Virus Genes
DOI 10.1007/s11262-014-1156-7

Comparative analysis of the complete genome sequences of Kenyan African swine fever virus isolates within p72 genotypes IX and X

Richard P. Bishop · Clare Fleischauer · Etienne P. de Villiers · Edward A. Okoth · Marisa Arias · Carmina Gallardo · Chris Upton

Received: 27 October 2014 / Accepted: 11 December 2014

Abstract Twelve complete African swine fever virus (ASFV) genome sequences are currently publicly available and these include only one sequence from East Africa. We describe genome sequencing and annotation of a recent pig-derived p72 genotype IX, and a tick-derived genotype X isolate from Kenya using the Illumina platform and comparison with the Kenya 1950 isolate. The three genomes constitute a cluster that was phylogenetically distinct from other ASFV genomes, but 98–99 % conserved within the group. Vector-based compositional analysis of the complete genomes produced a similar topology. Of the 125 previously identified 'core' ASFV genes, two ORFs of unassigned function were absent from the genotype IX sequence which was 184 kb in size as compared to 191 kb for the genotype X. There were multiple differences among East African genomes in the 360 and 110 multicopy gene families. The gene corresponding to 360-19R has transposed to the 5′ variable region in both genotype X isolates. Additionally, there is a 110 ORF in the tick-derived genotype X isolate formed by fusion of 13L and 14L that is unique among ASFV genomes. In future, functional analysis based on the variations in the multicopy families may reveal whether they contribute to the observed differences in virulence between genotpye IX and X viruses.

Keywords African swine fever virus · p72 Genotype · Genome sequence · Illumina · Multicopy gene family

Edited by William Dundon.

R. P. Bishop (✉) · E. P. de Villiers · E. A. Okoth
International Livestock Research Institute (ILRI),
PO Box 30709, Nairobi 00100, Kenya
e-mail: r.bishop@cgiar.org

C. Fleischauer · C. Upton
Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC V8W 3P6, Canada

M. Arias · C. Gallardo
EU Reference Laboratory for ASF, CISA-INIA, Valdeolmos, 28130 Madrid, Spain

Introduction

African swine fever virus (ASFV) is a large DNA virus that causes a rapidly lethal disease in domestic pigs and has become a significant threat to the global pig industry [1]. The ancestral hosts are Argasid ticks in the *Ornithodoros moubata* complex and African wild suids, most importantly warthogs and bushpigs. To date, 12 fully annotated complete genome sequences are available in the public domain [2, 3], but these only represent six of the 22 currently published genotypes, which are defined by the sequence of the 3′ end of the p72 surface capsid protein gene [3, 4].

In Kenya and Uganda, two p72 genotypes, IX and X, have been identified as being prevalent in domestic pigs, warthogs and ticks [5–7]. The sequences of the 3′ ends of the p72 gene sequences defining genotypes IX and X are the most closely related of any of the 22 identified genotypes [4]. In addition, genotype I, which is typically West African in distribution, has also been identified in Kenya from the p72 sequence of a single isolate [8]. Currently, a single Kenyan ASFV isolate collected from a domestic pig in 1950 is the only fully sequenced and annotated genome that is classified within the East African genotype X [3]. We describe the complete sequencing and comparative analysis of a genotype X virus derived from a tick that was extracted from a warthog burrow in central Kenya and a genotype IX virus associated with an ASF outbreak in domestic pigs in Western Kenya in 2006. When used for

Published online: 03 February 2015

experimental infection of Spanish Landrace pigs these two viruses differed in virulence, with the genotype IX virus being highly virulent, whereas the tick-derived genotype X virus was only moderately virulent, two infected pigs surviving infection [9]. Analysis of these two ASFV genomes and comparison both with one another, and to previously determined ASFV genomes [3, 10], provided insight into genetic diversity among East African ASFV isolates and revealed features that may help explain the molecular basis of the observed differences in virulence.

Materials and methods

Virus isolation and DNA preparation

The primary virus isolation of Ken06.Bus [5] and Ken05/Tk1 [6], from a domestic pig in Western Kenya (Busia district) and a tick from a warthog burrow at Kapiti plains ranch in central Kenya, respectively, are described in the above references. In order to provide virus stocks for extraction of DNA, pigs were inoculated intramuscularly with ten 50 % hemadsorbing doses (HAD_{50}) of each ASFV isolate. Virus was purified from the red blood cell fraction of infected pig blood essentially as described [11], except that the virus was fractionated on 25–50 % rather than 25–60 % sucrose gradients… In addition, virus preparations were treated with DNase (50 pg/ml), followed by 1 % Tween 80 in order to remove contaminating cellular DNA and lipids before loading onto sucrose gradients. DNA was prepared from isolated virus by phenol extraction and ethanol precipitation, following lysis of virus with SDS and proteinase K treatment.

DNA amplification and sequencing

Purified DNA preparations were quantified using a nanodrop spectrophotometer. The concentrations were 3.45 and 14.2 ng/μl for Ken06.Bus and Ken05/Tk1, respectively. Because of the low concentration of DNA in the Ken06.Bus preparation, the DNA from this sample was non-specifically amplified using a commercial kit (Genomphi) prior to genome sequencing. The DNA isolated from Ken05/Tk1 was used directly as a template for sequencing without amplification. Preparation of True-Seq libraries and sequencing on an Illumina hi-seq sequencer were performed using the manufacturer's protocols, at the University of Liverpool Centre for Genomics Research. 11,019,903 and 8,373,530 paired end reads (2 × 150 bp) were generated for Ken06.Bus and Ken05/Tk1, respectively. The raw data were processed to remove low quality reads and residual adapter sequences using the Trim Sequence tool available in CLC Bio Genomics Workbench Version 5.5.1.

Genome assembly and annotation

GenBank files containing previously sequenced genomes were obtained from the viral orthologous clusters application (VOCs) [12, 13] reference strains (Kenya 1950 AY261360, Malawi Lil-20-1 AY261361, Tengani62 AY261364, Warthog AY261366, Warmbaths AY261365, Pretorisuskop-96-4 AY261363, Georgia 2007/1 FR682468, Mkuzi 1979 AY261362, BA71V-U18466, OurT88/3 AM712240, E75 FN557520, Benin97/1 AM712239). The complete Ken05/Tk1 and Ken06.bus genome sequences were submitted to GenBank with accession numbers ASFV-Ken05.TkKM111294 and ASFV-Ken06.BusKM111295, respectively.

The two ASFV genomes were individually assembled with the de novo assembly tool available within CLC Bio Genomics Workbench Version 5.5.1, without filtering out the contaminating porcine sequences. They were subsequently annotated using the genome annotation transfer utility (GATU) at the Viral Bioinformatics Resource Center [14]. The genome sequences of the virulent ASFV isolates Spain E75 and Kenya 1950 were used for the in-depth comparative analysis, but the gene nomenclature follows that of the reference genome BA117 [15] as modified in [16]. Exceptions to this were made for two members of the 360 multicopy family loci in genotype X Ken05/Tk1. ORF MGF 360-19R whose name was retained in order to highlight its identity to the corresponding gene in other ASFV genomes, although it was uniquely present at the left terminus and not the right. For 360-1L, the nomenclature was retained for an additional copy of the gene that was duplicated and also transposed to the right hand end of the virus. A conservative approach to annotation was adopted, in which only open reading frames (ORFs) considered to have a high probability of encoding a functional polypeptide were annotated. Criteria used were that the ORFs were non-overlapping and longer than 180 nucleotides. For example, the DP63R and multigene 300-3L loci were not included due to their complete overlap with other annotated ORFs. An exception to the length restriction was made for gene C62L which contained 45 amino acids in both Kenyan isolates and also ORF C84L that was truncated in East African and Malawi LIL20 isolates. These ORFs were included due to their high level of conservation with annotated genes in other ASFV genomes. Putative unassigned ORFs were identified by comparing the nucleotide and amino acid sequences against all fully sequenced ASFV genomes using two applications, viral genome organizer [17] and base-by-base [18, 19], in synchrony with BLAST searches [20]. ORFs

with multiple potential start codons were manually adjusted in length to match the consensus length.

Alignments and phylogenetic analysis

A multiple sequence alignment was created using the online version of MAFFT [21, 22] with 14 ASFV genomes. The output from this alignment contained several large gaps and a variety of misalignments that rendered the complete genome alignment impractical for phylogenetic analyses. To create a reliable phylogenetic tree, several smaller conserved regions within the central portion of the ASFV genome were isolated and assessed for gaps and alignment quality. Each of these isolated segments generated similar phylogenetic trees with only minor differences in the arrangement of the most closely related ASFV isolates. There was one exception, in which the Georgian and Tengani isolates were isolated from distinct groupings observed in all other trees, but the East African isolates were always placed in a discrete and closely related cluster. The largest of these isolated regions was chosen for construction of a final tree and remaining minor misalignments were corrected manually. This selected region was 22 kb in length and contained nineteen core genes, from the BA71V-D1133 locus in the helicase superfamily II) to BA71V-E184L, (an ORF of unknown function). To compare the genomic sequences of this region between isolates, any columns containing a gap were removed so only single nucleotide polymorphisms (SNPs) contributed toward the percent similarity scores between sequences. The resulting genomic alignment without gaps was approximately 21 kb in length. Additionally, we extracted and concatenated the sequences of the 19 proteins that fell within the 22 kb region and determined the percent similarity between isolates within the ORFs. A phylogenetic tree was created using the MEGA5 software [23] with this 21 kb alignment using the Neighbor-joining method [24] with 1,000 bootstrap replicates. A similar topology of the ASFV isolates resulted when Minimum-Evolution was used to generate the tree. A dot plot was produced using JDotter [25] to identify whether any rearrangements had occurred in the genomes of the two newly sequenced Kenyan isolates. The number of SNPs was determined for each genome and displayed on the corresponding branch.

To provide independent confirmation of the sequence identity analysis, a composition vector-based method (compdist software; J. Valverde, personal communication) was used for comparison with the analysis using homology-based algorithms. The composition vector (CV) of each sequence was obtained by determining the frequencies of short DNA strings, and then the pairwise distance between the CVs was calculated. Finally, a dendrogram using the neighbor-joining (NJ) method was generated based on the distance matrix. The only parameter in this method is the length K of the nucleotides or predicted peptides derived from them [26].

Results

Comparison of newly determined Kenyan ASFV genomes to other ASFV genomes

We have determined the first complete genome sequence from a p72 genotpye IX genotype and the second from genotpye X. The genomes of these two East African isolates clustered geographically together with genotype X isolate Kenya 1950 and were distinct from other ASFV genomes.

The Kenyan ASFV genomes assembled as single contigs of 191,058 bp (Ken05/Tk1) and 184,368 bp (Ken06.Bus) against a background of excess porcine DNA. Within the 22 kb region selected for in-depth analysis, Ken05/Tk1 exhibited greater than 99 % DNA identity with the 1950 Kenyan pig-derived virus, whereas the percentage identity between the tick-derived genotype X (Ken05/Tk1) and genotype IX (Ken06.Bus) was approximately 98 % (Table 1). The percentage identity at the protein level following translation of the 19 ORFS was 99.6 % between the two genotype X isolates, and 99.2 and 99.3 %, respectively, between the individual genotype X isolates and the genotype IX Ken06.Bus isolate. Viruses from regions outside East Africa were less closely related, for example, an ASFV from Malawi Lil-20-1 was between 94.8 and 94.99 % similar to the Kenyan isolates and the E75 isolate from Spain exhibited approximately 94 % DNA similarity (Table 1). All three of the Kenyan isolates, although genetically similar to one another, were relatively distinct from the other ASFV genomes.

Comparison of Kenyan genotype IX and X genomes

As mentioned in the previous section, the Ken05/Tk1 genome is the longest of the two isolates with a length of approximately 191 kb, while Ken06.Bus is approximately 184 kb in length. Ken05/Tk1 contains all 125 of the conserved ASFV genes identified in previous analyses [2], while Ken06.Bus contains only 123 of these, with a further five being truncated. Of the 125 conserved loci, two genes BA71V-E111R and BA71V-DP96R, both of unknown function, are absent from Ken06.Bus. Five more conserved genes are truncated in Ken06.Bus, these consist of 360 multigene family 8L, A240L, 0174L, C84L and I267L. All of the differences were due to 1–2 bp indels or single nucleotide substitutions. All of the truncated ORFs were generally similar to the other ASFV isolates with Ken05/

Table 1 Percentage identity of five African swine fever virus isolates within a 21 kb nucleotide sequence alignment derived from a 22 kb region with gaps removed

	Kenya 1950	Ken05/Tk1	Ken06.Bus	Malawi Lil	E75
Kenya 1950		99.37	98.55	94.96	94.21
Ken05/Tk1	99.6		98.46	94.99	94.13
Ken06.Bus	99.21	99.29		94.81	93.98
Malawi Lil	96.32	96.34	96.13		95.03
E75	95.09	95.13	94.91	95.87	

Values above the diagonal represent the DNA sequence similarity. Those below represent the concatenated protein sequence identity

Tk1 (genotype X) contained no genes with major truncations when compared to genotype X Kenya 1950. There were minor differences, including SNPs, which in some instances were non-synonymous and resulted in amino acid changes. A dot plot also revealed additional indels and small tandem repeats that differentiate the contemporary genomes of Kenyan isolates IX and X (data not shown). No rearrangements were observed between the genotype IX and X genomes and all repeated regions consisted of previously described multigene family members [2].

Several differences within the 360 and 110 gene families that have previously been implicated in determining host range and virulence were observed between the Kenyan Genotype IX and X genomes. The system used for naming of the members of these multigene families follows that described previously [2, 16]. Specific differences were (1) Ken05/Tk1 contains a novel gene created by fusion between two 110 multigene loci 13L and 14L; (2) the recent Kenyan genotype IX and X isolate genomes both contain 110 5L, which is absent from the Kenya 1950 genome; (3) Multigenes 110-7L and 8L are absent and the 360 multigene 13L is fragmented in KenBus.06, but all are present in the genotype X Kenyan isolates; (4) Kenya 1950 contains the 110 multigene 11L ORF, which is fragmented in Ken05/Tk1, and also in KenBus.06; (5) the Ken05/Tk1 and the Kenya 1950 viruses contain three 360 multigene family members located in the left terminal region (1L, 19R, 2L) that appear to be primarily absent from KenBus.06, although a fragment of 360 2L is present and contains 76 of the 365 amino acids corresponding to the protein in Kenya 1950 and Ken05Tk/1 exhibiting 89.5 and 77 % identity to these isolates, respectively; (6) the location of the 360 multigene 19R at the 5′ end of the two Kenyan genotype X isolates was unique among ASFV genomes and suggested recent transposition of the gene from the right to the left terminus; (7) Kenya 1950 (genotype X) contains a triplet of 360 multigenes (4L-6L-7L), whereas Ken05/Tk1 lacked the 7L copy, while the 6L copy was absent and 4L was truncated in Ken06.Bus; (8) Within the right terminal region genotype X isolates contain the 360 multigene 21R, whereas Ken06.Buscontains 22R; (9) The 3′ terminus in Ken05/Tk1 contains an additional copy of the 360 multigene 1L which has approximately 66 % amino acid similarity to the 1L gene located at the left end of the genome.

A final difference was the presence of a p22 structural protein repeat with 61.88 % protein identity, in both the tick-derived Ken05/Tk1 and the 1950 genotype X viruses, that is absent from the genotype IX virus. The occurrence of repeats of this gene has been observed in some other ASFV genomes and may be involved in antigenic variation [2]. Given the close genetic relationship of genotypes X and IX and separation of other isolates, duplication of p22 may have occurred independently on more than one occasion.

Phylogenetic analysis

DNA sequence and concatenated protein sequence percent identities of sequences derived from the 22 kb region selected for phylogenetic analysis for five of the ASFV genomes included in this study are outlined in Table 1. Both at the DNA and protein levels, Ken05/Tk1 is closest to the reference genotype X isolate Kenya 1950, while Ken06.Bus is also closely related to both Ken05/Tk1 and Kenya 1950. Malawi Lil-20-1 was less closely related to the genomes of Kenyan ASFV, and E75 was the least similar (Table 1). Overall, the data indicate that the three East African isolates are significantly closer genetically to one another than they are to any other ASFV isolates whose genomes have been sequenced to date. This is illustrated in the unrooted neighbor-joining tree shown in Fig. 1, that is derived from alignment of the same relatively conserved central 21 kb region with the gaps removed, using all 14 publicly available ASFV isolates. Phylograms generated using composition vector methods for building trees that were derived from ORFs spanning the complete genome produced similar topologies to that in Fig. 1 (data not shown).

The phylogenetic tree confirms that Ken05/Tk1 is most closely related to Kenya 1950, and that Ken06.Bus clusters closely with these two isolates in a clade that is distinct from other ASF genomes (Fig. 1). Malawi Lil-20-1 is most

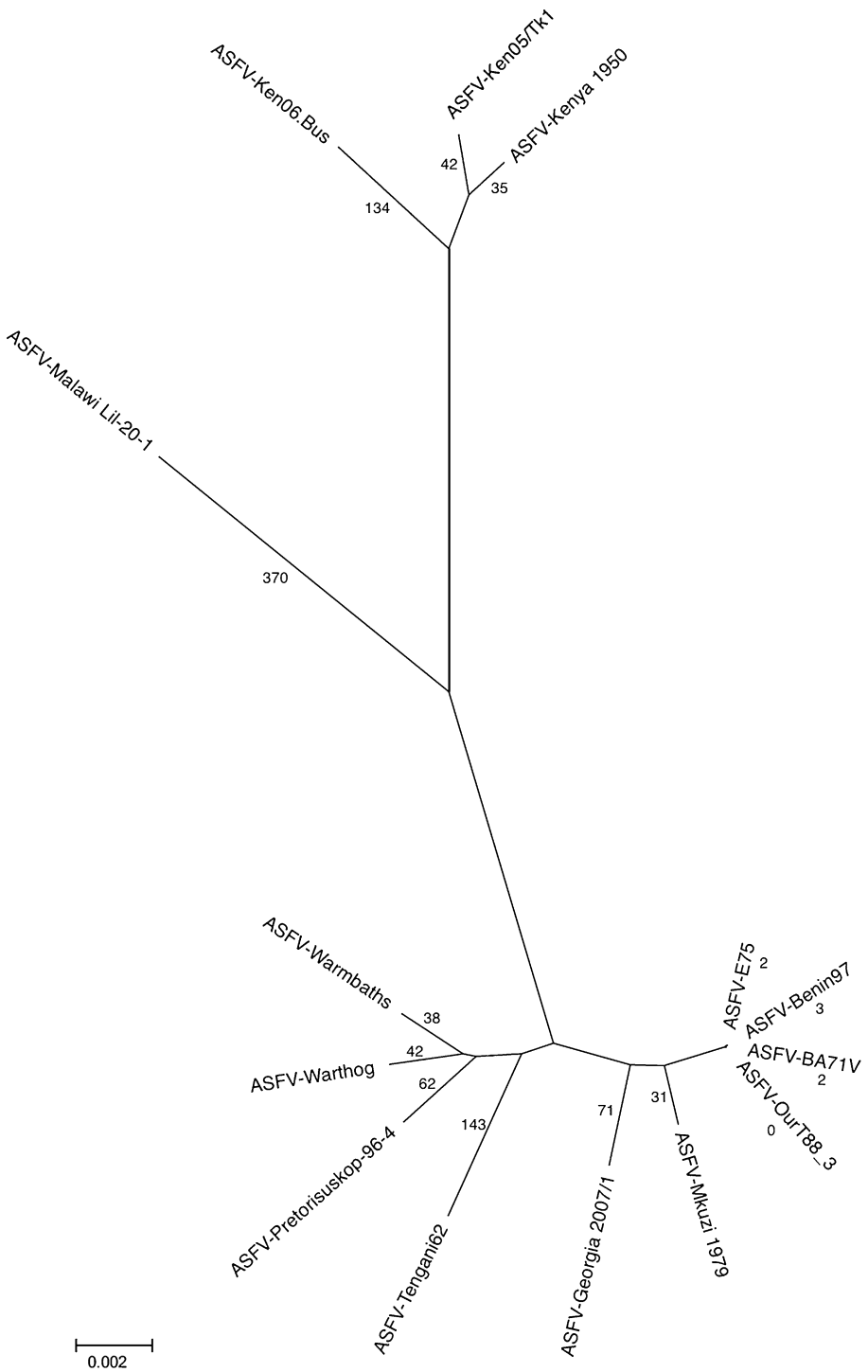

Fig. 1 Unrooted neighbor-joining tree obtained from an alignment of the central conserved region present in the genomes of 14 African swine fever virus isolates. The tree was generated using MEGA5 with 1,000 bootstrap replications. Bootstrap values at all nodes were 100 except for those between E75 and Benin (63), and Warthog and Warmbaths (98). The *scale bar* represents the number of nucleotide substitutions per site

distant from all other isolates, including the East African genome cluster and the two main clades, one comprising South Africa-Warmbaths, Namibia-Warthog, South Africa-Pretorisuskop and Malawi-Tengani, and the other containing Georgia 2007/1, South Africa-Mkuzi, Portugal-OurT88, Spain-BA71V (attenuated), Benin 97 and Spain-E75 (virulent).

Discussion

We report here, the first complete genome sequence of an ASFV genotype IX isolate and a second genotype X genome sequence from East Africa. The three East African genome sequences cluster geographically and are distinct from other ASFV genomes.

The two newly determined genome sequences reported here represent the first to be sequenced using Illumina next generation technology using relatively small amounts of template DNA containing a significant excess of contaminating porcine DNA. The amplified genotype IX sequence may have contained a very small number of probable sequencing errors as indicated by the apparent truncation of two housekeeping genes, thymidylate kinase and DNA polymerase beta-like protein. These genes will need to be verified by Sanger sequencing. Overall, the genome sequence was nonetheless of relatively good quality based on the high level of overall similarity to the genotype X sequence. Confirmation of classification within p72 genotype IX based on a high quality BLAST match with the previously determined p72 3′ end sequence of a genotype IX isolate also confirmed the overall quality of the genome sequence.

Although the genomes of the recently isolated genotype IX and X viruses are relatively similar in overall sequence, there is evidence that pig-derived genotype IX and tick-derived genotype X differ in virulence in domestic pigs. Genotype IX was associated with a lethal outbreak of ASF in Western Kenya [5], while genotype X was derived from a tick from central Kenya. Additionally, genotype X isolates were extracted from twelve apparently asymptomatic pigs at the Nairobi Farmers Choice slaughterhouse in 2005 [6] and also detected by PCR in indigenous 'carrier' pigs in Homa Bay district Western Kenya [7]. The difference in virulence was directly confirmed by experimental infections of Spanish pigs in the BSL3 pig unit at CISA-INIA in which two out of four pigs infected with the tick-derived genotype X survived, while all four pigs infected with genotype Ken06.Bus IX died [9].

Differences between ASFV genomes are frequently defined by variation in the number and sequence of multicopy gene family sequences located in the variable left and right hand termini of the virus, particularly the 360 and 110 gene families, named for the average number of amino acids that they encode, reviewed in [2]. Comparison of Ken05/Tk1 and Ken06.Bus indicated that a number of 360 and 110 gene family members were apparently absent from the more virulent genotype IX virus, Ken06.Bus when compared to genotype X. In addition, there were more subtle differences in both copy number and location of members both the 110 and 360 multigene families.

A previous comparison of virulent pig-derived and avirulent tick-derived isolates classified within ASFV genotype I from West Africa and the Iberian Peninsula, implicated an 8 kb deletion containing several members of the 360 and 505 multicopy families that was observed in the avirulent virus and presumed to underpin the phenotype [16]. This interpretation correlated with earlier experimental observations that deletion of members of the 360 gene family affected macrophage host range and virus replication in *O. porcinus ticks* [27, 28]. Differences in these two gene families may also be involved in the differences in virulence between genotypes IX and X in East Africa. However, the precise molecular mechanism may be different, since there is no increase in the multigene family 110 or 360 copy number in the genome of the more virulent genotype IX virus. The pig-derived Kenya 1950 genotype X isolate was described as virulent [3]. The genome of the 1950 isolate was very similar to that of the tick-derived genotype X virus from central Kenya, as indicated by the current comparative study. Detailed analysis of the differences might be informative in providing insight into why genotype X now appears to typically produce less virulent infections in pigs. Specific differences from Kenya 1950, that could be reproduced experimentally using functional genomics approaches, included the absence of multigene family 360 member 7L, the presence of 110 family 5L, and the fusion of the two family 110 genes 13L and 14L in Ken05/Tk1.

The absence of the DP96R gene from the genotype IX genome is shared only with Malawi Lil-20-1 and suggests that the gene is not essential for a functional virus, since the genotype IX isolate that we have sequenced in this study is readily transmissible between pigs and produces lethal infections of swine.

In conclusion, recent molecular epidemiological analyses [5, 6, 8] indicate that only genotypes IX and X appear widespread in Kenya and Uganda. The nucleotide sequences encoding the C-terminal region of p72 confirmed that Ken06.Bus and Ken05/Tk1, respectively, fall into these closely related genotypes. The genome sequences indicate that the diversity of the virus gene pool may be relatively limited in these two countries.

Statement of author contributions

RB, CG, and MA conceived the study. CG generated the purified ASFV DNA. EdV, CF, and CU analyzed the data. RB, EdV, CF, and CU wrote the manuscript.

Acknowledgments We are grateful to the CSIRO-AusAID Africa-Australia Food Security Initiative and to INIA, Spain under research agreement TF069018 for funding. We are particularly grateful to Dr. Bruce Pengelly of CSIRO for his support of this project. CU and CF were supported by an NSERC Discovery grant. We thank Dr. Harry Noyes and the Liverpool University Centre for Genomics technical advice relating to genome sequencing and template amplification. This is ILRI publication number 92–9146–365–5.

References

1. S. Costard, B. Wieland, W. de Glanville, F. Jori, R. Rowlands, W. Vosloo, F. Roger, D.U. Pfeiffer, L.K. Dixon, Philos. Trans. R. Soc. B **364**, 2683–2696 (2009)

2. L.K. Dixon, D.A.G. Chapman, C.L. Netherton, C. Upton, Virus Res. **173**, 3–14 (2013)
3. E.P. de Villiers, C. Gallardo, M. Arias, M. da Silva, C. Upton, R. Martin, R.P. Bishop, Virology **400**, 128–136 (2010)
4. A.D. Bastos, M.L. Penrith, C. Crucieire, J.L. Edrich, G. Hutchings, F. Roger, E. Couacy-Hymann, G.R. Thompson, Arch. Virol. **148**, 693–706 (2003). doi:10.1007/s00705-002-0946-8
5. C. Gallardo, D.M. Mwaengo, J.M. Macharia, A. Soler, E. Taracha, E. Okoth, E. Martin, R.P. Bishop, Virus Genes **38**, 85–95 (2009)
6. C. Gallardo, E. Okoth, V. Pelayo, R. Anchuelo, E. Martín, A. Simón, A. Llorente, R. Nieto, A. Soler, R. Martín, M. Arias, R.P. Bishop, J. Gen. Virol. **92**, 432–444 (2011)
7. E. Okoth, C. Gallardo, J.M. Macharia, A. Omore, V. Pelayo, D.W. Bulimo, M. Arias, P. Kitala, K. Baboon, I. Lekolol, D. Mijele, R.P. Bishop, Prev. Vet. Med. **110**, 198–205 (2013)
8. B.A. Lubisi, A.D. Bastos, R.M. Dwarka, W. Vosloo, Arch. Virol. **150**, 2439–2452 (2005). doi:10.1007/s00705-005-0602-1
9. M. Tignon, C. Gallardo, C. Iscaro, E. Hutet, Y. Van der Stede, D. Kolbasov, G. Mario Da Mia, M.F. Le Potier, R.P. Bishop, M. Arias, F. Koenen, J. Virol. Methods **178**, 161–170 (2011)
10. D.A.G. Chapman, A.C. Darby, M. da Silva, C. Upton, A.D. Radford, L.K. Dixon, Emerg. Inf. Dis. **17**, 599–605 (2011)
11. R.D. Wesley, A.E. Tuthill, Prev. Vet. Med. **2**, 53–62 (1984)
12. A. Ehlers, J. Osborne, S. Slack, R.L. Roper, C. Upton, Bioinformatics **18**, 1544–1545 (2002)
13. C. Upton, S. Slack, A.L. Hunter, A. Ehlers, R.L. Roper, J. Virol. **77**, 7590–7600 (2003)
14. V. Tcherepanov, A. Ehlers, C. Upton, BMC Gen. **7**, 1–10 (2006)
15. R.J. Yanez, J.M. Rodriguez, M.L. Nogal, L. Yuste, C. Enriquez, J.F. Rodriguez, E. Vinuela, Virology **208**, 249–278 (1995)
16. D.A.G. Chapman, V. Tcherepanov, C. Upton, L.K. Dixon, J. Gen. Virol. **89**, 397–408 (2008)
17. C. Upton, D. Hogg, D. Perrin, M. Boone, N.L. Harris, Virus Res. **70**, 55–64 (2000)
18. W. Hillary, S.H. Lin, C. Upton, Microb. Inform. Exp. **1**, 1–6 (2011)
19. R.L. Roper, V. Tcherepanov, C. Upton, Bioinformatics **5**, 1–9 (2004)
20. S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, J. Mol. Biol. **215**, 403–410 (1990)
21. K. Katoh, G. Asimenos, H. Toh, Methods Mol. Biol. **537**, 39–64 (2009)
22. K. Katoh, D.M. Standley, Mol. Biol. Evol. **30**, 772–780 (2013)
23. K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, S. Kumar, Mol. Biol. Evol. **28**, 2731–2739 (2011)
24. N. Saitou, M. Nei, Mol. Biol. Evol. **4**, 406–425 (1987)
25. R. Brodie, R.L. Roper, C. Upton, J. Dotter, Bioinformatics **20**, 279–281 (2004)
26. B.L. Hao, J. Qi, B. Wang, Mod. Phys. Lett. **2**, 103–108 (2003)
27. L. Zsak, Z. Lu, T.G. Burrage, J.G. Neilan, G.F. Kutish, D.M. Moore, D.L. Rock, J. Virol. **7**, 3066–3076 (2001)
28. T.G. Burrage, Z. Lu, J.G. Neilan, D.L. Rock, L. Zsak, J. Virol. **78**, 2445–2453 (2004)