JVI Journals.ASM.org

New Insights into the Evolution of *Entomopoxvirinae* from the Complete Genome Sequences of Four Entomopoxviruses Infecting *Adoxophyes honmai*, *Choristoneura biennis*, *Choristoneura rosaceana*, and *Mythimna separata*

Julien Thézé,[a] Jun Takatsuka,[b] Zhen Li,[c] Julie Gallais,[a] Daniel Doucet,[c] Basil Arif,[c] Madoka Nakai,[d] Elisabeth A. Herniou[a]

Institut de Recherche sur la Biologie de l'Insecte, Unité Mixte de Recherche (UMR) 7261, Centre National de la Recherche Scientifique (CNRS), Université François-Rabelais, UFR Sciences et Techniques, Tours, France[a]; Forestry and Forest Products Research Institute, Tsukuba, Japan[b]; Great Lakes Forestry Centre, Sault Sainte Marie, Ontario, Canada[c]; Department of Applied Biological Science, Institute of Agriculture, Tokyo University of Agriculture and Technology, Fuchu, Tokyo, Japan[d]

Poxviruses are nucleocytoplasmic large DNA viruses encompassing two subfamilies, the *Chordopoxvirinae* and the *Entomopoxvirinae*, infecting vertebrates and insects, respectively. While chordopoxvirus genomics have been widely studied, only two entomopoxvirus (EPV) genomes have been entirely sequenced. We report the genome sequences of four EPVs of the *Betaentomopoxvirus* genus infecting the Lepidoptera: *Adoxophyes honmai* EPV (AHEV), *Choristoneura biennis* EPV (CBEV), *Choristoneura rosaceana* EPV (CREV), and *Mythimna separata* EPV (MySEV). The genomes are 80% AT rich, are 228 to 307 kbp long, and contain 247 to 334 open reading frames (ORFs). Most genes are homologous to those of *Amsacta moorei* entomopoxvirus and encode several protein families repeated in tandem in terminal regions. Some genomes also encode proteins of unknown functions with similarity to those of other insect viruses. Comparative genomic analyses highlight a high colinearity among the lepidopteran EPV genomes and little gene order conservation with other poxvirus genomes. As with previously sequenced EPVs, the genomes include a relatively conserved central region flanked by inverted terminal repeats. Protein clustering identified 104 core EPV genes. Among betaentomopoxviruses, 148 core genes were found in relatively high synteny, pointing to low genomic diversity. Whole-genome and spheroidin gene phylogenetic analyses showed that the lepidopteran EPVs group closely in a monophyletic lineage, corroborating their affiliation with the *Betaentomopoxvirus* genus as well as a clear division of the EPVs according to the orders of insect hosts (Lepidoptera, Coleoptera, and Orthoptera). This suggests an ancient coevolution of EPVs with their insect hosts and the need to revise the current EPV taxonomy to separate orthopteran EPVs from the lepidopteran-specific betaentomopoxviruses so as to form a new genus.

Poxviruses are large double-stranded DNA (dsDNA) viruses infecting a wide range of animals. They belong to the phylogenetically related group of viruses termed nucleocytoplasmic large DNA viruses (NCLDV) (1). They harbor linear dsDNA genomes with inverted terminal repeats (ITRs) (2). Poxvirus genomes are 130 to 375 kbp long and replicate in the cytoplasm (3). The family *Poxviridae* includes two subfamilies: the *Chordopoxvirinae*, infecting vertebrates, and the *Entomopoxvirinae*, infecting insects. The chordopoxviruses are classified into nine genera, including *Orthopoxvirus* and *Avipoxvirus* (4), and have been the subjects of the main body of research on poxviruses (5, 6). The entomopoxviruses (EPVs) are currently divided into three genera based on host range and virion morphology: *Alphaentomopoxvirus*, infecting coleopterans; *Betaentomopoxvirus*, infecting lepidopterans and orthopterans; and *Gammaentomopoxvirus*, infecting dipterans (4). However, the lack of genomic data has precluded the integration of unifying genetic criteria into this classification. That is why the orthopteran EPV *Melanoplus sanguinipes entomopoxvirus* was removed from the *Betaentomopoxvirus* genus (4) and why *Diachasmimorpha entomopoxvirus*, infecting both a braconid parasitic wasp and its tephritid fruit fly dipteran host, remains unclassified (7, 8). Reports of entomopoxviruses from bumblebees (9) and cockroaches (10) further show that the taxonomic biodiversity of EPV remains largely undescribed.

EPV virions are embedded within a matrix protein, termed a spheroid, forming typical oval-shaped occlusion bodies (OBs) composed mainly of the spheroidin protein (11). Spheroidin is a functional homolog of the baculovirus polyhedrin (12) in that it affords the virions some protection against inactivating environmental agents such as heat, desiccation, and UV light (12, 13). The OBs dissolve in the alkaline-reducing environment of the insect midgut with the aid of an endogenous alkaline protease and release the virions to initiate infection in columnar epithelial cells prior to systemic infection (14). Virus replication occurs principally in the fat tissue, but other tissues are also affected (15). Interestingly, while baculoviruses spread within larval tissues through the tracheal system (16), these tissues are rarely infected by EPVs. Apparently, EPVs use hemocytes to spread within susceptible tissues (11). The course of EPV infection is generally slow (13); insects can survive as long as several weeks after the initial infection and can even remain in the larval stage longer than an uninfected host (17). OBs are disseminated in the environment through regurgitation, defecation, and, ultimately, the disintegra-

Received 14 February 2013 **Accepted** 5 May 2013
Published ahead of print 15 May 2013
Address correspondence to Julien Thézé, theze.julien@gmail.com.

doi:10.1128/JVI.00453-13

tion of dead hosts (11, 18). There is also one report on transmission via parasitoids (7).

EPVs have been studied mainly because of their potential as microbial biocontrol agents. Field studies on important Asian and North American lepidopteran pests revealed that EPVs could be found in diseased larvae of the smaller tea tortrix, *Adoxophyes honmai* (Lepidoptera: Tortricidae) (19, 20), the 2-year-cycle budworm moth, *Choristoneura biennis* (Lepidoptera: Tortricidae) (21), the oblique-banded leafroller moth *Choristoneura rosaceana* (Lepidoptera: Tortricidae), and the oriental armyworm, *Mythimna separata* (Lepidoptera: Noctuidae) (22). In contrast to baculoviruses, which can kill insect hosts shortly after infection and could be used in place of a chemical insecticide (23), EPVs are slow-acting pathogens and may be more appropriate for reducing the growth rate of the pest population via epizootics that affect the frequency of insect outbreaks. It has been suggested that combining fast- and slow-killing strategies could contribute to better insect pest control and diminish the need for chemical insecticides (24).

To date, only two EPV genomes have been completely sequenced: those of *Melanoplus sanguinipes entomopoxvirus* (MSEV) (25), infecting the North American migratory grasshopper (Orthoptera: Acrididae), and *Amsacta moorei entomopoxvirus* (AMEV) (26), infecting the red hairy caterpillar (Lepidoptera: Arctiidae). AMEV and MSEV have similar genome sizes of 232 kb and 236 kb, with 294 and 267 open reading frames (ORFs), respectively. However, they share little overall genome homology in terms of gene content or order. With only 106 genes in common, AMEV and MSEV share less than half of their gene content, which is the reason for the removal of MSEV from the *Betaentomopoxvirus* genus. However, the other orthopteran EPVs remain in the *Betaentomopoxvirus* genus. A single genome (AMEV) is, indeed, not sufficient to allow the proposal of unifying genomic characters for a genus. More lepidopteran EPV sequences could allow us to discriminate between several taxonomic hypotheses, as follows. (i) The genus *Betaentomopoxvirus* contains orthopteran and lepidopteran EPVs, and MSEV is a peculiar, divergent virus. In this case, we should not be able to find unifying characteristics for the regrouping of orthopteran viruses. (ii) Lepidopteran and orthopteran EPVs are phylogenetically interrelated. In this case, comparative genomics should show high genome structure divergence, which could encompass the diversity already observed between AMEV and MSEV. (iii) The genus *Betaentomopoxvirus* contains only lepidopteran EPVs, and orthopteran EPVs belong to a different genus. In this case, we expect to find unifying genomic and phylogenetic criteria excluding orthopteran EPVs from the genus *Betaentomopoxvirus*.

The current paucity of EPV genomic data hinders both functional and evolutionary studies. Here we present the complete genome sequences of four EPVs isolated from Lepidoptera with the aim of defining common features for betaentomopoxviruses (BetaEPVs). We sequenced EPVs isolated from *Adoxophyes honmai* (AHEV), *Choristoneura biennis* (CBEV), *Choristoneura rosaceana* (CREV), and *Mythimna separata* (MySEV). We performed genome colinearity and gene content analyses both within the BetaEPV lineage and with more distantly related poxviruses. We combined comparative genomic analyses with phylogenetic analyses in order to understand the evolution of the subfamily *Entomopoxvirinae* at the genomic level.

MATERIALS AND METHODS

DNA isolation and sequencing. AHEV, CBEV, CREV, and MySEV were isolated from diseased larvae of *Adoxophyes honmai* (collected from a tea field in Tokyo, Japan) (19, 20), *Choristoneura biennis* (from the province of Ontario, Canada) (21), *Choristoneura rosaceana* (collected in Eastern Canada), and *Mythimna separata* (obtained from Fulin Sun, Chinese Center for Virus Culture Collection, Wuhan, China) (22), respectively. Viruses were propagated in their respective hosts except for CBEV, which was propagated in *Choristoneura fumiferana*.

OBs were purified by homogenization and density gradient centrifugation using a 0.25 M sucrose-Percoll (GE Healthcare) solution (19). The purified OB suspensions were dissolved with an alkali buffer containing a reducing agent (1 M sodium carbonate and 0.4 M sodium thioglycolate). Undissolved OBs and heavy debris were pelleted by centrifugation at 900 × *g* for 3 min, and the supernatants were centrifuged at 20,400 × *g* for 10 min. Viral genomic DNA was extracted using a Puregene tissue purification kit (Qiagen). The 454 high-throughput sequencing technology was used to sequence AHEV, CBEV, and CREV in 454 single reads and MySEV in 454 paired-end reads.

Genome assembly and annotation. The genomes were assembled *de novo* using Newbler, version 2.6 (27). Overlapping contigs were assembled using Geneious, version 5.5. To fill the gaps between contigs, resolve ambiguities, and position inverted terminal repeat (ITR) regions, PCR primers were designed at contig extremities, and amplicons were subjected to Sanger sequencing (28).

The annotations were performed in three steps. First, Glimmer3 (29) was used for *de novo* prediction of ORFs encoding more than 50 amino acids (aa) with a methionine as the start codon. Second, the protein sequences encoded by each ORF were aligned to the Viral Orthologous Clusters (VOCs) of the Viral Bioinformatics Resource Center (30, 31) and to NCBI's nonredundant protein database by using BLASTp (32) to identify functional homologies. Third, both the delimitation of ITR regions and the correction of 454 homopolymer ambiguities in coding regions were carried out manually.

Comparative genomic analyses. Reciprocal best-hit alignments using BLASTp (32) were performed to identify orthologous proteins between the AMEV genome and the four new genomes, those of AHEV, CBEV, CREV, and MySEV. Similarly, orthologous proteins were identified between MSEV and the five BetaEPV genomes and between the vaccinia virus Western Reserve (VACV) genome and the six EPV genomes. Orthologous gene positions were retrieved on each genome and were integrated into the Circos visualization program (33). The AMEV, MSEV, and VACV genomes were set as references for the visualization of genome colinearity maps among BetaEPVs, EPVs, and poxviruses.

A clustering based on "profile hidden Markov model" alignments using the jackhmmer program of the HMMER 3 package (34) was performed on all EPV proteins to identify potentially inherited conserved genes within the EPV and BetaEPV lineages. Among these genes, we determined gene order conservation within a lineage by using the GeneSyn program (35).

Phylogenetic analyses. A phylogenomic approach was used to position AHEV, CBEV, CREV, and MySEV within the whole-genome poxvirus phylogeny. To date, poxviruses appear to possess 49 core genes (31) that have been identified in the AHEV, CBEV, CREV, and MySEV genomes and in the genomes of representative species of all poxvirus genera. Multiple amino acid alignments were performed on the 49 poxvirus core genes, including those of AHEV, CBEV, CREV, MySEV, and 12 additional poxvirus species, by using the Clustal Omega program (36). In order to ascertain that the poxvirus core genes used for phylogenetic analyses shared the same evolutionary history and could be used as a proxy for the evolution of the virus species, we performed phylogenetic congruence tests to detect any possible conflict in phylogenetic signals between poxvirus core genes. These tests did not show any conflicting phylogenetic signal between genes (data not shown), and therefore, all the multiple amino acid alignments were concatenated prior to phylogenetic recon-

TABLE 1 General features of entomopoxvirus genomes

Genome	Size (bp)	No. of ORFs	No. of singletons	ITR size (bp)	GC content (%)	Coding capacity (%)
Melanoplus sanguinipes entomopoxvirus	236,120	267	144	7,201	18.3	91.6
Amsacta moorei entomopoxvirus "L"	232,392	294	73	9,458	17.8	95.4
Adoxophyes honmai entomopoxvirus "L"	228,750	247	27	5,617	21	89.8
Choristoneura biennis entomopoxvirus "L"	307,691	334	19	23,817	19.7	91
Choristoneura rosaceana entomopoxvirus "L"	282,895	296	11	13,406	19.5	90.2
Mythimna separata entomopoxvirus "L"	281,182	306	64	7,347	19.7	90.5

struction. A maximum likelihood (ML) phylogenetic inference was performed on the concatenated multiple amino acid alignments with the substitution model and model parameters WAG+G, selected using ModelGenerator (37) under the Akaike information criterion. ML analysis was performed with the RAxML program (38), and support for nodes in ML trees was obtained from 100 bootstrap iterations.

A multiple amino acid alignment of the spheroidin gene was performed, including amino acid sequences from the AHEV, CBEV, CREV, and MySEV genomes and all the sequences available from the GenBank public database (25, 26, 39–45). An ML phylogenetic inference was performed on the multiple amino acid alignment for spheroidin with the RAxML program (38) by using the substitution model and model parameters WAG+G. The root of the tree was determined by midpoint rooting.

Nucleotide sequence accession numbers. The AHEV, CBEV, CREV, and MySEV genomes have been deposited in EMBL under accession numbers HF679131, HF679132, HF679133, and HF679134, respectively.

RESULTS

Features of the AHEV, CBEV, CREV, and MySEV genomes. AHEV, CBEV, CREV, and MySEV OB particles were isolated from diseased larvae of *Adoxophyes honmai*, *Choristoneura biennis*, *Choristoneura rosaceana*, and *Mythimna separata*, respectively. The four EPV genomes were assembled in contiguous sequences ranging from 229 kb for the smallest, AHEV, to 308 kb for the largest, CBEV (Table 1). This size range is somewhat similar to that of the previously sequenced EPV genomes, AMEV and MSEV (25, 26). As expected for poxviruses (4), the genomes include a central region flanked by inverted terminal repeat (ITR) regions at the extremities. Due to the repetitive nature of the ITRs, their sequences retain a number of ambiguities. As with other EPV genomes, the nucleotide composition of the four genomes is AT rich, at approximately 80% of the total nucleotide content (Table 1).

Genome contents. The genome annotations predicted 247 and 334 ORFs encoding proteins of more than 50 aa for AHEV and CBEV, respectively (Table 1), with few overlaps between ORFs. This corresponds to about 90% of the genome coding capacity (Table 1). Homology searches in public databases were performed to assign a functional annotation to each ORF. Homologs could be found for approximately 80% of the ORFs and corresponded mostly to genes already found in AMEV. Overall, these conserved proteins are encoded in the central regions of genomes and are essential to virus structure and replication.

Several large gene families of unknown functions, with many members repeated in tandem, were found in EPV genomes. The N1R/p28 gene family is by far the largest, with more than 20 copies per genome and as many as 48 in CBEV. This gene family, based on the VOCs (30, 31) database, regroups the ALI, MTG, and 17K/KilA-N domain proteins previously described separately (11, 26). The tryptophan repeat and leucine-rich gene families are more modest than the N1R/p28 gene family, with copy numbers ranging from 2 to 10. Differences in genome size could be explained in part by differences in N1R/p28 gene copy numbers. Indeed, 21 gene copies represent 8% of the AHEV ORFs, while 48 copies represent 14% of the coding capacity of CBEV. The number of ORFs encoding hypothetical proteins, for which no homologs are found in the databases, was also higher in larger genomes. It is worth mentioning that a number of unknown ORFs showed similarities to proteins found in other large DNA viruses of insects, most notably to those encoded by the baculovirus antiapoptotic *iap* and p35 gene families. The majority of these less conserved, repeated, hypothetical, and singleton ORFs are present mostly in the terminal regions of the genomes and, remarkably, in isolated regions located right in the middle of the genome.

Genome colinearity. In order to compare the global genome synteny conservation among poxviruses, reciprocal best-hit alignments were performed to determine gene orthology among AMEV, MSEV, VAVC, and the four new EPVs. Genes normally have only one ortholog per genome. However, since ITRs are identical, genes located in the ITRs have two orthologs in the other genomes. We mapped the orthologous gene positions in circular colinearity maps (Fig. 1) and found high colinearity all along five lepidopteran BetaEPV genomes (Fig. 1a). Central regions are highly conserved, while extremities lose orthology and synteny conservation. This finding suggests strong gene content and order conservation within the central regions of lepidopteran BetaEPV genomes. Interestingly, AHEV had a large genomic inversion located at kbp 120 to 175. This indicates that central regions are composed of two independent parts that may undergo inversions without any apparent effect on replication.

Similarly, we looked for colinearity conservation among EPVs (including the orthopteran EPV MSEV) (Fig. 1b). The five lepidopteran BetaEPV genomes showed less colinearity with the MSEV genome than with each other, as illustrated by fewer connecting lines in Fig. 1b than in Fig. 1a. The loss of gene content and order conservation between the five BetaEPV genomes and the orthopteran MSEV genome indicates that MSEV is evolutionarily divergent from lepidopteran EPVs, suggesting that orthopteran and lepidopteran EPVs indeed belong to different genera.

At the *Poxviridae* family level, the comparison of EPV genomes to the historical chordopoxvirus model, the VACV genome (Fig. 1c), highlighted the sparseness of colinearity. The few orthologous genes are located in the central region and correspond mostly to poxvirus core genes. In summary, as the genomes become more divergent, fewer orthologs are found between genomes. However, the central regions of poxvirus genomes retained certain levels of conservation, but with many inversions and rearrangements, corroborating previous studies (26, 46).

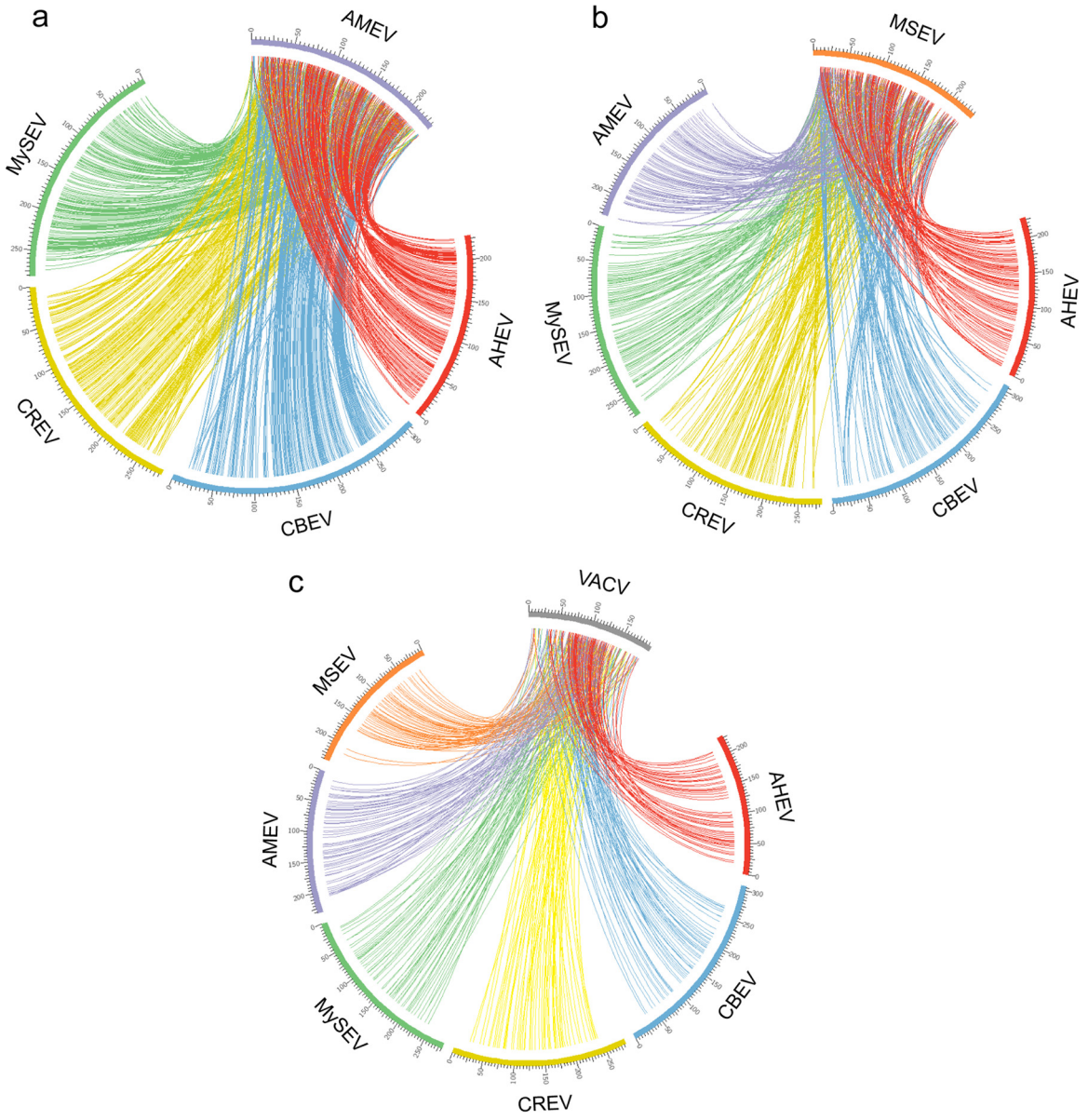

FIG 1 Poxvirus genome colinearity maps. The maps are based on the identification of orthologs by reciprocal best-hit analyses. The circles indicate genome colinearity conservation among lepidopteran BetaEPV genomes, with AMEV set as a reference (a), EPV genomes, with MSEV set as a reference (b), and poxvirus genomes, with VACV set as a reference (c).

EPV core genome. Protein clustering was performed on all EPV proteins to identify core genes for the *Entomopoxvirinae* subfamily and the *Betaentomopoxvirus* genus. This analysis grouped together ORFs sharing homologous domains. The size of the cluster corresponded to the number of times a particular homolog group was found in the genomes. Clusters representing gene families, such as the N1R/p28 gene family, contained more than a hundred genes. It was not possible to assign orthology between gene copies for such large clusters. They were, therefore, removed from the analyses, and we concentrated on genes present only once per genome. Core genes were defined as single-copy-number genes in the genomes of all members of a particular group. We determined that 104 genes are conserved in all EPV genomes and 148 in all BetaEPV genomes (Fig. 2 and Table 2). The 104 EPV core genes include the 49 poxvirus core genes (31) and 55 EPV-specific genes. Among these 55 genes, we identified the spheroidin, DNA photolyase, ubiquitin, putative thioredoxin, protein tyrosine phosphatase 2, protein phosphatase 1B, protein phosphatase 2C, lipase, and Ca^{2+} binding protein (BP) genes, as well as 46 ORFs of unknown function initially identified in the genome of AMEV. The 44 supplementary ORFs defining the BetaEPV core genes include those encoding the Cu/Zn superoxide dismutase, thymidine kinase, and a second poly(A) polymerase small subunit VP39, as well as 41 genes of unknown function. Although not included as core genes, the N1R/p28, leucine-rich, and tryptophan repeat gene families are present in all EPV genomes.

To determine if there were strict physical constraints on the order of the core genes, we analyzed the relative positions of the

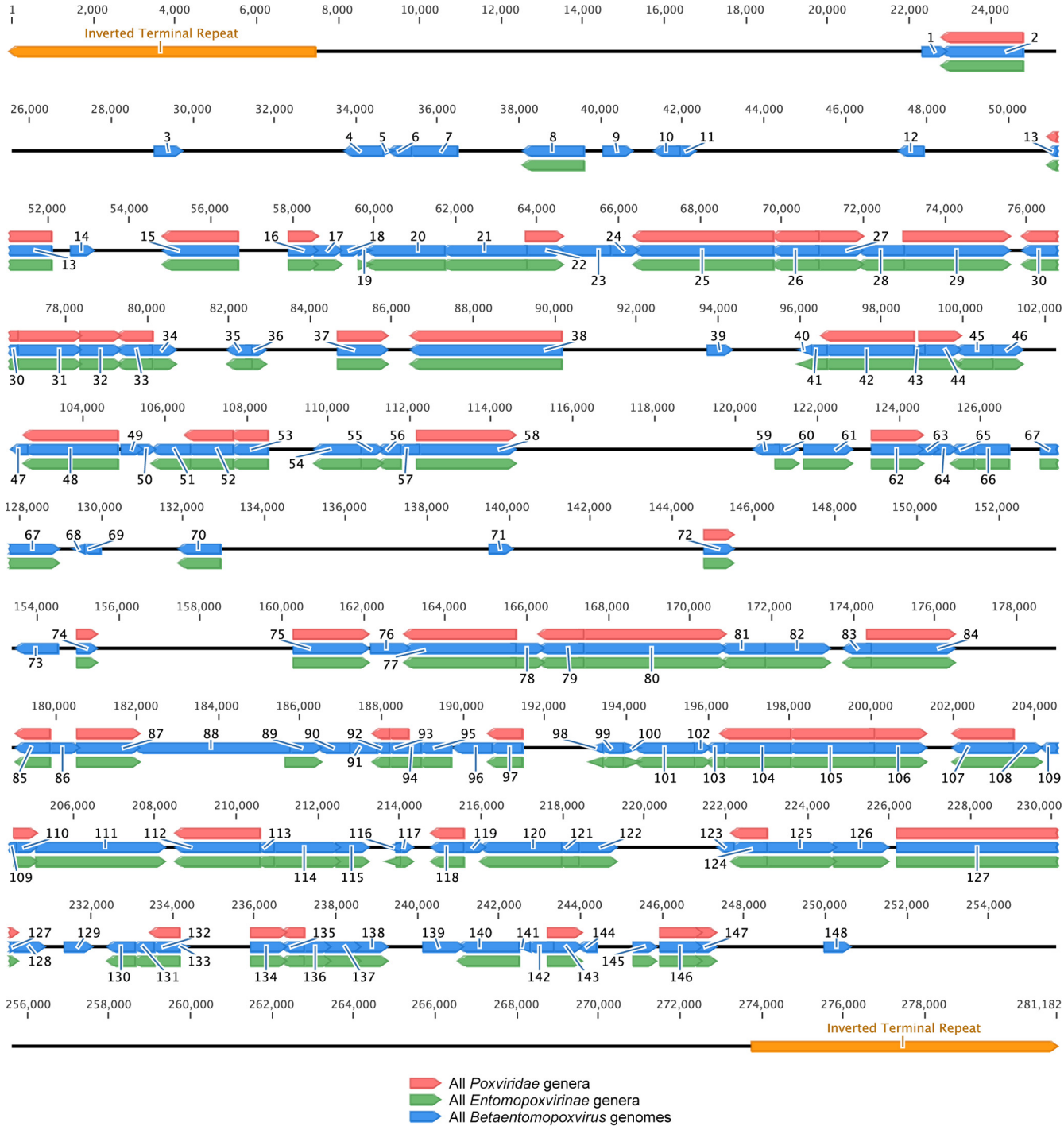

FIG 2 Localization of the 148 betaentomopoxvirus core genes in the genome of MySEV. Red, green, and blue arrows represent genes conserved in all poxvirus genera, in all EPV genera, and in all BetaEPV genomes, respectively. Orange arrows represent ITRs. MySEV noncore genes are not displayed. Genes are numbered as in Table 2.

clusters in all poxvirus genomes. We were not able to identify any colocalized core genes at the level of the *Poxviridae* family. In contrast, within the BetaEPVs, we found seven clusters of strict gene order conservation containing 2 to 17 adjacent genes (clusters B1 [$n = 2$], B2 [$n = 2$], B3 [$n = 9$], B4 [$n = 17$], B5 [$n = 5$], B6 [$n = 5$], and B7 [$n = 2$]) (Table 2). Cluster B1 includes genes involved in metal ion cell detoxification. Cluster B2 contains genes of unknown function. Cluster B3 includes genes involved in transcription/mRNA modification. Cluster B4 includes genes involved in DNA replication, transcription/mRNA modification, and virus-host interactions. Cluster B5 includes the RNA polymerase RPO147 and the Ca^{2+} binding protein. Cluster B6 includes the uracil-DNA glycosylase, DNA polymerase processivity factor and the putative late 16-kDa membrane protein (Cop-J5L).

TABLE 2 Betaentomopoxvirus core genome

No.	Predicted function or similarity[a]	ORF no. in the genome of:							Cluster
		AHEV	AMEV	CBEV	CREV	MySEV	MSEV	VACV	
1	Cu/Zn superoxide dismutase	AHEV230	AMEV255	CBEV291	CREV257	MySEV026			B1
2	Metalloprotease (Cop-G1L)	AHEV231	AMEV256	CBEV292	CREV258	MySEV027	MSEV056	VACV078	B1
3	Thymidine kinase	AHEV022	AMEV016	CBEV044	CREV024	MySEV035			
4	Unknown; similar to AMEV004	AHEV024	AMEV004	CBEV065	CREV045	MySEV039			
5	Unknown; similar to AMEV017	AHEV025	AMEV017	CBEV046	CREV026	MySEV040			B2
6	Unknown; similar to AMEV018	AHEV026	AMEV018	CBEV047	CREV027	MySEV041			B2
7	Unknown; similar to AMEV261	AHEV015	AMEV261	CBEV030	CREV011	MySEV042			
8	DNA photolyase	AHEV028	AMEV025	CBEV053	CREV035	MySEV044	MSEV235		
9	Unknown; similar to AMEV022	AHEV027	AMEV022	CBEV055	CREV036	MySEV046			
10	Unknown; similar to AMEV028	AHEV031	AMEV028	CBEV049	CREV029	MySEV048			
11	Unknown; similar to AMEV123	AHEV114	AMEV123	CBEV173	CREV141	MySEV049			
12	Unknown; similar to AMEV032	AHEV034	AMEV032	CBEV068	CREV047	MySEV056			
13	Entry-fusion complex component, myristylprotein	AHEV036	AMEV035	CBEV088	CREV067	MySEV060	MSEV121	VACV087	
14	Unknown; similar to AMEV034	AHEV035	AMEV034	CBEV086	CREV065	MySEV062			
15	Poly(A) polymerase catalytic subunit VP55	AHEV038	AMEV038	CBEV085	CREV064	MySEV064	MSEV143	VACV057	
16	NlpC/P60 superfamily protein (Cop-G6R)	AHEV040	AMEV041	CBEV082	CREV061	MySEV066	MSEV039	VACV084	
17	Unknown; similar to AMEV040	AHEV041	AMEV040	CBEV083	CREV062	MySEV067	MSEV138		
18	Unknown; similar to AMEV042	AHEV042	AMEV042	CBEV081	CREV060	MySEV068			
19	Unknown; similar to AMEV043	AHEV043	AMEV043	CBEV080	CREV059	MySEV069	MSEV188		
20	Unknown; similar to AMEV044	AHEV044	AMEV044	CBEV079	CREV058	MySEV070	MSEV140		
21	Unknown; similar to AMEV045	AHEV045	AMEV045	CBEV078	CREV057	MySEV071	MSEV077		
22	Late transcription factor VLTF-2	AHEV046	AMEV047	CBEV077	CREV056	MySEV072	MSEV187	VACV119	
23	Unknown; similar to AMEV048	AHEV047	AMEV048	CBEV071	CREV050	MySEV073			
24	Unknown; similar to AMEV049	AHEV048	AMEV049	CBEV070	CREV049	MySEV074			
25	DNA polymerase	AHEV049	AMEV050	CBEV069	CREV048	MySEV075	MSEV036	VACV065	
26	RNA polymerase RPO35	AHEV051	AMEV051	CBEV094	CREV070	MySEV076	MSEV149	VACV152	B3
27	DNA topoisomerase type I	AHEV052	AMEV052	CBEV095	CREV071	MySEV077	MSEV130	VACV104	B3
28	Unknown; similar to AMEV053	AHEV053	AMEV053	CBEV096	CREV072	MySEV078	MSEV120		B3
29	RNA polymerase-associated protein RAP94	AHEV054	AMEV054	CBEV097	CREV073	MySEV079	MSEV118	VACV102	B3
30	mRNA-decapping enzyme (Cop-D9R)	AHEV055	AMEV058	CBEV103	CREV080	MySEV081	MSEV150	VACV115	B3
31	DNA helicase, transcript release factor	AHEV056	AMEV059	CBEV104	CREV081	MySEV082	MSEV148	VACV138	B3
32	Poly(A) polymerase small subunit VP39	AHEV057	AMEV060	CBEV105	CREV082	MySEV083	MSEV041	VACV095	B3
33	ssDNA/dsDNA binding protein VP8 (Cop-L4R)	AHEV061	AMEV061	CBEV108	CREV083	MySEV084	MSEV158	VACV091	B3
34	Unknown; similar to AMEV062	AHEV062	AMEV062	CBEV109	CREV084	MySEV085	MSEV160		B3
35	Unknown; similar to AMEV072	AHEV063	AMEV072	CBEV124	CREV092	MySEV088	MSEV044		
36	Unknown; similar to AMEV071	AHEV064	AMEV071	CBEV125	CREV093	MySEV089	MSEV049		
37	Internal virion protein (Cop-L3L)	AHEV066	AMEV069	CBEV127	CREV095	MySEV092	MSEV180	VACV090	
38	RNA polymerase RPO132	AHEV068	AMEV066	CBEV129	CREV097	MySEV094	MSEV155	VACV144	
39	Protein tyrosine phosphatase 2	AHEV070	AMEV078	CBEV134	CREV102	MySEV098			B4
40	Putative thioredoxin	AHEV073	AMEV079	CBEV136	CREV104	MySEV102	MSEV087		B4
41	Unknown; similar to AMEV080	AHEV074	AMEV080	CBEV137	CREV105	MySEV103	MSEV085		B4, E1
42	RNA helicase, DExH-NPH-II domain	AHEV079	AMEV081	CBEV138	CREV106	MySEV104	MSEV086	VACV077	B4, E1
43	Unknown; similar to AMEV082	AHEV080	AMEV082	CBEV139	CREV107	MySEV105			B4
44	Entry and fusion IMV protein (Cop-L5R)	AHEV081	AMEV083	CBEV140	CREV108	MySEV106	MSEV129	VACV092	B4
45	Ser/Thr kinase (Cop-B1R)	AHEV082	AMEV084	CBEV141	CREV109	MySEV107	MSEV154		B4
46	Unknown; similar to AMEV085	AHEV083	AMEV085	CBEV142	CREV110	MySEV108	MSEV088		B4, E2
47	Unknown; similar to AMEV086	AHEV084	AMEV086	CBEV143	CREV111	MySEV109			B4
48	NTPase, DNA primase	AHEV085	AMEV087	CBEV144	CREV112	MySEV110	MSEV089	VACV110	B4, E2
49	Unknown; similar to AMEV088	AHEV086	AMEV088	CBEV145	CREV113	MySEV111			B4
50	Unknown; similar to AMEV089	AHEV087	AMEV089	CBEV146	CREV114	MySEV112			B4
51	Unknown; similar to AMEV090	AHEV088	AMEV090	CBEV147	CREV115	MySEV113	MSEV116		B4
52	Intermediate transcription factor VITF-3 45-kDa subunit (Cop-A23R)	AHEV089	AMEV091	CBEV148	CREV116	MySEV114	MSEV052	VACV143	B4
53	mRNA-capping enzyme small subunit	AHEV090	AMEV093	CBEV149	CREV117	MySEV115	MSEV124	VACV117	B4
54	Unknown; similar to AMEV096	AHEV092	AMEV096	CBEV151	CREV119	MySEV117	MSEV213		B4
55	Unknown; similar to AMEV098	AHEV093	AMEV098	CBEV152	CREV120	MySEV118	MSEV136		B4
56	Unknown; similar to AMEV102	AHEV108	AMEV102	CBEV157	CREV125	MySEV119	MSEV092		
57	Unknown; similar to AMEV104	AHEV107	AMEV104	CBEV158	CREV126	MySEV120			
58	Early transcription factor large subunit VETF-L	AHEV106	AMEV105	CBEV159	CREV127	MySEV121	MSEV063	VACV126	
59	Unknown; similar to AMEV107	AHEV105	AMEV107	CBEV160	CREV128	MySEV126			
60	Unknown; similar to AMEV101	AHEV109	AMEV101	CBEV156	CREV124	MySEV127	MSEV079		
61	Protein phosphatase 1B	AHEV097	AMEV119	CBEV167	CREV135	MySEV129	MSEV081		
62	Myristylated protein, essential for entry/fusion (Cop-A16L)	AHEV099	AMEV118	CBEV165	CREV133	MySEV131	MSEV090	VACV136	
63	Unknown; similar to AMEV117	AHEV100	AMEV117	CBEV164	CREV132	MySEV132			
64	Unknown; similar to AMEV116	AHEV101	AMEV116	CBEV163	CREV131	MySEV133			
65	Unknown; similar to AMEV120	AHEV096	AMEV120	CBEV168	CREV136	MySEV134	MSEV082		
66	Unknown; similar to AMEV121	AHEV111	AMEV121	CBEV169	CREV137	MySEV135	MSEV064		
67	Unknown; similar to AMEV099	AHEV095	AMEV099	CBEV154	CREV122	MySEV138	MSEV071		
68	Conotoxin-like protein	AHEV128	AMEV267	CBEV182	CREV150	MySEV140			
69	Unknown; similar to AMEV126	AHEV196	AMEV126	CBEV175	CREV143	MySEV141			
70	Lipase	AHEV192	AMEV133	CBEV184	CREV152	MySEV144	MSEV048		
71	Unknown; similar to AMEV075	AHEV012	AMEV075	CBEV193	CREV161	MySEV151			
72	Entry-fusion complex essential component (Cop-H2R)	AHEV194	AMEV127	CBEV181	CREV149	MySEV156	MSEV060	VACV100	
73	Poly(A) polymerase small subunit VP39	AHEV102	AMEV115	CBEV162	CREV130	MySEV168			
74	Sulfhydryl oxidase, FAD linked (Cop-E10R)	AHEV103	AMEV114	CBEV161	CREV129	MySEV170	MSEV093	VACV066	
75	Trimeric virion coat protein; rifampin resistance	AHEV112	AMEV122	CBEV171	CREV139	MySEV176	MSEV069	VACV118	
76	Unknown; similar to AMEV128	AHEV154	AMEV128	CBEV180	CREV148	MySEV177			
77	mRNA-capping enzyme large subunit	AHEV190	AMEV135	CBEV186	CREV154	MySEV178	MSEV066	VACV106	

(Continued on following page)

TABLE 2 (Continued)

No.	Predicted function or similarity[a]	ORF no. in the genome of: AHEV	AMEV	CBEV	CREV	MySEV	MSEV	VACV	Cluster
78	Unknown; similar to AMEV137	AHEV189	AMEV137	CBEV187	CREV155	MySEV179	MSEV068		
79	Viral membrane formation (Cop-A11R)	AHEV188	AMEV138	CBEV188	CREV156	MySEV180	MSEV151	VACV130	
80	P4a precursor	AHEV187	AMEV139	CBEV189	CREV157	MySEV181	MSEV152	VACV129	
81	Unknown; similar to AMEV140	AHEV186	AMEV140	CBEV190	CREV158	MySEV182	MSEV170		
82	Unknown; similar to AMEV141	AHEV185	AMEV141	CBEV191	CREV159	MySEV183	MSEV050		
83	Unknown; similar to AMEV145	AHEV183	AMEV145	CBEV195	CREV163	MySEV185	MSEV167		
84	P4b precursor	AHEV182	AMEV147	CBEV194	CREV162	MySEV186	MSEV164	VACV122	
85	ATPase/DNA-packaging protein	AHEV180	AMEV150	CBEV196	CREV164	MySEV190	MSEV171	VACV155	
86	Unknown; similar to AMEV151	AHEV179	AMEV151	CBEV197	CREV165	MySEV191			
87	Essential Ser/Thr kinase morph (Cop-F10L)	AHEV178	AMEV153	CBEV198	CREV166	MySEV192	MSEV173	VACV049	
88	Unknown; similar to AMEV156	AHEV177	AMEV156	CBEV199	CREV167	MySEV193			
89	Unknown; similar to AMEV157	AHEV176	AMEV157	CBEV200	CREV168	MySEV194	MSEV169		
90	Unknown; similar to AMEV159	AHEV175	AMEV159	CBEV201	CREV169	MySEV195			
91	Unknown; similar to AMEV160	AHEV174	AMEV160	CBEV202	CREV170	MySEV196			
92	Viral membrane-associated early morphogenesis protein (Cop-A9L)	AHEV173	AMEV161	CBEV203	CREV171	MySEV197	MSEV108	VACV128	
93	Holliday junction resolvase	AHEV172	AMEV162	CBEV204	CREV172	MySEV198	MSEV106	VACV142	
94	Unknown; similar to AMEV163	AHEV171	AMEV163	CBEV205	CREV173	MySEV199	MSEV112		
95	Unknown; similar to AMEV164	AHEV170	AMEV164	CBEV206	CREV174	MySEV200	MSEV107		
96	Unknown; similar to AMEV165	AHEV169	AMEV165	CBEV207	CREV175	MySEV201			
97	RNA polymerase RPO19	AHEV168	AMEV166	CBEV208	CREV176	MySEV202	MSEV101	VACV124	
98	Ubiquitin	AHEV166	AMEV167	CBEV209	CREV177	MySEV206	MSEV144		
99	Unknown; similar to AMEV168	AHEV165	AMEV168	CBEV210	CREV178	MySEV207	MSEV165		
100	Unknown; similar to AMEV169	AHEV164	AMEV169	CBEV212	CREV180	MySEV209	MSEV163		
101	Unknown; similar to AMEV173	AHEV160	AMEV173	CBEV216	CREV184	MySEV210	MSEV157		
102	Unknown; similar to AMEV171	AHEV162	AMEV171	CBEV214	CREV182	MySEV211	MSEV166		
103	Unknown; similar to AMEV172	AHEV161	AMEV172	CBEV215	CREV183	MySEV212	MSEV098		
104	Virion protein (Cop-E6R)	AHEV163	AMEV170	CBEV213	CREV181	MySEV213	MSEV145	VACV062	
105	Morph, early transcription factor small subunit (VETF-s)	AHEV159	AMEV174	CBEV217	CREV185	MySEV214	MSEV113	VACV111	
106	FEN1-like nuclease (Cop-G5R)	AHEV157	AMEV179	CBEV218	CREV186	MySEV215	MSEV115	VACV082	
107	Virion core cysteine protease	AHEV156	AMEV181	CBEV223	CREV191	MySEV217	MSEV189	VACV076	
108	Unknown; similar to AMEV183	AHEV155	AMEV183	CBEV224	CREV192	MySEV218	MSEV190		
109	Unknown; similar to AMEV185	AHEV153	AMEV185	CBEV225	CREV193	MySEV219			
110	IMV MP/virus entry (Cop-A28L)	AHEV152	AMEV186	CBEV226	CREV194	MySEV220	MSEV132	VACV151	
111	Spheroidin	AHEV151	AMEV187	CBEV227	CREV195	MySEV221	MSEV073		
112	ATPase, NPH1	AHEV150	AMEV192	CBEV228	CREV196	MySEV222	MSEV053	VACV116	
113	Unknown; similar to AMEV198	AHEV149	AMEV198	CBEV229	CREV197	MySEV223	MSEV161		
114	NAD-dependent DNA ligase	AHEV148	AMEV199	CBEV230	CREV198	MySEV224	MSEV162		
115	Unknown; similar to AMEV200	AHEV147	AMEV200	CBEV231	CREV199	MySEV225	MSEV159		
116	Unknown; similar to AMEV203	AHEV145	AMEV203	CBEV233	CREV201	MySEV228	MSEV168		
117	Unknown; similar to AMEV204	AHEV144	AMEV204	CBEV234	CREV202	MySEV229	MSEV095		
118	Late transcription factor VLTF-3	AHEV142	AMEV205	CBEV235	CREV203	MySEV231	MSEV065	VACV120	
119	Unknown; similar to AMEV206	AHEV141	AMEV206	CBEV236	CREV204	MySEV232			
120	DNA polymerase-beta/AP polymerase	AHEV140	AMEV210	CBEV237	CREV205	MySEV233	MSEV117		
121	Unknown; similar to AMEV211	AHEV139	AMEV211	CBEV239	CREV207	MySEV234	MSEV137		
122	Unknown; similar to AMEV219	AHEV138	AMEV219	CBEV245	CREV212	MySEV235	MSEV072		
123	Unknown; similar to AMEV218	AHEV136	AMEV218	CBEV244	CREV213	MySEV239			
124	IMV membrane protein (Cop-L1R)	AHEV135	AMEV217	CBEV246	CREV211	MySEV240	MSEV183	VACV088	
125	Unknown; similar to AMEV216	AHEV134	AMEV216	CBEV247	CREV210	MySEV241	MSEV099		
126	Unknown; similar to AMEV214	AHEV133	AMEV214	CBEV248	CREV209	MySEV242	MSEV184		
127	RNA polymerase RPO147	AHEV199	AMEV221	CBEV256	CREV222	MySEV244	MSEV042	VACV098	B5
128	Unknown; similar to AMEV224	AHEV201	AMEV224	CBEV260	CREV226	MySEV245			B5
129	Unknown; similar to AMEV225	AHEV203	AMEV225	CBEV261	CREV228	MySEV248			B5
130	Unknown; similar to AMEV226	AHEV204	AMEV226	CBEV262	CREV229	MySEV251	MSEV031		B5
131	Ca^{2+} BP	AHEV205	AMEV228	CBEV263	CREV230	MySEV252	MSEV097		B5
132	RNA polymerase RPO18	AHEV207	AMEV230	CBEV265	CREV232	MySEV253	MSEV245	VACV112	
133	Unknown; similar to AMEV229	AHEV206	AMEV229	CBEV264	CREV231	MySEV254			
134	Uracil-DNA glycosylase, DNA polymerase processivity factor	AHEV210	AMEV231	CBEV267	CREV234	MySEV257	MSEV208	VACV109	B6
135	Putative late 16-kDa membrane protein (Cop-J5L)	AHEV211	AMEV232	CBEV268	CREV235	MySEV258	MSEV142	VACV097	B6
136	Unknown; similar to AMEV233	AHEV212	AMEV233	CBEV269	CREV236	MySEV259	MSEV033		B6
137	Protein phosphatase 2C	AHEV213	AMEV234	CBEV272	CREV238	MySEV260	MSEV135		B6
138	Unknown; similar to AMEV235	AHEV214	AMEV235	CBEV273	CREV239	MySEV261	MSEV123		B6
139	Unknown; similar to AMEV240	AHEV218	AMEV240	CBEV276	CREV243	MySEV264			
140	Unknown; similar to AMEV238	AHEV217	AMEV238	CBEV275	CREV242	MySEV265	MSEV055		
141	Unknown; similar to AMEV241	AHEV219	AMEV241	CBEV277	CREV244	MySEV266			
142	Unknown; similar to AMEV245	AHEV224	AMEV245	CBEV280	CREV247	MySEV267			
143	S-S bond formation pathway protein substrate (Cop-F9L)	AHEV221	AMEV243	CBEV279	CREV246	MySEV268	MSEV094	VACV048	
144	Unknown; similar to AMEV242	AHEV220	AMEV242	CBEV278	CREV245	MySEV269			
145	Unknown; similar to AMEV247	AHEV222	AMEV247	CBEV283	CREV250	MySEV272	MSEV139		
146	IMV heparin binding surface protein	AHEV225	AMEV248	CBEV285	CREV252	MySEV274	MSEV206	VACV101	B7
147	IMV membrane protein entry/fusion complex component (Cop-A21L)	AHEV226	AMEV249	CBEV286	CREV253	MySEV275	MSEV209	VACV140	B7
148	Unknown; similar to AMEV013	AHEV019	AMEV013	CBEV061	CREV042	MySEV279			

[a] ssDNA, single-stranded DNA; FAD, flavin adenine dinucleotide; morph, morphogenesis; MP, membrane protein; AP, apurinic/apyrimidinic.

TABLE 3 Poxvirus genomes used in the phylogenomic analysis

Subfamily	Genus	Genome	Abbreviation	Genome accession no.
Chordopoxvirinae	*Avipoxvirus*	Fowlpox virus strain Iowa	FWPV	NC_002188
	Capripoxvirus	Sheeppox virus strain 17077-99	SPPV	NC_004002
	Cervidpoxvirus	Deerpox virus strain W-848-83	DPV	NC_006966
	Leporipoxvirus	Myxoma virus strain Lausanne	MYXV	NC_001132
	Molluscipoxvirus	Molluscum contagiosum virus strain subtype 1	MOCV	NC_001731
	Orthopoxvirus	Vaccinia virus strain Western Reserve	VACV	NC_006998
	Parapoxvirus	Orf virus strain OV-SA00	ORFV	NC_005336
	Suipoxvirus	Swinepox virus strain Nebraska 17077–99	SWPV	NC_003389
	Yatapoxvirus	Yaba monkey tumor virus strain Amano	YMTV	NC_005179
	Crocodylipoxvirus	Nile crocodile poxvirus strain Zimbabwe	CRV	NC_008030
Entomopoxvirinae	*Betaentomopoxvirus*	*Adoxophyes honmai* entomopoxvirus"L" strain Japan	AHEV	HF679131
		Amsacta moorei entomopoxvirus "L" strain Moyer	AMEV	NC_002520
		Choristoneura biennis entomopoxvirus "L" strain Canada	CBEV	HF679132
		Choristoneura rosaceana entomopoxvirus "L" strain Canada	CREV	HF679133
		Mythimna separata entomopoxvirus "L" strain China	MySEV	HF679134
	Unclassified	*Melanoplus sanguinipes* entomopoxvirus strain Tucson	MSEV	NC_001993

Finally, cluster B7 includes two surface/membrane proteins of the intracellular mature virion (IMV). None of these clusters are conserved in MSEV. At the EPV level, only two clusters of two adjacent genes could be found. The first cluster (E1) includes the RNA helicase DExH-NPH-II domain and an unknown gene, AMEV080, and the second cluster (E2) includes the nucleoside triphosphatase (NTPase), DNA primase, and an unknown gene, AMEV085.

Whole-genome poxvirus phylogeny. Phylogenetic analysis was conducted on the 49 poxvirus core genes (31) for which homologs were identified in 12 poxvirus species representative of each poxvirus genus and in AHEV, CBEV, CREV, and MySEV (Table 3). A concatenated multiple alignment of the 49 poxvirus core genes was used to reconstruct the poxvirus phylogeny by maximum likelihood inference. In accordance with previous studies (47–49), we obtained a highly supported phylogeny (Fig. 3) showing two major monophyletic clades corresponding to the chordopoxvirus and EPV subfamilies. AMEV, AHEV, CBEV, CREV, and MySEV grouped in a well-supported monophyletic lineage corroborating their affiliation within a single genus. Within the BetaEPVs, AHEV, CBEV, and CREV are closer to AMEV than MySEV. Moreover, CBEV and CREV, infecting hosts belonging to the same genus, are very closely related, even though *C. biennis* is a forestry pest while *C. rosaceana* is a pest of apple orchards.

Spheroidin phylogeny. The spheroidin amino acid sequence phylogeny based on a larger sampling of EPV taxa (Fig. 4) showed strong phylogenetic similarity in terms of tree topology as well as branch length with the whole-genome EPV phylogeny (Fig. 3). This suggests that the spheroidin gene bears a good phylogenetic signal reflecting EPV species phylogeny. The phylogeny of all the spheroidin proteins available in public databases included sequences from coleopteran EPVs of the genus *Alphaentomopoxvirus*. Strikingly, the tree showed a clear division of the EPVs according to the orders of their insect hosts.

DISCUSSION

Here we report the complete genome sequences of four entomopoxviruses. This is long overdue, since the previous two EPV genomes were published more than 10 years ago (25, 26). The AHEV, CBEV, CREV, and MySEV genomes have general characteristics similar to those of the two EPV genomes sequenced previously. They are extremely AT rich, a reason why obtaining and assembling their sequences had been problematic (11).

EPV comparative genomics. Like other poxvirus genomes, EPV genomes possess a central region encoding essential core proteins and terminal regions containing less conserved, nonessential, and orphan proteins, possibly involved in virus-host responses. Colinearity analyses showed that the five lepidopteran BetaEPV genomes are similar and that the orthopteran EPV MSEV is evolutionarily divergent (Fig. 1).

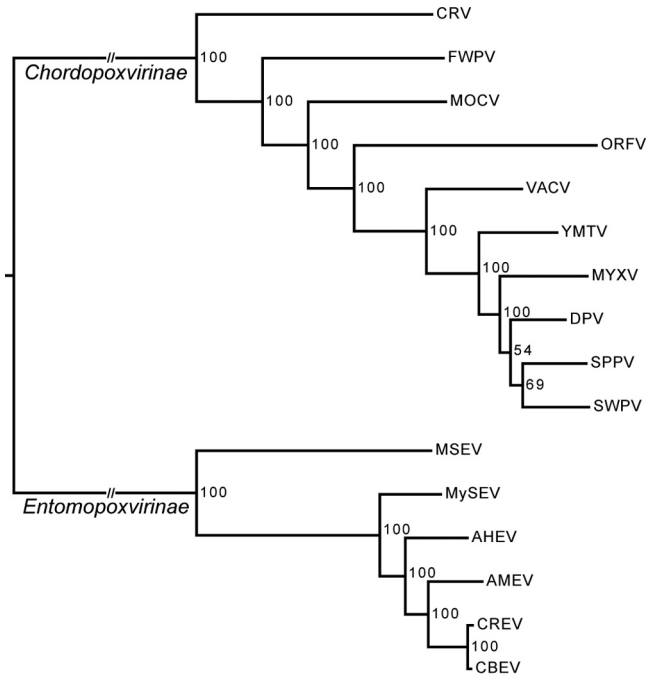

FIG 3 Whole-genome poxvirus phylogeny. The tree was obtained from maximum likelihood inference analysis of a concatenated amino acid multiple alignment of the 49 poxvirus core genes. Support for nodes indicates maximum likelihood nonparametric bootstraps (100 replicates). Full virus names are listed in Table 3.

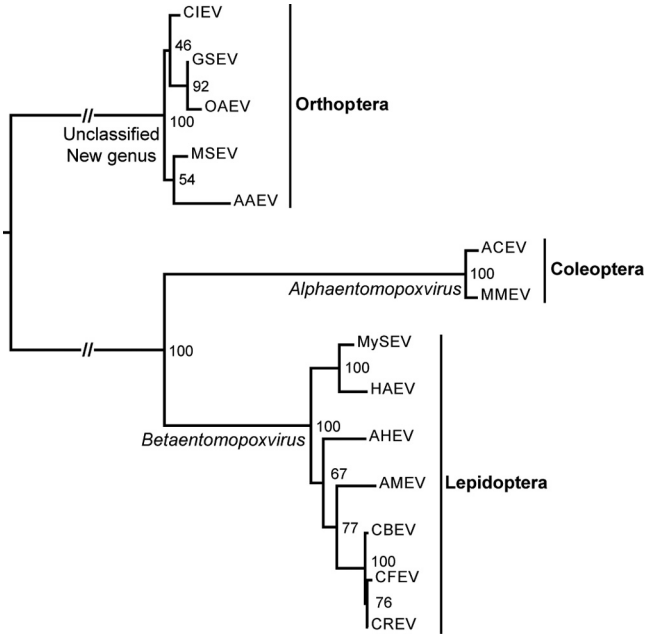

FIG 4 Spheroidin gene phylogeny. The tree was obtained from maximum likelihood inference analysis of the spheroidin gene amino acid alignment. Support for nodes indicates maximum likelihood nonparametric bootstraps (100 replicates). Abbreviations: CIEV, *Calliptamus italicus* entomopoxvirus; GSEV, *Gomphocerus sibiricus* entomopoxvirus; OAEV, *Oedaleus asiaticus* entomopoxvirus; AAEV, *Anacridium aegyptium* entomopoxvirus; ACEV, *Anomala cuprea* entomopoxvirus; MMEV, *Melolontha melolontha* entomopoxvirus; HAEV, *Heliothis armigera* entomopoxvirus "L"; CFEV, *Choristoneura fumiferana* entomopoxvirus "L."

The sizes of BetaEPV genomes are extremely variable; CBEV, CREV, and MySEV are at least 50 kbp larger than the average size of other known EPVs. Larger genome sizes are due mainly to large protein families of unknown functions with many members repeated in tandem and predominantly clustered in terminal regions but also dispersed all along the genomes. The N1R/p28 genes are the most abundant gene family (>150 members found in all 5 BetaEPV genomes). These genes have also been identified in other NCLDV, such as iridoviruses and mimiviruses (50, 51), and some contain baculovirus repeated ORF (*bro*) domains. Considering the number of repeated members present in genomes, they could have important adaptive roles as virulence factors (52). Moreover, we identified several orphan genes found in other insect viruses, notably in baculoviruses, that could be involved in adaptation.

As observed within the *Chordopoxvirinae* subfamily (46, 53), global genome synteny is highly conserved among lepidopteran EPVs but less conserved at the level of the *Entomopoxvirinae* subfamily. There is, however, no gene synteny between chordopoxviruses and EPVs, pointing to significant gene rearrangements after the division and radiation of the two subfamilies. In contrast, with 49 conserved genes shared by all poxvirus genomes (31) and 104 shared by all EPV genomes, conservation of gene content remains remarkably substantial (Fig. 2; Table 2). This suggests that poxviruses need a relatively large number of core genes to perform complex functions. Yet gene order conservation does not appear to be crucial. The minimum poxvirus gene set of 49 is doubled for the EPV subfamily and encompasses additional genes related to EPV ecology, such as the spheroidin and DNA photolyase genes, both protecting virions from environmental degradation (11, 54). The number of BetaEPV core genes is 148, accounting for half to two-thirds of the overall number of genes predicted in each genome. Many of these genes, notably those encoding replication, transcription/mRNA modification, and envelope proteins, are arranged in a strict order within this genus, which may indicate that strong conservative selection pressure has kept the genes in this particular order. A similar trend has been observed in chordopoxviruses (48). The poxvirus linear genome structure could support sequential gene expression to ensure essential morphogenesis pathways, which may still be perceptible at the genus level but may be lost at higher taxonomic levels.

EPV phylogeny and taxonomy. Phylogenetic analyses of the 49 poxvirus core genes (Fig. 3) show that the four new genomes are more closely related to AMEV than to any other poxvirus. This confirms that AHEV, CBEV, CREV, and MySEV, isolated from lepidopteran hosts, belong to the genus *Betaentomopoxvirus*. The spheroidin phylogeny, including more EPV isolates, indicates that EPVs infecting insects from the same taxonomic order (Lepidoptera, Orthoptera, or Coleoptera) group together. There is thus a clear partition of the EPVs according to the orders of their insect hosts. The EPV genera were historically based on host range and virion morphology. The *Betaentomopoxvirus* genus was established as comprising viruses infecting Orthoptera and Lepidoptera. However, based on genomic divergence, the species *Melanoplus sanguinipes entomopoxvirus* "*O*," infecting Orthoptera, was removed from the genus (4). Our phylogenetic analyses showed that orthopteran EPVs are excluded from the *Betaentomopoxvirus* genus. This suggests that host order could be a good criterion for defining EPV genera and that a new genus should be established for orthopteran EPVs (Fig. 4). This implies an ancient coevolution of EPVs with their insect hosts similar to that observed with baculoviruses and other large DNA viruses of insects (55, 56). The genome tree also shows that the subfamily *Chordopoxvirinae* is phylogenetically structured according to the taxonomic class of the host (mammals, birds, and reptiles). The coevolution between poxviruses and their hosts that culminated in their present distribution and host range suggests a remote virus origin, presumably going back to the common ancestors of vertebrates and insects: the first bilaterian Metazoa (57, 58).

Although the *Entomopoxvirinae* are structured according to the orders of the insect hosts, this virus clustering according to host taxonomy is not observable within the genus *Betaentomopoxvirus.* The phylogenies show an entanglement of EPVs infecting different lepidopteran host families (Arctiidae, Noctuidae, and Tortricidae) (59). EPVs, and large DNA viruses in general, tend to exhibit a fairly narrow host range (60), but the close phylogenetic relationships of EPVs infecting distant hosts suggest that large host shifts can occur. Current pathology data on EPVs show their relative host specificities (e.g., AHEV) (24). But generalists, such as the *Heliothis armigera entomopoxvirus* "*L*" (HAEV) (David Dall, personal communication), could promote host shifts, explaining the tangled phylogenetic relationships within the BetaEPVs.

Comparison of CREV and CBEV. Within the *Betaentomopoxvirus* genus, AHEV, AMEV, MySEV, and HAEV are phylogenetically well differentiated, as should be expected for viruses belonging to different species (Fig. 3 and 4). In contrast, CBEV and CREV are quite closely related phylogenetically, calling for a closer examination to determine if they are the same or distinct species.

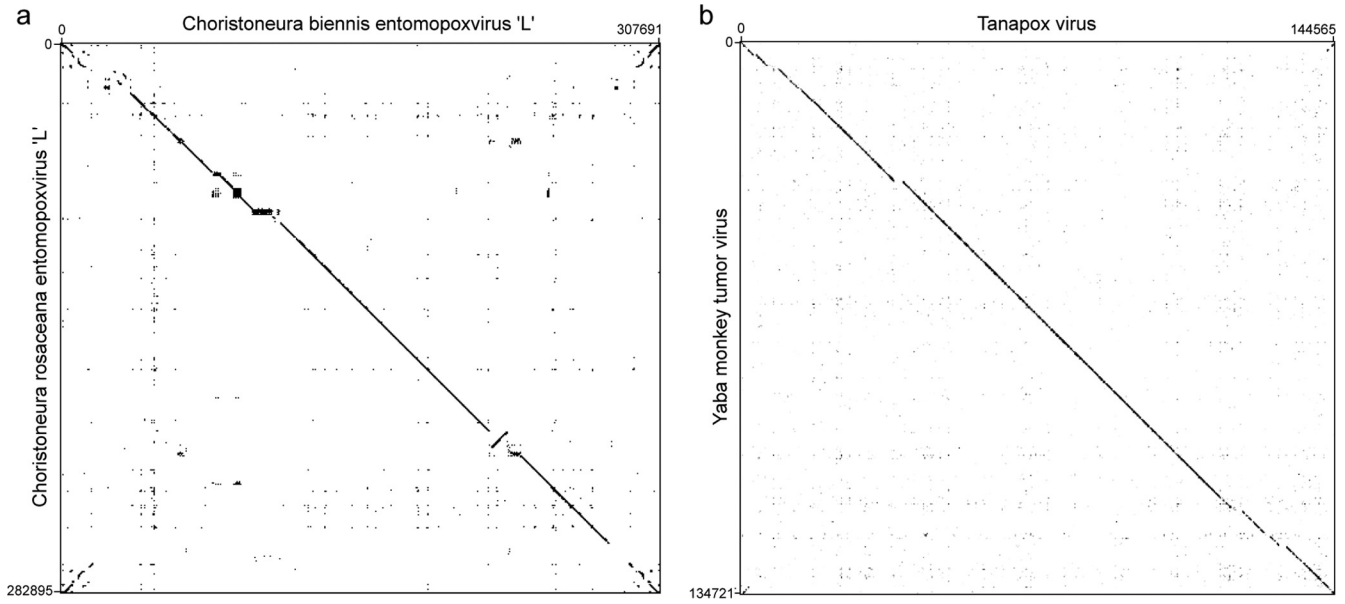

FIG 5 Genome synteny visualization by dot plots. The dot plots were obtained from whole-genome DNA homology alignments, using the Gepard program, between CBEV and CREV (a) and between the *Tanapox virus* and *Yaba monkey tumor virus* species (b).

CBEV and CREV were both isolated in Canada from phytophagous pests belonging to the same genus. CBEV was isolated from *C. biennis*, the 2-year-cycle budworm, a forest pest feeding mostly on spruce trees, and CREV from *C. rosaceana*, the oblique-banded leafroller, a pest of orchard trees, such as apples, prunes, and cherries, and some hardwood. If the two viruses infect closely related hosts and share the same geographical range, they appear to be linked to different ecological habitats. The 49 core poxvirus gene nucleotide sequences are 97.2% identical in CREV and CBEV. This is well within the 96% identity proposed to differentiate among orthopoxvirus species but below the 98% accepted within-strain variation (4), suggesting that CREV and CBEV could be different strains of the same viral species.

The genomes of CREV and CBEV are, however, quite different in size and gene content (Table 1). CBEV is ~25 kb larger than CREV; the difference is mostly explained by the large CBEV ITRs containing several N1R/p28 gene copies and genes coding for hypothetical proteins. The remaining difference corresponds to genes coding for hypothetical proteins spread all along both genomes. Overall, 35 genes are different in the two genomes, corresponding to around 10% of both genomes. Furthermore, using dot plots (created with the Gepard program [61]), we compared genome synteny between the two genomes infecting *Choristoneura* species (CBEV and CREV) (Fig. 5a) and between genomes of two different chordopoxvirus species belonging to the same genus (*Tanapox virus* and *Yaba monkey tumor virus*, both species of the *Yatapoxvirus* genus) (Fig. 5b). We observed more rearrangements, deletions, and insertions between the CBEV and CREV genomes than between the *Yatapoxvirus* species (Fig. 5). These differences in genomic content and organization suggest that CBEV and CREV should be classified into different species, even if this classification was not corroborated by phylogenetic relationships and core gene nucleotide distances.

This discrepancy implies that we cannot apply the orthopoxvirus species genetic distance to define entomopoxvirus species. Although phylogenetic relationships and core gene nucleotide distances show the closeness of CBEV and CREV, they infect different hosts of the same genus and are specialized to clearly different ecological niches, implying that the two viruses are very likely to belong to two separate species.

Conclusions. The genome sequences of AHEV, CBEV, CREV, and MySEV have provided new insights into EPV genomic organization and evolution. Our results allow certain generalizations on the structure of poxvirus genomes. Like those of chordopoxviruses, EPV genomes are structured in two parts, which appear to have evolved quite differently: the central core region and the more divergent terminal regions. Genetic diversity within the central core is relatively low in the BetaEPVs, resulting in high genome colinearity, both in terms of gene content and in terms of synteny conservation. However, the central core is much less diverse at the *Entomopoxvirinae* subfamily and *Poxviridae* family levels. The terminal regions, containing large gene families, as well as orphan genes, could play an important role in the adaptation of viruses to their hosts. In particular, the N1R/p28 gene family could play an adaptive role similar to that of the K3L antihost factor in orthopoxviruses, which was recently described as forming adaptive genomic accordions (62, 63).

Phylogenies showed the long history of coevolution between poxviruses and their hosts. The *Entomopoxvirinae* are grouped based on the orders of their insect hosts, suggesting that taxonomic revision is necessary. Basic pathological and genomic knowledge of EPVs, however, remains sparse, particularly for alpha- and gammaentomopoxviruses. This diverse, understudied group of viruses could find new applications as microbial biocontrol agents for sustainable agriculture. Finally, a better understanding of the early origin and evolution of the *Poxviridae* could shed new light on the evolutionary history of all large DNA viruses.

ACKNOWLEDGMENTS

European Research Council grant 205206 GENOVIR funded J. Thézé, J. Gallais, and E. A. Herniou. The sequencing of the AHEV and MySEV genomes was supported by JSPS KAKENHI grant 21380038, and that of CBEV and CREV genomes by a grant from Genome Canada through the Ontario Genomics Institute.

We thank David Dall for discussions on entomopoxvirus host range.

REFERENCES

1. **Iyer LM, Aravind L, Koonin EV.** 2001. Common origin of four diverse families of large eukaryotic DNA viruses. J. Virol. **75:**11720–11734.
2. **Wittek R, Menna A, Müller HK, Schümperli D, Boseley PG, Wyler R.** 1978. Inverted terminal repeats in rabbit poxvirus and vaccinia virus DNA. J. Virol. **28:**171–181.
3. **Schramm B, Locker JK.** 2005. Cytoplasmic organization of poxvirus DNA replication. Traffic **6:**839–846.
4. **Skinner MA, Buller RM, Damon IK, Lefkowitz EJ, McFadden G, McInnes CJ, Mercer AA, Moyer RW, Upton C.** 2011. *Poxviridae*, p 291–309. *In* King AM, Lefkowitz E, Adams MJ, Carstens EB (ed), Virus taxonomy. Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier, Amsterdam, Netherlands.
5. **Mercer AA, Schmidt A, Weber OF.** 2007. Poxviruses. Birkhäuser, Basel, Switzerland.
6. **McFadden G.** 2005. Poxvirus tropism. Nat. Rev. Microbiol. **3:**201–213.
7. **Lawrence PO.** 28 May 2002. Purification and partial characterization of an entomopoxvirus (DLEPV) from a parasitic wasp of tephritid fruit flies. J. Insect Sci. **2:**10. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC355910/.
8. **Lawrence PO.** 2005. Morphogenesis and cytopathic effects of the *Diachasmimorpha longicaudata* entomopoxvirus in host haemocytes. J. Insect Physiol. **51:**221–233.
9. **Clark TB.** 1982. Entomopoxvirus-like particles in three species of bumblebees. J. Invertebr. Pathol. **39:**119–122.
10. **Radek R, Fabel P.** 2000. A new entomopoxvirus from a cockroach: light and electron microscopy. J. Invertebr. Pathol. **75:**19–27.
11. **Perera S, Li Z, Pvalik L, Arif B.** 2010. Entomopoxviruses, p 83–115. *In* Asgari S, Johnson KN (ed), Insect virology. Caister Academic Press, Norfolk, United Kingdom.
12. **Rohrmann GF.** 1986. Polyhedrin structure. J. Gen. Virol. **67:**1499–1513.
13. **Arif BM.** 1995. Recent advances in the molecular biology of entomopoxviruses. J. Gen. Virol. **76:**1–13.
14. **Bilimoria SL, Arif BM.** 1979. Subunit protein and alkaline protease of entomopoxvirus spheroids. Virology **96:**596–603.
15. **Roberts DW, Granados RR.** 1968. A poxlike virus from *Amsacta moorei* (Lepidoptera: Arctiidae). J. Invertebr. Pathol. **12:**141–143.
16. **Volkman LE.** 2007. Baculovirus infectivity and the actin cytoskeleton. Curr. Drug Targets **8:**1075–1083.
17. **Ishii T, Takatsuka J, Nakai M, Kunimi Y.** 2002. Growth characteristics and competitive abilities of a nucleopolyhedrovirus and an entomopoxvirus in larvae of the smaller tea tortrix, *Adoxophyes honmai* (Lepidoptera: Tortricidae). Biol. Control **23:**96–105.
18. **Goodwin RH.** 1991. Replacement of vertebrate serum with lipids and other factors in the culture of invertebrate cells, tissues, parasites, and pathogens. In Vitro Cell. Dev. Biol. **27A:**470–478.
19. **Nakai M, Sakai T, Kunimi Y.** 1997. Effect of entomopoxvirus infection of the smaller tea tortrix, *Adoxophyes* sp. on the development of the endoparasitoid, *Ascogaster reticulatus.* Entomol. Exp. Appl. **84:**27–32.
20. **Nakai M, Kunimi Y.** 1998. Effects of the timing of entomopoxvirus administration to the smaller tea tortrix, *Adoxophyes* sp. (Lepidoptera: Tortricidae) on the survival of the endoparasitoid, *Ascogaster reticulatus* (Hymenoptera: Braconidae). Biol. Control **13:**63–69.
21. **Bird FT, Sanders CJ, Burke JM.** 1971. A newly discovered virus disease of the spruce budworm, *Choristoneura biennis* (Lepidoptera: Tortricidae). J. Invertebr. Pathol. **18:**159–161.
22. **Hukuhara T, Xu JH, Yano K.** 1990. Replication of an entomopoxvirus in two lepidopteran cell lines. J. Invertebr. Pathol. **56:**222–232.
23. **Cory J.** 1997. Use of baculoviruses as biological insecticides. Mol. Biotechnol. **7:**303–313.
24. **Takatsuka J, Okuno S, Ishii T, Nakai M, Kunimi Y.** 2010. Fitness-related traits of entomopoxviruses isolated from *Adoxophyes honmai* (Lepidoptera: Tortricidae) at three localities in Japan. J. Invertebr. Pathol. **105:**121–131.
25. **Afonso CL, Tulman ER, Lu Z, Oma E, Kutish GF, Rock DL.** 1999. The genome of *Melanoplus sanguinipes* entomopoxvirus. J. Virol. **73:**533–552.
26. **Bawden AL, Glassberg KJ, Diggans J, Shaw R, Farmerie W, Moyer RW.** 2000. Complete genomic sequence of the *Amsacta moorei* entomopoxvirus: analysis and comparison with other poxviruses. Virology **274:**120–139.
27. **Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437:**376–380.
28. **Sanger F, Coulson AR.** 1978. The use of thin acrylamide gels for DNA sequencing. FEBS Lett. **87:**107–110.
29. **Salzberg SL, Delcher AL, Kasif S, White O.** 1998. Microbial gene identification using interpolated Markov models. Nucleic Acids Res. **26:**544–548.
30. **Ehlers A, Osborne J, Slack S, Roper RL, Upton C.** 2002. Poxvirus Orthologous Clusters (POCs). Bioinformatics **18:**1544–1545.
31. **Upton C, Slack S, Hunter AL, Ehlers A, Roper RL.** 2003. Poxvirus orthologous clusters: toward defining the minimum essential poxvirus genome. J. Virol. **77:**7590–7600.
32. **Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang ZQ, Miller W, Lipman DJ.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
33. **Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA.** 2009. Circos: an information aesthetic for comparative genomics. Genome Res. **19:**1639–1645.
34. **Eddy SR.** 1998. Profile hidden Markov models. Bioinformatics **14:**755–763.
35. **Pavesi G, Mauri G, Iannelli F, Gissi C, Pesole G.** 2004. GeneSyn: a tool for detecting conserved gene order across genomes. Bioinformatics **20:**1472–1474.
36. **Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG.** 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. **7:**539. doi:10.1038/msb.2011.75.
37. **Keane TM, Creevey CJ, Pentony MM, Naughton TJ, Mclnerney JO.** 2006. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol. Biol. **6:**29. doi:10.1186/1471-2148-6-29.
38. **Stamatakis A.** 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22:**2688–2690.
39. **Sriskantha A, Osborne RJ, Dall DJ.** 1997. Mapping of the *Heliothis armigera* entomopoxvirus (HaEPV) genome, and analysis of genes encoding the HaEPV spheroidin and nucleoside triphosphate phosphohydrolase I proteins. J. Gen. Virol. **78:**3115–3123.
40. **Li X, Barrett JW, Yuen L, Arif BM.** 1997. Cloning, sequencing and transcriptional analysis of the *Choristoneura fumiferana* entomopoxvirus spheroidin gene. Virus Res. **47:**143–154.
41. **Sanz P, Veyrunes JC, Cousserans F, Bergoin M.** 1994. Cloning and sequencing of the spherulin gene, the occlusion body major polypeptide of the *Melolontha melolontha* entomopoxvirus (MmEPV). Virology **202:**449–457.
42. **Mitsuhashi W, Saito H, Sato M, Nakashima N, Noda H.** 1998. Complete nucleotide sequence of spheroidin gene of *Anomala cuprea* entomopoxvirus. Virus Res. **55:**61–69.
43. **Hernandez-Crespo P, Veyrunes JC, Cousserans F, Bergoin M.** 2000. The spheroidin of an entomopoxvirus isolated from the grasshopper *Anacridium aegyptium* (AaEPV) shares low homology with spheroidins from lepidopteran or coleopteran EPVs. Virus Res. **67:**203–213.
44. **Zhao C, Wang L, Li Y, Yun G.** 2003. Cloning and analysis of *Oedaleus asiaticus* entomopoxvirus spheroidin gene. Virol. Sin. **18:**593–596.
45. **Li YD, Wang LY, Gaol XW, Zhao CY, Tian ZF.** 2004. Complete nucleotide sequence of spheroidin genes of *Calliptamus italicus* entomopoxvirus (CiEPV) and *Gomphocerus sibiricus* entomopoxvirus (GsEPV). Insect Sci. **11:**173–182.

46. **McLysaght A, Baldi PF, Gaut BS.** 2003. Extensive gene gain associated with adaptive evolution of poxviruses. Proc. Natl. Acad. Sci. U. S. A. **100:** 15655–15660.
47. **Xing K, Deng R, Wang J, Feng J, Huang M, Wang X.** 2006. Genome-based phylogeny of poxvirus. Intervirology **49:**207–214.
48. **Bratke KA, McLysaght A.** 2008. Identification of multiple independent horizontal gene transfers into poxviruses using a comparative genomics approach. BMC Evol. Biol. **8:**67. doi:10.1186/1471-2148-8-67.
49. **Wu GA, Jun SR, Sims GE, Kim SH.** 2009. Whole-proteome phylogeny of large dsDNA virus families by an alignment-free method. Proc. Natl. Acad. Sci. U. S. A. **106:**12826–12831.
50. **Wong CK, Young VL, Kleffmann T, Ward VK.** 2011. Genomic and proteomic analysis of invertebrate iridovirus type 9. J. Virol. **85:**7900–7911.
51. **Legendre M, Santini S, Rico A, Abergel C, Claverie J-M.** 2011. Breaking the 1000-gene barrier for Mimivirus using ultra-deep genome and transcriptome sequencing. Virol. J. **8:**99. doi:10.1186/1743-422X-8-99.
52. **Nicholls R, Gray T.** 2004. Cellular source of the poxviral N1R/p28 gene family. Virus Genes **29:**359–364.
53. **Lefkowitz EJ, Wang C, Upton C.** 2006. Poxviruses: past, present and future. Virus Res. **117:**105–118.
54. **Nalcacioglu R, Dizman YA, Vlak JM, Demirbag Z, van Oers MM.** 2010. *Amsacta moorei* entomopoxvirus encodes a functional DNA photolyase (AMV025). J. Invertebr. Pathol. **105:**363–365.
55. **Herniou EA, Olszewski JA, O'Reilly DR, Cory JS.** 2004. Ancient coevolution of baculoviruses and their insect hosts. J. Virol. **78:**3244–3251.
56. **Thézé J, Bézier A, Periquet G, Drezen JM, Herniou EA.** 2011. Paleozoic origin of insect large dsDNA viruses. Proc. Natl. Acad. Sci. U. S. A. **108:** 15931–15935.
57. **Ruiz-Trillo I, Riutort M, Littlewood DTJ, Herniou EA, Baguna J.** 1999. Acoel flatworms: earliest extant bilaterian metazoans, not members of Platyhelminthes. Science **283:**1919–1923.
58. **Peterson KJ, Lyons JB, Nowak KS, Takacs CM, Wargo MJ, McPeek MA.** 2004. Estimating metazoan divergence times with a molecular clock. Proc. Natl. Acad. Sci. U. S. A. **101:**6536–6541.
59. **Mutanen M, Wahlberg N, Kaila L.** 2010. Comprehensive gene and taxon coverage elucidates radiation patterns in moths and butterflies. Proc. R. Soc. B **277:**2839–2848.
60. **Villarreal LP, Defilippis VR, Gottlieb KA.** 2000. Acute and persistent viral life strategies and their relationship to emerging diseases. Virology **272:**1–6.
61. **Krumsiek J, Arnold R, Rattei T.** 2007. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics **23:**1026–1028.
62. **Elde NC, Child SJ, Eickbush MT, Kitzman JO, Rogers KS, Shendure J, Geballe AP, Malik HS.** 2012. Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses. Cell **150:**831–841.
63. **Anderson RP, Roth JR.** 1977. Tandem genetic duplications in phage and bacteria. Annu. Rev. Microbiol. **31:**473–505.