JOURNAL OF VIROLOGY, May 1990, p. 2073–2081
0022-538X/90/052073-09$02.00/0

Vol. 64, No. 5

Multigene Families in African Swine Fever Virus: Family 360

ANTONIO GONZÁLEZ,† VICTOR CALVO,‡ FERNANDO ALMAZÁN, JOSÉ M. ALMENDRAL, JUAN CARLOS RAMÍREZ, INMACULADA DE LA VEGA, RAFAEL BLASCO,§ AND ELADIO VIÑUELA*

Centro de Biología Molecular (Consejo Superior de Investigaciones Científicas-UAM), Facultad de Ciencias, Universidad Autónoma, Canto Blanco, 28049 Madrid, Spain

Received 15 September 1989/Accepted 2 January 1990

A group of cross-hybridizing DNA segments contained within the restriction fragments RK′, RL, RJ, and RD′ of African swine fever virus DNA were mapped and sequenced. Analysis of these sequences revealed the presence of a family of homologous open reading frames in regions close to the DNA ends. The whole family is composed of six open reading frames with an average length of 360 coding triplets (multigene family 360), four of which are located in the left part of the genome and two of which are in the right terminal *Eco*RI fragment. In close proximity to the right terminal inverted repeat, we found an additional small open reading frame which was homologous to the 5′-terminal portion of the other open reading frames, suggesting that most of that open reading frame has been deleted. These repeated sequences account for the previously described inverted internal repetitions (J. M. Sogo, J. M. Almendral, A. Talavera, and E. Viñuela, Virology 133:271–275, 1984). Most of the genes of multigene family 360 are transcribed in African swine fever virus-infected cells. A comparison of the predicted protein sequences of family 360 indicated that several residues are conserved, suggesting that an overall structure is maintained for every member of the family. The transcription direction of each open reading frame, as well as the evolutionary relationships among the genes, suggests that the family originated by gene duplication and translocation of sequences between the DNA ends.

African swine fever (ASF) virus DNA is a double-stranded molecule of about 170 kilobase pairs. This molecule shares several structural features with the DNA of poxviruses, such as the presence of hairpin loop structures at the DNA ends (13) and terminal inverted repetitions (TIR) (2).

Sogo et al. (26) have described the presence of internal inverted repetitions which consist of about 0.13 kilobases (kb) of sequence located 2.5 kb from the left TIR and 0.4 kb from the right TIR in the terminal fragments RK′ and RD′. ASF virus DNA can vary as much as 20 kb in length by means of deletions or additions in regions located close to the DNA ends (4).

In the accompanying paper (1), we described both a number of cross-hybridizations between ASF virus DNA restriction fragments and the DNA sequence of a multigene family (family 110) located in several small *Eco*RI fragments. In this paper, we describe a second group of cross-hybridizing fragments which includes restriction fragments RK′, RL, RJ, and RD′. The DNA sequences obtained from these fragments revealed the existence in ASF virus DNA of a multigene family (family 360) which is not homologous to multigene family 110.

MATERIALS AND METHODS

Cells, viruses, and recombinant clones. The Vero cell line (CCL 81) was obtained from the American Type Culture Collection (Rockville, Md.). BA71V strain of ASF virus, derived from the BA71 field isolate by adaptation to Vero cell cultures, has been previously described (7). Recombinant clones containing restriction fragments of BA71V DNA (15) were used.

Oligonucleotide synthesis. Oligodeoxynucleotides were synthesized by the phosphoramidite method in an Applied Biosystems 381A DNA synthesizer. The oligonucleotides were subsequently subjected to polyacrylamide gel electrophoresis, and the bands corresponding to full-size products were excised and eluted.

DNA sequencing. DNA sequencing was performed by either the chemical degradation procedure (17) or the chain termination procedure (24). Nucleotide sequences were obtained from both strands of ASF virus DNA restriction fragments cloned in pUC plasmids or M13mp phages (18, 19).

Northern (RNA) blotting. RNA from Vero cells infected with the BA71 strain of ASF virus was purified and probed with 5′-end-labeled oligonucleotides as described previously (1). The oligonucleotides were specific for genes K′360 (GA CTTATACTTTTCTTCATCTAGTAAGGCG), K′362 (ATC TCGCCCACCGTATTATTTTCGGACACA), L356 (ACAA TGATAGAATAACCGTATATATCTGCT), J319 (ACTCTC GATGAATCTCCTCTCCCATTTCCT), D′311 (CAAGGTA TCTGAAAGATGTATAAGAGCATC), and D′363 (CCCTT TTGTTTCGCCAGGGTCTTACCCTCT).

Computer analysis. Compilation and routine analysis of DNA sequences were performed with the software package of the University of Wisconsin genetics computer group (5) in a DEC VAX 730 computer running under the VMS operating system. The programs COMPARE, GAP, and BESTFIT were used to carry out dot matrix comparisons (16), global alignments of amino acid sequences (21), and alignment of homologous DNA sequences (25). Multiple alignments were carried out by the progressive alignment method of Feng and Doolittle (10). Searches of the protein data bank of the National Biomedical Research Foundation were done with the WORDSEARCH program of the University of Wisconsin genetics computer group package (29) and the FASTA program (22).

* Corresponding author.
† Present address: Instituto de Parasitología "López Neyra," CSIC c/ Ventanilla, 11, 18001 Granada, Spain.
‡ Present address: Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA 02115.
§ Present address: Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, Bethesda, MD 20892.

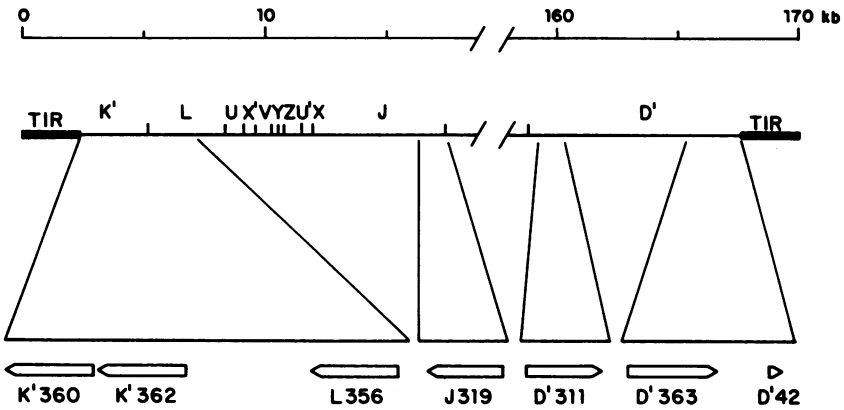


FIG. 1. Arrangement of multigene family 360 in ASF virus DNA. Only the terminal regions of the ASF virus genome are shown. The *Eco*RI restriction fragments are named as described previously (2). Regions containing cross-hybridizing sequences which were fully sequenced are expanded. Positions of different ORFs in the sequences are indicated (▭).

Genealogic tree construction. A genealogic tree based on the multiple alignment of amino acid sequences was derived by the method of Feng and Doolittle (10). This and alternative tree topologies were tested by both distance matrix programs (12) and the protein parsimony method (6, 11) with the programs of the Phyllip phylogeny inference package (9).

RESULTS

Mapping of cross-hybridizing DNA sequences. Hybridization studies with cloned ASF virus DNA fragments revealed a complex pattern of repeated sequences in the proximities of the TIR at both ends of the ASF virus genome (1). One group of cross-hybridizing fragments included the restriction fragments RK', RL, RJ, and RD'.

To determine the precise location of these repeated sequences within the fragments, hybridization experiments were carried out with subfragments RJ and RD' (data not shown), and results indicated that the repeated sequence in fragment RJ (5.3 kb) was located in the rightmost 1.5 kb of the fragment. Repeated sequences in fragment RD' (10.7 kb) were found in the leftmost 1 kb of the fragment and in the 2.5 kb immediately adjacent to the TIR.

Sequencing of cross-hybridizing segments. To determine the nucleotide sequence of the repeated sequences, restriction fragments RK', RL, RJ, and RD' were sequenced in the regions indicated in Fig. 1. The total length of the sequence determined was 9,540 base pairs. This sequence was distributed in four noncontiguous stretches (Fig. 2). The G+C content was 32%, significantly lower than the 41% reported for ASF virus DNA (7). DNA sequences obtained from the central part of the genome of ASF virus have a higher G+C content than the terminal regions (C. Simón-Mateo, personal communication).

Multigene family 360. A search for open reading frames (ORFs) showed the presence of six ORFs with lengths greater than 300 codons. Figure 1 shows the locations of these ORFs in the restriction map. Each ORF was named according to the restriction fragment in which it was contained and the length of the ORF in coding triplets.

Four ORFs were located in the left part of the genome, three of them (K'360, K'362, and L356) in the 5 kb adjacent to the left TIR (Fig. 2a). The right terminal *Eco*RI fragment (RD') contained two more ORFs (D'311 and D'363) and a small one (D'42) which is discussed below. ORFs located in the left part of the ASF virus restriction map were transcribed to the left, whereas ORFs located in the right part of the restriction map were transcribed to the right.

An additional ORF (177 codons in length) was detected in the region between K'362 and L356 and will be described elsewhere (A. Camacho and E. Viñuela, unpublished results).

Homology between ORFs. Figure 3 shows the result of a dot matrix analysis (16). A number of diagonal lines are evident, showing that these ORFs were homologous. ORF D'42 was clearly homologous to the 5' ends of the six remaining ORFs. The regions of similarity defined by the diagonal lines in the dot matrix were aligned by using the algorithm of Smith and Waterman (25). The degree of similarity varied from 52 to 81% (approximately). The percentage of similarity observed is in good agreement with the hybridization results (with the exception of the results from fragments RL and RD', which did not hybridize under stringent conditions), despite the 70.4% similarity between L356 and D'363. The observed internal inverted repetitions previously detected by electron microscopy at 2.5 and 0.4 kb from the left and right TIR, respectively (26), are in agreement with the positions of the reading frames K'362 and D'42. The high similarity of this repetition (81.3%) probably favored the appearance of the derived electron microscopy structures.

Transcription of family 360. To determine whether all of the genes are expressed and, if so, the time of expression during infection, probes specific for each gene were required to avoid cross-hybridization with homologous transcripts. For this purpose, 30-nucleotide-long sequences which were unique for each gene were chosen, and the corresponding oligonucleotides were synthesized and used to probe Northern blots containing RNA from ASF virus-infected cells. Transcripts for genes K'360, K'362, J319, D'311, and D'363 were present at early times in infected cells (Fig. 4). Some of the probes hybridized to multiple species of RNA, showing the existence of heterogeneity of the transcripts. The lengths of the transcripts (1.4 to 1.5 kb) are in good agreement with the lengths of the ORFs (0.9 to 1.1 kb). We were not able to detect an RNA for gene L356. Transcripts for genes K'360, K'362, and D'363 were also present at late times, although with sizes different from those of the early transcripts.

Comparison of putative protein products. The predicted

a/

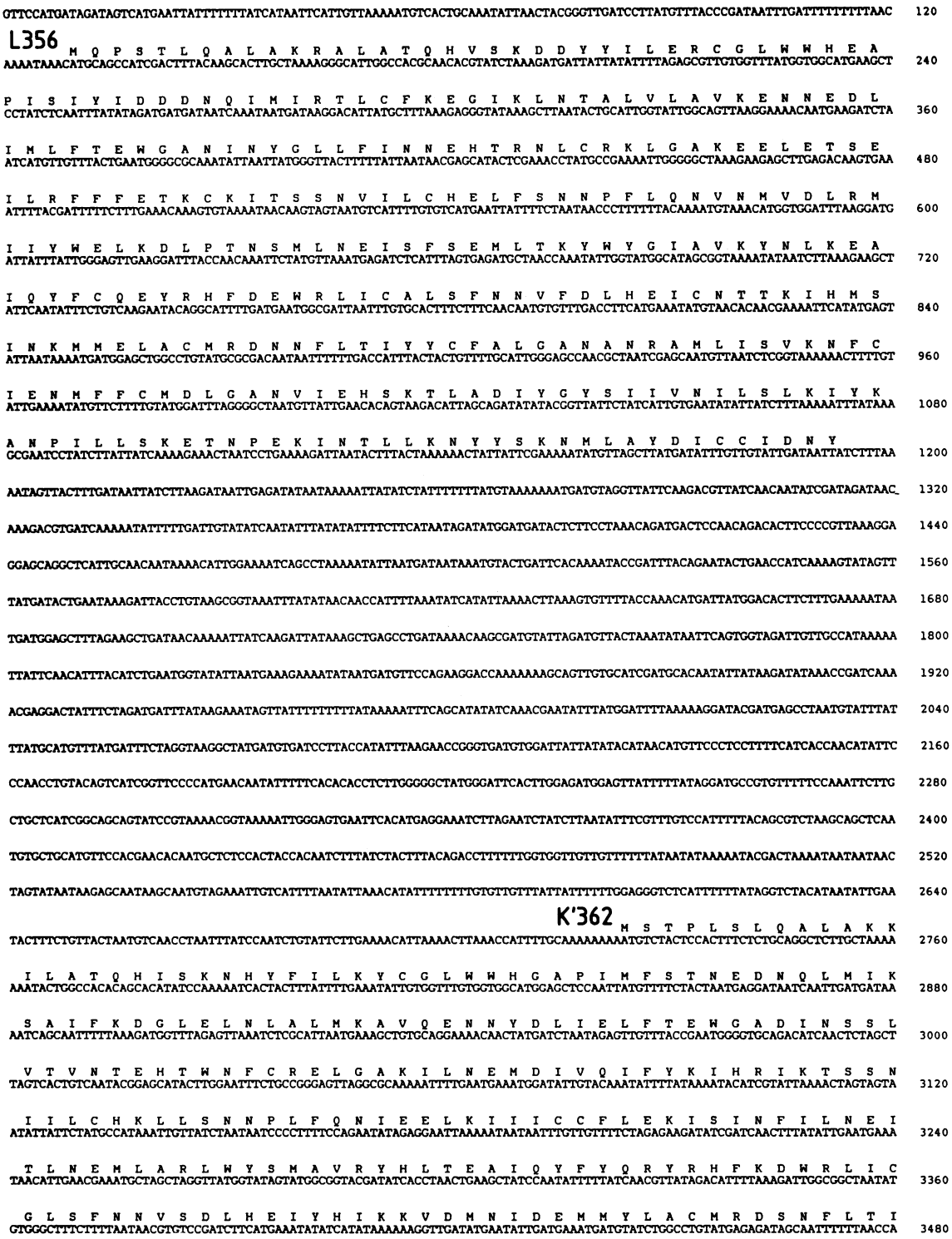
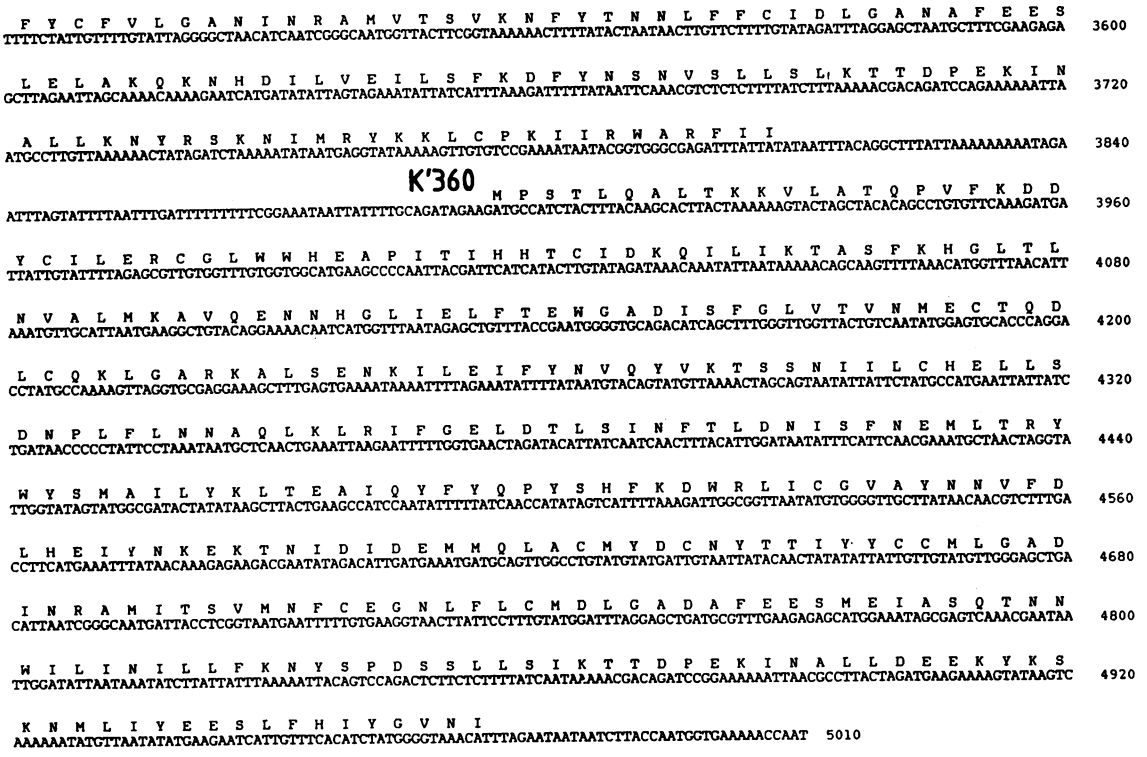


FIG. 2. Nucleotide sequence of four cross-hybridizing segments. Predicted ORFs are shown with amino acids in single-letter code. The sequences correspond precisely with the expanded regions in Fig. 1. (a) Partial sequence of fragments RL and RK′, displayed from right to left according to the restriction map. (b) Partial sequence of the RJ fragment, displayed from right to left according to the restriction map. (c) DNA sequence of the leftmost part of RD′ fragment, displayed from left to right according to the restriction map. (d) Partial sequence of the RD′ fragment adjacent to the right TIR, displayed from left to right. Sequence continues on the following pages.



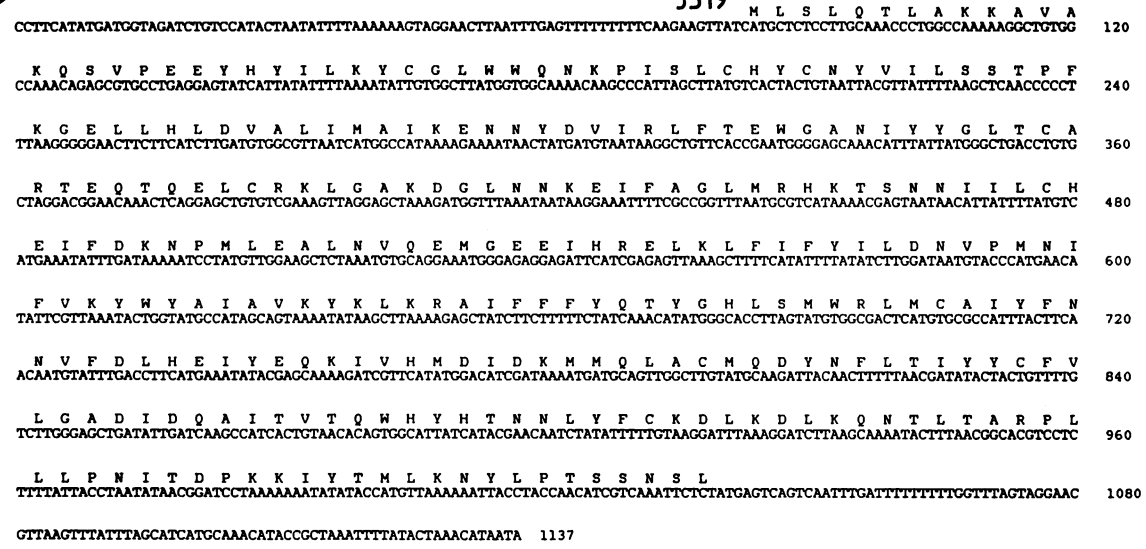


FIG. 2—*Continued.*

amino acid sequences translated from the ORFs were aligned by the progressive alignment procedure of Feng and Doolittle (10). Two conserved stretches are apparent in this alignment (Fig. 5): GLWW at position 40 and LFTEWG at position 94. This, along with the striking conservation of some residues, probably reflects structural or functional constraints (or both) in the evolution of these sequences. The motif CXXLGA is found three times in the sequence and probably reflects an ancient internal duplication.

A search of the National Biomedical Research Foundation protein sequence data base was carried out by using the algorithm of Wilbur and Lipman (29) and the FASTA program (22). Both methods failed to identify any entry in the protein data bank with significant similarity to the sequences derived from the ORFs of family 360. Also, a search with the sequence motifs conserved among the sequences did not render any significant results.

Evolutionary relationships among different genes of multi-

c)

D'311 M L S L Q T I A K M A V A T N T
CTAGGACAAAGAAAGTATATATAGCCAATAATTATTCCACTAAATTGATTTCCATACTGATGGGTATGGAGCCATGTTGTCTCTGCAGACAATCGCGAAAATGGCCGTAGCAACAAACAC 120

Y S K Y H Y P I L K V F G L W W K N S T L N G P I K I C N H C N N I M V G E Y P
CTACTCCAAGTATCACTATCCAATACTGAAGGTCTTTGGGCTGTGGTGGAAAAACAGCACGCTAAATGGCCCTATTAAAATATGTAACCATTGCAACAACATAATGGTAGGAGAATATCC 240

M C Y N H G M S L D I A L I R A V K E R N I S L V Q L F T E W G G N I D Y G A L
TATGTGTTACAATCATGGAATGAGCCTGGATATAGCTTTGATTCGGGCGGTAAAAGAGCGCAATATATCCTTAGTCCAGCTTTTCACCGAATGGGGGGAAATATTGACTATGGGGCACT 360

C R N T P S M Q R L C K S L G A K P P K G R M Y M D A L I H L S D T L N D N D L
TTGTCGTAACACTCCATCTATGCAAAGATTATGTAAAAGTTTGGGAGCCAAACCACCAAAGGGCCGAATGTATATGGATGCTCTTATACATCTTTCAGATACCTTGAATGATAATGATCT 480

I R G Y E I F D D N S V L D C V N L I R L K I M L T L K A R I P L M E Q L D Q I
GATTAGGGGGTATGAGATTTTTGATGATAATAGCGTGTTGGATTGTGTCAATCTCATACGACTCAAAATAATGCTTACCTTGAAGGCCCGTATACCTCTCATGGAACAACTAGACCAAAT 600

A L K Q L L Q R Y W Y A M A V Q H N L T T A I H Y F D N H I P N I K P F S L R C
TGCCTTAAAACAACTTCTGCAGCGATACTGGTATGCCATGGCTGTACAACACAACTTAACAACAGCTATCCACTATTTTGATAATCATATTCCTAATATAAAGCCATTTAGTCTGCGCTG 720

A L Y F N D P F K I H D A C R T V N M D P N E M M N I A C Q Q D L N F Q S I Y Y
TGCTTTGTATTTTAATGATCCCTTTAAAATCCATGATGCTTGCAGGACTGTAAATATGGATCCTAATGAGATGATGAACATTGCTTGTCAACAGGATTTAAACTTTCAAAGCATTTACTA 840

S Y I L G A D I N Q A M L M S L K Y G N L S N M W F C I D L G A D A F K E A G A
TAGTTATATTTTAGGGGCTGATATTAATCAGGCTATGCTAATGTCTTTAAAGTATGGTAATCTTTCTAATATGTGGTTTTGCATAGATTTGGGGGCGGATGCCTTTAAAGAGGCAGGGGC 960

L A G K K K K S V T A H I R S
GCTTGCTGGGAAAAAAAAAAAGAGTGTTACAGCACATATTAGGTCTTAATATCTTTAAGCGGGAGTTGATTCCCCCCTGTAAAGATCCTGATCCTTATCAAATCCAAATTCTGTTAAAAA 1080

ACTACATTCTAAAAAATGTC 1100

d)

D'363 M P S T L Q V L A K K
TTTTAATAATTGATGACTAAAATCATATTATAATGCCGTGCAAAAAATAATTATTTTTCGGTTAAAGGATACCTTTAAATAAAAAACGATGCCATCCACTCTACAAGTGCTTGCTAAAAA 120

V L A L G E H K E N E H I S R E Y Y Y H I L K C C G L W W H E A P I I L C F D G
GGTGTTGGCCTTAGGGGAACATAAAGAAAATGAACATATATCTAGAGAATATTATTATCATATATTAAAGTGTTGCGGTTTATGGTGGCATGAGGCTCCGATTATACTTTGTTTTGATGG 240

S E Q M M I K T P I F E E G I L L N T A L M K A V Q E N N Y E L I N L F T E W G
GAGTGAGCAAATGATGATAAAGACTCCAATCTTTGAAGAAGGCATATTACTTAATACTGCATTAATGAAAGCTGTACAGGAGAACAATTATGAATTAATAAACTTGTTCACTGAATGGGG 360

A N I N Y G L I S I N T E H A R D L C R K L G A K E M L E R N E V I Q I V F K T
AGCAAACATCAATTATGGATTAATTTCCATTAATACTGAGCATGCCCGGGACCTATGTCGAAAATTAGGAGCTAAAGAAATGCTTGAAAGAAATGAAGTTATACAAATTGTATTTAAAAC 480

L D D I T S S N I I L C H E L F T N N P L L E N V N M G E M R M I I H W R M K N
ATTAGATGATATCACCAGTAGTAATATAATTTTATGTCATGAATTATTTACCAACAATCCTCTTTTAGAGAATGTAAATATGGGGGAAATGAGGATGATAATTCATTGGAGGATGAAAAA 600

L T N L L L N N D S I S E I L T K F W Y G I A V K Y N L K D A I Q Y F Y Q R F M
TTTAACGAACCTATTATTAAATAATGACTCTATTAGTGAAATATTAACTAAATTCTGGTATGGTATAGCAGTAAAATATAATCTTAAGGATGCAATCCAATATTTTTACCAGAGATTCAT 720

D F N E W R V T C A L S F N N V N D L H K M Y I T E K V H T N N D E M M N L A C
GGACTTCAACGAGTGGCGAGTAACATGTGCTCTTTCTTTTAATAATGTGAATGATCTTCATAAGATGTATATAACAGAGAAGGTTCATACGAATAATGACGAAATGATGAATCTAGCCTG 840

S I Q D R N L S T I Y Y C F L L G A N I N Q A M L T S V L N Y N I F N L F F C I
CAGCATTCAAGACAGAAATTTATCAACCATTTACTATTGTTTTCTATTGGGGGCTAACATCAATCAAGCAATGTTGACCTCAGTATTAAATTATAATATTTTTAACTTATTCTTTTGTAT 960

D L G A D A F E E G K T L A K Q K G Y N E I V E I L S L D I I Y S P N T D F S S
AGACTTAGGGGCTGATGCCTTTGAAGAGGGTAAGACCCTGGCGAAACAAAAGGGGTATAATGAAATAGTGGAAATCTTATCATTAGATATCATTTATAGTCCAAATACTGACTTCTCATC 1080

K I E P E H I S S L L K N F Y P K N L F A F D R C N P G L Y Y S
AAAAATAGAACCTGAACATATTAGTTCTTTGTTAAAAAACTTTTATCCAAAAAATCTGTTCGCTTTTGATCGTTGCAATCCTGGTTTATATTATTCTTAGAGGACCGCTACAAAAATTAT 1200

TTTTTTTTCTTGATCAAAGCTCCAAAATAATTATTAGATTAAAGTCGCCTATAGCAGCTGCCCACTCCAAAAAAAGTATTTTATAGTACAAAAAACACGAAAAATAATTTGCGGCCGGCG 1320

GCAACTATTTGTTGTTGTCCAAAACTTAATGTTTTTTTAATATTTTTAAATGCAACCATGGATTGTTGGACTATCAGGGAGAAGAACTATAGCTACATCATATTGTCAATACTGGTAATA 1440

CTATTAATATGGTATCTTATACTTAACTATTGTCGATCGAAAAAAAAATGCAGTTACAAACAACATGCCGCCACCAGCGTACACGGTGTCAAGTAGCTGTTCTCAATAATAGGGTTGATCG 1560

ACGCTCTTCGTAATAATATGTTGATTGGCGCATCATAAAATGCTGAGGAACTATCCGTTAATATATGTTGATTGACGCCCACTTATGAATGGAAGTAATGATTTTGTATAAAATACGGGT 1680

TTGTGAGGGCCTTATTTTTTTCTTATTAGAACAAAGCGTGTATTTTTAAGGCCTATAGCAAGAGTATGTTTAATTAACACCTACAACAGTAATATTTAAGGCCAGTAAAATAATGTTTAA 1800

D'42 M P T P L S L Q A P A K K V L A T Q H I S K D H L Y
TTAAGGCCTGACCACTAAAACTTAAACGATTTTGTAAAAAAATATGCCTACTCCACTTTCTCTACAGGCTCCCGCTAAAAAAGTACTGGCCACACAGCACATATCTAAAGATCACCTTTA 1920

F E I L W F M V A F F D A Y S L
TTTTGAAATATTGTGGTTTATGGTGGCATTCTTTGACGCCTACAGTCTTTAACGCCTGCAGTAATAATTGATATCTCCAGCGCTAATAATAATTGATACCCACAGCGGTAATAATAATTG 2040

ATACCCACAGCGGTAATGTCGTCCATCTGGCCGCCCCAAAAAAAAGTATTTACGGTGGGGTTTATTACCGGCGGCGTAACACCAGTTATGGTTAGTTTTGTCTGGCCCGCCGCCCAGCCG 2160

CAAAAAAAAATCAATTACAGCCGCAAAAAAAAATATTTCCGGCCGCGGAGTTTCTACAAAAAAAATGTTAGCTTTTAATTGTTGTAGACCAGGTTCTTAGGGAAGGTTCAACTCCTTTTT 2280

CTCGAATAAAAAA 2293

FIG. 2—*Continued.*

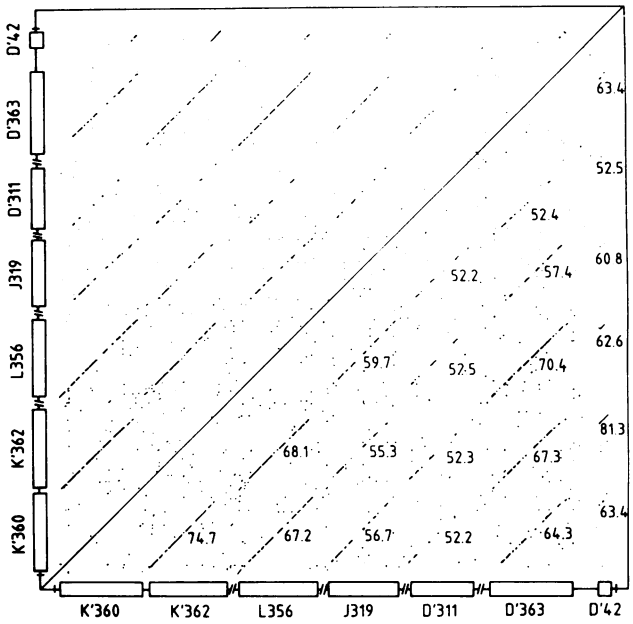


FIG. 3. Dot matrix comparison of ORFs. The sense strand of each ORF was compared with all remaining ORFs with a window comparison of 20 and a stringency of 15. The percent identity of each diagonal was obtained by aligning the involved DNA segments by using the algorithm of Smith and Waterman (25).

gene family 360. Table 1 shows the percent identity derived from the multiple alignment shown in Fig. 5. Pairs K'360-K'362 and L356-D'363 were the ones most related. This can also be seen at the DNA level (Fig. 3).

TABLE 1. Percent identity between predicted protein products derived from multiple alignment

Protein product	% Identity with:				
	K'362	L356	J319	D'311	D'363
K'360	61.8	55.8	44.4	33.5	48.3
K'362		57.1	47.6	34.8	52.8
L356			48.1	36.1	61.4
J319				37.5	45.1
D'311					37.2

Figure 6 shows a genealogic tree derived from the multiple alignment of the predicted protein products of each gene. Clearly, there was no grouping of the sequences according to their location at the left or right part of the genome. Instead, the branching pattern of the tree suggests that genes located at opposite positions in the restriction map have a common origin. To confirm this further, we used the sequence of the ORF D'42, which was not included in the genealogic tree because of its very different size. Pairwise alignments of D'42 with all the other protein sequences clearly showed that D'42 is very similar to K'362 (data not shown). This relationship can also be seen in the dot matrix comparison of the respective DNA segments (Fig. 3).

DISCUSSION

A search for repetitive sequences in ASF virus DNA has led us to determine the nucleotide sequence of several regions located close to the ends of the DNA molecule. The analysis of these sequences indicated that the repetitions were related to six genes with lengths of 311 to 363 coding triplets. These genes were homologous, as is shown by comparisons both at the DNA level (Fig. 3) and at the

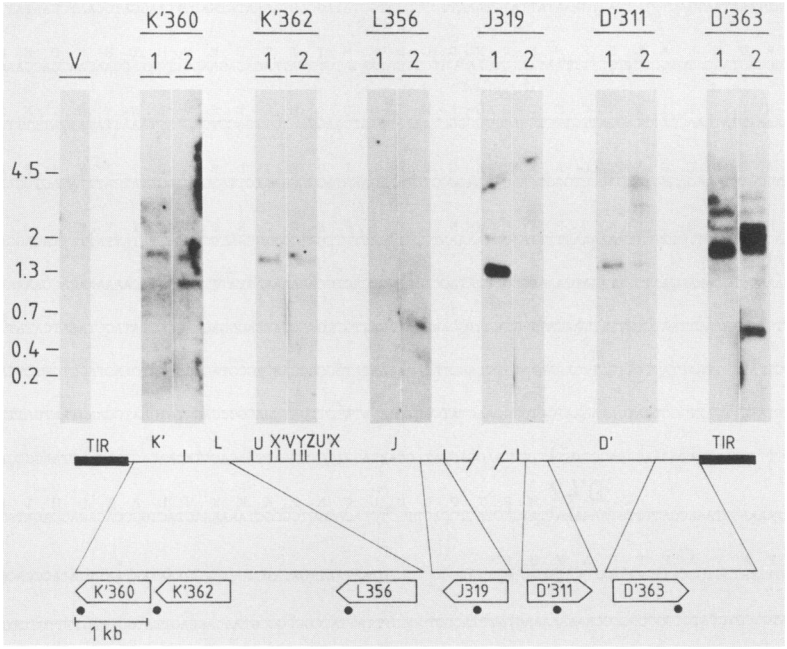


FIG. 4. Transcription of the members of multigene family 360. RNA from infected cells at early (lanes 1) or late (lanes 2) times was hybridized with oligonucleotide probes specific for each gene. Sizes are indicated in kilobases. V, RNA from uninfected Vero cells; ⊏⊐, positions of different ORFs in the sequences. ●, Locations of the sequences of each probe within the genes.

```
                                                                                      70
K'362   MSTPLSLQALAKKILA      TQHISKNHYF ILKYCGLWWHGA      PI MFSTNEDNQLMIKSAIFK
K'360     MPSTLQALTKKVLA      TQPVFKDDYC ILERCGLWWHEA      PITIHHTCIDKQILIKTASFK
L356    MQPSTLQALAKRALA       TQHVSKDDYY ILERCGLWWHEA      PISIY IDDDNQIMIRTLCFK
D'363     MPSTLQVLAKKVLALGEHKENEHISREYYYHILKCCGLWWHEA      PI ILCFDGSEQMMIKTPIFE
J319        MLSLQTLAKKAVA      KQSVPEEYHYILKYCGLWWQNK      PISLCHYC NYVILSSTPFK
D'311       MLSLQTIAKMAVA      TNTYSKYHYPILKVFGLWWKNSTLNGPIKICNHCNNIMVGEYPMCYN
D'42    MPTPLSLQAPAKKVLATQHISKDHLYFEILW FMVAFFDAYSL
                                                                                     140
K'362   DG LELNLALMKAVQENNYDLIELFTEWGADINSSLVTVNTEHTWNFCRELGA KILNEMDIVQIFYKIH
K'360   HG LTLNVALMKAVQENNHGLIELFTEWGADISFGLVTVNMECTQDLCQKLGARKALSENKILEIFYNVQ
L356    EG IKLNTALVLAVKENNEDLIMLFTEWGANINYGLLFINNEHTRNLCRKLGAKEELETSEILRFFFETK
D'363   EG ILLNTALMKAVQENNYELINLFTEWGANINYGLISINTEHARDLCRKLGAKEMLERNEVIQIVFKTL
J319    GELLHLDVALIMAIKENNYDVIRLFTEWGANIYYGLTCARTEQTQELCRKLGAKDGLNNKEI   FAGLM
D'311   HG MSLDIALIRAVKERNISLVQLFTEWGGNIDYGALCRNTPSMQRLCKSLGAKPP KGRMYMDALIHLS
                                                                                     210
K'362   RIKTSSNIILCHKLLSNNPLF  QNIEELKI IICCFLEKISINFILNEITLNEMLARLWYSMAVRYHLTE
K'360   YVKTSSNIILCHELLSDNPLF  LNNAQLKL RIFGELDTLSINFTLDNISFNEMLTRYWYSMAILYKLTE
L356    CKITSSNVILCHELFSNNPFLQNVNMVDLRM IIYWELKDLPTNSMLNEISFSEMLTKYWYGIAVKYNLKE
D'363   DDITSSNIILCHELFTNNPLLENVNMGEMRM IIHWRMKNL TNLLLNNDSISEILTKFWYGIAVKYNLKD
J319    RHKTSNNIILCHEIFDKNPMLEALNVQEMGE EIHRELK LFIFYILDNVPM NIFVKYWYAIAVKYKLKR
D'311   DTLNDNDLIRGYEIFDDNSVLDCVNLIRLKI MLTLKAR  IPLMEQLDQIALKQLLQRYWYAMAVQHNLTT
                                                                                     280
K'362   AIQYFYQRYRHFKDWRLICGLSFNNVSDLHE IYHIKKVDMNIDEMMYLAC MRDSNFLTIFYCFVLGANI
K'360   AIQYFYQPYSHFKDWRLICGVAYNNVFDLHE IYNKEKTNIDIDEMMQLAC MYDCNYTTIYYCCMLGADI
L356    AIQYFCQEYRHFDEWRLICALSFNNVFDLHE ICNTTKIHMSINKMMELAC MRDNNFLTIYYCFALGANA
D'363   AIQYFYQRFMDFNEWRVTCALSFNNVNDLHK MYITEKVHTNNDEMMNLACSIQDRNLSTIYYCFLLGANI
J319    AIFFFYQTYGHLSMWRLMCAIYFNNVFDLHE IYEQKIVHMDIDKMMQLAC MQDYNFLTIYYCFVLGADI
D'311   AIHYFDNHIPNIKPFSLRCALYFNDPFKIHD AC  RTVNMDPNEMMNIAC QQDLNFQSIYYSYILGADI
                                                                                     350
K'362   NRAMVTSVKNFYTNNLFFCIDLGANAFEESL ELAKQKNHDILVEILSFKDFYNSNVSLLSLKTTDPEKIN
K'360   NRAMITSVMNFCEGNLFLCMDLGADAFEESM EIASQTNNWILINILLFKN YSPDSSLLSIKTTDPEKIN
L356    NRAMLISVKNFCIENMFFCMDLGANVIEHSK TLADIYGYSIIVNILSLKI YKANPILLS KETNPEKIN
D'363   NQAMLTSVLNYNIFNLFFCIDLGADAFEEGK TLAKQKGYNEIVEILSLDIIYSPNTDFSS  KIEPEHIS
J319    DQAITVTQWHYHTNNLYFCKDL        K DLKQNT                LTARPLLLP NITDPKKIY
D'311   NQAMLMSLKYGNLSNMWFCIDLGADAFKEAG ALAGKK                           KKSVTAHIR
                                  381
K'362   ALL  KNYRSKNIMRYKKLCPKIIRWARFII
K'360   ALLDEEKYKSKNMLIYEE   SLFHIYGVNI
L356    TLL  KNYYSKNMLAYD    ICCIDNYL
D'363   SLL  KNFYPKNLFAFD    RCNPGLYYS
J319    TML  KNYLPTSSNSL
D'311   S
```

FIG. 5. Multiple alignment of predicted protein products for multigene family 360. Residues which are conserved among all the proteins of multigene family 360 are boxed in black.

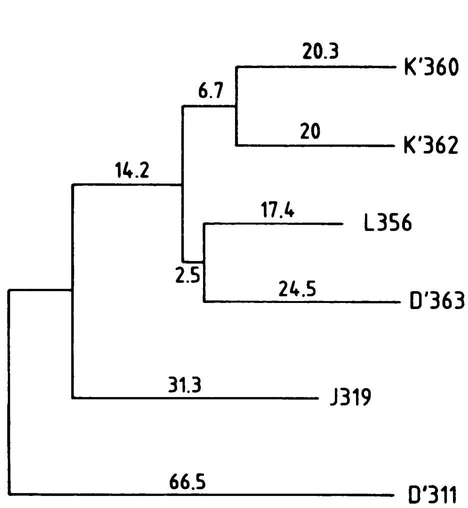


FIG. 6. Genealogic tree for proteins of multigene family 360. Tree construction was carried out as described in Materials and Methods. Evolutionary distance for each branch is indicated.

protein level (Fig. 5 and Table 1). In addition, a small ORF (D'42) homologous to the 5' end of the remaining members of the family was present. The sequences obtained account for both the cross-hybridization between restriction fragments RK', RL, RJ, and RD' (1) and the previously reported internal inverted repetitions (26).

The hybridization of RNA of infected cells with specific probes for each gene indicated that most members of multigene family 360 are transcribed during infection, although to different extents (Fig. 4). In addition, the different sizes and intensities of the bands obtained after hybridization of RNA isolated at late times suggest that early and late transcription are independently controlled for each gene.

Comparison of the putative protein products of the genes

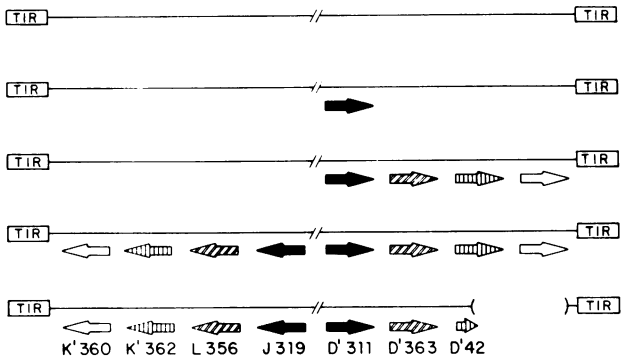


FIG. 7. Model for evolution of multigene family 360. Several steps in the evolution are shown schematically. Arrows represent genes, and parentheses indicate deletion events. The first steps involve gene duplication at one DNA end, followed by sequence divergence of duplicated genes. In a second step, translocation of sequences from one DNA end to the opposite one would take place. Present genes and the evolutionary relationships among them can be accounted for by additional sequence divergence and a single deletion event.

(Fig. 5) revealed a striking conservation of some amino acid stretches. This suggests the existence of selective pressure to maintain a common overall structure in each ORF of the family since the time of their divergence from a common ancestor.

The arrangement of the genes in the viral genome as well as their transcription direction is compatible with the hypothesis that they once constituted an enlarged TIR. It has been well documented that poxvirus genomes undergo deletions at one end and duplication of sequences from the other end, thus generating TIR of larger sizes (3, 8, 14, 20, 23). The evolutionary relationships among members of multigene family 360 also support the idea that duplication and translocation of sequences between ASF virus DNA ends have occurred.

A simplified model for the evolution of family 360 (Fig. 7) implies gene duplication within one DNA end and translocation of sequences between both DNA ends. Also, a single deletion event accounts for the presence of the truncated ORF D'42. This model may be rather simplistic, since the larger evolutionary distance between D'311 and J319, compared with the evolutionary distance between L356 and D'363, indicates that multiple translocation events may have taken place.

Multigene families have been described before for other large DNA-containing viruses such as human cytomegalovirus (28) and Shope fibroma virus (27). The multigene family of Shope fibroma virus is organized similarly to multigene family 360 in ASF virus, since its members are located near the ends of the genome and are transcribed towards the DNA ends. In fact, part of the multigene family in Shope fibroma virus is contained within the inverted terminal repetitions and therefore could represent one intermediate step in our evolutionary model (Fig. 7).

The role of multigene family 360 in ASF virus biology is unknown. Experiments to detect the protein products in infected cells are in progress.

ACKNOWLEDGMENTS

We are grateful to R. F. Doolittle and J. Felsestein for providing computer programs for genealogic analysis. We thank J. A. Baquedano for advice in designing computer graphic programs and R. Sánchez and A. Zazo for technical assistance.

This work was supported by grants from the Comisión Interministerial de Ciencia y Tecnología, European Economic Community, and Fundación Ramón Areces.

LITERATURE CITED

1. **Almendral, J. M., F. Almazán, R. Blasco, and E. Viñuela.** 1990. Multigene families in African swine fever virus: family 110. J. Virol. **64**:2064–2072.
2. **Almendral, J. M., R. Blasco, V. Ley, A. Beloso, A. Talavera, and E. Viñuela.** 1984. Restriction site map of African swine fever virus DNA. Virology **133**:258–270.
3. **Archard, L. C., M. Mackett, D. E. Barnes, and K. R. Dumbell.** 1984. The genome structure of cowpox virus white pock variants. J. Gen. Virol. **65**:875–886.
4. **Blasco, R., M. Agüero, J. M. Almendral, and E. Viñuela.** 1989. Variable and constant regions in African swine fever virus DNA. Virology **168**:330–338.
5. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
6. **Eck, R. V., and M. O. Dayhoff.** 1966. Atlas of protein sequence and structure. National Biomedical Research Foundation, Silver Spring, Md.
7. **Enjuanes, L., A. L. Carrascosa, M. A. Moreno, and E. Viñuela.** 1976. Titration of African swine fever (ASF) virus. J. Gen. Virol. **32**:471–477.
8. **Esposito, J. J., C. D. Cabradilla, J. H. Nakano, and J. F. Obijeski.** 1981. Intragenomic sequence transposition in monkeypox virus. Virology **109**:231–243.
9. **Felsestein, J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**:783–791.
10. **Feng, D. F., and R. F. Doolittle.** 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J. Mol. Evol. **25**:351–360.
11. **Fitch, W. M.** 1971. Toward defining the course of evolution: minimum change for a specified tree topology. Syst. Zool. **20**:406–416.
12. **Fitch, W. M., and E. Margoliash.** 1967. Construction of phylogenetic trees. Science **155**:279–284.
13. **González, A., A. Talavera, J. M. Almendral, and E. Viñuela.** 1986. Hairpin loop structure of African swine fever virus DNA. Nucleic Acids Res. **14**:6835–6844.
14. **Kotwal, G. J., and B. Moss.** 1988. Analysis of a large cluster of nonessential genes deleted from a vaccinia virus terminal transposition mutant. Virology **167**:524–537.
15. **Ley, V., J. M. Almendral, P. Carbonero, A. Beloso, E. Viñuela, and A. Talavera.** 1984. Molecular cloning of African swine fever virus DNA. Virology **133**:249–257.
16. **Maizel, J. V., and R. P. Lenk.** 1981. Enhanced graphic matrix analysis of nucleic acid and protein sequences. Proc. Natl. Acad. Sci. USA **78**:7665–7669.
17. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA **74**:560–564.
18. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
19. **Messing, J., and J. Vieira.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.
20. **Moyer, R. W., R. L. Graves, and C. T. Rothe.** 1980. The white pock (μ) mutants of rabbit poxvirus. III. Terminal DNA sequence duplication and transposition in rabbit poxvirus. Cell **22**:545–553.
21. **Needleman, S. B., and C. D. Wunsch.** 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. **48**:443–453.
22. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequences comparison. Proc. Natl. Acad. Sci. USA **85**:2444–2448.
23. **Pickup, D. J., B. S. Ink, B. L. Parsons, W. Hu, and W. J. Joklik.** 1984. Spontaneous deletions and duplications of sequences in

the genome of cowpox virus. Proc. Natl. Acad. Sci. USA **81**:6817–6821.
24. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
25. **Smith, T. F., and M. S. Waterman.** 1981. Identification of common molecular subsequences. J. Mol. Biol. **147**:195–197.
26. **Sogo, J. M., J. M. Almendral, A. Talavera, and E. Viñuela.** 1984. Terminal and internal inverted repetitions in African swine fever virus DNA. Virology **133**:271–275.
27. **Upton, C., and G. McFadden.** 1986. Tumorigenic poxviruses: analysis of viral DNA sequences implicated in the tumorigenicity of Shope fibroma virus and malignant rabbit virus. Virology **152**:308–321.
28. **Weston, K., and B. G. Barrel.** 1986. Sequence of the short unique region, short repeats, and part of the long repeats of human cytomegalovirus. J. Mol. Biol. **192**:177–208.
29. **Wilbur, W. J., and D. J. Lipman.** 1983. Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. USA **80**:726–730.