WPI

Patterns of Viral Signal Peptide Sequences and their Corresponding Hosts

A Thesis Submitted to the Faculty of
WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the
Degree in Master of Science
in
Bioinformatics and Computational Biology

Submitted by:
Ann-Elizabeth Le

Submitted on:
April 2021

Approved by:
Professor Dmitry Korkin, PhD - Thesis Director/Advisor
Professor Fyodor Kondrashov, External Committee Member

This thesis report represents work of WPI graduate students submitted to the faculty as evidence of a degree requirement. WPI routinely publishes these reports on its web site without editorial or peer review. For more information about the projects program at WPI, see https://digital.wpi.edu/col

ABSTRACT

A signal peptide (SP) is a short sequence located in the N-terminal of prokaryotic and eukaryotic proteins that helps mediate targeting of newly synthesized proteins to the endoplasmic reticulum (ER). SPs are variable in length and amino acid composition, which suggests that ER targeting and other co-translational steps are affected by the SP itself (Burdukiewicz et al., 2018). Viruses typically stay in a host that can harbor them without harming them unless they problematically cross into a new species (Healthline, 2019). As we experience the COVID-19 pandemic, we realize more than ever the importance of studying potential viral signaling patterns, which may shed light upon predictive models for viral protein replication between species and hosts. In this exploratory study, we compared the viral SPs and their corresponding genes of seven viral families to host genomes using a semi-automated approach with Protein BLAST. Viruses have the ability to encode functions required for viral replication by completely relying on the protein synthesis machinery present of a host cell (Goulding, *Virus replication*). Therefore, we proposed three different mechanisms viruses use to exploit a host: (1) only the signal peptide is captured from a host gene, (2) a part of the host gene is captured, and (3) the full gene is captured for the virus to have a function. For these three proposed evolutionary mechanisms, we identified functional clusters of all viral SP/host gene matches and found both known and potentially novel functional clusters pertaining to viral signal peptides. We also constructed ancestral sequences to detect selective pressure within *Herpesviridae* and *Baculoviridae* in which we found that both viral and host SP have amino acid conservation through their corresponding extant and ancestral sequences. As the initial steps on the first comprehensive study of SP in virions while linking virus-host phylogeny, this research can enhance our understanding of viral-host evolution across viral species and families.

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to Professor Dmitry Korkin, for giving me the opportunity to partake in his lab's research from the beginning of this past summer and for being my advisor for the past five years, as well as the entire Korkin lab for their support. I want to extend my gratitude to Professor Fyodor Kondrashov from the Institute of Science and Technology in Austria for providing wonderful evolutionary insight and guidance during every weekly meeting despite the Austrian time zone difference. I would like to thank Hui Mao for creating the sequences from the Reference Viral Database for use in this project. I am also grateful to Jocelyn Petitto for helping me brainstorm different questions to tackle for this project and offering sound scientific advice in the very beginnings of this project. Last but not least, I am thankful for my family and friends who have endlessly supported me through the completion of this project and my degree.

TABLE OF CONTENTS

I. BACKGROUND

A. Infectious Diseases

Infectious diseases are caused by microscopic organisms − such as bacteria, viruses, fungi, or parasites − that breach the body's exterior and multiply to cause symptoms that can range from mild to deadly. Mild infections may involve fever and fatigue, which usually are usually treated with rest and over-the-counter medication, while some life-threatening diseases may require hospitalization (Infectious Diseases Society of America, n.d.). In particular, viral infections occur when a virus invades a host and uses components of that host's cells to help them multiply, which often damages or destroys infected cells. Although not all viral diseases are contagious, the majority of them are. Some of the most common respiratory viral illnesses include the flu, the common cold, and measles, all of which are spread from person to person through aerosolized transmission. Some viruses that are spread through bodily contact or fluids include human immuno-deficiency virus (HIV), herpes simplex virus, Ebola, and hepatitis. Another form of viral transmission can involve a virus travelling from one organism to another, which can cause diseases like Dengue Fever, Yellow Fever, Middle East Respiratory Syndrome (MERS), and Severe Acute Respiratory Syndrome (SARS) (Healthline, 2019). Viruses may travel from vertebrate animals to humans, which cause zoonotic diseases. Zoonoses comprise a large percentage of all newly identified diseases along with existing ones. Some diseases, such as HIV which originated in chimpanzees, begin as a zoonosis and later mutate into human-only strains. Zoonoses cause recurring outbreaks such as the Ebola virus, whereas others such as SARS-CoV-2 which has caused a global pandemic (WHO, 2020).

B. Treatments of Infectious Diseases

Prevention of infectious diseases involves safety protocols and guidelines for individuals and protective standards when dealing with the natural environment can span from a national or global scale when dealing with methods of detection, prevention, and treatment. One of the best preventative measures for infectious diseases are vaccines that contain an agent that resembles the disease-causing microorganism. This foreign antigen goes on to stimulate the body's adaptive immune system so that the body is able to recognize the pathogen, produce antibodies, and destroy the microorganism in a potential future infection. Antibiotics and antivirals are other forms of powerful therapeutics that fight infections. Although they are effective in killing and preventing bacterial and viral growth, antibiotic and antiviral resistance is a growing issue. Modern medicine has a constant need to create new kinds of antibiotics and antivirals to treat newly drug-resistant strains due to the high mutation rate of bacteria and viruses. For pharmaceutical companies to partake in antibiotic research and development, it requires a large expense, risk, and time-consuming procedures. The return on investment for antibiotics and antivirals is unpredictable and less profitable given patients only receive a short course of treatment, which is on the opposite spectrum from the profitable research into long-term ailments such as diabetes or high cholesterol. Although there are a large number of drugs to treat HIV, there is a great need for novel drugs to combat antiviral resistance for other viral diseases such as influenza and hepatitis B and C. Because of the slower turn-around for antibiotic and antiviral medicine

development, it is advantageous to find potential alternative methods of fighting infectious diseases that circumvent the issue of continuous antibiotic and antiviral resistance (Drexler, 2010).

C. Signal Peptides in Eukaryotes/Prokaryotes

Eukaryotic proteins encoded by the nuclear genome are synthesized in the cytosol or on ribosomes attached to the endoplasmic reticulum (ER) and are then transported to specific subcellular or extracellular compartments. Localization of a protein is important for proper function, such as those related to cellular replication or homeostasis, and relies on a short amino acid sequence called a targeting or sorting signal (Burdukiewicz et al., 2018).

In Figure 1, the pathway for mRNA targeting into the ER in eukaryotes is shown within the signal-recognition particle (SRP)-ribosome cycle. To begin, large (L) and small (S) cytosolic ribosome subunits engage in the translation of mRNA that encode for secretory or membrane proteins and are targeted through the SRP pathway to the ER. After the SRP receptor recognizes the SRP, the signal peptide sequence engages with the translocon in the ER membrane and are usually cleaved off by signal peptidase. Near the end, termination of protein synthesis leads to the release of ribosomal subunits from the ER membrane to the cytosol (Kapp et al., 2009).

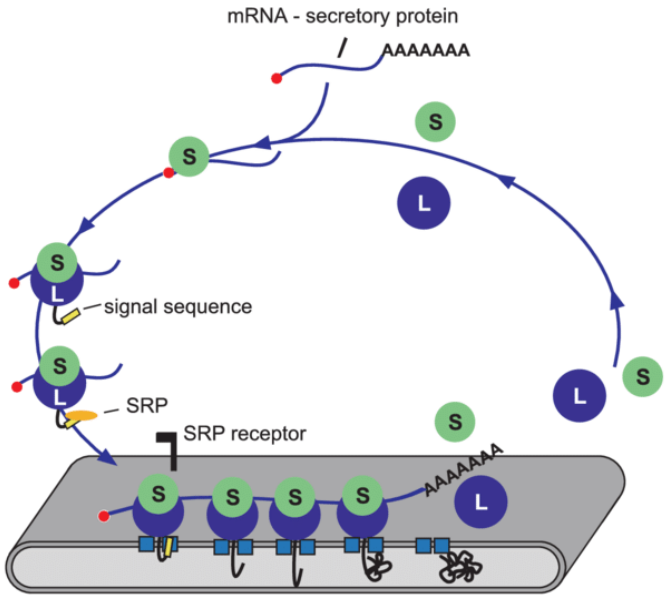


Figure 1. The signal-recognition particle (SRP)-ribosome cycle that shows a pathway for mRNA translocation to the endoplasmic reticulum (Kapp et al., 2009).

The function of a eukaryotic signal sequence is to mediate targeting of newly synthesized proteins to the endomembrane system, including the endoplasmic reticulum (ER) and the Golgi apparatus where proteins undergo post-translational modifications and folding (Burdukiewicz et al., 2018). SP-bearing proteins can remain inside of these compartments, become membrane-inserted and can do this without being part of a protein complex, or be released from the ER membrane, exported outside the cell, and later degraded but still

have some functions on their own (Hedge and Bernstein, 2006). Some functions that SP-bearing proteins may have include metabolism such as beta-galactosidase or pepsins (Hofmann et al., 1991), maintenance of tissue structure such as collagen (Chan et al., 2001), immune response such as interferons or interleukins (Zhang et al., 2005), and regulation of other homeostatic molecules such as prolactin or glucagon (Huang et al., 2010).

Signal peptides (SP) are short sequences located in the N-terminal of prokaryotic and eukaryotic proteins normally ranging within 30 to 60 amino acids in length, though they can be in very rare cases more than 60 amino acids in length. The structure of a typical signal peptide consists of 3 main regions: the n-region, h-region, and c-region (Hedge and Bernstein, 2006). As shown in Figure 2, the SP begins at the positively-charged n-region that ranges from 5 to 8 amino acid residues indicated in yellow. This region has been proposed to enforce proper topology on a polypeptide during translocation through the ER membrane (von Heijne et al., 1988). Right next to the n-region is the hydrophobic region, or h-region, indicated in blue, which is a stretch of about 8 to 12 amino residues, that is the main region of the SP as a whole. Following that region is the c-region, indicated in green, which consists of about 6 amino acid residues that are normally polar and uncharged and contains the cleavage site, indicated in red. At the cleavage site, a signal peptidase cleaves the SP off during or after protein translocation into the ER (Paetzel et al., 2002). However, some proteins do not have a cleavage site, and the SP acts as the first transmembrane domain. In these cases, the domain is sometimes referred to as a signal anchor (Hegde et al., 2006). This knowledge on the SP structure within a protein leads to questions of what the diverse functions are of specific signal peptide sequences, either together with or independent of their corresponding mature proteins.

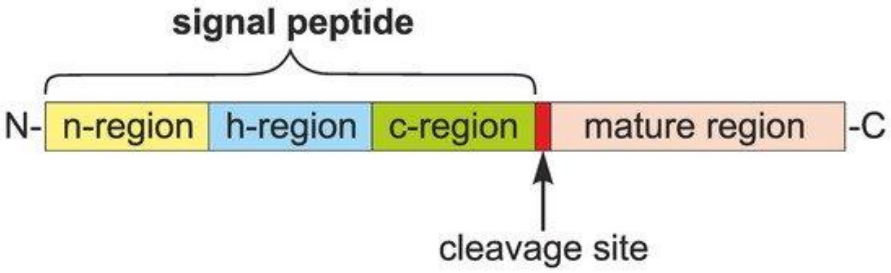


Figure 2. Organization of a typical signal peptide (SP). The lengths of SP regions are not drawn to scale (Burdukiewicz et al., 2018).

Signal sequences are variable in length and amino acid composition and play a large role in influencing the efficiency of protein translocation (Hegde et al., 2006). In one study, to conduct an analysis on signal peptides from eukaryotic toxins across kingdoms of eukaryotes, a tool called Razor was built to identify using the first 23 N-terminal residues. From this, it was seen that many defensive proteins across eukaryotic kingdoms have a toxin-like signal peptide, suggesting that many of the defensive proteins have emerged through convergent evolution. Understanding the presence or absence of SPs from eukaryotic proteins from this study helps in the development of recombinant protein expression systems since intracellular accumulation of secretory proteins and toxins may be toxic to corresponding eukaryotic cells (Bhandari et al., 2021).

One study expanded upon the understanding of signal peptide cleavage through detecting peptides being uncleaved for several types of proteins such as mammalian cell entry (Mce) proteins and PE-PGRs proteins. Through proteomic screenings in culture filtrate, membrane fraction, and whole cell lysate of *Mycobacterium tuberculosis,* it was found that the majority of these proteins with uncleaved signal peptides are enriched in a specific lipid phase. Some of these proteins were identified to be either located in the inner membrane or outer ER membrane. Because these proteins were able to travel to ER by the SRP pathway, this can suggest the post-targeting quality of the specific signal peptides within this study (De Souza et al., 2011). Knowledge of certain signal peptides that elicit proteins to undergo ER modifications can influence protein targeting to specific subcellular or extracellular compartments. Because of this, there is a great importance in studying SP structure and predictive qualities in order to improve novel drugs and therapies that may utilize SP signaling (Hegde et al., 2006).

D. Signal Peptide Predictions in Viruses

Viruses are intracellular molecules that consist of a single nucleic acid, RNA, or DNA. A typical virus, shown in Figure 3, consists of a protective protein coat known as a capsid that varies in shape − from helical forms to more complex structures. The capsid functions as a protective layer against the environment and acts in recognizing receptors that target the virus to a specific host. Some viruses have a phospholipid envelope which is acquired from the infected host's cell membrane and surrounds the protein capsid. Embedded in the lipid envelope, viral encoded proteins referred to as spike projections are glycoproteins that are involved in receptor recognition and viral tropism (Goulding, *Viruses: Introduction*).

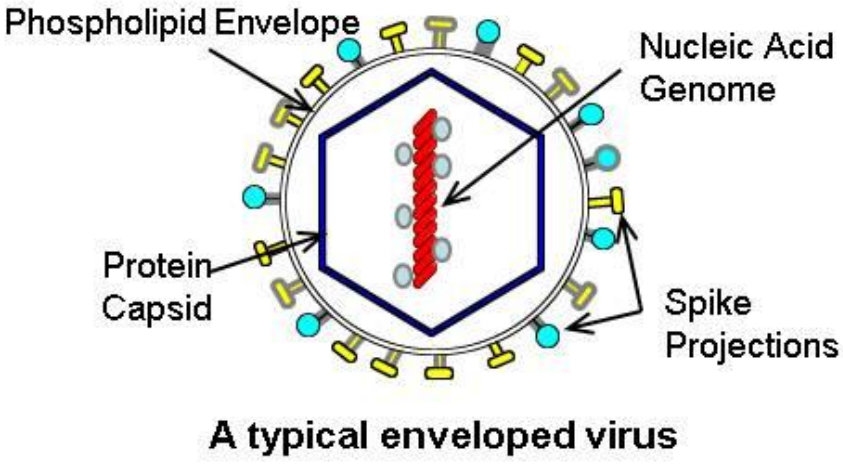


Figure 3. A typical structure of an enveloped virus.

Viruses have the ability to encode functions required for viral replication by completely relying on the protein synthesis machinery present of a host cell. Although the replication cycles of viruses can vary, there are six basic stages that are essential, which are shown in Figure 4. The first stage is attachment during which viral proteins on the capsid or phospholipid envelope engage with receptors on the host cell surface, which

determines the host range or tropism of a virus. Viral tropism involves determination of the infectious route needed to breach the specific host tissue. Some examples of viral tropisms are the targeting of T cells and monocytes with corresponding CD4 or CCR5 receptors by HIV or the targeting of the nasal epithelium with corresponding ICAM-1 receptors by the influenza virus. After viral proteins attach to a specific receptor on the host cell that can induce conformational changes, fusion of viral and cellular membranes occur. The third step of uncoating involves the viral capsid being removed and degraded by enzymes while viral genomic nucleic acid is released. After uncoating the viral genome, transcription or translation of the viral genome is initiated, which varies depending on the strandedness and sense of RNA and DNA. In the fifth step of assembly, viral proteins that may or may not be post-translationally modified are prepared for release from the host cell. Finally, lysis, resulting in the death of the infected host cell, or budding, resulting in the release from the host cell, occurs. Some viral proteins may remain within the host's cell membrane or residual viral proteins may remain within the cytoplasm of the host cell, either of which could potentially be processed and presented at the cellular surface via antigen processing and presentation pathways (Goulding, *Virus replication*).

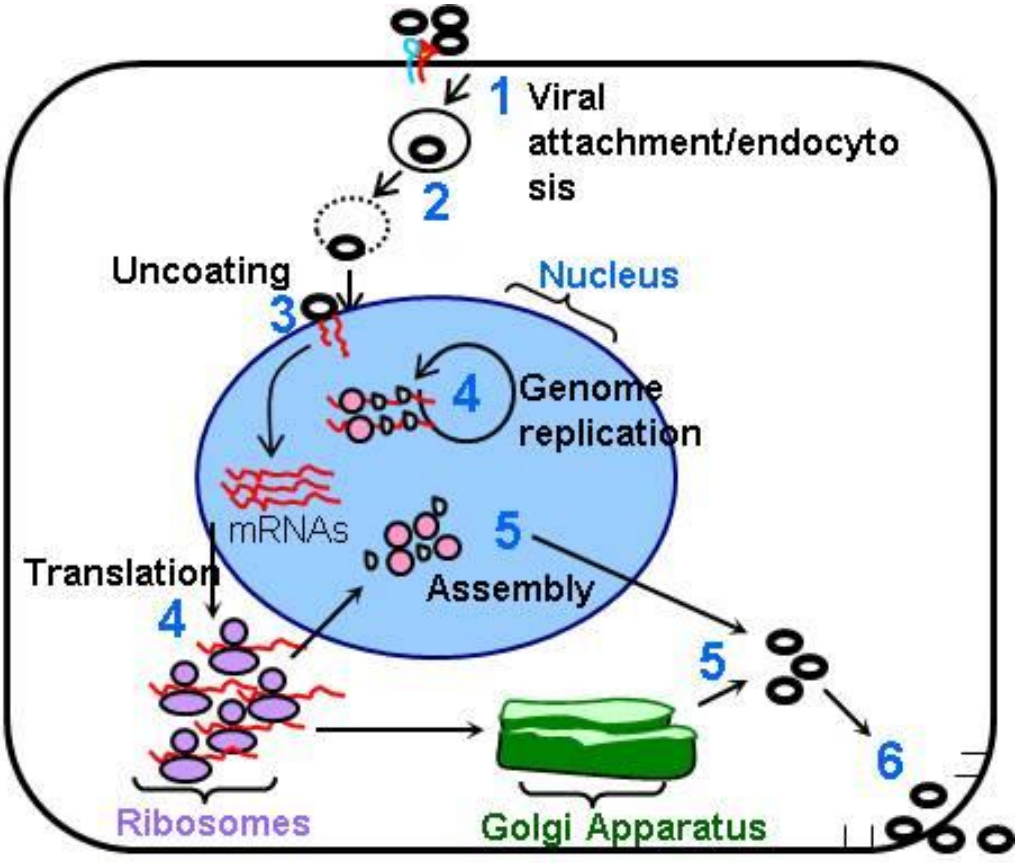


Figure 4. The six basic steps that are essential for viral replication.

In one experiment, the signal peptide of Human foamy virus (HFV) envelope glycoprotein was studied due to its unusually long length of 170 amino acid residues in length, much longer than typical 30-60 amino acid signal peptides, as indicated below in Figure 5.

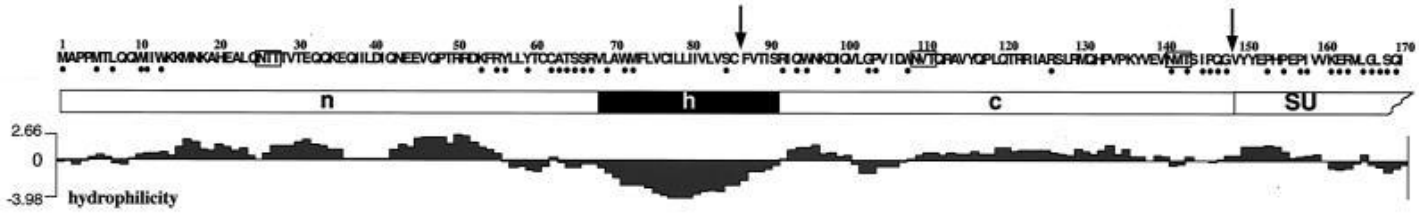


Figure 5. Schematic illustration of the N-terminal 170 aa of the HFV Env protein (Lindemann et al., 2001).

Amino acid residues that are conserved between Env proteins of different FV species are marked with black dots below the amino acid sequence. The SPC cleavage sites at positions 86 and 148 have been suggested by previous studies, indicated by the black arrows, and N-glycosylated sites are boxed. It was found that mutants with N-terminal deletions from 5 to 40 amino acid residues do not support FV expressed proteins at the cellular surface, suggesting that the N-terminal truncations of the mutant Env proteins affect targeting to the secretory pathway and proper envelope development. In the N-terminal budding domain of the h-region, deletion of two conserved tryptophan residues were shown to influence particle release and infectivity (Lindemann et al., 2001).

In one study that investigated the post-transcription regulation of exogenous and endogenous beta-retroviruses of sheep, it was found that the sheep genome contains around 27 copies of JSRV-related endogenous retroviruses, (enJSRVs) which play an essential role in the host reproductive biology and act as virus restriction factors. JSRV is an exogenous pathogenic virus that causes a transmissible lung adenocarcinoma in sheep. It was shown that sheep beta-retroviruses use a strategy to regulate exogenous posttranscriptional viral gene expression. More specifically, the signal peptide of the envelope glycoprotein of JSRV enhances nuclear export of full-length viral RNA and increases viral particle production by acting at a post-translational step of the replication cycle (Caporale et al., 2009).

Although some features of signal sequences still remain obscure, studies in eukaryotic and prokaryotic cells as well as viruses have shown that there is a predictive post-targeting quality to signal sequences and their importance in protein synthesis. These same questions that are posed to eukaryotic and prokaryotic signal peptides in proteins can be posed regarding viral proteins. Because of this, scientific questions that focus on signal peptide architecture and prediction for identifying and understanding multiple targeting and post-targeting functions can be of use for studying in not only eukaryotic and prokaryotic cells but also in viruses.

II. METHODOLOGY

A. Signal Peptide Detection and Processing of Reference Viral Database Protein Sequences

To tackle the virology-related question of where viral SPs are derived from, we retrieved viral proteins from a raw unclustered protein FASTA file (U-RVDBv18.0-prot.fasta.bz2) from the Reference Viral Database (Bigot et al., 2020). A schematic of the processing workflow is shown in Figure 6.

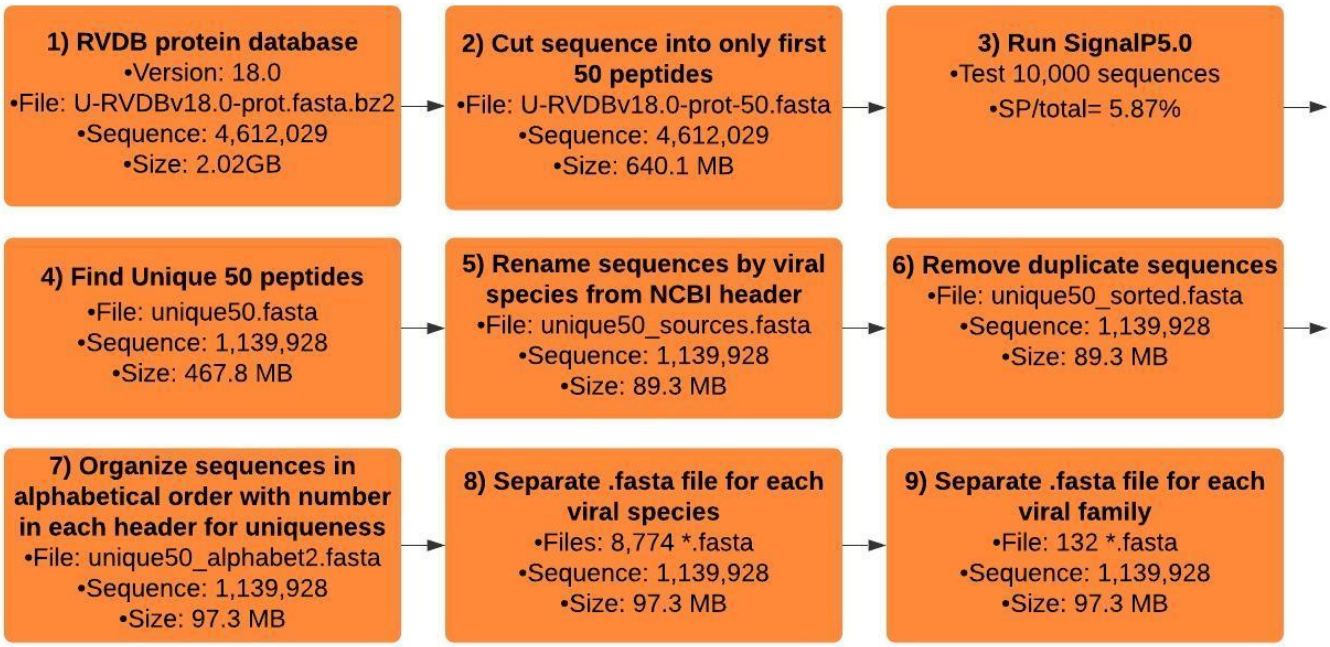


Figure 6. Schematic of workflow for processing viral protein sequences for data analysis preparation.

After their retrieval, we trimmed each protein sequence down to its first 50 amino acid residues to reflect typical eukaryotic cell SP regions (Hofmann et al., 1991). SignalP 5.0 is a software that uses deep convolutional and recurrent neural network architectures, including a conditional random field to predict the presence of signal peptides and the location of their cleavage sites in proteins from *Archaea*, *Bacteria*, and *Eukarya*. In *Bacteria* and *Archaea*, SignalP 5.0 can discriminate between three types of signal peptides: Sec/SPI, Sec/SPII, and Tat/SPI. We ran SignalP 5.0 on the FASTA file of viral proteins in order to retrieve sequences that were predicted to display signal peptide signalling factors based on patterns seen in known *Archaea*, *Bacteria*, and *Eukarya*. This produced a result of 5.87% in signal peptide detection, which was a promising indication that these particular sequences may show potential viral signal peptide functionality similar to *Prokarya* and *Eukarya* (Armenteros et al., 2019). There were some protein sequences less than 50 peptide sequences which we kept in our study for now for the purpose of prevention in any potential exclusion bias. These sequences were then put into a new FASTA file called unique50.fasta where they were reduced to non-duplicate sequences. From there, we processed the unique50.fasta file further to prepare for later data

analysis by species and family patterns. We renamed the sequences by species indicated in brackets from the traditional NCBI headers as shown.

>2acc|GENBANK|AYD68779.1|GENBANK|MH171300|non-structural polyprotein [Marine RNA virus BC-4]**acc|REFSEQ|YP_009667031.1|REFSEQ|NC_043682|non-structural polyprotein [Marine RNA virus BC-4]
MNRVNREPLSSNFKNPPLYEPKVIIRLKRVIGGNIDNMFRNSIDQNYGAG

→

>Marine RNA virus BC-4
MNRVNREPLSSNFKNPPLYEPKVIIRLKRVIGGNIDNMFRNSIDQNYGAG

We then checked for duplicate sequences again to confirm that they are all unique, put them in alphabetical order, and added a number after the header name for uniqueness in naming as shown.

>Marine RNA virus BC-4
MNRVNREPLSSNFKNPPLYEPKVIIRLKRVIGGNIDNMFRNSIDQNYGAG

→

>Marine RNA virus BC-4_797110
MNRVNREPLSSNFKNPPLYEPKVIIRLKRVIGGNIDNMFRNSIDQNYGAG

We then separated the unique50.fasta file into 8,774 FASTA files, each named with its corresponding viral species name. The next processing step was grouping the 8,774 FASTA files by viral species into 132 separate FASTA files based on their corresponding viral family (Bigot et al., 2020).

B. Basic Local Alignment of Viral Signal Peptides

To understand any potential functional similarities that the viral signal peptides may share with their corresponding host origin, the Protein Basic Local Alignment Search Tool (Protein BLAST) was used (Camacho et al., 2009). However, the corresponding hosts of the viral species were identified before using Protein BLAST. For timing and efficiency purposes, seven viral families were chosen for their connections to some commonly well-known diseases with a set of hosts that show some overlap in species, which are shown in Table 1.

Table 1. Viral families used in this project along with host names and associated diseases (Hulo et al., 2011).

Viral Family	**Host Names**	**Associated diseases**
Adenoviridae	Vertebrates (taxid:7742)	Mainly respiratory diseases, croup, and bronchitis, pneumonia Human adenoviruses
Baculoviridae	Decapoda (taxid:6683), Diptera (taxid:7147), Hymenoptera (taxid:7399), Lepidoptora (taxid:7088)	Infects insects
Coronaviridae	Vertebrates (taxid:7742)	Mainly respiratory diseases (pneumonia) and gastroenteritis SARS-CoV: (Severe acute respiratory syndrome
Flaviviridae	Humans (taxid:9606), Mammals (taxid:40674), Mosquitos (taxid:7157), Ticks (taxid:6935)	Hepaciviruses: hepatitis Pestiviruses: hemorrhagic syndromes, abortion, fatal mucosal disease Flavivirus: hemorrhagic fever, encephalitis
Herpesviridae	Vertebrates (taxid:7742)	HHV-1 and HHV-2: skin vesicles, mucosal ulcers, encephalitis and meningitis HHV-3: chickenpox (Varicella) and shingles GaHV-2: Marek's disease HHV-4/HHV-8: B-lymphocytes HHV-5: congenital CMV infection HHV-6/HHV-7: roseola infantum, exanthem subitum
Orthomyxoviridae	Birds (taxid:8782), Humans (taxid:9606), Horses (taxid:9788), Pigs (taxid:9821), Seals (taxid:97097)	Influenza A, B, C, D
Paramyxoviridae	Humans (taxid: 9606), Vertebrates (taxid:7742)	Measles, mumps, respiratory tract infections

Two methods that were used in this study due to the exploratory nature of this project. A manual approach was the initial method and an automated approach was taken overtime as a faster method. In the first approach, we used the Viral-Host Database to search for the corresponding host name of the viral species. An example in Figure 7 is shown with *Adoxophyes honmai nucleopolyhedrovirus* as the viral species and *Adoxophyles honmai* as its corresponding host name (Mihara et al., 2016).

Virus (species) name	Virus lineage	Host name	Host lineage
Adoxophyes honmai nucleopolyhedrovirus [TAX:224399]	Viruses; Baculoviridae; Alphabaculovirus	Adoxophyes honmai [TAX:85585]	Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Apoditrysia; Tortricoidea; Tortricidae; Tortricinae; Archipini; Adoxophyes

Figure 7. Identification of host name for viral species in Viral-Host Database (Mihara et al., 2016).

With this information, the Protein BLAST web interface was utilized by inputting the viral peptide sequences within each viral species in a FASTA format and comparing it to non-redundant (nr) sequences for the corresponding host organism name that was found by the Virus-Host Database (Mihara et al., 2016).

In the automated approach, the corresponding host organism names were identified by the family in order for an entire viral family FASTA file to be used on the command-line version of BLAST. Therefore, each family's typical host was identified through ViralZone, with Figure 7 shown as an example (Hulo et al., 2011). With the known host names' taxonomy IDs, the NCBI BLAST command line tool was used to retrieve all species-level taxonomy IDs associated with these higher classification taxonomy IDs. An example of retrieving species taxonomy IDs from a higher classification taxonomy name such as vertebrates would be as follows on the command line:

```
$ get_species_taxids.sh -n Vertebrates
```

With the list of taxonomy IDs for each host, seven text files were created for each viral family with species-level taxonomy IDs, depending on their host organism. The next step was using the command-line version of Protein BLAST, in which the following was a bash command used for the *Baculoviridae* family:

```
$ blastp
-task blastp -db nr -taxidlist hosts/bacu.txids \
-query Seq_by-fam/Baculoviridae.fasta \
-outfmt "10 qseqid qacc length qseq sseq qstart qend sstart send sseqid sacc staxid
evalue score ppos mismatch gaps nident pident sscinames stitle" -out
blastp_results/Baculoviridae.csv \
-max_target_seqs 1 \
-subject_besthit
```

In the above command, the argument for '-db' is 'nr' for non-redundant sequences and the argument for '-taxidlist' is 'hosts/bacu.txids', which is the text file of all of the species-level taxonomy IDs created before. This would allow for the input of viral signal peptides in *Baculoviridae*, indicated by the '-query' argument, to be searched against non-redundant sequences for the given host taxonomy IDs. For the argument 'outfmt', the query sequence ID, query accession, alignment length, query sequence, subject sequence, query start, query end, subject start, subject end, subject sequence ID, subject accession, subject taxonomy ID, expect value, raw score, percentage of positive-scoring matches, total number of gaps, number of of identical matches percentage of identical matches, subject scientific name, and subject title were all specifiers that were outputted into a .csv

file. The argument for 'subject_besthit' would ensure that the Protein BLAST algorithm produces only the best hit (smallest e-value; if e-values are identical, then the hit with the longest alignment length is chosen) for that maximum target of producing only one hit per query from the 'max_target_seq' argument. The 'max_target_seq' argument was provided to reduce redundancy in possible multiple matches for any viral signal peptides (Camacho et al., 2009).

C. Functional Clustering of Host Gene Matches

Once the viral signal peptides were found to either produce a hit or not on Protein BLAST, those that did were in results from the BLAST online web interface in the initial approach or within the .csv file outputted by the automated approach. From these BLAST results, the next question would pertain to the type of host gene matches (hits) that were produced. This question was related to whether these host genes had functions that could be grouped or classified in any way that would lead to any suggestions of functional preference in viruses. Functionally annotating the host genes would shed insight into potential functional evolutionary explanations of viruses mimicking host organisms.

Therefore, we used the Database for Annotation, Visualization and Integrated Discovery (DAVID) to give biological meaning to the large list of host genes produced by Protein BLAST. In this bioinformatics database tool, we uploaded a gene list that contained the accession numbers of the corresponding host genes that produced a hit to viral SP on Protein BLAST (retrieved from the server in the manual approach and from the 'sacc' column in the automated .csv output). This list was then used as input to the functional annotation tool that batch annotates and performs gene-GO term enrichment analysis to highlight the most relevant GO terms associated with the given gene list. These annotation categories include GO terms, protein-protein interactions, protein functional domains, disease associations, bio-pathways, homologous genes, and many more. After a complete annotation, the same gene list was put through the functional annotation clustering tool, which uses a novel algorithm to measure relationships among the annotation terms based on degrees of their co-association genes to group the similar, redundant, and heterogeneous annotation content into annotation groups. The results from this analysis would give us a better understanding of whether the matching host genes are of similar function(s) to their viral counterparts that can be classified into related clusters (Huang et al., 2009).

D. Reconstructing Ancestral Sequences of Viral Peptides and Host Genes

Another question to consider after getting Protein BLAST results would be whether there could be some phylogenetic inference made for the instances where the viral signal peptide matched with a host gene. This would be an inquiry regarding potential selective pressure in viral preference for specific host genes during viral hijacking of the host genome (Goulding, *Virus replication*).

To begin making an inference about phylogeny, we turned toMrBayes (version 3.2.7a), a command-line tool used for Bayesian phylogenetic analysis. This bioinformatics program uses Markov Chain Monte Carlo (MCMC) techniques to sample from the posterior probability distribution. The Metropolis-coupling method accelerates convergence between "cold" chains and "heated" chains that are automatically created from the sequences inputted. In this case, the sequences used for this analysis were identified viral SP sequences that

had a Protein BLAST hit along with their matching host gene. From this, phylogeny tree models are created based on probability of evolutionary convergence (Ronquist et al., 2012).

To start using MrBayes, a NEXUS-formatted multiple sequence alignment (MSA) using CLUSTAL OMEGA (Madeira et al., 2019) of the identified viral SP sequences was created. This NEXUS file was then inputted into MrBayes and then run through the software algorithm within the directory that the NEXUS file was located, which looks like the following:

```
mb > mb
mb > execute [file_name].nex
mb > mcmcp samplefreq=100 printfreq=500
mb > mcmc
```

Multiple files are output into the same directory in which the MrBayes program was run. One of the outputs is a tree file (.t) where multiple unrooted tree generations with homogeneity statistics are produced from phylogenetic inference, portraying the relationship among the species without depicting their common ancestor (Ronquist et al., 2012).

From these unrooted trees, ancestral sequences can be reconstructed for later comparison between virus and host. Ancestral sequence reconstruction (ASR) is a technique used in molecular evolution which is based on a phylogenetic analysis of extant homologous amino acid sequences. The ancestral sequences can reveal mechanisms and dynamics of protein evolution, which is key to answering the question of how viral signal peptides and their corresponding host gene have co-evolved. We used the Phylogenetic Analysis for Maximum Likelihood (PAML) graphical interface tool that is able to estimate parameters, test hypotheses, and reconstruct ancestral sequences to study evolutionary processes. To begin, this tool inputs a multiple sequence alignment of the intended sequences in Phylip format (Yang et al., 2007). Again, we used the same viral signal peptide and/or host gene sequences from the previous step using MrBayes to produce the Phylip-formatted multiple sequence alignment with CLUSTAL OMEGA (Madeira et al., 2019). The second input PAML takes is a generated tree file, which would be the tree file that MrBayes produced through phylogenetic inference in the previous step. After the program is set to run with the ancestral sequence reconstruction option checked, there are multiple outputs that PAML produces. For this study, the reconstructed ancestral sequences are in the file named “rst”. Within this text file, reconstructed ancestral sequences are created for every tree generation from the MrBayes tree file input. For each tree in the “rst” file, ancestral sequences are predicted to be at certain nodes within the tree along with the corresponding conserved amino acid sequence. For each tree of the virus or host, if their reconstructed ancestral sequences are more distant, then that could suggest independent capture of the sequences in potential convergent evolution. This could be a case where the viral species themselves may not be closely related but independently evolve similar traits as a result of having to adapt to similar environments based on their host. If the ancestral sequences are closer, then potential divergent evolution is suggested where the viruses or hosts shared a common ancestor and ended up developing different signal peptide regions based on the host (Yang et al., 2007). This methodology sought to support a co-evolutionary

explanation while in search for possible indications of viral signal preference when viruses are reproduced through host replicative machinery.

III. RESULTS/ DISCUSSION

A. Viral Signal Peptides Alignments with Host Genes

From the manual Protein BLAST approach, many viral signal peptide sequences produced a hit with a host gene match. The table below summarizes the results from the initial approach.

Table 2. Protein BLAST statistics for each viral family analyzed in this project.

Viral Family	Total SP seq.	BLAST seq. (A)	Total viral SP | Host gene matches (B)	Ratio (B/A)
Adenoviridae	509,577	7,579	27	0.00356
Baculoviridae	22,301	11,821	30	0.00253
Coronaviridae	505,465	8,617	2	0.00023
Flaviviridae	590,978	49,245	19	0.00039
Herpesviridae	502,017	11,619	43	0.00370
Orthomyxoviridae	60,865	1,520	1	0.00066
Paramyxoviridae	508,859	2,692	3	0.00111

From Table 2, the seven viral family statistics are shown with corresponding total signal peptide sequences after processing viral protein sequences from the Reference Viral Database (Bigot et al., 2020) within each family. The number of viral signal peptide sequences that had corresponding viral species identified with a corresponding host name from the Virus-Host Database (Mihara et al., 2016), is found in column A (labeled 'BLAST seq. (A)'). From this column, Flaviviridae had the largest number of identified virus-host relationships from the Virus-Host Database at 49,245 sequences that were able to be BLASTed while Orthomyxoviridae had the smallest number of identified virus-host relationships at 1,520 sequences. Column B reports the number of viral signal peptides from column A that produced a host gene match from Protein BLAST. From this column, *Herpesviridae* produced the largest number of host gene matches at 43 while Orthomyxoviridae produced the smallest number at just 1. In the last column, the ratio of signal peptide similarity was calculated as the amount of sequences producing Protein BLAST hits from column B out of the total amount of sequences identified with a host from column A. From highest to lowest ratios of signal peptide similarity was *Herpesviridae* at 0.00370 (0.37%), *Adenoviridae* at 0.00356 (0.36%), *Baculoviridae* at 0.00253 (0.25%), *Paramyxoviridae* at 0.00111 (0.11%), *Orthomyxoviridae* at 0.00066 (0.07%), *Flaviviridae* at 0.00039

(0.04%), and *Coronaviridae* at 0.00023 (0.02%). The differences in occurrence of signal peptide similarity to a host gene is unknown, but these initial numbers provide some initial insight into the differences of signal peptides across the viral families.

From the results in the initial Protein BLAST approach, it was important to find the significance of the viral signal peptides that produced a match to a host gene as this method became the primary aim to understand this phenomenon within this project. Table 5 in the Appendix shows detailed annotations about the results produced from the viral signal peptides that produced a host gene match. From that table, the third column labeled 'viral gene' is the full viral gene that the viral signal peptide comes from, which was identified from the original U-RVDBv18.0-prot.fasta.bz2 downloaded from the Reference Viral Database (Bigot et al., 2020). After retrieving the original full viral gene, the next question was whether or not the full viral gene would match with the same host gene. To answer this question, the full viral gene was also put through Protein BLAST against the same host name. The last column in Table 5 has 'Y' if the full viral gene also matched with the same full host gene and 'N' if that is not the case. The following number ranges in that same column indicates the amino acid residues at which the viral signal peptide matched with the host gene in cases for 'N' and at which the full viral gene matched with the host gene in cases for 'Y'. This column was also highlighted in either red or green for 'N' cases where only the viral signal peptide produced a match with a host gene. The boxes highlighted in green indicates that the matched host gene, or subject frame, was within the host's signal peptide region. This would be the host gene's first 50 amino acid residues (Hedge and Bernstein, 2006). The boxes highlighted in red indicates that the matched host gene or subject frame was not within the host's signal peptide region.

From the annotated results in Table 5 from the Appendix detailing patterns of how the viral signal peptide and/or gene matches with a host gene, three different categories were created to propose mechanisms by which the virus uses host machinery to capture host genes in order to mimic their functions. The first mechanism is that only the virus steals the host signal peptide, which is suggested in cases where only the viral signal peptide sequence matches with a host gene, with an example shown in Figure 8 using the viral signal peptide/gene sequence for a large T antigen from *Bovine polyomavirus 2a*.

Host: Bos taurus[TAX:9913]
>acc|GENBANK|AIT68757.1|NEIGHBOR|KM496324|large T antigen [Bovine polyomavirus 2a]
MDSVLTRQEKRELCDLLEIPHHCYGNIPMMKAQYKKMCLVYHPDKGGDGSKMENEPIYGTPQFKAWWYQQHGSYFSESQSSSKPTRGGP
SRFSRAASPDSTTSTSTTGSGIFASESSQRSDVGASGSQTSSAKTSSSSGGGRATGGGGTRVFGGTAYSREFRSESQDSDLYCDETLGSS
PESAGEDDGPSSSTSQPSTSTPQKPRKMPPGAARKRAFDDGDSSAGSSRTFDSTPPKTKKNESFHHPTDFPSCLLDYLSHAVYSNKTVSA
FAVYSTLEKVTLLYEKCDKFKVDFKSRHKYKAPDGSLAGILFLITLTKHRVSAVKNYCSMFCTISFLIVKGVNKSPEFYRALCNEPFALL
EENRNGVWSYEFDSKNEKAESVSWTAIAEFAETYELEDPLIIMAHYLDFANPFPCAKCHGKGLKAHKDHEKQHNNAKLFKNAKAQKNICQ
QAADVVLAKKRLKILESSREELLAEKFKKQLSKLRELHTIPLLEHMAGVAWYCCLFSDFEEKLVRVLQLLTENCPKHRNSLFIGPINSGK
TSFAAAILDLIEGKSLNVNCPADKLNFELGCAIDHFAVVFEDVKGQTSLNKTLQPGQGIHNLDNLREHLDGAVSVNLERKHVNKRTQIFP
PCIVTANEYVFPQTLLARFAYTLKFEPRKVLKKALDANKDLSKHRILQQGLTLLLALIWLCPASKFHSSIREEVASWKSILNSEIGNEKF
CVMIENIEKGMDPLKDFIEEESTDDSGRFTQSQ

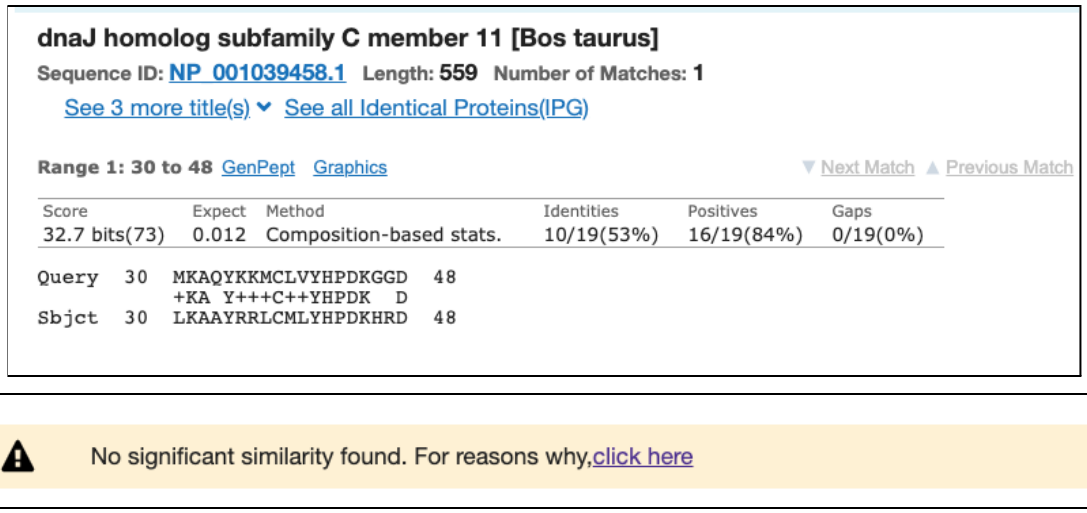
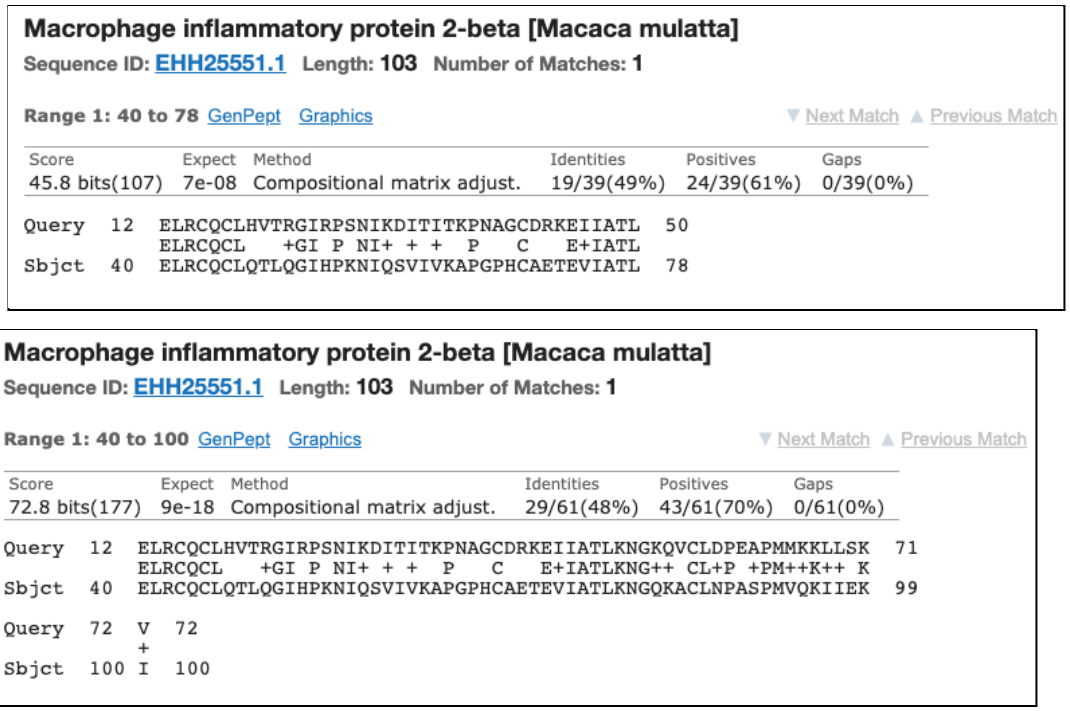


Figure 8. First proposed mechanism of viral evolution in which only the viral signal peptide sequences captures part of a host gene (the host is *Bos taurus*). Only the viral SP sequence highlighted in yellow is able to produce a hit on Protein BLAST (top), but not the full corresponding viral gene (bottom).

The second mechanism is that the virus manages to replicate part of a host gene, which is suggested in cases where only the partial viral gene matches a host gene, with an example shown in Figure 9 using the viral signal peptide/gene sequence for RhUL146b from *Macacine betaherpesvirus 3*.

```
Host: Macaca mulatta [TAX:9544]
>acc|GENBANK|ABS84093.1|GENBANK|EU003822|RhUL146b [Macacine betaherpesvirus 3]
MTLIAYHQTEAELRCQCLHVTRGIRPSNIKDITITKPNAGCDRKEIIATLKNGKQVCLDPEAPMMKKLLSKVPEGKYPSFWEQYKEHFLK
MFTE
```



Macrophage inflammatory protein 2-beta [Macaca mulatta]

Sequence ID: EHH25551.1 **Length:** 103 **Number of Matches:** 1

Range 1: 40 to 78 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
45.8 bits(107)	7e-08	Compositional matrix adjust.	19/39(49%)	24/39(61%)	0/39(0%)

```
Query  12  ELRCQCLHVTRGIRPSNIKDITITKPNAGCDRKEIIATL  50
           ELRCQCL   +GI P NI+ + +  P   C   E+IATL
Sbjct  40  ELRCQCLQTLQGIHPKNIQSVIVKAPGPHCAETEVIATL  78
```


Macrophage inflammatory protein 2-beta [Macaca mulatta]

Sequence ID: EHH25551.1 **Length:** 103 **Number of Matches:** 1

Range 1: 40 to 100 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
72.8 bits(177)	9e-18	Compositional matrix adjust.	29/61(48%)	43/61(70%)	0/61(0%)

```
Query  12  ELRCQCLHVTRGIRPSNIKDITITKPNAGCDRKEIIATLKNGKQVCLDPEAPMMKKLLSK  71
           ELRCQCL   +GI P NI+ + +  P   C   E+IATLKNG++ CL+P +PM++K++ K
Sbjct  40  ELRCQCLQTLQGIHPKNIQSVIVKAPGPHCAETEVIATLKNGQKACLNPASPMVQKIIEK  99

Query  72  V  72
           +
Sbjct  100 I  100
```

Figure 9. Second proposed mechanism of viral evolution in which the viral gene captures part of a host gene (the host is *Macaca mulatta*). Both the viral SP sequence highlighted in yellow (top) and corresponding viral gene (bottom) are able to produce a hit on Protein BLAST. However, only part of the viral gene matches with a gene host (bottom).

The third mechanism would be that the virus manages to replicate the full host gene function in cases where the full viral gene matches with the host gene, with an example shown in Figure 10 using the viral signal peptide/gene sequence for a polyprotein from *Aedes flavivirus*.

Host: Aedes albopictus [TAX:7160]
>acc|GENBANK|BAH83670.1|GENBANK|AB488411|polyprotein [Aedes flavivirus]
RTIILTPTRVVMAEVMEALGQAGITCDRNLMYCRRNLVTVACHATFTKFVLSHGVKKIGVALIVMDECHFMDPMSIAARGVMEHLHEKGTKLMYL
SATPPGHSPDGGSNFPIHDQAIAFPSWMTPAWINSVRKSRNSKKAIMFVPSHTQANSLAATIPGAVPLHRANFTANYARASSDETSLVVSTDISEMGAN

cell fusing agent virus polyprotein-like protein [Aedes albopictus]

Sequence ID: AAQ03217.1 Length: 1557 Number of Matches: 1

Range 1: 961 to 1009 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
90.9 bits(224)	2e-23	Composition-based stats.	41/49(84%)	44/49(89%)	0/49(0%)

```
Query  1     RTIILTPTRVVMAEVMEALGQAGITCDRNLMYCRRNLVTVACHATFTKF  49
             RTIILTPTRVVMAEV++AL   GI  DRNLMYC+RNLVTVACHATFTKF
Sbjct  961   RTIILTPTRVVMAEVIDALAAVGINSDRNLMYCKRNLVTVACHATFTKF  1009
```

cell fusing agent virus polyprotein-like protein [Aedes albopictus]

Sequence ID: AAQ03217.1 Length: 1557 Number of Matches: 1

Range 1: 961 to 1154 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
370 bits(949)	4e-119	Compositional matrix adjust.	172/194(89%)	182/194(93%)	0/194(0%)

```
Query  1     RTIILTPTRVVMAEVMEALGQAGITCDRNLMYCRRNLVTVACHATFTKFVLSHGVKKIGV  60
             RTIILTPTRVVMAEV++AL   GI  DRNLMYC+RNLVTVACHATFTKFVLSHGVKKIGV
Sbjct  961   RTIILTPTRVVMAEVIDALAAVGINSDRNLMYCKRNLVTVACHATFTKFVLSHGVKKIGV  1020

Query  61    ALIVMDECHFMDPMSIAARGVMEHLHEKGTKLMYLSATPPGHSPDGGSNFPIHDQAIAFP  120
              LI+MDECHFMDPMSIAARG+MEHLHEKGTKLMYLSATPPGH+PDGGSNFPIHDQAIAFP
Sbjct  1021  TLIIMDECHFMDPMSIAARGIMEHLHEKGTKLMYLSATPPGHTPDGGSNFPIHDQAIAFP  1080

Query  121   SWMTPAWINSVRKSRNSKKAIMFVPSHTQANSLAATIPGAVPLHRANFTANYARASSDET  180
             SWMTPAWIN VRKSRNS+KAIMFVPSHTQANSLA +IPGAV LHR NF+ NYARA SDET
Sbjct  1081  SWMTPAWINGVRKSRNSRKAIMFVPSHTQANSLAGSIPGAVSLHRGNFSTNYARAGSDET  1140

Query  181   SLVVSTDISEMGAN  194
             +LVVSTDISEMGAN
Sbjct  1141  TLVVSTDISEMGAN  1154
```

Figure 10. Third proposed mechanism of viral evolution in which the viral gene captures the full host gene (the host is *Aedes albopictus*). Both the viral SP sequence highlighted in yellow (top) and corresponding viral gene (bottom) are able to produce a hit on Protein BLAST.

Using these three proposed viral-host mechanisms, these totals were calculated for each viral family and shown in Table 3.

Table 3. Grouping of viral signal peptide match or origin from host gene. Bracketed numbers indicate that the match includes that host SP.

Viral Family	Viral SP only = host gene [in host SP]	Viral gene (partial) = host gene [in host SP]	Viral gene (full) = host gene [in host SP]
Adenoviridae	2 [0]	1 [1]	1 [0]
Baculoviridae	1 [0]	4 [1]	25 [22]
Coronaviridae	-	-	2 [1]
Flaviviridae	-	6 [0]	13 [2]
Herpesviridae	4 [4]	16 [3]	26 [22]
Orthomyxoviridae	-	1 [1]	-
Paramyxoviridae	1 [1]	-	2 [1]
TOTAL	8 [5]	28 [6]	69 [48]

From the results above, the totals for each mechanism indicate that the most common mechanism is where the full viral gene matches the host gene at 69 cases and that the least common mechanism is where only the viral signal peptide matches the host gene at 8 cases. *Herpesviridae* has 4 occurrences of only viral signal peptide matching with a host gene, with the same 4 cases being matched to the host signal peptide region which is indicated in the bracketed number. At 16 cases, *Herpesviridae* had the highest occurrences of a partial viral gene match to a host gene, with 3 of the 16 cases also matching to the host signal peptide region. For the full viral gene matching to a host gene, *Baculoviridae* and *Herpesviridae* had the highest number of cases, with the majority of them also occurring within the host signal peptide region. Overall, the results from the initial approach of Protein BLAST indicate that the mechanism where a viral signal peptide matches to a host gene, much less to a corresponding host signal peptide region, is seen much more rarely than the occurrence where a viral gene partially or fully matches to a host gene. These specific viral signal peptides and genes along with their corresponding host genes were further analyzed in this project to better understand when these cases occur.

From the second automated Protein BLAST approach, many viral signal peptide sequences produced a hit with a host gene match. In this more recent approach, *Baculoviridae* was used due to its faster availability of Protein BLAST results. Since this viral family had the smallest number of signal peptides at 22,301 sequences, *Baculoviridae* produced the first Protein BLAST .csv output to perform analyses after about five

days. The following results summarized in Table 4 are solely from these 4,976 *Baculoviridae* signal peptide sequences that produced a host gene match.

Table 4. Grouping of *Baculoviridae* signal peptide sequences that produce a hit from Protein BLAST.

Proposed viral mechanism	# of viral SP producing host gene match	# in host SP
Viral SP only = host gene	4,204	1,063
Viral gene (partial) = host gene	686	346
Viral gene (full) = host gene	86	57
TOTAL	4,976	1,466

From the table above describing Protein BLAST results for *Baculoviridae*, out of the 22,301 total *Baculoviridae* signal peptide sequences, 4,976 (22.31%) viral signal peptide sequences produced a hit on the command-line version of Protein BLAST. There are 4,204 cases where only the viral signal peptide sequence matches with a host gene, 686 cases where part of the corresponding viral gene of the SP matches with a host gene, and 86 cases where the full corresponding viral gene of the SP matches with a host gene. The total number of cases where the viral signal peptide or corresponding gene produces a match in the host signal peptide region is calculated in the last column. In total, there are 1,466 cases out of the total 4,976 Protein BLAST hits where this occurs, which is almost ⅓ of all cases.

When looking at the automated approach results, we observed more instances where only the viral signal peptide sequence matches with the host gene than originally thought from the manual Protein BLAST approach results, which was surprising to see considering it seemed like the opposite from the initial BLAST approach. Because the initial approach excluded many viral species without identified hosts from the Virus-Host Database, the viral SPs excluded added to the bias of this approach and possibly missed several gene matches. From the results seen in both Protein BLAST approaches, these particular viral signal peptide/gene sequences, along with their corresponding host genes, are further analyzed in search of potential function and selective pressure.

B. Functional Clustering of Host Genes

From the subset of initial Protein BLAST results from the manual approach that showed the viral signal peptides producing a match to a host gene, we analyzed the corresponding host gene in search of potential functional preference that the virus may have when replicating through hosts. From the Protein BLAST results in Table 3, the host genes' accession IDs were compiled into a text file and submitted into the DAVID gene conversion tool. From the input of 78 unique gene accession IDs, the online tool was able to recognize only 33 of them, which are listed in Table 6 in the Appendix. The list of 33 recognized genes was submitted to the DAVID functional annotation tool and the functional annotation clustering tool to search for common

functional clusters. From the functional annotation clustering chart shown collectively in Figures 11 and 12, there were 5 clusters that were found. The classification stringency parameter can be adjusted depending on how strict one may want the clustering tool to perform. In order to keep the greatest number of terms used for the functional annotation clustering tool, the low classification stringency was used. With this parameter, only five terms were omitted: tumor necrosis factor-mediated signalling pathway, cell proliferation, negative regulation of apoptotic process, protein heterodimerization activity, and metabolic pathways.

Annotation Cluster 1	Enrichment Score: 2.51			Count
GOTERM_BP_DIRECT	regulation of transcription involved in G1/S transition of mitotic cell cycle	RT		3
GOTERM_BP_DIRECT	G1/S transition of mitotic cell cycle	RT		3
GOTERM_CC_DIRECT	nucleoplasm	RT		7
Annotation Cluster 2	**Enrichment Score: 2.42**			**Count**
GOTERM_BP_DIRECT	immune response	RT		6
INTERPRO	CXC chemokine	RT		3
INTERPRO	CXC chemokine, conserved site	RT		3
GOTERM_BP_DIRECT	inflammatory response	RT		5
GOTERM_BP_DIRECT	response to lipopolysaccharide	RT		4
UP_KEYWORDS	Cytokine	RT		4
INTERPRO	Chemokine interleukin-8-like domain	RT		3
KEGG_PATHWAY	Cytokine-cytokine receptor interaction	RT		5
SMART	SCY	RT		3
GOTERM_MF_DIRECT	chemokine activity	RT		3
GOTERM_BP_DIRECT	chemokine-mediated signaling pathway	RT		3
GOTERM_CC_DIRECT	extracellular space	RT		6
GOTERM_BP_DIRECT	defense response to bacterium	RT		3
UP_SEQ_FEATURE	disulfide bond	RT		8
GOTERM_MF_DIRECT	growth factor activity	RT		3
GOTERM_BP_DIRECT	regulation of cell proliferation	RT		3
UP_KEYWORDS	Disulfide bond	RT		8
KEGG_PATHWAY	Chemokine signaling pathway	RT		3
UP_KEYWORDS	Signal	RT		8
GOTERM_CC_DIRECT	extracellular region	RT		5
UP_SEQ_FEATURE	signal peptide	RT		6
GOTERM_BP_DIRECT	G-protein coupled receptor signaling pathway	RT		3
UP_KEYWORDS	Secreted	RT		4
Annotation Cluster 3	**Enrichment Score: 2.29**			**Count**
GOTERM_BP_DIRECT	nucleotide-excision repair, DNA gap filling	RT		3
GOTERM_BP_DIRECT	DNA damage response, detection of DNA damage	RT		3
GOTERM_BP_DIRECT	translesion synthesis	RT		3
GOTERM_BP_DIRECT	nucleotide-excision repair, DNA incision, 5'-to lesion	RT		3
GOTERM_BP_DIRECT	nucleotide-excision repair, DNA incision	RT		3
GOTERM_BP_DIRECT	transcription-coupled nucleotide-excision repair	RT		3
GOTERM_CC_DIRECT	nucleoplasm	RT		7
UP_KEYWORDS	Nucleus	RT		8
GOTERM_CC_DIRECT	nucleus	RT		8
UP_KEYWORDS	Methylation	RT		3

Figure 11. Functional annotation clustering of host genes that had a viral signal peptide match from Protein BLAST (clusters 1-3).

Annotation Cluster 4	Enrichment Score: 1.03		Count
GOTERM_BP_DIRECT	DNA strand elongation involved in DNA replication	RT	3
GOTERM_BP_DIRECT	telomere maintenance via recombination	RT	3
KEGG_PATHWAY	DNA replication	RT	3
UP_SEQ_FEATURE	zinc finger region:C4-type	RT	3
UP_KEYWORDS	DNA replication	RT	3
GOTERM_BP_DIRECT	DNA replication	RT	3
GOTERM_BP_DIRECT	DNA repair	RT	3
GOTERM_CC_DIRECT	nucleoplasm	RT	7
GOTERM_MF_DIRECT	chromatin binding	RT	3
GOTERM_MF_DIRECT	protein binding	RT	12
UP_KEYWORDS	Zinc	RT	5
UP_KEYWORDS	Nucleus	RT	8
UP_KEYWORDS	Disease mutation	RT	5
UP_KEYWORDS	Ubl conjugation	RT	4
UP_KEYWORDS	Zinc-finger	RT	4
GOTERM_MF_DIRECT	DNA binding	RT	4
GOTERM_CC_DIRECT	nucleus	RT	8
UP_KEYWORDS	DNA-binding	RT	4
GOTERM_MF_DIRECT	zinc ion binding	RT	3
UP_SEQ_FEATURE	mutagenesis site	RT	4
GOTERM_CC_DIRECT	cytoplasm	RT	7
UP_KEYWORDS	Metal-binding	RT	5
UP_KEYWORDS	Transferase	RT	3
GOTERM_CC_DIRECT	extracellular exosome	RT	4
UP_KEYWORDS	Acetylation	RT	4
GOTERM_MF_DIRECT	metal ion binding	RT	3
GOTERM_CC_DIRECT	cytosol	RT	4
UP_SEQ_FEATURE	sequence variant	RT	11
UP_KEYWORDS	Phosphoprotein	RT	7
UP_KEYWORDS	Polymorphism	RT	9
UP_KEYWORDS	Cytoplasm	RT	3
Annotation Cluster 5	**Enrichment Score: 0.67**		**Count**
GOTERM_BP_DIRECT	cell surface receptor signaling pathway	RT	5
UP_KEYWORDS	Signal	RT	8
UP_SEQ_FEATURE	topological domain:Extracellular	RT	5
UP_SEQ_FEATURE	glycosylation site:N-linked (GlcNAc...)	RT	6
GOTERM_CC_DIRECT	integral component of membrane	RT	7
UP_KEYWORDS	Transmembrane helix	RT	7
UP_KEYWORDS	Transmembrane	RT	7
UP_SEQ_FEATURE	topological domain:Cytoplasmic	RT	5
UP_KEYWORDS	Receptor	RT	3
UP_SEQ_FEATURE	transmembrane region	RT	6
GOTERM_CC_DIRECT	plasma membrane	RT	5
UP_KEYWORDS	Glycoprotein	RT	5
UP_KEYWORDS	Membrane	RT	7
UP_KEYWORDS	Alternative splicing	RT	5

Figure 12. Functional annotation clustering of host genes that had a viral signal peptide match from Protein BLAST (clusters 4-5).

From the functional annotation classification tool, cluster 1, 3, and 4 all have functional terminology related to cellular division and replication such as nucleoplasm, regulation of transcription involved in G1/S transition of mitotic cell cycle, DNA damage response, detection of DNA damage, nucleus, DNA replication, nucleus, acetylation, and cytoplasm as core examples. In the literature, it is seen that viruses hijack DNA damage responses to enhance viral replication. In one study, it was found that DNA virus infection can cause DNA damage responses in host cells, some of which include ATM kinase activation of H2AX phosphorylation. The activation of components within the DNA damage signalling response is an active viral kinase-driven strategy that has been found for a requirement in efficient gamma-herpesvirus replication (Tarakanova et al., 2007). Another article describes Marseilleviridae that initiate their replication by transiently recruiting the nuclear transcription machinery to their cytoplasmic viral factor (Fabre et al., 2017). Therefore, these three clusters are consistent with the types of genes that viruses hijack in order to replicate efficiently. This alludes to the significance of the large number of gene counts, indicated on the right-most column labeled 'Count', that are correlated with the functional annotations from those three clusters.

For cluster 2, some of the annotations that correlate with the largest amount of gene counts are immune response, CXC chemokine, inflammatory response, chemokine activity, disulfide bond, signal [peptide], and cytokine-cytokine receptor interaction. In the literature, it is indicated that DNA viruses can encode homologues of cytokines, chemokines, and their receptors, which are all molecules that play an important role in the control of the immune response. Viruses have the ability to capture host genes or evolved genes that target specific immune pathways (Alcami, 2003), which helps explain the chemokine and cytokine-related functional annotation key words. However, on our literature review for specific inflammatory-targeting genes that the virus may capture, we were unable to find a specific article that described an instance where a viral signal peptide captured a host gene related to inflammatory response. Having 5 genes correlate to this functional annotation indicates an area of interest to study with a potential novel indication into inflammatory pathways of the host.

For cluster 5, the main keywords that encompassed this functional group were cell surface receptor signalling pathway, signal, glycosylation site: N-linked (GlcNAc), receptor, and transmembrane region. This was a promising indication with the term 'signal' seen in this cluster and within cluster 2. These results may suggest that viral signal peptides are capturing 'signal peptide' functions of their corresponding hosts. It is known that viruses trigger metabolic reprogramming in host cells to hijack cellular metabolism and efficiently reproduce (for a review, see Thaker et al., 2019). Therefore, the acceptance of glycosylation, which is linked to the metabolic pathway as a keyword, indicates that this set of host genes were mimicked by the viral signal peptides to possibly support viral replication. Functional annotation terms involving locations within a eukaryotic cell such as 'transmembrane region' from cluster 5 or 'nucleus' and 'cytoplasm' from cluster 4 may point towards translocative properties that these viral signal peptides are capturing from the host. In one study, it has been proposed that endoplasmic reticulum signal peptides lead hepatitis C virtual polyproteins to the endoplasmic reticulum membrane, which suggests that these signal peptides aid in the compartment localization of matured viral proteins. These specific signal peptides identified in this study direct polyprotein processing in the ER membrane before virion assembly, which may indicate that this is a general mechanism adopted by hepatitis viruses for virion replication (Wu, 2001). This suggests the hypothesis that those particular

viral signal peptides may hijack translocative host gene functions. As some of the clusters created through this seven-family-wide functional gene study present types of functions that viruses have been known to hijack within the host to facilitate optimal replication, it is important to note that the many of the corresponding viral genes of the SP are similar to the matching host gene's function..

Using the subset of Protein BLAST results from the automated approach that showed the Baculoviridae viral signal peptides producing a match to a host gene, we also analyzed the corresponding host genes through DAVID tools. 433 unique genes were submitted to the DAVD gene ID conversion tool, of which only 79 were recognized. No clusters were able to be found amongst these 79 genes, possibly due to the high functional variability in these host genes. Therefore, a different strategy was used to identify functional gene clusters. For the three proposed mechanisms that were used to categorize the Protein BLAST results for *Baculoviridae*, the viral signal peptide or gene sequences along with the host gene matches were separated into the three separate groups. Using CLANS, a bioinformatics tool that performs clustering based on all-against-all BLAST+ similarities (Zimmermann et al., 2018), the viral SP for each host-gene matching mechanism was used in order to group similar viral SPs together. The respective host gene matches of the largest cluster of similar viral SPs in each of the three mechanisms were studied for functional similarities.

Out of the 4,204 cases where only the *Baculoviridae* signal peptide region matched with a host gene, a cluster of 1,064 similar sequences produced a set of 688 unique gene matches. From the cluster of 1,064 viral peptides, the most common functional properties from their corresponding host match was ubiquitin-related and superoxide genes. 19 viral signal peptides matched with ubiquitin-60S ribosomal protein L40 from *Eufriesea mexicana*, 15 viral signal peptides matched with ubiquitin from *Penaeus vannamei*, and 6 viral signal peptides matched with Cu-Zn superoxide dismutase from *Lymantria dispar*. It has previously been shown that viruses can take over the host ubiquitin system. Many DNA viruses can encode their own ubiquitin modifying machinery to dysregulate the host cell cycle (Gustin et al., 2011).

Out of the 686 cases where part of the *Baculoviridae* gene matched with a host gene, a cluster of 200 similar sequences produced a set of 38 unique gene matches. From the cluster of 200 genes, 167 of them matched with BRO-d from *Bombyx mori*. In previous *Baculoviridae* studies, it has been shown the BRO-d gene is involved in the regulatory process of different viral gene transcription phases. In one experiment in particular, it was shown that BRO-d deficiency strengthens viral pathogenicity and that this specific gene up-regulates the gene expression levels of increased apoptotic level of host cells (Zhang et al., 2005).

Out of the 86 cases where the full *Baculoviridae* gene matched with a host gene, a cluster of 54 similar sequences produced a set of 22 unique gene matches. From these 54 viral genes, 30 of them matched with a glycogenin isoform from various host species, with the most from *Wasmannia auropunctata*. In the search for glycogenin functions that viruses may capture from hosts, it was difficult to find a specific study that identified this particular function for viral signal peptides. Glycogen in human tissues have been identified, but glycogen function in these tissues remain unknown. Glycogenin catalyzes the formation of glucose polymers; mutations in genes encoding enzymes such as glycogenin can cause glycogen storage diseases (Adeva-Andany et al., 2016). From this review of glycogen metabolism in humans, it is a curious aspect as to why *Baculoviridae* signal peptides match this gene function.

Through this family-wide study of Baculoviridae signal peptides and genes that partially and fully match with a host gene, it can be seen that a large majority of them, when clustered into categorical mechanisms of capturing host genes, share similar functions.

C. Signal Peptides and Corresponding Hosts' Ancestral Sequences Show Similarities

To further understand the evolutionary method of how certain viral signal peptides have captured host genes, we conducted a phylogenetic analysis and reconstructed ancestral sequences of the viral sequences along with their corresponding host gene matches. In the manual Protein BLAST approach, this analysis initially focused on cases in which only the viral signal peptide matched with a host gene in the signal peptide region. MrBayes needs at least four sequences as an input so we can only use *Herpesviridae* since that was the only family that had at least four sequences that fulfilled this requirement. The output of the reconstructed sequences from PAML is shown in Figure 13.

Viral SP Tree and List of extant and reconstructed sequences

((Murid_betaherpesvirus, Murine_roseolovirus) **6** , Human_gammaherpesvirus, Papio_ursinus_cytomegalovirus) **5**;

```
Murid_betaherpesvirus          LAFCCVTAWI VGVGIEVTFC VRTQRLP
Murine_roseolovirus            VIFYLILSFV TIIGNKEHTS VQTENEP
node #6                        MPFYTISEWL TGVGWLEIFC VRTERMP
Human_gammaherpesvirus         MPRYTESEWL TGVGWLEIFT VRRERMP
Papio_ursinus_cytomegalovirus  TPDLTISLYS LGIGGNLILC VCLTRRP
node #5                        MPRYTISEWL TGVGWLEIFC VRRERMP
```

Host SP Tree and List of extant and reconstructed sequences

((QCI30416.1_chemokine, AAL15040.1_prostaglandin) **6**, AAH19172.1_Cd200_Mus_musculus, AAB37258.1_mum1_Homo_sapiens) **5**;

```
QCI30416.1_chemokine           VGTEVLQESI C
AAL15040.1_prostaglandin       VGNLVAIVVL C
node #6                        VGNLVAIVVL C
AAH19172.1_Cd200_Mus_musculus  MGSLVFRRPF C
AAB37258.1_mum1_Homo_sapiens   RGEFGMSAVS C
node #5                        VGNLVAIVVL C
```

Figure 13. PAML output for extant and reconstructed ancestral sequences for viral SP and host SP. The highlighted nodes indicate the ancestral sequences and the red bold letters indicate amino acid conservation.

The results for *Herpesviridae* signal peptide sequences and the corresponding matched host's signal peptide are shown in Figure 13. The unrooted phylogenetic trees for both the virus and host were created by MrBayes. The extant sequences in regular font from both the viral and host SP are the amino acid residues in which there were no gaps from the MSA. The ancestral sequences created from the extant sequences for viral and host SP are highlighted in yellow and labelled "node #5" and "node #6".

Between the unrooted trees that infer phylogeny, it can be seen that the two ancestral sequences for both the virus and corresponding host have the same tree order, which is an indication of their co-evolutionary similarity patterns within their respective signal peptide regions. Within both the viral and gene list of extant and reconstructed sequences, their reconstructed ancestral sequences are also fairly conserved to their respective tree's extant sequences, as indicated by the bold red font, save for a few amino acid residues. This result shows that the viral signal peptides as well as their corresponding host signal peptides have converged better in some viral and host species than others when comparing extant and reconstructed sequences respectively within each tree. However, since this tree is so small it is hard to confidently claim an evolutionary explanation yet since although viral and host SP ancestral sequences show some amino acid similarities, there may be many more that are missed from this initial Protein BLAST approach.

From the results in the automated Protein BLAST approach, the larger amount of viral and host sequences enable this analysis to be more reliable and allows us to have the ability to delve into different case studies. Therefore, the viral and matching host genes' ancestral sequences were analyzed by three case studies based on the three proposed viral mechanisms. However, because MrBayes infers phylogeny, we relied on the viral signal peptide clustering method from CLANS to use for this study. From this, the three separate groups were analyzed, both on the viral signal peptide and full host gene side. Due to the large tree sizes and amount of extant sequences used for each case study for *Baculoviridae*, only the ancestral sequences are shown in Figure 14.

Case 1. Only viral signal peptide sequence matches with host gene. Nodes 80, 156 are viral SP and nodes 16, 28 are host gene ancestral sequences.

```
node #80     MQIFVKTLTG KTITVEVESA DTIGTLK
node #156    MQIFVKTLTG KTITVDVESS DTVATLK
node #16     MQIFVKTLTG KTITLEVEPS DTIENVKAKI QDKEGIPPDQ QRLIFAGKQL EDGRTLSDYN IQKESTLHLV
node #28     MQIFVKTLTG KTITLEVESS DTIENVKAKI QDKEGIPPDQ QRLIFAGKQL EDGRTLSDYN IQKESTLHLV
```

Case 2. Partial viral gene matches with host gene. Nodes 183, 201 are viral SP and nodes 21, 38 are host gene ancestral sequences.

```
node #183    MSRVKVGFFK DTFALRYVLV RFVAKDVASL K
node #201    MAQVKIKFFK DTFTLRYVLV KFVAKDIASL K
node #21     MAQVKIGEFK FGEDTFTLRY VLEQGNQQVK FVAKDIASSL KYGNCKKAIR DHVDDKYKYT YEARLVTHTP
NSVVKKGDPL YLHPHTVLIT KSGVIQLIMK SKLPYAIELQ EWLLEEVIPQ VLCTGKYDPA IKQQEEKNKQ LVTKLIATFT
EHTNALQAVV AQKTEELFKK QEFIERIIAI KDKQIEAKDL QVTRVMTDLN RMYTGFQETM QRKDELLQVK DTQVSNLIAK
MIDLSDRAVQ YPADKRKHPV LCVTRDGTTF TAITGQKTYV ESQKHKRNID AANIVVENIR PNPTVDWNNA TDRLQSKRSK
RSISFDSLEE AQQFENRIKY LLKNANNIN
node #38     MAQVKIGEFK FGEDTFTLRY VLEQGNQQVK FVAKDIASSL KYGNCKKAIR DHVDDKYKYT YEARLVTHTP
NSVVKKGDPL YLHPHTVLIT KSGVIQLIMK SKLPYAIELQ EWLLEEVIPQ VLCTGKYDPA IKQQEEKNKQ LVTKLIATFT
EHTNALQAVV AQKTEELFKK QEFIERIIAI KDKQIEAKDL QVTRVMTDLN RMYTGFQETM QRKDELLQVK DTQVSNLIAK
MIDLSDRAVQ YPADKRKHPV LCVTRDGTTF TAITGQKTYV ESQKHKRNID AANIVVENIR PNPTVDWNNA TDRLQSKRSK
RSISFDSLEE AQQFENRIKY LLKNANNIN
```

Case 3. Full viral gene matches with host gene. Nodes 55, 106 are viral SP and nodes 23, 42 are host gene ancestral sequences.

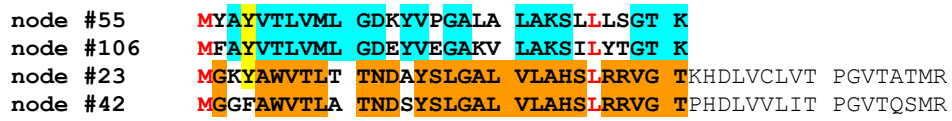


Figure 14. ASR of viral SP and corresponding host gene matches for all three viral mechanisms. Bold red font indicates conservation across all sequences. Yellow highlights indicate conservation across almost all sequences. Blue and orange highlights indicate conservation amongst viral or host sequences, respectively.

In each of the cases from Figure 14, it can be seen that the viral signal peptide ancestral sequences are very similar to their host gene ancestral sequences. It can also be seen that two pairing viral SP or host nodes within their respective trees are similar to each other, as indicated in blue and green highlighted residues, respectively (i.e. node #80 is highly similar to node #156 and node #16 is highly similar to node #28). It can also be seen that there is also high conservation across all viral SP and host ancestral sequences, indicated in red bold font or through a majority of the ancestral sequences in yellow highlighted residues. There are only a few amino acid residues that are different when comparing viral signal peptide and host gene ancestral nodes. However, the co-evolutionary mechanism can not be determined from just the ancestral nodes as further analysis into their respective trees must be done. The viral SPs from CLANS span across different viral species with a diverse set of host species for its corresponding host gene matches. With these ancestral sequences, the conservation of a majority of amino acid residues through phylogenetic inference suggests that there is convergence among these viral and host extant sequences, however diverse they may be in terms of *Baculoviridae* species.

IV. CONCLUSION

The results of this project suggest that signal peptides do show some kind of functional role, not only in eukaryotic proteins but in viral proteins as well. It was also demonstrated that viruses exploit all evolutionary mechanisms to hijack genes from their host in any way possible, as seen in the Protein BLAST results. The three proposed mechanisms by which they do so involve only the viral signal peptide being able to replicate from a host gene, part of the viral gene being captured from a host gene, and the full viral gene reproducing from a host gene.

Although the exact mechanism by which the virus SP could capture host genes and its preference in genes of particular functions are still unknown, the functional clustering that we have created for some of those host genes has provided insight into what some of those functions may be. The five clusters created from the manual Protein BLAST approach showed functional clusters related to inflammatory response as well as glycogenin, both functions of which are not yet fully elucidated. Although the other known gene functions within the functional clusters, such as glycosylation and DNA replication, have been studied for viral mimicry,

it is not yet well understood in cases where only the viral signal peptide shows similarity and how it may have evolved to "prefer" these specific genetic functions to hijack.

From the manual BLAST results for *Herpesviridae* viral family, reconstructed ancestral sequences were created for the viral and host signal peptides. Both the viral and host trees have different species within their respective trees. Both their respective ancestral sequences shared high sequence similarity; node #5 was similar to node #6 in both trees. When comparing their ancestral sequences to their extant sequences, the viral and host SPs shared high sequence similarity in many instances, which can indicate convergence by which both the viral and host species, despite being in distant groups, have evolved similar traits as a result of adapting to similar environments (i.e. selective pressure of the virus to capture a specific mechanism from the host). However, this can also indicate divergence by which both the viral and host species become different from their ancestors following gene transfer (i.e. viruses jump between hosts and horizontally transferred genes to eukaryotic genomes). When looking at both viral and host gene ancestral sequences, MrBayes created phylogenetic trees in which their ancestral sequences or nodes #5 and #6 were at the same location, which leads to speculation of coevolution. However, due to the small amount of sequences used in this case, it cannot be confidently determined as to whether, in the event of viral capture of host genes, it is coevolution or independent convergence evolution. Therefore, the automated BLAST results for *Baculoviridae* were used for phylogenetic analysis in hopes of understanding ancestry of viral SPs on a family-wide scale. This analysis sought an evolutionary explanation within the three proposed mechanisms by which viruses capture host genes. For all three cases, it was seen that the ancestral sequences for the viral SP region were extremely similar to the ancestral sequences of its corresponding host genes, indicating some form of amino acid sequence conservation for particular regions of both the virus and host. Amino acid conservation in a particular residue from the viral and host gene ancestral sequences consisted of the following occurrences: all sequences in both trees have the same amino acid, most sequences in both trees have the same amino acid, and both of the viral, both sequences in their own tree have the same amino acid, and all sequences in both trees have a differing amino acid. Calculating the distance between the viral and host sequences would further our understanding in whether convergent evolution occurred in a single event or spanned across multiple events.

As we continue this exploratory research, it is evident that there are many questions still waiting to be examined. One possible question could be posed towards the derivation for the remainder of the full viral gene aside from the signal peptide region in the cases where only the viral signal peptide or the partial viral gene is similar to the host gene. A future direction would be the comparison of the evolution of these viral signal peptides across other categorical schema as well. In terms of Baltimore classification, which is a classification system that places viruses into one of seven groups depending on their DNA or RNA properties (strandedness, sense, and method of replication) (Hulo et al., 2011); this was another layer of comparison we were considering in our initial methodology but were not able to implement. However, it is a useful categorical tool that could potentially play a role in defining viral and host SP functionality patterns. Host groups could also be compared in the study of particular genes that these viral signal peptides have more similar functions to, which could lead to further convergent and divergent evolutionary questions as well.

As one of the first comprehensive evolutionary studies that links the viral-host interaction to investigate signal peptides, these findings have the ability for us to further understand this post-targeting quality of signal

peptides that is being seen in literature, which has the potential to later create more predictive models for viral replication. Therefore, the culmination of this preliminary study is extremely exciting, especially as it will eventually evolve into more questions that will expand upon our understanding of which evolutionary mechanisms drive viral signalling and replication patterns and thus pave paths to a fuller comprehension of viral-host co-evolution.

V. REFERENCES

Adeva-Andany, M. M., González-Lucán, M., Donapetry-García, C., Fernández-Fernández, C., & Ameneiros-Rodríguez, E. (2016). Glycogen metabolism in humans. *BBA clinical*, *5*, 85-100.

Alcami, A. (2003). Viral mimicry of cytokines, chemokines and their receptors. *Nature Reviews Immunology*, *3*(1), 36-50.

Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., ... & Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature biotechnology*, *37*(4), 420-423.

Bhandari, B. K., Gardner, P. P., & Lim, C. S. (2021). Razor: annotation of signal peptides from toxins. *bioRxiv*, 2020-11.

Bigot T, Temmam S, Pérot P and Eloit M. RVDB-prot, a reference viral protein database and its HMM profiles [version 2; peer review: 2 approved]. F1000Research 2020, 8:530.

Burdukiewicz, M., Sobczyk, P., Chilimoniuk, J., Gagat, P., & Mackiewicz, P. (2018). Prediction of signal peptides in proteins from malaria parasites. *International journal of molecular sciences*, *19*(12), 3709.

Camacho C. *et al.* (2009) "BLAST+: architecture and applications"

Caporale, Marco, et al. "The signal peptide of a simple retrovirus envelope functions as a posttranscriptional regulator of viral gene expression." *Journal of virology* 83.9 (2009): 4591-4604.

Chan, D., Ho, M. P., & Cheah, K. E. (2001). Aberrant signal peptide cleavage of collagen X in Schmid metaphyseal chondrodysplasia: implications for the molecular basis of the disease. *Journal of Biological Chemistry*, *276*(11), 7992-7997.

De Souza, G. A., Leversen, N. A., Målen, H., & Wiker, H. G. (2011). Bacterial proteins with cleaved or uncleaved signal peptides of the general secretory pathway. *Journal of proteomics*, *75*(2), 502-510.

Drexler, M. (2010). Prevention and Treatment. In *What you need to know about infectious disease*. essay, National Academies Press (US).

Fabre, E., Jeudy, S., Santini, S., Legendre, M., Trauchessec, M., Couté, Y., ... & Abergel, C. (2017). Noumeavirus replication relies on a transient remote control of the host nucleus. *Nature communications*, *8*(1), 1-12.

Goulding, J. (n.d.). *Viruses: Introduction*. Pathogens and Disease. https://www.immunology.org/public-information/bitesized-immunology/pathogens-and-disease/viruses-introduction.

Goulding, J. (n.d.). *Virus replication*. Pathogens and Disease. https://www.immunology.org/public-information/bitesized-immunology/pathogens-and-disease/virus-replication.

Gustin, J. K., Moses, A. V., Früh, K., & Douglas, J. L. (2011). Viral takeover of the host ubiquitin system. *Frontiers in microbiology*, *2*, 161.

Healthline. (2019, May 9). *Viral disease definition*. Viral Diseases 101.

Hegde, R. S., & Bernstein, H. D. (2006). The surprising complexity of signal sequences. *Trends in biochemical sciences*, *31*(10), 563-571.

Hofmann, K. J., & Schultz, L. D. (1991). Mutations of the α-galactosidase signal peptide which greatly enhance secretion of heterologous proteins by yeast. *Gene*, *101*(1), 105-111.

Huang, Y., Wilkinson, G. F., & Willars, G. B. (2010). Role of the signal peptide in the synthesis and processing of the glucagon‐like peptide‐1 receptor. *British journal of pharmacology*, *159*(1), 237-251.

Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc. 2009;4(1):44-57.*

Infectious Diseases Society of America. (n.d.). *Facts about ID*. IDSA Home. https://www.idsociety.org/public-health/facts-about-id/.

Kapp, K., Schrempf, S., Lemberg, M. K., & Dobberstein, B. (2009). Post-targeting functions of signal peptides. *Protein transport into the endoplasmic reticulum*, 1-16.

Lindemann, D., Pietschmann, T., Picard-Maureau, M., Berg, A., Heinkelein, M., Thurow, J., ... & Rethwilm, A. (2001). A particle-associated glycoprotein signal peptide essential for virus maturation and infectivity. *Journal of virology*, *75*(13), 5762-5771.

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., ... & Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research*, 47(W1), W636-W641.

Mihara, T., Nishimura, Y., Shimizu, Y., Nishiyama, H., Yoshikawa, G., Uehara, H., Hingamp, P., Goto, S., and Ogata, H.; Linking virus genomes with host taxonomy. *Viruses* 8, 66 doi:10.3390/v8030066 (2016).

Paetzel, M., Karla, A., Strynadka, N. C., & Dalbey, R. E. (2002). Signal peptidases. *Chemical reviews*, *102*(12), 4549-4580.

Ronquist, F., M. Teslenko, P. van der Mark, D.L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M.A.Suchard, and J.P. Huelsenbeck. 2012. MRBAYES 3.2: Efficient Bayesian phylogenetic inference and model selection across a large model space. Syst. Biol. 61:539-542.

Tarakanova, V. L., Leung-Pineda, V., Hwang, S., Yang, C. W., Matatall, K., Basson, M., ... & Virgin IV, H. W. (2007). γ-herpesvirus kinase actively initiates a DNA damage response by inducing phosphorylation of H2AX to foster viral replication. *Cell host & microbe*, *1*(4), 275-286.

Thaker, S. K., Ch'ng, J., & Christofk, H. R. (2019). Viral hijacking of cellular metabolism. *BMC biology*, *17*(1), 1-15.

ViralZone: a knowledge resource to understand virus diversity. Hulo C, de Castro E, Masson P, Bougueleret L, Bairoch A, Xenarios I, Le Mercier P. Nucleic Acids Res. 2011 Jan;39(Database issue):D576-82.

von HEIJNE, G., & GAVEL, Y. (1988). Topogenic signals in integral membrane proteins. *European Journal of Biochemistry*, *174*(4), 671-678.

World Health Organization. (2020, July 29). Zoonoses. https://www.who.int/news-room/fact-sheets/detail/zoonoses.

Wu, J. Z. (2001). Internally located signal peptides direct hepatitis C virus polyprotein processing in the ER membrane. *IUBMB life*, *51*(1), 19-23.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution*, *24*(8), 1586-1591.

Zhang, L., Leng, Q., & Mixson, A. J. (2005). Alteration in the IL‐2 signal peptide affects secretion of proteins in vitro and in vivo. *The Journal of Gene Medicine: A cross‐disciplinary journal for research on the science of gene transfer and its clinical applications*, *7*(3), 354-365.

Zhang, Y., Wei, M., Li, J., Zhang, T., Quan, Y., Shu, T., & Yu, W. (2016). Effect of bro-d deletion on the viral transcription and cell apoptosis in silkworm (Bombyx mori). *Journal of Agricultural Biotechnology*, *24*(7), 1073-1082.

Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. (2018). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J Mol Biol.* S0022-2836(17)30587-9.

VI. APPENDICES

Table 5. Viral signal peptide sequences that produced a host match hit on Protein BLAST along with further annotation.

Viral family	Viral SP	Viral full gene	Host gene [host]	Full gene match? Query:Subject frame
Adenoviridae	>Murine mastadenovirus A_824457 MCFKVRLLGGVPGDMGLG GVPGDMGLGGVPGDMGLG GVPGDMGLGGVPGD	>acc|REFSEQ|AP_000364.1|REFSEQ|AC_000012|E4 ORFB [Murine mastadenovirus A] MCFKVRLLGGVPGDMGLGGVPGDMGLGGVPGDMGLG GVPGDMGLGGVPGDMGLGGVPGNIGSRPGNTWRCARE HWQSSREHWQSSRKHLAACQGFFHLQSLYIKLCGLLKC ILDRF	Collagen alpha-1 XXIV chain isoform 1 precursor [Mus musculus]	N, 10-73, 1035-1100
Adenoviridae	>Simian mastadenovirus C_955385 MALSVQDCARLTGQSVPTM ERFRPLRNIWNRVREFTRAA TTSAGITWLSR	>acc|REFSEQ|YP_213967.1|REFSEQ|NC_006879|pTP [Simian adenovirus 1] MALSVQDCARLTGQSVPTMERFRPLRNIWNRVREFTRA ATTSAGITWLSRYIYHYHRLMLDDLSPGAPATVGWPLYR EPPPHFLVGYQYLVRTCNDYVFESRAYSRLKYTEITQPG MQVVNWSVMANCTYTINTGAYHRFVDLDDFQTTLTQI QQAVLAERVVADLALLQPLRGFGSTRMADRGEAEVPVE RLMQDYYKDLRRCQSEAWGMAERLRIQQAGPKDVVLL ATIRRLKTAYFNYIISSITSQLPADAPQRPTVLSLPCDCDW IDAFLEKFSDPVDLDVLRSLHGVPTQQLIKCIVSAVSLPD GPNHLPSLHGGALRGGVFELRPRENGRAVTETMRRRRG EMIERFVDRLPVRRRRRRPVPVAELPEEPVPLEEEEPEEE AEAPRGAFEQEVRDTVADLIRLLQEELTVAARNSQFFNF AVDFYEAMERLEAIGDINEMTLRRWIMYFFVCEHIATTL NYLFQRLRNYAVFARHVELNVAQVVMRARDSFGNVVY SRVWNENGLNAFSQLMRRISNDLAATVERAGHGDLREE EVEQFMAEIAYQDNSGDVQEILRQAAVNDADIDSVELSF RFRTRGPVVFTQRRHIQDLNRRVVAHASSLRARHQPLPN PHENVPLPPLPPGAEPPLPPGARPRRVR	kelch-like protein 32 isoform X1 [Papio anubis]	N, 14-49, 411-451
Adenoviridae	>Bat mastadenovirus WIV17_38246 MDPFSTMFSQRQDIYVPPRY SAPSEGRNSITYSQLPPLYDT TKIYLIDNK	>acc|REFSEQ|YP_009362843.1|REFSEQ|NC_034626|penton base [Bat mastadenovirus WIV17] MDPFSTMFSQRQDIYVPPRYSAPSEGRNSITYSQLPPLYD TTKIYLIDNKSADISSLNYQNDHSNFLTSVVQNSDFSPIE AATQTIKFDDRSRWGANLKTILYMNMPNVTDFMFSNTF KVKLMISKKDNVPVYDWVDLKIPEGNFSVMKIIDLMNN AIVDHYMSGPRANGVKEEDIGVKIDTRNFMLGYDPTTE VVTPGVYTYEAFHPDIVLLPGCAIDFTLSRLNNVLGIRK KLPFQHGFIISYEDLGGGNIPGLMNLVSYKNGQGEIEVV TKDEKDRSYHIGEDNSVPKTFTSYRSWYLSYNYGDEEQ SIRNSTLLTNPDITCGVEQVYWSLPDLACEPVTFKASYN VNNYPVVGTEVLPMQSRSFFNAQAVYSQLVSERTNQTH VFNRFPENQILVRPPAPTISSISENVPAVTDHGVVPIKNTLS GVQRVTVTDARRRICPYVYKAVGVLSPKVLSSRTL	lysosomal H+-transporting ATPase V1 subunit A, partial [Rousettus leschenaulti]	N, 11-46, 123-152
Adenoviridae	>Porcine adenovirus 5_884640 MVLLFNHGNVDYLVSPGQR VCQLILEKISTPEVLEVASLD DTDRGASGFG	>acc|GENBANK|AAK00135.1|GENBANK|AF221544|ORF2 [Porcine adenovirus 5] MSSDRLLFVRLCRDAVPPRRASPGAAGYDLCSTNYVVV RARDKALVPTGLRLQIPPGHYGRIAPRSGLAVTSGLDVG AGVIDADYRGEVMVLLFNHGNVDYLVSPGQRVCQLILE KISTPEVLEVASLDDTDRGASGFGSTG	Swine dUTPase in complex with alpha,beta-iminodUTP and magnesium ion [Sus scrofa]	Y, 2-133, 4-135
Adenoviridae	>Bat AAV SC2991_36054 FMLPQYGYLTLNGPGTNNN NLSTPSSAFYCLEYFPSQML RTGNNFVFTYE	>acc|GENBANK|AFH02748.1|GENBANK|JN857334|capsid [Bat AAV SC2991] FMLPQYGYLTLNGPGTNNNNLSTPSSAFYCLEYFPSQML RTGNNFVFTYEFEKVPFHSMFMHNQALDRLMNPLVDQ YLWYLDATNGNNLTFRKAGAKNFPEYFRNWIPGPGCRN QQWNKVGTK	hypothetical protein mRhiFer1_008815 [Rhinolophus ferrumequinum]	N, 17-94, 77-221
Adenoviridae	>Bat adeno-associated virus_36359 LPAYNNHLYRQIQSSGTGDG TYFGYSTPWGYFDFNRFHC HFSPRDWQRLI	>acc|GENBANK|ADD26595.1|GENBANK|GU226880|capsid protein [Bat adeno-associated virus] LPAYNNHLYRQIQSSGTGDGTYFGYSTPWGYFDFNRFHC HFSPRDWQRLINNHWGIRPKRLHFKLFNIQVKEVTTTDG TTTIANNLTSTIQVFADTEYQLPYVLGNAHEGCLPPFPAD VFMLPQYAYLTLNANPTNQGAAALSLPQSAFYCLEY	hypothetical protein mRhiFer1_009420 [Rhinolophus ferrumequinum]	N, 1-92, 47-139
Adenoviridae	>Bat parvovirus_38665 LGGTEFRIDRKHKDSQLLPQ	>acc|GENBANK|AIF74234.1|GENBANK|KJ641679|NS1 [Bat parvovirus] LGGTEFRIDRKHKDSQLLPQTPVIISTNNNVYEVTGGNT	uncharacterized protein LOC114235279 [Eptesicus fuscus]	N, 1-77, 13-89

	TPVIISTNNNVYEVTGGNTV TGVHSKPLRE	VTGVHSKPLRERIVQLNMMKQLPSTFGEISCEEVAAWLL ACYARFTITLQGFYEQWGLEVTPNDFPLNDPCVGHLQD LILDDRGICVTCGNRLPLDADYIPEPTAVGDVPSTPTGTS SFKILTPTKLHFWCTEFDLSYLDSPGPAAGKRGLVPEDSE PSASRGPDPPAAPKKPKRRLNLEEVSIRDDWSSQPRDEH EARWHAEHVEREERAREEPVARAAGLTPSEWGEKMGV VPQGEGRSPLVLHCFETLEESEDDTN		
Adenoviridae	>Bat parvovirus_38673 MAPTNRKPGGWVIPGYKYL GPFNSLDNGEPVNAADKAA RDHDFAYQSYIE	>acc|GENBANK|AIF74218.1|GENBANK|KJ641669|VP1 [Bat parvovirus] MAPTNRKPGGWVIPGYKYLGPFNSLDNGEPVNAADKA ARDHDFAYQSYIEKGQNPYLNFNKADLQFIDDLRNDWS FGAIVGSAAFKAKQAIAPSLSNDNERPGTSKQAAAADR AAKRKLYFARSNKGSKVSKMDNPGGDNNEPQDNGSGP SSGQDRAGGGGGGGGGGANPGGGSGGHGVGVSTGGW KAGCVFTDKYIITTNTRQWYCPIYNGHKYTKMTSDDAG QNGYTMWEGIATPWGHFNFNAYTSHFSPQDWQRLTNE YKRWRPVKMRVKIYNLQIKQIVNLGLDTLYNNDLTAGV HIFCDGSHQFPYSQHPWDEGTMPELPNEIWRLTQYAYYQ DNRDLVDQSHDEVTPLDVEKLLRVNTPLFILESASHQVL RTGEETSFDFDFSSGWVFNDRAYAPPQADYNPLIATRRH YATWNNQTKEYAYNRYSPYAKPSQWVPGPSLQYKGDTR TSVDPAKERGPLTVTYGPPGTHRQDEDNQEGQFMPSED PTKPQSVQRIGYSIAPVNGACSGLDPHTLAYDSSPHSRED RYMTVRNIDLDMTRYNALEVFDGSTINGGRARLKNMW MYPNQVWNSTPISRANPIWLKTPRTDNHTLLDTSDGTLP MEHPPGIIFIKVAKIPIPTENNADSYLNLYVTGQVTCEILW ECERYQTKNWRPEIRNTAQVFKEGGLFDFN	hypothetical protein mRhiFer1_009419 [Rhinolophus ferrumequinum]	N, 165-301, 4-139
Adenoviridae	>Bat rhabdovirus_38842 FGLKAKERELKTEGRFFALM TLSLRIYFVATEWLIAKYILP VFPEITMGD	>acc|GENBANK|QCI56506.1|GENBANK|MK004996|RNA-dependent RNA polymerase [Bat rhabdovirus] FGLKAKERELKTEGRFFALMTLSLRIYFVATEWLIAKYIL PVFPEITMGDSFVSLQKKIFNITKTQTQSERSCVNFVVSL DYEKWNAH	endogenous bornavirus-like L-1 [Eptesicus nilssonii]	Y, 3-87, 466-546
Adenoviridae	>Bat betaherpesvirus B7D8_37040 MSSPDLADTCNSKSRFIMLN IYGPVDWSSEIPGLSEAFYTD MDNRDFEAD	>acc|GENBANK|AFK83966.1|GENBANK|JQ805139|B125 [Bat betaherpesvirus B7D8] MSSPDLADTCNSKSRFIMLNIYGPVDWSSEIPGLSEAFYT DMDNRDFEADLEERRDLNRRWLSVCEKILPIVLTYWKA KILTSEPAQWFAQAECYDDRPLMKLMLQIGAGTSLGSSL SRHLTHMRRLIIARVYEDHHPMMGEWFFVSRTGLSGGN VRVFELDYITSTLLQVTVPIYQWSWTNITTLTDAQLSAR NHLFLDRAEEGEDTGSLSEKYRLKPRHRKRQSKGARKE PFRRLVLHLVKLGISNDAQWANKDVSLYMQTMATEYRR NRGKRALDTAVKYILAKKTILDYIVPDECVVREPSNLEI MYNPIFTLLVQQGYDPKEVGVFMLAWMQSSVLGRNVI WLQSPFDAELAAFVSGVLENVPLICILPDKPSQKELEDC VEKMVIWWRSARISKRVSATLIAILSGQCTRIFNGNEFVT VKQTPCIVTSKDDVLSVDIQEHRLYGVRRYVTKLCFEKS VDVKLSAWSVIEFFQWAKGKIFACDLSETLEISNGVAQL KRYGEEELKSRMEDYI	PREDICTED: protein OSCP1 isoform X2 [Hipposideros armiger]	N, 718-854, 4-139
Adenoviridae	>Bat bocavirus_37080 MAPINRKPRGWVLPGYKYL RPFNPLENGEPVNKADEAAR KHDIAYNQYLE	>acc|REFSEQ|YP_009229910.1|REFSEQ|NC_029300|VP1 protein [Bat bocavirus] MAPINRKPRGWVLPGYKYLRPFNPLENGEPVNKADEAA RKHDIAYNQYLEKGQNPYLNFNKADQTFIDDLRHDNSF GGAIGKTYFSIKKTLFPHLAGEGEPPAKRSKPGERAAKR KQYFARSNRQAKQQRMEQPDAAAADETSTTPHSRSARS AGGGGGGPGGSGGSGGGGNSVGFSTGGWDAGSIFTDNI IITTNTRQWYAPIYDGHKYKKHKAGDNGGTDSDWSGIS TPWGVFNFNCYASHFSPQDWQRLTNEYVKWRPKSMTV KIYNLQIKQIVKLGADTLYNNDLTAGVHVFCDGSHRYPY SQHPWDKDTMPELPDEIWKLPQYAYYQIQDDLAEKASG DNVINPDDMSKNLIRDAPLYILETSTHEVLRTGEDTSFNF TFSSGWVMNDRAFCPPQADFNPLSQTRRYSAHRDNNK NIFGRYNPYKKPSNWFPGPALSYTGNLKKNNGYQTGKT TAPITVTYQHPHEASRNSTATATGVSMPNEEGARKSGYN VTPVNHAYAAAENITLGYDSAPQSEAIENITTVDVDRDM TRWGSLWACDGVEGTYAQIKHSWMFPMQAWNGIPIAR YTPIWVKIPRTDFHTLLDSPDGTLPMTHPPGNIFCKVAKI PVPSTDTESFLNIYVTGQVTCEIVWEAERYQTKNWRPEI RNDVKSFKDAHLYDFDTAGKYNTPENFNEEMPTRMGIN RVL	hypothetical protein mRhiFer1_009419 [Rhinolophus ferrumequinum]	N, 170-295, 15-139
Adenoviridae	>Bat bornavirus_37097 GLPEGCIGKMAEDLISTCDW INIKGAVSQHLATLLSTLLSE KGVHTPELS	>acc|GENBANK|AGN73381.1|GENBANK|KF170228|RNA-dependent RNA polymerase [Bat bornavirus] GLPEGCIGKMAEDLISTCDWINIKGAVSQHLATLLSTLLS EKGVHTPELSVIMGGTLTHSLPTASDDRAGLA	uncharacterized protein LOC118671920 [Myotis myotis]	Y, 1-72, 314-385

Adenoviridae	>Bat gammaretrovirus_37886 DLREVNKRVQDIHPTVPNPY NLLSSIPPSHVWYSVLDLKD AFFCLRLHPA	>acc|GENBANK|AGN73375.1|GENBANK|KF170221|pol protein [Bat gammaretrovirus] DLREVNKRVQDIHPTVPNPYNLLSSIPPSHVWYSVLDLK DAFFCLRLHPASQSIFAFEWRDPDSGITXQLTWTRLPQGF KNSPTIFDEALHQDLAAFRADHPQVTLLQYVDDLLLAG ATKEDCETGTGAXLLELSRLGYRVSAKKAQICQRRVTY LGYALEGGKRWLTEARKSTVTQIPRPQTPRQVREFLGTA GFCRLWIPGFATLAEPLYPLTKQGNPFEWGETHQKAFDN IKKALLSAPALALPDVAKPFTLFIDERKGVARGVLTQTLG PWKRPVAYLSKRLD	LOW QUALITY PROTEIN: uncharacterized protein LOC118676587, partial [Myotis myotis]	Y, 1-287, 512-798
Adenoviridae	>Bat gammaretrovirus_37887 GNMAFTALPTNWTGLCTLA MVIPDIDIIPGDEPVPLPSLDY IASRQRRAV	>acc|GENBANK|AHA85401.1|GENBANK|KF547869|envelope glycoprotein [Bat gammaretrovirus] GNMAFTALPTNWTGLCTLAMVIPDIDIIPGDEPVPLPSLD YIASRQRRAVQLIPLLIGLGVARAVASGTSGLGVALHSYN KLSNQLIEDVQALSGTISDIQDQIDSLAEVVLQNRRGLDL LTAEQGGICLALQERCCFYANKSGIVRDKIR	syncytin-1-like [Myotis myotis]	Y, 1-151, 349-499
Adenoviridae	>Bat gammaretrovirus_37889 VAEGKGVAKGVLTQKLGPW KRPVAYLSKRLDPVAAEWPG CLRIIAAAALL	>acc|GENBANK|AHA85400.1|GENBANK|KF547868|polymerase [Bat gammaretrovirus] VAEGKGVAKGVLTQKLGPWKRPVAYLSKRLDPVAAEWP GCLRIIAAAALLVKEATKLTFGQTLQVTSAHNLEGLLRA PPDRWMTNARVTQYQILLLDLPRITFHQTVALNPATLLPE ADESQPLHCCRDTLDTLTSLFVDTFSGRVEAYPSRREIA	LOW QUALITY PROTEIN: uncharacterized protein LOC118723079 [Pipistrellus kuhlii]	N, 1-140, 1008-1147
Adenoviridae	>Bat herpesvirus_37972 NSVYGFTGVASGMLPCLKIA ETVTFQGRRMLEMSKDLIES ITPTRLQHII	>acc|GENBANK|ALH21064.1|GENBANK|KR261855|DNA polymerase [Bat herpesvirus] CNSVYGFTGVASGMLPCLKIAETVTFQGRRMLEMSKDL IESITPTRLQHIIGRPLSCTPSAYFRVIYGDTDSV	DNA polymerase delta catalytic subunit [Pipistrellus kuhlii]	Y, 2-73, 648-709
Adenoviridae	>Bovine gammaherpesvirus 4_50699 KKKLCRGHYLWALGCYMLL AVVSLRLSLRFKCDVDSLDL ESRDFQSQHCR	>acc|GENBANK|AAN18276.1|GENBANK|AY143156|Bo17 [Bovine gammaherpesvirus 4] KKKLCRGHYLWALGCYMLLAVVSLRLSLRFKCDVDSL DLESRDFQSQHCRDILYNNLKLPAKRSINCSGITRGDQEA VAQALLDNLEVKKKRPPLTDTYYLNITRDCERFRAQRKF IQFPLSKEELDFPIAYSMVVHEKIENFERLLRAVYAPQNIY CVHVDVKSPETFKEAVKAIISCLPNVFMASKLVPVVYAS WSRVQADLNCMEDLLQSSVPWKYLLNTCGTDFPIKTNA EMVLALKMLNGKNSMESEIPPESKKNRWKYSYEVTDTL YPTSKMKDPPPDNLPMFTGNAYFVASRAFVQHVLDNPK SQRLVEWVKDTYSPDEHLWATLQRAPWMPGSVPSHPK YHISDMTAIARLVKWQYHEGDVSMGAPYAPCSGIHRRAI CIYGAGDLYWILQNHHLLANKFDPRVDDNVLQCLEEYL RHKA	mucus-type core 2 beta-1,6-N-acetylglucosaminyltransferase [Bos taurus]	Y, 1-428, 7-434
Adenoviridae	>Bovine gammaherpesvirus 4_50774 MKMAGWKKKLCPGHHLWA LGCYMLLAVVSLRLSLRFKC DVDSLDLESRDF	>acc|GENBANK|AEL29824.1|NEIGHBOR|JN133502|viral beta-1,6-N-acetylglucosaminyltransferase [Bovine gammaherpesvirus 4] MKMAGWKKKLCPGHHLWALGCYMLLAVVSLRLSLRF KCDVDSLDLESRDFQSQHCRDMLYNSLKLPAKRSINCSG ITRGDQEAVVQALLDNLEVKKKRPPLTDTYYLNITRDCE RFKAQRKFIQFPLSKEELDFPIAYSMVVHEKIENFERLLR AVYAPQNIYCVHVDVKSPETFKEAVKAIISCFPNVFMAS KLVPVVYASWSRVQADLNCMEDLLQSSVSWKYLLNTC GTDFPIKTNAEMVLALKMLKGKNSMESEVPSESKKNR WKYRYEVTDTLYPTSKIKDPPPDNLPMFTGNAYFVASRA FVQHVLDNPKSQILVEWVKDTYSPDEHLWATLQRAPW MPGSVPSHPKYHISDMTAIARLVKWQYHEGDVSMGAPY APCSGIHRRAICIYGAGDLYWILQNHHLLANKFDPRVDD NVLQCLEEYLRHKAIYGTEL	mucus-type core 2 beta-1,6-N-acetylglucosaminyltransferase [Bos taurus]	Y, 1-440, 1-440
Adenoviridae	>Bovine gammaherpesvirus 4_50701 KLCRGHHLWALGCYMLLAV VSLRLSLRFKCDVDSLDLES RDFQSQHCRDI	>acc|GENBANK|AAN18277.1|GENBANK|AY143157|Bo17 [Bovine gammaherpesvirus 4] KLCRGHHLWALGCYMLLAVVSLRLSLRFKCDVDSLDLE SRDFQSQHCRDILYNNLKLPAKRSINCSGITRGDKEAVV QALLDNLEVKKRRPPLTDTYYLNITRDCERFRAQRKFIQ FPLSKEELDFPIAYSMVVHEKIENFERLLRAVYAPQNIYC VHVDVKSPETFKEAVKAIISCLPNVFMASKLVPVVYASW SRVQADLNCMEDLLQSSVPWKYLLNTCGTDFPIKTNAE MVLALKMLNGKNSMESEIPPESKKNRWKYSYEVTDTLY PTSKMKDPPPYNLPMFTGNAYFVASRAFVQHVLENPKS QRLVEWVKDTYSPDEHLWATLQRAPWMPGSVPSHPKY HISDMTAIARLVKWQYHEGDVSMGAPYAPCSGIHRRAIC IYGAGDLYWILQNHHLLANKFDPRVDDNVLQCLEEYLR HKA	beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase 3 [Bos taurus]	Y, 1-426, 9-434
Adenoviridae	>Bovine gammaherpesvirus 4_50839 MYAECWICKGSEGIIDVKYC	>acc|GENBANK|AAA96267.1|GENBANK|M60043|unknown [Bovine gammaherpesvirus 4] MYAECWICKGSEGIIDVKYCHCIGDLQYVHSECLVHWIR	probable E3 ubiquitin-protein ligase MARCH10 isoform X3 [Bos	N, 5-50, 669-722

	HCIGDLQYVHSECLVHWIRV SGTKQCKFCQ	VSGTKQCKFCQYTYILKEKERRRPRVPHYWTEAQTATF YRWGCGFLFIIFSWVVSFMLTVLSACYSWSTALIVLVVS VVCIYPILFGVYFVVMVLCIANLLTEFDSYNSYEMSPCV SDECLSIKID	taurus]	
Adenoviridae	>Bovine gammaherpesvirus 6_50884 DGQQIRVFSCLLAAAKNRN YILPMPSAGNTDGYQGATVI NPLSGFYNTPV	>acc|GENBANK|AIT68770.1|GENBANK|KM437997|DNA polymerase [Bovine gammaherpesvirus 6] DGQQIRVFSCLLAAAKNRNYILPMPSAGNTDGYQGATVI NPLSGFYNTPVLVVDFASLYPSIIQAHNLCYSTMIEHDQL SRFPELKKADYETFVISSGPVHFVKKHKTESLLASLLTA WLAKRKAIRKELATCQDSKLKTILDKQQLAIKVTCNAV YGFTGVASGLLPCLKIAETVTLQ	DNA polymerase delta catalytic subunit isoform X1 [Bos taurus]	Y, 2-177, 549-720
Adenoviridae	>Bovine leukemia virus_51406 LVERTDGLLELLLSKYHLDE PHLPMTQALSRALWTHNQI NLLPIL	>acc|GENBANK|AAB62995.1|GENBANK|U45273|reverse transcriptase [Bovine leukemia virus] LVERTDGLLELLLSKYHLDEPHLPMTQALSRALWTHNQI NLLPIL	cysteine-rich secretory protein LCCL domain-containing 2 precursor [Bos taurus]	N, 9-39, 31-66
Adenoviridae	>Bovine leukemia virus_51592 MRTIQRGWAPNHIWQADITH YKYKQFTYALHVFVDTYSG ATHASAKRGLT	>acc|GENBANK|ABB90624.1|GENBANK|DQ288977|integrase [Bovine leukemia virus] MRTIQRGWAPNHIWQADITHYKYKQFTYALHVFVDTYS GATHASAKRGLTTQMTMEGLLEAIVHLGRPKKLNTDQG ANYTSKTFVRFCQQFGISLSHHVPYNPTSSGLVERTNGL LKLLLSKYHLDEPHLPMTQALSRALWTHNQINLLPILKT RWELHHSPPLAVISEGGETPKGSDKLFLYKLPGQNNRRW LGPLPALVEASGGALLATNPPVWVPWRLLKAFKCPKND GPEDAK	TPA: Pro-Pol-like [Bos taurus]	N, 1-148, 562-705
Adenoviridae	>Bovine polyomavirus 2_52571 MDSILSRQEKRELCDLLEIPH HCYGNIPMMKAQYKKMCL LYHPDKGGDGA	>acc|REFSEQ|YP_009110706.1|REFSEQ|NC_025811|large T antigen [Bovine polyomavirus 2] MDSILSRQEKRELCDLLEIPHHCYGNIPMMKAQYKKMC LLYHPDKGGDGAKMVKLNSLWTSFQDEVTKLRAEVNF FSYQENEPIYGTPQFKTWWYQQHGSYFSESQSTSKPSRG GPSRRFSRATSPDSTTSTSTTGSGIFTESSQHSDVGPAGSQ TSSAKTTSSSGRSRAPSGGGTGVFGSAAYSREFRSETQDS DLYCDETLGSSPESTGEDDGPSTSASQPTTSTPQKPRKAP PGAARKRAFDDGDSSAGSSRTFDSTPPKTKKNENMHHP ADFPSCLSDYLSHAVYSNKTVSSFAVYSTLEKITLLYEKC DKFKVDFKSRHKYRSPDGSLAGILFLITLTKHRVSAVKN YCSMFCTISFLIVKGVNKSPEFYRALCNEPFQLLEENRN GVWSYEFDSKNEKAESVSWTAIAEFAETYELEDPLIIMA HYLDFANPFPCAKCHGKGLKAHKDHEKQHNNAKLFKN AKAQKNICQQAADVVLAKKRLKILESSREELLAEKFKK QLSKLKELSAIPLLEHMAGVAWYCCLFNDFEEKLVKVL QLLTENCPKHRNCLFIGPINSGKTSFAAAILDLIEGKSLNV NCPADKLNFELGCAIDHFAVVFEDVKGQTSLNKTLQPGQ GIHNLDNLREHLDGAVSVNLERKHVNKRTQIFPPCIVTA NEYVFPQTLLARFAYTLKFEPRKVLRKALEANKDLGKH RILQQGLTLLLALIWLCPSHKFHSSIREDISTWKSILNSEI GNEKFCIMIENIEKGLDPLKDFIEEEDANTDDSGRFTQSQ	dnaJ homolog subfamily C member 11 [Bos taurus]	N, 30-48, 30-48
Adenoviridae	>Bovine polyomavirus 2a_52577 MDSVLTRQEKRELCDLLEIP HHCYGNIPMMKAQYKKMC LVYHPDKGGDGS	>acc|GENBANK|AIT68757.1|NEIGHBOR|KM496324|large T antigen [Bovine polyomavirus 2a] MDSVLTRQEKRELCDLLEIPHHCYGNIPMMKAQYKKM CLVYHPDKGGDGSKMENEPIYGTPQFKAWWYQQHGSY FSESQSSSKPTRGGPSSRFSRAASPDSTTSTSTTGSGIFASE SSQRSDVGASGSQTSSAKTSSSSGGGRATGGGGTRVFGG TAYSREFRSESQDSDLYCDETLGSSPESAGEDDGPSSSTS QPSTSTPQKPRKMPPGAARKRAFDDGDSSAGSSRTFDST PPKTKKNESFHHPTDFPSCLLDYLSHAVYSNKTVSAFAV YSTLEKVTLLYEKCDKFKVDFKSRHKYKAPDGSLAGILF LITLTKHRVSAVKNYCSMFCTISFLIVKGVNKSPEFYRAL CNEPFALLEENRNGVWSYEFDSKNEKAESVSWTAIAEFA ETYELEDPLIIMAHYLDFANPFPCAKCHGKGLKAHKDH EKQHNNAKLFKNAKAQKNICQQAADVVLAKKRLKILE SSREELLAEKFKKQLSKLRELHTIPLLEHMAGVAWYCCL FSDFEEKLVRVLQLLTENCPKHRNSLFIGPINSGKTSFAA AILDLIEGKSLNVNCPADKLNFELGCAIDHFAVVFEDVK GQTSLNKTLQPGQGIHNLDNLREHLDGAVSVNLERKHV NKRTQIFPPCIVTANEYVFPQTLLARFAYTLKFEPRKVLK KALDANKDLSKHRILQQGLTLLLALIWLCPASKFHSSIRE EVASWKSILNSEIGNEKFCVMIENIEKGMDPLKDFIEEES TDDSGRFTQSQ	dnaJ homolog subfamily C member 11 [Bos taurus]	N, 30-48, 30-48
Adenoviridae	>Bovine retrovirus CH15_52705 MGQCASTAPYLQMLRHFLK AYGICIDMVDINKCIQILKEY NPWFPEEGTM	>acc|REFSEQ|YP_009243639.1|REFSEQ|NC_029852|putative group-specific antigen [Bovine retrovirus CH15] MGQCASTAPYLQMLRHFLKAYGICIDMVDINKCIQILKE YNPWFPEEGTMDSELWLQAKDNVEKAHRQGEKIPIRFW SVWSVIFSLLRAVHGQLPLTTVQRLNGPVLDELRLDVDT	TPA: Gag-like [Bos taurus]	N, 254-533, 13-283

		KQAVHEDLKDNIYEPTPPAPVDGERKIMESSKGAVSEQP LAVCSFEEQIQKRFSDLSQQIEKLTLRLARTEANKTLQAK PKAALSLAAAALDEGDFGDYEFSDDEEIPSLAVFPVIRPP PVQGQAVPARYEGIDIDVVGKLKKAVTLYGPQSHYVKE MLTGVAKHFGNFAPHDWKTLARSLLKEPEYLQWNMWF SDLAAQQAAENAQSGNPNVRLITYPMLTGTGNFEDVNV QAQQTPIEIHDQLVDLAMEAWDRIRPTGEHYGSWTKVM QRNNEPYVEFIARLRVVLERTVVGEKARDQLLKMLAFE NANEDCRRALLPVKETGDVNAYLKACKDIGSETRKMQ MFAETMVSTWKALNEKSAMKCYGCGQEGHLKRNCQK VNKEKRIGRKDINKQAPGICPKCKRGRHWARECRTKVS FIGLDQEEQAGNEQGGYLRGPKTKTEGIYFPYPMPPNTV PR		
Adenoviridae	>Bovine retrovirus CH15_52706 MGRSSLTMKGVTVHPGVID ADYQGEIQIMMSAVANISFK KGDRIAQLLIL	>acc|REFSEQ|YP_009243640.1|REFSEQ|NC_029852|putative protease [Bovine retrovirus CH15] MGRSSLTMKGVTVHPGVIDADYQGEIQIMMSAVANISFK KGDRIAQLLILPYFSLQTSNNKRIGGFGSTNPINVFWAQK ITQQQPLVTCYIDDDKKGFSALIDTGSDITIVAQQHWPKS KSYSIIPCQVKGISQQPVSSIGRSDTFVTLKGEEGQFAVLK PYVLNNSFSLIGRDTLEQWGAHLEF	TPA: protease-like [Bos taurus]	Y, 1-185, 276-461
Baculoviridae	>Antheraea pernyi nucleopolyhedrovirus_19811 GFHVHEFGDTSNGCTSAGE HFNPTHQQHGAPDAAERHV GDLGNVRSAGCT	>acc|GENBANK|BAX08892.1|NEIGHBOR|LC194889|superoxide dismutase [Antheraea pernyi nucleopolyhedrovirus] MKAICIISGDVYGEVHFEQSAPGETVYITGHLLNLPRGL HGFHVHEFGDTSNGCTSAGEHFNPTHQQHGAPDAAER HVGDLGNVRSAGCTALTPVDMSDNVISLYGPLSILGRSL VVHVDRDDLGLTDHPLSKITGNSGGRLGCGIIGVK	Cu/Zn superoxide dismutase, partial [Antheraea pernyi]	Y, 35-149, 1-115
Baculoviridae	>Bombyx mandarina nucleopolyhedrovirus S2_48170 MKAICIISGDVHGKIYFRQES ANRPLKISGYLLNLPRGLHG FHVHEYGDT	>acc|GENBANK|BBN66070.1|NEIGHBOR|LC500465|superoxide dismutase [Bombyx mori nucleopolyhedrovirus] MKAICIISGDVHGKIYFRQESANRPLKISGYLLNLPRGLH GFHVHEYGDTSNGCTSAGEHFNPTDEDHGAPDAEIRHV GDLGNIKSVGYNSLTEINMMDSVMSLYGPHNIIGRSLVV HTDKDDLGLTDHPLSKTTGNSGGRLGCGIIAICK	superoxide dismutase [Cn-Zn] [Bombyx mandarina]	Y, 11-147, 103-238
Baculoviridae	>Maruca vitrata nucleopolyhedrovirus_801132 MKAICIISGDVHGQIHFQQES PSHSLKISGHLLNLPRGLHGF HVHEYGDM	>acc|REFSEQ|YP_950752.1|REFSEQ|NC_008725|superoxide dismutase [Maruca vitrata nucleopolyhedrovirus] MKAICIISGDVHGQIHFQQESPSHSLKISGHLLNLPRGLH GFHVHEYGDMSNGCTSAGEHFNPTDEDHGAPDAEIRHV GDLGNIKSAGYNALTDVNIIDNVMSLYGPHNVIGRSLVV HTDRDDLGLTDHPLSKTTGNSGGRLGCGIIAIAKS	Cn-Zn superoxide dismutase [Lymantria dispar]	Y, 1-151, 2-152
Baculoviridae	>Operophtera brumata nucleopolyhedrovirus_841557 MKALCVLTGDIDGKVYFTQ ETPTHLVKITGYVMNLPKGL HGIHIHEFGDL	>acc|REFSEQ|YP_009552660.1|REFSEQ|NC_040621|SOD [Operophtera brumata nucleopolyhedrovirus] MKALCVLTGDIDGKVYFTQETPTHLVKITGYVMNLPKG LHGIHIHEFGDLSNGCTSAGEHFNPTHLDHGAPNALVRH VGDLGNIEARVSNALTEVNKIDNVMSLFGPHSILGRSLVI HTQRDDLGLTDHPLSKTTGNSGGRLGCGIIGVCTN	Superoxide dismutase [Cn-Zn] [Operophtera brumata]	Y, 1-129, 3-150
Baculoviridae	>Plodia interpunctella granulovirus_883117 MKGLCILTGDVEGEVEFWQ DKPSMPMRILGFVRNLPRGN HGIHIHEFGDT	>acc|REFSEQ|YP_009330183.1|REFSEQ|NC_032255|SOD [Plodia interpunctella granulovirus] MKGLCILTGDVEGEVEFWQDKPSMPMRILGFVRNLPRG NHGIHIHEFGDTTNGCTSAGEHFNPHHRDHGGPNSAER HLGDLGNIYSRGEHVATPIDRIDPMISLYGPYNILGRSLVI HAMEDDYGRGSNELSKITGNAGSRLGCGVIGVKREKTI GIGDDKVVSSSTNHKC	superoxide dismutase, partial [Plodia interpunctella]	N, 1-118, 3-122
Baculoviridae	>Artogeia rapae granulovirus_23516 MKGVCVMVGDVTGTIVFHQ EKPTCPVHITGYLYNLPYGY HGFHIHEFGDT	>acc|REFSEQ|YP_003429374.1|REFSEQ|NC_013797|sod [Artogeia rapae granulovirus] MKGVCVMVGDVTGTIVFHQEKPTCPVHITGYLYNLPYG YHGFHIHEFGDTTNGCTSAGEHFNPFHKEHGGPHDEER HLGDLGNIYSVGGNATRIDIVDNMISLYGAYNILGRSLVV HTMKDDYGKGDNKSSKITGNAGGRLGCGVIGVKHESQ VAI	superoxide dismutase [Cu-Zn] [Pieris rapae]	Y, 1-148, 3-150
Baculoviridae	>Plutella xylostella granulovirus_883977 MKGVCVMSGGKGVVTFTQ DLPHHVTRMSGHLIDLPKGE HGIHIHEFGDVT	>acc|GENBANK|AMQ35658.1|NEIGHBOR|KU529791|PxGV-Corf46 protein [Plutella xylostella granulovirus] MKGVCVMSGGKGVVTFTQDLPHHVTRMSGHLIDLPKG EHGIHIHEFGDVTNGCTSAGEHLNPFDEEHGDAEHGHL GDLGNVRSYGYKYTHFSMISPRISLYGEHSVLGRSLVVH EQRDDLGRGDDVTSKTTGNSGGRLCCGVIGVKREELSS F	superoxide dismutase, partial [Plutella xylostella]	Y, 2-147, 3-150
Baculoviridae	>Spodoptera litura nucleopolyhedrovirus II_963094 MKAICLLSGDVHGRVTFEQ QSPEHLLYIKGYVVNLPKGF HGFHVHEFGDT	>acc|REFSEQ|YP_002332749.1|REFSEQ|NC_011616|SOD [Spodoptera litura nucleopolyhedrovirus II] MKAICLLSGDVHGRVTFEQQSPEHLLYIKGYVVNLPKGF HGFHVHEFGDTSNGCTSAGEHFNPLGRNHGAPNAVDRH VGDLGNIEAKKSNSLTEIDKIDNVMSLFGEYSVIGRSLVI HSDRDDLGLGNHPLSKTTGNSGGRIACGIIGYAI	superoxide distmutase [Cu-Zn] [Spodoptera litura]	Y, 2-150, 4-151

Baculoviridae	>Trichoplusia ni single nucleopolyhedrovirus_989471 MKAVCMISGDVTGQVIFLQE SPEHLLKITGFILNLPQGLHG FHVHEFGDT	>acc|REFSEQ|YP_002332749.1|REFSEQ|NC_011616|SOD [Spodoptera litura nucleopolyhedrovirus II] MKAICLLSGDVHGRVTFEQQSPEHLLYIKGYVVNLPKGF HGFHVHEFGDTSNGCTSAGEHFNPLGRNHGAPNAVDRH VGDLGNIEAKKSNSLTEIDKIDNVMSLFGEYSVIGRSLVI HSDRDDLGLGNHPLSKTTGNSGGRIACGIIGYAI	superoxide dismutase [Cu-Zn] [Trichoplusia ni]	Y, 2-150, 4-152
Baculoviridae	>Bombyx mori nucleopolyhedrovirus_48308 MAQVKIGEFKFGEDMFTLR YVLEQGNQQVKFVAKDIAN KLNFKNTKKAIR	>acc|GENBANK|AKN59256.1|GENBANK|KR139833|Bro-d [Bombyx mori nucleopolyhedrovirus] MAQVKIGEFKFGEDMFTLRYVLEQGNQQVKFVAKDIAN KLNFKNTKKAIRDHVDDKYKTTYERGARLVPPAPNSVA KQGDPLYLQPHTVLITKSGVIQLIMKSKLPYAIELQEWLL EEVIPQVLCTGKYDPAIKQQEEKNKQLVTKLIATFTEHTN ALQAVVAQKTEELFKKQEFIERIIAIKDKQIEAKDLQVTR VMTDLNRMYTGFQETMQRKDEMMHKKDELLQVKDTQ VSNLIAKMIDLSDRAVQYPADKRKHPVLCVTRDGTTFTA ITGQKTYVESQKHKRNIDAANIVVENIRPNPTVDWNNAT DRLQSKRSKRSISFDSLEEAQQFENRIKYLLKNANNIN	BRO-d [Bombyx mori]	Y, 1-348, 1-348
Baculoviridae	>Agrotis segetum nucleopolyhedrovirus A_14222 MPQVKIGVFKFGEDKFKLRY VVGNDKDVLFVAKDIASVL KYEKPANAVAK	>acc|REFSEQ|YP_529720.1|REFSEQ|NC_007921|BRO-A [Agrotis segetum nucleopolyhedrovirus A] MPQVKIGVFKFGEDKFKLRYVVGNDKDVLFVAKDIASV LKYEKPANAVAKHVDKKYKCYFLEKGPRIEDPSFGDNG SVGVEVSIIKKDLIKKGHPLFLYDQTILITKSGVIQLIMKS KLPYAVELQEWLLEEVIPQVLCTGKYQPAVANNSECLSK SNEMILKMSQELILAKQNSDAMIQEMIVARRDAETARRD MVVLSTRIADIAQDVITKPSNPQLLHTLAVCEIGNNEFAF LRPQKRSLQRSLNNLRRNGQADLVYANDYVPNSMNVL NKVKEHVPKDKFKAKNNKITLLKEYDKQKLIEIINKSLT ARQLSLVQQKLSE	BRO-d [Lepidoptera]	N, 1-164, 1-148
Baculoviridae	>Hyphantria cunea nucleopolyhedrovirus_706503 MCQVKIGEFKFGEDTFALRY VLERDQQVKFVAKDVAASL KYQDTKHAVKS	>acc|REFSEQ|YP_473245.1|REFSEQ|NC_007767|BRO-c [Hyphantria cunea nucleopolyhedrovirus] MCQVKIGEFKFGEDTFALRYVLERDQQVKFVAKDVAAS LKYQDTKHAVKSHVDDKYKCTFERGCINISKENSVKQG DPLYLSPQTILIDKIGVIQLFMRSKLHNAAELQNWFYERV LPQCTARQSALSLLQDAQATVKFNSAPVEGHFYAATTLL YAERNLFKIGQTTNLTRRLAALNCGRADDDQMRYVLQT EPTVHHTLLEKLMKQELRPYRNSGEVYCTDFEHIKRALE ACLRRCSQN	BRO-d [Bombyx mori]	N, 1-123, 1-128
Baculoviridae	>Agrotis segetum nucleopolyhedrovirus A_14226 MQIFVKTLTGKTVTVDIEST DTVEQLKQKITDKEGIPPDQ QRLIYAGKQL	>acc|REFSEQ|YP_529805.1|REFSEQ|NC_007921|ubiquitin [Agrotis segetum nucleopolyhedrovirus A] MQIFVKTLTGKTVTVDIESTDTVEQLKQKITDKEGIPPDQ QRLIYAGKQLEDSRTMSDYNIQKESTIHLVLRLRGGSW	ubiquitin, partial [Ostrinia furnacalis]	Y, 1-77, 1-77
Baculoviridae	>Spodoptera frugiperda granulovirus_962217 MQIFVKTLTGKTITVEVESSD TVATLKQKIMDKESVPSDQQ RLIFAGKQL	>acc|GENBANK|AXS01062.1|NEIGHBOR|MH170055|v-ubi [Spodoptera frugiperda granulovirus] MQIFVKTLTGKTITVEVESSDTVATLKQKIMDKESVPSD QQRLIFAGKQLDDERTLADYNIQKESTLHLVLRLRGGNK EMIV	ubiquitin-60S ribosomal protein L40 [Spodoptera frugiperda]	Y, 1-79, 1-79
Baculoviridae	>Spodoptera litura nucleopolyhedrovirus II_963125 MQIFVKTLTGKTVTVEVEST DTVEQLKQKITDKEGIPPDQ QRLIYAGKQL	>acc|REFSEQ|YP_002332826.1|REFSEQ|NC_011616|V-ubiquitin [Spodoptera litura nucleopolyhedrovirus II] MQIFVKTLTGKTVTVEVESTDTVEQLKQKITDKEGIPPD QQRLIYAGKQLDDSRTMGDYNIQKESTIHLVLRLRGGAV	ubiquitin-60S ribosomal protein L40 [Spodoptera litura]	Y, 1-78, 1-78
Baculoviridae	>Bombyx mori nucleopolyhedrovirus_48536 MQIFIKTLTGKTITAETEPAET VADLKQKIADKEGVPVDQQ RLIFAGKQL	>acc|REFSEQ|NP_047441.1|REFSEQ|NC_001962|Ubiquitin [Bombyx mori nucleopolyhedrovirus] MQIFIKTLTGKTITAETEPAETVADLKQKIADKEGVPVDQ QRLIFAGKQLEDSKTMADYNIQKESTLHMVLRLRGG	ubiquitin and ribosomal protein S27a [Bombyx mori]	Y, 1-76, 1-76
Baculoviridae	>Sucra jujuba nucleopolyhedrovirus_964896 MQIFVKTLTGKSITLEVEPSD SIEMVKQKIQDKEGIPPDQQ RLIYAGKQL	>acc|REFSEQ|YP_009186717.1|REFSEQ|NC_028636|ubiquitin [Sucra jujuba nucleopolyhedrovirus] MQIFVKTLTGKSITLEVEPSDSIEMVKQKIQDKEGIPPDQ QRLIYAGKQLEDSSTVSDYNIQKESTLHLVLRLRGGITNT FVICNALR	putative NEDD8, partial [Operophtera brumata]	Y, 7-76, 1-70
Baculoviridae	>Operophtera brumata nucleopolyhedrovirus_841586 MQIFVKTLTGKSITLDVEPSD TIESVKSKITDKEGVPPEQQR	>acc|REFSEQ|YP_009552587.1|REFSEQ|NC_040621|ubiquitin [Operophtera brumata nucleopolyhedrovirus] MQIFVKTLTGKSITLDVEPSDTIESVKSKITDKEGVPPEQ QRLIFSGKQLEDSMTISDYNIQKESTIHLVLRLRGG	putative NEDD8, partial [Operophtera brumata]	Y, 7-76, 1-70

	LIFSGKQL			
Baculoviridae	>Chrysodeixis chalcites nucleopolyhedrovirus_77600 QVDAHNVVPCWLTFIADRH DYDEFKSKIDAELENLLTPFS LVIQHPYKSV	>acc|GENBANK|ACD01433.1|GENBANK|EU401910|DNA photolyase I [Chrysodeixis chalcites nucleopolyhedrovirus] QVDAHNVVPCWLTFIADRHDYDEFKSKIDAELENLLTPF SLVIQHPYKSVVSIESSTNTSIDWSNLLSSRNVDHSVKRI KWTEAGYNAAILRLATFIQCYIYNYKNSIHNPVSSKQSD LSPFFHFGFISAQRVIYHLRFCVTKKSVLQKTVFIRKLKN NIEKFIENCLYRREFADNFCYFNLNYITFNAASPQIKRYIA KHLRYYTYSLNELEYSQTHDNIWNKAQEDLRENGKIYP FIRVYWAKKI	CPD photolyase, partial [Chrysodeixis chalcites]	Y, 8-247, 5-230
Baculoviridae	>Helicoverpa armigera nucleopolyhedrovirus_165412 MYKQIITILLLVLFLSVLDGA RILCVFPVPSYSHHAVFEAYT NALALRGH	>acc|GENBANK|AJP07822.1|NEIGHBOR|KJ701033|ecdysteroid UDP-glucosyltransferase [Helicoverpa armigera nucleopolyhedrovirus] MYKQIITILLLVLFLSVLDGARILCVFPVPSYSHHAVFEAY TNALALRGHTIVRITPFPTKRNDSSNVTDVDVSLSKDYF KSLVDRSRLFKKRGVISETSSVTARNYISLVHMLIDQFSM ESVRQLIESNNVFDLLVTEAFLDYPLVFSHLFGDVPVIQIS SGHALAENFETMGAVSRHPIYYPNLWRNKFQNLNVWEI ITEIYTELVLYLEFARLADEQTKMLRHQFGPNTPSVEELR QRVQLLFVNTHPLFDNNRPVPPSVQYLGSLHLDRNNDV DEQQTMDYNLMQFLNNSTNGVVYVSFGTSIRVSDMDD EFLFEFITAFKQLPYNILWKTDGMPMEHVLPKNVLTQTW LPQHHVLKHSNVVAFVTQGGMQSTDEAIDACVPLIGIPF MGDQAYNTNKYEELGIGRNLDPVTLTSHILVSAVLDVTV NNKSRYTSNIKALNRSTNYRTRKPMEKAIWYTEHVIDN GKNPILKTKAANVSYSKYYMSDIIVPVITFLVMTHLGQA IRRLVVI	UDP-glycosyltransferase UGT33B1 [Helicoverpa armigera]	Y, 6-498, 9-486
Baculoviridae	>Agrotis segetum nucleopolyhedrovirus A_14162 MESFEERLMSFESWPASNHV SPQKLAAAGFYYLNRGDEV RCAFCKVEIMR	>acc|REFSEQ|YP_529789.1|REFSEQ|NC_007921|IAP-3 [Agrotis segetum nucleopolyhedrovirus A] MESFEERLMSFESWPASNHVSPQKLAAAGFYYLNRGDE VRCAFCKVEIMRWRPGDDPLADHKRWAPHCKYVCQID GEEKLRVGEDECGSRSGDNAPKHPAFVSYDARIETYKN KWPRALTQTPHQLASAGFYYTGIGDAVLCFYNDCRLSE WNAGDDPWREHARWFAECPYVRKFKGLDYIQKIATEA CLIRGEESATSSTEPTRGLSVKDEKAAQLENEELVCKICF EGRRNVCFMPCGHVVACRECSLNVERCPLCRDKFTSIQR LFY	baculoviral IAP repeat-containing protein 3-like isoform X3 [Zerene cesonia]	Y, 5-270, 87-367
Baculoviridae	>Hyphantria cunea nucleopolyhedrovirus_706528 MEVTFATAGSLKNIVDALRG LLTYATFDCNADGLHLQSM DSEHVALVDLR	>acc|GENBANK|BAC02931.1|GENBANK|AB069856|proliferating cell nuclear antigen [Hyphantria cunea nucleopolyhedrovirus] MEVTFATAGSLKNIVDALRGLLTYATFDCNADGLHLQS MDSEHVALVDLRLKRAGFARYTCERKLSFSVPMRGLHK IVRAATTNKQLTMRASARDDQVHFAFKTAERTVTCALS QISLDVERLGVPDDDEYDCVLAVASDAWARVCSDLAQL DATVVELSSGAAGLCFAADAGDGVRANVLLRAAPRRPL TQAFACRYLNAFGQTAPLSKFVNVCMSANAPLRLRFCL ERLGKLDLYLAPQVSSGAKDGQ	proliferating cell nuclear antigen [Pieris rapae]	Y, 11-244, 12-249
Baculoviridae	>Spodoptera frugiperda granulovirus_962134 MESTFTSLRQANSGLKVDIR RTRLIHESTTTKPTTKGVVY WMSRDSRIED	>acc|REFSEQ|YP_009121836.1|REFSEQ|NC_026511|DNA [Spodoptera frugiperda granulovirus] MESTFTSLRQANSGLKVDIRRTRLIHESTTTKPTTKGVV YWMSRDSRIEDNWAFLYAQQMAIEHKVPLHVCFCLVSS FCNAGIRQFHFLLEGLKFVQQECRRLNVSFHVLDGSGD QVLLDWCNKHDVDAIVCDFNPTTEPLKWVDNVKKTIPS HMSLTQVDAHNVVPCWLQSSAYSAHTFRTAINRKLDEY LTQYPDAKLHPYGDATYPHIDWDALLQSRNMHADVQP VTWAEPGYHGAIKVFTKFAGERIKHYHELRNKPTENVC SNLSPWLHFGQISAQRVILHIKSLTNVGKQNVDTFIDECL IRRELADNFCYHNKEYDSIEASAPQWAIDTLEAHKFDRR AYVYSLNALETASTHDPLWNAAQTQLLREGKMHQYMR MYWCKKILEWSEDAQTALVHAIYLNDKYSLDGRDPNG YVGCMWSICGVHDRAWRELPIYGKVRYMNYNGCVRKF DVKTYINRYK	deoxyribodipyrimidine photo-lyase-like [Spodoptera frugiperda]	Y, 16-465, 104-555
Baculoviridae	>Helicoverpa zea single nucleopolyhedrovirus_165624 MKAICILSGDISGEICFSQESP LHLIKITGFILNLPRGLHGIH VHEFGDT	>acc|GENBANK|AKN50497.1|GENBANK|KM596835|Superoxide dismutase (SOD) [Helicoverpa zea single nucleopolyhedrovirus] MKAICILSGDISGEICFSQESPLHLIKITGFILNLPRGLHGI HVHEFGDTSNGCTSAGEHFNPTGQTHGAPNATVRHVGD LGNVESFGINSLTEVNIVDNVMSLFGPHSILGRSLVVHTD RDDLGLTDHPLSRITGNSGGRLGCGIIGVTNSYKEASVK	hypothetical protein B5X24_HaOG204842 [Helicoverpa armigera]	Y, 2-151, 4-152
Baculoviridae	>Helicoverpa armigera nucleopolyhedrovirus_165303 MQIFVKTLTGKTITVDVESS	>acc|GENBANK|QBM79000.1|NEIGHBOR|MK507817|ubiquitin [Helicoverpa armigera nucleopolyhedrovirus] MQIFVKTLTGKTITVDVESSDSVETVKEKIAAKEGVPVD QQRLIYAGKQLEDSMTMNDYSIQKEATLHLVLRLRGGQ	hypothetical protein B5X24_HaOG204842 [Helicoverpa armigera]	Y, 1-76, 1-76

	DSVETVKEKIAAKEGVPVD QQRLIYAGKQL	SIRTGF		
Baculoviridae	>Spodoptera frugiperda granulovirus_962157 MIAKAILSGDVRGEILFVQPT PASLMRITGVLHGLPFGCHG LHIHEFGDT	>acc|GENBANK|AXS01071.1|NEIGHBOR|MH170055|sod [Spodoptera frugiperda granulovirus] MIAKAILSGDVRGEILFVQPTPASLMRITGVLHGLPFGCH GLHIHEFGDTSNGCTSAGEHLNPHHQPHGGQFSAQRHL GDLGNVCASTFEKIDHMLSLYGKHSILGRSIVVHTMEDD LGLSGDETGNSGGRLACGIIGVVVE	superoxide dismutase [Cu-Zn]-like [Spodoptera frugiperda]	Y, 3-140, 5-151
Baculoviridae	>Neodiprion lecontei nucleopolyhedrovirus_829456 MFTVFASFFIITVAVVSSDESI ANVSPTGRIVGGSPTSIDEIP YQVSLQV	>acc|REFSEQ|YP_025203.1|REFSEQ|NC_005906|hypothetica l protein [Neodiprion lecontei nucleopolyhedrovirus] MFTVFASFFIITVAVVSSDESIANVSPTGRIVGGSPTSIDEI PYQVSLQVYSTHICGASIISDSWILTAAHCITYPVTLYRIR SGSTLSISGGVVTRVESAYVHHAYYTNNYGIPVNDIALL KLTNSLILGITSAAVPLYDKNEIIPDESTAXITGWGTLTEN GNTPVVLYSVNIPVIPTSTCAQIFRNWGGLPENQICAASP GGGKDACQGDSGGPMVVNDRLAGIVSWGNGCGRNGW PGVYTEVAAYREWITSLTGV	PREDICTED: trypsin delta/gamma-like [Neodiprion lecontei]	Y, 24-259, 24-250
Baculoviridae	>Neodiprion lecontei nucleopolyhedrovirus_829462 MLDHRRIGAAVISWSKTLSV PIDVYEEDDVLDEALHPYER FKTRNARSCY	>acc|REFSEQ|YP_025280.1|REFSEQ|NC_005906|hypothetica l protein [Neodiprion lecontei nucleopolyhedrovirus] MLDHRRIGAAVISWSKTLSVPIDVYEEDDVLDEALHPYE RFKTRNARSCYSYKVLTCLEFSRIF	PREDICTED: venom serine protease Bi-VSP-like [Neodiprion lecontei]	N, 1-46, 41-86
Coronaviridae	>Porcine epidemic diarrhea virus_887079 ASDTTINGFSSFCVATRQFTIT LFYNVTNSYGYVSKSQDSN CPFTLQSVN	>acc|GENBANK|QEQ12560.1|GENBANK|MK578558|spike protein [Porcine epidemic diarrhea virus] ASDTTINGFSSFCVATRQFTITLFYNVTNSYGYVSKSQDS NCPFTLQSVNDYLSFSKFCVSTSLLAGACTIDLFGYPEFG SGVKFTSLYFQFTKGELITGTPKPLQGVTDVSFMTLDVC TKYTIYGFKGEGIITLTNSSFLAGVYYTSDSGQLLAFKNV TSGAVYSVTPCSFSEQAAY	spike protein [Sus scrofa]	Y, 1-178, 531-708
Coronaviridae	>Swine enteric coronavirus_970158 MKFLIHCWLFLPLLSALSLT QDVTRCGVNTNFRRFFSKFN VQAPATVILG	>acc|REFSEQ|YP_009199242.1|REFSEQ|NC_028806|spike protein [Swine enteric coronavirus] MKFLIHCWLFLPLLSALSLTQDVTRCGVNTNFRRFFSKF NVQAPATVILGGYLPSMNSSSWYCGQGIETASGVHGIFL SYIDSGQGFEIGIAQEPFDPSGYQLYLYKDNGNTGATARL RICQFPHNKTLGPSVGDVTTGRNCLVNKAIPAHLQDNKN IVVGLTWDNDRVTVFADKIYHFYLKNEWSRVATRCYNK RSCAMQYVYTPIYYMLNVTSSGEDGIYYEPCISNCIGYA ANVFATESNGHIPDSFSFNNWFLLSNDSTLLHGKVVSNQ PLLVNCLWAIPKIYGLGQFFSFNQTLGGSCNGAAAERAP EALRFNINDTSVILADGSIVLHTALGTNFSFVCSNSSNPH TATFTLPLGATKVPYYCFLKVDTYNSTVYKFLAVLPPTV REIVITKYGDVYVNGFGYLHIGLLDAVTINFTGHGTDGD VSGFWTIASTNFVDALVEVQGAAIQRILYCDDPVSQLKC SQVSFDLDNGFYPISSRNLLSHEQPISFVTLPSFNDHSFVN ITVSAAFGGHSGANLVASDTTINGFSSFCVETRQFTITLFY NVTNSYGYVSKSQTSNCPFTLQSVNDYLSFNKFCISTSL LAGACTIDIFGYPDFGSGVKLTSLYFQFTKGELITGTPKPL VGVTDVSFMTLDVCTKYTIYGFKGEGIITLTNSSFLAGV YYTSDSGQLLAFKNVTSGAVYSVTPCSFSEQAAYVNDDI VGVISSLSNSTFNNTRELPGFFYHSNDGSNCTEPVLVYSN IGVCTSGSIGYVPYQHGQVKIAPTVTGNISIPTNFSMSIRT EYLQLYNTPVSVDCATYVCNGNSRCKQLLSQYTAACKT IESALQLSARLESVEVNSMLTISEEALQLADISSFNGDGY NFTSVLGVSVYDPTSGRVVQKRSLIEDLLFNKVVTNGLG TVDEDYKRCTNGRSVADLVCAQYYSGVMVLPGVVDAE KLHMYSASLIGGMALGGLTAAVALPFSYAVQARLNYLA LQTDVLQRNQQLLAESFNAAIGNITSAFDSVKEAISQTSQ GLNTVAHALNKVQEVVNSQGAALSQLTLQLQHNFQAIS SSIDDIYSRLDSLSADVQVDRLITGRLSALNSFVSQTLAK YTEVQASRKLAQQKVNECVKSQSQRYGFCGGDGEHIFS LVQAAPQGLLFLHTVLVPGDFVNVIAIAGLCVNDEIALTL REPGLVLFTHELQTYNATEYFVSSRRMFEPRKPTVSDFV QIESCVVTYVNLTSDQLPDVIPDYIDVNKTFDEIVASLPN RTGPSLPIDVFNATYLNLTVEIADLEQRSESLHNITEELRS LIYNINNTLVDLEWLNRVETYIKWPWWVWLIIFIVLIFVV SLLVLCCISTGCCGCCGCCGACFSGCCRGPRLQPYEAIE KVHVQ	spike protein [Sus scrofa]	Y, 1-1382-1-1386
Flavivirdae	>Kamiti River virus_776563 KKEKKPSLAGEAKGSRTIWY MWLGSRFLEFEALGFLNAD HWVSRENFPGG	>acc|GENBANK|ABW77748.1|GENBANK|EU074051|NS5 [Kamiti River virus] KKEKKPSLAGEAKGSRTIWYMWLGSRFLEFEALGFLNA DHWVSRENFPGGVGGLGVNYFGYYLKDIAGRGKYLFA D	AAEL017001-PA [Aedes aegypti]	Y, 1-76, 224-299

Flavivirdae	>Zika virus_1009620 KKQGEFGKAKGSRAIWYM WLGARFLEFEALGFLNEDH WMGRENSGGGVEG	>acc|GENBANK|ABW77724.1|GENBANK|EU074027|NS5 [Zika virus] KREKKQGEFGKAKGSRAIWYMWLGARFLEFEALGFLN EDHWMGRENSGGGVEGLGLQRLGYVLEEMSRAPGGK MYAD	AAEL017001-PA [Aedes aegypti]	Y, 1-75, 224-296
Flavivirdae	>Cell fusing agent virus_69862 ETLTSVVGFSMTDVSTYSQQ KVLREKVDTVIPPHPQHIRR VNRTITKHLI	>acc|GENBANK|ABA19276.1|GENBANK|AH015271|c [Cell fusing agent virus] ETLTSVVGFSMTDVSTYSQQKVLREKVDTVIPPHPQHIR RVNRTITKHLIRFFKNRNLRPRILSKEEFIANVRNDAAVG SWSRDVPWRDVQEAIQDQCFWDLIGKERALHLQGKCE MCIYNTMGKKEKKPSLAGEAKGSRTIWYMWLGSRFLEF EALGFLNADHWVSREHFPGGVGGVGVNYFGYYLKDIA SRGK	AAEL017001-PA [Aedes aegypti]	Y, 1-195, 100-294
Flavivirdae	>West Nile virus_1001794 AADGKTVMDVISREDQRGS GQVVTYALNTFTNLAVQLVR MMEGEGVIGPD	>acc|GENBANK|SNU18590.1|GENBANK|LT903701|polyprotein [West Nile virus] PAADGKTVMDVISREDQRGSGQVVTYALNTFTNLAVQL VRMMEGEGVIGPDDVEKLGKGKGPKVRTWLFENGEER LSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQ EWKPSTGWYDWQQVPFCSNHFTELIMKDGRTLVVPCR GQDELIGRARISPGAGWNVRDTACLAKSYAQMWLLLYF HRRDLRLMANAICSAVP	AAEL017001-PA [Aedes aegypti]	N, 7-53, 359-405
Flavivirdae	>Wesselsbron virus_1001753 KPFEFGKAKGSRAIWYMWL GARFLEFEALGFLNEDHWV SRENSGGGVEGT	>acc|GENBANK|AFK88565.1|GENBANK|JN226799|RNA-dependent RNA polymerase [Wesselsbron virus] KPFEFGKAKGSRAIWYMWLGARFLEFEALGFLNEDHW VSRENSGGGVEGTGLQYLGYILKELGGKT	AAEL017001-PA [Aedes aegypti]	Y, 1-65, 228-292
Flavivirdae	>Yellow fever virus_1007462 FYADDTAGWDTRITEADLD DEQEILNYMSPHHKKLAQA VMEMTYKNKVVK	>acc|GENBANK|BBA54494.1|GENBANK|LC227567|polyprotein [Yellow fever virus] FYADDTAGWDTRITEADLDDEQEILNYMSPHHKKLAQA VMEMTYKNKVVKVLRPAPGGKAYMDVISRRDQRGSGQ VVTYALNTIT	AAEL017001-PA [Aedes aegypti]	Y, 1-85, 296-384
Flavivirdae	>Aroa virus_23323 KLGEFGKAKGSRAIWYMWL GARFLEFEALGFLNEDHWLS RENSGAGVEGI	>acc|GENBANK|AAC58750.1|GENBANK|AF013362|NS5 protein [Aroa virus] KLGEFGKAKGSRAIWYMWLGARFLEFEALGFLNEDHW LSRENSGAGVEGIGLQRLGYVLRDMGRGGGKIYADDVA GWDTRITEKDLDNEMIVLEHMEPAHKKLAHAIFTLTYR HKVVRVMRPGPNGRTYMDVISREDQRGSGQVVTYALN TFTNAIVQLIRSAEAEGIITAFSIEEVSESVLDELLKWLED FVWERLKMMAISGDDCAVKACDERFATALNFLNAMSK VRKDIPEWKPSSGWSDWQQVPFCSHHFVEIRMKDGREL VVPCRHQDELVGRARVSPGATWTIRESACMAKAYAQM WMLMYFHRRDLRIMANAICSAVPVDWVPTGRTTWSIH GKGE	AAEL017001-PA [Aedes aegypti]	N, 6-171, 233-402
Flavivirdae	>Israel turkey meningoencephalomyelitis virus_774346 KMGEFGKAKGSRAIWYMW LGSRFLEFEALGFLNEDHW MGRENTLGGVEGM	>acc|REFSEQ|YP_009664837.1|REFSEQ|NC_043115|NS5 protein [Israel turkey meningoencephalomyelitis virus] KMGEFGKAKGSRAIWYMWLGSRFLEFEALGFLNEDHW MGRENTLGGVEGMGLQKLGYVLRDMAGKEGGLMYAD DTAGWDTRITKADLENEALILEKMDPEHRRLAESLIKFA YMNKVVKVMRPGREGVTVMDVISREDQRGSGQVVTY ALNTFTNLCVQLIRCMEGEGLLKPEEVEGLERGKHKKIQ DWLGKNGRERLAAMAVSGDDCVVKPMDDRFSTALHFL NSMSKIRKDIPEWKPSTGWRNWQDVPFCSHHFHELNM KDGRTIVVPCRHQDELIGRARLSPGSGWSLTETACLSKA YGQMWLLMYFHRRDLRLMANAICSSVPISWVPTGRTP WSIHGKGE	AAEL017001-PA [Aedes aegypti]	N, 6-175, 233-405
Flavivirdae	>Koutango virus_779946 KPGEFGKAKGSRAIWFMWL GARFLEFEALGFLNEDHWL GRKNSGGGVEGL	>acc|REFSEQ|YP_009664838.1|REFSEQ|NC_043116|NS5 protein [Koutango virus] KPGEFGKAKGSRAIWFMWLGARFLEFEALGFLNEDHW LGRKNSGGGVEGLGLQKLGYVLREVGGKPGGKIYADD TAGWDTRITKADLENEGKVLRFLEGEHRLLARAIIELTY RHKVVKVMRPAANGKTVMDVISREDQRGSGQVVTYAL NTFTNLAVQLVRMMEGEGVVGPEDVEKLGKGKEARVK SWLSENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLN AMSKVRKDIQEWKPSVGWHDWQYVPFCSNHFAELIMK DGRTLVVPCRGQDELIGRARISPGAGWNVKETACLAKS YAQMWLLLYFHRRDLRLMANAICSAVPMNWVPTGRTT WSIHAKGD	AAEL017001-PA [Aedes aegypti]	N, 1-175, 228-405
Flavivirdae	>Kedougou virus_777164 KREKKTSEFGKAKGSRAIW YMWLGARFLEFEALGFLNE DHWFSRENSHAG	>acc|GENBANK|ABW77692.1|GENBANK|EU073995|NS5 [Kedougou virus] KREKKTSEFGKAKGSRAIWYMWLGARFLEFEALGFLNE DHWFSRENSHAGVEGMGLQKLGYVLEEISKRHGGLMY AD	AAEL017001-PA [Aedes aegypti]	Y, 1-77, 224-299

Flavivirdae	>Saint Louis encephalitis virus_935290 IWYMWLGARFLEFEALGFL NEDHWMSRENSYGGVEGK GLQKLGYILQEIS	>acc|GENBANK|AAX33351.1|GENBANK|AY825077|NS5 protein [Saint Louis encephalitis virus] IWYMWLGARFLEFEALGFLNEDHWMSRENSYGGVEGK GLQKLGYILQEISQIPGGKMYADDTAGWDTRITKEDLKN EAKITKRMDERHRRLAEAIIDLTYRHKVVKVMRPGPDG KTYMDVISREDQRGSGQVVTYALNTFTNLAVQLIRCME AEGVVDEDDIMRVRLGRLAKAVEWLRKNGPERLSRMA VSGDDCVVKPLDDRFATALHFLNNMSKIRKDIQEWKPS TGWHNWQEVPFCSHHFNELMLKDGRTIVVPCRSQDELI GRARISPGAGWNVKETACLSKSYAQMWLLMYFHRRDL RMMANAICSAVPVNWVPTGRTTWSIHGKGEWMTTEDM L	AAEL017001-PA [Aedes aegypti]	N, 1-163, 241-406
Flavivirdae	>Quang Binh virus_912612 NMMGKKEKKPSSFGEARGS RIIWYMWLGSRFLEYEALGF LNEDHWVAREN	>acc|GENBANK|AXG21693.1|GENBANK|MG602499|NS5 protein [Quang Binh virus] NMMGKKEKKPSSFGEARGSRIIWYMWLGSRFLEYEAL GFLNEDHWVARENFPCGVGGVGVNYFGYYLKEIAQKG QWLIADDTAGWDT	AAEL017001-PA [Aedes aegypti]	Y, 1-87, 220-306
Flavivirdae	>West Nile virus_1001798 ADVRSYCYLASVSELSTRAA CPTMGEAHNEKRADPAFVC KQGVVDRGWGN	>acc|GENBANK|AAW80621.1|GENBANK|AY839589|polypr otein [West Nile virus] ADVRSYCYLASVSELSTRAACPTMGEAHNEKRADPAFV CKQGVVDRGWGNGCGLFGKGSIDTCAKFACTTKATGW IIQKENIKYEVAIFVHGPTTVESHGNYPTQVGATQAGRFS ITPSAPSYTMKLGDYGEVTIDCEPRSGIDTGAYYVMSVG AKSFLVHREWFMDLNLPWSSAGGTTWRNRETLMEFEE	E-protein, partial [Anas platyrhynchos]	Y, 1-191, 57-247
Flavivirdae	>Duck Tembusu virus_107332 ATELAVVRSYCYEPKVSDVT TESRCPTMGEAHNPKATYAE YICKKDFVDR	>acc|GENBANK|ALE71322.1|GENBANK|KR061334|envelop e protein [Duck Tembusu virus] ATELAVVRSYCYEPKVSDVTTESRCPTMGEAHNPKATY AEYICKKDFVDRGWGNGCGLFGKGSIQTCAKFDCSKKA EGRIVQKENVQFEVAVFIHGSTEASTYHNYSAQQSLKHA ARFVITPKSPVYTAEMED	E-protein, partial [Anas platyrhynchos]	Y, 1-333, 53, 185
Flavivirdae	>Tembusu virus_972432 FSCLGMQNRDFVEGVNGVE WIDVVLEGGSCVTITAKDKP TIDVKMMNMEA	>acc|GENBANK|AEL99350.1|GENBANK|HQ833330|E protein [Tembusu virus] FSCLGMQNRDFVEGVNGVEWIDVVLEGGSCVTITAKDK PTIDVKMMNMEATELAVVRSYCYEPKVSDVTTESRCPT MGEAHNPKATYAEYICKKDFVDRGWGNGCGLFGKGSI QTCAKFDCTKKAEGRIVQKENVQFEVAVFIHGSTEASTY HNYSAQQSLKHAARFVITPKSPVYTAEMEDYGTVTLEC EPRSGVDMGQFYVFTMNTKSWLVNRDWFHDLNLPWT GSSAGTWQNKESLIEFEEAHATKQSVVALASQEGALHA ALAGAIPVKYSGSKLEMTSGHLKCRVKMQGLKLKGMT YPMCSNTFSLVKNPTDTGHGTVVVELSYAGTDGPCRVPI SMSADLNDMTPVGRLITVNPYVSTSSTGAKIMVEVEPPF GDSFILVGSGKGQIRYQWHRSGSTIGKAFTSTLKGAQRM VALGDTAWDFGSVGGVLTSIGKGIHQVFGSAFKSLFGG MSWITQGMLGALLLWMGLNARDRSISMTFLAVGGILVF LAVNV	E-protein, partial [Anas platyrhynchos]	Y, 1-499, 4-502
Flavivirdae	>Ilheus virus_707597 MNCLGISNRDFVEGLSGGT WVDIVLEGGSCVTVMAKDK PTLDIKLIRMEA	>acc|GENBANK|ABQ88001.1|GENBANK|EF396941|polypro tein [Ilheus virus] MNCLGISNRDFVEGLSGGTWVDIVLEGGSCVTVMAKD KPTLDIKLIRMEAKDLATVRSYCYQATVTDSSTEARCPT MGEAHNSKSLDASYVCKSSYVDRGWGNGCGLFGKGSI QTCVKFSCPGKATGKSIQRENLNYDVAVYVHGPTSAAA HGNYTAQLTGKYAAKFSITPSAPTYTANLGEYGEATMEC EPRAALDIDNYYVMSLNNKHWLVNRDWFHDLDLPWT GPATESWKNRESLIEFEEPHATRQTVVALGNQEGALHTA LAGAIPVEVSSTTLTLNSGHLKCRLKLDKLKIKGTTYAM CKGTFAFAQTPVDTGHGTIVAELTYTGTDGPCKIPISMTA DLRDMTPIGRLVTVNPIIPSSANSQKILVELEPPFGSSFILV GQENNQIKYQWHKTGSTIGNALKTTWKGAQRFAVLGD TAWDFGSVGGIFNSIGKTIHGVFGTAFRSLFGGMSWVTQ ALMGALLLWLGISARERTVSLIMLSVGGILLFLAVNVHA DTGCAIDMARRELKCGSGIFIHNDVETWRNNYKYHPLT PRGFAKVIQMSKDKGVCGIRSVGRLEHEMWEAIAPELN AIFEDNGVDLSVVVKGQTG	E-protein, partial [Anas platyrhynchos]	Y, 1-508, 4-511
Flavivirdae	>Aedes flavivirus_10621 RTIILTPTRVVMAEVMEALG QAGITCDRNLMYCRRNLVT VACHATFTKFV	>acc|GENBANK|BAH83670.1|GENBANK|AB488411|polypro tein [Aedes flavivirus] RTIILTPTRVVMAEVMEALGQAGITCDRNLMYCRRNLVT VACHATFTKFVLSHGVKKIGVALIVMDECHFMDPMSIAA RGVMEHLHEKGTKLMYLSATPPGHSPDGGSNFPIHDQAI AFPSWMTPAWINSVRKSRNSKKAIMFVPSHTQANSLAAT IPGAVPLHRANFTANYARASSDETSLVVSTDISEMGAN	cell-fusing agents virus polyprotein-like protein [Aedes albopictus]	Y, 1-194, 961-1154

Flavivirdae	>Usutu virus_995882 YWGSVKEDRITYGGPWKFD RKWNGLDDVQLIVVAPGKA AINIQTKPGIFK	>acc|GENBANK|AFU97144.1|GENBANK|JX473241|non-structural protein 3 [Usutu virus] YWGSVKEDRITYGGPWKFDRKWNGLDDVQLIVVAPGK AAINIQTKPGIFKTPQGEIGAVSLDYPEGTSGSPILDKNGD IVGLYGNGVILGNG	cell-fusing agents virus polyprotein-like protein [Aedes albopictus]	Y, 3-86, 822-903
Flavivirdae	>Bovine viral diarrhea virus 1_53583 ADFEERHYKKIFIREGNFEGP FRQEAMNTMMCSRCQGKH RRFEMDREPKS	>acc|GENBANK|CAA91137.1|GENBANK|Z54332|viral nonstructural protein p125 [Bovine viral diarrhea virus 1] ADFEERHYKKIFIREGNFEGPFRQEAMNTMMCSRCQGK HRRFEMDREPKSARYCAECDRLHPAEEGDFWAESSMLG LKITYFALMDGKVYDITEWAGCQRVGISPDTHRVPSHISF GSRMPGTSGRQEYNGFVQYTARGKLFLRNLPILATKVK MLMVGNLGEEIGDLEHLGWILRGPAVCKKITEHEKCHV NILDKLTAFFGIMPRGTTPRAPVRFPTALLKIRRGLETGW AYTHQ	dnaJ homolog subfamily C member 14 [Bos taurus]	N, 24-127, 526-629
Orthomyxoviriae	>Influenza A virus_719512 MKTIIALSYVLCLVFAQKIPG NDNSTATLCLGHHAVPNGT MVKTITNDRI	>acc|GENBANK|QFQ22733.1|GENBANK|MN573911|hemagglutinin [Influenza A virus] MKTIIALSYVLCLVFAQKIPGNDNSTATLCLGHHAVPNGT MVKTITNDRIEVTNATELVQNSSIGEICDSPHQILDGENC TLIDALLGDPQCDGFQNKKWDLFVERSKAYSNCYPYDV PDYASLRSLVASSGTLEFNNESFNWTGVKQNGTSSACIR KSSSSFFSRLNWLTHLNYTYPALNVTMPNNEQFDKLYIW GVHHPGTDKDQIFLYAQSSGRITVSTKRSQQAVIPNIGSR PRIRDIPSRISIYWTIVKPGDILLINSTGNLIAPRGYFKIQS GKSSIMRSDAPIGKCKSECITPNGSIPNDKPFQNVNRITY GACPRYVKHSTLKLATGMRNVPEKQTRGIFGAIAGFIEN GWEGMVDGWYGFRHQNSEGRGQAADLKSTQAAIDQIN GKLNRLIGKTNEKFHQIEKEFSEVEGRIQDLEKYVEDTKI DLWSYNAELLVALENQHTIDLTDSEMNKLFEKTKKQLR ENAEDMGNGCFKIYHKCDNACIGSIRNGTYDHNVYRDE ALNNRFQIKGVELKSGYKDWILWISFAISCFLLCVALLGF IMWACQKGNIRCNICI	Cryo-EM structure of the human glucagon-like peptide-2 receptor-Gs protein complex [Homo sapiens]	N, 1-39, 1-43
Papilloavirida	>Human papillomavirus_684690 MEDGEMADIGYGNLNFKAL QENRPDVSLDIVNETCKYPD FLKMQNDVYGD	>acc|GENBANK|CAA61969.1|GENBANK|X89877|capsid protein L1 [Human papillomavirus] MEDGEMADIGYGNLNFKALQENRPDVSLDIVNETCKYP DFLKMQNDVYGDSCFFFARREQCYARHFFVRGGNVGD DIPGEQIDAGTYKNDFYIPGASGQTQNKIGNSMYFPTVS GSLVSSDAQLFNRPYWLQRAQGHNNGV	L1 protein [Homo sapiens]	Y, 1-141, 222-352
Papillomaviridae	>Human papillomavirus_685097 MHGPKATLQDIVPHLEPQNE IPVDLLCHEQLSDSEEANDEI DGVNHQHLP	>acc|GENBANK|AQZ41090.1|GENBANK|KY502120|transforming protein E7 [Human papillomavirus] MHGPKATLQDIVPHLEPQNEIPVDLLCHEQLSDSEEAND EIDGVNHQHLPARRAEPPRHTLPCMCCPCEARIELVVESS APDPRAFQQLFLNTLSFVCPWCASQQ	E7 protein [Human papillomavirus type 18 [Homo sapiens]	Y, 1-105, 1-105
Paramyxoviridae	>Nipah henipavirus_830383 MDKLELVNDGLNIIDFIQKN QKEIQKTYGRSSIQQPSIKDQ TKAWEDFLQ	>acc|REFSEQ|NP_112022.1|REFSEQ|NC_002728|Phosphoprotein [Nipah henipavirus] MDKLELVNDGLNIIDFIQKNQKEIQKTYGRSSIQQPSIKD QTKAWEDFLQCTSGESEQVEGGMSKDDGDVERRNLED LSSTSPTDGTIGKRVSNTRDWAEGSDDIQLDPVVTDVVY HDHGGECTGYGFTSSPERGWSDYTSGANNGNVCLVSD AKMLSYAPEIAVSKEDRETDLVHLENKLSTTGLNPTAVPF TLRNLSDPAKDSPVIAEHYYGLGVKEQNVGPQTSRNVN LDSIKLYTSDDEEADQLEFEDEFAGSSSEVIVGISPEDEEP SSVGGKPNESIGRTIEGQSIRDNLQAKDNKSTDVPGAGP KDSAVKEEPPQKRLPMLAEEFECSGSEDPIIRELLKENSLI NCQQGKDAQPPYHWSIERSISPDKTEIVNGAVQTADRQR PGTPMPKSRGIPIKKGTDAKYPSAGTENVPGSKSGATRH VRGSPPYQEGKSVNAENVQLNASTAVKETDKSEVNPVD DNDSLDDKYIMPSDDFSNTFFPHDTDRLNYHADHLGDY DLETLCEESVLMGVINSIKLINLDMRLNHIEEQVKEIPKII NKLESIDRVLAKTNTALSTIEGHLVSMMIMIPGKGKGER KGKNNPELKPVIGRDILEQQSLFSFDNVKNFRDGSLTNE PYGAAVQLREDLILPELNFEETNASQFVPMADDSSRDVI KTLIRTHIKDRELRSELIGYLNKAENDEEIQEIANTVNDII DGNI	AngRem104 [Homo sapiens]	N, 5-46, 11-52
Paramyxoviridae	>Measles morbillivirus_801652 KDECFMYMFLLGVVEDSDP LGPPIGRAFGSLPLGVGRSTA KPEELLKEAT	>acc|GENBANK|BAA20206.1|GENBANK|D87566|matrix protein [Measles morbillivirus] KDECFMYMFLLGVVEDSDPLGPPIGRAFGSLPLGVGRST AKPEELLKEATELDIVVRRTAGLNEKLVFYNNTPLTLLTP WRKVLTTGSVFNANQVCNAVNLIPLDTPQRFRVVYMSIT RLSDNGYYTVPRRMLEFRSVNAVAFNLLVTLRIDKAIGP GKII	AngRem52 [Homo sapiens]	Y, 1-151, 49-201

Paramyxoviridae	>Beilong virus_43013 MAGNAGTADFLNSSWEEGG TLTAIDPEADERGRLIPKYRV INPGTNSRKS	>acc|GENBANK|QLI42698.1|GENBANK|MN598981|matrix protein [Beilong virus] MAGNAGTADFLNSSWEEGGTLTAIDPEADERGRLIPKYR VINPGTNSRKSAGYMYLLVHGIIEEKDVAGASAKNNKTF AAFPLGVGRSNALPQDLLDSVVGLDVTVRRTAGSSEML VFGTSNISPVLTPWKDILTTGAIFPAIKVCNNVDMVAVDR QQKFRCIFLTITMLTDAGVYKVPKTILDFRMANAVSFNL LVELLIGADMTASGIKGIINDYGERVTTFMIHVGNFLRRR NKEYSVDYCRQKIDKMDLRFSLGAVGGLSLHVCIKGKM SNALKAQLGYKTTICYSLMDTNPYLNKLMWKAECKIN KVTAVLQPSVPKEFKVYEDVLIDHTGKIMK	AngRem52 [Homo sapiens]	Y, 1-340, 1-340
Herpesviridae	>Human gammaherpesvirus 4_277161 MERRFMVTLQCLVLLYLAPE CGGTDQCDNFPQMLRDLRD AFSRVKTFFQT	>acc|GENBANK|QCF51234.1|NEIGHBOR|MK540313|interleukin-10 [Human gammaherpesvirus 4] MERRFMVTLQCLVLLYLAPECGGTDQCDNFPQMLRDLR DAFSRVKTFFQTKDEVDNLLLKESLLEDFKGYLGCQALS EMIQFYLEEVMPQAENQDPEAKDHVNSLGENLKTLRLR LRRCHRFLPCENKSKAVEQIKNAFNKLQEKGIYKAMSEF DIFINYIEAYMTIKAR	Cytokine Synthesis [Homo sapiens]	Y, 29-170, 18-159
Herpesviridae	>Human gammaherpesvirus 4_277162 MERRFMVTLQCLVLLYLAPE CGSTDQCDNFPQMLRDLRD AFSRVKTFFQT	>acc|GENBANK|QCF50673.1|NEIGHBOR|MK540302|interleukin-10 [Human gammaherpesvirus 4] MERRFMVTLQCLVLLYLAPECGSTDQCDNFPQMLRDLR DAFSRVKTFFQTKDEVDNLLLKESLLEDFKGYLGCQALS EMIQFYLEEVMPQAENQDPEAKDHVNSLGENLKTLRLR LRRCHRFLPCENKSKAVEQIKNAFNKLQEKGIYKAMSEF DIFINYIEAYMTIKAR	IL10, partial [Homo sapiens]	Y, 9-170, 7-177
Herpesviridae	>Human gammaherpesvirus 4_277552 MSGGLFYNPFLRPNKGLLK KSDKEYLRLIPKCFQTPGAA GVVDVRGPQPP	>acc|GENBANK|QCF52148.1|NEIGHBOR|MK540331|DNA polymerase catalytic subunit [Human gammaherpesvirus 4] MSGGLFYNPFLRPNKGLLKKSDKEYLRLIPKCFQTPGAA GVVDVRGPQPPLCFYQDSLTVVGGDEDGKGMWWRQR AQEGTARPEADTHGSPLDFHVYDILETVYTHEKCAVIPS DKQGYVVPCGIVIKLLGRRKADGASVCVNVFGQQAYFY ASAPQGLDVEFAVLSALKASTFDRRTPCRVSVEKVTRRSI MGYGNHAGDYHKITLSHPNSVCHVATWLQDKHGCQIFE ANVDATRRFVLDNDFVTFGWYSCRRAIPRLQHRDSYAE LEYDCEVGDLSVRREDSSWPSYQALAFDIECLGEEGFPT ATNEADLILQISCVLWSTGEEAGRYRRILLTLGTCEDIEG VEVYEFPSELDMLYAFFQLIRDLSVEIVTGYNVANFDWP YILDRARHIYSINPASLGKIRAGGVCEVRRPHDAGKGFL RANTKVRITGLIPIDMYAVCRDKLSLSDYKLDTVARHLL GAKKEDVHYKEIPRLFAAGPEGRRRLGMYCVQDSALV MDLLNHFVIHVEVAEIAKIAHIPCRRVLDDGQQIRVFSCL LAAAQKENFILPMPSASDRDGYQGATVIQPLSGFYNSPV LVVDFASLYPSIIQAHNLCYSTMITPGEEHRLAGLRPGED YESFRLTGGVYHFVKKHVHESFLASLLTSWLAKRKAIK KLLAACEDPRQRTILDKQQLAIKCTCNAVYGFTGVANGL FPCLSIAETVTLQGRTMLERAKAFVEALSPANLQALAPS PDAWAPLNPEGQLRVIYGDTDSLFIECRGFSESETLRFAE ALAAHTTRSLFVAPISLEAEKTFSCLMLITKKRYVGVLTD GKTLMKGVELVRKTACKFVQTRCRRVLDLVLADARVK EAASLLSHRPFQESFTQGLPVGFLPVIDILNQAYTDLREG RVPMGELCFSTELSRKLSAYKSTQMPHLAVYQKFVERN EELPQIHDRIQYVFVEPKGGVKGARKTEMAEDPAYAER HGVPVAVDHYFDKLLQGAANILQCLFDNNSGAALSVLQ NFTARPPF	DNA polymerase catalytic subunit [Homo sapiens]	Y, 1-995, 1-980
Herpesviridae	>Cercopithecine betaherpesvirus 5_70287 MFFNPYLAGARKPPACVAK RPAQKTFLELVPRGAMYDG QSGLIKHKTGR	>acc|GENBANK|AEW46232.1|GENBANK|JQ264771|DNA polymerase catalytic subunit [Cercopithecine betaherpesvirus 5] MFFNPYLAGARKPPACVAKRPAQKTFLELVPRGAMYDG QSGLIKHKTGRVPIMFYRDIKHVLDNDMAWPCPLPPSPP PIESIAGRVFGPLKFHTYDQVDGVLTYDTSEGLSPRYRHH ITPSGNVLRFFGATEQGHSICVNVFGQRSYFYCEYQDGE LLRDLLASASELVAEPRMSYALTIVQATKTSIYGYGTRPV PNLYQVSMSNWSMAKKIGEYLLEQGVPVYEIRVDPLTR LVIDKKMTTFGWCCVNRYEWRHKNKSSTCDFEIDCDVA DLMAIADDTSWPVYRCLSFDIECMSASGGFPVAEQVDDI IIQISCVCYETGGTGREVSTVFGTSGMHLFTIGRGGGVGT ADVYEFPSEYEMLLGFLIFFKRYAPCFVTGYNINAFDFK YILTRLETVYKMSVGPFSKLPSHGRFYTYTPPKRHYAMT TATKVFISGCVVIDMYPVCMAKTNSPNYKLNTMAELYL NQQKEDLSYKEIPVRFIAGSEGRADVGKYCMQDALLVR DLFNIINFHYEAGAIARLAKIPMRRVVFDGQQIRIYTSLL DECACRDFILPNHKGADASAENPEVSYQGATVFEPQIGY YSDPVVVFDFASLYPSIIMAHNLCYSTLIMPDGECPPDTQ VFTVELENGVTHRFAKNTARLSVLSELLTKWVSQRRAV RETMRGCQDPVKRMLLDKEQLALKVTCNAFYGFTGVV	DNA polymerase catalytic subunit [Homo sapiens]	Y, 2-1026, 6-1005

		NGMMPCLPIAASITRIGRDMLTRTSTFVEENFTEPCFLHN FFNREDYSGDPVEVKVIYGDTDSVFVCYRGITATALINR GPSLAAYVTQCLFVDPIKLEFEKVFVSLMMICKKRYIGK VSGESELSMKGVDLVRKTSCEFVKNVTRDIIQLLFDDPE VSRAAVQLSRLTFDELKAQGVPLGFWRVIQRLCKARDD LYTSRARVEELVLSSVLSKDVSLYKQANLPHIAVIKRLAA RSEELPTVGDRVFYVLTYPGDGRTTGVCNYEIAEDPAYV REHGVPIHAEKYFDQVIKAVTNVLSPVFPPETLKKDKFLI GVLPRRVYLESAFLPYSVKANEHC		
Herpesviridae	>Human gammaherpesvirus 8_278287 KVIYGDTDSLFICCMGFNMD SVSDFAEELASITTNTLFRSPI KLEAEKIF	>acc|GENBANK|CAA07599.1|GENBANK|AJ007663|DNA polymerase [Human gammaherpesvirus 8] KVIYGDTDSLFICCMGFNMDSVSDFAEELASITTNTLFRS PIKLEAEKIFKCLLLLTKKRYVGVL	DNA polymerase catalytic subunit [Homo sapiens]	Y, 1-63, 735-797
Herpesviridae	>Suid betaherpesvirus 2_966815 WLRKRKEVKARMKRCEDP MLVLILDKQQLALKVTCNA FYGFTGAVHGLLP	>acc|GENBANK|ADI48288.1|GENBANK|HM113361|DNA polymerase [Suid betaherpesvirus 2] WLRKRKEVKARMKRCEDPMLVLILDKQQLALKVTCNA FYGFTGAVHGLLPCLPLAASITSIGRDMLRQTSDFINNVL SS	DNA polymerase delta catalytic subunit isoform X3 [Sus scrofa]	Y, 2-75, 652-725
Herpesviridae	>Human gammaherpesvirus 8_278175 EDEDVPVCWICNEELGNERF RACGCTGELENVHRSCLST WLTISRNTACQ	>acc|GENBANK|ADQ57888.1|NEIGHBOR|HQ404500|E3 ubiquitin ligase [Human gammaherpesvirus 8] MEDEDVPVCWICNEELGNERFRACGCTGELENVHRSCL STWLTISRNTACQICGVVYNTRVVWRPLREMTLLPRLTY QEGLELIVFIFIMTLGAAGLAAATWVWLYIVGGHDPEID HVAAAAYYVFFVFYQLFVVFGLGAFFHMMRHVGRAYA AVNTRVEVFPYRPRPTSPECAVEEIELQEILPRGDNQDEE GPAGAAPGDQDGPADGAPVHRDSEESVDEAAGYKEAG EPTHNDGRDDNVEPTAVGCDCNNLGAERYRATYCGGY VGVQSGDGAYSVSCHNKAGPSSLVDILPQGLPGGGYGS MGVIRKRSAVSSALMFH	E3 ubiquitin-protein ligase MARCHF2 isoform 2 [Homo sapiens]	N, 5-71, 60-127
Herpesviridae	>Human gammaherpesvirus 8_278382 MASKDVEEGVEGPICWICRE EVGNEGIHPCACTGELDVVH PQCLSTWLTV	>acc|GENBANK|ADQ57893.1|NEIGHBOR|HQ404500|membrane protein [Human gammaherpesvirus 8] MASKDVEEGVEGPICWICREEVGNEGIHPCACTGELDV VHPQCLSTWLTVSRNTACQMCRVIYRTRTQWRSRLNLW PEMERQEIFELFLLMSVVVAGLVGVALCTWTLLVILTAPA GTFSPGAVLGFLCFFGFYQIFIVFAFGGICRVSGTVRALYA ANNTRVTVLPYRRPRRPTANEDNIELTVLVGPAGGTDEE PTDESSEGDVASGDKERDGSSGDEPDGGPNDRAGLRGT ARTDLCAPTKKPVRKNHPKNNG	E3 ubiquitin-protein ligase MARCHF2 isoform 1 [Homo sapiens]	N, 11-81, 69-139
Herpesviridae	>Human gammaherpesvirus 4_277326 MKSSKNDTFVYRTWFKTLV VYFVMFVMSAAVPITAMFP NLGYPCYFNALV	>acc|GENBANK|QCF51015.1|NEIGHBOR|MK540309|envelope glycoprotein M [Human gammaherpesvirus 4] MKSSKNDTFVYRTWFKTLVVYFVMFVMSAAVPITAMFP NLGYPCYFNALVDYGALNLTNYNLAHHLTPTLYLEPPE MFVYITLVFIADCVAFIYYACGEVALIKARKKVSGLTDLS AWVSAVGSPTVLFLAILKLWSIQVFIQVLSYKHVFLSAFV YFLHFLASVLHACACVTRFSPVWVVKAQDNSIPQDTFL WWVVFYLKPIVTNLYLGCLALETLVFSLSVFLALGNSFY FMVGDMVLGAVNLFLVLPIFWYILTEVWLASFLRHNFGF YCGMFIASIILILPLVRYEAVFVSAKLHTTVAINVAIIPILCS VAMLIRICRIFKSMRQGTDYVPVSETVELELESEPRPRPS RTPSPGRNRRRSSTSSSSSRSTRRQRPVSTQALISSVLPMT TDSEEEIFP	integral membrane protein [Homo sapiens]	N, 1-349, 1-344
Herpesviridae	>Human herpesvirus 4 type 2_279452 MKSSKNDTFVYRTWFKTLV VYFVMFVMSAVVPITAMFP NLGYPCYFNALV	>acc|REFSEQ|YP_001129479.1|REFSEQ|NC_009334|BBRF3 [Human herpesvirus 4 type 2] MKSSKNDTFVYRTWFKTLVVYFVMFVMSAVVPITAMFP NLGYPCYFNALVDYGALNLTNYNLAHHLTPTLYLEPPE MFVYITLVFIADCVAFIYYACGEVALIKARKKVSGLTDLS AWVSAVGSPTVLFLAILKLWSIQVFIQVLSYKHVFLSAFV YFLHFLASVLHACACVTRFSPVWVVKAQDNSIPQDTFL WWVVFYLKPIVTNLYLGCLALETLVFSLSVFLALGNSFY FMVGDMVLGAVNLFLVLPIFWYILTEVWLASFLRHNFGF YCGMFIASIILILPLVRYEAVFVSAKLHTTVAINVAIIPILCS VAMLIRICRIFKSMRQGTDYVPVSETVELELESEPRPRPS RTPSPGRNRRRSSTSSSSSRSTRRQRPVSTQALISSVLPMT TDSEEEIFP	integral membrane protein [Homo sapiens]	N, 1-349, 1-344
Herpesviridae	>Cercopithecine betaherpesvirus 5_70238 MAPSKVDMMNSRIWGVSVL LVFLTFFNICGHTTMMNVPG VGYPCSYFNVV	>acc|GENBANK|AEV80636.1|NEIGHBOR|FJ483969|envelope glycoprotein M [Cercopithecine betaherpesvirus 5] MAPSKVDMMNSRIWGVSVLLVFLTFFNICGHTTMMNVP GVGYPCSYFNVVDFMEMNMSQYNVMHLHTPMLFTDV VQTILYVIFSMLVFLCVVIYYVCCWLKITYRKEEGLNLN QRTRDIAFMGDSMSTFIFILVMDTFELFTLAMTFRLASVI AFMACIHFFCLTIYSVSLITNYQSFKNSMFQLARIHPKLR	integral membrane protein [Homo sapiens]	Y, 1-344, 1-342

		GTVQYRTMVVNLVQMLLGFNTAMLAMALCLGFGNNFF VQTGHMVMAVFFVFAFVSIIYFLLLEVIFFQYVKVQFGY HAGTFFGLCGLIYPIIKYDLAYTPDYTNNVAAAFAVCFII WAGFTACRCIRYFRHHHTVRYKSLATTEEIASLKEAATS DVEEIAT		
Herpesviridae	>Cercopithecine betaherpesvirus 5_70334 MKALYNPRFIGVALLLMSLI AYSECVHELHCECPNTRSGI YPGHIKTVLV	>acc|REFSEQ|YP_004936092.1|REFSEQ|NC_012783|chemok ine vCXCL7 [Cercopithecine betaherpesvirus 5] MKALYNPRFIGVALLLMSLIAYSECVHELHCECPNTRSGI YPGHIKTVLVKKPGVNCPVTEVIATLKNGQKVCLDPDAP MVKNKILTKVSI	C-X-C motif chemokine 1, partial [Macaca mulatta]	N, 28-89, 40-100
Herpesviridae	>Gallid alphaherpesvirus 2_146879 MQALLLVLVLFIVQIYLLPG NGISLESLAADKRCKCVKVT NRPTGLGPII	>acc|GENBANK|ABG22896.1|GENBANK|DQ534538|vIL8 [Gallid alphaherpesvirus 2] MQALLLVLVLFIVQIYLLPGNGISLESLAADKRCKCVKV TNRPTGLGPIIAVDVIPPGIHCRRTEIIFALKKNRKVCVDP EAPWVQQFIKKLERQHRTRKENLMVGEDGGKSTVGPV KNTIEPTPPTIGSHICL	C-X-C motif chemokine 13 precursor [Gallus gallus]	N, 19-116, 19-116
Herpesviridae	>Macacine betaherpesvirus 3_792203 MNRAIFNPRVLGVALLLMTL IAHHQTAAELRCQCLQVMK GIPPSNIQRIS	>acc|GENBANK|AEI59319.1|GENBANK|JN039372|UL146b [Macacine betaherpesvirus 3] MNRAIFNPRVLGVALLLMTLIAHHQTAAELRCQCLQVM KGIPPSNIQRISITRPNAGCERREIIATLKNGKQVCLDPEA PMMKKMCQKFPGGTYPSFWEHLMTLFRDWMLTPQA	C-X-C motif chemokine 1, partial [Macaca mulatta]	N, 27-89, 38-100
Herpesviridae	>Macacine betaherpesvirus 3_792370 MTLIAYHQTEAELRCQCLHV TRGIRPSNIKDITITKPNAGC DRKEIIATL	>acc|GENBANK|ABS84093.1|GENBANK|EU003822|RhUL1 46b [Macacine betaherpesvirus 3] MTLIAYHQTEAELRCQCLHVTRGIRPSNIKDITITKPNAG CDRKEIIATLKNGKQVCLDPEAPMMKKLLSKVPEGKYP SFWEQYKEHFLKMFTE	growth-regulated alpha protein precusor [Macaca mulatta]	N, 12-72, 40-100
Herpesviridae	>Cynomolgus cytomegalovirus_98670 MRKIFFFFSFISYCTFLNHAA NPCCSHPCQNRGICMTVGFH QYKCDCTHT	>acc|REFSEQ|YP_009337452.1|REFSEQ|NC_033176|Cy10 [Cynomolgus cytomegalovirus] MRKIFFFFSFISYCTFLNHAANPCCSHPCQNRGICMTVGF HQYKCDCTHTGFYDHDCTTPELLTRIKLLLKPDPDTVYY TLTHFSTIWWLINNIPFLRHAVMRYILTSRSDLIDSPPIYN ADYEYPSWEAFSNLSYYARILPPVPHDCPTPVGVKGKTK LPDAAHVVEKTLLRRTFTPDPQGTNLMFAFFGQHFTHQF FRTDLARGPAFTKGLGHGVDLNHIYGETPEKQTKLRLFK DGKMKYQMIDGEMYPPSVKDVQVHMIYPPEVPENMRL AVGQEAFGLVPGLMMYATIWLREHNRVCDVLKQEHPE WSDEQLFQTTRLIIIGETINIIVTQYVQQLSGYHFQLLFKP ELLFGTSFQYQNRIFSEFNMLYHWHSLMPDAFEIDKTKY DYVNFVYNNSILMTHGITQLVESFTKQIAGRISGGRNVPP ALRRVSRAAIEHGREMRFQSLNQYRKRFRLKPYESFEEL TGEKEIAAELRALYGDVEAVELYTGFLVEKPRPNAIFGES ILELGAPFSLKGLMANVICSPGYWKPSTFGGDVGFGIVK SATIQSLVCSNVKGCPLAAFRVPDEELLKARNGSSSSSSS SPPSPSSSSSSSSSPSVQQHVDPGVVIKREHAEL	PREDCITED: prostaglandin G/H synthase 2 [Macaca fascicularis]	Y, 19-623, 16-604
Herpesviridae	>Macacine betaherpesvirus 3_792286 MRRIFLLFSFIFYCTFLNHAA NPCCSHPCQNRGICTTVGFH HYKCDCTHT	>acc|GENBANK|APT39924.1|NEIGHBOR|KX689267|Rh10 [Macacine betaherpesvirus 3] MRRIFLLFSFIFYCTFLNHAANPCCSHPCQNRGICTTVGF HHYKCDCTHTGFYDHDCTTPELFTRIKLFLKPDPDTVYY TLTHFSTIWWLVNNIPFLRHAVMRYILTSRSDLIDSPPIYN ADYEYPSWEAFSNLSYYARILPPVPHDCPTPLGVKGKKN LPDAAHVVEKTLLRRTFTPDPQGTNLMFAFFGQHFTHQF FRTDLARGPAFTKGLGHGVDLNHIYGETPEKQSKLRLFR DGKMKYQIIDGEMYPPSVKDVQVHMIYPPEVPENMRLA VGQEAFGLVPGLMMYATIWLREHNRVCDVLKREHPEW SDEQLFQTTRLIIIGETINIIVTQYVQQLSGYHFQLLFKPEL LFGTSFQYQNRIFSEFNMLYHWHSLMPDAFEIDKTKYD YINFVYNNSILMTHGITQLVESFTKQIAGRISGGRNVPPA LRRVSRAAIEHGREMRFQSLNQYRKRFSLKPYESFEELT GEKEIAAELRALYGDIEAVELYTGFLVEKPRPNAIFGESIL ELGAPFSLKGLMANVICSPGYWKPSTFGGDVGFGIVKS ATIQSLVCSNVKGCPLAAFRVPNKELLKARNGSSSSSSS SPSPSSSSSSPSVQQHVDPGVVIKREHAEL	prostaglandin G/H synthase 2 [Macaca mulatta]	Y, 19-621, 16-604
Herpesviridae	>Cercopithecine betaherpesvirus 5_70323 MIAFACVSLLFLIPFGYTANP CCSHPCQNRGVCMTVGFDR YWCDCTRTGF	>acc|GENBANK|AEV80548.1|NEIGHBOR|FJ483969|prostagl andin G/H synthase 2 [Cercopithecine betaherpesvirus 5] MIAFACVSLLFLIPFGYTANPCCSHPCQNRGVCMTVGFD RYWCDCTRTGFYDRHCSTPEFLTRIKLLLKPTPDTVHYV LTHFPTLWWIVNNVFFLRNAVMRYILLTRSEMIDSPPLYN LDYGYKSWEAFSNLSYYTRTLPPVPRDCPTPLGVKGKA RLPDSTQIVDKVFIRRTFIPDPQGTNLMFAFFGQHFTHQF FRTDMVRGPAFTTGLGHGVDLSHIYGETLEKQHKLRLF	prostaglandin G/H synthase 2 precursor (Homo sapiens)	Y, 16-604, 15-604

		KDGKMKYQVIDGEMYPPTVKETQVHMLYAPEVPDDLR FAVGNEVFGLVPGLMMYATIWLREHNRVCDVLKHEHPE WGDEQLFQTTRLIITGETINIIITQYVQQLSGYHFQLIFDP ELLFGTAFQYQNRIAAEFNTLYRWHSLLPDDFRIGDAKY GYEHFVYNNSILIKHGIAQLVESFTRQIAGRISGGRNIPPA LRRVFREGIEHGRKMRYQSLNEYRKRFRLKPYESFEELT GEKEIAAGLEALYGDVEAVELYTGFIVEKPRPGAIFGESI MELGAPFSLKGLMANVICSPAYWKPSTFGGDVGLDIVK TATIQSLICANVKGCPLAAFRVSNAELLKAFNGSSPSSLY SEQKAGSRVLINSRTEL		
Herpesviridae	>Papio ursinus cytomegalovirus_860272 MTGEVCHVNDTMKAYGMT PDLTISLYSLGMILGIGGNML ILCVICLTRRP	>acc|REFSEQ|YP_009137431.1|REFSEQ|NC_027016|US27D [Papio ursinus cytomegalovirus] MTGEVCHVNDTMKAYGMTPDLTISLYSLGMILGIGGNM LILCVICLTRRPWFAHDILFMNMIFADLLMLLTIPAWVYY LLNYTQVSRLECVGLSFAFYVPLFLHSDMIVAIAVERYQ NLVRNTPVSCKTAVISCTLLWTTVTLVTSPYYMLRETQE PDSCILGNYTWHVSSPYRTVMDVAINLWTFVVPAGVILF LSIRIHSCSWGNRKLNTRTSRHLDAMAINMLFFNGLFNL VMFRDIVTDFSSGLSCHYLRQEHFFRMMGIAFVFARPIF NPILYVSISKKLLRSIFLLFMRVPYDTLDSEHAELMVQGK INNEQVPDPFPRECESML	prostaglandin E2 subtype EP4 receptor [Papio ursinus]	N, 23-49, 21-47
Herpesviridae	>Cercopithecine betaherpesvirus 5_70221 GNMLLLPVVWTVVVASGSA ASPCKPHEYAVGSECCPKCG NAGYRVKTNCS	>acc|GENBANK|ACQ55249.1|GENBANK|FJ883003|membrane glycoprotein UL144 [Cercopithecine betaherpesvirus 5] GNMLLLPVVWTVVVASGSAASPCKPHEYAVGSECCPKC GNAGYRVKTNCSEYSGTVCEPCPPGSYNNNTDRTNCTA CDTCNASSVAVNNCTATQNVRCRPINSTVTHLDARPGHH GGNHSWESDYDSTRDWWFILVFVVFLFSVILLILGCRLIL LHTTYGKMLTAYMYERIT	U35 [Homo sapiens]	N, 23-98, 142-219
Herpesviridae	>Cercopithecine betaherpesvirus 5_70261 MDFSIFCTFEQKLSLTDVGK LAKLTGAIIPVPSRHHLIKHH HLGLHQYID	>acc|GENBANK|AEV80622.1|NEIGHBOR|FJ483969|capsid triplex subunit 2 [Cercopithecine betaherpesvirus 5] MDFSIFCTFEQKLSLTDVGKLAKLTGAIIPVPSRHHLIKH HHLGLHQYIDSTRGYVRLRSLLRNMTPTILRRIEGNQLA MQVPTHGQLYTVLNTGPVLWEKGDGLCLLPPLFQGPLV REYPVSVGQWNLVLPWLLPSQLAVEINQRILIMALYSLD RSYEEVKAAVHQLQHVTFRDATFTIPDPVIENHLLLDMK TACLSMSMVANLASDLTMNYVRKLALEDSSMLLVKCQE LLMRLDREHADGPRGRPHHVSPDEEISRLSALFVMLRQ LDDLIHEQVMFTVCDVSPDNKSATCIFKG	U56 [Homo sapiens]	Y, 4-300, 3-296
Herpesviridae	>Human gammaherpesvirus 4_276644 LASYPVFQSIRGQTFHLLFV DEANFIKKEALPAILGFMLQ KDAKIIFISS	>acc|GENBANK|AWG87531.1|NEIGHBOR|MG298823|BDRF1 [Human gammaherpesvirus 4] LASYPVFQSIRGQTFHLLFVDEANFIKKEALPAILGFMLQ KDAKIIFISSVNSADQATSFLYKLKDAQERLLNVVSYVC QEHRQDFDMQDSMVSCPCFRLHIPSYITMDSNIRATTNL FLDGAFSTELMGDTSSLSQGSLSRTVRDDAINQLELCRV DTLNPRVAGRLASSLYVYVDPAYTNNTSASGTGIAAVTH DRADPNRVIVLGLEHFFLKDLTGDAALQIATCVVALVSSI VTLHPHLEEVKVAVEGNSSQDSAVAIASIIGESCPLPCAFV HTKDKTSSLQWPMYLLTNEKSKAFERLIYAVNTASLSAS QVTVSNTIQLSFDPVLYLISQIRAIKPIPLRDGTYTYTGKQ RNLSDDVLVALVMAHFLATTQKHTFKKVH	U60/66 [Homo sapiens]	Y, 8-376, 291-656
Herpesviridae	>Cercopithecine betaherpesvirus 5_70364 MLLLPVVWTVVVASGSAAS PCKPHEYAVGSECCPKCGNA GYRVKTNCSEY	>acc|GENBANK|AEV80668.1|NEIGHBOR|FJ483969|membrane glycoprotein UL144 [Cercopithecine betaherpesvirus 5] MLLLPVVWTVVVASGSAASPCKPHEYAVGSECCPKCGN AGYRVKTNCSEYSGTVCEPCPPGSYNNNTDRTNCTACD TCNASSVAVNNCTATQNVRCRPINSTVTHLDARPGHHGG NHSWESDYDSTRDWWFILVFVVFLFSVILLILGCRLILLH TTYGKMLTAYMYERIT	TNFRSF14 protein variant, partial [Homo sapiens]	N, 21-96, 142-219
Herpesviridae	>Cercopithecine betaherpesvirus 5_70221 GNMLLLPVVWTVVVASGSA ASPCKPHEYAVGSECCPKCG NAGYRVKTNCS	>acc|GENBANK|ACQ55249.1|GENBANK|FJ883003|membrane glycoprotein UL144 [Cercopithecine betaherpesvirus 5] GNMLLLPVVWTVVVASGSAASPCKPHEYAVGSECCPKC GNAGYRVKTNCSEYSGTVCEPCPPGSYNNNTDRTNCTA CDTCNASSVAVNNCTATQNVRCRPINSTVTHLDARPGHH GGNHSWESDYDSTRDWWFILVFVVFLFSVILLILGCRLIL LHTTYGKMLTAYMYERIT	TNFRSF14 protein variant, partial [Homo sapiens]	N, 23-82, 43-102
Herpesviridae	>Human gammaherpesvirus 4_277447 MQEQEQQMREQEQQVRKQ EGQVRXQEGQVREQEGQVR EQEGQVREQEGQV	>acc|GENBANK|QAD58893.1|GENBANK|MH590447|EBNA-1 [Human gammaherpesvirus 4] MQEQEQQMREQEQQVRKQEGQVRXQEGQVREQEGQV REQEGQVREQEGQVREQEGQXGEQEGQMGEQEGQMG EQEEQMGEQEEQMQEQEEQMGEQKEQMRKQEEQMGE QEEQVQKQEEQVQKQEEQMRKQEEQMXEQEEQMREQ EEQMLKQKEQTEQEEQTGEQEEQMREQEEQMREQEEQ MREQEEQMGKQEEQMWEQKEQMWEQKEQMWKQEEQ MGEQEEQMQKQEEQVRKQEEQVRKQEEQMRKQEEQM	golgin subfamily A member 6-like protein 6 isoform X1 [Homo sapiens]	N, 22-252, 381-640

<table>
<tr><td></td><td></td><td>RKQEEQMRKQEEQMRKQEEXMGEQEEXMGEQEEQMGEQKEQMRKQEEQMGXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXAGGAGGAGGAGGAGAGGAGAGGAGGAGAGGAGAGGAGGAGAGGAGAGGAGGAGAGGAGAGGAGGAGAGGAGGAGGGGAGGAGGGGRGRGGSGGRGRGGSGGRGRGGSGGRRGRGRERARGRSRERARGRGRGRGEKRPRSPSSQSSSSGSPPRRPPPGRRPFFHPVGDADYFEYLQEGGPDGEPDVPPGAIEQGPTDDPGEGPSTGPRGQGDGGRRKKGGWFGKHRGQGGSNPKFENIAEGLRVLLARSHVERTTEEGNWVAGVFVYGGSKTSLYNLRRGIALAVPQCRITPLSRLPFGMAPGPGPQPGPLRESIVCYFMVFLQTHIFAEVLKDAIKDLVMTKPAPTCNIKVTVCSFDDGVDLPPWFPPMVEGAAAEGDDGDDGDEGGDGDEGEEGQE</td><td></td><td></td></tr>
<tr><td>Herpesviridae</td><td>>Human gammaherpesvirus 8_278778
MPRYTESEWLTDFIIDALDSGRFWGVGWLDEQKRIFTVPGRNRRERMPEG</td><td>>acc|GENBANK|QAX88127.1|NEIGHBOR|MK143395|K11 [Human gammaherpesvirus 8]
MPRYTESEWLTDFIIDALDSGRFWGVGWLDEQKRIFTVPGRNRRERMPEGFDDFYEAFLEERRRHGLPEIPETETGLGCFGRLLRTANRARQERPFTIYKGKMKLNRWIMTPRPYKGCEGCLVYLTQEPAMKNMLKALFGIYPHDDKHREKALRRSLRKKAQREAARKQAAAVATPTTSSAAEVSSRSQSEDTESSDSENELWVGAQGFVGRDMHSLFFEEPEPSGFGSSGQSSSLLAPDSPRPSTSQVQGPLHVHTPTDLCLPTGGLPSPVIFPHETQGLLAPPAGQSQTPFSPEGPVPSHVSGLDDCLPMVDHIEGCLLDLLSDVGQELPDLGDLGELLCETASPQGPMQSEGGEEGSTESVSVLPATHPLESSAPGASVMGSGQELPDLGDLSELLCETASPQGPMQSEGGEEGSTESVSVLPATHPLESSAPGASVMGSSFQASDNVDDFIDCIPPLCRDDRDVEDQEKADQTFYWYGSDMRPKVLTATQSVAAYLSKKQAIYKVGDKLVPLVVEVYYFGEKVKTHFDLTGGIVICSQVPEASPEHICQTVPPYKCLLPRTAHCSVDANRTLEQTLDRFSMGVVAIGTNMGIFLKGLLEYPAYFVGNASRRRIGKCRPLSHRHEIQQAFDVERHNREPEGSRYASLFLGRRPSPEYDWDHYPVILHIYLAPFYHRD</td><td>mum1 [Homo sapiens]</td><td>N, 10-42, 24-56</td></tr>
<tr><td>Herpesviridae</td><td>>Human gammaherpesvirus 8_278796
MRRFRQLTTDCAPGKQNMVVMGRRTWLSIPAGCRPLAGRINVVLSRTLET</td><td>>acc|GENBANK|AYC63009.1|GENBANK|MH261317|dihydrofolate reductase [Human gammaherpesvirus 8]
MRRFRQLTTDCAPGKQNMVVMGRRTWLSIPAGCRPLAGRINVVLSRTLETPPPGAHFLASSLDAALGLARSPELAQQIDKVWVIGGGNLYREALTGPWPVRLFLTRVLHDFACDVFLSHDSLAAYARVNPKPGEQERVFQERGIFYMFETYIKVTQSSDTALPDLERPRPATPPFSETS</td><td>Crystal structure of a methotrexate-resistant mutant of human dihydrofolate reductase [Homo sapiens]</td><td>N, 5-153, 37-185</td></tr>
<tr><td>Herpesviridae</td><td>>Cynomolgus cytomegalovirus_98642
MNCAIFNPRVLGVALLLMTLIAHHQTAASELRCQCLQVTQGIHPKNIQSM</td><td>>acc|REFSEQ|YP_009337564.1|REFSEQ|NC_033176|Cy158.3 [Cynomolgus cytomegalovirus]
MNCAIFNPRVLGVALLLMTLIAHHQTAASELRCQCLQVTQGIHPKNIQSMTITKPNGGCDRREIIATLKNGQKVCLNPEAPMMKKVLSKFPGGTYSSFWQHFMTLFTD</td><td>Macrophage inflammatory protein n2-beta [Macaca fascicularis]</td><td>Y, 1-90, 6-100</td></tr>
<tr><td>Herpesviridae</td><td>>Murine roseolovirus_824832
MYYIVIFYLILSFVTGDSKISIIGNKEQKLHTSVTLKCSIQTENEPLIIT</td><td>>acc|REFSEQ|YP_009344844.1|REFSEQ|NC_033620|cell surface glycoprotein [Murine roseolovirus]
MYYIVIFYLILSFVTGDSKISIIGNKEQKLHTSVTLKCSIQTENEPLIITWQRKKGSESPENICTYSKTHGAVTQPKFLNKFTLTELNLYNTSINIKNITMDDACCYLCLYNIFGTEKKSAEICITVYIQPMGYLQNMTYKNYSIISCIVTSYPESSVQWILGNFSLKNVTNKIVNKNMTTTVNNSIIVYNENMSGCEVKCQIKFKNNITEYPYVFEHSLVNNVTSKYTSFKDSTYSFIAVFLFLMIVVAFLCIVIYWKRYSAILNNLINI</td><td>Cd200 protein [Mus musculus]</td><td>N, 18-136, 30-148</td></tr>
<tr><td>Herpesviridae</td><td>>Murid betaherpesvirus 8_823924
MNLMFILFYIIISINPTLSQVEVVTQDVNSSLRSPASLRCSLKTTQEPLI</td><td>>acc|REFSEQ|YP_007016518.1|REFSEQ|NC_019559|e127 [Murid betaherpesvirus 8]
MNLMFILFYIIISINPTLSQVEVVTQDVNSSLRSPASLRCSLKTTQEPLIVTWQKKKAVGPENMATYSKAHGVVLQPTYKDRINITELGLLNSSITFWNATLDDEGCYMCLFNMFGSGKVSGTSCLTLRYTPSVFISGNLSDNILNVTCSAVSRPNVSITWVPLGDGLKNFTTSQSQSDGTTKVTSVLYGNLTLHNKNVTCRIIYLNTVYNYSVVLDGFYPTPPTSPVSFTSDNSTRTTRTILLLIIIFMILTTVIALLYWRKQRYSRLMRFYRRSVVPLCPKL</td><td>OX-2 membrane glycoprotein isoform X1 [Rattus norvegicus]</td><td>N, 5-224, 48-270</td></tr>
<tr><td>Herpesviridae</td><td>>Murid betaherpesvirus 8_823942
MRLLFILAFCCVTAWIVEGVGIEVLQETFCVSLRTQRLPIQKIKTYTIKE</td><td>>acc|GENBANK|AKB93337.1|NEIGHBOR|KP202868|b156.5 [Murid betaherpesvirus 8]
MRLLFILAFCCVTAWIVEGVGIEVLQETFCVSLRTQRLPIQKIKTYTIKEGAMRAVIFVTKRGLRICADPNAGWTQAAITTLDKKNKRNKHKFNTTTAIPTQVPASANETSTLY</td><td>chemokine XCL1 [Rattus norvegicus]</td><td>Y, 1-108, 1-107</td></tr>
</table>

Herpesviridae	>Bovine gammaherpesvirus 4_50774 MKMAGWKKKLCPGHHLWALGCYMLLAVVSLRLSLRFKCDVDSLDLESRDF	>acc|GENBANK|AAO22158.1|GENBANK|AF465331|Bo17 protein [Bovine gammaherpesvirus 4] MKMAGWKKKLCPGHHLWALGCYMLLAVVSLRLSLRFKCDVDSLDLESRDFQSQHCRDMLYNSLKLPAKRSINCSGITRGDQEAVVQALLDNLEVKKKRPPLTDTYYLNITRDCERFKAQRKFIQFPLSKEELDFPIAYSMVVHEKIENFERLLRAVYAPQNIYCVHVDVKSPETFKEAVKAIISCFPNVFMASKLVPVVYASWSRVQADLNCMEDLLQSSVSWKYLLNTCGTDFPIKTNAEMVLALKMLKGKNSMESEVPSESKKNRWKYRYEVTDTLYPTSKMKDPPPDNLPMFTGNAYFVASRAFVQHVLDNPKSQILVEWVKDTYSPDEHLWATLQRAPWMPGSVPSHPKYHISDMTAIARLVKWQYHEGDVSMGAPYAPCSGIHRRAICIYGAGDLYWILQNHHLLANKFDPRVDDNVLQCLEEYLRHKAIYGTEL	core 2 beta-1, 6-N-acetylglucosaminyltransferase [Bos taurus]	Y, 1-440, 1-440
Herpesviridae	>Bovine gammaherpesvirus 4_50699 KKKLCRGHYLWALGCYMLLAVVSLRLSLRFKCDVDSLDLESRDFQSQHCR	>acc|GENBANK|AAN18276.1|GENBANK|AY143156|Bo17 [Bovine gammaherpesvirus 4] KKKLCRGHYLWALGCYMLLAVVSLRLSLRFKCDVDSLDLESRDFQSQHCRDILYNNLKLPAKRSINCSGITRGDQEAVAQALLDNLEVKKKRPPLTDTYYLNITRDCERFRAQRKFIQFPLSKEELDFPIAYSMVVHEKIENFERLLRAVYAPQNIYCVHVDVKSPETFKEAVKAIISCLPNVFMASKLVPVVYASWSRVQADLNCMEDLLQSSVPWKYLLNTCGTDFPIKTNAEMVLALKMLNGKNSMESEIPPESKKNRWKYSYEVTDTLYPTSKMKDPPPDNLPMFTGNAYFVASRAFVQHVLDNPKSQRLVEWVKDTYSPDEHLWATLQRAPWMPGSVPSHPKYHISDMTAIARLVKWQYHEGDVSMGAPYAPCSGIHRRAICIYGAGDLYWILQNHHLLANKFDPRVDDNVLQCLEEYLRHKA	mucus-type core 2 betat-1,6-N-acetylglucosaminyltransferase [Bos taurus]	Y, 1-428, 1-434
Herpesviridae	>Suid betaherpesvirus 2_966676 AVPGLSIAGLSLPGLSIPGLSLPGLSIPGLITSNVSAGTCRHVTANNVNS	>acc|GENBANK|QBA83945.1|GENBANK|MK377741|IE-A [Suid betaherpesvirus 2] AVPGLSIAGLSLPGLSIPGLSLPGLSIPGLITSNVSAGTCRHVTANNVNSRGGHLVTSTEADTTVNSEEPFSAVCANNTSGIVTSGKQSTSAETGTISIVTGDDIITVDESDCDNTTQPRHETPAPLRPSASGSGPCRSPSPRCPSPKPSTSRTNDMRQGDDPGPDDNIILVIKELQSMSDKCMKKDPVKKIGLKYGPRKNPAPQVQIPSSCDDDVPEPWPIESKDQT	LOW QUALITY PROTEIN: glutamine-rich protein 2 [Sus scrofa]	N,120-165, 561-609
Herpesviridae	>Common bottlenose dolphin gammaherpesvirus 1 strain Sarasota_83824 MLLCFLALVAFWPCLLPAFTVTVPKDLYVVEYGGNVTLECRFPVDKQLNL	>acc|REFSEQ|YP_009388506.1|REFSEQ|NC_035117|membrane protein De2 [Common bottlenose dolphin gammaherpesvirus 1 strain Sarasota] MLLCFLALVAFWPCLLPAFTVTVPKDLYVVEYGGNVTLECRFPVDKQLNLLALVVYWEMEDKKIIQFVNGEEDVNVQHDSYKNRVTLLKNQLPLGKAALQITDVKLQDAGIYCCLISYGGADYKRITLKVNASYRKINGSVAAVPGSSGHELTCQSEGYPEAEVIWTRGNASGKGGPALSGKTTVTSSKREEKLFNVTSVLRINTTANEIFYCIFRRLGHEENSTAELVIPGPDWAPGRWTYPRSRLGVPGAFLAVLVIVVILIFYVK	programmed cell death 1 ligand 1 [Tursiops truncatus]	Y, 12-268, 72-260
Herpesviridae	>Common bottlenose dolphin gammaherpesvirus 1 strain Sarasota_83839 MSAAVIHQLEEALDEDEKERLLFLCRDIAADSAPGTIKDLLGVLRERGHL	>acc|REFSEQ|YP_009388505.1|REFSEQ|NC_035117|apoptosis inhibitor FLIP De1 [Common bottlenose dolphin gammaherpesvirus 1 strain Sarasota] MSAAVIHQLEEALDEDEKERLLFLCRDIAADSAPGTIKDLLGVLRERGHLSPAHLAELMYRIRRFDLLKRALHINAATARAILGQQPPFISAYRVLLTDIDDGLEKSDVASLAFLLRNYTGGKQVDNDKGFLSLVIELEKLNLIAPDHLDLLESCLEKIHRPDLRTKIQKYNRVAPR	CASP8 and FADD-like apoptosis regulator isoform X11 [Tursiops truncatus]	Y, 1-175, 28-202
Herpesviridae	>Gallid alphaherpesvirus 2_146580 AEISNSIYQHPFPDGLSEEEKQKLERRRKRNRDAARRRRREQTYYVDKLH	>acc|GENBANK|ACE62906.1|GENBANK|EU725859|MEQ protein [Gallid alphaherpesvirus 2] AEISNSIYQHPFPDGLSEEEKQKLERRRKRNRDAARRRREQTYYVDKLHETCEELRRANEHLRKEIRDLRTECTSLRVQLACHEPVCPMAVPLTVTLGLLTN	hypoxia-inducible factor 1-alpha [Gallus gallus]	Y, 17-102, 22-104
Herpesviridae	>Gallid alphaherpesvirus 2_146625 HPFPDGLSEEEKQKLERRRKRNRDAARRRREQTYYVDKLHETCEELRRA	>acc|GENBANK|ACE62903.1|GENBANK|EU725856|MEQ protein [Gallid alphaherpesvirus 2] HPFPDGLSEEEKQKLERRRKRNRDAARRRREQTYYVDKLHETCEELRRANEHLRKEIRDLRTECTSLRVQLACHEPVCPMAVPLTVTLDLLT	basic leucine zipper transcriptional factor ATF-like 3 isoform X2 [Gallus gallus]	Y, 8-93, 22-104
Herpesviridae	>Gallid alphaherpesvirus 2_146706 MEANMSFENDYYSPIQLFAEIEAYANTMDKSPDLDILRTIEEFDETLLSE	>acc|REFSEQ|YP_001033977.1|REFSEQ|NC_002229|transactivating tegument protein VP16 [Gallid alphaherpesvirus 2] MEANMSFENDYYSPIQLFAEIEAYANTMDKSPDLDILRTIEEFDETLLSEIEVRTQSIPSPLVAPSVTKMSLPSPSPAPPNSLYTRLLHELDFVEGPSILARLEKINVDLFSCFPHNKHLYEHAKILSVSPSEVLEELSKNTWTYTALNLNEHGEMALPMPPTTKADLPSYVDDIQNFYLGELEAREKSYATMFYGYCR	MDV alpha TIF [Gallus gallus]	Y, 1-427, 1-427

		ALAEYIRQSAIKDLRDARVEDKNIGACSKARQYIAERYY REAARFAKLLYVHLYLSTTRDVSQRLEASQMGRQNIFV YLKCEWLQERHFHCLFQPVIFNHGVVIVEGRVLTAPELR AQNYIRSEFGLPLIRCKLVEEPDMPLISPPPFSGDAPRASV YLLQCIRSKLEVYSLSHPPNPQLHVHKEHVHVQKLESPP NYGTTVEALLMDSSDRNSISPGDPVATTISTL		
Herpesviridae	>Cercopithecine betaherpesvirus 5_70249 MASSVLSDLASLAMTVDFG LGCLEAYLRLNSGQQIPLPW PRGWCLVLQEL	>acc|REFSEQ|YP_004936001.1|REFSEQ|NC_012783|tegument protein UL24 [Cercopithecine betaherpesvirus 5] MASSVLSDLASLAMTVDFGLGCLEAYLRLNSGQQIPLP WPRGWCLVLQELEADDDYGPEDIKRLGQYLCCETRLTF VGRVVTEPVLTHGQRKKTLIILTCGEGQVFCYVEEDCCL YYVARSFLELQRVGLRAVDAIYCPRYMRSSAVKRGFRGL LRAWDISGDALSRYVTRRHGDYLHLPYPEGACLRLCNL KCFETSSGYGFMLRSLRQHFGKQVICLGSVDLRGEVPSS PPLLRWPATRVPVVLADGGRVFCFDLMNLRYIRLADDLA AFMCLGLRQLLDNRRFTGEHELHDSIPECPRGRQHDD	DR1 [Homo sapiens]	N, 20-109, 60-153
Herpesviridae	>Cercopithecine betaherpesvirus 5_70292 MFQFMFLLIFYIPLGYAANPC CSHPCQNRGVCMTVGFDRY WCDCTRTGFY	>acc|REFSEQ|YP_004935976.1|REFSEQ|NC_012783|prostaglandin G/H synthase 2 [Cercopithecine betaherpesvirus 5] MFQFMFLLIFYIPLGYAANPCCSHPCQNRGVCMTVGFD RYWCDCTRTGFYDRHCSTPEFLTRIKLLLKPTPDTVHYV LTHFPTLWWIVNNVPFLRNAVMRYILLTRSEMIDSPPLYN LDYGYNSWEAFSNLSYYTRTLPPVPRDCPTPLGVKGKA RLPDSTQIVDKVFIRRTFIPDPQGTNLMFAFFGQHFTHQF FRTDMVRGPAFTTGLGHGVDLSHIYGETLEKQHKLRLF KDGKMKYQVIDGEMYPPTVKETQVHMLYAPEVPDDLR FAVGNEVFGLVPGLMMYATIWLREHNRVCDVLKHEHPE WGDEQLFQTTRLIITGETINIIITQYVQQLSGYHFQLIFDP ELLFGTAFQYQNRIAAEFNTLYRWHSLLPDDFRIGDAKY GYEHFVYNNSILIKHGIAQLVESFTRQIAGRISGGRNIPPA LRRVSREGIEHGRKMRYQSLNEYRKRFRLKPYESFEELT GEKEIAAELEALYGDVEAVELYTGFIVEKPRPGAIFGESII ELGAPFSLKGLMANVICSPAYWKPSTFGGDVGLDIVKTA TIQSLICANVKGCPLAAFRVSNAELLKAFNGSSPSSLYSE QKAGSRVLINSRTEL	hypothetical protein EGK_01845 [Macaca mulatta]	Y, 15-603, 15-604
Herpesviridae	>Cercopithecine betaherpesvirus 5_70340 MKIYQARYDQTHSRFGPRA GSQCVCNCFMYIHALHLKG AHTTLGTDTLDQ	>acc|REFSEQ|YP_004936023.1|REFSEQ|NC_012783|large tegument protein [Cercopithecine betaherpesvirus 5] MKIYQARYDQTHSRFGPRAGSQCVCNCFMYIHALHLKG AHTTLGTDTLDQILTEGAHLDTAVEATLKQKRPGERLPV FRLGEEVPNIITSSFGKTAHALSRPFNGTAETRDLDGYTC YGIFDFLFYAYQKPRPVYVVVTVNALARAVILLEHDIFVF DPHASDRSRYAAVYQCDSLYDVVMLLTYFGTRLADFYY DALFVYMIDLSIQNVPESEINGIIISLFRDPDIALPVSALPP VPDSPVTVLPSLPAPPDSTPKKTPEKRKATPNVSHGGKK KPTTAKDKGSSRVSTTAPYNCADALAALTRYERLITQAE RETHNIVIRAPPTSGWILFSHSGLPFDETFLTDRMEQIVM SHIEHASCLLNRWRPSTPMYQQVRQLRGFSEEIDRFLSM WLGHELHLLGVYETIKSFNTSHLTPLQRAMIDKLRAVFI HYGAIHGPKVIKWVQQILKAVEKTNYTHLAVRLLEYAE ENPLDVDETFVCLRQQDFSAVVNAVNGRRQITSQQQEQ VQLALQQLSAAIYGIDSHDLERISFDNKDTKQTLSKLDS QAQTILLQRGNTKITELQGDLKRQISAMLQRRYNQIISGS LPVEELQAMQKRLEQAANLAQEMSELHLCDVNLQAPF KEMHEQLSYLITGHTATNSMSFSDELLQLRSQFTYATQV KEDTESKIHDLMLSIETAIQEPTTRSSNIAMAMVQEQLNE LQQLGGANIPEIATRLEKVHKVLNSLQQEEQAVRVFVHG LTYDNLPNDQTLKRQARLSDLLREDDSLHEIYIQKILDVF NTLLDRVSDKAFPKPQVFDIINSLINQLPQTSTLMKDLHT ANAALCQLSKQLEALAKVPRDKRLEALTELVQYFVSNS SLLVNLMNLQVGKTTLPLLYEHLKMELQDKHIQQAEAA WLQEAKKLTVTSADMVERFLQTAPSSTAADLARPDLQA KLRAFLEQEAKKQEEERKTVIKEQQGMVTTELARITDA VKAQTLSVIPTLNLGSLQDVISSLGADGRDILEKFNRNLL ASLSNLVKLMEERVSLCIQDLLSGQDTHYQQYQEDART MRQALTHIGTMMDSQLSQETLRTLSDLVRRSMFIEKCQL RNATSVFSNTDYAEDYKRYKEAQRQLDTQLREARMELH QQSATVQRALHQPDTRITPQNLTVGKDLPDRLNTESAGP FQRPVFQDVLKQQLETYKKTIKDETELMNTKLKSENEL RQAKLTSLSEQWSDLVTKHKMDAMEVVTPDAKNLIQN PLEAMTGLLAKAYTQMSYLDAQKVLNWALLFLHDAYN QIQNNPGHPCYAQLPNFPTLIQQAQSRLETVSIHVNNNA SCENFIAQHESASTATDREAIQAVETAWATLEAKRVAGGE ARYKKVQEVLLRMKQSLSDVELQDTLATAYYQLLNSIQ AFAYSLDFQTQLHKIRDLKARFADLIKQQHLNASEEVPL PMPHLPGNTVIASPLSFAKGLAALERYVLGGYQWLTECI NRQPLVCQRIDDIPAVLPSTDIDRKRVALDRLKRLSFSTK	large tegument protein [Homo sapiens]	Y, 1-2061, 1-2038

		NEMCYEVVDVFGLHQLMSKNGVPIHFVLSYGNVFFKY MALNNDDKQLAKKFAQVKNLVTGRYKVVTVNVAVAQT LKTFWSQISQYDLKPLLSGQTLIGLGETNSLVNLKIFIYIV VSAWNLQLDVLQEHRGPVIRIPIDDLCIAITTFYPEYIYGI VKHPIQNTLSSLVRVLKKDIVQEAINNVTQMPPVYSADEI KGFCINPKIWSSVNLSRVMWDQSLIRQLCDVGPRKNGA QKLWQYAVAIMVFPQDLLQCLWLELRPKFAEEFATLFDF FQALFVLFTHQYDITRESNMNTHLATGEPIVQTVGIRRK DHTDKSLLDIFIETDTAIDYVLGSWVFGIPVCCAIYVSEIL SGSRLLLARHIEYTVRDADFIHVQRAKDLNLNHVITQTW TNTPLEQCWFQAQIQRIKDHLRTPLELDFIPLVIYNAHDK TVHSVIRPPTTAERDVSRILVENPFPTLPLIDVPESDLVIFD RVPINTDFLREDPPPVYRQRGRSSGSGVSKGGKAAKASS AAPKKVQVVKTPQVDQVQAEGSSVDVPEVSDDEEDYEI LTEDESYELIELSSEEEQQSLPSVRTFVEPVFRQAALQTIV RKEDVTTVSPTAASSSFPNPPMTTLTQNVVNAIQILRSVR VDLQSMAKSVNETINRLRFLYLL		
Herpesviridae	>Cercopithecine betaherpesvirus 5_70422 MQYSLSHLLVATLLGTLLAS TMVFADKEERCLCPKTIQGI HPKNIQSVEL	>acc|GENBANK|ACQ55260.1|GENBANK|FJ883005|chemokine vCXCL4 [Cercopithecine betaherpesvirus 5] MQYSLSHLLVATLLGTLLASTMVFADKEERCLCPKTIQGI HPKNIQSVELHEPRDMCPNVEVIAKLKNGNEVCLNTEG PMVKKIIEKMRDREIERIQQQSQ	Neutrophil Activating Peptide-2 Variant Form M6I with Five Additioinal Amino Terminal Residues (Dsdly) [Homo sapiens]	Y, 28-92, 6-70
Herpesviridae	>Cercopithecine betaherpesvirus 5_70225 MADPVYVGGFLVRYDEPPG EAELFLPSGVVDRWLRDCR GPLPLNVNHDES	>acc|GENBANK|AEV80617.1|NEIGHBOR|FJ483969|capsid maturation protease [Cercopithecine betaherpesvirus 5] MADPVYVGGFLVRYDEPPGEAELFLPSGVVDRWLRDCR GPLPLNVNHDESATVGYVAGLQNVRAGLFCLGRVTSPK FLDIVQKASEKSELVSRGPPSESSLRPDGVLEFLSGSYSG LSLSSRRDINAADGAAGDAETACFKHVALCSVGRRRGT LAVYGRQPDWVMERFPDLTEADREALRNQLSGSGEVAA KESAESSAAAAVDPFQSDSYGLLGNSVDALYIQERLPKL RYDKRLVGVTARESYVKASVSPAEQETCDIKVEKERPKE PEQSHVPTESMSHPMSAVATPAASTVAPSQAPLALAHDG VYLPKDAFFSLIGASRPLAEAAGARAAYPAVPPPPAYPV MNYEDPSSRHFDYSAWLRRPAYDAVPPLPPPPVMPMPY RRRDPMMEEAERAAWERGYAPSAYDHYVNNGSWSRSR SGALKRRRERDASSDEEEDMSFPGEADHGKARKRLKA HHGRDNNNSGSDAKGDRYDDIREALQELKREMLAVRQI APAALLAPAQLATPVASPTTTTSHQAEASEPQASTAAAAP STASSHGSKSAERGVVNASCRVAPPLEAVNPPKDMVDL NRRLFVAALNKME	protease/assembly protein [Homo sapiens]	Y, 5-589, 4-527
Herpesviridae	>Cercopithecine betaherpesvirus 5_70502 MTLVLFATEYDSAHIVANVL SKSASEHCVFPLLVKHHASN LIYFCLQTQK	>acc|REFSEQ|YP_004936035.1|REFSEQ|NC_012783|helicase-primase primase subunit [Cercopithecine betaherpesvirus 5] MTLVLFATEYDSAHIVANVLSKSASEHCVFPLLVKHHAS NLIYFCLQTQKCTDSQRVSPVFVVNNDVLNLAHYLHTR QPIPLSALVDSLNEDETRPIYNHLFRTVISPEHGGEVREFK HLVYFHHTAIVRYLNLIFLCPTSPSWFISVFGHTEGQVLL TMFYYLLERQYSTISTVEEYVRSFSRDLGVIIPTHATMSE FTRLLLGSPFRSRIPQFVQYAMARNQRDYDELLHVDTRI NTFREHARLPDTVCVHYIYLAYRTALSRSRLLKYREVVA YDESANGNGNGSSGEQCKKQPLFLGRHLEEDLLDVMK KYFSLPNFLQDYIETRMLLTELHDLRLSGYNYRTESPTLT GFFGTSSQVMRKLDHINSMSDAVFPPLERSLSGLLRLCA SLKTANTYATGTLAQYSQRQFLLPKESHHDYPIPLFRVQL PSDHHVFCAVTSENWHQSLFPSDLLKHVPDSHFSDEALT DMIWLHDDDVASSNPETQFYYTRHEIFNERLPTHNFVAD FDLRLRDGVNGLSKEVIFEICRGLRRVWVTVWESLFGDT DPERHPVYFFKSACKTSTPDFYDDDNPPPSYEIRTDYCK CTDKLGLRIITPFPERTLVINPSVLKAIAQVLNHAICLDIPL HEHLDPISHPESSLDTGIYHHGRSVRLPFMYKMDQEDG YFMHRRLLPIFIIPEGFRDHPLGFVRAQLDVRNLLHHHPP RPPPGSSSVQTTPSPVTRIVLSVRDKLCPSPEINFMETRSV NVTRYEKRTLSDVISYHLHGVAGRRPEHGHADDTTDLQ RLVVTRVWPLLLEHLTQHYDSKVSEQFTSQHTLTFQPHG PHCVSVKRLDGSRTRDFRCLNYTHRNPQETVQVFIDLRT EHSYALWASLWSRCFTKKCHSNAKNVHISVKIRPPTADQ	helicase/primase complex protein [Homo sapiens]	Y, 1-897, 1-855

Table 6. List of host genes recognized by DAVID (highlighted in blue boxes) and unable to be identified (highlighted in yellow boxes) by DAVID's gene conversion tool.

Accession ID	Host species	Gene name
J9E9X3_AEDAE	Aedes aegypti	AAEL017001-PA(AaeL_AAEL017001)
DYR_HUMAN	Homo sapiens	dihydrofolate reductase(DHFR)
GCNT3_BOVIN	Bos taurus	glucosaminyl (N-acetyl) transferase 3, mucin type(GCNT3)
PD1L1_HUMAN	Homo sapiens	CD274 molecule(CD274)
NC2B_HUMAN	Homo sapiens	down-regulator of transcription1(DR1)
CXCR1_MACMU	Macaca mulatta	C-X-C motif chemokine receptor1(CXCR1)
DJC14_BOVIN	Bos taurus	DnaJ heat shock protein family(Hsp40) member C14(DNAJC14)
OX2G_RAT	Rattus norvegicus	Cd200 molecule(Cd200)
H9FLI1_MACMU	Macaca mulatta	prostaglandin-endoperoxide synthase 2(PTGS2)
GG6L6_HUMAN	Homo sapiens	golgin A6 family-like 6(GOLGA6L6)
IL10_HUMAN	Homo sapiens	interleukin 10(IL10)
D6W5Y8_HUMAN	Homo sapiens	melanoma associated antigen (mutated) 1(MUM1)
BIRC3_HUMAN	Homo sapiens	baculoviral IAP repeat containing 3(BIRC3)
SODC_HUMAN	Homo sapiens	superoxide dismutase 1, soluble(SOD1)
PCNA_HUMAN	Homo sapiens	proliferating cell nuclear antigen(PCNA)
DPOD1_HUMAN	Homo sapiens	DNA polymerase delta 1, catalytic subunit(POLD1)
MARH2_HUMAN	Homo sapiens	membrane associated ring-CH-type finger 2(MARCH2)
A0A096MNR5_PAPAN	Papio anubis	kelch like family member 32(KLHL32)
Q6FGW4_HUMAN	Homo sapiens	interleukin 10(IL10)
G137C_HUMAN	Homo sapiens	G protein-coupled receptor 137C(GPR137C)
DPOLA_HUMAN	Homo sapiens	DNA polymerase alpha 1, catalytic subunit(POLA1)
MIF_MACFA	Macaca fascicularis	macrophage migration inhibitory factor (glycosylation-inhibiting factor)(MIF)
Q90603_CHICK	Gallus gallus	MDV alpha TIF(LOC396264)
TNR14_HUMAN	Homo sapiens	TNF receptor superfamily member14(TNFRSF14)
COOA1_MOUSE	Mus musculus	collagen, type XXIV, alpha 1(Col24a1)
CXCL7_HUMAN	Homo sapiens	pro-platelet basic protein(PPBP)

GLP2R_HUMAN	Homo sapiens	glucagon like peptide 2 receptor(GLP2R)
Q05D16_MOUSE	Mus musculus	CD200 antigen(Cd200)
CXL13_HUMAN	Homo sapiens	C-X-C motif chemokine ligand 13(CXCL13)
F1NJK2_CHICK	Gallus gallus	hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)(HIF1A)
SODC_BOMMO	Bombyx mori	Cu/Zn superoxide dismutase(SOD)
GROA_HUMAN	Homo sapiens	C-X-C motif chemokine ligand 1(CXCL1)
RL40_HUMAN	Homo sapiens	ubiquitin A-52 residue ribosomal protein fusion product 1(UBA52)
VE7_HPV18 A0A6J0BBU4_NEOLC A0A7E5WV88_TRINI AAK76747.1 A0A3Q2U701_CHICK R0JMS0_ANAPL A0A126LAY4_HUMAN AMD82184.1 A0A0F7JHQ2_PLOIN XP_035442936.1 A0JKF3_BOMMA Q86XW6_HUMAN Q2PQR1_OSTFU 6LJJ_A A0A1S6L094_BOMMO A0A287A4F2_PIG Q7JGJ7_PAPHU A0A650FP55_PIG SFRICE_015877 A0A126LAX2_HUMAN Q962S3_SPOFR A0A126LAZ9_HUMAN A0A482IBS2_PLUXY 6A42_A Q0PQY0_PIERA A0A2K5UM46_MACFA A0A0L7LNP9_9NEOP A0A4D6YMU3_RAT CAB44705.1 F2VYD2_CHRCC AAQ03217.1 G7MF20_MACMU G9LPN2_HELAM A0A2W1BUB5_HELAM A0A6J0C5C8_NEOLC A0A126LB10_HUMAN A0A2S0RQQ5_PIERA D7NY61_ROULE AMD82173.1 A0A0L7KZE6_9NEOP A0A2U4BNB2_TURTR H9BE67_ANTPE P04218.1		