Related strains of African swine fever virus with different virulence: genome comparison and analysis

Raquel Portugal,1 João Coelho,2 Dirk Höper,1 Nicole S. Little,4 Chad Smithson,4 Chris Upton,4 Carlos Martins,2 Alexandre Leitão2,3 and Günther M. Keil1

Correspondence
Raquel Portugal
raquel.portugal@fli.bund.de

1Institut für molekulare Virologie und Zellbiologie, Friedrich-Loeffler-Institut, Greifswald-Insel Riems 17493, Germany
2Centro de Investigação Interdisciplinar em Sanidade Animal (CIISA), Faculdade de Medicina Veterinária, Universidade de Lisboa, Avenida da Universidade Técnica, Lisboa 1300-477, Portugal
3Instituto de Investigação Científica Tropical, CVZ, FMV, Avenida da Universidade Técnica, Lisboa 1300-477, Portugal
4Department of Biochemistry and Microbiology, University of Victoria, Victoria BC, Canada

Two strains of African swine fever virus (ASFV), the high-virulence Lisboa60 (L60) and the low-virulence NH/P68 (NHV), which have previously been used in effective immunization/protection studies, were sequenced. Both were isolated in Portugal during the 11-year period after the introduction of ASFV to the European Continent in 1957. The predicted proteins coded by both strains were compared, and where differences were found these were also compared to other strains of known virulence. This highlighted several genes with significant alterations in low-virulence strains of ASFV that may constitute virulence factors, several of which are still uncharacterized regarding their function. Phylogenetic analysis grouped L60 and NHV closest to other P72 genotype I ASFV strains from Europe and West Africa, consistent with the assumed West African origin of all European strains. Interestingly, a relatively lower genomic identity exists between L60 and NHV, both isolated in a similar geographical location 8 years apart, than with other European and west African strains isolated subsequently and in more distant locations. This may reflect the intensive passage in tissue culture, during the early 1960s, of a Portuguese isolate to obtain an attenuated vaccine, which may have led to NHV. This study contributes to a better understanding of the evolution of ASFV, and defines additional potential virulence genes for future studies of pathogenesis towards the development of effective vaccines.

INTRODUCTION

African swine fever virus (ASFV) is the agent of African swine fever (ASF), an important and highly lethal contagious disease of domestic swine and wild boar, listed as notifiable by the World Organisation for Animal Health (OIE). It still lacks an effective vaccine for prevention, or measures for treatment. Disease control is based on culling of animals, and control of trade and animal movements, with severe socio-economic impacts on affected countries (Costard et al., 2013a).

ASFV also infects African wild suids (warthogs, bushpigs) which do not develop the disease but can act as asymptomatic carriers (Jori et al., 2013). Furthermore, soft ticks of the Ornithodoros moubata complex also serve as a natural reservoir in Africa and transmit the disease to suids, thus contributing to the difficulty in eradicating ASF (Burrage, 2013; Costard et al., 2013b).

The disease was first identified in Africa following the introduction of European domestic swine in Kenya (Montgomery, 1921). It is now present in most of sub-Saharan Africa. Portugal had the first outbreak recorded in Europe in 1957. This was effectively controlled, but after a second introduction in 1960, the disease persisted and in the following decades spread to other European countries, and also to Cuba, Brazil, the Dominican Republic and Haiti (Sanchez-Vizcaino et al., 2012). ASF has been eradicated...
from these locations with the exception of Sardinia, Italy, where it remains endemic. The serious threat of ASF for pig husbandry worldwide triggered increased awareness after its introduction to the Georgian Republic in 2007, spreading in the following years to other Caucasian countries, the Russian Federation, Ukraine and Belarus (Gogin et al., 2013; Rahimi et al., 2010). Since early 2014, further outbreaks of ASF have been reported in Lithuania, Poland, Latvia and Belarus (OIE, 2014).

ASFV, the sole arbovirus with a dsDNA genome, is classified as the only member of the family Asfarviridae and is included in the nucleocyttoplasmic large DNA virus superfamily, or as recently proposed, the Megavirales order (Colson et al., 2013; Dixon et al., 2005). It replicates mainly in the cytoplasm of mononuclear phagocytic system cells. Depending on the strain, genome size varies between 170 and 193 kbp and is predicted to contain 150 to 167 genes, which are involved in viral replication and morphogenesis as well as in modulation of host cell functions and immune evasion (Correia et al., 2013; Dixon et al., 2004, 2013).

Circulating strains with different levels of virulence cause a broad range of clinical symptoms in susceptible domestic swine and European wild boar, ranging from peracute and acute forms, with haemorrhagic fever and death within a few days, to chronic or even inapparent disease (Blome et al., 2013). The underlying mechanisms are far from being fully understood. Animals surviving infection with low-virulence strains are only protected against highly virulent related strains, and the cellular component of the immune response is fundamental for this protection (Takamatsu et al., 2013).

Comparison of the genomes of highly and weakly virulent strains may contribute to defining virulence genes, and help in the development of strategies to create effective vaccines. The first complete genome sequence available was derived from cell culture-adapted Ba71V (Yañez et al., 1995). In more recent years, sequences from strains of different origin and genotypes have become available along with comparative analyses (Chapman et al., 2008, 2011; de Villiers et al., 2010). Most of the variation among ASFV genomes results from the presence of different numbers of multi-gene family (MGF) genes in the left and right variable regions (LVR and RVR, respectively). MGFs are characteristic of ASFV and have no obvious homology with other known genes. Five families have been recognized – MFG 100, 110, 300, 360 and 505/530 – named according to the average (mean) number of encoded aa. Often found in tandem, MGF genes are thought to have evolved by duplication (Dixon et al., 2013). Although the presence of multiple members suggests an advantage for ASFV, the function of many is still unknown. The LVR and RVR flank a highly conserved central region (CCR) containing most of the genes known to be involved in virus replication, assembly and host cell function modulation (Dixon et al., 2004, 2013; Yañez et al., 1995). The ASFV genome also possesses terminal inverted repeats (TIR) and covalently cross-linked termini (González et al., 1986; Yáñez et al., 1995).

Here we report the sequence and comparison of two ASFV strains collected 8 years apart from infected domestic pigs in similar geographical locations (central Portugal): the highly virulent L60, isolated in 1960 after its introduction from Angola (Manso Ribeiro & Azevedo, 1961), and the low-virulence NHV, isolated in 1968 from a pig presenting with chronic ASF (Vigário et al., 1974). Since NHV protects pigs after challenge with L60 (Leitão et al., 2001), these viruses provide a window on the virulence and evolution of ASFV.

**RESULTS**

**Major ORF differences between the L60 and NHV genomes**

Sequence analysis revealed 182 362 bp and 163 ORFs for L60 (accession no. KM262844) and 172 051 bp and 158 ORFs for NHV (accession no. KM262845) (Table S1, available in the online Supplementary Material), with the majority of differences being located in the LVR and RVR of the genomes.

Differences in the intergenic areas of the genomes may also have an impact on virulence, for example by affecting promoter or other regulatory sequences. However, since there is only scanty information regarding such sequence elements in ASFV, we focused on aa differences between L60- and NHV-encoded proteins (Alignment S1) and also compared these differences to other strains of known virulence, presented in Table 1. Of note, a striking similarity exists between the genomes of NHV and OURT88/3 (see below), and hence differences relating to NHV usually apply also to OURT88/3, unless otherwise stated.

The LVRs of the L60 and NHV genomes are approximately 41 and 31 kbp, respectively, delimited by the conserved A224L gene (Table S1). Numerous MGF members are located in this area (Table 2), accounting for the majority of alterations.

The first key difference occurs following approximately 6.8 kbp of sequence, with an insertion of 4458 bp occurring in NHV between MGF 110-2L and -13L, which in L60 is fused and annotated as MGF 110-13L+2L. This insertion comprises MGF 110 and 100 members – MGF 110-4L, -5L, -9L, MGF 100-1R – and ORFs 285L and 86R. Their function is currently unknown, but orthologues can be found in other ASFV strains. MGF 110-2L exists in Ba71V (U104L), with two conservative aa substitutions in its predicted protein in comparison with NHV. In Tengani62, Pretorisuskop/96/4 and Georgia2007/1, the predicted proteins have a couple of non-conservative aa substitutions in comparison with NHV or Ba71V, in positions that are conserved in the three highly virulent strains. MGF 110-4L and -5L are very similar,
indicating a recent duplication. Both have orthologues in Lisbon57 (LIS 124-1, LIS 124-2, respectively), Ba71V, Tengani62, Pretoriuskop/96/4, Georgia2007/1, Malawi Lil-20/1 and Kenya1950, among which this aa sequence is strongly conserved. A frameshift mutation in NHV leads to a C terminus 25 aa shorter, with the final 41 aa not conserved. NHV pMGF 100-1R is identical to Lisbon57, and differs from Tengani62, Pretoriuskop/96/4 and Georgia2007/1 orthologues by a single non-conserved aa. Orthologues are also present in Kenya1950 and Malawi Lil-20/1, with several non-conserved aa substitutions. p285L is found in Lisbon57, Tengani62, Pretoriuskop/96/4, Malawi Lil-20/1 and Kenya1950, with six non-conserved aa positions.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Virulence</th>
<th>Genotype (p72)</th>
<th>Origin/host</th>
<th>Reference</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ba71V</td>
<td>Avirulent</td>
<td>I §§</td>
<td>1971 Spain/Vero cell culture adapted</td>
<td>Yáñez et al. (1995)</td>
<td>NC_001659</td>
</tr>
<tr>
<td>E75</td>
<td>High</td>
<td>I §§</td>
<td>1975 Spain/Domestic pig</td>
<td>de Villiers et al. (2010)</td>
<td>FN557520</td>
</tr>
<tr>
<td>Tengani62</td>
<td>High</td>
<td>V / I §</td>
<td>1962 Malawi/Domestic pig</td>
<td>Pan (1992)</td>
<td>AY261364</td>
</tr>
<tr>
<td>Pretoriuskop/96/4</td>
<td>High</td>
<td>XX / I §</td>
<td>1996 South Africa/Tick</td>
<td>Zsak et al. (2001)</td>
<td>AY261363</td>
</tr>
<tr>
<td>Malawi Lil-20/1</td>
<td>High</td>
<td>VIII §§</td>
<td>1983 Malawi/Tick</td>
<td>Haresnape &amp; Wilkinson (1989)</td>
<td>AY261361</td>
</tr>
<tr>
<td>OURT88/3</td>
<td>Low</td>
<td>I §§</td>
<td>1988 Portugal/Tick</td>
<td>Chapman et al. (2008)</td>
<td>AM712240</td>
</tr>
<tr>
<td>Benin97/1</td>
<td>High</td>
<td>I §</td>
<td>1997 Benin/Domestic pig</td>
<td>Chapman et al. (2008)</td>
<td>AM712239</td>
</tr>
<tr>
<td>Georgia2007/1</td>
<td>High</td>
<td>II</td>
<td></td>
<td></td>
<td>2007 Georgia/Domestic pig</td>
</tr>
<tr>
<td>Lisbon57*</td>
<td>High</td>
<td>I ‰</td>
<td>1957 Portugal/Domestic pig</td>
<td>Blasco et al. (1989)</td>
<td>M58155, M34948</td>
</tr>
<tr>
<td>Mkuzi1979</td>
<td>Unknown</td>
<td>I §/ VII §</td>
<td>1979 South Africa/Tick</td>
<td>Complete genome</td>
<td>AY261362</td>
</tr>
<tr>
<td>Warthog</td>
<td>Unknown</td>
<td>IV §</td>
<td>1980 Namibia/Warthog</td>
<td>Zsak et al. (2005)</td>
<td>AY261366</td>
</tr>
<tr>
<td>Warmbaths</td>
<td>Unknown</td>
<td>III †/ I §</td>
<td>1987 South Africa/Tick</td>
<td>Zsak et al. (2005)</td>
<td>AY261365</td>
</tr>
</tbody>
</table>

*Table 1. ASFV strains used for comparison studies*

*Partial sequences of the LVR available.
†Bastos et al. (2003).
§Gallardo et al. (2009).
§de Villiers et al. (2010).
‖Rowlands et al. (2008).*

Table 2. MGF members present in the LVR and RVR of the L60 and NHV genomes

<table>
<thead>
<tr>
<th>Region</th>
<th>MGF family member</th>
<th>L60</th>
<th>NHV</th>
</tr>
</thead>
<tbody>
<tr>
<td>LVR</td>
<td>MGF 100</td>
<td>-</td>
<td>1R†</td>
</tr>
<tr>
<td></td>
<td>MGF 110</td>
<td>1L, 13L + 2L, (E75) 11L, (Benin) 12L, (Benin) 13L, 14L</td>
<td>1L, 2L*, 4L†, 5L†, 9L*, 13L§, 14L§</td>
</tr>
<tr>
<td></td>
<td>MGF 300</td>
<td>1L, 2R, 4L</td>
<td>1L, 2R, 4L</td>
</tr>
<tr>
<td></td>
<td>MGF 360</td>
<td>1L, 2L, 3L, 4L, 6L, 8L, 9L, 10L, 11L, 12L, 13L, 14L</td>
<td>1L, 2L, 3L, 4L, 8L, 9L*</td>
</tr>
<tr>
<td>RVR</td>
<td>MGF 100</td>
<td>2L</td>
<td>2L*†</td>
</tr>
<tr>
<td></td>
<td>MGF 360</td>
<td>16R, 18R</td>
<td>16R, 17R*, 18R*</td>
</tr>
<tr>
<td></td>
<td>MGF 505</td>
<td>11L</td>
<td>11L*</td>
</tr>
</tbody>
</table>

*With non-conservative aa substitutions or other significant aa sequence alteration when comparing NHV to L60 and/or other highly virulent strains.
†Absent from L60 but present and conserved in other highly virulent strains.
‡Truncated.
§Significant aa sequence alteration in comparison to OURT88/3.
mostly in the Kenya1950 sequence. However, differing aa positions in NHV p285L always have an identical counterpart in another of the highly virulent strains. p86R can be found similarly only in Lisbon57, with a couple of non-conservative aa substitutions.

The fused MGF 110-13L + 2L ORF in L60 is similar to E75 MGF 110-13L and Benin97/1 MGF 110-11L, which contain ORFs that are seven and two codons shorter, respectively, due to a homopolymeric G-stretch of varying length near the 5’ ends. Since this ORF is annotated under different names in other strains, for L60 it will be annotated as MGF 110-13L + 2L for clarity of comparison with NHV. NHV MGF 110-15L may be fused to the adjacent MGF 110-14L in comparison with OURT88/3, due to a frameshift mutation caused by the different-length homopolymeric G-stretch (9-G in OURT88/3, 10-G in NHV), leading to a stop in NHV-encoded protein 3 aa downstream of methionine. This frameshift can result in fusion of the residual pMGF 110-13L sequence to pMGF 110-14L. However, we have to admit the possibility of 454 sequencing ambiguity in such homopolymeric areas, even though the accuracy of the obtained sequence was high. A comparable fused ORF can be found in the Georgia2007/1 (MGF 110-13L) and Kenya1950 strains, but encoding several non-conservative aa substitutions.

In comparison with L60, NHV lacks the nt 7244–8634 of the former. The deletion begins after nt 11696, within the MGF 110-14L sequence, and leads to loss of three MGF 110 members from the NHV genome: E75 MGF 110-11L, Benin MGF 110-12L and Benin MGF 110-13L. Besides L60, these ORFs are present in E75 and Benin97/1, reflecting their close relation (see below), but are not always present in other strains. Of note, we detected discrepancies in the current annotation of MGF 110-11L to -14L among strains OURT88/3, Benin97/1, E75 and Georgia2007/1, where similar ORFs are termed differently or are not annotated (e.g. E75 MGF 110-13L is highly similar to Benin97/1 MGF 110-11L as stated above, but less so to Benin97/1 MGF 110-13L; an ORF similar to E75 MGF 110-11L exists in Benin97/1 but is not annotated, and in Georgia2007/1 is annotated as MGF 110-14L). A revision of the annotations in these areas of the genomes will be necessary.

Following MGF 110-14L are MGF 360-4L and -6L in the L60 genome. MGF 360-6L is absent from the NHV genome, due to a second deletion of 2173 bp (nt 13529 of NHV, nt 10465–12638 of L60). It is identical in L60, Benin97/1 and E75, and highly conserved across virulent strains of different genotypes – Pretoriuskop/96/4, Georgia2007/1, Kenya1950 and Malawi Lil-20/1. It is also absent from Ba71V.

A third deletion of 10068 bp occurs further downstream in the LVR of the NHV genome (nt 20698 of NHV, nt 19809–29877 of L60). This leads to the loss of several adjacent MGF 360 and 505 members – MGF 360-10L, -11L, MGF 505-1R, MGF 360-12L, -13L, -14L, MGF 505-2R – and a 5’ truncation of MGF 360-9L and MGF 505-3R flanking the deletion, identical to OURT88/3. A similar but slightly shorter deletion exists in the Ba71V genome. Importantly, these ORFs are present in the highly virulent strains of ASFV. Consecutive MGF 505 members -4R, -5R, -7R, -8R and -10R follow this large deletion and terminate the LVR. The encoded proteins are mostly conserved when comparing L60 and NHV, except for pMGF 505-5R and -8R, each having a single non-conservative aa difference between both strains. For NHV, neither have equivalent substitutions at the same aa positions in orthologues of other highly virulent strains, where these aa positions are always conserved.

The CCR of the genomes, starting with the A224L ORF, consists of approximately 141 kbp comprising 111 ORFs. It contains most of the ASFV genes known to be involved in viral DNA replication, virion assembly and host cell function modulation (Dixon et al., 2013; Yáñez et al., 1995). Some differences of note exist between L60 and NHV proteins in this area (Table 3). Major differences occur in two consecutive ORFs for proteins involved in haemadsorption: pEP153R/8CR, a C-type lectin-like protein, and pEP402R/8DR, a homologue of CD2-type receptors of T lymphocytes. A nt deletion close to the 5’ terminus of EP153R in NHV leads to the generation of a stop codon truncating the encoded protein. pEP402R of L60 contains a deletion of 29 aa after aa 148, located in the extracellular region, close to the predicted transmembrane domain (Rodriguez et al., 1993) and only present in this strain. The presence of this oligopeptide is thus not required for haemadsorption, since L60 displays this characteristic. In NHV, three separate nt deletions occur in this ORF in comparison with L60, with the first close to the 5’ end leading to a stop codon after 21 aa, explaining its non-haemadsorbing phenotype. In the sequence of M1249L, a large ORF of unknown function, a 1 nt deletion in the L60 sequence in a homopolymeric T-stretch, shared with M1249L of E75, may generate a stop codon after codon 110. A start codon downstream may lead to expression of a protein with 1107 aa. The sequence of this ORF is highly conserved across strains of different virulence, such as NHV, OURT88/3, Benin97/1 and Georgia2007/1, with all encoded aa substitutions in the predicted proteins being conservative. Comparison to more distant African strains such as Tenganis62, Pretoriuskop/96/4, Kenya1950 and Malawi Lil-20/1 also shows a high degree of conservation. Hence, the difference regarding L60 and E75, if not a result of sequencing ambiguity, does not seem to affect virulence. Of note, the pC147L homologue of RNA polymerase subunit 6 (Yáñez et al., 1995) has a non-conservative aa substitution in OURT88/3 that is exclusive to this strain (Ser in position 74, Phe in all other strains), which is one of the few differences between NHV. L60 and NHV pC147L have only one aa difference, but it is conservative. pB119L, a thiol oxidase, has a non-conservative aa substitution present only in NHV and OURT88/3 (Ser in position 18, Phe in all other strains), predicted to be located at the border of an z1 helix structure (Hakim & Fass,
D *Specific substitution for NHV and OURT88/3, other strains have another aa.
412 Journal of General Virology
pCP312R and pI215L contain one non-conservative aa present in OURT88/3 pCP204L), extending the encoded lead to a frameshift near the end of the NHV ORF (not homopolymeric stretch (9-A in NHV, 10-A in L60) may nalization into host cells. Of note, deletion of 1 nt in a viral phosphoprotein, P30/P32, necessary for virus inter-

3 aa pattern repetition exist in the various strains of ASFV duplicated 5 aa pattern ‘SICSD’ in NHV. Different types of pB407L differs between L60 and NHV by the presence of one only by one extra ‘TCAS’ repeat present in OURT88/3. Benin97/1

Three ORFs present differences due to repetitive aa patterns in their encoded proteins: B169L, B602L and B407L. pB169L displays three consecutive units of the aa sequence ‘PAGPK’ in NHV, and two in L60. Benin97/1 pB169L is identical to that of NHV, and hence this difference is unlikely to be related to virulence. pB602L contains tandem repetitive aa sequences (‘TCAS’, ‘TGAS’, ‘TCAD’ and ‘TNVD’) which differ in number among the different strains and have been used for virus genotyping. Comparing L60 and NHV, L60 lacks several of these repeats, shortening its protein by 80 aa. NHV and OURT88/3 differ only by one extra ‘TCAS’ repeat present in ORT88/3. pB407L differs between L60 and NHV by the presence of one duplicated 5 aa pattern ‘SICSD’ in NHV. Different types of 5 aa pattern repetition exist in the various strains of ASFV regarding this protein.

CP204L codes for a highly immunogenic early-expressed viral phosphoprotein, P30/P32, necessary for virus internalization into host cells. Of note, deletion of 1 nt in a homopolymeric stretch (9-A in NHV, 10-A in L60) may lead to a frameshift near the end of the NHV ORF (not present in OURT88/3 pCP204L), extending the encoded protein by 10 aa, a condition also found in Ba71V. pCP312R and pI215L contain one non-conservative aa substitution each, exclusive to NHV and OURT88/3, in comparison with Ba71V, L60 and other highly virulent strains. In pI215L, a ubiquitin-conjugating enzyme, the non-conservative aa substitution in NHV and OURT88/3 (aa 95, Gly rather than Glu) is 10 aa distant from the active site Cys85 (Hingamp et al., 1992).

The RVR can be considered to extend from the first MGF members present on the right side of the genomes, following the CCR. Once again, the main differences reside in MGF genes (see Table 2). MGF 360-16R starts after nt 171288 and 162483 of the L60 and NHV genomes, respectively. In L60, due to an extra nt in a homopolymeric T-stretch close to the 3’ terminus, the encoded protein is 2 aa shorter in comparison with NHV. L60 MGF 360-16R is identical to that of Benin97/1; in NHV, it is identical in OURT88/3 and E75. African strain orthologues in Tengani62, Pretoriuskop/96/4, Malawi Lil-20/1 and Kenya1950 have a 41 aa extended C terminus. In the following pMGF 505-11L, NHV and OURT88/3 have 4 aa substitutions, with one non-conservative and being unique to these strains, in a position that is conserved in all strains of known virulence except Kenya1950. L60 pMGF 505-11L is identical to E75, Benin97/1 and Ba71V (DP542L) and is an interesting MGF member, because pressure appear to exist to conserve its aa sequence. Of a total of 99 substitutions in the aligned aa sequences, only 15 are non-conservative, mainly in African strains Malawi Lil-20/1 and Kenya1950. In the following ORF, MGF 100-2L, a 1 nt deletion in NHV, results in a frameshift mutation after 55 nt and the emergence of a stop codon after 28 codons.

Table 3. Differences between L60 and NHV in regard to CCR ORFs

<table>
<thead>
<tr>
<th>ORF</th>
<th>Difference in encoded proteins</th>
<th>Functions/characteristics</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EP402R</td>
<td>Truncation in NHV</td>
<td>CD2 homologue</td>
<td>Borca et al. (1994, 1998); Rodriguez et al. (1993)</td>
</tr>
<tr>
<td>M1249L</td>
<td>Frameshift mutation in L60 – N- and C-terminus separation†</td>
<td>Contains ubiquitin-like domain</td>
<td>Iyer et al. (2006)</td>
</tr>
<tr>
<td>B119L (9GL)</td>
<td>aa – non-conservative</td>
<td>Thiol oxidase; interacts with other viral proteins</td>
<td>Hakim &amp; Fass (2009); Lewis et al. (2000); Rodriguez et al. (2006)</td>
</tr>
<tr>
<td>B169L</td>
<td>Repetitive aa pattern ‘PAGPK’ unit number</td>
<td>Predicted transmembrane protein</td>
<td>Yáñez et al. (1995)</td>
</tr>
<tr>
<td>B602L (9RL)</td>
<td>Different number of tandem repetitive aa sequences</td>
<td>CAP80, viral chaperone necessary for folding of the major viral capsid protein P72</td>
<td>Cobbold et al. (2001); Epifanio et al. (2006)</td>
</tr>
<tr>
<td>B407L</td>
<td>Duplicated ‘SICSD’ aa pattern in NHV</td>
<td>Unknown function</td>
<td>Yáñez et al. (1995)</td>
</tr>
<tr>
<td>CP204L</td>
<td>Frameshift mutation in NHV – 10 aa extended C terminus</td>
<td>P30/P32 viral phosphoprotein; involved in virus internalization</td>
<td>Afonso et al. (1992); Gómez-Puertas et al. (1998); Prados et al. (1993)</td>
</tr>
<tr>
<td>CP312R*</td>
<td>aa – non-conservative</td>
<td>Unknown function; potential serological epitope of ASFV</td>
<td>Kolnberger et al. (2002); Reis et al. (2007)</td>
</tr>
<tr>
<td>I215L*</td>
<td>aa – non-conservative</td>
<td>Ubiquitin-conjugating enzyme</td>
<td>Bulimo et al. (2000); Hingamp et al. (1995); Rodriguez et al. (1992)</td>
</tr>
</tbody>
</table>

*Specific substitution for NHV and OURT88/3, other strains have another aa.
†ORFs annotated in L60 as M1107L and M110L (corresponding to M1249L C- and N-termini, respectively).
However, an alternative start codon downstream of the original AUG may result in the translation of a 16 aa N-terminal truncated protein with an additional Val after the Met, but which is otherwise identical to the remaining L60 encoded protein. pMGF 100-2L is not only highly conserved in virulent strains, but also in Ba71V (DP141L), in which it is identical to L60. Downstream, MGF 360-18R is divided into two ORFs in NHV, similarly to OURT88/3 (Chapman et al., 2008), as a result of one extra nt in a homopolymeric C-stretch which introduces a stop codon after 72 codons. Another start codon downstream may recover the C-terminal 148 aa translation, and hence the 5′ segment of the ORF is annotated as MGF 360-17R and the 3′ segment as MGF 360-18R. L60 pMGF 360-18R is identical to that of Benin97/1. In E75 a similar ORF exists although it is annotated with another start codon further downstream, identical to NHV and OURT88/3. A revision of this annotation in E75 should be considered.

Other more distantly related strains – Georgia2007/1, Tengani62, Pretorisuskop/96/4 and Malawi Lil-20/1 – are 17 aa shorter in the N-terminus in comparison with L60, and Kenya1950 and Malawi Lil-20/1 have a 2 aa extended C-terminus. Ba71V DP148R is identical to NHV and OURT88/3 MGF 360-18R.

MGF 360-19R is the final MGF member of the sequenced L60 and NHV genomes, and the final ORF in NHV. The N terminus of the encoded protein is conserved in several strains except for Kenya1950 and Malawi Lil-20/1, where a similar ORF sequence was not detected. The C terminus is of variable variable size in the different strains, with the longest ORFs in Benin97/1, Ba71V (identical), Tenganii62 and Pretorisuskop/96/4 (363 aa), followed by NHV and OURT88/3 (303 aa and final 15 aa not conserved), L60, E75 and Georgia2007/1 (277, 275 and 269 aa, respectively). The codons for the final 15 aa at the end of the encoded protein in NHV are included in a terminal sequence of 69 nt, which is repeated in inverted orientation at the beginning of the genome (nt 275–343), hence forming part of the TIR. In L60 the right end of the genome extends further in comparison with NHV, with an extra 1496 bp being sequenced in our conditions. The last sequenced ORF in L60 was BA71V-DP60R, which is already part of the TIR sequence.

**Genome comparison and phylogenetic analysis**

Alignments performed between the sequences of L60, NHV and other ASFV strains for which full genome sequences were available revealed 98.17 % identity between L60 and NHV at the nt level (Table 4). NHV has a striking resemblance to OURT88/3, isolated 20 years later from infected Ornithodorus erraticus ticks collected in Ourique, Portugal, with which it shares 99.99 % identity. L60 conversely has the greatest identity with E75 (99.97 %), isolated in Lerida, Spain in 1975. When considering the predicted proteins, NHV and L60 are 99.65 % identical (Table 5), implying that many of the existing genomic differences are silent or reside in non-coding DNA. L60 has the higher degree of identity with Benin97/1 (99.9 %).

![Table 4. Identity % between ASFV strains in global alignments; identity at the nt level of the 14 genomic sequences available.](http://vir.sgmjournals.org)
A phylogenetic analysis of the 14 sequenced genomes clearly shows separate branching between most ASFV strains from southern and eastern Africa (a cluster formed by Malawi Lil/20/1 and Kenya1950, almost as an out-group, and another cluster grouping Tengani62, Pretorisuskop96/4, Warthog and Warmbaths), and European (NHV, OURT88/3, Ba71V, E75, L60 and Georgia2007/1), West African (Benin97/1) and one South African strain (Mkuzi1979) (Fig. 1a). There is also much less variation in this latter group, mainly including p72 genotype I genomes, with the exception of the more separated Georgia2007 genotype II and Mkuzi1979 genotype I/VII (see Table 1). The phylogenetic tree also confirms what was observed in terms of identity between the genomes presented in Tables 4 and 5, with L60 grouping almost indistinguishably with E75 and Benin97/1, and NHV with OURT88/3.

To obtain a clearer separation in the upper branch of the phylogenetic tree between the genotype I strains, considering the variation in identity percentages from nt to aa (Tables 4 and 5), we used the predicted aa sequences from the annotated ORFs concatenated into a single sequence (Georgia2007/1 and Mkuzi1979 were also used as an out-group). These were aligned and a similar phylogenetic analysis was performed, in which we were indeed able to obtain a higher resolution between NHV and OURT88/3 and among L60, E75 and Benin97/1 (Fig. 1b). In this tree, L60 branches before Benin97/1 and E75 and is more similar to Benin97/1, as confirmed by the identity values (99.90 % relative to Benin97/1, 99.58 % relative to E75).

**DISCUSSION**

In this work, we present the full genome sequences of two ASFV strains of different virulence, L60 and NHV. L60 was obtained during the ASF outbreak in Portugal in 1960 after introduction from Angola, and was the origin of the widespread epidemic that lasted in the Iberian Peninsula until the 1990s and still persists in Sardinia. NHV is a non-haemadsorbing low-virulent strain isolated in 1968 from a pig presenting with a chronic form of ASF. However, considering the high level of nt identity between L60 and strains obtained in 1997 from Benin, West Africa (99.95 %), or in 1975 from Lerida, Spain (99.97 %), the identity observed between L60 and NHV (98.17 %) is comparably low considering that only 8 years separate their isolation.

Table 5. Identity (in percentages) between ASFV strains in global alignments; identity between concatenated sets of translated ORFs from eight of the strains.

<table>
<thead>
<tr>
<th></th>
<th>NHV</th>
<th>OURT88/3</th>
<th>BA71V</th>
<th>L60</th>
<th>Benin97/1</th>
<th>E75</th>
<th>Mkuzi1979</th>
<th>Georgia2007/1</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHV</td>
<td>99.97</td>
<td>99.97</td>
<td>99.50</td>
<td>99.65</td>
<td>99.57</td>
<td>99.35</td>
<td>98.05</td>
<td>96.82</td>
</tr>
<tr>
<td>OURT88/3</td>
<td>99.97</td>
<td>99.97</td>
<td>99.50</td>
<td>99.65</td>
<td>99.57</td>
<td>99.35</td>
<td>98.05</td>
<td>96.82</td>
</tr>
<tr>
<td>BA71V</td>
<td>99.50</td>
<td>99.49</td>
<td>99.49</td>
<td>99.76</td>
<td>99.76</td>
<td>99.61</td>
<td>98.61</td>
<td>96.70</td>
</tr>
<tr>
<td>Benin97/1</td>
<td>99.57</td>
<td>99.57</td>
<td>99.71</td>
<td>99.71</td>
<td>99.90</td>
<td>99.61</td>
<td>98.35</td>
<td>97.11</td>
</tr>
<tr>
<td>E75</td>
<td>99.35</td>
<td>99.34</td>
<td>99.38</td>
<td>99.58</td>
<td>99.61</td>
<td>98.61</td>
<td>98.05</td>
<td>96.70</td>
</tr>
<tr>
<td>Mkuzi1979</td>
<td>98.05</td>
<td>98.05</td>
<td>98.47</td>
<td>98.37</td>
<td>98.35</td>
<td>98.05</td>
<td>98.05</td>
<td>96.88</td>
</tr>
<tr>
<td>Georgia2007/1</td>
<td>96.82</td>
<td>96.90</td>
<td>97.05</td>
<td>97.10</td>
<td>97.11</td>
<td>96.70</td>
<td>96.88</td>
<td>96.88</td>
</tr>
</tbody>
</table>

**Fig. 1.** Phylogenetic comparison of L60, NHV and other ASFV strains. Maximum-likelihood phylogenetic trees were reconstructed from (a) a multiple nt sequence alignment of 14 complete genomic sequences and (b) a multiple aa sequence alignment of concatenated protein sets of eight different ASFV strains. Node values show percentage bootstrap support. Scale bars represents an estimation of the percentage of aa substitutions per site.
This divergence may be attributed to the vaccination of pigs in 1962 with a virus passaged 150 times in bone marrow primary cultures (Manso Ribeiro et al., 1963), which may have contributed to the development of NHV. Nevertheless, both strains are related since vaccination with NHV confers protection from subsequent lethal infection with L60 (Leitão et al., 2001). Interestingly, only minute differences are found between NHV and OURT88/3, obtained from ticks 20 years later in the south of Portugal (Boinas et al., 2004), and NHV also replicates in ticks (Rowlands et al., 2009), thus being well adapted to a natural cycle of transmission. It seems to be the most direct ancestor of OURT88/3. This study also confirms that most European strains of p72 genotype I, with the exception of the recently introduced Georgia2007/1 (genotype II), probably derive from a single-source introduction of ASFV into Portugal and the European Continent.

An analysis of the differences between the genomes of related ASFV strains of different virulence should give clues concerning factors important for virulence. Our comparison of the L60 and NHV genomes focused on aa differences between translated ORFs, especially deletions and non-conservative substitutions which may change the protein properties, making them good candidates for in-depth studies with respect to their role in pathogenesis and immune escape. The majority of differences found between both strains are located in the LVRs of the genomes, resulting mostly from MGF gene rearrangements. One of the main differences is an insertion in the NHV genome comprising MGF 100-1R and MGF 110-4L. -5L and -9L genes, together with ORFs 285L and 86R, all of unknown function. Interestingly, these are present in another highly virulent Portuguese strain, Lisbon57, isolated during the first European outbreak of ASF. However, the full sequence of Lisbon57 is not yet available, and the limited information regarding this isolate does not allow further interpretations. MGF-4L, -5L, MGF 100-1R and 285L do not seem to correlate directly with lower virulence, since they are also present and conserved in other highly virulent strains. On the other hand, significant differences exist in the pMGF 110-2L, -9L and, to a lesser extent, p86R, which could lead to lower virulence. A study with isolates lacking all MGF 110 members suggested that these may neither be essential for replication nor have a role in virulence (Aguiero et al., 1990). However, this could depend on the genomic context and may be different in other strains.

Due to deletions in the LVR of the genome, NHV loses MGF 360-6L, which is highly conserved across virulent strains of different genotypes and is also absent from Ba71V and OURT88/3. It may constitute another virulence factor, but its function still lacks characterization. A larger deletion of approximately 10 kb further downstream in the NHV genome, similar to OURT88/3, leads to loss or truncation of several MGF 505 and 360 genes: MGF 360-9L, -10L, -11L, MGF 505-1R, MGF 360-12L, -13L, -14L, MGF 505-2R and MGF 505-3R. Several of these have been implicated in replication in macrophages, virulence, tick host infection and type I IFN immune response (Afonso et al., 2004; Burrage et al., 2004; Neilan et al., 2002; Zsak et al., 2001). In Vero cell-adapted Ba71V, also with a deletion of 8.2 kb in this area of the genome, replication in macrophages was restored after the introduction of consecutive 3CL, 3DL and 3EL (MGF 360-9L, -10L and -11L, respectively); 3FR (MGF 505-1R); and 3HL, 3IL and 3LL (MGF 360-12L, -13L and -14L, respectively) (Zsak et al., 2001). When the same stretch of genes, and 3NR (MGF 505-2R/A489R), were deleted from Pretoriuskop/96/4 (Pr4A35), growth in macrophages was impaired (Zsak et al., 2001). The same ORFs are absent or truncated in the NHV genome, even with additional genes adjacent to this area deleted or truncated, yet this strain replicates effectively in swine macrophages (Gil et al., 2008; Keil et al., 2014). Further genomic factors have to be contemplated. Ba71V has an extra deletion of 3.2 kb in the RVR of its genome, resulting in loss of several ORFs conserved in all natural strains of ASFV sequenced to date (Chapman et al., 2008), and also in the present L60 and NHV -17L, -18L, -19R, -110L and -111L. These may have an important role for replication in macrophages and/or virulence, but they are still mostly uncharacterized, except for 110L encoding an extra copy of structural viral protein p22, similar to KP177R present in the LVR (Dixon et al., 1994). Other differences existing between the Ba71V and NHV or OURT88/3 genomes, principally in the LVRS, can cause defective replication: ORFs absent from the Ba71V genome, ORFs with non-conservative aa substitutions or deletions, and ORFs absent from NHV (Table 6).

The MGF 360 and 505 genes also have a role in suppressing type I IFN responses, since the Pr4A35 strain induced a strong type I IFN-related transcriptional response in pig macrophages, contrarily to the parental Pretoriuskop/96/4, and supernatants of infected macrophages with Pr4A35 contained IFNz (Afonso et al., 2004). This was confirmed in macrophages infected with L60 and NHV, where higher levels of IFNz were produced in the early stages of infection with NHV than with L60, in addition to other cytokines involved in inflammatory responses and cellular immunity (Gil et al., 2008). Pr4A35 is also attenuated in swine (Afonso et al., 2004). NHV and OURT88/3 have the same MGF 505 and 360 genes deleted or truncated, supporting their effect in immunosuppression and virulence. It would be interesting to restore such genes in these strains and verify the resulting virulence phenotype.

In the CCR, the main differences occur in genes involved in haemadsorption. As referred to above for OURT88/3 (Chapman et al., 2008), NHV is also non-haemadsorbing, with similar deletions as in OURT88/3 for both the lectin-like EP153R and CD2-like EP402R. Haemadsorption is not directly related to virulence, since there are virulent field strains lacking this feature (Gonzague et al., 2001; Pan & Hess, 1984), and deletion of EP402R (8DR) or EP153R (8CR) from Malawi Lil-20/1 did not reduce virulence for domestic pigs, although deletion of 8DR caused a delay in viral spread (Borca et al., 1998; Neilan et al., 1999). Interestingly, haemadsorption enhanced replication in ticks, as observed when it was restored in the NHV strain.
but it did not restore virulence for pigs (Rowlands et al., 2009). Both proteins exert other functions: EP402R has immunosuppressor activity (Dixon et al., 2004) and EP153R inhibits apoptosis and down-modulates MHC I/SLA I membrane expression (Hurtado et al., 2004, 2011). Their simultaneous absence from NHV, as well as OURT88/3, may contribute to the attenuated phenotypes. Also in the CCR, pB119L viral thiol oxidase, pI215L ubiquitin-conjugating enzyme and pCP312R of unknown function all contain non-conservative aa substitutions that are exclusive to OURT88/3 and NHV. B119L is necessary for correct virion maturation and replication in macrophages, and its deletion from Malawi Lil-20/1 led to attenuation in swine (Lewis et al., 2000). The aa substitutions in both pB119L and pI215L are located outside their functional domains, but it is possible that they alter the protein structure, interfering with its function in low-virulence strains. Serological studies identified pCP312R as an antigen of ASFV (Reis et al., 2007), and the difference in its sequence could have consequences for immunogenicity.

Overall, our comparison of the genomes of two related strains of ASFV with differing virulence confirmed or highlighted further genes that may be of importance in virulence: (i) LVR genes present in the low-virulence strains NHV and OURT88/3 without conserved counterparts in L60 or other highly virulent strains – MGF 110-9L, -13L, -14L; MGF 360-4L; X64R; MGF 300-4L (N terminus similar to J182L, C terminus similar to J104L); MGF 360-9L (125 aa C terminus corresponds to Ba71V-A125L); MGF 505-8R; (ii) CCR B119L thiol oxidase, pI215L ubiquitin-conjugating enzyme and antigenic CP312R, with non-conservative aa substitutions present only in NHV and OURT88/3; (iii) RVR MGF 505-11L and MGF 100-2L, of unknown function, with significant differences shared by NHV and OURT88/3 when compared with L60 and other highly virulent strains. These genes constitute interesting targets for future studies on the pathogenesis and evasion mechanisms of ASFV.

**METHODS**

**Cells and viral infections.** Primary swine macrophage cultures, obtained as previously described (Portugal et al., 2009), were used to propagate L60 and NHV for up to a maximum of eight passages.

**DNA extraction.** DNA was extracted from macrophage cultures at 42 h post-infection with L60 or NHV by ‘salting-out’ to minimize DNA shearing, as described (Schmitt & Keil, 1996).

**Long PCR amplification of genomic segments.** Primers were designed using the Ba71V genome for amplification of a series of 10–21 kb segments (Table S2), each overlapping by at least 100 bp and covering the entire genome with the exception of the TIRs, which were only partially sequenced. Optimal annealing temperatures were determined by gradient PCR. Fifty nanogram aliquots of whole infected-cell DNA were used in 100 μl PCR with 300 mM of each primer, using high-fidelity KOD Xtreme Hot Start DNA Polymerase (Novagen) according to the manufacturer’s instructions. Amplified products were size separated by 0.6 % agarose gel electrophoresis, and amplicons of correct size were purified and quantified.

**Genome sequencing.** Equimolar amounts (0.03 pmol) of consecutive PCR segments were pooled to cover three areas of the genomes: (i) the left-hand 50 kb; (ii) the central 50–110 kb; (iii) to the right of 110 kb. Each pool of PCR products was sequenced using 454 technology (Life Sciences). Small sequence gaps and regions of uncertain sequence, mainly due to homopolymeric tracts, were sequenced via conventional Sanger methodology.

**Sequence annotation, comparative analysis and phylogeny.** The novel ASFV genomes were annotated with the aid of Genome Annotation Transfer Utility software (Tcherepanov et al., 2006), using the OURT88/3 strain as a reference and following the ORF nomenclature scheme proposed by Chapman et al. (2008). Analysis and

<table>
<thead>
<tr>
<th>ORF</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absent from Ba71V</td>
<td>L60L; 285L; 86R; MGF 100-1R; MGF 110-9L; -13L; -14L; MGF 360-4L; X64R; MGF 300-4L (N terminus similar to J182L, C terminus similar to J104L); MGF 360-9L (125 aa C terminus corresponds to Ba71V-A125L); MGF 505-8R</td>
</tr>
<tr>
<td>With non-conservative substitutions or deletions</td>
<td>DP86L (11 aa shorter C terminus in NHV); KP93L (aa 61-71 deletion in NHV); MGF 360-1L/KP360L; MGF 110-4L/KP124L; MGF 300-1L/KP124L; MGF 300-2R/KP124L (6 aa shorter C terminus in Ba71V); MGF 505-5R/A498R, MGF 505-7R/A528R, MGF 505-9R/A506R</td>
</tr>
<tr>
<td>Absent from NHV</td>
<td>V82L; Y118L (MGF 110); UP60L (MGF 360); J104L, J182L (MGF 300); A125L (MGF 360); A489R (MGF 505)</td>
</tr>
</tbody>
</table>
comparison of genome sequences was performed using Jalview
(Waterhouse et al., 2009), Base-By-Base (Hillary et al., 2011) and viral
Genome Organizer (Upton et al., 2000). CLUSTAL Omega was used for
protein sequence alignments (Sievers et al., 2011). For phylogenetic
analysis, whole-genome DNA alignments were done using CLUSTAL_X
(Larkin et al., 2007), with default parameters. Concatenated protein
sequences were aligned using CLUSTAL_W2, available at http://www.
ebi.ac.uk/Tools/msa/clustalw2/ (Goujon et al., 2010), also with default
parameters, and further subjected to manual optimization using Jalview.
ProtTest 3.0 and jModelTest 2.1.2 (Darriba et al., 2011, 2012)
were used to select the best model for phylogenetic tree reconstruc-
tion based on protein and DNA alignment, respectively. Maximum-
likelihood (ML) trees with 1000 bootstrap replicates were constructed
using a combination of PhyML (Guindon et al., 2010), with settings
indicated by previous determination of the optimum phylogenetic
model, and the tools SEQBOOT, CONSENSE and retree in PHYLIP package
version 3.695. Trees were edited using MEGA 5 (Tamura et al., 2011).

ACKNOWLEDGEMENTS

We thank K. Giesow for expert technical assistance. This work was
funded through EU Project no. 211691, FP7-KBEE-2007-1-3-05 and the
Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ),
contract no. 81157484.

REFERENCES

Afonso, C. L., Alcaraz, C., Brun, A., Sussman, M. D., Onisk, D. V.,
highly antigenic membrane and secreted protein of African swine
fever virus. Virology 189, 368–373.

Afonso, C. L., Piccone, M. E., Zaffuto, K. M., Neilan, J., Kutish, G. F.,
African swine fever virus multigene family 360 and 530 genes affect

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.

Bastos, A. D., Penrith, M. L., Crucière, C., Edrich, J. L., Hutchings, G.,
field strains of African swine fever virus by partial p72 gene

Blasco, R., de la Vega, I., Almazán, F., Agüero, M. & Viñuela, E.
near the ends of the viral DNA. Virology 173, 251–257.

swine fever in domestic pigs and European wild boar. Virus Res 173,
122–130.

Characterization of pathogenic and non-pathogenic African swine
fever virus isolates from Ornithodoros erraticus inhabiting pig

Borca, M. V., Kutish, G. F., Afonso, C. L., Irusta, P., Carrillo, C., Brun,
gene with similarity to the T-lymphocyte surface antigen CD2

Borca, M. V., Carrillo, C., Zsak, L., Laegreid, W. W., Kutish, G. F.,
like gene, 8-DR, from African swine fever virus affects viral infection

protein binds to the African swine fever virus encoded ubiquitin

Burrage, T. G. (2013). African swine fever virus infection in

African swine fever virus multigene family 360 genes affect virus
replication and generalization of infection in Ornithodoros porcinus

Comparison of the genome sequences of non-pathogenic and

Chapman, D. A., Darby, A. C., Da Silva, M., Upton, C., Radford, A. D. &

chaperone specialized for folding of the major capsid protein of

Colson, P., De Lamballerie, X., Yutin, N., Asgari, S., Bigot, Y.,
authors (2013). “Megavirales”, a proposed new order for eukaryotic

Correia, S., Ventura, S. & Parkhouse, R. M. (2013). Identification and
utility of innate immune system evasion mechanisms of ASFV. Virus
Res 173, 87–100.

Costard, S., Jones, B. A., Martinez-López, B., Mur, L., de la Torre, A.,
Martínez, M., Sánchez-Vizcaino, F., Sánchez-Vizcaino, J. M., Pfeiffer,
the European Union through illegal importation of pork and pork

Costard, S., Mur, L., Lubroth, J., Sanchez-Vizcaino, J. M. & Pfeiffer,

fast selection of best-fit models of protein evolution. Bioinformatics
27, 1164–1165.

jModelTest 2: more models, new heuristics and parallel computing.
Nat Methods 9, 772.

de Villiers, E. P., Gallardo, C., Arias, M., da Silva, M., Upton, C.,
complete African swine fever virus genome sequences. Virology
400, 128–136.

Dixon, L. K., Twigg, S. R., Baylis, S. A., Vydelingum, S., Bristow, C.,
kbp region from the right end of the genome of a pathogenic African

Dixon, L. K., Abrams, C. C., Bowick, G., Goatley, L. C., Kay-Jackson,
African swine fever virus proteins involved in evading host defence

135–143. Edited by C. M. FAUQUET, M. A. MAYO, J. MANILOFF,


Epifano, C., Krijnse-Locker, J., Salas, M. L., Rodriguez, J. M. & Salas,


Fast, scalable generation of high-quality protein multiple sequence alignments using CLUSTAL Omega. Mol Syst Biol 7, 539.


