JOURNAL OF VIROLOGY, Jan. 1999, p. 533–552 Vol. 73, No. 1
0022-538X/99/$04.00+0

The Genome of *Melanoplus sanguinipes* Entomopoxvirus

C. L. AFONSO, E. R. TULMAN, Z. LU, E. OMA, G. F. KUTISH, AND D. L. ROCK*

Plum Island Animal Disease Center, Agricultural Research Service, U.S. Department of Agriculture, Greenport, New York 11944

Received 3 August 1998/Accepted 25 September 1998

The family *Poxviridae* contains two subfamilies: the *Entomopoxvirinae* (poxviruses of insects) and the *Chordopoxvirinae* (poxviruses of vertebrates). Here we present the first characterization of the genome of an entomopoxvirus (EPV) which infects the North American migratory grasshopper *Melanoplus sanguinipes* and other important orthopteran pests. The 236-kbp *M. sanguinipes* EPV (MsEPV) genome consists of a central coding region bounded by 7-kbp inverted terminal repeats and contains 267 open reading frames (ORFs), of which 107 exhibit similarity to previously described genes. The presence of genes not previously described in poxviruses, and in some cases in any other known virus, suggests significant viral adaptation to the arthropod host and the external environment. Genes predicting interactions with host cellular mechanisms include homologues of the inhibitor of apoptosis protein, stress response protein phosphatase 2C, extracellular matrixin metalloproteases, ubiquitin, calcium binding EF-hand protein, glycosyltransferase, and a triacylglyceride lipase. MsEPV genes with putative functions in prevention and repair of DNA damage include a complete base excision repair pathway (uracil DNA glycosylase, AP endonuclease, DNA polymerase β, and an NAD^+-dependent DNA ligase), a photoreactivation repair pathway (cyclobutane pyrimidine dimer photolyase), a LINE-type reverse transcriptase, and a *mutT* homologue. The presence of these specific repair pathways may represent viral adaptation for repair of environmentally induced DNA damage. The absence of previously described poxvirus enzymes involved in nucleotide metabolism and the presence of a novel thymidylate synthase homologue suggest that MsEPV is heavily reliant on host cell nucleotide pools and the de novo nucleotide biosynthesis pathway. MsEPV and lepidopteran genus B EPVs lack genome colinearity and exhibit a low level of amino acid identity among homologous genes (20 to 59%), perhaps reflecting a significant evolutionary distance between lepidopteran and orthopteran viruses. Divergence between MsEPV and the *Chordopoxvirinae* is indicated by the presence of only 49 identifiable chordopoxvirus homologues, low-level amino acid identity among these genes (20 to 48%), and the presence in MsEPV of 43 novel ORFs in five gene families. Genes common to both poxvirus subfamilies, which include those encoding enzymes involved in RNA transcription and modification, DNA replication, protein processing, virion assembly, and virion structural proteins, define the genetic core of the *Poxviridae*.

The *Poxviridae* family consists of large cytoplasmic double-stranded DNA viruses separated into two subfamilies: the *Entomopoxvirinae* (poxviruses of insects) and the *Chordopoxvirinae* (poxviruses of vertebrates) (130). The entomopoxvirus (EPV) subfamily is divided into three genera based primarily on differences in viral host range and virion morphology. Genus A viruses infect coleopterans, genus B viruses infect lepidopterans and orthopterans, and genus C viruses infect dipterans (7, 54). Insects are the only known hosts of EPVs, and observed viral host range is restricted to one or a few related species (7).

A detailed genetic comparison of the two subfamilies has been limited by the lack of information on EPV genomics. Restriction endonuclease analysis and DNA cross-hybridization studies have, however, suggested large genomic differences between lepidopteran group B EPVs and chordopoxviruses (ChPVs) (63, 102). Limited gene comparisons have also shown that at certain loci, lepidopteran EPV gene order is distinct from that of ChPVs (66, 175) and that the degree of amino acid similarity between EPV and ChPV enzymatic and structural proteins is low (6, 66, 175).

EPVs have been studied mainly because they are potential insect biocontrol agents and expression vectors (7, 41, 180). However, EPV genomic organization and molecular mechanisms of replication, pathogenesis, and host range are largely unknown. Few EPV genes have been characterized in detail, and additional information on the viral genome and virus-host interactions is necessary to further develop and improve these viruses as biocontrol agents (6, 175).

Melanoplus sanguinipes EPV (MsEPV) infects the North American migratory grasshopper *M. sanguinipes*, an agriculturally important insect pest, as well as two related grasshopper species (*M. differentialis* and *M. packardii*), the desert locust (*Schistocerca gregaria*) (179), and the African migratory locust (*Locusta migratoria*) (82, 108). MsEPV produces a large ellipsoid virion (250 to 300 nm in length) with a rectangular core. Grasshopper nymphs are infected by MsEPV after oral ingestion of virus-containing occlusion bodies. Presumably, the virus infects cells of the midgut prior to generalization of infection to the major target organ, the fat body (40, 72). Infection results in a slow and debilitating disease with high mortality, occurring 25 to 30 days postinfection. High titers of infectious spheroids, which can number up to 8×10^7 per grasshopper, are evident at 12 to 15 days postinfection (72, 127, 208).

Here, we present a genomic analysis of MsEPV. These data represent the first characterization of an EPV genome; further, they define the genetic core of the *Poxviridae*.

MATERIALS AND METHODS

MsEPV DNA isolation and cloning. MsEPV genomic viral DNA was extracted from gradient-purified viral occlusion bodies obtained from the North American migratory grasshopper, *M. sanguinipes*, as previously described (102). Random DNA fragments were obtained by incomplete enzymatic digestion with *Tsp*509I

* Corresponding author. Mailing address: Plum Island Animal Disease Center, P.O. Box 848, Greenport, NY 11944-0848. Phone: (516) 323-2500, ext. 330. Fax: (516) 323-2507. E-mail: drock@cshore.com.

endonuclease (New England Biolabs, Beverly, Mass.). DNA fragments of 1.5 to 2.5 kbp were isolated after separation on agarose gels, cloned into the dephosphorylated *Eco*RI site of plasmid pUC19, and grown in *Escherichia coli* DH10B cells (GIBCO BRL, Gaithersburg, Md.). Double-stranded plasmids pUC19 was purified by the rapid boiling method (162). DNA templates were sequenced from both ends with M13 forward and reverse primers, using dideoxy chain terminator sequencing chemistries (163) and an Applied Biosystems PRISM 377 automated DNA sequencer (Perkin-Elmer, Foster City, Calif.). Applied Biosystems sequence software (version 3.0) was used for lane tracking and trace extraction. Chromatogram traces were base called with Phred software (43), which also produced a quality file containing a predicted probability of error at each base position. The sequences were assembled with Phrap software (42), using the quality files and default settings to produce a consensus sequence. Subsequent manual editing was done with the Consed sequence editor (56). The final DNA consensus sequence represented on average an eightfold redundancy at each base position.

MsEPV genome organization was confirmed by comparing observed *Bam*HI, *Hin*dIII, and *Sca*I restriction fragments to the consensus sequence data. Right and left ends of the genome were confirmed by using *Alu*I, *Bgl*II, *Cla*I, *Nhe*I, *Pml*I, and *Sau*3AI restriction digests (New England Biolabs).

DNA sequence analysis. Genome DNA composition, structure, repeats, and restriction enzyme patterns were analyzed with the Wisconsin Genetics Computer Group (GCG) programs (33). Open reading frames (ORFs) consisting of more than 60 amino acids, and with a methionine start codon and codon usage consistent with known EPV gene sequences from GenBank were considered to be protein encoding (176, 177). DNA and protein comparisons with entries in genetic databases were performed with FASTA (141) and BLAST programs (2). Multiple sequence alignments were performed with the CLUSTAL (74, 187), GCG Pileup (33), MSA (116), and Macaw (166) computer programs. Motif searches were done against the SBASE release 5 (44) and Prosite release 14 (9) databases. Other protein patterns were determined with the profile scanning programs PROBE (134), GIBBS, ASSET (133), and Hidden Markov Model (37, 38). Prediction of transmembrane domains was accomplished with MEMSAT (84, 85) and TopPred (24) software. Signal peptides were predicted with the program Sigseq (195). Physical descriptions of proteins were obtained by using SAPS software (19). Phylogenetic analysis was done with the Phylip computer programs (45) and Phylo_Win graphic tools (51). Gene families were identified by using the following criteria: (i) similarity based on BLAST scores (3) and pairwise clustering with CLUS (94); (ii) cluster profiles produced by PROBE (134); (iii) statistical significance, determined by using PRDF (142); and (iv) the presence of unique motifs, determined by using Pileup, GIBBS, and MACAW (33, 88, 133, 166).

Abbreviations. Organisms have been abbreviated as follows: *Amsacta moorei* EPV, AmEPV; *Autographa californica* nuclear polyhedrosis virus, AcNPV; African swine fever virus, ASFV; *Choristoneura biennis* EPV, CbEPV; *Choristoneura fumiferana* EPV, CfEPV; cowpox virus, CPV; fowlpox virus, FPV; *Heliothis armigera* EPV, HaEPV; *Melolontha melolontha* EPV, MmEPV; *Molluscum contagiosum* virus, MCV; *Orgyia pseudosugata* NPV, OpNPV; rabbit fibroma virus, RFV; swinepox virus, SPV; and variola virus, VAR.

Nucleotide sequence accession number. The MsEPV genome sequence has been deposited in GenBank under accession no. AF063866.

RESULTS AND DISCUSSION

Organization of the MsEPV genome. The MsEPV genome was assembled into a contiguous sequence of 236,120 bp, similar in size to a previous estimate of 235 kbp (108). Because genomic termini were not sequenced, the left-most nucleotide of the assembled sequence was arbitrarily designated base no. 1.

The nucleotide composition is 81.7% A+T, as previously estimated for MsEPV (101), and is uniformly distributed over the entire length of the MsEPV genome. The total amino acid composition of all MsEPV ORFs reflects a bias for residues with A+T-rich codons. As previously noted in DNAs of other A+T-rich organisms (182), MsEPV preferentially encodes the 6 amino acids specified by codons exclusively composed of A and/or T (Lys, Asn, Ile, Leu, Tyr, and Phe). These amino acids represent the majority (61%) of all those encoded.

Two hundred and sixty-seven ORFs defined as methionine-initiated ORFs of greater than 60 amino acids are present (Fig. 1). The high A+T content, which results in a paucity of start codons (2.1%) and a large number of stop codons (14.5%), facilitates ORF identification. Predicted ORFs represent a 99% coding density, with an average ORF length of 854 nucleotides. Forty-four ORFs overlap other ORFs, and 28 smaller ORFs are completely contained within larger ORFs. Only 98 of the 267 MsEPV ORFs have been assigned a putative similarity or function based on homologies with other viral or cellular genes (Tables 1 and 2). Of the 155 most centrally located ORFs (MSV036 to MSV190), 45 (29%) are ChPV homologues. In contrast, of 112 ORFs in the terminal genomic regions (MSV001 to MS035 and MSV191 to MSV267), only 4 (3.5%) are identifiable ChPV homologues (Fig. 1).

MsEPV has a genomic organization similar to that of other known chordopoxviruses (53, 122, 169, 170). There is no evidence for introns, both strands are protein encoding, and there are few overlapping ORFs. ORFs frequently occur in head-to-tail tandem arrays (Fig. 1). Within the terminal 50 kbp of the genome, most ORFs are transcriptionally oriented toward their respective termini. As seen in other poxviruses, the MsEPV genome contains a central coding region bounded by two inverted terminal repeat (ITR) regions of approximately 7 kbp each (Fig. 1). The first 126-bp repeat marks the boundary between the ITR and the central coding region (Fig. 1). There are also regions internal to the ITR containing additional tandem repeats and several gene families.

ITRs. Although lacking sequence identity, MsEPV and ChPV ITR tandem repeats are similar (Fig. 1) (121, 190, 204). MsEPV ITRs contain a series of tandemly repeated sequences, 21-bp repeats (27 and 34 copies in the left and right ITRs, respectively) followed by blocks of 7-bp repeats (33 and 32 copies), 69-bp repeats (58 and 52 copies), 26-bp repeats (5 copies each), 72-bp repeats (2 copies each), and a 126-bp repeat (1 copy each). A variable number of incomplete 69-bp repeats (56 to 59 bp) separate the four terminal blocks of repeats. Nucleotide identity within sets of repeats is approximately 60% for the 21-bp repeats and 80 to 98% for the others. Comparison of 4-kbp noncoding ITR regions of vaccinia virus and MsEPV shows that the most numerous repeats, 69 or 70 bp long, are accompanied by less-abundant repeats of 125 to 126 bp and 26 bp and incomplete forms of the 69- to 70-bp repeats (54 to 58 bp). As in orthopoxviruses, there is a nonrepetitive spacer region within each ITR (53, 122). The coding capacity of MsEPV ITRs is limited. Four ORFs are present in the left ITR, and six ORFs are found in the right ITR. Within each ITR, four ORFs are present as single-copy genes. Three copies of the most-terminal ORF are present in the right ITR (MSV265, MSV266, and MSV267).

Gene expression regulatory elements. DNA sequences upstream, within, and downstream of MsEPV ORFs exhibited similarity to promoter and regulatory elements described for other poxviruses, thus suggesting some degree of conservation of gene regulatory mechanisms. Of the 14 MsEPV ORFs with homology to known poxvirus early genes, 8 contain a TGAA AxxxxA motif in the region 5′ of the putative translational start codon (Table 1), and an additional four genes contain this motif with only a single nucleotide substitution (data not shown). This putative early-type promoter element resembles the early-promoter core consensus sequence found in other ChPVs (130, 170). This motif has previously been found upstream of the thymidine kinase (TK) genes of other EPVs (58, 119), and a similar motif (TGAATxxxxA) is found upstream of the CbEPV DNA polymerase gene (131). Interestingly, the upstream sequence of the MmEPV fusolin gene, which demonstrates early promoter activity (111), also contains the TGA AAxxxxA motif (52). The vaccinia virus consensus early transcriptional stop sequence (TTTTTxT), which has also been observed downstream of EPV TK and other EPV gene sequences (58, 110, 175, 216), is found within 100 bases of the 3′ ends of 49 of the 84 MsEPV ORFs preceded by the TGAAA xxxxA motif.

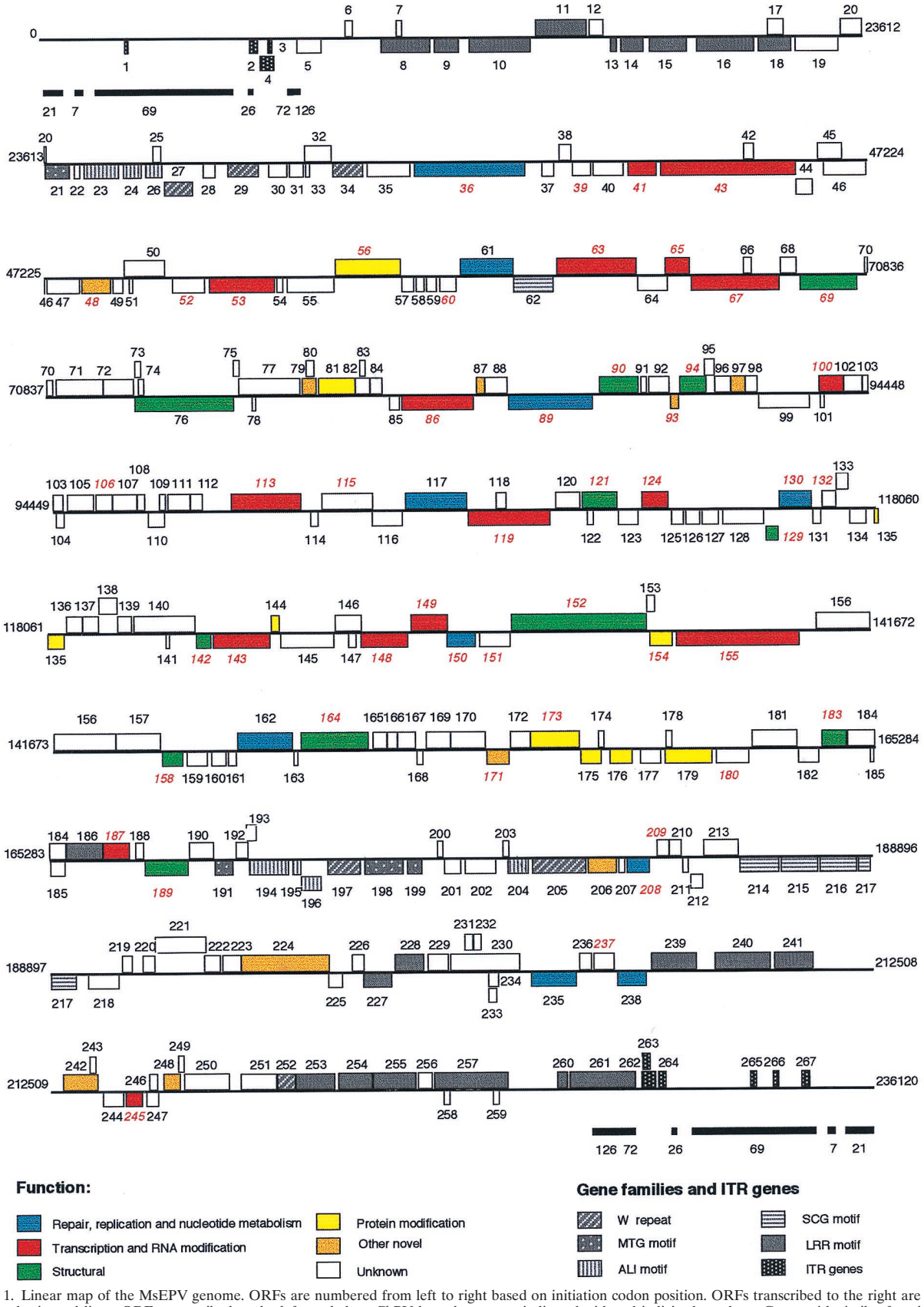


FIG. 1. Linear map of the MsEPV genome. ORFs are numbered from left to right based on initiation codon position. ORFs transcribed to the right are located above the horizontal lines; ORFs transcribed to the left are below. ChPV homologues are indicated with red italicized numbers. Genes with similar functions and members of gene families are colored according to the figure key. ITRs are represented as heavy black bars underneath the ORF map (numbers indicate sizes [in base pairs] of nucleotide repeats).

TABLE 1. MsEPV ORFs

ORF	Position (length, aa)[a]	Best match[b]	BlastP score	% Identity	Length, aa[a]	Predicted structure and/or function[c]	Promoter type[d]
MSV001	2631–2446 (62)					ITR, 62 aa	
MSV002	6320–6054 (89)					ITR, 89 aa	
MSV003	6756–6577 (60)					ITR, 60 aa, TM	
MSV004	6814–6350 (155)					ITR, 155 aa	E
MSV005	8171–7434 (246)					TM	
MSV006	8831–9055 (75)						
MSV007	10282–10473 (64)						
MSV008	11327–9828 (500)	[P28854, AmEPV ORF Q3]	333	40	205	LRR	E
MSV009	12138–11377 (254)	[P28854, AmEPV ORF Q3]	286	38	204	LRR	E
MSV010	14199–12367 (611)	[P28854, AmEPV ORF Q3]	298	35	218	LRR	E
MSV011	14293–15807 (505)	[P28854, AmEPV ORF Q3]	298	39	226	LRR	E
MSV012	15855–16244 (130)						E
MSV013	16654–16439 (72)	[P28854, AmEPV ORF Q3]	77	36	71	LRR	E
MSV014	17462–16731 (244)	[P28854, AmEPV ORF Q3]	256	42	171	LRR, TM	E
MSV015	18692–17535 (386)	[P28854, AmEPV ORF Q3]	325	38	207	LRR	E
MSV016	20651–18936 (572)	[P28854, AmEPV ORF Q3]	315	40	218	LRR	
MSV017	20990–21430 (147)						
MSV018	21717–20701 (339)	[P28854, AmEPV ORF Q3]	315	37	213	LRR	
MSV019	23074–21764 (437)						
MSV020	23121–23645 (175)						E
MSV021	24438–23659 (260)	AF003534, Chilo iridescent virus ORF 074R	128	22	193	MTG motif	E
MSV022	24696–24478 (73)						E
MSV023	25850–24756 (365)	L44593, bacteriophage BK5-T ORF266	135	27	144	ALI motif	E
MSV024	26483–25875 (203)	AF003534, Chilo iridescent virus OFR 011L	239	37	154	ALI motif	
MSV025	26794–26997 (68)					TM	
MSV026	27095–26526 (190)	AF003534, Chilo iridescent virus ORF 011L	201	32	154	ALI motif	
MSV027	27979–27089 (297)	M96361, AcNPV 41.6-kDa protein	190	32	148	Tryptophan repeat	E
MSV028	28594–28172 (141)	AF003534, Chilo iridescent virus ORF 011L	125	28	98		E
MSV029	29846–28881 (322)	M96361, AcNPV 41.6-kDa protein	133	29	167	Tryptophan repeat	E
MSV030	30674–30081 (198)						E
MSV031	31112–30690 (141)					TM	L
MSV032	31140–31895 (252)					SP	
MSV033	31309–31112 (66)					TM, SP	
MSV034	32866–31925 (314)	M96361, AcNPV 41.6-kDa protein	159	28	150	Tryptophan repeat	E
MSV035	34198–32882 (439)						
MSV036	37490–34254 (1079)	[P30319, CbEPV DNA polymerase] (E9L)	1,668	40	958	DNA polymerase	E
MSV037	38330–37944 (129)						E
MSV038	38456–38821 (122)						L
MSV039	39408–38830 (193)	[U60315, MCV MCV062R protein] (G6R)	117	25	127		L
MSV040	40359–39412 (316)						E
MSV041	41293–40409 (295)	[L22579, VAR poly(A) polymerase regulatory] (J3R)	299	35	244	Poly(A) polymerase (small subunit) PAP_S	E
MSV042	43777–44037 (87)						
MSV043	45280–41324 (1319)	[P20504, vaccinia virus RNA polymerase RPO147] (J6R)	1,313	30	1,185	RNA polymerase, RPO147	E
MSV044	45770–45276 (165)						L
MSV045	45881–46579 (233)					TM, SP	
MSV046	47237–45897 (447)					TM, SP	E
MSV047	48296–47304 (331)					TM, SP	E
MSV048	49188–48325 (288)	D12680, *Rhizopus niveus* lipase	142	29	158	Lipase, TM	L
MSV049	49565–49218 (116)					TM	L
MSV050	49581–50717 (379)					TM	L
MSV051	49855–49661 (65)					TM	
MSV052	51931–50897 (345)	[P20998, vaccinia virus A23R protein] (A23R)	116	26	324		
MSV053	53902–51962 (647)	[P24486, CbEPV NPH-1] (D11L)	1,997	58	648	Nucleoside phosphohydrolase, NPH-1	L
MSV054	54163–53918 (82)						
MSV055	55645–54248 (466)					TM	L
MSV056	55659–57545 (629)	[X76267, VAR (Garcia 66) F2L] (G1L)	98	23	237	Metalloprotease, TM	L
MSV057	57947–57552 (132)						L
MSV058	58207–57947 (87)						L
MSV059	58573–58241 (111)						E
MSV060	59204–58623 (194)	[P20496, vaccinia virus H2 late protein] (H2R)	316	37	183	TM	L
MSV061	59208–60800 (531)	Z83109, *Caenorhabditis elegans* F44G3.3 gene product	308	28	338	RT	
MSV062	61995–60793 (401)					SCG motif, TM	L
MSV063	62037–64316 (760)	[X76265, VAR 82-kDa subunit] (A7L)	410	25	583	Early transcription factor, $VETF_S$	L
MSV064	65158–64319 (280)						E
MSV065	65228–65881 (218)	[P07609, vaccinia virus late transactivator protein] (A2L)	194	22	198	Late transcription factor, VLTF-3	L
MSV066	67390–67614 (75)					TM, SP	
MSV067	68454–65875 (860)	[P20979, vaccinia virus mRNA capping enzyme, large subunit] (D1R)	639	28	750	mRNA capping	L
MSV068	68466–68945 (160)					TM, SP	L
MSV069	70676–68952 (575)	[U44841, HaEVP rifampicin resistance protein] (D13L)	1,697	54	572	Morphogenesis, rifampin resistance	L

Continued on facing page

TABLE 1—*Continued*

ORF	Position (length, aa)[a]	Best match[b]	BlastP score	% Identity	Length, aa[a]	Predicted structure and/or function[c]	Promoter type[d]
MSV070	70701–71036 (112)						
MSV071	71179–72504 (442)						
MSV072	72504–73397 (298)						L
MSV073	73510–73707 (66)						
MSV074	73619–73843 (75)						
MSV075	76302–76490 (63)						
MSV076	76309–73400 (970)	[P29815, AmEPV spheroidin]	439	25	707	Spheroidin	L
MSV077	76411–78204 (598)					TM	E
MSV078	76944–76744 (67)					TM	
MSV079	78257–78670 (138)		70	26	88	C_2H_2 zinc finger	L
MSV080	78363–78590 (76)						TM
MSV081	78693–79763 (357)	P36993, *Mus musculus* PP2C, beta isoform	329	25	323	Protein phosphatase, PP2C	L
MSV082	79775–80191 (139)						L
MSV083	79878–80060 (61)						
MSV084	80206–80574 (123)						E
MSV085	81072–80719 (118)					TM	L
MSV086	83232–81082 (717)	[P20502, vaccinia virus RNA helicase] (I8R)	776	33	651	RNA helicase, NPH-II	L
MSV087	83258–83485 (76)	D45892, *Neurospora crassa* thioredoxin	60	26	67	Thioredoxin	L
MSV088	83508–84122 (205)						L
MSV089	86650–84149 (834)	[G41700, RFV C5 protein] (D5R)	548	27	643	Nucleic acid-dependent NTPase, TM	
MSV090	86786–87925 (380)	[U60315, MCV MC121L protein] (A16L)	389	29	274	Potential membrane protein, TM	
MSV091	87971–88189 (73)						E
MSV092	88220–88807 (196)						L
MSV093	89112–88792 (107)	[P33821, VAR E10R protein] (E10R)	194	45	90	Potential redox, yeast ERV1	L
MSV094	89128–89850 (241)	[P24361, vaccinia virus F9 protein] (F9L)	262	32	186	Potential membrane protein, TM	L
MSV095	89843–90082 (80)					TM	L
MSV096	90112–90564 (151)						E
MSV097	90573–90992 (140)	PRF:1906390A, *Atriplex nummularia* caltractin-like protein	120	28	128	Calcium binding protein	L
MSV098	91014–91337 (108)					L	
MSV099	92883–91327 (519)					TM, SP	E
MSV100	93123–93812 (230)	P33813, VAR RNA polymerase RPO19 (A5R)	68	26	126	RNA polymerase subunit, RPO19, TM	
MSV101	93298–93095 (68)						
MSV102	93834–94364 (177)						E
MSV103	94376–94828 (151)						
MSV104	94827–94513 (105)					TM	
MSV105	94873–95634 (254)					TM	E
MSV106	95701–96189 (163)	[P20997, vaccinia virus A22 protein] (A22R)	134	29	154		E, L
MSV107	96207–96884 (226)					TM, SP	L
MSV108	96891–97118 (76)					TM, SP	L
MSV109	97531–97743 (71)					TM, SP	
MSV110	97709–97167 (181)					TM, SP	E
MSV111	97802–98404 (201)					TM, SP	E
MSV112	98404–98793 (130)					TM	
MSV113	99618–101639 (674)	[P04308 vaccinia virus 70 kDa subunit (D6R)	734	44	379	Early transcription factor, $VETF_S$	L
MSV114	102136–101864 (91)					TM, SP	E
MSV115	102182–103696 (505)	[J03399, vaccinia virus G5R protein] (G5R)	189	28	270		
MSV116	104571–103621 (317)					TM	E
MSV117	104621–106429 (603)	M13961, *Rattus norvegicus* DNA polymerase β	185	31	186	DNA polymerase β, SP	L
		U40707, *Caenorhabditis elegans* AP endonuclease	162	31	180	AP endonuclease	
MSV118	107228–107512 (95)					TM	
MSV119	108840–106420 (807)	[P33067, VAR RNA polymerase-associated protein] (H4L)	564	26	797	RNA polymerase-associated factor, RAP94, TM	L
MSV120	108922–109674 (251)						E
MSV121	109692–110690 (333)	[P15909, FPV protein FP1] (G9R)	380	30	317	Potential membrane protein, TM	L
MSV122	110059–109847 (71)					SP	
MSV123	111375–110686 (230)						
MSV124	111409–112209 (267)	[S42252, FPV mRNA capping enzyme, small subunit] (D12L)	95	18	221	mRNA capping, TM	E, L
MSV125	112651–112220 (144)						E, L
MSV126	113119–112676 (148)						
MSV127	113641–113126 (172)						
MSV128	114960–113695 (422)						
MSV129	115345–114947 (133)	[P07615, vaccinia virus L5R protein] (L5R)	90	31	114	TM	L
MSV130	115362–116345 (328)	[U80056, AmEPV DNA topoisomerase 1] (H6R)	986	59	330	Type I topoisomerase L	
MSV131	116598–116344 (85)					TM	
MSV132	116604–117029 (142)	[P29816, AmEPV G4R protein] (A28L)	463	58	139	AmEPV G4R, TM, SP	L
MSV133	117022–117405 (128)						L
MSV134	117881–117408 (158)					TM	L
MSV135	118619–117903 (239)	P40371, *Schizosaccharomyces pombe* PP2C	128	22	246	Protein phosphatase, PP2C	E, L
MSV136	118649–119098 (150)						L
MSV137	119105–119551 (149)						
MSV138	119548–120117 (190)						L
MSV139	120121–120537 (139)						E, L
MSV140	120541–122364 (608)						L
MSV141	121630–121433 (66)						
MSV142	122781–122365 (139)	[M17418, FPV 15.6-kDa protein] (J5L)	171	33	132	Potential membrane protein	L

Coninued on following page

TABLE 1—*Continued*

ORF	Position (length, aa)[a]	Best match[b]	BlastP score	% Identity	Length, aa[a]	Predicted structure and/or function[c]	Promoter type[d]
MSV143	124506–122794 (571)	[P33809, VAR poly(A) polymerase catalytic] (E1L)	249	29	322	Poly(A) polymerase (PAP_L), TM	
MSV144	124539–124778 (80)	U01220, *Neurospora crassa* ubiquitin	356	86	80	Ubiquitin	
MSV145	126345–124771 (525)						L
MSV146	126360–127091 (244)						
MSV147	127018–126704 (105)					TM	
MSV148	128490–127078 (471)	[P20534, vaccinia virus DNA helicase] (A18R)	423	30	376	DNA helicase, TM	L
MSV149	128526–129569 (348)	P21087, vaccinia virus RNA polymerase RP035 (A29L)	90	20	213	RNA polymerase subunit, RPO35	L
MSV150	130430–129564 (289)	[P32817, FPV D10 protein] (D10R)	186	31	200	NTP pyrophosphohydrolase, MutT	
MSV151	131400–130462 (313)	[P33836, VAR A11R protein] (A11R)	158	22	314		L
MSV152	131427–135344 (1306)	[P33817, VAR major core protein precursor P4a] (A10L)	256	22	840	Core protein, P4a	L
MSV153	135325–135537 (71)					TM	
MSV154	136112–135357 (252)	AB000449, *Homo sapiens* Ser/Thr protein kinase 1 (B1R)	206	32	170	Ser/Thr protein kinase, VRK1	E
MSV155	139728–136159 (1190)	[P17474, CPV RNA polymerase RPO132] (A24R)	925	30	770	RNA polymerase, RPO132	E, L
MSV156	140126–143506 (1127)						E, L
MSV157	143533–144822 (430)						E
MSV158	145485–144811 (225)	[X76267, VAR core protein VP8 precursor] (L4R)	133	28	219	Core protein, VP8	L
MSV159	146202–145528 (225)						L
MSV160	146701–146231 (157)						L
MSV161	147002–146724 (93)	AFO019224, HaEPV ORF F2 protein	154	39	78	HaEPV F2, TM, SP	L
MSV162	147019–148584 (522)	P26996, *Thermus thermophilus* NAD^+-DNA ligase	178	25	356	NAD^+-dependent DNA ligase	
MSV163	148808–148593 (72)					TM	L
MSV164	148850–150793 (648)	[P17355 FPV major core protein precursor P4b] (A3L)	297	24	519	Core protein, P4b, TM	L
MSV165	150915–151292 (126)					TM, SP	L
MSV166	151313–151600 (96)					TM, SP	L
MSV167	151629–152162 (178)						E
MSV168	152391–152176 (72)						
MSV169	152435–153124 (230)					TM	
MSV170	153165–154136 (324)						E
MSV171	154863–154132 (244)	[P21055, vaccinia virus A32L protein] (A32L)	168	30	204	ATP/GTP binding motif	
MSV172	154891–155442 (184)						L
MSV173	155472–156842 (457)	[P32216, SPV Ser/Thr protein kinase C20L] (F10L)	247	26	375	Ser/Thr protein kinase, KRF1	L
MSV174	157386–157568 (61)					TM	
MSV175	157502–156840 (221)	U90931, *Bacteroides fragilis* metalloprotease toxin 2	85	29	48	Metalloprotease, SP	
MSV176	158396–157683 (238)	X89576, *Homo sapiens* MT2-MMP protein	103	29	130	Metalloprotease, SP	
MSV177	159202–158573 (210)					TM, SP	E
MSV178	159328–159507 (60)						
MSV179	160698–159304 (465)	U82541, *Xenopus laevis* matrix metalloprotease	163	25	179	Metalloprotease, TM, SP	
MSV180	161788–160760 (343)	[P07614, vaccinia virus protein L3L protein] (L3L)	188	32	162		
MSV181	161820–163121 (434)					TM, SP	
MSV182	163771–163121 (217)						
MSV183	163810–164535 (242)	[U60315, MCV MC069R protein] (L1R)	503	48	224	Myristylated membrane protein, TM	
MSV184	164558–165802 (415)					TM	L
MSV185	165438–165253 (62)						
MSV186	165805–166824 (340)	[P28854, AmEPV ORF Q3]	124	30	164	LRR	
MSV187	166845–167627 (261)	[S42254, FPV transactivator protein] (A1L)	98	30	106	Late transcription factor, VLTF-2	
MSV188	167805–168008 (68)						E, L
MSV189	169331–168003 (443)	[P29817, AmEPV G1L protein] (I7L)	1,074	48	452	Core protein, I7L	
MSV190	169350–170030 (227)	[P29818 AmEPV G2R protein]	268	31	217	AmEPV G2R, TM	
MSV191	170628–170020 (203)					MTG motif	E
MSV192	170671–171021 (117)					L	
MSV193	170978–171286 (103)					TM	
MSV194	172239–171013 (409)	L44593, bacteriophage BK5-T ORF 266 protein	262	40	157	ALI motif	E
MSV195	172550–172290 (87)	P24655, AcNPV 38-kDa protein, orf2	91	33	84	ALI motif	
MSV196	173145–172540 (202)	AF003534, Chilo iridescent virus ORF 011L	275	35	190	ALI motif	E
MSV197	174289–173282 (336)	M96361, AcNPV 41.6-kDa protein	166	29	162	Tryptophan repeat	
MSV198	175515–174319 (399)	AF003534, Chilo iridescent virus ORF 074R	150	23	240	MTG motif	E
MSV199	176049–175576 (158)	AF003534, Chilo iridescent virus ORF 074R	120	25	148	MTG motif	E
MSV200	176444–176632 (63)						
MSV201	177155–176625 (177)						
MSV202	178207–177224 (328)						E
MSV203	178373–178555 (61)					TM	
MSV204	179133–178468 (222)	AF003534, Chilo iridescent virus ORF 011L	137	27	161	ALI motif	E
MSV205	180752–179163 (530)	M96361, AcNPV 41.6-kDa Protein	182	33	146	Tryptophan repeat	E
MSV206	181634–180774 (287)	U94833, *Haemophilus sommus* lipooligosaccharide biosynthesis	91	25	236	Glycosyltransferase, TM	L
MSV207	181871–181638 (78)						
MSV208	182595–181900 (232)	U20824, equine herpesvirus uracil DNA glycosylase (D4R)	264	34	214	UNG	E
MSV209	182790–183128 (113)	[P20996, vaccinia virus A21 protein] (A21L)	69	24	115	SP	
MSV210	183142–183474 (111)						

Continued on facing page

TABLE 1—*Continued*

ORF	Position (length, aa)[a]	Best match[b]	BlastP score	% Identity	Length, aa[a]	Predicted structure and/or function[c]	Promoter type[d]
MSV211	183743–183501 (81)						
MSV212	184138–183740 (133)						L
MSV213	184168–185160 (331)						L
MSV214	186301–185144 (386)					SCG motif, TM	L
MSV215	187434–186325 (370)					SCG motif, TM, SP	L
MSV216	188543–187434 (370)					SCG motif, TM	L
MSV217	189685–188543 (381)					SCG motif, TM	L
MSV218	190939–189980 (320)						
MSV219	190981–191469 (163)						
MSV220	191566–191925 (120)						E
MSV221	191912–193381 (490)						
MSV222	193368–193838 (157)					TM	
MSV223	193896–194408 (171)					TM	E
MSV224	194435–196933 (833)		141	23	627	NTPase/Helicase	E, L
MSV225	197342–196917 (142)						E
MSV226	197604–197930 (109)						E
MSV227	198805–197888 (306)	[P28854, AmEPV ORF Q3]	165	30	200	LRR	E
MSV228	198841–199677 (279)	[P28854, AmEVP ORF Q3]	148	32	167	LRR	L
MSV229	199808–200353 (182)	AF003534 Chilo iridescent virus ORF 011L	195	37	123		E
MSV230	200414–202423 (670)						E
MSV231	200853–201053 (67)						
MSV232	201060–201341 (94)						
MSV233	201796–201521 (92)					TM	E
MSV234	201830–201480 (117)						
MSV235	204120–202723 (466)	D31902, *Monodelphis domestica* CPD photolyase	1,377	57	429	CPD photolyase	
MSV236	204132–204530 (133)						E
MSV237	204558–205133 (192)	[P20999, vaccinia virus B2R protein (B2R)	434	44	189		E
MSV238	206102–205227 (292)	L08594, *Arabidopsis thaliana* thymidylate synthase	930	59	286	Thymidylate synthase	E
MSV239	206210–207496 (429)	[P28854, AmEPV ORF Q3]	341	41	217	LRR	E, L
MSV240	208070–209650 (527)	[P28854, AmEPV ORF Q3]	330	38	219	LRR	
MSV241	209732–210892 (387)	[P28854 AmEPV ORF Q3]	310	38	213	LRR	E
MSV242	212889–213875 (329)	L22858 AcNPV apoptosis inhibitor protein	129	25	132	IAP, TM	
MSV243	213634–213822 (63)						
MSV244	214649–214005 (215)						
MSV245	215215–214658 (186)	[P33058, variola virus RNA polymerase RPO18] (D7R)	93	23	156	RNA polymerase, RPO18	E
MSV246	215380–215583 (68)						
MSV247	215660–215232 (143)					TM	E
MSV248	215803–216252 (150)	U75930 OpNPV inhibitor of apoptosis protein	217	29	144	IAP	E
MSV249	216210–216398 (63)					TM	L
MSV250	216350–217672 (441)						E
MSV251	218047–219033 (329)					$C_3H_2C_3$ RING finger	
MSV252	219048–219530 (161)					Tryptophan repeat	E
MSV253	219557–220756 (400)	[P28854, AmEPV ORF Q3]	275	36	213	LRR	E
MSV254	220798–221796 (333)	[P28854, AmEPV ORF Q3]	258	41	160	LRR, TM	E
MSV255	221829–223037 (403)	[P28854, AmEPV ORF Q3]	341	42	208	LRR	E
MSV256	223107–223508 (134)						
MSV257	223560–225680 (707)	[P28854, AmEPV ORF Q3]	238	35	207	LRR	E
MSV258	224029–223835 (65)						
MSV259	225490–225224 (89)					SP	
MSV260	227136–227414 (93)	[P28854, AmEPV ORF Q3]	116	38	81	LRR	
MSV261	227470–229341 (624)	[P28854, AmEPV ORF Q3]	341	42	219	LRR	E
MSV262	229505–229969 (155)					ITR, 155 aa	E
MSV263	229563–229742 (60)					ITR, 60 aa; TM	
MSV264	229999–230265 (89)					ITR, 89 aa	
MSV265	232644–232829 (62)					ITR, 62 aa	
MSV266	233334–233519 (62)					ITR, 62 aa	
MSV267	234162–234347 (62)					ITR, 62 aa	

[a] aa, amino acids.
[b] Accession numbers are from the GenBank or SwissProt database (unless otherwise indicated). Poxvirus data are in brackets; vaccinia virus data are in parentheses.
[c] Function was deduced from the degree of amino acid similarity to either known genes or Prosite signatures. TM, Z score of ≥ 1.96 for the prediction of transmembrane domains by using the Memsat computer program; SP, Z score of ≥ 2.5 for the prediction of signal peptides by using Sigcleave.
[d] Putative promoter type. E, early; L, late.

Twenty-four of 36 MsEPV homologues of late ChPV genes (170) contain the consensus poxvirus-late-promoter sequence (TAAATG) at the translational start site (15, 157, 198). This late-promoter sequence has been previously described for other EPV genes, including those encoding spheroidin, the nucleoside triphosphatase (NTPase) hydrolase I (NPH-I), and topoisomerase (64, 66, 110, 164, 175). Eleven of the 12 remaining putative MsEPV late genes contain upstream sequences that have been found at the start of poxvirus late genes. Such sequences include TAAAT upstream of the translational start site (eight ORFs) (158), TAAAAT (one ORF) (95), and TAATG (three ORFs) (157, 158). Similar to the fusolin genes of other EPVs, 33 MsEPV ORFs contain the TAATG motif at the translational start site (27, 140, 215).

TABLE 2. Chordopoxvirus homologues in MsEPV

Function	MsEPV ORF	Vaccinia virus ORF	% Amino acid identity (length)	Poxvirus with highest degree of homology	% Amino acid identity (length)	Gene name and/or function
Transcription/mRNA modification						
RNA polymerase						
	MSV043	J6R	30 (1,185)			RPO147
	MSV100	A5R	25 (162)	VAR	26 (162)	RPO19
	MSV119	H4L	26 (797)	VAR	26 (797)	RAP94
	MSV149	A29L	20 (213)			RPO35
	MSV155	A24R	30 (770)	CPV	30 (770)	RPO132
	MSV245	D7R	23 (133)	VAR	23 (156)	RPO18
Transcription factors						
	MSV063	A7L	25 (583)	VAR	25 (583)	$VETF_L$
	MSV065	A2L	22 (198)			VLTF-3
	MSV113	D6R	44 (379)			$VETF_S$
	MSV187	A1L	26 (78)	FPV	30 (106)	VLTF-2
NTPase/helicase						
	MSV053	D11L	36 (633)	FPV	37 (633)	NPH-I
	MSV086	I8R	33 (651)			RNA helicase/NPH-II
	MSV148	A18R	30 (376)			DNA helicase
mRNA modification						
	MSV041	J3R	35 (244)	VAR	35 (244)	PAP_S
	MSV067	D1R	28 (750)			Capping enzyme, large subunit
	MSV124	D12L	20 (177)	FPV	18 (221)	Capping enzyme, small subunit
	MSV143	E1L	29 (322)	VAR	29 (322)	PAP_L
DNA replication/repair						
	MSV036	E9L	29 (943)	FPV	30 (926)	DNA polymerase
	MSV089	D5R	29 (560)	RFV	27 (643)	NTPase
	MSV130	H6R	34 (315)	FPV	37 (321)	Topoisomerase
	MSV150	D10R	35 (91)	FPV	31 (200)	*mutT* motif
	MSV208	D4R	18 (199)	FPV	30 (117)	UNG
Structural						
	MSV069	D13L	28 (445)			Rifampicin resistance
	MSV090	A16L	27 (368)	MCV	29 (274)	Putative membrane protein
	MSV094	F9L	32 (186)			Putative membrane protein
	MSV121	G9R	25 (335)	FPV	30 (317)	Putative membrane protein
	MSV129	L5R	31 (114)			Putative membrane protein
	MSV142	J5L	29 (131)	FPV	33 (132)	Putative membrane protein
	MSV152	A10L	22 (718)	VAR	22 (840)	Core protein, P4a
	MSV158	L4R	27 (219)	VAR	28 (219)	Core protein, VP8
	MSV164	A3L	23 (532)	FPV	24 (519)	Core protein, P4b
	MSV183	L1R	38 (224)	MCV	48 (224)	Membrane protein
	MSV189	I7L	26 (416)	FPV	26 (427)	Core protein
Enzymes						
	MSV048			CPV	25 (75)	Lipase
	MSV093	E10R	42 (100)	VAR	45 (90)	Potential redox, ERV1
	MSV154	B1R	32 (161)			Protein kinase
	MSV171	A32L	30 (204)			ATP/GTP binding motif
	MSV173	F10L	24 (407)	SPV	26 (375)	Protein kinase
	MSV056	G1L	23 (591)	VAR	23 (591)	Metalloprotease
Unknown						
	MSV039	G6R	21 (146)	MCV	25 (127)	
	MSV052	A23R	26 (324)			
	MSV060	H2R	37 (183)			
	MSV106	A22R	29 (154)			
	MSV115	G5R	28 (270)			
	MSV132	A28L	33 (146)	MCV	38 (141)	
	MSV151	A11R	22 (314)	VAR	22 (314)	
	MSV180	L3L	32 (162)			
	MSV209	A21L	24 (115)			
	MSV237	B2R	44 (189)			

Transcription and mRNA biogenesis. MsEPV contains homologues of 18 of the 26 vaccinia virus genes thought to be involved in transcriptional processes (130) (Fig. 1; Table 2). This suggests the presence of conserved mechanisms for generating functional mRNA among the two poxvirus subfamilies. Vaccinia virus RNA polymerase is encoded by at least eight viral genes ranging in size from 7 to 147 kDa. MsEPV homologues of the two largest vaccinia virus subunits, RPO147 (J6R) and RPO132 (A24R), and the smaller subunits RPO35 (A29L), RPO19 (A5R), and RPO18 (D7R) are MSV043, MSV155, MSV149, MSV100, and MSV245, respectively (Table 2). Homologues of the three remaining vaccinia virus RNA polymerase subunits are not identifiable in MsEPV. Amino acid variability within homologous ChPV RPO subunits sug-

gests that other MsEPV subunits may also be highly variable and, thus, undetectable by current computer search and analysis algorithms. In addition, MsEPV contains a homologue (MSV119) of the RNA polymerase-associated protein RAP94 (H4L), which is specifically required for transcription of early-promoter templates (Table 2) (130).

Four homologues of vaccinia virus transcription factors are encoded in MsEPV. MSV113 and MSV063 are homologues of the two subunits of the vaccinia virus early transcription factor, $VETF_S$ (D6R) and $VETF_L$ (A7L), respectively. MSV187 and MSV065 are homologues of the two late transcription factors VLTF-2 (A1L) and VLTF-3 (A2L) (Table 2). While MSV113 has 44% amino acid identity to vaccinia virus $VETF_S$, the levels of amino acid identity to $VETF_L$, VLTF-2, and VLTF-3 homologues are much lower (22 to 24%). The vaccinia virus late transcription factor VLTF-1 (G8R), VLTF-4 (H5R), and G2R gene product homologues are either absent from the MsEPV genome or unidentifiable. The absence of a VLTF-1 homologue is surprising, since this gene is essential for vaccinia virus replication and is conserved among ChPV genera (130, 170).

Four MsEPV ORFs, MSV053, MSV086, MSV113, and MSV148, are homologues of four NTPase-helicase genes found in vaccinia virus (Table 2). These include the NPH-I homologue (D11L), the RNA-DNA helicase (NPH-II) homologue (I8R), the small subunit of the early transcription factor $VETF_S$ (D6R), and the DNA helicase (A18R), respectively (Table 2). These MsEPV ORFs contain motifs conserved among the NTPase and helicase enzymes of the RNA and DNA helicase superfamily II (55, 93). Only NPH-I homologues have been previously described in other EPVs (65, 110, 175, 217). As expected, this gene (MSV053) has a higher level of amino acid identity to EPV genes (58%) than to orthopox-, molluscipox-, and leporipoxvirus homologues (33 to 37%). Another ORF (MSV224) also contains carboxy-terminal helicase and NTPase motifs in addition to a cysteine-rich amino terminus, but it lacks homology to other poxvirus genes.

MSV148 encodes a homologue of the vaccinia virus A18R gene. A18R encodes a late virion-associated DNA helicase that is essential for correct viral gene expression and productive infection (12, 174). The essential nature of A18R suggests a similar function for MSV148 in MsEPV.

MsEPV contains homologues of vaccinia virus genes involved in transcriptional termination, capping, and polyadenylation. MSV067 and MSV124 are homologues of the large and small subunits of the vaccinia virus capping enzyme, D1R and D12L, respectively (Table 2). MSV143 and MSV041 are homologues of the large and small polyadenylation polymerase (PAP) subunits, PAP_L (E1L) and PAP_S (J3R), respectively (Table 2).

Nucleotide metabolism. MsEPV lacks all previously described poxvirus genes involved in nucleotide metabolism (130). Absent are genes encoding TK, thymidylate kinase, the large and small subunits of ribonucleotide reductase, dUTPase, glutaredoxin, and guanylate kinase and the cytidine kinase gene found in FPV (92). The absence of a TK gene in MsEPV is surprising given that TK genes have been identified in other group B EPVs (AmEPV, CbEPV, and CfEPV) (58, 119). This paucity of viral enzymes suggests that MsEPV replication is heavily dependent on host cell nucleotide biosynthesis. These differences in nucleotide metabolism must be of significance for viral cell and/or tissue tropism within the grasshopper host.

Interestingly, and unlike other known poxviruses, MSV238 encodes a thymidylate synthase (TSY) homologue. MSV238 is very similar to TSY genes from eukaryotes (52 to 59% amino acid identity). The 29-amino-acid TSY Prosite motif (PS00091), which contains the catalytic cysteine residue, is also conserved in MSV238 with the exception of a single conservative substitution (leucine to isoleucine) at position 170. Homodimeric TSY catalyzes the methylation of dUMP to the nucleotide precursor dTMP, thus representing an important part of the de novo pathway of pyrimidine biosynthesis (21). Despite its ubiquitous distribution in nature, a viral TSY gene has been observed only in a few herpesviruses and bacteriophages (13, 79, 80).

DNA replication. MsEPV contains homologues of most vaccinia virus genes involved in DNA replication, including DNA polymerase (E9L), ATP-GTP binding protein (D5R), DNA topoisomerase (H6R), and replication essential protein kinase (B1R) (Table 2). However, it lacks the processivity factor (A20R) and an ATP-dependent DNA ligase (A50R). Notably, and unlike any other known virus, MsEPV encodes an NAD^+-dependent DNA ligase homologue (Table 1).

The MsEPV DNA polymerase (MSV036) is homologous to family B replicative DNA polymerases found in CbEPV (40% identity over 958 amino acids) and other ChPVs (29 to 30% amino acid identity over 900 amino acids). MSV036 also exhibits similarity to the DNA polymerase genes of chlorella virus PBCV-1 (39% identity over 247 amino acids; GenBank accession no. S35209) and ASFV (21% identity over 452 amino acids; GenBank accession no. U27575). MSV036 identity to family B DNA polymerases includes the highly conserved region I, in which the Prosite family signature (PS00116) is 100% conserved (5, 207).

MSV089 is homologous to the vaccinia virus ATP-GTP binding protein D5R (29% identity over 560 amino acids). Although more divergent than other ChPV D5R homologues, MSV089 contains regions of similarity throughout the protein, including the extended I (A) type of nucleoside triphosphate binding motif (55). D5R is known to be essential for viral DNA replication and is involved in homologous recombination (159).

Like AmEPV, MSV130 encodes a eukaryotic type I DNA topoisomerase with homology to vaccinia virus H6R. In MsEPV, the type I DNA topoisomerase Prosite motif (PS00176) and the active site residue (Tyr-292) are 100% conserved (172). In addition, critical residues required for transesterification by H6R are also conserved (205). Of six critical DNA recognition motifs described in the vaccinia virus protein, three are conserved and two are conservatively substituted in MSV130 (168).

MSV162 exhibits similarity to bacterial NAD^+-dependent DNA ligases (Fig. 2B). To our knowledge, this is the first NAD^+-dependent DNA ligase found in a virus genome. Eukaryotic organisms and other DNA viruses, including all known poxviruses, encode ATP-dependent DNA ligases (26, 67, 89, 103, 115, 213). MSV162 is most similar to bacterial *Thermus thermophilus* ligase and includes 12 of the 16 Prosite signature residues (PS01055) and the active-site motif Lys-X-Asp-Gly (118). Residues essential for adenylation and deadenylation steps are conserved in MSV162 (Lys-112 and Asp-118, respectively) (118). In addition, MSV162 exhibits partial conservation of a second NAD^+-ligase Prosite signature (PS01056) but lacks residues typically conserved in the carboxyl terminus. Given that NAD^+-dependent DNA ligases have been found only in bacteria, the presence of this gene in a eukaryotic virus is surprising and suggests that MSV162 has a prokaryotic origin.

DNA repair. MsEPV encodes at least seven genes with putative DNA repair functions (Table 1). These include homologues of genes encoding a uracil DNA glycosylase (UNG) (MSV208), DNA polymerase β (MSV117), AP endonuclease (MSV117), DNA helicase (MSV148), a ChPV *mutT* homo-

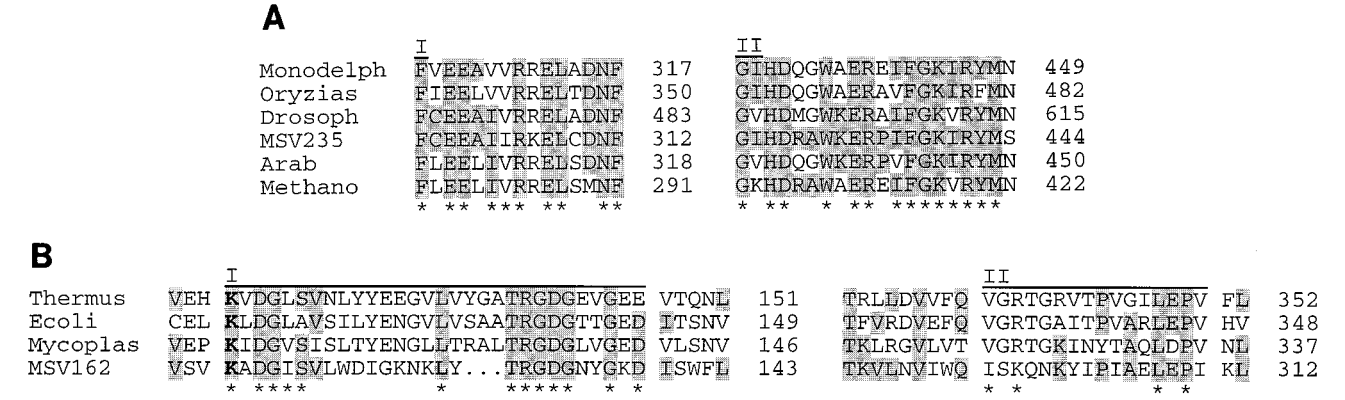


FIG. 2. Multiple amino acid sequence alignments of MsEPV ORFs with DNA repair and replication enzymes. Boldfaced letters represent active site residues, asterisks mark residues that match Prosite signatures, and shaded residues represent amino acids with identity to those of the corresponding MsEPV ORF. Amino acid positions are indicated on the right. (A) Alignment of MSV235 with class 2 CPD photolyases; regions I and II represent class 2 CPD photolyase Prosite signatures PS01083 and PS01084, respectively. Abbreviations: Monodelph, *Monodelphis domestica*, accession no. D31902; Oryzias, *Oryzias latipes*, accession no. S52048; Drosoph, *Drosophila melanogaster*, accession no. S52047; Arab, *Arabidopsis thaliana*, accession no. X99301; Methano, *Methanobacterium thermoautotrophicum*, accession no. D30752. (B) Alignment of MSV162 with NAD^+-dependent DNA ligases; regions I and II represent Prosite signatures PS01055 and PS01056 for NAD^+-dependent DNA ligases, respectively. Abbreviations: Thermus, *Thermus aquaticus*, accession no. P26996; Ecoli, *E. coli*, accession no. P15042; Mycoplas, *Mycoplasma pneumoniae*, accession no. AE000047.

logue (MSV150), cyclobutane pyrimidine dimer (CPD) photolyase (MSV235), and a LINE-type reverse transcriptase (RT) (MSV061). Genes for DNA polymerase β and AP endonuclease have not been previously described in poxviruses. Further, the CPD photolyase and the LINE-type RT have not been previously found in any virus genome. The presence of this complement of genes in MsEPV suggests that virally encoded DNA repair functions are important for virus survival in nature.

Although other large DNA viruses, such as vaccinia virus and ASFV, contain some genes of the base excision repair (BER) pathway (181, 191, 212), MsEPV provides the first example of a virus potentially encoding all genes required for BER. DNA damage in eukaryotic cells arises spontaneously from hydrolytic events, oxygen free-radical attack, or methylation of ring nitrogen by endogenous agents (114). This pathway is also essential for resistance to DNA damage inflicted by exogenous DNA-damaging agents such as ionizing radiation and other radical-inducing agents (167). UNG, AP endonuclease, DNA polymerase β, and DNA ligase act sequentially through the BER pathway to repair damaged DNA (30). UNG removes deaminated cytosine (uracil) to generate apurinic or apyrimidinic sites (AP sites), class II AP endonucleases remove AP sites after cleaving the DNA strand 5′ to the AP site, DNA polymerase β fills the gap by DNA repair synthesis, and DNA ligase finishes the repair process by closing the gap (20, 30, 124, 203).

MSV208 is most similar to the equine herpesvirus UNG (34% identity over 214 amino acids); it has significant identity to UNG in bacteria (*Bacillus subtilis* PIR accession no. S39712), lower eukaryotes (slime mold [*Dictostelium* sp.]; GenBank accession no. U32866), and higher eukaryotes (mouse [*Mus musculus*]; GenBank accession no. X99018) (29 to 34% identity over 195 to 216 amino acids). Surprisingly, MSV208 exhibits much less similarity to the poxvirus UNG homologues in vaccinia virus (19% identity over 78 amino acids) and FPV (30% identity over 117 amino acids). The MsEPV gene does, however, contain amino acid substitutions at the predicted UNG active site. Most notably, an aspartic acid residue has been replaced by an arginine at the N-glycosylic bond cleavage site (residue 60 in MSV208). The significance of these residue changes for protein function is not known. In vaccinia virus and human cytomegalovirus, UNG enzymes have been implicated in other functions, including establishing the correct temporal progression of DNA synthesis and viral replication (126, 146, 181). Thus, it is possible that MSV208 performs other functions unrelated to BER.

Homologues of both class II AP endonucleases and DNA polymerase β are encoded by MSV117. This gene has homology to eukaryotic and viral class II AP endonucleases at its amino terminus (amino acids 1 to 296) and homology to DNA polymerase β at its carboxy terminus (amino acids 296 to 607). MSV117 contains most residues of AP endonuclease class II Prosite signatures 2 and 3 (PS00730 and PS00731, respectively), which include conserved and potentially metal-binding cysteine and histidine residues. The degree of identity of MSV117 to eukaryotic and viral DNA polymerase β enzymes is highest at the catalytic region (31% identity, over 186 amino acids, to rat DNA polymerase β (GenBank accession no. M13961) and includes the DNA polymerase X Prosite signature (PS00522). Rat DNA polymerase β is a smaller protein (335 amino acids) consisting of two domains connected by a protease-sensitive region (96). The 31-kDa carboxyl-terminal domain contains the residues critical for catalytic activity as defined by the crystal structure (29). These residues are present in MsEPV (Arg-469, Asp-476, and Asp-478). ASFV, another cytoplasmic DNA virus with an arthropod host, encodes both an AP endonuclease and a DNA polymerase β in separate ORFs (212). The ASFV DNA polymerase β (174 amino acids) is the smallest functional DNA polymerase β enzyme that has been described (137). MSV117 has homology to both ASFV genes.

The fusion of the AP endonuclease and DNA polymerase β genes into one gene has not been previously described. This fusion is reasonable, however, since the activities of both enzymes are coordinately required for DNA BER (14). Although a common strategy for RNA viruses and retroviruses, polyprotein processing has been observed for only a few vaccinia virus and ASFV structural proteins (4, 130, 173). The absence of vaccinia virus and ASFV proteolytic cleavage consensus sequences (Ala-Gly/Ala-Ser and Gly-Gly-X, respectively) at the AP endonuclease-DNA polymerase β junction in MSV117 suggests that this gene product may have a dual enzymatic function.

A role in DNA repair or recombination is possible for

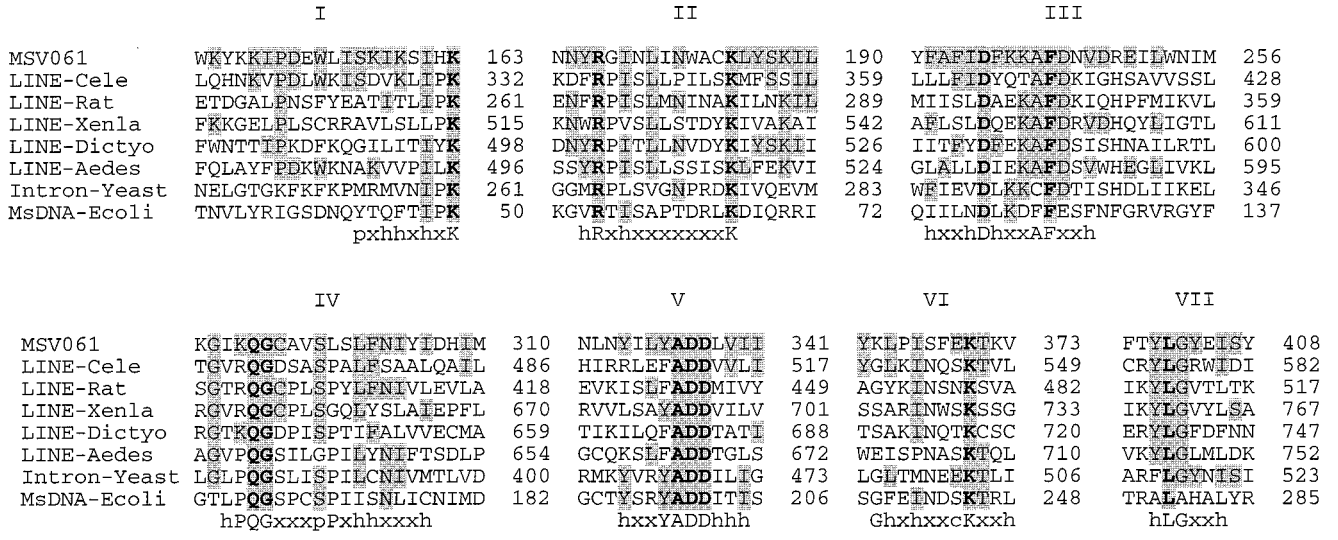


FIG. 3. Multiple amino acid sequence alignments of MSV061 with RTs. The seven RT motifs are indicated with roman numbers I to VII (210). Boldfaced letters indicate invariant amino acids, shaded letters indicate amino acids that are identical to corresponding ones in MSV061, and consensus residues are indicated at the bottom as follows: h, hydrophobic; p, small polar; c, charged; and x, any amino acid. Uppercase letters indicate the one-letter amino acid code. Amino acid positions are indicated on the right. Abbreviations: LINE, LINE type of RT; Intron, group II intron; MsDNA, multicopy single-stranded DNA; Cele, *Caenorhabditis elegans*, accession no. U00063; Rat, *Rattus norvegicus*, accession no. X61294; Xenla, *Xenopus laevis*, accession no. P14381; Dictyo, *Dictyostelium discoideum*, accession no. X57031; Aedes, *Aedes aegypti*, accession no. M95171; Yeast, *Saccharomyces cerevisiae*, accession no. P21325; Ecoli, *E. coli*, accession no. V00694.

MSV117. DNA polymerase β is the simplest naturally occurring DNA polymerase known, and it is thought to function in a variety of repair mechanisms, including mismatched base repair (203), AP lesion repair (123), and monofunctional adduct repair (34). DNA polymerase β also seems to be involved in a repair-type DNA synthesis associated with recombination (68, 77) and with replicative DNA synthesis (183).

MSV235 shares similarity to class II CPD photolyases from marsupials, fish, insects, plants, and bacteria (38 to 57% identity over 428 to 445 amino acids) (Fig. 2B). This gene represents the first photolyase homologue found in a viral genome. CPD photolyase is a photoreactive enzyme that mediates repair of UV-induced CPDs in DNA (71, 87). The predicted protein of 466 residues exhibits 119 of 141 conserved class II residues (214). Both class II Prosite signatures (PS01083 and PS01084) in the carboxyl-terminal region (residues 298 to 312 and 425 to 444, respectively) are present, except for a conservative arginine-to-lysine substitution at position 306. Eukaryotic photolyases possess a protruding amino terminus with three regions of clustered positively charged amino acids which have been proposed to contain sequences for nuclear or mitochondrial transport (214). Consistent with a cytoplasmic mode of replication, these regions are absent from MSV235.

The importance of light-dependent DNA repair mechanisms in maintaining virus populations in nature has recently been demonstrated. Host cell light-dependent repair mechanisms have been reported to restore infectivity in up to 52% of sunlight-damaged viruses in natural marine virus communities (197). The ubiquity of CPD enzymes in nature (they are found in bacteria, plants, and mammals), the efficiency of light energy to repair UV-induced DNA damage, the unienzymatic nature of the system (71), and the detrimental effects of UV damage on survival of insect DNA viruses (8) suggest that a photolyase gene might be found in an insect virus. A virus-encoded photorepair system may thus confer a selective advantage for MsEPV in nature, where long periods of environmental exposure may occur.

MSV150 is a homologue of vaccinia virus genes D9R and D10R (Table 2) (91). All three genes contain the Prosite signature (PS00893) for MutT proteins. The amino acid identity of MSV150 to Shope fibroma, vaccinia, and molluscum contagiosum virus D10R homologues is 28 to 35% over 87 to 91 residues. MSV150 colinearity to D10R is interrupted by two regions (amino acids 65 to 117 and 146 to 180) which are absent in the vaccinia virus homologue. Although the specific function of the vaccinia virus D9R and D10R homologues is unknown (91), bacterial *mutT* pyrophosphohydrolase genes help prevent DNA damage and assure fidelity of RNA transcription within the GO error avoidance system that is responsible for removing an oxidatively damaged form of guanine (8-hydroxyguanine or 7,8-dihydro-8-oxoguanine) from both DNA and the nucleotide pools (125, 184).

MSV061 has significant homology to LINE-type RTs. The seven conserved regions (domains I to VII) characteristic of diverse retroelements, which include the two critical RT-identifying motifs Asp-h-2X-Ala-Phe and Tyr-h-Asp-Asp-X-3h (where h is any hydrophobic amino acid and X is any amino acid), are also present in MSV061 (Fig. 3) (210). Domains II, III, V, and VII are perfectly conserved, while domains I, IV, and VI contain one, two, and one substituted residues, respectively (Fig. 3). Multiple alignments and phylogenetic trees generated by the neighbor-joining method with 1,000 bootstrap replicates (161) show that MSV061 is most closely related (98% bootstrap support) to the LINE-type transposable elements and is least closely related to RT from yeast introns and *E. coli* MsDNA (Fig. 3 and data not shown). Other distinctive features of LINE retrotransposons are missing in MSV061, suggesting that it may be the remnant of an old transposition or a truncated LINE. Genes normally adjacent to LINE RTs, such as ORF1 or zinc finger-containing ORFs, are not found adjacent to MSV061 (46, 152, 209). A triple 21-bp repeat located immediately 3′ of the MSV061 translational stop codon may be the remnant of a transpositional event. All available data suggest that MSV061 is a functional viral gene: critical RT motifs are conserved (Fig. 3), the gene shows normal MsEPV base composition with typical MsEPV codon usage (data not shown), and a potential late promoter (TAATG) is located at the translational start site of the ORF.

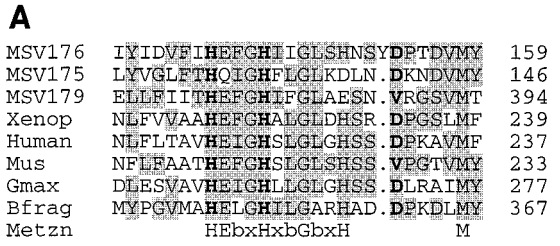
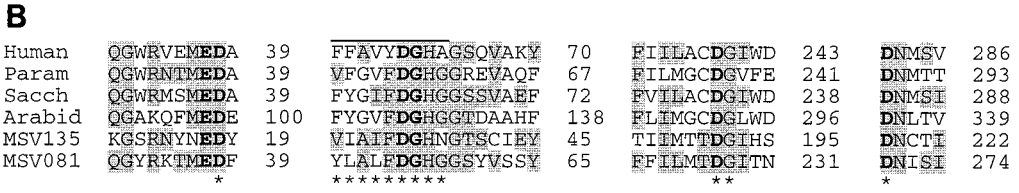


FIG. 4. Multiple amino acid sequence alignments of MsEPV ORFs with protein modification enzymes. (A) Alignment of MSV175, MSV176, and MSV179 with the catalytic or zinc-binding regions of zinc-dependent proteases. Boldfaced letters represent amino acids which are either histidine zinc ligands or glutamic acid catalytic residues, and shaded residues represent amino acids with identity to the corresponding MsEPV ORF. The consensus for the metzincin (Metzn) subfamily (81) is exhibited underneath. (where b is any bulky hydrophobic amino acid and x is any amino acid). Abbreviations: Xenop, *Xenopus laevis*, accession no. L49412; Human, *Homo sapiens*, accession no. P39900; Mus, *Mus musculus*, accession no. L36244; Gmax, *Glycine max*, accession no. U63725; Bfrag, *Bacteroides fragilis*, accession no. U90931. (B) Alignment of MSV081 and MSV135 with eukaryotic PP2C proteins. Boldfaced letters represent metal-coordinating residues (28), asterisks mark highly conserved residues, shaded residues represent amino acids with identity to MsEPV, and overlined residues mark the Prosite signature (PS00142). Abbreviations: Human, *Homo sapiens*, accession no. P35813; Param, *Paramecium tetraurelia*, accession no. Z36985; Sacch, *Saccharomyces cerevisiae*, accession no. U72346; Arabid, *Arabidopsis thaliana*, accession no. U78721.



Roles for MSV061 in DNA repair, viral DNA replication, or possibly gene acquisition are all plausible. LINE RT-mediated repair of double-strand chromosomal breaks has recently been demonstrated (129, 185). RTs from human L1 or yeast Ty1 or from the trypanosomatid protozoan *Crithidia* sp. (CRE1 transposon) can repair double-strand breaks by the insertion of complementary DNA at the break site. In the absence of homologous recombination, RTs repair double-strand breaks by nonhomologous end joining with capture of DNA within the cleavage site (129, 185).

Long terminal repeat (LTR)-containing retrotransposons have been found integrated in other DNA viruses. Integration of LTR-type retrotransposons into baculovirus AcNPV DNA has been described previously (49). This retroelement is flanked by LTRs and contains three ORFs similar in size and location to the *gag*, *pol*, and *env* genes of retroviruses (107). Also, integrated sequences of avian reticuloendotheliosis virus have been recently identified in field and vaccine strains of FPV, thus demonstrating that retroviral genomes can be integrated into the DNA of large cytoplasmic viruses (73). MSV061 does not, however, resemble the RT found in either of these LTR-type transposable elements, and the MsEPV genome does not contain any other retroviral elements associated with LTR transposons, such as *gag*, *pol*, *env*, RNase H, integrase, or LTR DNA sequences.

Protein modification. Active participation of MsEPV in protein modification is indicated by the presence of eight viral and cellular gene homologues. These homologues include two protein kinases (MSV154 and MSV173), two type 2C cellular protein phosphatases (MSV081 and MSV135), ubiquitin (MSV144), and three metalloproteases (MSV175, MSV176, and MSV179).

MSV154 and MSV173 are similar to the two vaccinia virion-associated serine/threonine protein kinases, VPK1 (B1R) and VPK2 (F10L), respectively (Table 2). Both ORFs contain the conserved catalytic region IV with an active-site motif of serine/threonine protein kinases (Prosite PS00108), and MSV173 has region I with a protein kinase ATP-binding signature (Prosite PS00107) (69). Although MSV154 lacks the glycine residues that are conserved in the ATP-binding region of other poxvirus and mammalian homologues, it does contain the lysine ATP-binding residue (Prosite PS00107) essential for the kinase activity found in vaccinia virus VPK1 (113). VPK1 is necessary for vaccinia virus DNA replication (151), and VPK2 also appears to be essential for virus viability (112).

MSV081 and MSV135 encode protein phosphatase 2C (PP2C) homologues which are similar to each other and to PP2Cs from a broad range of organisms (Fig. 4A). To our knowledge, this is the first report of a *PP2C* gene in a viral genome. PP2C is the prototypic member of a large family of Mg^{2+}/Mn^{2+}-dependent protein serine/threonine phosphatases (PPM family) present in both eukaryotes and prokaryotes (11). The six invariant metal-coordinating residues common to all PP2C amino-terminal catalytic domains are conserved in MSV081 and MSV135 (Fig. 4A) (11). In addition, MSV081 contains all 8 amino acids present in the PP2C Prosite signature (PS01032) while MSV135 has only a single substitution (Fig. 4A). The different sizes of MSV081 and MSV135 (357 and 239 residues, respectively) indicate that the two are isoforms. MSV081 contains a signal peptide and a cleavage site at the amino terminus, suggesting that it is a secreted protein. Neither ORF contains the 90-amino-acid carboxyl-terminal region characteristic of mammalian PP2Cs (28). Among other functions, PP2C reverses stress-activated protein kinase cascades in the fission yeast (171), inactivates cystic fibrosis transmembrane conductance regulation in humans (189), determines cell fate in bacteria (36), and promotes sex determination in *Caenorhabditis elegans* (22). Although the pleiotropic functions of this enzyme preclude predictions of specific roles during viral infection, a role in regulation of host intracellular signaling pathways is likely.

MSV144 encodes a ubiquitin homologue (Table 1). Ubiquitin is a highly conserved protein which forms covalent attachments to protein substrates and induces degradation of targeted proteins by the 26S proteasome complex (23). Amino acid identity between MSV144 and eukaryotic ubiquitin (83 to 88%) includes residues required for protein ubiquitination (data not shown). This percentage of identity is lower than that observed among eukaryotic ubiquitin genes (approximately 96%). Several baculoviruses also encode ubiquitin genes (*v-ubi*) which are among the most divergent known (76% amino acid identity to the mammalian ubiquitin consensus) (61, 160, 194).

The presence of ubiquitin and ubiquitin-conjugating enzymes in different arthropod viruses (59, 75, 150, 194) and the

role of ubiquitination in insect development (60) suggest that MSV144 performs an insect-host-related function. ASFV, another cytoplasmic DNA virus with an arthropod host (144), encodes a ubiquitin-conjugating enzyme and incorporates ubiquitinated proteins into the virion (75, 76, 154). The baculovirus v-*ubi* product is a nonessential structural protein that affects viral growth in cell culture (150). Covalent attachment of cellular ubiquitin to specific targets and their subsequent degradation affect numerous processes, including regulation of gene expression, cell cycle, signal transduction, apoptosis, receptor-mediated endocytosis, and antigen processing (23, 199). Indeed, over 45 confirmed or putative cellular substrates for ubiquitination have been identified, and many are from independent cellular regulatory pathways (199).

MsEPV encodes homologues of two types of metalloproteases (Table 1). The first type, represented by MSV056, is a homologue of vaccinia virus G1L. Like G1L, this gene contains an amino-terminal His-2X-Glu-His inverted metalloprotease motif and downstream glutamate residues (201). The presence of a homologue for G1L, a protein known to be involved in virion core protein processing, and the presence of potential proteolytic cleavage sites in virion core protein homologues in MsEPV suggest conservation in poxvirus structural protein processing and morphogenesis (193, 200).

A second type of metalloprotease catalytic domain, His-Glu-2X-His (86), characterizes two of three similar ORFs (Fig. 4B). MSV176 and MSV179 each contain a perfect His-Glu-2X-His consensus, while MSV175 has a glutamic acid-to-glutamine substitution at the active-site residue (Fig. 4B). These three ORFs also contain residues, including a third histidine zinc ligand and Met turn region downstream of the core His-Glu-2X-His domain, which are consistent with the metzincin subfamily of zinc-dependent metalloproteases (16, 81). In addition, all three ORFs contain putative amino-terminal signal peptides which are common among extracellular metalloendopeptidases (149). The presence of the His-Glu-2X-His motif and potential signal peptide and the significant degree of similarity to matrixins (mammalian extracellular matrix metalloproteinases) suggest that at least MSV176 and MSV179 may function as extracellular metalloproteases.

Baculoviruses have been shown to encode a metalloprotease (enhancin) which enhances virulence during infection by digesting the proteins of the host midgut peritrophic membrane (106, 156). A protein with enhancin-like activity has also been reported in *Pseudaletia separata* EPV (211). Thus, MSV176 and MSV179 may perform a similar host-related function in MsEPV infection.

Cellular functions. MSV048 has significant homology to triacylglyceride lipases found in fungi (*Rhizopus* spp.), eubacteria (*Synechocystis* spp.), protozoa (*Plasmodium* spp.), and higher plants (*Ipomoea* spp.) and similarity at the potential lipase active site of previously described cowpox and ectromelia virus ORFs (196) (Fig. 5D). MSV048 contains a potential catalytic triad (Ser-173, Asp-227, and His-265), the Prosite signature (PS00120), and a high overall degree of amino acid similarity to known lipases (Fig. 5D; Table 1) (18, 31, 32). The predicted size of MSV048 (288 amino acids) is similar to that of most fungal lipases (265 to 297 amino acids), and it exhibits 29% identity over 158 amino acids to the most closely related lipase (from *Rhizopus niveous*). The presence of a potential signal peptide suggests that the protein is secreted. Given that the grasshopper fat body is the major organ infected by MsEPV (72) and is also the main site of triacylglycerol storage (50, 188), this viral lipase could conceivably be involved in the hydrolysis of lipids, perhaps functioning as a virulence factor.

MSV242 and MSV248 are similar to viral and cellular inhibitor of apoptosis genes (*iap*) (Table 1) (35, 186). Both predicted MsEPV IAP proteins contain an amino-terminal baculovirus IAP repeat motif (BIR motif; Prosite PS01282) and one C_3HC_4 RING finger motif (Prosite PS00518) at the carboxyl terminus. Like baculovirus genes, MSV242 contains two BIR motifs while the smaller gene, MSV248, contains only a single BIR. *iap* genes were initially described in baculoviruses, where they were shown to inhibit apoptosis of infected cells and to increase viral infectivity (25). *iap*-like genes have been identified in only three virus families, the *Baculoviridae* (25), *Iridoviridae* (GenBank accession no. P40629), and *Asfarviridae* (ASFV) (132), all of which have arthropod hosts. The presence of *iap* genes in these viruses suggests that an apoptotic cellular response to viral infection may be an important host defense mechanism in diverse arthropods.

MSV097 encodes a protein with homology to the EF-hand superfamily of calcium binding proteins (Fig. 5B). These include regulatory and structural proteins such as calmodulin and caltractin (135). The 12-residue EF-hand loop motif (Prosite PS00018) responsible for calcium binding is represented twice in the amino terminus of MSV097 (Fig. 5B). The level of conservation within these two EF-hand motifs indicates high-affinity calcium binding. The carboxyl terminus is, however, less similar to other calcium binding proteins. Because calcium binding proteins control multiple intracellular processes, a role for this gene in virus-cell interactions is likely.

MSV206 has similarity to bacterial glycosyltransferases involved in lipopolysaccharide capsule biosynthesis and pathogenicity (Fig. 5C). These enzymes, which transfer sugar residues to lipid moieties or other sugar residues, have not been described previously in poxviruses (83, 145, 178). The presence of a transmembrane domain at the carboxyl terminus of MSV206 (amino acids 252 to 276) suggests that the protein is membrane associated. In bacteria, sugar polymerization is catalyzed by an inner-membrane-bound transferase complex. In MsEPV-infected grasshoppers, changes in the distribution of cell membrane carbohydrates on hemocytes have been observed (72, 127). Thus, MSV206 may modify surface polysaccharides on infected cells. MSV206 has no similarity to the baculovirus UDP-glucosyltransferase, which interferes with normal molting of virus-infected larvae by catalyzing the transfer of glucose to ecdysteroids (138).

MSV087 and MSV093 contain conserved cysteine residues indicative of redox-active centers found in glutaredoxin and thioredoxin (Fig. 5A) (39, 78, 105). MSV087 has similarity to thioredoxin of the fungus *Neurospora crassa* (23% amino acid identity over 67 amino acids) and shows partial conservation at the thioredoxin Prosite signature PS00194 (Fig. 5A). Prolines 27 and 65, which are necessary for maintenance of the *E. coli* thioredoxin structure, are conserved in MSV087 (39). Thioredoxins are small proteins of approximately 100 amino acids which participate in redox reactions via reversible oxidation of a redox-active disulfide bond (78). These enzymes are multifunctional, performing roles in DNA replication, protein synthesis, protein folding, and photosynthesis (78).

MSV093 has significant homology to the vaccinia virus gene E10R (53) and a lower level of identity to potential E10R homologues found in other cytoplasmic DNA viruses and eukaryotes (62, 97, 117, 170, 212). The yeast *ERV1* gene has been shown to function in oxidative phosphorylation and appears to function in eukaryotic cell growth (47, 117). All of these genes contain the pair of conserved cysteine residues typical of glutaredoxin and thioredoxin redox-active centers (78).

Structural proteins. Four ChPV virion core protein homologues are present in MsEPV (Table 2). MSV152, MSV164, MSV158, and MSV189 are homologues of vaccinia virus A10L,

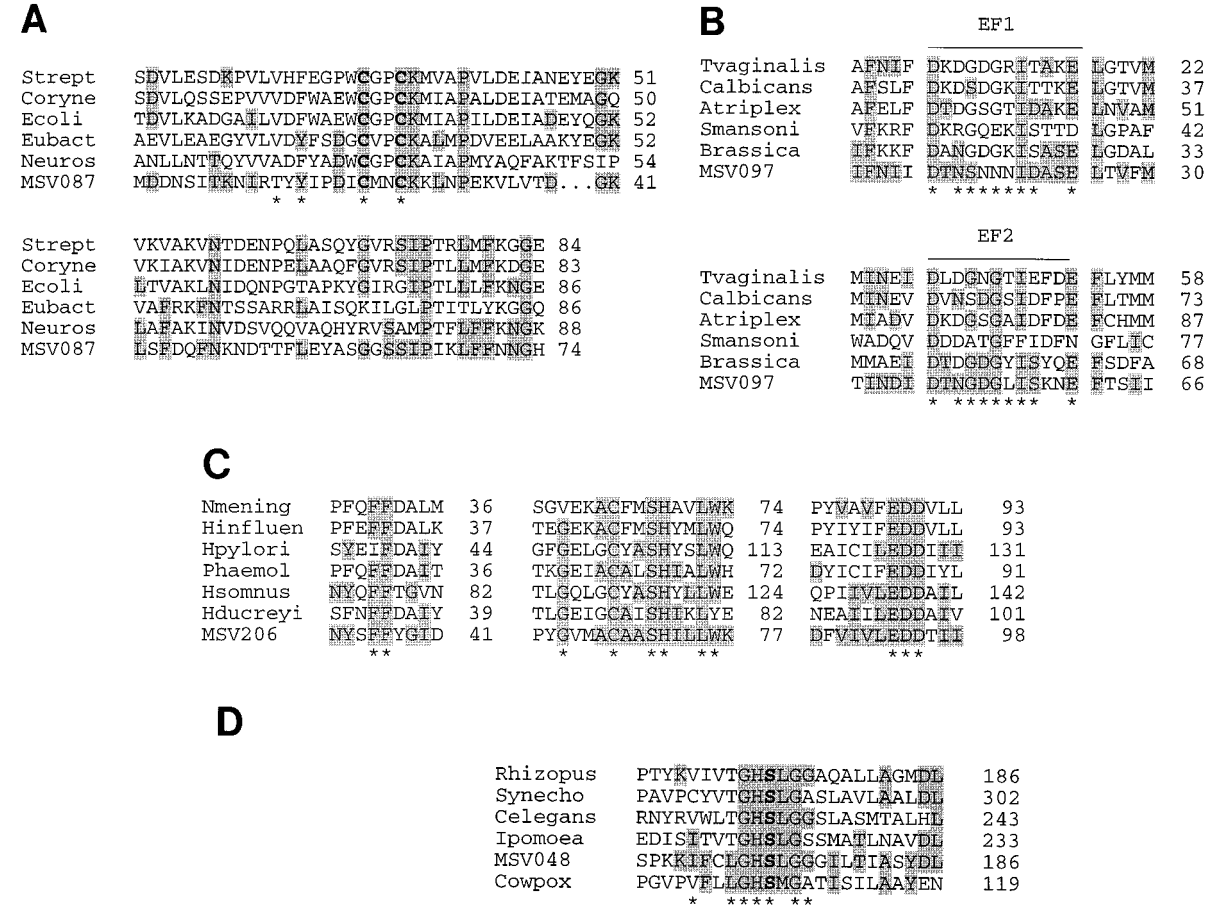


FIG. 5. Multiple amino acid sequence alignments of MsEPV ORFs with cellular and viral homologues. Boldfaced letters represent active site residues, asterisks mark residues from a Prosite signature (when indicated) or those exhibiting ≥85% conservation, and shaded residues represents amino acids with identity to MsEPV. Amino acid positions are indicated on the right. (A) Alignment of MSV048 at the active site of triacylglycerol lipases. Abbreviations: Rhizopus, *Rhizopus niveus*, accession no. D12680; Synecho, *Synechocystis* sp., accession no. D64004; Celegans, *Caenorhabditis elegans*, accession no. U97001; Ipomoea, *Ipomoea nill*, accession no. U55867; and Cowpox, CPV putative lipase, accession no. X94355. The Prosite signature is PS00120. (B) Alignment of MSV097 with two EF-hand motifs from calcium-binding proteins. Abbreviations: Tvaginalis, *Trichomonas vaginalis*, accession no. U38786; Calbicans, *Candida albicans*, accession no. P23286; Atriplex, *Atriplex nummularia*, accession no. PRF: 1906390A; Smansoni, *Schistosoma mansoni*, accession no. P15845; Brassica, *Brassica napus*, accession no. D63152. The Prosite signature is PS00018. (C) Alignment of MSV206 with bacterial glycosyltransferase genes. Abbreviations: Nmening, *Neisseria meningitidis*, accession no. U65788; Hinfluen, *Haemophilus influenzae*, accession no. U36398; Hpylori, *Helicobacter pylori*, accession no. AE000592; Phaemol, *Pasteurella haemolytica*, accession no. U15958; Hsomnus, *Haemophilus somnus*, accession no. U94833; Hducreyi, *Haemophilus ducreyi*, accession no. U58147. (D) Alignment of MSV087 with thioredoxin genes. Abbreviations: Strept, *Streptomyces aureofaciens*, accession no. P33791; Coryne, *Corynebacterium nephridii*, accession no. P00275; Ecoli, *E. coli*, accession no. M54881; Eubact, *Eubacterium acidaminophilum*, accession no. P21610; Neuros, *Neurospora crassa*, accession no. D45892. The Prosite signature is PS00194.

A3L, L4R, and I7L, which encode the virion core precursor proteins P4a, P4b, and VP8 and the core-associated I7L protein, respectively (130). Interestingly, MSV152 and MSV158, like their vaccinia virus homologues, contain potential proteolytic cleavage sites (104, 192, 193), which suggests that aspects of structural protein processing may be conserved between MsEPV and ChPVs. The proteolysis of P4a, P4b, and VP8 precursor proteins is intimately associated with normal vaccinia virus morphogenesis and production of infectious virions (104, 130, 202).

The overall degree of amino acid similarity between MsEPV and ChPV core protein homologues is low (22 to 28% identity) compared to the similarity observed among ChPV homologues (45 to 65% amino acid identity). Additionally, homologues of the following vaccinia virus structural genes are not found in MsEPV, including A4L core protein, F17L and I3L DNA-binding phosphoproteins, and the G7L, D2L, D3R, and A12L proteins associated with internal parts of intracellular mature virions (130).

Of the 14 known membrane proteins in vaccinia virus (130), only L1R is conserved in MsEPV (MSV183), perhaps reflecting the closer relationship between these proteins and host-specific functions. The vaccinia virus L1R gene, which encodes a major myristylated membrane protein that is associated exclusively with the primary membrane surrounding the virion core, is involved in virion assembly (48, 147, 148).

Homologues of five genes representing two conserved ChPV gene families are present in MsEPV. Invariant cysteine residues and putative transmembrane domains unique to each family are conserved in these MsEPV ORFs (170). MSV183 and MSV094, homologues of vaccinia virus L1R and F9L, respectively, comprise one gene family. MSV090, MSV121, and MSV142 are homologues of the vaccinia virus genes A16L, G9R, and J5L and comprise the second gene family. Although most of the genes in these two ChPV gene families remain poorly characterized, G9R and A16L have been shown to be myristylated and potentially soluble proteins (120). J5L is thought to be an essential gene (218). The presence of these two gene

families in both subfamilies of the *Poxviridae* suggests that they may provide highly conserved replicative or structural functions.

MSV069 is 25 to 28% identical to ChPV rifampin resistance proteins (vaccinia virus D13L) and 54% identical to the HaEPV D13L homologue (139). In vaccinia virus, this essential gene is associated with virion assembly and may direct the formation of Golgi complex-derived viral crescents, which are the first morphologically distinct structures observed during poxvirus assembly (130). Viral crescents have been observed in cells infected with MsEPV and other genus B EPVs (54, 57, 72, 98). Thus, a similar role for MSV069 in EPV morphogenesis is likely. Interestingly, homologues of other vaccinia virus genes associated with early events of virus morphogenesis, such as A14L and A17L (153, 155, 206), were not identified.

ChPV homologues of unknown function. MsEPV also encodes homologues of 10 ChPV genes of unknown function (Table 2). Four of these homologues exhibit a high degree of amino acid identity to their ChPV counterparts. MSV237 and vaccinia virus B2R, which exhibit 44% amino acid identity, are located in the right termini of their respective genomes (53). MSV060, which contains a putative signal peptide, and MSV132, which contains a putative amino-terminal transmembrane and signal peptide, exhibit up to 37 and 44% amino acid identity with ChPV homologues of vaccinia virus H2R and A28L, respectively (Table 2). Although MSV115 and vaccinia virus G5R exhibit only 28% amino acid identity, they contain two prominent protein motifs (Asp-Ala-Glu-Phe-X-Met-Cys-2X-Ala and Trp-Pro-4X-Asp-Gln-Asp) which are also conserved in the MCV homologue MC060R.

Gene families of unknown function. MsEPV contains 43 novel ORFs grouped into five gene families. Genes from these families are asymmetrically distributed in regions terminal to the conserved central part of the genome. Family members form tandem arrays occasionally interrupted by other genes (Fig. 1). The lack of similarity between gene family members and known ChPV genes and the presence of host range genes in similar genomic locations in ChPVs suggest that MsEPV gene families perform host range functions. Analysis of sequences upstream of the translational initiation sites suggests that four of the five gene families may be expressed early in infection (Table 1).

The leucine-rich repeat (LRR) family contains 21 ORFs which range in size from 72 to 707 amino acids. The distinguishing feature of this family is a 44-amino-acid repeat that contains regularly spaced leucine residues at 22-amino-acid intervals. A similar motif is present in the AmEPV Q3 ORF (58) and accounts for the homology between LRR family ORFs and the Q3 ORF (Table 1). MsEPV LRR regions are also similar to *Listeria* internalin proteins (SwissProt accession no. P25146), *Trypanosoma* adenylate cyclase regulatory protein (SwissProt accession no. P23799), and yeast protein phosphatase SDS22 regulatory protein (SwissProt accession no. P22194). Other known LRR domain-containing proteins participate in protein-protein interactions, and most are involved in signal transduction (90, 109, 136).

Seven MsEPV ORFs belong to the alanine-leucine-isoleucine (ALI) motif family. The distinguishing feature of this gene family is an amino-terminal motif which contains invariant alanine, leucine, and isoleucine residues. ORFs containing the ALI motif are divided into two subgroups. Subgroup I includes four ORFs (MSV024, MSV026, MSV196, and MSV204) which range in size from 190 to 222 amino acids and contain Ile-Ile-X-Cys-Phe and Ile-Asp-Leu-Try/Phe-Phe motifs between residues 100 and 150. Subgroup II includes two longer proteins (MSV023 and MSV194) which lack the carboxyl-terminal motifs of the first group and instead contain a different, highly conserved (55% over 365 amino acids) carboxyl terminus. An additional subgroup II ALI motif protein (MSV195) has only 87 amino acids. Both subgroups of ALI family ORFs demonstrated amino-terminal homology to motif regions in putative genes from Chilo iridescent virus (GenBank accession no. AF003534); AcNPV (GenBank accession no. L22858); bacteriophages BK5-T (GenBank accession no. L44593), N15 (GenBank accession no. AF064539), and A2 (GenBank accession no. Y12813); and the bacterium *Haemophilus influenzae* (GenBank accession no. U32821). While the Chilo iridescent virus ORF showed extensive homology over the length of subgroup I ALI family ORFs, subgroup II ORFs shared an extended amino-terminal motif with the invertebrate virus, bacteriophage, and bacterial ORFs (data not shown).

MsEPV ORFs MSV027, MSV029, MSV034, MSV197, MSV205, and MSV252 comprise the tryptophan (W) repeat family. The distinguishing feature of this family is the presence of a repetitive 23-amino-acid motif that contains tryptophan, leucine, and isoleucine residues. These ORFs contain 3 to 12 copies of the motif, with a tryptophan residue spaced every 23 amino acids and a leucine or isoleucine residue spaced every 11 or 12 amino acids. Three of these ORFs (MSV027, MSV197, and MSV205) contain a carboxyl-terminal $C_3H_2C_3$ (RING-H2) variation of the C_3HC_4 RING finger motif (17, 165), and one ORF (MSV029) contains a partial RING-H2 motif. The similarity of MsEPV W-repeat ORFs to an uncharacterized 41.6-kDa protein from AcNPV (Table 1) is based primarily on the periodically repeated tryptophan residues present in both proteins.

The methionine-threonine-glycine (MTG) family contains four ORFs (MSV021, MSV191, MSV198, and MSV199) and is defined by a 50-amino-acid amino-terminal motif containing these three invariant residues and by an internal motif [SxWxI(5x)FK]. MTG motif ORFs range in size from 158 to 399 amino acid residues. Interestingly, MTG family ORFs demonstrate similarity to ORFs C72R and C74R from Chilo iridescent virus (Table 1).

The serine-cysteine-glycine (SCG) family contains five ORFs (MSV214, MSV215, MSV216, MSV217, and MSV062). The distinguishing feature of this family is a 37- to 38-amino-acid motif with invariant serine, cysteine, and glycine residues around position 170 and a predicted transmembrane domain. SCG ORFs are relatively uniform in length (386 to 401 amino acids) and exhibit 34 to 46% amino acid identity.

Relationship of MsEPV to other genus B EPVs. MsEPV resembles other EPVs in genome size, DNA composition, and the presence of conserved genes (6, 7). However, the absence of some genes present in other group B EPVs, low levels of amino acid identity to gene homologues, and extensive rearrangements in gene order suggest that MsEPV is distantly related to other described genus B EPVs. Genomic differences among EPVs may reflect the diversity within the hexopod class of arthropods and the long evolutionary presence of insects.

The size of the MsEPV genome (236 kbp) is similar to those reported for AmEPV (225 kbp) and HaEPV (233 kbp) (63, 175), and the high A+T content of MsEPV (82%) is comparable to that of the AmEPV genome (81.5%) (99, 100). ITRs of approximately 7 kbp may also be present in HaEPV (175).

MsEPV contains 10 of 16 ORFs previously described in other EPVs. These include the following: AmEPV G1L (vaccinia virus I7L), G2R, G4R (vaccinia virus A28), G5R (spheroidin), and G6L (vaccinia virus NPH-1) (66) and their homologues in CfEPV, CbEPV, and HaEPV (110, 175, 217); CbEPV DNA polymerase (131); AmEPV topoisomerase (143); HaEPV rifampicin resistance protein (vaccinia virus D13L) (139; and

the F2 ORF from HaEPV (175). In addition, the 21 ORFs comprising the MsEPV LRR family are similar to the Q3 ORF that is located adjacent to the AmEPV TK gene (58). The LRR gene family and the Q3 ORF contain similar repetitive leucine-rich sequences, and both are located in terminal genomic locations.

Known EPV inclusion proteins are divergent or absent in MsEPV. The MsEPV spheroidin protein exhibits only about 20% amino acid identity to spheroidins of genus A (MmEPV) and genus B (HaEPV, AmEPV, and CbEPV) viruses. This contrasts with the 76 to 92% amino acid identity observed among other spheroidins of genus B viruses (6). In addition, MsEPV spheroidin contains approximately one-half (20 residues) the number of cysteines present in AmEPV and CbEPV (10, 215). As might be expected given the lack of observed spindle-shaped inclusions in MsEPV-infected cells (72), MsEPV does not encode a homologue of the major spindle body protein (fusolin, spindolin) that has been identified in group B and group A EPVs (6, 52, 70, 128), nor does it encode a gene with similarity to the AmEPV filament-associated protein (1). MsEPV lacks a TK gene, which is present in AmEPV, CbEPV, and CfEPV (6).

The MsEPV genome is not colinear with other described genus B EPV genomes. Of six known genes that are grouped adjacently in the AmEPV genome, MsEPV contains five homologues that are spread over 115 kbp (58, 66). This includes the spheroidin and NPH-I homologues, which despite their juxtaposition in all known genus B EPVs are separated by approximately 20 kbp in MsEPV (66). Similarly, the HaEPV F2L ORF is only 3 kbp from the HaEPV spheroidin gene, while MsEPV homologues of these genes are separated by 70 kbp (175). The MsEPV topoisomerase is adjacent to the vaccinia virus A28L homologue, while in AmEPV these genes are separated by over 100 kbp (143). Thus, the conserved colinear core of genes that has been proposed to exist for EPV genomes (65, 175) may in fact be conserved only among the lepidopteran viruses. Given the above discussion, orthopteran and lepidopteran EPVs may represent two distinct genera of entomopoxviruses.

Conclusions. The MsEPV DNA sequence provides the first view of EPV genomics. Comparison of MsEPV with ChPVs establishes the genetic core of the *Poxviridae*. EPV genome analysis provides basic knowledge of viral functions, including response to DNA damage, nucleotide metabolism, manipulation of cellular stress responses, and virulence, which underlie viral interactions with the arthropod host and the environment. An improved understanding of these interactions will lead to the design of insect biocontrol strategies with enhanced efficacy and versatility.

ACKNOWLEDGMENTS

We thank A. Ciupryk, S. Mireilles, and J. R. Emanuelli for excellent technical assistance; T. Lewis for helpful discussion and assistance; and F. Barany, K. Becker, J. D. Boeke, J. Cox, T. H. Eickbush, A. Gabriel, B. D. Hammock, R. Kretsinger, H. Krokan, S. Parikh, and A. Sancar for comments on EPV gene homologues.

REFERENCES

1. **Alaoui-Ismaili, M. H., and C. D. Richardson.** 1996. Identification and characterization of a filament-associated protein encoded by *Amsacta moorei* entomopoxvirus. J. Virol. **70:**2697–2705.
2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
3. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
4. **Andres, G., C. Simon-Mateo, and E. Vinuela.** 1993. Characterization of two African swine fever virus 220-kDa proteins: a precursor of the major structural protein p150 and an oligomer of phosphoprotein p32. Virology **194:** 284–293.
5. **Argos, P.** 1988. A sequence motif in many polymerases. Nucleic Acids Res. **16:**9909–9916.
6. **Arif, B. M.** 1995. Recent advances in the molecular biology of entomopoxviruses. J. Gen. Virol. **76:**1–13.
7. **Arif, B. M., and E. Kutstak.** 1991. The entomopoxviruses, p. 179–195. *In* E. Kutstak (ed.), Viruses of invertebrates. Marcell Dekker, Inc., New York, N.Y.
8. **Bailey, L.** 1973. Control of invertebrates by viruses, p. 534–553. *In* A. J. Gibbs (ed.), Viruses and invertebrates, vol. 31. Elsevier Publishing Company, New York, N.Y.
9. **Bairoch, A., P. Bucher, and K. Hofmann.** 1997. The PROSITE database: its status in 1997. Nucleic Acids Res. **25:**217–221.
10. **Banville, M., F. Dumas, S. Trifiro, B. Arif, and C. Richardson.** 1992. The predicted amino acid sequence of the spheroidin protein from *Amsacta moorei* entomopoxvirus: lack of homology between major occlusion body proteins of different poxviruses. J. Gen. Virol. **73:**559–566.
11. **Barford, D.** 1996. Molecular mechanisms of the protein serine/threonine phosphatases. Trends Biochem. Sci. **21:**407–412.
12. **Bayliss, C. D., and R. C. Condit.** 1995. The vaccinia virus A18R gene product is a DNA-dependent ATPase. J. Biol. Chem. **270:**1550–1556.
13. **Belfort, M., A. Moelleken, G. F. Maley, and F. Maley.** 1983. Purification and properties of T4 phage thymidylate synthetase produced by the cloned gene in an amplification vector. J. Biol. Chem. **258:**2045–2051.
14. **Bennett, R. A. O., D. M. Wilson III, D. Wong, and B. Demple.** 1997. Interaction of human apurinic endonuclease and DNA polymerase β in the base excision repair pathway. Proc. Natl. Acad. Sci. USA **94:**7166–7169.
15. **Bertholet, C., P. Stocco, E. Van Meir, and R. Wittek.** 1986. Functional analysis of the 5′ flanking sequence of a vaccinia virus late gene. EMBO J. **5:**1951–1957.
16. **Bode, W., F. X. Gomis-Ruth, and W. Stockler.** 1993. Astacins, serralysins, snake venom and matrix metalloproteinases exhibit identical zinc-binding environments (HEXXHXXGXXH and Met-turn) and topologies and should be grouped into a common family, the 'metzincins'. FEBS Lett. **331:** 134–140.
17. **Borden, K. L. B., and P. S. Freemont.** 1996. The RING finger domain: a recent example of a sequence-structure family. Curr. Opin. Struct. Biol. **6:** 395–401.
18. **Brady, L., A. M. Brzozowski, Z. S. Derewenda, E. Dodson, G. Dodson, S. Tolley, J. P. Turkenburg, L. Christiansen, B. Huge-Jensen, L. Norskov, L. Thim, and U. Menge.** 1990. A serine protease triad forms the catalytic centre of a triacylglycerol lipase. Nature **343:**767–770.
19. **Brendel, V., P. Bucher, I. R. Nourbakhsh, B. E. Blaisdell, and S. Karlin.** 1992. Methods and algorithms for statistical analysis of protein sequences. Proc. Natl. Acad. Sci. USA **89:**2002–2006.
20. **Cappelli, E., R. Taylor, M. Cevasco, A. Abbondandolo, K. Caldecott, and G. Frosina.** 1997. Involvement of XRCC1 and DNA ligase III gene products in DNA base excision repair. J. Biol. Chem. **272:**23970–23975.
21. **Carreras, C. W., and D. V. Santi.** 1995. The catalytic mechanism and structure of thymidylate synthase. Annu. Rev. Biochem. **64:**721–762.
22. **Chin-Sang, I. D., and A. M. Spence.** 1996. *Caenorhabditis elegans* sex-determining protein FEM-2 is a protein phosphatase that promotes male development and interacts directly with FEM-3. Genes Dev. **10:**2314–2325.
23. **Ciechanover, A., and A. L. Schwartz.** 1998. The ubiquitin-proteasome pathway: the complexity and myriad functions of proteins death. Proc. Natl. Acad. Sci. USA **95:**2727–2730.
24. **Claros, M. G., and G. von Heijne.** 1994. TopPred II: an improved software for membrane protein structure predictions. Comput. Appl. Biosci. **10:**685–686.
25. **Clem, R. J., and L. K. Miller.** 1994. Control of programmed cell death by the baculovirus genes *p35* and *iap*. Mol. Cell. Biol. **14:**5212–5222.
26. **Colinas, R. J., S. J. Goebel, S. W. Davis, G. P. Johnson, E. K. Norton, and E. Paoletti.** 1990. A DNA ligase gene in the Copenhagen strain of vaccinia virus is nonessential for viral replication and recombination. Virology **179:** 267–275.
27. **Dall, D., A. Sriskantha, A. Vera, J. Lai-Fook, and T. Symonds.** 1993. A gene encoding a highly expressed spindle body protein of *Heliothis armigera* entomopoxvirus. J. Gen. Virol. **74:**1811–1818.
28. **Das, A. K., N. R. Helps, P. T. W. Cohen, and D. Barford.** 1996. Crystal structure of the protein serine/threonine phosphatase 2C at 2.0 Å resolution. EMBO J. **15:**6798–6809.
29. **Davies, J. F. I., R. J. Almassy, Z. Hostomska, R. A. Ferre, and Z. Hostomsky.** 1994. 2.3 Å crystal structure of the catalytic domain of DNA polymerase β. Cell **76:**1123–1133.
30. **Demple, B., and L. Harrison.** 1994. Repair of oxidative damage to DNA: enzymology and biology. Annu. Rev. Biochem. **63:**915–948.
31. **Derewenda, Z. S., U. Derewenda, and G. G. Dodson.** 1992. The crystal and molecular structure of the *Rhizomucor miehei* triacylglyceride lipase at 1.9 Å resolution. J. Mol. Biol. **227:**818–839.
32. **Derewenda, Z. S., and A. M. Sharp.** 1993. News from the interface: the

molecular structures of triacylglyceride lipases. Trends Biochem. Sci. **18:**20–25.
33. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
34. **Dianov, G., A. Price, and T. Lindahl.** 1992. Generation of single-nucleotide repair patches following excision of uracil residues from DNA. Mol. Cell. Biol. **12:**1605–1612.
35. **Duckett, C. S., V. E. Nava, R. W. Gedrich, R. J. Clem, J. L. Van Dongen, M. C. Gilfillan, H. Shiels, J. M. Hardwick, and C. B. Thompson.** 1996. A conserved family of cellular genes related to the baculovirus *iap* gene and encoding apoptosis inhibitors. EMBO J. **15:**2685–2694.
36. **Duncan, L., S. Alper, F. Arigoni, R. Losick, and P. Stragier.** 1995. Activation of cell-specific transcription by a serine phophatase at the site of asymmetric division. Science **270:**641–646.
37. **Eddy, S. R.** 1996. Hidden Markov models. Curr. Opin. Struct. Biol. **6:**361–365.
38. **Eddy, S. R., G. Mitchison, and R. Durbin.** 1995. Maximum discrimination hidden Markov models of sequence consensus. J. Comput. Biol. **2:**9–23.
39. **Ellis, L. B. M., P. Saurugger, and C. Woodward.** 1992. Identification of the three-dimensional thioredoxin motif: related structure in the ORF3 protein of the *Staphylococcus aureus mer* operon. Biochemistry **31:**4882–4891.
40. **Erlandson, M.** 1991. Protease activity associated with occlusion body preparations of an entomopoxvirus from *Melanoplus sanguinipes*. J. Invertebr. Pathol. **57:**255–263.
41. **Erlandson, M. A., and D. A. Streett (ed.).** 1997. Entomopoxviruses associated with grasshoppers and locusts: biochemical characterization, p. 131–146. *In* M. S. Goettel and D. L. Johnson (ed.), Memoirs of the Entomological Society of Canada, vol. 171. Entomological Society of Canada, Ottawa, Ontario, Canada.
42. **Ewing, B., and P. Green.** 1998. Base-calling of automated sequencer traces using *Phred*. II. Error probabilities. Genome Res. **8:**186–194.
43. **Ewing, B., L. Hillier, M. C. Wendl, and P. Green.** 1998. Base-calling of automated sequencer traces using *Phred*. I. Accuracy assessment. Genome Res. **8:**175–185.
44. **Fabian, P., J. Murvai, Z. Hatsagi, K. Vlahovicek, H. Hegyi, and S. Pongor.** 1997. The SBASE protein domain library, release 5.0: a collection of annotated protein sequence segments. Nucleic Acids Res. **25:**240–243.
45. **Felsenstein, J.** 1989. PHYLIP—phylogeny inference package (version 3.2). Cladistics **5:**164–166.
46. **Feng, Q., J. V. Moran, H. H. Kazazian, Jr., and J. D. Boeke.** 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell **87:**905–916.
47. **Francavilla, A., N. L. Vujanovic, L. Polimeno, A. Azzarone, A. Iacobellis, A. Deleo, M. Hagiya, T. L. Whiteside, and T. E. Starzl.** 1997. The *in vivo* effect of hepatotrophic factors augmenter of liver regeneration, hepatocyte growth factor, and insulin-like growth factor-II on liver natural killer cell functions. Hepatology **25:**411–415.
48. **Franke, C. A., E. M. Wilson, and D. E. Hruby.** 1990. Use of a cell-free system to identify the vaccinia virus L1R gene product as the major late myristylated virion protein M25. J. Virol. **64:**5988–5996.
49. **Friesen, P. D., and M. S. Nissen.** 1990. Gene organization and transcription of TED, a lepidopteran retrotransposon integrated within the baculovirus genome. Mol. Cell. Biol. **10:**3067–3077.
50. **Gade, G., and A. M. T. Beenakkers.** 1977. Adipokinetic hormone-induced lipid mobilization and cyclic AMP accumulation in the fat body of *Locusta migratoria* during development. Gen. Comp. Endocrinol. **32:**481–487.
51. **Galtier, N., M. Gouy, and C. Gautier.** 1996. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. CABIOS **12:**543–554.
52. **Gauthier, L., F. Cousserans, J. C. Veyrunes, and M. Bergoin.** 1995. The *Melolontha melolontha* entomopoxvirus (*Mm*EPV) fusolin is related to the fusolins of lepidopteran EPVs and to the 37K baculovirus glycoprotein. Virology **208:**427–436.
53. **Goebel, S. J., G. P. Johnson, M. E. Perkus, S. W. Davis, J. P. Winslow, and E. Paoletti.** 1990. The complete DNA sequence of vaccinia virus. Virology **179:**247–266.
54. **Goodwin, R. H., R. J. Milner, and C. D. Beaton.** 1991. *Entomopoxvirinae*, p. 259–285. *In* J. R. Adams and J. R. Bonami (ed.), Atlas of invertebrate viruses, vol. 8. CRC Press, Inc., Boca Raton, Fla.
55. **Gorbalenya, A. E., E. V. Koonin, A. P. Donchenko, and V. M. Blinov.** 1989. Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. Nucleic Acids Res. **17:**4713–4730.
56. **Gordon, D., C. Abajian, and P. Green.** 1998. *Consed*: a graphical tool for sequence finishing. Genome Res. **8:**192–202.
57. **Granados, R. R., and D. W. Roberts.** 1970. Electron microscopy of a poxlike virus infecting an invetebrate host. Virology **40:**230–243.
58. **Gruidl, M. E., R. L. Hall, and R. W. Moyer.** 1992. Mapping and molecular characterization of a functional thymidine kinase from *Amsacta moorei* entomopoxvirus. Virology **186:**507–516.
59. **Guarino, L. A., G. Smith, and W. Dong.** 1995. Ubiquitin is attached to membranes of baculovirus particles by a novel type of phospholipid anchor. Cell **80:**301–309.
60. **Haas, A. L., O. Baboshina, B. Williams, and L. M. Schwartz.** 1995. Coordinated induction of the ubiquitin conjugation pathway accompanies the developmentally programmed death of insect skeletal muscle. J. Biol. Chem. **270:**9407–9412.
61. **Haas, A. L., D. J. Katzung, P. M. Reback, and L. A. Guarino.** 1996. Functional characterization of the ubiquitin variant encoded by the baculovirus *Autographa californica*. Biochemistry **35:**5385–5394.
62. **Hagiya, M., A. Francavilla, L. Polimeno, I. Ihara, H. Sakai, T. Seki, M. Shimonishi, K. A. Porter, and T. E. Starzl.** 1994. Cloning and sequence analysis of the rat augmenter of liver regeneration (ALR) gene: expression of biologically active recombinant ALR and demonstration of tissue distribution. Proc. Natl. Acad. Sci. USA **91:**8142–8146.
63. **Hall, R. L., and W. F. Hink.** 1990. Physical mapping and field inversion gel electrophoresis of *Amsacta moorei* entomopoxvirus DNA. Arch. Virol. **110:**77–90.
64. **Hall, R. L., Y. Li, J. Feller, and R. W. Moyer.** 1996. The *Amsacta moorei* entomopoxvirus spheroidin gene is improperly transcribed in vertebrate poxviruses. Virology **224:**427–436.
65. **Hall, R. L., and R. W. Moyer.** 1993. Identification of an *Amsacta* spheroidin-like protein within the occlusion bodies of *Choristoneura* entomopoxviruses. Virology **192:**179–187.
66. **Hall, R. L., and R. W. Moyer.** 1991. Identification, cloning, and sequencing of a fragment of *Amsacta moorei* entomopoxvirus DNA containing the spheroidin gene and three vaccinia virus-related open reading frames. J. Virol. **65:**6516–6527.
67. **Hammond, J. M., S. M. Kerr, G. L. Smith, and L. K. Dixon.** 1992. An African swine fever virus gene with homology to DNA ligases. Nucleic Acids Res. **20:**2667–2671.
68. **Hammond, R. A., J. K. McClung, and M. R. Miller.** 1990. Effect of DNA polymerase inhibitors on DNA repair in intact and permeable human fibroblasts: evidence that DNA polymerase δ and β are involved in DNA repair synthesis induced by *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine. Biochemistry **29:**286–291.
69. **Hanks, S. K., A. M. Quinn, and T. Hunter.** 1988. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science **241:**42–52.
70. **Hayakawa, T., J. Xu, and T. Hukuhara.** 1996. Cloning and sequencing of the gene for an enhancing factor from *Pseudaletia separata* entomopoxvirus. Gene **177:**269–270.
71. **Hearst, J. E.** 1995. The structure of photolyase: using photon energy for DNA repair. Science **268:**1858–1859.
72. **Henry, J. E., B. P. Nelson, and J. W. Jutila.** 1969. Pathology and development of the grasshopper inclusion body virus in *Melanoplus sanguinipes*. J. Virol. **3:**605–610.
73. **Hertig, C., B. E. H. Coupar, A. R. Gould, and D. B. Boyle.** 1997. Field and vaccine strains of fowlpox virus carry integrated sequences from the avian retrovirus, reticuloendotheliosis virus. Virology **235:**367–376.
74. **Higgins, D. G., J. D. Thompson, and T. J. Gibson.** 1996. Using CLUSTAL for multiple sequence alignments. Methods Enzymol. **266:**383–402.
75. **Hingamp, P. M., J. E. Arnold, R. J. Mayer, and L. K. Dixon.** 1992. A ubiquitin conjugating enzyme encoded by African swine fever virus. EMBO J. **11:**361–366.
76. **Hingamp, P. M., M. L. Leyland, J. Webb, S. Twigger, R. J. Mayer, and L. K. Dixon.** 1995. Characterization of a ubiquitinated protein which is externally located in African swine fever virions. J. Virol. **69:**1785–1793.
77. **Hirose, F., Y. Hotta, M. Yamaguchi, and A. Matsukage.** 1989. Difference in the expression level of DNA polymerase β among mouse tissues: high expression in the pachytene spermatocyte. Exp. Cell Res. **181:**169–180.
78. **Holmgren, A.** 1985. Thioredoxin. Annu. Rev. Biochem. **54:**237–271.
79. **Honess, R. W., W. Bodemer, K. R. Cameron, H.-H. Niller, B. Fleckenstein, and R. E. Randall.** 1986. The A+T-rich genome of *Herpesvirus saimiri* contains a highly conserved gene for thymidylate synthase. Proc. Natl. Acad. Sci. USA **83:**3604–3608.
80. **Honess, R. W., M. A. Craxton, L. Williams, and U. A. Gompels.** 1989. A comparative analysis of the sequence of the thymidine kinase gene of a gammaherpesvirus, herpesvirus saimiri. J. Gen. Virol. **70:**3003–3013.
81. **Hooper, N. M.** 1994. Families of zinc metalloproteases. FEBS Lett. **354:**1–6.
82. **Jaeger, B., and W. H. R. Langridge.** 1984. Infection of *Locusta migratoria* with entomopoxviruses from *Arphia conspersa* and *Melanoplus sanguinipes* grasshoppers. J. Invert. Pathol. **43:**374–382.
83. **Jennings, M. P., D. W. Hood, I. R. A. Peak, M. Virji, and R. E. Moxon.** 1995. Molecular analysis of a locus for the biosynthesis and phase-variable expression of the lacto-N-neotetraose terminal lipopolysaccharide structure in *Neisseria meningitidis*. Mol. Microbiol. **18:**729–740.
84. **Jones, D. T., W. R. Taylor, and J. M. Thornton.** 1994. A model recognition approach to the prediction of all-helical membrane protein structure and topology. Biochemistry **33:**3038–3049.
85. **Jones, D. T., W. R. Taylor, and J. M. Thornton.** 1994. A mutation data matrix for transmembrane proteins. FEBS Lett. **339:**269–275.

86. **Jongeneel, C. V., J. Bouvier, and A. Bairoch.** 1989. A unique signature identifies a family of zinc-dependent metallopeptidases. FEBS Lett. **242:** 211–214.
87. **Kanai, S., R. Kikuno, H. Toh, H. Ryo, and T. Todo.** 1997. Molecular evolution of the photolyase-blue-light photoreceptor family. J. Mol. Evol. **45:**535–548.
88. **Karlin, S., V. Brendel, and P. Bucher.** 1992. Significant similarity and dissimilarity in homologous proteins. Mol. Biol. Evol. **9:**152–167.
89. **Kerr, S. M., and G. L. Smith.** 1989. Vaccinia virus encodes a polypeptide with DNA ligase activity. Nucleic Acids Res. **17:**9039–9050.
90. **Kobe, B., and J. Deisenhofer.** 1993. Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. Nature **366:**751–756.
91. **Koonin, E. V.** 1993. A highly conserved sequence motif defining the family of MutT-related proteins from eubacteria, eukaryotes and viruses. Nucleic Acids Res. **21:**4847.
92. **Koonin, E. V., and T. G. Senkevich.** 1993. Fowlpox virus encodes a protein related to human deoxycytidine kinase: further evidence for independent acquisition of genes for enzymes of nucleotide metabolism by different viruses. Virus Genes **7:**289–295.
93. **Koonin, E. V., and T. G. Senkevich.** 1992. Vaccinia virus encodes four putative DNA and/or RNA helicases distantly related to each other. J. Gen. Virol. **73:**989–993.
94. **Koonin, E. V., R. L. Tatusov, and K. E. Rudd.** 1996. Protein sequence comparison at genome scale. Methods Enzymol. **266:**295–322.
95. **Kotwal, G. J., A. W. Hugin, and B. Moss.** 1989. Mapping and insertional mutagenesis of a vaccinia virus gene encoding a 13,800-Da secreted protein. Virology **171:**579–587.
96. **Kumar, A., S. G. Widen, K. R. Williams, P. Kedar, R. L. Karpel, and S. H. Wilson.** 1990. Studies of the domain structure of mammalian DNA polymerase β. J. Biol. Chem. **265:**2124–2131.
97. **Kutish, G. F., Y. Li, Z. Lu, M. Furuta, D. L. Rock, and J. L. Van Etten.** 1996. Analysis of 76 kb of the chlorella virus PBCV-1 330-kb genome: map positions 182 to 258. Virology **223:**303–317.
98. **Lai-Fook, J., E. M. Brooks, and D. J. Dall.** 1994. Intracellular structures associated with in vitro replication of *Heliothis armigera* entomopoxvirus. J. Invertebr. Pathol. **64:**269–272.
99. **Langridge, W. H. R., and D. W. Roberts.** 1977. Molecular weight of DNA from four entomopoxviruses determined by electron microscopy. J. Virol. **21:**301–308.
100. **Langridge, W. H. R., R. F. Bozarth, and D. W. Roberts.** 1977. The base composition of entomopoxvirus DNA. Virology **76:**616–620.
101. **Langridge, W. H. R., and J. E. Henry.** 1981. Molecular weight and base composition of DNA isolated from *Melanoplus sanguinipes* entomopoxvirus. J. Invertebr. Pathol. **37:**34–37.
102. **Langridge, W. H. R., E. Oma, and J. E. Henry.** 1983. Characterization of the DNA and structural proteins of entomopoxviruses from *Melanoplus sanguinipes*, *Arphia conspirsa*, and *Phoetaliotes nebrascensis* (Orthoptera). J. Invertebr. Pathol. **42:**327–333.
103. **Lasko, D. D., A. E. Tomkinson, and T. Lindahl.** 1990. Eukaryotic DNA ligases. Mutat. Res. **236:**277–287.
104. **Lee, P., and D. E. Hruby.** 1993. *trans* processing of vaccinia virus core proteins. J. Virol. **67:**4252–4263.
105. **LeMaster, D. M., P. A. Springer, and C. J. Unkefer.** 1997. The role of the buried aspartate of *Escherichia coli* thioredoxin in the activation of the mixed disulfide intermediate. J. Biol. Chem. **272:**29998–30001.
106. **Lepore, L. S., P. R. Roelvink, and R. R. Granados.** 1996. Enhancin, the granulosis virus protein that facilitates nucleopolyhedrovirus (NPV) infections, is a metalloprotease. J. Invertebr. Pathol. **68:**131–140.
107. **Lerch, R. A., and P. D. Friesen.** 1992. The baculovirus-integrated retrotransposon TED encodes *gag* and *pol* proteins that assemble into viruslike particles with reverse transcriptase. J. Virol. **66:**1590–1601.
108. **Levin, D. B., D. Adachi, L. L. Williams, and T. G. Myles.** 1993. Host specificity and molecular characterization of the entomopoxvirus of the lesser migratory grasshopper, *Melanoplus sanguinipes*. J. Invertebr. Pathol. **62:**241–247.
109. **Li, J., and J. Chory.** 1997. A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. Cell **90:**929–938.
110. **Li, X., J. W. Barrett, L. Yuen, and B. M. Arif.** 1997. Cloning, sequencing and transcriptional analysis of the *Choristoneura fumiferana* entomopoxvirus spheroidin gene. Virus Res. **47:**143–154.
111. **Li, Y., R. L. Hall, and R. W. Moyer.** 1997. Transient, nonlethal expression of genes in vertebrate cells by recombinant entomopoxviruses. J. Virol. **71:** 9557–9562.
112. **Lin, S., and S. S. Broyles.** 1994. Vaccinia protein kinase 2: a second essential serine/threonine protein kinase encoded by vaccinia virus. Proc. Natl. Acad. Sci. USA **91:**7653–7657.
113. **Lin, S., W. Chen, and S. S. Broyles.** 1992. The vaccinia virus B1R gene product is a serine/threonine protein kinase. J. Virol. **66:**2717–2723.
114. **Lindahl, T.** 1993. Instability and decay of the primary structure of DNA. Nature **362:**709–715.
115. **Lindahl, T., and D. E. Barnes.** 1992. Mammalian DNA ligases. Annu. Rev. Biochem. **61:**251–281.
116. **Lipman, D. J., S. F. Altschul, and J. D. Kececioglu.** 1989. A tool for multiple sequence alignment. Proc. Natl. Acad. Sci. USA **86:**4412–4415.
117. **Lisowsky, T.** 1994. ERV1 is involved in the cell-division cycle and the maintenance of mitochondrial genomes in *Saccharomyces cerevisiae*. Curr. Genet. **26:**15–20.
118. **Luo, J., and F. Barany.** 1996. Identification of essential residues in *Thermus thermophilus* DNA ligase. Nucleic Acids Res. **24:**3079–3085.
119. **Lytvyn, V., Y. Fortin, M. Banville, B. Arif, and C. Richardson.** 1992. Comparison of the thymidine kinase genes from three entomopoxviruses. J. Gen. Virol. **73:**3235–3240.
120. **Martin, K. H., D. W. Grosenbach, C. A. Franke, and D. E. Hruby.** 1997. Identification and analysis of three myristylated vaccinia virus late proteins. J. Virol. **71:**5218–5226.
121. **Massung, R. F., J. C. Knight, and J. J. Esposito.** 1995. Topography of variola smallpox virus inverted terminal repeats. Virology **211:**350–355.
122. **Massung, R. F., L. Liu, J. Qi, J. C. Knight, T. E. Yuran, A. R. Kerlavage, J. M. Parsons, J. C. Venter, and J. J. Esposito.** 1994. Analysis of the complete genome of smallpox variola major virus strain Bangladesh-1975. Virology **201:**215–240.
123. **Matsumoto, Y., and D. F. Bogenhagen.** 1989. Repair of a synthetic abasic site in DNA in a *Xenopus laevis* oocyte extract. Mol. Cell. Biol. **9:**3750–3757.
124. **Matsumoto, Y., and K. Kim.** 1995. Excision of deoxyribose phosphate residues by DNA polymerase B during DNA repair. Science **269:**699–702.
125. **Michaels, M. L., and J. H. Miller.** 1992. The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine). J. Bacteriol. **174:**6321–6325.
126. **Millns, A. K., M. S. Carpenter, and A. M. DeLange.** 1994. The vaccinia virus-encoded uracil DNA glycosylase has an essential role in viral DNA replication. Virology **198:**504–513.
127. **Miranpuri, G. S., M. A. Erlandson, J. P. Gillespie, and G. G. Khachatourians.** 1992. Changes in hemolymph of the migratory grasshopper, *Melanoplus sanguinipes*, infected with an entomopoxvirus. J. Invertebr. Pathol. **60:** 274–282.
128. **Mitsuhashi, W., H. Saito, and M. Sato.** 1997. Complete nucleotide sequence of the fusolin gene from an entomopoxvirus in the cupreous chafer, *Anomala cuprea Hope* (Coleoptera: Scarabaeidae). Insect Biochem. Mol. Biol. **27:**869–876.
129. **Moore, J. K., and J. E. Haber.** 1996. Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks. Nature **383:**644–646.
130. **Moss, B.** 1996. Poxviridae: the viruses and their replication, p. 2637–2671. *In* B. N. Fields, D. M. Knipe, and P. M. Howley (ed.), Fields virology. Lippincott-Raven, Philadelphia, Pa.
131. **Mustafa, A., and L. Yuen.** 1991. Identification and sequencing of the *Choristoneura biennis* entomopoxvirus DNA polymerase gene. J. DNA Sequencing Mapping **2:**39–45.
132. **Neilan, J. G., Z. Lu, G. F. Kutish, L. Zsak, T. G. Burrage, M. V. Borca, C. Carrillo, and D. L. Rock.** 1997. A BIR motif containing gene of African swine fever virus, 4CL, is nonessential for growth *in vitro* and viral virulence. Virology **230:**252–264.
133. **Neuwald, A. F., and P. Green.** 1994. Detecting patterns in protein sequences. J. Mol. Biol. **239:**698–712.
134. **Neuwald, A. F., J. S. Liu, D. J. Lipman, and C. E. Lawrence.** 1997. Extracting protein alignment models from the sequence database. Nucleic Acids Res. **25:**1665–1677.
135. **Niki, I., H. Yokokura, T. Sudo, M. Kato, and H. Hidaka.** 1996. Ca^{2+} signalling and intracellular Ca^{2+} binding proteins. J. Biochem. **120:**685–698.
136. **Ohkura, H., and M. Yanagida.** 1991. *S. pombe* gene sds22+ essential for a midmitotic transition encodes a leucine-rich repeat protein that positively modulates protein phosphatase-1. Cell **64:**149–157.
137. **Oliveros, M., R. J. Yanez, M. L. Salas, J. Salas, E. Vinuela, and L. Blanco.** 1997. Characterization of an African swine fever virus 20-kDa DNA polymerase involved in DNA repair. J. Biol. Chem. **272:**30899–30910.
138. **O'Reilly, D. R., and L. K. Miller.** 1989. A baculovirus blocks insect molting by producing ecdysteroid UDP-glucosyl transferase. Science **245:**1110–1112.
139. **Osborne, R. J., T. M. Symonds, A. Sriskantha, J. Lai-Fook, C. A. Fernon, and D. J. Dall.** 1996. An entomopoxvirus homologue of the vaccinia virus D13L-encoded 'rifampicin resistance' protein. J. Gen. Virol. **77:**839–846.
140. **Pearson, A., C. Richardson, and L. Yuen.** 1991. The 5′ noncoding region sequence of the *Choristoneura biennis* entomopoxvirus spheroidin gene functions as an efficient late promoter in the mammalian vaccinia expression system. Virology **180:**561–566.
141. **Pearson, W. R.** 1990. Rapid and sensitive sequence comparison with FASTP and FASTA. Methods Enzymol. **183:**63–98.
142. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85:**2444–2448.
143. **Petersen, B. O., R. L. Hall, R. W. Moyer, and S. Shuman.** 1997. Characterization of a DNA topoisomerase encoded by *Amsacta moorei* entomopoxvirus. Virology **230:**197–206.
144. **Plowright, W., J. Parker, and M. A. Peirce.** 1969. African swine fever virus

in ticks (*Ornithodoros moubata*, Murray) collected from animal burrows in Tanzania. Nature **221:**1071–1073.
145. **Potter, M. D., and R. Y. C. Lo.** 1996. Cloning and characterization of the *galE* locus of *Pasteurella haemolytica* A1. Infect. Immun. **64:**855–860.
146. **Prichard, M. N., G. M. Duke, and E. S. Mocarski.** 1996. Human cytomegalovirus uracil DNA glycosylase is required for the normal temporal regulation of both DNA synthesis and viral replication. J. Virol. **70:**3018–3025.
147. **Ravanello, M. P., and D. E. Hruby.** 1994. Characterization of the vaccinia virus L1R myristylprotein as a component of the intracellular virion envelope. J. Gen. Virol. **75:**1479–1483.
148. **Ravanello, M. P., and D. E. Hruby.** 1994. Conditional lethal expression of the vaccinia virus L1R myristylated protein reveals a role in virion assembly. J. Virol. **68:**6401–6410.
149. **Rawlings, N. D., and A. J. Barrett.** 1995. Evolutionary families of metallopeptidases. Methods Enzymol. **248:**183–228.
150. **Reilly, L. M., and L. A. Guarino.** 1996. The viral ubiquitin gene of *Autographa californica* nuclear polyhedrosis virus is not essential for viral replication. Virology **218:**243–247.
151. **Rempel, R. E., M. K. Anderson, E. Evans, and P. Traktman.** 1990. Temperature-sensitive vaccinia virus mutants identify a gene with an essential role in viral replication. J. Virol. **64:**574–583.
152. **Ring, M., T. A. Pfeifer, and T. A. Grigliatti.** 1996. Identification of a 5′ truncated non-LTR-retrotransposon, YAKPs1, from the variegated cutworm, *Peridroma saucia*, using PCR. Insect Biochem. Mol. Biol. **26:**511–518.
153. **Rodríguez, D., M. Esteban, and J. R. Rodríguez.** 1995. Vaccinia virus A17L gene product is essential for an early step in virion morphogenesis. J. Virol. **69:**4640–4648.
154. **Rodriguez, J. M., M. L. Salas, and E. Vinuela.** 1992. Genes homologous to ubiquitin-conjugating proteins and eukaryotic transcription factor SII in African swine fever virus. Virology **186:**40–52.
155. **Rodríguez, J. R., C. Risco, J. L. Carrascosa, M. Esteban, and D. Rodríguez.** 1998. Vaccinia virus 15-kilodalton (A14L) protein is essential for assembly and attachment of viral crescents to virosomes. J. Virol. **72:**1287–1296.
156. **Roelvink, P. W., B. G. Corsaro, and R. R. Granados.** 1995. Characterization of the *Helicoverpa armigera* and *Pseudaletia unipuncta* granulovirus enhancin genes. J. Gen. Virol. **76:**2693–2705.
157. **Rosel, J., and B. Moss.** 1985. Transcriptional and translational mapping and nucleotide sequence analysis of a vaccinia virus gene encoding the precursor of the major core polypeptide 4b. J. Virol. **56:**830–838.
158. **Rosel, J. L., P. L. Earl, J. P. Weir, and B. Moss.** 1986. Conserved TAAATG sequence at the transcriptional and translational initiation sites of vaccinia virus late genes deduced by structural and functional analysis of the *Hin*dIII H genome fragment. J. Virol. **60:**436–449.
159. **Roseman, N. A., and D. E. Hruby.** 1987. Nucleotide sequence and transcript organization of a region of the vaccinia virus genome which encodes a constitutively expressed gene required for DNA replication. J. Virol. **61:**1398–1406.
160. **Russell, R. L. Q., and G. F. Rohrmann.** 1993. Nucleotide sequence of the ubiquitin-39K region from the *Orgyia pseudotsugata* multinucleocapsid nuclear polyhedrosis virus genome. J. Gen. Virol. **74:**1191–1195.
161. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for recontructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.
162. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
163. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
164. **Sanz, P., J.-C. Veyrunes, F. Cousserans, and M. Bergoin.** 1994. Cloning and sequencing of the spherulin gene, the occlusion body major polypeptide of the *Melolontha melolontha* entomopoxvirus (*Mm*EPV). Virology **202:**449–457.
165. **Saurin, A. J., K. L. B. Borden, M. N. Boddy, and P. S. Freemont.** 1996. Does this have a familiar RING? Trends Biochem. Sci. **21:**208–214.
166. **Schuler, G. D., S. F. Altschul, and D. J. Lipman.** 1991. A workbench for multiple alignment construction and analysis. Proteins **9:**180–190.
167. **Seeberg, E., L. Eide, and M. Bjoras.** 1995. The base excision repair pathway. Trends Biochem. Sci. **20:**391–397.
168. **Sekiguchi, J., and S. Shuman.** 1997. Mutational analysis of vaccinia virus topoisomerase identifies residues involved in DNA binding. Nucleic Acids Res. **25:**3649–3656.
169. **Senkevich, T. G., J. J. Bugert, J. R. Sisler, E. V. Koonin, G. Darai, and B. Moss.** 1996. Genome sequence of a human tumorigenic poxvirus: prediction of specific host response-evasion genes. Science **273:**813–816.
170. **Senkevich, T. G., E. V. Koonin, J. J. Bugert, G. Darai, and B. Moss.** 1997. The genome of *Molluscum contagiosum* virus: analysis and comparison with other poxviruses. Virology **233:**19–42.
171. **Shiozaki, K., and P. Russell.** 1995. Counteractive roles of protein phosphatase 2C (PP2C) and a MAP kinase kinase homolog in the osmoregulation of fission yeast. EMBO J. **14:**492–502.
172. **Shuman, S., E. M. Kane, and S. G. Morham.** 1989. Mapping the active-site tyrosine of vaccinia virus DNA topoisomerase I. Proc. Natl. Acad. Sci. USA **86:**9793–9797.
173. **Simon-Mateo, C., G. Andres, and E. Vinuela.** 1993. Polyprotein processing in African swine fever virus: a novel gene expression strategy for a DNA virus. EMBO J. **12:**2977–2987.
174. **Simpson, D. A., and R. C. Condit.** 1995. Vaccinia virus gene A18R encodes an essential DNA helicase. J. Virol. **69:**6131–6139.
175. **Sriskantha, A., R. J. Osborne, and D. J. Dall.** 1997. Mapping of the *Heliothis armigera* entomopoxirus (HaEPV) genome, and analysis of genes encoding the HaEPV spheroidin and nucleoside triphosphate phosphohydrolase 1 proteins. J. Gen. Virol. **78:**3115–3123.
176. **Staden, A., and A. D. McLachlan.** 1982. Codon preference and its use in identifying protein coding regions in long DNA sequences. Nucleic Acids Res. **10:**141–156.
177. **Staden, R.** 1982. An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. Nucleic Acids Res. **10:**2951–2961.
178. **Stevens, M. K., J. Klesney-Tait, S. Lumbley, K. A. Walters, A. M. Joffe, J. D. Radolf, and E. J. Hansen.** 1997. Identification of tandem genes involved in lipooligosaccharide expression by *Haemophilus ducreyi*. Infect. Immun. **65:**651–660.
179. **Streett, D. A., E. A. Oma, and J. E. Henry.** 1990. Cross infection of three grasshopper species with the *Melanoplus sanguinipes* entomopoxvirus. J. Invertebr. Pathol. **56:**419–421.
180. **Streett, D. A., S. A. Woods, and M. A. Erlandson (ed.).** 1997. Entomopoxviruses of grasshoppers and locusts: biological control potential, p. 115–130. *In* M. S. Goettel and D. L. Johnson (ed.), Memoirs of the Entomological Society of Canada, Vol. 171. Entomological Society of Canada, Ottawa, Ontario, Canada.
181. **Stuart, D. T., C. Upton, M. A. Higman, E. G. Niles, and G. McFadden.** 1993. A poxvirus-encoded uracil DNA glycosylase is essential for virus viability. J. Virol. **67:**2503–2512.
182. **Sueoka, N.** 1961. Correlation between base composition of deoxyribonucleic acid and amino acid composition of protein. Biochemistry **47:**1141–1149.
183. **Sweasy, J. B., and L. A. Loeb.** 1992. Mammalian DNA polymerase β can substitute for DNA polymerase I during DNA replication in *Escherichia coli*. J. Biol. Chem. **267:**1407–1410.
184. **Taddei, F., H. Hayakawa, M.-F. Bouton, A.-M. Cirinesi, I. Matic, M. Sekiguchi, and M. Radman.** 1997. Counteraction by MutT protein of transcriptional errors caused by oxidative damage. Science **278:**128–130.
185. **Teng, S.-C., B. Kim, and A. Gabriel.** 1996. Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks. Nature **383:**641–644.
186. **Teodoro, J. G., and P. E. Branton.** 1997. Regulation of apoptosis by viral gene products. J. Virol. **71:**1739–1746.
187. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
188. **Tietz, A., and H. Weintraub.** 1978. Hydrolysis of glycerides by lipases of the fat body of the locust, *Locusta migratoria*. Insect Biochem. **8:**11–16.
189. **Travis, S. M., H. A. Berger, and M. J. Welsh.** 1997. Protein phosphatase 2C dephosphorylates and inactivates cystic fibrosis transmembrane conductance regulator. Proc. Natl. Acad. Sci. USA **94:**11055–11060.
190. **Upton, C., and G. McFadden.** 1986. DNA sequence homology between the terminal inverted repeats of Shope fibroma virus and an endogenous cellular plasmid species. Mol. Cell. Biol. **6:**265–276.
191. **Upton, C., D. T. Stuart, and G. McFadden.** 1993. Identification of a poxvirus gene encoding a uracil DNA glycosylase. Proc. Natl. Acad. Sci. USA **90:**4518–4522.
192. **VanSlyke, J. K., C. A. Franke, and D. E. Hruby.** 1991. Proteolytic maturation of vaccinia virus core proteins: identification of a conserved motif at the N termini of the 4b and 25K virion proteins. J. Gen. Virol. **72:**411–416.
193. **VanSlyke, J. K., S. S. Whitheead, E. M. Wilson, and D. E. Hruby.** 1991. The multistep proteolytic maturation pathway utilized by vaccinia virus P4a protein: a degenerate conserved cleavage motif within core proteins. Virology **183:**467–478.
194. **van Strien, E. A., B. J. H. Jansen, R. M. W. Mans, D. Zuidema, and J. M. Vlak.** 1996. Sequence and transcriptional analysis of the ubiquitin gene cluster in the genome of *Spodoptera exigua* nucleopolyhedrovirus. J. Gen. Virol. **77:**2311–2319.
195. **von Heijne, G.** 1986. A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. **14:**4683–4690.
196. **Wall, E. M., J.-X. Cao, N. Chen, R. M. L. Buller, and C. Upton.** 1997. A novel poxvirus gene and its human homolog are similar to an *E. coli* lysophospholipase. Virus Res. **52:**157–167.
197. **Weinbauer, M. G., S. W. Wilhelm, C. A. Suttle, and D. R. Garza.** 1997. Photoreactivation compensates for UV damage and restores infectivity to natural marine virus communities. Appl. Environ. Microbiol. **63:**2200–2205.
198. **Weir, J. P., and B. Moss.** 1984. Regulation of expression and nucleotide sequence of a late vaccinia virus gene. J. Virol. **51:**662–669.

199. **Weissman, A. M.** 1997. Regulating protein degradation by ubiquitination. Immunol. Today **18:**189–198.
200. **Whitheead, S. S., and D. E. Hruby.** 1994. Differential utilization of a conserved motif for the proteolytic maturation of vaccinia virus proteins. Virology **200:**154–161.
201. **Whitehead, S. S., and D. E. Hruby.** 1994. A transcriptionally controlled *trans*-processing assay: putative identification of a vaccinia virus-encoded proteinase which cleaves precursor protein P25K. J. Virol. **68:**7603–7608.
202. **Wilcock, D., and G. L. Smith.** 1994. Vaccinia virus core protein VP8 is required for virus infectivity, but not for core protein processing or for INV and EEV formation. Virology **202:**294–304.
203. **Wilson, S. H.** 1990. Gene regulation and structure-function studies of mammalian DNA polymerase β, p. 199–234. *In* P. Strauss and S. H. Wilson (ed.), The eukaryotic nucleus, vol. 1. Telford Press, Caldwell, N.J.
204. **Wittek, R., A. Menna, H. K. Müller, D. Schümperli, P. G. Boseley, and R. Wyler.** 1978. Inverted terminal repeats in rabbit poxvirus and vaccinia virus DNA. J. Virol. **28:**171–181.
205. **Wittschieben, J., and S. Shuman.** 1994. Mutational analysis of vaccinia DNA topoisomerase defines amino acid residues essential for covalent catalysis. J. Biol. Chem. **269:**29978–29983.
206. **Wolffe, E. J., D. M. Moore, P. J. Peters, and B. Moss.** 1996. Vaccinia virus A17L open reading frame encodes an essential component of nascent viral membranes that is required to initiate morphogenesis. J. Virol. **70:**2797–2808.
207. **Wong, S. W., A. F. Wahl, P.-M. Yuan, N. Arai, B. E. Pearson, K. Arai, D. Korn, M. W. Hunkapiller, and T.-S. Wang.** 1988. Human DNA polymerase α gene expression is cell proliferation dependent and its primary structure is similar to both prokaryotic and eukaryotic replicative DNA polymerases. EMBO J. **7:**37–47.
208. **Woods, S. A., D. A. Streett, and J. E. Henry.** 1992. Temporal patterns of mortality from an entomopoxvirus and strategies for control of the migratory grasshopper (*Melanoplus sanguinipes* F.). J. Invertebr. Pathol. **60:**33–39.
209. **Wright, D. A., N. Ke, J. Smalle, B. M. Hauge, H. M. Goodman, and D. F. Voytas.** 1996. Multiple non-LTR retrotransposons in the genome of *Arabidopsis thaliana*. Genetics **142:**569–578.
210. **Xiong, Y., and T. H. Eickbush.** 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. **9:**3353–3362.
211. **Xu, J., and T. Hukuhara.** 1994. Biochemical properties of an enhancing factor of an entomopoxvirus. J. Invertebr. Pathol. **63:**14–18.
212. **Yanez, R. J., J. M. Rodriguez, M. L. Nogal, L. Yuste, C. Enriquez, J. F. Rodriguez, and E. Vinuela.** 1995. Analysis of the complete nucleotide sequence of African swine fever virus. Virology **208:**249–278.
213. **Yanez, R. J., and E. Vinuela.** 1993. African swine fever virus encodes a DNA ligase. Virology **193:**531–536.
214. **Yasui, A., A. P. M. Eker, S. Yasuhira, H. Yajima, T. Kobayashi, M. Takao, and A. Oikawa.** 1994. A new class of DNA photolyases present in various organisms including aplacental mammals. EMBO J. **13:**6143–6151.
215. **Yuen, L., J. Dionne, B. Arif, and C. Richardson.** 1990. Identification and sequencing of the spheroidin gene of *Choristoneura biennis* entomopoxvirus. Virology **175:**427–433.
216. **Yuen, L., and B. Moss.** 1987. Oligonucleotide sequence signaling transcriptional termination of vaccinia virus early genes. Proc. Natl. Acad. Sci. USA **84:**6417–6421.
217. **Yuen, L., M. Noiseux, and M. Gomes.** 1991. DNA sequence of the nucleoside triphosphate phosphohydrolase I (NPH I) of the *Choristoneura biennis* entomopoxvirus. Virology **182:**403–406.
218. **Zajac, P., D. Spehner, and R. Drillien.** 1995. The vaccinia virus J5L open reading frame encodes a polypeptide expressed late during infection and required for viral multiplication. Virus Res. **37:**163–173.