Analysis of the genome of leporid herpesvirus 4

Bobby Babra[1], Gregory Watson[1], Wayne Xu[2,5], Brendan Jeffrey [1], Jia-Rong Xu[1,3], Dan Rockey[1,4], George Rohrmann[4], and Ling Jin[1,4*]

[1]Department of Biomedical Sciences, College of Veterinary Medicine, Oregon State University, Corvallis, OR 97331. [2]Supercomputing Institute for Advanced Computational Research, University of Minnesota, Minneapolis, Minnesota 55455. [3]College of Veterinary Medicine, Nanjing Agricultural University, Jiangsu 210095, China. [4]Department of Microbiology, College of Science, Oregon State University, Corvallis, OR 97331. [5]Department of Veterinary and Biomedical Sciences, University of Minnesota, 1971 Commonwealth Avenue, Saint Paul, MN 55108

Running title: LHV-4 genome sequencing
Keywords: LHV-4, genome, sequencing
Abstract: 150
Figures: 6
Tables: 3
Supplemental Tables and Figures: 6
Accession No. JQ596859

*Address Correspondence to:

Ling Jin
Department of Biomedical Sciences
College of Veterinary Medicine,
Oregon State University,
Corvallis, OR 97331
Email: ling.jin@oregonstate.edu
Phone: 541-737-9893
Fax: 541-737-2730

Abstract

The genome of a herpesvirus highly pathogenic to rabbits, leporid herpesvirus 4 **(**LHV-4), was analyzed using high-throughput DNA sequencing technology and primer walking. The assembled DNA sequences were further verified by restriction endonuclease digestion and Southern blot analyses. The total length of the LHV-4 genome was determined to be about 124 kb. Genes encoded in the LHV-4 genome are most closely related to herpesvirus of the *Simplexvirus* genus, including human herpesviruses (HHV -1 and HHV-2), monkey herpesviruses including cercopithicine (CeHV-2 and CeHV-16), macacine (McHV-1), bovine herpesvirus 2 (BHV-2), and a lineage of wallaby (macropodid) herpesviruses (MaHV -1 and -2). Similar to other simplexvirus genomes, LHV-4 has a high overall G+C content of 65%-70% in the unique regions and 75-77% in the inverted repeat regions. Orthologs of ICP34.5 and US5 were not identified in the LHV-4 genome. This study shows that LHV-4 has the smallest simplexvirus genome characterized to date.

Introduction

Herpesviridae is a widely distributed family of large DNA viruses that have enveloped spherical to pleomorphic virions of 120-200 nm and isometric capsids of 100-110 nm in diameter. The genomes are linear, double-stranded DNA ranging from 125-241 kb and have a guanine + cytosine content of 32-75 % (McGeoch et al., 2006). They are pathogenic for many different types of vertebrates, ranging from humans to reptiles and birds. Currently there are three subfamilies within the *Herpesviridae*: *Alphaherpesvirinae*, *Betaherpesvirinae*, and *Gammaherpesvirinae*. Within the *Alphaherpesvirinae* there are five genera: *Simplexvirus*, *Varicellovirus*, *Mardivirus*, *Scutavirus*, and *IItovirus*. With the exception of bovine herpesvirus 2 (BHV-2), LHV-4, and two closely related viruses of marsupials (wallabies), macropodid herpesvirus 1 and 2 (MaHV-1, -2) (Guliani et al., 1999; Jin et al., 2008a; Johnson and Whalley, 1990; Mahony et al., 1999; McGeoch et al., 2006), all other members of the genus *Simplexvirus* are pathogenic to primates, including human herpesviruses 1 and 2 (HHV-1, and HHV-2,), cercopithecine herpesvirus -1, -2 and -16 (CeHV-1, -2 and -16), macacine herpesvirus-1

(McHV1), ateline (spider monkey) herpesvirus 1 (AtHV-1), and saimiriine (squirrel monkey) herpesvirus 1 (SaHV-1). The genomes of *Simplexviruses* are about 150 kb on average (McGeoch and Cook, 1994), and contain two unique regions called the unique long (UL) and unique short (US), which are both flanked by a pair of inverted repeat sequences: for UL, the flanking inverted repeat is called RL, whereas for US, the inverted repeat is called RS (Fig. 1). Virions contain four isomeric forms of the genome and have genomic sequence similarity greater than 50% to that of HHV-2 (McGeoch, 1989).

In addition to LHV-4, three naturally occurring herpesviruses of rabbits and hares (leporids) called Leporid herpesvirus (LHV-1, -2 and -3) have been identified (Roizman and Pellett, 2001). All three viruses are members of the *Rhadinovirus* genus of *Gammaherpesvirinae,* but only LHV-3 is associated with lymphoproliferative disease in cottontail rabbits (Hesselton et al., 1988; Jin et al., 2008a). In contrast, LHV-4 is a simplexvirus that was associated with a disease outbreak in domestic rabbits near Anchorage, Alaska in the US (Jin et al., 2008a; Jin et al., 2008b). It causes an acute infection similar to ocular infections produced by HHV-1 and is characterized by conjunctivitis, corneal epithelial keratitis and periocular swelling, ulcerative dermatitis, progressive weakness, anorexia, respiratory distress, and abortion (Jin et al., 2008b). LHV-4 is highly virulent in newborn and pre-weaned rabbits and caused about 28% mortality in the Alaska outbreak. Experimental LHV-4 infection in ten-week-old rabbits resulted in high morbidity with severe ocular disease and high fever, but no mortality was observed. Following primary infection, LHV-4 DNA was detected in the trigeminal ganglia (TG) but not in other tissues after 20 days post-infection in experimentally infected rabbits and mice (Jin et al., 2008a). These findings suggest that LHV-4 is able to establish latent infections in TG, which is one of the unique features of alphaherpesvirus infections.

In this report we describe the sequence of the genome of LHV-4. Our genome assembly revealed that LHV-4 is about 124 kbp and has similar overall genomic architecture to other annotated simplexviruses including the inverted repeat regions. Although most genes from the unique regions have an average of 40-79% sequence similarity to HHV-1 or HHV-2, the genes in the inverted repeats contain conserved regions with 68-82% sequence similarity to those of HHV-1 and HHV-2. Orthologs of US5 and ICP34.5 were not identified in the LHV-4 genome.

Results and Discussion

Sequencing of the LHV-4 genome: Viral DNA extracted from purified virions was sequenced by high-throughput DNA sequencing technology using the GS FLX+ System from 454 Life Technology (Roche). The GS FLX+ System was used to produce longer reads to avoid the difficulty in genome assembly due to short repeat sequences that frequently are present in the genome. 65,597 sequences were generated from GS FLX with average read length of 340 bp and mode read length at 450bp. More than 99% of the sequences have an average base Phred quality score of greater than 20 (Supplemental Fig.1). These sequences produced a coverage depth of 38x across the viral genome. 553 contigs were assembled by Newbler software from 454 Life Technology with N50 of 17890 bases in length. Two large contigs (96902 bp and 11908 bp) were determined to be the unique long (UL) and unique short (US) regions, respectively, by comparing them to annotated herpesvirus genomes. Only partial inverted repeat sequences were assembled by *de novo* assembly. Using primer walking and sequencing of numerous PCR DNA amplification products, the long repeat region (RL) of 3159 bp and the short repeat region (RS) of 4541 bp were assembled using MIRA and Geneious software. RL2 and RS1 are also predicted in the RL and RS, respectively (Table 1 and Fig. 2). The "a" sequence was obtained by comparing the end sequences of both RL and RS and is estimated to be 379 bp. To verify the assembled DNA sequence, the LHV-4 genome was examined by restriction endonuclease digestion and Southern blot analyses. To resolve large DNA fragments, the digested DNA was separated by either 0 .7 or 1.0% agarose gel electrophoresis. The expected DNA fragments from *Hind*III, *Bam*HI, and *Eco*RI digestions within UL were produced as predicted (Supplemental Fig. 2). Based on the assembled genome sequence and endonuclease restriction digestion results, the LHV-4 genome was found to be about 124 kbp. This is close to the size of between 112 and 130 kbp estimated by field inversion gel electrophoresis (FIGE) (Supplemental Fig. 3).

Confirmation of the LHV-4 inverted repeat size: Both RL and RS of LHV-4 are much shorter than those of other annotated simplexviruses. To confirm the size of the inverted repeats flanking the unique regions, both virion and nucleocapsid DNA were digested by *Bam*HI and *Eco*RI to ensure all possible isomeric forms of the viral genome were included (Fig. 1). There are 12

*Bam*HI and 14 *Eco*RI restriction sites within the UL. Since no *Eco*RI site was predicted in RL, RS and US, the terminal RL and internal RL should be included in *Eco*RI fragments (*Eco*RIa and *Eco*RIb) at about 4 kb and 31 kb, respectively, when UL1 is adjacent to the terminal RL (Fig. 1A), or about 10 kb and 25 kb, respectively, when UL56 is adjacent to terminal RL (Fig. 1B). To confirm this prediction, *Eco*R I digested viral DNA fragments were hybridized by DNA probes selected from RL (ICP0) and RS (ICP4) (Fig. 1). When the *Eco*RI digested viral DNA was hybridized with the ICP0 probe, the predicted *Eco*RI DNA fragments at 4, 10, 25, and 31 kb were all hybridized (Fig. 3). When the ICP4 probe was used, only the predicted *Eco*RI fragment at 25 kb and 31 kb hybridized (Fig. 3). Collectively, these results agree with our predictions based on bioinformatic assembly (Fig. 1). To confirm these hybridization results, the digested viral DNA was probed again with probes ICP0, ICP4, UL56, and US1 sequentially on a single blot membrane. The UL1 probe which was hybridized to a different blot under the same conditions as the others (shown in Fig. 4). UL1 probe is selected within UL1 gene before the *Eco*RIa restriction site (Fig. 1). When the UL1 probe was used, only two bands around the predicted sizes were hybridized in viral DNA digested with *Eco*RI: the large band is about 25 kb (open arrow), while the smaller band is around 4 kb (solid arrow) (Fig. 4, UL1). UL56 probe is selected within UL56 gene after the *Eco*RIb restriction site (Fig. 1). When the UL56 probe was used, only a 10 kb (solid arrow) or a 31 kb (open arrow) *Eco*RI fragment was hybridized (Fig.4-UL56, lane with nucleocapsid DNA). Both UL1 and UL56 hybridization results agree with the predicted fragment lengths as shown in Fig. 1. The first *Eco*RIa site is about 1 kb from the RL, while the last *Eco*RIb site is about 7 kb from the RL (Fig. 1); therefore, it is calculated that *Eco*RI fragments hybridized by the UL56 probe are about 6 kb larger than those hybridized by the UL1 probe (Fig. 4). When US1 and ICP4 probes were used, only the predicted *Eco*RI fragments at 25 kb and 31 kb were hybridized, which suggest that ICP0, ICP4, and US1 are all hybridized to the same *Eco*RI fragments. It also suggests that, the internal RL, both RS repeats, and US are on a single *Eco*RI fragment. Since the digested DNA was run in 1% agarose at 2.5V/cm for only 16h, the 25 kb and 31 kb fragments were close to each other as single big thick band (Fig. 4, arrow on panels ICP0, ICP4 and US1), however they are seen separately in 0.7% agarose gel run at 2.5V/cm for 24 h (Fig. 3, bands at 25kb and 31kb).

The terminal *Bam*HIa and *Bam*HIb sites are 4496 bp and 4450 bp from the ends of the LHV-4 genome (Figs. 1C and 1D), respectively. In addition, there are *Bam*HI restriction sites at the beginning of RS; therefore, two fragments at 5526 bp and 5543 bp containing the IRL should be produced following *Bam*HI digestion (Fig. 1C and 1D). When the ICP0 probe was used, only *Bam*HI fragments ranging from 4 to 5.5 kb were hybridized which were close to the predicted size. When ICP4 probe was used, only a 19 kb *Bam*HI fragment was hybridized. Probe UL56, which is located between the two *Bam*HI sites at 95979 and 98273 downstream of the last *Eco*RIb site, hybridized to the expected 2294 bp *Bam*HI fragment. When the US1 probe was used, only the 19 kb *Bam*HI fragment was hybridized as the ICP4 probe did. Again, this suggests that both RS and US are presenton the same *Bam*HI fragment. These results again agreewith predictions based on the assembled DNA sequence. It was also observed that the ICP0 probe hybridized to the heterogeneous *Bam*HI fragments between 4 to 6 kb (Fig. 4), which suggests that the inverted repeats may harbor a repeat array that is heterogeneous in the number of repeats near the end of the genome. In addition, it is possible that the end of genome is heterogeneous. No report is available that explains how the end of the viral genome is protected from being recognized or repaired as "damaged DNA." It is also possible that during replication, some of the newly synthesized genomes were not fully protected and were mistakenly recognized as damaged DNA, resulting in heterogeneous terminal sequences.

Genome characteristics: The size of the LHV-4 genome was previously estimated by pulse field gel electrophoresis to be between 112 and 130 kbp (Jin et al., 2008a), and the DNA sequence analysis and mapping resulted in a size within this range. The average genome size of simplexviruses is about 150 kb. Most simplexviruses have a UL at about 118 kb, whereas the LHV-4 UL is only about 96 kb. In addition, RL is about 35% the size of RL of primate simplexviruses (Table 2) and RS is about 30% smaller than RS of primate simplexviruses. Taken together these contribute to the 16 kb difference in observed genome size between LHV-4 and other simplexviruses. Although LHV-4 has a smaller genome, orthologs of 69 open reading frames (ORFs) known to encode proteins in *Simplexvirus* were predicted from the LHV-4 genome by Glimmer3 software (Table 1 and Fig. 2). The comparison of LHV-4 is made to HHV-2 and CeHV-2, since LHV-4 is a little closer to them based on distance matrix analysis of UL40 (Table 3). Both LHV-4 ICP0 and ICP4 are smaller than those of primate

simplexviruses (Tables 1 and 2). LHV-4 ICP0 shares 42.7% and 43.2% sequence identity with CeHV-2 ICP0 and HHV-2 ICP0, respectively. The conserved Ring domain of ICP0 shares 82% identity with the same region- in CeHV-2 and HHV-2. LHV-4 ICP4 has 51.9% and 45.1% homology to CeHV-2 ICP4 and HHV-2 ICP4, respectively; however, LHV-4 ICP4 also contains conserved herpes ICP4 N terminal and C terminal regions. Both conserved N terminal and C terminal regions of LHV-4 ICP4 have 68-76% and 76-77% homology to those of CeHV-2 ICP4 and HHV-2 ICP4, respectively (Table 1). ~~Although latency associated transcript (LAT) genes are not fully characterized from LHV-4, transcripts from the RL can be detected by RT-PCR in the trigeminal ganglion collected at 48 days post-infection (data not shown), which suggests that LAT-like RNA is also present during latency.~~ Most of the UL genes of LHV-4 are ~~about~~ 6-20% shorter than their orthologs in other simplexviruses, except for some genes of tegument proteins (UL14, UL51), a major capsid protein (UL38), genes involved in DNA packaging (UL15, UL29, UL31, UL32, UL33), structural proteins (UL49A, US2, US4, US6), and regulatory genes (UL48, US 3) (Table 1). Similar to other simplexviruses, the gene encoding the large tegument protein (UL36) is the largest gene of LHV-4, yet it is still significantly smaller (8.7% to 10.2%) than that of CeHV-2 and HHV-2 (Table 1).

In addition to having smaller predicted ORFs, the LHV-4 genome has smaller intergenic regions when compared to HHV-2 (Supplemental Table 1). This contributes to almost a 4.7 kb reduction in genome size. The predicted proteins encoded by LHV-4 UL genes have 40-79% amino acid sequence identity compared to those of CeHV-2 and HHV-2 (Table 1) and 50-85% identity with those of BHV-2 that have been sequenced so far (Table 1). The US is slightly shorter than that of CeHV-2 and HHV-2. The LHV-4 US has 11 ORFs designated US1 through US12 (Table 1) (A homolog of the US5 ORF is not present). The predicted proteins from LHV-4 US share only 32-63% identity with those of CeHV-2 and HHV-2 (Table 1). Because genes in the US encode mostly glycoproteins that are located on the surface of the virion envelope and play major roles in virus entry and immune evasion, it is not surprising that these LHV-4 glycoproteins are less conserved than other genes in UL because they are under selective pressure by the host immune system.

An ortholog of the US5 was not identified in the LHV-4 genome, which may not be so surprising. Within primate simplexviruses, the US5 is less conserved and has only 26.5% amino acid identity between cercopithecine herpesviruses 1 (CeHV-1) and HHV-1 (Perelygina et al., 2003). The US5 encodes glycoprotein J (gJ) (Ghiasi et al., 1998) which may have anti-apoptosis activity during infection (Jerome et al., 2001). ~~Interestingly, LHV-4 does not replicate well in Vero cells, and apoptosis is induced by 48 hpi in these cells (data not shown).~~ An ortholog of ICP34.5 (RL1) was also not identified; ICP34.5 is required for neuronal virulence for HHV-1 (Bolovan et al., 1994; Orvedahl et al., 2007; Perng et al., 1995; Thompson and Stevens, 1983). The lack of ICP34.5 has been observed in some simplexviruses from primates, such as cercopithecine herpesviruses (CeHV-1, CeHV-2, and CeHV-16) (Perelygina et al., 2003; Tyler et al., 2005; Tyler and Severini, 2006). This supports the hypothesis that a different pathogenic mechanism may have developed in human simplexviruses after their divergence from monkey simplexviruses (Tyler et al., 2005).

The LHV-4 has an "a" sequence predicted to be 379 bp, and its repeat profile is different from other simplexvirus "a" sequences (Umene et al., 2008). It has six direct repeat arrays (DR) of 20 bp, two long DRb of 59 bp and two DRc of 18 bp, which contains the predicted *Pac*2 sites (CGCCGCG) (Fig. 5). The LHV4 "a" sequence does not have long unique sequences, such as Ub and Uc, seen in other simplexviruses (Fig. 5B). Since there are many GC repeats in the "a" sequence, it made the final assembly of this region in the LHV-4 genome very difficult. Although there may be two "a" sequence present in the RL-RS junction based on the Southern hybridization data obtained with ICP0 probe (Fig. 4), only one copy of the "a" sequence was included in the final genome assembly, which may cause the whole genome to be slightly smaller than the actual size. The overall nucleotide composition of the LHV-4 genome is about 67% G+C. This study shows that LHV-4 has the smallest simplexvirus genome characterized to date.

Gene conservation in simplexviruses: The most conserved genes in the LHV-4 genome are the helicase-primase subunit (UL5); the DNA packaging terminase subunit 2 (UL15 exon 2); the capsid protein (UL18) and major capsid protein (UL19); and the small subunit of ribonucleotide reductase (UL40). The amino acid sequences of these genes have above 70% identity to those of

primate simplexviruses (Table 1). The UL15 exon2 of LHV-4 shares approximately 79% and 75.5% identity with that of HHV-2 and CeHV-2, respectively. The amino sequence of UL40, ribonucleotide reductase, has about 75% identity to CeHV-2 and HHV-2, and 83% identity to BHV-2.

Phylogenetic analyses of LHV4 glycoprotein B (gB): Orthologs of gB (UL27) have been identified in all subfamilies of the *Herpesviridae*. Studies involving HHV-1 gB have shown that it plays essential roles in virus attachment, penetration, membrane fusion, and cell-to-cell spreading (Cai et al., 1988; Pereira, 1994). In addition, gB serves as a major antigenic determinant of host tropism (Gerdts et al., 2000). Analyses of the amino acid sequences of gB orthologs of herpesviruses show that the carboxy-terminal hydrophobic region is conserved. LHV-4 encodes a predicted ortholog of gB of 878 aa, which has 60% or more identity to various simplexviruses (Table 1).

Although it is thought that many viruses including members of the *Herpesviridae* undergo host dependent evolution, the relationship of simplexviruses supports a more complex phylogeny (Fig 6). The simplexviruses clearly form a distinct lineage with a probability value of 1, distinguishing them from a representative example of a mardivirus (GaHV-2) and a distant betaherpesvirus, tupaia herpesvirus 1 (tree shrew) (THV-1). Within the simplexviruses, the cercopithicine (old world monkeys) (CeHV-2), CeHV-16 , macacine (McHV-1) and human herpesviruses (HHV-1 and HHV-2) form a well-supported lineage. However, other primate viruses such as the new world ateline (AtHV-1) and saimiriine (SaHV-1) monkey lineage, are distinct from the other simplexviruses. In contrast, BHV-2, LHV-4 and MaHV-1 are grouped with the old world human-cercopithicine lineage with a high degree of confidence. The phylogenetic analyses might reflect a complex evolution that could have involved cross-infection between disparate species (marsupials, rabbits, cattle, and primates) to result in the pattern observed.

LHV-4 has not been found to be associated with any major disease outbreak in wild lagomorphs or other animals (Jin et al., 2008b). However, in addition to the original outbreak in domestic rabbits, in Alaska in 2006, an isolated LHV-4 infection has been recently reported in a pet rabbit

from Ontario, Canada (Brash et al., 2010). It is possible that LHV-4 does not normally produce disease in leporid populations. External stressors may reactivate the virus and contribute to clinical disease. The infection in rabbits caused by LHV-4 is similar to ocular infections caused by HHV-1 in humans. LHV-4 infection in rabbits, therefore, may be useful as a natural host model for herpesvirus latency reactivation studies.

Materials and methods

Cell culture and virus isolation: Rabbit kidney cells (RK) (RK-13B, American Type Culture Collection, Rockville, MD) were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS) (Invitrogen), penicillin (100 U/ml), and streptomycin (100 µg/ml) (Sigma-Aldrich, Inc.,) at 37°C with 5% CO_2 in a humidified incubator. LHV-4 was initially isolated from frozen skin samples with hemorrhagic lesions from an affected rabbit from the 2006 outbreak in Wasilla, Alaska (Jin et al., 2008b). We named this LHV-4 strain "Wasilla".

Purification of viral DNA: Viral DNA was obtained from the plaque-purified virus that was only passed once to avoid genetic variations between multiple passages. No difference in DNA sequence was observed in PCR products amplified from the plaque purified virus and the second passage virus. LHV-4 was propagated in RK cells maintained in DMEM supplemented with 5% calf serum and antibiotics as above. Confluent cell monolayers were infected with plaque-purified virus at an MOI of 0.1. Viral DNA was extracted from either purified virions or from purified intracellular nucleocapsids as previously described (Jin et al., 2000).

Determination of the LHV-4 genome size by Field Inversion Gel Electrophoresis (FIGE). Purified LHV-4 virions were washed once with PBS, and mixed with 1% low-melting-temperature agarose and poured into a plug mold apparatus. Agarose plugs were treated with 10 mM Tris-HCl (pH 8.0), 100 mM EDTA, 1% *N*-lauroyl sarcosine, and 200 µg/ml proteinase K at 50°C overnight. The plugs were then washed several times with 10 mM Tris-HCl (pH 8.0) and inserted into the loading wells of a 1% agarose gel in 0.5x TBE buffer (45 mM Tris-borate, pH 8.0, 1 mM EDTA). Viral DNA were separated by FIGE using an MJ Research PPI-200 programmable pulse inverter with program 4 (initial reverse time, 0.05 min; reverse increment,

0.01 min; initial forward time, 0.15 min; forward increment, 0.03 min; number of steps, 81; reverse increment, 0.001 min; forward increment, 0.003) run at 8 V/cm for 17 h at 4°C. The Mid-Range PFG marker I (New England Biolabs) was used as a DNA size marker.

454 sequencing: Sample preparations for 454 sequencing were carried out using protocols provided by the manufacturer. The viral genome, total 5μg of purified viral DNA, was nebulized to produce fragments less than 800 bp before sequencing. DNA was sequenced by the GS FLX+ System from 454 Life Sciences/Roche.

Primers and PCR amplification: Selection of primers for LHV-4 sequence amplification was based on the DNA sequences assembled from the 454 sequence data. Numerous PCR primers were used to fill gaps and verify portions of the LHV-4 genome assembly (Supplemental Table. 2). PCR amplification with LHV-4 specific primers was performed as follows: a 25 μl reaction solution containing 1X Pfx amplification buffer (Invitrogen), 1X PCR Enhancer solution (Invitrogen), 0.5 μM $MgSO_4$, 0.4 μM dNTP's, 0.4 μM primers (Forward and Reverse), 1.0 U of Platinum Pfx DNA polymerase (Invitrogen), and 0.01-0.1 μg of viral DNA, was subjected to 94°C for 2 min, 30 cycles of 94°C for 30 s, 55°C for 45 s, and 72°C for 45 s, followed by a 5 min elongation reaction at 72°C after the final cycle.

PCR DNA Sequencing: PCR products were sequenced directly following purification with a ChargeSwitch PCR Clean-Up kit (Invitrogen). All sequencing was carried out by the Center for Gene Research and Biocomputing at Oregon State University using an ABI Prism®3730 Genetic Analyzer with a BigDye® Terminator v. 3.1 Cycle Sequencing Kit employing ABI Prism®3730 Data Collection Software v. 3.0 and ABI Prism® DNA Sequencing Analysis Software v. 5.2.

***De novo* assembly:** *De novo* assembly of the UL and US region of the LHV-4 genome was primarily performed using 454 Newbler *de novo* assembler (version 2.5). 65597 sequence reads were assembled with Newbler software (454 Life Technology) into 553 contigs in total with a maximum length of 97095 and N50 of 17890. Long contigs were searched against known herpesvirus genomes at NCBI and compared with significant expected values to other simplexviruses. The repetitive, reiterated contigs within the IR region were also compared to other sequenced viruses and used as a template to assemble larger contigs with lower coverage at

bridge points. The MIRA program provided algorithmic parameters to work with repetitive sequence data (Chevreux, 1999). In addition, 454 data was combined with Sanger sequence data from PCR reactions to assemble the inverted repeat sequence. Overlapping contigs were bridged using custom computer scripts to incorporate raw sequence reads onto the overlaps and imported into the Geneious software platform (Biomatters Ltd). Contigs were also compared to longer assemblies using the Cap3 assembly package with agreement (Huang and Madan, 1999). Basic alignment and parsing was facilitated by the Oregon State University Center for Genome Research and Biocomputing (OSU CGRB) server clusters (http://bioinfo.cgrb.oregonstate.edu/index.html).

Comparative Genomics: LHV-4 ORFs and genes were predicted by using Gilmmer3 v3.02 (Delcher, 2007) and GeneMarkS (Besemer, 2001). The gene functions were annotated by blastp searching against the NCBI non-redundant protein database (nr). Protein families were determined by pfam database search. Unique LHV-4 nucleotide sequences were compared by Smith-Waterman local alignment percentage identity to annotated alphaherpesviruses. Sequences were translated and aligned using standard parameters in ClustalW. Aligned, non-overlapping sequence ends were culled. Phylogenetic analysis of gB of LHV-4 was compared with betaherpesvirus tupaia herpesvirus 1 (THV-1) (NC_002794), selected simplexviruses including AtHV-1 (AAA43839.1); BHV-2 (AAA46053.1); CeHV-2 (AY714813.1), HHV-1 (ADG45180.1); HHV-2 (ADG45180.1), MaHV-2 (AAD11960.1); McHV-1 (AF061754); CeHV-16 (U14662.1); SaHV-1 (AAA43841); and a mardiviris GaHV-2 (JQ314003). The phylogenetic tree was created by Bayesian Phylogenetic method using RAxML program through the OSU CRGB server, and the tree was viewed using FigTree viewing software.

Southern Blotting: Genomic DNA was digested with *Hind*III, *Bam*HI, and *Eco*RI respectively, electrophoresed through either 1.0% or 0.7% agarose gel, and transferred to a nylon membrane (Jin et al., 2000). The genomic DNA was then UV cross-linked to the membrane and probed with digoxigenin-labeled PCR products as shown in Fig. 1. To make digoxigenin-labeled PCR products, digoxigenin-labeled deoxynucleoside triphosphates (Roche Diagnostics, Indianapolis, Ind.) were added to the PCR mixtures using primers listed in Supplemental Table 3 for each probe. The digoxigenin-labeled PCR products were then cleaned with a ChargeSwitch PCR Clean-Up kit (Invitrogen) before using. After incubation with the probe, membranes were

washed with 0.1% sodium dodecyl sulfate and 10% 20x SSC (1X SSC, 0.15 M NaCl plus 0.015 M sodium citrate) before incubation with an anti-digoxigenin antibody conjugated with alkaline phosphatase. The membrane was then developed by incubation with a chemiluminescent peroxidase substrate (Roche). The blots were exposed to film, and the molecular masses of the resulting bands were determined by using a digoxigenin- labeled DNA Molecular Weight Marker VII (Roche). The membrane was stripped with 0.1% SDS and 0.2M NaOH before probing with a different probe.

Accession Number

The sequences in this study have been deposited in GenBank database (Accession No. JQ596859).

Acknowledgments

This work was supported by NIH RO3AI080999 and the College of Veterinary Medicine at Oregon State University. We thank Drs. Adam Vanarsdall for the pulse field gel data and Aimee Reed for proof reading the manuscript.

Disclosure statement

The authors declare that they have no conflict of interest.

Figures and Legends

Fig. 1. Schematic of isomeric forms of the LHV-4 genome and location of the DNA probes used in hybridization. UL and US indicates the long and short unique regions, respectively. The solid arrow represents the RL, the open arrow represents the RS. The "a" sequence is the boundary between RL and RS. Locations of DNA probes are shown in striped boxes (UL1, UL56, ICP0, ICP4, and US1) above the maps and are drawn in an approximate scale with respect to their genome locations. A and B) Two possible *Eco*R I sites near the end of the genome containing RL and RS. The *Eco*RIa and b sites are near UL1 *and* UL56, respectively. C and D)

Two possible *Bam*HI sites near the end of RL. The *Bam*HI a and b sites are near UL1and UL56, respectively.

Fig. 2. Gene layout of LHV-4 genome. The locations of predicted protein-coding ORFs are shown as defined as colors in the key. The number above each ORF corresponds to the nt location on the genome map. The open boxes are inverted repeats (RL and RS) flanking UL and US. UL and US are between RL and RS.

Fig. 3. Southern blot analysis of LHV-4 genomic structure with only ICP0 and ICP4 probes. LHV-4 viral DNA was digested with *Bam*HI and *Eco*RI. After processing, the blot was hybridized with Dig-labeled ICP0 DNA probe. The membrane was stripped and re-probed using Dig-labeled ICP4 DNA probe. The digested LHV-4 DNA was electrophoresed through a 0.7% agarose gel run at 2.5V/cm for 24 h. Lanes V: virion, N: nucleocapsid DNA, M: DIG-labeled DNA Molecular Weight Marker VII (Roche Applied Science).

Fig. 4. Southern blot analysis of LHV-4 genomic structure with multiple probes. LHV-4 viral DNA was digested with *Bam*HI and *Eco*RI. After processing, the blots were hybridized with Dig-labeled ICP0 DNA probe. The membrane was stripped and reprobed using Dig-labeled ICP4, US1, UL1, and UL56 DNA probes. Digested LHV-4 DNA was electrophoresed through a 1.0% agarose gel at 2.5 V/cm for 16 h. Lanes V: virion DNA, N: nucleocapsid DNA, M: DIG-labeled DNA Molecular Weight Marker VII (Roche Applied Science). The marker sizes are indicated in bp. Note: the UL1 probe was hybridized to a different blot obtained under the same conditions as the others. Open arrow: *Eco*RI fragments at either 25kb or 31 kb. Solid arrow: *Eco*RI fragment at either 4kb or 10kb. Arrow: *Eco*RI fragments at 25 kb and 31kb.

Fig. 5. Schematics of the "a" sequence. A) Location of direct repeats (DR) within the "a" sequence of LHV-4. B) Consensus structure of the "a" sequence seen in HHV-1. DR1: direct repeat. Ub: unique b sequence. DR2 array: various copy number of DR2 elements. DR 4: direct repeat. C) Nucleotide sequence of the DRa, DRb, DRc and Pac 2 sites. Pac 2 site is within DRc of LHV4, Uc of HHV1, respectively.

Fig.6. A phylogenetic tree of selected simplexvirus gB. The tree was rooted with a betaherpesvirus tupaia herpesvirus 1 (THV-1) (NC_002794), and LHV-4 was compared to selected simplexviruses including AtHV-1 (AAA43839.1); BHV-2 (AAA46053.1); CeHV-2 (AY714813.1), HHV-1 (ADG45180.1); HHV-2 (ADG45180.1), MaHV-2 (AAD11960.1); McHV-1 (AF061754); CeHV-16 (U14662.1); SaHV-1 (AAA43841); and a mardiviris GaHV-2 (JQ314003). The scale bar represents the genetic distance (nucleotide substitutions per site). 1 stands for the bootstrap value of 100. For details see Materials and Methods.

Supplemental Fig. 1. Average quality score per sequence. Reads with low scores were filtered out of initial drafts at arbitrary scores through several contig draft iterations.

Supplemental Fig. 2. Endonuclease restriction analysis of LHV-4 with *Bam*HI, *Eco*RI and *Hind*III. A) LHV4 DNA digests run in 1% of agarose at 2.5V/cm for 16h. B) Digested DNA run in 0.7% agarose gel run at 2.5V/cm for 24 h. Lanes V: virion DNA, N: nucleocapsid DNA, MW1: 1 kb plus DNA ladder (Invitrogen), MW2: DNA Molecular Weight Marker VII, DIG-labeled DNA Molecular Weight Marker VII (Roche Applied Science), MW3: Lambda Mix Marker, 19 (Fermentas). The marker sizes are indicated in bp.

Supplemental Fig. 3. Analysis of LHV-4 DNA by FIGE. MW: Mid Range DNA maker (New England Biolabs). LHV-4: viral DNA. For details see Materials and Methods.

References:

Bolovan, C.A., Sawtell, N.M., Thompson, R.L., 1994, ICP34.5 mutants of herpes simplex virus type 1 strain 17syn+ are attenuated for neurovirulence in mice and for replication in confluent primary mouse embryo cell cultures. J Virol 68, 48-55.

Brash, M.L., Nagy, E., Pei, Y., Carman, S., Emery, S., Smith, A.E., Turner, P.V., 2010, Acute hemorrhagic and necrotizing pneumonia, splenitis, and dermatitis in a pet rabbit caused by a novel herpesvirus (leporid herpesvirus-4). Can Vet J 51, 1383-1386.

Cai, W.H., Gu, B., Person, S., 1988, Role of glycoprotein B of herpes simplex virus type 1 in viral entry and cell fusion. J Virol 62, 2596-2604.

Chevreux, B., Wetter, T. and Suhai, S. 1999. Genome Sequence Assembly Using Trace Signals and Additional Sequence Information
In Proceedings of the German Conference on Bioinformatics (GCB)

Gerdts, V., Beyer, J., Lomniczi, B., Mettenleiter, T.C., 2000, Pseudorabies virus expressing bovine herpesvirus 1 glycoprotein B exhibits altered neurotropism and increased neurovirulence. J Virol 74, 817-827.

Ghiasi, H., Nesburn, A.B., Cai, S., Wechsler, S.L., 1998, The US5 open reading frame of herpes simplex virus type 1 does encode a glycoprotein (gJ). Intervirology 41, 91-97.

Guliani, S., Smith, G.A., Young, P.L., Mattick, J.S., Mahony, T.J., 1999, Reactivation of a macropodid herpesvirus from the eastern grey kangaroo (Macropus giganteus) following corticosteroid treatment. Vet Microbiol 68, 59-69.

Hesselton, R.M., Yang, W.C., Medveczky, P., Sullivan, J.L., 1988, Pathogenesis of Herpesvirus sylvilagus infection in cottontail rabbits. Am J Pathol 133, 639-647.

Huang, X., Madan, A., 1999, CAP3: A DNA sequence assembly program. Genome Res 9, 868-877.

Jerome, K.R., Chen, Z., Lang, R., Torres, M.R., Hofmeister, J., Smith, S., Fox, R., Froelich, C.J., Corey, L., 2001, HSV and glycoprotein J inhibit caspase activation and apoptosis induced by granzyme B or Fas. J Immunol 167, 3928-3935.

Jin, L., Lohr, C.V., Vanarsdall, A.L., Baker, R.J., Moerdyk-Schauwecker, M., Levine, C., Gerlach, R.F., Cohen, S.A., Alvarado, D.E., Rohrmann, G.F., 2008a, Characterization of a novel alphaherpesvirus associated with fatal infections of domestic rabbits. Virology 378, 13-20.

Jin, L., Schnitzlein, W.M., Scherba, G., 2000, Identification of the pseudorabies virus promoter required for latency-associated transcript gene expression in the natural host. J Virol 74, 6333-6338.

Jin, L., Valentine, B.A., Baker, R.J., Lohr, C.V., Gerlach, R.F., Bildfell, R.J., Moerdyk-Schauwecker, M., 2008b, An outbreak of fatal herpesvirus infection in domestic rabbits in Alaska. Vet Pathol 45, 369-374.

Johnson, M.A., Whalley, J.M., 1990, Structure and physical map of the genome of parma wallaby herpesvirus. Virus Res 18, 41-48.

Mahony, T.J., Smith, G.A., Thomson, D.M., 1999, Macropodid herpesviruses 1 and 2 occupy unexpected molecular phylogenic positions within the Alphaherpesvirinae. J Gen Virol 80 (Pt 2), 433-436.

McGeoch, D.J., 1989, The genomes of the human herpesviruses: contents, relationships, and evolution. Annu Rev Microbiol 43, 235-265.

McGeoch, D.J., Cook, S., 1994, Molecular phylogeny of the alphaherpesvirinae subfamily and a proposed evolutionary timescale. J Mol Biol 238, 9-22.

McGeoch, D.J., Rixon, F.J., Davison, A.J., 2006, Topics in herpesvirus genomics and evolution. Virus Res 117, 90-104.

Orvedahl, A., Alexander, D., Talloczy, Z., Sun, Q., Wei, Y., Zhang, W., Burns, D., Leib, D.A., Levine, B., 2007, HSV-1 ICP34.5 confers neurovirulence by targeting the Beclin 1 autophagy protein. Cell Host Microbe 1, 23-35.

Pereira, L., 1994, Function of glycoprotein B homologues of the family herpesviridae. Infect Agents Dis 3, 9-28.

Perelygina, L., Zhu, L., Zurkuhlen, H., Mills, R., Borodovsky, M., Hilliard, J.K., 2003, Complete sequence and comparative analysis of the genome of herpes B virus (Cercopithecine herpesvirus 1) from a rhesus monkey. J Virol 77, 6167-6177.

Perng, G.C., Thompson, R.L., Sawtell, N.M., Taylor, W.E., Slanina, S.M., Ghiasi, H., Kaiwar, R., Nesburn, A.B., Wechsler, S.L., 1995, An avirulent ICP34.5 deletion mutant of herpes

simplex virus type 1 is capable of in vivo spontaneous reactivation. J Virol 69, 3033-3041.

Roizman, B., Pellett, P.E. 2001. Fields Virology. In The family Herpesviridae: a brief introduction, Knipe, D.M., Howley, P.M., eds. (Philadelphia, Lippincott Williams & Wilkins), pp. 2381-2397.

Thompson, R.L., Stevens, J.G., 1983, Biological characterization of a herpes simplex virus intertypic recombinant which is completely and specifically non-neurovirulent. Virology 131, 171-179.

Tyler, S.D., Peters, G.A., Severini, A., 2005, Complete genome sequence of cercopithecine herpesvirus 2 (SA8) and comparison with other simplexviruses. Virology 331, 429-440.

Tyler, S.D., Severini, A., 2006, The complete genome sequence of herpesvirus papio 2 (Cercopithecine herpesvirus 16) shows evidence of recombination events among various progenitor herpesviruses. J Virol 80, 1214-1221.

Umene, K., Oohashi, S., Yoshida, M., Fukumaki, Y., 2008, Diversity of the a sequence of herpes simplex virus type 1 developed during evolution. Journal of General Virology 89, 841-841.

Figure-1

Click here to download Figure: Fig. 1 new.ppt

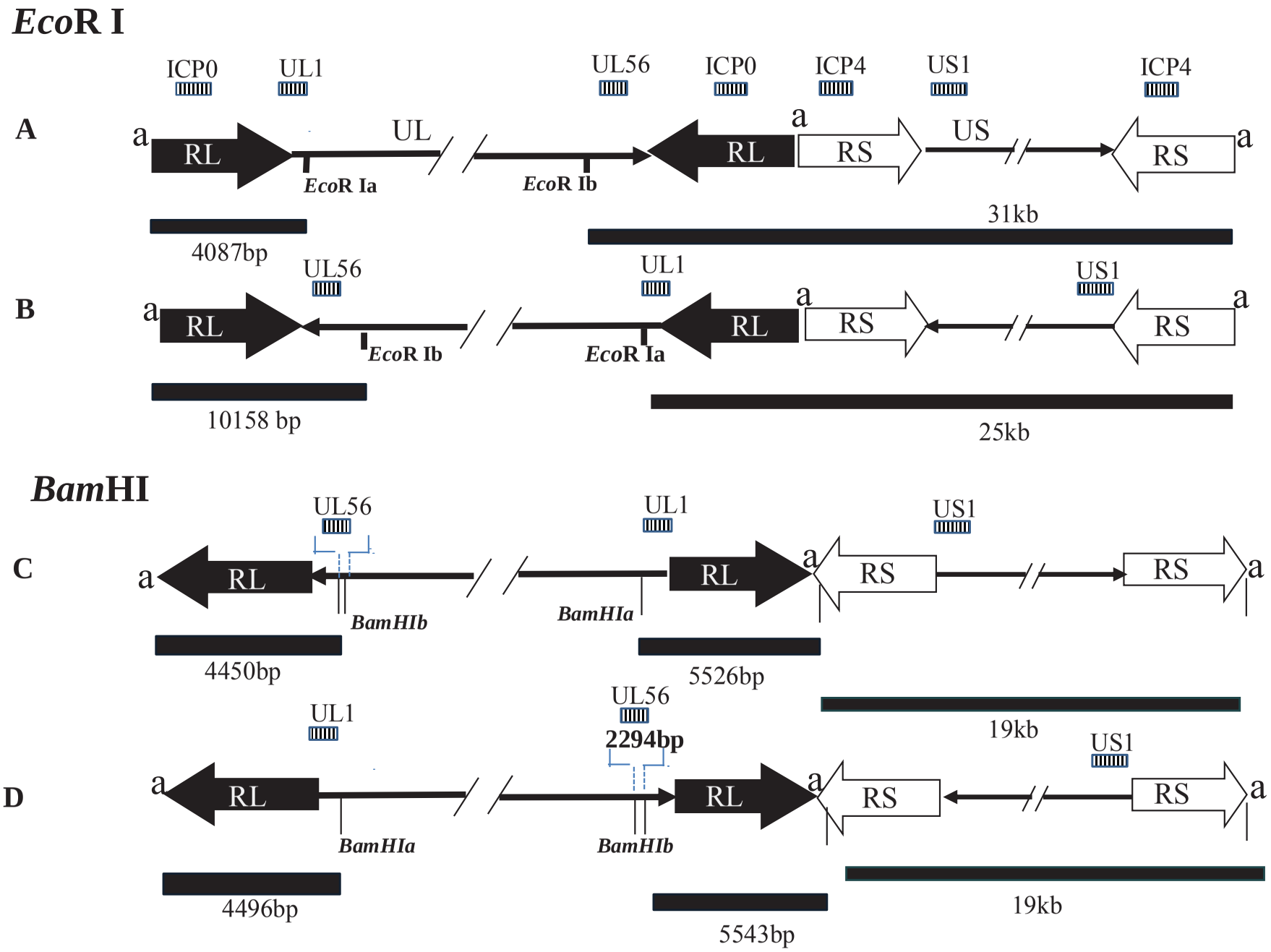


Figure-2

Click here to download Figure: Fig_2 070712.ppt

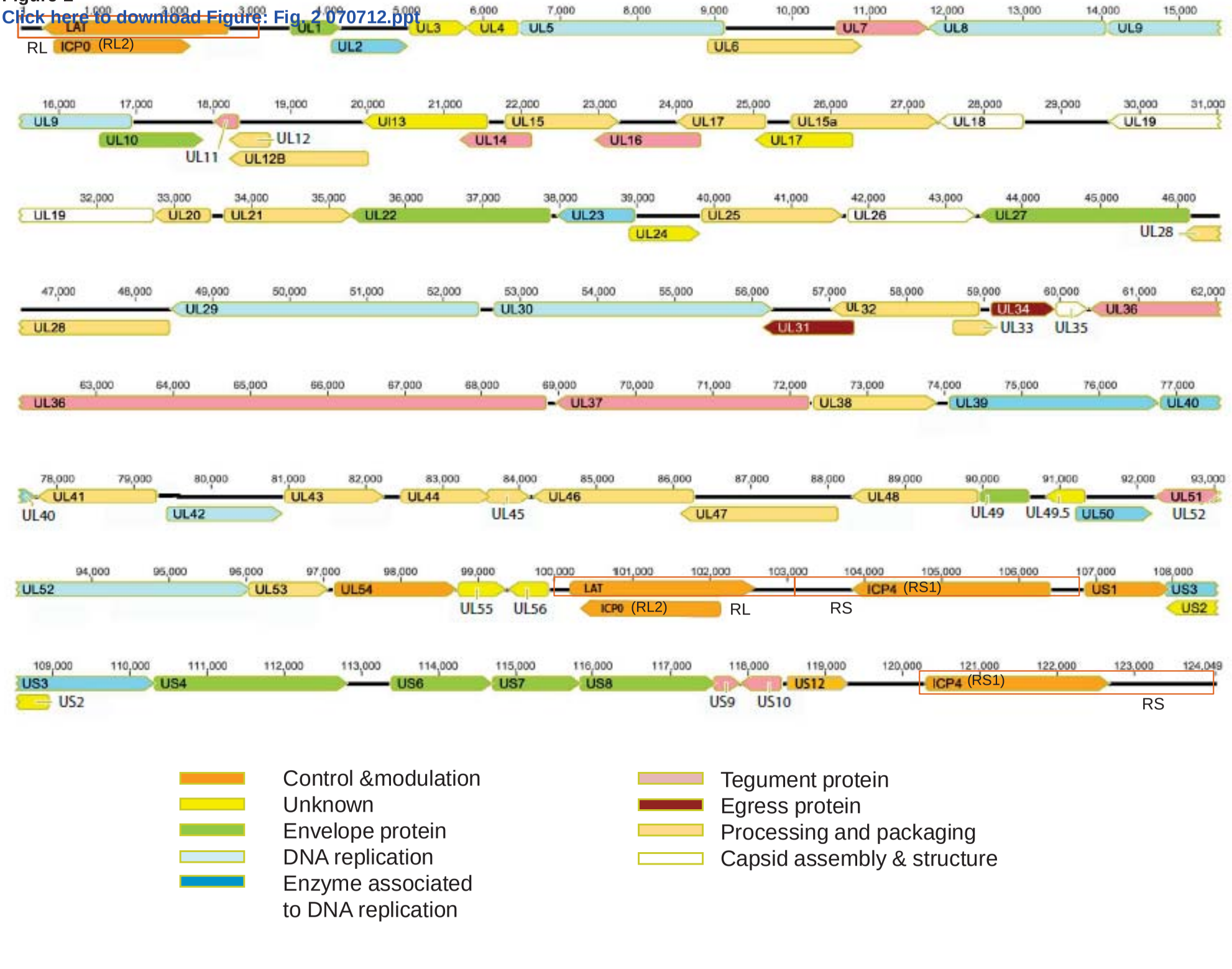


Figure-3

Click here to download Figure: Fig. new3.ppt

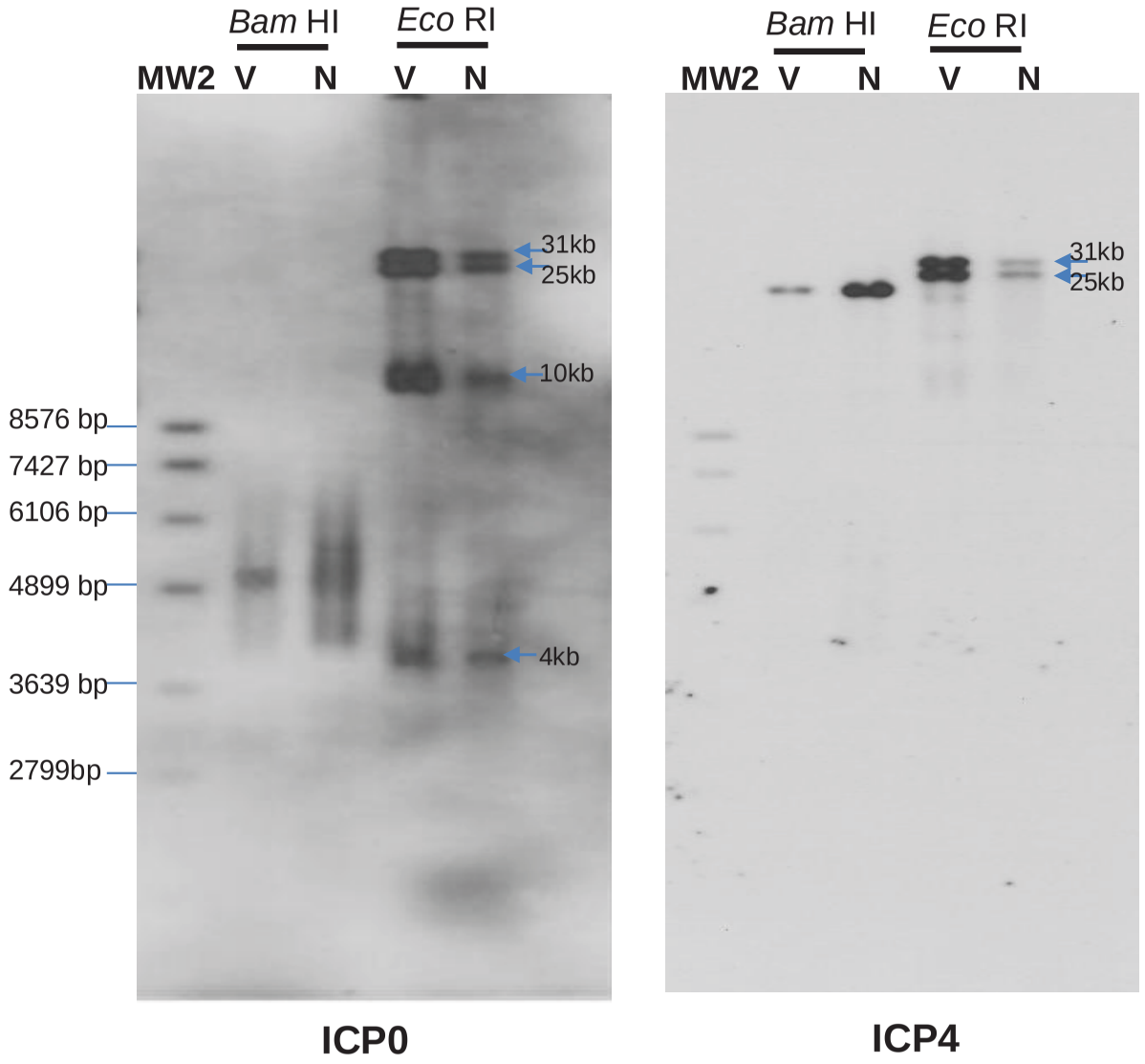


Figure-4

Click here to download Figure: Fig. new4.ppt

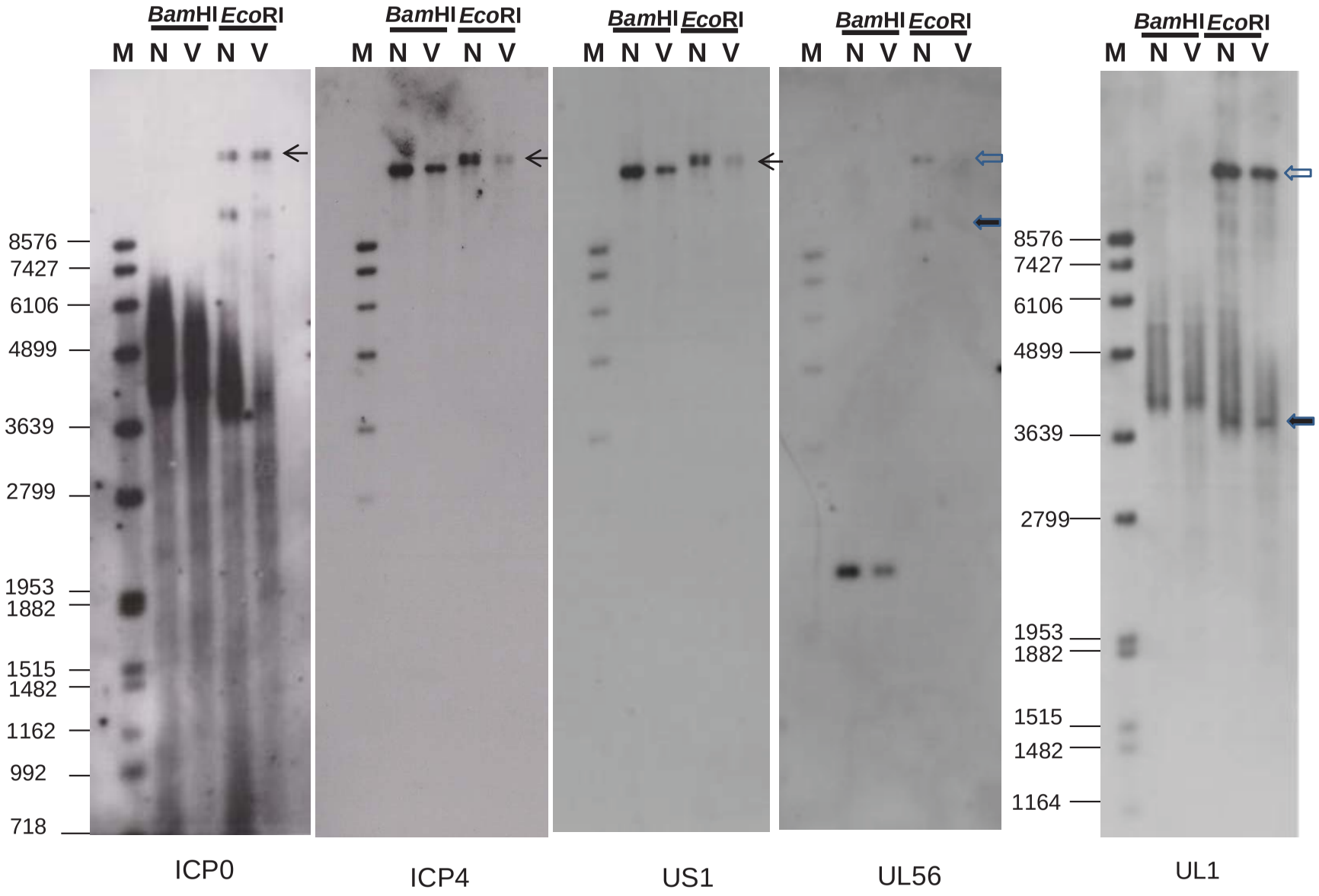
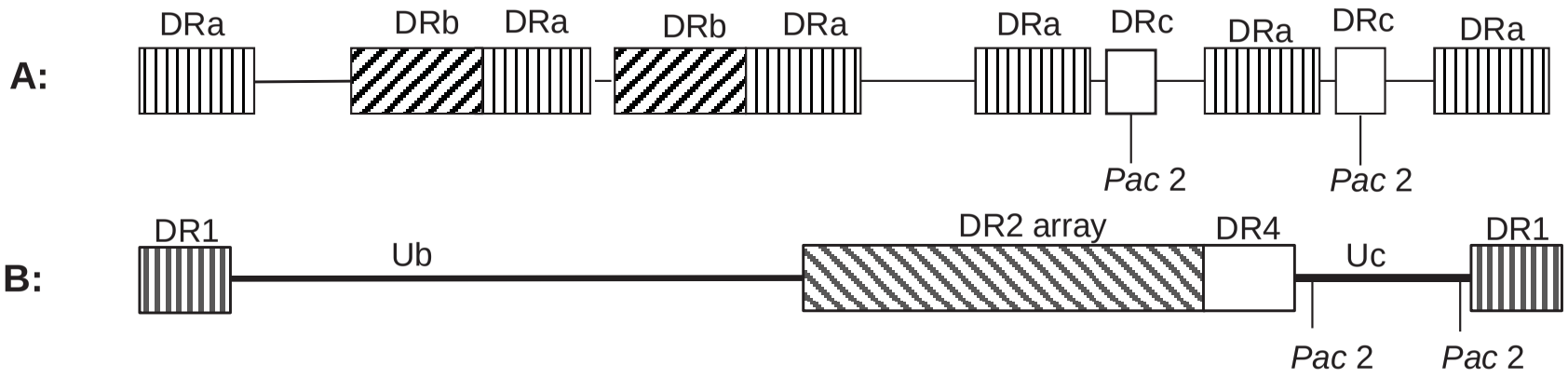


Figure-5

Click here to download Figure: Fig._5new.ppt



C:

DRa: TATGCAGATGGGGTTCCGC

DRb:GGCGGAAGGGAGGCCCCCGGGCCCCCCGTCCCACATCCGCC

DRc:CGCGGCGAATCGGGGCCC

Pac 2 sites (CGCCGCG)

Figure-6

Click here to download Figure: Fig. 6new.ppt

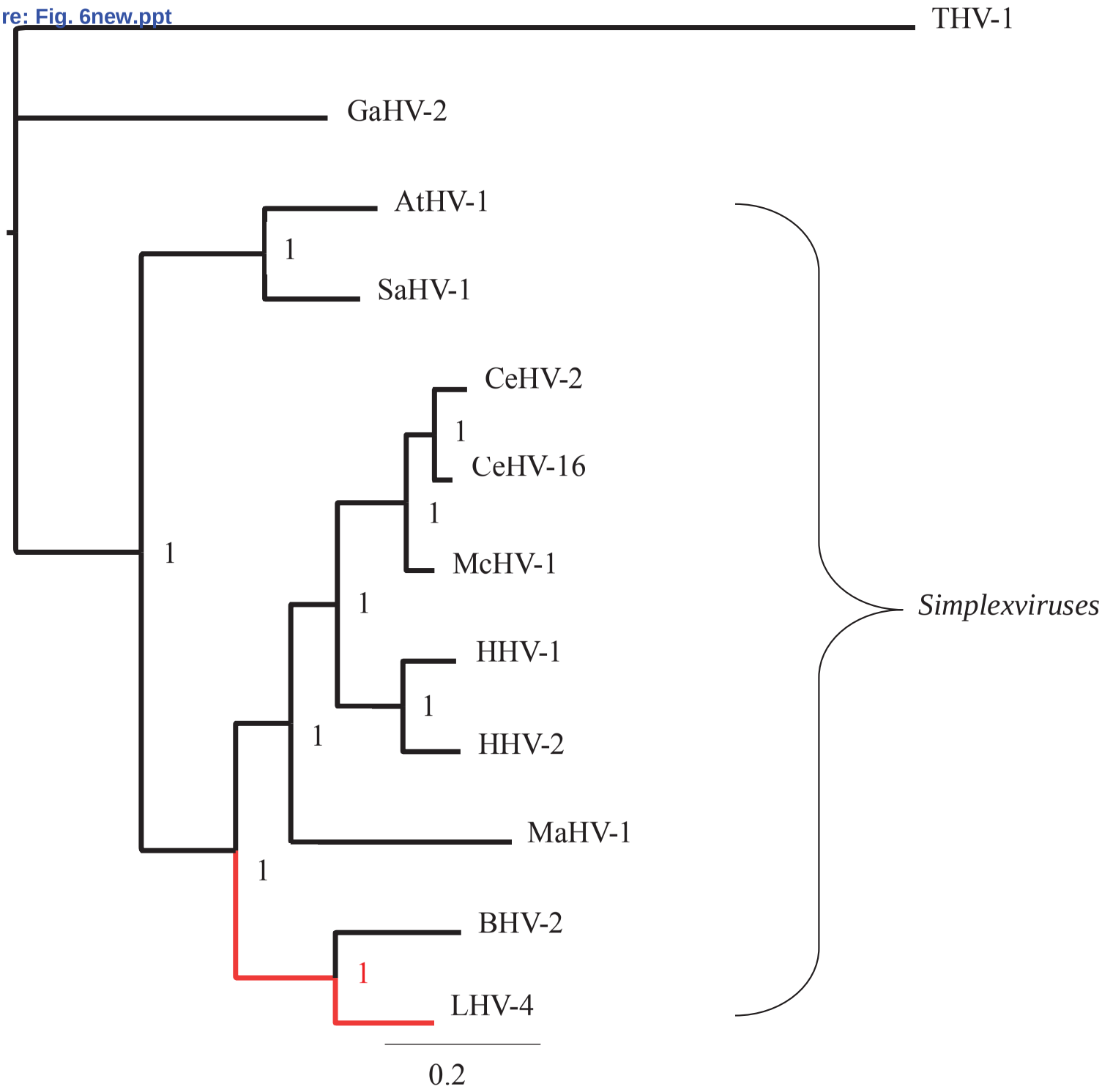


Table 1

Table 1. LHV4 genes for which predictions were made regarding functional properties and comparison with other simplexviruses

		LHV-4[a]	CeHV-2[a]	HHV-2[a]	BHV-2[a#]
Gene	**Predicted Functions****	aa	%*/ aa	%*/ aa	%*/ aa
RL2	immediate early protein	569	42.7/688	43.2/702	
orf7	Hypothetical protein	132			
orf8	Hypothetical protein	70			
UL1	glycoprotein L	156	53.6/232	54 /204	
UL2	uracil-DNA glycosylase	245	57.4/316	64/255	
UL3	nuclear phosphoprotein	202	55.6/227	62/233	70/213
UL4	nuclear protein	206	48.2/204	46/201	62/210
UL5	helicase-primase helicase subunit	885	78.3/875	74/881	
UL6	capsid portal protein	651	57.9/678	61/678	
UL7	tegument protein	283	46.4/296	48/296	
UL8	helicase-primase subunit	733	37.6/758	39/752	52/733
UL9	DNA replication-origin replication helicase	831	64.5/879	69/867	
UL10	glycoprotein M	410	45.2/450	51/467	
UL11	myristylated tegument protein	86	45.9/87	42/96	
UL12	alkaline exonuclease	577	53/615	49/620	
UL13	tegument serine/threonine protein kinase	504	51.7/514	54/518	
UL14	tegument protein	285	52.6/214	60/219	
UL15 exon1	DNA packaging terminase subunit 1	466	65.5/343	66/343	
UL15 exon2	DNA packaging terminase subunit 1	609	75.4/392	79/391	
UL16	tegument protein	347	47.7/362	49/449	
UL17	DNA packaging tegument protein	672	55/703	56/702	
UL18	capsid protein	314	72.4/317	70/318	80/316
UL19	major capsid protein	1025	70.9/1377	74/1374	84/1385
UL20	envelope protein	223	50.7/225	48/222	64.4/223
UL21	tegument protein	521	41.8/526	44/532	56/522
UL22	glycoprotein H	845	40.5/863	42/838	57/867
UL23	thymidine kinase	310	42.4/378	47/376	56/323
UL24	nuclear protein	258	51.8/260	50/281	64/265
UL25	DNA packaging tegument protein	580	59.3/577	65/585	72/579
UL26	capsid maturation protease	539	61.1/594	52/636	50/562
UL26.5	capsid morphogenesis; capsid	336	43.7/292	38.6/329	49/292

	scaffold protein				
UL27	glycoprotein B	878	68/885	66.7/901	79/917
UL28	DNA packaging terminase subunit 2	771	60.6/784	64/785	71/664
UL29	ICP8 single-stranded binding protein	1295	65.7/1196	74/1196	80/1186
orf54	Hypothetical protein	119			
UL30	DNA polymerase	1178	68/1226	65/1240	73/1211
UL31	nuclear egress laminar protein	375	66.9/304	70/305	
UL32	DNA packaging protein	612	57.6/590	62/598	
UL33	DNA packaging protein	159	54.7/135	59/130	
UL34	nuclear egress protein	258	64.9/266	69/276	
UL35	small capsid protein	105	54/114	56/112	
UL36	large tegument protein	2803	51/3070	50/3122	
UL37	tegument protein	1071	54.5/1271	59/1114	
UL38	triplex capsid protein VP19C	501	48.8/459	59/466	
orf73	Hypothetical protein	49			
UL39	ribonucleotidereductase subunit 1	890	63.8/975	61/1142	75/784
UL40	Ribonucleotide reductase subunit 2	314	74.4/325	75/337	83/313
UL41	tegument host shut off protein	484	51.9/484	54/492	66/487
orf79	Hypothetical protein	94			
UL42	DNA polymerase processivity factor	454	44.9/438	59/470	52/449
UL43	envelope protein	401	26.1/381	30/414	51/391
orf83	Hypothetical protein	42			
UL44	glycoprotein C	429	43.3/464	40/480	56/425
UL45	Envelope protein UL45	158	42.3/174	43/172	55/166
UL46	tegument protein VP11/12	632	45.3/680	49/721	
UL47	tegument protein VP13/14	676	46.9/674	51/696	
UL48	transactivating tegument protein VP16	523	52.6/488	53/490	
orf92	Hypothetical protein	99			
UL49	Tegument protein VP22	206	38.2/275	42.1/300	
UL49A	Glycoprotein N	97	50/78	33.3/87	
UL50	deoxyuridine triphosphatase	360	39.2/367	40/369	
UL51	tegument protein	239	66/228	58/244	
UL52	helicase primase subunit	1038	57/1053	56/1066	
UL53	glycoprotein K	334	53/335	56/338	
UL54	ICP27	502	30/510	47/512	
UL55	nuclear protein	181	40.9/191	47/186	65/193

UL56	membrane protein	150	27.8/226	38.5/235	
orf109	Hypothetical protein	150			
RL2	ICP0	569	42.7/688	43.2/702	
orf116	Hypothetical protein	600			
RS1	transcription activator,ICP4	771	51.9/1185	45.1/1277	
orf120	Hypothetical protein	682			
orf123	Hypothetical protein	144			
orf125	Hypothetical protein	86			
US1	ICP22	384	40/431	43/413	
US2	virion protein	325	48/276	43/291	
US3	protein kinase	784	57/454	61/481	
US4	glycoprotein G	808	36/606	33/699	
orf130	Hypothetical protein				
US6	glycoprotein D	403	60/395	63/393	
US7	Virion glycoprotein I	335	32.7/399	35/372	
US8	glycoprotein E	504	35/540	38/545	
US8.5	membrane protein	96	48.9/102	54.5/146	
US9	type 2 membrane protein; tegument-associated; localizes envelope proteins	88	32.4/91	39/89	
US10	virion protein	153	41/276	34/302	
US11	translational regulation	349	42.3/115	38/151	
US12	inhibits antigen presentation # TAP transporter inhibitor ICP47	133	59.2/78	42.1/86	
orf140	Hypothetical protein	86			
orf142	Hypothetical protein	47			
orf143	Hypothetical protein	369			
RS1	transcription activator,ICP4	771	51.9/1185	45.1/1277	

*percentage of pair ID;

** predicted functions derived from simplexviruses

[a] accession number for LHV-4,CeHV-2, HHV-2, and BHV-2 are JQ596859, AY714813, JN561323 and AAA46053, respectively. # BHV-2 genome has not been fully sequenced yet.

Table 2

□able 2. Comparison of genes in the inverted repeats □ithin □imple□viruses

strain	LHV4		HHV2		CeHV16		CeHV2	
	Length□bp	□C□	Length□bp	□C□	Length□bp	□C□	Length□bp	□C□
□C□0	1710	80	3321	76.3	3029	78	2684	78.9
□L	3159	77.4	9263	75.4	10103	78.9	9207	79.9
□C□4	2313	79.2	4097	81	3570	84	3560	84.1
□□	4541	75.8	6711	80.1	7428	82	6625	82.4

Table 3

□able 3. □istance □ atri□Analysis of □L40

	CeHV-16	CeHV-2	HHV-2	HHV-1	BHV-2	LHV-4
CeHV-16	100	95.88	77.57	77.40	73.39	70.49
CeHV-2		100	81.76	81.56	74.74	71.31
HHV-2			100	89.48	76.07	75.68
HHV-1				100	77.37	70.02
BHV-2					100	78.23
LHV-4						100

Supplementary Fig. 1

Click here to download Supplementary Material (To be Published): Supplemental Fig. 1.ppt

Supplementary Fig. 2

Click here to download Supplementary Material (To be Published): Supplemental Fig. 2.ppt

Supplementary Fig. 3

Click here to download Supplementary Material (To be Published): Sup Fig.3.ppt

Supplementary table -1

Click here to download Supplementary Material (To be Published): Supplemental Table 1.doc

Supplementary Table 2

Click here to download Supplementary Material (To be Published): Supplemetnal Table 2new.doc

Supplementary Table 3

Click here to download Supplementary Material (To be Published): Supplemental table 3.ppt

Feature Table

Click here to download Supplementary Material (Not to be Published): a_table.txt