Journal of General Virology (2002), **83**, 855–872. Printed in Great Britain

The sequence of camelpox virus shows it is most closely related to variola virus, the cause of smallpox

Caroline Gubser[1,2] and Geoffrey L. Smith[1,2]

[1] Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK

[2] Department of Infectious Diseases, Division of Investigative Science, Faculty of Medicine, Imperial College, St Mary's Campus, Norfolk Place, London W2 1PG, UK

Camelpox virus (CMPV) and variola virus (VAR) are orthopoxviruses (OPVs) that share several biological features and cause high mortality and morbidity in their single host species. The sequence of a virulent CMPV strain was determined; it is 202 182 bp long, with inverted terminal repeats (ITRs) of 6045 bp and has 206 predicted open reading frames (ORFs). As for other poxviruses, the genes are tightly packed with little non-coding sequence. Most genes within 25 kb of each terminus are transcribed outwards towards the terminus, whereas genes within the centre of the genome are transcribed from either DNA strand. The central region of the genome contains genes that are highly conserved in other OPVs and 87 of these are conserved in all sequenced chordopoxviruses. In contrast, genes towards either terminus are more variable and encode proteins involved in host range, virulence or immunomodulation. In some cases, these are broken versions of genes found in other OPVs. The relationship of CMPV to other OPVs was analysed by comparisons of DNA and predicted protein sequences, repeats within the ITRs and arrangement of ORFs within the terminal regions. Each comparison gave the same conclusion: CMPV is the closest known virus to variola virus, the cause of smallpox.

Introduction

Poxviruses are complex viruses that replicate in the cytoplasm and encode many enzymes and immunomodulatory proteins (Moss, 1996). They are classified into vertebrate (*Chordopoxvirinae*) and insect (*Entomopoxvirinae*) subfamilies. Chordopoxviruses (ChPVs) are subdivided into eight genera and members of six of these have been sequenced: These are vaccinia virus (VV) strains Copenhagen (COP) (Goebel *et al.*, 1990), Tian Tan (accession no. AF095689), modified virus Ankara (MVA) (Antoine *et al.*, 1998) and Western Reserve (Smith *et al.*, 1991) (and references therein), variola virus (VAR) strains Bangladesh-1975 (BSH) (Massung *et al.*, 1993a, 1994), India-1967 (IND) (Shchelkunov *et al.*, 1995) and Garcia-1966 (Shchelkunov *et al.*, 2000), myxoma virus (MYX) strain Lausanne (Cameron *et al.*, 1999), Shope fibroma virus (SFV) strain Kaza (Willer *et al.*, 1999), molluscum contagiosum virus (MCV) (Senkevich *et al.*, 1997), fowlpox virus (FPV) (Afonso *et al.*, 2000), lumpy skin disease virus (Tulman *et al.*, 2001) and Yaba-like disease virus (YLDV) (Lee *et al.*, 2001). In addition, the sequence of 50 kb from each end of the genome of cowpox virus (CPV) strain GRI-90 (Shchelkunov *et al.*, 1998), and parts of the Yaba monkey tumour virus (accession nos AB025319, AB018404 and AB015885), swinepox virus (Massung *et al.*, 1993b) and ectromelia virus (Chen *et al.*, 2000) genomes have been reported.

Camelpox virus (CMPV) is a poorly characterized orthopoxvirus (OPV) that causes a severe and economically important disease in camels, especially young animals (Mc Grane & Higgins, 1986). The discovery of CMPV caused concern during the WHO-sponsored smallpox eradication campaign due to its description as smallpox-like (Baxby, 1972). Both CMPV and VAR, the cause of smallpox, cause a systemic illness in a single host species, form a small, white pock on the chorioallantoic membrane of a fertile hen's egg, have a similar ceiling temperature for growth, and are restricted for replication in rabbit skin (Fenner *et al.*, 1989). Despite these similarities, CMPV and VAR are distinguishable (Bedson,

Author for correspondence: Geoffrey Smith (at Imperial College). Fax +44 207 594 3973. e-mail glsmith@ic.ac.uk

Accession numbers: AY009089 and AY037935.

1972; Baxby, 1974; Esposito & Knight, 1985) and CMPV has rarely, if ever, caused disease in man (Jezek *et al.*, 1983). Likewise, VAR is unable to cause disease in camels, although camels immunized with VAR are resistant to subsequent infection with CMPV (Baxby *et al.*, 1975).

Smallpox was eradicated by vaccination. Originally cowpox virus (CPV) was the vaccine, but vaccinia virus (VV) (Fenner *et al.*, 1988), a virus of unknown origin (Baxby, 1981), was used in the 20th century. CMPV, VAR, CPV and VV are all OPVs, a ChPV genus that also includes monkeypox virus and ectromelia virus (Fenner *et al.*, 1989). To increase our understanding of CMPV and of OPV phylogeny we sequenced the genome of CMPV strain CMS (CMPV-CMS), a virulent virus isolated in 1970 from Iran (Baxby, 1972), and also the termini of CMPV strain 903 (CMPV-903) (Douglass & Dumbell, 1996) isolated from Somalia. Hitherto, CMPV sequence data were restricted to a few individual genes (Binns, 1992; Meyer & Rziha, 1993; Douglass & Dumbell, 1996).

Analyses of the CMPV genome sequence, the arrangement of open reading frames (ORFs), the protein sequences and the nature of the repeats within the inverted terminal repeats (ITRs) all showed that CMPV was more closely related to VAR than to any other virus.

Methods

■ **Virus and cells.** Human TK$^-$143 cells and monkey kidney BS-C-1 cells were grown as described (Mathew *et al.*, 2001). CMPV-CMS and CMPV-903 were kindly provided by Keith Dumbell (Cape Town, South Africa). Viruses were plaque purified twice on monolayers of BS-C-1 cells and virus stocks were grown in TK$^-$143 cells.

■ **Virus purification and DNA extraction.** CMPV was purified from the cytoplasm of infected TK$^-$143 cells by sedimentation in two successive sucrose density gradients (Mackett *et al.*, 1985) and DNA was extracted from purified virions as for other OPVs (Esposito *et al.*, 1981).

■ **Construction of CMPV DNA libraries and shotgun sequencing.** The CMPV DNA sequence was determined using the random shotgun sequencing method (Bankier *et al.*, 1987). Virus DNA was sheared by sonication and DNA fragments of 2–5 kb and 5–10 kb were cloned into pUC118 (Yanisch-Perron *et al.*, 1985). Plasmid DNA was extracted using a Qiagen Biorobot 9600. The DNA library containing fragments of 2–5 kb was used for most sequencing, whereas fragments of 5–10 kb were useful for gap filling. DNA was sequenced with M13 forward and reverse universal primers or specially designed oligonucleotides (BioLabs) on an Applied Biosystems model 373 Sequencer using the cycle sequencing method with fluorescent dye terminators and AmpliTaq DNA polymerase FS (ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction, Perkin-Elmer). Applied Biosystems sequence software was used for lane tracking and trace extraction and data were transferred to UNIX workstations for all further processing.

■ **Genomic DNA assembly.** Raw data were processed using the program Pregap4, including Phred (Ewing & Green, 1998) and Phrap (Ewing *et al.*, 1998). A consensus sequence was produced and edited on the graphical interface of Gap4 (Bonfield *et al.*, 1995; Bonfield & Staden, 1996). Oligonucleotide primer design, genomic DNA composition, inverted repeats and restriction enzyme patterns were determined with the Wisconsin Genetic Computer Group (GCG) program (Devereux *et al.*, 1984).

■ **Bioinformatic analysis.** ORFs were identified with NIP4 software (Staden & McLachlan, 1982). Protein comparisons with sequence databases and amino acid sequence analyses were processed and viewed using PIX (http://www.hgmp.mrc.ac.uk/Registered/Webapp/pix/). Related proteins were aligned with CLUSTALW (Thompson *et al.*, 1994) and were edited with GeneDoc (Nicholas & Nicholas, 1997). Phylogeny studies were carried out using the maximum likelihood analysis program PUZZLE (Strimmer & von Haeseler, 1996, 1997) version 5.0 (VT model of substitution) (Müller & Vingron, 2000) and the PHYLIP package version 3.5 (Felsenstein, 1989) using the programs SEQBOOT, PROTDIST, NEIGHBOR and CONDENSE. Phylogenetic trees were viewed with TREEVIEW (Page, 1996). OPV genome sequences were aligned using DOTTER (Sonnhammer & Durbin, 1995).

Results and Discussion

Genome sequence

Raw sequence data (1641717 nucleotides) were assembled into a 202182 bp contiguous sequence (average density of 8·12 readings per nucleotide). The entire sequence was read on both strands and had a base composition of 66·9% A + T. The genome was slightly longer than reported but the calculated *Hin*dIII restriction map (Fig. 1a) was consistent with previously published maps (Esposito & Knight, 1985) except for the 718 bp U fragment that was missed previously. The sequence obtained is predicted to extend very close to the terminal hairpins because: (i) the size of terminal restriction fragments determined by *Hin*dIII or *Eco*RI digestion followed by denaturation, snap-back hybridization and gel electrophoresis showed these to be indistinguishable from those predicted from nucleotide sequence by computer; and (ii) the sequence contains the nucleotide motif 5′ TTTTTTTCTAGACACTAA-AT 3′ that is identical to the sequence present in VV and that is needed for the resolution of concatemeric DNA replication intermediates. In sequenced OPVs this motif is present very close the terminal hairpin. Compared to other OPVs (Mackett & Archard, 1979; DeFilippes, 1982) the CMPV genome has a distinctive *Hin*dIII restriction map and is clearly a separate OPV species.

Inverted terminal repeats

The OPV genome consists of a linear dsDNA molecule with covalently linked termini and ITRs. ITRs studied previously contain two unique sequences, non-repeated (NR) 1 and 2, and one or more blocks of tandem repeats. Two sets of repeats are present in the ITRs of VV-COP (Goebel *et al.*, 1990), CPV-GRI (Shchelkunov *et al.*, 1998) and RCN (Parsons & Pickup, 1987) but not in most VAR strains (Massung *et al.*, 1995) (see Fig. 4). NR1 is adjacent to the terminal hairpin loop and contains the motif essential for the resolution of DNA concatemeric replication intermediates (DeLange *et al.*, 1986; Merchlinsky & Moss, 1986). In contrast, NR2 is located after the first set of repeats and its function remains unknown.

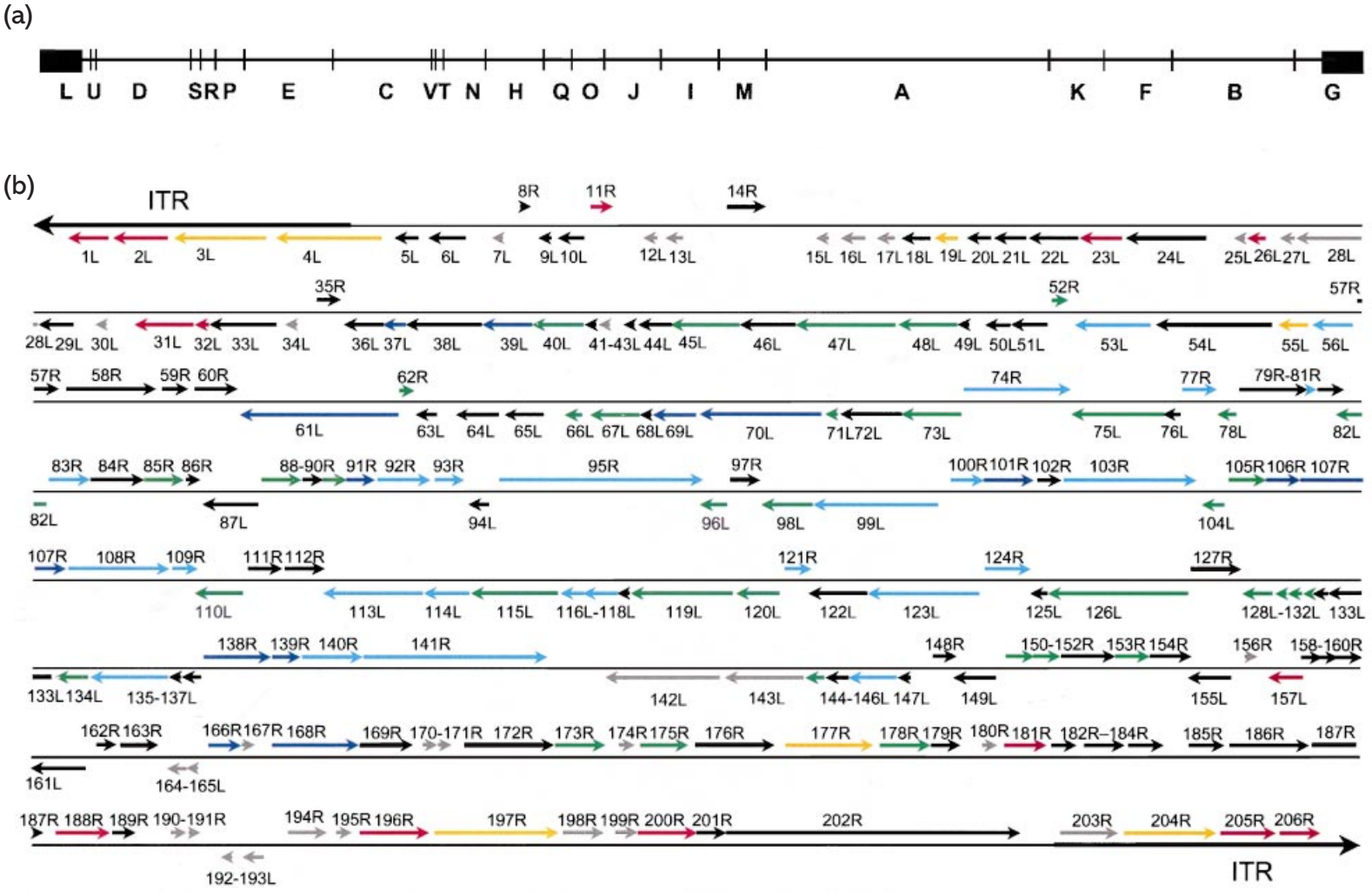


Fig. 1. (a) *Hin*dIII restriction map of the CMPV genome predicted for DNA sequence. Fragments are lettered A to U in decreasing size according to existing convention (Mackett & Archard, 1979). ITRs are indicated with open boxes. (b) ORF map of the CMPV genome. ORFs are represented by coloured arrows (green, assembly/structural; light blue, RNA metabolism; dark blue, DNA metabolism; yellow, host range; red, immunomodulators; black, unknown; grey, gene fragments) and are numbered from the left to the right end of the genome based on the position of the first methionine codon (Table 1). They were assigned the designation L (left) or R (right) to represent the direction of transcription. The ITRs are indicated by horizontal bars at each end of the genome. Each line represents just under 26 kb.

The CMPV-CMS ITR is 6045 bp and encodes protein to within 650 bp of the terminus. This situation is unlike VV (Goebel *et al.*, 1990) and CPV (Shchelkunov *et al.*, 1998) but very similar to VAR (Massung *et al.*, 1994; Shchelkunov *et al.*, 1995, 2000). In the terminal 630 bp, there is a single block of tandem repeats consisting of three 70 bp units, followed by a 52 bp incomplete unit and a 27 bp related sequence. However, no equivalent of NR2 was identified within the CMPV genome showing that NR2 is not essential for OPV replication.

General arrangement of ORFs

Translation of the CMPV DNA sequence identified 206 ORFs $\geqslant$ 195 nucleotides starting with ATG or smaller ORFs that are conserved in other ChPVs (Fig. 1b, Table 1). Two of the smaller ORFs are the RNA polymerase subunit rpo7 (Amegadzie *et al.*, 1992) and the IMV membrane protein A14.5L (Betakova *et al.*, 2000). In addition, there are 61 minor ORFs that fall wholly or partly within larger ORFs (data not shown). Major ORFs have an average length of 964 nucleotides, encode proteins of 53 to 1869 amino acids and 101 are transcribed leftwards and 105 rightwards.

CMPV has a genome organization very similar to other ChPVs. Protein coding sequences are contiguous and are on both DNA strands. ORFs are tightly packed and there is little noncoding DNA. Some ORFs towards the middle of the genome are even slightly overlapping. Blocks of ORFs are transcribed in the same direction, most notably at the ends of the genome (Fig. 1b).

The CMPV ITRs contain three ORFs that are consequently diploid (1L/206R, 2L/205R and 3L/204R). A fourth ORF crosses the left ITR and represents a complete version of the CPV-GRI D4L ORF (Shchelkunov *et al.*, 1998). Within the right ITR, ORF 203R represents the C-terminal 376 amino acids of 4L. Upstream of the initiating methionine of 203R an additional 85 codons are conserved until the ITR internal boundary is reached. This suggests that during CMPV evolution, DNA was copied by a terminal transposition event from the left of the genome to the right, rather then the converse. Other poxviruses also contain ORFs that cross only one ITR (Willer *et al.*, 1999).

Table 1 shows CMPV major ORFs, their nucleotide coordinates, number of amino acids and the best protein matches in public databases. Most proteins have a putative function based on amino acid similarities with other viral or cellular proteins. The general arrangement of CMPV ORFs is collinear with other OPVs. Fig. 1(b) illustrates a genetic map of the CMPV genome with predicted functions represented using a colour code. The central region encodes mainly proteins involved in DNA and RNA metabolism (dark and light blue) or in virion assembly or structure (green). In contrast, the terminal regions encode proteins with known or predicted virulence (red) or host range functions (yellow). ORFs encoding proteins of unknown function are shown in black and incomplete ORFs in grey. The latter are fragments of larger ORFs in other OPVs.

Minimal gene complement

Computational analyses identified 87 genes that are conserved in all sequenced ChPVs (Table 1, asterisks) and we define these as the ChPV minimal gene complement. These lie within the central genomic region (107 kb) between CMPV 44L (VV-COP F9L) and CMPV 151R (A34R) and mostly encode proteins for RNA transcription, DNA replication or virion structure (Table 1) that are essential for virus replication. Conversely, genes within the terminal regions are non-essential and encode proteins involved in host range, virulence or immunomodulation (Fig. 1b). These non-essential genes probably were acquired later during poxvirus evolution. Consonant with this hypothesis, whereas the average A + T content of the genome is 66·9%, 6/11 ORFs with an A + T content of less than 59% are clustered near the termini (1L/206R, 2L/205R, 5L and 6L) and several of these encode immunomodulators (see below). In addition, CMPV terminal ORFs have a codon usage that differs from the average codon usage for the whole genome (data not shown). Collectively, this suggests these terminal genes were acquired more recently.

Gene fragments

Within 60 kb at either end of the CMPV genome, 33 ORFs are incomplete versions of 23 larger ORFs found in other poxviruses due to frameshift or nonsense mutations. The role, if any, of gene fragments is unknown and many are unlikely to be expressed. The retention of so many gene fragments in OPV genomes is intriguing and in CMPV these represent 16% of all CMPV ORFs. Incomplete ORFs are also found in other OPVs (Smith *et al.*, 1991; Aguado *et al.*, 1992). To compare fragmented genes, 94 ORFs present within the terminal genomic regions of VV-COP, VAR-BSH, CPV-GRI or CMPV are listed in Table 2. ORFs are named as in VV-COP where possible. Complete ORFs are marked with a tick, fragmented ORFs with the letter F, deleted ORFs with a horizontal dash, and the 26 ORFs that are present in all four OPVs are highlighted. Although conserved in these viruses, some of the latter genes are non-essential for virus replication (Perkus *et al.*, 1991). CPV contains the greatest number of complete ORFs in these terminal regions. Comparing fragmented genes in VAR-BSH and CMPV, there are 20 CMPV ORFs that are absent or broken in VAR, while 11 VAR ORFs are broken in CMPV (Table 3).

ORFs specific to CMPV

Hitherto, CPV contained all known OPV genes from terminal genomic regions, suggesting CPV was the closest virus to the ancestral OPV (Shchelkunov *et al.*, 1998). However, five ORFs (8R, 9L, 182R, 183R and 184R) are unique to CMPV. ORFs 8L and 9L are short (65 and 81 amino acids, respectively) and are absent from other OPVs due to

Table 1. Major ORFs encoded by CMPV

Grey ORFs represents CMPV fragmented ORFs. CMPV ORFs marked with an asterisk are part of the ChPV minimal gene complement. CPV, cowpox virus strain GRI-90; VV, vaccinia virus strain Copenhagen; WR, Western Reserve; VAR, variola virus strain Bangladesh-1975; CAPV, capripox virus.

ORF†	Translation‡ Start	Translation‡ Stop	Size (aa)	Matches in public protein databases§	Putative function/structure	Amino acid identity‖
1L	1420	653	255	VV Lister 35 kDa protein	35 kDa secreted chemokine-binding protein	237/246 (96)
2L	2596	1547	349	CPV D2L	Soluble TNF receptor, CrmB	311/351(94)
3L	4441	2684	585	CPV D3L	Ankyrin-like protein	543/585 (92)
4L	6687	4669	672	CPV D4L	Ankyrin-like protein	620/672 (92)
5L	7308	6847	153	VV C16L	Unknown	143/153 (94)
6L	8260	7547	237	CPV S1R	Unknown	195/205 (91)
7L	8938	8729	69	CPV K3R f	Soluble TNF receptor, CrmE, fragment	67/69 (97)
8R	9233	9430	65		Unknown, histidine-rich protein	
9L	9861	9568	81		Unknown	
10L	10417	9911	168	CPV C4L	Similarity to VV C7L	157/170 (92)
11R	10554	10973	139	VV C11R	Epidermal growth factor	132/142 (92)
12L	11401	11123	92	VV C10L f2	Unknown, fragment	75/92 (81)
13L	11883	11545	112	VV C10L f1	Unknown, fragment	101/105 (96)
14R	12612	13340	242	VAR D6R	RING zinc-finger motif	235/242 (97)
15L	14682	14434	82	VV C9L f3	Ankyrin-like protein, fragment	64/79 (81)
16L	15436	14981	151	VV C9L f2	Ankyrin-like protein, fragment	93/150 (62)
17L	16053	15715	112	VV C9L f1	Ankyrin-like protein, fragment	81/99 (81)
18L	16720	16172	182	VV C8L	Unknown	171/184 (94)
19L	17244	16792	150	VV C7L	Host range protein	139/150 (92)
20L	17944	17474	156	VV C6L	Unknown	146/150 (97)
21L	18699	18097	200	VV C5L	Unknown	184/193 (95)
22L	19719	18772	315	VV C4L	Unknown	304/316 (96)
23L	20583	19786	265	VV C3L	Complement control protein	256/265 (96)
24L	22189	20651	512	VV C2L	Kelch-like protein	499/512 (97)
25L	22890	22681	69	VV C1L f	Unknown, fragment	66/69 (95)
26L	23311	22958	117	VV N1L	Virulence factor, secreted protein	111/117 (94)
27L	23947	23663	94	VV N2L f	Alpha-amanitin sensitive protein, fragment	82/89 (92)
28L	25316	23976	419	VV M1L f	Ankyrin-like protein, fragment	405/419 (95)
29L	26023	25361	220	VV M2L	Unknown	212/220 (96)
30L	26679	26476	67	VV K1L f	Ankyrin-like protein, host range protein, fragment	66/67 /(99)
31L	28313	27192	373	VV K2L	Serine protease inhibitor-like protein, SPI-3, inhibits cell fusion	350/373 (93)
32L	28630	28364	88	VV K3L	IFN resistance protein, eIF2α-like protein	76/88 (86)
33L	29964	28690	424	VV K4L	Phosopholipase-like protein	408/424 (96)
34L	30502	30278	74	VV K5L f	Unknown	53/61 (86)
35R	30840	31286	148	VV K7R	Unknown	141/148 (95)
36L	32102	31353	249	VV F1L	Unknown	214/245 (87)
37L	32545	32102	147	VV F2L	dUTPase	145/147 (99)
38L	34011	32569	480	VV F3L	Kelch-like protein	464/480 (96)
39L	34984	34022	319	VV F4L	Ribonucleotide reductase, small subunit	312/319 (97)
40L	35983	35012	323	VV F5L	36.5 kDa membrane protein precursor	291/321 (97)
41L	36236	36018	72	VV F6L	Unknown	63/70 (90)
42L	36495	36280	71	VV F7L f	Unknown, fragment	62/65 (95)

Table 1 (*cont.*)

ORF†	Translation‡ Start	Translation‡ Stop	Size (aa)	Matches in public protein databases§	Putative function/structure	Amino acid identity‖
43L	36844	36647	65	VV F8L	Unknown	62/65 (95)
44L*	37542	36904	212	VV F9L	Unknown	208/212 (98)
45L*	38848	37529	439	VV F10L	Serine-threonine protein kinase 2	435/439 (99)
46L	39935	38871	354	VV F11L	Unknown	344/354 (97)
47L*	41886	39979	635	VV F12L	IEV protein	615/635 (96)
48L*	43046	41928	372	VV F13L	EEV antigen, phospholipase, IMCBH sensitivity	363/372 (97)
49L	43284	43063	73	VV F14L	Unknown	60/73 (82)
50L*	44029	43553	158	VV F15L	Unknown	134/135 (99)
51L	44730	44029	233	VV F16L	Unknown	223/230 (96)
52R*	44792	45097	101	VV F17R	IMV core phosphoprotein, VP11, DNA binding protein	99/101 (98)
53L*	46533	45094	479	VV E1L	Poly(A) polymerase catalytic subunit	472/479 (98)
54L*	48743	46530	737	VV E2L	Unknown	727/737 (98)
55L	49436	48864	190	VV E3L	IFN resistance protein, binds dsRNA,	184/190 (97)
56L*	50268	49489	259	VV E4L	RNA polymerase subunit rpo30, VITF-1	257/259 (99)
57R	50345	51334	329	VV E5R	Unknown	312/329 (94)
58R*	51478	53181	567	VV E6R	Unknown	555/567 (97)
59R	53264	53761	165	VV E7R	Myristyl protein	151/165 (91)
60R*	53887	54708	273	VV E8R	Unknown	265/273 (97)
61L*	57734	54714	1006	VV E9L	DNA polymerase	987/1006 (98)
62R*	57766	58053	95	VV E10R	ERV1/ALR protein, disulphide bond formation	91/95 (95)
63L	58437	58048	129	VV E11L	Unknown	126/129 (97)
64L	59257	58424	277	VV O1L f2	Unknown, fragment	255/277 (92)
65L	60110	59346	254	VV O1L f1	Unknown, fragment	264/291 (90)
66L	60770	60444	108	VV O2L	Glutaredoxin 1, thioltransferase	106/108 (99)
67L*	61854	60916	312	VV I1L	DNA-binding protein, late stage virion morphogenesis	307/312 (98)
68L*	62082	61864	72	VV I2L	Unknown	71/72 (98)
69L*	62898	62083	271	VV I3L	Phosphoprotein, binds ssDNA	263/271 (97)
70L	65294	62979	771	VV I4L	Ribonucleotide reductase, large subunit	763/771 (98)
71L*	65560	65321	79	VV I5L	IMV structural protein, VP13K	77/79 (98)
72L*	66727	65579	382	VV I6L	Unknown	378/382 (98)
73L*	67991	66720	423	VV I7L	IMV core protein	422/423 (99)
74R*	67997	70027	676	VV I8R	Nucleoside triphosphate phosphohydrolase II, RNA helicase, NTPase	666/676 (98)
75L*	71806	70031	591	VV G1L	Metallo-endoproteinase, virion morphogenesis	586/591 (99)
76L*	72138	71803	111	VV G3L	Unknown	111/111 (100)
77R*	72132	72794	220	VV G2R	Late elongation factor, IBT-dependent protein	217/220 (98)
78L*	73138	72764	124	VV G4L	Glutaredoxin 2, membrane protein, virion morphogenesis	122/124 (98)
79R*	73141	74445	434	VV G5R	Unknown	425/434 (97)
80R*	74453	74644	63	VV G5.5R	RNA polymerase subunit rpo7	63/63 (100)
81R*	74646	75143	165	VV G6R	Unknown	161/165 (97)
82L*	76223	75108	371	VV G7L	IMV core protein, VP16K	370/371 (99)
83R*	76254	77036	260	VV G8R	Late transcription factor, VLTF-1	258/260 (99)
84R*	77056	78078	340	VV G9R	Myristyl protein	334/340 (98)
85R*	78079	78831	250	VV L1R	Myristylated IMV protein	246/250 (98)

Table 1 (*cont.*)

ORF†	Translation‡ Start	Stop	Size (aa)	Matches in public protein databases§	Putative function/structure	Amino acid identity‖
86R	78863	79126	87	VV L2R	Unknown	85/87 (97)
87L*	80165	79116	349	VV L3L	Unknown	335/349 (95)
88R*	80190	80945	251	VV L4R	IMV core protein VP8, DNA & RNA binding protein	249/251 (99)
89R*	80955	81341	128	VV L5R	Unknown	127/128 (99)
90R*	81298	81759	153	VV J1R	Dimeric virion protein	151/153 (98)
91R	81775	82308	177	VV J2R	Thymidine kinase	172/177 (97)
92R*	82374	83375	333	VV J3R	Poly(A) polymerase stimulatory subunit, VP39, 2'-*O*-methyltransferase	323/333 (96)
93R*	83290	83847	185	VV J4R	RNA polymerase subunit rpo22	183/185 (98)
94L*	84312	83911	133	VV J5L	Unknown	128/133 (96)
95R*	84417	88277	1286	VV J6R	RNA polymerase subunit rpo147	1274/1286 (99)
96L*	88789	88274	171	VV H1L	Tyrosine-serine phosphatase, virion maturation	169/171 (98)
97R*	88803	89372	189	VV H2R	Unknown	186/189 (98)
98L*	90352	89375	325	VV H3L	Immunodominant IMV envelope protein p35	315/325 (96)
99L*	92740	90353	795	VV H4L	RNA polymerase-associated transcription specificity factor, RAP94	785/795 (98)
100R*	92926	93558	210	VV H5R	Late transcription factor, VLTF-4	195/210 (92)
101R*	93559	94503	314	VV H6R	DNA topoisomerase I	313/314 (99)
102R*	94540	94980	146	VV H7R	Unknown	142/146 (97)
103R*	95024	97558	844	VV D1R	mRNA capping enzyme, large subunit	832/844 (98)
104L*	97957	97517	146	VV D2L	IMV core protein	142/146 (97)
105R*	97950	98663	237	VV D3R	IMV core protein	229/237 (96)
106R*	98663	99319	218	VV D4R	Uracil DNA glycosylase	214/218 (98)
107R*	99351	101708	785	VV D5R	Nucleoside triphosphatase	777/785 (98)
108R*	101749	103662	637	VV D6R	Early transcription factor small subunit, VETF-1	630/637 (98)
109R*	103689	104174	161	VV D7R	RNA polymerase subunit rpo18	158/161 (98)
110L	105051	104137	304	VV D8L	IMV surface protein, cell-binding protein, carbonic anhydrase-like protein	292/304 (96)
111R*	105093	105734	213	VV D9R	25 kDa mutT-like protein	211/213 (99)
112R*	105731	106477	248	VV D10R	29 kDa mutT-like protein, negative regulator of gene expression	243/248 (97)
113L*	108373	106478	631	VV D11L	Nucleoside triphosphate phosphohydrolase I	621/631 (98)
114L*	109271	108408	287	VV D12L	mRNA capping enzyme small subunit, intermediate transcription factor, VITF	286/287 (99)
115L*	110957	109302	551	VV D13L	IMV protein, rifampicin resistance	547/551 (99)
116L*	111433	110981	150	VV A1L	Late transcription factor, VLTF-2	148/150 (98)
117L*	112127	111453	224	VV A2L	Late transcription factor, VLTF-3	222/224 (99)
118L	112354	112124	76	VAR A2.5L	Unknown	75/76 (98)
119L*	114303	112369	644	VV A3L	IMV major core protein, P4b	636/644 (98)
120L	115183	114356	275	VV A4L	IMV protein	263/275 (95)
121R*	115221	115715	164	VV A5R	RNA polymerase subunit rpo19	163/164 (99)
122L*	116830	115712	372	VV A6L	Unknown	365/372 (98)
123L*	118986	116854	710	VV A7L	Early transcription factor large subunit, VETF	703/710 (99)
124R*	119040	119906	288	VV A8R	Intermediate transcription factor, VITF-3	285/288 (98)
125L*	120269	119940	109	VV A9L	IMV protein, role in morphogenesis	84/97 (86)
126L*	122948	120270	892	VV A10L	IMV major core protein P4a	870/892 (97)
127R*	122963	123922	319	VV A11R	Unknown	314/319 (98)

Table 1 (*cont.*)

ORF†	Translation‡ Start	Translation‡ Stop	Size (aa)	Matches in public protein databases§	Putative function/structure	Amino acid identity‖
128L*	124499	123924	191	VV A12L	IMV protein	186/191 (97)
129L*	124735	124523	70	VV A13L	IMV membrane protein, p8	64/69 (92)
130L*	125115	124843	90	VV A14L	IMV protein, p16	89/90 (98)
131L*	125293	125132	53	VV A14.5L	IMV protein	53/53 (100)
132L*	125567	125283	94	VV A15L	Unknown	91/94 (96)
133L*	126684	125551	377	VV A16L	Myristyl protein	375/377 (99)
134L*	127298	126687	203	VV A17L	IMV membrane protein, morphogenesis factor	202/203 (99)
135R*	127313	128794	493	VV A18R	DNA helicase, DNA dependent ATPase, transcript release factor	473/493 (95)
136L*	129005	128775	76	VV A19L	Unknown	75/76 (98)
137L*	129356	129006	116	VV A21L	Unknown	116/116 (100)
138R*	129355	130635	426	VV A20R	DNA polymerase processivity factor	419/426 (98)
139R*	130598	131131	177	VV A22R	Holliday junction resolvase	176/176 (100)
140R*	131148	132296	382	VV A23R	Intermediate transcription factor, VITF-3	378/382 (98)
141R*	132293	135787	1164	VV A24R	RNA polymerase subunit rpo132	1154/1163 (99)
142L	138797	136617	726	VAR A29L	CPV ATI protein	666/723 (92)
143L	140338	138842	498	VAR A30L	CPV ATI protein	470/498(94)
144L	140720	140367	117	VV A27L	Membrane IMV protein, fusion protein	106/110 (96)
145L*	141161	140721	146	VV A28L	Unknown	140/146 (95)
146L*	142079	141162	305	VV A29L	RNA polymerase subunit rpo35	301/305 (98)
147L*	142275	142042	77	VV A30L	Unknown	77/77 (100)
148R	142435	142884	149	VV A31R	Unknown	122/149 (81)
149L*	143663	142851	270	VV A32L	ATP & GTP binding motif A, DNA packaging	296/300 (98)
150R	143781	144335	184	VV A33R	EEV glycoprotein	176/184 (95)
151R*	144359	144865	168	VV A34R	EEV glycoprotein	166/168 (98)
152R	144907	145437	176	VV A35R	Unknown	174/176 (98)
153R	145504	146160	218	VV A36R	IEV membrane protein	208/216 (96)
154R	146230	147021	263	VV A37R	Unknown	257/262 (98)
155L	148123	147290	277	VV A38L	Integral membrane glycoprotein	262/277 (94)
156R	148915	149148	77	VV A39R f	Semaphorin-like protein, fragment	67/70 (95)
157L	150164	149505	219	VV A41L	Secreted immunomodulator	207/219 (94)
158R	150336	150737	133	VV A42R	Profilin-like protein	129/133 (96)
159R	150775	151362	195	VV A43R	Type I membrane glycoprotein	188/195 (96)
160R	151364	151600	78	VV MVA 156R	Unknown	47/48 (97)
161L	152734	151694	346	VV A44L	Hydroxysteroid dehydrogenase	337/346 (97)
162R	152781	153158	125	VV A45R	IMV core protein, superoxide dismutase-like protein	120/125 (96)
163R	153148	153870	240	VV A46R	Intracellular signalling inhibitor	232/240 (96)
164L	154402	154034	122	VV A47L f2	Unknown, fragment	116/122 (95)
165L	154617	154399	72	VV A47L f1	Unknown, fragment	65/71 (91)
166R	154781	155395	204	VV A48R	Thymidylate kinase	200/204 (98)
167R	155444	155689	81	VV A49R f	Unknown, fragment	57/62 (91)
168R	155961	157619	552	VV A50R	DNA ligase	535/552 (96)
169R	157672	158676	334	VV A51R	Unknown	315/334 (94)
170L	159143	158871	90	Genomic region encoding VV A53R f1	Unknown	

Table 1 (*cont.*)

ORF†	Translation‡ Start	Translation‡ Stop	Size (aa)	Matches in public protein databases§	Putative function/structure	Amino acid identity‖
172R	159712	161406	564	VV A55R	Kelch-like protein	531/564 (94)
173R	161456	162403	315	VV A56R	Virus haemagglutinin	285/315 (89)
174R	162733	163026	97	VV A57R f	Guanylate kinase fragment	87/97 (89)
175R	163184	164086	300	VV B1R	Serine-threonine protein kinase I, IMV protein	291/300 (97)
176R	164177	165685	502	CPV B2R	Schlafen-like protein	458/503 (91)
177R	165906	167579	557	VV B4R	Ankyrin-like protein	527/557 (94)
178R	167683	168636	317	VV B5R	EEV glycoprotein	294/317 (92)
179R	168718	169263	181	VV B6R	Unknown	151/181 (83)
180R	169567	169848	93	VV B7R f	Virulence factor, fragment	83/92 (90)
181R	169900	170700	266	VV B8R	Soluble IFN-γ receptor	244/263 (92)
182R	170794	171282	162	Unknown	Similarity to CMPV 202R	
183R	171455	172222	255	Unknown	Similarity to CMPV 202R	
184R	172355	173026	223	Unknown	Similarity to CMPV 202R	
185R	173647	174324	225	CPV B8R	Similarity to CAP-T4	145/225 (64)
186R	174470	175975	501	VV B10R	Kelch-like protein	145/161 (90)
187R	176326	177186	286	VV B12R	Protein kinase-like	270/286 (94)
188R	177278	178312	344	VV WR B13R	Serine protease inhibitor-like, SPI-2, crmA	337/344 (97)
189R	178420	178869	149	VV B15R	Unknown	143/149 (95)
190R	179266	179565	99	VV WR B15R f1	Soluble interleukin 1-β receptor, fragment	88/92 (95)
191R	179632	179844	70	VV WR B15R f2	Soluble interleukin 1-β receptor, fragment	57/64 (89)
192L	180481	180266	71	VV B17L f2	Unknown, fragment	90/97 (92)
193L	181010	180600	136	VV B17L f1	Unknown, fragment	119/121 (98)
194R	181344	182081	245	VV B18R f1	Ankyrin-like protein, fragment	224/235 (95)
195R	182263	182559	97	VV B18R f2	Ankyrin-like protein, fragment	71/74 (95)
196R	182695	183762	355	VV B19R	IFN-α/β receptor	316/353 (89)
197R	183877	186228	783	VAR B18R	Ankyrin-like protein	757/791 (95)
198R	186307	187077	256	CPV B19R f1	Kelch-like protein, fragment	235/256 (91)
199R	187319	187741	140	CPV B19R f2	Kelch-like protein fragment	122/131 (93)
200R	188225	189343	372	VAR B21R	Serine protease inhibitor-like, SPI-1	364/362 (97)
201R	189526	190095	189	CPV B21R	Similarity to VV C13L and C14L	170/193 (88)
202R	190354	195963	1869	VAR B22R	Putative membrane-associated glycoprotein	1803/1906 (94)
203R	196437	197537	366	CPV H2R f	Ankyrin-like protein, fragment	355/360 (93)
204R	197765	199522	585	VAR G1R	Ankyrin-like protein	554/585 (92)
205R	199610	200659	349	CPV H4R	Soluble TNF receptor, CrmB	311/351(94)
206R	200786	201553	255	VV Lister 35 kDa protein	35 kDa secreted chemokine-binding protein	237/246 (96)

† The ORFs are numbered from left to right of the genome and the direction of transcription is indicated by letter code L (for left) and R (for right).
‡ Including the last base of the stop codon.
§ Usually only the related VV-COP gene is described. Alternatively, another poxvirus and/or cellular counterpart is identified. For references to stated gene functions see see Goebel *et al.* (1990); Smith *et al.* (1991); Johnson *et al.* (1993); Massung *et al.* (1994); Shchelkunov *et al.* (1998); Bowie *et al.* (2000); Price *et al.* (2000); Almazán *et al.* (2001); Gardner *et al.* (2001); Lee *et al.* (2001); Moss (2000); Ng *et al.* (2001); Saraiva & Alcamí (2001); van Eijl *et al.* (2002).
‖ Percentage shown in parentheses.

rearrangement of corresponding DNA. No related proteins were detected in public databases. In contrast, ORFs 182R, 183R and 184R encode polypeptides with 23 to 31% amino acid identity to the very large VAR-BSH protein B22R. Complete ORFs related to VAR-BSH B22R are present in CMPV (202R), ectromelia virus (Chen *et al*., 2000), CPV-GRI (Shchelkunov *et al*., 1998), MCV (Senkevich *et al*., 1997), FPV (Afonso *et al*., 2000), SFV (Willer *et al*., 1999), MYX (Cameron *et al*., 1999), YLDV (Lee *et al*., 2001) and lumpy skin disease virus (Tulman *et al*., 2001) but are absent in VV (Goebel *et al*., 1990). These ORFs encode the largest OPV proteins (~ 214 kDa), predicted membrane glycoproteins of unknown function. The similarity of CMPV ORFs 182R, 183R and 184R with 202R suggests these smaller ORFs are remnants of another member of this family.

Immunomodulatory proteins

CMPV infection of camels can produce severe disease, suggesting CMPV may interfere with the host response to infection. CMPV expresses soluble proteins that bind IFN-γ (Alcamí & Smith, 1995), IFN-α/β (Symons *et al*., 1995), CC chemokines (Alcamí *et al*., 1998a) and tumour necrosis factor (TNF) (Alcamí *et al*., 1999), and ORFs 181R, 196R, 1L/206R and 2L/205R, respectively, are predicted to encode these activities. In addition, ORFs 11R and 23L encode proteins that are very similar to the VV epidermal growth factor (Blomquist *et al*., 1984) and soluble inhibitor of complement (Kotwal *et al*., 1990). Proteins encoded by ORFs 31L, 188R and 200R have similarity to serpins that have anti-fusion or anti-apoptotic activity; for review see Turner *et al*. (1995). Proteins encoded by ORFs 32L and 55L are similar to VV proteins K3L and E3L that mediate resistance to IFN; for review see Smith *et al*. (1998). Additionally, bioinformatic studies suggest that ORFs 6L, 176R and 201R may have immunomodulatory or host range function.

ORF 6L

The only counterpart of protein 6L among sequenced poxviruses is a slightly shorter protein (S1R) (210 amino acids) in CPV-GRI (Shchelkunov *et al*., 1998). Hydropathy plots of 6L predict an integral membrane protein with five or six transmembrane domains and a putative signal peptide. Protein 6L is closely related to an uncharacterized human protein of family UPF0005 (72% identity, 82% similarity) whose members contain several membrane-spanning domains and share a signature from the third to fourth transmembrane domain (Walter *et al*., 1995). Protein 6L is also related (33–35% identity and 50–59% similarity) to the rat glutamate binding protein (Kumar *et al*., 1991) and the Bax inhibitor-1 (BI-1) family of anti-apoptotic integral membrane proteins (Xu & Reed, 1998; Kawai *et al*., 1999). When overexpressed in mammalian cells, BI-1 suppressed apoptosis induced by Bax, etoposide, staurosporine and growth factor deprivation, indicating BI-1 is a regulator of cell death pathways controlled by Bcl-2 and Bax (Xu & Reed, 1998). Possibly 6L regulates apoptosis in CMPV-infected cells.

ORF 176R

Protein 176R has similarity (36% identity, 56% similarity) with members of the Schlafen (SLFN) protein family that are expressed preferentially in lymphoid tissues and are regulated differentially during thymocyte maturation. Family members are grouped by size: a short form of about 350 amino acids (SLFN1, SLFN2) and a longer form of about 550 amino acids (SLFN3 to SLFN7). The prototype of the family, SLFN1, inhibits T cell growth and development (Schwarz *et al*., 1998). Related proteins are encoded by several OPVs; however, only CMPV 176R and CPV-GRI B2R (Shchelkunov *et al*., 1998) are undisrupted genes, whereas in VV-COP, VV-WR, VAR-BSH, VAR-IND and VAR-GAR, the ORF is fragmented. SLFN proteins are intracellular and so if 176R had a similar location and function, it is unclear how it would regulate T cell development other than after infection of T cells.

ORF 201R

Protein 201R contains a signal peptide, a RGD motif and shows amino acid similarity to the C-terminal domain of OPV TNF receptors CrmB and CrmD (Alcamí *et al*., 1998b). RGD motifs mediate the binding of proteins to cell surface integrins: therefore, 201R might be a secreted protein that binds back to infected and/or uninfected cells. A similar protein is encoded by CPV-GRI gene B21R (Shchelkunov *et al*., 1998) and VV-COP contains a disrupted version of this gene (C13L and C14L) (Goebel *et al*., 1990).

Phylogeny

The relationship of CMPV to other OPVs was analysed by comparison of DNA sequences, predicted protein sequences, repeats within the ITRs and ORFs in the terminal regions. Each comparison gave the same conclusion: CMPV and VAR are more closely related to each other than either is to any other known virus.

DNA sequence comparisons

The central region of OPV genomes is highly conserved between different viruses and this close relationship allowed the pairwise alignment of complete genomes using the program DOTTER (Sonnhammer & Durbin, 1995). Comparison of CMPV and VAR-BSH showed their genomes are collinear except for the differing length of the ITRs and four insertions of 1·5–2·9 kb in CMPV relative to VAR (Fig. 2a, arrows). An apparent fifth gap (Fig. 2a, arrowhead) represents a region where there are several smaller rearrangements. The CMPV ITRs are longer than VAR-BSH ITRs because they

contain three ORFs that are present outside the VAR-BSH ITR near the right end of the genome. A line running perpendicular to the diagonal (Fig. 2a, asterisk) illustrates the presence of this oppositely orientated region present at both ends of the CMPV genome and at only the right end of the VAR-BSH genome.

In contrast to these similarities between CMPV and VAR, more breaks were found when the genomes of VV-COP and VAR-BSH, or VV-COP and CMPV were compared (data not shown). In the terminal regions CPV-GRI could also be compared. Here too, CMPV and VAR were most closely related and this was illustrated by (i) fewer breaks in the aligned genomes and (ii) closer nucleotide sequence identity (data not shown).

The four significant insertions in the CMPV genome compared to VAR-BSH encode ORFs that are absent from all VAR strains analysed. Near the left end of the genome, ORFs 6L and 7L are present in CPV-GRI, whereas 8R and 9L are unique to CMPV. Near the right end of the CMPV genome, most of the region encoding ORFs 182R, 183R and 184R (see above) is missing in VAR-BSH, although parts of 182R containing several frameshift mutations can be identified (region encoding genes B9R and B10R).

The relationships between CMPV, VAR-BSH and VV-COP were analysed further by comparison of the percentage nucleotide identity in the conserved central 110 kb of these genomes (encoding CMPV genes 44L to 151R, VV genes F9L to A34R and VAR-BSH genes C13L to A37R) (Fig. 2b). The region was divided into blocks of approximately 20 kb. Alignments of CMPV, VAR and VV showed that throughout this region nucleotide identity was $\geqslant$ 91%. However, VAR and CMPV are more closely related (96·6–98·6%, average 98·0%) than are CMPV and VV (91·9–98·3%, average 96·7%) or VV and VAR (91·4–97·9%, average 96·0%) (Fig. 2b). For comparison, VAR-BSH and VAR-IND shared 99·8% nucleotide identity in this region.

A DNA distance matrix, constructed using programme PUZZLE 5.0 (Strimmer & von Haeseler, 1997), was used to analyse the genetic distance between CMPV, VAR-BSH and VV-COP (Fig. 2c). In this method, identical sequences give a score of zero and unrelated sequences a score of 1. The analysis showed that the genetic distance between CMPV and VAR (0·0166) was lower than between CMPV and VV (0·0220), and VAR and VV (0·0267). For comparison, the corresponding regions of SFV and MYX were included. This showed the genetic distance between these leporipoxviruses was 0·1277 (8-fold greater than between CMPV and VAR), whereas the distances between each leporipoxvirus and all three OPVs ranged from 0·5642 to 0·5696.

Phylogenetic analyses of protein sequences

To investigate further the evolutionary relationships between CMPV and other poxviruses, phylogenetic trees for specific proteins were constructed from CLUSTALW alignments of protein sequences. In the terminal regions, a comparison of CMPV, VAR, CPV and VV identified 26 ORFs conserved in all these viruses (Table 2, highlighted). The percentage amino acid identities of the CMPV, CPV and VV proteins to VAR-BSH (Fig. 3a) show that in most cases (20/26) the CMPV protein was more closely related to VAR than were proteins from CPV or VV. In 3/26 cases the CMPV protein and the corresponding protein from either or both VV and CPV were equally closely related to VAR, and in only 3/26 cases was the CPV or VV protein more closely related than the CMPV protein to the VAR protein. Consensus phylogenetic trees were also constructed using the maximum likelihood program PUZZLE (Methods) from 11 of these protein sequences (asterisks in Fig. 3a) and are shown in Fig. 3(b, c) together with the corresponding, more distantly related proteins from MYX or SFV to root the tree. In Fig. 3(c) the OPV grouping is expanded to show the relationships together with the quartet puzzling support values. To obtain an independent analysis of the grouping of these four OPVs, the dataset was bootstrapped using the PHYLIP package version 3.5 (Felsenstein, 1989) with programs SEQBOOT, PROTDIST, NEIGHBOR and CONDENSE (Fig. 3e). Both these analyses placed VAR and CMPV together and distinct from CPV and VV.

Seventeen members of the ChPV minimal gene complement were compared next. Phylogenetic trees constructed using PUZZLE for these individual proteins did not give consistent relationships between OPVs and three different topologies were observed. In 53% of the trees, CMPV and VAR-BSH were grouped together on the same branch independently of VV-COP. In 29% of the cases, VAR-BSH and VV-COP grouped together and were independent of CMPV. Finally, in 18% of the cases, CMPV and VV-COP grouped together independently of VAR-BSH. The inconsistent relationship obtained is explained by the very high conservation in these proteins (up to 99% amino acid identity). Similar results have been reported with analysis of single genes from closely related species (Huelsenbeck & Bull, 1996). Therefore, to obtain a reliable phylogenetic relationship, 20 proteins from the central regions were compared simultaneously as for the terminal ORFS (above). This showed that CMPV was most closely related to VAR, and VV was more distantly related (Fig. 3e, f, g). A comparison of the scale bars in Fig. 3(b, e) shows the greater conservation of the proteins encoded in the centre of the genome.

Collectively, the phylogenetic analysis of proteins showed VAR and CMPV are most closely related, consistent with analysis of two CMPV genes (Binns, 1992; Douglass & Dumbell, 1996), which showed CMPV, VAR and taterapoxvirus are closely related.

Comparisons of inverted terminal repeats

OPV ITRs vary in length; for instance, VV-COP, CMPV and VAR-BSH have ITRs of 12068, 6045 and 725 bp,

Table 2. Comparison of OPV ORFs near the genome termini

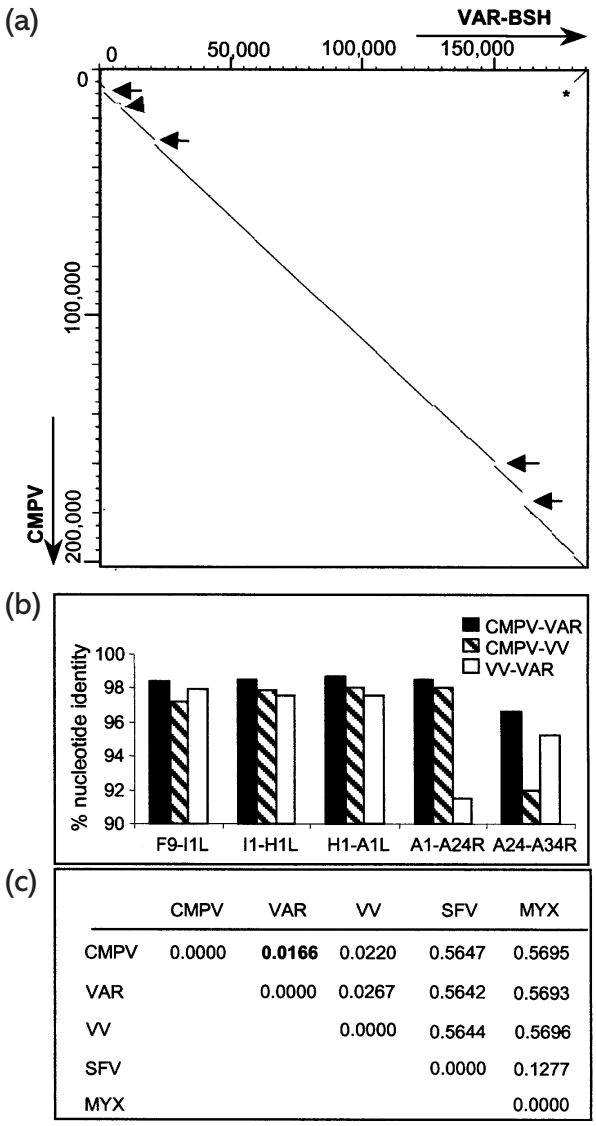
Where possible, each ORF was named after its VV-COP (VV) counterpart, or else by an equivalent ORF that is present in one of the other OPVs examined. Complete ORFs are marked with a tick, fragmented ORFs with the letter F, and deleted ORFs with a horizontal dash. A shaded background highlights the ORFs that are present in all four OPVs. CMPV, camelpox virus strain CMS; CPV, cowpox virus strain GRI-90; VV, vaccinia virus strain Copenhagen; VAR, variola virus strain Bangladesh-1975.

Left end					Right end				
ORFs	VV	VAR	CMPV	CPV	ORFs	VV	VAR	CMPV	CPV
VV C23L	F	–	✓	✓	VV F2L	✓	✓	✓	✓
VV C22L	F	–	✓	✓	VV F3L	✓	F	✓	✓
VV C21-C19L	F	–	✓	✓	VV F4L	✓	✓	✓	✓
VV C18-C17L	F	F	✓	✓	VV F5L	✓	✓	✓	✓
VV C16L	✓	✓	✓	✓	VV F6L	✓	✓	✓	✓
VV C12L	✓	✓	✓	✓	VV F7L	✓	✓	F	✓
CPV S1R	–	–	✓	F	VV F8L	✓	✓	✓	✓
CPV K3R	–	–	F	✓	VV A51R	✓	✓	✓	✓
CMPV 8L	–	–	✓	–	VV A52R	✓	–	–	✓
CMPV 9L	–	–	✓	–	VV A53R	F	–	F	✓
CPV D6L	–	–	–	✓	VV A54R	✓	–	–	✓
CPV D7L	–	–	–	✓	VV A55R	✓	F	✓	✓
CPV D8L	–	–	–	✓	VV A56R	✓	✓	✓	✓
CPV D9L	–	–	–	✓	VV A57R	F	F	F	F
CPV D10L	–	–	–	✓	VV B1R	✓	✓	✓	✓
CPV D11L	–	–	–	✓	VV B2-B3R	F	F	✓	✓
CPV D12L	–	–	–	✓	VV B4R	✓	✓	✓	✓
CPV D13L	–	–	–	✓	VV B5R	✓	✓	✓	✓
CPV D14L	–	–	–	✓	VV B6R	✓	F	✓	✓
CPV C1L	–	–	–	✓	VV B7R	✓	–	F	✓
CPV C2L	–	–	–	✓	VV B8R	✓	✓	✓	✓
CPV C3L	–	–	–	✓	CMPV 182R	–	F	✓	–
CPV C4L	–	F	✓	✓	CMPV 183R	–	–	✓	–
VV C11R	✓	✓	✓	✓	CMPV 184R	–	–	✓	–
VV C10L	✓	✓	F	✓	VV B9R	F	–	✓	✓
CPV C7R	–	✓	✓	✓	VV B10R	F	–	✓	✓
CPV C8L	–	✓	F	✓	VV B11R	F	–	–	✓
CPV C9L	–	F	–	✓	VV B12R	✓	F	✓	✓
CPV C10L	–	–	–	✓	VV B13-B14R	F	✓	✓	✓
VV C9L	✓	F	F	✓	VV B15R	✓	✓	✓	✓
VV C8L	✓	–	✓	✓	VV B16R	✓	F	F	✓
VV C7L	✓	✓	✓	✓	VV B17L	✓	✓	F	✓
VV C6L	✓	✓	✓	✓	VV B18R	✓	✓	F	✓
VV C5L	✓	F	✓	✓	VV B19R	✓	✓	✓	✓
VV C4L	✓	✓	✓	✓	VV B20R	F	✓	✓	✓
VV C3L	✓	✓	✓	✓	CPV B19R	–	F	F	✓
VV C2L	✓	F	✓	✓	CPV B20R	–	✓	✓	✓
VV C1L	✓	✓	F	✓	CPV B21R	F	–	✓	✓
VV N1L	✓	✓	✓	✓	CPV B22R	–	✓	✓	✓
VV N2L	✓	✓	F	✓	CPV K1R	F	–	–	✓
VV M1L	✓	✓	F	✓	CPV K2R	–	–	–	✓
VV M2L	✓	✓	✓	✓	VV B22R	✓	–	–	✓
VV K1L	✓	F	F	✓	VV B23-B24R	F	F	F	✓
VV K2L	✓	✓	✓	✓	VV B25-B27R	F	✓	✓	✓
VV K3L	✓	✓	✓	✓	VV B28R	F	✓	✓	✓
VV K4L	✓	–	✓	✓	VV B29R	F	✓	✓	✓
VV K5-K6L	F	–	F	✓					
VV K7R	✓	✓	✓	✓					
VV F1L	✓	✓	✓	✓					

respectively. Although CMPV and VAR ITRs differ in length, and therefore might appear divergent, a terminal transposition event (Moyer & Graves, 1981) could create larger VAR ITRs similar to CMPV. Alternatively, if previously VAR had ITRs of similar size to CMPV, a deletion of sequences from within the VAR left ITR could have created the present structure. Evidence supporting this possibility comes from analysis of CMPV ORF 4L, which crosses the left ITR boundary and is repeated in part within the right ITR (ORF 203R). In VAR, sequences related to CMPV ORF 4L are present at each end of the genome (D1* and f* Fig. 4a, b), suggesting a longer ITR at one stage. In contrast, ORFs related to CMPV 1L/206R, 2L/205R and 3L/204R are found at only the right end of the VAR genome (G3R, G2R and G1R) (Fig. 4a). VAR strain Somalia is unusual in that its ITR is longer than other VAR strains and the repeated sequences, which are outside the ITR in other VAR strains, are included in the Somalia ITR. This might suggest that VAR strain Somalia represents a structure intermediate between most VAR strains and that of other OPVs (Massung *et al.*, 1995).

Generally, ORFs in the terminal regions are variable between OPVs and distinctive for each virus. However, the arrangement of ORFs close to and within the ITR of CMPV and VAR shows a higher degree of similarity (Fig. 4a, b). Firstly, in each virus, ORFs extend to within 650 bp of the termini, a feature that distinguishes CMPV and VAR from other OPVs. Secondly, outside the ITR the gene pattern is conserved except for the absence of counterparts of CMPV 6L to 9L from the left and 201R from the right end of VAR.

All OPV ITRs sequenced hitherto contain blocks of tandem repeats that vary in sequence, number and arrangement. A comparison of OPV terminal repeats (Fig. 4c) shows that CMPV is most similar to VAR. Firstly, CMPV strains CMS and 903 and all five sequenced strains of VAR, but not VV, CPV and racoonpoxvirus (Parsons & Pickup, 1987; Massung *et al.*, 1995), show a block of repeats containing three 70 bp repeats (yellow) followed by related sequences of 52 or 64 (red) and 27 (green) bp. Although the absolute length of these repeats varies slightly between viruses, their sequences are highly conserved, confirming their relationship. Secondly, some repeats (pink and dark blue symbols) that are shared by VV and CPV are absent from all VAR and CMPV strains. CMPV strains CMS and 903 have different numbers of 70 bp repeats consistent with terminal length heterogeneity in CMPV isolates (Pfeffer *et al.*, 1996) and each lack the NR2 sequence (Fig. 4c).



(c)

	CMPV	VAR	VV	SFV	MYX
CMPV	0.0000	**0.0166**	0.0220	0.5647	0.5695
VAR		0.0000	0.0267	0.5642	0.5693
VV			0.0000	0.5644	0.5696
SFV				0.0000	0.1277
MYX					0.0000

Fig. 2. (a) Graphical dotplot alignment of the CMPV-CMS and VAR-BSH genomes using the program DOTTER. Regions of high similarity are shown by a diagonal line. Nucleotide numbering for each genome is indicated on the corresponding axis. Arrows indicate regions of DNA $\geqslant$ 1·5 kb that are present in CMPV but not VAR. The arrowhead indicates a region of sequence rearrangement. The asterisk illustrates the CMPV ITR that is present at only the right end of the VAR genome. (b) Pairwise DNA alignment [GCG program GAP (Devereux *et al.*, 1984)] of the central 100 kb of the CMPV, VAR and VV genomes divided into blocks of approximately 20 kb, labelled using VV gene nomenclature. Nucleotide identities (%) are shown. (c) DNA distance matrix. DNA sequences between counterparts of VV genes F9L to A34R were aligned using the program CLUSTALW (Thompson *et al.*, 1994) and a DNA distance matrix was constructed using the program PUZZLE 5.0 (Strimmer & von Haeseler, 1996). CMPV, CMPV-CMS; VAR, VAR-BSH; VV, VV-COP; MYX, MYX Lausanne; SFV, SFV Kaza.

Evolution of VAR and CMPV

All the above comparisons established that VAR and CMPV are more closely related to each other than to any other virus. This suggests either that one virus has evolved from the other, or that they have each evolved from a closely related ancestral virus distinct from VV and CPV. The first possibility seems unlikely because of DNA sequences unique to either virus (insertions of 1·5–2·9 kb in CMPV and the presence of NR2 in the VAR ITR). So evolution from a closely related ancestor, possibly a rodent virus (Fenner *et al.*, 1988, 1989),

(a)

ORFs	CMPV	CPV	VV	Putative function
C16L	97	95	91	Unknown
C12L	96	92	92	SPI-1
C11R	92	85	89	EGF-like growth factor
C7L*	98	99	99	Host range
C6L	98	94	96	Unknown
C4L	96	95	95	Unknown
C3L*	95	91	95	Complement control
N1L*	97	95	94	Virulence factor
M2L*	97	96	96	Unknown
K2L	95	95	93	SPI-3
K3L*	90	82	81	IFN resistance
K7R	97	100	96	Unknown
F1L	90	87	85	Unknown
F2L*	98	97	96	dUTPase
F4L	97	97	97	Ribonucleotide reductase
F5L	94	88	88	36.5 kDa precursor
F6L	95	88	88	Unknown
F8L*	96	95	95	Unknown
A51R*	96	95	92	Unknown
A56R	89	84	84	Haemagglutinin
B1R*	97	96	96	Protein kinase
B4R	96	94	94	Ankyrin-like
B5R	96	92	93	EEV glycoprotein
B8R*	98	92	92	Soluble IFN-γ receptor
B15R*	95	97	96	Unknown
B19R	94	90	88	Soluble IFN-α/β receptor

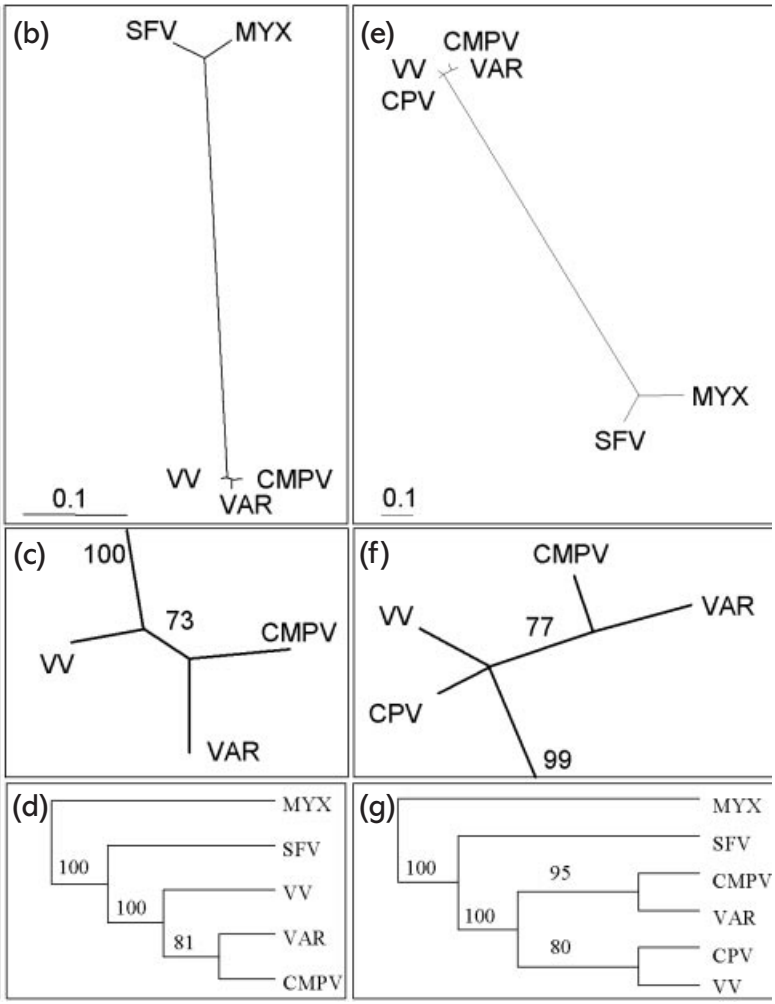


Fig. 3. (a) VAR-BSH terminal ORFs that are conserved in CMPV-CMS, CPV-GRI-90 and VV-COP and the amino acid identities of the encoded proteins compared with VAR-BSH proteins. Consensus phylogenetic trees for 20 proteins (VV-COP F12L, E1L, E9L, I8R, J3R, H6R, D1R, D4R, D5R, D6R, D7R, D10R, D11L, D12L, A1L, A2L, A5R, A7L, A18R and A24R) encoded in the central region (b, c, d) or for 11 proteins [marked with an asterisk in (a)] encoded in the terminal regions (e, f, g) of the OPV genomes were constructed with program PUZZLE version 5.0 (Strimmer & von Haeseler, 1996, 1997) (b, c, e, f) and with the PHYLIP package version 3.5 (Felsenstein, 1989) (d, g). Amino acid sequences of OPV proteins were aligned using the program CLUSTALW version 1.8 (Thompson *et al*., 1994) and phylogenic trees were viewed using TREEVIEW version 1.6.0 (Page, 1996). The scale bar indicates 0·1 substitutions per site (b, e) and the quartet puzzling support values for each branch are indicated in (c) and (f) (VT model of substitution, 25000 puzzling steps). The more distantly related proteins from MYX or SFV were included to root the tree. The branch lengths shown in (d) and (g) are arbitrary and the root position was forced using MYX. The numbers at the forks show the number of bootstrap repetitions, out of 100, in which the given topology was observed. Bootstrapping values were calculated using the modules SEQBOOT (random number seed 123, 100 replicates), PRODIST (Dayhoff PAM matrix, analysis of 100 data sets), NEIGHBOR (neighbour-joining analysis of 100 data sets) and CONDENSE.

seems more probable. When this took place is uncertain, but highly infectious diseases, such as measles and smallpox, require human populations of between 100000 and 300000 to retain transmission between susceptible (non-immune) hosts. During human evolution, populations of this size within a reasonably defined geographical area arose when man adopted intensive agriculture rather than being an isolated hunter gather, between 5000 and 10000 years BC. The presence of camels in areas of human and associated rodent population expansion such as the Nile, Tigris, Euphrates, Ganges and Indus river basins makes it possible that an ancestral OPV might have spread to camels at a similar time. The presence of many broken genes in VAR and CMPV, which are non-essential for virus replication but not jettisoned from these virus genomes, suggests that VAR and CMPV are relatively recent pathogens of man and camels, respectively. Given longer, these viruses might have become adapted better to man and camels to become less virulent and possibly lose some of these non-essential gene fragments.

Whatever the precise origin of VAR and CMPV, the

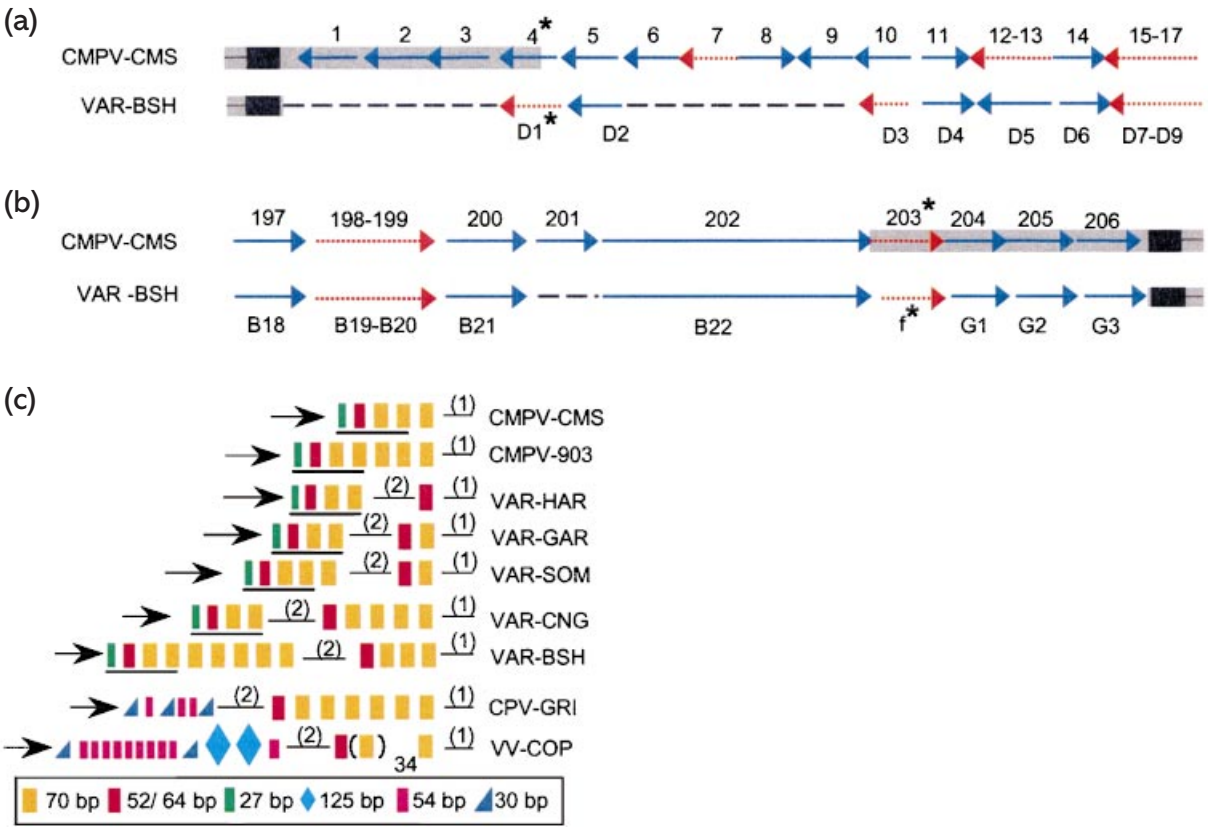


Fig. 4. Comparison of CMPV-CMS and VAR-BSH terminal ORFs and repeats within the ITRs. Leftmost 16 kb (a) and rightmost 18 kb (b) of the CMPV genome and the corresponding region of VAR are shown. ITRs are represented with a grey background, ORFs by blue arrows, fragmented ORFs by orange arrows and regions absent in VAR with a dashed line. Asterisks represent sequences related to CMPV 4L. The block of terminal repeats is illustrated by a black box. (c) OPV terminal repeats. Different repeats are illustrated in colour as indicated. ORFs are shown by arrows. NR1 (1) and NR2 (2) sequences are represented by horizontal lines. Underlined repeats indicate the block of repeats unique to CMPV and VAR strains in which there are two or more 70 bp repeats in VAR strains Harvey-1947 (HAR), Garcia (GAR), Somalia-1970 (SOM) and Congo-1970 (CNG) (Massung *et al.*, 1995).

collinearity of their genomes (except for a few insertions in CMPV), their DNA sequences, ITRs and encoded proteins all show they are the closest known viruses to each other. In addition, CMPV and VAR share other distinctive biological properties, such as their ability to induce high morbidity and mortality in a single host species, the similar pock morphology and ceiling temperature for growth in the chorioallantoic membrane of the fertile hen's egg and the inability to grow in rabbit skin (Fenner *et al.*, 1989).

Although the disease smallpox has been eradicated, there are concerns about the potential use of VAR in bioterrorism and the WHO has scheduled and postponed (until 31/12/2002) destruction of the last VAR stocks held in Russia and USA. CMPV has not caused disease in man, but the possibility of an OPV such as VAR, monkeypox, CMPV or taterapox virus emerging or re-emerging as a threat to human health increases as the proportion of the world's population that is immunologically naïve for OPVs increases. The parallel increase in those immunosuppressed due to HIV infection potentiates the chance of OPVs jumping species and adapting to mankind. This possibility and the threat of bioterrorism justify the retention of adequate stocks of vaccine (VV) to combat OPV infections. Finally, it is unclear whether all, only a few, or just one of the differences between the CMPV and VAR genomes are responsible for the inability of CMPV to cause human disease. Consequently, genetic modification of CMPV to delete genes that are present in CMPV but absent in VAR might be unwise. It might also be unwise to insert into CMPV genes encoding Th2 cytokines, which caused a dramatic change in ectromelia virus virulence (Jackson *et al.*, 2001).

We thank Bela Tiwari, Nicki Gray and Rory Bowden for computing help. This work was supported by the UK Medical Research Council, The Wellcome Trust, the Swiss National Science Foundation (83 EU-56142), Roche, and the University of Lausanne. G.L.S. is a Wellcome Trust Principal Research Fellow.

Table 3. CMPV ORFs that are broken in or absent from the VAR-BSH genome, and VAR-BSH ORFs that are broken in the CMPV genome

The putative functions of the encoded proteins are indicated. ORFs are named using the CMPV (top) or the VAR-BSH (bottom) nomenclature. F, fragmented ORFs; –, deleted ORFs.

CMPV ORFs broken in or absent from VAR–BSH		
CMPV	**VAR–BSH**	**Putative function**
4L	F	Ankyrin-like
6L	–	Unknown
8L	–	Unknown
9L	–	Unknown
10L	F	Unknown
18L	–	Unknown
21L	F	Unknown
24L	F	Kelch-like
33L	–	Unknown
38L	F	Kelch-like
152L	F	Unknown
154R	F	Unknown
161L	F	3-β-HSD
172R	F	Kelch-like
176R	F	Schlafen-like
179R	F	Unknown
185R	–	Unknown
186R	–	Kelch-like
187R	F	Protein kinase-like
201R	–	Unknown
VAR–BSH ORFs broken in CMPV		
VAR–BSH	**CMPV**	**Putative function**
D5L	F	Unknown
D7L	F	IL-18 binding protein
D18L	F	Unknown
P2L	F	α-Amanitin sensitive
O1L	F	Ankyrin-like
C12L	F	Unknown
Q1L	F	Unknown
J1L	F	Unknown
J3R	F	Unknown
B15L	F	Unknown
B16R	F	Ankyrin-like

References

Afonso, C. L., Tulman, E. R., Lu, Z., Zsak, L., Kutish, G. F. & Rock, D. L. (2000). The genome of fowlpox virus. *Journal of Virology* **74**, 3815–3831.

Aguado, B., Selmes, I. P. & Smith, G. L. (1992). Nucleotide sequence of 21·8 kbp of variola major virus strain Harvey and comparison with vaccinia virus. *Journal of General Virology* **73**, 2887–2902.

Alcamí, A. & Smith, G. L. (1995). Vaccinia, cowpox, and camelpox viruses encode soluble gamma interferon receptors with novel broad species specificity. *Journal of Virology* **69**, 4633–4639.

Alcamí, A., Symons, J. A., Collins, P. D., Williams, T. J. & Smith, G. L. (1998a). Blockade of chemokine activity by a soluble chemokine binding protein from vaccinia virus. *Journal of Immunology* **160**, 624–633.

Alcamí, A., Symons, J. A., Khanna, A. & Smith, G. L. (1998b). Poxviruses: capturing cytokines and chemokines. *Seminars in Virology* **8**, 419–427.

Alcamí, A., Khanna, A., Paul, N. L. & Smith, G. L. (1999). Vaccinia virus strains Lister, USSR and Evans express soluble and cell surface tumour necrosis factors receptor. *Journal of General Virology* **80**, 949–959.

Almazán, F., Tscharke, D. C. & Smith, G. L. (2001). The vaccinia virus superoxide dismutase-like protein (A45R) is a virion component that is nonessential for virus replication. *Journal of Virology* **75**, 7018–7029.

Amegadzie, B. Y., Ahn, B. Y. & Moss, B. (1992). Characterization of a 7-kilodalton subunit of vaccinia virus DNA-dependent RNA polymerase with structural similarities to the smallest subunit of eukaryotic RNA polymerase II. *Journal of Virology* **66**, 3003–3010.

Antoine, G., Scheiflinger, F., Dorner, F. & Falkner, F. G. (1998). The complete genomic sequence of the modified vaccinia Ankara strain: comparison with other orthopoxviruses. *Virology* **244**, 365–396.

Bankier, A. T., Weston, K. M. & Barrell, B. G. (1987). Random cloning and sequencing by the M13/dideoxynucleotide chain termination method. *Methods in Enzymology* **155**, 51–93.

Baxby, D. (1972). Smallpox-like viruses from camels in Iran. *Lancet* ii, 1063–1065.

Baxby, D. (1974). Differentiation of smallpox and camelpox viruses in cultures of human and monkey cells. *Journal of Hygiene (London)* **72**, 251–254.

Baxby, D. (1981). *Jenner's Smallpox Vaccine. The Riddle of the Origin of Vaccinia Virus*. London: Heinemann.

Baxby, D., Hessami, M., Ghaboosi, B. & Ramyar, H. (1975). Response of camels to intradermal inoculation with smallpox and camelpox viruses. *Infection and Immunity* **11**, 617–621.

Bedson, H. S. (1972). Camelpox and smallpox. *Lancet* **ii**, 1253.

Betakova, T., Wolffe, E. J. & Moss, B. (2000). The vaccinia virus A14.5L gene encodes a hydrophobic 53-amino-acid virion membrane protein that enhances virulence in mice and is conserved among vertebrate poxviruses. *Journal of Virology* **74**, 4085–4092.

Binns, M. (1992). Analysis of the camelpox virus thymidine kinase gene. *British Veterinary Journal* **148**, 541–546.

Blomquist, M. C., Hunt, L. T. & Barker, W. C. (1984). Vaccinia virus 19-kilodalton protein: relationship to several mammalian proteins, including two growth factors. *Proceedings of the National Academy of Sciences, USA* **81**, 7363–7367.

Bonfield, J. K. & Staden, R. (1996). Experiment files and their application during large-scale sequencing projects. *DNA Sequence* **6**, 109–117.

Bonfield, J. K., Smith, K. & Staden, R. (1995). A new DNA sequence assembly program. *Nucleic Acids Research* **23**, 4992–4999.

Bowie, A., Kiss-Toth, E., Symons, J. A., Smith, G. L., Dower, S. K. & O'Neill, L. A. (2000). A46R and A52R from vaccinia virus are antagonists of host IL-1 and toll-like receptor signaling. *Proceedings of the National Academy of Sciences, USA* **97**, 10162–10167.

Cameron, C., Hota-Mitchell, S., Chen, L., Barrett, J., Cao, J. X., Macaulay, C., Willer, D., Evans, D. & McFadden, G. (1999). The complete DNA sequence of myxoma virus. *Virology* **264**, 298–318.

Chen, N., Buller, R. M., Wall, M. & Upton, C. (2000). Analysis of host response modifier ORFs of ectromelia virus, the causative agent of mousepox. *Virus Research* **66**, 155–173.

DeFilippes, F. M. (1982). Restriction enzyme mapping of vaccinia virus DNA. *Journal of Virology* **43**, 136–149.

DeLange, A. M., Reddy, M., Scraba, D., Upton, C. & McFadden, G. (1986). Replication and resolution of cloned poxvirus telomeres in vivo generates linear minichromosomes with intact viral hairpin termini. *Journal of Virology* **59**, 249–259.

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

Douglass, N. J. & Dumbell, K. R. (1996). DNA sequence variation as a clue to the phylogenesis of orthopoxviruses. *Journal of General Virology* **77**, 947–951.

Esposito, J. J. & Knight, J. C. (1985). Orthopoxvirus DNA: a comparison of restriction profiles and maps. *Virology* **143**, 230–251.

Esposito, J. R., Condit, R. C. & Obijeski, J. (1981). The preparation of orthopoxvirus DNA. *Journal of Virological Methods* **2**, 175–179.

Ewing, B. & Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research* **8**, 186–194.

Ewing, B., Hillier, L., Wendl, M. C. & Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research* **8**, 175–185.

Felsenstein, J. (1989). PHYLIP–Phylogeny Inference Package (Version 3.2). *Cladistics* **5**, 164–166.

Fenner, F., Anderson, D. A., Arita, I., Jezek, Z. & Ladnyi, I. D. (1988). *Smallpox and its Eradication*. Geneva: World Health Organization.

Fenner, F., Wittek, R. & Dumbell, K. R. (1989). *The Orthopoxviruses*, chapter 4. London: Academic Press.

Gardner, J. D., Tscharke, D. C., Reading, P. C. & Smith, G. L. (2001). Vaccinia virus semaphorin A39R is a 50–55 kDa secreted glycoprotein that affects the outcome of infection in a murine intradermal model. *Journal of General Virology* **82**, 2083–2093.

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.

Huelsenbeck, J. P. & Bull, J. J. (1996). A likelihood ratio test to detect conflicting phylogenetic signal. *Systematic Biology* **45**, 92–98.

Jackson, R. J., Ramsay, A. J., Christensen, C. D., Beaton, S., Hall, D. F. & Ramshaw, I. A. (2001). Expression of mouse IL-4 by a recombinant ectromelia virus suppresses cytolytic lymphocyte responses and overcomes genetic resistance to mousepox. *Journal of Virology* **75**, 1205–1210.

Jezek, Z., Kriz, B. & Rothbauer, V. (1983). Camelpox and its risk to the human population. *Journal of Hygiene Epidemiology Microbiology and Immunology* **27**, 29–42.

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

Kawai, M., Pan, L., Reed, J. C. & Uchimiya, H. (1999). Evolutionally conserved plant homologue of the Bax inhibitor-1 (BI-1) gene capable of suppressing Bax-induced cell death in yeast(1). *FEBS Letters* **464**, 143–147.

Kotwal, G. J., Isaacs, S. N., McKenzie, R., Frank, M. M. & Moss, B. (1990). Inhibition of the complement cascade by the major secretory protein of vaccinia virus. *Science* **250**, 827–830.

Kumar, K. N., Tilakaratne, N., Johnson, P. S., Allen, A. E. & Michaelis, E. K. (1991). Cloning of cDNA for the glutamate-binding subunit of an NMDA receptor complex. *Nature* **354**, 70–73.

Lee, H.-J., Essani, K. & Smith, G. L. (2001). The genome sequence of Yaba-like disease virus, a yatapoxvirus. *Virology* **281**, 170–192.

McGrane, J. J. & Higgins, A. J. (1986). Infectious diseases of the camel: viruses, bacteria and fungi. In *The Camel in Health and Disease*, pp. 92–110. Edited by A. J. Higgins. London: Ballière Tindall.

Mackett, M. & Archard, L. C. (1979). Conservation and variation in orthopoxvirus genome structure. *Journal of General Virology* **45**, 683–701.

Mackett, M., Smith, G. L. & Moss, B. (1985). The construction and characterization of vaccinia virus recombinants expressing foreign genes. In *DNA Cloning: a Practical Approach*, pp. 191–211. Edited by D. M. Glover. Oxford: IRL Press.

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

Massung, R. F., Jayarama, V. & Moyer, R. W. (1993b). DNA sequence analysis of conserved and unique regions of swinepox: identification of genetic elements supporting phenotypic observations including a novel G protein-coupled receptor homologue. *Virology* **197**, 511–528.

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

Massung, R. F., Knight, J. C. & Esposito, J. J. (1995). Topography of variola smallpox virus inverted terminal repeats. *Virology* **211**, 350–355.

Mathew, E. C., Sanderson, C. M., Hollinshead, R. & Smith, G. L. (2001). A mutational analysis of the vaccinia virus B5R protein. *Journal of General Virology* **82**, 1199–1213.

Merchlinsky, M. & Moss, B. (1986). Resolution of linear minichromosomes with hairpin ends from circular plasmids containing vaccinia virus concatemer junctions. *Cell* **45**, 879–884.

Meyer, H. & Rziha, H. J. (1993). Characterization of the gene encoding the A-type inclusion protein of camelpox virus and sequence comparison with other orthopoxviruses. *Journal of General Virology* **74**, 1679–1684.

Moss, B. (1996). *Poxviridae*: the viruses and their replication. In *Fields Virology*, 3rd edn, pp. 2637–2671. Edited by B. N. Fields, D. M. Knipe & P. M. Howley. Philadelphia: Lippincott–Raven.

Moss, B. (2001). *Poxviridae*: the viruses and their replication. In *Fields Virology*, 4th edn, pp. 2849–2883. Edited by D. M. Knipe & P. M. Howley. Philadelphia: Lippincott Williams & Wilkins.

Moyer, R. W. & Graves, R. L. (1981). The mechanism of cytoplasmic orthopoxvirus DNA replication. *Cell* **27**, 391–401.

Müller, T. & Vingron, M. (2000). Modeling amino acid replacement. *Journal of Comparative Biology* **7**, 761–776.

Ng, A., Tscharke, D. C., Reading, P. C. & Smith, G. L. (2001). The vaccinia virus A41L protein is a soluble 30 kDa glycoprotein that affects virus virulence. *Journal of General Virology* **82**, 2095–2105.

Nicholas, K. B. & Nicholas, H. B. (1997). GeneDoc: a tool for editing and annotating multiple sequence alignments, 2.5.000 edn. Distributed by the authors.

Page, R. D. (1996). TreeView: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**, 357–358.

Parsons, B. L. & Pickup, D. J. (1987). Tandemly repeated sequences are present at the ends of the DNA of raccoonpox virus. *Virology* **161**, 45–53.

Perkus, M. E., Goebel, S. J., Davis, S. W., Johnson, G. P., Norton, E. K. & Paoletti, E. (1991). Deletion of 55 open reading frames from the termini of vaccinia virus. *Virology* **180**, 406–410.

Pfeffer, M., Meyer, H., Wernery, U. & Kaaden, O. R. (1996). Comparison of camelpox viruses isolated in Dubai. *Veterinary Microbiology* **49**, 135–146.

Price, N., Tscharke, D. C., Hollinshead, M. & Smith, G. L. (2000). Vaccinia virus gene B7R encodes an 18-kDa protein that is resident in the endoplasmic reticulum and affects virus virulence. *Virology* **267**, 65–79.

Saraiva, M. & Alcamí, A. (2001). CrmE, a novel soluble tumor necrosis factor receptor encoded by poxviruses. *Journal of Virology* **75**, 226–233.

Schwarz, D. A., Katayama, C. D. & Hedrick, S. M. (1998). Schlafen, a new family of growth regulatory genes that affect thymocyte development. *Immunity* **9**, 657–668.

Senkevich, T. G., Koonin, E. V., Bugert, J. J., Darai, G. & Moss, B. (1997). The genome of molluscum contagiosum virus: analysis and comparison with other poxviruses. *Virology* **233**, 19–42.

Shchelkunov, S. N., Massung, R. F. & Esposito, J. J. (1995). Comparison of the genome DNA sequences of Bangladesh-1975 and India-1967 variola viruses. *Virus Research* **36**, 107–118.

Shchelkunov, S. N., Safronov, P. F., Totmenin, A. V., Petrov, N. A., Ryazankina, O. I., Gutorov, V. V. & Kotwal, G. J. (1998). The genomic sequence analysis of the left and right species-specific terminal region of a cowpox virus strain reveals unique sequences and a cluster of intact ORFs for immunomodulatory and host range proteins. *Virology* **243**, 432–460.

Shchelkunov, S. N., Totmenin, A. V., Loparev, V. N., Safronov, P. F., Gutorov, V. V., Chizhikov, V. E., Knight, J. C., Parsons, J. M., Massung, R. F. & Esposito, J. J. (2000). Alastrim smallpox variola minor virus genome DNA sequences. *Virology* **266**, 361–386.

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

Smith, G. L., Symons, J. A. & Alcamí, A. (1998). Poxviruses; interfering with interferon. *Seminars in Virology* **8**, 409–418.

Sonnhammer, E. L. & Durbin, R. (1995). A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* **167**, GC1–10.

Staden, R. & McLachlan, A. D. (1982). Codon preference and its use in identifying protein coding regions in long DNA sequences. *Nucleic Acids Research* **10**, 141–156.

Strimmer, K. & von Haeseler, A. (1996). Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Molecular Biology and Evolution* **13**, 964–969.

Strimmer, K. & von Haeseler, A. (1997). Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proceedings of the National Academy of Sciences, USA* **94**, 6815–6819.

Symons, J. A., Alcamí, A. & Smith, G. L. (1995). Vaccinia virus encodes a soluble type I interferon receptor of novel structure and broad species specificity. *Cell* **81**, 551–560.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). Improved sensitivity of profile searches through the use of sequence weights and gap excision. *Computer Applications in the Biosciences* **10**, 19–29.

Tulman, E. R., Afonso, C. L., Lu, Z., Zsak, L., Kutish, G. F. & Rock, D. L. (2001). Genome of lumpy skin disease virus. *Journal of Virology* **75**, 7122–7130.

Turner, P. C., Musy, P. Y. & Moyer, R. W. (1995). Poxvirus serpins. In *Viroceptors, Virokines and Related Immune Modulators Encoded by DNA Viruses*, pp. 67–88. Edited by G. McFadden. Austin: R. G. Landes.

van Eijl, H., Hollinshead, M., Rodger, G., Zhang, W.-H. & Smith, G. L. (2002). The vaccinia virus F12L protein is associated with intracellular enveloped virus particles and is required for their egress to the cell surface. *Journal of General Virology* **83**, 195–207.

Walter, L., Marynen, P., Szpirer, J., Levan, G. & Gunther, E. (1995). Identification of a novel conserved human gene, TEGT. *Genomics* **28**, 301–304.

Willer, D. O., McFadden, G. & Evans, D. H. (1999). The complete genome sequence of Shope (rabbit) fibroma virus. *Virology* **264**, 319–343.

Xu, Q. & Reed, J. C. (1998). Bax inhibitor-1, a mammalian apoptosis suppressor identified by functional screening in yeast. *Molecular Cell* **1**, 337–346.

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119.

Received 30 August 2001; Accepted 13 November 2001