2007

Comparative Genetic Analysis of Genomic DNA Sequences of Two Human Isolates of *Tanapox virus*

Steven H. Nazarian
Biotherapeutics Research Group, Robarts Research Institute, and Department of Microbiology and Immunology, University of Western Ontario, London, Ontario N6G 2V4, Canada

John W. Barrett
Biotherapeutics Research Group, Robarts Research Institute, and Department of Microbiology and Immunology, University of Western Ontario, London, Ontario N6G 2V4, Canada

A. Michael Frace
Biotechnology Core Facility Branch, Division of Scientific Resources, National Center for Preparedness, Detection, and Control of Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30329, USA

Melissa Olsen-Rasmussen
Biotechnology Core Facility Branch, Division of Scientific Resources, National Center for Preparedness, Detection, and Control of Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30329, USA

Marina Khristova
Biotechnology Core Facility Branch, Division of Scientific Resources, National Center for Preparedness, Detection, and Control of Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30329, USA

See next page for additional authors

Available online at www.sciencedirect.com
ScienceDirect
Virus Research 129 (2007) 11–25
Virus Research
www.elsevier.com/locate/virusres

Comparative genetic analysis of genomic DNA sequences of two human isolates of *Tanapox virus*☆

Steven H. Nazarian[a], John W. Barrett[a], A. Michael Frace[b], Melissa Olsen-Rasmussen[b], Marina Khristova[b], Mae Shaban[a], Sarah Neering[d], Yu Li[c], Inger K. Damon[c], Joseph J. Esposito[b], Karim Essani[d], Grant McFadden[a,*]

[a] *Biotherapeutics Research Group, Robarts Research Institute, and Department of Microbiology and Immunology, University of Western Ontario, London, Ontario N6G 2V4, Canada*
[b] *Biotechnology Core Facility Branch, Division of Scientific Resources, National Center for Preparedness, Detection, and Control of Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30329, USA*
[c] *Poxvirus and Rabiesvirus Branch, Division of Viral and Rickettsial Diseases, National Center for Zoonotic, Vector-Borne, and Enteric Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30329, USA*
[d] *Laboratory of Virology, Department of Biological Science, Western Michigan University, Kalamazoo, MI 49008, USA*

Received 10 March 2007; received in revised form 1 May 2007; accepted 1 May 2007
Available online 14 June 2007

Abstract

Members of the genus *Yatapoxvirus*, which include *Tanapox virus* (TPV) and *Yaba monkey tumor virus*, infect primates including humans. Two strains of TPV isolated 50 years apart from patients infected from the equatorial region of Africa have been sequenced. The original isolate from a human case in the Tana River Valley, Kenya, in 1957 (TPV-Kenya) and an isolate from an infected traveler in the Republic of Congo in 2004 (TPV-RoC). Although isolated 50 years apart the genomes were highly conserved. The genomes differed at only 35 of 144,565 nucleotide positions (99.98% identical). We predict that TPV-RoC encodes 155 ORFs, however a single transversion (at nucleotide 10241) in TPV-Kenya resulted in the coding capacity for two predicted ORFs (11.1L and 11.2L) in comparison to a single ORF (11L) in TPV-RoC. The genomes of TPV are A + T rich (73%) and 96% of the sequence encodes predicted ORFs. Comparative genomic analysis identified several features shared with other chordopoxviruses. A conserved sequence within the terminal inverted repeat region that is also present in the other members of the *Yatapoxviruses* as well as members of the *Capripoxviruses*, *Swinepox virus* and an unclassified *Deerpox virus* suggests the existence of a conserved near-terminal sequence secondary structure. Two previously unidentified gene families were annotated that are represented by ORF TPV28L, which matched homologues in certain other chordopoxviruses, and TPV42.5L, which is highly conserved among currently reported chordopoxvirus sequences.

Keywords: *Tanapox*; *Yatapoxvirus*; Poxvirus; Comparative genomics

1. Introduction

Poxviruses constitute two sub-families, *Chordopoxvirinae* and *Entomopoxvirinae*, which infect a wide range of vertebrate and insect hosts, respectively (Buller et al., 2005). Characteristic features of poxviruses include a cytoplasmic life cycle, a large virion size and large genome compared to other viruses (Moss, 2007). Poxviruses contain a linear, double-stranded DNA genome with palindromic, covalently-closed ends. Sequenced poxvirus genomes vary from ~134 to ~360 kbp in length and 130 to 328 open reading frames (ORFs) can be predicted from the sequences. At the ends of poxvirus DNA genomes are mirror image terminal inverted repeat (TIR) regions, however, among different strains the lengths of the TIR regions vary from a few hundred nucleotides, such as in *Variola virus* (VARV), to approximately 12 kbp, such as in *Shope fibroma virus* (SHFV). In general, the chordopoxvirus

☆ Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the funding agencies. Use of trade names or commercial sources is for identification only and does not imply endorsement by the funding agencies.

* Corresponding author. Present address: University of Florida, 1600 SW Archer Road, ARB Room R4-295, P.O. Box 100332, Gainesville, FL 32610, USA. Tel.: +1 352 273 6852; fax: +1 352 273 6849.
E-mail address: grantmcf@ufl.edu (G. McFadden).

0168-1702/$ – see front matter
doi:10.1016/j.virusres.2007.05.001

genome is organized so that the essential housekeeping genes, including those required for transcription, replication and morphogenesis, are located within the central region of the genome. Genes nearer to the DNA ends are generally more variable and encode for a wide variety of functions, including genes dedicated to ensure virus replication within the host by modulating the host innate and adaptive immune response (Seet et al., 2003).

The genomic DNA sequences of over 100 different poxvirus strains have been determined. The particular poxvirus sequences used in the present study are listed in Table 1. All of the sequences used are available through GenBank and two curated poxvirus sites—www.poxvirus.org/ and www.biovirus.org/.

There are two species in the genus *Yatapoxvirus*: *Yaba monkey tumor virus* (YMTV) and *Tanapox virus* (TPV). Both species have caused human infection. A previously sequenced poxvirus, *Yaba-like disease virus* (YLDV) (Lee et al., 2001), is a TPV from an infected non-human primate (Brunetti et al., 2003; Espana et al., 1971; Esposito and Fenner, 2001; McNulty et al., 1968). TPV and YLDV are suspected to be transmitted by arthropod vectors and both produce a similar rash illness, fever with prodromal symptoms that is followed by the development of few nodular skin lesions (Downie and Espana, 1972; Damon, 2007; Knight et al., 1989). In contrast, YMTV produces a very distinct disease, primarily in non-human primates, which is characterized by epidermal histiocytomas of the head and limbs (Downie and Espana, 1972; Knight et al., 1989). The observed biological differences between YMTV and YLDV are likely explained by the 82% nucleotide identity and an approximately 10 kbp deletion from YMTV compared to YLDV (Brunetti et al., 2003; Downie and Espana, 1972; Espana et al., 1971; Knight et al., 1989; Lee et al., 2001).

A previous study, in which the genome of YMTV was sequenced, examined the conservation of certain gene families that were found to be below the usual 50 codon cutoff (Brunetti et al., 2003). To further this research, two isolates of TPV were sequenced; one is the first isolate from a 1957 human outbreak of TPV in the Tana River valley in Kenya (TPV-Kenya) and the other was isolated from an infected college student traveling in the Congo Basin in the Republic of Congo (TPV-RoC) (Dhar et al., 2004). The current study is a comparative genomic analysis of these two isolates of TPV that are from discrete geographic regions of Africa and isolated 50 years apart.

Table 1
Summary information of poxvirus sequences used in this study

Genus	Virus	Strain	Virus short form	Genome size (bp)	Accession number	Reference
Yatapox	Tanapox virus	RoC	TPV-RoC	144553	EF420157	This study
	Tanapox virus	Kenya	TPV-Kenya	144565	EF420156	This study
	Yaba-like disease virus		YLDV	144575	NC_002642	Lee et al. (2001)
	Yaba monkey tumor virus	Roswell Park-Yohn	YMTV	134721	NC_005179	Brunetti et al. (2003)
Capripox	Goatpox virus	Pellor	GTPV	149599	NC_004003	Tulman et al. (2002)
	Lumpy skin disease virus	Neethling 2490	LSDV	150773	NC_003027	Tulman et al. (2001)
	Sheeppox virus	A	SHPV	150057	AY077833	Tulman et al. (2002)
Suipox	Swinepox virus	Nebraska 17077-99	SWPV	146454	NC_003389	Afonso et al. (2002)[a]
Leporipox	Myxoma virus	Lausanne	MYXV	161773	NC_001132	Cameron et al. (1999)
	Shope fibroma virus	Kasza	SHFV	159857	NC_001266	Willer et al. (1999)
Molluscipox	Molluscum contagiosum virus	Subtype 1	MOCV	190289	NC_001731	Senkevich et al. (1996)
Orthopox	Camelpox virus	CMS	CMPV	202205	AY009089	Gubser and Smith (2002)
	Cowpox virus	Brighton Red	CPXV	224499	C_003663	
	Ectromelia virus	Moscow	ECTV	209771	NC_004105	
	Horsepox virus	MNR-76	HSPV	212633	DQ792504	Tulman et al. (2006)
	Monkeypox virus	Zaire	MPXV	196858	NC_003310	Shchelkunov et al. (2001)
	Raccoonpox virus		RCNV	ND[a]	M23018	
	Taterapox virus	Dahomey 1968	TATV	198050	NC_008291	
	Vaccinia virus	Western Reserve	VACV	194711	NC_006998	
	Variola virus	India 1964 7125 Vellor	VARV	186127	DQ437586	Esposito et al. (2006)
Avipox	Canarypox virus	ATCC VR111	CNPV	359853	NC_005309	Tulman et al. (2004)
	Fowlpox virus	Iowa	FWPV	288539	NC_002188	Afonso et al. (2000)
Parapox	Bovine papular stomatitis virus	BV-AR02	BPSV	134431	NC_005337	Delhon et al. (2004)
	Orf virus	NZ2	ORFV	137820	DQ184476	Mercer et al. (2006)
Unclassified	Crocodilepox virus		CRV	190054	NC_008030	Afonso et al. (2006)
	Deerpox virus	W-1170-84	DPV	170560	AY689437	Afonso et al. (2005)

[a] Not determined.

Comparative genomics reported here reveals the similarities and differences within and without *Yatapoxviruses*. In particular, the genetic relationships of TPV with sequenced isolates of the genera *Capripoxvirus* and *Suipoxvirus* and an unclassified *Deerpox virus* are explored.

2. Materials and methods

2.1. Genomic sequencing

Sequencing was performed essentially as described elsewhere (Esposito et al., 2006). Briefly, TPV genomic DNA was extracted from cells infected with TPV-Kenya (Knight et al., 1989) and TPV-RoC (Dhar et al., 2004). The genomic DNA was used as template for production of a set of 14 overlapping polymerase chain reaction (PCR) amplicons that span virtually the entire viral genome. Amplicons of 10–12 kbp each were produced using the Expand High Fidelity PCR System (Roche Applied Science, Indianapolis, IN, USA). The product of eight identical PCR mixtures for each amplicon were pooled and treated with ExoSap-IT (USB Corporation, Cleveland, OH, USA) to reduce PCR errors in the amplicon templates, which were used for primer-walking cycle-sequencing reactions. Cycle sequencing reactions used Applied Biosystems (PE Biosystems, Foster City, CA, USA) Big-Dye 3.1 dye chemistry and ABI 3730XL DNA sequencers and the sequencing primers (Integrated DNA Technologies, Coralville, IA, USA) were designed to anneal approximately at every 400 bases across the templates, which enabled a nine-fold average sequence redundancy. To verify certain sequences, additional cycle sequencing involved direct sequencing from the full-length extracted genome DNA. Chromatogram data was assembled using Seqmerge (Wisconsin Package Version 10.3, Accelrys Inc., San Diego, CA, USA) and Phred/Phrap base-calling and assembly software and Consed for sequence editing (Balbas and Gosset, 2001; Domi and Moss, 2002). ORFs were identified and alignments performed using MacVector 6.5.3 (Oxford Molecular Ltd.).

2.2. Estimation of nucleotide substitution

TPV-Kenya and TPV-RoC were compared 25,000 bp at a time by using a base-by-base pairwise comparison matrix containing 144,565 nucleotide positions. Nucleotide differences were analyzed for transversions and transitions.

3. Results

Two TPV isolates from infected humans either living (TPV-Kenya) or traveling (TPV-RoC) through equatorial Africa were sequenced, which provided an opportunity to investigate the evo-

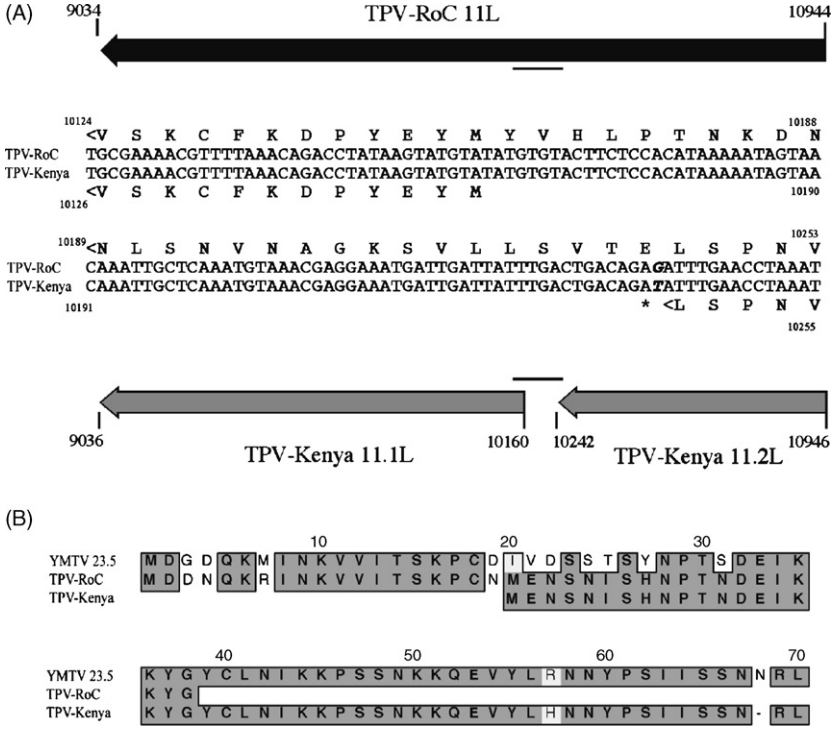

Fig. 1. ORF differences between TPV-RoC and TPV-Kenya. (A) The 11L gene from TPV-RoC (solid black arrow) is a single ORF. The same region from TPV-Kenya encodes two ORFs (solid gray arrows). The sequences between the comparable regions are identical except for a single transversion at position 10239 of TPV-RoC. This change results in a termination codon (* [TAG]) in the predicted transcript of TPV-Kenya instead of the incorporation of a glutamic acid (E [GAG]) as predicted for TPV-RoC. The single nucleotide change is bolded and italized. The sequence presented represents the minus strand. The predicted amino acid sequences for TPV-RoC or TPV-Kenya are indicated above or below the corresponding nucleotide sequence. Numbers indicate position with the respective genomes. The single black lines above or below the solid arrows show the region that is represented by the sequence comparison. (B) ClustalW alignment of the 23.5L ORF from YMTV, TPV-Kenya and TPV-RoC. Similar amino acids are shaded light grey and identical amino acids are shaded dark grey.

lutionary diversity of two TPVs that spanned 50 years and were from two different African countries.

3.1. Genome architecture of TPV

In GenBank there are sequences of TPV-Kenya that represent approximately 8 kbp of the total genome (GenBank accession numbers AY253325, AF245394 and AF153912); these sequences are about 98% identical to cognate sequences in a reported YLDV genome sequence (Lee et al., 2001). In order to sequence the two TPV isolates described here, PCR amplicon and cycle sequencing primers were designed by using the reported YLDV sequence. The determined sequences of TPV-Kenya and TPV-RoC comprised 144,565 and 144,553 bp, respectively, 96% of which encode for putative ORFs. Both viruses are 73% A + T-rich, which is consistent with the other sequenced yatapoxviruses (YLDV 73% A + T and YMTV 70% A + T) (Brunetti et al., 2003; Lee et al., 2001). By comparison with the YLDV sequences, the TPV sequences lack the putative concatemer resolution domain proximal to the hairpin-loop termini. However, the two TPV isolates were sequenced to within 20 bp of cognate reported YLDV genomic sequences (Lee et al., 2001).

TPV-Kenya and TPV-RoC encode 156 and 155 distinct ORFs, respectively (Table 2). All ORFs that were reported for YLDV are present in both isolates of TPV, with two exceptions—ORF 11L and ORF 23.5L (Fig. 1). TPV-Kenya 11L has a premature stop codon at codon 236, which results in a truncated ORF (Fig. 1A). Approximately 80 bp downstream of the 11L stop codon in TPV-Kenya, a putative ORF corresponding to the second half of the 11L ORF is present and may be transcribed as a distinct gene product. The two ORFs in TPV-Kenya are denoted 11.1L and 11.2L (Fig. 1A and Table 2). The two predicted ORFs have been identified previously and were annotated in GenBank. TPV-Kenya 11.1L was previously labeled TPV ORFL7R (accession number AAD46181) and TPV ORFL8R (accession number AAD46182). The ORF 11.2L is identical to TPV ORFL4R (accession number AAD46179), which indicates that this truncated ORF has been independently identified. In contrast, the 11L ORF in TPV-RoC is not truncated. The TPV-RoC 11L-encoded protein is an ankyrin repeat protein that contains a predicted F-box domain (Fig. 2B) (Mercer et

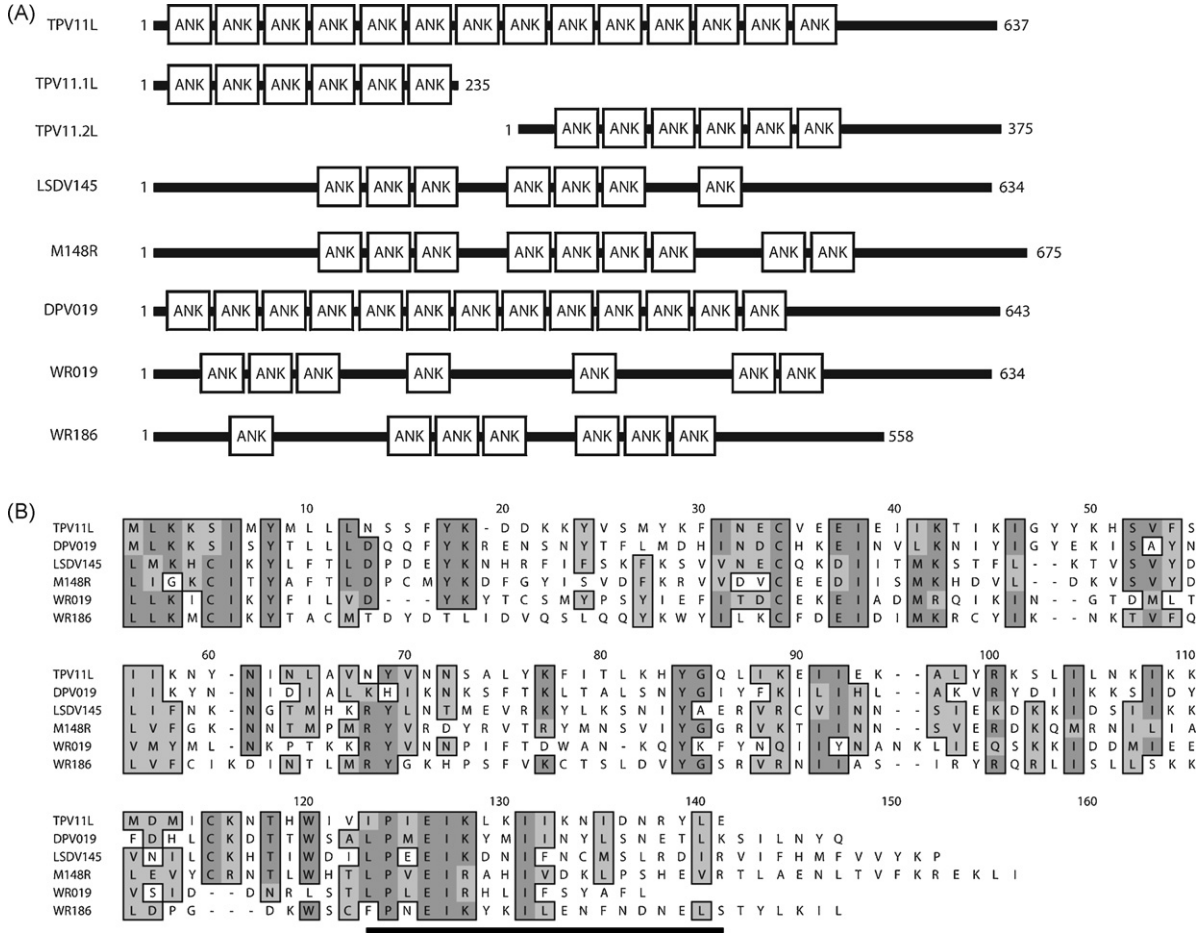

Fig. 2. Structural analysis of multiple ankyrin repeat-containing proteins. (A) Various homologous ankyrin repeat-containing proteins were identified using the c-terminal end of the TPV 11L protein as a query sequence. Predicted ankyrin repeats are indicated by the boxes (ANK). Numbers indicate the amino acid length of the various proteins from various poxviruses, including: TPV-RoC (TPV11L), TPV-Kenya (TPV11.1L and TPV11.2L), LSDV, MYXV (M148R), DPV and VACV (WR019 and WR186). (B) The C-terminal end of each of the proteins in the top panel is aligned. DPV019 and LSDV145 had an approximately 30 amino acid stretch, N-terminal to the aligned sequence that did not match and these residues are not included. The bold line underneath the alignment indicates the putative F-box domain that is complete in all sequences but WR019.

Table 2
Identification of the predicted open reading frames (ORFs) of TPV-Kenya and TPV-RoC

ORF	TPV-Kenya			TPV-RoC			Predicted structure or function	DPV[a]		LSDV[b]		VACV[c]		MYXV[d]		YMTV[e]	
	Codon		aa[f]	Codon		aa		ORF	Identity/ similarity	ORF	Identity/ similarity	ORF	Identity/ similarity	ORF	Identity/ similarity	ORF	Identity/ similarity
	Start	Stop		Start	Stop												
1L	1738	740	333	1738	740	333		DPV007	38/60	LSDV007	33/55	WR010	28/49			1L	71/84
2L	2868	1855	338	2868	1855	338	TNF binding protein	DPV008	38/53							2L	73/82
3L	3583	2918	222	3583	2918	222		DPV009	42/66	LSDV150		WR039	38/60				
4L	4329	3616	238	4329	3616	238		DPV011	35/62	LSDV009	35/56	WR029	33/62			4L	71/83
5L	4840	4373	156	4840	4373	156	LAP/PHD domain			LSDV010	45/59			M153R	39/56	5L	64/82
6L	5354	4908	149	5353	4907	149				LSDV001	37/57	WR196	41/63			6L	81/93
7L	6473	5421	351	6471	5419	351	Chemokine inhibitor	DPV013	49/70	LSDV011	40/66					7L	74/87
8L	7131	6493	213	7129	6491	213	Ankyrin repeat	DPV014	32/58	LSDV012	43/65			M149R	28/53	11L	27/53
9L	7848	7171	226	7846	7169	226	Virulence gene factor	DPV017	39/62			WR031	34/51	M154L	26/46		
10L	9021	7873	383	9019	7871	383	SERPIN/Spi3ortholog	DPV018	45/61			WR034	29/48	M008.1	31/52		
11L				10944	9034	637	Ankyrin repeat	DPV019	51/69			WR186	27/45			11L	79/93
11.1L	10160	9036	375				Ankyrin repeat	DPV019	51/69	LSDV148	23/45	WR186	27/46	M149R	29/49	11L	79/93
11.2L	10946	10242	235				Ankyrin repeat	DPV019	52/73	LSDV145	31/52	WR019	29/47	M148R	23/45	11L	82/94
12L	11232	10969	88	11230	10967	88	IF2α-like PKR inhibitor	DPV020	51/73	LSDV014	47/66	WR034	37/57			12L	75/88
13L	12121	11267	285	12119	11265	285	Monoglyceride lipase				WR038		50/70			13L	76/87
14L	12554	12147	136	12552	12145	136	IL-18 binding protein	DPV021	46/68	LSDV015	40/65					14L	55/72
15L	12815	12585	77	12813	12583	77	EGF-like growth factor			LSDV016	41/55			M010L	36/52		
16L	13345	12818	176	13343	12816	176	Mitochondria anti-apoptotic factor	DPV022	35/59	LSDV017	30/54			M011L	27/47	16L	64/80
17L	13821	13393	143	13819	13391	143	dUTPase	DPV023	62/74	LSDV018	64/76	WR041	58/73	M012L	63/80	17L	86/95
18L	14246	13866	127	14244	13864	127	Pyrin domain	DPV024	40/59					M013L	30/55		
19L	15846	14281	522	15844	14279	522	Kelch protein	DPV025	37/62	LSDV019	34/54	WR042	27/48	M014L	24/57	19L	76/89
20L	16848	15874	325	16846	15872	325	Ribonucleotide reductase	DPV026	77/89	LSDV020	78/89	WR043	77/87	M015L	75/86	20L	91/95
21L	17133	16879	85	17131	16877	85		DPV027	30/61	LSDV021	38/60			M016L	42/66	21L	65/84
22L	17480	17172	103	17478	17170	103										22L	34/55
23L	17703	17485	73	17701	17483	73											
23.5L	17949	17800	50	18005	17892	38		DPV029	59/68	LSDV023	68/85	WR047	43/69	M018L	58/76	23.5L	86/95
24L	18721	18080	214	18719	18078	214		DPV031	56/82	LSDV024	50/73	WR048	52/75	M019L	46/73	24L	84/93
25L	20036	18702	445	20034	18700	445	Serine/threonine protein kinase	DPV032	77/90	LSDV025	78/90	WR049	72/83	M020L	73/87	25L	90/96
26L	21988	20063	642	21986	20061	642	EEV maturation	DPV034	48/69	LSDV027	45/61	WR051	36/56	M021L	41/61	26L	70/86
27L	23133	22024	370	23131	22022	370	Palmitylated EEV envelope protein	DPV035	72/86	LSDV028	77/87	WR052	58/74	M022L	72/83	27L	90/94
28L	23332	23189	48	23330	23187	48											
28.5L	23529	23359	57	23527	23357	57										28.5L	67/84

Table 2 (*Continued*)

ORF	TPV-Kenya			TPV-RoC			Predicted structure or function	DPV[a]		LSDV[b]		VACV[c]		MYXV[d]		YMTV[e]	
	Codon		aa[f]	Codon		aa		ORF	Identity/ similarity	ORF	Identity/ similarity	ORF	Identity/ similarity	ORF	Identity/ similarity	ORF	Identity/ similarity
	Start	Stop		Start	Stop												
29L	24027	23584	148	24025	23582	148		DPV037	66/80	LSDV029	67/81	WR054	56/73	M024L	52/72	29L	82/92
30L	24738	24094	215	24736	24092	215		DPV038	41/61	LSDV030	38/59	WR055	41/59	M025L	33/55	30L	77/92
31R	24797	25111	105	24795	25109	105	DNA-binding virion core protein	DPV039	70/86	LSDV031	73/83	WR056	55/75	M026R	70/83	31R	90/95
32L	26523	25114	470	26521	25112	470	Poly(A) polymerase	DPV040	72/86	LSDV032	72/85	WR057	67/83	M027L	71/84	32L	91/98
33L	28597	26543	685	28595	26541	685		DPV041	50/72	LSDV033	44/66	WR058	40/62	M028L	45/66	33L	72/88
34L	29202	28660	181	29200	28658	181	dsRNA-binding	DPV042	45/64	LSDV034	49/69	WR059	38/55	M029L	58/75	34L	68/83
35L	29811	29245	189	29809	29243	189	RNA polymerase subunit rpo30	DPV043	68/83	LSDV036	69/81	WR060	71/85	M030L	67/82	35L	86/93
36R	29928	30983	352	29926	30981	352		DPV044	35/51	LSDV035	37/56	WR061	25/44	M031R	31/53	36R	74/88
37R	30987	32687	567	30985	32685	567		DPV045	70/85	LSDV037	70/85	WR062	61/81	M032R	63/81	37R	87/95
38R	32710	33513	268	32708	33511	268		DPV046	80/91	LSDV038	79/91	WR064	70/81	M033R	76/87	38R	93/97
39L	36533	33516	1006	36531	33514	1006	DNA polymerase	DPV047	70/84	LSDV039	70/85	WR065	69/82	M034R	70/84	39L	86/93
40R	36566	36847	94	36564	36845	94		DPV048	70/88	LSDV040	72/86	WR066	66/80	M035R	66/89	40R	88/97
41L	37236	36850	129	37234	36848	129				LSDV041	54/74	WR067	44/71			41L	74/87
42L	39262	37226	679	39260	37224	679		DPV050	49/69	LSDV042	42/65	WR068	39/61	M036L	41/65		
42.5L	39413	39324	30	39411	39322	30								M037L	67/82		
43L	40365	39433	311	40363	39431	311	DNA binding core protein	DPV051	73/87	LSDV043	69/86	WR070	68/83	M038L	69/87	43L	86/96
44L	40590	40369	74	40588	40367	74		DPV052	50/63	LSDV044	49/66	WR071	45/62	M039L	45/68	44L	79/89
45L	41391	40594	266	41389	40592	266	ssDNA-binding phosphoprotein	DPV053	64/83	LSDV045	61/77	WR072	55/72	M040L	58/81	45L	87/94
46L	41694	41458	79	41692	41456	79	Structural	DPV055	78/93	LSDV046	68/81	WR074	46/76	M041L	46/71	46L	86/96
47L	42872	41715	386	42870	41713	386		DPV056	55/78	LSDV047	54/72	WR075	54/74	M042L	50/71	47L	83/93
48L	44158	42872	429	44156	42870	429	Topoisomerase II	DPV057	77/90	LSDV048	76/88	WR076	69/85	M043L	73/85	48L	92/97
49R	44164	46191	676	44162	46189	676	Helicase	DPV058	62/78	LSDV049	65/80	WR077	58/76	M044R	57/75	49R	84/93
50L	47963	46194	590	47961	46192	590	Metalloprotease	DPV059	64/79	LSDV050	60/78	WR078	55/75	M45L	58/75	50L	82/92
51L	48295	47963	111	48293	47961	111		DPV061	55/75	LSDV052	43/66	WR079	44/65	M46L	54/73	51L	73/86
52R	48289	48954	222	48287	48952	222	Transcriptional elongation factor	DPV060	50/72	LSDV051	49/70	WR080	49/71	M47R	46/66	52R	85/94
53L	49301	48927	125	49299	48925	125	Glutaredoxin	DPV062	73/88	LSDV053	76/88	WR081	44/65	M48L	68/85	53L	99/100
54R	49304	50620	439	49302	50618	439		DPV063	53/69	LSDV054	51/70	WR082	44/61	M49R	47/67	54R	76/89
55R	50626	50814	63	50624	50812	63	RNA polymerase subunit rpo7	DPV064	85/93	LSDV055	85/95	WR083	79/88	M50R	85/93	55R	96/98
56R	50817	51335	173	50815	51333	173		DPV065	55/77	LSDV056	56/73	WR084	46/70	M51R	55/78	56R	80/89
57L	52483	51362	374	52481	51360	374	Virion core protein	DPV066	64/79	LSDV057	60/75	WR085	51/68	M52L	55/69	57L	82/91
58R	52513	53292	260	52511	53290	260	Late transcription factor VLTF-1	DPV067	88/98	LSDV058	87/97	WR086	83/94	M53R	83/94	58R	97/99
59R	53308	54309	334	53306	54307	334		DPV068	64/80	LSDV059	60/77	WR087	50/70	M54R	56/74	59R	80/90
60R	54313	55053	247	54311	55051	247	IMV membrane protein	DPV069	81/91	LSDV060	81/93	WR088	69/83	M55R	74/92	60R	91/96
61R	55075	55347	91	55073	55345	91		DPV070	45/67	LSDV061	45/70	WR089	30/55	M56R	24/50	61R	71/84
62L	56276	55329	316	56274	55327	316		DPV071	65/83	LSDV062	66/84	WR090	54/74	M57L	60/80	62L	86/95
63R	56301	57047	249	56299	57045	249	Core protein VP8	DPV072	78/87	LSDV063	78/89	WR091	60/82	M58R	76/87	63R	91/95

64R	57066	57449	128	57064	57447	128	IMV membrane protein	DPV073	51/75	LSDV064	59/77	WR092	46/67	M59R	52/70	64R	78/90
65R	57406	57882	159	57404	57880	159	Virion protein	DPV074	63/81	LSDV065	67/78	WR093	49/69	M60R	59/71	65R	83/92
66R	57882	58430	183	57880	58428	183	Thymidine kinase	DPV075	67/78	LSDV066	63/77	WR094	70/78	M61R	66/78	66R	82/90
67R	58481	59014	178	58479	59012	178	Host-range protein	DPV076	43/67	LSDV067	42/59	WR021	36/66	M62R	42/64	67R	80/92
68R	59084	60082	333	59081	60079	333	Poly-A polymerase small subunit	DPV077	73/89	LSDV068	76/90	WR095	71/87	M65R	73/87	68R	91/95
69R	60000	60554	185	59997	60551	185	RNA polymerase subunit rpo22	DPV078	76/87	LSDV069	77/90	WR096	74/87	M66R	71/85	69R	92/95
70L	60947	60537	137	60944	60534	137		DPV079	67/81	LSDV070	62/80	WR097	61/80	M67L	66/81	70L	82/94
71R	61038	64892	1285	61035	64889	1285	RNA polymerase subunit rpo147	DPV080	87/94	LSDV071	84/93	WR098	80/92	M68R	85/94	71R	93/97
72L	65407	64895	171	65404	64892	171	Dual specificity Ser/Thr and Tyr phosphatase	DPV081	81/91	LSDV072	75/89	WR099	63/83	M69L	76/88	72L	88/97
73R	65423	65992	190	65420	65989	190		DPV082	69/86	LSDV073	64/83	WR100	62/80	M70R	63/81	73R	84/94
74L	66969	66001	323	66966	65998	323	IMV envelope protein p35	DPV083	54/74	LSDV074	52/73	WR101	35/61	M71L	50/74	74L	80/91
75L	69366	66973	798	69363	66970	798	RNA polymerase-associated RAP94	DPV084	78/89	LSDV075	77/88	WR102	69/84	M72L	75/85	75L	91/97
76R	69525	70064	180	69522	70061	180	Late transcription factor VLTF-4	DPV085	48/68	LSDV076	43/64	WR103	41/57	M73R	44/65	76R	77/86
77R	70080	71024	315	70077	71021	315	DNA topoisomerase	DPV086	68/85	LSDV077	69/86	WR104	63/82	M74R	62/82	77R	84/93
78R	71043	71486	148	71040	71483	148		DPV087	51/67	LSDV078	52/69	WR105	39/68	M75R	51/69	78R	80/91
79R	71526	74045	840	71524	74043	840	mRNA capping enzyme large subunit	DPV088	71/87	LSDV079	72/86	WR106	68/84	M76R	69/84	79R	88/95
80L	74471	74013	153	74469	74011	153	Virion protein	DPV090	41/64	LSDV080	36/59	WR107	48/68	M77L	40/63	80L	74/91
81R	74470	75204	245	74468	75202	245	Virion protein	DPV089	41/61	LSDV081	36/61	WR108	38/56	M78R	30/54	81R	67/84
82R	75204	75857	218	75202	75855	218	Uracil DNA glycosylase	DPV091	76/88	LSDV082	72/87	WR109	69/88	M79R	72/87	82R	82/93
83R	75924	78281	786	75922	78279	786	NTPase	DPV092	80/92	LSDV083	78/91	WR110	70/85	M80R	77/90	83R	94/98
84R	78281	80185	635	78279	80183	635	Early transctription factor VETFs	DPV093	88/94	LSDV084	88/94	WR111	80/90	M81R	86/93	84R	95/99
85R	80221	80700	160	80219	80698	160	RNA polymerase subunit rpo18	DPV094	77/92	LSDV085	80/93	WR112	71/85	M82R	77/90	85R	94/98
86R	80748	81383	212	80746	81381	212	mutT motif	DPV095	70/84	LSDV086	68/82	WR114	61/77	M84R	58/77	86R	87/95
87R	81383	82147	255	81381	82145	255	mutT motif	DPV096	65/81	LSDV087	65/80	WR115	50/70	M85R	59/78	87R	89/96
88L	84038	82146	631	84036	82144	631	Transcription termination factor NPH-1	DPV097	75/89	LSDV088	75/88	WR116	70/86	M86L	71/86	88L	92/97

Table 2 (*Continued*)

ORF	TPV-Kenya			TPV-RoC			Predicted structure or function	DPV[a]		LSDV[b]		VACV[c]		MYXV[d]		YMTV[e]	
	Codon		aa[f]	Codon		aa		ORF	Identity/ similarity	ORF	Identity/ similarity	ORF	Identity/ similarity	ORF	Identity/ similarity	ORF	Identity/ similarity
	Start	Stop		Start	Stop												
89L	84933	84073	287	84931	84071	287	mRNA capping enzyme VITF	DPV098	80/91	LSDV089	77/88	WR117	74/89	M87L	77/88	89L	95/98
90L	86620	84965	552	86618	84963	552	Rifampin resistance protein	DPV099	80/92	LSDV090	80/91	WR118	73/86	M88L	77/90	90L	93/97
91L	87096	86647	150	87094	86645	150	Late transcription factor VLTF-2	DPV100	64/85	LSDV091	66/84	WR119	63/85	M89L	72/88	91L	87/95
92L	87794	87123	224	87792	87121	224	Late transcription factor VLTF-3	DPV101	88/95	LSDV092	84/93	WR120	84/95	M90L	86/94	92L	95/98
93L	88018	87794	75	88016	87792	75	Redox virion protein	DPV102	60/82	LSDV093	64/85	WR121	55/76	M91L	68/82	93L	84/93
94L	90001	88031	657	89999	88029	657	4b core protein	DPV103	79/90	LSDV094	73/87	WR122	64/80	M92L	75/87	94L	93/97
95L	90516	90061	152	90514	90059	152	Virion core protein	DPV104	44/62	LSDV095	36/55			M93L	31/55	95L	68/82
96R	90556	91059	168	90554	91057	168	RNA polymerase subunit rpo19	DPV105	64/81	LSDV096	68/85	WR124	62/78	M94R	62/81	96R	87/94
97L	92174	91062	371	92172	91060	371		DPV106	72/89	LSDV097	75/87	WR125	57/78	M95L	69/86	97L	92/98
98L	94333	92201	711	94331	92199	711	Early transcription factor, VETF1	DPV107	78/90	LSDV098	77/90	WR126	71/86	M96L	76/89	98L	91/96
99R	94386	95255	290	94384	95253	290	Intermediate transcription factor VITF-3	DPV108	65/81	LSDV099	67/82	WR127	61/80	M97R	68/83	99R	90/95
100L	95494	95258	79	95492	95256	79	IMV membrane protein	DPV109	86/91	LSDV100	75/84	WR128	72/81	M98L	78/90	100L	91/94
101L	98203	95498	902	98201	95496	902	Core protein P4a	DPV110	69/85	LSDV101	67/82	WR129	54/72	M99L	61/79	101L	92/96
102R	98218	99150	311	98216	99148	311		DPV111	78/89	LSDV102	76/89	WR130	55/75	M100R	75/88	102R	91/96
103L	99672	99166	169	99670	99164	169	Core protein	DPV112	58/72	LSDV103	55/70	WR131	46/63	M101L	77/86	103L	77/86
104L	99920	99717	68	99918	99715	68	IMV membrane protein	DPV113	57/75	LSDV104	56/79			M102L	50/70	104L	83/94
105L	100248	99970	93	100246	99968	93	IMV phosphoprotein	DPV114	86/94	LSDV105	78/90	WR133	61/77	M103L	73/83	105L	98/94
106L	100426	100268	53	100424	100266	53	IMV membrane virulence factor	DPV115	84/94	LSDV106	79/88	WR134	66/80	M104L	79/88	106L	100/100
107L	100700	100419	94	100698	100417	94		DPV116	50/69	LSDV107	52/71	WR135	52/67	M105L	51/72	107L	78/88
108L	101829	100687	381	101827	100685	381	IMV protein	DPV117	64/81	LSDV108	63/80	WR136	51/70	M106L	55/73	108L	79/87
109L	102418	101846	191	102416	101844	191	IMV membrane phosphoprotein	DPV118	75/90	LSDV109	61/78	WR137	41/63	M107L	57/73	109L	92/97
110R	102433	103869	479	102431	103867	479	DNA helicase	DPV119	62/81	LSDV110	58/76	WR138	58/76	M108R	61/79	110R	85/94
111L	104077	103856	74	104075	103854	74		DPV120	76/90	LSDV111	72/87	WR139	62/77	M109L	81/90	111L	81/94
112L	104410	104081	110	104408	104079	110	Fusion protein	DPV122	57/73	LSDV113	57/72	WR140	57/70	M110L	56/74	112L	82/91
113R	104409	105683	425	104407	105681	425	DNA polymerase processivity factor	DPV121	50/71	LSDV112	51/66	WR141	46/66	M111R	47/66	113R	76/89
114R	105695	106165	157	105693	106163	157	DNA processing	DPV123	72/88	LSDV114	67/83	WR142	67/86	M112R	63/83	114R	77/89
115R	106190	107338	383	106188	107336	383	Intermediate transcription factor VITF-3	DPV124	60/79	LSDV115	63/78	WR143	60/77	M113R	60/76	115R	84/91

116R	107343	110837	1165	107341	110835	1165	RNA polymerase subunit rpo132	DPV125	88/96	LSDV116	89/96	WR144	82/92	M114R	85/94	116R	94/98
117L	111283	110840	148	111281	110838	148	Fusion protein	DPV126	39/61	LSDV117	41/64	WR150	43/61	M115L	43/66	117L	41/64
118L	111703	111287	139	111701	111285	139	Viral replication A28-like	DPV127	70/85	LSDV118	61/78	WR151	54/71	M116L	61/80	118L	78/92
119L	112618	111719	300	112616	111717	300	RNA polymerase subunit rpo35	DPV128	64/79	LSDV119	66/77	WR152	57/76	M117L	61/77	119L	85/92
120L	112814	112590	75	112812	112588	75	IMV membrane	DPV129	69/82	LSDV120	58/77	WR153	54/84	M118L	63/75	120L	90/94
120.5L	112978	112847	44	112976	112845	44										120.5L	69/81
121L	113777	113019	253	113775	113017	253	GTPase; DNA packaging	DPV131	83/94	LSDV121	83/92	WR155	59/76	M120L	81/92	121L	89/96
122R	113889	114446	186	113887	114444	186	EEV glycoprotein	DPV132	44/59	LSDV122	41/55	WR156	33/57	M121R	39/57	122R	66/79
123R	114472	114981	170	114470	114979	170	C-type lectin-like domain; EEV glycoprotein	DPV133	64/82	LSDV123	54/75	WR157	48/68	M122R	57/81	123R	79/92
124R	115022	115558	179	115020	115556	179		DPV134	44/66	LSDV124	40/60	WR158	40/57	M123R	41/60	124R	72/84
125R	115597	116451	285	115595	116449	285		DPV135	38/59	LSDV125	37/62			M124R	35/58	125R	81/92
126R	116512	117180	223	116510	117178	223	EEV glycoprotein	DPV136	34/49							126R	35/54
127R	117228	118037	270	117226	118035	270		DPV137	39/61	LSDV127	36/58	WR160	27/51	M126R	35/55	127R	70/85
128L	118847	118038	270	118839	118036	268	CD47-like	DPV139	32/57	LSDV128	31/52	WR162	23/40	M128L	31/54	128L	66/83
129R	118851	119264	138	118843	119256	138	Myristylprotein	DPV138	50/67			WR063	28/45	M129R	36/55	129R	76/88
130L	119813	119256	186	119805	119248	186											
131R	119889	120092	68	119881	120084	68										131R	38/65
132R	120123	120368	82	120115	120360	82		DPV141	46/66	LSDV130	45/68					132R	63/80
133L	121408	120380	343	121400	120372	343	3-Beta hydroxysteroid dehydrogenase	DPV142	53/69			WR170	43/63				
134R	121447	121914	158	121439	121906	156	IL-24-like			LSDV005	26/47						
135R	121993	127701	1903	121985	127693	1903	VARV B22R-like	DPV146	48/65	LSDV134	48/63			M134R	43/59	135R	78/89
136R	127701	128753	351	127693	128745	351	Type-I IFN receptor	DPV147b	28/51	LSDV135	33/53	WR200	26/44	M135R	23/41		
137R	128783	129241	153	128775	129233	153	A52R-family	DPV148	37/60	LSDV136	42/68	WR022	22/45	M136R	28/55	137R	72/86
138R	129270	130283	338	129262	130275	338		DVP149	43/61	LSDV137	42/60	WR177	35/59	M137R	32/55	138R	64/80
139R	130362	130931	190	130354	130923	190	A52R-family	DPV152	50/69			WR039	26/44	M139R	44/66	139R	68/81
140R	130966	132675	570	130958	132667	570	Kelch-like	DPV160	27/47	LSDV151	28/47	WR180	31/56	M140R	43/62		
141R	132705	133064	120	132697	133056	120	CD200-like	DPV153	44/63	LSDV138	52/69			M141R	38/60	141R	72/82
142R	133101	134027	309	133093	134019	309	Serine/threonine protein kinase	DPV154	58/76	LSDV139	60/80	WR183	47/64	M142R	56/74	142R	84/92
143R	134063	134764	234	134055	134756	234	Kila-N/RING finger	DPV155	38/61	LSDV140	38/60	WR208	25/45	M143R	46/64	143R	80/91
144R	134798	135601	268	134790	135593	268	vCCP/EEV host range	DPV156	42/59	LSDV141	37/56	WR025	33/53	M144R	34/52	144R	64/76
145R	135900	136868	323	135892	136860	323	vCCR8	DPV162	36/60	LSDV011	37/62					145R	66/78
146R	137021	138427	469	137013	138431	473	Ankyrin repeat	DPV164	41/63	LSDV147	37/59	WR186	21/41	M149R	34/56	146R	75/86
147R	138465	139937	491	138456	139928	491	Ankyrin repeat	DPV164	29/52	LSDV148	35/54	WR186	23/39	M149R	23/58	147R	78/89

Table 2 (*Continued*)

ORF	TPV-Kenya			TPV-RoC			Predicted structure or function	DPV[a]		LSDV[b]		VACV[c]		MYXV[d]		YMTV[e]	
	Codon		aa[f]	Codon		aa		ORF	Identity/similarity	ORF	Identity/similarity	ORF	Identity/similarity	ORF	Identity/similarity	ORF	Identity/similarity
	Start	Stop		Start	Stop												
148R	139951	141378	476	139942	141369	476	Ankyrin repeat	DPV166	32/51	LSDV145	27/49	WR186	22/45	M005R	23/50	148R	72/85
149R	141392	142393	334	141383	142384	334	SERPIN/crmA ortholog	DPV167	48/68	LSDV149	44/67	WR195	38/56	M151R	45/62	149R	80/93
150R	142434	142733	100	142425	142724	100		DPV168	28/49	LSDV153	31/62			M004.1	33/51	150R	74/88
151R	142825	143823	333	142816	143814	333		DPV007	32/54	LSDV007	34/56	WR209	25/46			151R	71/84

[a] Ortholog in DPV.
[b] Ortholog in LSDV.
[c] Ortholog in VACV.
[d] Ortholog in MYXV.
[e] Ortholog in YMTV.
[f] Number of amino acids.

al., 2005). A BLAST search of the intact 11L protein revealed homologues only in YLDV (11L), YMTV (11L), *Deerpox virus* (DPV; DPV019) and *Vaccinia virus* (VACV; VACV WR186). However, when 11.2L was used as the query sequence, additional homologues in *Myxoma virus* (MYXV; M148R), *Lumpy skin disease virus* (LSDV; LSDV145) and an additional VACV protein encoded by VACV WR019 were detected. The proteins encoded by TPV11L, DPV019, M148R, LSDV145, WR186 and WR019 range from 558 to 675 amino acids and contain 7–14 predicted ankyrin repeats (Fig. 2A). While all proteins except for VACV WR019 contain the entire predicted F-box domain (Mercer et al., 2005), there is significant sequence similarity outside of the domain (Fig. 2B). It may be that the sequence, found between the last predicted ankyrin repeat and the start of the F-box domain, acts as an important functional determinant of the proteins. The fact that 11L is truncated in TPV-Kenya suggests that all 14 ankyrin domains are not required to remain functional. Alternatively, the potential gene products from ORFs 11.1L and 11.2L might interact and form a functional complex.

A previously unidentified ORF was annotated between ORFs 23L and 24L of YMTV and denoted 23.5L (Brunetti et al., 2003). A truncated ortholog was found in YLDV. We find that neither isolate of TPV contains a full-length copy of this predicted ORF, as compared with YMTV. However each isolate encodes for a truncated version of the 23.5L (Fig. 1B). TPV-Kenya encodes a 50 aa ORF that aligns to the carboxy half of YMTV 23.5L and is 98% identical. In contrast, TPV-RoC encodes a 38 aa ORF which is 71% identical to the amino terminus of YMTV 23.5L (Fig. 1B). A transversion at position 17890 changes a tyrosine (TPV-Kenya) to a stop codon (TPV-RoC) causing premature termination of TPV-RoC 23.5L. As well, an insertion at position 17994 changes a string of thymines from T5 (TPV-RoC) to T6 (TPV-Kenya) and disrupts the coding from the downstream start codon on the minus strand.

3.2. Overall nucleotide comparative analysis

Comparison of the two TPV isolates on a nucleotide-by-nucleotide basis indicates 35 changes across a pairwise sequence alignment of 144,565 nucleotide positions. Thirty-one of the changes were within predicted coding regions and could be divided into 13 transitions, 12 transversions and 6 deletions. Six transitions cause only synonymous codon changes. The other seven transitions resulted in non-synonymous substitutions within the coding sequence resulting in a single amino acid difference between the comparable protein sequences between the two TPV isolates. Six of these non-synonymous events resulted in relatively non-conserved changes. In contrast, 11 of the 12 transversions were non-synonymous and 10 of the 11 non-synonymous changes were to non-conserved amino acids. An A to C transversion at position 10241 changes a stop codon (TAG) on the minus strand template of 11.1L of TPV-Kenya to a glutamic acid in TPV-RoC resulting in a full-length 11L ORF, comparable in length to the other poxvirus 11L orthologs. The 6 deletions represent the absence of one of four hexanucleotide direct repeats (CATATA) present at the

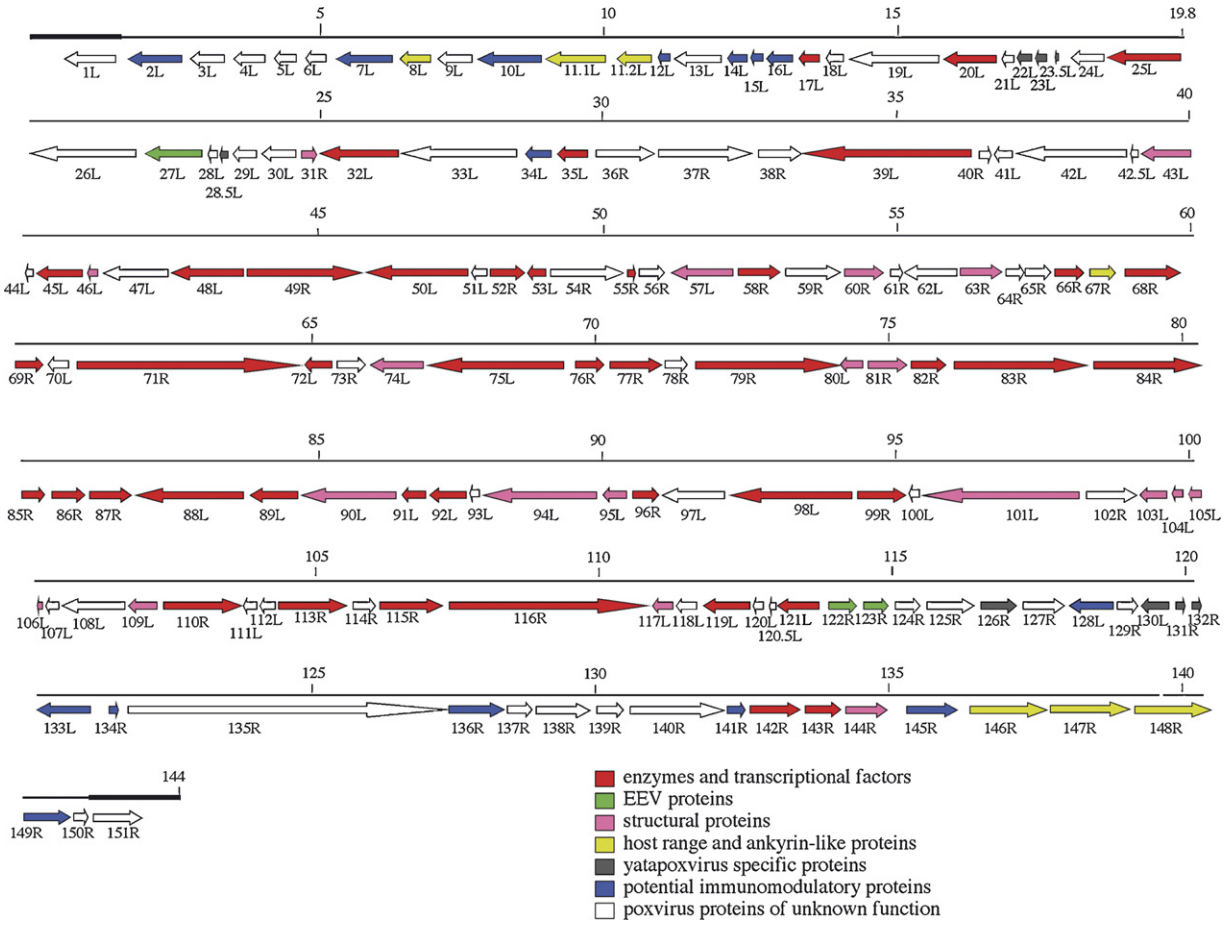

Fig. 3. TPV-Kenya genomic map. ORFs are displayed as arrows that also indicate the direction of transcription. The arrows are coloured to indicate a specific functional category. At either end of the genome is a bolded section that indicates the terminal inverted repeat.

5′ end of ORF 128L in TPV-RoC. The result of this hexanucleotide deletion results in a shortened 128L amino acid sequence in TPV-RoC (MYMYMYNY) compared to TPV-Kenya (MYMYMYMYNY). The other sequence differences that distinguish the TPV isolates include two transitions in non-coding sequences and an insertion in the intergenic region between ORFs 23L and 23.5L of a thymidine (T) repeat; T8 in TPV-RoC compared to T7 in TPV-Kenya.

3.3. Conserved DNA sequence near the termini

While examining the intergenic regions of TPV, a predicted 58-codon ORF was found within the TIR, located between the extreme terminus and 1L/151R. The ORF is transcribed toward the center of the genome and in the opposite direction of all ORFs 20 kbp from either end of the DNA (Fig. 3). The ORF is also present in the YLDV sequence in GenBank but was not described in the publication, possibly due to the ORF mirror-image orientation, which might contribute to dsRNA production (Lee et al., 2001). The comparable region in YMTV was previously described as a pseudogene (Brunetti et al., 2003). To determine the likelihood that the ORF encodes a functional protein, a translated BLAST search (tBLASTx) was used to find homologous amino acid sequences. Several poxviruses, including YLDV, YMTV, LSDV and DPV, have sequences that show significant homology to the TPV query sequence; however, the nucleic acid sequences in YMTV, LSDV and DPV lacked a start codon. Therefore, the cognate sequences appear to represent either a pseudogene or a terminal DNA sequence conserved across genera. The region of nucleotide conservation consists of a 300 nucleotide segment that surrounds the predicted 58-codon ORF. The TIR regions of all chordopoxviruses were compared for similarity to these conserved sequences. The 300

Table 3
TIR conserved sequence positions in various poxvirus species

Virus	Left end		Right end	
	Start	End	Start	End
TPV-Kenya	400	714	144164	143850
TPV-RoC	400	714	144154	143840
YLDV	418	732	144158	143844
YMTV	437	751	134285	133971
GTPV	1018	1329	148582	148271
LSDV	1286	1592	149488	149182
SHPV	1225	1530	148833	148528
SWPV	1062	1381	145393	145074
DPV	4756	5070	165805	165491

nucleotide conserved sequence was found in *Swinepox virus* (SWPV) and overlapped with a hypothetical gene (designated 002) in SWPV, DPV and *Capripoxviruses* (Table 3). While this sequence exhibits homology across genera, the highest identities were found between members within a particular genus (Fig. 4).

A highly conserved sequence, present in the TIR region of orthopoxviruses, has been described previously (Shchelkunov et al., 1998). However, the sequence appears to be distinct from the 300-bp *Yatapoxvirus* sequences and their cognates (Fig. 4). Since new sequence information has become available from the time that this DNA region was last compared and reported (Baroudy et al., 1982; Shchelkunov et al., 1998), selected orthopoxvirus sequences from CPXV, VACV, VARV, HSPV, TATV, ETCV, MPXV, CMLV and RCNV (acronyms described in Table 1; data not shown) were aligned and compared. Interestingly, CMLV lacks this entire sequence and RCNV encodes for only 137 nucleotides of the ~300-bp conserved sequences. As previously described, these orthopoxvirus nucleotide sequences are highly homologous, sharing 87–100% sequence identity ((Shchelkunov et al., 1998) and data not shown).

	TPV	YLDV	YMTV	DPV	SWPV	LSDV	GTPV	SHPV
TPV								
YLDV	98%							
YMTV	84%	85%						
DPV	54%	55%	54%					
SWPV	45%	45%	47%	62%				
LSDV	36%	36%	35%	39%	40%			
GTPV	36%	36%	37%	39%	40%	95%		
SHPV	36%	36%	34%	39%	39%	94%	93%	

Fig. 4. Identity matrix of a conserved sequence found in the TIR of various poxvirus species. An approximately 300 bp region within the TIR of the poxviruses listed was aligned using ClustalW and percent identity was determined. Poxviruses are listed in order of relatedness for this particular sequence.

3.4. Identification of two conserved poxvirus gene families

Two TPV intergenic regions contained potential ORFs below the commonly used codon limit of 50. The ORFs are located between 27L and 28.5L, and 42L and 43L; they were designated 28L and 42.5L, respectively (Table 2). To determine if these ORFs likely encode functional proteins, tBLASTx was used to find homologous sequences. However, due to the small size of these ORFs, BLAST searches were unable to detect any homologous sequences and thus the search for homologues was performed manually.

The ORF 28L is present in both TPV and YLDV; the predicted ORF encodes for a potential protein of 48 aa. The region between orthologues of 27L and 28.5L of the genomes of species of chordopoxviruses currently available were examined. Orthologues were identified in all orthopoxviruses, parapoxviruses, SWPV, and the unclassified poxviruses DPV and *Crocodilepox virus* (CRV) (Fig. 5a). However, the closely related capripoxviruses lacked a homologous gene in this region. Members of the genera *Avipoxvirus*, *Leporipoxvirus* and *Molluscum contagiosum virus* also lacked the sequence (Table 4).

The ORF 42.5L is predicted to encode a 30 amino acid protein. A search of all poxvirus genomes for orthologues showed that ORF 42.5L is highly conserved among the *Chordopoxvirinae* and orthologues in all vertebrate poxviruses currently

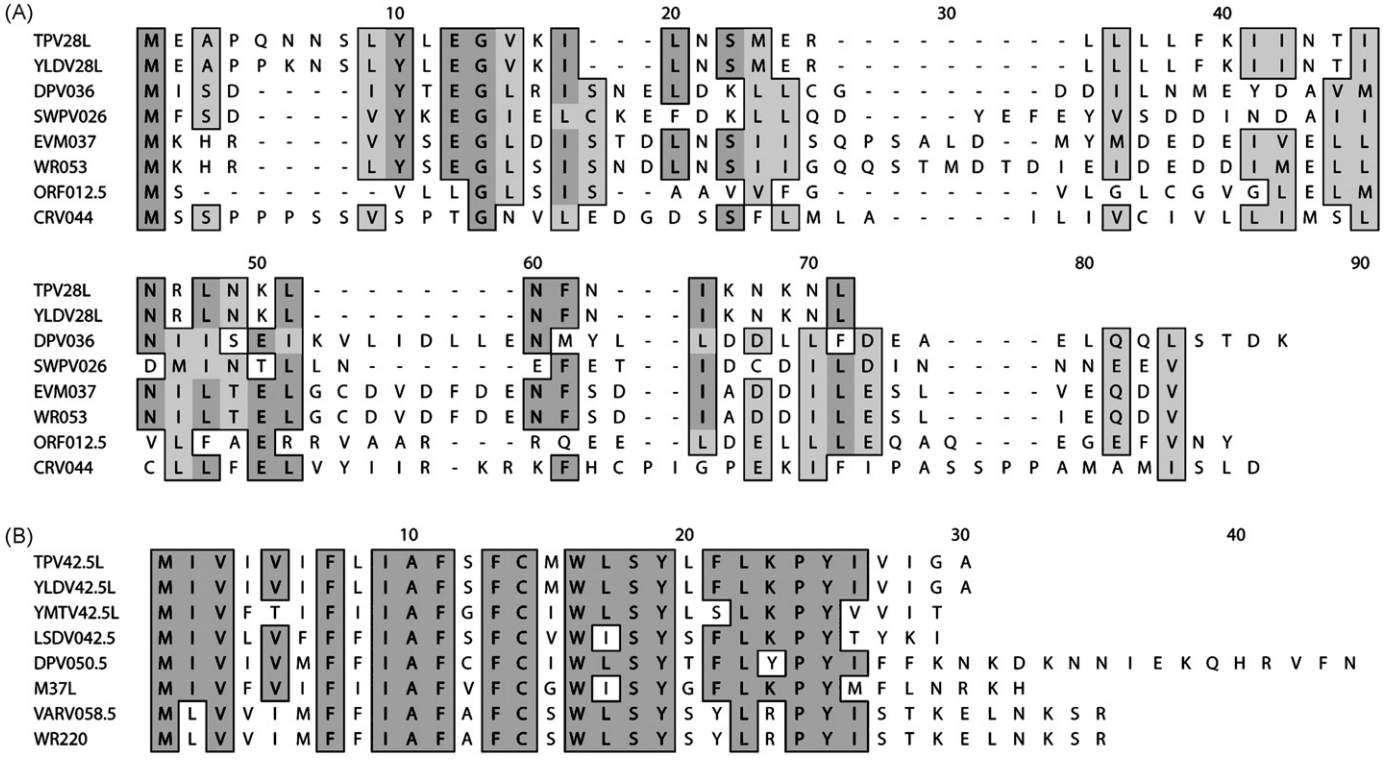

Fig. 5. Alignments of the predicted 28L and 42.5L gene families. (a) Orthologs of TPV 28L are aligned and include YLDV, DPV, SWPV, ECTV (EVM037), VACV (WR053), ORFV, CRV. (b) Orthologs of TPV 42.5L are aligned and include YLDV, YMTV, LSDV, DPV, MYXV (M37L), VARV, VACV (WR220), as annotated at www.poxvirus.org.

Table 4
Members of two new poxvirus gene families

Genus	Virus	28L gene family			42.5L gene family		
		Designation	Start	Stop	Designation	Start	Stop
Yatapox	TPV-RoC	28L	23330	23187	42.5L	39411	39322
	TPV-Kenya	28L	23332	23189	42.5L	39413	39324
	YLDV	28L	23342	23199	42.5L	39425	39336
	YMTV	NP[a]			42.5L	33938	33852
Capripox	GTPV	NP			038.5	37590	37504
	SHPV	NP			038.5	37805	37719
	LSDV	NP			042.5	38189	38103
Suipox	SWPV	026	19915	19724	038.5	34712	34620
Leporipox	MYXV	NP			37L	39385	39290
	SHFV	NP			37L	39399	39316
Molluscipox	MOCV	NP			043.5L	64055	63933
Orthopox	ECTV	037	50964	50749	053.5	68659	68555
	MPXV	C20L	42461	42240	I0.5L	59903	59799
	CPXV	062	58869	58648	079.5	76579	76475
	VARV	041	33516	33295	058.5	51129	51025
	HSPV	054	53482	53261	070.5	71114	71010
	TATV	054	42022	41807	072.5	59706	59602
	CMPV	49L	43284	43063	66.5L	60901	60797
	VACV	053	42188	41967	220[b]	59851	59744
Avipox	FWPV	NP			090.5	90681	90782
	CNPV	NP			117.5	120153	120254
Parapox	ORFV	012.5[b]	11760	11578	029.5	32396	32244
	BPSV	011.5[b]	12820	12650	028.5	33354	33214
Unclassified	CRV	044	58267	58025	064.5	86842	86762
	DPV	036	32435	32226	050.5	49191	49060

[a] Not present.
[b] Annotated at www.poxvirus.org/.

sequenced were found (Fig. 5b). The nucleotide sequence has previously only been reported as a putative ORF of 32 codons for sequenced leporipoxviruses (Cameron et al., 1999; Willer et al., 1999). Additionally, an orthologue, VACV ORF WR220, has been annotated in sequences at http://www.poxvirus.org/ (Table 4).

4. Discussion

The virtually complete genome sequences were determined for two isolates of TPV recovered from clinical cases that occurred about 50 years apart. Annotation of the determined sequences revealed a single ORF difference between the two genomes. The ORF 11L is truncated in TPV-Kenya compared to TPV-RoC, which suggests that even the small genetic variability present between the two genomes has possibly resulted in changes to the proteome. A TPV nucleotide sequence is conserved in the TIR region of several poxviruses closely related to the yatapoxviruses. An analogous but distinct DNA sequence located in the correlate region of the genome is also present in the orthopoxviruses. Finally, two novel gene families are proposed following identification using comparative genomics.

One of the difficulties that arise when limited sequences are available for comparison is that, ORFs that do not meet the standard search parameters can be difficult to assign. Many poxvirus ORFs are quite small and it is unlikely that they will achieve a significant match using BLAST. Selected available sequences from chordopoxviruses were used to determine significantly conserved sequences in TPV. One approach was to compare a tentatively designated ORF and examine areas of several poxviruses that contained highly conserved ORFs flanking this region.

Using this method, two previously unidentified gene families, 28L and 42.5L, which are clearly present in members of several other poxvirus genera, were identified. The region between ORFs 27L and 28.5L was previously assigned to a large but overlapping ORF (28R) that was identified in YLDV (Lee et al., 2001); therefore ORF 28L was not originally identified as a putative gene. However, the evidence that orthologs of 28L are encoded by a variety of poxviruses suggests that 28L encodes a protein product (Table 4). Conversely, 42.5L had previously only been identified in the leporipoxviruses but this ortholog is highly conserved among the *Chordopoxvirinae*. Due to its extremely small size (30–44 codons), it is unlikely to have been considered an ORF previously. On close inspection, however, 42.5L has a conserved early and late promoter (data not shown; www.poxvirus.org/) and the putative amino acid sequence shares 62–77% identity with orthologs among other chordopoxviruses.

These properties should be sufficient to designate 42.5L as a putative ORF. Other related predictive methods, such as analyzing purine skew of the ORFs (Da Silva and Upton, 2005), were not used since both of these ORFs had clear orthologues in other poxviruses.

In addition to identifying new putative ORFs, a conserved DNA sequence in the TIR of several genera of poxviruses was described. An analogous sequence that exhibited a similar organization pattern, but did not show significant homology, has been identified previously in the orthopoxviruses. A possible role in DNA replication for the orthopoxvirus conserved sequence in the TIR region has been proposed (Shchelkunov et al., 1998), but since there is considerable divergence of the sequence across genera, a precise mechanism is unclear and may be structural rather than sequence-specific. Structural elements such as Holliday Junctions and cruciform structures have been shown to be important for resolution of concatenated DNA into unit length DNA molecules (Palaniyar et al., 1999). This sequence may fulfill its function in a similar way, relying on a structural motif.

If the conserved sequence does, in fact, play a role in DNA replication, then a sequence that performs a similar function must be present in all poxviruses. Therefore, an attempt to define this sequence in other poxvirus species was made. It is possible to find sequences in other poxvirus TIRs that share some homology to the conserved sequence; however, without more sequence information from viral members within genera it is difficult to clearly define since there is a lack of sequence information for viral members within other poxvirus genera, including two genera composed of only a single member each (*Suipoxvirus* and *Molluscipoxvirus*).

Through a comparative genomics approach, we have identified important additional features of yatapoxviruses noted by prior sequencing of YMTV and YLDV. The results presented indicate a relatively slow evolutionary rate, which suggests a relatively stable, confined evolutionary niche. From this standpoint, the primate host-range of TPV and YLDV in the central region of the rainforest of Africa appears to have remained the same, at least for 50 years, despite extensive ecological changes, particularly urbanization of forested areas. There have been suggestions that an insect vector might be involved in *Yatapoxvirus* transmission because TPV and YLDV infection are localized to one or two lesions and not systemic like smallpox (Damon, 2007). Maintaining a lifecycle that includes a potential non-human primate reservoir, an insect reservoir, as well as a human reservoir suggests that a constant genetic selective pressure might be maintained on the TPV and YLDV genome, which would lower the likelihood of sequence divergence. However, this may not explain the lack of nucleotide changes in the third bp position and it is unlikely that the DNA polymerase encoded by *Yatapoxviruses* has a high enough fidelity to explain this phenomenon. The codon bias present in many *Yatapoxvirus* genes represent the most rarely used codons in mammalian cells (Barrett et al., 2006). An alternative explanation to explain the third position conservation is that this codon bias is required for efficient gene expression in a variety of distinct host species. In contrast, a poxvirus that is able to infect several different hosts, e.g. *Cowpox virus*, which appears to be parental to the orthopoxviruses, has a sequence that is more amenable to changing with different hosts.

Acknowledgements

This work was supported by the Canadian Institutes of Health Research (CIHR) and National Cancer Institute of Canada (NCIC). SN was supported by an Ontario Graduate Scholarship and Western Graduate Research Scholarship. GM held a Canada Research Chair in Molecular Virology and is an International Scholar of the Howard Hughes Medical Institute.

References

Afonso, C.L., Delhon, G., Tulman, E.R., Lu, Z., Zsak, A., Becerra, V.M., Zsak, L., Kutish, G.F., Rock, D.L., 2005. Genome of deerpox virus. J. Virol. 79 (2), 966–977.

Afonso, C.L., Tulman, E.R., Delhon, G., Lu, Z., Viljoen, G.J., Wallace, D.B., Kutish, G.F., Rock, D.L., 2006. Genome of crocodilepox virus. J. Virol. 80 (10), 4978–4991.

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 (8), 3815–3831.

Afonso, C.L., Tulman, E.R., Lu, Z., Zsak, L., Osorio, F.A., Balinsky, C., Kutish, G.F., Rock, D.L., 2002. The genome of swinepox virus. J. Virol. 76 (2), 783–790.

Balbas, P., Gosset, G., 2001. Chromosomal editing in *Escherichia coli*. Vectors for DNA integration and excision. Mol. Biotechnol. 19 (1), 1–12.

Baroudy, B.M., Venkatesan, S., Moss, B., 1982. Incompletely base-paired flip-flop terminal loops link the two DNA strands of the vaccinia virus genome into one uninterrupted polynucleotide chain. Cell 28 (2), 315–324.

Barrett, J.W., Sun, Y., Nazarian, S.H., Belsito, T.A., Brunetti, C.R., McFadden, G., 2006. Optimization of codon usage of poxvirus genes allows for improved transient expression in mammalian cells. Virus Genes 33 (1), 15–26.

Brunetti, C.R., Amano, H., Ueda, Y., Qin, J., Miyamura, T., Suzuki, T., Li, X., Barrett, J.W., McFadden, G., 2003. Complete genomic sequence and comparative analysis of the tumorigenic poxvirus Yaba monkey tumor virus. J. Virol. 77 (24), 13335–13347.

Buller, R.M., Arif, B.M., Black, D.N., Dumbell, K.R., Esposito, J.J., Lefkowitz, E.J., McFadden, G., Moss, B., Mercer, A.A., Moyer, R.W., Skinner, M.A., Tripathy, D.N., 2005. Poxviridae. In: Fauquet, C., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A. (Eds.), Virus Taxonomy: Classification and Nomenclature of Viruses; Eighth Report of the International Committee on Taxonomy of Viruses. Elsevier/Academic Press, Oxford, pp. 117–133.

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 (2), 298–318.

Da Silva, M., Upton, C., 2005. Using purine skews to predict genes in AT-rich poxviruses. BMC Genomics 6 (1), 22.

Damon, I.K., 2007. Poxviruses. In: Knipe, D.M., Howley, P.M. (Eds.), Fields Virology, Vol.2, 5th ed. Lippincott, Williams & Wilkins, New York, pp. 2947–2976.

Delhon, G., Tulman, E.R., Afonso, C.L., Lu, Z., de la Concha-Bermejillo, A., Lehmkuhl, H.D., Piccone, M.E., Kutish, G.F., Rock, D.L., 2004. Genomes of the parapoxviruses ORF virus and bovine papular stomatitis virus. J. Virol. 78 (1), 168–177.

Dhar, A.D., Werchniak, A.E., Li, Y., Brennick, J.B., Goldsmith, C.S., Kline, R., Damon, I., Klaus, S.N., 2004. Tanapox infection in a college student. N. Engl. J. Med. 350 (4), 361–366.

Domi, A., Moss, B., 2002. Cloning the vaccinia virus genome as a bacterial artificial chromosome in Escherichia coli and recovery of infectious virus in mammalian cells. Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12415–12420.

Downie, A.W., Espana, C., 1972. Comparison of Tanapox virus and Yaba-like viruses causing epidemic disease in monkeys. J. Hyg. (Lond.) 70 (1), 23–32.

Espana, C., Brayton, M.A., Ruebner, B.H., 1971. Electron microscopy of the Tana poxvirus. Exp. Mol. Pathol. 15 (1), 34–42.

Esposito, J.J., Sammons, S.A., Frace, A.M., Osborne, J.D., Olsen-Rasmussen, M., Zhang, M., Govil, D., Damon, I.K., Kline, R., Laker, M., Li, Y., Smith, G.L., Meyer, H., Leduc, J.W., Wohlhueter, R.M., 2006. Genome sequence diversity and clues to the evolution of variola (smallpox) virus. Science 313 (5788), 807–812.

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 (Pt 4), 855–872.

Knight, J.C., Novembre, F.J., Brown, D.R., Goldsmith, C.S., Esposito, J.J., 1989. Studies on Tanapox virus. Virology 172 (1), 116–124.

Lee, H.J., Essani, K., Smith, G.L., 2001. The genome sequence of Yaba-like disease virus, a yatapoxvirus. Virology 281 (2), 170–192.

McNulty Jr., W.P., Lobitz Jr., W.C., Hu, F., Maruffo, C.A., Hall, A.S., 1968. A pox disease in monkeys transmitted to man. Clinical and histological features. Arch. Dermatol. 97 (3), 286–293.

Mercer, A.A., Fleming, S.B., Ueda, N., 2005. F-box-like domains are present in most poxvirus ankyrin repeat proteins. Virus Genes 31 (2), 127–133.

Mercer, A.A., Ueda, N., Friederichs, S.M., Hofmann, K., Fraser, K.M., Bateman, T., Fleming, S.B., 2006. Comparative analysis of genome sequences of three isolates of Orf virus reveals unexpected sequence variation. Virus Res. 116 (1–2), 146–158.

Moss, B., 2007. Poxviridae: the viruses and their replication. In: Knipe, D.M., Howley, P.M. (Eds.), Fields Virology, vol. 2, fifth ed. Lippincott, Williams & Wilkins, New York, pp. 2905–2946.

Palaniyar, N., Gerasimopoulos, E., Evans, D.H., 1999. Shope fibroma virus DNA topoisomerase catalyses holliday junction resolution and hairpin formation in vitro. J. Mol. Biol. 287 (1), 9–20.

Seet, B.T., Johnston, J.B., Brunetti, C.R., Barrett, J.W., Everett, H., Cameron, C., Sypula, J., Nazarian, S.H., Lucas, A., McFadden, G., 2003. Poxviruses and immune evasion. Annu. Rev. Immunol. 21, 377–423.

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., 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 (2), 432–460.

Shchelkunov, S.N., Totmenin, A.V., Babkin, I.V., Safronov, P.F., Ryazankina, O.I., Petrov, N.A., Gutorov, V.V., Uvarova, E.A., Mikheev, M.V., Sisler, J.R., Esposito, J.J., Jahrling, P.B., Moss, B., Sandakhchiev, L.S., 2001. Human monkeypox and smallpox viruses: genomic comparison. FEBS Lett. 509 (1), 66–70.

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., Zsak, L., Kutish, G.F., Rock, D.L., 2004. The genome of Canarypox virus. J. Virol. 78 (1), 353–366.

Tulman, E.R., Afonso, C.L., Lu, Z., Zsak, L., Sur, J.H., Sandybaev, N.T., Kerembekova, U.Z., Zaitsev, V.L., Kutish, G.F., Rock, D.L., 2002. The genomes of sheeppox and Goatpox viruses. J. Virol. 76 (12), 6054–6061.

Tulman, E.R., Delhon, G., Afonso, C.L., Lu, Z., Zsak, L., Sandybaev, N.T., Kerembekova, U.Z., Zaitsev, V.L., Kutish, G.F., Rock, D.L., 2006. Genome of Horsepox virus. J. Virol. 80 (18), 9244–9258.

Willer, D., McFadden, G., Evans, D.H., 1999. The complete genome sequence of Shope (rabbit) fibroma virus. Virology 264 (2), 319–343.