Journal of General Virology (1995), **76**, 729–740. *Printed in Great Britain*

A set of African swine fever virus tandem repeats shares similarities with SAR-like sequences

F. Almazán, J. R. Murguía, J. M. Rodríguez, I. de la Vega and E. Viñuela*

Centro de Biología Molecular 'Severo Ochoa' (CSIC-UAM), Facultad de Ciencias, Universidad Autónoma, Cantoblanco, 28049 Madrid, Spain

A group of cross-hybridizing DNA segments contained within the *Eco*RI restriction fragments U′, X and J of a Vero cell-adapted strain (BA71V) of African swine fever virus (ASFV) were mapped and sequenced. Analysis of the nucleotide sequence revealed the presence of a set of long internal repeated sequences composed of five types of tandemly repeat units of about 200 bp. These tandem repeats contain a G-rich core of 10–14 nucleotides surrounded by regions with a high A+T content distributed in oligo(dA).oligo(dT) tracts. Next to the repeated sequences we detected two related open reading frames that are members of a new multigene family (multigene family 300). Comparison of DNA sequences from several virus isolates indicated that this region undergoes frequent rearrangements leading to either duplications or deletions of the repeat units. These ASFV repeated sequences share similarities with chromosomal α satellite DNA, the scaffold-associated region and satellite III of *Drosophila*. Similar tandemly repeated sequences have not been described in other viruses.

Introduction

African swine fever virus (ASFV), a large enveloped icosahedral deoxyvirus, is the causative agent of an important disease of domestic pigs and related species of the *Suidae* family (reviewed in Viñuela, 1987; Costa, 1990). Unfortunately, although a number of strategies have been explored including immunization with inactivated virus particles, inoculation with attenuated ASFV strains and immunization with different purified virus polypeptides, reliable protection against ASFV has never been achieved (Wardley *et al.*, 1987; Escribano *et al.*, 1993). The difficulties encountered in the search for an ASFV vaccine along with the proficiency of the virus to establish persistent infections have led to speculation that ASFV might use mechanisms to counteract the host immune defences.

The ASFV genome is a single molecule of double-stranded DNA of approximately 170 kb, with covalently linked ends (Ortín *et al.*, 1979) and terminal inverted repetitions (TIR) (Sogo *et al.*, 1984; de la Vega *et al.*, 1994) similar to those of poxviruses (Wittek & Moss, 1980; Baroudy *et al.*, 1982). Crosslinks are composed of partially unpaired and A–T-rich sequences that are found as flip-flop forms at the DNA ends (González *et al.*, 1986). Also in common with the poxviruses (Moss, 1990), ASFV particles contain the enzymic machinery required for the synthesis of mature virus early mRNAs (Kuznar *et al.*, 1980; Salas *et al.*, 1981, 1986).

Two types of internal repetitions have been described within the genome of ASFV: long (over 200 bp) repeats associated with multigene families located next to the TIR at both ends of the virus genome; and short (10–50 bp) repeats detected in both intergenic and intragenic regions. So far, four multigene families have been described and analysed in detail. Multigene families 110 (Almendral *et al.*, 1990), 360 (González *et al.*, 1990) and 505 (Rodríguez *et al.*, 1994) were initially detected in the BA71V strain of ASFV; multigene family 100 was first identified in the virulent strain Malawi LIL20/1 (Vydelingum *et al.*, 1993). Genetic variation in ASFV DNA takes place mainly through deletion or addition of DNA sequences in regions located close to the genome ends (Dixon & Wilkinson, 1988; Blasco *et al.*, 1989*a*). The main variation detected when comparing different virus field isolates is a change in the number of genes belonging to multigene families. Additionally, new genes can be generated by recombination events between homologous genes (Blasco *et al.*, 1989*b*; de la Vega *et al.*, 1990; Dixon *et al.*, 1993). Nothing is known about the

* Author for correspondence. Fax +34 1 3974 799. e-mail FALMAZAN@MVAX.CBM.UAM.ES

The nucleotide sequence data reported have been deposited in GenBank under accession number U13763.

biological relevance of the multigene families in ASFV and their genetic variation. The significance of the intergenic and intragenic short tandemly repeated sequences, which have been found at different locations in the ASFV genome (Dixon *et al.*, 1990; Angulo *et al.*, 1992; Rodríguez *et al.*, 1992), remains unclear.

In this report, we describe a third group of internal repeated sequences that are not related to multigene families. This group is formed by long tandemly repeated sequences found within the *Eco*RI (R) U′, X and J restriction fragments at the left end of the BA71V genome. These repeated sequences show a number of properties that are similar to those described for the eukaryotic scaffold-associated region (SAR) of the histone gene loop, and for chromosomal α satellite DNA and satellite III of *Drosophila*. Contiguous to these repeated sequences we have detected two related open reading frames (ORFs) that are similar to genes from the multigene family 300 of the ASFV strain Malawi LIL20/1 (Yozawa *et al.*, 1994).

Methods

Cells and viruses. Vero cells were obtained from the ATCC and cultured in Dulbecco modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum. The concentration of serum was reduced to 2% during virus infections. Culture of porcine macrophages and purification of viral DNA were carried out as described (Blasco *et al.*, 1989*a*). The Vero cell-adapted ASFV strain BA71V was grown as described (Enjuanes *et al.*, 1976; Carrascosa *et al.*, 1985). Virus isolate LIS57 was originally obtained from J. Vigario and cloned in porcine macrophages as described (Blasco *et al.*, 1989*a*).

Plasmids and bacterial strains. Plasmids containing restriction fragments RJ, RU′ and RX from BA71V DNA have been described elsewhere (Ley *et al.*, 1984; Almendral *et al.*, 1990). Restriction fragments RU′, RX and *Eco*RI–*Hin*dIII (R–H) from the left end of fragment RJ, as well as the corresponding fragments from LIS57 virus DNA, were cloned into the polylinker region of phage vectors M13mp18 and M13mp19 (Messing, 1983). *Escherichia coli* JM109 was used as the host for plasmids and bacteriophage M13.

Southern blot hybridization. DNA samples were digested with restriction endonucleases, subjected to electrophoresis in agarose gels and transferred to nylon membranes following standard procedures (Southern, 1975). DNA probes were labelled with [α-^{32}P]dATP by nick-translation (Rigby *et al.*, 1977) to a specific activity of 10^8 c.p.m./μg. Hybridizations were done under either stringent (68 °C) or relaxed (37 °C and 20% formamide) conditions as described elsewhere (Almendral *et al.*, 1990).

DNA sequencing and computer analysis. DNA sequencing was carried out on single-stranded DNA templates by the dideoxynucleotide chain termination method (Sanger *et al.*, 1977). The M13 sequencing primer and oligonucleotide primers corresponding to inner sequences were used. Nucleotide sequences were determined on both strands at least once. Analysis of DNA sequences was performed with the software package of the University of Wisconsin Genetics Computer Group (UWGCG, version 7.3-AXP) (Devereux *et al.*, 1984) with a DEC 3500/S computer running under the AXP operating system. Programs COMPARE and DOTPLOT (Devereux *et al.*, 1984) were used to carry out dot-matrix comparisons (Maizel & Lenk, 1981), while alignments of nucleotide and amino acid sequences were done with the programs GAP and BESTFIT (Devereux *et al.*, 1984). Program BEND (DNASTAR package) was used to predict the curvature of the DNA fragments analysed. Searches of the PIR (release 40), SWISSPROT (release 28), EMBL (release 38) and GenBank (release 83) databases were made with the programs FASTA and TFASTA (Pearson & Lipman, 1988). Sequence motifs and signature patterns in the amino acid sequences were searched by using the program MacPattern (Fuchs, 1991) and the PROSITE library (Bairoch, 1991) (release 11.1).

Isolation and analysis of RNAs from infected Vero cells. A set of the different types of RNA synthesized in Vero cell cultures infected with the BA71V strain of ASFV was obtained as follows. Whole cell RNA was prepared by the guanidinium isothiocyanate-caesium chloride method (Chirgwin *et al.*, 1979) from mock-infected cells, cells infected for 8 h in the presence of either 150 μg/ml cycloheximide (immediate early RNA) or 40 μg/ml cytosine arabinoside (early RNA), and cells infected for 18 h in the absence of inhibitors (late RNA).

For Northern blot analyses, 20 μg of each of the different RNAs was fractionated in formaldehyde–agarose gels, transferred onto nitrocellulose paper, and hybridized with synthetic oligonucleotides labelled with [γ-^{32}P]ATP and T4 polynucleotide kinase, according to standard protocols (Sambrook *et al.*, 1989). The probe for J268L mRNA was the synthetic oligonucleotide 5′ CCAGCGTAGCCCTAAATGGTATAA-AAGGAC 3′, complementary to the region between nucleotides 67 and 96 of the ORF. The probe for J154R mRNA was the oligonucleotide 5′ AGAATATGAGTTTTATCCCAGTACAACCCC 3′, complementary to nucleotides 96 to 125 of the ORF.

Primer extension analyses were carried out as described (Sambrook *et al.*, 1989), using the same oligonucleotides as for the Northern blot analyses. After hybridization of the 5′ end-labelled primers to 20 μg of the different RNAs, the samples were extended with avian myeloblastosis virus reverse transcriptase for 2 h at 37 °C, and then subjected to electrophoresis in 6% polyacrylamide sequencing gels.

Results

Mapping of cross-hybridizing DNA sequences

Hybridization studies with cloned ASFV DNA fragments revealed a complex pattern of repeated sequences next to the TIR at both ends of the ASFV genome, grouped in three main clusters of cross-hybridizing fragments (Almendral *et al.*, 1990). Two of these groups, corresponding to multigene families 110 and 360, have been discussed previously (Almendral *et al.*, 1990; González *et al.*, 1990). The third group of cross-hybridizing fragments includes fragments RU′, RX and RJ, located at the left end of the ASFV genome (Fig. 1). Fragment RJ hybridizes with fragment RU′ under stringent hybridization conditions, and with fragment RX under relaxed conditions. However, fragments RU′ and RX do not cross-hybridize, indicating that their hybridization with fragment RJ may be due to two unrelated sequences (Almendral *et al.*, 1990).

Due to the large size of fragment RJ (5·3 kb) a finer mapping of the repeated sequences was required before nucleotide sequence determination. Hybridization experiments were carried out with subfragments of RJ

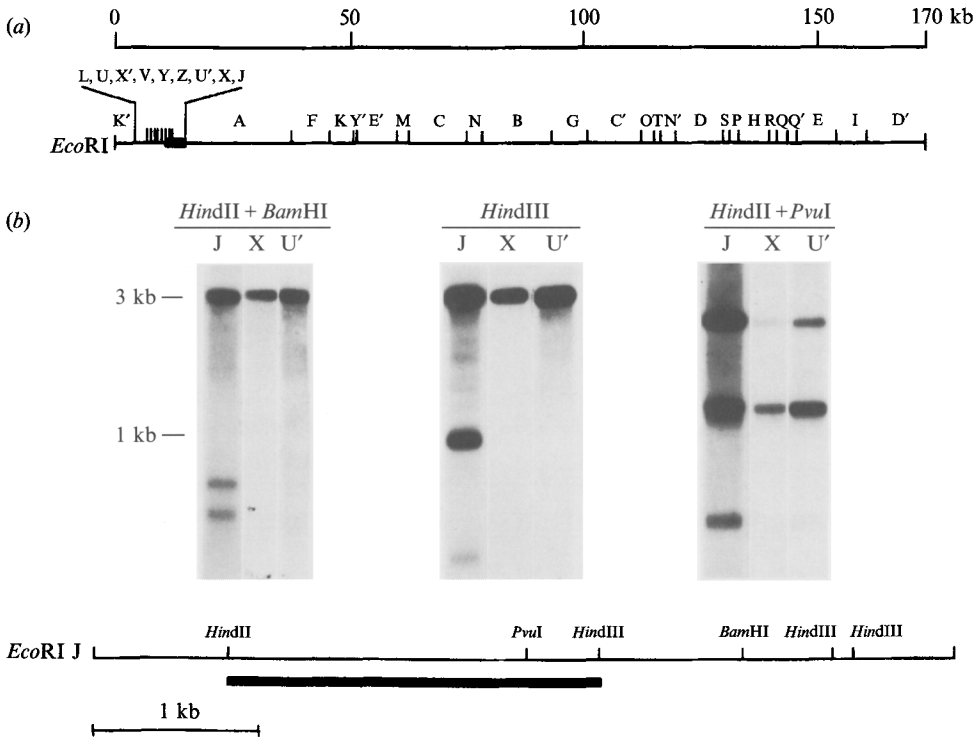


Fig. 1. Mapping of cross-hybridizing sequences in fragment RJ. (*a*) Location of fragments RU′, RX and RJ (indicated by a black box) in the ASFV genome. (*b*) Southern blots. Purified fragment RJ was digested with the indicated restriction endonucleases. Samples were electrophoresed in agarose gels, blotted onto nylon membranes and hybridized with 0·1 μg of purified fragments RU′, RX and RJ uniformly labelled with ^{32}P. Hybridizations were carried out as described (Almendral *et al.*, 1990). The black bar indicates the location of the repeated sequences within fragment RJ.

using as probes ^{32}P-labelled fragments RU′ and RX. The results obtained indicated that the repeated sequences were located in a 2·4 kb long *Hin*dII–*Hin*dIII restriction fragment at the left end of fragment RJ (Fig. 1).

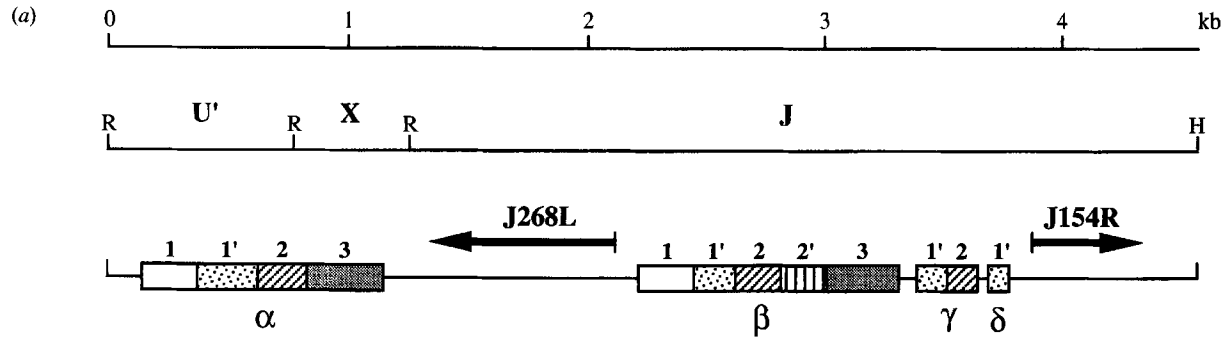
Nucleotide sequence of fragments RU′, RX and EcoRI–HindIII within RJ

Nucleotide sequencing of the restriction fragments RU′, RX and the R–H fragment at the left end of RJ was done by the dideoxynucleotide chain termination method (Sanger *et al.*, 1977), using either the M13 universal primer or synthetic 17-mer oligonucleotide primers corresponding to inner sequences. A total of 4591 bp were determined on both strands (Fig. 2).

A dot-matrix analysis of the DNA sequence was performed to detect the repeated sequences (Fig. 3). This analysis revealed a number of diagonal lines, indicating the existence of four direct repetitions (α, β, γ and δ) separated by stretches of unique sequence with varying lengths (Figs 2 and 3). These direct repetitions are composed of five different tandemly repeat units about 200 bp long, named units 1, 1′, 2, 2′ and 3. Repeat α contains four tandem repeat units in a 11′23 arrangement, and repeat β is composed of five tandem repeat units with a 11′22′3 organization. Repeat γ contains two truncated repeat units in a 1′2 arrangement, while repeat δ contains a single truncated repeat unit of type 1′. The identity between the different repeated units ranges from 39·2% to 84·6% (see Table 1).

These results account for the previously described cross-hybridization between restriction fragments in this DNA region (Almendral *et al.*, 1990). Hybridization under stringent conditions between fragments RU′ and RJ is explained by the presence of repeat units 1, 1′, 2 and 2′ in these fragments. Hybridization between fragments RX and RJ under relaxed conditions is consistent with the presence of units 3 in both fragments. On the other hand, fragments RU′ and RX do not hybridize because the unit 3 present in RX is only distantly related to units 1, 1′ and 2 in RU′ (see Table 1).

The repeated sequences have a high A+T content (70% in units 1, 1′, 2 and 2′; 76% in units 3) that appears to be distributed in oligo(dA).oligo(dT) tracts, including palindromic sequences of variable length. In the repeat units 1, 1′, 2 and 2′ of the repeated sequences α and β, these arrays of oligo(dA).oligo(dT) tracts surround a run of 10–14 G residues. Repeat units 3 and



(b)

GAATTCATCGTATTATCATGGATAAAAGCAGTAAATGTAATCAATTATAAAATTGACTTATTGAAGAGAAATGTTAAATGAGTGAAATCGGTGTTTACGATGATGTGCATGATCATACGA 120

AGAAACAGGTTCACTGGTGTCCATGATCAAAATTTAATATTCAACGTAAAAAGTACAGACGTTAACTGTTTAGTTTAAACATTAGTTTACACTAAAGGTTAATCTTTAGATTAAGCCTAG 240

TCAACAATGTTTGACATTTCTTCCATTCTTATACGGGGGGGGGGGGTCTTAATCGTTGTAATACTATTGTTGTGGATTGTTGAACATAATGAAGACTTTATAGATGCTAAATCAATGAAT 360

TATAATAATCAAACTGTTTAATGTGAGCTATAAATCTTATAAAAGTATTTATCAAAATTTATTTTTTTTTATAAAAAGTACAGGAATTAATCATTTAGTTTAAACTTTAGTTTGGCTATT 480

TGAGTTAACGTAGTTTAATATTTCTTCCATGCTTATCCAAAGGGGGGGGGGGGGGTCTTAATCGTTCTAATACTATTGTTGTGGATTGTTGAACATAATGAAAACTTTATAGATGCCTAA 600

AATAATTGAATCTAAAAATAGTAGTATACTTTAATCAAAATTTGATTTCTCATAAAAAAGGTACACATCATTCATTTAAGTTTCATGATGTTTAGTTCATTACTTCTACAATTACTGGGG 720

GGGGGGGTCCCTCATAGCTTTAGTATTATTGTGGCTTACTGACTATTATGTAGAAGAATTCATAGAAGCAACAATATATGTTACTAATGCTACATTGTAAATTATAAAATACATACTAAA 840

CTAATTTCAGTATTTTTTTTTGTTCATATAAGTTAAGGTACAAAAATGATTAAACACTGCAAAAAAAGAAAATCATAATGCTATTATACATAGTGATCATAGTGGCTTGTATCATTTCTA 960

AACTAGTTCCAAATGAATATTGGGCAATACATCTATTTTTCATCATTATGATTTTTATGGTATATATGTATGAAAAGTTAGATATACATCAAAAATATCAGTTCTGGAATTATACGATGT 1080

CAGGCTTATCTGGACATAACGTACAGGTAATATGTAAGTGTTACTAAATACTATGAAGTATCTATTTTTTTTACAACAAAAAAAATAGATGTAATCGGTTTTATCATTTAGGTGTGTATT 1200

TCTTTAGCATCTTTCTAAGAATTCATTGTTTAGCGGTATAAAAACAATGAATGATCATCATATTCTATTTAACAATCTCCTAAATAAATGAACATCTTTTTCATTTTAACATGATCACCA 1320
W

ATAGTCATTTTGCGAAAAGGCATACATATGATCAATATCAGACCTACAATGAATGTTTCCATAATATCCCTTTATCGCAATAATTCTATTTTTGCATTCCGATATCTCATCATCTGTGCT 1440
Y D N Q S F A Y M H D I D S R C H I N G Y Y G K I A I I R N K C E S I E D D T S

ATTATATGTTTCCATAACTGTTTCATCATCAAACATAAATCCTATTAAATAAGCAAAGCGCTTTAATCCCGGATAGATTTTCACCATTTTTCTAAGAGCCGTGTATAGCTTGTAATAAAT 1560
N Y T E M V T E D D F M F G I L Y A F R K L G P Y I K V M K R L A T Y L K Y Y I

GGCCAGAAATATGCAATAAAGCGTAGAAAGAGAGTAATTTTTGGCATAAAAGGTTTTGAAGGTTTGATGAATGGCTAAATCGCAGATAATATAAGATACGATATTAAAGCGCACCTGTTC 1680
A L F I C Y L T S L S Y N K A Y F T K F T Q H I A L D C I I Y S V I N F R V Q E

ACGCAGATTTGTTGAAAACTTCGTAGAAAGATTCAACAAATAAAAGGTTATTAATAGTTGCTCATCATTCTCCTTATACGACTTCGTCAGACACTCTAAGATTTTACTACTGGACACATC 1800
R L N T S F K T S L N L L Y F T I L L Q E D N E K Y S K T L C E L I K S S S V D

TGCCACATGTTGAACATTTAAAGCCTGTTCTTCTTCTGTGTTACGACAAAAGAGCCGCGCATATTCAGGTGAAGCTCCCCAGGATAACAACGTCCGTGTCACGGCTAAATTTTTTTTGAC 1920
A V H Q V N L A Q E E E T N R C F L R A Y E P S A G W S L L T R T V A L N K K V

GATGACTTTCATCAGAAATAAGTCTTTATTTTTGCATTGATCACTATGCGAATTTGTATAGTTGACGCCGTTGCATTGAGTACATTGATATAATGTTTTGCAATTCCAGCGTAGCCCTAA 2040
I V K M L F L D K N K C Q D S H S N T Y N V G N C Q T C Q Y L T K C N W R L G L

ATGGTATAAAAGGACTGTATTTTCGACGCAAGCATGCTGATTAACGATGTTTTTGAGACAACACGTTGTTAAGGACACCATATTGTCTCCAATTTGTTGGATAAAAGTCTTTACTAAAAA 2160
H Y L L V T N E V C A H Q N V I N K L C C T T L S V M ← J268L

AATAGGTTTTTAGTTTTAACAATCGAGATTTTATTATTTGGATATAGCATCATCAAAAAGATTTATGTTAAAATTTAATTTGTCTTACAAAAAGTACAGATTCAATAATTGGTTTAACTA 2280

TTTAGTTTAACTATTTAGTTAAATTACTTGGTTAGCAATGTTTGAGCTTTCTTCCATTCTTATTCGAGGGGGGGGGGGGGTCCTAATAGTTTTGATACTTTTGTTGTGGATTGTTGATGA 2400

AAACTGTACAGATGCTAAAGCAATGGCATATAATATTAATTGTACTGTTTAATGTGAGTTTATAAGTATGTATCAAAATTTATTTTTTTTTATAAAAGTACAGAAATTAATGTTAGCTTA 2520

AACTTTAGTTTAGCTATTTGAGTTAATGATGTTTAACCTATCTTCCATGCTTATCAGGGGGGGGGGGGTCCTAATTGTTTTACGTTATAAATTTGATCAAAATTTAATTTCTTATAAAAA 2640

AGGTACACATCAACATTGCTCATTTAAGTTTCATAATGTTTGATTCATTACTTCCTACAATTACTGGGGCGGGGGGTCTTTAATAGCATTAATATTATTATGGTTTGCTGACTATTTGGA 2760

ATTTATAGAAGCAAAAATAGATACTTTTACTAATAGCAGTGCATTATAAGTGATCAAAATTTAAATATCTCATAAAAAAGGTACACATCAACATCATTCATTTGTTTCATGATGTTTGAT 2880

TCATTACTTTCTACAATCACTCGGGGGGGGGCTCTCTCATAGCTTTAGTATTGCTATGGTTTACTAATTATTATGTAGAATTTATAGGAGCAGATATATATGGTGGTAATTGTACTCTATA 3000

Fig. 2. For legend see opposite.

```
AGTTTATAAAAATTTCAGTATATTTTTTTTGTTAATATAACAGAAATGATTTAAATTACAAAAAAAAGAGAATCATCATGCTATTATACATTGTTATTATAGTAGCATATGTAAGTTATA  3120

AATTGGTTCCAAAACATATTGGCCAATACTCATGTTTATGGCTTACATGGTATATACCCATGAAAAATTAGATATAAATGAAAGATCCGGATTTTGGAAATATATTATAGCAAAACTATT  3240

CAGATGCCACGGATGTGAAATATGTAAGTGATATTAAAAACATAATACGAAGTGTAATAATTGACTATTTTTTGTTGAATGGAAACTGTTTTTGAGTAATATGTTGAACATAATGGTTTT  3360

TATAGATGTTAAATCATTTGATTGTAGTAATTGTACTAAATGATGTGAGATATAAACCTAGTTCAATCAAAATTTAATTTTTTTATAAAAAGCTACAGAGTGTTTTATTAAACGTAGCTT  3480

ATTTAAAAAGTTACACAATGTTAAAATCTCTACTTACTTTAATTCTTTGTGCGGGTTTTATTAACTTTATCCATATTATGGCTTATTACTTACCATGTAGAACTTATAGAAGCAATAGATG  3600

ATTTCTACGACTGAAATATAGAATAGTTCATTTTCTATTTGTAAAATAATGATTTATATTCTCTCCTAAAAATACGATCGATCCTTTATATGCTTTGAAAACGAATATTAACAACTTGAT  3720

TTTTTTTTCTATAAATAAACTATAAATGAAAATAGTAAAACTCATAGAGTCTTATAAGTGAACATCCATCATGTTACTCAAACGTTGGACTATTAAAAGATACTCCGTGTGCATTATTGC  3840

TTTTAATCAGTATGATTACTTTATACGAAGCCGCTATTAAAACGCTTATCACACACCGAAAACAGATTTTAAAACACCCCGATAGCCGTGAAATTTTACTAGCTTTGGGGTTGTACTGGG  3960
J154R→M  I  T  L  Y  E  A  A  I  K  T  L  I  T  H  R  K  Q  I  L  K  H  P  D  S  R  E  I  L  L  A  L  G  L  Y  W  D

ATAAAACTCATATTCTTGTTAAATGTCGTGAATGTGGGAAAATGAGTCTTACCGGAAAACACAGTACAAAATGTATTAACATTAATTGTCTACTTATTCTTGCCATAAAAAAAAAGAATA  4080
  K  T  H  I  L  V  K  C  R  E  C  G  K  M  S  L  T  G  K  H  S  T  K  C  I  N  I  N  C  L  L  I  L  A  I  K  K  K  N  K

AGCGTATGGTTGATACCTTGATAAGAATGGGTGCGGATGTAACATATATACATCTTTTAAAGAATAAGATAAAACTGTCATACAACCAGCTGTCTATGCTTAAAAGCAACTCGCAGATTT  4200
  R  M  V  D  T  L  I  R  M  G  A  D  V  T  Y  I  H  L  L  K  N  K  I  K  L  S  Y  N  Q  L  S  M  L  K  S  N  S  Q  I  S

CGTTGAAGGAGCTTCATGCTATATGCTACCTTTTATATGGTCGGCTTCCCAAAAAAATTAAACAAGGGATGCAACTGTGTAAAACAATGGCGGGACTATGTGGTGAACTTTTATGAGCAT  4320
  L  K  E  L  H  A  I  C  Y  L  L  Y  G  R  L  P  K  K  I  K  Q  G  M  Q  L  C  K  T  M  A  G  L  C  G  E  L  L

TTTTAGCTCCGTAAATGTAATATATGTATATAAAACAAATAGATAATACCAAAATATATTCTATGTACATAACATCCGGAAAATTATTTTTTTTTTCTCATACCCTTAAATATAAAAATAT  4440

TGGGTTTCTTCACTAAACTTTAGAGGTAAAAATTTTTCTTTGTTTTGCATCATCATGTATGGGTTTAGGTTATCCCAGGGATTGTTTATTTGAATATTTCCTAAATAGGAACACAACGCC  4560

ATGATCATATATCTTTCATTCTGGTAAGCTT  4591
```

Fig. 2. Nucleotide sequence and deduced amino acid sequences of fragments RU′, RX and *Eco*RI–*Hin*dIII within RJ. (*a*) Location and organization of the repeated sequences and ORFs within the sequenced region. Repeated sequences are indicated by boxes and the ORFs by arrows. The endonuclease restriction sites relevant to the cloning procedures are indicated (R, *Eco*RI; H, *Hin*dIII). (*b*) Nucleotide and predicted amino acid sequences. The nucleotide sequence is shown in the 5′ to 3′ direction and from left to right according to the restriction map. Repeated sequences are underlined with different patterns according to the nomenclature used in (*a*). The G-rich core of the repeated sequences is indicated by black boxes and the *Eco*RI and *Hin*dIII restriction sites by open boxes. The amino acid sequences are shown below the DNA sequence in the single-letter amino acid code, and the name of each ORF is given beside the initiating ATG. The accession number for this sequence in the GenBank database is U13763.

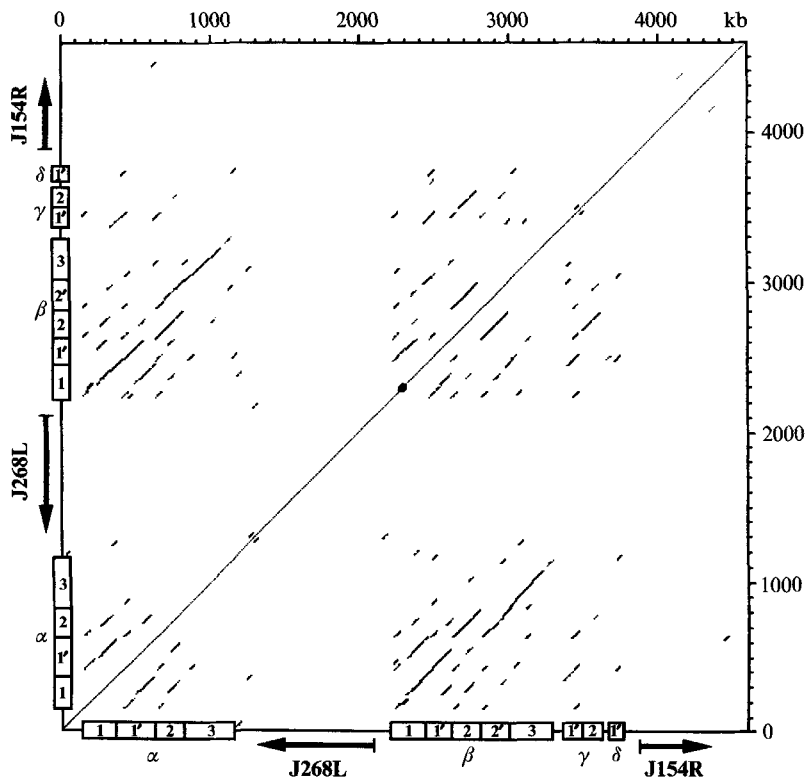


Fig. 3. Repeated sequences within the sequenced region. A dot-matrix analysis was performed with a window size of 30 nucleotides, allowing a maximum of nine mismatches. The location of the repeated sequences and ORFs is indicated on both the horizontal and vertical axes.

Table 1. *Percentage identity between the different repeated units derived from the dot matrix in Fig. 3*

Repeated unit	Percentage identity with:*										
	α1′	α2	α3	β1	β1′	β2	β2′	β3	γ1′	γ2	δ1′
α1	72·9	50·8	41·6	69·4	57·5	58·2	51·3	43·6	46·1	50·9	46·4
α1′		60·9	44·4	75·9	83·5	63·4	52·4	41·2	59·8	49·1	58·9
α2			48·4	46·7	44·3	84·6	73·7	45·3	39·2	64·3	44·6
α3				39·7	39·8	43·2	43·8	66·7	50·5	46·6	51·8
β1					74·8	49·2	53·5	46·3	55·9	48·3	46·4
β1′						41·3	49·6	40·9	74·1	39·7	55·4
β2							82·7	43·9	46·2	68·1	46·4
β2′								48·7	39·2	66·4	46·4
β3									49·5	50·9	49·9
γ1′										42·2	48·2
γ2											46·4

* Calculated using the program GAP (Devereux *et al.*, 1984).

the truncated repeats γ and δ lack the G-rich core, except the repeat unit 1′ found in the repeated sequence γ that contains a run of four G residues.

Open reading frames

Nucleotide sequence analysis revealed the existence of two ORFs located next to the tandemly repeated sequences within the R–H fragment at the left end of RJ (Fig. 2). These ORFs are named J268L and J154R according to the previously established nomenclature (Rodríguez *et al.*, 1992). The ASFV ORFs are designated by the letter of the *Eco*RI fragment in which translation is started, followed by the number of amino acids of the encoded protein and by L or R depending on whether the reading frame is leftward or rightward on the standard map. ORF J154R is located between nucleotides 3852 and 4316 of the sequence shown in Fig. 2, while ORF J268L starts at position 2121 and ends at position 1315 and is located in the lower strand.

The deduced molecular masses of the putative proteins J268L and J154R are 31·3 and 17·6 kDa, and their theoretical isoelectric points are 8·0 and 10·6, respectively. No hydrophobic sequences that may constitute signal sequences or transmembrane domains have been found in either protein.

Searches against the GenBank, EMBL, PIR and SWISSPROT databases with the amino acid sequences of J268L and J154R, using programs TFASTA and FASTA (Pearson & Lipman, 1988), revealed no significant similarities to any sequences in the databanks except the expected similarity to their counterparts in the ASFV strain Malawi LIL20/1. These ORFs present in the Malawi LIL20/1 isolate have been defined as members of a new multigene family, named multigene family 300 (Yozawa *et al.*, 1994). We have found that the putative proteins J154R and J268L contain the motif '$CX_2CX_8HX_3C$' between residues 45 to 61 and 40 to 56, respectively, which may constitute a zinc-finger domain of the C_2HC form (Berg, 1986*a*, *b*). This motif is also present in all three members of multigene family 300 of the isolate Malawi LIL20/1 described by Yozawa *et al.* (1994). On the other hand, adjacent to this sequence protein J154R contains the sequence 'KKKNK' between residues 73 and 77, which shares a remarkable degree of similarity with the consensus sequence for the nuclear targeting signal '(RKTA)KK(RQNTSG)K' (Dingwall & Laskey, 1986; Gómez-Márquez & Segade, 1988).

Transcriptional analysis of ORFs J268L and J154R

Transcription of ORFs J268L and J154R was analysed by Northern blot hybridization and primer extension. For these experiments, RNA was prepared from Vero cells that were either mock-infected, or infected with ASFV in the presence or absence of cycloheximide and cytosine arabinoside, as described in Methods.

Northern blot hybridizations were performed with oligonucleotide probes specific for each gene. The probe for J268L mRNA recognized two major virus-induced transcripts of approximately 1·1 and 1·4 kb, synthesized during the immediate early and late phases of infection, respectively. The probe also hybridized with minor transcripts of 1·5 kb in cycloheximide RNA and 1·9 kb in late RNA (Fig. 4*b*). The probe for J154R mRNA detected a major transcript of 0·6 kb in the late RNA sample. Other minor transcripts were detected in all the RNA samples from virus-infected cells (Fig. 4*b*).

To map the 5′ ends of the J268L and J154R transcription products, primer extension analyses were performed by using as primers the oligonucleotides previously used for the Northern blots. The J268L probe

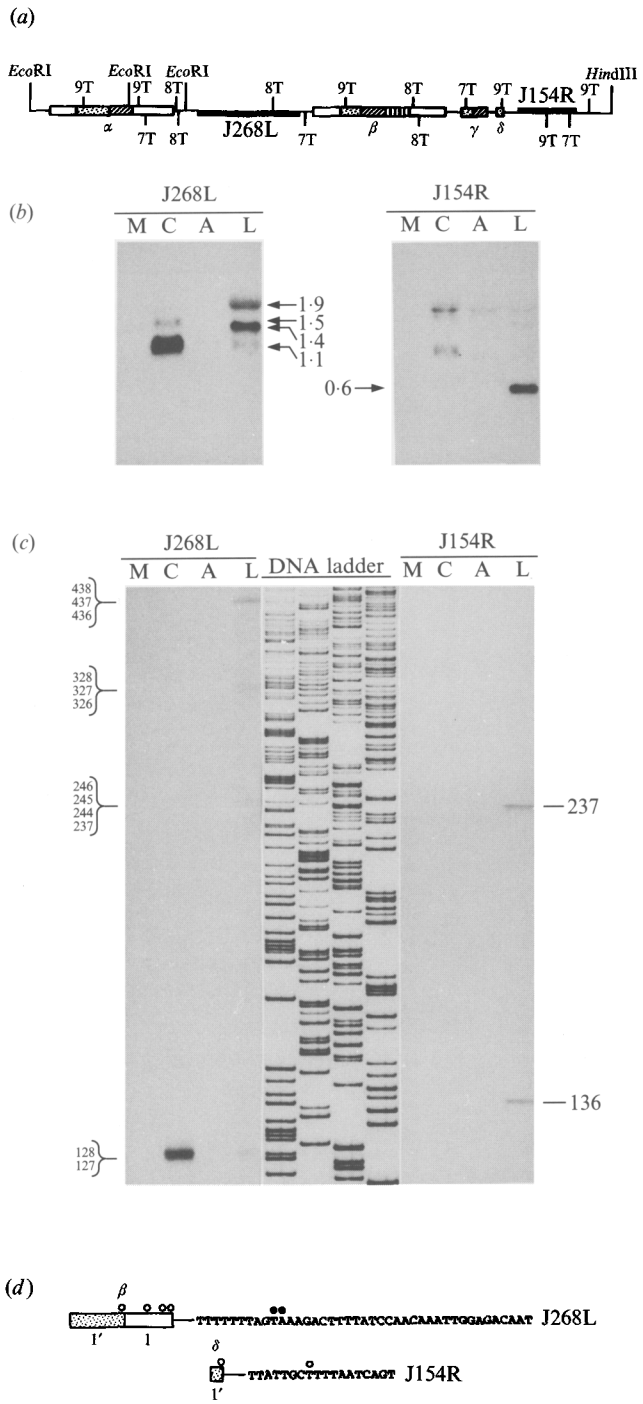


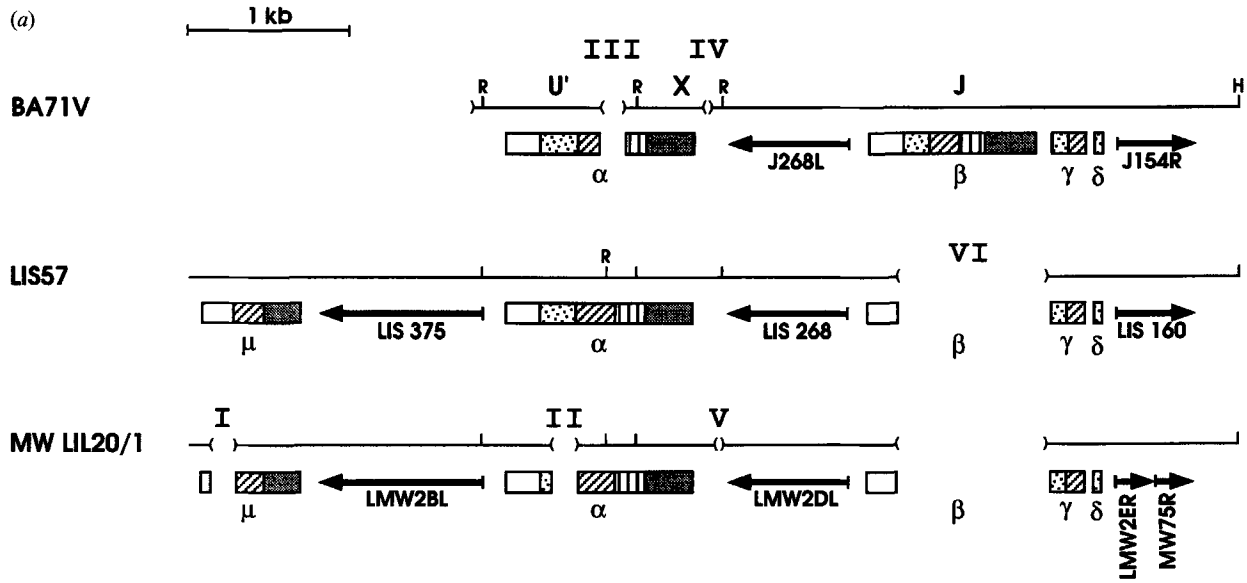
Fig. 4. Transcriptional analysis of the genes J268L and J154R. (*a*) Location within the sequenced region of the repeated sequences (boxes with different patterns) and ORFs (black boxes). The distribution of stretches of seven or more consecutive thymidylate residues (T) in either strand are also shown. (*b*, *c*) Autoradiograms of RNA hybridizations (*b*) and primer extension analyses (*c*) of four types of RNA from Vero cells mock-infected (lanes M), or infected in the presence of either cycloheximide (lanes C) or cytosine arabinoside (lanes A), or in the absence of inhibitors (lanes L) are shown. In the autoradiograms corresponding to Northern blot hybridizations, the sizes (in kb) of the most prominent transcripts are indicated. The samples corresponding to the primer extension analyses were electrophoresed alongside an unrelated DNA sequencing reaction (DNA ladder) used as a size marker, and the numbers correspond to the sizes (in nucleotides) of the major bands detected. (*d*) Precise location of transcriptional initiation sites within the nucleotide sequence of the noncoding strand. The sequence is shown in the 5′ to 3′ direction and from right to left in the case of J268L and from left to right in the case of J154R, according to the restriction map. Repeated sequences and ORFs are indicated by boxes. ●, ○, Early and late transcriptional initiation sites, respectively.

hybridized mainly with the immediate early RNA sample, generating after extension two products of 127 and 128 nucleotides, corresponding to transcriptional initiation sites located 31 and 32 nucleotides, respectively, upstream of the first nucleotide of the predicted translation initiation codon (Fig. 4*c*, *d*). In agreement with the Northern blot results, additional extensions were detected in the late RNA sample, corresponding to transcriptional initiation sites into the tandemly repeated sequence *β* (Fig. 4*c*, *d*). The more intense extension detected in the late RNA sample (436 to 438 nucleotides) could correspond to the late transcript of 1·4 kb detected in the Northern blot. The probe for J154R mRNA hybridized exclusively with late RNA, giving rise to two bands of 136 and 237 nucleotides, corresponding to transcriptional initiation sites located 11 and 112 nucleotides, respectively, upstream of the ORF J154R (Fig. 4*c*, *d*). This last transcriptional initiation site is located into the repeated sequence *δ* (Fig. 4*d*). The absence of extension products corresponding to the less abundant transcripts detected by Northern blot could be explained by a lower sensitivity of the primer extension compared with the Northern blot analysis or, alternatively, to the large size of these extension products, which would not be detected within the gels used for these experiments.

Recently, a stretch of seven or more thymidylate residues (7T) has been identified as the signal for the 3′ end formation of ASFV early and late RNAs (Almazán *et al.*, 1992, 1993). Stretches of eight or nine thymidylate residues are found downstream of the stop codons of genes J268L and J154R (Fig. 4*a*). The sizes of the main transcripts detected by Northern blot and the positions of the 5′ ends of these transcripts are consistent with the use of the 7T motifs as signals for 3′ end formation.

Structure of the tandemly repeated sequences in other isolates of ASFV

To analyse the genetic variation of the tandemly repeated sequences described here, we determined the nucleotide sequence of the corresponding region in the virulent ASFV isolate LIS57. The nucleotide sequence of this



(b)

I	LIS57	TTTTCTATGTAAAAAGTACAGACTT.....TATACAGACGTCGCTTGTTTAAGCT
	MW LIL20/1	TTTTCTATGTAAAAAGTACA GACGTCGCTTGTTTAAGCT
II	LIS57	AAAAAATTTCATCAAAATTTATT.....GTTTTCATCAAAATTTAATTTCT
	BA71V	AAAAGTATTTATCAAAATTTATT.....ACTTTAATCAAAATTTGATTTCT
	MW LIL20/1	AAATGTTTTCATC AAAATTTAATTTCT
III	LIS57	CTTCCTACAATTACTGGGGGGGGGTA.....ATAATTACTGGGGGGGGGGGGTCTTT
	MW LIL20/1	CTTCCTACAATTACTGGGGGGGG.TC.....ACAATTACT.GGGGGGGGGGGTCTTT
	BA71V	CTT.CTACAATTACTGG GGGGGGGGTCCCT
IV	LIS57	CTA.TTTTTTTTGTTGTAAAAAAAGA.....CCAAAAAAAATAGATGTAATCGGTTT
	MW LIL20/1	CTATTTTTTTTTGTTGTAAAAAAAAC.....AAAAATAGATGTAATCGGTTT
	BA71V	CTA.TTTTTTTTACAACAAAA AAAATAGATGTAATCGGTTT
V	LIS57	AGATGTAATCGGTTTCATCATTTAGG.....TTGTTTAGTGATATAAAAACAATGA
	BA71V	AGATGTAATCGGTTTTATCATTTAGG.....TTGTTTAGCGGTATAAAAACAATGA
	MW LIL20/1	AGATGTAATCGGTTTCATCATTTA GTTTATTGATATAAAAACAATAA
VI	BA71V	TGGATTGTTG.....ATGAAAACTG.....TTTTTATAGATGTTAAATCA
	LIS57	TGGATTATTGAATATAATGAAAAC TTTATAGATGCTAAAGCA
	MW LIL20/1	TGGATTGTTGAATATAATGAAAAC TTTATAGATGCTAGAATG

Fig. 5. Genetic structure of the sequenced region in three different ASFV isolates. (*a*) Distribution of the repeated sequences and ORFs in the sequenced region in the ASFV isolates BA71V, LIS57 and Malawi LIL20/1 (MWLIL20/1). Restriction sites are: R, *Eco*RI; H, *Hind*III. The ORFs are indicated by arrows and the tandemly repeat sequences by boxes. Deletions are given roman numbers. (*b*) DNA sequences around the deletion boundaries are shown for each deletion after alignment with the undeleted sequence obtained in different viral isolates. Lines show the positions of direct repetitions at the deletion end points. The nucleotide sequences for ASFV isolate LIS57 have been submitted to the GenBank database and assigned the accession numbers M34948 and U13764.

region in the virulent ASFV isolate Malawi LIL20/1 was obtained from GenBank (Yozawa *et al.*, 1994; accession number U03762). As shown in Fig. 5, comparison of the nucleotide sequences from these isolates revealed the existence of several differences due to six deletions. In three cases (I, II and III), deletions appear to be the result of homologous recombination between tandemly repeat units resulting in the generation of new repeated sequences with different arrangements. Deletions IV and V also seem to be produced by

homologous recombination, where the tandemly repeated sequences are not implicated. In the remaining deletion (VI), which accounts for the loss of most of the repeated sequence β in the isolates LIS57 and Malawi LIL20/1, no significant similarity was found around the deletion end points (Fig. 5*b*).

The two ORFs (J268L and J154R) detected next to the repeated sequences in the BA71V isolate present several differences in the other ASFV isolates analysed (Fig. 5*a*). The counterparts of J268L in the ASFV strains Malawi LIL20/1 (LMW2DL) and LIS57 (LIS268) encode polypeptides of 268 amino acids that are more than 90% identical to J268L. The counterpart of J154R in the isolate LIS57 (LIS160) encodes a polypeptide of 160 amino acids with a 98·7% identity to J154R. However, in the Malawi LIL20/1 strain this gene appears to be truncated giving rise to a shorter ORF, LMW2ER, that encodes a polypeptide of 76 amino acids with 94·7% identity to the N-terminal region of J154R. An examination of the nucleotide sequence downstream of this ORF revealed the presence of a second ORF encoding a 75 amino acid polypeptide that is 87% identical to the C-terminal region of J154R. We have named this ORF MW75R.

Discussion

We have determined the sequence of a cluster of direct repeats located in *Eco*RI fragments U′, X and J at the left end of the ASFV genome. Analysis of the sequence by dot-matrix comparisons showed that the repeated sequences consist of five types of repeat units. These repeats consist of long units (about 200 bp) of repeated sequences tandemly reiterated that have a high A+T content distributed in oligo(dA).oligo(dT) tracts, which, in the case of repeat units 1, 1′, 2 and 2′, surround a G-rich core of 10–14 residues.

The distribution of the oligo(dA).oligo(dT) tracts has similar characteristics to those of DNA sequences implicated in curvature of DNA. Intrinsically bent or curved DNA molecules result when special base sequences are repeated in phase with the DNA helical repeat. Many base sequences can impart systematic curvature to DNA, but most such bends are small compared with the special effect produced by runs of oligo(dA).oligo(dT) tracts, each tract being about half a helical turn long and repeated at 10–11 bp intervals (reviewed in Trifonov, 1985; Hagerman, 1990; Crothers *et al.*, 1990). Analysis of these ASFV repeated sequences with the program BEND, which predicts the theoretical curvature of DNA fragments on the basis of data obtained by E. N. Trifonov (Bolshoy *et al.*, 1991), indicated the existence of long curved DNA regions within these repeated sequences (data not shown). It has been proposed by many authors that DNA curvature is of functional importance in a wide variety of biological processes. Curved DNA is found in regions upstream of prokaryotic promoters, the promoters themselves and origins of replication, and seems to play a role in chromatin organization. Curved DNA is frequently involved in all protein–DNA complexes in which DNA is wrapped around the protein or in which the protein binds specifically to only one side of the DNA molecule (reviewed in Travers, 1989; Hagerman, 1990). With regard to the function of these curved repeated sequences in ASFV, it is possible that they might contain recognition sequences for transcriptional regulatory proteins, in a similar way to other repeated sequences that function as transcriptional enhancers (Rosenthal *et al.*, 1983; Reeder, 1984).

These tandemly repeated sequences also share a number of properties with chromosomal α satellite DNA, both consisting of 180 bp long units with a high A+T content distributed in oligo(dA).oligo(dT) tracts (Manueldis, 1976; Wu & Manueldis, 1980). α satellites are found mainly at the centromeres of all chromosomes of many primates, and some regions have been shown to bind the heterochromatin centromere protein CENP-B via a 17 bp motif, the CENP-B box. The centromere is a critical chromosomal element responsible for the proper segregation of chromosomes in mitosis and meiosis, where α satellites may be involved in the structure and/or function of the centromere (reviewed in Willard, 1990).

As mentioned above, the principal characteristic of these ASFV repeated sequences is the presence of a G-rich core embedded in very A+T-rich regions. Similar sequences have been detected in satellite III repeats of *Drosophila*, a centromeric SAR-like DNA sequence with a repeat length of 359 bp located in the centromeric heterochromatin of the X chromosome (reviewed in Brutlag, 1980). As in the case of SARs, most A–T base pairs of satellite III are found in the form of oligo(dA).oligo(dT) tracts, which constitute the essential determinants for the specific SAR-scaffold interaction (Käs *et al.*, 1989). Recent evidence shows that topoisomerase II is involved in the cleavage of satellite III repeats in *Drosophila* DNA. Sequencing of *in vivo* topoisomerase II cleavage sites in the histone SAR region as well as in satellite III repeats reveals very precise cutting in G-rich spikes embedded in very A+T-rich regions (Käs & Laemmli, 1992). Topoisomerase II has an essential function in the segregation of the intertwined daughter chromosomes at the end of DNA replication (DiNardo *et al.*, 1984). Also, genetic evidence obtained in yeast has suggested a role for topoisomerase II in chromosome condensation as well as deconden-

sation (Uemura *et al.*, 1987; Newport, 1987; Newport & Spann, 1987; Adachi *et al.*, 1991).

Although long complex repeat arrays have been characterized in other viruses, such as herpesviruses (reviewed in McGeoch, 1989) and iridoviruses (Schnitzler *et al.*, 1987; Fischer *et al.*, 1988), only ASFV contains large multiple internal repeats that share similarity with repeated sequences containing the cleavage sites of topoisomerase II. The function of these ASFV SAR-like sequences is unknown. It would be of interest to determine whether the ASFV topoisomerase II (García-Beato *et al.*, 1992) shows some specificity of cleavage for such repeated sequences found at the left end of the viral genome and to study the possible role of these repeats in genome packaging into virions and/or in DNA replication.

We have studied the variation of these repeated sequences in diverse virus isolates. The sequence comparisons indicate that this region of the ASFV genome undergoes DNA rearrangements leading to duplication and deletion of repeat units to form tandem repeat arrays. The greater variability observed in the left DNA end of ASFV field isolates (Fig. 5) could be a consequence of a lower selective pressure to maintain these sequences. However, other studies suggest that the repeated sequences and multigene families present near the DNA ends might play some role in the infection of the soft tick vector of ASFV, since the ends of the genome of virus isolated from soft ticks inhabiting warthog burrows in Africa are less variable than those of virus isolated from domestic pigs (Dixon & Wilkinson, 1988; Blasco *et al.*, 1989*a*, *b*).

Next to the tandemly repeated sequences we have detected the existence of two ORFs named J268L and J154R, which are members of the multigene family 300. The Northern hybridization and primer extension analyses indicate that both ORFs are transcribed during viral infection. J268L is expressed at both immediate early and late times of the infection cycle, whereas transcripts for J154R are detected only after virus DNA replication. Searches against the databanks revealed similarities only to their counterparts and the other members of the multigene family 300 in the ASFV strain Malawi LIL20/1. The members of this multigene family contain a cysteine-rich domain that may constitute a zinc-finger motif. Regions containing cysteine and aromatic residues have been shown to form metal-binding structures. In some cases, they generate a zinc-finger domain (Miller *et al.*, 1985), but in other cases, the metal ion serves to stabilize protein–protein interactions (Frankel *et al.*, 1988). Adjacent to this cysteine-rich domain, ORF J154R contains the sequence 'KKKNK', which corresponds with the consensus sequence for the nuclear targeting signal '(RKTA)KK(RQNTSG)K' (Dingwal & Laskey, 1986; Gómez-Márquez & Segade, 1988). Nothing is known about the role of the multigene family 300 in viral infection. One possibility is that it could operate as a transcriptional modulator of some virus or cellular gene(s).

We are grateful to Paco Rodríguez, Rafael Yañez, María L. Salas and José Salas for critical reading of the manuscript. This work was supported by grants from Comisión Interministerial de Ciencia y Tecnología, Junta de Extremadura, and by an institutional grant from Fundación Ramón Areces.

References

ADACHI, Y., LUKE, M. & LAEMMLI, U. K. (1991). Chromosome assembly *in vitro*: topoisomerase II is required for condensation. *Cell* **64**, 137–148.

ALMAZÁN, F., RODRÍGUEZ, J. M., ANDRÉS, G., PÉREZ, R., VIÑUELA, E. & RODRÍGUEZ, J. F. (1992). Transcriptional analysis of multigene family 110 of African swine fever virus. *Journal of Virology* **66**, 6655–6667.

ALMAZÁN, F., RODRÍGUEZ, J. M., ANGULO, A., VIÑUELA, E. & RIDRÍGUEZ, J. F. (1993). Transcriptional mapping of a late gene coding for the p12 attachment protein of African swine fever virus. *Journal of Virology* **67**, 553–556.

ALMENDRAL, J. M., ALMAZÁN, F., BLASCO, R. & VIÑUELA, E. (1990). Multigene families in African swine fever virus: family 110. *Journal of Virology* **64**, 2064–2072.

ANGULO, A., VIÑUELA, E. & ALCAMÍ, A. (1992). Comparison of the sequence of the gene encoding African swine fever virus attachment protein p12 from field virus isolates and viruses passaged in tissue culture. *Journal of Virology* **66**, 3869–3872.

BAIROCH, A. (1991). PROSITE: a dictionary of sites and patterns in proteins. *Nucleic Acids Research* **19** (Suppl.), 2241–2245.

BAROUDY, B. M., VENKATESAN, S. & MOSS, B. (1982). Incompletely base-paired flip-flop terminal loops link the two DNA strands of the vaccinia virus genome into one uninterrupted polynucleotide chain. *Cell* **28**, 315–324.

BERG, J. M. (1986*a*). Potential metal-binding domains in nucleic acid binding proteins. *Science* **232**, 485–487.

BERG, J. M. (1986*b*). More metal-binding fingers. *Nature* **319**, 264–265.

BLASCO, R., AGÜERO, M., ALMENDRAL, J. M. & VIÑUELA, E. (1989*a*). Variable and constant regions in African swine fever virus DNA. *Virology* **168**, 330–338.

BLASCO, R., DE LA VEGA, I., ALMAZÁN, F., AGÜERO, M. & VIÑUELA, E. (1989*b*). Genetic variation of African swine fever virus: variable regions near the ends of the viral DNA. *Virology* **173**, 251–257.

BOLSHOY, A., MCNAMARA, P., HARRINGTON, R. E. & TRIFONOV, E. N. (1991). Curved DNA without A–A: experimental estimation of all 16 DNA wedge angles. *Proceedings of the National Academy of Sciences, USA* **88**, 2312–2316.

BRUTLAG, D. L. (1980). Molecular arrangement and evolution of heterochromatic DNA. *Annual Review of Genetics* **14**, 121–144.

CARRASCOSA, A. L., DEL VAL, M., SANTARÉN, J. F. & VIÑUELA, E. (1985). Purification and properties of African swine fever virus. *Journal of Virology* **54**, 337–344.

CHIRGWIN, J. M., PRZYBYLA, A. E., MACDONALD, R. J. & RUTTER, W. J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18**, 5294–5299.

COSTA, J. V. (1990). African swine fever virus. In *Molecular Biology of Iridoviruses*, pp. 247–270. Edited by G. Darai. Boston: Kluwer Academic Publishers.

CROTHERS, D. M., HARAN, T. E. & NADEAU, J. G. (1990). Intrinsically bent DNA. *Journal of Biological Chemistry* **265**, 7093–7096.

DE LA VEGA, I., VIÑUELA, E. & BLASCO, R. (1990). Genetic variation and multigene families in African swine fever virus. *Virology* **179**, 234–246.

DE LA VEGA, I., GONZÁLEZ, A., BLASCO, R., CALVO, V. & VIÑUELA, E. (1994). Nucleotide sequence and variability of the inverted terminal repetitions of African swine fever virus DNA. *Virology* **201**, 152–156.

DEVEREUX, J., HAEBERLI, P. & SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* **12**, 387–395.

DINARDO, S., VOELKEL, K. & STERNGLANZ, R. (1984). DNA topoisomerase II mutant of *Saccharomyces cerevisiae*: topoisomerase II is required for segregation of daughter molecules at the termination of DNA replication. *Proceedings of the National Academy of Sciences, USA* **81**, 2616–2620.

DINGWALL, C. & LASKEY, R. A. (1986). Protein import into the cell nucleus. *Annual Review of Cell Biology* **2**, 367–390.

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. *Journal of General Virology* **69**, 2981–2993.

DIXON, L. K., BRISTOW, C., WILKINSON, P. J. & SUMPTION, K. J. (1990). Identification of a variable region of the African swine fever virus genome that has undergone separate DNA rearrangements leading to expansion of minisatellite-like sequences. *Journal of Molecular Biology* **216**, 677–688.

DIXON, L. K., BAYLIS, S. A., VYDELINGUM, S., TWIGG, S. R. F., HAMMOND, J. M., HINGAMP, P. M., BRISTOW, C., WILKINSON, P. J. & SMITH, G. L. (1993). African swine fever virus genome content and variability. *Archives of Virology* **7** (Suppl.), 185–199.

ESCRIBANO, J. M., ALCARAZ, C. & RUÍZ-GONZALVO, F. (1993). New perspectives in African swine fever virus protection. In *Coordination of Agricultural Research African Swine Fever*, pp. 113–129. Edited by A. Galo. Luxembourg: Commission of the European Communities.

ENJUANES, L., CARRASCOSA, A. L., MORENO, M. A. & VIÑUELA, E. (1976). Titration of African swine fever (ASF) virus. *Journal of General Virology* **32**, 471–477.

FISCHER, M., SCHNITZLER, P., SCHOLZ, J., ROSEN WOLFF, A., DELIUS, H. & DARAI, G. (1988). DNA nucleotide sequence analysis of the *Pvu*II DNA fragment L of the genome of insect iridescent virus type 6 reveals a complex cluster of multiple tandem, overlapping, and interdigitated repetitive DNA elements. *Virology* **167**, 497–506.

FRANKEL, A. D., BREDT, D. S. & PABO, C. O. (1988). Tat protein from human immunodeficiency virus forms a metal-linked dimer. *Science* **240**, 70–73.

FUCHS, R. (1991). MacPattern: protein pattern searching on the Apple Macintosh. *Computer Applications in the Biosciences* **7**, 105–106.

GARCÍA-BEATO, R., FREIJE, J. M., LOPEZ OTIN, C., BLASCO, R., VIÑUELA, E. & SALAS, M. L. (1992). A gene homologous to topoisomerase II in African swine fever virus. *Virology* **188**, 938–947.

GÓMEZ-MÁRQUEZ, J. & SEGADE, F. (1988). Prothymosin alpha is a nuclear protein. *FEBS Letters* **226**, 217–219.

GONZÁLEZ, A., TALAVERA, A., ALMENDRAL, J. M. & VIÑUELA, E. (1986). Hairpin loop structure of African swine fever virus DNA. *Nucleic Acids Research* **14**, 6835–6844.

GONZÁLEZ, A., CALVO, V., ALMAZÁN, F., ALMENDRAL, J. M., RAMÍREZ, J. C., DE LA VEGA, I., BLASCO, R. & VIÑUELA, E. (1990). Multigene families in African swine fever virus: family 360. *Journal of Virology* **64**, 2073–2081.

HAGERMAN, P. J. (1990). Sequence-directed curvature of DNA. *Annual Review of Biochemistry* **59**, 755–781.

KÄS, E., IZAURRALDE, E. & LAEMMLI, U. K. (1989). Specific inhibition of DNA binding to nuclear scaffolds and histone H1 by distamycin. The role of oligo(dA).oligo(dT) tracts. *Journal of Molecular Biology* **210**, 587–599.

KÄS, E. & LAEMMLI, U. K. (1992). *In vivo* topoisomerase II cleavage of the *Drosophila* histone and satellite III repeats: DNA sequence and structural characteristics. *EMBO Journal* **11**, 705–716.

KUZNAR, J., SALAS, M. L. & VIÑUELA, E. (1980). DNA-dependent RNA polymerase in African swine fever virus. *Virology* **101**, 169–175.

LEY, V., ALMENDRAL, J. M., CARBONERO, P., BELOSO, A., VIÑUELA, E. & TALAVERA, A. (1984). Molecular cloning of African swine fever virus DNA. *Virology* **133**, 249–257.

MCGEOCH, D. J. (1989). The genomes of the human herpesviruses: contents, relationships, and evolution. *Annual Review of Microbiology* **43**, 235–265.

MAIZEL, J. V. J. & LENK, R. P. (1981). Enhanced graphic matrix analysis of nucleic acid and protein sequences. *Proceedings of the National Academy of Sciences, USA* **78**, 7665–7669.

MANUELIDIS, L. (1976). Repeating restriction fragments of human DNA. *Nucleic Acids Research* **3**, 3063–3076.

MESSING, J. (1983). New M13 vectors for cloning. *Methods in Enzymology* **101**, 20–78.

MILLER, J., MCLACHLAN, A. D. & KLUG, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO Journal* **4**, 1609–1614.

MOSS, B. (1990). Regulation of vaccinia virus transcription. *Annual Review of Biochemistry* **59**, 661–688.

NEWPORT, J. (1987). Nuclear reconstitution *in vitro*: stages of assembly around protein-free DNA. *Cell* **48**, 205–217.

NEWPORT, J. & SPANN, T. (1987). Disassembly of the nucleus in mitotic extracts: membrane vesicularization, lamin disassembly, and chromosome condensation are independent processes. *Cell* **48**, 219–230.

ORTÍN, J., ENJUANES, L. & VIÑUELA, E. (1979). Cross-links in African swine fever virus DNA. *Journal of Virology* **31**, 579–583.

PEARSON, W. R. & LIPMAN, D. J. (1988). Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences, USA* **85**, 2444–2448.

REEDER, R. H. (1984). Enhancers and ribosomal gene spacers. *Cell* **38**, 349–351.

RIGBY, P. W., DIECKMANN, M., RHODES, C. & BERG, P. (1977). Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. *Journal of Molecular Biology* **113**, 237–251.

RODRÍGUEZ, J. M., SALAS, M. L. & VIÑUELA, E. (1992). Genes homologous to ubiquitin-conjugating proteins and eukaryotic transcription factor SII in African swine fever virus. *Virology* **186**, 40–52.

RODRÍGUEZ, J. M., YAÑEZ, R. J., PAN, R., RODRÍGUEZ, J. F., SALAS, M. L. & VIÑUELA, E. (1994). Multigene families in African swine fever virus: family 505. *Journal of Virology* **68**, 2746–2751.

ROSENTHAL, N., KRESS, M., GRUSS, P. & KHOURY, G. (1983). BK viral enhancer element and a human cellular homolog. *Science* **222**, 749–755.

SALAS, M. L., KUZNAR, J. & VIÑUELA, E. (1981). Polyadenylation, methylation, and capping of the RNA synthesized *in vitro* by African swine fever virus. *Virology* **113**, 484–491.

SALAS, M. L., REY CAMPOS, J., ALMENDRAL, J. M., TALAVERA, A. & VIÑUELA, E. (1986). Transcription and translation maps of African swine fever virus. *Virology* **152**, 228–240.

SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. (1989). *Molecular cloning: A Laboratory Manual*, 2nd edn. New York: Cold Spring Harbor Laboratory.

SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, USA* **74**, 5463–5467.

SCHNITZLER, P., DELIUS, H., SCHOLZ, J., TOURAY, M., ORTH, E. & DARAI, G. (1987). Identification and nucleotide sequence analysis of the repetitive DNA element in the genome of fish lymphocystis disease virus. *Virology* **161**, 570–578.

SOGO, J. M., ALMENDRAL, J. M., TALAVERA, A. & VIÑUELA, E. (1984). Terminal and internal inverted repetitions in African swine fever virus DNA. *Virology* **133**, 271–275.

SOUTHERN, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* **98**, 503–517.

TRAVERS, A. A. (1989). DNA conformation and protein binding. *Annual Review of Biochemistry* **58**, 427–452.

TRIFONOV, E. N. (1985). Curved DNA. *CRC Critical Reviews in Biochemistry* **19**, 89–106.

UEMURA, T., OHKURA, H., ADACHI, Y., MORINO, K., SHIOZAKI, K. & YANAGIDA, M. (1987). DNA topoisomerase II is required for condensation and separation of mitotic chromosomes in *S. pombe*. *Cell* **50**, 917–925.

VIÑUELA, E. (1987). Molecular biology of African swine fever virus. In *African Swine Fever*, pp. 31–49. Edited by Y. Becker. Boston: Martinus Nijhoff Publishing.

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. *Journal of General Virology* **74**, 2125–2130.

WARDLEY, R. C., NORLEY, S. G., MARTINS, C. V. & LAWMAN, M. J. (1987). The host response to African swine fever virus. *Progress in Medical Virology* **34**, 180–192.

WILLARD, H. F. (1990). Centromeres of mammalian chromosomes. *Trends in Genetics* **6**, 410–416.

WITTEK, R. & MOSS, B. (1980). Tandem repeats within the inverted terminal repetition of vaccinia virus DNA. *Cell* **21**, 277–284.

WU, J. C. & MANUELIDIS, L. (1980). Sequence definition and organization of a human repeated DNA. *Journal of Molecular Biology* **142**, 363–386.

YOZAWA, T., KUTISH, G. F., AFONSO, C. L., LU, Z. & ROOK, D. L. (1994). Two novel multigene families, 530 and 300, in the terminal variable regions of African swine fever virus genome. *Virology* **202**, 997–1002.

(*Received 24 August 1994; Accepted 22 November 1994*)