Virology 400 (2010) 128–136

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Phylogenomic analysis of 11 complete African swine fever virus genome sequences

Etienne P. de Villiers [a,*], Carmina Gallardo [b], Marisa Arias [b], Melissa da Silva [c], Chris Upton [c], Raquel Martin [b], Richard P. Bishop [a]

[a] International Livestock Research Institute, PO Box 30709, Nairobi 00100, Kenya
[b] EU reference Laboratory for ASF, CISA-INIA, Crta Algete el Cesar s/n 28130 Valdeolmos, Madrid, Spain
[c] Biochemistry and Microbiology, University of Victoria, Victoria, BC, Canada V8W 3P6

ARTICLE INFO

Article history:
Received 4 November 2009
Returned to author for revision
22 November 2009
Accepted 13 January 2010
Available online 20 February 2010

Keywords:
African swine fever virus
Comparison
Phylogenetics
Full-genome sequences

ABSTRACT

Viral molecular epidemiology has traditionally analyzed variation in single genes. Whole genome phylogenetic analysis of 123 concatenated genes from 11 ASFV genomes, including E75, a newly sequenced virulent isolate from Spain, identified two clusters. One contained South African isolates from ticks and warthog, suggesting derivation from a sylvatic transmission cycle. The second contained isolates from West Africa and the Iberian Peninsula. Two isolates, from Kenya and Malawi, were outliers. Of the nine genomes within the clusters, seven were within p72 genotype 1. The 11 genomes sequenced comprised only 5 of the 22 p72 genotypes. Comparison of synonymous and non-synonymous mutations at the genome level identified 20 genes subject to selection pressure for diversification. A novel gene of the E75 virus evolved by the fusion of two genes within the 360 multicopy family. Comparative genomics reveals high diversity within a limited sample of the ASFV viral gene pool.

Introduction

African swine fever (ASF) is an acute, highly contagious and often fatal disease of domestic pigs (Hess, 1981; Ley et al., 1984) caused by African swine fever virus (ICTVdB Type Species 00.002.0.01.001 [ASFV]). This is a large cytoplasmic virus and is the only currently recognized member of the family Asfarviridae (Dixon 1988; Dixon et al. 2004). The genome is a single molecule of linear double-stranded DNA (dsDNA) and is between approximately 170 and 190 kbp in size, depending on the isolate. The primary reservoir of the virus is probably soft ticks of the genus *Ornithodoros* and ASFV is the only known arbovirus with a DNA genome. Wild African suids, most importantly warthogs and bush pigs can be infected but do not exhibit clinical symptoms (reviewed by Penrith et al. 2004). The epidemiology of ASF is complex and varies according to location, both a sylvatic cycle involving ticks associated with wild suids and direct pig to pig transmission are important in different regions. Acute disease caused by the virus is characterized by high fever, hemorrhages in the reticuloendothelial system, and a high mortality rate. Infectious virus can survive for several months in fresh and salted dried meat products. Montgomery first formally described the ASFV from Kenya in 1921 (Montgomery 1921). Recent studies indicate that the molecular diversity of the virus, as defined by partial sequencing of the major capsid protein p72 (Bastos et al. 2003), appears to be highest in central and eastern Africa (Lubisi et al. 2005). The disease spread from West Africa to the Iberian Peninsula, initially to Portugal in 1957 and 1960, and subsequently to several other countries in Europe and Latin America (reviewed by Penrith et al. 2004). In 2007 there was also a serious outbreak in Georgia, subsequently spreading to adjacent countries including Russia, which appears to have originated in south east Africa (Rowlands et al. 2008). These outbreaks have been extremely costly, since there is no current control measure, other than slaughter of infected pig herds. In Europe the disease remains endemic in Sardinia. ASFV is also endemic in most countries of Sub-Saharan Africa, where it is widely believed to constrain the development of the smallholder pig industry (Penrith et al. 2004).

The first complete ASFV genome was generated from the avirulent VERO cell culture adapted isolate BA71V (Yanez et al. 1995). The genome sequences of a virulent isolate from Benin and an avirulent tick isolate from Portugal were recently determined and a detailed comparison of the virulent Benin and the two avirulent isolates, derived from the p72 gene sequence group I was presented (Chapman et al. 2008). Chapman et al. (2008) also briefly describe automated annotation of seven additional complete genomes of southern and eastern African origin that had previously been deposited in GenBank. These seven isolates are described in Table 1, and the preliminary annotation is available on a publicly accessible website (http://athena.bioc.uvic.ca/database.php?db=asfarviridae).

We have assembled and annotated the complete genome sequence of E75, a second virulent isolate classified within p72 genotype I, originating from Spain. This was achieved using raw

* Corresponding author. Fax: +254 20 4223001.
E-mail address: e.villiers@cgiar.org (E.P. de Villiers).

0042-6822/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2010.01.019

Table 1
Summary of ASF virus genomes used in the study.

Abbreviation	GenBank accession	Strain name	Genome name	Genome size (bp)	No. ORFs	Notes[a]
ASFV-BA71V	NC_001659	BA71V	African swine fever virus strain BA71V	170101	160	Yanez et al. 1995; Country: "Spain"; Tissue culture adapted
ASFV-Benin97/1	AM712239	Benin97	African swine fever virus strain ASFV-Benin97	182284	156	Chapman et al. 2008; Country: "Benin"; Host: "Domestic pig"; Virulence High
ASFV-Ken	AY261360	Kenya 1950	African swine fever virus strain Kenya 1950	193886	161	Zsak et al. 2005; Country: "Kenya"; Host: "Domestic pig"; Virulence High
ASFV-Mal	AY261361	Malawi Lil-20-1 1983	African swine fever virus strain Malawi Lil-20-1 1983	187612	160	Haresnape and Wilkinson 1989; Country: "Malawi"; Host: "Tick"; Virulence High
ASFV-Mku	AY261362	Mkuzi 1979	African swine fever virus strain Mkuzi 1979	192714	167	Zsak et al. 2005; Country: "Zululand"; Host: "Tick"; Virulence Unknown
ASFV-OurT88/3	AM712240	OurT88_3	African swine fever virus strain ASFV-OurT88_3	171719	157	Boinas et al. 2004; Country: "Portugal"; Host: "Tick"; Virulence Low
ASFV-Pret	AY261363	Pretorisuskop-96-4	African swine fever virus strain Pretorisuskop-96-4	190324	167	Zsak et al. 2005; Country: "Republic of South Africa"; Host: "Tick"; Virulence High
ASFV-Teng	AY261364	Tengani62	African swine fever virus strain Tengani62	185689	162	Pan 1992; Country: "Malawi"; Host: "Domestic pig"; Virulence High
ASFV-War	AY261366	Warthog	African swine fever virus strain Warthog	186528	164	Zsak et al. 2005; Country: "Namibia"; Host: "Warthog"; Virulence Unknown
ASFV-Warm	AY261365	Warmbaths	African swine fever virus strain Warmbaths	190244	167	Zsak et al. 2005; Country: "Republic of South Africa"; Host: "Tick"; Virulence Unknown
ASFV-E75	FN557520	E75	African swine fever virus strain Spanish isolate	181187	166	This study; Country: "Spain"; Host: "Domestic pig"; Virulence High

[a] Notes indicate reference for virus, geographical origin, host and virulence of virus isolates.

sequence data generated by a commercial company and represents the first ASFV genome to be determined using 'next generation' pyrosequencing. We provide a phylogenetic analysis of the 11 publicly available complete ASFV genomes and contrast the results with genotyping based on sequence data from three single copy ASF genes, p72, p54 and central variable region (CVR) within the *B602L* gene.

Results

Sequencing of the ASFV E75 genome

The ASF Spanish isolate E75 (ASFV-E75) is a virulent and highly infective hemadsorbing virus, isolated from domestic pigs during outbreaks that occurred in Lerida (Spain) in 1975. The genome was sequenced using a 454 Life Sciences GS-20 sequencer and assembly produced a genome with 64× coverage and following PCR to fill in small gaps, the complete assembly resulted in a 181,187 bp genome. The genome was annotated and compared to the other sequenced ASFV isolates that were available in GenBank (see Table 1 for the origin of these sequences).

Identification of a core set of common genes among ASFV isolates

Concatenating large multigene datasets to improve the accuracy of phylogenetic inference is an accepted technique (Gontcharov et al. 2004; Rokas et al. 2003; Sanderson et al. 2003). We determined the core set of orthologous genes for each of the 11 ASF virus genomes using OrthoMCL (Li et al. 2003). A set of 123 orthologous ORFs comprising all conserved single copy genes and members of paralogous gene families common to all 11 genomes were identified. There was a high degree of conservation among these 123 genes, with an average BlastP similarity of 92%. The core set of 123 genes were regarded as orthologous and concatenated for each genome to create an input for phylogenetic analysis. The VOCs database identified 118 orthologous genes from the 11 genomes analyzed, compared to 120 genes when only 2 genomes are used. Chapman et al. (2008) reported 109 conserved genes in their study with 10 genomes. VOCs groups orthologous genes into families based on BLASTP scores set by a human database curator and this might explain the discrepancy in the number of conserved genes identified in the two studies. The larger number of orthologous genes predicted by OrthoMCL compared to VOCs are due to the fact that it was designed to identify both orthologous and paralogous genes which is not the case with VOCs.

Phylogenetic analyses of ASFV isolates

In order to determine the genetic relationship at the whole genome level between the ASFV isolates, we performed multiple sequence alignments of the concatenated core conserved set of genes from the 11 ASFV isolates. Several studies have shown that a concatenated multi gene approach can resolve ambiguities in phylogenetic reconstructions based on single genes (Gontcharov et al. 2004; Rokas et al. 2003). The amino acid sequences of the core set of orthologous genes from each of the 11 ASFV isolates were therefore concatenated into a single pseudo-sequence, and a neighbor-joining phylogenetic tree was constructed from a multiple amino acid sequence alignment of the concatenated sequences (Fig. 1A). This tree topology was used in subsequent analyses of the genetic relatedness of the isolates.

The phylogenetic tree analysis derived from the 123 concatenated genes separated the viruses into two major clusters that correlate with their geographical distribution. One cluster consists of four closely related isolates from West Africa and the Iberian Peninsula classified within p72 genotype group I (Bastos et al. 2003). This clade is in not visually obvious as a discrete cluster due to the geometry of the tree in which the four isolates form two sub-clusters located at the top and bottom, respectively (Fig. 1A). However according to the genetic distance these four isolates are very close to one another when compared to the other seven genomes. ASFV-Benin97/1 and ASFV-E75, whose sequence was determined in this study, form one sub-cluster while the culture adapted Spanish laboratory strain ASFV-BA17V and the non-pathogenic *Ornithodoros erraticus*-tick derived isolate ASFV-OURT88/3 from Portugal represented a second sub-cluster. The second major cluster consists of several Southern African ASFV isolates. Two South African tick-derived isolates (ASFV-Warm and ASFV-Pret) together with a warthog isolate from Namibia (ASFV-War) and the Tengani porcine isolate (ASFV-Ten) from Malawi form a single sub-cluster. A domestic pig isolate from Kenya (ASFV-Ken) and a tick-derived isolate from Malawi (ASFV-Mal) appear to be outliers and are not very closely related in terms of the overall phylogenomic analysis. Bootstrap support for the 123 gene concatenated tree was