# 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 3 $\beta$-hydroxysteroid dehydrogenase 

Nicola J. Douglass, ${ }^{1,2}$ Neil W. Blake, ${ }^{1} \dagger$ Jeffrey J. Cream, ${ }^{3}$ Bambos A. Soteriou, ${ }^{1}$ Hong Yi Zhang, ${ }^{1}$ Agapi Theodoridou' and Leonard C. Archard ${ }^{1}$<br>${ }^{\dagger}$ Department of Biochemistry, Charing Cross and Westminster Medical School, St Dunstan's Road, London W6 8RP, UK<br>${ }^{2}$ Department of Medical Microbiology, University of Cape Town, Observatory, 7925 Cape Town, South Africa<br>${ }^{3}$ Department of Dermatology, Charing Cross Hospital, Fulham Palace Road, London W6 8RF, UK


#### Abstract

Molluscum contagiosum virus (MCV) and vaccinia virus (W) are serologically unrelated poxviruses with a disparate genome composition (MCV, $66 \%$ $\mathbf{G}+\mathbf{C} ; \mathbf{V V}, \mathbf{3 3 \%} \mathbf{G}+\mathbf{C}$ ). Molecular studies of MCV have been hindered by the inability to propagate the virus in cells cultured in vitro. We sequenced 7765 bp of MCV DNA cloned from four widely spaced regions throughout the MCV genome and identified a total of 11 potential open reading frames (ORF), designated CX1-11. These include MCV homologues of the VV genes encoding protein kinase 2, structural protein VP8, RNA polymerase 35 kDa subunit and $3 \beta$-hydroxysteroid dehydrogenase. The position and orientation of the MCV ORFs was collinear to the V genome, with the exception of the region around ORF CX11 which is inverted in the MCV genome.


Molluscum contagiosum virus (MCV) is the only species of the genus Molluscipoxvirus and does not cross-react serologically with other poxviruses (Mitchell, 1953). MCV causes raised umbilicated lesions in the epidermis of human skin (Postlethwaite, 1970), specifically infecting differentiating keratinocytes. The virus has not been cultured in vitro although

[^0]MCV replication has been demonstrated in human foreskin grafted to athymic mice (Buller et al., 1995). MCV infection is common in children, sexually transmitted between adults and is frequently seen in immunocompromised individuals (Cotton et al., 1987). In contrast to most poxviruses, the MCV genome has a high $G+C$ content ( $66 \%$ ). Restriction endonuclease analysis of the MCV genome shows no similarity to other poxviruses and there is considerable variability in restriction enzyme maps between three MCV subtypes (Porter \& Archard, 1992). The genome has inverted terminal repeats (Porter \& Archard, 1987; Bugert et al., 1989) containing tandemly repeated sequences (Bugert et al., 1993). Initial sequencing studies on MCV type I (MCV-I) identified an open reading frame (ORF) potentially encoding a 43 kDa protein, homologous to the vaccinia virus (VV) major envelope antigen (Blake et al., 1991). Further studies identified a number of ORFs potentially encoding limited homology to VV proteins (Bugert et al., 1993; Hadasch et al., 1993) and apparent MCV homologues of the VV genes encoding poly $(\mathrm{A})$ polymerase and small ( 22 kDa ) and large ( 147 kDa ) subunits of the DNAdependent RNA polymerase (Sonntag et al., 1995). The latter occur within BamHI fragment A of MCV-I DNA and their arrangement is co-linear with the equivalent VV locus.

We now report the sequence of four widely spaced regions of the MCV-I genome and the identification of putative MCV genes by comparison to the VV genomic sequence (Goebel et al., 1990; Johnson et al., 1993).

MCV-I was purified from human skin lesions and virus DNA extracted as described by Porter \& Archard (1987). The HindIII, BamHI and ClaI restriction enzyme maps of MCV-I DNA are described by Porter \& Archard (1992). HindIII fragments Q, C, E, I, $\mathrm{J}_{\mathrm{L}}, \mathrm{N}, \mathrm{M}, \mathrm{H}, \mathrm{K}_{\mathrm{R}}$ and $\mathrm{J}_{\mathrm{R}}$ were eluted from agarose gels and cloned in the Hindill site of plasmid pUC1318 (Kay \& McPherson, 1987) by standard methods. The recombinant HindIII C plasmid was digested with BamHI and sub-


Fig. 1. Hindill restriction enzyme maps of $V V$ and $M C V-I$, showing the regions of MCV-I sequenced. Open reading frames are indicated by arrows. The MCV-I ORFs are designated $\mathrm{CX1}-\mathrm{CX} 11$ reading from left to right in the genome and are compared with the corresponding ORFs in W.
fragments were recloned to give clones of MCV-I BamHI fragments M, N and L. Genomic MCV-I DNA was digested with ClaI and cloned randomly in the ClaI site of Bluescript $\mathrm{KS}(+$ ) (Stratagene): recombinant plasmids containing ClaI fragments F, H and E were isolated. Each cloned MCV-I fragment was sequenced using commercially available pUC forward and reverse primers with the PRISM Ready Reaction DyeDeoxy Terminator Cycle sequencing kit and the Applied Biosystems model 373A DNA sequencing system. ORFs were identified using DNA Strider 1.2 and potential protein homologies were identified using BLAST searches. Sequences of interest were extended by custom primer walking or sequencing appropriate subclones with pUC forward and reverse primers. The sequences reported here were determined in both directions. Where sequences spanned two adjacent cloned fragments such as BamHI L and N, HindIII I and E or HindIII N and M , primers were designed to span the restriction sites. PCR sequencing using genomic DNA for the HindIII I and E or HindIII N and M junctions and a plasmid containing the HindIII C fragment including the BamHI L and N junction (see Fig. 1) confirmed that these fragments were indeed contiguous within the MCV genome. Repetitive sequencing in both directions revealed no ambiguities.

Four regions of MCV-I DNA totalling 7765 bp were sequenced (Fig. 1) and analysis of these revealed marked similarity between $M C V$ and $V V$ at the amino acid level. These regions of MCV-I DNA correspond to VV ORFs F9L, F10L, L3L to J3R, A28L-30L and A44L which include the VV genes known to code for protein kinase 2 (F10L), structural protein VP8 (L4R), RNA polymerase 35 kDa subunit (A29L) and the $3 \beta$-hydroxysteroid dehydrogenase ( $3 \beta$-HSD) (A44L: Goebel et al., 1990; Johnson et al., 1993). The sequences of the expanded regions of Fig. I are shown in Fig. 2. The 1850 bp MCV sequence shown in Fig. 2(a) comprises the entire BamHI N fragment plus 251 bp of the left end of BamHI M and reads from right to left in the genome. A complete ORF of 1329 bp (CXI) is present in position 110-1438, encoding a putative
protein of 443 amino acids (aa) which shows $60 \cdot 2 \%$ identity and $80.5 \%$ similarity to VV protein kinase 2 (VVPK2). The catalytic domains I, II and VI of VVPK2 (Lin \& Broyles, 1994) are conserved in the putative MCV protein sequence (Fig. 3a), suggesting that a protein expressed from this ORF would have kinase activity. A partial ORF (CX1) extending from position 1422-1850 shows $44 \cdot 4 \%$ identity and $72 \cdot 2 \%$ similarity at the amino acid level to VV ORF F9L.

The 3158 bp sequence at the right end of the centrally located MCV-I ClaI F fragment is shown in Fig. 2(b). This contains three complete ORFs of 762 bp (CX4), 438 bp (CX5) and 549 bp (CX6) and two partial ORFs of 88 bp (CX3) and 863 bp (CX7). These MCV ORFs correspond to VV genes L3L-J3R (Goebel et al., 1990), with the exception of the thymidine kinase (TK) gene, J2R, not found at this position in MCV-I. The complete ORFs code for putative polypeptides of 254 (CX4), 146 (CX5) and 183 (CX6) aa, resembling the products of VV ORFs L4R ( $42.8 \%$ identity; $75.3 \%$ similarity), L5R ( $31.8 \%$ identity; $60.6 \%$ similarity) and JIR ( $44.5 \%$ identity; $74 \cdot 2 \%$ similarity) respectively.

The partial 863 bp ORF (CX7) shows $57 \cdot 2 \%$ aa identity and $80.7 \%$ aa similarity to VV J3R. VV ORF L4R encodes the structural protein VP8 (Yang et al., 1988) and VV J3R encodes the poly(A) polymerase stimulatory subunit (Gershon et al., 1991). The VV TK gene is located between ORFs JIR and J3R (Hruby \& Ball, 1982) but an MCV-I homologue was not found in the corresponding location. Thymidine kinase activity has not been demonstrated in MCV and it is not known whether this gene is absent or situated in a different position in the MCV-I genome. Fowlpox virus (FPV) also lacks a TK gene in this position (Drillien et al., 1987) but it has been located in the region corresponding to VV I4L (Binns et al., 1992). Fig. 2 (c) shows 1251 bp of MCV-I DNA, located at the left end of HindIII I and the right end of HindIII E and reading from right to left in the genome. It includes two complete ORFs of 201 bp (CX10) and 978 bp (CX9) at positions $80-280$ and 183-1160 respectively. CXIO encodes a putative polypeptide of 67 aa


Fig. 2. (a) For legend see page 3117.
which is $44.9 \%$ identical and $75 \cdot 4 \%$ similar to that of VV ORF A30L. CX9 encodes a putative 303 aa protein which shows $44.5 \%$ identity and $69.5 \%$ similarity to the RNA polymerase 35 kDa subunit encoded by VV A29L (Amegadzie et al., 1991). Downstream of the MCV-I homologue of A29L is the start of ORF CX8 at position 1206, which shows $66 \cdot 7 \%$ identity to VV ORF A28L over the first 15 aa .

Fig. 2(d) shows 1506 bp of sequence from the right end of MCV-I HindIII fragment N and the left end of HindIII M. This contains one complete ORF of 1248 bp at positions $70-1317$ (CX1I) encoding a putative product of 354 aa which is $44 \%$ identical and $70.3 \%$ similar to VV $3 \beta$-HSD, the product of VV ORF A44L (Moore \& Smith, 1992). This MCV-I ORF is in the opposite orientation to the VV gene and a sequence to the left of CXII (data not shown) is similar to VV A45R, which is located to the right of VV A 44 L .

Analysis of the sequences $5^{\prime}$ to MCV ORFs reveals motifs consistent with known poxvirus early and late promoters (Davison \& Moss, 1989a, b) which are conspicuous amongst
the $\mathrm{G}+\mathrm{C}$-rich coding regions of MCV. Elements with homology to VV early promoters were present upstream of ORFs CX9 and CXII and the VV late motif TAAAT was present upstream of ORFs CXI-7 and CX10 (see Fig. 2). It has yet to be determined whether these motifs are active in transcription of MCV genes.

These results demonstrate collinearity in gene organization between MCV-I and VV in the region between VV ORFs F9L and A45R. Seven complete MCV-I ORFs have been identified which, when translated, would show $32-60 \%$ aa identity and $61-81 \%$ aa similarity to their respective VV homologues (see Fig. 3). Such a high degree of protein conservation in two viruses which do not cross-react serologically and which have genomes of disparate base composition implies the conservation of functions essential to poxvirus infectivity, such as the kinase, structural protein and RNA polymerase subunit. Conservation of a potential virulence factor, the $3 \beta-\mathrm{HSD}$, thought to reduce host inflammatory response by inducing host steroid hormone synthesis (Moore \& Smith, 1992), is
(b)

R S E L F H A H P K P A D L H A H P K $Q$ G D G G A N O K
1 cgCGACTCGAGGAAGTGCGCGTGCGGITTGGGCGCGTCCAAGTGCGCGTGAGGCTITTGCCCGTCGCCGCCCGCGTICTGCITGTTCÄITITACACGCAAA $\xrightarrow{\mathrm{CXC}_{4}}$
 M N A L F E N L F D E D A V C A G S V S R E D F L L V V A 201 CGGCGCCAAGGTCAAGTTCCCGCGTTCGCTGCTGTCCATGTACCGCGTGGTGCCGCGCACCATGTCGCGTTACGAGCTGGCGCTCGTGCAGICCGAGACG

301 GTCACGGGCGTGGTCTTCACCACCGTGTACAACGTGCGCCGCAACCTGGGGCTGGAGGAGCGCGAGGCGCTGAGCCTGCCCGCGCTCGAGAAGTACTACC
63 V T G V V F T T V Y N V R R N L G L E E R E A L S L P A 401 tgGaCAAGGCGAACGACGTGCTTACGCTCATGGTCAACAACACCAACCTGGAGCACATAGCGGCCTACCGCATGCGCAGCCGGCGCCTGCTCAACCCCGT
97 D K A N D V L T L M V N N T N L E H I A A Y R M R S R R L L 501 gGtgirccgcgcgggcgcagtgccgctggcgctggictttacctcgcgcaagaagctcagcaictatcgcgaggacaccagccagccggccgaagacagc
130 V F R A G A V P L A L V F T
g01 acctacaccaagaicgcggccaatgtggcgctggcggecaatacgcaggectgctgctgctigatgigcacacgcccgggacggcgctcatgctgaccg
163 T $Y$ T K I A A N V A L A G K Y A G L L L L D V 701 cCgtgtacgggctggacgatcggcgcgagctgcgcaagctcgcggacagcaccgccctcgagaaccaccagcagagcggcgcgctctcggaagccatgaa

801 ACTCAGCGACTTCCGCGCCGTCTTCGAAGGGCTGAAGAAGAGCGITCCCCTTACGAATTIGGAGATGATCAATGAGTGAGTGCCCCCCTCTTCAAGCGGT L S D F R A V F
901 CAGGCCGGGCGCGGTAGCTCGTGCCTGTCGGCACATCTCTCAGGGCAGCGAAAAGTAGGGGGGTGGGGAGACTtGAGCCCAGAACATGICGAGACCTICC
1001 tCCAGgGCACGATCGCGAGACCCCCGTtCCGAAAAGGAAACAACGAACGTtATCCCCTCCAAGAACGAGAAGAGCGTCCCCGTCCGTGTCGAGCTGCCCC CX5
1101 cctcgantttcgagagaatañ̂gagcagcaccgcacgccagcagcagaggagtagcgigcgcctgcagcctgitittatcgagccgegcttcgagcacg
1 M S S T A R O O Q R S S V R L Q P V F I E P

28 F L C G E R Y L W I V F F E A L V A L L L L R W F L
1301 gItCTCGCGTCGTCCGCGCGCGCCGGAGGACCCGTTACGTCAGATGGTCGCGGGCAAGCGCCTGGCCTGCGCGGGCGAGCGGTtGATGATCCTCGGGTIG

1401 CACGGCGGCCCCCAGGCCGCGCTGAACCTGGACGGCAGCGAGGTGAGACTGCCCAACTGCGAGGCTTTTCTGCGCGGCGCGGGGCGTGCAAACACTGACA
93 H G G P O A A L N L D G S E V R L P N C E A F L R G A G R A N T D
1501 GCGCGGGCGAGGCAGGAACGGATGCGGGGACCGGCACGGACGTGGGCGCTCTCCTTGTCTAACGCCCTCCTCTCACGCCCTCTCGAGGCGGAAAGTCGTG
127 A G E A G T D A G T G T D $V$ V G A L $L$ L $V$ *
1601 CGCGTtCGGAACGCGTGAGGAGGCGGCACTGCTAGCGACGCGgACACATtGCgaggcaccgcccgcccgcgctcgacaacgacaggaacgcaccacgecc CX6
1701 GACGGCACGGGCACGCGCGCGCCTTTCCGGCCTGIAAATIGGACCACAAGCAGTACCTGCTAACTATGTTCTICGCAGAGGACAGCTCCTICTTCAAGTAC
1 M D H K O Y L L T M F F A E D S S F
1801 ctgtcgeagcaggacgacgacacagccctggacgacgtaatgattgtcaagcactacaiggacgigctgctggccctgctagtgcgcgccaagaacaagc
22 L S E D D D D T A L D D V M I V K H Y
1901 tCGAGGCGCTTGGGCACTGCTACGAGCCGCTGTCCGAGGACtTCCGCGCGCTCtigcaigtgcgccagctgcgagaactacgccaggtgcacgaccgcgc
56 E A L G H C Y E P L S E D F R A L L H V R O L R
2001 gCtgCtgcgcctggacgcagagcccgigcaigtgagccatggctacctigcggacticgigctcagcctagtacggctggcgegcgagctgegggagctg

2101 tgCgtgccaccgcgcacgcgctacgtggacccgcgcgacgaccctacgctggcctacgtgctggagaicctgcacggcacggacgicgactctggegcgg
 CX7
2201 GAGCTTACGCTCTCGCGCGCCCGGAAGCCGAAAAAATAAGTCCTGTCCGGAGAGCGCTGCCTGGCTGCAGTCCTTCTCGCCGCCCTIAAATCGAGATGGA
156 A $Y$ A L A R P
2301 gGCCCAGGCCATGGAGCGCCCGCTGCTCTACTICCACGAGCTGACGCAGACGCAAGAGTACGACGCGGAGGTCGAGCGCGCCGCGCGCTCGCGCTICCCC
3 A O A M E R P L L Y F H E L T O T O E Y D A E V E R A A R
2401 gCacaagggcagctcangctgctcaicggcgagctctitticctgaacaagctgcaccggcgcgagatgctcgccggcaccacggtggittacatcggct
36 A O G O L K L L I G E L F F L N K L H R R E M L A G T T V V V
2501 CCGCGCCCGGCGGGCACATCCGCTACCTGGTGGAGCACtTCCGCGCGCTAGGGGTGCCGCTGCGCTGGATGCTGCTCGACGGGCGCAGCCACGACTCGCG
70 A P G G H I R Y L V E H F R A L G V P L
2601 tCTGCAAGGCCTGTCCGACGTCACGCTGGTCACGCGCTtCGTGGACGAGCGCTACCTGAIGCGCAIGCGGCAGGCGCTGCGCGGCGCGCGCGTGGtGCTG

2701 atctccgacatccgctcgcgicgcggcagcgagcccagcacggaggacctgctgtacgactacgcgttgcaaaactccatgctgagcatcctgaagcccg
 2801 tgGCgtccagtctgangiggcgctgcccgittcccgaccagtggctgcacaactictacgicgigtgcgacaaggaactgctgcagcccttcgcgecgec

2901 GTTCTCGGCAGAGCTGCGATTGCTGAGCGTGCACGCGGGCGCGCCGCGCTTGCGCTGCATCACGCTGGCGGCCGCACGCGATTATGAAAAAAAAATGTTC
203 F S A E L R L L S V H A G A P R L R C I
3001 tatctcantancgigatccggcgccgcatcgigctcancticgactacccgaaccaggagtacgactictttcacatgitccacctcctaaatacggtac
236 Y L N N V I R R R I V L N F D Y P N O E Y D F F H M F H L L N
3101 tgigtcctcgcagcticgacagccccaccaaganggtgctertcttgcagcantcgat
270 C P R S F D S P T K K V L F L O 0 S
Fig. 2. (b) For legend see facing page.
S F I K H G F R H H V R W P T L A L G V V L A
401 gCtCACAGCCGTGGAGTCCATGCCCACGCGCAAGGTCTTCCACGCCTTCGTGGAGCCCGTGCTGGAGGGCACGCTGCACATGTGCGTGCACCTCAAGAAG
$L T$ A V E S M P T R K V F H A F V E P V L
501 aCgCagaccgagggcgacgcgiacgigiccatgcacgactttgactactacgiggtgcgcgacgacggcacgctgagcaagctcaagaágcccaaggatc

601 IGCGCGAGACGCTGCTGCACAGCTTCCTGGAGTACCGGCTCAAGAACACCAAGAGCATCGAGCTGGTGGCCTTIAGCAGCGGCACGCAGATCCGCGAGGA

701 GCTGCTGACGCACCTGGCCGGGGTGCTGGACATCGAGGTCrtCACGCgCgAGCACGCCAACGTGAAGGTGACCTtCCCCGAGGAGCCTCGCAGCACGTGT
L L T H L A G V L D I E V F T R E H A N V K V V T F P E E P P R
801 CCGTTIGGCGTGATCGCGCCGCGCGGGCAGCTGCGCAICTTCTTCGAGGCCTACCCCTGGGTTGACACGCACCAGCACCTGCACGCCGTGCTGCGCCTGC
P F G V I A P R G Q L R I F F E A Y P
901 tggagaggangcttgigcccgacgtgcgcagcagccagaiccttgicacgccggaactggacttcgagggcggcgigtccaagtacgaccccgcgagccg

1001 catgctgctggiccgcgacatggtgaccatgagcatcgitaactititcggcgegcgigcgcagctcgacacctaccacgatticgacatgcgtgicgtg

1101 gacaccgagcgcticctgtccgcgctcgcggaggctitigctaccctgcgcgcgctcgictaatggtgcgcgctagagcgcgecgcacaatctcgccgec
O T E R F L S A L A E A F A T L R A L V *
CX8
1201 tgiahairggaccegctctcgatcticticetcgtggtcgecgegeccgecg
M D P L S V F F LVVAAAA
(d)

1 agtctcatggrcartittgittcgangittcacgcgigagtagcaacaangaggaaangaancccctagcagacanctgcgictcaagacaagtcacact
101 CITGItTGITGTCCCACICTCGAAAGTCGCGCICTITGCITICCTIGGIGAAGAGCCACGCCGCACTTGACAGGCACGITCTTCCGAGACTGTCAAGIGT CXII
201 CCGGCTCCAAAGTGAAAAAGAACTICTCTTAGCACGGCACGCCCAAACTGCGGGCATGAAGGTGTACGCGGTGACGGGTGGCGGTGGCTICATCGGCAGC $M \quad K \quad V \quad Y \quad A \quad V \quad T \quad G \quad G \quad G \quad G \quad F \quad I \quad G \quad S$
301 tacattgtacgcgcactgctgcaatgrgagcgcacgctcattgagctgcgcgigatigacgtccgatggggacaaabictcctcgcggancgigancg Y I V R A L L Q C E R T L I E 401 tggictacatctactgcgacgigtgcgacactgcgcgcctgigtgccgcgctcgagggagtggacgtgctcaiccacactgcggggctagicgacgtgat

501 gggggagtatagcgaggacgagaittaccgcgcgancgigcacgggacacacagcgcgcttagcgcctgcgictgcgcgggcgigcgctitgiggtgiac

601 accagcagtatggaggicgitggcccgaicatgcgcgcggagcccticgitgccgacgagaiaaccgagtacgagicttgccaccaccactgctacccgc
 701 gCagCanggcggangccgaggagctagigctcagticcaacgggcgccgagtacgcgggggicagcgcatgctracatgcgcgttgcgcccgccgggtgi S K A E A E E L V L S S N G R R V $\quad$ R G G O R M L
801 CTACGGTGAGGGCAACCAGCTGCTGCTGCGGCTAGCAAAGAACTACGTGCGCATGGGCTTGCACGTGCCACGTACCGTGTGCGAGAACGCTCTGCAAAGC

901 agGgTCTAcgitggcaacgiggcctggatgcacgtactigccgcacgcgcgctgcaggaiccggactcgcgcctgccggecaacgcatatitctgtiacg

1001 accactctccgigcatggactacgangctttcantgtgatgctcttacgctcgitcggcgiggagctgggcggtccgcggctcccgcgcgctitactgac

1101 CGTGGCGGCGTATACCAACGCCGCACTGCAGTGGCTGCTCCGCCAGCTGGGCATCCGCTTCTCGCCTCTGCTCAACCCTTACACGCTCGCCGITGCCAAT

1201 gCCTGCTTCGTCATACGCACGCGCAAGGCACGCGAGCACATGGGCTACGAGCCGATCCACAACTGGAAGCAGTCGCGCAAAAACACCACGCGATGGCTGC A C F V I R T R K A R E H M G Y E P I
1301 gCtCGCAGCTCGCGAGCTAACTGCACCTTGCTAGAGCGAGCGCGGGCGGCGGACACTGGCACCGGAAGCATCGACAAGGAAAGGCACCCACCAGGAACGA $S$ a L A $\mathrm{S}^{\text {* }}$
1401 AAGTCGCTACCAAGAACAGAGAAAGGCACCCACCAGGAACAGGGAACAATCCCCCGCCAGGAACGAGGAAAGGCACACCCCTGTAACAAGGAAAAGGCAC
1501 accecg
Fig. 2. MCV-I DNA sequences with deduced amino acid sequences below. Potential promoter sequences are underlined. (a) 1850 bp of MCV-I DNA spanning the left end of BamHI L and the entire BamHIN fragment, reading from right to left in the genome. (b) 3158 bp from the right end of MCV-I Clal F, reading from left to right. (c) 1251 bp from the left end of MCV-I HindIll I and right end of HindIII E, reading from right to left in the genome. (d) 1506 bp from the right end of Hindlll $N$ and the left end of Hindlll $M$, reading from left to right in the MCV-I genome.
(a)

MCV CXI MASA-ALSTLYGAFVARYLRKLSLYSTINSVTCAI HVGRIVGTLQNCSVRILNRCNNHDQLSFRLLLEAFAETVNLLPPKORAEIAAQVGVDLEAASHEE WV F9L MAETKEFKTLYNLFIDSYLOKLAQHSIPTNVTCAIHIGEVIGQFKNCALRITNKCMSNSRLSFTLMVESFIEVISLLPEKDRRAIAEEIGIDLDDVPSAV ** . . *** *. **.**. * .******.* ..* ...*...** *.* .* .*** *.....* * ...****.* ** ..*.**.

MCV CXI SRLERKCRAHAALVQSIDVQTLNVGTCIAPPGRSLAIQVVNSGS
VV F9L SKLEKNCNAYAEVNNIIDIQKLDIGECSAPPGQHMLLQIVNTGSAEANCGLOTIVKSLNKIYVPPIIENRLPYYDPWFLVGVAIILVIFTVAICSIRRNL *.**..* * * . . **.* *..******. . .*.**.**

MCV Cx1
W F9L ALKYRYGTFLYV
kinase
MCV CX2 MAFSDSASADAPUSAVPAPRRDETTVLGDEIYFNYVYGQLELSDSHIPHVRMLRYFRNFSRAALLRIASTEYVNPSYFQOKDKRFAPVNNDFYHLSTGGY W F1OL. MGVANDSSPEYQW--MSPHRLSDTVILGDCLYFNNIMSQLDLHONWAPSVRLLNYFKNFNRETLLKIEENDYINSSFFQQKDKRFYPINDDFYHISTGGY *. ....... * ....* .*..*** .*** . .**.* ..****.***.**.* .**.* ....*.*.******** *.*.****.*****

MCV CX2 GIVFRVEEYVVKFVFEPGSQFHPMDLTSEYTVPRFLYNNLRGDERLLUVRALAMGLNYXIGFLYTLYKRVLHMVLLLARILDGGPLSLAYSRRQVAKLFA W F10L GIVFKIDNYVVKFVFEATKLYSPMETTAEFTVPKFLYNNLKCDEKKL IVCAUAMGLNYKLTFLHTLYKRVLHMLLLLIOTMDGOELSLRYSSKVFLKAFN

MCV CX2 ERKOSAKFVRLLSYFYPAVIKSNLNVINHFGHMIHFFEHEKRANYTYORGNIIVFPLARCSAEXVTAANCAEFGFASVVHYVKFLFLOMALLYIKIYELS W F10L ERKDSIKFVKLLSHFYPAVINSNINVINYFNRMFHFFEHEKRTKYEYERGNIIIFPLALYSADKVDTELAIKLGFKSLVQYIKFIFLQMALLYIKIYELP


MCV CX2 C-HNFIHVDLKPDNILLFDSEREMRIHVGERSYVFREPVRSALNDFDFSQVSEIPNKKITASLRVEQNWFYDFHFFVHTLLKVYPELERDAAWSKALGEF W F1OL CCDNFLHADLKPDNILLFDSNEPIIIHLKDKKFVFNERIKSALNDFDFSQVAGIINKKIKNNFKVKHNWYYDFHFFVHTLLKTYPEIEKDIEFSTALEEF


MCV CX2 LVCCNRNTCEKFRLRVRRLHPISFLVRFVARDLFSDWINGERRP
WV F10L IMC-TKTDCDKYRLKVSILHPISFLEKFIMRDIFSDWINGGN--
..* . *.*.**.* ******* .*. **.*******
(b)

VP8
MNALFENLFDEDAVC-AGSVSR-EDFLLVVAGAKVKFPRSLLSMYRVVPRTMSRYELALVQSETVTGVVFTTVYNVRRNLGLEEREALSLPALEKYYLDK MSLLLENL IEEDTIFFAGSISEYDDLQMVIAGAKSKFPRSMLSIFNIVPRTMSKYELELIHNENITGAMFTTMYNIRNNLGLGD-DKLTIEAIENYFLDP


MCV CX4 ANDVLILMVNNTNLEHIAAYRMRSRRLLNPVVFRAGAVPLALVFTSRKKLSIYREDTSQAAEDSTYTKIAANVALAGKYAGLLLLDVHTPGTALMLTAVY WV L4R NNEVMPLIINNTDMTAVIPKKSGRRKNKNMVIFROGSSPILCIFETRKKINIYKENMESASTE--YTPIGDNKALISKYAGINVLNVYSPSTSMRLNAIY *.*. *...**.. . . . *. * *.** *. *. .* .***...*.*. *. . ***. *** .****. .*.* .*.*.. **.*

MCV CX4 GLDORRELRKLADSTALENHQQSGALSEAMKLSDFRAVFEGLKKSVPLTNLEMINE WV L4R GFTNKNKLEKLSTNKELESYS-SSPLQEPIRLNDFLGLLECVKKNIPLTDIP-TKD *. .. ***. . **. *..**...*.** ...* .**...***..

MCV CX5 MSSTAROOQRSSVRLQPVFIEPRFEHAFLCGERYLWIVFFEALVALLLLRUFLGSELRAAFSRRPRAPEDPLROMVAGKRLACAGERLMILGLH--GGPG WV LSR M-----ENVPNVYFNPVFIEPIFKHSLLSVYKHRLIVLFEVFVVFILIYVFFRSELNMFFMPKRKIP-DPIDRLRRAN-LACEDDKLMIYGLPLMTTQT * .. ..* ..****** **... . **.** .* ..*. *. *** * . . ***. .. .. *** ...*****

```
MCV CX5 AALNLDGSEVRLPNCEAFLRGAGRANTDSAGEAGTDAGTGTDVGALLV
WV LSR SALSINSKPIVYKDCAKLLRSIN-........-GSQPVSLNDV--LRR
    **.... . .* .**. *... . ** *
```

MCV CX6 MDHKOYLLTMFFAEDSSFFKYLSEQDDDTALDDVMIVKHYMDVLLALLVRAKNKLEALGHCYEPLSEDFRALLHVRQLRELRQVHDRALLRLDAEPVHVS
WV J1R MDHNQYLLTMFFADDDSFFKYLASQDDESSLSDILQITQYLDFLLLLLIOSKNKLEAVGHCYESLSEEYRQLTKFTDSQDFKKLFNKVPI-VTDGRVKLN

MCV CX6 HGYLADFVLSLVRLARE--LGELCVPPRTRYVDPRDDPTLAYVLEILHGTDVDSGAGAYALARPEAEKISPVRRALPGCSPSRRP
WV J1R KGYLFDFVISLMRFKKESSLATTAIDP-IRYIDPRRDIAFSNVMDILKSNKVNNN
*** ***.**.*. .* *. . * **.*** *.... *...** . *..

Fig. 3. ( $a, b$ ) For legend see facing page.


MCV CX11 MGLHVPRTVCENALQSRVYVGNVAMMHVLAARALQEPDSRLPGNAYFCYDHSPCMOYEAFNVMLLRSFGVELGGPRLPRALLTVAAYTNAALQULLRQLG WV A4OL HGNIMYRTVDDNAVHSRVYVGNAAWHVLLAAKYIQYPGSKIKGNAYFCYDYSPSCSYDMFNLLLMKPLGIE-QGSRIPRWMLKMYACKND-----MKRIL * . *** . **...***************. .* *.*.. ******** ** *.**..*....*.**.*.** .*.*** ....

MCV CX11 IRFSPLLNPYTLAVANACFVIRTRKAREHMGYEPIHNWKOSRKNTTRWLRSQLAS
WV A4OL FRKPSLLNNYTLKISNTTFEVRTNNAELDFNYSPIFDVOVAFKRTRKWL--EESE
.* ..*** *** ..*. * .** .* . *** . . ** .** . .
Fig. 3. Amino acid alignments of the translated sequences from Fig. 2 and their W counterparts. MCV-I coding sequences of $<100 \mathrm{bp}$ have not been included in the alignments. Asterisks indicate identical amino acid residues and dots similar amino acid substitutions. Dashes represent deletions. MCV-I CX1 and CX7 are partial ORFs. Amino acid alignments were performed using ClustalV. The conserved catalytic domains of VV protein kinase 2 are shown in (a).
particularly interesting. A $3 \beta$-HSD gene has also been reported in FPV (Skinner et al., 1994) and its presence in three distinct genera suggests a fundamental role for this enzyme in poxvirus replication in vivo.

MCV has a restricted host range and cell tropism and is less pathogenic than the orthopoxviruses. Genes responsible for host range and pathogenicity are generally located near the termini of poxvirus genomes and sequencing of MCV terminal regions (Bugert et al., 1993, and our unpublished data for HindIII fragments $\mathrm{C}, \mathrm{H}$ and K ) indicates that the terminal regions of MCV and VV differ considerably in coding potential, suggesting that the determinants of restricted host range and limited pathogenicity of MCV are located there.

The development of a system to culture MCV will aid the functional analysis of these putative MCV proteins and may reveal the presence of novel and biologically interesting genes in this unusual virus.

We wish to thank the Advanced Biotechnology Centre at Charing Cross and Westminster Medical School for use of the automated sequencing facilities and Dr Colin Porter (Institute of Child Health) for providing plasmid pUC1318.

## References

Amegadzie, B. Y., Ahn, B.-Y. \& Moss, B. (1991). Identification, sequence, and expression of the gene encoding a Mr 35,000 subunit of the vaccinia virus DNA-dependent RNA polymerase. Journal of Biological Chemistry 76, 13712-13718.
Blake, N. W., Porter, C. D. \& Archard, L. C. (1991). Characterization of a Molluscum contagiosum virus homolog of the vaccinia virus p 37 K major envelope antigen. Journal of Virology 65, 3583-3589.
Binns, M. M., Boursnell, M. E. G. \& Skinner, M. A. (1992). Gene translocation in poxviruses: the fowlpox virus thymidine kinase gene is flanked by 15 bp direct repeats and occupies the locus which in vaccinia virus is occupied by the ribonucleotide reductase large subunit gene. Virus Research 24, 161-172.
Bugert, J., Rosen-Wolff, A. \& Darai, G. (1989). Genomic characterization of Molluscum contagiosum virus type $I$ : identification of the repetitive DNA sequences in the viral genome. Virus Genes 3, 159-173.
Bugert, J. J., Raab, K., Rosen-Wolff, A., Janssen, W. \& Darai, G. (1993). Determination of the position of the boundaries of the terminal repetitive sequences within the genome of Molluscum contagiosum virus type I by DNA nucleotide sequence analysis. Virology 192, 391-396.
Buller, R. M. L., Burnett, J., Chen, W. \& Kreider, J. (1995). Replication of Molluscum contagiosum virus. Virology 213, 655-659.
Cotton, D. W. K., Cooper, C., Barrett, D. F. \& Leppard, B. J. L. (1987). Severe atypical Molluscum contagiosum infection in an immunocompromised host. British Journal of Dermatology 116, 871-876.
Davison, A. J. \& Moss, B. (1989a). Structure of vaccinia virus early promoters. Journal of Molecular Biology 210, 749-769.
Davison, A. J. \& Moss, B. (1989 b). The structure of vaccinia virus late promoters. Journal of Molecular Biology 210, 771-784.

Drillien, R., Spehner, D., Villeval, D. \& Lecocq, J.-P. (1987). Similar genetic organization between a region of fowlpox virus DNA and the vaccinia virus HindIII J fragment despite divergent location of the thymidine kinase gene. Virology 160, 203-209.
Gershon, P. D., Ahn, B.-Y., Garfield, M. \& Moss, B. (1991). Poly(A) polymerase and a dissociable polyadenylation stimulatory factor encoded by vaccinia virus. Cell 66, 1269-1278.
Goebel, S. J., Johnson, G. P., Perkus, M. E., Davis, S. W., Winslow, J. P. \& Paoletti, E. (1990). The complete DNA sequence of vaccinia virus. Virology 179, 247-266.
Hadasch, R. P., Bugert, J. J., Janssen, W. \& Darai, G. (1993). Characterization of the genome of Molluscum contagiosum virus type I between the genome coordinates 0.045 and 0.075 by DNA nucleotide sequence analysis of a $5 \cdot 6-\mathrm{kb}$ HindiII/MluI DNA fragment. Intervirology 36, 32-43.
Hruby, D. E. \& Ball, L. A. (1982). Mapping and identification of the vaccinia virus thymidine kinase gene. Journal of Virology 43, 403-409.
Johnson, G. P., Goebel, S. J. \& Paoletti, E. (1993). An update on the vaccinia virus genome. Virology 196, 381-401.
Kay, R. \& McPherson, J. (1987). Hybrid pUC vectors for the addition of new restriction enzyme sites to the ends of DNA fragments. Nucleic Acid Research 15, 2778.
Lin, S. \& Broyles, S. S. (1994). Vaccinia protein kinase 2: a second essential serine/threonine protein kinase encoded by vaccinia virus. Proceedings of the National Academy of Sciences, USA 91, 7653-7657.
Mitchell, J.C. (1953). Observations on the virus of Molluscum contagiosum. British Journal of Experimental Pathology 34, 44-49.
Moore, J. B. \& Smith, G. L. (1992). Steroid hormone synthesis by a vaccinia enzyme : a new type of virus virulence factor. EMBO Journal 11, 1973-1980.
Porter, C. D. \& 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. Journal of General Virology 68, 673-682.
Porter, C. D. \& Archard, L. C. (1992). Characterization by restriction mapping of three subtypes of Molluscum contagiosum virus. Journal of Medical Virology 38, 1-6.
Postlethwaite, R. (1970). Molluscum contagiosum: a review. Archives of Environmental Health 21, 432-452.
Skinner, M. A., Moore, J. B., Binns, M. M., Smith, G. L. \& Boursnell, M. E. G. (1994). Deletion of fowlpox virus homologues of vaccinia virus genes between the $3 \beta$-hydroxysteroid dehydrogenase (A44L) and DNA ligase (A50R) genes. Journal of General Virology 75, 2495-2498.
Sonntag, K.-C., Clauer, U., Bugert, J. J., Schnitzler, P. \& Darai, G. (1995). Identification and properties of the genes encoding the poly(A) polymerase and a small ( 22 kDa ) and the largest subunit ( 146 kDa ) of the DNA-dependent RNA polymerase of Molluscum contagiosum virus. Virology 210, 471-478.
Yang, W.-P., Kao, S.-Y. \& Bauer, W. R. (1988). Biosynthesis and posttranslational cleavage of vaccinia virus structural protein VP8. Virology 167, 585-590.

[^1]
[^0]:    Author for correspondence: Leonard C. Archard.
    Fax +44181846 7099. e-mail RabcbOO@cxwms.ac.uk
    tPresent address: CRC Institute for Cancer Studies, University of Birmingham Medical School, Birmingham, UK.

    The GenBank accession numbers of sequences reported in this paper are: U32423 (region CX1-CX2); U32424 (region CX3-CX7); U32425 (region CX8-CX10); U32426 (region CX11).

[^1]:    Received 4 March 1996; Accepted 13 August 1996

