, 13L, and 14L ORFs in the ASFV_Kyiv_2016_131 isolate encode proteins with two fragments, and the MGF 360-3L_2 and 360-13L_2 ORFs are found at the RVR. The MGF 360-17R ORF is only found in the OURT 88_3 and NHV strains, and all five isolates from Uganda. The MGF 360-18R ORF is not present in isolates from Uganda; however, it is found in all other isolates, but it is located at the LVR in the Estonia isolate. The MGF 360-19R ORF is missing only in the Malawi_Lil_20_1, Ken06.Bus, N10, and ASFV_Kyiv_2016_131 strains.

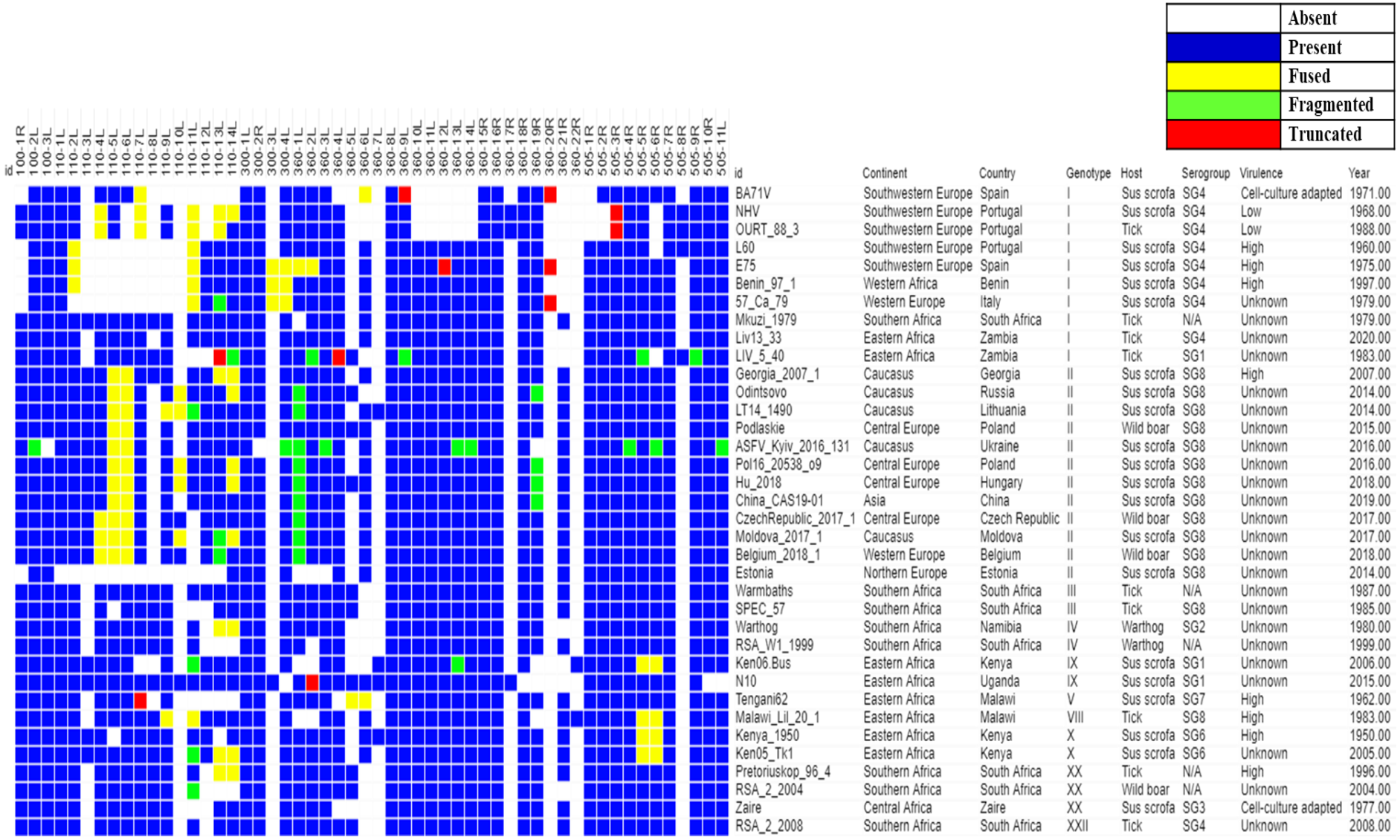
The MGF 360-19R is found at the LVR in the Kenya_1950, Ken05.Tk1 and Estonia isolates. The MGF 360-19R ORF in isolates Czech Republic_2017_1, Hu_2018, China_CAS19-01, and strains from Poland, encode a protein with two fragments. A truncated MGF 360-20R ORF is found in the 57_Ca_79, BA71V, and E75, not present in any other isolates. The MGF 360-21R is located at the LVR in the Malawi_Lil_20_1 and Estonia isolate and found at the RVR in other strains. The MGF 360-21R ORF is missing in the LIV_5_40 and RSA_W1_1999 isolates. The MGF 360-22R ORF is only found in the Ken06.Bus and Malawi_Lil_20_1 isolates (Figure 4.6).

MGF 505: Deletion of some members of the MGF 360 and 505/530 from the genome of pathogenic ASFV isolates results in attenuation**.** The NHV and OURT_88_3 strains encode a truncated 505-3R ORF, not found in other isolates. The MGF 505-4R ORF is present in the N10 strain but not annotated. The MGF 505-5R ORF in the LIV_5_40 encodes a protein with two fragments, while the MGF 505-9R ORF encodes a protein with three fragments, not found in other isolates. The RSA_2_2004 isolate encodes duplicate ORFs of the MGF 505-6R, while it is missing in LIV_5_40 and reference strains NHV, OURT_88_3, and L60. A fusion of the MGF 505-6R and 7R ORFs is only found in the Malawi_Lil_20_1, Kenya_1959, Ken05.Tk1, Ken06.Bus and BA71V strains. The MGF 505-8R ORF is only found in the LIV_5_40 isolate and isolates from Uganda. The MGF 505-10R and 11L are only missing in isolates from Uganda, and the MGF 505-9R is found at the RVR in these

strains. The MGF 505-4R, 6R, and 11L ORFs in the ASFV_Kyiv_2016_131 isolate encode proteins with two fragments (Figure 4.6).

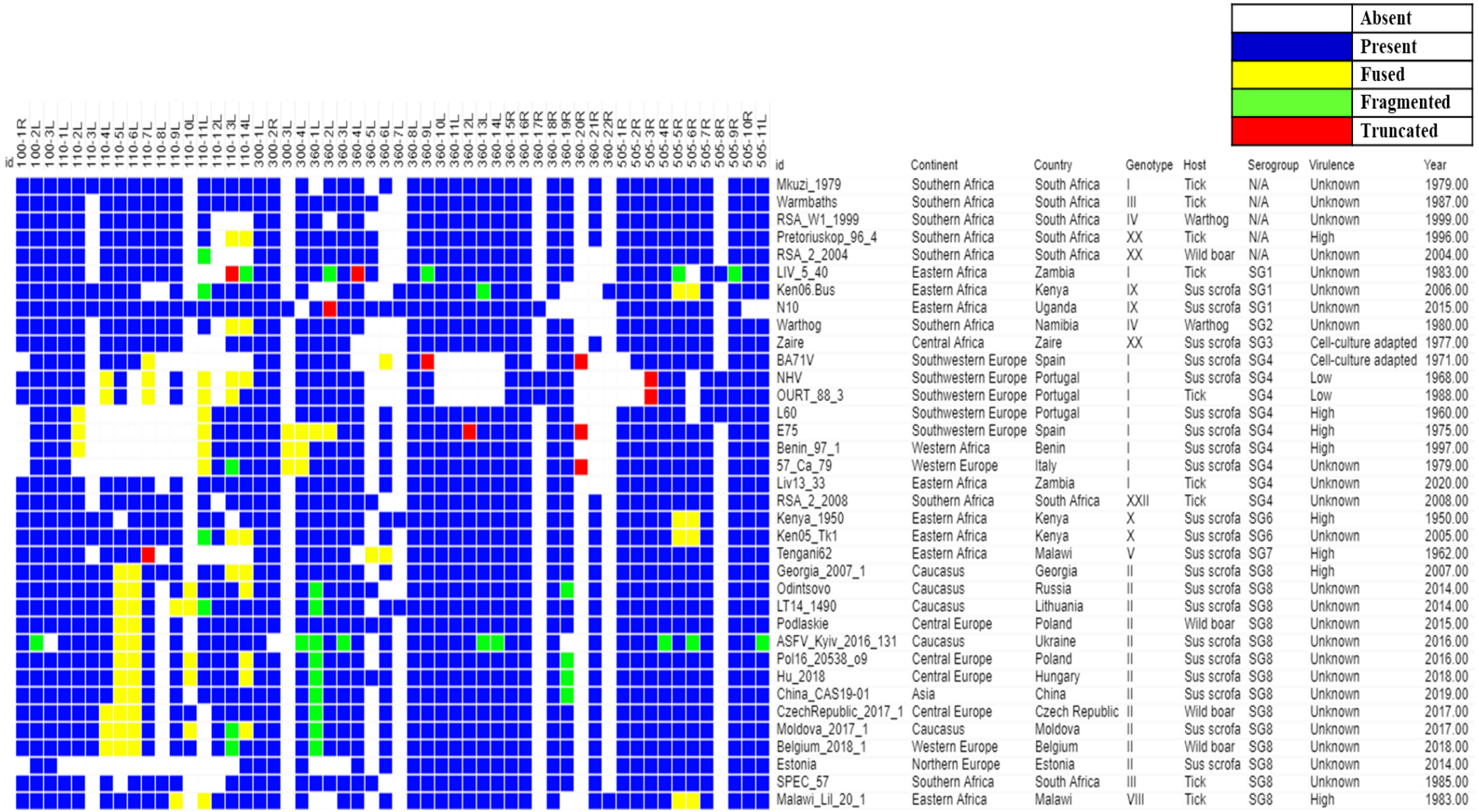
Sorting and Clustering of MGFs: The five MGFs were investigated by performing hierarchical clustering based on the state of each of the MGF genes (present, absent, truncated, fragmented, fused) using Morpheus software, and results were compared with results obtained by sorting the isolates by genotype, serogroup, country, host, virulence and year, to identify which variables may be associated with arrangement patterns of the MGFs found among ASFV isolates. Sorting of the isolates by genotype (Figure 4.4) and serogroup (Figure 4.5) reproduced similar clustering patterns observed with the hierarchical method. The European genotype I isolates clustered close to each other with the exception of the Benin_97_1 isolate, which was isolated in west Africa. European genotype II isolates also grouped close to each other. Although Liv13_33 and LIV_5_40 isolates represent genotype I, they do not cluster closely and consistently to European genotype I strain as the Benin_97_1 isolate does. However, these patterns were lost when the MGFs were sorted by country, host, virulence, and year (Data not shown).

The hierarchical method clustered ASFV isolates with similar MGF arrangement patterns into clades (Figure 4.6). Clade A is formed by BA71V, NHV, and OURT 88_3 strains, and clade B by L60, E75, Benin_97_1, 57_Ca_79, and Estonia isolates. Clade C is comprised of sub-clades formed by Odintsovo, Podlaskie, HU_2018, and Pol16_20538_o9, and China_CAS19-01 isolates, while clade D is formed by LT14_1490, Moldova_2017_1, Czech Republic_2017_1, and Belgium_2018_1. Clade E is comprised of Mkuzi_1979, Georgia_2007_1, Ken05_Tk1, and Kenya_1950. Clade F is comprised of sub-clades formed by RSA_W1_1999, Warthog, and RSA_2_2004 isolates, and SPEC_57, Warmbaths, Pretoriuskop_96_4, RSA_2_2008, and Zaire isolates. Clade G was formed by Malawi_Lil_20_1 and Ken06.Bus strains. The ASFV_Kyiv_2016_131 and Liv13_33 clustered in the same grouping as clades C, D and E, while LIV_5_40 and Tengani62 grouped close to clade F strains. The N10 strain grouped did not group with any of the major clades. The hierarchical clustering method maintained the close clustering pattern of the European genotype I and II isolates.



id	Continent	Country	Genotype	Host	Serogroup	Virulence	Year
BA71V	Southwestern Europe	Spain	I	Sus scrofa	SG4	Cell-culture adapted	1971.00
NHV	Southwestern Europe	Portugal	I	Sus scrofa	SG4	Low	1968.00
OURT_88_3	Southwestern Europe	Portugal	I	Tick	SG4	Low	1988.00
L60	Southwestern Europe	Portugal	I	Sus scrofa	SG4	High	1960.00
E75	Southwestern Europe	Spain	I	Sus scrofa	SG4	High	1975.00
Benin_97_1	Western Africa	Benin	I	Sus scrofa	SG4	High	1997.00
57_Ca_79	Western Europe	Italy	I	Sus scrofa	SG4	Unknown	1979.00
Mkuzi_1979	Southern Africa	South Africa	I	Tick	N/A	Unknown	1979.00
Liv13_33	Eastern Africa	Zambia	I	Tick	SG4	Unknown	2020.00
LIV_5_40	Eastern Africa	Zambia	I	Tick	SG1	Unknown	1983.00
Georgia_2007_1	Caucasus	Georgia	II	Sus scrofa	SG8	High	2007.00
Odintsovo	Caucasus	Russia	II	Sus scrofa	SG8	Unknown	2014.00
LT14_1490	Caucasus	Lithuania	II	Sus scrofa	SG8	Unknown	2014.00
Podlaskie	Central Europe	Poland	II	Wild boar	SG8	Unknown	2015.00
ASFV_Kyiv_2016_131	Caucasus	Ukraine	II	Sus scrofa	SG8	Unknown	2016.00
Pol16_20538_o9	Central Europe	Poland	II	Sus scrofa	SG8	Unknown	2016.00
Hu_2018	Central Europe	Hungary	II	Sus scrofa	SG8	Unknown	2018.00
China_CAS19-01	Asia	China	II	Sus scrofa	SG8	Unknown	2019.00
CzechRepublic_2017_1	Central Europe	Czech Republic	II	Wild boar	SG8	Unknown	2017.00
Moldova_2017_1	Caucasus	Moldova	II	Sus scrofa	SG8	Unknown	2017.00
Belgium_2018_1	Western Europe	Belgium	II	Wild boar	SG8	Unknown	2018.00
Estonia	Northern Europe	Estonia	II	Sus scrofa	SG8	Unknown	2014.00
Warmbaths	Southern Africa	South Africa	III	Tick	N/A	Unknown	1987.00
SPEC_57	Southern Africa	South Africa	III	Tick	SG8	Unknown	1985.00
Warthog	Southern Africa	Namibia	IV	Warthog	SG2	Unknown	1980.00
RSA_W1_1999	Southern Africa	South Africa	IV	Warthog	N/A	Unknown	1999.00
Ken06.Bus	Eastern Africa	Kenya	IX	Sus scrofa	SG1	Unknown	2006.00
N10	Eastern Africa	Uganda	IX	Sus scrofa	SG1	Unknown	2015.00
Tengani62	Eastern Africa	Malawi	V	Sus scrofa	SG7	High	1962.00
Malawi_Lil_20_1	Eastern Africa	Malawi	VIII	Tick	SG8	High	1983.00
Kenya_1950	Eastern Africa	Kenya	X	Sus scrofa	SG6	High	1950.00
Ken05_Tk1	Eastern Africa	Kenya	X	Sus scrofa	SG6	Unknown	2005.00
Pretoriuskop_96_4	Southern Africa	South Africa	XX	Tick	N/A	High	1996.00
RSA_2_2004	Southern Africa	South Africa	XX	Wild boar	N/A	Unknown	2004.00
Zaire	Central Africa	Zaire	XX	Sus scrofa	SG3	Cell-culture adapted	1977.00
RSA_2_2008	Southern Africa	South Africa	XXII	Tick	SG4	Unknown	2008.00

Figure 4. 4. Sorting of the MGFs by genotype in Morpheus.



id	Continent	Country	Genotype	Host	Serogroup	Virulence	Year
Mkuzi_1979	Southern Africa	South Africa	I	Tick	N/A	Unknown	1979.00
Warmbaths	Southern Africa	South Africa	III	Tick	N/A	Unknown	1987.00
RSA_W1_1999	Southern Africa	South Africa	IV	Warthog	N/A	Unknown	1999.00
Pretoriuskop_96_4	Southern Africa	South Africa	XX	Tick	N/A	High	1996.00
RSA_2_2004	Southern Africa	South Africa	XX	Wild boar	N/A	Unknown	2004.00
LIV_5_40	Eastern Africa	Zambia	I	Tick	SG1	Unknown	1983.00
Ken06.Bus	Eastern Africa	Kenya	IX	Sus scrofa	SG1	Unknown	2006.00
N10	Eastern Africa	Uganda	IX	Sus scrofa	SG1	Unknown	2015.00
Warthog	Southern Africa	Namibia	IV	Warthog	SG2	Unknown	1980.00
Zaire	Central Africa	Zaire	XX	Sus scrofa	SG3	Cell-culture adapted	1977.00
BA71V	Southwestern Europe	Spain	I	Sus scrofa	SG4	Cell-culture adapted	1971.00
NHV	Southwestern Europe	Portugal	I	Sus scrofa	SG4	Low	1968.00
OURT_88_3	Southwestern Europe	Portugal	I	Tick	SG4	Low	1988.00
L60	Southwestern Europe	Portugal	I	Sus scrofa	SG4	High	1960.00
E75	Southwestern Europe	Spain	I	Sus scrofa	SG4	High	1975.00
Benin_97_1	Western Africa	Benin	I	Sus scrofa	SG4	High	1997.00
57_Ca_79	Western Europe	Italy	I	Sus scrofa	SG4	Unknown	1979.00
Liv13_33	Eastern Africa	Zambia	I	Tick	SG4	Unknown	2020.00
RSA_2_2008	Southern Africa	South Africa	XXII	Tick	SG4	Unknown	2008.00
Kenya_1950	Eastern Africa	Kenya	X	Sus scrofa	SG6	High	1950.00
Ken05_Tk1	Eastern Africa	Kenya	X	Sus scrofa	SG6	Unknown	2005.00
Tengani62	Eastern Africa	Malawi	V	Sus scrofa	SG7	High	1962.00
Georgia_2007_1	Caucasus	Georgia	II	Sus scrofa	SG8	High	2007.00
Odintsovo	Caucasus	Russia	II	Sus scrofa	SG8	Unknown	2014.00
LT14_1490	Caucasus	Lithuania	II	Sus scrofa	SG8	Unknown	2014.00
Podlaskie	Central Europe	Poland	II	Wild boar	SG8	Unknown	2015.00
ASFV_Kyiv_2016_131	Caucasus	Ukraine	II	Sus scrofa	SG8	Unknown	2016.00
Pol16_20538_o9	Central Europe	Poland	II	Sus scrofa	SG8	Unknown	2016.00
Hu_2018	Central Europe	Hungary	II	Sus scrofa	SG8	Unknown	2018.00
China_CAS19-01	Asia	China	II	Sus scrofa	SG8	Unknown	2019.00
CzechRepublic_2017_1	Central Europe	Czech Republic	II	Wild boar	SG8	Unknown	2017.00
Moldova_2017_1	Caucasus	Moldova	II	Sus scrofa	SG8	Unknown	2017.00
Belgium_2018_1	Western Europe	Belgium	II	Wild boar	SG8	Unknown	2018.00
Estonia	Northern Europe	Estonia	II	Sus scrofa	SG8	Unknown	2014.00
SPEC_57	Southern Africa	South Africa	III	Tick	SG8	Unknown	1985.00
Malawi_Lil_20_1	Eastern Africa	Malawi	VIII	Tick	SG8	High	1983.00

Figure 4. 5. Sorting of the MGFs by serogroup in Morpheus.

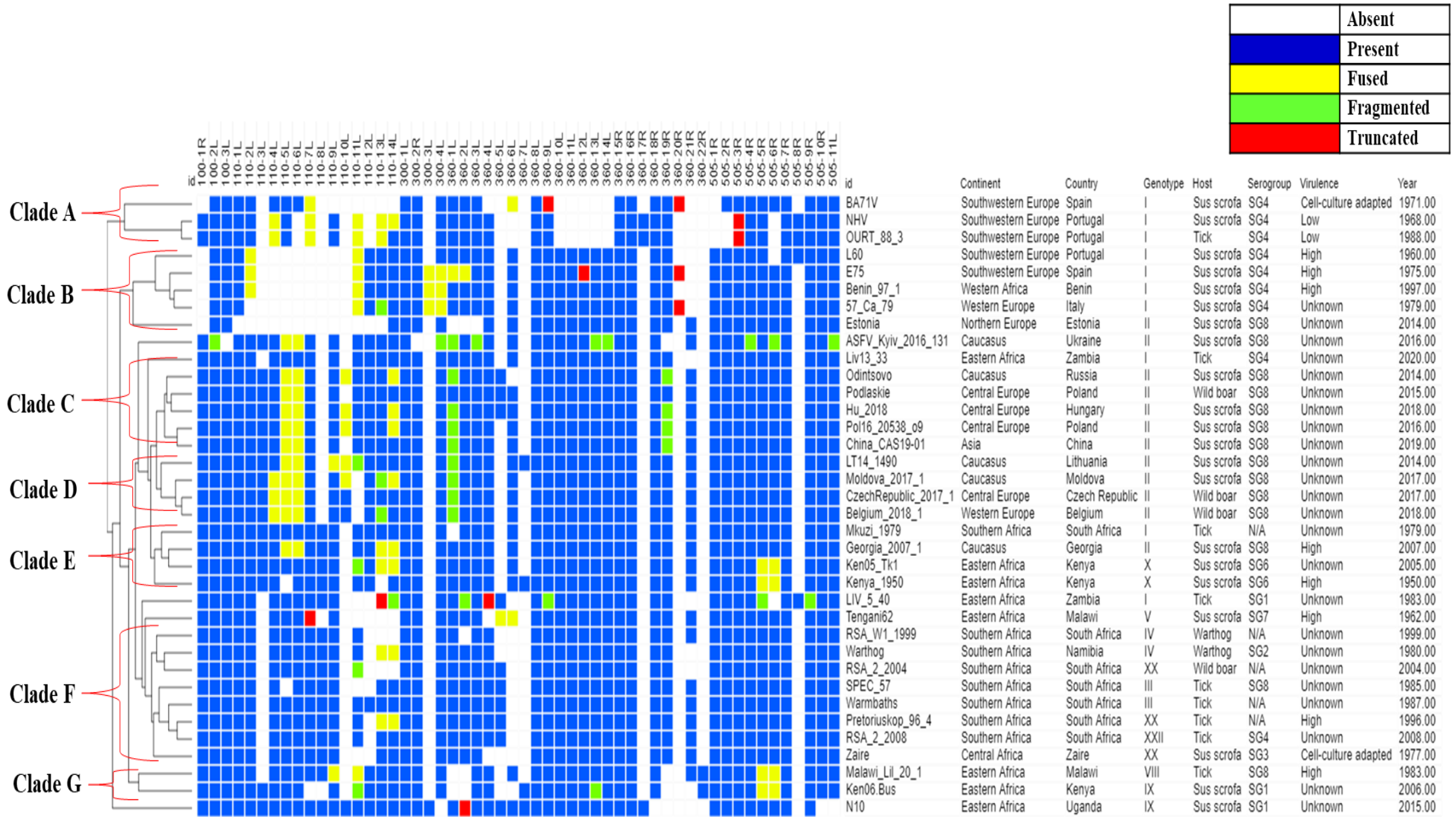


Figure 4. 6. Hierarchical clustering of the MGFs, using the Euclidean distance metric and average linkage methods in Morpheus.

Of interest was the similarity of the MGF arrangement patterns maintained between the high virulence and low virulence ASFV strains isolated in Portugal and Spain (Figure 4.7A). The L60 strain (Portugal, 1960) and E75 strain (Spain, 1975), both of high virulence, are missing MGF 100-1R, 110-3L, -10L, while the low virulence strains NHV (Portugal, 1968) and OURT_88_3 (Portugal, 1988) encode the MGF 100-1R, 110-4L, 7L, and 9L. The low virulence strains and the cell-culture adapted isolates BA71V do not encode the MGF 360-10L, -14L.

The European genotype II isolates (with the exception of a naturally attenuated Estonia isolate) encode more copies of the MGF 110 compared to European genotype I isolates. The more recent genotype II isolates, Czech Republich_2017_ Moldova_2017_1, and Belgium_2018_1, from different hosts, encode fused MGF 110-4L, 5L, and 6L ORFs (Figure 4.7B). In Africa, where diversity of ASFV strains exists, some level of similarity in the MGF arrangement patterns is observed. To some extent, the MGF arrangement patterns seem to be associated with the geographical origin of the isolates (Figure 4.7C). Most isolates from Southern Africa cluster together as do strains from Eastern Africa. However, a cluster formed by both Eastern and Southern African isolates is also observed.

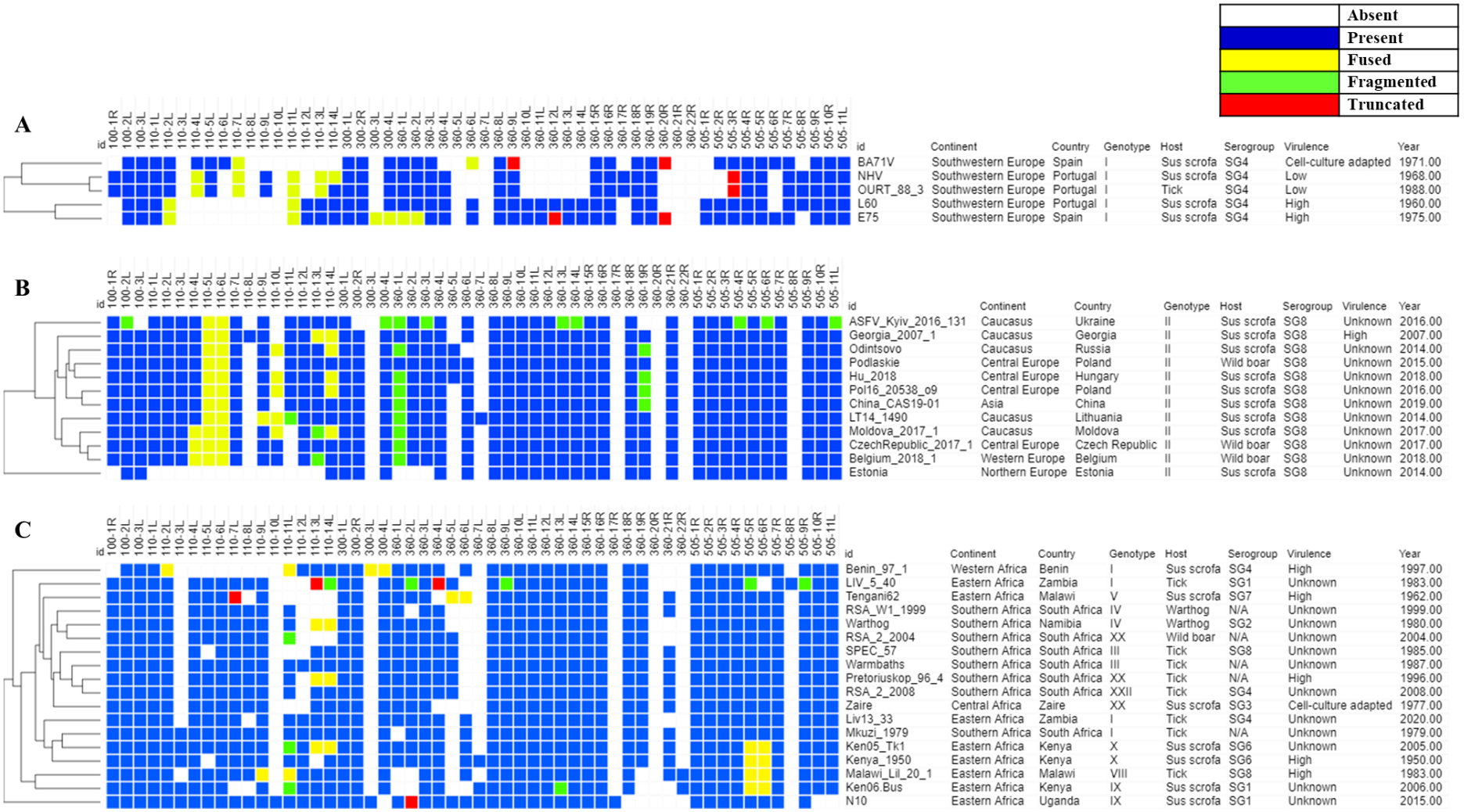


Figure 4. 7. Hierarchical clustering of the MGFs in (**A**) European genotype I strains, (**B**) European genotype II strains, and (**C**) Southern African strains, using the Euclidean distance metric and average linkage methods in Morpheus.

4.4. DISCUSSION

The ASFV genome contains three regions, namely, the left variable region (LVR), central conserved region (CCR), and the right variable region (RVR). Significant differences in ASFV isolates are found at the terminal ends (LVR and RVR) of the genome, where MGFs are located, and the difference in the number of MGFs results in variation in genome length among ASFV isolates (Chapman *et al*., 2008). The most variable MGF is the MGF 110 encoding 14 members. In contrast, the MGF 100, 300, and 505/530 are comprised of fewer members, and less variation is observed in the absence and presence of different members among ASFV isolates. The MGF 100 and 300 are considered highly conserved among ASFV isolates, while the MGF 505/530 and 360 are moderately conserved (Imbery & Upton, 2017). In this study, we have used the framework by Imbery and Upton (2017) to conduct an updated MGF analysis of six novel ASFV isolates and other publicly available fully sequenced isolates from GenBank.

Of the fourteen members of the MGF 110, the European genotype I strains encode between five (BA71V) and seven (NHV) MGF 110 members, while African isolates encode between 6 (Benin_97_1) and 13 (Mkuzi_1979, Kenya_1950, N10, R7, R8, R9, R25, and R35) members. Although there are variations in the presence and absence of the MGF 110 members between ASFV isolates, similar variations are observed in isolates that cluster together. The MGF 110 has not been shown to play a role in virulence, as the deletion of all members of the MGF 110 did not reduce the virulence of the virus in pigs (Aguero *et al*., 1990).

The MGF 110-6L contains an ER retention sequence (KDEL), which is used by many cellular and luminal ER proteins. The MGF 110-4L contains a homologous retention motif (KEDL); however, in biological systems, the homologous sequence has not been described as a functional retention motif (Netherton *et al*., 2004). The MGF 110-6L is localised to ER, while the MGF 110-4L is found in the heterogeneous vesicular structures that are concentrated to one side of the nucleus. The motif sequences (KDEL and KEDL) were deleted to determine whether the localisation of the MGF 110-4L and 6L proteins depend on the retention motif sequences found at their C-terminus (Netherton *et al*., 2004). It was demonstrated that the MGF 110-6L and MGF 110-6LΔKDEL were both retained in the ER; thus, the KDEL retention motif alone does not determine ER retention of the MGF 110-6L. The absence of the homologous sequence (KEDL) leads to the loss of the ER-Golgi

intermediate compartment and the redistribution of the protein to the ER (Netherton *et al*., 2004). Due to amino acid substitutions, the European genotype II isolates do not contain the ER retention motif sequence (KDEL) at their C-terminus, and the homologous sequence (KEDL) is not fully encoded. This observation could be a result of selection pressure exerted, such as adapting to a new host.

Of the 22 members of the MGF 360, the six novel study isolates encode between 15 and 16 members. The number varies among strains, from a minimum of 12 in OURT_88_3 to 18 in Kenya_1950, of the 22 members described. Our study isolates encode 10 of the 11 members of the MGF 505/530. The OURT_88_3 strain is the only tick isolate that is missing the MGF 360-10L, 11L, 12L, 13L, and 14L ORFs. Genotype I, tick isolates, Mkuzi_1979, Liv13_33 and LIV_5_40, are missing the MGF 360-20R, 21R, and 22R ORFs. Deletions found at the terminal ends can result in truncation or absence of some of the MGFs. The deletions observed in the LVR of the genome resulted in LIV_5_40 isolate losing the MGF 505-6R ORF. The absence of this ORF is also found in NHV, OURT_88_3, and L60 strains.

Previous studies have shown that the MGF 360 and 505/530 play a role in host range and virulence by promoting the survival of infected cells. Deletion of the MGF 360-9L, 10L, 11L, 12L, 13L, and 14L, and the MGF 505-1R and 2R ORFs, in the Pretoriuskop_96_4 isolates, induced a type I interferon response, which is not observed with the parental strain. It was further demonstrated that the deletion of the MGF 360-12L, 13L, and 14L ORFs in Pretoriuskop_96_4 isolates, reduced the viral titre in infected soft ticks by 100- to 1,000 fold (Burrage *et al*., 2004). The deletion of the MGF 360-10L, 11L, 12L, 13L, and 14L ORFs, in BA71V, NHV, and OURT_88_3, is associated with the attenuation of these isolates. The MGF 110-13L and the MGF 360-4L ORFs are truncated in the LIV_5_40 strain. The MGF 360-2L and 9L, and MGF 505-5R and 9R ORFs are fragmented in LIV_5_40 isolate, and this strain replicates effectively in swine macrophages.

When transposition of the ITR occurs, it can result in duplication of sequences contiguous to the ITR, resulting in an enlarged ITR (Fleming *et al*., 1995). The MGF 360-1L ORF in the Ken05.Tk1 isolate and 360-2L ORF in the Warthog isolate are located in the RVR, while located in the LVR in other ASFV isolates. The MGF 360-18R and -21R ORFs in the Estonia isolate, the MGF 360-19R ORF in Kenya_1950 and Ken05.Tk1 isolates, as well as the MGF 360-21R ORF in Malawi_Lil_20_1, are located in the LVR but are located in the RVR in other strains. This could mean that some of the ASFV isolates use a transposition mechanism

to adapt to a new host. Although members of the MGF 110 are of low conservation, all genotype II isolates except for Estonia contain fused MGF 110-5L and 6L ORFs, and this fusion is missing in Estonia due to a large deletion observed at the LVR. These observed fusions show that there is a selection pressure that exists in these isolates to maintain these fused ORFs.

The number of paralogs present in each MGF in a viral genome may change, and at least one paralog is dispensable for viral replication (Nishiyama, 1996). It is thought that the ASFV MGFs have originated through gene duplication and sequence divergence (de la Vega *et al*., 1990). The sequence similarity varies from 23% to 88% among MGF 360 and 46% to 74% for MGF 505/530 in ASFV isolates, revealing that gene duplication within MGFs is a continuous process (Zsak *et al*., 2001). Although the basic genome organisation of ASF viruses and poxviruses are similar, the genetic content of the terminal ends has not shown any apparent similarity. However, the placement of the MGFs at the terminal ends seems to be advantageous for the virus, and it has been demonstrated that changes at the terminal ends in poxvirus genomes result in changes in the biological features of the virus (Haller *et al*., 2014). The location of the MGFs at the genome ends may enhance the opportunity of generating successful variants. It has been shown that, under different constraints, poxviruses undergo successive cycle gene duplication and gene loss to adapt to a new host (Elde *et al*., 2012). This could be the same mechanism used by the ASF viruses since they use a similar replication mechanism as the poxviruses.

ASF virus was reported outside Africa for the first time in Portugal in 1957 and again in 1960-1993, and it was subsequently found in Spain in 1960-1995 and then again in Portugal in 1999 (Cwynar *et al*., 2019). Previous genotyping studies have shown that ESAC-WA isolates, representing genotype I, cluster together, and were introduced to Portugal in 1957 from west Africa (Bastos *et al*., 2003). We observed that Benin_97_1 representing genotype I and isolated in west Africa, clustered closely and consistently with European genotype I strains. This could be caused by a single introduction of the virus from the west-coast shipping route between Africa and Europe (Figure 3.6). However, the hierarchical clustering further grouped the LIV_5_40 close to eastern and southern African isolates, while Liv13_33 and Mkuzi_1979 clustered close to European genotype II isolates. In this study, we have demonstrated that European genotype I and II strains cluster closely together. The hierarchical clustering method further grouped ASFV isolates with similar MGF arrangement into sub-clades.

African isolates are genotypically diverse, and this could be due to the fact that these isolates have been circulating in Africa for a more extended period compared to European strains. However, some level of MGF arrangement pattern seems to be maintained among ASFV isolates from Africa, mostly in strains isolated in eastern and southern Africa where the sylvatic cycle exists. Clades E, F, and G are formed mainly by ASFVs from regions where the sylvatic cycle exits. The Benin_97_1 strain was isolated in a geographical region where the sylvatic cycle is absent and encodes less MGF 110 members compared to other African isolates. This diversification could be associated with the co-evolution of the ASFV in its natural host.

ASFV isolates (L60, E75, and Benin_97_1) that are known to be of high virulence do not encode the MGF 110-3L, 4L, 5L, 6L, 7L, 8L, 9L, and 10L, while the low virulence strains (NHV and OURT_88_3), forming part of the ESAC-WA clade, encode some of the missing ORFs. The Spanish cell-culture adapted stain BA71V lost more MGF copies compared to the African cell-culture adapted isolate Zaire. Mostly, variations that are observed are as a result of insertions or deletions.

Sorting of the MGFs by genotype and serogroup reproduce some of the clades that were obtained by the hierarchical clustering, which could suggest that genotype and serogroup play a role in the MGF arrangement patterns. Although some of the hierarchical clustering patterns are lost when the isolates are sorted by country, host, virulence, and year, this could be due to the lack of fully sequenced ASFV isolates. At least ten strains per genotype have to be fully sequenced and annotated to rule out if the country, host, virulence, and year do not contribute to the MGF arrangement patterns among ASFV isolates.

5. Conclusions

A fundamental tool of investigating virus evolution, relationship and assigning of genotypes is through molecular phylogeny. The commonly used method for phylogenetic analysis of viruses is based on sequencing a gene or a partial genome region and inferring a phylogenetic tree using those gene sequences (Meier-Kolthoff and Göker, 2017). The B646L gene encoding the p72 protein has historically been used in phylogenetic analysis of ASF viruses. The C-terminus region of this gene has been successfully used to define genotypes (Bastos *et al*., 2003), and 24 genotypes have been described. The p72 gene allows relatively fast and easy approach to discriminate between genotypes. However, it lacks the capacity for higher resolution between isolates belonging to the same genotype. Finer genotypic resolutions were achieved by using other various genes which include p30, p54, and B602L. The tetrameric amino acid repeats in the central variable region (CVR) of the protein encoded by B602L have proven useful to discriminate between closely related viruses or disaggregating viruses that are grouped in the same genotype based on B646L (Gallardo *et al*., 2014; Phologane *et al*., 2005). It was further demonstrated that the intergenic region (IGR) tetrameric repeat sequences have provided valuable information in determining the origin and the spread of closely related ASFV isolates (de Villiers *et al*., 2010; Gallardo *et al*., 2014; Gallardo *et al*., 2009; Ge *et al*., 2018).

The benefit of using additional genes in phylogenetic analysis is to add to limited repository of genes that are good phylogenetic markers present across all genomes, which can be readily used for finer discrimination among closely related isolates and serve as the quick estimation of congruency between closely related isolate. The MGFs are formed through a variety of evolutionary events that generate paralogs. Phylogenetic analysis based on the MGFs has been considered as one of the approaches that can be used to estimate the existence of ancient intra-genomic paralongs (Ambreen *et al*., 2014). It was previously shown that a phylogenetic tree constructed by the concatenation of p30, p54, p72, and the B602L gene sequences grouped the ASFV isolates with similar MGF characteristics together (Imbery and Upton, 2017). Thus, MGF analysis should be explored further in terms of virus genotyping and potentially included in future phylogenetic analysis. The used of the MGFs in phylogenetic analysis and a broader investigation of the genetic characteristic may make it possible to understand the effects of concerted evolution.

6. Future directions

Recombination is one of the significant driving forces in evolution in double-stranded DNA viruses, which is characterised by the ability to combine favourable mutations into one genome. Recombination events are usually observed between species with high sequence similarity, and they have been studies in DNA viruses such as the herpesviruses and adenoviruses. Recombination events have also occurred in other DNA viruses which include poxviruses, geminiviruses, and anelloviruses (Bugert, and Darai, 2000; Monci *et al*., 2002). Recombination events in ASFV genomes have not been well annotated. However, it was previously demonstrated using sixteen ASFV isolates that the p54 gene sequences, revealed recombination events in viruses isolated in Italy and one from South Africa. A recombination study showed the absence of recombination events in the p72 and the p30 genes, neither non-synonymous bias nor a small number of codons under positive selection were detected, which may indicate the natural virus evolution (Michaud *et al*., 2013).

It has been postulated that the genetic determinants of ASFV variability are potential hot-spots for recombination. Thus, 70 nucleotide sequences of the ASFV serotype-specific locus (EP153R and EP402R), were assessed in order to identify recombination breakpoints and reveal driving forces of virus evolution (Nefedeva *et al*., 2020). They assessed ratio of divergence at non-synonymous and synonymous sites (dN/dS), measured the direct effect of natural selection by fixed polymorphism analysis (pS/pN), detection of amino acids sites under positive selection by fixed effect likelihood analysis (FEL), and they computed the phylogenetic networks and performed pairwise genetic identity calculations. Recombination occurred in both protein and mostly among ASFV isolates from East Africa, where multiple virus transmission cycles are found (Nefedeva *et al*., 2020). However, more genomes sequences are needed in studies looking at identifying recombination events, because working with limited genome sequences will be of little value. To date, there are 24 genotypes based on the partial sequencing of the p72 gene. However, the vast majority of publicly available complete genome sequences represent genotypes I, II and IX, thus, limiting the understanding of geographical origin and outbreak patterns. We recommend that at least ten strains per genotype have to be fully sequenced and analysed to have a better understanding of the ASFV genome diversity in Africa where multiple genotypes exists, as compared to Asia where genotype II isolates are circulating.

7. References

Abergel, C., Legendre, M., Claverie, J.M., 2015. The rapidly expanding universe of giant viruses: Mimivirus, Pandoravirus, Pithovirus and Mollivirus. FEMS Microbiol Rev 39, 779-796.

Abrams, C.C., Chapman, D.A., Silk, R., Liverani, E., Dixon, L.K., 2008. Domains involved in calcineurin phosphatase inhibition and nuclear localisation in the African swine fever virus A238L protein. Virology 374, 477-486.

Achenbach, J.E., Gallardo, C., Nieto-Pelegrin, E., Rivera-Arroyo, B., Degefa-Negi, T., Arias, M., Jenberie, S., Mulisa, D.D., Gizaw, D., Gelaye, E., Chibssa, T.R., Belaye, A., Loitsch, A., Forsa, M., Yami, M., Diallo, A., Soler, A., Lamien, C.E., Sanchez-Vizcaino, J.M., 2017. Identification of a New Genotype of African Swine Fever Virus in Domestic Pigs from Ethiopia. Transbound Emerg Dis 64, 1393-1404.

Afonso, C.L., Piccone, M.E., Zaffuto, K.M., Neilan, J., Kutish, G.F., Lu, Z., Balinsky, C.A., Gibb, T.R., Bean, T.J., Zsak, L., Rock, D.L., 2004. African swine fever virus multigene family 360 and 530 genes affect host interferon response. J Virol 78, 1858-1864.

Afonso, C.L., Tulman, E.R., Delhon, G., Lu, Z., Viljoen, G.J., Wallace, D.B., Kutish, G.F., Rock, D.L., 2006. Genome of crocodilepox virus. J Virol 80, 4978-4991.

Aguero, M., Blasco, R., Wilkinson, P., Vinuela, E., 1990. Analysis of naturally occurring deletion variants of African swine fever virus: multigene family 110 is not essential for infectivity or virulence in pigs. Virology 176, 195-204.

Alcami, A., Angulo, A., Lopez-Otin, C., Munoz, M., Freije, J.M., Carrascosa, A.L., Vinuela, E., 1992. Amino acid sequence and structural properties of protein p12, an African swine fever virus attachment protein. J Virol 66, 3860-3868.

Alejo, A., Matamoros, T., Guerra, M., Andres, G., 2018. A Proteomic Atlas of the African Swine Fever Virus Particle. J Virol 92.

Almendral, J.M., Almazan, F., Blasco, R., Vinuela, E., 1990. Multigene families in African swine fever virus: family 110. J Virol 64, 2064-2072.

Alonso, C., Borca, M., Dixon, L., Revilla, Y., Rodriguez, F., Escribano, J.M., Ictv Report, C., 2018. ICTV Virus Taxonomy Profile: Asfarviridae. J Gen Virol 99, 613-614.

Ambreen, S., Khalil, F., Abbasi, A.A., 2014. Integrating large-scale phylogenetic datasets to dissect the ancient evolutionary history of vertebrate genome. Mol Phylogenet Evol 78, 1–13.

Andreani, J., Khalil, J.Y.B., Sevvana, M., Benamar, S., Di Pinto, F., Bitam, I., Colson, P., Klose, T., Rossmann, M.G., Raoult, D., La Scola, B., 2017. Pacmanvirus, a New Giant Icosahedral Virus at the Crossroads between Asfarviridae and Faustoviruses. J Virol 91.

Andres, G., 2017. African Swine Fever Virus Gets Undressed: New Insights on the Entry Pathway. J Virol 91.

Andres, G., Alejo, A., Salas, J., Salas, M.L., 2002. African swine fever virus polyproteins pp220 and pp62 assemble into the core shell. J Virol 76, 12473-12482.

Andres, G., Garcia-Escudero, R., Simon-Mateo, C., Vinuela, E., 1998. African swine fever virus is enveloped by a two-membraned collapsed cisterna derived from the endoplasmic reticulum. J Virol 72, 8988-9001.

Andres, G., Simon-Mateo, C., Vinuela, E., 1997. Assembly of African swine fever virus: role of polyprotein pp220. J Virol 71, 2331-2341.

Andres, G., Charro, D., Matamoros, T., Dillard, R. S., Abrescia, N. G. A., 2020. The cryo-EM structure of African swine fever virus unravels a unique architecture comprising two icosahedral protein capsids and two lipoprotein membranes. J Biol Chem 29, 1–12.

Argilaguet, J.M., Perez-Martin, E., Nofrarias, M., Gallardo, C., Accensi, F., Lacasta, A., Mora, M., Ballester, M., Galindo-Cardiel, I., Lopez-Soria, S., Escribano, J.M., Reche, P.A., Rodriguez, F., 2012. DNA vaccination partially protects against African swine fever virus lethal challenge in the absence of antibodies. PLoS One 7, e40942.

Bacciu, D., Deligios, M., Sanna, G., Madrau, M.P., Sanna, M.L., Dei Giudici, S., Oggiano, A., 2016. Genomic analysis of Sardinian 26544/OG10 isolate of African swine fever virus. Virology Reports 6, 81-89.

Bajrai, L.H., Benamar, S., Azhar, E.I., Robert, C., Levasseur, A., Raoult, D., La Scola, B., 2016. Kaumoebavirus, a new virus that clusters with faustoviruses and Asfarviridae. Viruses 278, 1-10.

Banjara, S., Shimmon, G.L., Dixon, L.K., Netherton, C.L., Hinds, M.G., Kvansakul, M., 2019. Crystal Structure of African Swine Fever Virus A179L with the Autophagy Regulator Beclin. Viruses 11.

Barasona, J.A., Gallardo, C., Cadenas-Fernandez, E., Jurado, C., Rivera, B., Rodriguez-Bertos, A., Arias, M., Sanchez-Vizcaino, J.M., 2019. First Oral Vaccination of Eurasian Wild Boar Against African Swine Fever Virus Genotype II. Front Vet Sci 6, 137.

Basta, S., Gerber, H., Schaub, A., Summerfield, A., McCullough, K.C., 2010. Cellular processes essential for African swine fever virus to infect and replicate in primary macrophages. Vet Microbiol 140, 9-17.

Bastos, A.D., Arnot, L.F., Jacquier, M.D., Maree, S., 2009. A host species-informative internal control for molecular assessment of African swine fever virus infection rates in the African sylvatic cycle Ornithodoros vector. Med Vet Entomol 23, 399-409.

Bastos, A.D., Penrith, M.L., Cruciere, C., Edrich, J.L., Hutchings, G., Roger, F., Couacy-Hymann, E., G, R.T., 2003. Genotyping field strains of African swine fever virus by partial p72 gene characterisation. Arch Virol 148, 693-706.

Bastos, A.D., Penrith, M.L., Macome, F., Pinto, F., Thomson, G.R., 2004. Co-circulation of two genetically distinct viruses in an outbreak of African swine fever in Mozambique: no evidence for individual co-infection. Vet Microbiol 103, 169-182.

Bellini, S., Rutili, D., Guberti, V., 2016. Preventive measures aimed at minimizing the risk of African swine fever virus spread in pig farming systems. Acta Vet Scand 58, 82.

Bishop, R.P., Fleischauer, C., de Villiers, E.P., Okoth, E.A., Arias, M., Gallardo, C., Upton, C., 2015. Comparative analysis of the complete genome sequences of Kenyan African swine fever virus isolates within p72 genotypes IX and X. Virus Genes 50, 303-309.

Blasco, R., Aguero, M., Almendral, J.M., Vinuela, E., 1989a. Variable and constant regions in African swine fever virus DNA. Virology 168, 330-338.

Blasco, R., de la Vega, I., Almazan, F., Aguero, M., Vinuela, E., 1989b. Genetic variation of African swine fever virus: variable regions near the ends of the viral DNA. Virology 173, 251-257.

Blome, S., Gabriel, C., Beer, M., 2014. Modern adjuvants do not enhance the efficacy of an inactivated African swine fever virus vaccine preparation. Vaccine 32, 3879-3882.

Blome, S., Gabriel, C., Dietze, K., Breithaupt, A., Beer, M., 2012. High virulence of African swine fever virus caucasus isolate in European wild boars of all ages. Emerg Infect Dis 18, 708.

Boinas, F.S., Hutchings, G.H., Dixon, L.K., Wilkinson, P.J., 2004. Characterization of pathogenic and non-pathogenic African swine fever virus isolates from Ornithodoros erraticus inhabiting pig premises in Portugal. J Gen Virol 85, 2177-2187.

Borca, M., Ramirez-Medina, E., Silva, E., Vuono, E., Rai, A., Pruitt, S., Holinka, L.G., Velazquez Salinas, L., Zhu, J., Gladue, D.P., 2020. Development of a highly effective African swine fever virus vaccine by deletion of the I177L gene results in sterile immunity against the current epidemic Eurasia strain.pdf. American Society for Microbiology. 94, 28.

Borca, M.V., Carrillo, C., Zsak, L., Laegreid, W.W., Kutish, G.F., Neilan, J.G., Burrage, T.G., Rock, D.L., 1998. Deletion of a CD2-like gene, 8-DR, from African swine fever virus affects viral infection in domestic swine. J Virol 72, 2881-2889.

Borca, M.V., Kutish, G.F., Afonso, C.L., Irusta, P., Carrillo, C., Brun, A., Sussman, M., Rock, D.L., 1994. An African swine fever virus gene with similarity to the T-lymphocyte surface antigen CD2 mediates hemadsorption. Virology 199, 463-468.

Boshoff, C.I., Bastos, A.D., Gerber, L.J., Vosloo, W., 2007. Genetic characterisation of African swine fever viruses from outbreaks in southern Africa (1973-1999). Vet Microbiol 121, 45-55.

Brown, V.R., Bevins, S.N., 2018. A Review of African Swine Fever and the Potential for Introduction into the United States and the Possibility of Subsequent Establishment in Feral Swine and Native Ticks. Front Vet Sci 5, 11.

Brun, A., Rivas, C., Esteban, M., Escribano, J.M., Alonso, C., 1996. African swine fever virus gene A179L, a viral homologue of bcl-2, protects cells from programmed cell death. Virology 225, 227-230.

Bugert, J. J., Darai, G., 2000. Poxvirus homologues of cellular genes. *Virus Genes* **21**, 111-133.

Burmakina, G., Malogolovkin, A., Tulman, E.R., Xu, W., Delhon, G., Kolbasov, D., Rock, D.L., 2019. Identification of T-cell epitopes in African swine fever virus CD2v and C-type lectin proteins. J Gen Virol 100, 259-265.

Burrage, T.G., Lu, Z., Neilan, J.G., Rock, D.L., Zsak, L., 2004. African swine fever virus multigene family 360 genes affect virus replication and generalization of infection in Ornithodoros porcinus ticks. J Virol 78, 2445-2453.

Carrascosa, A.L., Bustos, M.J., de Leon, P., 2011. Methods for growing and titrating African swine fever virus: field and laboratory samples. Curr Protoc Cell Biol Chapter 26, Unit 26 14.

Carrascosa, A.L., del Val, M., Santaren, J.F., Vinuela, E., 1985. Purification and properties of African swine fever virus. J Virol 54, 337-344.

Carrascosa, J.L., Carazo, J.M., Carrascosa, A.L., Garcia, N., Santisteban, A., Vinuela, E., 1984. General morphology and capsid fine structure of African swine fever virus particles. Virology 132, 160-172.

Chapman, D.A., Darby, A.C., Da Silva, M., Upton, C., Radford, A.D., Dixon, L.K., 2011. Genomic analysis of highly virulent Georgia 2007/1 isolate of African swine fever virus. Emerg Infect Dis 17, 599-605.

Chapman, D.A., Tcherepanov, V., Upton, C., Dixon, L.K., 2008. Comparison of the genome sequences of non-pathogenic and pathogenic African swine fever virus isolates. J Gen Virol 89, 397-408.

Cobbold, C., Brookes, S.M., Wileman, T., 2000. Biochemical requirements of virus wrapping by the endoplasmic reticulum: involvement of ATP and endoplasmic reticulum calcium store during envelopment of African swine fever virus. J Virol 74, 2151-2160.

Cobbold, C., Wileman, T., 1998. The major structural protein of African swine fever virus, p73, is packaged into large structures, indicative of viral capsid or matrix precursors, on the endoplasmic reticulum. J Virol 72, 5215-5223.

Cobbold, C., Windsor, M., Wileman, T., 2001. A virally encoded chaperone specialized for folding of the major capsid protein of African swine fever virus. J Virol 75, 7221-7229.

Coggins, L., 1974. African swine fever virus. Pathogenesis. Prog Med Virol 18, 48-63.

Costard, S., Wieland, B., de Glanville, W., Jori, F., Rowlands, R., Vosloo, W., Roger, F., Pfeiffer, D.U., Dixon, L.K., 2009. African swine fever: how can global spread be prevented? Philos Trans R Soc Lond B Biol Sci 364, 2683-2696.

Cwynar, P., Stojkov, J., Wlazlak, K., 2019. African Swine Fever Status in Europe. Viruses 11.

de Carvalho Ferreira, H.C., Weesendorp, E., Elbers, A.R., Bouma, A., Quak, S., Stegeman, J.A., Loeffen, W.L., 2012. African swine fever virus excretion patterns in persistently infected animals: a quantitative approach. Vet Microbiol 160, 327-340.

De Cock, G., RObinson, E.M., Keppel, J.J.G., 1940. Swine fever in South Africa. Onderstepoort Journal of Veterinary Science and Animal Industry 14, 31-39.

de la Vega, I., Vinuela, E., Blasco, R., 1990. Genetic variation and multigene families in African swine fever virus. Virology 179, 234-246.

de Leon, P., Bustos, M.J., Carrascosa, A.L., 2013. Laboratory methods to study African swine fever virus. Virus Res 173, 168-179.

de Villiers, E.P., Gallardo, C., Arias, M., da Silva, M., Upton, C., Martin, R., Bishop, R.P., 2010. Phylogenomic analysis of 11 complete African swine fever virus genome sequences. Virology 400, 128-136.

Detray, D.E., 1957. Persistence of viremia and immunity in African swine fever. Am J Vet Res 18, 811-816.

Dixon, L.K., 1988. Molecular cloning and restriction enzyme mapping of an African swine fever virus isolate from Malawi. J Gen Virol 69 (Pt 7), 1683-1694.

Dixon, L.K., Abrams, C.C., Bowick, G., Goatley, L.C., Kay-Jackson, P.C., Chapman, D., Liverani, E., Nix, R., Silk, R., Zhang, F., 2004. African swine fever virus proteins involved in evading host defence systems. Vet Immunol Immunopathol 100, 117-134.

Dixon, L.K., Chapman, D.A., Netherton, C.L., Upton, C., 2013. African swine fever virus replication and genomics. Virus Res 173, 3-14.

Dixon, L.K., Stahl, K., Jori, F., Vial, L., Pfeiffer, D.U., 2019a. African Swine Fever Epidemiology and Control. Annu Rev Anim Biosci.

Dixon, L.K., Sun, H., Roberts, H., 2019b. African swine fever. Antiviral Res 165, 34-41.

Dixon, L.K., Wilkinson, P.J., 1988. Genetic diversity of African swine fever virus isolates from soft ticks (Ornithodoros moubata) inhabiting warthog burrows in Zambia. J Gen Virol 69 (Pt 12), 2981-2993.

Domenech, J., Lubroth, J., Eddi, C., Martin, V., Roger, F., 2006. Regional and international approaches on prevention and control of animal transboundary and emerging diseases. Ann N Y Acad Sci 1081, 90-107.

Elde, N.C., Child, S.J., Eickbush, M.T., Kitzman, J.O., Rogers, K.S., Shendure, J., Geballe, A.P., Malik, H.S., 2012. Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses. Cell 150, 831-841.

Epifano, C., Krijnse-Locker, J., Salas, M.L., Rodriguez, J.M., Salas, J., 2006a. The African swine fever virus nonstructural protein pB602L is required for formation of the icosahedral capsid of the virus particle. J Virol 80, 12260-12270.

Epifano, C., Krijnse-Locker, J., Salas, M.L., Salas, J., Rodriguez, J.M., 2006b. Generation of filamentous instead of icosahedral particles by repression of African swine fever virus structural protein pB438L. J Virol 80, 11456-11466.

Fleming, S.B., Lyttle, D.J., Sullivan, J.T., Mercer, A.A., Robinson, A.J., 1995. Genomic analysis od a transposition-deletion variantof orf virus reveals a 3.3 kbp region of non-essential DNA. Journal of General Virology 76, 2969-2978.

Forman, A.J., Wardley, R.C., Wilkinson, P.J., 1982. The immunological response of pigs and guinea pigs to antigens of African swine fever virus. Arch Virol 74, 91-100.

Forth, J.H., Forth, L.F., King, J., Groza, O., Hubner, A., Olesen, A.S., Hoper, D., Dixon, L.K., Netherton, C.L., Rasmussen, T.B., Blome, S., Pohlmann, A., Beer, M., 2019. A Deep-Sequencing Workflow for the Fast and Efficient Generation of High-Quality African Swine Fever Virus Whole-Genome Sequences. Viruses 11.

Freitas, F.B., Frouco, G., Martins, C., Ferreira, F., 2019. The QP509L and Q706L superfamily II RNA helicases of African swine fever virus are required for viral replication, having non-redundant activities. Emerg Microbes Infect 8, 291-302.

Gabriel, C., Blome, S., Malogolovkin, A., Parilov, S., Kolbasov, D., Teifke, J.P., Beer, M., 2011. Characterization of African swine fever virus Caucasus isolate in European wild boars. Emerg Infect Dis 17, 2342-2345.

Galindo-Cardiel, I., Ballester, M., Solanes, D., Nofrarias, M., Lopez-Soria, S., Argilaguet, J.M., Lacasta, A., Accensi, F., Rodriguez, F., Segales, J., 2013. Standardization of pathological investigations in the framework of experimental ASFV infections. Virus Res 173, 180-190.

Galindo, I., Alonso, C., 2017. African Swine Fever Virus: A Review. Viruses 9.

Gallardo, C., Fernandez-Pinero, J., Pelayo, V., Gazaev, I., Markowska-Daniel, I., Pridotkas, G., Nieto, R., Fernandez-Pacheco, P., Bokhan, S., Nevolko, O., Drozhzhe, Z., Perez, C., Soler, A., Kolvasov, D., Arias, M., 2014. Genetic variation among African swine fever genotype II viruses, eastern and central Europe. Emerg Infect Dis 20, 1544-1547.

Gallardo, C., Mwaengo, D.M., Macharia, J.M., Arias, M., Taracha, E.A., Soler, A., Okoth, E., Martin, E., Kasiti, J., Bishop, R.P., 2009. Enhanced discrimination of African swine fever virus isolates through nucleotide sequencing of the p54, p72, and pB602L (CVR) genes. Virus Genes 38, 85-95.

Gallardo, C., Soler, A., Nieto, R., Cano, C., Pelayo, V., Sanchez, M.A., Pridotkas, G., Fernandez-Pinero, J., Briones, V., Arias, M., 2017. Experimental Infection of Domestic Pigs with African Swine Fever Virus Lithuania 2014 Genotype II Field Isolate. Transbound Emerg Dis 64, 300-304.

Gallardo, C., Soler, A., Nieto, R., Sanchez, M.A., Martins, C., Pelayo, V., Carrascosa, A., Revilla, Y., Simon, A., Briones, V., Sanchez-Vizcaino, J.M., Arias, M., 2015. Experimental Transmission of African Swine Fever (ASF) Low Virulent Isolate NH/P68 by Surviving Pigs. Transbound Emerg Dis 62, 612-622.

Gallardo, C., Fernández-Pinero, J., Arias, M., 2019. African swine fever (ASF) diagnosis, an essential tool in the epidemiological investigation. Virus Research 271, 1-10.

Gallot-Lavallee, L., Blanc, G., 2017. A Glimpse of Nucleo-Cytoplasmic Large DNA Virus Biodiversity through the Eukaryotic Genomics Window. Viruses 9.

Garcia-Escudero, R., Andres, G., Almazan, F., Vinuela, E., 1998. Inducible gene expression from African swine fever virus recombinants: analysis of the major capsid protein p72. J Virol 72, 3185-3195.

Gaudreault, N.N., Richt, J.A., 2019. Subunit Vaccine Approaches for African Swine Fever Virus. Vaccines (Basel) 7.

Ge, S., Li, J., Fan, X., Liu, F., Li, L., Wang, Q., Ren, W., Bao, J., Liu, C., Wang, H., Liu, Y., Zhang, Y., Xu, T., Wu, X., Wang, Z., 2018. Molecular Characterization of African Swine Fever Virus, China, 2018. Emerg Infect Dis 24, 2131-2133.

Giammarioli, M., Gallardo, C., Oggiano, A., Iscaro, C., Nieto, R., Pellegrini, C., Dei Giudici, S., Arias, M., De Mia, G.M., 2011. Genetic characterisation of African swine fever viruses from recent and historical outbreaks in Sardinia (1978-2009). Virus Genes 42, 377-387.

Gil, S., Sepulveda, N., Albina, E., Leitao, A., Martins, C., 2008. The low-virulent African swine fever virus (ASFV/NH/P68) induces enhanced expression and production of relevant regulatory cytokines (IFNalpha, TNFalpha and IL12p40) on porcine macrophages in comparison to the highly virulent ASFV/L60. Arch Virol 153, 1845-1854.

Goatley, L.C., Dixon, L.K., 2011. Processing and localization of the african swine fever virus CD2v transmembrane protein. J Virol 85, 3294-3305.

Golding, J.P., Goatley, L., Goodbourn, S., Dixon, L.K., Taylor, G., Netherton, C.L., 2016. Sensitivity of African swine fever virus to type I interferon is linked to genes within multigene families 360 and 505. Virology 493, 154-161.

Gomez-Puertas, P., Rodriguez, F., Oviedo, J.M., Brun, A., Alonso, C., Escribano, J.M., 1998. The African swine fever virus proteins p54 and p30 are involved in two distinct steps of virus attachment and both contribute to the antibody-mediated protective immune response. Virology 243, 461-471.

Gomez-Puertas, P., Rodriguez, F., Oviedo, J.M., Ramiro-Ibanez, F., Ruiz-Gonzalvo, F., Alonso, C., Escribano, J.M., 1996. Neutralizing antibodies to different proteins of African swine fever virus inhibit both virus attachment and internalization. J Virol 70, 5689-5694.

Gontcharov, A.A., Marin, B., Melkonian, M., 2004. Are combined analyses better than single gene phylogenies? A case study using SSU rDNA and rbcL sequence comparisons in the Zygnematophyceae (Streptophyta). Mol Biol Evol 21, 612-624.

Gonzalez, A., Calvo, V., Almazan, F., Almendral, J.M., Ramirez, J.C., de la Vega, I., Blasco, R., Vinuela, E., 1990. Multigene families in African swine fever virus: family 360. J Virol 64, 2073-2081.

Grau, R. R., Schroeder, M. E., Mulhern, E. L., McIntosh, M. T., Bounpheng, M. A., 2015. Detection of African swine fever, classical swine fever, and foot-and-mouth disease viruses in swine oral fluids by multiplex reverse transcription real-time polymerase chain reaction. Journal of Veterinary Diagnostic Investigation 27, 140-149.

Guinat, C., Gogin, A., Blome, S., Keil, G., Pollin, R., Pfeiffer, D.U., Dixon, L., 2016. Transmission routes of African swine fever virus to domestic pigs: current knowledge and future research directions. Vet Rec 178, 262-267.

Haller, S.L., Peng, C., McFadden, G., Rothenburg, S., 2014. Poxviruses and the evolution of host range and virulence. Infect Genet Evol 21, 15-40.

Hendrickson, R.C., Wang, C., Hatcher, E.L., Lefkowitz, E.J., 2010. Orthopoxvirus genome evolution: the role of gene loss. Viruses 2, 1933-1967.

Hernaez, B., Alonso, C., 2010. Dynamin- and clathrin-dependent endocytosis in African swine fever virus entry. J Virol 84, 2100-2109.

Hernaez, B., Guerra, M., Salas, M.L., Andres, G., 2016. African Swine Fever Virus Undergoes Outer Envelope Disruption, Capsid Disassembly and Inner Envelope Fusion before Core Release from Multivesicular Endosomes. PLoS Pathog 12, e1005595.

Hurtado, C., Granja, A.G., Bustos, M.J., Nogal, M.L., Gonzalez de Buitrago, G., de Yebenes, V.G., Salas, M.L., Revilla, Y., Carrascosa, A.L., 2004. The C-type lectin homologue gene (EP153R) of African swine fever virus inhibits apoptosis both in virus infection and in heterologous expression. Virology 326, 160-170.

Hutchings, G.H., Ferris, N.P., 2006. Indirect sandwich ELISA for antigen detection of African swine fever virus: comparison of polyclonal and monoclonal antibodies. J Virol Methods 131, 213-217.

Imbery, J., Upton, C., 2017. Organization of the multigene famalies of the African swine fever virs. Fine Focus 154, 156-171.

Jancovich, J.K., Chapman, D., Hansen, D.T., Robida, M.D., Loskutov, A., Craciunescu, F., Borovkov, A., Kibler, K., Goatley, L., King, K., Netherton, C.L., Taylor, G., Jacobs, B., Sykes, K., Dixon, L.K., 2018. Immunization of Pigs by DNA Prime and Recombinant Vaccinia Virus Boost To Identify and Rank African Swine Fever Virus Immunogenic and Protective Proteins. J Virol 92.

Jenson, J.S., Childerstone, A., Takamatsu, H., Dixon, L.K., Parkhouse, R.M., 2000. The cellular immune recognition of proteins expressed by an African swine fever virus random genomic library. J Immunol Methods 242, 33-42.

Jia, N., Ou, Y., Pejsak, Z., Zhang, Y., Zhang, J., 2017. Roles of African Swine Fever Virus Structural Proteins in Viral Infection. J Vet Res 61, 135-143.

Jo, Y. S., Gortázar C., 2020. African swine fever in wild boar, South Korea, 2019. Transbound Emerg Dis 67, 1776-1780.

Jori, F., Bastos, A.D., 2009. Role of wild suids in the epidemiology of African swine fever. Ecohealth 6, 296-310.

Jori, F., Vial, L., Penrith, M.L., Perez-Sanchez, R., Etter, E., Albina, E., Michaud, V., Roger, F., 2013. Review of the sylvatic cycle of African swine fever in sub-Saharan Africa and the Indian ocean. Virus Res 173, 212-227.

Kim, H. J., Cho, K. H., Lee, S. K., Kim, D. A., Nah, J. J., Kim, H. J., Kim, H. J., Hwang, J. Y., Sohn, H. J., Choi, J. G., Kang, H. E., Kim, H. E., 2020a. Outbreak of African swine fever in South Korea, 2019. Transbound Emerg Dis 67, 473-475.

King, D.P., Reid, S.M., Hutchings, G.H., Grierson, S.S., Wilkinson, P.J., Dixon, L.K., Bastos, A.D., Drew, T.W., 2003. Development of a TaqMan PCR assay with internal amplification control for the detection of African swine fever virus. J Virol Methods 107, 53-61.

King, K., Chapman, D., Argilaguet, J.M., Fishbourne, E., Hutet, E., Cariolet, R., Hutchings, G., Oura, C.A., Netherton, C.L., Moffat, K., Taylor, G., Le Potier, M.F., Dixon, L.K., Takamatsu, H.H., 2011. Protection of European domestic pigs from virulent African isolates of African swine fever virus by experimental immunisation. Vaccine 29, 4593-4600.

Kleiboeker, S.B., Burrage, T.G., Scoles, G.A., Fish, D., Rock, D.L., 1998. African swine fever virus infection in the argasid host, Ornithodoros porcinus porcinus. J Virol 72, 1711-1724.

Kollnberger, S.D., Gutierrez-Castaneda, B., Foster-Cuevas, M., Corteyn, A., Parkhouse, R.M., 2002. Identification of the principal serological immunodeterminants of African swine fever virus by screening a virus cDNA library with antibody. J Gen Virol 83, 1331-1342.

Koonin, E.V., Yutin, N., 2018. Multiple evolutionary origins of giant viruses. F1000Research 7, 1-13.

Krug, P.W., Holinka, L.G., O'Donnell, V., Reese, B., Sanford, B., Fernandez-Sainz, I., Gladue, D.P., Arzt, J., Rodriguez, L., Risatti, G.R., Borca, M.V., 2015. The progressive adaptation of a georgian isolate of African swine fever virus to vero cells leads to a gradual attenuation of virulence in swine corresponding to major modifications of the viral genome. J Virol 89, 2324-2332.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol 35, 1547-1549.

Lacasta, A., Ballester, M., Monteagudo, P.L., Rodriguez, J.M., Salas, M.L., Accensi, F., Pina-Pedrero, S., Bensaid, A., Argilaguet, J., Lopez-Soria, S., Hutet, E., Le Potier, M.F.,

Rodriguez, F., 2014. Expression library immunization can confer protection against lethal challenge with African swine fever virus. J Virol 88, 13322-13332.

Laidlaw, S.M., Skinner, M.A., 2004. Comparison of the genome sequence of FP9, an attenuated, tissue culture-adapted European strain of Fowlpox virus, with those of virulent American and European viruses. J Gen Virol 85, 305-322.

Leitao, A., Cartaxeiro, C., Coelho, R., Cruz, B., Parkhouse, R.M., Portugal, F., Vigario, J.D., Martins, C.L., 2001. The non-haemadsorbing African swine fever virus isolate ASFV/NH/P68 provides a model for defining the protective anti-virus immune response. J Gen Virol 82, 513-523.

Letunic, I., Bork, P., 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47, W256-W259.

Lewis, T., Zsak, L., Burrage, T.G., Lu, Z., Kutish, G.F., Neilan, J.G., Rock, D.L., 2000. An African swine fever virus ERV1-ALR homologue, 9GL, affects virion maturation and viral growth in macrophages and viral virulence in swine. J Virol 74, 1275-1285.

Lopera-Madrid, J., Osorio, J.E., He, Y., Xiang, Z., Adams, L.G., Laughlin, R.C., Mwangi, W., Subramanya, S., Neilan, J., Brake, D., Burrage, T.G., Brown, W.C., Clavijo, A., Bounpheng, M.A., 2017. Safety and immunogenicity of mammalian cell derived and Modified Vaccinia Ankara vectored African swine fever subunit antigens in swine. Vet Immunol Immunopathol 185, 20-33.

Lubisi, B.A., Bastos, A.D., Dwarka, R.M., Vosloo, W., 2005. Molecular epidemiology of African swine fever in East Africa. Arch Virol 150, 2439-2452.

Lubisi, B.A., Bastos, A.D., Dwarka, R.M., Vosloo, W., 2007. Intra-genotypic resolution of African swine fever viruses from an East African domestic pig cycle: a combined p72-CVR approach. Virus Genes 35, 729-735.

Lubisi, B.A., Dwarka, R.M., Meenowa, D., Jaumally, R., 2009. An investigation into the first outbreak of African swine fever in the Republic of Mauritius. Transbound Emerg Dis 56, 178-188.

Malmquist, W.A., Hay, D., 1960. Hemadsorption and cytopathic effect produced by African Swine Fever virus in swine bone marrow and buffy coat cultures. Am J Vet Res 21, 104-108.

Malogolovkin, A., Burmakina, G., Titov, I., Sereda, A., Gogin, A., Baryshnikova, E., Kolbasov, D., 2015a. Comparative analysis of African swine fever virus genotypes and serogroups. Emerg Infect Dis 21, 312-315.

Malogolovkin, A., Burmakina, G., Tulman, E.R., Delhon, G., Diel, D.G., Salnikov, N., Kutish, G.F., Kolbasov, D., Rock, D.L., 2015b. African swine fever virus CD2v and C-type lectin gene loci mediate serological specificity. J Gen Virol 96, 866-873.

Manual, J., 2008. Oversight Without Obstruction: The challenge for high-containment Labs. Environmental Health Perspective 116, 486-489.

Mason-D'Croz, D., Bogard, J. R., Herrero, M., Robinson, S., Sulser, T. B., Wiebe, K., Willenbockel, D., Charles, H., Godfray, J., 2020. Modelling the global economic consequences of a major African swine fever outbreak in China. Nature Food 1, 221-228.

Mazloum, A., Zinyakov, N.G., Pershin, A., Shevchenko, I., Zhukov, I., Fedoseyeva, N.D., Sharypova, D.V., Igolkin, A., Vlasov, N.A., 2018. Analysis of the changes in African swine fever virus genetic structure and biological properties during adaptation to continuous cell culture. Veterinary Science today 27, 21-25.

Mebus, C.A., 1988. African swine fever. Adv Virus Res 35, 251-269.

Meier-Kolthoff, J. P., Göker, M., (2017). VICTOR: genome-based phylogeny and classification of prokaryotic viruses. Bioinformatics 33, 3396-3404.

Michaud, V., Randriamparany, T., Albina, E., 2013. Comprehensive phylogenetic reconstructions of African swine fever virus: proposal for a new classification and molecular dating of the virus. PLoS One 8, e69662.

Miskin, J.E., Abrams, C.C., Dixon, L.K., 2000. African swine fever virus protein A238L interacts with the cellular phosphatase calcineurin via a binding domain similar to that of NFAT. J Virol 74, 9412-9420.

Monci, F., Sánchez-Campos, S., Navas-Castillo, J., Moriones, E. A., 2002. Natural recombinant between the geminiviruses Tomato yellow leaf curl Sardinia virus and Tomato yellow leaf curl virus exhibits a novel pathogenic phenotype and is becoming prevalent in Spanish populations. *Virology* **303**, 317–326.

Monteagudo, P.L., Lacasta, A., Lopez, E., Bosch, L., Collado, J., Pina-Pedrero, S., Correa-Fiz, F., Accensi, F., Navas, M.J., Vidal, E., Bustos, M.J., Rodriguez, J.M., Gallei, A., Nikolin, V., Salas, M.L., Rodriguez, F., 2017. BA71DeltaCD2: a New Recombinant Live Attenuated African Swine Fever Virus with Cross-Protective Capabilities. J Virol 91.

Montgomery, R.E., 1921. On a from of swine fever occurring in British East Africa (Kenya colony). Journal of comparative Pathology and Therapeutics 34, 159-191.

Mulumba-Mfumu, L.K., Saegerman, C., Dixon, L.K., Madimba, K.C., Kazadi, E., Mukalakata, N.T., Oura, C.A.L., Chenais, E., Masembe, C., Stahl, K., Thiry, E., Penrith,

M.L., 2019. African swine fever: Update on Eastern, Central and Southern Africa. Transbound Emerg Dis 66, 1462-1480.

Ndlovu, S., Williamson, A.L., Heath, L., Carulei, O., 2020a. Genome Sequences of Three African Swine Fever Viruses of Genotypes IV and XX from Zaire and South Africa, Isolated from a Domestic Pig *(Sus scrofa domesticus*), a Warthog (*Phacochoerus africanus*), and a European Wild Boar (*Sus scrofa*). Microbiology Resource Announcements 9, 1-3.

Ndlovu, S., Williamson, A.L., Malesa, R., van Heerden, J., Boshoff, C.I., Bastos, A.D.S., Heath, L., Carulei, O., 2020b. Genome Sequences of Three African Swine Fever Viruses of Genotypes I, III, and XXII from South Africa and Zambia, Isolated from Ornithodoros Soft Ticks. Microbiol Resour Announc 9, 1-3.

Nefedeva, M., Titov, I., Tsybanov, S., Malogolovkin, A., (2020). Recombination shapes African swine fever virus serotype-specific locus evolution. Nature Scientific Reports volume 10, 1-10.

Nei, M., Rooney, A.P., 2005. Concerted and Birth-and-Death Evolution of Multigene Families. Annual Review of Genetics 39, 121-152.

Neilan, J.G., Lu, Z., Kutish, G.F., Zsak, L., Lewis, T.L., Rock, D.L., 1997. A conserved African swine fever virus IkappaB homolog, 5EL, is nonessential for growth in vitro and virulence in domestic swine. Virology 235, 377-385.

Neilan, J.G., Zsak, L., Lu, Z., Burrage, T.G., Kutish, G.F., Rock, D.L., 2004. Neutralizing antibodies to African swine fever virus proteins p30, p54, and p72 are not sufficient for antibody-mediated protection. Virology 319, 337-342.

Neilan, J.G., Zsak, L., Lu, Z., Kutish, G.F., Afonso, C.L., Rock, D.L., 2002. Novel swine virulence determinant in the left variable region of the African swine fever virus genome. J Virol 76, 3095-3104.

Netherton, C., Rouiller, I., Wileman, T., 2004. The subcellular distribution of multigene family 110 proteins of African swine fever virus is determined by differences in C-terminal KDEL endoplasmic reticulum retention motifs. J Virol 78, 3710-3721.

Netherton, C.L., Connell, S., Benfield, C.T.O., Dixon, L.K., 2019. The Genetics of Life and Death: Virus-Host Interactions Underpinning Resistance to African Swine Fever, a Viral Hemorrhagic Disease. Front Genet 10, 402.

Nishiyama, Y., 1996. Herpesvirus genes: Molecular basis of viral replication and pathogenecity Nagoya Journal of Medical Science 59, 107-119.

Nix, R.J., Gallardo, C., Hutchings, G., Blanco, E., Dixon, L.K., 2006. Molecular epidemiology of African swine fever virus studied by analysis of four variable genome regions. Arch Virol 151, 2475-2494.

Nogal, M.L., Gonzalez de Buitrago, G., Rodriguez, C., Cubelos, B., Carrascosa, A.L., Salas, M.L., Revilla, Y., 2001. African swine fever virus IAP homologue inhibits caspase activation and promotes cell survival in mammalian cells. J Virol 75, 2535-2543.

Nurmoja, I., Petrov, A., Breidenstein, C., Zani, L., Forth, J.H., Beer, M., Kristian, M., Viltrop, A., Blome, S., 2017. Biological characterization of African swine fever virus genotype II strains from north-eastern Estonia in European wild boar. Transbound Emerg Dis 64, 2034-2041.

O'Donnell, V., Holinka, L.G., Gladue, D.P., Sanford, B., Krug, P.W., Lu, X., Arzt, J., Reese, B., Carrillo, C., Risatti, G.R., Borca, M.V., 2015a. African Swine Fever Virus Georgia Isolate Harboring Deletions of MGF360 and MGF505 Genes Is Attenuated in Swine and Confers Protection against Challenge with Virulent Parental Virus. J Virol 89, 6048-6056.

O'Donnell, V., Holinka, L.G., Krug, P.W., Gladue, D.P., Carlson, J., Sanford, B., Alfano, M., Kramer, E., Lu, Z., Arzt, J., Reese, B., Carrillo, C., Risatti, G.R., Borca, M.V., 2015b. African Swine Fever Virus Georgia 2007 with a Deletion of Virulence-Associated Gene 9GL (B119L), when Administered at Low Doses, Leads to Virus Attenuation in Swine and Induces an Effective Protection against Homologous Challenge. J Virol 89, 8556-8566.

O'Donnell, V., Holinka, L.G., Sanford, B., Krug, P.W., Carlson, J., Pacheco, J.M., Reese, B., Risatti, G.R., Gladue, D.P., Borca, M.V., 2016. African swine fever virus Georgia isolate harboring deletions of 9GL and MGF360/505 genes is highly attenuated in swine but does not confer protection against parental virus challenge. Virus Res 221, 8-14.

O'Donnell, V., Risatti, G.R., Holinka, L.G., Krug, P.W., Carlson, J., Velazquez-Salinas, L., Azzinaro, P.A., Gladue, D.P., Borca, M.V., 2017. Simultaneous Deletion of the 9GL and UK Genes from the African Swine Fever Virus Georgia 2007 Isolate Offers Increased Safety and Protection against Homologous Challenge. J Virol 91.

Ogata, H., Toyoda, K., Tomaru, Y., Nakayama, N., Shirai, Y., Claverie, J.M., Nagasaki, K., 2009. Remarkable sequence similarity between the dinoflagellate-infecting marine girus and the terrestrial pathogen African swine fever virus. Virol J 6, 178.

Olasz, F., Meszaros, I., Marton, S., Kajan, G.L., Tamas, V., Locsmandi, G., Magyar, T., Balint, A., Banyai, K., Zadori, Z., 2019. A Simple Method for Sample Preparation to Facilitate Efficient Whole-Genome Sequencing of African Swine Fever Virus. Viruses 11.

Oliveros, M., Garcia-Escudero, R., Alejo, A., Vinuela, E., Salas, M.L., Salas, J., 1999. African swine fever virus dUTPase is a highly specific enzyme required for efficient replication in swine macrophages. J Virol 73, 8934-8943.

Onisk, D.V., Borca, M.V., Kutish, G., Kramer, E., Irusta, P., Rock, D.L., 1994. Passively transferred African swine fever virus antibodies protect swine against lethal infection. Virology 198, 350-354.

Oura, C., 2013. African swine fever virus: on the move and dangerous. Vet Rec 173, 243-245.

Oura, C.A., Denyer, M.S., Takamatsu, H., Parkhouse, R.M., 2005. In vivo depletion of CD8+ T lymphocytes abrogates protective immunity to African swine fever virus. J Gen Virol 86, 2445-2450.

Penrith, M.L., 2009. African swine fever. Onderstepoort J Vet Res 76, 91-95.

Penrith, M.L., Thomson, G.R., Bastos, A.D., Phiri, O.C., Lubisi, B.A., Du Plessis, E.C., Macome, F., Pinto, F., Botha, B., Esterhuysen, J., 2004. An investigation into natural resistance to African swine fever in domestic pigs from an endemic area in southern Africa. Rev Sci Tech 23, 965-977.

Penrith, M.L., Vosloo, W., Jori, F., Bastos, A.D., 2013. African swine fever virus eradication in Africa. Virus Res 173, 228-246.

Penrith, M. L., 2020. Current status of African swine fever. CABI Agriculture and Bioscience 11, 1-26.

Phologane, S.B., Bastos, A.D., Penrith, M.L., 2005. Intra- and inter-genotypic size variation in the central variable region of the 9RL open reading frame of diverse African swine fever viruses. Virus Genes 31, 357-360.

Pietschmann, J., Guinat, C., Beer, M., Pronin, V., Tauscher, K., Petrov, A., Keil, G., Blome, S., 2015. Course and transmission characteristics of oral low-dose infection of domestic pigs and European wild boar with a Caucasian African swine fever virus isolate. Arch Virol 160, 1657-1667.

Pires, S., Ribeiro, G., Costa, J.V., 1997. Sequence and organization of the left multigene family 110 region of the Vero-adapted L60V strain of African swine fever virus. Virus Genes 15, 271-274.

Portugal, R., Coelho, J., Hoper, D., Little, N.S., Smithson, C., Upton, C., Martins, C., Leitao, A., Keil, G.M., 2015. Related strains of African swine fever virus with different virulence: genome comparison and analysis. J Gen Virol 96, 408-419.

Quembo, C.J., Jori, F., Vosloo, W., Heath, L., 2018. Genetic characterization of African swine fever virus isolates from soft ticks at the wildlife/domestic interface in Mozambique and identification of a novel genotype. Transbound Emerg Dis 65, 420-431.

Ramirez-Medina, E., Vuono, E., O'Donnell, V., Holinka, L.G., Silva, E., Rai, A., Pruitt, S., Carrillo, C., Gladue, D.P., Borca, M.V., 2019. Differential Effect of the Deletion of African Swine Fever Virus Virulence-Associated Genes in the Induction of Attenuation of the Highly Virulent Georgia Strain. Viruses 11.

Reis, A.L., Abrams, C.C., Goatley, L.C., Netherton, C., Chapman, D.G., Sanchez-Cordon, P., Dixon, L.K., 2016. Deletion of African swine fever virus interferon inhibitors from the genome of a virulent isolate reduces virulence in domestic pigs and induces a protective response. Vaccine 34, 4698-4705.

Reis, A.L., Goatley, L.C., Jabbar, T., Sanchez-Cordon, P.J., Netherton, C.L., Chapman, D.A.G., Dixon, L.K., 2017a. Deletion of the African Swine Fever Virus Gene DP148R Does Not Reduce Virus Replication in Culture but Reduces Virus Virulence in Pigs and Induces High Levels of Protection against Challenge. J Virol 91.

Reis, A.L., Netherton, C., Dixon, L.K., 2017b. Unraveling the Armor of a Killer: Evasion of Host Defenses by African Swine Fever Virus. J Virol 91.

Reis, A.L., Parkhouse, R.M., Penedos, A.R., Martins, C., Leitao, A., 2007. Systematic analysis of longitudinal serological responses of pigs infected experimentally with African swine fever virus. J Gen Virol 88, 2426-2434.

Reteno, D.G., Benamar, S., Khalil, J.B., Andreani, J., Armstrong, N., Klose, T., Rossmann, M., Colson, P., Raoult, D., La Scola, B., 2015. Faustovirus, an asfarvirus-related new lineage of giant viruses infecting amoebae. J Virol 89, 6585-6594.

Rodriguez, C.I., Nogal, M.L., Carrascosa, A.L., Salas, M.L., Fresno, M., Revilla, Y., 2002. African swine fever virus IAP-like protein induces the activation of nuclear factor kappa B. J Virol 76, 3936-3942.

Rodriguez, J.M., Moreno, L.T., Alejo, A., Lacasta, A., Rodriguez, F., Salas, M.L., 2015. Genome Sequence of African Swine Fever Virus BA71, the Virulent Parental Strain of the Nonpathogenic and Tissue-Culture Adapted BA71V. PLoS One 10, e0142889.

Rodriguez, J.M., Salas, M.L., 2013. African swine fever virus transcription. Virus Res 173, 15-28.

Rodriguez, J.M., Yanez, R.J., Almazan, F., Vinuela, E., Rodriguez, J.F., 1993. African swine fever virus encodes a CD2 homolog responsible for the adhesion of erythrocytes to infected cells. J Virol 67, 5312-5320.

Rodriguez, J.M., Yanez, R.J., Pan, R., Rodriguez, J.F., Salas, M.L., Vinuela, E., 1994. Multigene families in African swine fever virus: family 505. J Virol 68, 2746-2751.

Roger, F., Ratovonjato, J., Vola, P., Uilenber, G., 2001. Ornithodoros porcinus ticks, bushpigs, and African swine fever in Madagascar. Exp Appl Acarol 25, 263-269.

Rouiller, I., Brookes, S.M., Hyatt, A.D., Windsor, M., Wileman, T., 1998. African swine fever virus is wrapped by the endoplasmic reticulum. J Virol 72, 2373-2387.

Rowlands, R.J., Duarte, M.M., Boinas, F., Hutchings, G., Dixon, L.K., 2009. The CD2v protein enhances African swine fever virus replication in the tick vector, Ornithodoros erraticus. Virology 393, 319-328.

Rowlands, R.J., Michaud, V., Heath, L., Hutchings, G., Oura, C., Vosloo, W., Dwarka, R., Onashvili, T., Albina, E., Dixon, L.K., 2008. African swine fever virus isolate, Georgia, 2007. Emerg Infect Dis 14, 1870-1874.

Ruiz-Gonzalvo, F., Rodriguez, F., Escribano, J.M., 1996. Functional and immunological properties of the baculovirus-expressed hemagglutinin of African swine fever virus. Virology 218, 285-289.

Saitou, N., Nei, M., 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution. Molecular Biology and Evolution 4, 406-425.

Salas, M.L., Andres, G., 2013. African swine fever virus morphogenesis. Virus Res 173, 29-41.

Salguero, F.J., 2020. Comparative Pathology and Pathogenesis of African Swine Fever Infection in Swine. Front Vet Sci 7, 282.

Salguero, F.J., Gil, S., Revilla, Y., Gallardo, C., Arias, M., Martins, C., 2008. Cytokine mRNA expression and pathological findings in pigs inoculated with African swine fever virus (E-70) deleted on A238L. Vet Immunol Immunopathol 124, 107-119.

Sanchez-Cordon, P.J., Jabbar, T., Berrezaie, M., Chapman, D., Reis, A., Sastre, P., Rueda, P., Goatley, L., Dixon, L.K., 2018a. Evaluation of protection induced by immunisation of domestic pigs with deletion mutant African swine fever virus BeninDeltaMGF by different doses and routes. Vaccine 36, 707-715.

Sanchez-Cordon, P.J., Montoya, M., Reis, A.L., Dixon, L.K., 2018b. African swine fever: A re-emerging viral disease threatening the global pig industry. Vet J 233, 41-48.

Sanford, B., Holinka, L.G., O'Donnell, V., Krug, P.W., Carlson, J., Alfano, M., Carrillo, C., Wu, P., Lowe, A., Risatti, G.R., Gladue, D.P., Borca, M.V., 2016. Deletion of the thymidine

kinase gene induces complete attenuation of the Georgia isolate of African swine fever virus. Virus Res 213, 165-171.

Sanna, G., Dei Giudici, S., Bacciu, D., Angioi, P.P., Giammarioli, M., De Mia, G.M., Oggiano, A., 2017. Improved Strategy for Molecular Characterization of African Swine Fever Viruses from Sardinia, Based on Analysis of p30, CD2V and I73R/I329L Variable Regions. Transbound Emerg Dis 64, 1280-1286.

Sanz, A., Garcia-Barreno, B., Nogal, M.L., Vinuela, E., Enjuanes, L., 1985. Monoclonal antibodies specific for African swine fever virus proteins. J Virol 54, 199-206.

Sereda, A.D., Balyshev, V.M., 2011. [Antigenic diversity of African swine fever viruses]. Vopr Virusol 56, 38-42.

Sereda, A.D., Solovkin, S.L., Fugina, L.G., Makarov, V.V., 1992. [Immune reactions to the African swine fever virus]. Vopr Virusol 37, 168-170.

Service, M.W., 1980. Soft ticks (Order Metastigmata: Family Argasidae), A Guide to Medical Entomology. Macmillan Education UK, London, pp. 154-159.

Shackelton, L. A., Hoelzer, K., Parrish, C. R., Holmes, E. C., 2012. Comparative analysis reveals frequent recombination in the parvoviruses. *J. Gen. Virol.* 88, 3294–3301.

Suarez, C., Andres, G., Kolovou, A., Hoppe, S., Salas, M.L., Walther, P., Krijnse Locker, J., 2015. African swine fever virus assembles a single membrane derived from rupture of the endoplasmic reticulum. Cell Microbiol 17, 1683-1698.

Steyn, D.G., 1932, 'East African virus disease in pigs', 18th Report of the Director of Veterinary Services and Animal Industry, Union of South Africa 1, 99–109.

Sumption, K.J., Hutchings, G.H., Wilkinson, P.J., Dixon, L.K., 1990. Variable regions on the genome of Malawi isolates of African swine fever virus. J Gen Virol 71 (Pt 10), 2331-2340.

Sur, J.H., 2019. How far can African swine fever spread? J Vet Sci 20, e41.

Tabares, E., Marcotegui, M.A., Fernandez, M., Sanchez-Botija, C., 1980a. Proteins specified by African swine fever virus. I. Analysis of viral structural proteins and antigenic properties. Arch Virol 66, 107-117.

Tabares, E., Martinez, J., Ruiz Gonzalvo, F., Sanchez-Botija, C., 1980b. Proteins specified by African swine fever virus. II. Analysis of proteins in infected cells and antigenic properties. Arch Virol 66, 119-132.

Takamatsu, H.H., Denyer, M.S., Lacasta, A., Stirling, C.M., Argilaguet, J.M., Netherton, C.L., Oura, C.A., Martins, C., Rodriguez, F., 2013. Cellular immunity in ASFV responses. Virus Res 173, 110-121.

Thiry, E., Meurens, F., Muylkens, B., McVoy, M., Gogev, S., Thiry, J., Vanderplasschen, A., Epstein, A., Keil, G., Schynts, F., 2005. Recombination in alphaherpesviruses. *Rev. Med. Virol.* 15, 89–103.

Thomson, G.R., 1985. The epidemiology of African swine fever: the role of free-living hosts in Africa. Onderstepoort J Vet Res 52, 201-209.

Tu, S.L., Staheli, J.P., McClay, C., McLeod, K., Rose, T.M., Upton, C., 2018. Base-By-Base Version 3: New Comparative Tools for Large Virus Genomes. Viruses 10.

Tulman, E.R., Rock, D.L., 2001. Novel virulence and host range genes of African swine fever virus. Curr Opin Microbiol 4, 456-461.

Vydelingum, S., Baylis, S.A., Bristow, C., Smith, G.L., Dixon, L.K., 1993. Duplicated genes within the variable right end of the genome of a pathogenic isolate of African swine fever virus. J Gen Virol 74 (Pt 10), 2125-2130.

Wang, N., Zhao, D., Wang, J., Zhang, Y., Wang, M., Gao, Y., Li, F., Wang, J., Bu, Z., Rao, Z., Wang, X., 2019. Architecture of African swine fever virus and implications for viral assembly. Science 366, 640-644.

Wardley, R.C., Norley, S.G., Wilkinson, P.J., Williams, S., 1985. The role of antibody in protection against African swine fever virus. Vet Immunol Immunopathol 9, 201-212.

Wilkinson, P.J., Wardley, R.C., Williams, S.M., 1981. African swine fever virus (Malta/78) in pigs. J Comp Pathol 91, 277-284.

Woźniakowski, G., Frączyk, M., Niemczuk, K., Pejsak, Z., 2016. Selected aspects related to epidemiology, pathogenesis, immunity, and control of African swine fever. Journal of Veterinary Research 60, 119-125.

Xiong, D., Zhang, X., Yu, J., Wei, H., 2019. Rapid phylogenetic analysis of African swine fever virus from metagenomic sequences. bioRxiv, 1-13.

Yanez, R.J., Rodriguez, J.M., Nogal, M.L., Yuste, L., Enriquez, C., Rodriguez, J.F., Vinuela, E., 1995. Analysis of the complete nucleotide sequence of African swine fever virus. Virology 208, 249-278.

Yu, M., Morrissy, C.J., Westbury, H.A., 1996. Strong sequence conservation of African swine fever virus p72 protein provides the molecular basis for its antigenic stability. Arch Virol 141, 1795-1802.

Yun, C.H., 2020. - Editorial - Unforeseen enemy: African swine fever. Asian-Australas J Anim Sci 33, 1-3.

Zani, L., Forth, J. H., Forth, L., Nurmoja, I., Leidenberger, S., Henke, J., Carlson, J., Breidenstein, C., Viltrop, A., Höper, D., Sauter-Louis, C., Beer, M., Blome, S., 2018.

Deletion at the 5’-end of Estonian ASFV strains associated with anattenuated phenotype. Nature Scientific Reports volume 8, 1-11.

Zhu, J.J., Ramanathan, P., Bishop, E.A., O'Donnell, V., Gladue, D.P., Borca, M.V., 2019. Mechanisms of African swine fever virus pathogenesis and immune evasion inferred from gene expression changes in infected swine macrophages. PLoS One 14, e0223955.

Zhu, Z., Xiao, C. T., Fan, Y., Cai, Z., Lu, C., Zhang, G., Jiang, T., Tan, Y., Peng, Y., 2019. Homologous recombination shapes the genetic diversity of African swine fever viruses. Veterinary Microbiology 236, 1-9.

Zsak, L., Borca, M.V., Risatti, G.R., Zsak, A., French, R.A., Lu, Z., Kutish, G.F., Neilan, J.G., Callahan, J.D., Nelson, W.M., Rock, D.L., 2005. Preclinical diagnosis of African swine fever in contact-exposed swine by a real-time PCR assay. J Clin Microbiol 43, 112-119.

Zsak, L., Caler, E., Lu, Z., Kutish, G.F., Neilan, J.G., Rock, D.L., 1998. A nonessential African swine fever virus gene UK is a significant virulence determinant in domestic swine. J Virol 72, 1028-1035.

Zsak, L., Lu, Z., Burrage, T.G., Neilan, J.G., Kutish, G.F., Moore, D.M., Rock, D.L., 2001a. African swine fever virus multigene family 360 and 530 genes are novel macrophage host range determinants. J Virol 75, 3066-3076.

Zsak, L., Lu, Z., Kutish, G.F., Neilan, J.G., Rock, D.L., 1996. An African swine fever virus virulence-associated gene NL-S with similarity to the herpes simplex virus ICP34.5 gene. J Virol 70, 8865-8871.

Zsak, L., Sur, J.H., Burrage, T.G., Neilan, J.G., Rock, D.L., 2001b. African Swine Fever Virus (Asfv) Multigene Families 360 and 530 Genes Promote Infected Macrophage Survival. ScientificWorldJournal 1, 97.