JOURNAL OF VIROLOGY, Nov. 2004, p. 12576–12590 Vol. 78, No. 22
0022-538X/04/$08.00+0 DOI: 10.1128/JVI.78.22.12576–12590.2004


# Functional Genomics Analysis of Singapore Grouper Iridovirus: Complete Sequence Determination and Proteomic Analysis

Wen Jun Song,[1] Qi Wei Qin,[2] Jin Qiu,[1] Can Hua Huang,[1] Fan Wang,[1] and Choy Leong Hew[1]*

*Department of Biological Sciences[1] and Tropical Marine Science Institute,[2] National University of Singapore, Singapore*

Received 19 March 2004/Accepted 29 June 2004

**Here we report the complete genome sequence of Singapore grouper iridovirus (SGIV). Sequencing of the random shotgun and restriction endonuclease genomic libraries showed that the entire SGIV genome consists of 140,131 nucleotide bp. One hundred sixty-two open reading frames (ORFs) from the sense and antisense DNA strands, coding for lengths varying from 41 to 1,268 amino acids, were identified. Computer-assisted analyses of the deduced amino acid sequences revealed that 77 of the ORFs exhibited homologies to known virus genes, 23 of which matched functional iridovirus proteins. Forty-two putative conserved domains or signatures were detected in the National Center for Biotechnology Information CD-Search database and PROSITE database. An assortment of enzyme activities involved in DNA replication, transcription, nucleotide metabolism, cell signaling, etc., were identified. Viruses were cultured on a cell line derived from the embryonated egg of the grouper *Epinephelus tauvina*, isolated, and purified by sucrose gradient ultracentrifugation. The protein extract from the purified virions was analyzed by polyacrylamide gel electrophoresis followed by in-gel digestion of protein bands. Matrix-assisted laser desorption ionization–time of flight mass spectrometry and database searching led to identification of 26 proteins. Twenty of these represented novel or previously unidentified genes, which were further confirmed by reverse transcription-PCR (RT-PCR) and DNA sequencing of their respective RT-PCR products.**

Iridoviruses are animal viruses that infect only invertebrates and poikilothermic vertebrates, such as fish, insects, amphibians, and reptiles (40). They have been implicated as causative agents of serious systemic diseases among cultured and ornamental fish, as well as wild fish. Within the family *Iridoviridae*, four genera of DNA-containing viruses are currently known to infect invertebrates (*Iridovirus* and *Chloriridovirus*) and cold-blooded vertebrates (*Ranavirus* and *Lymphocystivirus*) (36). Major characteristic features of all iridoviruses are the large icosahedral viral particles (120 to 300 nm) present in the cytoplasm. Generally, isolates from fish tend to be larger (200 to 300 nm) in size than both amphibian and invertebrate viruses (120 to 200 nm). To date, genome sequences of five iridovirus genomes have been published: *Lymphocystis disease virus* (LCDV) (genus *Lymphocystivirus*) (33), *Chilo iridescent virus* (CIV) (genus *Iridovirus*) (18), *Tiger frog virus* (TFV) (genus *Ranavirus*) (13), *Infectious spleen and kidney necrosis virus* (ISKNV) (genus unassigned) (14), and *Ambystoma tigrinum virus* (ATV) (genus *Ranavirus*) (19).

Iridovirus pathogens have been regarded as a cause of serious systemic diseases among feral, cultured, and ornamental fish in the recent years. Mortalities of fish due to systemic iridovirus infection reaching 30 to 100% were observed. Histopathological signs in iridovirus-infected fish may include enlargement of cells and necrosis of the renal and splenic hematopoietic tissues (28). In 1994, a novel viral disease called sleepy grouper disease (SGD) resulted in significant economic losses in Singapore marine net cage farms. Finally, this novel iridovirus of the genus *Ranavirus*, designated *Singapore grouper iridovirus* (SGIV), was successfully isolated in 1998 from brown-spotted grouper (6, 29). Further, it was successfully grown in an alternate grouper embryonated egg (*Epinephelus tauvina*) cell line, with good resultant titers (9) and was used as a source to purify SGIV. The physiochemical properties of SGIV have been reported previously (28). At the molecular level, only a partial sequence encoding the highly conserved major capsid protein in SGIV has been reported (28). Due to its relevance in the aquaculture industry, it is important to study the molecular mechanism of viral infection and virus-host interaction in grouper. As an initial part of these studies, we have determined the complete genomic sequence of SGIV. We have also confirmed the authenticity of some open reading frames (ORFs) using the proteomic approach and reverse transcription-PCR (RT-PCR).

## MATERIALS AND METHODS

**Virus infection, purification, and genomic DNA extraction.** Grouper embryonic cells from the brown-spotted grouper *Epinephelus tauvina* (5) were cultured in Eagle's minimum essential medium containing 10% fetal bovine serum, 0.116 M NaCl, 100 IU of penicillin G/ml and 100 μl of streptomycin sulfate/ml. Culture media were equilibrated with HEPES to the final concentration of 5 mM and adjusted to pH 7.4 with $NaHCO_3$. Virus was inoculated onto confluent monolayers of the grouper cell line at a multiplicity of infection of approximately 0.1. When the cytopathic effect was sufficient, the medium containing SGIV was harvested and centrifuged at 12,000 × *g* for 30 min at 4°C. The pellet comprising the virus was resuspended with the culture medium and ultrasonicated. The suspension containing the lysate, virus, and cellular debris was then centrifuged at 4,000 × *g* for 20 min at 4°C. The supernatant was layered onto a cushion of 35% sucrose and centrifuged at 210,000 × *g* for 1 h at 4°C. The pellet, resuspended with the TN buffer (50 mM Tris-HCl [pH 7.4], 150 mM NaCl), was

* Corresponding author. Mailing address: Department of Biological Sciences, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260, Singapore. Phone: 65-68742692. Fax: 65-67795671. E-mail: dbshewcl@nus.edu.sg or dbshead@nus.edu.sg.

overlaid with 30, 40, 50, and 60% (m/v) sucrose gradients and centrifuged at 210,000 × *g* for 1 h at 4°C. Virus bands, present in 50% sucrose, were aspirated, sonicated briefly, and reloaded onto sucrose gradients. The lowest band (50% sucrose) was individually aspirated and spun down at 100,000 × *g*. The purity of virus was examined by negative staining under transmission electron microscopy (JEOL 100 CXII) and was shown to be sufficiently pure for isolation of the genomic DNA, construction of shotgun and restriction libraries, and proteomic analysis. The genomic DNA of the SGIV was treated with protease K and *N*-lauroylsarcosine, followed by phenol-chloroform extraction and alcohol precipitation (16).

**Construction of libraries.** Soluble genomic DNA was quantified by spectrophotometry (UV-1600; Shimadzu). Sixty micrograms of genomic DNA was diluted with TM buffer (5 mM Tris-HCl [pH 8.0], 1.5 mM $MgCl_2$) to a final volume of 200 μl and ultrasonicated (3-s bursts) using an ultrasonic liquid processor (model XL2020; Misonix Inc., Farmingdale, N.Y.). The appropriate viral DNA fragments (500 to 800 bp) were excised from the 1.0% agarose gel and extracted using the QIAquick gel extraction kit (QIAGEN). Genomic DNA fragments were end repaired with T4 DNA polymerase, followed by phosphorylation with T4 polynucleotide kinase. DNA fragments were purified using a High Pure PCR product purification kit (Roche) before the next enzymatic reaction. Sonicated fragments were ligated by incubation at 16°C overnight to the pUC19 vector, which had been prelinearized by SmaI followed by dephosphorylation. After purification, chimerical plasmids were transformed into electrocompetent-cell DH5α. More than 1,000 recombinants were selected from the library by the blue/white screening assay. To construct the restriction library, DNA fragments were obtained by restriction digestion with BamHI and cloned into the corresponding site of pBluescriptII KS(+) vector. Both libraries were used to scaffold the SGIV genome.

**Assembly and analysis of SGIV genome.** Sequencing of the viral fragments was carried out following the standard protocol supplied by Applied Biosystems. All cycle sequencing products were loaded onto the ABI PRISM 3100 genetic analyzer to acquire nucleotide sequences from both directions. Before the scaffolds were created, high-throughput BLAST analysis was performed for all nucleotide sequences to eliminate contamination reads, followed by vector screening with the InterPhace program (University of Washington). A software package, Vector NTI Suite 7.1 (InforMax Inc., Frederick, Mass.), was applied to create the contigs, assemble the genome, identify ORFs, analyze presumptive genes, and draw the genomic map. The whole genome was also submitted to http://www.softberry.com (Softberry Inc., Mount Kisco, N.Y.) for identification of all potential ORFs. These ORFs were searched against the mirror site of National Center for Biotechnology Information (NCBI) nucleotide database at the Singapore Bioinformatics Institute. The presumptive genes were submitted to the NCBI network service to search for conserved domains. Protein motifs were analyzed by using the PROSITE database, release 18.17 (8). Signal peptides and signal anchors were predicted with SignalP V2.0 (24, 25). Signal anchors exist in certain membrane proteins (type II membrane proteins) attaching to the membrane by an N-terminal sequence which shares many characteristics with a signal peptide sequence but is not cleaved. Transmembrane domains were predicted with TMpred (15).

**Mass spectrometric analysis of SGIV proteins.** The protein pellet of the lower band from sucrose gradient ultracentrifugation was separated by one-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Thirty-nine well-separated protein bands were excised, reduced, alkylated, and digested with trypsin (31). To extract the peptides, the gel particles were twice treated with 20 mM $NH_4HCO_3$ and 5% formic acid in 50% acetonitrile, respectively. All supernatants were combined and dried in a vacuum centrifuge. Dried peptides were dissolved in 3 to 20 μl of 0.1% trifluoroacetic acid in 50% acetonitrile. Dissolved peptides (0.5 μl) were spotted onto a target plate, followed by an equal volume of 10-mg/ml α-cyano-4-hydroxycinnamic acid in 50% acetonitrile–0.1% trifluoroacetic acid. After the spots had dried, the target plate was loaded into a Voyager-DE STR BioSpectrometry workstation mass spectrometer (PerSeptive Biosystems, Framingham, Mass.). Mass spectra were acquired with 20.5 kV, 73.5% of grid, and a delayed time of 380 ns under a positive-ion reflector mode. The resulting peptide mass fingerprints were searched against the SGIV ORF database using the AutoMS-Fit search program (version 1.2.18; PerSeptive Biosystem).

**RT-PCR.** Total RNA was extracted from viral cultures at different infective stages using an RNeasy Mini kit (QIAGEN). After the treatment of the total RNA with the RNase-free DNase I (QIAGEN), gene-specific primers were used to amplify the target genes by using the OneStep RT-PCR kit (QIAGEN). All the steps were followed according to the manufacturer's manual. Briefly, cDNA was reverse transcribed at 50°C for 30 min. The PCR amplification segment was started with an initial heating step at 95°C for 15 min (in order to simultaneously deactivate omniscript and sensiscript reverse transcriptases). After the activation of the HotStarTaq DNA polymerase, PCR amplification reactions were performed for 30 cycles under conditions of 95°C for 30 s, 51 to 58°C for 15 s, and 72°C for 1 min per cycle. The annealing temperature was optimized for different target genes. RT-PCR products were analyzed with 1% agarose gel and also subjected to nucleotide sequencing.

**Virus abbreviations.** ALIV, African lampeye iridovirus; ATV, *Ambystoma tigrinum* virus; BIV, Bohle iridovirus; BVDV, bovine viral diarrhea virus; CIV, *Chilo* iridescent virus; CV, chlorella virus; CZIV, *Costelytra zealandica* iridescent virus; EHDV, epizootic hemorrhagic disease virus; EHNV, epizootic hematopoietic necrosis virus; EHV-1, equine herpesvirus; FPV, fowlpox virus; FV3, frog virus 3; GIV, grouper iridovirus; GSIV, giant seaperch iridovirus; HVAV, *Heliotis virescens* ascovirus; IMRV, *Ictalurus melas* ranavirus; ISKNV, infectious spleen and kidney necrosis virus; LBIV, largemouth bass iridovirus; LCDV-1, lymphocystis disease virus 1; LYCIV, large yellow croaker iridovirus; MSEPV, *Melanoplus sanguinipes* entomopoxvirus; OMRV, *Oncorhynchus mykiss* ranavirus; PBCV, *Paramecium bursaria* chlorella virus; RGV, *Rana grylio* virus; RRV, *Regina* ranavirus; RSBI, Red Sea bream iridovirus; SBIV, sea bass iridovirus; SCV, *Siniperca chuatsi* virus; SFAV, *Spodoptera frugiperda* ascovirus; SGIV, Singapore grouper iridovirus; SIV, *Simulium* iridescent virus; SOV, *Sciaenops ocellatus* virus; TFV, tiger frog virus; TIV, *Tipula* iridescent virus; WIV, *Wiseana* iridescent virus.

**Nucleotide sequence accession number.** The complete SGIV genome sequence has been deposited in GenBank under accession no. AY521625. Accession numbers of 162 annotated ORFs are from AAS18016 to AAS18177, consecutively.

## RESULTS AND DISCUSSION

**Determination of the SGIV genome sequence.** We set out to generate 8× to 9× genome coverage of the SGIV genome. The bulk of the sequence coverage (2,065 passing reads) resulted from the shotgun library. However, 214 passing reads from the restriction library provided important intermediate-range linking information for assembly. Thirteen contigs ranging from 28,106 to 651 bp were scaffolded with the Contig Express program (CEP) of the Vector NTI suite 7.1. Final gaps were directly sequenced off the genomic DNA with custom synthetic primers and closed by 50 passing reads. In total, 2,329 cycle sequencing reaction products (free of contamination reads) from both random shotgun and restriction libraries were used to assemble the SGIV genome. Most of the genome (98.4%) was compiled by sequencing at least three times. Only 1.6% of the genome was assembled from a single recombinant. One hundred percent of the genome sequence was constructed from sequencing in both directions. Like other iridoviruses, SGIV was made up of a double-stranded DNA which is circularly permuted (30, 11). The whole SGIV genome consists of 140,131 bp with a G+C content of 48.64% (Fig. 1), which is slightly less than that of TFV (55.01%), ISKNV (54.78%), and ATV (54.02%) but substantially more than that of LCDV-1 (29.07%) and CIV (28.63%).

**Coding capacity of the viral genomic DNA sequence.** Prediction of presumptive genes was carried out by using the viral gene prediction program under the website http://www.softberry.com supplemented with Vector NTI suite 7.1. One hundred sixty-two presumptive ORFs were identified to code for proteins ranging from 41 to 1,268 amino acids on the sense (R) and antisense (L) DNA strands (Table 1). Computer-assisted analyses of the deduced amino acid sequences revealed that 23 of the ORFs share high levels of identity to iridovirus proteins which have been described previously to have specific biological functions. Fifty-one ORFs are homologous to other iridovirus genes, for which the corresponding proteins and their

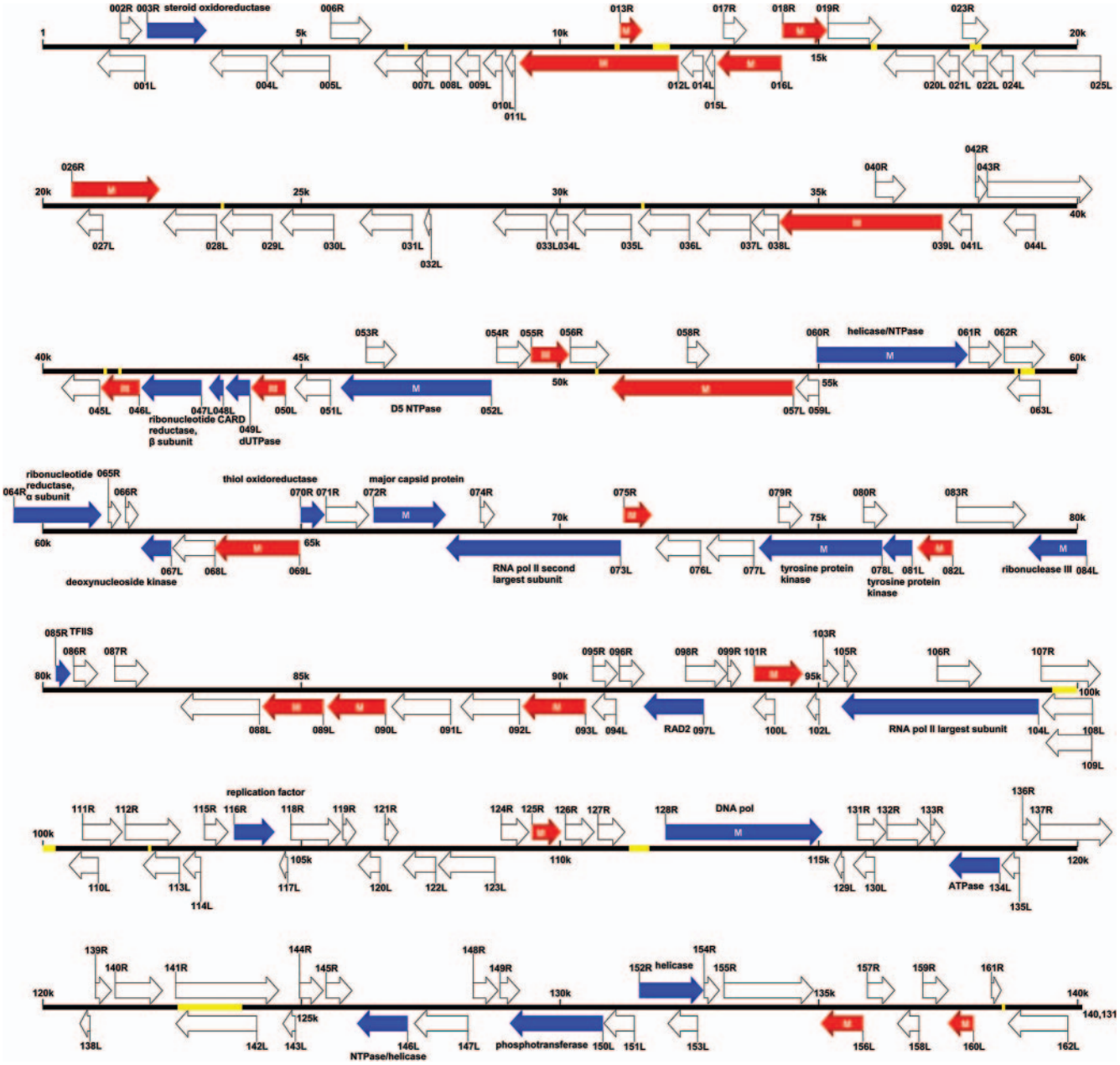


FIG. 1. Organization of the SGIV genome. The SGIV genome is shown in a linear format. A total of 162 ORFs, predicted by the FGENESV program (available through: http://www.softberry.com), supplemented with Vector NTI suite 7.1, are indicated by their locations, orientations, and putative sizes. Blue arrows represent ORFs with known function, while red arrows represent ORFs detected by RT-PCR. "M" represents an ORF whose expressed product was identified by MALDI-TOF mass spectrometry. Yellow lines represent repetitive sequence regions. The scale is in 5 kbp.

respective functions remain unknown. Additionally, three ORFs show weak homologies to genes of other viruses. Forty-two conserved domains, motifs, or signatures are identified from the NCBI CD-Search database and the PROSITE database (Table 1). A number of genes of SGIV are shown to be present in the ATV, TFV, LCDV, CIV, and ISKNV genomes. These include genes for the DNA polymerase, the DNA repair protein, the two largest subunits of DNA-dependent RNA polymerase II, the TFIIS, RNase III, ATPase, etc. (Table 1). There is no evidence of introns, and both strands are shown to contain ORFs. Three pairs of ORFs partially overlap other ORFs.

**Repetitive regions.** The analysis of the genome showed the presence of 17 repetitive regions distributed throughout the genome. In total. these occupy 2.6% of the SGIV genome, varying in size from 31 to 1,119 bp. These regions encompass eight perfect and nine imperfect repetitive sequences whose match percentages range from 80 to 99% (Table 2). No homologies between those repeats were detected. The base composition of 12 repeats is found to be more than 65% G+C. The

TABLE 1. Listing of potential expressed ORFs in SGIV

| ORF | Nucleotide position (length [aa[a]]) | MW[b] (kDa) | pI[c] | Conserved domain or signature[d] (CD/Prosite accession no.) | Match | | | | Predicted structure and/or function[g] |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | BlastP score | % Identity | Accession no.[e] | Species[f] | |
| ORF001L | 1971–1057 (304) | 34.65 | 9.19 | | | | | | TM |
| ORF002R | 1502–1903 (133) | 14.20 | 8.86 | | | | | | |
| ORF003R | 2018–3163 (381) | 43.16 | 8.99 | 3-Beta hydroxysteroid dehydrogenase (pfam01073) | 386 | 53 | AAL77802 | TFV054L | 3-beta-hydroxy-delta 5-C27-steroid oxidoreductase-like protein, TM |
| ORF004L | 4332–3235 (365) | 41.67 | 8.26 | | 45 | 23 | AAP33232 | ATV053R | |
| ORF005L | 5542–4400 (380) | 40.40 | 8.72 | Transmembrane amino acid transporter protein (pfam01490) C-type lectin domain signature (PS00615) | 76.3 | 23 | AAF61849 | *Mus musculus* | N system amino acids transporter NAT-1, TM, SA |
| ORF006R | 5570–6349 (259) | 29.15 | 5.25 | | 324 | 67 | AAP33234 | ATV055R | Early 31-kDa protein, TM |
| | | | | | 318 | 69 | CAA07475 | EHNV | |
| | | | | | 318 | 68 | AAL77797 | TFV025R | |
| | | | | | 261 | 67 | CAA37177 | FV3 | |
| | | | | | 115 | 32 | NP_078713 | LCDV122R | |
| | | | | | 39.3 | 22 | AAL98842 | ISKNV118L | |
| ORF007L | 7339–6416 (307) | 30.12 | 6.96 | | | | | | TM |
| ORF008L | 7886–7194 (230) | 22.15 | 9.52 | | | | | | |
| ORF009L | 8444–7980 (154) | 16.43 | 6.14 | | | | | | TM |
| ORF010L | 8888–8517 (123) | 13.61 | 5.13 | | | | | | |
| ORF011L | 9132–8944 (62) | 6.63 | 8.44 | | | | | | TM, SP |
| ORF012L | 12293–9219 (1024) | 117.36 | 9.46 | | 64 | 35 | AAP33240 | ATV061R | |
| ORF013R | 11173–11595 (140) | 14.49 | 6.08 | | | | | | TM |
| ORF014L | 12773–12348 (141) | 15.85 | 7.98 | | 79 | 35 | AAP33239 | ATV060R | |
| ORF015L | 13000–12821 (59) | 6.57 | 9.11 | | 33 | 51 | AAP33181 | ATV004R | TM, SA |
| ORF016L | 14289–13048 (413) | 46.26 | 5.08 | | 325 | 40 | AAP33180 | ATV003R | TM |
| | | | | | 70.5 | 26 | AAK82090 | CIV229L | |
| ORF017R | 13172–13609 (145) | 16.07 | 7.04 | | | | | | TM |
| ORF018R | 14317–15174 (285) | 32.32 | 6.02 | | 178 | 38 | AAP33179 | ATV002L | |
| | | | | | 61.2 | 34 | NP_078687 | LCDV093R | |
| ORF019R | 15196–16224 (342) | 36.82 | 7.44 | Poxvirus proteins of unknown function (pfam03003) | 381 | 70 | AAP33178 | ATV001L | TM |
| | | | | Glycoprotein hormones beta chain signature (PS00261) | 164 | 37 | NP_078745 | LCDV160L | |
| | | | | | 145 | 33 | AAK82199 | CIV337L | |
| ORF020L | 17246–16278 (322) | 35.37 | 8.66 | | 63.20 | 40 | AAP33250 | ATV070L | |
| ORF021L | 17725–17306 (139) | 16.42 | 6.19 | | 161 | 59 | AAP33251 | ATV071L | |
| | | | | | 149 | 59 | AAK54494 | RRV | |
| | | | | | 62.8 | 35 | NP_078640 | LCDV036R | |
| ORF022L | 18277–17777 (166) | 18.57 | 12.55 | | | | | | |
| ORF023R | 17793–18290 (165) | 18.22 | 5.43 | | | | | | TM |
| ORF024L | 18774–18319 (151) | 17.15 | 5.75 | | 45 | 27 | AAP33254 | ATV073L | |
| ORF025L | 20488–18956 (510) | 56.49 | 7.31 | SAP, putative DNA-binding (bihelical) motif (smart00513) | 161 | 53 | AAP33256 | ATV075L | |
| | | | | | 153 | 48 | AAK54496 | RRV | |
| | | | | | 100 | 36 | NP_078703 | LCDV110L | |
| ORF026R | 20567–22267 (566) | 63.32 | 6.40 | | 375 | 37 | AAP33257 | ATV076R | TM |
| | | | | | 78.2 | 21 | NP_078649 | LCDV048R | |
| ORF027L | 21162–20671 (163) | 17.08 | 8.94 | | | | | | TM |
| ORF028L | 23363–22350 (337) | 36.67 | 7.99 | Ig-like domain (PS50385) | | | | | TM, SP |
| ORF029L | 24445–23447 (332) | 36.58 | 6.42 | Neural cell adhesion molecule L1 (KOG3513) Ig-like domain (PS50385) | | | | | TM, SP |
| ORF030L | 25635–24610 (341) | 37.90 | 9.07 | | 33 | 37 | NP_041033 | EHV 1 | Tegument protein, TM, SP |
| ORF031L | 27160–26144 (338) | 37.62 | 8.46 | Ig-like domain (PS50385) | | | | | TM, SP |
| ORF032L | 27516–27391 (41) | 4.64 | 9.82 | | | | | | TM, SP |
| ORF033L | 29760–28726 (344) | 37.56 | 8.75 | Ig-like domain (PS50385) | | | | | TM, SP |
| ORF034L | 30161–29823 (112) | 12.66 | 8.90 | | | | | | TM, SP |
| ORF035L | 31388–30261 (375) | 42.26 | 8.99 | Neural cell adhesion molecule L1 (KOG3513) Ig-like domain (PS50385) | 48.1 | 25 | BAC11344 | *Homo sapiens* | Unnamed protein product, TM, SA |
| ORF036L | 32515–31526 (329) | 37.29 | 4.95 | | | | | | TM, SP |
| ORF037L | 33696–32668 (342) | 37.12 | 8.77 | Neural cell adhesion molecule L1 (KOG3513) | 43.1 | 27 | CAA40912 | *Drosophila melanogaster* | Fibroblast growth factor receptor, TM, SP |
| ORF038L | 34236–33724 (170) | 19.04 | 9.72 | | 121 | 64 | AAP33260 | ATV079L | TM, SA |
| | | | | | 69.3 | 53 | NP_078769 | LCDV194R | |
| | | | | | 36.2 | 41 | AAB94443 | CIV117L | |

*Continued on following page*

TABLE 1—*Continued*

| ORF | Nucleotide position (length [aa[a]]) | MW[b] (kDa) | pI[c] | Conserved domain or signature[d] (CD/Prosite accession no.) | Match | | | | Predicted structure and/or function[g] |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | BlastP score | % Identity | Accession no.[e] | Species[f] | |
| ORF039L | 37417–34262 (1051) | 118.22 | 8.38 | Protein kinase domain (PS50011) | 425 | 36 | AAP33261 | ATV080L | TM |
| | | | | | 176 | 45 | NP_078619 | LCDV010L | |
| | | | | | 150 | 28 | NP_078677 | LCDV080R | |
| | | | | | 96.3 | 24 | AAK82240 | CIV380R | |
| | | | | | 65.1 | 23 | AAL98779 | ISKNV055L | |
| ORF040R | 36123–36698 (191) | 20.51 | 9.56 | | | | | | TM, SP |
| ORF041L | 37978–37547 (143) | 15.24 | 6.05 | | | | | | |
| ORF042R | 38058–38285 (75) | 8.47 | 6.23 | | | | | | TM, SP |
| ORF043R | 38285–40288 (667) | 73.66 | 5.36 | | 179 | 38 | AAP33262 | ATV081R | |
| ORF044L | 39213–38608 (201) | 21.71 | 9.65 | | | | | | TM |
| ORF045L | 41090–40362 (242) | 22.94 | 9.80 | | | | | | |
| ORF046L | 41866–41120 (248) | 23.77 | 9.88 | | | | | | |
| ORF047L | 43063–41909 (384) | 43.58 | 5.69 | Ribonucleotide reductase, beta subunit (COG0208) | 580 | 71 | AAP33216 | ATV038R | Ribonucleoside-diphosphate reductase beta subunit, TM |
| | | | | | 567 | 70 | AAL77807 | TFV071L | |
| | | | | | 331 | 48 | NP_078636 | LCDV027R | |
| ORF048L | 43489–43214 (91) | 10.50 | 8.73 | Caspase recruitment domain (pfam00619) | 57 | 42 | AAP33218 | ATV040L | CARD-like caspase |
| ORF049L | 44002–43535 (155) | 17.06 | 6.58 | dUTPase (KOG3370) | 123 | 46 | AAL77806 | TFV068R | dUTPase |
| | | | | | 120 | 47 | AAP33220 | ATV042L | |
| | | | | | 121 | 40 | AAK82298 | CIV438L | |
| ORF050L | 44695–44033 (220) | 23.55 | 8.68 | Tumor necrosis factor receptor domain (cd00185) | 82 | 32 | P25119 | *Mus musculus* | Tumor necrosis factor receptor superfamily member 1B precursor, TM, SP |
| ORF051L | 45563–44868 (231) | 26.12 | 7.43 | Tumor necrosis factor receptor domain (cd00185) | 72 | 30 | AAO89081 | *Mus musculus* | Tumor necrosis factor receptor superfamily member 14 precursor, TM, SP |
| ORF052L | 48673–45767 (968) | 109.88 | 7.71 | Predicted ATPase (COG3378) | 1414 | 69 | AAP33258 | ATV077L | D5 family NTPase, TM |
| | | | | Poxvirus D5 protein (pfam03288) | 607 | 35 | NP_078717 | LCDV128L | |
| | | | | | 192 | 39 | AAB94479 | CIV184R | |
| | | | | | 365 | 46 | AAL98833 | ISKNV109L | |
| ORF053R | 46254–46832 (192) | 20.35 | 9.99 | Lipoprotein lipid attachment site (PS00013) | | | | | TM |
| ORF054R | 48777–49424 (215) | 25.13 | 4.98 | | 265 | 68 | AAP33259 | ATV078R | TM |
| | | | | | 101 | 34 | AAL98780 | ISKNV056L | |
| | | | | | 81.3 | 41 | NP_078618 | LCDV006L | |
| | | | | | 80.5 | 27 | AAB94419 | CIV067R | |
| ORF055R | 49447–50169 (240) | 22.77 | 10.65 | Collagens (type XV) (KOG3546) | 80.9 | 33 | ZP_00122019 | *Haemophilus somnus* 129PT | Hypothetical protein |
| ORF056R | 50198–50938 (246) | 22.93 | 6.48 | | | | | | |
| ORF057L | 54510–51004 (1168) | 131.30 | 8.83 | | 1176 | 52 | AAP33249 | ATV069R | TM |
| | | | | | 1165 | 52 | AAK37740 | RRV | |
| | | | | | 340 | 26 | NP_078748 | LCDV163R | |
| | | | | | 134 | 23 | AAL98800 | ISKNV076L | |
| | | | | | 80.1 | 31 | AAK82156 | CIV295L | |
| ORF058R | 52463–52876 (137) | 13.75 | 9.22 | | | | | | |
| ORF059L | 55000–54560 (146) | 16.32 | 9.10 | | 45 | 25 | AAP33185 | ATV008R | TM, SP |
| ORF060R | 54967–57879 (970) | 109.65 | 8.98 | DNA repair protein, SNF2 family (KOG0390) | 1214 | 61 | AAL77795 | TFV009L | Putative NTPase 1, TM |
| | | | | Helicase conserved C-terminal domain (pfam00271) | 1212 | 61 | AAP33184 | ATV007L | |
| | | | | | 940 | 58 | AAK53744 | RRV | |
| | | | | | 749 | 42 | NP_078720 | LCDV132L | |
| | | | | | 414 | 32 | AAL98787 | ISKNV063L | |
| | | | | | 81.3 | 30 | AAD48148 | CIV022L | |
| ORF061R | 57914–58528 (204) | 23.26 | 9.54 | Catalytic domain of CTD phosphatases (smart00577) | 187 | 50 | AAP33244 | ATV064R | |
| | | | | TFIIF-interacting CTD phosphatase (KOG1605) | 91.7 | 34 | AAL98729 | ISKNV005L | |
| | | | | | 76.6 | 33 | NP_078678 | LCDV082L | |
| | | | | | 53.5 | 27 | AAK82216 | CIV355R | |
| ORF062R | 58593–59363 (256) | 27.69 | 5.18 | Insulin-like growth factor (smart00078) | 91.3 | 38 | BAC67672 | *Cyanidioschyzon merolae* | DNA-directed RNA polymerase II largest subunit, SP |
| | | | | Predicted DNA-binding protein (KOG2588) | | | | | |

*Continued on facing page*

TABLE 1—*Continued*

| ORF | Nucleotide position (length [aa[a]]) | MW[b] (kDa) | pI[c] | Conserved domain or signature[d] (CD/Prosite accession no.) | Match | | | | Predicted structure and/or function[g] |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | BlastP score | % Identity | Accession no.[e] | Species[f] | |
| ORF063L | 59278–58649 (209) | 22.16 | 3.88 | Transcription elongation factor (COG5164) | 56.2 | 47 | XP_220553 | *Rattus norvegicus* | Similar to charcot-marie-tooth duplicated region transcript I, TM |
| ORF064R | 59415–61133 (572) | 63.72 | 8.15 | Ribonucleotide reductase, alpha subunit (COG0209) | 822 | 68 | AAL77800 | TFV041R | Ribonucleoside-diphosphate reductase, alpha subunit-like protein, TM |
| | | | | | 827 | 68 | AAP33245 | ATV065R | |
| | | | | | 636 | 54 | NP_078756 | LCDV176L | |
| ORF065R | 61268–61510 (80) | 9.35 | 8.85 | | | | | | |
| ORF066R | 61603–61845 (80) | 8.70 | 9.30 | | | | | | TM, SA |
| ORF067L | 62482–61907 (191) | 21.58 | 6.49 | Mitochondrial thymidine kinase 2 (KOG4235) | 158 | 43 | AAP33196 | ATV019L | Deoxynucleoside kinase, TM |
| | | | | Deoxynucleoside kinase (pfam01712) | 96.7 | 30 | NP_078725 | LCDV136R | |
| | | | | | 56.2 | 26 | AAF44495 | FPV | |
| | | | | | 53.1 | 25 | CAC84464 | SFAV-1 | |
| | | | | | 48.9 | 24 | CAC84481 | HVAV-3c | |
| ORF068L | 63334–62516 (272) | 29.61 | 8.59 | Immunoglobulins and major histocompatibility complex proteins signature (PS00290) | 209 | 50 | AAP33197 | ATV020L | TM |
| | | | | | 100 | 27 | AAL98836 | ISKNV112R | |
| | | | | | 100 | 27 | NP_078615 | LCDV003L | |
| | | | | | 40.4 | 32 | AAK82296 | CIV436L | |
| ORF069L | 64967–63321 (548) | 61.88 | 9.62 | | 211 | 36 | AAP33194 | ATV017R | |
| | | | | | 69.7 | 21 | NP_078643 | LCDV039R | |
| ORF070R | 64994–65452 (152) | 17.03 | 9.32 | Erv1/Alt family (pfam04777) | 143 | 47 | AAP33193 | ATV016L | Thiol oxidoreductase, TM |
| | | | | | 69.3 | 32 | NP_078699 | LCDV106L | |
| | | | | | 65.5 | 37 | AAL98767 | ISKNV043L | |
| | | | | | 38.9 | 26 | AAK82208 | CIV347L | |
| ORF071R | 65483–66307 (274) | 31.61 | 8.58 | | 50 | 36 | AAP33192 | ATV015L | |
| ORF072R | 66404–67795 (463) | 50.53 | 6.33 | Iridovirus major capsid protein (pfam04451) | 895 | 99 | AAM00286 | GIV | Major capsid protein, TM |
| | | | | | 681 | 71 | AAK55105 | TFV096R | |
| | | | | | 676 | 70 | AAO32315 | EHNV | |
| | | | | | 675 | 70 | AAP33191 | ATV014L | |
| | | | | | 670 | 71 | AAB01722 | FRG3V | |
| | | | | | 481 | 50 | AAC24486.2 | LCDV147L | |
| | | | | | 417 | 45 | AAK82135 | CIV274L | |
| | | | | | 374 | 43 | AAL72276 | ISKNV006L | |
| ORF073L | 71185–67874 (1103) | 123.39 | 8.28 | RNA polymerase II, second largest subunit (KOG0214) | 1530 | 65 | AAL77805 | TFV065L | DNA-directed RNA polymerase II second-largest subunit, TM |
| | | | | | 1533 | 65 | AAP33221 | ATV043R | |
| | | | | | 1295 | 66 | AAK84400 | RRV | |
| | | | | | 1018 | 46 | NP_078633 | LCDV025L | |
| | | | | | 755 | 41 | AAL98758 | ISKNV034R | |
| ORF074R | 68472–68738 (88) | 9.16 | 8.52 | | 71 | 73 | AAP33222 | ATV043bL | |
| | | | | | 51 | 47 | NP_149893 | CIV430R | |
| ORF075R | 71239–71775 (178) | 19.96 | 4.50 | | 44.3 | 37 | NP_078763 | LCDV185R | |
| ORF076L | 72715–71858 (285) | 30.35 | 7.17 | Purine nucleoside phosphorylase (KOG3984) | 305 | 50 | NP_000261 | *Homo sapiens* | Purine nucleoside phosphorylase, TM |
| ORF077L | 73747–72839 (302) | 34.12 | 9.05 | | | | | | TM |
| ORF078L | 76227–73855 (790) | 88.16 | 8.75 | Putative lipopolysaccharide-modifying enzyme (smart00672) | 552 | 43 | AAL77799 | TFV029R | Putative tyrosine protein kinase, TM |
| | | | | | 553 | 43 | AAP33237 | ATV058R | |
| | | | | | 231 | 33 | NP_078770 | LCDV195R | |
| | | | | | 177 | 40 | AAB94478 | CIV179R | |
| ORF079R | 74231–74680 (149) | 15.13 | 7.78 | | | | | | TM, SP |
| ORF080R | 75872–76324 (150) | 15.93 | 9.10 | | | | | | TM |
| ORF081L | 76809–76246 (187) | 21.79 | 9.05 | | 198 | 49 | AAL77799 | TFV029R | Putative tyrosine protein kinase |
| | | | | | 191 | 47 | AAP33237 | ATV058R | |
| | | | | | 143 | 50 | AAK54490 | RRV | |
| | | | | | 86.3 | 32 | NP_078770 | LCDV195R | |
| | | | | | 70.9 | 28 | AAB94478 | CIV179R | |
| ORF082L | 77592–76924 (222) | 24.32 | 9.02 | | | | | | |
| ORF083R | 77672–79009 (445) | 50.48 | 9.26 | | 308 | 42 | AAP33203 | ATV026L | TM |

*Continued on following page*

TABLE 1—*Continued*

| ORF | Nucleotide position (length [aa[a]]) | MW[b] (kDa) | pI[c] | Conserved domain or signature[d] (CD/Prosite accession no.) | Match | | | | Predicted structure and/or function[g] |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | BlastP score | % Identity | Accession no.[e] | Species[f] | |
| ORF084L | 80193–79066 (375) | 41.64 | 9.17 | dsRNA-specific ribonuclease (COG0571) | 294 | 45 | AAL77809 | TFV085L | Ribonuclease III, TM |
| | | | | Ribonuclease III family (smart00535) | 287 | 45 | AAP33202 | ATV025R | |
| | | | | | 215 | 45 | NP_078726 | LCDV137R | |
| | | | | | 120 | 35 | AAB94459 | CIV142R | |
| | | | | | 108 | 37 | AAL98811 | ISKNV087R | |
| ORF085R | 80251–80529 (92) | 10.57 | 8.27 | Transcription elongation factor TFIIS (COG1594) | 98.2 | 55 | AAL77810 | TFV086R | Transcription elongation factor SII |
| | | | | | 95.9 | 55 | AAP33201 | ATV024L | |
| | | | | | 57 | 60 | BAA04187 | CV | |
| | | | | | 57 | 60 | AAC96492 | PBCV-1 | |
| | | | | | 56.6 | 42 | NP_078754 | LCDV171R | |
| ORF086R | 80591–81055 (154) | 17.11 | 8.08 | | 108 | 40 | AAA43825 | FV3 | Putative immediate-early protein, TM |
| | | | | | 100 | 35 | AAB47251 | IMRV | |
| | | | | | 100 | 35 | AAP33200 | ATV023L | |
| | | | | | 99.8 | 38 | AAL77811 | TFV087R | |
| | | | | | 99.4 | 37 | AAB47252 | OMRV | |
| ORF087R | 81385–82032 (215) | 25.22 | 6.64 | | | | | | SP |
| ORF088L | 84187–82667 (506) | 54.01 | 4.89 | | 546 | 55 | AAK54492 | RRV | TM |
| | | | | | 542 | 55 | AAP33230 | ATV051L | |
| | | | | | 176 | 31 | AAL98731 | ISKNV007L | |
| | | | | | 176 | 30 | BAC66967 | RSBI | |
| | | | | | 172 | 29 | NP_078665 | LCDV067L | |
| | | | | | 108 | 29 | AAB94444 | CIV118L | |
| | | | | | 82.4 | 26 | CAC19148 | Ascovirus DpA V4 | |
| | | | | | 59.3 | 21 | AAK82318 | CIV458R | |
| ORF089L | 85420–84248 (390) | 45.59 | 8.17 | | 49.3 | 23 | AAP33232 | ATV053R | |
| ORF090L | 86627–85506 (373) | 43.48 | 7.74 | | 56 | 22 | AAP33232 | ATV053R | |
| ORF091L | 87886–86750 (378) | 44.54 | 6.93 | | | | | | |
| ORF092L | 89216–88086 (376) | 44.12 | 7.18 | | 33.5 | 19 | AAP33232 | ATV053R | |
| ORF093L | 90497–89280 (405) | 47.59 | 7.30 | | 43.9 | 21 | AAP33232 | ATV053R | |
| ORF094L | 91083–90622 (153) | 16.24 | 7.92 | | | | | | TM |
| ORF095R | 90635–91111 (158) | 16.61 | 5.12 | | | | | | TM, SP |
| ORF096R | 91148–91618 (156) | 16.88 | 7.53 | Tumor necrosis factor receptor domain (cd00185) | 45.4 | 35 | AAB53707 | *Ranus norvegicus* | Tumor necrosis factor receptor superfamily, member 11b Osteoprotegerin, TM, SP |
| ORF097L | 92774–91626 (382) | 43.18 | 9.03 | Xeroderma pigmentosum G, N, and I regions (cd00128) | 400 | 51 | AAL77816 | TFV101R | Putative DNA repair protein RAD2, TM |
| | | | | | 394 | 52 | AAP33187 | ATV010L | |
| | | | | | 382 | 54 | AAK53745 | RRV | |
| | | | | | 161 | 33 | NP_078767 | LCDV191R | |
| | | | | | 130 | 28 | AAL98751 | ISKNV027L | |
| | | | | | 128 | 28 | BAA82754 | RSBI | |
| | | | | | 94.7 | 23 | AAK82229 | CIV369L | |
| ORF098R | 92428–93231 (267) | 30.51 | 9.49 | | 191 | 56 | AAP33188 | ATV011R | TM, SP |
| | | | | | 187 | 56 | AAK53746 | RRV | |
| | | | | | 106 | 51 | CAC19143 | Ascovirus DpAV4 | |
| | | | | | 102 | 40 | NP_078627 | LCDV019R | |
| | | | | | 101 | 50 | AAL98810 | ISKNV086L | |
| | | | | | 90.5 | 42 | AAK82168 | CIV307L | |
| ORF099R | 93244–93492 (82) | 9.07 | 5.13 | | | | | | |
| ORF100L | 94153–93740 (137) | 14.89 | 9.13 | | | | | | TM |
| ORF101R | 93753–94694 (313) | 35.03 | 6.13 | | | | | | TM |
| ORF102L | 95007–94774 (77) | 8.54 | 6.93 | Ubiquitin/60s ribosomal protein L40 fusion (KOG0003) Ubiquitin/40s ribosomal protein S27a fusion (KOG0004) | 134 | 81 | AAD44040 | BVDV-2 | Polyprotein |
| ORF103R | 95092–95385 (97) | 11.05 | 5.04 | | 42 | 46 | AAP33269 | ATV088L | TM, SA |
| ORF104L | 99252–95446 (1268) | 139.16 | 8.30 | RNA polymerase II, large subunit (KOG0260) | 1477 | 60 | AAL77794 | TFV008R | DNA-dependent RNA polymerase largest subunit-like protein, TM |
| | | | | | 1476 | 59 | AAP33183 | ATV006R | |
| | | | | | 947 | 42 | AAA92868 | LCDV016L | |
| | | | | | 719 | 37 | BAA82753 | RSBI | |
| | | | | | 709 | 37 | AAL98752 | ISKNV028L | |
| | | | | | 332 | 31 | AAB33907 | CIV176R | |

*Continued on facing page*

TABLE 1—*Continued*

| ORF | Nucleotide position (length [aa[a]]) | MW[b] (kDa) | pI[c] | Conserved domain or signature[d] (CD/Prosite accession no.) | Match | | | | Predicted structure and/or function[g] |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | BlastP score | % Identity | Accession no.[e] | Species[f] | |
| ORF105R | 95498–95731 (77) | 8.11 | 10.02 | | 40 | 45 | AAK82205 | CIV344R | TM |
| ORF106R | 97298–98146 (282) | 29.94 | 10.94 | | | | | | TM |
| ORF107R | 99308–100453 (381) | 39.06 | 3.90 | | | | | | TM, SP |
| ORF108L | 100309–99329 (326) | 33.49 | 11.74 | | | | | | TM |
| ORF109L | 100305–99400 (301) | 27.58 | 3.54 | | | | | | TM, SP |
| ORF110L | 101067–100504 (187) | 21.21 | 6.29 | | | | | | TM |
| ORF111R | 100766–101533 (255) | 29.23 | 5.47 | | 148 | 33 | AAP33270 | ATV089R | |
| | | | | | 56 | 22 | NP_078768 | LCDV193L | |
| ORF112R | 101588–102655 (355) | 34.66 | 6.09 | Collagens (type XV) (KOG3546) | 52.4 | 64 | BAB34267 | *Escherichia coli* O157:H7 | Putative tail fiber protein, TM, SP |
| ORF113L | 102633–101944 (229) | 21.61 | 6.54 | | | | | | TM |
| ORF114L | 103050–102712 (112) | 11.98 | 8.92 | | | | | | TM, SP |
| ORF115R | 103122–103580 (152) | 17.24 | 6.78 | B-Cell lymphoma (smart00337) | 40 | 32 | AAF89533 | *Ovis aries* | Bak protein, TM |
| ORF116R | 103700–104476 (258) | 30.13 | 9.11 | | 277 | 58 | AAP33272 | ATV091R | Putative replication factor |
| | | | | | 144 | 40 | NP_078747 | LCDV162L | |
| | | | | | 99.4 | 33 | AAK82143 | CIV282R | |
| | | | | | 45.8 | 26 | AAL98785 | ISKNV061L | |
| ORF117L | 104733–104575 (52) | 6.28 | 8.46 | | | | | | |
| ORF118R | 104795–105754 (319) | 35.67 | 8.77 | | 275 | 55 | AAP33268 | ATV087R | |
| | | | | | 85 | 27 | NP_078701 | LCDV108L | |
| | | | | | 40.4 | 23 | AAK82148 | CIV287R | |
| | | | | | 39.3 | 29 | AAL98820 | ISKNV096L | |
| ORF119R | 105799–106050 (83) | 9.09 | 9.36 | | 36 | 29 | AAP33205 | ATV028L | |
| ORF120L | 106525–106103 (140) | 15.96 | 8.89 | | 110 | 47 | AAP33204 | ATV027R | |
| | | | | | 108 | 47 | AAK84402 | RRV | |
| | | | | | 55.5 | 32 | NP_078638 | LCDV032R | |
| ORF121R | 106615–106869 (84) | 9.84 | 8.23 | | | | | | TM, SP |
| ORF122L | 107599–106967 (210) | 24.25 | 9.48 | | | | | | |
| ORF123L | 108740–107652 (362) | 41.48 | 8.65 | | 169 | 29 | AAL13097 | RGV9807 | TM |
| | | | | | 169 | 28 | AAP33224 | ATV045R | |
| ORF124R | 108863–109399 (178) | 20.19 | 6.54 | | | | | | TM, SP |
| ORF125R | 109474–110028 (184) | 21.08 | 5.92 | | | | | | TM, SP |
| ORF126R | 110101–110658 (185) | 20.62 | 8.00 | | | | | | TM, SP |
| ORF127R | 110731–111252 (173) | 19.88 | 7.14 | | | | | | TM, SP |
| ORF128R | 112041–115070 (1009) | 114.95 | 8.63 | DNA polymerase family B (pfam00136) | 1374 | 65 | AAL77804 | TFV063R | DNA polymerase, TM |
| | | | | | 1373 | 65 | AAP33223 | ATV044L | |
| | | | | | 644 | 37 | NP_078724 | LCDV135R | |
| | | | | | 540 | 63 | AAK54493 | RRV | |
| | | | | | 531 | 34 | AAL98743 | ISKNV019R | |
| | | | | | 531 | 34 | BAA28669 | RSBI | |
| | | | | | 454 | 34 | CAC84133 | Iridovirus RMIV (IV22) | |
| | | | | | 362 | 30 | CAC19127 | Ascovirus DpAV4 | |
| | | | | | 315 | 31 | AAD48150 | CIV037L | |
| | | | | | 263 | 25 | CAC84471 | HVAV-3c | |
| | | | | | 261 | 26 | CAC19170 | SFAV-1 | |
| | | | | | 258 | 68 | AAK84401 | RRV | |
| ORF129L | 115490–115308 (60) | 6.68 | 4.57 | | | | | | |
| ORF130L | 116083–115673 (136) | 14.89 | 6.91 | TonB-dependent receptor proteins signature (PS00430) | | | | | |
| ORF131R | 115749–116303 (184) | 19.93 | 4.53 | IG-like domain (PS50385) | | | | | TM, SP |
| ORF132R | 116321–117148 (275) | 31.36 | 9.50 | | 198 | 48 | AAP33263 | ATV082R | TM |
| ORF133R | 117168–117440 (90) | 9.56 | 3.57 | | | | | | |
| ORF134L | 118498–117527 (323) | 36.50 | 8.33 | ATPases (smart00382) | 428 | 74 | AAL77796 | TFV016R | ATPase |
| | | | | | 369 | 66 | AAA43823 | FV3 | |
| | | | | | 430 | 74 | AAP33264 | ATV083L | |
| | | | | | 288 | 58 | NP_078656 | LCDV054R | |
| | | | | | 259 | 53 | AAL98847 | ISKNV122R | |
| | | | | | 259 | 53 | AAL68652 | GIV | |
| | | | | | 259 | 53 | BAA28670 | RSBI | |
| | | | | | 259 | 53 | BAA96406 | SBIV | |
| | | | | | 259 | 53 | BAA96407 | GIV | |
| | | | | | 259 | 53 | AAL68653 | GSIV | |
| | | | | | 259 | 53 | AAL68654 | LBIV | |
| | | | | | 259 | 53 | AAO16492 | LYCIV | |
| | | | | | 259 | 53 | BAA96408 | ALIV | |
| | | | | | 217 | 55 | AAL73346 | SCV | |
| | | | | | 216 | 55 | AAN77575 | SOV | |
| | | | | | 201 | 56 | AAM00905 | RSBI | |
| | | | | | 194 | 42 | AAB94422 | CIV075L | |

*Continued on following page*

TABLE 1—*Continued*

| ORF | Nucleotide position (length [aa[a]]) | MW[b] (kDa) | pI[c] | Conserved domain or signature[d] (CD/Prosite accession no.) | Match | | | | Predicted structure and/or function[g] |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | BlastP score | % Identity | Accession no.[e] | Species[f] | |
| ORF135L | 118885–118547 (112) | 12.95 | 5.98 | | 85 | 43 | AAP33265 | ATV084L | |
| | | | | | 33 | 22 | NP_078646 | LCDV042L | |
| ORF136R | 118946–119260 (104) | 11.61 | 7.66 | Possible membrane-associated motif in LPS-induced tumor necrosis factor alpha factor (smart00714) | 97.8 | 61 | AAP33206 | ATV029R | TM |
| ORF137R | 119282–120667 (461) | 49.69 | 10.07 | | 114 | 35 | AAP33207 | ATV030R | |
| ORF138L | 120907–120713 (64) | 7.38 | 4.72 | | | | | | |
| ORF139R | 121013–121324 (103) | 11.34 | 10.02 | | | | | | |
| ORF140R | 121397–122311 (304) | 32.14 | 4.85 | | | | | | |
| ORF141R | 122567–124558 (663) | 69.45 | 4.59 | Extracellular matrix glycoprotein Laminin subunit beta (KOG0994) | 80.1 | 25 | AAK01205 | *Mus musculus* | Mage-d3, TM |
| ORF142L | 124134–122572 (520) | 54.73 | 3.46 | | | | | | TM, SP |
| ORF143L | 124882–124643 (79) | 8.87 | 8.55 | | 34 | 49 | AAP33212 | ATV035L | TM, SA |
| ORF144R | 124963–125421 (152) | 16.67 | 9.07 | Fibroblast growth factor (KOG3885) | 67.8 | 33 | NP_570107 | *Rattus norvegicus* | Fibroblast growth factor, SP |
| ORF145R | 125480–125977 (165) | 18.07 | 9.20 | Acidic and basic fibroblast growth factor family (cd00058) | 47.4 | 29 | NP_032028 | *Mus musculus* | Fibroblast growth factor, TM, SP |
| ORF146L | 127052–126078 (324) | 36.71 | 6.72 | Predicted E3 ubiquitin ligase (KOG1814) | 269 | 42 | AAP33208 | ATV031R | NTPase/helicase |
| | | | | | 265 | 43 | AAL77808 | TFV078L | |
| | | | | | 105 | 30 | NP_078700 | LCDV107R | |
| | | | | | 60.8 | 35 | AAC97709 | MSEPV | |
| ORF147L | 128221–127187 (344) | 39.40 | 5.49 | | 42 | 27 | AAP33224 | ATV045R | TM |
| | | | | | 40 | 25 | AAL13097 | RGV 9807 | |
| ORF148R | 128324–128803 (159) | 17.63 | 6.99 | | 117 | 45 | AAP33210 | ATV033L | |
| | | | | | 54.3 | 28 | NP_078659 | LCDV059L | |
| ORF149R | 128843–129220 (125) | 14.59 | 5.10 | | | | | | |
| ORF150L | 130827–129301 (508) | 57.28 | 7.88 | | 283 | 34 | AAP33226 | ATV047L | Phosphotransferase, TM |
| | | | | | 171 | 27 | NP_078729 | LCDV143L | |
| | | | | | 65.9 | 35 | AAL98737 | ISKNV013R | |
| | | | | | 39.7 | 27 | AAK82240 | CIV380R | |
| ORF151L | 131435–130848 (195) | 22.26 | 6.48 | | 75 | 35 | AAP33227 | ATV048L | TM |
| | | | | | 45 | 29 | NP_078686 | LCDV091R | |
| ORF152R | 131534–132772 (412) | 46.57 | 9.46 | DNA or RNA helicases of superfamily II (COG1061) | 334 | 44 | AAP33229 | ATV050R | Helicase, SP |
| | | | | DEAD-like helicases superfamily (smart00487) | 328 | 44 | CAB37349 | EHNV | |
| | | | | | 324 | 44 | AAK55107 | TFV056L | |
| | | | | | 160 | 30 | AAB94470 | CIV161L | |
| | | | | | 64.7 | 24 | NP_042815 | ASFV | |
| ORF153L | 132661–132089 (190) | 21.96 | 8.93 | | 58 | 30 | CAB37350 | EHNV | 40kDa protein |
| | | | | | 57 | 29 | CAA58035 | FV3 | |
| ORF154R | 132788–133081 (97) | 10.87 | 9.73 | | | | | | TM |
| ORF155R | 133172–134899 (575) | 64.63 | 5.02 | Semaphorins (KOG3611) | 135 | 26 | NP_035482 | *Mus musculus* | Sema domain, immunoglobulin domain (Ig), and GPI membrane anchor, (semaphorin) 7A; H-Sema K1; Semaphorin K1, TM, SP |
| ORF156L | 135860–135048 (270) | 31.12 | 5.61 | | | | | | |
| ORF157R | 135948–136472 (174) | 19.73 | 8.80 | | 180 | 55 | AAP33225 | ATV046L | |
| ORF158L | 136944–136528 (138) | 15.77 | 6.15 | | | | | | |
| ORF159R | 137020–137511 (163) | 17.42 | 9.69 | | | | | | TM, SP |
| ORF160L | 137996–137508 (162) | 18.90 | 5.60 | | 122 | 40 | AAP33238 | ATV059R | |
| | | | | | 77.4 | 37 | NP_078685 | LCDV090R | |
| ORF161R | 138345–138533 (62) | 6.96 | 6.28 | | | | | | TM, SP |
| ORF162L | 139822–138674 (382) | 44.12 | 6.52 | | 407 | 50 | AAP33190 | ATV013L | Immediate-early protein ICP-46 |
| | | | | | 403 | 50 | AAK53747 | RRV | |
| | | | | | 401 | 49 | AAL77815 | TFV097R | |
| | | | | | 362 | 47 | PI4358 | FV3 | |
| | | | | | 131 | 27 | NP_078648 | LCDV047L | |
| | | | | | 48.5 | 29 | AAK82253 | CIV393L | |
| | | | | | 45.4 | 20 | AAL98839 | ISKNV115R | |

[a] aa, amino acid.
[b] MW, molecular mass.
[c] pI, isoelectric point.
[d] Accession numbers starting with PS are Prosite-derived numbers, while those starting with cd, smart, pfam, COG, or KOG are CD-Search within BlastP-derived numbers.
[e] Accession numbers are from GenBank or SwissProt database.
[f] All species corresponding virus ORFs are listed; only best match is listed if species are not related to virus.
[g] Function was deduced from the degree of amino acid similarity to or products of known genes or by the presence of Prosite signatures; TM, transmembrane domains; SP, N-terminal signal peptide; SA, N-terminal signal anchor.

TABLE 2. Positions of repetitive sequences in SGIV genome

| Nucleotide position | Period size (bp) | Copy no. | Matches (%) | G + C (%) |
|---|---|---|---|---|
| 7019–7062 | 18 | 2.4 | 96 | 85 |
| 11054–11156 | 51 | 2.0 | 96 | 52 |
| 11798–12126 | 54 | 6.1 | 99 | 69 |
| 16006–16132 | 27 | 4.7 | 87 | 71 |
| 17912–18143 | 36 | 6.4 | 96 | 55 |
| 23472–23505 | 9 | 3.8 | 100 | 53 |
| 31605–31635 | 15 | 2.1 | 100 | 48 |
| 41205–41242 | 18 | 2.1 | 90 | 71 |
| 41489–41528 | 9 | 4.4 | 93 | 75 |
| 50701–50751 | 18 | 2.8 | 100 | 73 |
| 58827–58871 | 15 | 3.0 | 100 | 74 |
| 58907–59192 | 27 | 10.6 | 100 | 74 |
| 99529–100248 | 63 | 11.4 | 100 | 73 |
| 102058–102093 | 18 | 2.0 | 100 | 78 |
| 111355–111749 | 26 | 15.2 | 95 | 65 |
| 122638–123756 | 42 | 26.6 | 80 | 64 |
| 138581–138629 | 17 | 2.9 | 100 | 29 |

longest perfect repetitive region, consisting of 11.4 copy numbers and 63 bp per period, is identified at positions 99529 to 100248 in the genome, where it is situated at the position 277 bp upstream of the start codon of the largest subunit of DNA-dependent RNA polymerase II (ORF104L). The biological function of these repetitive sequences remains to be determined. However, "junk DNA" intergenic sequences have been found to exert control over recombination, DNA replication, and gene expression. Many repeats act as binding sites for proteins or as structural elements on the level of RNA (35).

**DNA replication and repair.** Iridovirus replication occurs in two phases: a nuclear phase and a cytoplasmic phase. A functional nucleus is an essential cellular component for virus replication. After viral DNA is synthesized in the cell nucleus, the majority of viral DNA is transported to the cytoplasm where the packaging of DNA into the viral capsid occurs (40). SGIV encodes homologs of proteins involved in DNA replication, such as DNA polymerase (ORF128R), DNA repair protein (ORF097L), ATP-GTP binding protein (ORF052L), DNA binding/packing protein (ORF116R), and two helicases (ORF060R and ORF152R) containing highly conserved domains for DNA recombination and repair besides replication (Fig. 2).

ORF146L encodes a putative NTPase/helicase-like protein which could be a primase whose continual activity is required at the DNA replication fork. It catalyzes the synthesis of short molecules used as primers for DNA polymerase. ORF025L encodes a putative DNA binding motif—the so-called SAP motif (named after SAF-A/B, Acinus and PIAS)—which is found in a number of chromatin-associated proteins. It binds specifically to DNA elements called scaffold/matrix attachment regions, which are chromatin regions that bind to the nuclear matrix. Two proteins containing the SAP motif, SAF-A and Acinus, are targets of caspase cleavage during apoptosis, followed by chromatin degradation typical of programmed cell death (3). During apoptosis, SAF-A is cleaved in a caspase-dependent way. The cleavage occurs within the bipartite DNA-binding domain, resulting in the loss of DNA-binding activity and the concomitant detachment of SAF-A from nuclear structural sites. On the other hand, the cleavage does not compromise the association of SAF-A with hnRNP complexes, indicating that the function of SAF-A in the RNA metabolism is not affected during apoptosis (10). It may be inferred that the detachment of SAF-A, caused by the apoptotic proteolysis of its DNA-binding domain, could contribute to nuclear breakdown during host cell apoptosis.

**Transcription and mRNA biogenesis.** The putative SGIV gene products that are related to DNA transcription comprise

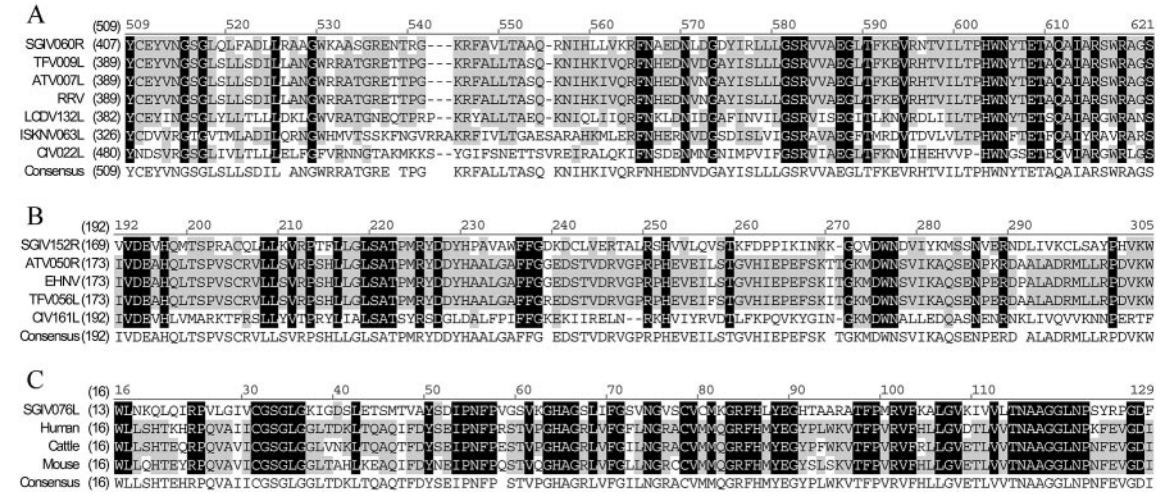


FIG. 2. Sequence alignment of selective SGIV ORF060R, ORF152L, and ORF076L with other known proteins. The homologous regions are shaded (black represents identical, grey represents conservative). The positions of the amino acid sequence are indicated on the left of the sequence. (A) Alignment of deduced amino acids of SGIV, ORF060R, accession no. AAS18075; TFV, ORF009L, accession no. NP_571991; ATV, ORF007L, accession no. AAP33184; RRV, accession no. AAK53744; LCDV, ORF132L, accession no. NP_078720; ISKNV, ORF063L, accession no. NP_612285; and CIV, ORF022L, accession no. NP_149485. (B) Alignment of deduced amino acids of SGIV, ORF152R, accession no. AAS18167; ATV, ORF050R, accession no. AAP33229; EHNV, accession no. CAB37349; TFV, ORF056L, accession no. NP_571999; and CIV, ORF161L, accession no. NP_149624. (C) Alignment of deduced amino acids of SGIV, ORF076L, AAS18091; human, *Homo sapiens*, accession no. NP_000261; cattle, *Bos taurus*, accession no. AAB34886; and mouse, *Mus musculus*, accession no. BAB25491.

the two largest subunits of DNA-dependent RNA polymerase II (ORF073L and ORF104L), one transcription elongation factor, TFIIS (ORF085R), and one RNase III enzyme (ORF084L; RNase III).

In addition, ORF063L exhibits similarity to one of the rat transcription factors which are important for transcriptional initiation. It may normally act to repress transcription at a variety of loci and may also play a role in chromatin structure or assembly (32). ORF061R encodes a TFIIF-interacting CTD phosphatase motif. It includes an NLI-interacting factor involved in RNA polymerase II regulation. ORF102L contains a fusion protein domain consisting of ubiquitin at the N terminus and ribosomal protein L40 at the C terminus. It also contains a zinc finger-like domain and is located in the cytoplasm (4). Ubiquitin is a highly conserved nuclear and cytoplasmic protein that has a major role in targeting cellular proteins for degradation by the 26S proteosome. It is also involved in the maintenance of chromatin structure, the regulation of gene expression, and the stress response.

**Nucleotide metabolism.** Predicted amino acid sequences of proteins required for the nucleotide transport and metabolism contain α and β subunits of ribonucleoside-diphosphate reductase (ORF064R and ORF047L), a ubiquitous cytosolic enzyme with a key role in DNA synthesis as it catalyzes the biosynthesis of deoxyribonucleotides. ORF049L encodes a dUTPase which is critical for the fidelity of DNA replication and repair. It also decreases the intracellular concentration of dUTP so that uracil cannot be incorporated into DNA (7). Purine nucleoside phosphorylase, which is involved in nucleotide transport and metabolism and encoded by ORF076L and which exists widely in mammals, was first identified in the family of *Iridoviridae* (Fig. 2).

**Cell signaling.** ORF078L and ORF081L encode two protein kinases that share a conserved catalytic core common with both serine/threonine and tyrosine protein kinases. There are a number of conserved regions in the catalytic domain of protein kinases. The protein corresponding to ORF067L belongs to the family of deoxynucleoside kinases that consists of various cytidine, guanosine, adenosine, and thymidine kinases (which also phosphorylate deoxyuridine and deoxycytosine). These enzymes catalyze the production of deoxynucleotide 5′-monophosphate from a deoxynucleoside.

**Immune evasion function.** ORF028L, ORF029L, ORF031L, ORF033L, ORF035L, and ORF131R encode homologs of the immunoglobulin (Ig)-like domains. Cellular members of the Ig superfamily include secreted and membrane-bound receptors and cell adhesion proteins (ORF029L and ORF035L) (39). ORF005L encodes a homolog of a mammalian amino acid transporter. It is also comprised of a C-type lectin signature which may bind to major histocompatibility complex (MHC) class I complex antigens and may promote or inhibit immune activity through intracellular signaling pathways. Thus, it is possible that ORF005L may interfere with normal immune surveillance or host responses (2). ORF068L is composed of an Ig-MHC signature ([FY]-x-C-x-[VA]-x-H). It is known that Ig constant domains and a single extracellular domain in each type of MHC chain are related. These homologous domains are approximately 100 amino acids long and include a conserved intradomain disulfide bond (26). These genes may function in host immune evasion, immune modulation, and aspects of cell and/or tissue tropism or perform other cellular functions (2).

ORF070R encodes a thiol oxidoreductase that impels the formation of disulfide bond. The correct formation of disulfide bonds is important for the folding and function of many secretory and membrane proteins. Organisms from all kingdoms of life have evolved a diverse range of thiol oxidoreductases (21).

ORF155R exhibits homology to mammalian semaphorin homologue. The sema domain occurs in semaphorins, which are a large family of secreted and transmembrane proteins, some of which function as repellent signals during axon guidance. Sema domains also occur in the hepatocyte growth factor receptor (41).

ORF053R encodes a prokaryotic membrane lipoprotein lipid attachment site found in prokaryotes. To our knowledge, this is a first report of this motif in iridovirus. Membrane lipoproteins are synthesized with a precursor signal peptide, which is cleaved by a specific lipoprotein signal peptidase (signal peptidase II). The peptidase recognizes a conserved sequence and cuts upstream of a cysteine residue to which a glyceride-fatty acid lipid is attached (12).

**Cellular function.** ORF003L is similar to 3-β-hydroxysteroid dehydrogenase from TFV and other poxviruses. It catalyzes the oxidative conversion of both 3-β-hydroxysteroid and ketosteroids, playing a critical role in biosynthesis of all classes of steroid hormones. ORF130L encodes a TonB-dependent receptor that interacts with outer membrane receptor proteins that carry out high-affinity binding and energy-dependent uptake of specific substrates into the periplasmic space. These substrates are either poorly permeative through porin channels or are encountered at very low concentrations. In the absence of TonB, these receptors bind to their substrates but do not carry out active transport. ORF115R encodes a homolog of a Bak protein, a member of the B-cell lymphoma (32% identity over 152 amino acids). Bcl-2 and related cytoplasmic proteins are key regulators of apoptosis, the cell suicide program critical for development, tissue homeostasis, and protection against pathogens. Bcl-2 family members are essential for maintenance of major organ systems to prevent a cellular apoptotic response to viral infection (1). ORF019R is composed of a glycoprotein hormone β chain signature. The function of ORF019R in the viral replication cycle is unknown.

**Phylogenetic analysis.** Iridoviruses are large cytoplasmic DNA viruses where each type has a specific insect or vertebrate host (38). One of the unifying features of this virus group is the presence of a major capsid protein (MCP) that is approximately 50 kDa in size. MCP is a suitable target for the study of viral evolution, since it contains highly conserved domains, but is sufficiently diverse to distinguish closely related iridovirus isolates (34). The amino acid sequences of the known MCPs are used in comparative analyses to elucidate the phylogenic relationships between different cytoplasmic DNA viruses.

ORF072R encodes SGIV MCP. Phylogenetic analysis indicated that SGIV is distinct from all known iridoviruses (Fig. 3), but it is much closer to the genus *Ranavirus*. Within the MCP, amino acid identities of 73.0 (BIV), 72.8 (TFV), 72.8 (FV3), 72.4 (ATV), and 72.1% (ENHV) are noted. However, it only shows amino acid identities of 52.2 (LCDV), 45.7 (CIV), and 44.4% (ISKNV). This suggested that SGIV is a novel member of the genus *Ranavirus* within the family *Iridoviridae*. Gener-

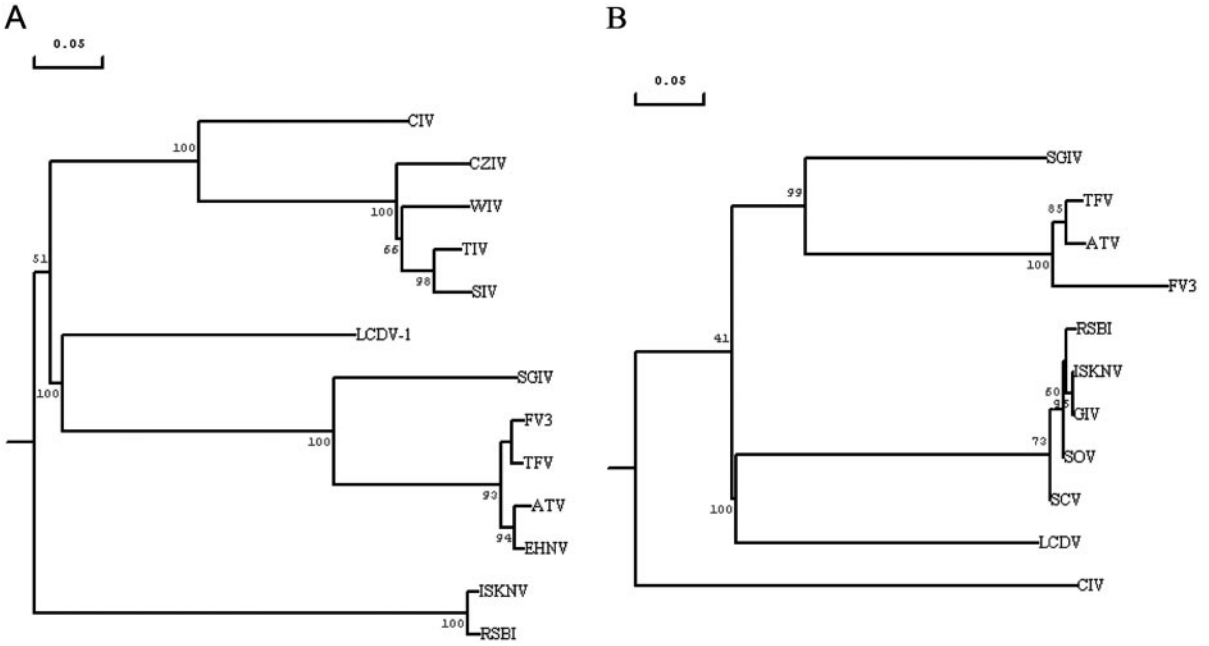


FIG. 3. Phylogenetic relationship of SGIV with representative iridoviruses. The analysis was based on the multiple alignments of the protein sequences of the major capsid protein and ATPase of iridoviruses. (A) SGIV, ORF072R, accession no. AAS18087; ATV ORF014L, accession no. AAP33191; ISKNV, ORF006L, accession no. AAL72276; TFV, ORF096R, accession no. AAK55105; FV3, accession no. AAB01722; EHNV, accession no. AAO32315; LCDV-1, ORF147L, accession no. AAC24486; CIV ORF274L, accession no. AAK82135; CZIV, accession no. AAB82569; RSBI, accession no. AAP74204; WIV, accession no. AAB82568; TIV, accession no. VCXFTI; and SIV, accession no. VCXFSI. (B) SGIV, ORF134L, accession no. AAS18149; TFV, ORF016R, accession no. AAL77796; ATV, ORF083L, accession no. AAP33264; FV3, accession no. AAA43823; SOV, accession no. AAN77575; ISKNV, ORF122R, accession no. 98847; GIV, accession no. AAL68652; RSBI, accession no. BAA28670; SCV, accession no. AAL73346; LCDV, ORF054R, accession no. NP_078656; and CIV, 075L, accession no. AAB94422.

ally, viruses with sequence identities within a given gene of less than 80% are considered members of different species rather than strains of the same species (37). The conserved protein sequence of the ATPase was also used to determine the relationship of SGIV with other iridoviruses (Fig. 3). The phylogenic tree of ATPase supports the view that SGIV is a novel species of the genus *Ranavirus*.

**Relationship of SGIV to other iridoviruses.** Conservation of synteny and of gene order can give insights to assess structural conservation among the viral genomes within the family *Iridoviridae*. Conservation of synteny refers to a pair of genomes in which at least some of the genes are located at similar map positions regardless of the gene order or the presence of intervening genes. When the evolutionary distance is large, scrambling of the gene order and the presence of nonsyntenic intervening genes become frequent (23). Therefore, it is necessary to account for these features when studying iridovirus evolution.

To make comparisons between SGIV and five other iridovirus genomes (ATV, TFV, LCDV, ISKNV, or CIV), we shifted

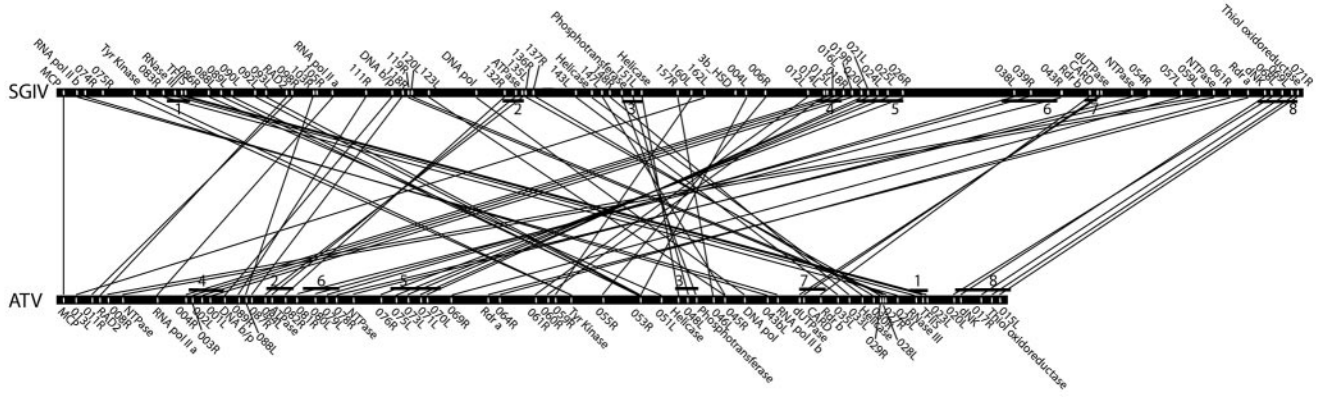


FIG. 4. Conserved segments between the SGIV and ATV genomes. Both genomes are linearized and shifted genes encoding MCP as the start point. Only linked genes or annotated ORFs are indicated. Straight lines represent the gene linkages between two species. Black bars indicate the conserved syntenic regions of both genomes.

the starting coordinates and set the start codon (ATG) of MCPs as the first base for all viral genomes. We also altered sense and antisense strands on ATV, LCDV, ISKNV, and CIV genomes in order to get the same nucleotide order on MCPs individually. However, none of the annotated ORFs were affected.

Comparing the SGIV genome to the LCDV, ISKNV, or CIV genome does not show possible clustering of genes in spite of the fact that SGIV shares 43, 22, or 29 real or annotated ORFs with the LCDV, ISKNV, or CIV genome, respectively. Although only 20 ORFs of SGIV reveal similarities to those of TFV genomes, it appears that some genes are located at similar map positions. In contrast, comparison of the SGIV genome with those of other iridoviruses shows that SGIV is much closer to ATV than other iridoviruses whose genomes are known. The sequenced genomes of the two closely related iridoviruses SGIV and ATV were compared with emphasis on genome organization and coding capacity (Fig. 4). The genome size and ORF numbers of the SGIV genome are much larger than those of ATV, which has a genome of 106,332 bp and contains 91 ORFs. Seventy-one ORFs of SGIV and ATV showed close homologies. There were some discrepancies in annotation, but inspection of DNA sequences showed that the corresponding genes are always present. Twenty-two corresponding ORFs between these two genomes are putative genes, but all remaining ORFs have no known function (Table 1). At least eight regions of conserved synteny containing more than three genes or annotated ORFs were also examined. Interestingly, TFIIS, RNase III, and one ORF (SGIV 086R, ATV 023L, and TFV 087R) are arranged in succession among SGIV, ATV, and TFV (Fig. 4). This cluster of genes may become a useful gene marker to distinguish unknown viruses from the genus *Ranavirus*. Scrambling of gene blocks was also observed between these two genomes. Two continuous conserved regions (blocks 4 and 5) in the SGIV genome were located at two separate gene blocks in the ATV genome, in which blocks 2 and 6 inserted. Orthologous genes between SGIV and ATV are quite similar in sequence conservation and also in gene order. Conserved linkages between SGIV and ATV indicate that they evolved from a common ancestor.

**Identification of SGIV proteins by MALDI-TOF MS and RT-PCR.** Purified viral proteins of SGIV extracted from the lowest band (50% sucrose) were separated by SDS-PAGE (Fig. 5). Thirty-nine clearly defined bands were excised and subjected to reduction, alkylation, tryptic digestion, and mass spectrometric analysis by matrix-assisted laser desorption ionization–time of flight (MALDI-TOF) mass spectrometry. Peak lists of tryptic peptide molecular weights of each band were searched against the 162-ORF database of SGIV to identify the proteins and corresponding genes. Twenty-six proteins, covering 5 to 67% of amino acid sequences, were matched with the theoretical SGIV ORF database by using the AutoMS-Fit search program (Table 3). Of those proteins matched in this study, only six are known viral proteins; these are MCP (ORF072R), DNA polymerase (ORF128R), two proteins relevant to DNA replication (ORFs 052L and 060R), RNase III (ORF084L), and tyrosine protein kinase (ORF078L). Several SDS-PAGE bands in the low-molecular-mass area (bands 35 to 39 and molecular masses around 10 kDa) matched ORF052L and ORF060R. However, the molecular weight search scores were quite low, and the identities of these proteins cannot be confirmed from the data. Matching of multiple numbers of SDS-PAGE bands to ORF052L and ORF060R may be explained by possible degradation of these large proteins during virus purification, since no protease inhibitors were used during these procedures. We were able to verify 12 SGIV genes which exhibited homologies to genes from other iridoviruses but whose biological function remain to be established. Another eight SGIV genes of unknown function, showing no homologies to any other viruses, were also verified.

Mass spectrometry is a powerful and a high-throughput technique used to identify proteins. It has been applied to analyze the proteome of white spot shrimp virus (17). The completion of the genomic DNA sequence of SGIV greatly

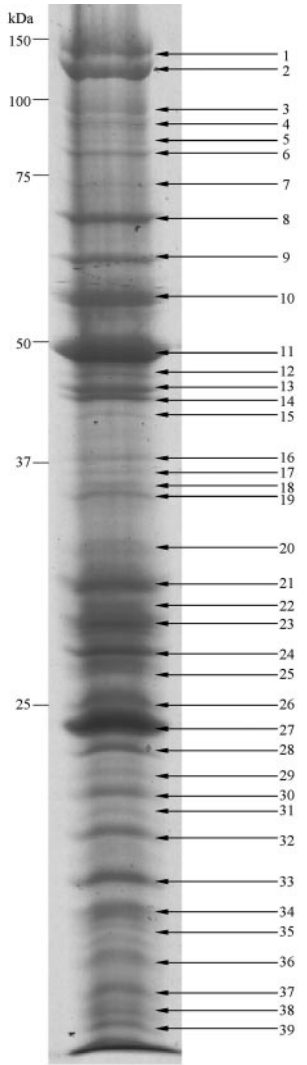


FIG. 5. SDS-PAGE of SGIV proteins. Viral proteins were purified and separated via one-dimensional SDS-PAGE. Thirty-nine visible gel-separated protein bands were excised and digested enzymatically, and their mass spectra were obtained and automatically searched against the SGIV ORF database. Twenty-six proteins were identified by MALDI-TOF mass spectrometry. However, peptide signals from bands 26, 35, 36, 37, 38, and 39 were too low to give satisfactory identification.

TABLE 3. Identification of SGIV proteins corresponding to ORFs by MS

| Band | ORF[a] | GenBank accession no. | Protein sequence coverage (%) | Confirmed by RT-PCR (48 h of infection) |
|---|---|---|---|---|
| 1 | 039L | AAS18054 | 32 | + |
| 2 | 012L | AAS18027 | 25 | + |
| 3 | 078L | AAS18093 | 20 | |
| | 039L | | 13 | |
| 4 | 060R | AAS18075 | 15 | |
| | 039L | | 15 | |
| 5 | 060R | | 15 | |
| | 052L | AAS18067 | 8 | |
| 6 | 052L | | 5 | |
| 7 | 012L | | 13 | |
| 8 | 069L | AAS18084 | 24 | + |
| 9 | 026R | AAS18041 | 34 | + |
| 10 | 052L | | 6 | |
| | 128R | AAS18143 | 9 | |
| 11 | 072R | AAS18087 | 27 | |
| 12 | 016L | AAS18031 | 24 | + |
| | 093L | AAS18108 | 22 | + |
| 13 | 090L | AAS18105 | 43 | + |
| | 089L | AAS18104 | 19 | + |
| 14 | 090L | | 6 | |
| 15 | 084L | AAS18099 | 22 | |
| 16 | 060R | | 7 | |
| | 089L | | 14 | |
| 17 | 072R | | 14 | |
| 18 | 072R | | 12 | |
| 19 | 101R | AAS18116 | 24 | + |
| 20 | 046L | AAS18061 | 25 | + |
| 21 | 018R | AAS18033 | 29 | + |
| 22 | 082L | AAS18097 | 47 | + |
| | 055R | AAS18070 | 37 | + |
| 23 | 055R | | 67 | |
| 24 | 156L | AAS18171 | 30 | + |
| 25 | 160L | AAS18175 | 24 | + |
| 26 | / | | | |
| 27 | 075R | AAS18090 | 15 | + |
| 28 | 075R | | 29 | |
| 29 | 075R | | 33 | |
| 30 | 075R | | 43 | |
| 31 | 057L | AAS18072 | 10 | + |
| 32 | 012L | | 8 | |
| | 125R | AAS18140 | 27 | + |
| 33 | 013R | AAS18028 | 23 | + |
| 34 | 050L | AAS18065 | 16 | + |
| 35 | / | | | |
| 36 | / | | | |
| 37 | / | | | |
| 38 | / | | | |
| 39 | / | | | |

[a] / refers to protein bands that did not identify any proteins reliably.

facilitated the discovery of new proteins by the proteomic approach, which proved to be an effective and sensitive way for discovering SGIV proteins. We have analyzed the SGIV proteome by one-dimensional gel. Furthermore, two-dimensional gel analysis will be used later to identify more novel proteins.

All 20 novel genes mentioned above were further checked and verified at the RNA level by RT-PCR. Total RNA (including virus and host) was extracted at 0-, 6-, 12-, 24-, 48-, and 72-h infective stages. Several genes started transcription early after the cell line was inoculated, 12 h (i.e., ORF090R and ORF093R) (data not shown). All novel genes were detected by RT-PCR after 48 h of infection (Fig. 6). Full lengths of 14 novel genes were amplified by reverse transcriptase and HotStarTaq DNA polymerase (Fig. 6A and 6). However, only partial sequences of ORF012L (2,107 to 3,075 bp), ORF039L (1 to 900 bp), ORF046L (17 to 747 bp), ORF050L (9 to 600 bp), ORF055R (12 to 588 bp), and ORF057L (7 to 832 bp) were amplified (Fig. 6C). Furthermore, RT-PCR products were used for DNA sequencing to confirm their respective authenticity.

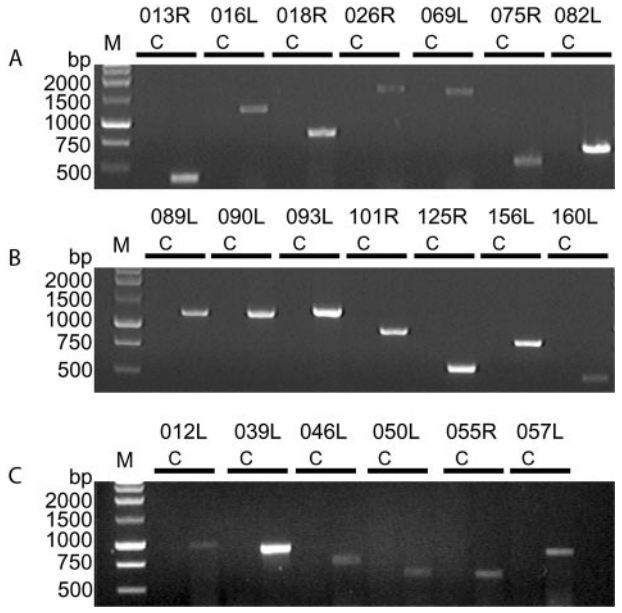


FIG. 6. Amplification of 20 novel genes of SGIV via RT-PCR. Total RNA (harvested after 48 h of infection) was isolated by using the RNeasy Mini kit and amplified by using the OneStep RT-PCR kit. Full lengths of 14 genes were amplified (A and B). Partial sequences were acquired from another six genes (C). Lanes C, control; lane M, 1-kb DNA ladder (Promega).

**Prediction of potential novel proteins.** The existence of an ORF in genomic data does not necessarily imply the existence of a functional gene. Despite the advances in bioinformatics, it is difficult to predict genes accurately from the genomic data alone (27). Although the genome sequence of the SGIV will ease the problem of gene prediction through comparative genomics, the success rate for correct prediction of the primary structure is still low. Therefore, verification of a gene product by proteomic methods is an important first step in annotating the genome. We predicted the secondary protein structures for these novel or unidentified proteins. Using a protein secondary structure predicting program, PSIPRED (20, 22), for the 20 novel proteins identified by MS in this study, we found that two proteins encoded by ORF046L and ORF050L consisted of random coils. ORF012L encoded a protein containing only $\alpha$ helices. Another 17 proteins were categorized as $\alpha/\beta$ proteins. The prediction of transmembrane regions and orientation was also done via TMpred on the ISREC server and is listed in Table 1. We intend to elucidate the three-dimensional structures of these novel proteins by analyzing structural biology and their functions by using small interfering RNA and other related technologies.

## CONCLUSION

We report a complete sequence of SGIV. Genomic analysis of SGIV provided fundamental knowledge of viral functions,

such as DNA replication and transcription, nucleotide metabolism, protein processing, manipulation of cellular responses, and virus-host interaction. We compared the SGIV genome with other five iridovirus genomes at the DNA and protein levels. Besides the conserved and known proteins, we also identified 20 novel proteins by using the proteomic approach. Proteomic analysis showed evidence of novel proteins detected at the posttranscriptional level. Our studies will provide important information on molecular mechanism of virus-host interactions and will have a broad impact on future strategies for the design of specific inhibitors or drugs to control these pathogens in general.

## ACKNOWLEDGMENTS

We greatly appreciate Shashikant Joshi for modification of the manuscript. We thank Yunhan Hong for helpful discussions. Swarup Sanjay's suggestions regarding the construction of the shotgun library are acknowledged. We are grateful to Xianhui Wang for advice on mass spectrometry and Yun Ping Lim for her assistance in the bioinformatic work.

This work was financially supported by the grant "Establishment of a Laboratory of Excellence in Aquatic and Marine Biotechnology (LEAMB)" to Choy Leong Hew.

## REFERENCES

1. **Adams, J. M., and S. Cory.** 1998. The Bcl-2 protein family: arbiters of cell survival. Science **281:**1322–1326.
2. **Afonso, C. L., E. R. Tulman, Z. Lu, L. Zsak, G. F. Kutish, and D. L. Rock.** 2000. The genome of fowlpox virus. J. Virol. **74:**3815–3831.
3. **Ahn, J. S., and M. C. Whitby.** 2003. The role of the SAP motif in promoting holliday junction binding and resolution by SpCCE1. J. Biol. Chem. **278:** 29121–29129.
4. **Chan, Y. L., K. Suzuki, and I. G. Wool.** 1995. The carboxyl extensions of two rat ubiquitin fusion proteins are ribosomal proteins S27a and L40. Biochem. Biophys. Res. Commun. **215:**682–690.
5. **Chew-Lim, M., G. H. Ngoh, M. K. Ng, J. M. Lee, P. Chew, J. Li, Y. C. Chan, and J. L. C. Howe.** 1994. Grouper cell line for propagating grouper viruses. Singap. J. Prim. Ind. **22:**113–116.
6. **Chua, F. H. C., M. L. Ng, K. L. Ng, J. J. Loo, and J. Y. Wee.** 1994. Investigation of outbreaks of a novel disease, 'sleepy grouper disease,' affecting the brown-spotted grouper, Epinephelus tauvina Forskal. J. Fish Dis. **17:**417–427.
7. **Eklunda, H., U. Uhlina, M. Färnegårdh, D. T. Loganb, and P. Nordlundb.** 2001. Structure and function of the radical enzyme ribonucleotides reductase. Prog. Biophys. Mol. Biol. **77:**177–268.
8. **Falquet, L., M. Pagni, P. Bucher, N. Hulo, C. J. Sigrist, K. Hofmann, and A. Bairoch.** 2002. The PROSITE database, its status in 2002. Nucleic Acids Res. **30:**235–238.
9. **Gibson-Kueh, S., P. Netto, G. H. Ngoh-Lim, S. F. Chang, L. L. Ho, Q. W. Qin, F. H. C. Chua, M. L. Ng, and H. W. Ferguson.** 2003. The pathology of systemic iridoviral disease in fish. J. Comp. Pathol. **129:**111–119.
10. **Gohring, F., B. L. Schwab, P. Nicotera, M. Leist, and F. O. Fackelmayer.** 1997. The novel SAR-binding domain of scaffold attachment factor A (SAF-A) is a target in apoptotic nuclear breakdown. EMBO J. **16:**7361–7371.
11. **Goorha, R., and K. G. Murti.** 1982. The genome of frog virus 3, an animal DNA virus, is circularly permuted and terminally redundant. Proc. Natl. Acad. Sci. USA **79:**248–252.
12. **Hayashi, S., and H. C. Wu.** 1990. Lipoproteins in bacteria. J. Bioenerg. Biomembr. **22:**451–471.
13. **He, J. G., L. Lu, M. Deng, H. H. He, S. P. Weng, X. H. Wang, S. Y. Zhou, Q. X. Long, X. Z. Wang, and S. M. Chan.** 2002. Sequence analysis of the complete genome of an iridovirus isolated from the tiger frog. Virology **292:**185–197.
14. **He, J. G., M. Deng, S. P. Weng, Z. Li, S. Y. Zhou, Q. X. Long, X. Z. Wang, and S. M. Chan.** 2001. Complete genome analysis of the mandarin fish infectious spleen and kidney necrosis iridovirus. Virology **291:**126–139.
15. **Hofmann, K., and W. Stoffel.** 1993. TMbase—a database of membrane spanning proteins segments. Biol. Chem. Hoppe-Seyler **374:**166. http://www.ch.embnet.org/software/TMPRED_form.html.
16. **Huang, C. H., L. R. Zhang, J. H. Zhang, L. C. Xiao, Q. J. Wu, D. H. Chen, and J. K. K. Li.** 2001. Purification and characterization of white spot syndrome virus (WSSV) produced in an alternate host: crayfish, *Cambarus clarkia.* Virus Res. **76:**115–125.
17. **Huang, C. H., X. B. Zhang, Q. S. Lin, X. Xu, Z. H. Hu, and C. L. Hew.** 2002. Proteomic analysis of shrimp white spot syndrome viral proteins and characterization of a novel envelope protein VP466. Mol. Cell. Proteomics **1:**223–231.
18. **Jakob, N. J., K. Muller, U. Bahr, and G. Darai.** 2001. Analysis of the first complete DNA sequence of an invertebrate iridovirus: coding strategy of the genome of *Chilo* iridescent virus. Virology **286:**182–196.
19. **Jancovich, J. K., J. H. Mao, V. G. Chinchar, C. Wyatt, S. T. Case, S. Kumar, G. Valente, S. Subramanian, E. W. Davidson, J. P. Collins, and B. L. Jacobs.** 2003. Genomic sequence of a ranavirus (family *Iridoviridae*) associated with salamander mortalities in North America. Virology **316:**90–103.
20. **Jones, D. T.** 1999. Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. **292:**195–202.
21. **Kadokura, H., and J. Beckwith.** 2001. The expanding world of oxidative protein folding. Nat. Cell Biol. **3:**E247–E249.
22. **McGuffin, L. J., K. Bryson, and D. T. Jones.** 2000. The PSIPRED protein structure prediction server. Bioinformatics **16:**404–405. http://bioinf.cs.ucl.ac.uk/psipred/psiform.html.
23. **Nadeau, J. H., and D. Sankoff.** 1998. The lengths of undiscovered conserved segments in comparative maps. Mamm. Genome **9:**491–495.
24. **Nielsen, H., and A. Krogh.** 1998. Prediction of signal peptides and signal anchors by a hidden Markov model, p. 122–130. *In* Proceedings of the 6th International Conference on Intelligent Systems for Molecular Biology. AAAI Press, Menlo Park, Calif.
25. **Nielsen, H., J. Engelbrecht, S. Brunak, and G. V. Heijne.** 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. **10:**1–6.
26. **Orr, H. T., D. Lancet, R. J. Robb, J. A. Lopez de Castro, and J. L. Strominger.** 1979. The heavy chain of human histocompatibility antigen HLA-B7 contains an immunoglobulin-like region. Nature **282:**266–270.
27. **Pandey, A., and M. Mann.** 2000. Proteomics to study genes and genomes. Nature **405:**837–846.
28. **Qin, Q. W., S. F. Chang, G. H. Ngoh-Lim, S. Gibson-Kueh, C. Shi, and T. J. Lam.** 2003. Characterization of a novel ranavirus isolated from grouper *Epinephelus tauvina.* Dis. Aquat. Org. **53:**1–9.
29. **Qin, Q. W., T. J. Lam, Y. M. Sin, H. Shen, S. F. Chang, G. H. Ngoh, and C. L. Chen.** 2001. Electron microscopic observations of a marine fish iridovirus isolated from brown-spotted grouper, *Epinephelus tauvina.* J. Virol. Methods **98:**17–24.
30. **Schnitzler, P., J. B. Soltau, M. Fischer, M. Reisner, J. Scholz, H. Delius, and G. Darai.** 1987. Molecular cloning and physical mapping of the genome of insect iridescent virus type 6 further evidence for circular permutation of the viral genome. Virology **160:**66–74.
31. **Shevchenko, A., M. Wilm, O. Vorm, and M. Mann.** 1996. Mass spectrometric sequencing of protein silver-stained gels. Anal. Chem. **68:**850–858.
32. **Swanson, M. S., E. A. Malone, and F. Winston.** 1991. SPT5, an essential gene important for normal transcription in *Saccharomyces cerevisiae*, encodes an acidic nuclear protein with a carboxy-terminal repeat. Mol. Cell. Biol. **11:** 3009–3019.
33. **Tidona, C. A., and G. Darai.** 1997. The complete DNA sequence of lymphocystis disease virus. Virology **230:**207–216.
34. **Tidona, C. A., P. Schnitzler, R. Kehm, and G. Darai.** 1998. Is the major capsid protein of iridoviruses a suitable target for the study of viral evolution? Virus Genes **16:**59–66.
35. **van Belkum, A., S. Scherer, L. van Alphen, and H. Verbrugh.** 1998. Short-sequence DNA repeats in prokaryotic genomes. Microbiol. Mol. Biol. Rev. **62:**275–293.
36. **van Regenmortel, M. H., C. M. Fauquet, and D. H. L. Bishop.** 2000. Virus taxonomy. Seventh report of the International Committee on Taxonomy of Viruses. Academic Press, New York, N.Y.
37. **Ward, C. W.** 1993. Progress towards a higher taxonomy of viruses. Res. Virol. **144:**419–453.
38. **Webby, R., and J. Kalmakoff.** 1998. Sequence comparison of the major capsid protein gene from 18 diverse iridoviruses. Arch. Virol. **143:**1949–1966.
39. **Williams, A. F., and A. N. Barclay.** 1988. The immunoglobulin superfamily—domains for cell surface recognition. Annu. Rev. Immunol. **6:**381–405.
40. **Williams, T.** 1996. The iridoviruses. Adv. Virus Res. **46:**345–412.
41. **Winberg, M. L., J. N. Noordermeer, L. Tamagnone, P. M. Comoglio, M. K. Spriggs, M. Tessier-Lavigne, and C. S. Goodman.** 1998. Plexin A is a neuronal semaphorin receptor that controls axon guidance. Cell **95:**903–916.