Journal of General Virology (2004), **85**, 105–117 DOI 10.1099/vir.0.19565-0

Poxvirus genomes: a phylogenetic analysis

Caroline Gubser,[1] Stéphane Hué,[2] Paul Kellam[2] and Geoffrey L. Smith[1]

[1]Department of Virology, Faculty of Medicine, Imperial College London, St Mary's Campus, Norfolk Place, London W2 1PG, UK

[2]Viral Genomics and Bioinformatics Group, Department of Virology, and Department of Immunology & Molecular Pathology, University College London, 46 Cleveland Street, London W1T 4JF, UK

Correspondence
Geoffrey Smith
glsmith@imperial.ac.uk

The evolutionary relationships of 26 sequenced members of the poxvirus family have been investigated by comparing their genome organization and gene content and by using DNA and protein sequences for phylogenetic analyses. The central region of the genome of chordopoxviruses (ChPVs) is highly conserved in gene content and arrangement, except for some gene inversions in *Fowlpox virus* (FPV) and species-specific gene insertions in FPV and *Molluscum contagiosum virus* (MCV). In the central region 90 genes are conserved in all ChPVs, but no gene from near the termini is conserved throughout the subfamily. Inclusion of two entomopoxvirus (EnPV) sequences reduces the number of conserved genes to 49. The EnPVs are divergent from ChPVs and between themselves. Relationships between ChPV genera were evaluated by comparing the genome size, number of unique genes, gene arrangement and phylogenetic analyses of protein sequences. Overall, genus *Avipoxvirus* is the most divergent. The next most divergent ChPV genus is *Molluscipoxvirus*, whose sole member, MCV, infects only man. The *Suipoxvirus*, *Capripoxvirus*, *Leporipoxvirus* and *Yatapoxvirus* genera cluster together, with *Suipoxvirus* and *Capripoxvirus* sharing a common ancestor, and are distinct from the genus *Orthopoxvirus* (OPV). Within the OPV genus, *Monkeypox virus*, *Ectromelia virus* and *Cowpox virus* strain Brighton Red (BR) do not group closely with any other OPV, *Variola virus* and *Camelpox virus* form a subgroup, and *Vaccinia virus* is most closely related to CPV-GRI-90. This suggests that CPV-BR and GRI-90 should be separate species.

Received 8 August 2003
Accepted 3 October 2003

INTRODUCTION

The poxviruses represent a family of large DNA viruses that replicate in the cytoplasm. *Variola virus* (VAR), the cause of the disease smallpox, is the most notorious poxvirus and was eradicated by vaccination with *Vaccinia virus* (VV) (Fenner *et al.*, 1988), a related *Orthopoxvirus* (OPV) of unknown origin (Baxby, 1981). The poxvirus family is subdivided into the entomopoxvirus (EnPV) and chordopoxvirus (ChPV) subfamilies (*Entomopoxvirinae* and *Chordopoxvirinae*), which infect insects and chordates, respectively (Moss, 2001). The ChPVs are further divided into eight genera (*Avipoxvirus*, *Molluscipoxvirus*, *Orthopoxvirus*, *Capripoxvirus*, *Suipoxvirus*, *Leporipoxvirus*, *Yatapoxvirus* and *Parapoxvirus*), whereas the EnPVs are divided into three genera (*A*, *B* and *C*). Although poxvirus genome organization, replication, host range and pathogenesis have been studied extensively (Moss, 2001), less is known about the evolutionary relationships of these viruses.

Recently, the number of poxvirus genome sequences has increased considerably (Table 1) (Upton *et al.*, 2003) and the sequences of two EnPVs and at least one member of each ChPV genus, except genus *Parapoxvirus*, are available for comparison. These data enable analysis of the evolutionary relationships of these viruses and the result of such an investigation is presented here. To study the evolutionary relationships of poxviruses, we have compared the size of virus genomes, the number of conserved and unique genes and their arrangement within the genome. The nucleotide or amino acid sequences of subsets of those genes were then used for phylogenetic analyses. Lastly, we have considered the OPV genus in more detail. The results showed that in addition to the close relationship of VAR and *Camelpox virus* (CMPV) that was noted previously (Gubser & Smith, 2002), *Monkeypox virus* (MPV), which causes a disease with clinical similarity to smallpox, is divergent from all OPVs, and so are *Ectromelia virus* (ECT) and *Cowpox virus* (CPV) Brighton Red (BR). Data presented also suggest that the classification of CPV-BR and CPV-GRI-90 as two strains of the same species should be reassessed.

METHODS

DNA sequence comparison from central genomic regions. The central ~100 kb of 11 OPVs (98 872 nucleotides in VV Copenhagen, VV-COP) were aligned with the program CLUSTALW version 1.8 (Thompson *et al.*, 1994) using the default parameters.

Table 1. Poxvirus complete genomic sequences

Poxvirus complete genomic sequences. The A+T contents are as described by authors or have been calculated using the program Composition from the GCG package. ITRs, Inverted terminal repeats.

Genus	Species		Strain	Genome (bp)	A+T (%)	ITRs (kbp)	Reference	GenBank accession no
Chordopoxviruses								
Orthopoxvirus	*Vaccinia virus*	VV	Copenhagen	191 636	66·6	12·0	Goebel *et al.* (1990)	M35027
			MVA	177 923	66·6	9·8	Antoine *et al.* (1998)	U94848
			Tian Tan	189 274	66·8	7·5	–	AF095689
	Variola virus	VAR	Bangladesh-1975	186 102	66·3	0·7	Massung *et al.* (1994)	L22579
			India-1967	185 578	67·3		Shchelkunov *et al.* (1995)	X69198
			Garcia-1966	186 986	67·3	0·6	Shchelkunov *et al.* (2000)	Y16780
	Monkeypox virus	MPV	Zaire-96-I-16	196 858	68·9	6·4	Shchelkunov *et al.* (2001)	AF380138
	Ectromelia virus	ECT	Moscow	209 771	66·8		–	AF012825
			Naval	207 620	66·9	7·4	–	*
	Camelpox virus	CMPV	CM-S	202 185	66·9	6	Gubser & Smith (2002)	AY009089
			M-96	205 719	66·8	7·7	Afonso *et al.* (2002b)	AF438165
	Cowpox virus	CPV	Brighton Red	224 501	66·6		–	AF482758
Leporipoxvirus	*Myxoma virus*	MYX	Lausanne	161 774	56·4	11·5	Cameron *et al.* (1999)	AF170726
	Shope fibroma virus	SFV	Kaza	159 857	60·5	12·4	Willer *et al.* (1999)	AF170722
Avipoxvirus	*Fowlpox virus*	FPV		288 539	69·0	9·5	Afonso *et al.* (2000)	AF198100
Capripoxvirus	*Lumpy skin disease virus*	LSDV	Neethling 2490	150 733	73·0	2·4	Tulman *et al.* (2001)	AF325528
	Goatpox virus	GTPV	Pellor	149 935	75·0	2·3	Tulman *et al.* (2002)	AY077835
			G20-LKV	149 695	75·0	2·2	Tulman *et al.* (2002)	AY077836
	Sheeppox virus	SPPV	TU-V02127	149 995	75·0	2·2	Tulman *et al.* (2002)	AY077832
			Strain A	150 057	75·0	2·3	Tulman *et al.* (2002)	AY077833
			NISKHI	149 662	75·0	2·1	Tulman *et al.* (2002)	AY077834
Suipoxvirus	*Swinepox virus*	SWPV	17077-99	146 454	72·0	3·7	Afonso *et al.* (2002a)	AF410153
Molluscipoxvirus	*Molluscum contagiosum virus*	MCV	Subtype 1	190 289	36·0	4·7	Senkevich *et al.* (1996)	U60315
Yatapoxvirus	*Yaba-like disease virus*	YLDV		144 575	73·0	1·9	Lee *et al.* (2001)	AJ293568
Entomopoxvirus								
EnPV B	*Melanoplus sanguinipes*	EnPVm		236 120	81·7	7	Afonso *et al.* (1999)	AF063866
	Amsacta moorei	EnPVa		232 392	81·5	9·4	Bawden *et al.* (2000)	AF250284

*ECT-NAV sequence data were produced at the Sanger Institute in collaboration with Antonio Alcamí of the Department of Pathology, University of Cambridge and can be obtained from ftp://ftp.sanger.ac.uk/pub/pathogens/ev/.

The nucleotide coordinates of the aligned sequences are: VV modified virus Ankara (VV-MVA, 30791–129436; Antoine *et al.*, 1998); VV-COP (38938–137809; Goebel *et al.*, 1990); VAR-India-1970 (VAR-IND, 26672–125468; Shchelkunov *et al.*, 1995); VAR-Garcia-1966 (VAR-GAR, 27657–126473; Shchelkunov *et al.*, 2000); VAR-Bangladesh-1975 (VAR-BSH, 27296–126098; Massung *et al.*, 1994); CMPV-M-96 (38694–137667; Afonso *et al.*, 2002b); CMPV-CMS (36904–135787; Gubser & Smith, 2002); MPV (36079–134701; Shchelkunov *et al.*, 2001); ECT-Naval (ECT-NAV, 42379–141280; www.sanger.ac.uk); and CPV-BR (52492–151512; accession no. AF482758). The transition/transversion ratio was estimated at 2·72 using the program Treepuzzle (Strimmer & von Haeseler, 1996) and this value was used for the construction of a maximum-likelihood distance matrix using the DNADIST program from the PHYLIP package version 3.6 (alpha2) (Felsenstein, 1989), with the F84 model of nucleotide substitution (Felsenstein, 1984).

Alignment of multiple protein sequences. Amino acid sequences of individual proteins were aligned by a method similar to that used previously (McGeoch *et al.*, 2000). Sequences of each protein from the different viruses were aligned separately by each of the programs CLUSTALW (Thompson *et al.*, 1994), Dialign2 (Morgenstern, 1999) and Multalin (Corpet, 1988) using the default program parameters. For each set of data, combined alignments were produced by re-extracting the individual sequences from these three alignments, with retention of the gapping characters introduced by each program. Then a new alignment was made from this triple set of sequences using the program CLUSTALW. All positions in the combined alignment that had a gap in any sequence were then excised, thus deleting both unanimously placed gaps and sections where the three primary alignments were in conflict.

Phylogenetic analysis of multiple protein sequences. The amino sequences of (i) 17 proteins that are conserved in all ChPVs or (ii) 12 proteins that are present in 12 OPVs were aligned individually and positions with gaps were excluded from the alignments as described above. The individual alignments were concatenated to form a single file of 10 451 (i) and 2316 (ii) amino acids, respectively. The most appropriate model of sequence evolution was determined

using the program Treepuzzle. For both concatenated alignments, neighbour-joining trees (Saitou & Nei, 1987) were constructed using the programs Prodist and Neighbor from the PHYLIP package version 3.0 (Felsenstein, 1989). These were then used as starting trees to construct the maximum-likelihood trees (Felsenstein, 1973) using the program ProML according to the Jones–Taylor–Thornton model (Jones *et al.*, 1992) with gamma distribution. Bootstrap analyses (Felsenstein, 1985) were performed on both trees using the programs SEQBOOT (1000 random replicates, random number seed = 133333), Protdist and CONSENSE.

Phylogenetic analysis of multiple OPV DNA sequences. The nucleotide sequences of 12 genes present in the terminal region of 12 OPVs were aligned individually with the program CLUSTALW version 1.8 (Thompson *et al.*, 1994). Positions with gaps were excluded from the alignment by manual inspection and individual alignments were concatenated to form a single file of 7233 (all genes), 4170 (genes present in the left end of the genome) and 3063 (genes present in the right end of the genome) nucleotides. For all concatenated alignments, neighbour-joining trees were constructed using the program PAUP* (Swofford, 2003) and these were used as starting trees for the construction of maximum-likelihood trees (Felsenstein, 1973) implemented using PAUP*. The model of nucleotide substitution used, as determined with Modeltest (Posada & Crandall, 1998), was the General Time Reversible (GTR) model with gamma distribution and proportion of invariable sites (shape parameter of the gamma distribution = 0·7082; proportion of invariable sites = 0·5676). The robustness of trees was evaluated by bootstrap analysis of the neighbour-joining trees, with 1000 rounds of replication, using PAUP*.

Similarity analysis. An analysis of the similarity of the terminal region of the CPV-BR and ECT-NAV genomes was done on the concatenated alignment used for phylogenetic alignment of OPV DNA sequences by using the program SimPlot (Ray, 1997). Genetic similarity was calculated according to the F84 model of evolution with a transition/transversion rate of 1·95.

RESULTS

Poxvirus gene content and genome organization

The sequences of 26 poxviruses have been determined (Table 1) and these include a member of each ChPV genus, except genus *Parapoxvirus*, and EnPVs *A* and *C*. A comparison showed that the general organization of the ChPV genome is conserved, with the central region (Fig. 1) encoding very similar proteins for RNA and DNA synthesis, protein processing, virion assembly and structural proteins. In contrast, genes encoded by terminal regions are more divergent between different ChPV genera, species within a genus and even strains of the same species. Many of these genes are non-essential for virus growth *in vitro*, and encode proteins affecting host range, virulence or interaction with the host immune system (Moss, 2001). Despite these similarities in ChPV genomes, the length varies from an estimated 139 kb in *Orf virus* to 289 kb in *Fowlpox virus* (FPV) and the A + T content varies from 75 % in the genus *Capripoxvirus* to 36 % in the genus *Parapoxvirus* (Moss, 2001) (Table 1). In contrast, EnPV genomes are more divergent and the gene order differs from ChPVs and between different EnPV genera (Afonso *et al.*, 1999; Bawden *et al.*, 2000).

Comparison of sequenced poxviruses identified 90 genes that are conserved in all ChPVs and this number is reduced to 49 by including two EnPVs (Table 2), consistent with a recent report (Upton *et al.*, 2003). Amongst ChPVs, these 90 conserved genes are all located within the central 100 kb region of the genome (Fig. 1). In contrast, no gene in the terminal region of any ChPV is conserved throughout the subfamily. If the sequence of VV strain Tian Tan (TT) (accession no. AF095689) is also included, the number of conserved genes drops to 80 due to gene fragmentation. The ten genes (*F12L*, *F15L*, *G6R*, *G8R*, *L5R*, *H4L*, *D4R*, *D11L*, *A4L*, *A28L* – named using VV-COP nomenclature) that are broken in VV-TT but conserved in all other ChPVs encode proteins that are essential for virus transcription (H4L, G8R and D11L), or form part of the intracellular mature virus (IMV) core (A4L), or are required for virus dissemination (F12L) (van Eijl *et al.*, 2002). These observations suggest strongly that there are sequencing errors in the VV-TT genome sequence. Consistent with this view, an alignment of the central 100 kb of 11 OPV genomes (VV-COP, VV-MVA, VV-TT, MPV-Zaire, VAR-BSH, VAR-IND, VAR-GAR, CMPV-CM-S, CMPV-M-96, ECT-NAV and CPV-BR) revealed many genes that contained multiple positions at which the sequence of 10 OPVs is identical but at which VV-TT differed (see supplementary data at JGV Online: http://vir.sgmjournals.org). For these reasons the sequence of VV-TT is excluded from most analyses. Recently, Upton *et al.* (2003) also reported errors in the VV-TT genome and demonstrated by resequencing that many of the genes reported to be disrupted were in fact complete.

An alignment of the central region of the genome from viruses from seven ChPV genera [VV, *Myxoma virus* (MYX), *Yaba-like disease virus* (YLDV), *Lumpy skin disease virus* (LSDV), *Swinepox virus* (SWPV), FPV and *Molluscum contagiosum virus* (MCV)] is shown in Fig. 1. The VV genome is taken as the reference against which others are compared and each VV gene is shown as a vertical bar. Genes that are conserved in all ChPVs (Table 2) are shown in grey. Genes shown in black are conserved in at least two ChPV genera. For viruses from other genera, only genes that differ from VV are illustrated and genes unique to individual viruses are shown in colour. This shows that, in this region of the genome, the *OPV*, *Leporipoxvirus*, *Yatapoxvirus*, *Suipoxvirus* and *Capripoxvirus* genera contain only 3, 2, 1, 1 and 0 unique genes, respectively, indicating that these viruses are closely related. In contrast, MCV and FPV encode 40 or 33 unique genes within the central part of their genome, suggesting that these viruses are more divergent.

Fig. 1 also demonstrates that in the central region of ChPV genomes, the overall gene order and content are very well conserved between the the OPV, *Leporipoxvirus*, *Yatapoxvirus*, *Capripoxvirus*, *Suipoxvirus* and *Molluscipoxvirus* genera. A notable feature is the presence of a gene (*C7L* in VV-COP,

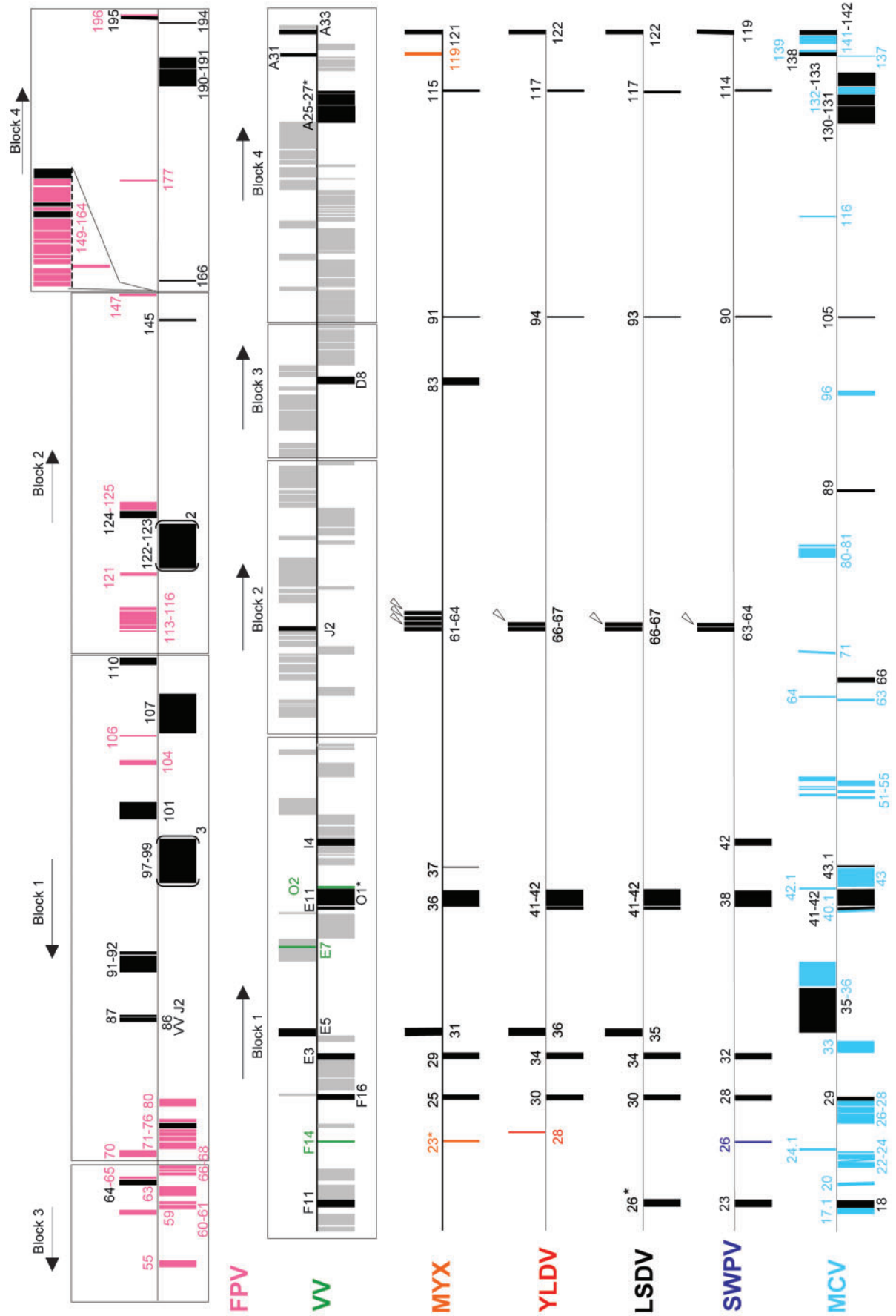
Block 4
Block 2
Block 1
Block 3
FPV
VV
MYX
YLDV
LSDV
SWPV
MCV

Fig. 1 arrowheads) in the central region of the *Leporipoxvirus*, *Yatapoxvirus*, *Capripoxvirus* and *Suipoxvirus* genomes, but in the terminal region of OPVs. In some viruses, this gene has been triplicated. This indicates that the genera *Leporipoxvirus*, *Yatapoxvirus*, *Capripoxvirus* and *Suipoxvirus* form a subgroup that is distinct from the OPVs.

In contrast, the genomes of FPV (*Avipoxvirus*) and MCV (*Molluscipoxvirus*) are divergent from other ChPV genera and contain many unique genes. FPV also shows rearrangement of the conserved genes. Whereas all the other ChPV genera have a conserved gene order, blocks of FPV genes have been transposed and/or inverted. The blocks of VV genes that run left to right 1, 2, 3 and 4 are present in FPV in the order 3, 1, 2 and 4 with blocks 1 and 3 in inverted orientation (Afonso *et al.*, 2000). This observation suggests that the FPV genome is most divergent compared to the other ChPVs.

Fig. 1 does not include the EnPVs because these are too divergent from ChPVs by genome size, gene arrangement and gene sequence similarity. Previous work showed that EnPVs are also divergent amongst themselves (Afonso *et al.*, 1999; Bawden *et al.*, 2000), which suggests that EnPVs diverged from ChPVs before the ChPVs evolved into distinct genera. These results also agree with previous suggestions that the orthopteran and lepidopteran members of genus *B* of EnPV might be split into separate genera (Afonso *et al.*, 1999; Bawden *et al.*, 2000). The overall amino acid identity of the 17 proteins we used for phylogenetic analysis of ChPVs (next section) is between 26·0 % and 29·9 % when comparing either EnPV with any ChPV, and only 55 % between the two EnPVs. For comparison, this value is 98·7 % between two members of the OPV genus (VV-COP and VAR-BSH), 94·8 % between the two leporipoxviruses (MYX and *Shope fibroma virus* (SFV)] and ~94 % between different capripoxviruses (data not shown). Because EnPVs are too divergent to be compared to ChPVs, or even to provide a reliable root, they were omitted from further analysis.

Phylogenetic relationships

Previously, the evolutionary relationships of poxviruses had been investigated based largely on genome collinearity and the nucleotide or amino acid sequence alignment of a few genes or proteins. However, rigorous phylogenetic studies using DNA and protein sequences from multiple genes for each virus are lacking. To establish phylogenetic relationships for ChPVs, we have compared (i) multiple protein sequences that are conserved in all sequenced ChPVs, (ii) DNA and protein sequences from terminal regions of the genomes that are conserved in OPVs, and (iii) DNA sequences from the central 100 kb of OPV genomes.

To compare the different ChPVs we selected 17 out of the 49 proteins conserved in all poxviruses. These were aligned for one or more member of each genus, nine viruses in total: SWPV, YLDV, VV-COP, VAR-BSH, MYX, SFV, MCV, LSDV and FPV. The use of several protein sequences to produces a single tree is more likely to represent the species tree accurately than a tree constructed with any single sequence. Previously, phylogenetic trees for single OPV proteins gave variable topologies (Afonso *et al.*, 2002b; Gubser & Smith, 2002). Similarly, others reported inconsistent tree topologies using single genes from closely related species (Huelsenbeck & Bull, 1996). The proteins chosen for analysis (VV-COP E9L, I7L, I8R, G9R, J3R, J6R, H2R, H4L, H6R, D1R, D5R, D6R, D11L, D13L, A7L, A16L and A24R) were selected to represent enzymes that are essential for transcription or DNA replication, and structural components of new virions (Table 2). All selected proteins are of similar length in the different viruses and are well conserved. Alignments were made for individual proteins, these alignments were edited and the sequences were concatenated into a single file that was used to construct a maximum-likelihood tree (Methods; Fig. 2). A tree drawn using the neighbour-joining method gave similar data (data not shown). The branch structure of the maximum-likelihood tree is unequivocal.

The phylogenetic tree shows that ChPVs divide into four main groupings. The first two, the *Molluscipox* (MCV) and *Avipox* (FPV) genera, each group separately. FPV has 113 unique genes, presumably derived from its avian host(s), and has the largest ChPV genome (288 kb), another feature that distinguishes it from the other ChPV members. The FPV genome is not collinear with other ChPVs and within the central region contains gene families and individual genes related to cellular genes that are found within terminal regions of other ChPVs (Fig. 1) (Afonso *et al.*, 2000). The 17 FPV proteins examined show a mean amino acid identity of 61·3 % to 62·0 % with all other ChPVs but show no greater similarity to any specific genus. A distinctive feature of FPV is that it is the only ChPV not to contain a counterpart of

Fig. 1. Gene content of the central region of seven ChPV genomes. VV-COP is used as the reference against which other viruses are compared. Only the prototype of each genus is represented. For each genome, the genes (vertical bars) are located on the top strand when transcribed rightwards and on the bottom strand when transcribed leftwards. Genes that are members of the minimal gene complement (the 90 genes conserved in all ChPVs) are shown in grey for VV-COP. Genes conserved between two or more viruses from different genera (but are not one of the 90 conserved genes) are shown in black and genes that are unique to one genus are represented in the same colour as the virus name. Genes that are present in some members only of one specific genus, but are fragmented or absent in other members, are indicated with an asterisk. Block 1, counterparts of VV-COP *F9L–G4L*; block 2, counterparts of VV-COP *G5R–D2L*; block 3, counterparts of VV-COP *D3R–A1L*; block 4, counterparts of VV-COP *A2L–A34R*. The arrows illustrate the orientation of the blocks within the VV-COP and FPV genomes. Arrowheads indicate the positions of orthologues of VV-COP gene *C7L*.

Table 2. Minimal gene complement of chordopoxviruses

The 90 genes that are present in all sequenced ChPVs are listed together with their function where known. Genes are named after their VV-COP counterpart. Genes also present in the two EnPVs are highlighted. IMV, intracellular mature virus; IEV, intracellular enveloped virus; EEV, extracellular enveloped virus.

ORF	Putative function	ORF	Putative function
F9L	Unknown	H6R	DNA topoisomerase I
F10L	IMV serine-threonine protein kinase	H7R	Unknown
F12L	IEV protein	D1R	mRNA capping enzyme, large subunit
F13L	EEV protein/phospholipase	D2L	IMV core protein
F15L	Unknown	D3R	IMV core protein
F17R	IMV core phosphoprotein, VP11/DNA-binding protein	D4R	Uracil-DNA glycosylase
E1L	Poly(A) polymerase catalytic subunit	D5R	Nucleoside triphosphatase
E2L	Unknown	D6R	Early transcription factor small subunit, VETF-1
E4L	Poly(A) polymerase catalytic subunit, rpo30/VITF-1	D7R	RNA polymerase subunit rpo18
E6R	Unknown	D9R	29 kDa mutT-like protein
E8R	Unknown	D10R	29 kDa mutT-like protein, negative regulator of gene expression
E9L	DNA polymerase	D11L	Nucleoside triphosphate phosphohydrolase I
E10R	IMV membrane-associated protein	D12L	mRNA capping enzyme small subunit, intermediate transcription factor, VITF
I1L	IMV core/DNA-binding protein	D13L	IMV protein, rifampicin resistance
I2L	Unknown	A1L	Late transcription factor/VLTF-2
I3L	Phosphoprotein, binds ssDNA	A2L	Late transcription factor/VLTF-3
I5L	IMV structural protein, VP13K	A2·5L	Thioredoxin-like protein
I6L	Unknown	A3L	IMV major core protein, P4b
I7L	IMV core protein	A4L	IMV core protein
I8R	Nucleoside triphosphate phosphohydrolase II, RNA helicase, NTPase	A5R	RNA polymerase subunit rpo19
G1L	Metallo-endoproteinase/virion morphogenesis	A6L	Unknown
G2R	Late transcription/IBT-dependent protein	A7L	Early transcription factor large subunit, VETF
G3L	Unknown	A8R	Intermediate transcription factor, VITF-3
G4L	Glutaredoxin 2, membrane protein, virion morphogenesis	A9L	IMV protein, role in morphogenesis
G5R	Unknown	A10L	IMV major core protein P4a
G5·5R	RNA polymerase subunit rpo7	A11R	Unknown
G6R	Unknown	A12L	IMV core protein
G7L	IMV core protein, VP16K	A13L	IMV membrane-associated protein/p8
G8R	Late transcription factor, VLTF-1	A14L	IMV protein, p16
G9R	Myristyl protein	A14·5L	IMV protein
L1R	Myristylated IMV protein	A15L	Unknown
L2R	Unknown	A16L	Myristyl protein
L3L	Unknown	A17L	IMV membrane protein, morphogenesis factor
L4R	IMV core protein VP8, DNA and RNA-binding protein	A18R	DNA helicase, DNA-dependent ATPase, transcript release factor
L5R	Unknown	A19L	Unknown
J1R	Dimeric virion protein	A20R	DNA polymerase processivity factor
J3R	Poly(A) polymerase stimulatory subunit, VP39	A21L	Unknown
J4R	RNA polymerase subunit rpo22	A22R	Holliday junction resolvase
J5L	Unknown	A23R	Intermediate transcription factor, VITF-3
J6R	RNA polymerase subunit rpo147	A24R	RNA polymerase subunit rpo132
H1L	Tyrosine-serine phosphatase, virion maturation	A28L	Unknown
H2R	Unknown	A29L	RNA polymerase subunit rpo35
H3L	Immunodominant IMV envelope protein p35	A30L	Unknown
H4L	RNA polymerase-associated transcription specificity factor, RAP94	A32L	ATP- and GTP-binding motif A, DNA packaging
H5R	Late transcription factor, VLTF-4	A34R	EEV glycoprotein

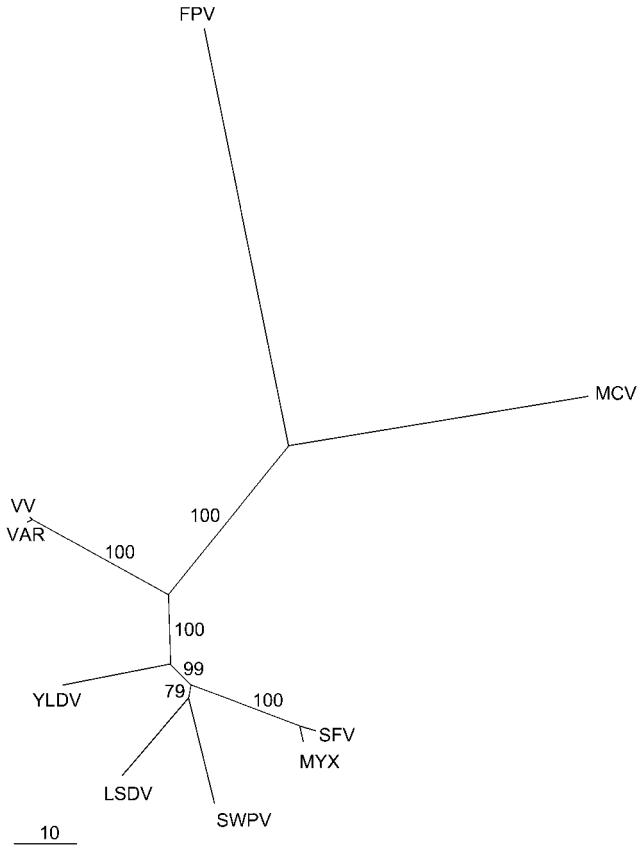

Fig. 2. Phylogenetic analysis of ChPVs. The amino acid sequences of 17 poxvirus proteins (VV-COP E9L, I7L, I8R, G9R, J3R, J6R, H2R, H4L, H6R, D1R, D5R, D6R, D11L, D13L, A7L, A16L and A24R) were aligned as described in Methods. The maximum-likelihood tree was obtained using the program ProML from the PHYLIP package version 3.0 (JTT model with gamma distribution) and is shown in an unrooted format. The bootstrap values from 1000 replica samplings and the divergence scale (substitutions per site) are indicated.

VV-COP gene *A33R*, which encodes an extracellular enveloped virus (EEV) envelope protein.

After the eradication of smallpox, MCV remains the only endemic human-specific poxvirus and it is divergent from other ChPVs. MCV is well adapted to humans (it survives long term and causes little morbidity) and this is reflected by 70 unique proteins (including several immunomodulators) and the lack of most of the immunomodulators encoded by other poxviruses (Senkevich *et al.*, 1996). Previous phylogenetic studies carried out using single MCV proteins resulted in different tree topologies depending on the gene and the method of tree construction, with MCV grouping individually in most cases but also together with FPV (Senkevich *et al.*, 1996). Data presented in Fig. 1 show that MCV and FPV are distinct ChPVs that have diverged from other ChPVs long ago. Like FPV, the conserved MCV proteins show a modest percentage amino acid identity with other ChPVs (range 61·7 % to 63·4 %) and MCV is no closer to any specific ChPV genus. MCV is the only ChPV not to contain a counterpart of VV-COP gene *J2R* (thymidine kinase).

The third and largest cluster of ChPVs includes the *Yatapoxvirus* (YLDV), *Capripoxvirus* (LSDV), *Suipoxvirus* (SWPV) and *Leporipoxvirus* (SFV and MYX) genera. Within this group, SFV and MYX, which cluster strongly together, also group with SWPV and LSDV, whereas YLDV is slightly more divergent. The genomes of all these viruses are relatively well conserved in gene content, gene arrangement (Fig. 1) and amino acid identity (data not shown). Notably, unlike OPVs, these viruses all contain at least one counterpart of VV gene *C7L* within the central region of their genomes (Fig. 1, arrowheads) between counterparts of VV-COP genes *J2R* and *J3R*. For these four ChPV genera, the overall amino acid identity is highest between SWPV and LSDV (79·2 %). This is consistent with the tree topology (Fig. 2) and suggests that SWPV and LSDV have evolved from a common ancestor.

The fourth ChPV group is genus OPV, illustrated by VV-COP and VAR-BSH, which group together tightly and separately from other ChPVs. The scale of the phylogenetic tree shows how closely related these viruses are compared to, for instance, the different leporipoxviruses and suggests that this group of viruses diverged more recently than members of other ChPV genera. Another feature that distinguishes the OPVs from other ChPVs is the presence of genes equivalent to VV-COP *F14L*, *E7L* and *O2L* within the central conserved region. ChPVs from outside the OPV genus lack these genes.

In summary, the comparison of poxviruses from different ChPV genera with each other using two different computational methods gave a robust phylogenetic tree. The only ChPV genus not represented here is *Parapoxvirus*, for which a complete genome sequence is awaited. The OPV genus is now considered in more detail.

Orthopoxvirus phylogeny

OPVs are the most intensively studied poxviruses. The reasons for this are largely historical: smallpox, caused by VAR, used to be a very serious disease of mankind; CPV is thought to have been used by Jenner in 1798 as the first human vaccine; and VV is the smallpox vaccine that was used in the modern era. Currently, there are 12 complete OPV genome sequences from 6 species (VAR, VV, CPV, MPV, ECT, CMPV) (Table 1). The origin and evolutionary relationships of these viruses are ill-defined, although it was demonstrated recently that CMPV-CMS and VAR are closely related (Gubser & Smith, 2002).

To compare the phylogeny of OPVs we selected genes that are conserved in the terminal genome regions of these viruses, where genes have greater divergence between species. Within the terminal regions of 10 sequenced OPVs, only 12 out of ~100 genes are present in every virus and this is due in part to mutations causing disruption of several genes in different

OPVs. The conserved genes are VV-COP *C7L* (host range function; Perkus *et al.*, 1990), *C6L* (unknown), *N1L* (intracellular virulence factor; Bartlett *et al.*, 2002), *K2L* (serine proteinase inhibitor, SPI-3; Law & Smith, 1992), *F2L* (dUTPase; McGeoch, 1990), *F4L* (ribonucleotide reductase, small subunit; Slabaugh *et al.*, 1988), *F6L* (unknown), *F8L* (unknown), *A56R* (the haemagglutinin glycoprotein that forms part of the extracellular enveloped virus outer envelope; Shida, 1986), *B1R* (protein kinase; Banham & Smith, 1992), *B5R* (EEV glycoprotein; Engelstad *et al.*, 1992) and *B15R* (unknown). These genes are known or likely to have an important function.

The proteins encoded by the selected genes were aligned and used to construct a maximum-likelihood tree (Methods, Fig. 3). Several facts may be deduced from the tree. First, the close relationship of CMPV and VAR is confirmed. The three strains of VAR each cluster together, as do the two strains of CMPV, and the VAR and CMPV clusters are more closely related to each other than to any other OPV species. Second, CPV-BR, MPV and ECT do not group closely with other OPVs. Lastly, although the two VV strains cluster closely together, the two CPV strains do not and they show remarkable divergence for two strains of the same virus species. It will be interesting to determine if other CPVs group predominantly with CPV-BR or GRI-90 or even form further divergent groups. These results suggest that classification of CPV as a single species within the genus OPV needs reconsideration.

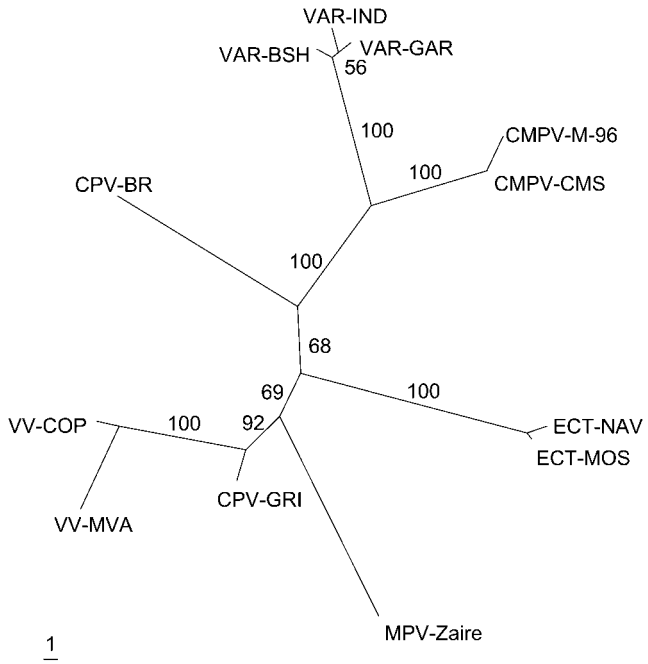

Fig. 3. Phylogenetic tree of 12 OPVs obtained by the maximum-likelihood method using protein sequences. The amino acid sequences of 12 OPV proteins (VV-COP C6L, C7L, N1L, K2L, F2L, F4L, F6L, F8L, A56R, B1R, B5R, B15R) encoded in the terminal regions of the genomes were aligned as described in Methods. The maximum-likelihood tree was obtained using the program ProML from the PHYLIP package version 3.0 (JTT model with gamma distribution) and is shown in an unrooted format. The bootstrap values from 1000 replica samplings and the divergence scale (substitutions per site) are indicated.

Next, we examined the DNA sequences of these genes from eight OPVs (Fig. 4a). Four OPVs were omitted because the extra strains of VAR, CMPV and ECT cluster closely with other members of their species (Fig. 3). This analysis confirmed the relationships observed for the protein sequences in Fig. 3. CMPV and VAR are closely related, whereas CPV-BR, ECT and MPV are divergent. As in Fig. 3, the two strains of CPV are sufficiently divergent from one another to justify being classified as independent species. We also analysed if the relationships of genes from different ends of the genome were the same (Fig. 4b, c). These trees confirm the overall relationships, and show that CPV-GRI-90 is always closest to VV and divergent from CPV-BR. However, CPV-BR and ECT-NAV group differently in the two cases. Genes from the left end of the genome place CPV-BR closer to the VV subgroup, and ECT-NAV closer to VAR and CMPV (Fig. 4b). Conversely, genes at the right end show the opposite relationships (Fig. 4c). These results might suggest that recombination events have occurred in these viruses.

To investigate this further, the nucleotide sequences of CPV-BR or ECT-NAV genes from the left or right end of the genome were compared using SimPlot with the equivalent genes from either CMPV-CMS and VAR-BSH or VV-MVA and VV-COP. Consistent with Fig. 4(b, c), genes from the left end of the CPV-BR genome are closer to the VV strains, whereas genes from the right end are more related to the CMPV/VAR subgroup (Fig. 5a). In contrast, genes from the left end of ECT-NAV show a similar relationship to each subgroup, whereas at the right end genes are closer overall to the VV subgroup (Fig. 5b). These results confirm the different topologies of the trees shown in Fig. 4(b, c), and also provide evidence of recombination within the right end of the genomes of CPV-BR and ECT-NAV.

Finally, we analysed the relatedness of the central ~100 kb of the OPV genomes. This region shows >90 % nucleotide identity between all OPVs and encodes all genes conserved throughout the ChPV subfamily. Remarkably, the maximum difference in length for this region was 382 nucleotides between CPV-BR (longest) and MPV (shortest), and CMPV-CMS and VAR-BSH differ in length by only 82 nucleotides. Most differences were intergenic caused by repeated oligonucleotides in one virus or another. Similar repeated oligonucleotides had been reported between VV-COP and VV-WR (Smith *et al.*, 1991) and VV-WR and VAR-Harvey (Aguado *et al.*, 1992). An alignment of this region is presented as supplementary data (JGV Online: http://vir.sgmjournals.org).

From this alignment, a distance matrix was constructed and is shown in Table 3. As expected the genetic distances

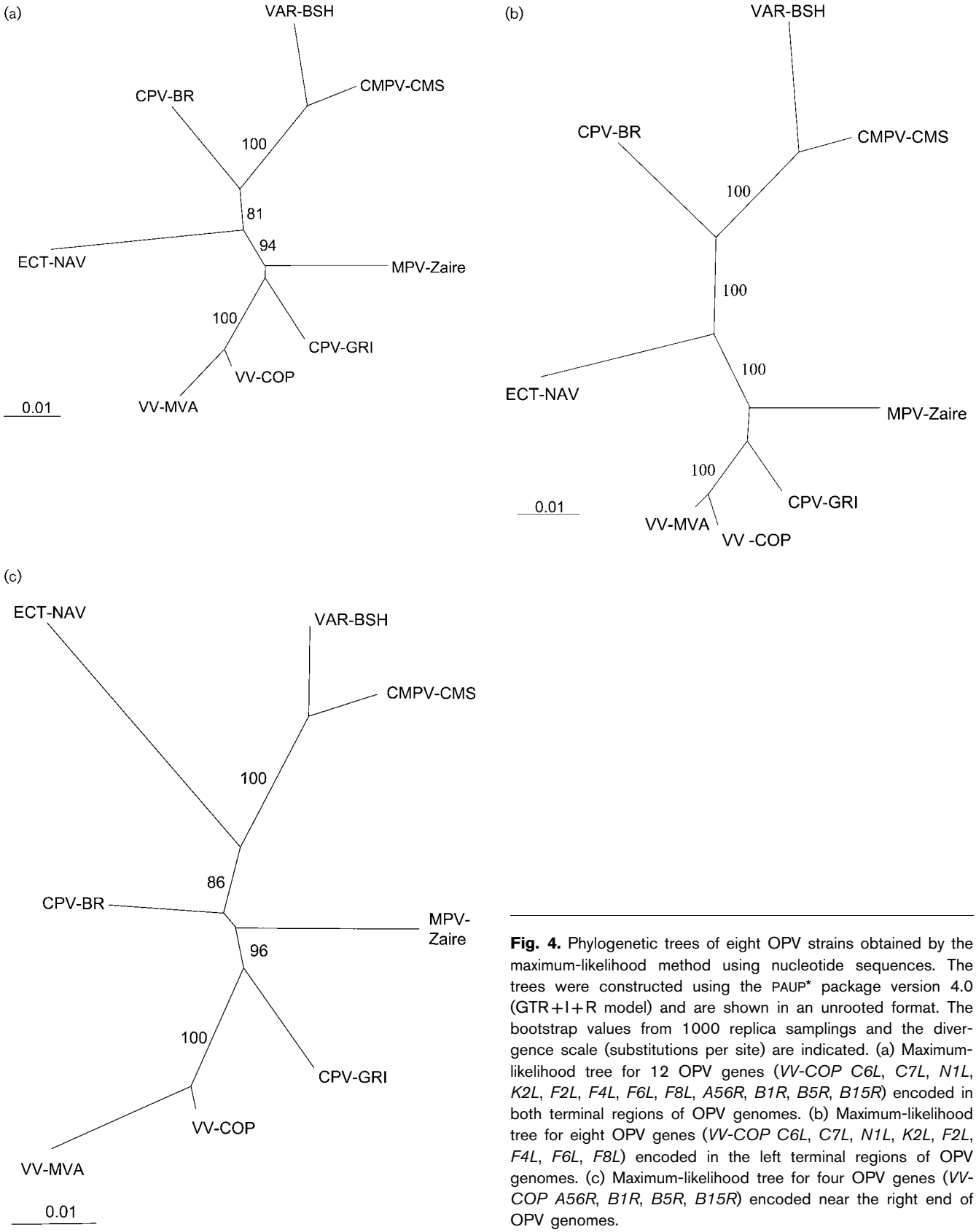

Fig. 4. Phylogenetic trees of eight OPV strains obtained by the maximum-likelihood method using nucleotide sequences. The trees were constructed using the PAUP* package version 4.0 (GTR+I+R model) and are shown in an unrooted format. The bootstrap values from 1000 replica samplings and the divergence scale (substitutions per site) are indicated. (a) Maximum-likelihood tree for 12 OPV genes (*VV-COP C6L*, *C7L*, *N1L*, *K2L*, *F2L*, *F4L*, *F6L*, *F8L*, *A56R*, *B1R*, *B5R*, *B15R*) encoded in both terminal regions of OPV genomes. (b) Maximum-likelihood tree for eight OPV genes (*VV-COP C6L*, *C7L*, *N1L*, *K2L*, *F2L*, *F4L*, *F6L*, *F8L*) encoded in the left terminal regions of OPV genomes. (c) Maximum-likelihood tree for four OPV genes (*VV-COP A56R*, *B1R*, *B5R*, *B15R*) encoded near the right end of OPV genomes.

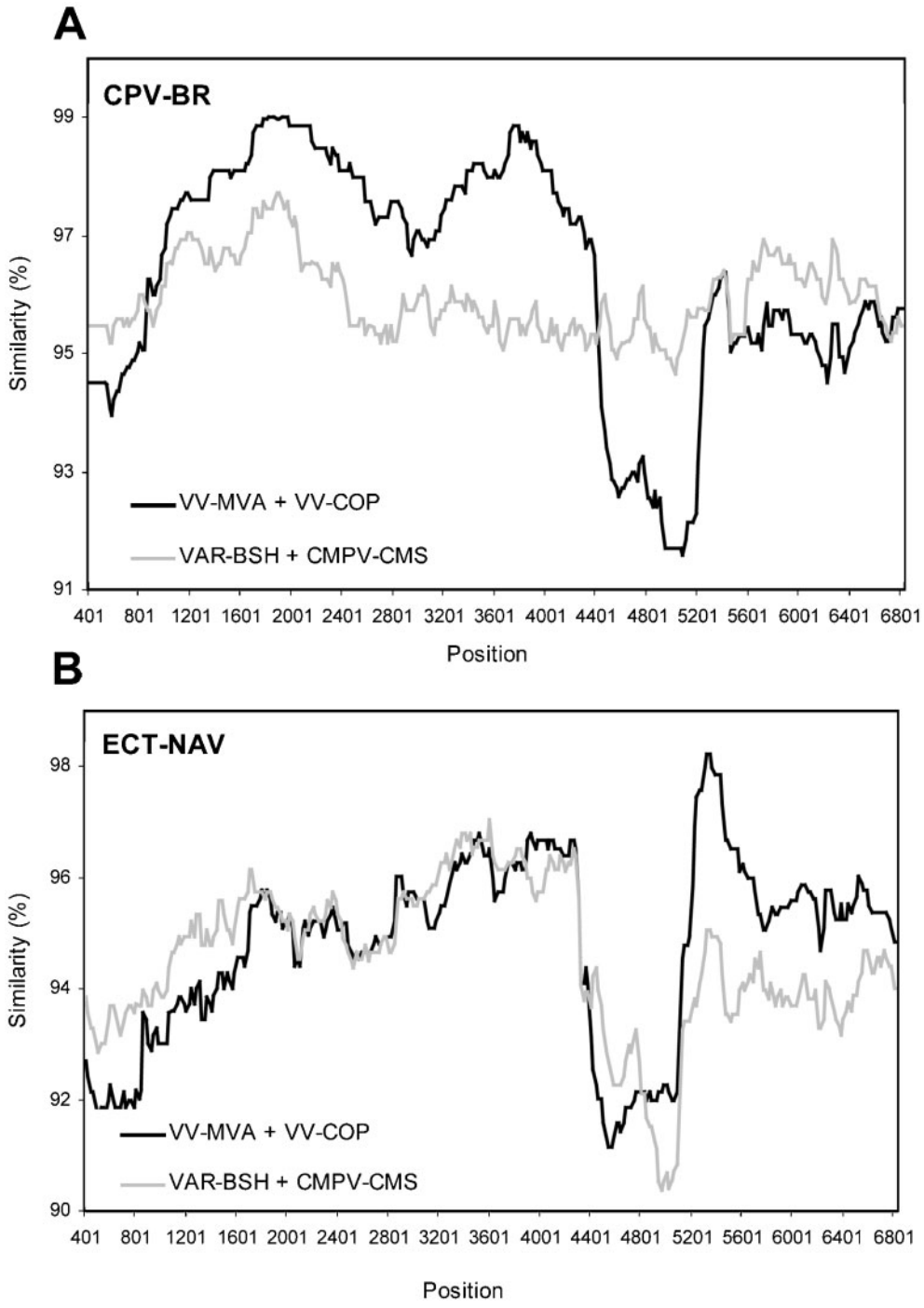

Fig. 5. Evidence for recombination between OPV genomes. The percent nucleotide similarity between each query sequence CPV-BR (A) and ECT-NAV (B) and the VAR-BSH/CMPV-CMS and the VV-MVA/VV-COP groups of sequences is represented by a grey and a black line, respectively. A sliding window of 800 bases was used along the nucleotide alignment, with an increment of 20 bases. In this alignment, the first 4170 and last 3063 nucleotides are from the left or right ends of the genome, respectively. The similarity values between the query and the reference sequences were computed according to the F84 model of evolution, with a transition/transversion ratio of 1·95.

between OPVs are low (0·0150–0·0354) compared to the genetic distance for the same region of SFV and MYX, which was eightfold greater than between CMPV and VAR (Gubser & Smith, 2002). This suggests that OPVs have diverged more recently than the leporipoxviruses from their common ancestor. When comparing different OPV species, the

Table 3. DNA distance matrix

DNA sequences between counterparts of VV-COP genes *F9L* and *A24R* were aligned using the program CLUSTALW version 1.8 (Thompson *et al.*, 1994. See supplementary data) and a DNA distance matrix was constructed using the program DNADIST from the PHYLIP package version 3.6 (alpha2) (Felsenstein, 1989) as described in Methods.

		VV			VAR			CMPV		MPV	ECT	CP
		TT	COP	MVA	IND	BSH	GAR	CMS	M-96	Zaire	NAV	BR
VV	TT											
	COP	0.0058										
	MVA	0.0076	0.0060									
VAR	IND	0.0268	0.0259	0.0255								
	BSH	0.0268	0.0259	0.0254	0.0012							
	GAR	0.0267	0.0258	0.0253	0.0036	0.0030						
CMPV	CMS	0.0222	0.0213	0.0209	0.0154	0.0150	0.0151					
	M-96	0.0223	0.0214	0.0209	0.0155	0.0151	0.0152	0.0002				
MPV	Zaire	0.0216	0.0205	0.0204	0.0307	0.0305	0.0304	0.0261	0.0262			
ECT	NAV	0.0285	0.0274	0.0271	0.0354	0.0351	0.0352	0.0313	0.0315	0.0298		
CPV	BR	0.0249	0.0238	0.0237	0.0329	0.0326	0.0325	0.0285	0.0286	0.0269	0.0281	

genetic distance was lowest between strains of CMPV and VAR (0·0151–0·0154), and highest between ECT and VAR (0·0352–0·0354) (Table 3).

DISCUSSION

Comparison of 26 sequenced poxviruses has identified 90 genes that are conserved in all ChPVs and 49 genes that are conserved in all poxviruses. These numbers are in agreement with Upton *et al.* (2003). All the conserved ChPVs genes lie within the central region of the genome.

The phylogenetic relationships of the sequenced poxviruses were examined. The genome organization, and percentage amino acid sequence identities, showed that the two sequenced EnPVs are distinct from ChPVs and quite divergent from each other so that they might be classified in separate genera (Afonso *et al.*, 1999; Bawden *et al.*, 2000). Within the ChPVs, the most divergent virus is FPV (genus *Avipoxvirus*) followed by MCV (genus *Molluscipoxvirus*). This overall conclusion is reached by comparison of the size of these genomes, the number of unique genes, the gene arrangement (Fig. 1) and phylogenetic analysis of the amino acid sequences of 17 conserved proteins (Fig. 2). Avipoxviruses are the only ChPVs to infect birds and MCV is a strictly human pathogen; both viruses have evolved unique immunomodulatory proteins that enable them to counteract the immune system of their hosts.

The other ChPVs show two clusters: the first includes the genera *Leporipoxvirus*, *Capripoxvirus*, *Suipoxvirus* and *Yatapoxvirus*, within which SWPV and LSDV share a common ancestor; the second is genus OPV. The first group of viruses have smaller genomes [range 144 575 (YLDV) to 161 774 (MYX); Table 1] and few unique genes in the central region of the genome (Fig. 1). These genomes also all have the orthologue of the VV-COP gene *C7L* within the central region of the genome, whereas in OPVs this is present near the left end. The relatively small size (139 kb) of the *Orf virus* genome (Mercer *et al.*, 1987) suggests genus *Parapoxvirus* might group within this cluster.

The OPVs represent a closely related group of viruses with larger genomes [177 923 (VV-MVA) to 224 501 (CPV-BR)]. Note that VV-MVA has lost about 30 kb compared to the parental Ankara strain (Meyer *et al.*, 1991).

Overall, the phylogenetic analyses of poxvirus protein sequences give relationships between genera (Fig. 2) that are consistent with relationships deduced from comparisons of genome organization and gene content (Fig. 1).

A more detailed analysis of OPVs revealed that the central regions of these genomes are very similar. Here the genes are collinear and over ~100 kb CMPV-CMS and VAR-BSH differ in length by only 82 nucleotides. Because of the high degree of similarity of genes and proteins from this region of OPVs, we compared the phylogenetic relationships of these viruses by using the DNA and protein sequences of genes from the terminal regions of the genome (Figs 3–5). These analyses established phylogenetic relationships but also indicated that these viruses have undergone recombination during their evolution. Data presented show that CMPV and VAR are closely related species, as reported previously (Gubser & Smith, 2002), but CPV-BR, MPV and ECT are divergent and do not group closely with other OPVs. For MPV, this is despite it causing a disease in man similar to smallpox. MPV occurs naturally in western and central Africa but is poorly transmitted from person to person and human infections tend to be limited local outbreaks. It has been proposed that rodents are the natural host for MPV (Fenner *et al.*, 1988).

The origin of the most extensively studied OPV, VV, is obscure. If Jenner used CPV as the first smallpox vaccine in 1796, sometime between then and 1939 when A. W. Downie reported that the available smallpox vaccine strains were a

distinct OPV that became known as VV (Downie, 1939a, b), CPV was replaced by VV as the smallpox vaccine. This probably had occurred by the late nineteenth century because the smallpox vaccine taken to the USA in 1856 and which became the New York City Board of Health Vaccine is VV not CPV. Similarly, pathologists who studied cells infected by smallpox vaccines used in the late nineteenth century reported the eosinophilic B type inclusion bodies made by VV and CPV, but failed to report the much more obvious A type inclusion bodies that are made by CPV but not VV. Therefore, by this time VV was probably already used as the smallpox vaccine. The possible origin of VV was discussed by Baxby (1981). He proposed VV was a distinct OPV species that was isolated from a species in which it was no longer endemic. Horsepox was one possibility. In support of this proposal, early vaccinators took vaccine from horses when the supply of CPV (a relatively rare disease) was scarce and one strain of VV (Ankara) was isolated from a horse. The recent demonstration that the VV-WR interferon-γ receptor binds and neutralizes equine interferon-γ (Symons *et al.*, 2002) is also consistent with this proposal. However, given the broad host range of VV and the broad species specificity of the VV interferon-γ receptor, these observations might be interpreted only as VV being a zoonosis in horses and that its natural host lies elsewhere. Phylogenetic comparisons indicate VV is not derived recently from either VAR or CPV but that it is closer to CPV-GRI-90 than CPV-BR.

An interesting feature of OPV genomes is the presence of many genes that are intact in one virus but fragmented in another. Comparison between OPV genomic sequences reveals that fragmented ORFs (i) are located mainly within 50 kbp of either terminus, (ii) represent 33/206 ORFs (16 %) in CMPV (Gubser & Smith, 2002) and (iii) are fewest in CPV (Shchelkunov *et al.*, 1998). Many of these gene fragments are unlikely to encode functional protein, so their retention is surprising. This might reflect the relative stability of OPV genomes and the apparent lack of a stringent packaging limit on poxvirus DNA (Smith & Moss, 1983; Perkus *et al.*, 1991). Another interpretation is that some OPVs are relatively recent pathogens of their respective hosts and have diverged from an ancestral virus recently in an evolutionary timescale, with disruption of some genes accompanying their divergence. The much closer relationship between OPVs than between leporipoxviruses is consistent with this interpretation (Fig. 2, Table 3). With time these viruses might have lost some of these non-essential gene fragments. In contrast to OPVs, within the leporipoxviruses SFV contains eight fragmented genes compared to MYX, but MYX contains only one fragmented gene compared to SFV (Cameron *et al.*, 1999; Willer *et al.*, 1999).

Finally, the two CPV strains are sufficiently divergent to justify their reclassification as different OPV species and it will be interesting to determine how other CPV strains isolated from different geographical locations compare with the two strains analysed to date. CPV-GRI-90 was proposed as a possible ancestral virus for OPVs because its genes in the terminal genome regions are mostly complete (Shchelkunov *et al.*, 1998), whereas other OPVs possess broken fragments of these genes. By this criterion, CPV-BR also might be considered close to a possible ancestral virus.

ACKNOWLEDGEMENTS

The work was supported by grants from The Wellcome Trust. G. L. S. is a Wellcome Trust Principal Research Fellow. S. H. is supported by the Health Protection Agency.

REFERENCES

Afonso, C. L., Tulman, E. R., Lu, Z., Oma, E., Kutish, G. F. & Rock, D. L. (1999). The genome of *Melanoplus sanguinipes* entomopoxvirus. *J Virol* **73**, 533–552.

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

Afonso, C. L., Tulman, E. R., Lu, Z., Zsak, L., Osorio, F. A., Balinsky, C., Kutish, G. F. & Rock, D. L. (2002a). The genome of swinepox virus. *J Virol* **76**, 783–790.

Afonso, C. L., Tulman, E. R., Lu, Z., Zsak, L., Sandybaev, N. T., Kerembekova, U. Z., Zaitsev, V. L., Kutish, G. F. & Rock, D. L. (2002b). The genome of camelpox virus. *Virology* **295**, 1–9.

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. *J Gen Virol* **73**, 2887–2902.

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.

Banham, A. H. & Smith, G. L. (1992). Vaccinia virus gene B1R encodes a 34-kDa serine/threonine protein kinase that localizes in cytoplasmic factories and is packaged into virions. *Virology* **191**, 803–812.

Bartlett, N., Symons, J. A., Tscharke, D. C. & Smith, G. L. (2002). The vaccinia virus N1L protein is an intracellular homodimer that promotes virulence. *J Gen Virol* **83**, 1965–1976.

Bawden, A. L., Glassberg, K. J., Diggans, J., Shaw, R., Farmerie, W. & Moyer, R. W. (2000). Complete genomic sequence of the *Amsacta moorei* entomopoxvirus: analysis and comparison with other poxviruses. *Virology* **274**, 120–139.

Baxby, D. (1981). *Jenner's Smallpox Vaccine*. London: Heinemann.

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.

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16**, 10881–10890.

Downie, A. W. (1939a). Immunological relationship of the virus of spontaneous cowpox to vaccinia virus. *Br J Exp Pathol* **20**, 158–176.

Downie, A. W. (1939b). A study of the lesions produced experimentally by cowpox virus. *J Pathol Bacteriol* **48**, 361–379.

Engelstad, M., Howard, S. T. & Smith, G. L. (1992). A constitutively expressed vaccinia gene encodes a 42-kDa glycoprotein related to complement control factors that forms part of the extracellular virus envelope. *Virology* **188**, 801–810.

Felsenstein, J. (1973). Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am J Hum Genet* **25**, 471–492.

Felsenstein, J. (1984). Distance methods for inferring phylogenies: a justification. *Evolution* **38**, 16–24.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.

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.

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.

Gubser, C. & Smith, G. L. (2002). The sequence of camelpox virus shows it is most closely related to variola virus, the cause of smallpox. *J Gen Virol* **83**, 855–872.

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

Jones, D. T., Taylor, W. R. & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* **8**, 275–282.

Law, K. M. & Smith, G. L. (1992). A vaccinia serine protease inhibitor which prevents virus induced cell fusion. *J Gen Virol* **73**, 549–557.

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

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.

McGeoch, D. J. (1990). Protein sequence comparisons show that the 'pseudoproteases' encoded by poxviruses and certain retroviruses belong to the deoxyuridine triphosphatase family. *Nucleic Acids Res* **18**, 4105–4110.

McGeoch, D. J., Dolan, A. & Ralph, A. C. (2000). Toward a comprehensive phylogeny for mammalian and avian herpesviruses. *J Virol* **74**, 10401–10406.

Mercer, A. A., Fraser, K., Barns, G. & Robinson, A. J. (1987). The structure and cloning of orf virus DNA. *Virology* **157**, 1–12.

Meyer, H., Sutter, G. & Mayr, A. (1991). Mapping of deletions in the genome of the highly attenuated vaccinia virus MVA and their influence on virulence. *J Gen Virol* **72**, 1031–1038.

Morgenstern, B. (1999). DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. *Bioinformatics* **15**, 211–218.

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.

Perkus, M. E., Goebel, S. J., Davis, S. W., Johnson, G. P., Limbach, K., Norton, E. K. & Paoletti, E. (1990). Vaccinia virus host range genes. *Virology* **179**, 276–286.

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.

Posada, D. & Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818.

Ray, S. C. (1997). SimPlot for Windows 95/NT, version 1.2.2. http://wwwwelchjhuedu/ ~sray/download.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.

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

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 Res* **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. & 7 other authors (2000). Alastrim smallpox variola minor virus genome DNA sequences. *Virology* **266**, 361–386.

Shchelkunov, S. N., Totmenin, A. V., Babkin, I. V. & 11 other authors (2001). Human monkeypox and smallpox viruses: genomic comparison. *FEBS Lett* **509**, 66–70.

Shida, H. (1986). Nucleotide sequence of the vaccinia virus hemagglutinin gene. *Virology* **150**, 451–462.

Slabaugh, M., Roseman, N., Davis, R. & Mathews, C. (1988). Vaccinia virus-encoded ribonucleotide reductase: sequence conservation of the gene for the small subunit and its amplification in hydroxyurea-resistant mutants. *J Virol* **62**, 519–527.

Smith, G. L. & Moss, B. (1983). Infectious poxvirus vectors have capacity for at least 25 000 base pairs of foreign DNA. *Gene* **25**, 21–28.

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. *J Gen Virol* **72**, 1349–1376.

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

Swofford, D. L. (2003). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sunderland, MA: Sinauer Associates.

Symons, J. A., Tscharke, D. C., Price, N. & Smith, G. L. (2002). A study of the vaccinia virus interferon-γ receptor and its contribution to virus virulence. *J Gen Virol* **83**, 1953–1964.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). Improved sensitivity of profile searches through the use of sequence weights and gap excision. *Comput Appl Biosci* **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. *J Virol* **75**, 7122–7130.

Tulman, E. R., Afonso, C. L., Lu, Z. & 7 other authors (2002). The genomes of sheeppox and goatpox viruses. *J Virol* **76**, 6054–6061.

Upton, C., Slack, S., Hunter, A. L., Ehlers, A. & Roper, R. L. (2003). Poxvirus orthologous clusters: towards defining the minimum essential poxvirus genome. *J Virol* **77**, 7590–7600.

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

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