VIROLOGY 233, 19–42 (1997)
ARTICLE NO. VY978607

The Genome of Molluscum Contagiosum Virus: Analysis and Comparison with Other Poxviruses

Tatiana G. Senkevich,* Eugene V. Koonin,† Joachim J. Bugert,*,‡ Gholamreza Darai,‡ and Bernard Moss*,[1]

*Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, and †National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20892; and ‡Institut für Medizinische Virologie der Universität Heidelberg, Im Neuenheimer Feld 324, 699120 Heidelberg, Federal Republic of Germany

Received February 18, 1997; returned to author for revision March 26, 1997; accepted April 15, 1997

Analysis of the molluscum contagiosum virus (MCV) genome revealed that it encodes approximately 182 proteins, 105 of which have direct counterparts in orthopoxviruses (OPV). The corresponding OPV proteins comprise those known to be essential for replication as well as many that are still uncharacterized, including 2 of less than 60 amino acids that had not been previously noted. The OPV proteins most highly conserved in MCV are involved in transcription; the least conserved include membrane glycoproteins. Twenty of the MCV proteins with OPV counterparts also have cellular homologs and additional MCV proteins have conserved functional motifs. Of the 77 predicted MCV proteins without OPV counterparts, 10 have similarity to other MCV proteins and/or distant similarity to proteins of other poxviruses and 16 have cellular homologs including some predicted to antagonize host defenses. Clustering poxvirus proteins by sequence similarity revealed 3 unique MCV gene families and 8 families that are conserved in MCV and OPV. Two unique families contain putative membrane receptors; the third includes 2 proteins, each containing 2 DED apoptosis signal transduction domains. Additional families with conserved patterns of cysteines and putative redox active centers were identified. Promoters, transcription termination signals, and DNA concatemer resolution sequences are highly conserved in MCV and OPV. Phylogenetic analysis suggested that MCV, OPV, and leporipoxviruses radiated from a common poxvirus ancestor after the divergence of avipoxviruses. Despite the acquisition of unique genes for host interactions and changes in GC content, the physical order and regulation of essential ancestral poxvirus genes have been largely conserved in MCV and OPV.

INTRODUCTION

Poxviruses are large, double-stranded DNA viruses that infect vertebrates and insects (Murphy *et al.,* 1995; Moss, 1996). The best known members of the family belong to the orthopoxvirus (OPV) genus, namely variola virus (VAR), the causative agent of smallpox, and the closely related vaccinia virus (VAC), which so successfully served as the smallpox vaccine. The complete genome sequences of VAC (192 kbp) and VAR (186 kbp) have been determined (Goebel *et al.,* 1990; Massung *et al.,* 1993; 1994; Shchelkunov *et al.,* 1993). The genome of only one other poxvirus, molluscum contagiosum virus (MCV), has been sequenced (Senkevich *et al.,* 1996).

VAR and MCV are unique among the poxviruses in their specific adaptation to humans and, with the eradication of smallpox, MCV is the only member of the family that commonly causes human disease. MCV produces small, benign, skin tumors primarily in children and young adults that persist for many months with only a weak immune response and almost no inflammation (Postlethwaite, 1970; Epstein, 1992; Porter *et al.,* 1992; Gottlieb and Myskowski, 1994). In immunodeficient individuals, however, the skin lesions can become extensive, and MCV is a common and untreatable opportunistic infection of AIDS patients (Cotton *et al.,* 1987; Schwartz and Myskowski, 1992).

MCV has no close relatives and is the only member of the molluscipoxvirus genus (Porter *et al.,* 1992; Murphy *et al.,* 1995). Two or three subtypes have been recognized by restriction endonuclease analyses (Parr *et al.,* 1977; Scholz *et al.,* 1989; Porter and Archard, 1992). Attempts to grow MCV in cell culture have been unsuccessful, though limited replication in human foreskin grafted to mice has recently been reported (Buller *et al.,* 1995; Fife *et al.,* 1996). The lack of an *in vitro* replication system has precluded detailed molecular biological characterization of MCV. The genome of MCV is similar to that of VAC in length and structure but has a much higher GC content (Parr *et al.,* 1977; Porter and Archard, 1987; Darai *et al.,* 1986). Until recently, only small segments of the MCV genome had been sequenced (Porter and Archard, 1987; Bugert *et al.,* 1993; Douglass *et al.,* 1996; Hadasch *et al.,* 1993; Sonntag *et al.,* 1995; Sonntag and Darai, 1996). Now, the complete sequence of the MCV type 1 genome (190 kbp), with the exception of the terminal nucleotides, is available (Senkevich *et al.,* 1996). An initial comparison of the genomes of MCV and OPV indicated that their central regions are conserved, whereas the ends are unique and contain genus-specific host response evasion genes. Here, we provide an in-depth

[1] To whom correspondence and reprint requests should be addressed. Fax: (301) 480-1147. E-mail: bmoss@nih.gov.

0042-6822/97 $25.00

analysis of the MCV genome and its comparison with those of OPV.

MATERIALS AND METHODS

MCV and OPV genome sequences

The sequence of the MCV genome (Senkevich *et al.,* 1996) was deposited with GenBank (Accession No. U60315). VAC and VAR sequences and gene designations are from the Copenhagen (Goebel *et al.,* 1990; Accession No. M35027) and Bangladesh (Massung, 1994; Accession No. L22579) strains, respectively, unless indicated otherwise.

Computer analysis of nucleotide and protein sequences

Database searches for sequence similarities were performed using the programs of the BLAST family (Altschul *et al.,* 1990, 1994). Nucleotide sequences, translated in all six reading frames, were compared to the nonredundant protein sequence database (National Center for Biotechnology Information, NIH) using the BLASTX program. Protein sequences were compared to the protein database using the BLASTP program and to the nucleotide database translated in six frames using the TBLASTN program. TBLASTN comparisons were additionally performed with the dbEST, the database of expressed sequence tags (Boguski *et al.,* 1993). Additionally, protein sequences were compared to the protein databases using a new version of the BLASTP program (BLAST2) that constructs alignments with gaps resulting in a greater search sensitivity (Altschul and Gish, 1996). BLAST searches were combined with the search for sequence signatures contained in the PROSITE library (Bairoch *et al.,* 1996) using the BLA program (Tatusov and Koonin, 1994).

Clustering of proteins based on the BLASTP scores was performed with the single-linkage method, using the CLUS program (Koonin *et al.,* 1996). Initially, a cut-off score of 70 was used for clustering; subsequently, multiple alignments were constructed for the resulting groups of proteins and checked for statistical significance and the presence of conserved motifs, in some cases resulting in removal of sequences from a group. Additional sequences were added to some of the clusters based on the results of BLAST2 searches.

Pairwise alignments of protein sequences were constructed using the ALIGN program (Myers and Miller, 1988). Multiple alignments were made using the MACAW program (Schuler *et al.,* 1991). Screening the sequence databases with motifs extracted from multiple alignments or directly from BLAST outputs was performed using the CAP and MoST programs (Tatusov *et al.,* 1994). Multiple alignments were also used to screen databases with the Hidden Markov Model (HMM) method implemented in the HMMer program (Eddy *et al.,* 1995).

Transmembrane segments in proteins were predicted using the PHDhtm program (Rost *et al.,* 1994). Signal peptides were predicted using the SIGNALP program (Nielsen *et al.,* 1997). Protein sequences were partitioned into putative globular and nonglobular domains using the SEG program (Wootton and Federhen, 1996). These methods were combined and applied to multiple protein sequences in batch mode using the UNIPRED program (Walker and Koonin, 1997).

Protein coding regions were predicted using the GeneMark program that derives nonhomogeneous Markov models for a learning set of coding sequences and ordinary Markov models for noncoding sequences and applies them to gene identification in uncharacterized nucleotide sequences (Borodovsky and McIninch, 1993; Borodovsky *et al.,* 1994). The putative MCV genes that have homologs among OPV genes were used as the learning set for coding regions, and MCV DNA sequences containing no long ORFs were used as the learning set for noncoding regions. Additionally, the MCV DNA was screened with models derived for GC-rich human coding sequences (M. Borodovsky, personal communication).

Conserved motifs in nucleotide sequences were detected using the DBSITE program (Claverie, 1994).

Phylogenetic trees were constructed using the neighbor-joining method (the NEIGHBOR program) and maximal parsimony method (the PROTPARS program) with boostrapping using the SEQBOOT program; all these programs are in the PHYLIP package (Felsenstein, 1996).

RESULTS

Terminal regions

The poxvirus genome consists of a linear double-stranded DNA molecule with covalently linked termini (Baroudy *et al.,* 1982). The two ends of the genome are identical but inverted in sequence (Garon *et al.,* 1978; Wittek *et al.,* 1978). Such inverted terminal repetitions (ITRs) vary in length from 0.7 kbp in the Bangladesh strain of VAR to 12 kbp in the Copenhagen strain of VAC (Goebel *et al.,* 1990; Massung *et al.,* 1994). In the sequenced isolate of MCV (Senkevich *et al.,* 1996), the ITR is 4.7 kbp but, as with other poxviruses, considerable variability between isolates occurs (Bugert and Darai, 1991; T. G. Senkevich, unpublished observations).

The ITRs of poxviruses contain a terminal AT-rich incompletely base-paired hairpin loop of about 100 nucleotides (Baroudy *et al.,* 1982; DeLange *et al.,* 1984), sets of short direct repeats (Wittek and Moss, 1980; Baroudy and Moss, 1982; Pickup *et al.,* 1982), and in some cases expressed genes (Wittek *et al.,* 1980). A covalent connection between the strands of MCV DNA was revealed by electron microscopy of partially denatured molecules (Parr *et al.,* 1977) and by rapid renaturation of restriction endonuclease fragments (Porter and Archard, 1987). This connection is not in the available sequence, however, since an estimated 30 to 50 terminal nucleotides were

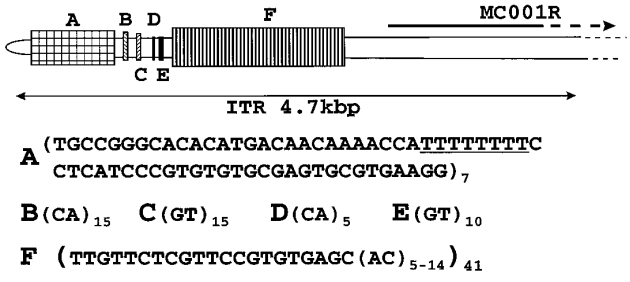

FIG. 1. Organization of unique and repetitive sequence elements in the MCV ITR. The left end of the genome including the predicted hairpin loop is shown. The boxes labeled A, B, C, D, E, and F represent tandem repetitive elements. The repetitive sequences are shown in parentheses with the numbers of repeats indicated by subscripts. Occasional variations were observed in some of the repeats. In element A, the number of T residues in the underlined sequence varies from 5 to 9; the first and two last repeat units in A are incomplete. In element F, approximately 1% of the AC doublets are replaced by GC. MC001R represents an ORF that starts in the left ITR and continues into the adjacent unique sequence.

removed by mung bean nuclease digestion during cloning of the ends of the MCV DNA (Senkevich *et al.,* 1996), T. G. Senkevich, unpublished results). Direct repeats are present within the MCV ITR (Bugert *et al.,* 1993). In Fig. 1, the repeats are represented as a combination of three distinct types of repetitive elements interspersed with unique sequences. A sequence similarity between the F and part of the A repeats was also discerned (not shown). Only the repetitive element closest to the terminal hairpin loop shows some similarity, in terms of size and composition, to the short tandem repeats found in the ITRs of other poxviruses.

The MCV gene complement

The MCV genome contains 64% GC and hence encodes a low frequency of the stop codons UAA, UAG, and UGA. Consequently, MCV DNA contains a greater number of long and overlapping open reading frames (ORFs) than VAC or VAR DNA with their 34% GC content. Indeed, there are 693 ORFs that start with ATG and contain at least 50 codons in the MCV genome, in contrast to 361 and 335 such ORFs in the similar-length VAC (Copenhagen strain) and VAR (Bangladesh strain) genomes, respectively. This made careful delineation of the gene complement of MCV a critical part of the genome analysis. For this purpose, we had employed a hierarchical strategy that resulted in a conservative estimate of 163 predicted genes (Senkevich *et al.,* 1996). Here, we undertook a more detailed analysis of all MCV ORFs and produced a list of 182 genes, 154 of which should be considered most highly likely (Table 1). For consistency, we retained the original gene reference numbers and used decimals to interpolate additional ORFs.

Statistically significant sequence similarity between the product of an ORF and a known protein(s) is the most reliable indication that the ORF in question is a real gene (Borodovsky *et al.,* 1994). Altogether, homologs were detected for 126 ORFs in the MCV genome; 105 have direct counterparts among OPV proteins, 11 have cellular homologs only, and 10 have the highest similarity to other MCV proteins. For 2 members of the last group, similarity to cellular proteins was also directly detectable in database searches, and for 5 others, a relationship with cellular proteins and/or other poxvirus proteins could be established secondarily through similarity to related MCV proteins (Tables 1 and 2).

The MCV coding regions that have homologs in OPV were used to train the GeneMark program for gene recognition (Borodovsky and McIninch, 1993), resulting in a fourth-order nonhomogeneous Markov model. This model recognized all but one of the genes whose products showed similarity to cellular proteins, with the exception being the small gene MC148R, as well as 23 additional putative genes. A nearly identical set of genes was predicted with the model for cellular genes with high GC content. All of these predictions were subsequently supported either by the identification of putative promoters (see below), by the prediction of protein structural features such as signal peptides and/or transmembrane helices, or by both (Table 1).

This two-step gene recognition procedure produced a set of 154 confidently predicted genes, but left several long intergenic regions. All ORFs containing at least 30 codons in these regions were explored in detail in search of possible transcription signals and protein structural features that would be indicative of actual genes. In an attempt to produce a gene set that was as complete as possible, we also included several ORFs that were not predicted by any of these criteria but appeared to be the longest ORFs in regions not containing more confidentally identified genes. We indicated the basis for the selection of individual ORFs, noting those that are most speculative, in Table 1.

Arrangement of ORFs

The ORFs in the MCV genome occur in blocks within which most are oriented in the same direction (Fig. 2). The largest blocks of leftward-oriented ORFs are on the left side of the genome and the largest blocks of rightward-oriented ORFs are in the middle and on the right side, an arrangement also present in OPV genomes (Goebel *et al.,* 1990; Massung *et al.,* 1994). We are unsure whether any functional genes are present within the ITRs of MCV. The ORF of 488 codons (MC001R), beginning proximal to the tandem repeats within the left ITR and extending past the junction into unique sequences, was considered only because of its location and length (Table 1 and Fig. 1). The corresponding ORF (MC164L) in the right ITR is even longer (680 codons) but overlaps with MC163R, a confidently predicted gene. Additional unlisted ORFs are contained entirely within the ITRs but they were not predicted to be *bona fide* coding regions

TABLE 1

The MCV Gene Products

MCV gene[a]	Location, bp, (codon length)	Identification criteria[b]	Promoter type[c]	Homologs (protein or alignment length and % identity)[d]		Predicted structural features and function[g,h]
				poxvirus	cellular	
MC001R ?	3270-4733(488)	L, size	-	MC164R (481)[e] 100%	-	NGL
MC002L	4849-6201(451)	SC, GM	E	MC162R (532) 22.7%	human signaling lymphocytic activating molecule, gi984969, (335) 20.4%	SP, TM-C, immunoglobulin domain
MC003L	6401-7735(445)	SC, GM	E	MC157R (452) 20.1%	human carcinoembryonic antigen homolog CGM2, gi1082275 (265) 16.8%	SP, TM-C, immunoglobulin domain
MC004L	7794-8234(147)	GM	E, L	MC005L (85) 24.4%	-	SP, TM-C, 1 S-S
MC004.1L	8331-8456 (42)	GM	E	-	-	TM
MC005L	8503-8757(85)	SC, GM	E	MC105L (70) 25.0%	-	thioredoxin-like protein?
MC006L	8801-12319(1175)	GM	E	-	-	large NGL-N, Gln-Ala repeats
MC006.1R	12432-12653(74)	GM?	E	-	-	-
MC007L	12614-13264(217)	GM	E	-	-	NGL
MC008L	13307-13834(176)	GM	E	-	-	-
MC009L	13928-14479(184)	GM	E	-	-	SP
MC009.1R?	15105-15320(72)	L	E?	-	-	NGL
MC009.2R?	15377-15673(99)	L	E?	-	-	NGL
MC010R?	15630-15848(73)	L	-	-	-	-
MC011L	15954-16844(297)	GM	E	-	-	SP
MC012L?	17186-17374(63)	L	E	-	-	-
MC013L	17403-18080(226)	SC, GM	L	-	DNJ2_HUMAN, human DnaJ protein homolog (397) 25.3%	molecular chaperone
MC014R?	18222-18662(147)	L	L	-	-	NGL
MC014.1L	18646-18897(84)	GM	E	-	-	-
MC015L	19063-19530(156)	GM	L	-	-	-
MC016L	19647-20285(213)	SC	L	F9L (212) 42.3%	-	TM-C, S-S
MC017L	20269-21597(443)	SC	L	F10L (439) 60.0%	CC21_ORISA(302)[e] 20.2%	Ser/Thr protein kinase
MC017.1L?	21614-21838(75)	L	L?	-	-	NGL
MC018L	21910-23949 (680)	SC	E	F11L (354) 21.3%	-	NGL-N
MC019L	24001-25989(663)	SC	E	F12L (635) 26.5%	-	Involved in EEV maturation/release in FPV (Ogawa et al., 1993)

[a] The predicted genes are numbered consecutively from the left end of the genome. "L" indicates genes transcribed leftward, and "R" indicates genes transcribed rightward. Question marks indicate tentatively identified genes as discussed in the text.

[b] L, location; SC, sequence conservation; GM, GeneMark prediction.

[c] E, early promoter; I, intermediate promoter; L, late promoter.

[d] Only the closest homolog is indicated; the gene names are for VAC strain Copenhagen unless indicated otherwise; VAR strain Bangladesh gene names are indicated when the corresponding gene is missing in VAC; the homologs are identified by their names in the SWISS-PROT database (names with underlines) or by the gene identification (gi) number in the nonredundant database; the alignments included the entire length of the shorter protein unless indicated otherwise.

[e] Only fragments of both proteins could be aligned; the alignment length is indicated.

[f] The EST encodes a partial protein sequence; the coding region contains frameshift errors, and the protein sequence was tentatively reconstructed so as to obtain an optimal alignment.

[g] SP, signal peptide; TM, TM-N, TM-I, TM-C, transmembrane helix, N-terminal, internal, and C-terminal, respectively; S–S, disulfide bond; NGL, predicted nonglobular protein; NGL-N, NGL-I, NGL-C, large N-terminal, internal, or C-terminal nonglobular domains, respectively.

[h] Functions of the VAC proteins are indicated according to Johnson *et al.* (1993) and Moss (1996) unless specific references are indicated in parentheses.

by sequence similarity, GeneMark, or transcription signal analyses. Moreover, these ORFs are not conserved in ITRs of different MCV isolates (T. G. Senkevich, unpublished results) and probably arise from the repetitive sequences in these regions (Fig. 1).

We previously noted that the majority of conserved genes are in the central portion of the genome and that, with only two exceptions, both gene order and arrangement are the same in MCV and OPV even though there are some uniques genes interspersed between conserved genes (Senkevich *et al.,* 1996).

Poxvirus homologs of MCV proteins

The 105 MCV genes with OPV homologs are included in Table 1. In all cases, structural features (signal peptide, transmembrane domains, and nonglobular regions) and functional motifs predicted for an MCV protein were also predicted for the VAC homolog. The MCV gene complement includes homologs of all OPV genes known to be essential for replication. One previously mentioned possible exception is the VAC B1R gene encoding a protein kinase (Senkevich *et al.,* 1996); mutations in this gene cause

TABLE 1—*Continued*

MCV gene[a]	Location, bp, (codon length)	Identification criteria[b]	Promoter type[c]	Homologs (protein or alignment length and % identity)[d] poxvirus	Homologs (protein or alignment length and % identity)[d] cellular	Predicted structural features and function[g,h]
MC020L	26216-26632(139)	GM	E	-	-	NGL
MC021L	26665-27888(388)	SC	L	F13L (372) 39.5%	human protein HU-K4, gi1575347 (437) 28.7%	major envelope antigen, phospholipase or phospholipid synthetase activity? (Koonin, 1996, Ponting, 1996)
MC022L	27860-28624(255)	GM	L	-	-	-
MC022.1L?	28628-28807(60)	GM	-	-	-	-
MC023L	28869-29492(208)	GM	E	-	-	-
MC024L	29617-29880(88)	GM	E	-	-	SP
MC024.1R?	29989-30303(105)	L	E	-	-	-
MC025L	30310-30753(148)	SC	-	F15L (147) 45.9%	-	-
MC026L	30753-31001(83)	SC, GM	E	-	hypothetical yeast protein gi1077251 (66)[e] 31%	modified RING finger domain
MC027L	31038-32639(534)	GM	E	-	-	-
MC028L	32642-33445(268)	GM	L	-	-	-
MC029L	33510-34199(230)	SC	E	F16L (231) 27.3%	-	-
MC030R	34262-34537(92)	SC	L	F17R (101) 51.4%	-	core DNA-binding phosphoprotein
MC031L	34654-36063(470)	SC	L	E1L (479) 54.4%	-	poly(A) polymerase catalytic subunit
MC032L	36063-38306(748)	SC	E	E2L (737) 31.9%	-	TM-C? S-S?
MC033L	38362-40098(579)	SC, GM	E	-	Xenopus class I histocompatibility antigen gi630959 (105)[e] 28%	SP, TM-C, immunoglobulin domain
MC034L	40131-41462(444)	SC	E, L	E4L (259) 57.8%	human TFSII gi443277 (187)[e] 21.4%	RNA polymerase subunit , intermediate transcription factor (VITF-I)
MC035R	41533-47799(2089)	SC	E	VAR B22R (1896) 26.7%	-	2 TM
MC036R	47966-51136(1057)	GM	E	-	-	NGL-N
MC037R	51506-53200(565)	SC	L	E6R (567) 56.4%	-	-
MC038R	53670-54497(276)	SC	L	E8R (273) 56.3%	-	2 TM
MC039L	54554-57589(1004)	SC	E, L	E9L (1006) 52.9%	DPOD_CANAL (1018) 19.1%	DNA polymerase
MC040R	57594-57896(101)	SC	L; E?	E10R (95) 57.4%	ERV1_YEAST (117) 19.5%	SP, S-S
MC040.1L?	57949-58269(107)	L	E	-	-	NGL
MC041L	58465-58860(132)	SC	L	E11L (129) 31.3%	-	-
MC042L	58850-61216(789)	SC	E	O1L (666) 25.2%	-	TM-C?, S-S?
MC042.1R?	61197-61433(79)	GM?, L	-	-	-	SP
MC043L	61479-63794(772)	GM	E	-	-	-
MC043.1L	63933-64055(41)	SC, GM	L	O3L (35)	-	TM-N

a temperature-sensitive phenotype when the virus is propagated in some cell lines (Rempel and Traktman, 1992).

The sequence relatedness of predicted MCV proteins and VAC homologs, listed in Table 1, varied from 15 to 79%. By grouping the VAC ORFs according to their similarity to MCV ORFs, an interesting and useful functional correlation was noted (Fig. 3). The proteins involved in transcription are the most highly conserved; of the 11 proteins with highest identity (75 and 65% groups), 9 are RNA polymerase subunits or transcription factors. The eight RNA polymerase subunits of VAC (J6R, A24R, A29R, E4L, J4R, A5R, D7R, and G5.5R) are from 45 to 79% identical to the corresponding MCV ORFs. Similarly, the VAC early and late transcription factor genes (D6R, A7L, H4L, A1L, A2L, and G8R) and their MCV counterparts are from 50 to 76% identical. Apparent exceptions are the recently described VAC factors, H5R and G2R, which are involved in late transcription (Condit *et al.,* 1996; Kovacs and Moss, 1996), but have only 26% identities with their MCV homologs; notably, both of these proteins are predicted to have nonglobular structures (Table 1) and to physically interact with each other (E. P. Black, N. Moussatche, and R. C. Condit, personal communication). The enzymes of VAC and MCV that are involved in mRNA modification (capping, methylation, and polyadenylylation) are also highly conserved with sequence identities that range from 52 to 59%. Similarly, VAC and MCV enzymes involved in replicating or modifying DNA (DNA polymerase, topoisomerase, uracil DNA glycosylase, and D5R encoded nucleoside triphosphatase) have sequence identities of 53 to 56%.

The VAC virion components with no known enzymatic activities or roles in transcription or replication vary from 15 to 62% in identity with the corresponding MCV ORFs, probably reflecting the functional diversity of this group. There are examples of highly conserved core (MC049L/

TABLE 1—*Continued*

MCV gene[a]	Location, bp, (codon length)	Identification criteria[b]	Promoter type[c]	Homologs (protein or alignment length and % identity)[d] poxvirus	cellular	Predicted structural features and function[g,h]
				26.8%		
MC044L	64275-65204(310)	SC	L	I1L (314) 56.5%	-	-
MC045L	65214-65429(72)	SC	L	I2L (73) 37.3%	-	TM-C
MC046L	64436-66299(288)	SC	E	I3L (269) 39.1%	-	DNA-binding phosphoprotein
MC047L	66341-66586(82)	SC	L	I5L 82/79/34.1%	-	SP, TM-C
MC048L	66753-67970(406)	SC	L	I6L (382) 34%	-	-
MC049L	68062-69606(515)	SC	L	I7L (423) 61.9%	-	core protein
MC050R	69615-71666(684)	SC	L	I8R (676) 46.4%	putative yeast helicase, gi1322677 (767) 18.8%	virion NTPase II, RNA helicase
MC051L	71853-72212(120)	SC	E	MC053L (133) 36%	-	SP, S-S
MC052R?	72329-72628(100)	L	E?	-	-	-
MC053L	72575-72973(133)	SC	E	MC051L (120) 36%	-	SP, S-S
MC053.1R?	72957-73166(70)	L	-	-	-	SP
MC053.2R?	73200-73427(76)	L	E?	-	-	TM-C
MC054L	73260-73964(235)	SC	E	MC053L (112) 31%	-	SP, S-S
MC055R?	73911-74558(216)	size, L	-	-	-	-
MC056L	74938-76716(593)	SC	L	G1L (591) 47.6%	-	Zn-dependent protease, virion morphogenesis (Whitehead and Hruby, 1994)
MC057L	76716-77039(108)	SC	L	G3L (111) 31.5%	-	SP
MC058R	77033-77770(246)	SC	L and/or I?	G2R (220) 25.9%	-	NGL?, isatin-b-semicarbazone dependence determinant , late transcription factor (VLTF) (Condit, 1996)
MC059L	77985-78362(126)	SC	L	G4L (124) 30.7%	GLRX_HUMAN (106) 26.4%	glutaredoxin (Gvakharia,1996)
MC060R	78366-79676(437)	SC	E	G5R (434) 32.1%	-	-
MC061R	79781-79969(63)	SC	E, L	G5.5R (63) 65%	DNA-dependent RNA polymerase subunit, Methanococcus jannaschii, gi1590941 (76) 27.0%	RNA polymerase subunit
MC062R	80069-80653(195)	SC	L	G6R (165) 32.5%	-	-
MC063L	80621-80869(83)	SC	E	-	HMGC_HUMAN , non-histone chromatin phosphoprotein HMGI-C	NGL, DNA-binding protein?

VAC I7L; 62%) and membrane (MC069/VAC L1R; 58%) proteins. However, of the five glycoprotein components of the outer membrane of the extracellular enveloped VAC, three are divergent (15 to 22% sequence identity) and two are missing in MCV.

Interestingly, the comparison of MCV ORFs to the nucleotide sequences of OPV genomes revealed two small, nonoverlapping, conserved ORFs that were not noted in the previous analyses of VAC and VAR genomes. These genes, designated A14.5L and O3L, may encode small hydrophobic proteins that are expressed late in infection and could be previously undetected components of the poxvirus virion membrane (Table 1, Fig. 4).

Cellular homologs of MCV proteins

Altogether, 36 MCV proteins have recognizable cellular homologs; conserved functional motifs shared with cellular proteins could be detected in several additional proteins, including the poly(A) polymerase catalytic subunit and putative ATPases (Table 1). Some of these proteins, such as the homologs of a glutathione peroxidase (MC066L), an MHC class 1 protein (MC080R), and a CC chemokine (MC148R), are likely to protect MCV against host responses to infection (Senkevich *et al.,* 1996). Typically, such homologous protein pairs have 20 to 30% identity (Table 1). One exception is the putative selenocysteinyl glutathione peroxidase (MC066L), which is 74% identical to its cellular homolog. The percentages of identical residues in the homologous MCV and cellular proteins are similar for the 20 proteins that are shared by MCV and OPV and for the 16 MCV proteins that have no counterparts in other poxviruses and are likely to be involved in virus–host interactions. Furthermore, proteins that are highly conserved between MCV and VAC are not much more likely than moderately conserved proteins to have cellular homologs. For example, a few of the RNA polymerase subunits but none of the similarly conserved early or late transcription factors have recognizable cellular homologs.

TABLE 1—*Continued*

MCV gene[a]	Location, bp, (codon length)	Identification criteria[b]	Promoter type[c]	Homologs (protein or alignment length and % identity)[d] poxvirus	cellular	Predicted structural features and function[g,h]
					(108) 21.8%	
MC064R	80987-81172(62)	L	E	-	-	prenylated, membrane-associated(?)
MC065L	81529-82734(402)	SC	L	G7L (371) 41.1%	-	core virion protein precursor
MC066L	83029-83469(211)	SC, GM	L	-	GHSC_HUMAN, human glutathione peroxidase (201) 73.8%	glutathione peroxidase, selenocysteine encoded by an UGA codon
MC067R	83825-84604(260)	SC	I	G8R (260) 71.5%	-	late transcription factor (VLTF-1)
MC068R	84640-85665(342)	SC	L	G9R (340) 38.7%	-	TM-C, S-S, myristylated
MC069R	85669-86397(243)	SC	L	L1R (250) 58.0%	-	TM-C, S-S , myristylated virion protein
MC070R	86507-86785(93)	SC	E	L2R (87) 23.7%	-	2 TM
MC071R	86791-87090(100)	GM	-	-	-	-
MC072L	87524-88453(310)	SC	L	L3L (350) 39.2%	-	-
MC073R	88479-89240(254)	SC	L	L4R (251) 42.6%	-	core DNA binding protein, VP8, important for early transcription (Wilcock and Smith, 1996)
MC074R	89486-89923(146)	SC	L	L5R (128) 28.4%	-	TM-I
MC075R	90102-90650(183)	SC	L	J1R (153) 45.2%	-	virion protein (Holzer, Dorner and Falkner, personal communication)
MC076R	90660-91688(343)	SC	E, L	J3R (333) 53.3%	-	poly(A) polymerase regulatory subunit, ribose 2'-O-methyltransferase
MC077R	91591-92151(187)	SC	E, L	J4R (185) 56.1%	-	RNA polymerase subunit
MC078L	92452-92853(134)	SC	L	J5L (133) 46.0%	-	TM-C, 4 S-S
MC079R	93012-96878(1289)	SC	E	J6R (1286) 79.2%	RPB1_DROME (1530) 21.5%	DNA-dependent RNA polymerase largest subunit
MC080R	96986-98170(395)	SC, GM	E	-	rat MHC I protein, gi11521, (353) 24.5%	3 TM-N, 1 TM-C, MHC I heavy chain homolog
MC081R	98241-98573(111)	GM	E, L	-	-	NGL
MC082L	98659-99165(169)	SC	L	H1L (172) 52.3%	PAC1_MOUSE (135)[e] 33.0%	protein tyrosine phosphatase, important for early transcription (Liu et al., 1996)
MC083R	99180-99752(191)	SC	L	H2R (189) 49.7%	-	TM-N
MC084L	99858-100751(298)	SC	L	H3L (324) 26.2%	-	TM-C, virion membrane protein
MC085L	100755-103127(791)	SC	L	H4L (795) 60.6%	-	RNA polymerase-associated protein RAP94
MC086R	103239-103898(220)	SC	E, L	H5R (203) 26.2%	-	NGL, late transcription factor, VLTF-4 (Kovacs, 1996)
MC087R	104017-104985(323)	SC	L	H6R (314)	TOP1_HUMAN (277)[e]	topoisomerase type I

MCV gene families

Leporipoxviruses and OPV possess a number of genes encoding related proteins, at least some of which may have evolved by duplication within the viral genome constituting multigene families (Upton and McFadden, 1986; Smith *et al.,* 1991; Senkevich *et al.,* 1993; E. V. Koonin, unpublished observations). In order to identify MCV gene families, we compared the protein coding sequences of MCV and other poxviruses to a database of all poxvirus protein sequences and performed clustering based on BLASTP scores (Koonin *et al.,* 1996). This procedure resulted in the delineation of 11 MCV gene families, 8 of which have one or more OPV homologs (Table 2). With some exceptions discussed below, the proteins included in each of the families show similarity to each other that is statistically highly significant in the context of the complete nonredundant protein sequence database.

The MCV counterparts for the two previously described OPV families will only be mentioned here. Thus, the four distantly related MCV genes, MC050R, MC095R, MC100R, and MC123R, are homologs of VAC DNA and RNA helicase family members (Koonin and Senkevich, 1992b; Table 2). Similarly, two MCV genes, MC098R and MC099R, with MutT pyrophosphohydrolase motifs are homologs of the previously described VAC genes D9R and D10R, respectively (Koonin, 1993; Table 2).

Unique MCV gene families: Putative membrane proteins and apoptosis inhibitors

Two of the three MCV-specific gene families encode predicted membrane proteins (Table 2). Each of the two proteins comprising one of these families (MC003L and MC157R) contains a putative signal peptide, a transmembrane helix, and an immunoglobulin domain and may

TABLE 1—*Continued*

MCV gene[a]	Location, bp, (codon length)	Identification criteria[b]	Promoter type[c]	Homologs (protein or alignment length and % identity)[d] poxvirus	cellular	Predicted structural features and function[g,h]
				52.9%	27.2%	
MC088R	105039-105467(143)	SC	L and/or I	H7R (129) 34.7%	-	-
MC089L?	105653-105994(114)	L	L	-	-	SP
MC090R	105999-108848(950)	SC	E, L	D1R (844) 51.8%	ABD1_YEAST (294)[e] 25.1%	mRNA capping enzyme large subunit
MC091L	108871-109380(170)	SC	L	D2L (146) 34.9%	-	virion protein
MC092R	109605-110408(268)	SC	L	D3R (237) 26.0%	-	virion protein
MC093R	110429-111106(226)	SC	E	D4R (218) 52.7%	UNG_BACSU (225) 26.7%	uracil-DNA glycosylase
MC094R	111133-113505(791)	SC	E, L	D5R (785) 56.1%	-	putative ATPase, DNA replication
MC095R	113703-115607(635)	SC	L	D6R (637) 74.4%	SECA_ANTSP (539)[e] 20.6%	early transcription factor (VETF), putative helicase
MC096L?	116018-116644(209)	size, L	-	-	-	-
MC097R	117205-117687(161)	SC	E, L	D7R (161) 60.9%	-	RNA polymerase subunit
MC098R	117721-118356(212)	SC	E	D9R (213) 44.7%	MUTT_ECOLI (129) 18.3%	putative NTP pyrophosphohydrolase (Koonin, 1993)
MC099R	118356-119042(229)	SC	L	D10R (248) 39.4%	MUTT_ECOLI (54)[e] 29%	putative NTP pyrophosphohydrolase (Koonin, 1993)
MC100L	119076-120977(634)	SC	L	D11L (631) 61.9%	Bacillus subtilis putative helicase , gi1303889 (557) 20.0%	virion NTPase I, putative helicase, involved in transcription termination (Deng, 1996; S. Shuman, personal communication).
MC101L	121064-121948(295)	SC	E, L and/or I	D12L (287) 58.6%	-	mRNA capping enzyme, small subunit
MC102L	122053-123693(547)	SC	L	D13L (551) 59.2%	-	rifampicin sensitivity factor, virion membrane protein
MC103L	123701-124207(169)	SC	I	A1L (150) 50.0%	-	late transcription factor (VLTF-2)
MC104L	124315-124998(228)	SC	I	A2L (224) 75.9%	-	late transcription factor (VLTF-3)
MC105L	125084-125293(70)	SC	L	VAR A3L (76) 40.8%	-	thioredoxin-like protein?
MC106L	125313-127337(675)	SC	L	A3L (644) 675/644/57.8%	-	major core protein
MC107L	127341-128603(421)	SC	L	A4L (281) 22.9%	-	core protein
MC108R	128642-129136(165)	SC	E, L	A5R (164) 53.6%	-	RNA polymerase subunit
MC109L	129181-130563(461)	SC	L and/or I	A6L (372) 39.4%	-	-
MC110L	130710-132830(707)	SC	L	A7L (710) 59.9%	-	early transcription factor subunit (VETF)
MC111R	133186-134490(435)	SC	E	A8R (288) 48.3%	-	-
MC112L	134525-134908(128)	SC	E, L	A9L (99)	-	SP, TM-C

be a plasma membrane receptor with as yet unknown specificities (Senkevich *et al.,* 1996). One protein in the second family, namely MC002L, has a similar domain organization; the first predicted transmembrane helix of each of the homologous proteins MC161R and MC162R, however, seems too far from the N-terminus to function as a cleavable signal peptide.

The third unique MCV gene family is of particular interest as its two members, MC159L and MC160L, contain a duplicated DED (MORT) protein interaction domain that is involved in FAS and TNFR1 signal transmission pathways leading to apoptosis (Boldin *et al.,* 1996; Muzio *et al.,* 1996). In standard database searches, the similarity between these MCV proteins and DED domains of FADD (also called MORT 1) and MACH 1-β (also called FLICE or caspase 8) was of limited statistical significance (*P* value of about 0.07 for MC159L and MACH 1-β proteins as computed using BLAST2). Nevertheless, MC159L and MC160L contain all the key amino acid residues that are conserved in DED domains (Fig. 5), and an HMM model derived from the alignment of known DED domains detected the two MCV proteins, together with an uncharacterized equine herpesvirus 2 protein, with scores significantly higher than those for any other proteins. When the BLAST2 search was repeated with the newly identified DED domains, related domains were detected in other herpesviruses (bovine herpesvirus 4, herpesvirus saimiri, human herpesvirus 6, human herpesvirus 7, and human herpesvirus 8 (Kaposi's sarcoma-associated virus)) proteins, as well as in four ESTs, thus expanding the known DED family (Fig. 5). The DED domains of the equine herpesvirus-2 E8 and MCV proteins were recognized independently and evidence was obtained that E8 and MC159 bind MACH 1-β and FADD, respectively, and block FAS- or TNFR1-induced apoptosis (Bertin *et al.,* 1997).

TABLE 1—*Continued*

MCV gene[a]	Location, bp, (codon length)	Identification criteria[b]	Promoter type[c]	Homologs (protein or alignment length and % identity)[d] poxvirus	cellular	Predicted structural features and function[g,h]
				49.5%		
MC113L	134912-137578(889)	SC	L	A10L (891) 44.9%	-	major core protein
MC114R	137593-138504(304)	SC	L	A11R (318) 48.9%	-	TM-C?
MC115L	138553-139086(178)	SC	E, L	A12L (192) 42%	-	virion protein
MC116R?	139102-139302(67)	L	-	-	-	SP
MC117L	139366-139569(68)	SC	L	A13L (70) 23.3%	-	SP, virion membrane protein
MC118L	139573-139854(94)	SC	L	A14L (90) 39.4%	-	2TM, virion membrane protein
MC119L	139877-140032(52)	SC	L	A14.5L (51) 45.0%	-	SP, TM-C
MC120L	140039-140326(96)	SC	L	A15L (94) 30.7%	-	-
MC121L	140465-141556(364)	SC	L	A16L (378) 39.3%	-	TM-C, S-S
MC122L	141570-142106(179)	SC	L	A17L (203) 38.2%	-	2-4 TM, virion membrane protein
MC123R	142121-144202(694)	SC	E?,L and/or I	A18R (493) 49.8%	YEJH_ECOLI (403) 17.7%	NGL-C, RNA helicase
MC124L	143954-144187(78)	SC	L	A19L (77) 42.5%	-	Zn finger protein
MC125L	144191-144532(114)	SC	L	A21L (117) 36.4%	-	SP
MC126R	144531-145958(476)	SC	E, L	A20R (426) 29.3%	-	NGL-I domain, DNA polymerase processivity factor (Klemperer, McDonald and Traktman, personal communication)
MC127R	145867-146712(282)	SC	-	A22R (176) 34.3%	-	-
MC128R	146773-147921(383)	SC	E	A23R (382) 52.6%	-	-
MC129R	147988-151482(1165)	SC	E, L	A24R (116) 75.8%	tomato RNA polymerase subunit 2, gi1049068 (1191) 23.2%	DNA-dependent RNA polymerase subunit
MC130L	151626-152978(451)	SC	L	VAR A29L (702) 23.6%	-	A type inclusion body protein homolog
MC131L	153007-154545(513)	SC	L and/or I?	VAR A30L (489) 26.7%	-	A type inclusion body protein homolog
MC132L	154658-155344(229)	SC, GM	E	-	human EST, GenBank N70419[f] (158)[e] 44%	-
MC133L	155452-157089(546)	SC	L	MC131L (513) 25.2%	-	A type inclusion body protein homolog; C-terminal domain corresponds to the 14 kDa fusion protein in VAC and VAR
MC134L	157093-157515(141)	SC	L	A28L (146) 51.7%	-	SP
MC135L	157561-158469(303)	SC	E	A29L (305)	-	RNA polymerase subunit

Conserved gene families of membrane proteins

Two of the eight families of MCV genes that are conserved in OPV encode putative membrane proteins with a generally similar domain organization, namely, the absence of an N-terminal signal peptide, a conserved domain with a distinctive pattern of potential disulfide bonds, and a single, C-terminal transmembrane domain. One of these families, consisting of MC069R and MC016L and their respective VAC homologs L1R and F9L, is conserved through the entire sequence length, with each protein containing six invariant cysteines potentially capable of forming three disulfide bonds, and a single predicted transmembrane helix near the C-terminus (Fig. 6A). These features are also conserved in the swine poxvirus homolog C19L (Fig. 6A). The L1R protein is a major myristylated component of the VAC intracellular mature virion envelope (Franke *et al.,* 1990; Ravanello and Hruby, 1994) and is a target for potent neutralizing monoclonal antibodies (Wolffe *et al.,* 1995). Furthermore, this protein contains disulfide bonds that affect conformation (Wolffe *et al.,* 1995; Ichihashi and Oie, 1996). MC069R shares the structural features of L1R including the N-terminal penultimate glycine, suggesting that it too is a myristylated envelope protein. Although their similarity to L1R is consistent with the VAC F9L protein and its MCV homolog (MC016L) being virion envelope components, they are unlikely to be myristylated because they lack the critical glycine (Towler *et al.,* 1988).

Members of the other family of putative membrane proteins with disulfide-bonded domains are encoded by MC068R, MC078L, and MC121L and their respective VAC homologs G9R, J5L, and A16L. These proteins all share a conserved C-terminal domain with eight invariant cysteines that may form four disulfide bonds and a single transmembrane helix (Fig. 6B). The family relationship is strongly supported by the similar domain organization and particularly the unique pattern of conserved cysteines, even though the similarity between these proteins

TABLE 1—*Continued*

MCV gene[a]	Location, bp, (codon length)	Identification criteria[b]	Promoter type[c]	Homologs (protein or alignment length and % identity)[d] poxvirus	cellular	Predicted structural features and function[g,h]
				44.7%		
MC136L	158441-158641(67)	SC	L	A30L (77) 40.3%	-	-
MC137L	158648-158812(56)	GM	L	-	-	TM-N
MC138R	158822-159172(117)	SC	E	A31R (124) 32.8%	-	-
MC139R?	159606-159956(117)	GM	-	-	-	NGL
MC140L	160039-160785(249)	SC	L	A32L (269) 53.4%	-	putative ATPase
MC141R	160865-161947(361)	GM	E	-	-	-
MC142R	161815-162363(183)	SC	E?	A33R (185) 19.4%	-	2 TM, extracellular enveloped virion glycoprotein in VAC
MC143R	162549-163025(159)	SC	L	A34R (168) 22.4%	MANR_HUMAN (121)[e] 29%	TM-N, extracellular enveloped virion glycoprotein in VAC; conserved lectin domain
MC144R	163178-163876(233)	SC	E	A35R (176) 23.6%	-	-
MC145R	163927-164790(288)	GM	E	-	-	-
MC145.1R?	164872-165108(79)	L	-	-	-	NGL
MC146R	165381-166598(406)	SC	E	A36R (221) 14.7%	-	TM-N, extracellular enveloped virion protein
MC147R?	166662-166955(98)	L	E	-	-	-
MC148R	166992-167303(104)	SC	E	-	SISF_MOUSE (92) 25.0%	SP, 2 S-S; "CC" chemokine
MC149R	167388-168278(297)	SC	E	A37R (263) 21.2%	-	-
MC149.1R?	168274-168525(84)	L	-	-	-	NGL
MC150R?	168753-169082(110)	L	E	-	-	-
MC151L?	168925-169314(130)	L	E?	-	-	-
MC152R	169622-170683(354)	SC	E	A44L (346) 43.4%	3BH1_HUMAN (373) 32.6%	3β-hydroxysteroid dehydrogenase/steroid D5-D4 isomerase
MC152.1R?	171835-172365(177)	L	-	-	-	-
MC153R	172396-173667(424)	GM	E	-	-	-
MC153.1R	173734-173910(59)	GM	E	-	-	TM-C
MC154R	173951-175015(355)	GM	E	-	-	-
MC155R	175136-176074(313)	GM	E	-	-	NGL, predicted prenylated, membrane-associated ?
MC156R ?	175996-176214(73)	L	-	-	-	SP, TM-C
MC157R	176253-177608(452)	SC	E	MC003L (445) 20.1%	-	SP, TM-C; immunoglobulin domain
MC158R	177769-178287(173)	GM	E	-	-	SP, 2 TM
MC159L	178374-179096(241)	SC; GM	E	MC160L (158)[e] 43.0%	MACH-β1, gi1403324 (235) 22%	contains a duplicated MORT domain; putative apoptosis regulator
MC159.1R?	179095-179346(114)	L	E	-	-	SP
MC160L	179150-180262(371)	SC; GM	E	MC159L (158)[e] 43.0%	MACH-β1, gi1403324 (235) 371/235/21.1%	NGL-C, contains a duplicated MORT domain; putative apoptosis regulator
MC161R[d]	180328-181761(478)	SC, GM	E	MC162R (532) 20.4%	-	TM-N, TM-C; immunoglobulin domain
MC162R[d]	182041-183636(532)	SC	E	MC002L (451) 22.7%	-	TM-N, TM-C; immunoglobulin domain
MC163R	183788-185647(620)	SC, GM	E?	-	SODC_IPOBA , sweet potato superoxide dismutase (151) 23.2%	TM-I, superoxide dismutase-related domain
MC164L?	184981-187020 (680)	size, L	-	MC001R (481)[e] 100%		-

was of moderate statistical significance when the complete database was screened (*P* values between 0.001 and 0.1). The MC078L/J5L gene products are small proteins that consist solely of the conserved domain; MC0121L/A16L and MC068R/G9R also contain an upstream domain with five additional conserved cysteines that may form two conserved disulfide bonds, and perhaps an additional disulfide bond with the position of the second cysteine varying in different proteins (Fig. 6B). The G9R protein contains an N-terminal penultimate glycine that is apparently myristylated (D. E. Hruby, cited in Johnson *et al.,* 1993), as does MC068R. The MC121L/A16L proteins also contain N-terminal penultimate glycines whereas the MC078L/J5L proteins do not. All the genes in this family have consensus late promoters (Table 1 and below) and J5L has been shown to be an essential protein that is expressed late in infection (Zajac *et al.,* 1995). Whether the proteins of this family are uncharacterized membrane components of virions or are involved in morphogenesis or other late functions remains to be determined.

Conserved gene families of putative redox active proteins

The existence of novel redox proteins is suggested by the presence of disulfide bonds in certain VAC proteins located within the usually reducing cytoplasmic environment (Rodriguez *et al.,* 1987). Two glutaredoxins encoded by VAC have been described. The first, O2L, is not essential for VAC replication and may be a cofactor for ribonu-

TABLE 2

MCV Gene Families

Genes: MCV/OPV conserved in MCV and OPV	Promoter type	Structural features and function
MC050R/I8R	Late	DNA and RNA helicases and nucleic acid-dependent ATPases; early transcription
MC095R/D6R	Late	
MC100R/D11L	Late	
MC123R/A18R	Early?, int.?, late	
MC098R/D9R	Early	Putative pyrophosphohydrolases
MC099R/D10R	Late	
MC032L/E2L	Early	Unknown
MC042L/O1L	Early	
MC130L, MC131L, MC133L/A29L, A30L, A31L (VAR)	Late	ATI/fusion proteins
MC069L/L1R	Late	Membrane proteins, unique disulfide-bonded domain
MC016L/F9L	Late	
MC068R/G9R	Late	Membrane proteins, unique disulfide-bonded domain
MC078L/J5L	Late	
MC121L/A16L	Late	
MC051L/(D7L/VAR)[a]	Late	Putative membrane (secreted?), glutaredoxin/thioredoxin-like (?) proteins, also distantly related to MC040R/E10R
MC053L/(D7L/VAR)[a]	Early	
MC054L/(D7L/VAR)[a]	Early	
MC004L/none	Early	Putative glutaredoxin/thioredoxin-like proteins; MC005L, predicted membrane or secreted protein
MC005L/none	Early	
MC105L/A3L (VAR)	Early	
Unique to MCV		
MC002L	Early	Predicted membrane proteins
MC161R	Early	
MC162R	Early	
MC003L	Early	Predicted membrane proteins
MC157R	Early	
MC159L	Early	Putative apoptosis regulators; contain duplicated DED domains
MC160L	Early	

[a] Low similarity, not orthologs.

cleotide reductase (Ahn and Moss, 1992; Rajagopal *et al.,* 1995). MCV has neither an O2L nor a ribonucleotide reductase homolog. MC059 is homologous to the second VAC glutaredoxin, G4L, which has an unknown role in virus replication (Gvakharia *et al.,* 1996).

MC040R is homologous to the VAC E10R gene, which has a pair of conserved cysteines separated by two residues, a configuration typical of redox active centers of glutaredoxins and thioredoxins (Holmgren, 1989). MC040R/E10R proteins show significant sequence similarity to African swine fever virus, yeast, and rat proteins as well (Fig. 6C). ERV1, one of the yeast homologs of E10R, is an essential protein required for mitochondrial biogenesis (Lisowsky, 1992; Lisowsky, 1994), and the rat homolog has been identified as a liver regeneration augmenting factor (Hagiya *et al.,* 1994). The characteristic cysteine pair is also present in proteins encoded by another gene family which consists of MC051L, MC053L, and MC054L, as well as proteins of VAR (but the gene is disrupted in the VAC WR strain and absent from the VAC Copenhagen strain), ectromelia, and swinepox (Table 2, Fig. 6C). The similarity among MC051L, MC053L, and MC054L is statistically significant in a complete database search using BLAST2 (P value $<10^{-4}$). By contrast, this protein family shows only a limited similarity to MC040R/E10R proteins and their cellular homologs, primarily in the motif which includes the glutaredoxin-like pair of cysteines (Fig. 6C). The alignment of this motif is very unlikely to be produced by chance (P value $<10^{-15}$) as computed using the MACAW program when the search space is limited to the poxvirus proteins and the cellular homologs of E10R. Therefore, we tentatively combine the E10R group with the MC051L–MC053L–MC054L family in a superfamily of proteins predicted to possess redox activity. The poxvirus proteins in this family contain N-terminal hydrophobic regions, whereas in the African swine fever virus, human, and one of the yeast proteins, these regions are considerably more hydrophilic (Fig. 6C). The hydrophobic regions may be either signal peptides or membrane anchors. The MC040/E10R proteins have late promoters and may have roles as membrane-anchored glutaredoxin-like proteins that facilitate the formation of disulfide bonds in other viral proteins. The members of the MC051L–

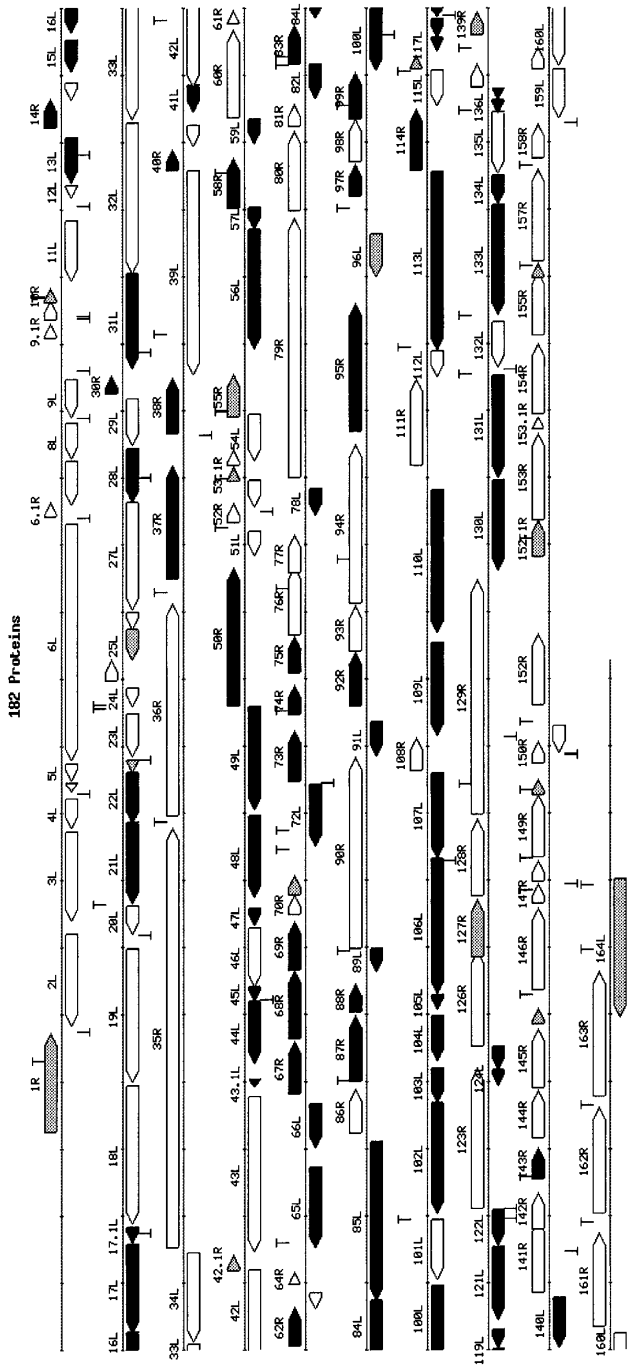

FIG. 2. Predicted MCV ORFs. ORFs are numbered as in Table 1, with rightward ORFs above the line and leftward ones below. Coding: white, predicted early genes; black, predicted intermediate or late genes; shaded, genes with no identifiable promoter sequence; T indicates consensus termination sequence (TTTTTNT) for early genes transcribed rightward, and overturned T for early genes transcribed leftward.

MC053L–MC054L family are predicted to be expressed early in infection and therefore are likely to have functions other than virion morphogenesis.

MCV encodes yet another putative family that consists of three proteins (of which only MC105L has a homolog in VAR and VAC WR strain) with a conserved pair of cysteines resembling a glutaredoxin/thioredoxin active site (Table 2, Fig. 6D). A four-way alignment (P value $<10^{-5}$) reveals a pattern of conserved residues, which, together with the similar size of these proteins and the presence of a putative redox cysteine pair, suggests a genuine relationship, although the similarity among these three MCV proteins is not statistically significant in a complete database search. One of the MCV proteins contains an N-terminal hydrophobic segment that may function as a signal peptide or a transmembrane domain, whereas the remaining proteins in this family lack long hydrophobic regions and are likely to be cytoplasmic. Like most of the MCV proteins with no counterparts in OPV, these MCV proteins are predicted to be expressed early in infection.

It is notable that MCV encodes eight proteins with putative redox sites compared to only three such proteins in VAC Copenhagen (G4R, O2L, and E10R) and five in VAR and VAC WR. Furthemore MC066L is a putative selenocysteinyl glutathione peroxidase. The abundance of potential redox proteins may indicate that inactivation of free radicals is critical for the replication of MCV.

Conserved family of inclusion proteins

Another conserved protein family consists of homologs of viral A type inclusion (ATI) proteins (Ichihashi *et al.,* 1971; Shida *et al.,* 1977; Patel *et al.,* 1986; Osterrieder *et al.,* 1994). The cytoplasmic intracellular mature virions of some poxviruses are embedded in the ATI matrix, presumably to protect them from the environment after cell lysis and release. The organization of genes for ATI protein homologs is highly variable in different poxviruses, indicative of a complex pattern of gene deletions and duplications (Table 2, Fig. 7A). MCV possesses three genes (MC130L, MC131L, and MC133L) coding for ATI-like proteins with conserved sequences indicated by the pairs of stippled and white boxes in Fig. 7A. We noted that two of the three ATI genes of MCV, namely MC131L and MC133L, as well as ATI genes of OPV, contain C-terminal domains (indicated by the black bars) that are homologous to the small fusion proteins of OPV, called p14 for VAC strain WR (Fig. 7A). The similarity between the C-terminal portion of MC133L and the fusion proteins was highly significant when the complete database was searched ($P < 10^{-8}$), whereas the corresponding values for MC131L were moderate but also significant ($P = 0.002$). MCV contains only this conserved domain associated with ATI proteins and does not encode a corresponding small fusion protein. Interestingly, this domain contains a conserved motif with a unique pattern of invariant amino acid residues (Fig. 7B), suggesting a specific function distinct from virus–cell membrane fusion that is mediated by the N-terminal portion of the VAC fusion protein (Gong *et al.,* 1990).

Promoter consensus sequences

AT-rich sequences preceding the MCV ORFs resemble the promoters of OPV and stand out against the GC-rich

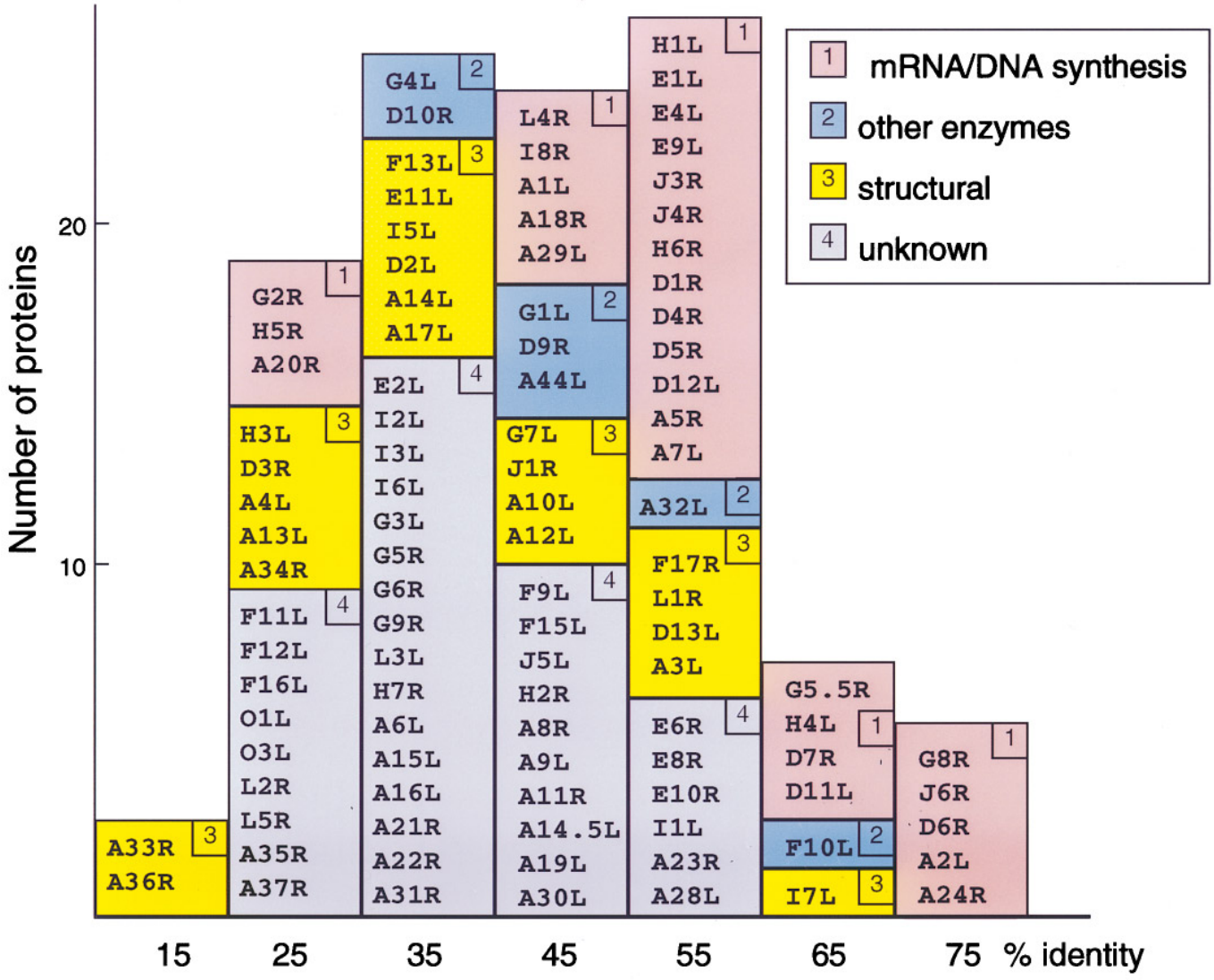

FIG. 3. Sequence conservation between MCV proteins and their homologs in VAC. Ordinate, number of homologous proteins; abscissa, percentage of amino acid identity of MCV and VAC (Copenhagen) proteins. Color coding of protein functional categories: pink, transcription, DNA replication, nucleic acid modification; blue, enzymes not predicted to be directly involved in synthesis of RNA or DNA; yellow, virion proteins without known enzymatic roles; lavender, unknown functions. D9R, D10R, and A32L are listed under enzymes (blue) but their activities have only been predicted. F13L (yellow) is a virion membrane component that is predicted to have enzymatic activity (Koonin, 1996; Ponting and Kerr, 1996).

background of the remainder of the MCV genome, making the prediction of promoters more straightforward than it is for the AT-rich VAC genome. Blake *et al.* (1991) had noted that the sequence upstream of one MCV ORF resembled an OPV late promoter and that another sequence resembled an early promoter. Our comparison of the 5′-untranslated regions of 29 VAC early genes with the corresponding regions of the homologous MCV genes revealed that the 15-bp core promoter sequences (Davison and Moss, 1989a) were located at nearly the same distance from the start codon, allowing a confident prediction of promoters for both viruses. For many of the VAC genes in this set, the transcription start sites had been identified experimentally, supporting the promoter position. The 29 putative MCV and VAC early promoters had nearly identical nucleotide frequencies in most positions (Fig. 8). This correspondence is remarkable, in view of the much higher overall GC content of the MCV genome compared to that of VAC. From an alignment of the 29 conserved MCV promoters, a position-dependent weight matrix (Claverie, 1994) was generated; this was used for an iterative screening of the 5′-untranslated sequences of the remaining MCV ORFs. Combined with visual inspection, this resulted in the prediction of early promoters for 90 MCV genes (Fig. 2 and Table 1).

The TAAAT transcriptional initiator sequence is the most prominent element of VAC late promoters (Rosel *et al.,* 1986), although it is also found associated with a minority of early and intermediate genes (Ahn *et al.,* 1990; Keck *et al.,* 1990). TAAAT is most frequently followed by G to form the first codon of the ORF, although in some cases the TAAAT is followed by an A and an ATG is located several nucleotides downstream. These sequences are present in most of the MCV genes that are homologs of VAC late genes as well as some additional genes (Table 1 and Fig. 2). One or more

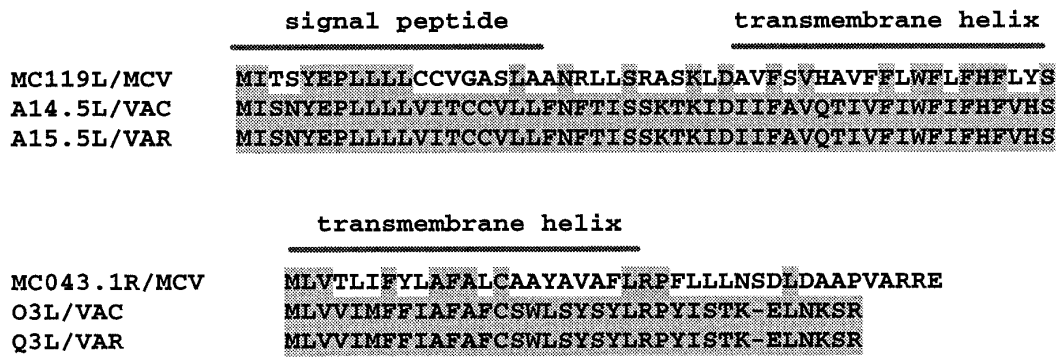

FIG. 4. Two previously unrecognized ORFs encoding small putative transmembrane proteins conserved in MCV and OPV. Alignment of predicted MC119L with VAC and VAR homologs A14.5L and A15.5L, respectively. Alignment of predicted MCV 043.1R with newly named VAC and VAR homologs O3L and Q3L, respectively. Conserved amino acids are shaded and predicted structural features indicated.

T tracts located between −8 and −24 relative to the first A of TAAAT are associated with strong late promoters (Davison and Moss, 1989b). Such sequences were found upstream of approximately one-half of the putative late genes of MCV. In most cases, both the length and the position of the T tract are well conserved in homologous MCV and VAC genes as shown by the examples in Fig. 9A, suggesting that the relative strengths of the promoters are similar.

Some MCV genes contain both early and late promoter consensus sequences (Table 1) and this combination of transcription signals is conserved in most of their OPV homologs.

An intermediate promoter sequence was defined by mutagenesis and an alignment of the regions just upstream of the three VAC genes A1L, A2L, and G8R encoding late transcription factors (Baldick *et al.*, 1992). Important features

Consensus		UU U U -U U+UU		U U UU UUU U RUDUU U	
MCV					
MC159Ln	11	FLRHLLEELDSHEDSLLLFLC	13	LCSLSQQRKLTLAALVEMLYVLQRMDLLKSR	
MC159Lc	98	LMVCVGEELDSSELRALRLFA	23	LENVGLVSPSSVSVLADMLRTLRRLDLCQQL	
MC160Ln	9	FLRNLLAELDASEHEVLRFLC	13	LRALQRRRLLTLSSMAELLCALRRFDVLKVR	
MC160Lc	96	QVAAINNMVGSEDLRVMCLCA	19	LEDAGAISPQDVSVLVTLLHAVCRYDLSVAL	
Cellular apoptosis signal transmitters					
MACH-beta-1n	5	NLYDIGEQLDSEDLASLKFLS	21	LQEKRMLEESNLSFLKELLFRINRLDLLITY	
MACH-beta-1c	104	MLYQISEEVSRSELRSFKFLL	8	LDDDMNLLDIFIEMEKRVILGEGKLDILKRV	
FADD	6	LLHSVSSSLSSSELTELKFLC	21	LLEQNDLEPGHTELLRELLASLRRHDLLRRV	
MCH4n	22	KLLIIDSNLGVQDVENLKFLC	21	LLAEDLLSEEDPFFLAELLYIIRQKKLLQHL	
MCH4c	119	LLYELSEGIDSENLKDMIFLL	17	LEKQGKIDEDNLTCLEDLCKTVVPK-LLRNI	
PEA-15	6	LFQDLTNNITLEDLEQLKSAC	21	LESHNKLDKDNLSIIEHIFEISRRPDLLTMV	
Herpesviruses					
E8-EHV2n	4	SMIDTYFSLDEDETETYLYLC	16	FKFLSDYACLSAANQMELLFRVGRLDLIRRI	
E8-EHV2c	95	LMALVNDFLSDKEVEEMYFLC	22	LEDLELLGGDKLTFLRHLLTTIGRADLVKNL	
BHV4n	5	VLLAIETHLNQNEKTFVMYFL	15	LENLHSKRKIRYPILIELMYILQRFDLLRSI	
BHV4c	97	LIFSIGQNIDDEDLISIKFIS	20	LEKVDMVGPDNLDLFETFFKQIHRMDIVKMI	
VG71_HSVSAn	5	TVLHITDSFTDEEMYCLLFLI	15	IIETLSKSTQWDICLMQCLYVLRKIGLLLNL	
VG71_HSVSAc	96	TLVNVNNNLTAKDEKRLCFIL	19	NMLCEMHVLECLCQLKKCLKQIGRSDLAKTV	
U15-HHV6	19	VLMSLSNMFSKIEIVYVKYLF	11	LPALTLSMTVTKSLVIEMLFILKRWEDIDQF	
U15-HHV7	19	ILMTLSNIISKVETIYIKYLF	11	FSGLTLTTTVTKSVVIEALFIIKRWQEIKQI	
K13-HHV8	5	VLCEVARKLGTDDREVVLFLL	15	LRALKEEGRLTFPLLAECLFRAGRRDLLRDL	
Human ESTs					
W23795		VIHQVEEALDTDEKEMLLFLC	15	LDILRERGKLSVGDLAELLYRVRRFDLLKRI	
N94588		----------VELEKLNLVAPDQLDLLEKC			
W52946		----------VELELKLNLVAPDQLDLLEKC			
R62438		----------LRLLDLGMTLVRVSLLQGLFLVQRLDL----			

FIG. 5. Alignment of putative apoptosis inhibitors encoded by MCV with homologous cellular and viral proteins. For proteins with duplicated DED (MORT) domains, the two copies are designated n (N-terminal) and c (C-terminal). The numbers on the left and center indicate the number of amino acids from the N-termini of the aligned proteins and between the conserved motifs, respectively. The consensus shows amino acid residues that are conserved in more than 50% of the aligned sequences. U indicates a bulky hydrophobic residue (I, L, V, M, F, Y, W). The minus sign in the consensus sequence indicates a negatively charged residue (D or E) and the plus sign indicates a positively charged residue (K or R). The residues that conform to the consensus are shown in boldface type. The most prominent conserved motif, located toward the C-terminus of the DED domain, is highlighted by inverse typing. The sequences of cellular (human) and herpesvirus proteins involved or implicated in apoptosis were from the NCBI nonredundant protein sequence database: MACH-beta-1, gi1403325; FADD, gi791038; MCH4, gi1498324; PEA-15, gi854322; E8-EHV2 (equine herpesvirus 2), gi1360817; U15-HHV6 (human herpesvirus 6), gi451938; U15-HHV7 (human herpesvirus 7), gi451938; K13-HHV8 (human herpesvirus 8, Kaposi's sarcoma associated herpesvirus), gi1718327; the BHV4 (bovine herpesvirus 4) sequence is from a previously uncharacterized 182 codon ORF encoded by nucleotides 2846 to 3381 in the nonconserved region E sequence (GenBank Z46385); the EST sequences are identified by their accession numbers in the dbEST database.

of this sequence include a TAAA transcription initiator site and an AAANAA core sequence 11 to 13 nucleotides upstream of the TAAA. This consensus is conserved in the MCV homologs of these genes as shown by the alignment in Fig. 9B. Several other VAC gene promoters that fit the intermediate consensus and lack T tracts typical of strong late promoters are also shown in Fig. 9B, even though their regulation has not been determined precisely. Generally, the 11- to 13-bp sequence between the core and the initiator of the MCV intermediate promoters are more GC-rich than their VAC homologs (Figs. 9A and 9B), consistent with the suggested spacer role of this region. A striking exception is the almost perfect identity of the AT-rich sequences comprising the entire VAC A18R promoter and its MCV homolog, raising the possibility of an additional regulatory role of the spacer in this case. One gene, D12L, encoding the large subunit of capping enzyme, with a predicted intermediate promoter, also contains an early promoter that is conserved in the MCV homolog.

Transcriptional termination sequences

Transcription stops 20 to 50 bp downstream of the TTTTTNT termination sequence of many VAC early genes (Yuen and Moss, 1987). This sequence was detected at the end of 46 predicted early genes of MCV (Fig. 2). In a number of cases, the terminator is far downstream of the end of the coding region, sometimes inside a late gene; furthermore, on 24 occasions, a single terminator occurs following 2 or more early genes, suggesting the formation of a 3′-coterminal nested set of mRNAs (Fig. 2). Examples of internal terminator sequences, thought to be nonfunctional because of secondary RNA structure, have been noted in a few VAC genes (Lee-Chen *et al.,* 1988; Luo and Shuman, 1991). Only two predicted early genes (MC094R and MC129R) contained a terminator signal within the coding region and in one case the signal was also present in the VAC homolog; in each of these genes, both early and late promoters were detected.

Concatemer resolution signal

The putative MCV DNA concatemer resolution signal (Fig. 10) was identified by its location near the end of the genome and by the presence of sequence elements similar to those of other poxviruses (DeLange and McFadden, 1987; Merchlinsky and Moss, 1989). These elements, TAAAT separated by seven to nine nucleotides from six to seven consecutive Ts (Merchlinsky and Moss, 1989; Merchlinsky, 1990), can function as a late promoter (Hu and Pickup, 1991; Stuart *et al.,* 1991). Although MCV has only five consecutive Ts, the late promoter function might be enhanced by the nearby run of four Ts (Davison and Moss, 1989b).

Phylogeny of MCV

Phylogenetic trees for 5 poxvirus genes were constructed using 2 methods based on different principles, namely neighbor-joining and protein parsimony. The genes chosen encode protein squences that are sufficiently conserved so as to allow an unambiguous alignment and were available for several vertebrate poxviruses as well as for homologous sequences in an entomopoxvirus or a bacterium as an outgroup. The genes satisfying these criteria and included in phylogenetic analysis were the DNA polymerase (MC039L), the uracil DNA glycosylase (MC093R), the early transcription factor subunit (MC095R), the virion NTPase I (MC100L), and the rifampicin-sensitivity factor (M102L) genes. Thus, 10 trees were constructed (2 methods X 5 genes), 3 of which are shown in Fig. 11. Depending on the gene and the method of tree construction, 2 alternative topologies were observed (compare Figs. 11A and 11B with 11C). Of the 10 trees, the first topology with MCV radiating from a common trunk with OPV (VAC and VAR) and parapoxvirus (ORF) or leporipoxvirus (SFV), after the divergence of avipoxvirus (FPV), was found 7 times. More importantly, in 3 cases this topology was strongly supported (>80%) by bootstrap analysis, as opposed to only 1 case of relatively strong support for the alternative topology which groups MCV with avipoxvirus. Thus, the scenario of poxvirus evolution depicted in Figs. 11A and 11B is the most likely one. This is compatible with the fact that OPV and leporipoxviruses share a number of conserved genes not found in MCV, particularly those that encode proteins involved in the interaction with the host immune system (McFadden and Graham, 1994; Alcami and Smith, 1995). Another corollary of this evolutionary scenario is that any genes that are conserved in avipoxviruses and in OPV, parapoxviruses, or leporipoxviruses but not in MCV may be ancestral genes that were eliminated in the lineage leading to MCV. At present, this is the situation for 3 genes, namely DNA ligase, C10L, and thymidine kinase; at least for thymidine kinase independent acquisition by avipoxvirus and OPV is possible (Koonin and Senkevich, 1992a). As more avipoxvirus sequences become available, additional ancestral genes may be revealed.

DISCUSSION

The essential poxvirus genes

The present report contains the first detailed comparison of the complete genomes of distantly related vertebrate poxviruses. Of the 182 predicted MCV genes, 105 have direct counterparts in OPV (Senkevich *et al.,* 1996). This shared set includes: (1) all 51 of the OPV genes considered to be essential for replication in cell culture, based on currently available data; (2) 15 OPV genes that are not absolutely essential, even though some enhance virus spread; and (3) 39 that are uncharacterized with regard to function. From this analysis, we suggest that between 51 and 90 genes comprise the minimal complement for replication of vertebrate poxviruses. The completion of the sequence of avipoxviruses and leporipoxviruses should allow a refinement of this number.

The physical arrangement of the common genes is virtually identical in MCV and OPV (Senkevich *et al.,*

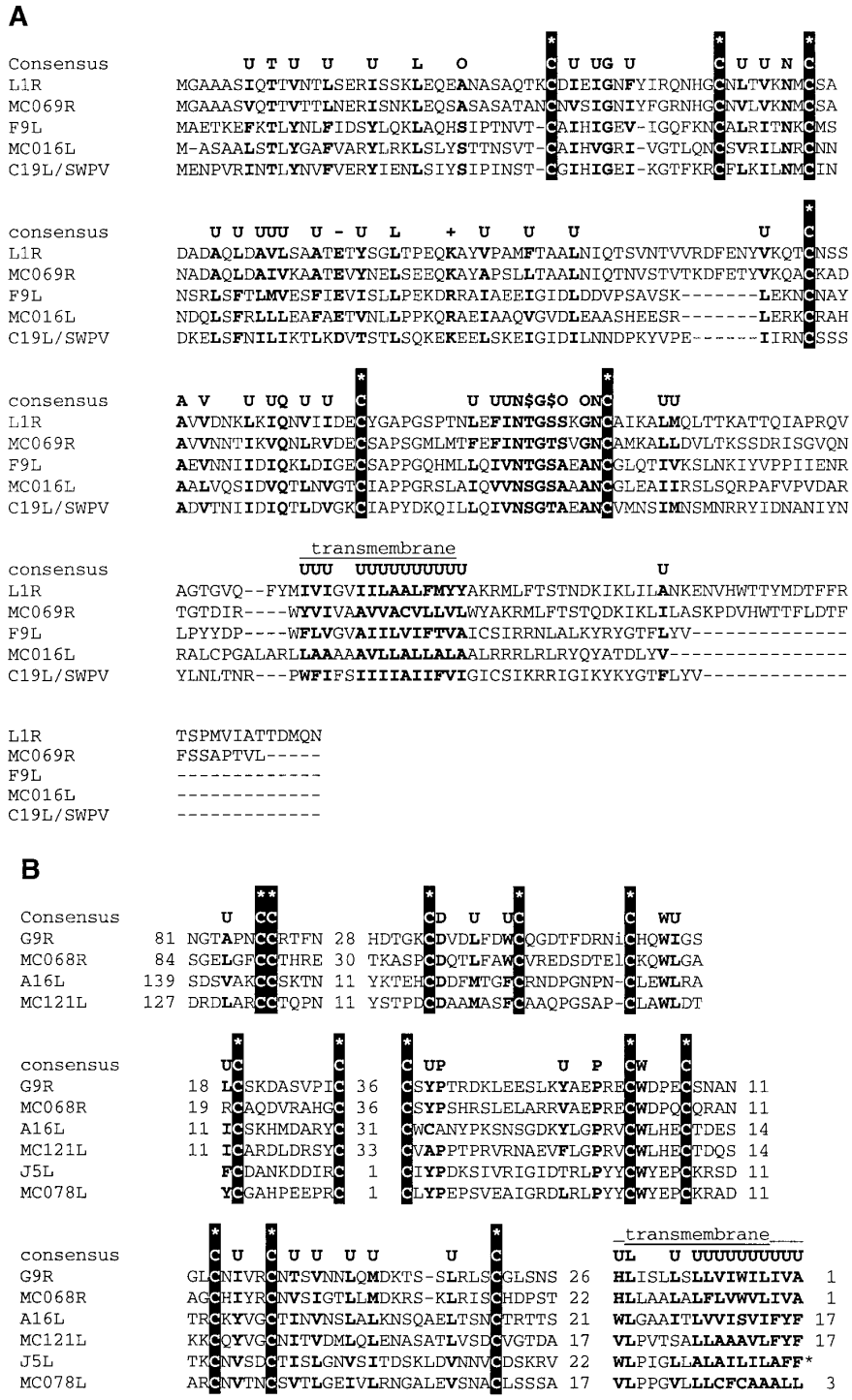

FIG. 6. Conserved poxvirus protein families with unique patterns of predicted disulfide bonds. (A) The major VAC myristylated protein L1R and VAC F9L are aligned with their MCV counterparts MC069R and MC016L, respectively. Also aligned is the swine poxvirus (SWPV) C19L ORF. Conserved cysteines are marked by asterisks and highlighted by reverse typing. Other consensus amino acid residues conserved in all of the aligned sequences are indicated in boldface; the other designations are as in Fig. 5. (B) Parts of the VAC myristylated protein G9R are aligned with its MCV counterpart MC068R and with the homologous VAC/MCV pairs A16L/ MC121L and J5L/ MC078L. The numbers on the left indicate the number of amino acids from the N-terminus; other numbers indicate the number of amino acids between domains. (C) The group of viral and cellular proteins homologous to VAC E10R protein and the MC051L-MC053l-MC054L family were aligned separately and superimposed using the MACAW program. The alignment within each group is highly reliable but the alignment between groups should be considered tentative, with the exception of the motif around the putative redox active cysteines (indicated by the bar). Consensus shows residues conserved in at least 50% of the aligned proteins. The sequences of the African swine fever virus and cellular proteins homologous to E10R were from the nonredundant protein database: pB119L/ASFV, gi780443; ALR/HUMAN (augmenter of liver regeneration), gi644888; ERV1_YEAST, gi119564; YP3085.03c/YEAST, gi1072405; yk51h9.3/CAEEL (*C. elegans*), gi1049356. (D) Alignment of MCV and VAR proteins comprising another family of putative thioredoxin-like proteins.

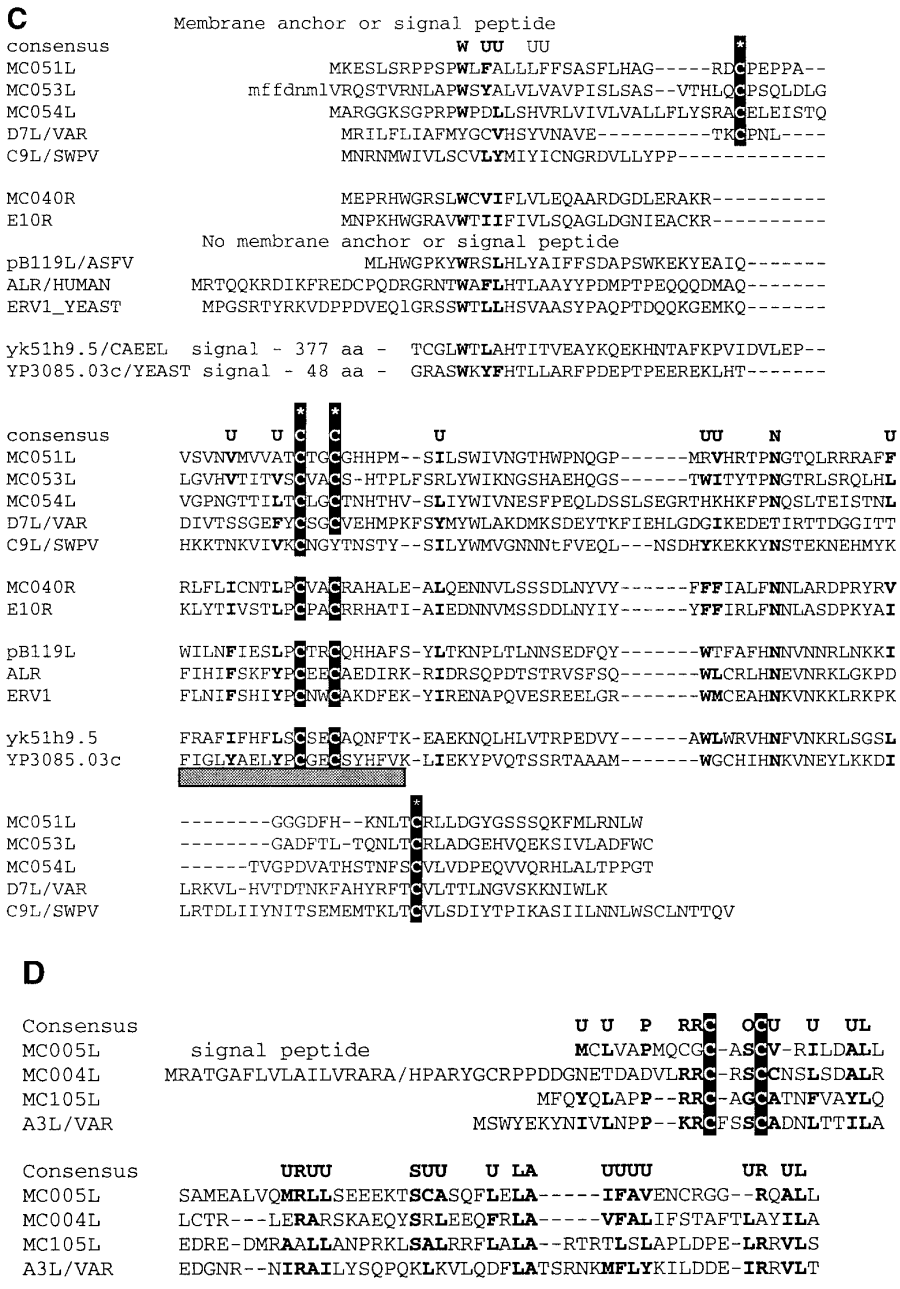

FIG. 6—*Continued*

1996). As the repertoires of genes implicated in virus–host interaction in MCV and OPV appear to be completely different, the evolution of the viral genomes should have involved a number of recombination events. It seems likely, therefore, that the conservation of gene order is at least partly due to selection pressure against shuffling of DNA sequences. Although based only on DNA hybridization or limited sequencing, the order of capripoxvirus and leporipoxvirus genes also appears to be conserved (Gershon *et al.,* 1989; Fleming *et al.,* 1993; Mercer *et al.,* 1995). The situation may be different in avipoxviruses, since there are blocks of fowlpox virus DNA sequence within which genes exist in the same relative position as OPV but the genomic location of those blocks differs widely between the two viruses (Binns *et al.,* 1988; Mockett *et al.,* 1992). Thus, comparisons of gene organization are consistent with our phylogenetic prediction and suggest that the gene order of MCV, OPV, and leporipoxviruses became fixed after the divergence of avipoxviruses.

We found a correspondence between gene function and sequence conservation. The VAC and MCV proteins involved in mRNA biogenesis and DNA replication were more similar to each other than were the poxvirus proteins of other functional classes. Since not all of the genes involved in the latter processes have been identified, the best candidates may be among those highly conserved.

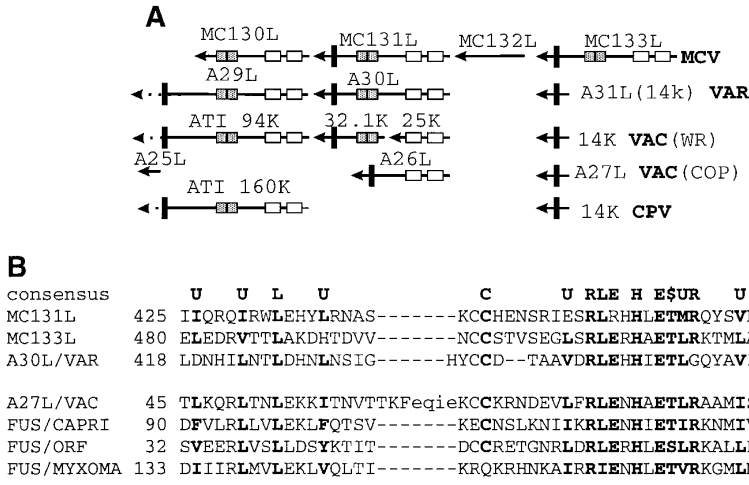

FIG. 7. Conservation and variability among poxvirus A type inclusion genes. (A) Organization of genes encoding ATI-like proteins in MCV and OPV. The MCV genes M130L, MC131L, and MC133L contain domains (stippled and white boxes) that are present in the homologous ATI genes of OPV. For strain WR of VAC and cowpox virus (CPV), the ATI proteins are designated by the size of the ORF in kDa (K) rather than by a formal ORF name. Two of the MCV ATI-like proteins, MC131L and MC133L, contain a third C-terminal domain (black bar) that is present in the ATI proteins of OPV as well as in short fusion proteins: A31L of VAR, 14K of VAC (WR), A27L of VAC (COP), and 14K of CPV. (B) Alignment of the conserved fusion protein motifs. MCV, VAC, and VAR proteins correspond to A. FUS/CAPRI (gi137868), FUS/ORF (gi137869), and FUS/MYXOMA (gi321384) refer to the putative fusion proteins of capripox, orf, and myxoma viruses, respectively. Numbers and consensus amino acids as in Fig. 5. $ indicates serine or threonine.

Only 36 MCV proteins have recognizable cellular homologs. Additional cellular homologs were not identified by screening databases with sequences conserved between MCV proteins and those of other poxviruses and with motifs and HMMs derived from multiple alignments. With the availability of the complete genome sequence of yeast (Goffeau *et al.,* 1996), more than 50% of the nematode *Caenorhabditis elegans* sequence (Waterston and Sulston, 1995), and the majority of human genes represented in dbEST (Adams *et al.,* 1995; Schuler *et al.,* 1996), it appears unlikely that cellular homologs for most poxvirus genes will ever be detected by sequence comparisons. Assuming that viral genes were originally derived from cellular genes, the acquisition of additional functions may have led to dramatic changes in sequence, thereby obscuring their origin. Determination of the three-dimensional structure of cellular and viral proteins, however, may help capture residual relationships.

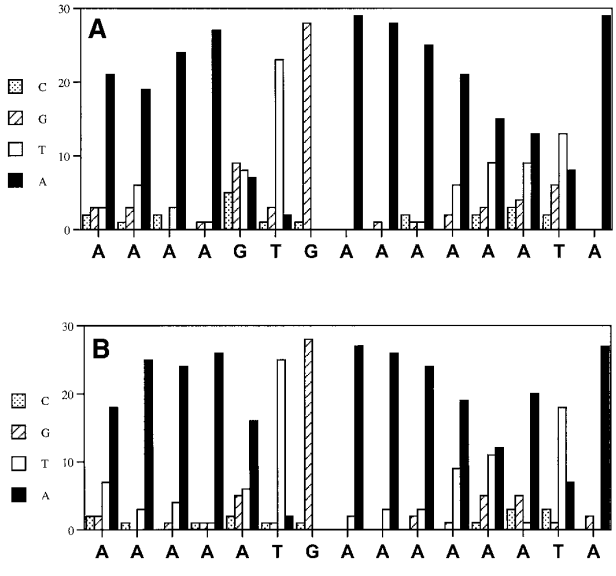

FIG. 8. Consensus early promoter sequences. The sequences upstream of 29 VAC early genes and their MCV homologs were aligned. The occurrence of each nucleotide at every position of the 15-bp core sequence is shown. The most frequent nucleotide at each position is indicated below the graphs. (A) MCV; (B) VAC (Copenhagen strain).

MCV gene expression

Although there is a paucity of experimental data regarding MCV gene expression (Porter *et al.,* 1992), we can confidently conclude that the regulation is very similar to that of VAC. MCV encodes highly conserved homologs of all of the OPV genes known to encode RNA polymerase subunits, transcription factors, and mRNA modification enzymes. Furthermore, early, intermediate, or late VAC promoter consensus sequences are present upstream of MCV genes. This promoter analysis produced a nearly complete, though hypothetical, picture of MCV gene expression (Fig. 2). With only a few possible exceptions, each pair of homologous genes in MCV and OPV possesses the same type of transcriptional signal, indicating that the temporal pattern of gene expression is conserved. In addition, the consensus termination signal TTTTTNT is strategically placed near the ends of many early MCV genes. A great majority of the unique MCV genes appear to have early promoters. Early promoters are also found next to many OPV genes that have no counterparts in MCV and such regulation is compatible with the role of these gene products in host cell interac-

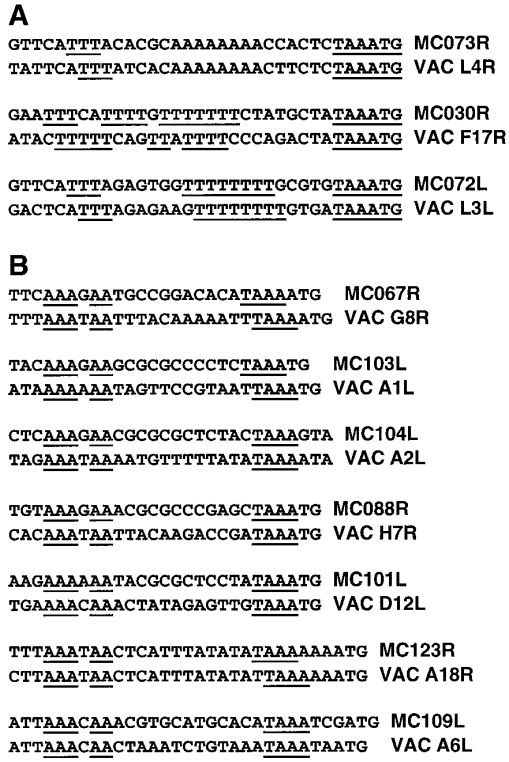

FIG. 9. Alignment of promoters of MCV late and intermediate genes with VAC homologs. (A) The late promoter sequences were aligned with respect to the underlined TAAATG transcriptional/translational start sites; T tracts, typical of strong promoters, are also underlined. (B) Alignment of promoters of VAC intermediate genes with MCV homologs. The sequences upstream of seven VAC genes (indicated on the right) and of their MCV homologs (Table 1) were aligned with respect to the AANAA consensus sequence. G8R, A1L, and A2L are bona fide intermediate promoters. The regulation of H7R, D12L , A18R, and A6L was inferred from the consensus sequence. In each case the ORF starts with the ATG on the right. The TAAA consensus sequence is underlined.

tions at an early stage of the infection. Two interesting exceptions are the deduced late promoters of MCV genes MC013L and MC066L, which encode a DnaJ molecular chaperone homolog and a glutathione peroxidase, respectively.

MCV genome replication

MCV DNA replication, like that of other poxviruses, occurs in cytoplasmic factory regions (Epstein and Fukayama, 1973). Although the details of poxvirus DNA replication remain largely unknown, it seems likely that MCV and OPV use similar mechanisms. A consensus concatemer resolution signal sequence was found near the hairpin termini of the MCV genome, as occurs in other poxviruses. Moreover, genes that have been associated with VAC DNA replication, namely DNA polymerase, uracil DNA glycosylase, topoisomerase, and the putative nucleoside triphosphatase encoded by the D5R gene, are highly conserved in MCV. Nevertheless, MCV lacks many nonessential genes of other poxviruses that may allow viral replication to occur in resting cells. The missing genes include a DNA ligase, two ribonucleotide reductase subunits, glutaredoxin (O2L), thymidine kinase, thymidylate kinase, guanylate kinase, and deoxyuridine triphosphatase. The absence of these enzymes may limit the replication of MCV to metabolically active keratinocytes. In this respect, there is enhanced proliferation of cells in the basal layer of *Molluscum contagiosum* lesions and labeling of viral inclusions with [^{3}H]thymidine appeared most intense in the suprabasal layer and followed a gradient of diminishing labeling through the stratum spinosum toward the granular layer (Epstein *et al.,* 1966).

MCV assembly and spread

MCV particles are similar to those of other poxviruses in size, appearance, and chemical composition (Goodpature and Woodruff, 1931; Pirie *et al.,* 1971; Oda *et al.,* 1982), though antigenic cross-reactivity could not be demonstrated (Mitchell, 1953; Shirodaria and Matthews, 1977). Moreover, the morphological sequence of MCV development into brick-shaped intracellular mature virions (IMV) resembles that of other poxviruses, as determined by electron microscopy (Dourmashkin and Bernhard, 1959; Sutton and Burnett, 1969; Epstein and Fukayama, 1973). Consistent with this, we found that MCV encodes homologs of all of the structural and membrane proteins that are known to be essential for the formation of IMV by OPV.

Subsequent steps, involving the formation of virus inclusions and the membrane wrapping and externalization of virions, appear to vary for different poxviruses. The polarity of virus development in cell strata and the absence of mature virus particles below the spinous layer are unique features of MCV (Sutton and Burnett, 1969). In cells infected with VAC, some IMV are wrapped with Golgi membranes to form intracellular enveloped

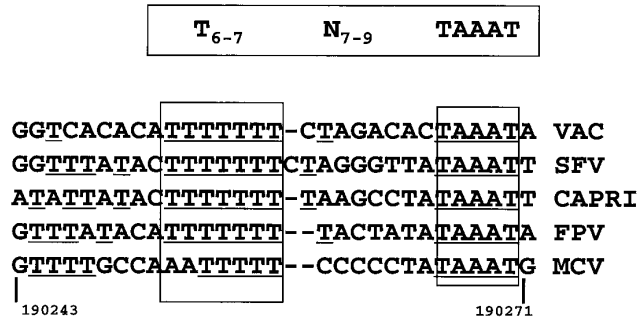

FIG. 10. Alignment of poxvirus concatemer junction resolution signals. The resolution consensus signal (Merchlinsky, 1990) is shown in a box at the top. VAC, Shope fibroma virus (SFV), capripox virus (CAPRI), fowlpox virus (FPV), and MCV sequences are aligned and the two parts of the resolution consensus signal are boxed. In each case the sequences are from the right ITR of the genome and are numbered for MCV. The late promoter consensus TAAAT and T runs, predicted to enhance late promoter function, are underlined.

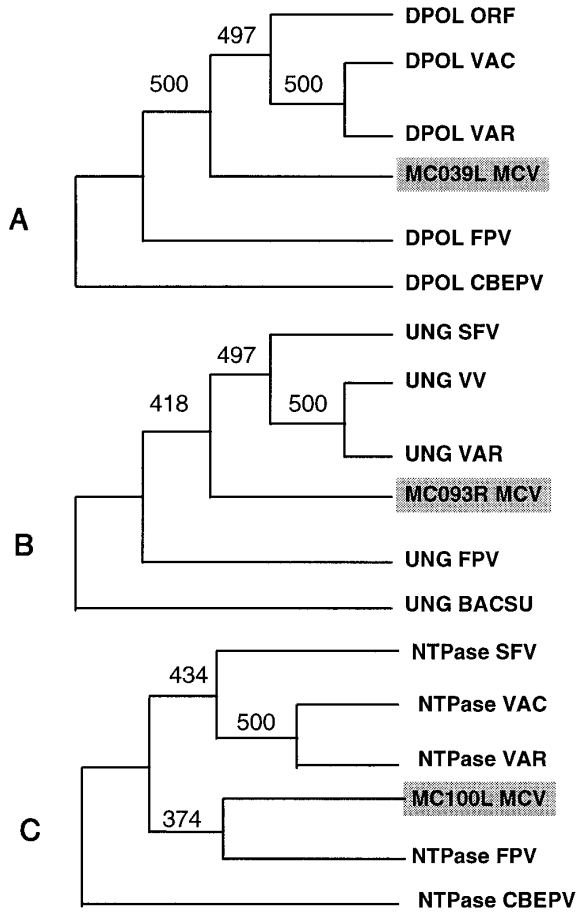

FIG. 11. Phylogeny of poxviruses. Branch lengths are arbitrary. The root position was forced using the respective outgroup. The numbers at the forks show the number of bootstrap repetitions, out of 500, in which the given topology was observed. (A) A PROTPARS tree derived for DNA polymerase (E9L and homologs). (B) A neighbor-joining tree derived for DNA uracil glycosylase (D4R and homologs). (C) A neighbor-joining tree derived for virion NTPase I (D11L and homologs). Protein sequence alignments used for tree construction were generated with the CLUSTALW program (Higgins *et al.,* 1996). The E9L alignment (1055 columns) and the D4R alignment (238 columns) included the entire lengths of all the compared proteins, whereas the D11L alignment (379 columns) included only the C-terminal portions of the proteins since the complete sequence of this gene from Shope Fibroma virus was not available. For each set of protein sequences, the alignment was unambiguous since the alignments of the vertebrate poxvirus sequences contained at least 50% identical residues, and the alignments with the sequences used as outgroups contained at least 30% identical residues. The alignments are available upon request.

virions (IEV) that fuse with the the plasma membrane to liberate extracellular enveloped virions (EEV) that mediate cell-to-cell spread (Schmelz *et al.,* 1994). In contrast, large numbers of MCV are packed into the cytoplasm and greatly distend the keratinocyte; the nucleus is compressed and cytoplasmic membranes and organelles are pushed against the plasma membrane. Some studies demonstrated a well-demarcated zone of lipid or a membranous sheath that physically separates the MCV inclusions from the plasma membrane (van Rooyen, 1938; Shelley and Burmeister, 1986; Heng *et al.,* 1989). Thus, Golgi wrapping of virions, if it occurs at all, is not a prominent feature of MCV morphogenesis, consistent with our determination that there are no MCV homologs for two of the three OPV proteins (B5R and A27R) known to be required for IEV and EEV formation and that others are highly divergent.

Suppression of host defense mechanisms

OPV and leporipoxviruses encode a large number of proteins that antagonize host immune defenses. These include secreted proteins that inhibit complement activation and bind interferons, tumor necrosis factor, and interleukin 1 and intracellular proteins that interfere with the actions of interferon, interleukin 1, and apoptosis signaling pathways (McFadden and Graham, 1994; Alcami and Smith, 1995; Moss, 1996). As MCV infections usually persist in the skin for many months without inflammation (Heng *et al.,* 1989; Viac and Chardonnet, 1990), the absence of MCV homologs of these poxvirus genes, with the exception of 3β-hydroxysteroid dehydrogenase/steroid D5-D4 isomerase, was completely unanticipated (Senkevich *et al.,* 1996). It is likely, therefore, that MCV encodes an entirely new set of proteins that limit the host response. Candidates for such proteins include the previously noted homologs of MHC class I molecules, CC chemokines, and glutathione peroxidase, as well as families of novel membrane proteins (Senkevich *et al.,* 1996) and apoptosis inhibitors (Bertin *et al.,* 1997; this paper). The possession of different host response evasion genes by MCV and OPV can be explained in two general ways. Either MCV and OPV acquired specific host response evasion genes independently after radiating from a hypothetical poxvirus ancestor or the ancestor had a "complete" set of such genes and some were deleted during the evolution of MCV and OPV.

While the production of proteins that antagonize the immune system may contribute to the prevention of inflammation, the restriction of MCV infection to the epidermis, the apparent low level of cell-to-cell spread, and the encapsidation of the lesions may minimize the signaling of the immune system.

Conclusions

The determination of the MCV genome sequence and its comparison with that of OPV have revealed more about the molecular biology of MCV than all previous studies combined. Indeed, the great similarity in organization of the essential genes of MCV and OPV and the almost complete difference in genes involved in host interactions were not anticipated. Moreover, some MCV gene products, e.g., DNA polymerase and topoisomerase, are potential targets for antiviral therapy that may be tested by recombinant methods. A full understanding of MCV replication and host interactions, however, will

require the propagation of the virus in tissue culture systems.

ACKNOWLEDGMENTS

We thank Dr. M. Borodovsky and W. S. Hayes for performing the GeneMark analysis and Dr. R. L. Tatusov and D. R. Walker for expert computer programming assistance. This research was supported in part by a fellowship from the Alexander von Humboldt Stiftun to J.J.B. and grant DA 142/10-2 from the Deutsche Forschungsgemeinschaft to G.D.

REFERENCES

Adams, M. D., Kerlavage, A. R., Fleischmann, R. D., Fuldner, R. A., Bult, C. J., Lee, N. H., Kirkness, E. F., Weinstock, K. G., Gocayne, J. D., White, O., *et al.* (1995). Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature* 377, 173–174.

Ahn, B.-Y., Jones, E. V., and Moss, B. (1990). Identification of the vaccinia virus gene encoding an 18-kilodalton subunit of RNA polymerase and demonstration of a 5′ poly(A) leader on its early transcript. *J. Virol.* 64, 3019–3024.

Ahn, B. Y., and Moss, B. (1992). Glutaredoxin homolog encoded by vaccinia virus is a virion-associated enzyme with thioltransferase and dehydroascorbate reductase activities. *Proc. Natl. Acad. Sci. USA* 89, 7060–7064.

Alcami, A., and Smith, G. L. (1995). Cytokine receptors encoded by poxviruses: A lesson in cytokine biology. *Immunol. Today* 16, 474–478.

Altschul, S. F., Boguski, M. S., Gish, W., and Wootton, J. C. (1994). Issues in searching molecular sequence databases. *Nature Genet.* 6, 119–129.

Altschul, S. F., and Gish, W. (1996). Local alignment statistics. *Methods Enzymol.* 266, 460–480.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.

Bairoch, A., Bucher, P., and Hofmann, K. (1996). The PROSITE database, its status in 1995. *Nucleic Acids Res.* 24, 189–196.

Baldick, C. J., Keck, J. G., and Moss, B. (1992). Mutational analysis of the core, spacer and initiator regions of vaccinia virus intermediate class promoters. *J. Virol.* 66, 4710–4719.

Baroudy, B. M., and Moss, B. (1982). Sequence homologies of diverse length tandem repetitions near ends of vaccinia virus genome suggest unequal crossing over. *Nucleic Acids Res.* 10, 5673–5679.

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

Bertin, J., Armstrong, R. C., Ottilie, S., Martin, D. A., Wang, Y., Banks, S., Wang, G.-H., Senkevich, T. G., Alnemri, E. S., Moss, B., Lenardo, M. J., Tomaselli, K. J., and Cohen, J. L. (1997). DED-containing herpesvirus and poxvirus proteins inhibit both Fas- and TNFR1-induced apoptosis. *Proc. Natl. Acad. Sci. USA* 94, 1172–1176.

Binns, M. M., Tomley, F. M., Campbell, J., and Boursnell, M. E. G. (1988). Comparison of a conserved region in fowlpox virus and vaccinia virus genomes and the translocation of the fowlpox virus thymidine kinase gene. *J. Gen. Virol.* 69, 1275–1283.

Blake, N. W., Porter, C. D., and Archard, L. C. (1991). Characterization of a molluscum contagiosum virus homolog of the vaccinia virus p37K major envelope antigen. *J. Virol.* 65, 3583–3589.

Boguski, M. S., Lowe, T. M., and Tolstoshev, C. M. (1993). dbEST — Database for "expressed sequence tags." *Nature Genet.* 4, 232–233.

Boldin, M. P., Goncharov, T. M., Goltsev, Y. V., and Wallach, D. (1996). Involvement of MACH, a novel MORT1/FADD-interacting protease, in Fas/APO-1- and TNF receptor-induced cell death. *Cell* 85, 803–815.

Borodovsky, M., and McIninch, J. (1993). GenMark: Parallel gene recognition for both DNA strands. *Computers Chem.* 17, 123–133.

Borodovsky, M., Rudd, K. E., and Koonin, E. V. (1994). Intrinsic and extrinsic approaches for detecting genes in a bacterial genome. *Nucleic Acids Res.* 22, 4756–4767.

Bugert, J. J., and Darai, G. (1991). Stability of molluscum contagiosum virus DNA among 184 patient isolates — Evidence for variability of sequences in the terminal inverted repeats. *J. Med. Virol.* 33, 211–217.

Bugert, J. J., Raab, K., Rosenwolff, A., Janssen, W., and Darai, G. (1993). Determination of the position of the boundaries of the terminal repetitive sequences within the genome of molluscum contagiosum virus type-1 by DNA nucleotide sequence analysis. *Virology* 192, 391–396.

Buller, R. M. L., Chen, J. B. W., and Kreider, J. (1995). Replication of molluscum contagiosum virus. *Virology* 213, 655–659.

Claverie, J.-M. (1994). Some useful statistical properties of position-weight matrices. *Comput. Chem.* 18, 287–294.

Condit, R. C., Xiang, Y., and Lewis, J. I. (1996). Mutation of vaccinia virus gene G2R causes suppression of gene A18R ts mutants: Implication for control of transcription. *Virology* 220, 10–19.

Cotton, D. W. K., Cooper, C., Barrett, D. F., and Leppard, B. J. L. (1987). Severe atypical molluscum contagiosum infection in an immunocompromised host. *Br. J. Dermatol.* 116, 871–876.

Darai, G., Reisner, H., Scholz, J., Schnitzler, P., and Lorbacher de Ruiz, H. (1986). Analysis of the genome of molluscum contagiosum virus by restriction endonuclease analysis and molecular cloning. *J. Med. Virol.* 18, 29–39.

Davison, A. J., and Moss, B. (1989a). The structure of vaccinia virus early promoters. *J. Mol. Biol.* 210, 749–769.

Davison, A. J., and Moss, B. (1989b). The structure of vaccinia virus late promoters. *J. Mol. Biol.* 210, 771–784.

DeLange, A. M., Futcher, B., Morgan, R., and McFadden, G. (1984). Cloning of the vaccinia virus telomere in a yeast plasmid vector. *Gene* 27, 13–21.

DeLange, A. M., and McFadden, G. (1987). Efficient resolution of replicated poxvirus telomeres to native hairpin structures requires two inverted symmetrical copies of a core target DNA sequence. *J. Virol.* 61, 1957–1963.

Douglass, N. J., Blake, N. W., Cream, J. J., Soterious, B. A., Zhang, H. Y., Theodoridou, A., and Archard, L. C. (1996). Similarity in genome organization between molluscum contagiosum virus (MCV) and vaccinia virus (VV): Identification of MCV homologues of the VV genes for protein kinase 2, structural protein VP8, RNA polymerase 35 kDa subunit and 3b-hydroxysteroid dehydrogenase. *J. Gen. Virol.* 77, 3113–3120.

Dourmashkin, R., and Bernhard, W. (1959). A study with the electron microscope of the skin tumour of molluscum contagiosum. *J. Ultrastructure Res.* 3, 11–38.

Eddy, S. R., Mitchison, G., and Durbin, R. (1995). Maximum discrimination models of sequence consensus. *Comput. Biol. J.* 2, 9–23.

Epstein, W. L. (1992). Molluscum contagiosum. *Semin. Dermatol.* 11, 184–189.

Epstein, W. L., Conant, M. A., and Krasnobrod, H. (1966). Molluscum contagiosum: Normal and virus infected epidermal cell kinetics. *J. Invest. Dermatol.* 46, 91–103.

Epstein, W. L., and Fukayama, K. (1973). Maturation of molluscum contagiosum virus (MCV) in vivo: Quantitative electron microscopic autoradiography. *J. Invest. Dermatol.* 60, 73–79.

Felsenstein, J. (1996). Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol.* 266, 418–427.

Fife, K. H., Whitfeld, M., Faust, H., Goheen, M. P., Bryan, J., and Brown, D. (1996). Growth of molluscum contagiosum virus in a human foreskin xenograft model. *Virology* 226, 95–112.

Fleming, S. B., Blok, J., Fraser, K. M., Mercer, A. A., and Robinson, A. J. (1993). Conservation of gene structure and arrangement between vaccinia virus and orf virus. *Virology* 195, 175–184.

Franke, C. A., Wilson, E. M., and Hruby, M. D. (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.

Garon, C. F., Barbosa, E., and Moss, B. (1978). Visualization of an inverted terminal repetition in vaccinia virus DNA. *Proc. Natl. Acad. Sci. USA* 75, 4863–4867.

Gershon, P. D., Ansell, D. M., and Black, D. N. (1989). A comparison of the genome organization of capripoxvirus with that of the orthopoxviruses. *J. Virol.* 63, 4703–4708.

Goebel, S. J., Johnson, G. P., Perkus, M. E., Davis, S. W., Winslow, J. P., and Paoletti, E. (1990). The complete DNA sequence of vaccinia virus. *Virology* 179, 247–266; 517–563.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., and Oliver, S. G. (1996). Life with 6000 genes. *Science* 274, 546, 563–567.

Gong, S. C., Lai, C. F., and Esteban, M. (1990). Vaccinia virus induces cell fusion at acid pH and this activity is mediated by the N-terminus of the 14-kDa virus envelope protein. *Virology* 178, 81–91.

Goodpature, E. W., and Woodruff, C. E. (1931). A comparison of the inclusion bodies of fowlpox and Molluscum contagiosum. *Am. J. Pathol.* 7, 1–7.

Gottlieb, S. L., and Myskowski, P. L. (1994). Molluscum contagiosum. *Int. J. Dermatol.* 33, 453–461.

Gvakharia, B. O., Koonin, E., and Mathews, C. K. (1996). Vaccinia virus G4L gene encodes a second glutaredoxin. *Virology* 226, 408–411.

Hadasch, P., Buget, J. J., Janssen, W., and Darai, G. (1993). Characterization of the genome of molluscum contagiosum virus type 1 between the genome coordinates 0.045 and 0.075 by DNA nucleotide sequence analysis of a 5.6-kb *Hin*dIII/*Mlu*I DNA fragment. *Intervirology* 36, 32–43.

Hagiya, M., Francavilla, A., Polimeno, L., Ihara, I., Sakai, H., Seki, T., Shimonishi, M., Porter, K. A., and Starzl, T. (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.

Heng, M. C. Y., Steuer, M. E., Levy, A., McMahon, S., Richmn, M., Allen, S. G., and Blackhart, B. (1989). Lack of host cellular immune response in eruptive molluscum contagiosum. *A. J. Dermatopathol.* 11, 248–254.

Higgins, D.G., Thompson, J.D., and Gibson, T.J. (1996). Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.* 266, 383–402.

Holmgren, A. (1989). Thioredoxin and glutaredoxin systems. *J. Biol. Chem.* 264, 13963–13966.

Hu, F.-Q., and Pickup, D. J. (1991). Transcription of terminal loop region of vaccinia virus DNA is initiated from telomere sequences directing DNA resolution. *Virology* 181, 716–720.

Ichihashi, Y., Matsumoto, S., and Dales, S. (1971). Biogenesis of poxviruses: Role of A-type inclusions and host cell membranes in virus dissemination. *Virology* 46, 507–532.

Ichihashi, Y., and Oie, M. (1996). Neutralizing epitopes on penetration protein of vaccinia virus. *Virology* 220, 491–494.

Johnson, G. P., Goebel, S. J., and Paoletti, E. (1993). An update on the vaccinia virus genome. *Virology* 196, 381–401.

Keck, J. G., Baldick, C. J., and Moss, B. (1990). Role of DNA replication in vaccinia virus gene expression: A naked template is required for transcription of three late transactivator genes. *Cell* 61, 801–809.

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.

Koonin, E. V. (1996). A duplicated catalytic motif in a new superfamily of phosphohydrolases and phospholipid synthases that includes poxvirus envelope proteins. *Trends Biochem. Sci.* 21, 242–243.

Koonin, E. V., and Senkevich, T. G. (1992a). Evolution of thymidine and thymidylate kinases — The possibility of independent capture of TK genes by different groups of viruses. *Virus Genes* 6, 187–196.

Koonin, E. V., and Senkevich, T. G. (1992b). Vaccinia virus encodes four putative DNA and/or RNA helicases distantly related to each other. *J. Gen. Virol.* 73, 989–993.

Koonin, E. V., Tatusov, R. L., and Rudd, K. E. (1996). Protein sequence comparison at genome scale. *Methods Enzymol.* 266, 295–322.

Kovacs, G. R., and Moss, B. (1996). The vaccinia virus H5R gene encodes late gene transcription factor 4: Purification, cloning and overexpression. *J. Virol.* 70, 6796–6802.

Lee-Chen, G. J., Bourgeois, N., Davidson, K., Condit, R. C., and Niles, E. G. (1988). Structure of the transcription initiation and termination sequences of seven early genes in the vaccinia virus *Hin*dIII D fragment. *Virology* 163, 64–79.

Lisowsky, T. (1992). Dual function of a new nuclear gene for oxidative phosphorylation and vegetative growth in yeast. *Mol. Gen. Genet.* 232, 58–64.

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.

Liu, K., Lemon, B., and Traktman, P. (1995). The dual-specificity phosphatase encoded by vaccinia virus, VH1, is essential for viral transcription in vivo and in vitro. *J. Virol.* 69, 7823–7834.

Luo, Y., and Shuman, S. (1991). Antitermination of vaccinia virus early transcription: Possible role of RNA secondary structure. *Virology* 185, 432–436.

Massung, R. F., Esposito, J. J., Liu, L.-L., Qi, J., Utterback, T. R., Knight, J. C., Aubom, L., Yuran, T. E., Parsons, J. M., Loparev, V. N., Selivanov, N. A., Cavallaro, K. F., Kerlavage, A. R., Mahy, B. W. J., and Venter, J. C. (1993). Potential virulence determinants in terminal regions of variola smallpox virus genome. *Nature* 366, 748–751.

Massung, R. F., Liu, L.-I., Qi, J., Knight, J. C., Yuran, T. E., Kerlavage, A. R., Parsons, J. M., Venter, J. C., and Esposito, J. (1994). Analysis of the complete genome of smallpox variola major virus strain Bangladesh-1975. *Virology* 201, 215–240.

McFadden, G., and Graham, K. (1994). Modulation of cytokine networks by poxviruses: The myxoma virus model. *Semin. Virol.* 5, 421–429.

Mercer, A. A., Lyttle, D. J., Whelan, W. M., Fleming, A. B., and Sullivan, J. T. (1995). The establishment of a genetic map of ORF virus reveals a pattern of genomic organization that is highly conserved among divergent poxviruses. *Virology* 212, 698–704.

Merchlinsky, M. (1990). Mutational analysis of the resolution sequence of vaccinia virus DNA — Essential sequence consists of 2 separate AT-rich regions highly conserved among poxviruses. *J. Virol.* 64, 5029–5035.

Merchlinsky, M., and Moss, B. (1989). Nucleotide sequence required for resolution of the concatemer junction of vaccinia virus DNA. *J. Virol.* 63, 4354–4361

Mitchell, J. C. (1953). Observations on the virus of *Molluscum contagiosum. Br. J. Exp. Pathol.* 34, 44–49.

Mockett, B., Binns, M. M., Boursnell, M. E. G., and Skinner, M. A. (1992). Comparison of the locations of homologous fowlpox and vaccinia virus genes reveals major genome reorganization. *J. Gen. Virol.* 73, 2661–2668.

Moss, B. (1996). *Poxviridae:* The viruses and their replication. *In* "Fields Virology" (B. N. Fields, D. M. Knipe, and P. M. Howley, Eds.), pp. 2637–2671. Lippincott-Raven, Philadelphia.

Murphy, F. A., Fauquet, C. M., Bishop, D. H. L., Ghabrial, S. A., Jarvis, A. W., Martelli, G. P., Mayo, M. A., and Summers, M. D. (1995). Virus taxonomy. Classification and nomenclature of viruses. *Arch. Virol. Suppl.* 10, 1–586.

Muzio, M., Chinnaiyan, A. M., Kischkel, F. C., O'Rourke, K., Shevchenko, A., Ni, J., Scaffidi, C., Bretz, J. D., Zhang, M., Gentz, R., Mann, M., Krammer, P. H., Peter, M. E., and Dixit, V. M. (1996). FLICE, a novel FADD-homologous ICE/CED-3-like protease, is recruited to the CD95 (Fas/APO-1) death-inducing signaling complex. *Cell* 85, 817–827.

Myers, E. W., and Miller, W. (1988). Optimal alignments in linear space. *Comput. Appl. Biosci.* 4, 11–17.

Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* 10, 1–6.

Oda, H., Ohayama, Y., Sameshima, T., and Hirakawa, K. (1982). Structural polypeptides of Molluscum contagiosum virus: Their variability in various isolates and locations within the virion. *J. Med. Virol.* 9, 19–25.

Ogawa, R., Calvert, J. G., Yanagida, N., and Nazerian, K. (1993). Insertional inactivation of a fowlpox virus homologue of the vaccinia virus F12L gene inhibits the release of enveloped virions. *J. Gen. Virol.* 74, 55–64.

Osterrieder, N., Meyer, H., and Pfeffer, M. (1994). Characterization of the gene encoding the A-type inclusion body protein of mousepox virus. *Virus Genes* 8, 125–135.

Parr, R. P., Burnett, J. W., and Garon, C. F. (1977). Structural characterization of the molluscum contagiosum virus genome. *Virology* 81, 247–256.

Patel, D. D., Pickup, D. J., and Joklik, W. K. (1986). Isolation of cowpox virus A-type inclusions and characterization of their major protein component. *Virology* 149, 174–189.

Pickup, D. J., Bastia, D., Stone, H. O., and Joklik, W. (1982). Sequence of terminal regions of cowpox virus DNA: Arrangement of repeated and unique sequence elements. *Proc. Natl. Acad. Sci. USA* 79, 7112–7116.

Pirie, G. D., Bishop, P. M., Burke, D. C., and Postlethwaite, R. (1971). Some properties of purified molluscum contagiosum virus. *J. Gen. Virol.* 13, 311–320.

Ponting, C. P., and Kerr, I. D. (1996). A novel family of phospholipase D homologues that includes phospholipid synthases and putative endonucleases: Identification of duplicated repeats and potential active site residues. *Protein Sci.* 5, 914–922.

Porter, C. D., and Archard, L. C. (1987). Characterization and physical mapping of Molluscum contagiosum virus DNA and location of a sequence capable of encoding a conserved domain of epidermal growth factor. *J. Gen. Virol.* 68, 673–682.

Porter, C. D., and Archard, L. C. (1992). Characterisation by restriction mapping of three subtypes of molluscum-contagiosum virus. *J. Med. Virol.* 38, 1–6.

Porter, C. D., Blake, N. W., Cream, J. J., and Archard, L. C. (1992). Molluscum contagiosum virus. *Mol. Cell. Biol. Hum. Dis.* 1, 233–257.

Postlethwaite, R. (1970). Molluscum contagiosum. A review. *Arch. Environ. Health* 21, 432–452.

Rajagopal, I., Ahn, B. Y., Moss, B., and Mathews, C. K. (1995). Roles of vaccinia virus ribonucleotide reductase and glutaredoxin in DNA precursor biosynthesis. *J. Biol. Chem.* 270, 27415–27418.

Ravanello, M. P., and Hruby, D. E. (1994). Conditional lethal expression of the vaccinia virus L1R myrstylated protein reveals a role in virus assembly. *J. Virol.* 68, 6401–6410.

Rempel, R. E., and Traktman, P. (1992). Vaccinia virus-B1 kinase — Phenotypic analysis of temperature-sensitive mutants and enzymatic characterization of recombinant proteins. *J. Virol.* 66, 4413–4426.

Rodriguez, J. F., Paez, E., and Esteban, M. (1987). A 14,000-Mr envelope protein of vaccinia virus is involved in cell fusion and forms covalently linked trimers. *J. Virol.* 61, 395–404.

Rosel, J. L., Earl, P. L., Weir, J. P., and Moss, B. (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, 436–439.

Rost, B., Sander, C., and Schneider, R. (1994). PHD — An automatic mail server for protein secondary structure prediction. *Comput. Appl. Biosci.* 10, 53–60.

Schmelz, M., Sodeik, B., Ericsson, M., Wolffe, E. J., Shida, H., Hiller, G., and Griffiths, G. (1994). Assembly of vaccinia virus: The second wrapping cisterna is derived from the trans Golgi network. *J. Virol.* 68, 130–147.

Scholz, J., Rosen-Wolff, A., Bugert, J., Reisner, H., White, M. I., Darai, G., and Postlethwaite, R. (1989). Epidemiology of molluscum contagiosum using genetic analysis of the viral DNA. *J. Med. Virol* 27, 87–90.

Schuler, G. D., Altschul, S. F., and Lipman, D. J. (1991). A workbench for multiple alignment construction and analysis. *Proteins: Structure, Function, Genet.* 9, 180–190.

Schuler, G. D., Boguski, M. S., Stewart, E. A., Stein, L. D., Gyapay, G., Rice, K., White, R. E., Rodriguez-Tome, P., Aggarwal, A., Bajorek, E., Bentolila, S., Birren, B. B., Butler, A., Castle, A. B., Chiannilkulchai, N., Chu, A., Clee, C., Cowles, S., Day, P. J., Dibling, T., Drouot, N., Dunham, I., Duprat, S., East, C., Hudson, T. J., *et al.* (1996). A gene map of the human genome. *Science* 274, 540–546.

Schwartz, J. J., and Myskowski, P. L. (1992). Molluscum contagiosum in patients with human immunodeficiency virus infection — A review of 27 patients. *J. Am. Acad. Dermatol.* 27, 583–588.

Senkevich, T. G., Bugert, J. J., Sisler, J. R., Koonin, E. V., Darai, G., and Moss, B. (1996). Genome sequence of a human tumorigenic poxvirus: Prediction of specific host response-evasion genes. *Science* 273, 813–816.

Senkevich, T. G., Muravnik, G. L., Pozdnyakov, S. G., Chizhikov, V. E., Ryazankina, O. I., Shchelkunov, S. N., Koonin, E. V., and Chernos, V. I. (1993). Nucleotide sequence of XhoI O fragment of ectromelia virus DNA reveals significant differences from vaccinia virus. *Virus Res.* 30, 73–88.

Shchelkunov, S. N., Marennikova, S. S., Blinov, V. M., Resenchuk, S. M., Totmenin, A. V., Chizhikov, V. E., Gutorov, V. V., Safronov, P. F., Kurmanov, R. K., and Sandakhchiev, L. S. (1993). [Entire coding sequence of the variola virus]. *Dokl. Akad. Nauk.* 328, 629–632.

Shelley, W. B., and Burmeister, V. (1986). Demonstration of a unique viral structure: The molluscum viral colony sac. *Br. J. Dermatol.* 115, 557–562.

Shida, H., Tanabe, K., and Matsumoto, S. (1977). Mechanism of virus occlusion into A-type inclusion during poxvirus infection. *Virology* 76, 217–233.

Shirodaria, P. V., and Matthews, R. S. (1977). Observations on the antibody responses in *Molluscum contagiosum. Br. J. Dermatol.* 96, 29–34.

Smith, G. L., Chan, Y. S., and Howard, S. T. (1991). Nucleotide sequence of 42 kbp of vaccinia virus strain WR from near the right inverted terminal repeat. *J. Gen. Virol.* 72, 1349–1376.

Sonntag, K.-C., Clauer, U., Bugert, J. J., Schnitzler, P., and Darai, G. (1995). Identification and properties of the genes encoding the poly(A) polymerase and a small (22 kDa) and the largest subunit (147 kDa) of the DNA-dependent RNA polymerase of molluscum contagiosum virus. *Virology* 210, 471–478.

Sonntag, K.-C., and Darai, G. (1996). Strategy for identifying the gene encoding the DNA polymerase of molluscum contagiosum virus type 1. *Virus Genes* 13, 31–44.

Stuart, D., Graham, K., Schreiber, M., Macaulay, C., and McFadden, G. (1991). The target DNA sequence for resolution of poxvirus replicative intermediates is an active late promoter. *J. Virol.* 65, 61–70.

Sutton, J. S., and Burnett, J. W. (1969). Ultrastructural changes in dermal and epidermal cells of skin infected with molluscum contagiosum virus. *J. Ultrastructure Res.* 26, 177–196.

Tatusov, R. L., Altschul, S. F., and Koonin, E. V. (1994). Detection of conserved segments in proteins: Iterative scanning of sequence databases with alignment blocks. *Proc. Natl. Acad. Sci. USA* 91, 12091–12095.

Tatusov, R. L., and Koonin, E. V. (1994). A simple tool to search for sequence motifs that are conserved in BLAST outputs. *Comput. Appl. Biosci.* 10, 457–459.

Towler, D. A., Adams, S. P., Eubanks, S. R., Towery, D. S., Jackson-Machelski, E., Glaser, L., and Gordon, J. I. (1988). Myristoyl CoA:protein *N*-myristoyltransferase activities from rat liver and yeast possess overlapping yet distinct peptide substrate specificities. *J. Biol. Chem.* 263, 1784–1790.

Upton, C., and McFadden, G. (1986). Tumorigenic poxviruses: Analysis of viral DNA sequences implicated in the tumorigenicity of Shope fibroma virus and malignant rabbit virus. *Virology* 152, 308–321.

van Rooyen, C. E. (1938). The micromanipulation and microdissection of the molluscum contagiosum inclusion body. *J. Pathol.* 46, 425–436.

Viac, J., and Chardonnet, Y. (1990). Immunocompetent cells and epithelial cell modifications in molluscum contagiosum. *J. Cutan. Pathol.* 17, 202–205.

Walker, D. R., and Koonin, E. V. (1997) SEALS: A system for easy analysis of lots of sequences. *ISMB* 5, in press.

Waterston, R., and Sulston, J. (1995). The genome of *Caenorhabditis elegans. Proc. Natl. Acad. Sci. USA* 92, 10836–10840.

Whitehead, S., and Hruby, D. (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.

Wilcock, D., and Smith, G. L. (1996). Vaccinia virions lacking core protein VP8 are deficient in early transcription. *J. Virol.* 70, 934–943.

Wittek, R., Cooper, J., Barbosa, E., and Moss, B. (1980). Expression of the vaccinia virus genome—Analysis and mapping of mRNAs encoded within the inverted terminal repetition. *Cell* 21, 487–493.

Wittek, R., Menna, A., Muller, K., Schumperli, D., Bosley, P. G., and Wyler, R. (1978). Inverted terminal repeats in rabbits poxvirus and vaccinia virus DNA. *J. Virol.* 28, 171–181.

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

Wolffe, E. J., Vijaya, S., and Moss, B. (1995). A myristylated, membrane protein encoded by the vaccinia virus L1R open reading frame is the target of potent neutralizing monoclonal antibodies. *Virology* 211, 53–63.

Wootton, J. C., and Federhen, S. (1996). Analysis of compositionally biased regions in sequence databases. *Methods Enzymol.* 266, 554–573.

Yuen, L., and Moss, B. (1987). Oligonucleotide sequence signaling transcriptional termination of vaccinia virus early genes. *Proc. Natl. Acad. Sci. USA* 84, 6417–6421.

Zajac, P., Spehner, D., and Drillien, R. (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.