The bioinformatic characterization of five novel poxviruses

by

Shin-Lin (Cindy) Tu
BSc, University of Victoria, 2015

A Thesis Submitted in Partial Fulfillment of
the Requirement for the Degree of

MASTER OF SCIENCE

in the Department of Biochemistry and Microbiology

© Shin-Lin (Cindy) Tu, 2018
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy or other means, without the permission of the author.

Supervisory Committee

The bioinformatic characterization of five novel poxviruses

by

Shin-Lin (Cindy) Tu
BSc, University of Victoria, 2015

Supervisory Committee

Dr. Chris Upton, Department of Biochemistry and Microbiology
Supervisor

Dr. Caroline Cameron, Department of Biochemistry and Microbiology
Departmental Member

Dr. John Taylor, Department of Biology
Outside Member

Abstract

Poxviruses are double stranded (ds) DNA viruses with large brick-shaped virions (~200x300nm) that can be seen by light microscopy. The *Chordopoxvirus* (ChPV) subfamily demonstrates a vast genetic diversity in poxvirus virulence and evolution, and infects a wide range of vertebrate hosts including human/primates, rodents, birds, squirrels, and many economically important ruminants. There are at least 14 distinct ChPV genera, whose members have genomes that range between 127-360 kbp, and can be either GC-rich (33-38% A+T base composition) or AT-rich (up to 76% A+T). My work in the assembly and annotation of novel poxviruses serves to enrich the poxvirus sequence repository and further virulence characterization, comparative analysis, and phylogenetic studies.

Using a variety of programs, as well as tools developed by the Virus Bioinformatics Research Centre, a protocol is created, refined, and applied to the assembly and annotation of novel poxviruses: Pteropox virus (PTPV) from a south Australian megabat *Pteropus scapulatus*, Eptesipox virus (EPTV) from a north American microbat *Eptesicus fuscus*, sea otter poxvirus (SOPV) from the north American *Enhydra lutris*, and two Kangaroopox viruses western and eastern Kangaroopox viruses (WKPV, EKPV) from the Australian *Macropus fuliginosus* and *Macropus giganteus*. This is the first time poxviruses from these vertebrate hosts are assembled in full, and the result supports the establishment of 4 new ChPV genera.

The two bat-isolated poxviruses, PTPV and EPTV, likely did not co-speciate with their hosts despite infection of related host species. Instead, EPTV forms a sister clade with the Clade II virus, and together forms a sister group with the orthopoxviruses. On the other hand, PTPV and SOPV are each other's closest extant relatives despite the distant

geographical location from which they were isolated; together they share a novel homolog of TRAIL (Tumor necrosis factor-Related Apoptosis-Inducing Ligand) never before seen in poxviruses. SOPV additionally encodes distinct interleukin (IL)-18 binding protein and tumor necrosis factor (TNF) receptor-like protein that could have novel immune-evasion roles. The KPVs present the first case of a putative viral cullin-like protein, which might be involved in regulating the host ubiquitination pathway. Altogether, these novel proteins can potentially serve as new virokines and viroceptors in the form of viromimicry pathogenesis; they demonstrate the capacity and diversity with which poxviruses modulate host immune responses in their favour, and should be studied further.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Å	Angstrom
aa	Amino acid
APC	Anaphase-promoting complex
APVs	Avipoxviruses
AT	Adenine + thymine
ATI	A-type inclusion
BLAST	Basic local alignment search tool
BP	Binding protein
CCTOP	Constrained consensus topology prediction
CD	Cluster of differentiation
CDC	Centre for disease control
CDD	Conserved domain database
CEV	Cell-associated enveloped virus
ChPV	Chordopoxvirus
CPU	Central processing unit
CRD	Cysteine-rich domain
CRL	Cullin RING E3 ligase
CTD	C-terminus domain
DBG	de Bruijn graph
DNA	Deoxyribonucleic acid
dsDNA	Double stranded DNA
EEV	Extracellular enveloped virus
FDA	Food and drug administration
GAG	Glycosaminoglycan
GC	Guanine + cytosine
HGT	Horizontal gene transfer
HIV	Human immunodeficiency virus
HMM	Hidden Markov models
HP	Hypothetical protein
I-TASSER	Iterative threading assembly refinement
ICTV	International committee on taxonomy of viruses
IEV	Intracellular enveloped virus
IFN	Interferon
Ig	Immunoglobulin
IκB	Inhibitor of kappa B
IL	Interleukin
IMV	Intracellular mature virus
indels	Insertions/deletions
ITR	Inverted terminal repeats
IUCN	International union for conservation of nature
kbp	Kilo base pairs
KPVs	Kangaroopox viruses
MAFFT	Multiple alignment using fast fourier transform
Mbp	Mega base pairs

MHC	Major histocompatibility complex
MIRA	Mimicking intelligent read assembly
ML	Maximum-likelihood
MSA	Multiple sequence alignment
MUSCLE	Multiple sequence comparison by log-expectation
MVA	Modified vaccinia Ankara
NK	Natural killer cells
N-WASP	Neural Wiskott-Aldrich syndrome protein
NCBI	National center for biotechnology information
NCLDV	Nucleo-cytoplasmic large DNA viruses
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NGS	Next generation sequencing
nt	Nucleotides
OLC	Overlapping-consensus
OPVs	Orthopoxviruses
ORF	Open reading frame
PACR	Poxvirus anaphase-promoting complex/cyclosome regulator
PCR	Polymerase chain reaction
PDB	Protein database
pI	Isoelectric point
PKR	Protein kinase R
PSI-BLAST	Position-specific iterative basic local alignment search tool
PSSM	Position-specific scoring matrix
RAP	RNA polymerase-associated protein
RAxML	Randomized axelerated maximum likelihood
RMSD	Root-mean-square deviation
RNA	Ribonucleic acid
RPO	RNA polymerase
RPS-BLAST	Reverse position-specific basic local alignment search tool
SAM	Sequence alignment map
SARS	Severe acute respiratory syndrome
SNP	Single nucleotide polymorphism
SPAdes	St. Petersburg genome assembler
STAT	Signal transducer and activator of transcription
TF	Transcription factor
TNF	Tumor necrosis factor
TNFR	Tumor necrosis factor receptor
TRAIL	Tumor necrosis factor-related apoptosis inducing ligand
VACV-Cop	Vaccinia Copenhagen strain
VBRC	Viral bioinformatics resource centre
VETF	Viral early transcription factor
VOCs	Viral orthologous clusters
WGS	Whole genomic sequences
WHO	World health organization

*For individual poxvirus species name and abbreviation, see **Table 1**.

Acknowledgement

I would like to thank my supervisor Dr. Chris Upton for his constant support throughout this Master's program. He has thought of me in numerous exciting research opportunities, invested countless meetings for my projects, and self-exemplified the diligence and precision of a scientist each time we go through our manuscripts. Every external seminar notices/podcasts/blog posts/online course materials he has shared with the lab has taught be to be resourceful with my learning. He has provided tremendous mentorship and patience to both the academic and personal growth aspects of this journey, and supported my ventures at expanding my portfolio with professional development workshops and science outreach programs. I would also like to thank my supervisory committee members: Dr. Caroline Cameron and Dr. John Taylor, for their guidance and all the critical pointers/questions brought up in our meetings. To my friendly collaborators Dr. Mark O'Dea, Dr. Mark Bennett, and Jessica Jacobs: thanks for sharing your expertise and cultivating young scientists like me in your projects. Likewise, I couldn't have done this without the people I met along the way: Ragha for all the protein wisdoms he conferred to me; Melinda for all the support and assistance in the very stressful time that is thesis and defence preparations; my lab mates (Kathleen, Deyvid, Simar, Navpreet, Luke, Kaegan, Andrea, David, Caity, Alex, Farzana) whom I had the pleasure to work/have fun together; and most importantly Chad and Jacob for all the crucial training and growth I've had: the journey would not be as fruitful without you.

Finally, to my family and Jonathan: thanks for having my back in my most trying and rewarding time, yet.

Chapter 1. INTRODUCTION

1.1 What are poxviruses?

Poxviruses are double stranded (ds) DNA viruses with genomes between 127-360kbp, encoding between 100-330 genes. Members of the *Poxviridae* family have large ovoid or brick-shaped virions (~200x300nm) that can be seen by light microscopy. Poxviruses fall under Group I with other dsDNA viruses under the classic Baltimore classification[1], which categorizes viruses into 7 different groups based on their modes of replication, mRNA synthesis, genome composition (DNA or RNA) and structure (ds or single stranded). Group I viruses transcribe mRNA directly from the DNA template, and include the herpesviruses, the adenoviruses, and the papillomaviruses. However, unlike the dsDNA viruses at the time, poxvirus was the first of this group found to replicate in the cytoplasm instead of the nucleus. Besides *Poxviridae,* 8 other large dsDNA virus families have since been found to replicate (fully or partially) in the cytoplasm as well: *Ascoviridae, Asfarviridae, Iridoviridae, Marseilleviridae, Megaviridae, Pandoraviridae, Phycodnaviridae,* and *Pithoviridae*. Poxviruses and these virus families are collectively termed as the nucleo-cytoplasmic large DNA viruses (NCLDV); they group into a monophyletic branch and share 5 genes[2-4]. The NCLDV include some of the largest viruses discovered to date: the mimivirus[5] discovered in 2003 (1.2 Mbp genome), the 2013 pandoraviruses[6] (2.5 Mbp genome), and the 2014 pithovirus[7] discovered and revived from permafrost (1.5μm virion size). These cytoplasm-replicating large viruses, which can exceed some bacteria both in terms of genome and virion sizes, sparked debates on domains of life[8-10], and inspired hypotheses of an ancient virus world[11,12], as well as the origin of eukaryotic cells[13,14].

The *Poxviridae* family itself is divided into two subfamilies: *Entomopoxvirinae*, whose members infect insects; and *Chordopoxvirinae*, whose members infect vertebrates. Today, the International Committee on Taxonomy of Viruses (ICTV) recognizes 11 genera of Chordopoxviruses (ChPVs) based on phylogenetic relationship from sequence data[15]. Most ChPV nomenclature follows the guideline outlined by the ICTV, and viruses are named with a prefix derived from the host in which the virus was originally/commonly isolated from, followed by an appendage of "pox". **Figure 1** shows the phylogenetic topology of the poxvirus family with varied base composition highlighted from high GC% (red) to neutral (light blue/red) to high AT% (blue).

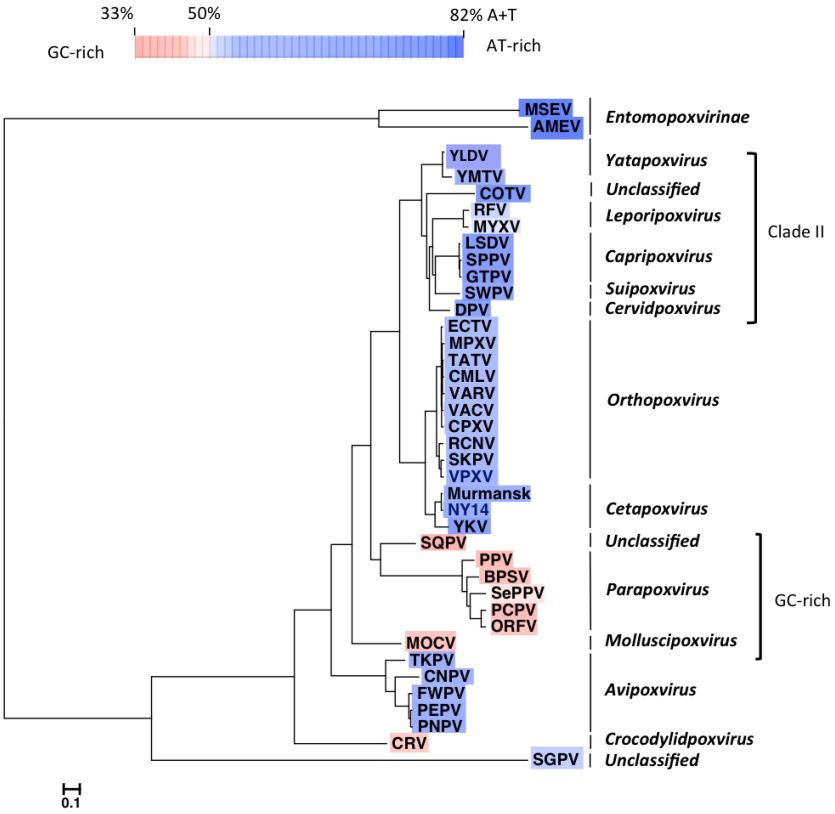

Figure 1. Phylogenetic relationship of Poxviridae family and its representative genera and species.

The topology and base composition of genera in the Chordopoxvirus subfamily (33-72% AT) using the Entomopoxvirus subfamily as outgroup (82% AT); maximum-likelihood phylogeny is generated from a multiple sequence alignment of concatenated amino acid sequences of seven conserved proteins: RPO147, RAP94, mRNA capping enzyme large subunit, P4a precursor, RPO132, VETF-L and DNA primase; the colour gradient is used to capture base compositions, whereby the intensity of red and blue proportionally represent the GC or AT richness of the genome, respectively.

ChPV species from their corresponding genera are referred to as *Avipoxvirus, Molluscipoxvirus, Leporipoxvirus, Orthopoxvirus, Centapoxvirus, Yatapoxvirus, Capripoxvirus, Suipoxvirus, Parapoxvirus, Cervidpoxvirus,* and *Crocodylidpoxvirus*. However, several newly sequenced and unclassified viruses such as the cotia virus (COTV), the squirrelpox virus (SQPV), and the salmon gill poxvirus (SGPV) will each require a new genus; bringing the minimum number of extant poxvirus genera up to 14. For most ChPV comparative analyses (including this thesis), the SGPV has been excluded due to its ancient divergence from the rest of the ChPVs. Of these poxviruses, the orthopoxviruses (OPVs) are extensively studied as the prototype models of poxvirus biology[16] because this group includes the variola (VARV) and vaccinia (VACV) viruses, which are the etiological agent of smallpox and the smallpox vaccine virus, respectively. In fact, 12,998 of the 17,669 PubMed "poxvirus" search results are associated with "orthopoxvirus". In contrast, the sister clade consisting of the COTV, cervidpoxviruses, leporipoxviruses, yatapoxviruses, suipoxviruses, and capripoxviruses and are referred to as the Clade II viruses[17]; members of this clades infect mice, deer, rabbits, monkeys, pigs, and the economically important ruminants sheep/cattle, respectively. Notably, the evolution and virulence of Myxoma virus (MYXV; a *Leporipoxvirus* member), is one that has been extensively tracked and characterized since its introduction as a pest control for the Australian rabbit population 60 years ago[18-23]. Exclusive to members of the Clade II, these viruses share a unique rearrangement of the C7L gene (Type I IFN inhibitor) and E7R gene (a myristylated protein) apart from the usual synteny seen in the rest of the ChPVs[24]. The ChPVs have a wide range of base compositions from 33-67% AT, but the drive behind this divergence in base composition is currently unknown[25,26]. The term "GC-rich viruses" is used in this thesis to denote the subset of viruses with 33-38% AT (SQPV, parapoxviruses, molluscum contagiosum virus (MOCV))[27] in contrast to the rest

of the neutral or more AT-rich viruses (56-76% AT); this GC-rich characteristic[28] is reflected to some extent on the amino acid-based phylogenetic tree (**Figure 1**). Among them, MOCV is a strict human pathogen[29], while the parapoxviruses , which largely infect even-toed ungulates, frequently demonstrate zoonotic infections[30-35]. SQPV, on the other hand, nearly wiped out the naïve UK red squirrel population upon introduction of the naturally immune US grey squirrels[36-38]. Finally, members of the *Avipoxvirus* (APV) genus demonstrate strict infection of domestic and wild bird species; their global distribution, prevalence, and effects on the poultry industry have led to more and more members being sequenced in recent years[39-43].

1.2 Poxviruses in history

The most infamous poxvirus in history is the VARV, which caused smallpox epidemics that killed 300-500 million people in the course of the human history (including at least 18 reigning monarchs). According to the Centre for Disease Control (CDC), smallpox as a disease was transmitted primarily through direct contact with the respiratory droplets from infected individuals, and underwent a long asymptomatic incubation period of 7-19 days. The clinical symptoms began with acute onset of fever and symptoms similar to the common cold, but ultimately manifested as pustular rashes all over the body. Altogether, any malignant and haemorrhagic rashes and/or derivative respiratory complications caused a fatality as high as 30% (VARV major strain).

Early archaeological evidence found pox-like lesions on the mummy of the Egyptian pharaoh Ramses V from 3000 years ago (1157 B.C.)[44]. A smallpox-like disease described as the "Plague of Athens" was imported into Greece in the Peloponnesian War (430 B.C.)[45] then

again later in Rome as the "Antonine Plague" (170 A.D.). Unambiguous records of smallpox emerged in 3rd century BC China, and 6th century A.D. in Europe. The emergence of a highly pathogenic virus, such as VARV, is theorized to have arisen after an ancestral virus (likely one with a broad host range) crossed the species-barrier to infect a new host, and underwent rapid "post-transfer adaptation" in order to fine-tune and optimize replication[46]. This phenomenon is widely observed in viruses that cause other pathogenic zoonotic diseases, such as HIV-1 from primates, influenza virus type A from birds, and SARS coronavirus and Ebola virus from bats. Various phylogenetic evidence dated the divergence of VARV from its nearest relatives, the taterpoxvirus (TATV) and the camelpoxvirus (CMLV), to around 3000-4000 years ago[47-51]. Subsequently, these researches attempted to map the region of VARV emergence. One interesting hypothesis suggested the eastern African continent as a likely region of VARV emergence[52], whereby the geographical distribution of the naked sole gerbils (the only host of TATV[53]) was met with the historical introduction of domesticated camels into Africa with a large human settlement also 3000-4000 thousand years ago. This co-localization of rodents, camels, and humans potentially enabled the ancestral, broad host range virus to make jumps into these respective hosts. Subsequent gene-loss events in individual hosts likely gave rise to the narrow host ranges seen in TATV, CMLV, and VARV today[54].

During its course of rampage, observation of induced immunity in individuals recovered from smallpox gave rise to the risky practice of "variolation", which involved the inoculation of scabs and puss directly from a smallpox lesion (the primary source). A similar effect was also observed in those infected with the "cowpox" from the namesake animal source, and the primitive inoculation experiments using cowpox lesions (a secondary source) became what would later be known as a "vaccination". Subsequently, it is the normalization of vaccination,

as practiced and promoted by Dr. Edward Jenner in 1796, that set up the fundamental basis of immunology[55]. The practice of vaccination was sustained and advanced with the laboratory VACV, which became the prototype model for vaccine development[16,56] as well as model systems for studying poxviral biology[57,58], virulence[59-64], and gene expression[65-67]. By convention, poxvirus orthologous genes are typically referenced against the VACV-Copenhagen strain name when referred in comparative analyses (e.g.: J6L would be used in reference to the VACV-Cop nomenclature for RPO147 orthologous genes, while J6 is used for the protein). Interestingly, historical records and later molecular studies show that VACV actually originated from a horsepox virus that caused an affliction called "grease" in horses, which subsequently infected cows as well[55,68,69]. According to records from the World Health Organization (WHO) and Centre for Disease Control (CDC), the increased understanding and development of effective vaccines freed North America and Europe of smallpox in 1952 and 1953, respectively. In contrast, Australia and New Zealand were never widely endemic with smallpox as they were likely protected by geographical distance from everyone else[70]. From the promises shown by these examples, an intensified global smallpox eradication campaign was announced by the WHO in 1967. Health professionals from around the world conducted massive door-to-door searching for remaining cases in endemic regions of South America, Asia, and Africa. Since smallpox was not a chronic infection, had a stable serotype and no animal reservoir, epidemic hotspots can be easily contained by quarantine of the infected individuals, and vaccination of those around them[71]. A decade and $300 million later, the global eradication of the disease was declared in 1980.

As demonstrated from the emergence of VARV, the historical success of smallpox complied with the three steps of viral disease emergence or re-emergence[72]: (1) introduction of a viral pathogen into a new host species, (2) establishment of the pathogen in the new host, and (3)

efficient dissemination of the pathogen in the new population that bring about outbreaks, epidemics or pandemics. At various stages of the human history, VARV was met with changes in the human demographics such as growth in population and density (agricultural expansion and industrialization), and its transmission enabled by the changes in human behaviours through colonialism, war, famine, or development of commerce[73]. Similarly, our modern world presents opportunities for pathogens through changes in climate, ecosystems, accessible international travel, or even changes in political landscapes and resistance against vaccines[74,75]. For these reasons, it remains crucial that we seek to understand poxvirus evolution and pathogenesis, and continue poxvirus research, surveillance, and advancement of vaccines.

1.3 Poxvirus concerns and research today

Today, the emergence of poxviral diseases is followed by public health institutes around the globe[76]. Three known agents are under surveillance at the CDC: MOCV, Monkeypox virus (MPXV), and Orf virus (ORFV); in contrast to the zoonotic nature of the latter two viruses, the former strictly infects humans. Typically, these poxviral diseases have been benign. However, according to WHO, at least four MPXV outbreaks have occurred in the last 20 years (once in the US), with one strain (Central African) causing up to 10% fatality. Considering our largely unvaccinated generation today[77], and the rate at which viruses can evolve (genome adaptability)[78], threats may be imminent.

In addition, other sporadic cases of poxvirus zoonotic transmissions have also been observed around the world. Cowpox virus (CPXV) infections reported worldwide have raised public health concerns for the post-smallpox vaccination population[79]. VACV has been circulating in Brazil since the 1960's, recent outbreaks from 2000 and 2010 observed VACV infections

of dozens of dairy workers that resulted in high fevers and painful pustules[80-84]. In 2004, a college student contracted tanapox virus (TANV) while doing animal research[85]. A case of sealpox (SePPV) infection was reported in 2005 after a marine mammal technician, who was bitten by a seal, developed ORFV infection-like symptoms[86]. In 2012 and 2013, two patients from the United States with equine exposures acquired novel poxvirus infections[87]. More recently, in 2014, an immunosuppressed patient in New York developed rash, and blister-like lesions, possibly from a feral cat, that led to a 15x15cm ulceration on the flank; subsequent analysis uncovered a novel poxvirus species (NY14 virus on **Figure 1**)[88,89]. These clinical cases demonstrate the prevalence of zoonotic poxviruses in our environment.

Lastly, there remains concerns for smallpox bioterrorism warfare against today's unvaccinated generation[90-92]. Sources of the smallpox agent include the remaining stocks of variola virus at two government institutes, other (modified) zoonotic poxviruses, or the *de novo* synthesis of infectious poxvirus virions (a method recently proved possible[93]). For reasons above, researchers today are propelled to further advance our understanding of poxvirus pathogenesis, map out poxviruses in our environment, and develop new diagnostic/analytic tools and vaccines[94-97].

Originally, the first generation smallpox vaccines demonstrated higher-than-ideal fatality rates[98] such that pre-emptive mass-vaccination is not a recommended solution. Consequently, the continual research effort of a modified VACV Ankara strain (MVA) led to a highly attenuated and promising third-generation smallpox vaccine that is currently under evaluation by the Food and Drug Administration (FDA)[99]. The MVA strain lost up to 15% (~30kb) of its genome (mostly virulence genes) during serial passage, and is unable to replicate fully in mammalian cells[100]. The loss of immunomodulatory genes involved in host evasion

(virostealth), such as type I and II interferons (IFNs), cytokines, and chemokines, also meant that MVA could elicit strong immunogenic response in hosts. These features, on top of the accumulated research done on VACV biology and gene expression, made recombinant MVA a top contender as a smallpox vaccine, as well as a vector for numerous other diseases and cancer immunotherapies[101-105]. Alternatively, since APVs can't replicate in non-avian hosts, but can efficiently express recombinant genes in mammalian cells to induce immune responses, other researches have looked at using APVs as another vaccine vector for human diseases without the risks for VACV infection[106-109].

Enabled by the advancement in sequencing and assembly technologies, other research has focused on the sequencing of ancient and novel poxvirus samples to map out the evolution of certain lineages and to fully elucidate the phylogeny of poxviruses. To date, full VARV genomic sequences have been extracted and assembled from a near 400 years old Lithuanian child mummy[110], as well as from two separate Czech museum samples dating back at least 60 and 160 years ago[111]. A recent re-assembly and analyses of the 400 years old VARV genome estimated an average mutation rate of 40 single nucleotide polymorphisms (SNPs) selected across VARV strains per 100 years[112]. In contrast, the detection and sequencing of novel poxviruses in new host populations today is met with the aforementioned urgency to map the prevalence of poxviruses in our environment. Sequencing and assembly of novel genomes will advance our understanding of poxvirus evolution, virulence, virus-host interactions, and/or provide the basis to establish surveillance program of those with zoonotic potentials[94]. Furthermore, genomic research that catalogues novel genes will form the basis for subsequent experiments.

1.4 Poxvirus biology

1.4.1 Genome organization and evolution

Recall that members of the *Poxviridae* virus family are large viruses with dsDNA genomes between 127-360kbp with varying AT base compositions. Neutral poxvirus mutation rates have been determined from OPV studies to be approximately on the scale of 10^{-6} substitutions/site/year[49], which is about 1000 magnitude more than average mammalian genomes[113], but 10-100 fold less than RNA viruses[114]. About 100-300 genes, transcribed from both strands of the DNA, are tightly packed into a poxvirus genome with little overlapping of the open reading frames (ORFs). The poxvirus genome is organized whereby a central core with 81 genes (important for replication and transcription) is conserved across all species. Moving outwards from this core, the terminal sequences become more susceptible to recombination events, and encode a variety of virulence genes among different viruses that contribute to unique host range, tropism, immunomodulation, and/or pathogenesis[89,115]. The ends of the genome contain characteristic "inverted terminal repeats" (ITRs), which vary in length between viruses, and form covalently closed hairpin termini.

In addition to SNPs, there are four types of genome evolution for poxviruses: (1) gene-loss, (2) gene-duplication, (3) recombination, and (4) foreign gene captures via horizontal gene transfer (HGT) events. It is thought from prototypic OPV studies that the main driver of poxvirus evolution was gene-loss events. It was speculated that the ancestral poxvirus was a virus with a broad host range that underwent an overall loss of genes rather than the gain of genes that led to speciation and gave rise to narrower host range[116,117]. Comparative analyses demonstrated that CPXV contain the most complete gene set most similar to the ancestral virus, whereas the rest of the OPVs had underwent sequential gene losses (and no gene-gains) since the last OPV common ancestor[17,118]. In contrast, gene-duplication (manifested in the

forms of paralogs or multi-gene families) is an opposite type of genome diversification, and is often lineage-specific[17,119,120]. This is particularly evident in canarypox virus (CNPV) and other APVs. With genome size up to 360kbp (approximately twice the size of the prototype VACV genome), 138 of CNPV genes form 14 multi-gene families (49% of the genome)[121]. Sequence recombinations have been extensively observed in poxvirus genomes, and are essential to the creation of new phenotypes and genetic diversity[122-126]. The observation of a block of unique single nucleotide polymorphisms (SNPs) found solely in the O1L virulence genes (activation of extracellular-signal regulated kinase) of VARV, TATV, and CMLV (each other's closest relatives) is thought to restrict host range seen in this subclade of the OPVs[54]; the source of the sequence block is, however, yet unknown. On the other hand, extant poxviruses have been shown to encode virulence genes which were horizontally transferred from hosts[127,128]. Examples of ones involved in host immune defense mechanisms include the major histocompatibility complex (MHC) class I[129], interleukin (IL)-10[130], the interferon gamma (IFN-γ) receptor[131,132], and tumor necrosis factor receptors (TNFR)[133-135]; others, like the glutaredoxin and glutathione peroxidase[136], are involved in resistance to cellular oxidative stress. Together, these recurrent features shape up the diverse poxvirus evolution and unique pathogenesis we see today.

1.4.2 Life cycle

Recall that poxviruses have large, characteristic ovoid or brick-shaped virions (~200x300nm) that can be seen even by light microscopy. Each virion core encloses a genome, along with several viral enzymes, early transcription factors (TFs), and an RNA polymerase complex. Recently, poxviruses virions have been found to incorporate transcripts as well[137]. The poxvirus life cycle is temporally controlled by 3 stages of gene expression: early, intermediate, and late; each stage is governed by a specific promoter activated *only* by TFs

expressed in the previous stage[66].

The poxvirus life cycle begins with the binding of the infectious virion particles to ubiquitious cellular surface elements such as the glycosaminoglycans (GAGs)[138] or laminin[139]; interestingly, no specific host-cell receptors have been identified for poxvirus entry to date[140-142]. Infectious virions come in three main forms: the IMV (intracellular mature virus), the CEV (cell-associated enveloped virus), and the EEV (extracellular enveloped virus). IMVs are thought to disseminate with the rupture of cellular membrane, whereas CEV is attached to the cell membrane, but can bud off as EEVs in the form of free particles, which is critical for rapid cell-cell spread[143-145]. IMVs have at least 20 non-glycosylated surface proteins, but EEVs have only about 6 on their additional membrane[142]. Through unclear mechanisms, these proteins form fusion/entry complexes at the cellular membrane. Due to the non-specificity of binding and fusion mechanisms, it follows that any host restriction mechanisms must happen after virus entry (elaborated in Chapter 1.4.3 Pathogenesis: virulence and host range). After the fusion event, the virion core enters the cell. Note that the hurdles to bypass the nuclear envelope barrier that most viruses face[146] does not apply to poxviruses, as their life cycle occurs entirely in the cytoplasm. Here, RNA polymerase and TFs bind to early promoters ("AAAxTxGAAAxxTA") to transcribe and express early products that are involved in "precursor metabolism" (thymidine kinase, ribonucleotide reductase, dUTPase), "replication" (DNA polymerase, helicase-primase, uracil DNA glycoslase, and DNA ligase etc…)[147], and "virulence" genes that modulate host responses[148]. After this early stage of transcription, the virion core "uncoats" and releases viral DNA into the cytoplasm. Replication proceeds in the cytoplasmic inclusion bodies ("virus factories"), and DNA synthesis is typically detected within 2 hours post-infection[147]. From here, the DNA progenies also serve as templates for subsequent intermediate gene

expression; during this process, the poxvirus utilizes host proteins and expresses enzymes involved in “DNA processing” (Holliday junction resolvase), and “DNA packaging” (ATPase, telomere-binding protein 1). In VACV, all intermediate and late mRNAs contain 5’-poly adenosine(A) leader sequences resulting from viral RNA polymerase slippage at the conserved promoter sequences (“TAAA”). The poly(A) sequences can have between 3-51 adenosines, and this feature is recently confirmed to confer translational advantage to poxviral transcripts[149,150]. Finally, the virus prepares for virion assembly by expressing the structural proteins in the last stage of its replication cycle. These viral late gene products accumulate for the assembly of IMVs, which are then trafficked through the Golgi membranes to form the IEVs (intracellular enveloped virus). IEVs later fuse with the cell membrane from which they either stay attached as CEV and protrude to the neighbouring cells by actin tail polymerization, or are released as free EEV particles[151].

1.4.3 Pathogenesis: virulence and host range

On the molecular level, the success of poxvirus infections in any host is determined *not* by the binding and entry of the virus, but by the completion of its replication[140]. Given that poxviruses attach to ubiquitous components on the cellular membrane (GAGs or laminin) instead of specific host receptors, it follows that any abortive replication must happen post-virion entry. There are cellular proteins that may restrict poxvirus replication, these include proteins that control: cell-cycle (e.g.: S-phase regulator), differentiation state (e.g.: cell lineage factors), protein folding (e.g.: heat shock protein 90), virion trafficking (e.g.: N-WASP), or signal transductions that induce antiviral responses (e.g.: interferons, protein kinases, STAT proteins, NF-κB)[140]. Therefore, in non-permissive hosts, replication halts because the poxvirus cannot circumvent certain checkpoints[152]. Throughout their life cycle, poxviruses rely on virulence proteins that modulate the intracellular and extracellular

environments against any antiviral defence triggers and responses. Virulence proteins can be characterized into 3 categories: virostealth, virotransduction, or the viromimicry of host cytokines and receptors[153,154]. Whereas "virostealth" describes the internal masking of viral infection of a cell[155] (e.g.: through the down regulation of MHC antigen presenting receptor genes[156]), "virotransduction" inhibits internal antiviral signals (e.g.: through the inactivation of apoptosis[157,158]). In contrast to these internal virulence factors, "virokines" and "viroceptors" are extracellular viral homologs that mimic host counterparts, and modulate extracellular responses by blocking communications usually in a competitive manner[154]. The virulence factors that are specifically associated with the virus's ability to replicate in a host are classically termed the "host range" proteins[159,160].

Poxvirus host range protein *in vitro* studies using cultured cell lines can differ markedly from *in vivo* conditions. For example, VARV only cause smallpox in human, but can replicate efficiently in most mammalian cell types *in vitro*[140]. Thus, host range studies typically cannot account for poxvirus pathogenesis, which is ultimately determined by tropisms at the cellular, tissue-specific, and organismal levels[140]. Consequently, the immune responses employed at the latter two levels are what determine the migration and dissemination of poxvirus within and between hosts that manifest the final disease outlook. Unfortunately, due to the complexity associated, understanding of tissue and organismal level tropism is still limited. On the cellular level, host range genes can usually be determined by mutagenesis or knock-out experiments. *In vitro* studies have shown that, aside from identifying host range genes through defective replication, rescue can also be made. For example, an insertion of an ankyrin-repeat gene (CHOhr) to VACV and mouse ECTV permits these viruses to replicate in what were previously non-permissive Chinese hamster ovary (CHO) cell-line[161,162]. The M-T5 gene encodes another ankyrin-repeat protein that plays a vital role in myxoma virus

replication, and is also characterized as a host range protein[163]. Other examples of host range proteins include the E3 and K3 proteins in VACV that deactivate protein kinase R (PKR) and, in turn, impede the induction of an antiviral state from host[164].

Realistically, host range determination is not merely the presence or absence of certain virulence genes. In addition to the various other components these virulence proteins may interact or associate with, it is likely that no poxvirus encodes the exact same version and combination as another. Therefore, the diversity and different combinations of virulence proteins in poxviruses account for the range of hosts poxviruses can infect as a family[117,165]. Consequently, each virulence factor reacts differently in different hosts, which in turn have specific immune responses and localization of their own. The poxvirus community today recognize difficulties in associating host infectivity with host range proteins, and appreciates the diversity and combinations of virulence proteins encoded by different poxviruses[117,165]. Thus, it remains crucial to annotate virulence genes encoded by different poxviruses. This allows researchers to map out viral pathogenesis as extensively as possible, and consequently prepare a catalogue of different pathways when faced with novel or re-emergence of zoonotic virus infections.

1.5 Bioinformatics

1.5.1 The sequencing era

The "Bioinformatics and Functional Genomics" textbook (3rd edition)[166] defines bioinformatics as "the science of managing and analyzing biological data using advanced computing techniques […] with the goal of revealing new insights and principles in biology". Most digital biological data originates from nucleotide sequences, which became accessible

in the 1980s with the growth of molecular technology. These genome nucleotide (nt) sequences serve as templates for RNA and protein sequences, and the range of biological data has since expanded to include whole genomes (DNA), transcriptome (RNA), and proteome (protein sequences and structures). Along with this is the development of computational tools and methods for data management (e.g.: sequence databases) and data retrieval (e.g.: sequence similarity searches, structure/function prediction tools)[167].

In the 1970s, Sanger sequencing allowed researchers to peek into the genomic blueprints of certain model organisms and set the foundations of comparative genomic analyses. However, classical and biomedical research have since been transformed by the next generation sequencing technology (NGS)[168], which accelerated the Human Genome Project in the 2000s. Today, 400 million sequences from whole genomic shotgun projects (WGS) are available at the National Center for Biotechnology Information (NCBI; this is 2800x more than 15 years ago). Metagenomics studies from organisms around the world overcame the problems of classical microbiology whereby 99% of the organisms are unculturable under laboratory conditions. For viruses, this also means that whole genomes can be easily sequenced from samples without growing the virus in cell cultures. Similarly, the amount of reads that can be sequenced from DNA has provided virologists with the ability to sequence and assemble the aforementioned ancient virus genomes from mummies and/or the permafrost layers (Ch. 1.3 and 1.1, respectively). Today, sequencing technology enters its third generation with the development of "long-read sequencers" such as PacBio or Oxford Nanopore technologies[169]. The decrease in costs, along with the introduction of bench-top and miniature sequencers, means improved accessibility for labs to conduct more sequencing projects. This increase in productivity (throughput) is accompanied by the production of more sequence reads (data) per project. As the sequencing technologies evolve, the field of bioinformatics is constantly

faced with the necessity to improve data management/analyses tools. To this end, efficient parsing of large amount of data in a timely manner as well as any new ways to explore/interpret information arise as some of the main challenges post-sequencing era (addressed in conclusion). Below, the bioinformatic analysis of poxviruses is described.

1.5.2 Poxviral bioinformatic analyses

1.5.2.1 Genome assembly

Complete genome assemblies serve as the basis of most comparative genomic works. Many assembly programs can generate a typical poxvirus genome (127-360kbp) from raw sequence data on a desktop computer. To that end, an assembler typically has to work with tens of millions of reads. However, contrary to common beliefs, more data or redundant coverage can sometimes impede productivity due to limitation on computing memory. Quality control protocols that remove redundant reads and contaminating sequences are thus applied to reduce raw read file sizes where needed.

There are two main types of genome assemblies: reference assembly and *de novo* assembly. The former assembly maps raw reads to a supplied reference genome, and is useful for assembly new virus strains against a previously published reference, whereas a *de novo* assembly is used for a novel virus to generate contiguous sequence(s) (contigs) without the bias of a template, and is often more computationally intensive. For both, the quality of a genome is manually validated by the sufficient and consistent read coverage (the number of reads mapped to a given position) across the assembled contig. Assemblers may fail to extend reads into ideal sized contigs when (1) adaptor sequences are still attached to the reads, or (2) coverage of reads dramatically increase or drops relative to the average of the contig. In such cases, manual removal of adaptors and manual extension of contigs may be

needed to improve assembly. To date, there are two prominent algorithms for assemblers: the overlapping-consensus (OLC) and the de Bruijn Graph (DBG) algorithms. In short, the OLC algorithm merges reads into extended contigs based on overlap sequences, whereas the DBG algorithm breaks reads into k-mers (sequences of k-unit size) and examines coverage statistics. The OLC algorithm was originally labour-intensive for NGS reads, whose runtime and computational complexity scaled drastically because of the multitudes of short reads produced, and was additionally unsuitable to assemble repeat regions (algorithm required to compare large data volume with a high chance of false positive overlaps due to short sequence lengths). The DBG algorithm, on the other hand, showed improved performance because it employed a different graph theory, and provided coverage statistics for error-correction. However, OLC-based assemblers have since incorporated the "string graph assemblers" and overcome the previous problems. Benchmarking of assembler programs have been difficult due to the variation of contigs generated. Recent research, while not praising one tool over the other, has predicted a comeback of OLC assemblers with the longer reads being produced by the third generation sequencers[170]. In short, algorithms continue to evolve with new demands, and this phenomenon demonstrates an example of the drive that grows the field of bioinformatics.

For poxviruses, genomes are often not extended to the hairpin termini due to the sudden drop in read coverage. It is thought that the dsDNA genome does not fragment properly at the hairpin terminal ends (stays covalently-joined), and consequently this region cannot be PCRed during the amplification step in sequencing[171]. Thus, poxvirus genomes are typically not assembled to the hairpin terminals. It is also possible that each termini fragments to different extent. However, because of ITR features of poxvirus genomes, terminal sequences

are inverted but identical, thus one end may be extended further based on the sequence of the other end to yield an equally extended genome.

Upon the completion of assembly, percentage of AT composition in the final genome may be calculated. The overall genome AT% is useful in grouping poxviruses (recall poxvirus classification by base composition in Chapter 1.4.1). Regions of AT% anomalies can also be detected using dot plots, whereby regions with irregular AT% generate different intensity of dots (nucleotide matches), and will appear as stripes against the overall genome dotplot. A dotplot experiment with the GC-rich MOCV identified two unusually AT-rich regions as "pathogenicity islands", that encode virulence genes derived from horizontal transfer events[172].

1.5.2.2 Poxvirus database

As of July 2017, there were 389 complete poxvirus genomes released on NCBI. The authors of the NCBI viral genomes project put it nicely: "as the number of viral records in the public sequence databases grows, retrieving a viral genomic sequence of interest with associated information is becoming increasingly complex. High redundancy in the databases is a common problem for all organisms; in the case of viruses, however, the large number of available strains, isolates, and mutants further exacerbates the problem"[173]. The Viral Orthologous Clusters (VOCs) database was developed to facilitate the management of dsDNA virus genomic data. It offers advantages for extracting poxviral data via organization of genes into orthologous clusters. The VOCs database itself has grown from 30 poxvirus sequences in 2004, to 114 genomes in 2012, to 361 genomes today. VOCs also easily generates data that shows that, currently, there are at least 740 orthologous gene families; 81 of these are conserved in all ChPVs, and more than 200 have unknown functions. (This

number of uncharacterized sequences is not unusual: of all the NCBI protein sequences, approximately 1 in 3 has no assigned function[174]). Identification of ORFs and annotation of genes in a poxvirus genome is important in that it serves as the basis for any subsequent comparative analyses. Orthologous sequences are used to create multiple sequence alignments (MSAs), which in turns are used to map out the molecular phylogenetic relationship between species. The pool of unique genes and hypothetical proteins (HPs) represents potential candidates for subsequent wet-lab experiments, as these may reveal novel important virulence or immunomodulatory roles.

1.5.2.3 Gene annotation

The putative annotation of a gene starts with the identification of ORFs. Some trends have been identified for *bona fide* poxviral ORFs and are used as guidelines when annotating genes. First, poxviral genomes are compacted with genes with little overlap. This means that there are few non-coding regions between genes, and that ORFs typically do not overlap regardless of the strand that the gene is encoded on. Second, most poxviral genes have conserved promoter motifs. Poxvirus promoters are typically found within 50 nucleotides upstream of the translational start site (position +1), and are composed of an upstream *core* region followed by a *spacer* region before the +1 *initiator* region. From VACV model studies, early poxviral promoters are found to be AT-rich with a core motif of "AAAxTxGAAAxxTA"[175]. In contrast, the intermediate and late promoters both have a "TAAA" initiator motif, and the late promoters are additionally followed by a T then a G[66]. Third, certain amino acid composition (such as low Asp/Glu and high Ser) and extreme isoelectric points (pI) have also been associated with ORFs thought not to be functional[176]. For AT-rich poxviruses, a "purine skew" has been found to be associated with ORFs on the coding strand of real genes[177]. Fourth, poxvirus genomes typically have same gene synteny in

their core region, and newly annotated ORFs should look to conserve this trend over annotating competing ORFs that may disrupt the conserved synteny.

The bioinformatic annotation of genes is usually based on the inferred homology from global alignment with previously annotated sequences, and/or the local similarity against characterized domains or motifs. For poxviruses, a gene that has sequence similarity with genes from another poxviral species is termed an ortholog (if they shared a vertical lineage), while a gene with similarity to an eukaryotic gene or an unusual poxvirus ortholog is termed a general homolog (a HGT origin). Because protein functions are determined by their structure, in general, a gene is selected on the amino acid sequence level during the course of evolution, thus the most common similarity search can be made by NCBI BLASTP search engine (Basic Local Alignment Search Tool); the result is a pairwise alignment with a percentage (%) identity score to be viewed in the context of sequence lengths, alignment coverage, and E-values.

At the current stage, no clear-cut % amino acid (aa) identity can be universally set to distinguish homologous sequences from non-homologous. When the % aa identity drops to below 20%, it corresponds to an average of 2.5 substitutions per site (accounting for multiple substitutions that had occurred at the same site), and a “twilight zone” is reached whereby the % identity score reaches an asymptote and barely changes with increasing genetic distances[166,178]. Therefore, % identity alone has its shortcomings when used as the sole indicator to establish homology. Instead, homology search results from tools such as BLASTP should be examined manually in the context of sequence lengths, alignment coverage, E-values, and the aa conserved. The former two paint a picture of global or local similarity, and provide information on potential sequence extensions or truncations. In

contrast, the E-value scores the probability (and thus the statistical significance) of the match occurring by chance in a BLAST database of a particular size; note that the shorter the sequence or the bigger the database, the higher the probability for random background hits. As reflected in the individual scoring matrices and BLAST scores, certain aa conservations are scored heavier than the others. For example, conservations of disulfide-bridge forming cysteines (a BLOSUM62 matrix score of 9) or strategically placed glycines (a score of 6; found in secondary turn and loop structures) can infer functional significance more so than the conservation of any small hydrophobic residues (Leu, Ile, Val, Ala all scored 4). So, unlike % aa identity (which scores the conservation of all amino acid the same weight), these type of scores would provide the statistically adjusted sum of all the substitution scores.

1.5.2.4 Phylogeny

Phylogeny is the inference of evolutionary relationship in the form of a tree that provides hypothesis on past biological events. Traditionally, phylogeny was based on morphological characterization of organisms; today, phylogenetic analysis uses molecular sequencing data to define the relationships between species or protein families. The simplest phylogenetic tree of the ChPVs can be used to capture the diversity and evolutionary relationships between the poxviruses that infect vertebrates. The phylogeny results can be additionally calibrated with epidemiological history and fossil records of host species to deduce virus origin and infer relationship with hosts, as in the hypotheses of VARV origin[52,179]. It can also be used to track evolutionary rates and distances of a virus species from samples across time, as in the cases of MYXV virulence evolution in Australia[20,23], or the analysis of ancient and modern VARV sample[112]. Phylogeny can also be used to demonstrate gene orthology. Ultimately, the accuracy of a phylogenetic tree relies on the input multiple sequence alignment (MSA), whose quality ultimately depends on the selection of sequences and software.

Poxvirus trees have been created based on criteria such as the presence/absence of gene families[17,122,180], gene order[180], or conserved sequences(s)[24,119,181], and overall produced consistent trees[49]. For the purpose of creating a ChPV tree that captures the wide diversity of the subfamily, input sequences made from concatenated conserved protein sequences (amino acid) may serve to be a better candidate for reasons listed below. The varied base composition between ChPVs (33-76% AT) can be too large to create a reliable nucleotide alignment used for phylogenetic tree. Protein sequences also capture a greater amount of permutation with 20 amino acids compared to 4 nucleotides in DNA, and evolve slower across diverged species, which better capture their homologous relationship. Amino acids evolve in a stepwise fashion, which may be modeled by substitution matrix models that can account for the number of intermediate substitutions a position has had to code for the extant residue[166,182]. Overall, these suggest that protein sequences are more phylogenetically informative than nucleotides for the purpose of creating a phylogenetic tree that captures the diversity of the ChPV family.

With that being said, DNA sequences can be more informative when working with closely related species within a genus, which allows the capture of subtle SNP changes in their genome evolution. In any case, it should be kept in mind that phylogenetic tree represents an *average* of the genetic differences regardless of the localization of these substitutions, as a protein could have different domains subjected to different mutational constraints. The Gamma rate of heterogeneity, which is a type of statistical model, can be applied to account for varied substitution rates across protein domains.

1.6 Research rationale and objectives

Here at the Upton Lab, we focus our research on poxviruses as well as other large DNA viruses. In parallel, our Viral Bioinformatics Resource Centre (VBRC) develops and promotes bioinformatics tools for various viral research applications. This consequently generates several collaborative opportunities, and veterinarian researchers and virologists (including those from CDC Atlanta) have approached us with sets of raw read files of what could be potential novel poxviruses. My work was set out to establish a working protocol that accurately assembles full-length poxvirus genomes from the raw sequence data, and subsequently characterize the virus on the genomic and phylogenetic levels using various bioinformatics tools. The animal sources of these poxvirus extractions include, from southern Australia: a megabat (*Pteropus scapulatus)*, a western grey kangaroo (*Macropus fuliginosus*) and an eastern grey kangaroo (*Macropus giganteus*), and from north America: a microbat (*Eptesicus fuscus*) and a sea otter (*Enhydra lutris*). Full poxvirus genomes have not been assembled from these sources before. In addition to their veterinary importance, the discoveries of novel poxviruses are of vital interest because the analysis of their genomes and genes will elucidate more branches on the poxviral phylogenetic tree, as well as expand the repertoire of poxviral virulence genes.

The sequence assembly and annotation generated from my work provides the necessary data for subsequent comparative genomic analyses. Novel unique genes, especially those with putative functions, expand the repertoire of poxvirus proteins, broaden mechanisms of viral pathogenesis, and should prompt for further experimental studies. If needed, the genomic characterization of these poxviruses could aid in the downstream development of veterinary diagnostics tools and/or epidemiological studies.

Chapter 2. MATERIALS AND METHODS

A summary of the protocol employed for this thesis is seen in **Figure 2**.

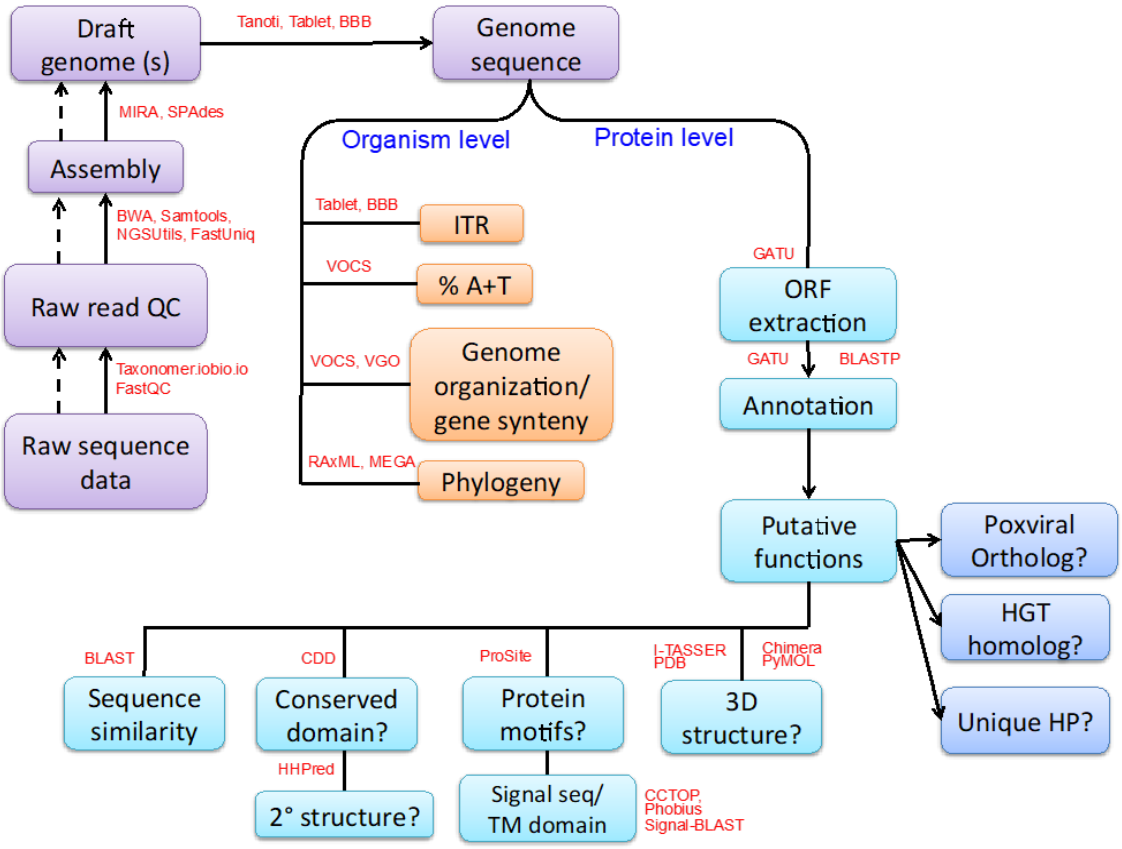

Figure 2. Workflow for the assembly and annotation of novel poxvirus genomes.

Protocol is divided into genome assembly (purple), analysis on the organism level (orange), and analysis on the protein level (blue), followed by putative annotation of genes (dark blue); bioinformatics tools utilized for each step are indicated in red.

For poxvirus-specific bioinformatic analyses, extensive use was made of the Viral Bioinformatics Resource Centre (https://virology.uvic.ca/). The following tools were used: Viral Orthologous Clusters (VOCs) database for sequence management[183,184]; Genome Annotation Transfer Utility (GATU) for annotation[185]; JDotter for creating a genome dotplot[186]; Base-By-Base (BBB) for editing genome/gene/protein alignments[187]; and Viral Genome Organizer (VGO) for genome organization comparisons[188].

2.1 Genome assembly

In the cases of Pteropox virus (PTPV), Eptesipox virus (EPTV), western and eastern Kangaroopox viruses (WKPV, EKPV), our collaborators identified, extracted, sequenced novel poxviruses, and supplied the raw sequence data that I assembled, annotated, and analyzed. Sea otterpox virus (SOPV) was the exception, where I assisted with the annotation of the genome straight from a fully assembled contig. Overviews of the veterinary background and wet-lab efforts written by our collaborators are provided in individual virus chapters under the "background section" for context purposes.

Current DNA extraction protocols and sequencing technologies are optimized such that, given quality samples, the sequencing of a single genome can create large datasets that can counter-intuitively impede productivity. Some of the raw datasets in this thesis range from 8.75GB to 15.27GB per single paired-end file (from PTPV and EKPV respectively), and consequently result in failed assemblies due to insufficient memory space on a typical computing station (the thesis work was performed on a Mac with intel Core i5 and 16 gigabytes of memory). These two particular datasets were eventually assembled following reduction in data size (to 40 megabytes and 3 gigabytes, respectively) using the quality control protocols below.

2.1.1 Quality control

Taxonomer, a metagenomics tool that characterizes sequencing reads to different taxonomic categories (http://taxonomer.iobio.io/)[189], was used to identify non-poxviral contaminants in the raw sequence files. Corresponding reference genome(s)/scaffold(s) of the major contaminant source(s) is downloaded from the NCBI database. The Burrows Wheeler

Aligner (BWA)[190] was then used to index the contaminant reference sequence into a database, and subsequently mapped to the sample raw reads using the "BWA-MEM" algorithm; reads unmapped to contaminant sequences were extracted for assembly using SAMtools[191], NGSUtils[192], and various bash scripts (see **Appendix 1**).

Sequential data reduction steps were taken, as needed, by repeating additional contaminant removal, shortening read names (as individual read names can sometimes exceed 40 characters), and/or removing duplicated reads using FastUniq[193]. The FastQC program (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and SAMtools' "flagstats" were used to generate quick overviews of raw read datasets pre- and post-filtering.

2.1.2 Assembly and validation

Filtered reads were inputted into the SPAdes assembler[194] and/or the MIRA assembler[195] to generate contiguous sequences or contigs. Contig(s) were examined in BBB sequence editor[187], and may be compared/joined/extrapolated as needed, to generate a preliminary fully extended genome. From here, Tanoti (http://www.bioinformatics.cvr.ac.uk/tanoti.php), a BLAST-guided reference based short read aligner, was used to map raw reads to the preliminary genome, and a SAM (sequence alignment map) file was created for visualization downstream. SAMtools[191] was used to convert file formats as needed. The Tablet software[196] was used to visualize raw reads mapped to assembled genome contig, examine coverage, and make manual base-calls to adjust/correct the preliminary genome into the final genome.

The quality of the final assembled genome was checked against poxvirus references through generation of dotplots using JDotter[186]. The query sequence and reference sequence are

placed on the x- or y-axis, and a dot is placed on a coordinate if the residues at x and y positions are identical. This effectively generates a dotplot that displays regions of similarity (a continuous diagonal line), indels (disjointed diagonal lines), and different types of repeats and rearrangements (various lines appearing in different directions) to be viewed in one glance. Self-plots were also effective for detecting regions of incongruent base compositions (different density of random background matches), which may be indicative of HGT regions[172].

2.2 Genome annotation

2.2.1 ORF identification

ORFs from the assembled genome were identified using the Genome Annotation Transfer Utility (GATU)[185], and annotated with its closest BLASTN result virus as the reference genome. Although GATU automated the annotation process between closely related species, for novel genomes (which can be quite diverged from other genomes), we utilized it more for the ORF identification ability and evaluated most of the decisions manually. Initially, ORFs greater than 50 codons with no more than 25% overlap with neighbouring genes were selected and annotated. Subsequently, smaller ORFs (between 30-50 codons) were examined and only annotated provided that a previously annotated poxvirus ortholog existed and/or if a poxvirus promoter-like motif was present immediately 5' to the ORF.

2.2.2 BLASTP searches

The extracted ORFs were placed into a FASTA file and inputted into a BLASTP search against databases using relaxed parameters; a word-size of 2 is used to improve the sensitivity of our searches. To offset the time consumption caused by the decrease in word-

size, preliminary sequence searches were limited to the *Poxviridae* family (taxid: 10240 and 40069) on NCBI or the locally downloaded VOCs poxvirus gene sequences (performed batch search and was much faster). The search was expanded to the entire non-redundant (nr) database when no significant hits were found (i.e. no poxviral orthologs were found). Any BLASP pairwise alignments with less than 40% aa identity or less than 90% coverage are examined further in terms of sequence lengths, sequence alignment, coverage, E-values, and the aa conserved. Additional tools are utilized to explore potential function of any diverged sequences.

2.2.3 Additional tools and resources

For the ORFs suspected to have undergone frame-shift, extensions/truncations, or fragmentation, BLASTX were performed. This uses the nucleotide region and translates it into the 3 different reading frames for detection of any partial genes. The Viral Genome Organizer (VGO) was also used to facilitate this process by displaying start and stop codon positions between homologous sequences, along with many other features that enabled a holistic survey of the local environment of each gene[188]. VGO, which had access to all the curated information from the VOCs database, enabled graphical comparisons and interactions of multiple genomes without the difficulty of sequence alignment or introductions of gaps. Genes are represented by size-proportional coloured blocks that can be dragged/aligned, and with desired orthologs highlighted across species. In one glance, VGO allowed comparison through the graphical displays of gene synteny, intergenic promoter spaces, a base compositional graph, distribution of start/stop codons, or un-annotated ORFs. Regions of interest were selected, and the DNA or aa sequences were extracted for further analyses.

The primary sequences of proteins have telling features such as domains or motif sequences that can be quickly scanned against curated databases and shed light onto its potential functions. The Conserved Domain Database (CDD) utilizes the reverse position-specific (RPS)-BLAST program, a variant of position-specific iterative (PSI)-BLAST, which scans the query sequence against position-specific scoring matrices (PSSMs) of annotated conserved protein domains[197]. These PSSMs have adjusted scores for conserved residues specific to each protein profiles (built from multiple alignments), and were a useful supplement to hone in and confirm BLASTP alignments with low % identity scores. As needed, protein sequences were additionally searched against known motifs collected in the PROSITE database using ScanProsite[198]. The search was quick and scanned through 1700+ documented entries and returned with curated profile hits linked to more information. Additional comprehensive but time-intensive search was at times conducted using InterProScan[199], which probed between multiple resources including the CDD and PROSITE databases.

For ORFs that failed to match homologs on the primary sequence level, sequences were searched with the HHPred program. HHPred is a sensitive tool that allows for remote homolog detection by comparing the Hidden Markov Models (HMMs) profiles between query and protein families, and offers insights by scoring the secondary structures between proteins[200,201]. In our experience, we have found the HHPred probability score to be very reliable. From here, any biological features were also taken into consideration for the validity of assigned annotation. Identification of known features such as similar gene synteny (using VOCs database and VGO), or congruence with known protein topology (literature search), transmembrane domains (using CCTOP server[202] or Phobius[203]), and/or signal peptides

(using Signal-BLAST[204]) can add weight to the annotation.

Lastly, since tertiary structure of a protein evolves the slowest, structural alignment and similarity may be useful to detect "twilight zone" homologs. Confirmation of interesting features recognized in a novel sequences were taken a step further via 3D structure modeling with I-TASSER[205]. The generated structures were subsequently visualized and manipulated in either PyMOL[206] or UCSF Chimera[207]. The "Tools > Sequence > Match->Align" function of UCSF Chimera was used to generate the structure-based sequence alignment. At this stage, any novel, unique ORFs lacking obvious homologs or known features were annotated as hypothetical proteins (HPs), recognizing that some may be non-functional and/or non protein-encoding until experimentally proven.

2.3 Phylogenetic trees

The sequences used for phylogenetic analysis of ChPVs were retrieved from the VOCs database[183,184]. DNA sequences were aligned using the MAFFT algorithm[208], and amino acid sequences were aligned using MUSCLE algorithm[209]. In most cases, gaps and poorly aligned regions are removed using the trimAl program[210]. This step reduces poorly aligned regions and potential artefacts, which can cause long branch attraction (an observed limitation of phylogeny algorithm, where by a highly divergent sequences creates a false positive convergence of branches)[211,212]. The constructed MSAs were manually checked using BBB[187]. The viruses mentioned in this thesis are displayed in **Table 1**.

Table 1. A list of poxviruses mentioned in this thesis

List is sorted by ascending AT%, and the intensity of red and blue proportionally represent the GC or AT richness of the genome, respectively. A reduced subset of species used in phylogenetic tree is marked by asterisk.

Rep?	Abbreviation	Virus	Strain	Genus	Accession number	A+T%
*	SQPV	Squirrelpox virus	Red squirrel UK	Unclassified	NC_022563	33.31
*	PCPV	Pseudocowpox virus	VR634	Parapoxvirus	NC_013804	35.01
*	PPV	Parapoxvirus red deer	HL953	Parapoxvirus	NC_025963	35.04
*	BPSV	Bovine papular stomatitis virus	AR02	Parapoxvirus	NC_005337	35.50
*	ORFV	Orf virus	SA00	Parapoxvirus	NC_005336	36.56
*	MOCV	Molluscum contagiosum virus	subtype 1	Molluscipoxvirus	NC_001731	36.64
*	CRV	Crocodilepox virus	Zimbabwe	Crocodylidpoxvirus	NC_008030	38.08
*	SePPV	Seal parapoxvirus	AFK76s1	Parapoxvirus	NC_035188	44.08
*	EKPV	Eastern grey kangaroopox virus	Sunshine Coast	Unclassified (Thylacopoxvirus)	MF467281	46.06
*	WKPV	Western grey kangaroopox virus	Western Australia	Unclassified (Thylacopoxvirus)	MF467280	46.10
*	MYXV	Myxoma virus	Lausanne	Leporipoxvirus	NC_001132	56.44
*	RFV	Rabbit fibroma virus	Kasza	Leporipoxvirus	NC_001266	60.47
	SGPV	Salmon gill poxvirus	2012-04-F277-L3G	Unclassified	NC_027707	62.45
*	PTPV	Pteropox virus	Australia	Unclassified	NC_030656	66.24
*	CPXV	Cowpox virus	Brighton Red	Orthopoxvirus	NC_003663	66.60
*	VACV	Vaccinia virus	Western Reserve	Orthopoxvirus	NC_006998	66.66
	TATV	Taterpox virus	Dahomey 1968	Orthopoxvirus	NC_008291	66.74
	ECTV	Ectromelia virus	Moscow	Orthopoxvirus	NC_004105	66.82
	CMLV	Camelpox virus	M96	Orthopoxvirus	NC_003391	66.83
	MPXV	Monkeypox virus	Zaire	Orthopoxvirus	NC_003310	66.91
*	VARV	Variola major virus	India 3 1967	Orthopoxvirus	NC_001611	67.27
*	RCNV	Raccoonpox virus	Herman	Orthopoxvirus	NC_027213	67.67
	SKPV	Skunkpox virus	WA	Orthopoxvirus	NC_031038	68.50
*	SOPV	Sea otterpox virus	ELK	Unclassified	N/A	68.72
*	VPXV	Volepox virus	CA	Orthopoxvirus	NC_031033	68.73
*	FWPV	Fowlpox virus	Iowa	Avipoxvirus	NC_002188	69.11
*	CNPV	Canarypox virus	VR111	Avipoxvirus	NC_005309	69.63
*	YMTV	Yaba monkey tumor virus	Amano	Yatapoxvirus	NC_005179	70.17
*	Murmansk	Murmansk poxvirus	LEIV-11411	Centapoxvirus	NC_035468	70.21
*	TKPV	Turkeypox virus	HU1124/2011	Avipoxvirus	NC_028238	70.22
	NY14	NY014 poxvirus	2013	Centapoxvirus	NC_035469	70.48
	PNPV	Pigeonpox virus	FeP2	Avipoxvirus	NC_024447	70.48
*	PEPV	Penguinpox virus	PSan92	Avipoxvirus	NC_024446	70.50
*	SWPV	Swinepox virus	Nebraska	Suipoxvirus	NC_003389	72.60
*	YLDV	Yaba-like disease virus	Davis	Yatapoxvirus	NC_002642	73.00
*	DPV	Deerpox virus	W-848-83	Cervidpoxvirus	NC_006966	73.84
	LSDV	Lumpy skin disease virus	Neethling 2490	Capripoxvirus	NC_003027	74.09
*	YKV	Yoka poxvirus	DakArB 4268	Centapoxvirus	NC_015960	74.42
*	GTPV	Goatpox virus	Pellor	Capripoxvirus	NC_004003	74.69
	SPPV	Sheeppox virus	TU	Capripoxvirus	NC_004002	74.99
*	COTV	Cotia virus	SPAn232	Unclassified	NC_016924	76.41
*	EPTV	Eptesipox virus	Washington	Unclassified (Chiropoxvirus)	NC_035460	76.44
	MSEV	Melanoplus sanguinipes entomopoxvirus	Tucson	Unassigned entomopoxvirus	NC_001993	81.73
	AMEV	Amsacta moorei entomopoxvirus	Moyer	Betaentomopoxvirus	NC_002520	82.22

As more species get included on a tree, the number of ways to arrange the branches increases substantially. The conservative method would be to survey all the possible tree arrangements. The maximum-likelihood (ML) method is computationally extensive, but produce stringent results having performed exhaustive search of all possible trees. RAxML v8.2.10[213] was used to generate a maximum-likelihood estimation of the base frequency, under the LG substitution model[182] and Gamma model of rate heterogeneity. The argument

"PROTGAMMAAUTO" selects the LG substitution model for our protein dataset as the one with the highest likelihood score on the parsimonious starting tree (out of 19 other models). The LG substitution model then computes the substitution probabilities that have occurred along each phylogeny branch, and predicts the tree with the highest likelihood of occurrence. Additionally, 1000 bootstrap replications are performed to evaluate the robustness of the positions of each branch. The resultant Newick file was subsequently visualized in MEGA6.06[214].

Chapter 3. RESULTS AND DISCUSSION

3.1 Phylogenetic tree of Chordopoxviruses

3.1.1 Abstract

The objective of this chapter is to update the ChPV phylogenetic tree with 5 novel poxviruses. Originally, the ChPV tree was made using the amino acid sequences from a set of 7 conserved core genes due to limited sequence availability at the time. Its topology and bootstrap values are later evaluated against that of the phylogenetic tree created using all 81 conserved genes. The result demonstrates that a phylogenetic tree made with these 7 conserved genes accurately reflects the tree topology made with 81 conserved genes. Importantly, it reproduces with high bootstrap values, and can be generated much faster. From the trees, we conclude that 4 novel genera should be established for the 5 novel poxviruses; their phylogenetic relationship is elucidated and discussed. Additionally, it appears that if the current genus classification was applied to all poxviruses, there would be a need to restructure the *Parapoxvirus* and *Avipoxvirus* genera.

3.1.2 Results

3.1.2.1 Concatenation of 7 and 81 gene sequences from 32 ChPV representative species

At the time of the South Australian Pteropox virus (PTPV) research in 2016, the only available bat-isolated poxvirus sequences were 7 genes from a partially sequenced North American bat-isolated poxvirus (EPTV; Eptesipox virus), and one partial P4b precursor gene from West African Eidolon helvum poxvirus 1 (AGL97805.2). It then made sense to compare the maximum available genes between PTPV and EPTV by creating a phylogenetic tree using these 7 genes: RNA polymerase 147kDa subunit (RPO147) (J6R in reference strain

VACV-Copenhagen), RNA polymerase-associated protein RAP94 (H4L), mRNA capping enzyme large subunit (D1R), P4a precursor (A10L), RPO132 (A24R), viral early transcription factor (VETF) large subunit (A7L), and DNA primase (D5R). To get a sense of the range of gene divergence captured by these 7 genes, **Figure 3** shows the % aa and % nt identity of these 7 EPTV genes when compared to the VACV-Copenhagen reference strain. This set of Eptesipox genes covers a significant range of gene variation, and shows % aa conservation ranging from 54% (between the A10L orthologs) to up to 83% (between the RPO147 orthologs).

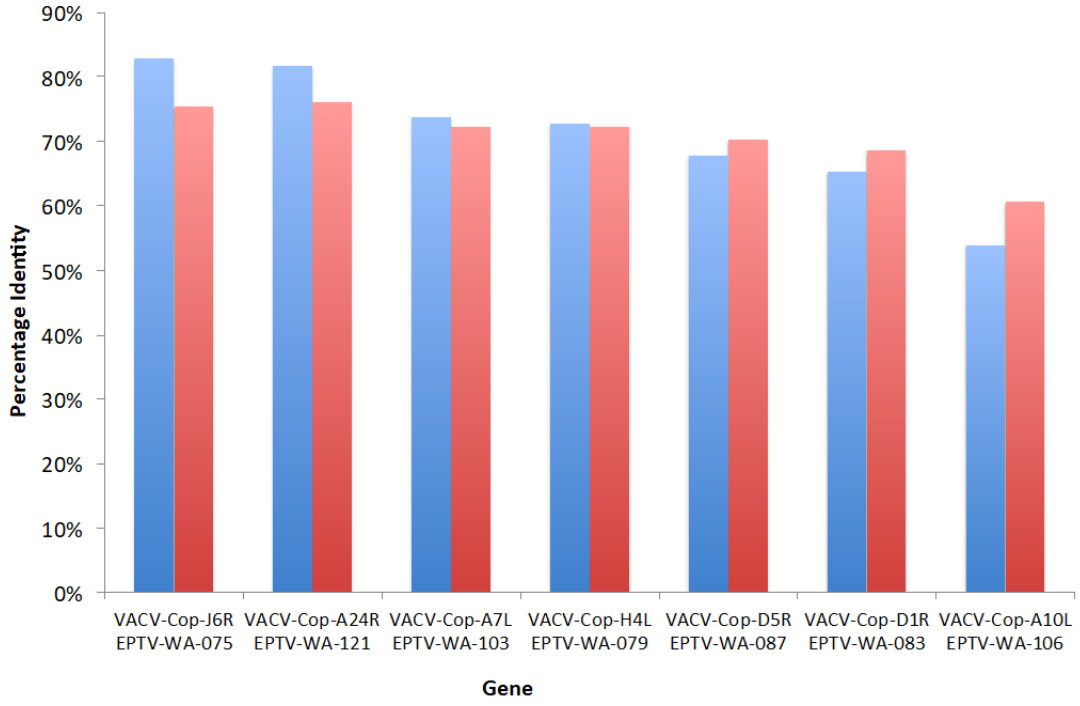

Figure 3. The degree of conservation between the set of 7 EPTV genes with VACV-Cop orthologs.

Amino acid and nucleotide % identities are displayed as blue and red bars, respectively.

Consequently, a maximum-likelihood (ML) phylogenetic tree for ChPVs was made from the concatenation of the separate alignments made with these 7 genes (**Figure 4**). For practical reasons, closely related genomes (such as those from the OPVs) that add no additional information to the placement of the new poxviruses were removed. Although, intuitively, it

may seem like a good idea to increase the data size by sampling from more viruses, in practice, however, large datasets can take extra time and memory to process, which impede productivity. The reduction to 32 ChPV representatives does not change the overall topology from a tree made with all 44 poxvirus species (data not shown). The final alignment, with gaps removed, is 6,430 nt long.

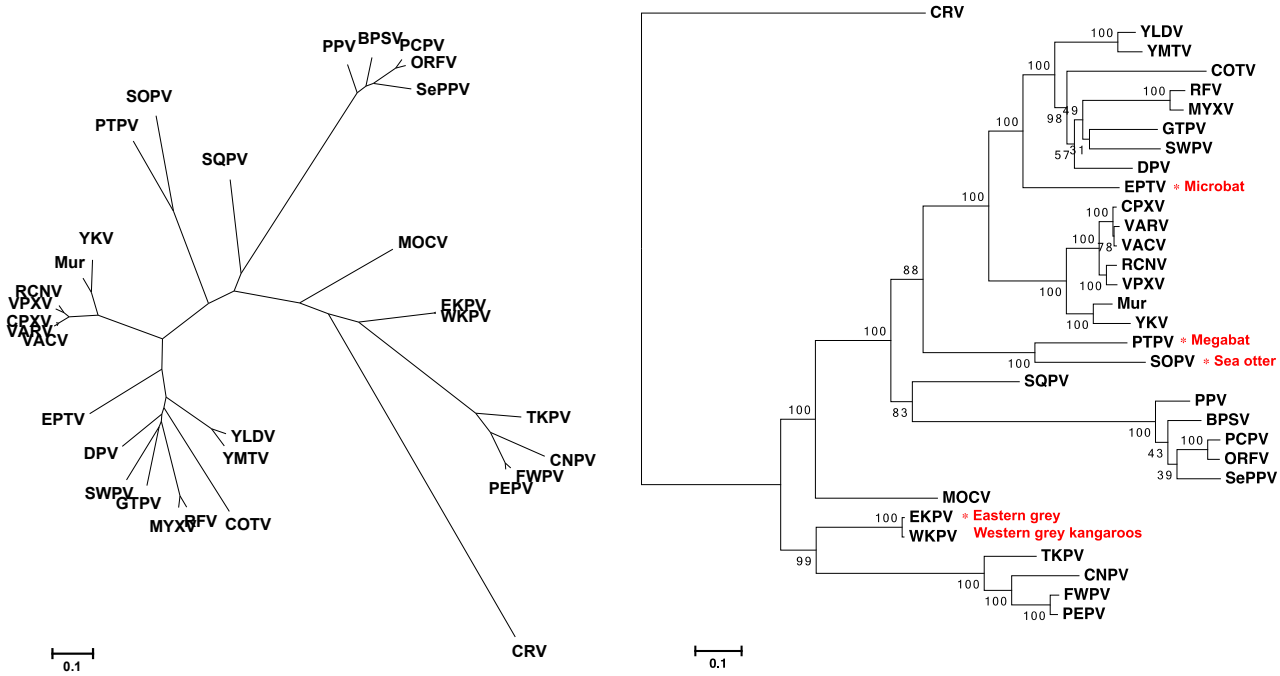

Figure 4. Phylogenetic trees of ChPV representative species using the amino acid MSA from concatenated sequences of 7 genes.

The 7 genes are RPO147 (J6R), RNA polymerase-associated protein RAP94 (H4L), mRNA capping enzyme large subunit (D1R), P4a precursor (A10L), RPO132 (A24R), viral early transcription factor (VETF) large subunit (A7L), and DNA primase (D5R). Scale bar represent the number of amino acid substitution per site.

When trees were created individually for each of these 7 genes, topologies only varied slightly at some branches with lower bootstrap values. It has been observed that different single-gene phylogeny can create varied topologies among the same set of OPV viruses[215,216] due to conflicting signals[217]. The outgroup CRV has a very long branch length, exceeding 1 amino acid substitution per site from the other ChPVs. Note that this is an *average* of substitution across the length of the alignment, and does not mean that every position of the sequence is changed. The LG model that is applied to our maximum-likelihood method

accounts for the multiple substitutions that have occurred at each site. Given the availability of up to 81 conserved ChPV genes after the full genome assembly of the EPTV (see subsequent chapter), a tree was made with all 81 conserved core genes from the same 32 representative ChPV species, and the topology for our new viruses were the same (**Figure 5**).

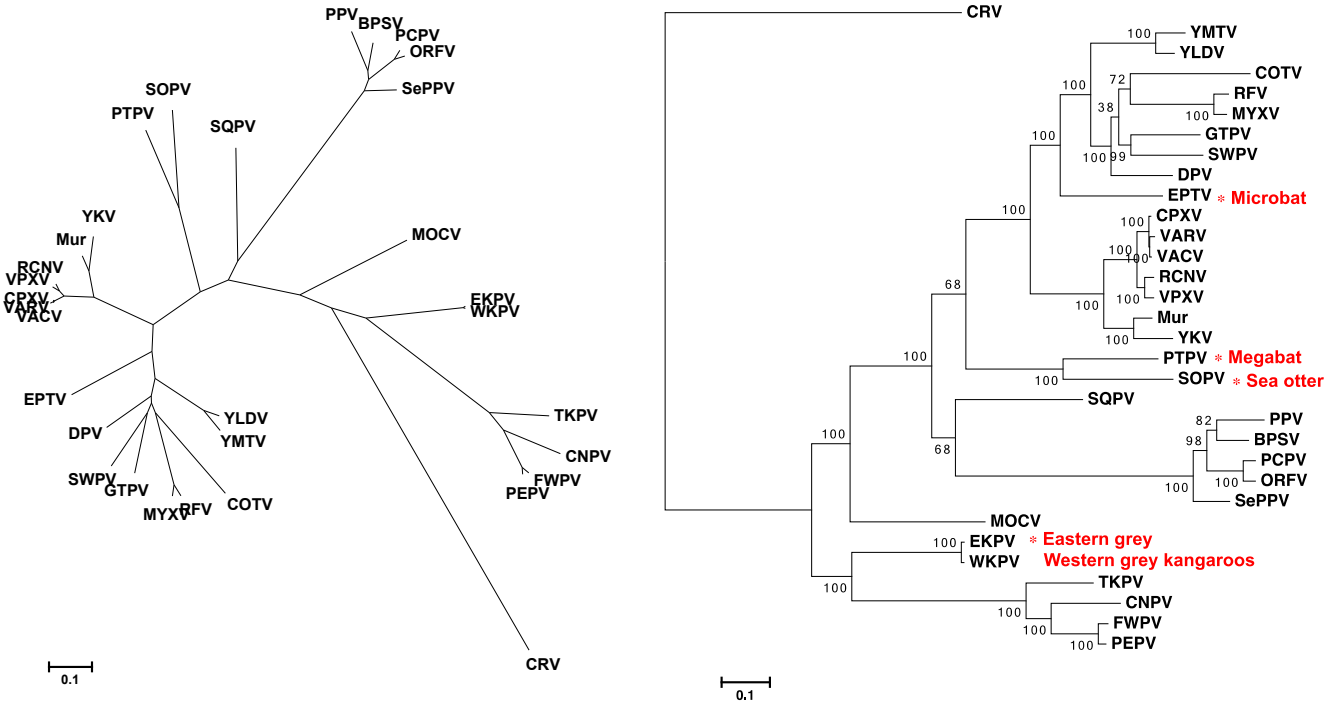

Figure 5. Phylogenetic trees of ChPV representative species using the amino acid MSA from concatenated sequences of all 81 conserved ChPV genes.

Scale bar represent the number of amino acid substitution per site.

3.1.2.2 Relationship between input file and runtime

Finally, **Figure 6** summarizes the total runtime for completing different phylogenetic analyses from the various trial runs with different starting files (trees not shown). The runtime recorded in this experiment is not equivalent to the actual elapsed time, but sum of the actual time the CPU (central processing unit) spent in the kernel and user-mode. This sum of CPU time is used for benchmarking purposes because it disregards the time spent on other processes that the user may be running simultaneously. Because our commands were executed with 4 threads of execution (i.e. the computation is performed in parallel between 4

separate instances of the running program), in our case, it seems like the actual elapsed time can be estimated to be roughly ¼ of the sum CPU time.

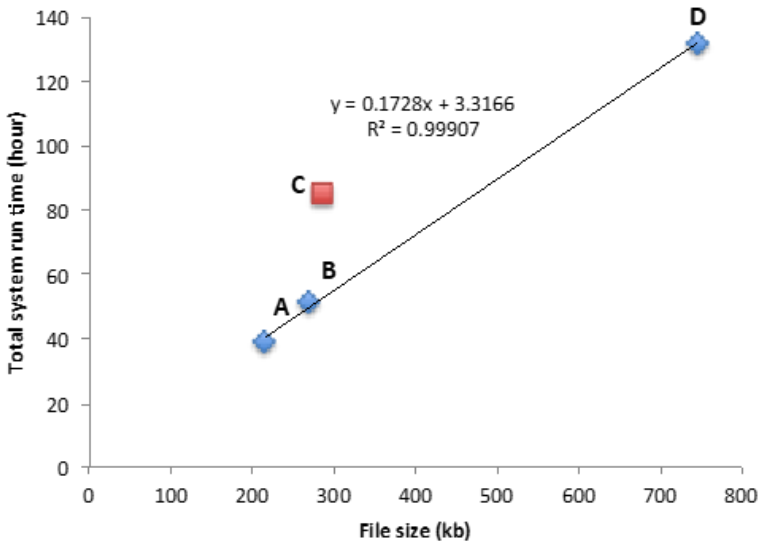

Figure 6. Relationship between different phylogeny input files and the resultant CPU run time per analysis.

Data points are made from phylogeny trials using different input files; A= 7 genes from 32 ChPV representative species; B = 7 genes from all 41 ChPV species; C= 7 genes from all 41 ChPV and 2 entomopoxvirus species; D = 81 genes from 32 ChPV representative species. A linear increase in time against file size is seen, where point C (data set that includes divergent species) is the outlier.

3.1.2.3 Distance and % identity matrix

In this section, the threshold for current genus classification is examined using the resultant distance matrix from the comprehensive 81 conserved genes phylogenetic tree. The scale bar of the phylogenetic tree is based on the average amino acid substitution per site across the alignment file, and the distance between the taxonomic nodes (virus species) is recorded on the matrix. The current ChPV classification include genus with intra-genus (within a genus) distances as large as 0.309 (seen in APVs). Here we present and examine the distance matrix of representative ChPVs made with 81 concatenated genes with: (1) the current genus borders boxed, and (2) values highlighted in different ways based on the different genus threshold values applied.

Table 2. A distance matrix between ChPV species with values highlighted based on current genus classification

Matrix created from the maximum-likelihood phylogenetic tree made with the concatenation of 81 conserved ChPV genes from 32 representative ChPV species; distance value is the average amino acid substitution per site under the LG substitution model; current genus classification is boxed in thick borders; values within the current intra-genus distance threshold (0.309 between TKPV and CNPV) are

Genus	Virus	CRV	YLDV	YMTV	COTV	RFV	MYXV	GTPV	SWPV	DPV	EPTV	VARV	VACV	CPXV	VPXV	RCNV	YKV	Mur	PTPV	SOPV	SQPV	BPSV	PPV	ORFV	PCPV	SePPV	MOCV	EKPV	WKPV	TKPV	CNPV	FWPV	PEPV
Crocodylidpox	CRV		1.322	1.328	1.384	1.326	1.308	1.343	1.341	1.329	1.327	1.308	1.302	1.302	1.311	1.313	1.341	1.312	1.346	1.387	1.249	1.420	1.426	1.416	1.422	1.423	1.125	1.122	1.123	1.285	1.282	1.282	1.278
Clade II	YLDV	1.322		0.094	0.404	0.362	0.360	0.314	0.330	0.293	0.389	0.455	0.450	0.451	0.450	0.455	0.470	0.455	0.666	0.700	0.644	0.926	0.951	0.941	0.939	0.916	0.772	0.828	0.827	0.911	0.914	0.906	0.901
	YMTV	1.328	0.094		0.418	0.377	0.374	0.330	0.349	0.310	0.410	0.470	0.465	0.466	0.465	0.468	0.489	0.475	0.673	0.705	0.651	0.926	0.954	0.942	0.942	0.919	0.778	0.836	0.836	0.924	0.926	0.913	0.911
	COTV	1.384	0.404	0.418		0.411	0.412	0.369	0.375	0.351	0.473	0.552	0.547	0.546	0.546	0.549	0.553	0.543	0.723	0.765	0.729	0.995	1.024	1.007	1.012	0.982	0.868	0.903	0.901	0.948	0.962	0.948	0.946
	RFV	1.326	0.362	0.377	0.411		0.053	0.327	0.327	0.309	0.437	0.503	0.497	0.495	0.501	0.500	0.524	0.505	0.687	0.728	0.662	0.923	0.944	0.935	0.934	0.914	0.784	0.844	0.846	0.943	0.955	0.940	0.938
	MYXV	1.308	0.360	0.374	0.412	0.053		0.323	0.323	0.305	0.435	0.494	0.488	0.487	0.493	0.492	0.517	0.497	0.680	0.723	0.652	0.903	0.927	0.917	0.915	0.901	0.771	0.836	0.837	0.937	0.942	0.931	0.930
	GTPV	1.343	0.314	0.330	0.369	0.327	0.323		0.272	0.261	0.401	0.479	0.475	0.475	0.476	0.475	0.488	0.474	0.668	0.714	0.666	0.951	0.977	0.962	0.964	0.939	0.802	0.859	0.859	0.924	0.936	0.920	0.919
	SWPV	1.341	0.330	0.349	0.375	0.327	0.323	0.272		0.261	0.413	0.481	0.476	0.475	0.472	0.473	0.489	0.476	0.674	0.715	0.661	0.931	0.960	0.944	0.947	0.927	0.790	0.838	0.837	0.909	0.914	0.903	0.902
	DPV	1.329	0.293	0.310	0.351	0.309	0.305	0.261	0.261		0.377	0.450	0.444	0.444	0.444	0.444	0.460	0.447	0.655	0.694	0.636	0.921	0.949	0.936	0.934	0.917	0.767	0.830	0.831	0.910	0.917	0.907	0.905
Unclassified	EPTV	1.327	0.389	0.410	0.473	0.437	0.435	0.401	0.413	0.377		0.450	0.446	0.446	0.446	0.448	0.473	0.453	0.657	0.702	0.645	0.917	0.933	0.919	0.919	0.905	0.783	0.847	0.847	0.908	0.911	0.905	0.901
Orthopox	VARV	1.308	0.455	0.470	0.552	0.503	0.494	0.479	0.481	0.450	0.450		0.013	0.015	0.063	0.063	0.218	0.181	0.676	0.706	0.616	0.887	0.905	0.898	0.897	0.881	0.757	0.832	0.830	0.935	0.935	0.923	0.921
	VACV	1.302	0.450	0.465	0.547	0.497	0.488	0.475	0.476	0.444	0.446	0.013		***0.009***	0.058	0.058	0.214	0.176	0.670	0.702	0.610	0.878	0.895	0.888	0.887	0.873	0.751	0.826	0.824	0.932	0.932	0.919	0.917
	CPXV	1.302	0.451	0.466	0.546	0.495	0.487	0.475	0.475	0.444	0.446	0.015	***0.009***		0.057	0.057	0.213	0.175	0.671	0.703	0.609	0.878	0.896	0.887	0.887	0.872	0.752	0.825	0.823	0.933	0.933	0.919	0.916
	VPXV	1.311	0.450	0.465	0.546	0.501	0.493	0.476	0.472	0.444	0.446	0.063	0.058	0.057		0.036	0.211	0.176	0.673	0.705	0.614	0.887	0.905	0.890	0.891	0.879	0.756	0.826	0.825	0.933	0.934	0.923	0.921
	RCNV	1.313	0.455	0.468	0.549	0.500	0.492	0.475	0.473	0.444	0.448	0.063	0.058	0.057	0.036		0.211	0.177	0.674	0.703	0.614	0.882	0.901	0.894	0.894	0.879	0.755	0.828	0.826	0.935	0.940	0.923	0.922
Centapox	YKV	1.341	0.470	0.489	0.553	0.524	0.517	0.488	0.489	0.460	0.473	0.218	0.214	0.213	0.211	0.211		0.113	0.689	0.721	0.659	0.922	0.948	0.933	0.934	0.916	0.796	0.858	0.857	0.935	0.943	0.928	0.927
	Mur	1.312	0.455	0.475	0.543	0.505	0.497	0.474	0.476	0.447	0.453	0.181	0.176	0.175	0.176	0.177	0.113		0.673	0.705	0.620	0.888	0.908	0.899	0.901	0.883	0.758	0.831	0.830	0.923	0.927	0.912	0.910
Unclassified	PTPV	1.346	0.666	0.673	0.723	0.687	0.680	0.668	0.674	0.655	0.657	0.676	0.670	0.671	0.673	0.674	0.689	0.673		0.403	0.670	0.839	0.870	0.852	0.853	0.840	0.800	0.873	0.875	0.976	0.980	0.967	0.968
Unclassified	SOPV	1.387	0.700	0.705	0.765	0.728	0.723	0.714	0.715	0.694	0.702	0.706	0.702	0.703	0.705	0.703	0.721	0.705	0.403		0.686	0.871	0.897	0.884	0.888	0.865	0.816	0.904	0.906	1.004	1.009	0.998	0.995
Unclassified	SQPV	1.249	0.644	0.651	0.729	0.662	0.652	0.666	0.661	0.636	0.645	0.616	0.610	0.609	0.614	0.614	0.659	0.620	0.670	0.686		0.748	0.759	0.756	0.755	0.752	0.608	0.746	0.747	0.989	0.983	0.966	0.965
Parapox	BPSV	1.420	0.926	0.926	0.995	0.923	0.903	0.951	0.931	0.921	0.917	0.887	0.878	0.878	0.887	0.882	0.922	0.888	0.839	0.871	0.748		0.154	0.163	0.161	0.161	0.864	0.985	0.986	1.217	1.201	1.189	1.182
	PPV	1.426	0.951	0.954	1.024	0.944	0.927	0.977	0.960	0.949	0.933	0.905	0.895	0.896	0.905	0.901	0.948	0.908	0.870	0.897	0.759	0.154		0.181	0.180	0.192	0.875	1.002	1.004	1.251	1.232	1.224	1.220
	ORFV	1.416	0.941	0.942	1.007	0.935	0.917	0.962	0.944	0.936	0.919	0.898	0.888	0.887	0.890	0.894	0.933	0.899	0.852	0.884	0.756	0.163	0.181		0.047	0.182	0.876	0.999	1.001	1.242	1.218	1.206	1.201
	PCPV	1.422	0.939	0.942	1.012	0.934	0.915	0.964	0.947	0.934	0.919	0.897	0.887	0.887	0.891	0.894	0.934	0.901	0.853	0.888	0.755	0.161	0.180	0.047		0.184	0.876	1.000	1.002	1.241	1.217	1.207	1.202
	SePPV	1.423	0.916	0.919	0.982	0.914	0.901	0.939	0.927	0.917	0.905	0.881	0.873	0.872	0.879	0.879	0.916	0.883	0.840	0.865	0.752	0.161	0.192	0.182	0.184		0.881	0.991	0.993	1.220	1.190	1.188	1.184
Molluscipox	MOCV	1.125	0.772	0.778	0.868	0.784	0.771	0.802	0.790	0.767	0.783	0.757	0.751	0.752	0.756	0.755	0.796	0.758	0.800	0.816	0.608	0.864	0.875	0.876	0.876	0.881		0.553	0.552	0.857	0.867	0.851	0.849
Unclassified	EKPV	1.122	0.828	0.836	0.903	0.844	0.836	0.859	0.838	0.830	0.847	0.832	0.826	0.825	0.826	0.828	0.858	0.831	0.873	0.904	0.746	0.985	1.002	0.999	1.000	0.991	0.553		0.010	0.672	0.683	0.670	0.664
	WKPV	1.123	0.827	0.836	0.901	0.846	0.837	0.859	0.837	0.831	0.847	0.830	0.824	0.823	0.825	0.826	0.857	0.830	0.875	0.906	0.747	0.986	1.004	1.001	1.002	0.993	0.552	0.010		0.671	0.681	0.670	0.664
Avipox	TKPV	1.285	0.911	0.924	0.948	0.943	0.937	0.924	0.909	0.910	0.908	0.935	0.932	0.933	0.933	0.935	0.935	0.923	0.976	1.004	0.989	1.217	1.251	1.242	1.241	1.220	0.857	0.672	0.671		***0.309***	0.279	0.278
	CNPV	1.282	0.914	0.926	0.962	0.955	0.942	0.936	0.914	0.917	0.911	0.935	0.932	0.933	0.934	0.940	0.943	0.927	0.980	1.009	0.983	1.201	1.232	1.218	1.217	1.190	0.867	0.683	0.681	***0.309***		0.246	0.243
	FWPV	1.282	0.906	0.913	0.948	0.940	0.931	0.920	0.903	0.907	0.905	0.923	0.919	0.919	0.923	0.923	0.928	0.912	0.967	0.998	0.966	1.189	1.224	1.206	1.207	1.188	0.851	0.670	0.670	0.279	0.246		0.035
	PEPV	1.278	0.901	0.911	0.946	0.938	0.930	0.919	0.902	0.905	0.901	0.921	0.917	0.916	0.921	0.922	0.927	0.910	0.968	0.995	0.965	1.182	1.220	1.201	1.202	1.184	0.849	0.664	0.664	0.278	0.243	0.035	

Table 3. A distance matrix between ChPV species with values highlighted based on a proposed genus classification (Centapox intra-genus threshold)

Matrix created from the maximum-likelihood phylogenetic tree made with the concatenation of 81 conserved ChPV genes from 32 representative ChPV species; distance value is the average amino acid substitution per site under the LG substitution model; current genus classification is boxed in thick borders; values less than the intra-genus distance threshold (0.113 between YKV and Murmansk virus) are highlighted in yellow as a genus.

Genus	Virus	CRV	YLDV	YMTV	COTV	RFV	MYXV	GTPV	SWPV	DPV	EPTV	VARV	VACV	CPXV	VPXV	RCNV	YKV	Mur	PTPV	SOPV	SQPV	BPSV	PPV	ORFV	PCPV	SePPV	MOCV	EKPV	WKPV	TKPV	CNPV	FWPV	PEPV
Crocodylidpox	CRV		1.322	1.328	1.384	1.326	1.308	1.343	1.341	1.329	1.327	1.308	1.302	1.302	1.311	1.313	1.341	1.312	1.346	1.387	1.249	1.420	1.426	1.416	1.422	1.423	1.125	1.122	1.123	1.285	1.282	1.282	1.278
Clade II	YLDV	1.322		0.094	0.404	0.362	0.360	0.314	0.330	0.293	0.389	0.455	0.450	0.451	0.450	0.455	0.470	0.455	0.666	0.700	0.644	0.926	0.951	0.941	0.939	0.916	0.772	0.828	0.827	0.911	0.914	0.906	0.901
	YMTV	1.328	0.094		0.418	0.377	0.374	0.330	0.349	0.310	0.410	0.470	0.465	0.466	0.465	0.468	0.489	0.475	0.673	0.705	0.651	0.926	0.954	0.942	0.942	0.919	0.778	0.836	0.836	0.924	0.926	0.913	0.911
	COTV	1.384	0.404	0.418		0.411	0.412	0.369	0.375	0.351	0.473	0.552	0.547	0.546	0.546	0.549	0.553	0.543	0.723	0.765	0.729	0.995	1.024	1.007	1.012	0.982	0.868	0.903	0.901	0.948	0.962	0.948	0.946
	RFV	1.326	0.362	0.377	0.411		0.053	0.327	0.327	0.309	0.437	0.503	0.497	0.495	0.501	0.500	0.524	0.505	0.687	0.728	0.662	0.923	0.944	0.935	0.934	0.914	0.784	0.844	0.846	0.943	0.955	0.940	0.938
	MYXV	1.308	0.360	0.374	0.412	0.053		0.323	0.323	0.305	0.435	0.494	0.488	0.487	0.493	0.492	0.517	0.497	0.680	0.723	0.652	0.903	0.927	0.917	0.915	0.901	0.771	0.836	0.837	0.937	0.942	0.931	0.930
	GTPV	1.343	0.314	0.330	0.369	0.327	0.323		0.272	0.261	0.401	0.479	0.475	0.475	0.476	0.475	0.488	0.474	0.668	0.714	0.666	0.951	0.977	0.962	0.964	0.939	0.802	0.859	0.859	0.924	0.936	0.920	0.919
	SWPV	1.341	0.330	0.349	0.375	0.327	0.323	0.272		0.261	0.413	0.481	0.476	0.475	0.472	0.473	0.489	0.476	0.674	0.715	0.661	0.931	0.960	0.944	0.947	0.927	0.790	0.838	0.837	0.909	0.914	0.903	0.902
	DPV	1.329	0.293	0.310	0.351	0.309	0.305	0.261	0.261		0.377	0.450	0.444	0.444	0.444	0.444	0.460	0.447	0.655	0.694	0.636	0.921	0.949	0.936	0.934	0.917	0.767	0.830	0.831	0.910	0.917	0.907	0.905
Unclassified	EPTV	1.327	0.389	0.410	0.473	0.437	0.435	0.401	0.413	0.377		0.450	0.446	0.446	0.446	0.448	0.473	0.453	0.657	0.702	0.645	0.917	0.933	0.919	0.919	0.905	0.783	0.847	0.847	0.908	0.911	0.905	0.901
Orthopox	VARV	1.308	0.455	0.470	0.552	0.503	0.494	0.479	0.481	0.450	0.450		0.013	0.015	0.063	0.063	0.218	0.181	0.676	0.706	0.616	0.887	0.905	0.898	0.897	0.881	0.757	0.832	0.830	0.935	0.935	0.923	0.921
	VACV	1.302	0.450	0.465	0.547	0.497	0.488	0.475	0.476	0.444	0.446	0.013		***0.009***	0.058	0.058	0.214	0.176	0.670	0.702	0.610	0.878	0.895	0.888	0.887	0.873	0.751	0.826	0.824	0.932	0.932	0.919	0.917
	CPXV	1.302	0.451	0.466	0.546	0.495	0.487	0.475	0.475	0.444	0.446	0.015	***0.009***		0.057	0.057	0.213	0.175	0.671	0.703	0.609	0.878	0.896	0.887	0.887	0.872	0.752	0.825	0.823	0.933	0.933	0.919	0.916
	VPXV	1.311	0.450	0.465	0.546	0.501	0.493	0.476	0.472	0.444	0.446	0.063	0.058	0.057		0.036	0.211	0.176	0.673	0.705	0.614	0.887	0.905	0.890	0.891	0.879	0.756	0.826	0.825	0.933	0.934	0.923	0.921
	RCNV	1.313	0.455	0.468	0.549	0.500	0.492	0.475	0.473	0.444	0.448	0.063	0.058	0.057	0.036		0.211	0.177	0.674	0.703	0.614	0.882	0.901	0.894	0.894	0.879	0.755	0.828	0.826	0.935	0.940	0.923	0.922
Centapox	YKV	1.341	0.470	0.489	0.553	0.524	0.517	0.488	0.489	0.460	0.473	0.218	0.214	0.213	0.211	0.211		***0.113***	0.689	0.721	0.659	0.922	0.948	0.933	0.934	0.916	0.796	0.858	0.857	0.935	0.943	0.928	0.927
	Mur	1.312	0.455	0.475	0.543	0.505	0.497	0.474	0.476	0.447	0.453	0.181	0.176	0.175	0.176	0.177	***0.113***		0.673	0.705	0.620	0.888	0.908	0.899	0.901	0.883	0.758	0.831	0.830	0.923	0.927	0.912	0.910
Unclassified	PTPV	1.346	0.666	0.673	0.723	0.687	0.680	0.668	0.674	0.655	0.657	0.676	0.670	0.671	0.673	0.674	0.689	0.673		0.403	0.670	0.839	0.870	0.852	0.853	0.840	0.800	0.873	0.875	0.976	0.980	0.967	0.968
Unclassified	SOPV	1.387	0.700	0.705	0.765	0.728	0.723	0.714	0.715	0.694	0.702	0.706	0.702	0.703	0.705	0.703	0.721	0.705	0.403		0.686	0.871	0.897	0.884	0.888	0.865	0.816	0.904	0.906	1.004	1.009	0.998	0.995
Unclassified	SQPV	1.249	0.644	0.651	0.729	0.662	0.652	0.666	0.661	0.636	0.645	0.616	0.610	0.609	0.614	0.614	0.659	0.620	0.670	0.686		0.748	0.759	0.756	0.755	0.752	0.608	0.746	0.747	0.989	0.983	0.966	0.965
Parapox	BPSV	1.420	0.926	0.926	0.995	0.923	0.903	0.951	0.931	0.921	0.917	0.887	0.878	0.878	0.887	0.882	0.922	0.888	0.839	0.871	0.748		0.154	0.163	0.161	0.161	0.864	0.985	0.986	1.217	1.201	1.189	1.182
	PPV	1.426	0.951	0.954	1.024	0.944	0.927	0.977	0.960	0.949	0.933	0.905	0.895	0.896	0.905	0.901	0.948	0.908	0.870	0.897	0.759	0.154		0.181	0.180	0.192	0.875	1.002	1.004	1.251	1.232	1.224	1.220
	ORFV	1.416	0.941	0.942	1.007	0.935	0.917	0.962	0.944	0.936	0.919	0.898	0.888	0.887	0.890	0.894	0.933	0.899	0.852	0.884	0.756	0.163	0.181		0.047	0.182	0.876	0.999	1.001	1.242	1.218	1.206	1.201
	PCPV	1.422	0.939	0.942	1.012	0.934	0.915	0.964	0.947	0.934	0.919	0.897	0.887	0.887	0.891	0.894	0.934	0.901	0.853	0.888	0.755	0.161	0.180	0.047		0.184	0.876	1.000	1.002	1.241	1.217	1.207	1.202
	SePPV	1.423	0.916	0.919	0.982	0.914	0.901	0.939	0.927	0.917	0.905	0.881	0.873	0.872	0.879	0.879	0.916	0.883	0.840	0.865	0.752	0.161	0.192	0.182	0.184		0.881	0.991	0.993	1.220	1.190	1.188	1.184
Molluscipox	MOCV	1.125	0.772	0.778	0.868	0.784	0.771	0.802	0.790	0.767	0.783	0.757	0.751	0.752	0.756	0.755	0.796	0.758	0.800	0.816	0.608	0.864	0.875	0.876	0.876	0.881		0.553	0.552	0.857	0.867	0.851	0.849
Unclassified	EKPV	1.122	0.828	0.836	0.903	0.844	0.836	0.859	0.838	0.830	0.847	0.832	0.826	0.825	0.826	0.828	0.858	0.831	0.873	0.904	0.746	0.985	1.002	0.999	1.000	0.991	0.553		0.010	0.672	0.683	0.670	0.664
	WKPV	1.123	0.827	0.836	0.901	0.846	0.837	0.859	0.837	0.831	0.847	0.830	0.824	0.823	0.825	0.826	0.857	0.830	0.875	0.906	0.747	0.986	1.004	1.001	1.002	0.993	0.552	0.010		0.671	0.681	0.670	0.664
Avipox	TKPV	1.285	0.911	0.924	0.948	0.943	0.937	0.924	0.909	0.910	0.908	0.935	0.932	0.933	0.933	0.935	0.935	0.923	0.976	1.004	0.989	1.217	1.251	1.242	1.241	1.220	0.857	0.672	0.671		0.309	0.279	0.278
	CNPV	1.282	0.914	0.926	0.962	0.955	0.942	0.936	0.914	0.917	0.911	0.935	0.932	0.933	0.934	0.940	0.943	0.927	0.980	1.009	0.983	1.201	1.232	1.218	1.217	1.190	0.867	0.683	0.681	0.309		0.246	0.243
	FWPV	1.282	0.906	0.913	0.948	0.940	0.931	0.920	0.903	0.907	0.905	0.923	0.919	0.919	0.923	0.923	0.928	0.912	0.967	0.998	0.966	1.189	1.224	1.206	1.207	1.188	0.851	0.670	0.670	0.279	0.246		0.035
	PEPV	1.278	0.901	0.911	0.946	0.938	0.930	0.919	0.902	0.905	0.901	0.921	0.917	0.916	0.921	0.922	0.927	0.910	0.968	0.995	0.965	1.182	1.220	1.201	1.202	1.184	0.849	0.664	0.664	0.278	0.243	0.035	

Table 4. An amino acid % identity matrix between ChPV species with values highlighted based on a proposed genus classification (Centapox intra-genus threshold)

Matrix created from the maximum-likelihood phylogenetic MSA made with the concatenation of 81 conserved ChPV genes from 32 representative ChPV species; current genus classification is boxed in thick borders; values less than the intra-genus distance threshold (89.7% between YKV and Murmansk virus) are highlighted in yellow as a genus.

Genus	Virus	CRV	YLDV	YMTV	COTV	RFV	MYXV	GTPV	SWPV	DPV	EPTV	VARV	VACV	CPXV	VPXV	RCNV	YKV	Mur	PTPV	SOPV	SQPV	BPSV	PPV	ORFV	PCPV	SePPV	MOCV	EKPV	WKPV	TKPV	CNPV	FWPV	PEPV
Crocodylidpox	CRV	100%	47.8%	47.8%	46.9%	48.1%	48.4%	47.4%	47.4%	47.8%	47.8%	48.2%	48.4%	48.3%	48.1%	48.1%	47.6%	48.2%	47.5%	46.8%	49.8%	47.0%	47.0%	47.0%	47.0%	46.7%	52.5%	52.2%	52.2%	48.6%	48.7%	48.7%	48.7%
Clade I	YLDV	47.8%	100%	91.2%	72.2%	74.2%	74.2%	76.7%	75.8%	77.6%	72.8%	69.7%	70.0%	69.9%	70.0%	69.8%	69.3%	69.9%	62.1%	61.0%	62.0%	53.9%	53.4%	53.6%	53.6%	54.3%	58.3%	57.1%	57.2%	56.0%	55.8%	56.0%	56.0%
	YMTV	47.8%	91.2%	100%	71.5%	73.5%	73.5%	75.8%	74.8%	76.6%	71.7%	69.0%	69.3%	69.2%	69.4%	69.2%	68.5%	69.0%	61.8%	60.9%	61.9%	54.0%	53.5%	53.7%	53.7%	54.4%	58.1%	56.9%	56.9%	55.7%	55.5%	55.8%	55.8%
	COTV	46.9%	72.2%	71.5%	100%	72.0%	71.9%	73.9%	73.6%	74.6%	69.4%	66.0%	66.2%	66.2%	66.3%	66.2%	66.3%	66.6%	60.4%	59.2%	59.5%	52.8%	52.2%	52.4%	52.4%	53.2%	55.8%	55.4%	55.4%	55.3%	54.8%	55.2%	55.1%
	RFV	48.1%	74.2%	73.5%	72.0%	100%	94.8%	75.9%	75.9%	76.9%	70.7%	67.8%	68.1%	68.1%	68.0%	68.0%	67.3%	67.9%	61.6%	60.2%	62.0%	54.6%	54.2%	54.3%	54.3%	54.8%	58.4%	57.1%	57.0%	55.1%	54.9%	55.3%	55.3%
	MYXV	48.4%	74.2%	73.5%	71.9%	94.8%	100%	76.1%	76.0%	77.0%	70.6%	68.1%	68.3%	68.4%	68.2%	68.2%	67.5%	68.2%	61.7%	60.4%	62.4%	55.2%	54.6%	54.7%	54.8%	55.2%	58.7%	57.3%	57.3%	55.2%	55.2%	55.5%	55.4%
	GTPV	47.4%	76.7%	75.8%	73.9%	75.9%	76.1%	100%	78.8%	79.5%	72.3%	68.7%	68.9%	68.8%	68.9%	69.0%	68.7%	69.2%	62.2%	60.6%	61.3%	53.5%	53.0%	53.3%	53.2%	53.9%	57.4%	56.3%	56.4%	55.9%	55.5%	56.1%	56.0%
	SWPV	47.4%	75.8%	74.8%	73.6%	75.9%	76.0%	78.8%	100%	79.5%	71.8%	68.7%	68.9%	68.9%	69.1%	69.0%	68.6%	69.1%	61.7%	60.7%	61.5%	53.9%	53.3%	53.5%	53.5%	54.0%	57.8%	56.8%	56.8%	56.0%	55.9%	56.2%	56.2%
	DPV	47.8%	77.6%	76.6%	74.6%	76.9%	77.0%	79.5%	79.5%	100%	73.4%	69.9%	70.2%	70.2%	70.3%	70.3%	69.9%	70.3%	62.5%	61.3%	62.4%	54.3%	53.7%	53.9%	54.0%	54.4%	58.5%	57.2%	57.1%	56.2%	56.0%	56.2%	56.2%
Unclassified	EPTV	47.8%	72.8%	71.7%	69.4%	70.7%	70.6%	72.3%	71.8%	73.4%	100%	70.0%	70.3%	70.2%	70.2%	70.1%	69.3%	70.1%	62.3%	61.0%	62.2%	54.4%	54.1%	54.3%	54.4%	54.8%	57.9%	56.5%	56.5%	56.0%	55.8%	56.0%	56.0%
Orthopox	VARV	48.2%	69.7%	69.0%	66.0%	67.8%	68.1%	68.7%	68.7%	69.9%	70.0%	100%	98.7%	98.5%	93.9%	93.9%	82.2%	84.6%	61.6%	60.7%	63.3%	55.5%	55.1%	55.2%	55.3%	55.5%	58.9%	57.2%	57.2%	55.2%	55.1%	55.4%	55.4%
	VACV	48.4%	70.0%	69.3%	66.2%	68.1%	68.3%	68.9%	68.9%	70.2%	70.3%	98.7%	100%	***99.1%***	94.4%	94.4%	82.4%	84.9%	61.8%	60.9%	63.5%	55.7%	55.4%	55.5%	55.6%	55.7%	59.1%	57.3%	57.4%	55.3%	55.2%	55.6%	55.5%
	CPXV	48.3%	69.9%	69.2%	66.2%	68.1%	68.4%	68.8%	68.9%	70.2%	70.2%	98.5%	***99.1%***	100%	94.5%	94.5%	82.4%	85.0%	61.8%	60.9%	63.5%	55.7%	55.4%	55.5%	55.5%	55.7%	59.0%	57.4%	57.4%	55.3%	55.2%	55.6%	55.5%
	VPXV	48.1%	70.0%	69.4%	66.3%	68.0%	68.2%	68.9%	69.1%	70.3%	70.2%	93.9%	94.4%	94.5%	100%	96.5%	82.6%	84.9%	61.7%	60.8%	63.3%	55.4%	55.0%	55.3%	55.4%	55.5%	58.9%	57.3%	57.4%	55.3%	55.2%	55.4%	55.3%
	RCNV	48.1%	69.8%	69.2%	66.2%	68.0%	68.2%	69.0%	69.0%	70.3%	70.1%	93.9%	94.4%	94.5%	96.5%	100%	82.6%	84.9%	61.7%	60.8%	63.3%	55.5%	55.1%	55.2%	55.3%	55.5%	58.9%	57.3%	57.3%	55.3%	55.0%	55.4%	55.3%
Centapox	YKV	47.6%	69.3%	68.5%	66.3%	67.3%	67.5%	68.7%	68.6%	69.9%	69.3%	82.2%	82.4%	82.4%	82.6%	82.6%	100%	***89.7%***	61.3%	60.4%	61.8%	54.6%	54.0%	54.2%	54.3%	54.7%	57.7%	56.4%	56.4%	55.4%	55.2%	55.7%	55.6%
	Mur	48.2%	69.9%	69.0%	66.6%	67.9%	68.2%	69.2%	69.1%	70.3%	70.1%	84.6%	84.9%	85.0%	84.9%	84.9%	***89.7%***	100%	61.7%	60.8%	63.2%	55.4%	55.0%	55.1%	55.1%	55.5%	58.8%	57.2%	57.3%	55.6%	55.4%	55.8%	55.8%
Unclassified	PTPV	47.5%	62.1%	61.8%	60.4%	61.6%	61.7%	62.2%	61.7%	62.5%	62.3%	61.6%	61.8%	61.8%	61.7%	61.7%	61.3%	61.7%	100%	71.9%	61.3%	56.5%	55.7%	56.2%	56.3%	56.5%	57.7%	56.0%	55.9%	54.1%	54.0%	54.3%	54.2%
Unclassified	SOPV	46.8%	61.0%	60.9%	59.2%	60.2%	60.4%	60.6%	60.7%	61.3%	61.0%	60.7%	60.9%	60.9%	60.8%	60.8%	60.4%	60.8%	71.9%	100%	61.0%	55.8%	55.4%	55.6%	55.5%	56.0%	57.3%	55.5%	55.4%	53.8%	53.7%	54.0%	54.0%
Unclassified	SQPV	49.8%	62.0%	61.9%	59.5%	62.0%	62.4%	61.3%	61.5%	62.4%	62.2%	63.3%	63.5%	63.5%	63.3%	63.3%	61.8%	63.2%	61.3%	61.0%	100%	60.2%	60.2%	60.1%	60.1%	60.0%	64.7%	60.2%	60.2%	53.3%	53.4%	53.8%	53.7%
Parapox	BPSV	47.0%	53.9%	54.0%	52.8%	54.6%	55.2%	53.5%	53.9%	54.3%	54.4%	55.5%	55.7%	55.7%	55.4%	55.5%	54.6%	55.4%	56.5%	55.8%	60.2%	100%	86.8%	86.1%	86.3%	86.2%	57.3%	54.0%	54.0%	48.8%	49.0%	49.2%	49.3%
	PPV	47.0%	53.4%	53.5%	52.2%	54.2%	54.6%	53.0%	53.3%	53.7%	54.1%	55.1%	55.4%	55.4%	55.0%	55.1%	54.0%	55.0%	55.7%	55.4%	60.2%	86.8%	100%	85.0%	85.1%	84.2%	57.2%	53.8%	53.7%	48.3%	48.5%	48.5%	48.6%
	ORFV	47.0%	53.6%	53.7%	52.4%	54.3%	54.7%	53.3%	53.5%	53.9%	54.3%	55.2%	55.5%	55.5%	55.3%	55.2%	54.2%	55.1%	56.2%	55.6%	60.1%	86.1%	85.0%	100%	95.5%	84.9%	57.0%	53.9%	53.8%	48.3%	48.8%	48.9%	49.0%
	PCPV	47.0%	53.6%	53.7%	52.4%	54.3%	54.8%	53.2%	53.5%	54.0%	54.4%	55.3%	55.6%	55.5%	55.4%	55.3%	54.3%	55.1%	56.3%	55.5%	60.1%	86.3%	85.1%	95.5%	100%	84.7%	57.0%	53.9%	53.8%	48.4%	48.9%	48.9%	49.0%
	SePPV	46.7%	54.3%	54.4%	53.2%	54.8%	55.2%	53.9%	54.0%	54.4%	54.8%	55.5%	55.7%	55.7%	55.5%	55.5%	54.7%	55.5%	56.5%	56.0%	60.0%	86.2%	84.2%	84.9%	84.7%	100%	56.7%	53.8%	53.8%	48.7%	49.4%	49.3%	49.4%
Molluscipox	MOCV	52.5%	58.3%	58.1%	55.8%	58.4%	58.7%	57.4%	57.8%	58.5%	57.9%	58.9%	59.1%	59.0%	58.9%	58.9%	57.7%	58.8%	57.7%	57.3%	64.7%	57.3%	57.2%	57.0%	57.0%	56.7%	100%	66.9%	66.9%	56.6%	56.3%	56.7%	56.7%
Unclassified	EKPV	52.2%	57.1%	56.9%	55.4%	57.1%	57.3%	56.3%	56.8%	57.2%	56.5%	57.2%	57.3%	57.4%	57.3%	57.3%	56.4%	57.2%	56.0%	55.5%	60.2%	54.0%	53.8%	53.9%	53.9%	53.8%	66.9%	100%	99.0%	61.7%	61.4%	61.7%	61.9%
	WKPV	52.2%	57.2%	56.9%	55.4%	57.0%	57.3%	56.4%	56.8%	57.1%	56.5%	57.2%	57.4%	57.4%	57.4%	57.3%	56.4%	57.3%	55.9%	55.4%	60.2%	54.0%	53.7%	53.8%	53.8%	53.8%	66.9%	99.0%	100%	61.8%	61.5%	61.7%	62.0%
Avipox	TKPV	48.6%	56.0%	55.7%	55.3%	55.1%	55.2%	55.9%	56.0%	56.2%	56.0%	55.2%	55.3%	55.3%	55.3%	55.3%	55.4%	55.6%	54.1%	53.8%	53.3%	48.8%	48.3%	48.3%	48.4%	48.7%	56.6%	61.7%	61.8%	100%	76.9%	78.5%	78.5%
	CNPV	48.7%	55.8%	55.5%	54.8%	54.9%	55.2%	55.5%	55.9%	56.0%	55.8%	55.1%	55.2%	55.2%	55.2%	55.0%	55.2%	55.4%	54.0%	53.7%	53.4%	49.0%	48.5%	48.8%	48.9%	49.4%	56.3%	61.4%	61.5%	76.9%	100%	80.5%	80.6%
	FWPV	48.7%	56.0%	55.8%	55.2%	55.3%	55.5%	56.1%	56.2%	56.2%	56.0%	55.4%	55.6%	55.6%	55.4%	55.4%	55.7%	55.8%	54.3%	54.0%	53.8%	49.2%	48.5%	48.9%	48.9%	49.3%	56.7%	61.7%	61.7%	78.5%	80.5%	100%	96.6%
	PEPV	48.7%	56.0%	55.8%	55.1%	55.3%	55.4%	56.0%	56.2%	56.2%	56.0%	55.4%	55.5%	55.5%	55.3%	55.3%	55.6%	55.8%	54.2%	54.0%	53.7%	49.3%	48.6%	49.0%	49.0%	49.4%	56.7%	61.9%	62.0%	78.5%	80.6%	96.6%	100%

In the first scenario, **Table 2** visualizes the current genus classification threshold, whereby any pair of viruses with distances less than 0.310 (the maximum intra-genus distance taken from CNPV and TKPV of the *Avipoxvirus* genus) are highlighted "yellow" to form a genus together; consequently, any distances less than 0.009 (taken from the minimum distance between distinct species CPXV and VACV) are classified into the same virus species in "red". It can be seen that the current genus classification does not correspond strictly with their corresponding distance values. The threshold value applied to the grouping of APVs (<0.310) is simply too big such that it falsely groups the *Centapovxirus* species (YKV, Murmansk) with the OPVs, groups swinepox virus (SWPV) and deerpox virus (DPV) into the *Capripoxvirus* genus, and lumps the DPV again with the *Yatapoxvirus*.

Alternatively, **Table 3** proposes and visualizes a more conserved threshold value from the *Centapoxvirus* genus (the genus with the largest distance after avipox- and parapoxviruses). In this scenario, the maximum intra-genus threshold between YKV and Murmansk virus (<0.113) is used and applied to the matrix. Together, the results suggest that the *Parapoxvirus* and *Avipoxvirus* genera be broken down further. **Table 4** shows the corresponding amino acid identity.

3.1.3 Discussion

3.1.3.1 Gene and sequence selection

Overall, the results have demonstrated that the tree made from this particular set of 7 conserved core genes captures a variation of gene divergence within poxviruses (**Figure 3**), and reproduce the same ChPV topology as a tree made from all 81 conserved genes (data not shown) with acceptable bootstrap values above 70[218]. Some branches with lower bootstrap values are observed for the Clade II viruses in our tree; this is also seen in the published

Clade II phylogenies[181], and is commonly found in phyla that underwent rapid divergence into many species groups in a short period. We also show that the phylogeny runtime benefits from the reductions of sampled genes and species representatives. From **Figure 6**, we see a correlation and linear growth between increasing file size and increasing phylogenetic analysis runtime. Specifically, the removal of about 29% of ChPV species (leaving just 32 representatives from a total of 44 species) resulted in a decrease in analysis runtime by about 27%. Similarly, by increasing the amount of genes sampled from 7 to 81 (increasing file size by 3.5 times), resulted in an approximately 3.5 times increase in runtime. We additionally observed that data point C, which had included divergent poxviruses (such as SGPV and the entompoxviruses), was the obvious outlier that exponentially increased runtime albeit having created the same topology. Consequently, we recommended against the inclusion of divergent and redundant poxviruses both for MSA quality/accuracy and efficiency sakes.

3.1.3.2 Inconsistency with previous tree

The goal of this section is to illustrate an accurate phylogenetic relationship for the 5 novel species and the extant ChPV tree. An inconsistency is noted for our EPTV position against the result in its pilot study[219]. Although bootstrap analysis strongly supports the position of EPTV in our tree, the previous study had indicated EPTV to be more closely related to COTV and that the pair branched off the backbone to form their own clade[219]. When we further analyzed the concatenated protein sequences used by Emerson *et al.* (2013) we found that the protein sequences of the P4a precursor gene (one of the 7 genes used in sequence concatenation) were placed under the wrong viruses. Importantly, this switched the COTV and EPTV proteins with orthologous sequences from ORFV and BPSV (bovine popular stomatitis virus), respectively, and this was sufficient, when the tree was generated, to create the impression that COTV and EPTV were within a single clade due to the close relationship

of the parapoxviruses. The correct phylogenetic in this work shows that EPTV branches off the common backbone of the phylogenetic tree as a sister group to the Clade II viruses.

3.1.3.3 Implications of viral phylogeny

As can be seen in **Figure 4**, the novel poxviruses in this thesis (KPVs, PTPV, EPTV, and SOPV) are very divergent from the extant ChPVs based on pairwise distances (measured in aa substitution per site) and % aa identity across their 81 core genes MSA, and should establish 4 novel genera in total.

The KPVs form a novel genus that is positioned on the phylogenetic tree as a sister group to the APVs, but have an average distance of 0.552 aa substitution (**Table 3**) and 67% aa identity (**Table 4**) with MOCV across the 81 core genes MSA. A genus name, *Thylacopoxvirus* (thylaco-: [Gr.] *thylakos* meaning sac or pouch), is proposed. Their short branch length (0.010) is similar to that of the CPXV and VACV (0.009), which is roughly predicted to have evolved about 10,000 years ago[51]. The crude estimated mutation accumulation rate of poxviruses is at (1.7-8.8) x 10^{-6} nucleotide substitution per site per year. Taking the regions bounded by ITR, KPVs have 5861 differences. Thus, KPV may be estimated to have diverged 4-20 thousand years ago. The KPVs' respective kangaroo hosts *M. fuliginosus* and *M. giganteus* are endemic in western and eastern Australia, respectively, and are thought to have diverged from a common macropodid ancestor during the Pliocene, approximately 2.5±1.5 million years before present[220]. Therefore, the KPVs do not support ancient host-virus co-speciation, and demonstrate (relatively recent) adaptation and colonization in kangaroo hosts found on opposite sides of the Australian continent.

Similarly, the discovery of multiple poxviruses in bats and marine mammal hosts leads to inquiries about the possibility of co-speciation between the host and virus[221]. The phylogenetic tree (**Figure 4**) shows PTPV and EPTV as very divergent from each other, each genome having very different base compositions (66 and 76% AT), and branching off into their own clades from the common backbone. While EPTV forms its own branch that is sister group to the Clade II viruses, PTPV groups with SOPV instead, and have 72% pairwise aa identity (**Table 4**) and an average of 0.403 aa substitution (**Table 3**) across their 81 core genes MSA. The marine mammal poxviruses (SOPV and SePPV) are equally as divergent, with the SePPV being GC-rich and grouping with the parapoxviruses, the two are 0.865 distances apart (**Table 3**) and share only 56% aa identity (**Table 4**). This was not surprising, as the molecular phylogeny shows seals and sea otters are quite diverged as well; in fact, recent molecular phylogeny suggested that the pinnipeds (which include seals) shared a more recent common ancestor with the bears than with the mustelids (include sea otters).

Eptesicus fuscsus (EPTV source) is part of the microbat suborder and is found in the United States and is geographically and markedly genetically separated from members of the megabat suborder (containing PTPV source), with an estimated divergence time of 58.9 million years between them[222]. A separate bat-isolated poxvirus, Eidolon helvum poxvirus 1, which is only identified in a metagenomic study of material from African straw-coloured fruit bats (a megabat), is excluded because no corresponding sequence data is available; however, it has previously been shown[223] to be most closely related (76 % aa identity of the P4b core protein) to MOCV, and both of these viruses have low AT %. Despite both being members of the *Pteropodidae* family (of the megabat order), *Eidolon helvum* and *Pteropus scapulatus* are markedly geographically separated species, being found in Africa and Australia respectively.

This low level of sequence similarity between characterized poxviruses from bats is not surprising given the genetic and geographical separation between the three species.

Given the long distance seen between the megabat and microbat poxviruses (which translates to a divergence time roughly on the scale of a million (10^6) years ago), it would have been possible to suggest that these bat poxviruses have evolved in the highly diverged bat species since an ancient divergence. However, the fact that the poxviruses found in megabat and sea otter are each other's closest relatives demonstrates that poxviral phylogeny does not always conform to host phylogeny. In this case, we cannot make any inferences about virus-host co-evolution, as it is difficult to distinguish between ancient and recent adaptation events from phylogenetic data. It is observed that neither the bat poxviruses nor the marine mammalian poxviruses share exclusive evolutionary lineages despite their isolations from related hosts. This is consistent with examples of the ChPV tree where the capripox- and parapoxviruses do not cluster into the same branch despite both infecting even-toed ungulates hosts[49].
Taken together, these novel poxviruses provide additional examples to show that most poxviruses do not always to co-speciate with the host lineages, and demonstrate the diversity of poxvirus in terms of adaptation in different hosts across time and geographical locations.

3.1.3.4 Problems with current genus classification

From the resultant distance matrices of the ChPV phylogenetic trees, it appears that the current ChPV genus classification (**Table 2**) had been a somewhat arbitrary process, as no consistent genetic distance is seen across and between different genera. It no longer seems appropriate to group the avian poxviruses or the GC-rich parapoxviruses into one genus by disregarding the great genetic distances observed between them (>0.310 average aa substitution per site). To allow the least amount of changes to the current structure, a

proposed distance threshold for genus can come from that of the *Centapoxvirus* genus, where the largest intra-genus distance is approximately 0.113 on the MSA made from 81 conserved core genes (**Table 3**). In such cases, the current avian poxviruses should at least be broken down into 3 different genera represented by TKPV, CNPV, and FWPV respectively. PEPV and PNPV have distances of about 0.035 with FWPV, and would be grouped under the same genus. Likewise, current parapoxviruses should be deconstructed into separate genera with the exception of ORFV and PCPV (0.047 pairwise distance). In summary, based on the MSA of 81 conserved proteins, viruses with < 0.113 pairwise distances (or >90% aa identity) should form a genus of their own. This proposed threshold would affect no additional poxvirus genera other than *Avipoxvirus* and *Parapoxvirus*, but establishes a more consistent distance threshold throughout the ChPV genera.

3.2 Pteropox virus

Genomic characterization of a novel poxvirus from a flying fox: evidence for a new genus?

Mark A. O'Dea[1*], Shin-Lin Tu[2*], Stanley Pang[1], Thomas De Ridder[3], Bethany Jackson[1], Chris Upton[2]

1. School of Veterinary and Life Sciences, Murdoch University, Perth, Western Australia, Australia
2. Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC, Canada
3. Department of Agriculture and Water Resources, Cairns, Queensland, Australia

* Co-first authors: these authors contributed equally to this work.

The work contained in this chapter constitutes the manuscript submitted to and accepted by the Journal of General Virology. Permission to use the copyrighted material is provided in **Appendix 2**.

MO, SP, TDR, BJ supplied the samples, performed the wet-lab components of the experiment. MO, SLT performed the genome assembly; SLT finalized the genome sequence.

SLT annotated the genes and performed sequence, structural, and phylogeny analyses. MO provided veterinary context and implications. MO, SLT, and CU drafted the paper.

3.2.1 Abstract

The carcass of an Australian little red flying fox (*Pteropus scapulatus)* which died following entrapment on a fence was submitted to the laboratory for Australian bat lyssavirus exclusion testing, which was negative. During post-mortem, multiple nodules were noted on the wing membranes, therefore degenerate PCR primers targeting the poxvirus DNA polymerase gene were used to screen for poxviruses. The poxvirus PCR screen was positive and sequencing of the PCR product demonstrated very low, but significant similarity with the DNA polymerase gene from members of the *Poxviridae* family. Next generation sequencing of DNA extracted from the lesions was assembled into the first full-length bat-isolated poxvirus genome of 133,492 nucleotides. Analysis of this Pteropox virus genome revealed it to be AT-rich with inverted terminal repeats of at least 1,314 nt and to contain 143 predicted genes. The genome contains a surprisingly large number (29) of genes not found in other poxviruses, one of which appears to be a homolog of the mammalian TRAIL gene. Phylogenetic analysis indicates that the poxvirus described here is not closely related to any other poxvirus isolated from bats or other species, and that it likely should be placed in a new genus.

3.2.2 Background

Bats are reservoirs of numerous viral agents that are transmissible to humans and animals, some of which are highly pathogenic in these species. Recent emerging or re-emerging diseases of public health significance that have their origin in bats include infections by SARS coronavirus[224], MERS coronavirus[225], Hendra virus[226], Nipah virus[227], Ebola virus[228,229] and Australian bat lyssavirus[230]. In addition to this, ongoing research into bat-borne viruses using next generation sequencing based protocols is continually revealing new

viral species[231,232]. Given the extremely large diversity of bat species in the world, with bats being second only to rodents in mammalian species diversity and making up some 20% of mammals[233], it is likely this trend of discovery will continue. With the notable exception of lyssaviruses, the majority of bat-borne viruses reported to date, do not appear to cause significant clinical disease in the bat host, and this ability to host viruses with no apparent ill-effect is hypothesised to be one of the factors contributing to the broad range of viruses harboured by bats[234]. It has recently been shown that a contracted IFN locus and constitutive expression of *IFN-α* is present in black flying foxes, and is hypothesised to be one of the mechanisms associated with unusual mammalian host-viral dynamics[235].

In 2013, three separate reports described the association of poxviruses with bats on 3 continents. Eptesipox virus (EPTV) was isolated in North America[219], Eidolon helvum poxvirus 1 in West Africa[223], and a third virus was identified only by electron microscopy in South Australia[236]. Since only small amounts of sequence data were obtained from the first 2 viruses, we present here the first complete genome of a bat-associated poxvirus.

An Australian little red flying fox (*Pteropus scapulatus*) was found injured, trapped on a fence in the Kimberley region of North Western Australia. These animals are the smallest flying foxes, a species of megabat, recently reclassified as Yinpterochiroptera (Springer, 2013), in mainland Australia. The animal subsequently died at a wildlife care facility. Upon gross examination, multiple (approximately 30), well circumscribed, papular, crusting lesions of approximately 0.5 cm diameter were present across the entire area of the wing membranes. There were no other abnormalities noted on examination of internal organs. A PCR test for the presence of Australian bat lyssavirus RNA in the brain was negative. Due to the gross pathology representing a pox-type lesion, DNA was extracted from the wing lesions and

tested positive for poxviruses by PCR with degenerate primers targeting the viral DNA polymerase. Unfortunately, limited samples were collected from the animal and no viral particles were observed in the homogenised lesion preparation under transmission electron microscopy.

3.2.3 Results

3.2.3.1 High throughput sequencing

Approximately 33 million MiSeq paired end reads were generated with an average read length of approximately 240 nt. In the absence of an available *Pteropus scapulatus* genome, the reads were mapped to the *Pteropus alecto* (black flying fox) scaffold genome using the Burrows Wheeler Aligner[190] in an attempt to filter out host sequences before attempting assembly. After filtering, approximately 0.05% of the reads remained unmapped and paired, these were extracted using SAMtools[191] and fastqutils[192] and then *de novo* assembled using MIRA[195]. The initial assembly generated a 132,353 bp contig that included both inverted terminal repeat (ITR) junctions. Manual extension to match the ITRs (each 1,314 bp) created a final contig of 133,492 bp with sequencing coverage of approximately 400x. In consideration of the species from which this virus was isolated, we have named it Pteropox virus (PTPV). However, since this is the only reported isolation of this virus, the little red flying fox may not be the sole or natural host of this virus.

3.2.3.2 Genome annotation

In the PTPV genome, 143 ORFs were annotated with 001 and 143 being identical because of their location entirely within the left and right ITRs, respectively (**Table 5**). Similar to other poxviruses, the PTPV ORFs are very tightly packed; the core genes are transcribed mostly

inward, whereas the terminal genes are transcribed towards the genome termini. The PTPV genome possesses the 81 core orthologs that are present in all Chordopoxviruses. Interestingly, PTPV contains 29 predicted genes that, at this time, appear to be unique to this virus. All but 3 of these ORFs would encode proteins larger than 80 aa, therefore we believe that this set of ORFs are very likely to be *bona fide* genes. Although, two of these predicted genes (004 and 142; 47% aa identity) are paralogs with very weak similarity to kelch-domain proteins, only one of the unique PTPV genes is predicted to encode a protein with any significant similarity to a protein present in the non-redundant protein database or the translated metagenomics databases. This ORF, PTPV-Aus-040, translates to a predicted 276 aa protein with very significant similarity (52 % aa identity over the ligand domain) to TNF-related apoptosis-inducing ligand (TRAIL), a member of the tumor necrosis factor ligand superfamily (discussed below).

Table 5. Summary of Pteropox virus (PTPV) genome annotations

Ortholog families are represented by VACV-Cop designations; where the ortholog is lacking in VACV-Cop, CPXV-BR gene numbers is used.

Gene #	ORF position	AA #	Gene function	Orthologs
PTPV-001	1193-447	248	Hypothetical protein (ITR)	Unique to Pteropox
PTPV-002	1700-1272	142	Hypothetical protein	Unique to Pteropox
PTPV-003	2328-1744	194	Partial schlafen protein	VACV-Cop-B2R
PTPV-004	2941-2414	175	Hypothetical protein	Unique to Pteropox
PTPV-005	4129-3047	360	IL-1 receptor antagonist	VACV-Cop-C4L/C10L
PTPV-006	5203-4202	333	Ribonucleotide reductase (small subunit)	VACV-Cop-F4L
PTPV-007	6514-5213	433	Phospholipase-D-like protein	VACV-Cop-K4L
PTPV-008	7862-6504	452	Ankyrin repeat protein	VACV-Cop-B4R
PTPV-009	8610-7954	218	Hypothetical protein	Unique to Pteropox
PTPV-010	9180-8665	171	Hypothetical protein	Unique to Pteropox
PTPV-011	10209-9223	328	Hypothetical protein	Unique to Pteropox
PTPV-012	10556-10218	112	Hypothetical protein	Unique to Pteropox
PTPV-013	10893-10588	101	Hypothetical protein	Unique to Pteropox
PTPV-014	11674-11105	189	Hypothetical protein	Unique to Pteropox
PTPV-015	12279-11737	180	Hypothetical protein	Unique to Pteropox
PTPV-016	12622-12407	71	Cytoplasmic protein	VACV-Cop-F8L
PTPV-017	13339-12701	212	S-S bond formation pathway protein	VACV-Cop-F9L
PTPV-018	14648-13326	440	Ser/Thr kinase	VACV-Cop-F10L
PTPV-019	15944-14706	412	RhoA signalling inhibitor	VACV-Cop-F11L
PTPV-020	17964-16003	653	EEV maturation protein	VACV-Cop-F12L
PTPV-021	19160-18000	386	EEV envelope phospholipase	VACV-Cop-F13L
PTPV-022	19411-19220	63	Hypothetical protein	Unique to Pteropox
PTPV-023	19641-19483	52	Hypothetical protein	Unique to Pteropox

PTPV-024	20159-19719	146	Hypothetical protein	VACV-Cop-F15L
PTPV-025	20367-20161	68	RING-H2 zinc finger	SQPV-019
PTPV-026	21999-20398	533	Hypothetical protein	SQPV-020
PTPV-027	22766-22056	236	Conserved non-functional serine recombinase	VACV-Cop-F16L
PTPV-028	22792-23166	124	DNA-binding phosphoprotein (VP11)	VACV-Cop-F17R
PTPV-029	24578-23163	471	Poly(A) polymerase catalytic subunit (VP55)	VACV-Cop-E1L
PTPV-030	26797-24578	739	IEV morphogenesis	VACV-Cop-E2L
PTPV-031	27540-26860	226	dsRNA-binding IFN resistance/PKR inhibitor	VACV-Cop-E3L
PTPV-032	28383-27640	247	RNA polymerase (RPO30)	VACV-Cop-E4L
PTPV-033	28471-28701	76	Hypothetical protein	Unique to Pteropox
PTPV-034	28754-29074	106	Hypothetical protein	Unique to Pteropox
PTPV-035	29080-30780	566	IMV protein, virion morphogenesis	VACV-Cop-E6R
PTPV-036	30786-31595	269	ER-localized membrane protein, virion core protein	VACV-Cop-E8R
PTPV-037	31634-32068	144	dUTPase	VACV-Cop-F2L
PTPV-038	32137-33264	375	3 β-hydroxysteroid dehydrogenase/δ5→4 isomerase	VACV-Cop-A44L
PTPV-039	33283-33708	141	Hypothetical protein	Unique to Pteropox
PTPV-040	33721-34551	276	TNFSF10-like protein (TRAIL, membrane-associated)	Unique to Pteropox
PTPV-041	34613-35101	162	Hypothetical protein	Unique to Pteropox
PTPV-042	38113-35102	1003	DNA polymerase	VACV-Cop-E9L
PTPV-043	38143-38433	96	Sulfhydryl oxidase (FAD-linked)	VACV-Cop-E10R
PTPV-044	38846-38430	138	Virion core protein	VACV-Cop-E11L
PTPV-045	40890-38833	685	Virulence, modulates Raf/MEK/ERK pathway	VACV-Cop-O1L
PTPV-046	43436-40938	832	Hypothetical protein	SQPV-035
PTPV-047	43598-43458	46	Virus entry/fusion complex component	VACV-Cop-O3L
PTPV-048	44545-43613	310	DNA-binding core protein	VACV-Cop-I1L
PTPV-049	44767-44546	73	IMV membrane protein	VACV-Cop-I2L
PTPV-050	45610-44771	279	ssDNA-binding phosphoprotein	VACV-Cop-I3L
PTPV-051	45886-45653	77	IMV protein (VP13)	VACV-Cop-I5L
PTPV-052	47056-45911	381	Telomere-binding protein	VACV-Cop-I6L
PTPV-053	48350-47040	436	Virion core cysteine protease	VACV-Cop-I7L
PTPV-054	48357-50417	686	RNA helicase, DExH-NPH-II domain	VACV-Cop-I8R
PTPV-055	52191-50410	593	Metalloprotease	VACV-Cop-G1L
PTPV-056	52523-52188	111	Entry/fusion complex component	VACV-Cop-G3L
PTPV-057	52517-53182	221	Viral late transcription elongation factor (VLTF)	VACV-Cop-G2R
PTPV-058	53523-53155	122	Glutaredoxin-like protein	VACV-Cop-G4L
PTPV-059	53526-54842	438	FEN1-like nuclease	VACV-Cop-G5R
PTPV-060	54844-55035	63	RNA polymerase (RPO7)	VACV-Cop-G5.5R
PTPV-061	55036-55587	183	NLPc/P60 superfamily protein	VACV-Cop-G6R
PTPV-062	56795-55584	403	Virion phosphoprotein, early morphogenesis	VACV-Cop-G7L
PTPV-063	56824-57606	260	VLTF-1	VACV-Cop-G8R
PTPV-064	57624-58637	337	Entry/fusion complex component, myristylprotein	VACV-Cop-G9R
PTPV-065	58638-59384	248	IMV membrane protein	VACV-Cop-L1R
PTPV-066	59431-59706	91	Crescent membrane/immature virion protein	VACV-Cop-L2R
PTPV-067	60738-59707	343	Internal virion protein	VACV-Cop-L3L
PTPV-068	60763-61524	253	DNA-binding virion core protein (VP8)	VACV-Cop-L4R
PTPV-069	61542-61913	123	Entry/fusion IMV membrane protein	VACV-Cop-L5R
PTPV-070	61867-62337	156	IMV membrane protein, virion morphogenesis	VACV-Cop-J1R
PTPV-071	62376-63377	333	Poly(A) polymerase small subunit (VP39)	VACV-Cop-J3R
PTPV-072	63292-63849	185	RNA polymerase (RPO22)	VACV-Cop-J4R
PTPV-073	64264-63851	137	IMV membrane protein	VACV-Cop-J5L
PTPV-074	64337-68197	1286	RNA polymerase (RPO147)	VACV-Cop-J6R
PTPV-075	68709-68194	171	Tyr/Ser phosphatase, IFN-gamma inhibitor	VACV-Cop-H1L
PTPV-076	68724-69299	191	Entry/fusion IMV membrane protein	VACV-Cop-H2R
PTPV-077	70441-69296	381	IMV p35 heparan binding surface protein	VACV-Cop-H3L
PTPV-078	72826-70442	794	RNA polymerase associated protein (RAP94)	VACV-Cop-H4L
PTPV-079	72944-73567	207	VLTF-4	VACV-Cop-H5R
PTPV-080	73573-74532	319	DNA topoisomerase type I	VACV-Cop-H6R
PTPV-081	74525-74959	144	Crescent membrane/immature virion protein	VACV-Cop-H7R
PTPV-082	74996-77542	848	mRNA capping enzyme (large subunit)	VACV-Cop-D1R
PTPV-083	77890-77504	128	Virion core protein	VACV-Cop-D2L
PTPV-084	77883-78680	265	Virion core protein	VACV-Cop-D3R
PTPV-085	78687-79346	219	Uracil-DNA glycosylase, DNA pol processivity factor	VACV-Cop-D4R
PTPV-086	79379-81742	787	NTPase, DNA primase	VACV-Cop-D5R

PTPV-087	81739-83631	630	VETF-small subunit	VACV-Cop-D6R
PTPV-088	83667-84188	173	RNA polymerase (RPO18)	VACV-Cop-D7R
PTPV-089	84217-84837	206	mRNA decapping enzyme	VACV-Cop-D9R
PTPV-090	84848-85579	243	mRNA decapping enzyme	VACV-Cop-D10R
PTPV-091	87475-85571	634	ATPase, NPH1	VACV-Cop-D11L
PTPV-092	88404-87532	290	mRNA capping enzyme (small subunit)	VACV-Cop-D12L
PTPV-093	90077-88428	549	Trimeric virion coat protein (rifampicin resistance)	VACV-Cop-D13L
PTPV-094	90555-90103	150	VLTF-2	VACV-Cop-A1L
PTPV-095	91260-90586	224	VLTF-3	VACV-Cop-A2L
PTPV-096	91487-91260	75	S-S bond formation pathway protein	VACV-Cop-A2.5L
PTPV-097	93518-91509	669	P4b precursor	VACV-Cop-A3L
PTPV-098	94352-93564	262	Hypothetical protein	Unique to Pteropox
PTPV-099	94391-94921	176	RNA polymerase (RPO19)	VACV-Cop-A5R
PTPV-100	96046-94922	374	Virion morphogenesis, core protein	VACV-Cop-A6L
PTPV-101	98224-96098	708	VETF-large subunit	VACV-Cop-A7L
PTPV-102	98275-99159	294	VITF-3, small subunit	VACV-Cop-A8R
PTPV-103	99443-99156	95	Membrane-associated, early morphogenesis protein	VACV-Cop-A9L
PTPV-104	102203-99444	919	P4a precursor	VACV-Cop-A10L
PTPV-105	102218-103249	343	Viral membrane formation	VACV-Cop-A11R
PTPV-106	103794-103246	182	Virion core and cleavage processing protein	VACV-Cop-A12L
PTPV-107	104039-103809	76	IMV membrane protein, virion maturation	VACV-Cop-A13L
PTPV-108	104321-104040	93	Essential IMV membrane protein	VACV-Cop-A14L
PTPV-109	104501-104340	53	Non-essential IMV membrane protein	VACV-Cop-A14.5L
PTPV-110	104778-104491	95	Core protein	VACV-Cop-A15L
PTPV-111	105874-104765	369	Myristylated protein, essential for entry/fusion	VACV-Cop-A16L
PTPV-112	106455-105889	188	IMV membrane protein	VACV-Cop-A17L
PTPV-113	106470-107891	473	DNA helicase, transcript release factor	VACV-Cop-A18R
PTPV-114	108149-107892	85	Zn finger-like protein	VACV-Cop-A19L
PTPV-115	108500-108150	116	IMV membrane protein, entry/fusion complex	VACV-Cop-A21L
PTPV-116	108499-109806	435	DNA polymerase processivity factor	VACV-Cop-A20R
PTPV-117	109796-110272	158	Holliday junction resolvase	VACV-Cop-A22R
PTPV-118	110292-111440	382	VITF-3, large subunit	VACV-Cop-A23R
PTPV-119	111442-114954	1170	RNA polymerase (RPO132)	VACV-Cop-A24R
PTPV-120	118099-114947	1050	A-type inclusion protein	VACV-Cop-A25L
PTPV-121	120067-118133	644	P4c precursor	VACV-Cop-A26L
PTPV-122	120422-120126	98	IMV surface protein, fusion protein	VACV-Cop-A27L
PTPV-123	120861-120433	142	IMV MP/virus entry	VACV-Cop-A28L
PTPV-124	121773-120862	303	RNA polymerase (RPO35)	VACV-Cop-A29L
PTPV-125	121948-121742	68	IMV protein, entry/fusion complex component	VACV-Cop-A30L
PTPV-126	122100-122693	197	Hypothetical protein	VACV-Cop-A31R
PTPV-127	123429-122671	252	ATPase/DNA packaging protein	VACV-Cop-A32L
PTPV-128	123536-124072	178	Hypothetical protein	Unique to Pteropox
PTPV-129	124201-124728	175	C-type lectin-like IEV/EEV glycoprotein	VACV-Cop-A34R
PTPV-130	124776-125309	177	MHC class II antigen presentation inhibitor	VACV-Cop-A35R
PTPV-131	125380-125970	196	Hypothetical protein	Unique to Pteropox
PTPV-132	126039-126380	113	Hypothetical protein	Unique to Pteropox
PTPV-133	126452-126709	85	Hypothetical protein	Unique to Pteropox
PTPV-134	126820-127542	240	Chemokine binding protein	VACV-Cop-A41L
PTPV-135	127602-127850	82	Hypothetical protein	Unique to Pteropox
PTPV-136	127855-128337	160	Hypothetical protein	Unique to Pteropox
PTPV-137	128401-129111	236	Zn finger-like protein	CPXV-BR-023
PTPV-138	129158-129694	178	Hypothetical protein	Unique to Pteropox
PTPV-139	129773-130303	176	Hypothetical protein	Unique to Pteropox
PTPV-140	130360-130962	200	Hypothetical protein	Unique to Pteropox
PTPV-141	131034-131531	165	Hypothetical protein	Unique to Pteropox
PTPV-142	131631-132164	177	Hypothetical protein	Unique to Pteropox
PTPV-143	132300-133046	248	Hypothetical protein (ITR)	Unique to Pteropox

3.2.3.3 TNF-Related Apoptosis-Inducing Ligand (TRAIL) homolog

PTPV-Aus-040 has no poxvirus orthologs, however, when a BLASTP search of the non-redundant protein database was performed, the predicted protein best matched TRAIL proteins from a wide variety of species (**Figure 7**).

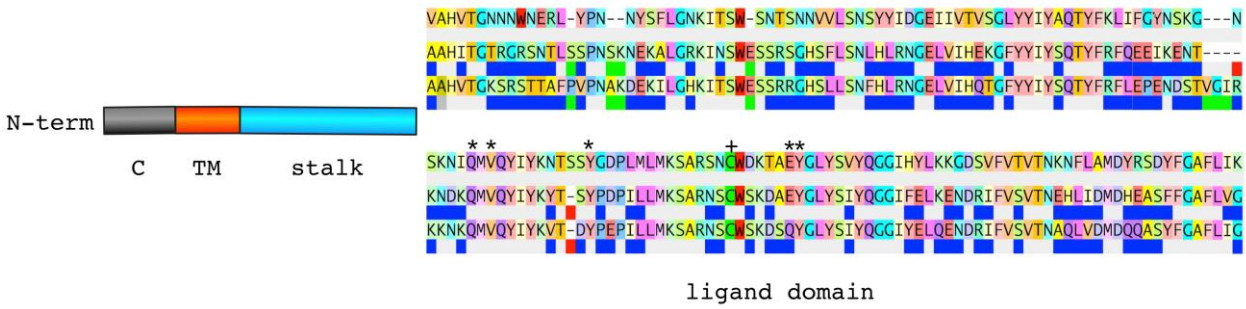

Figure 7. Domain organization of the PTPV-Aus-040 protein.

Cytoplasmic (1–28 aa) and transmembrane (29–50 aa) regions are labelled C and TM, respectively. The TNF-like ligand domain (135–276 aa) is shown as a multiple sequence alignment with PTPV-Aus-040 (top), human TRAIL (middle) and little brown bat TRAIL (bottom) sequences. Residues important for human TRAIL–receptor interactions and zinc ion binding are indicated by (*) and (+), respectively. The alignment was constructed with Base-By-Base; blue, green and red boxes underneath the human and little brown bat sequences illustrate aa changes, insertions and deletions, respectively compared to the PTPV-Aus-040 sequence (top).

TRAIL, which is also known as TNF superfamily member 10, is a type II transmembrane (TM) protein that has a relatively short cytoplasmic domain (N-terminus) followed by a TM domain and an extracellular ligand domain (C-terminus)[237]. CCTOP[202] predicted that this topology of functional domains is very likely conserved in the PTPV-Aus-040 protein. The eukaryotic TRAIL ligand domain is approximately 56 % identical (aa) to PTPV-Aus-040, suggesting a high likelihood to be its cellular homolog via HGT. When the predicted ligand domain of PTPV-Aus-040 was compared to all TRAIL proteins in the database, including several from bats, there was no indication that one, or a group, matched significantly better than any other. This suggests that the acquisition of the cellular counterpart is an ancient

event. I-TASSER, a protein structure and function prediction server[205], generated a predicted structure for the PTPV-Aus-040 protein with a root-mean-square deviation (RMSD) of atomic positions (protein backbone) of 0.76Å when threaded on to a structure of human TRAIL (PDB: 1D4V_B; **Figure 8**). The Robetta protein structure prediction server[238] separately produced very similar results (data not shown). Taken together, these results support the hypothesis that the PTPV-Aus-040 protein is likely to fold into a TRAIL like structure with the ability to interact with TRAIL receptors.

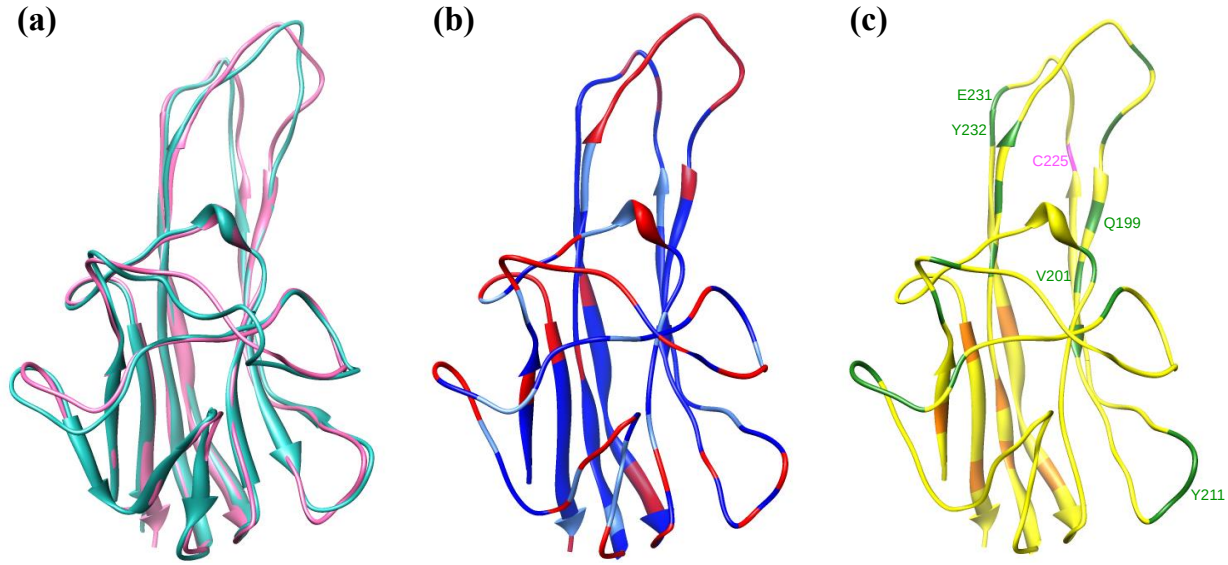

Figure 8. Predicted structure of Pteropox virus TRAIL protein.

(a) Superposition of I-TASSER modelled PTPV-Aus-040 protein (pink) and human TRAIL (PDB : 1D4V; cyan). (b) Predicted structure of the PTPV-Aus-040 protein with identical aa, function- ally conserved aa and non-conserved aa highlighted in blue, light blue and red, respectively. (c) Key aa on the PTPV-Aus-040 protein are highlighted to show Zn-binding cysteine (pink), hydrophobic homotrimer interface (orange) and DR5 receptor contact sites (green).

TRAIL shares a characteristic of other TNF superfamily protein in that it forms trimers, with receptor proteins binding at monomer-monomer interfaces. The DR5 receptor is the most extensively studied TRAIL receptor to date. The structure of a ligand-receptor complex has been solved[239-241] and the interacting amino acids determined. When the ligand portion of the

PTPV-Aus-040 protein is aligned with the human and bat TRAIL protein sequences, the essential residues involved in DR5 binding and apoptotic activity are conserved (**Figure 7** and **Figure 8**).

3.2.3.4 Schlafen-like protein

A number of orthopoxviruses, including Ectromelia virus (ECTV), CPXV and Raccoonpox virus (RCNV) encode a schlafen-like protein of approximately 500 aa[242]. Although the deletion of this gene from ECTV-Moscow generated a highly attenuated virus (Dr Mark Buller, personal communication), the gene is very likely non-functional in VARV because of multiple indels that cause frameshifts scattered throughout the gene, and thus the gene is not associated with smallpox virulence. The VACVs orthologs, however, are disrupted by the introduction of a premature stop-codon approximately at the mid-point of the gene, which is presumed to result in the synthesis of the N-terminal half of the schlafen-like polypeptide since the promoter region is unchanged[243]. Although mammalian schlafens, which comprise a broad family of proteins of various sizes, are known to be regulated by interferons and expressed in a variety of immune tissues, their role in virus infection, if any, is not clear. Interestingly, the PTPV-Aus-003 ORF also encodes the N-terminal half of a schlafen-like poxvirus protein without any DNA region present that resembles the 3' end of a schlafen-like ortholog. However, since a similar arrangement is also found in Yoka poxvirus (YKV)[244] and the Melanoplus sanguinipes entomopoxvirus strain Tucson (MSEV-Tuc)[245]these ORFs, including the VACV gene fragment, which represents the N-terminal half of the ECTV schlafen-like gene may, in fact, be encoding functional proteins.

3.2.3.5 Ankyrin-like proteins

Only one ORF (008) in the PTPV genome is predicted to encode an ankyrin-like protein. Many of the larger poxviruses contain multiple members of this group of proteins[246]. The PTPV-Aus-008 predicted protein has several ankyrin-like motifs within the N-terminal domain and a cellular F-box-like sequence[247] at the C-terminus, an arrangement known as a Pfam PRANC domain. The PTPV protein is highly diverged from all other poxvirus ankyrin-like proteins, it shares approximately 25 % aa identity with a variety of ankyrin-like proteins from a number of other poxviruses, but it is not possible to determine which is most closely related with any certainty. The poxvirus ankyrin-like proteins are not part of the essential set of core poxvirus proteins, but several have been shown to play a role in host range and modulation of the host's anti-viral immune response[246].

3.2.3.6 Genome organization

The large majority of PTPV genes, where they exist in other genomes, are syntenic with the majority of Chordopoxviruses, although some genes may appear to be in a different location simply due to large insertions or deletions in PTPV relative to the genome it is compared to. However, there are a few examples of genes that are in quite different positions. For example, the schlafen-like gene is at the left end of the PTPV genome, but at the right in the orthopoxviruses. The block of DNA (3.5 kb) comprising the PTPV-037 to -041 genes has a number of unusual features. In other genomes, PTPV-037 (dUTPase) and PTPV-038 (3' beta-hydroxysteroid dehydrogenase) are 1) in positions that are much further to the left and right, respectively, of their PTPV positions and 2) transcribed in the opposite direction. In addition, the PTPV-039, -040 (TRAIL homolog) and -041 genes are, to date, unique to this poxvirus and no matches to PTPV-039 and -041 were detected in any protein database. PTPV-039 and -041 might have been a larger ORF that was interrupted by the HGT of

PTPV-038 (TRAIL); either way, no homologous sequences have so far been found for these putative genes despite their probable ORF sizes. The genes that flank this block in PTPV, VACV-Cop-E8R (virion core protein) and VACV-Cop-E9R (DNA polymerase), are immediately adjacent to each other in all poxviruses except the related group of Orf virus (ORFV), Bovine papular stomatitis virus (BPSV) and SQPV, which have a single ORF between them. Since the PTPV genome was assembled from short reads, all insertion/deletion junction regions breaking expected synteny patterns were checked manually to ensure that normal sequencing reads spanned these regions confirming correct genome assembly.

When the relationships of the PTPV non-core/non-unique genes were compared to the other poxvirus genera, it was noted that several were only present in the GC-rich poxviruses. For example, orthologs of the PTPV-Aus-025/26 genes, which are inserted between the orthologs of VACV-Cop-F15L/F16L, are present only in GC-rich viruses such as MOCV, ORFV, SQPV and BPSV. The predicted zinc finger protein encoded by PTPV-Aus-025 is 40-45 % identical (aa) to the orthologs encoded by the GC-rich viruses. Similarly, PTPV-Aus-046 (**Table 1**) is found in an equivalent position to SQPV-035. Another group of PTPV genes that appear to be unique to this virus, may in fact be orthologs with unusually low sequence similarity because they are situated in a position that replaces an otherwise missing ortholog in other poxviruses. Examples are PTPV-Aus-022, -023, -098 and -128, which may be orthologs of VACV-Cop-F14L, -F14.5L, -A4L and -A33R, respectively.

Many of the biochemical processes that are required to replicate poxviruses involve multiple gene products. In some processes, the viral proteins are part of a protein complex (e.g. transcription machinery) and in others the proteins may be part of a multi-step pathway (e.g.

formation of extracellular virus). The prediction of virus characteristics based on the presence or absence of viral genes may be more difficult when they are part of such processes, and if host proteins are involved, the operation becomes even more problematic (e.g. ribonucleotide reductase subunits)[248] because an effect may only be present in certain host cells. However, when it is possible to observe a pattern in the presence/absence of genes among sets of viruses some biological information may be obtained. Of the 7 genes (A33R, A34R, A36R, A56R, B5R, F12L, F13L; VACV-Cop nomenclature) known to involved in the formation of extracellular enveloped vaccinia virus (EEV)[151], A36R, A56R and B5R are clearly absent from PTPV and PTPV-128 possesses only a short A33R-like motif. Reviewing the status of the 7 genes in all known poxvirus genomes reveals that these 4 genes are variably present, whereas the remaining 3 genes A34R, F12L, F13L are present in all ChPVs with the exception of SGPV which is perhaps the most divergent virus[249]. However, experiments using VACV deletion mutants do not create clearly distinguishable phenotypes for these 2 sets of genes. Furthermore, the presence of such a large number of unique genes of unknown function in PTPV makes it equally difficult to compare the intertwined characteristics of virulence and host range of PTPV with other poxviruses. A particular PTPV protein may have different activities in different hosts due to interactions with host-specific proteins or with other PTPV-specific proteins. For example, with respect to genes involved with interferon resistance[250], PTPV has orthologs of E3L and H1L but not C7L, K1L, K3L, B8R or B18R; however, with only genome sequences available, it is impossible to know whether the unique PTPV genes or different expression levels are compensating for the missing genes.

3.2.4 Discussion

In this study, we describe the predicted full-length coding region, including ITRs, of a poxvirus genome from a little red flying-fox, *Pteropus scapulatus*, tentatively named Pteropox virus. The genome of PTPV is significantly different from all known poxviruses sequenced to date, including the bat-derived poxviruses Eptesipox virus and Eidolon helvum poxvirus 1, such that it does not fall within any current poxvirus genus. This low level of sequence similarity with other characterised poxviruses from bats is not surprising given the genetic and geographical separation between the three species. *Eptesicus fuscsus* is part of the *Vespertillionidae* family and is found in the United States and Canada and is geographically and markedly genetically separated from the two members of the *Pteropodidae*, with an estimated divergence time between Yinpterochiroptera (containing the *Pteropodidae*) and Yangochiroptera (containing the *Vespertillionidae*) being in the order of 58.9 million years[222]. Despite both being members of the *Pteropodidae* family, *Eidolon helvum* and *Pteropus scapulatus* are markedly geographically separated species, being found in Africa and Australia respectively. It may still be possible that these three bat poxviruses have evolved in the three highly diverged bat species since an ancient divergence. The PTPV genome is 66% AT, which is not unusual among poxviruses that vary widely in nucleotide composition. Poxvirus genomes range from 33-82% AT, although only insect poxviruses have AT > 80%; however, the reason for this disparity and mechanism that produces and maintains it is unknown. Within a genus, values are similar, but again there is no obvious trend associated with the organization of the genera on the complete phylogenetic tree. However, the nucleotide composition of PTPV is interesting in light of the presence of several genes that are otherwise restricted to the GC-rich viruses. This indicates that these two groups of viruses may be more closely related than simple sequence identity and the phylogenetic tree might suggest.

The majority of recent viral discoveries in bats have been from apparently healthy specimens. For example diverse coronavirus species[251,252]and reovirus species[253] have been isolated and characterised from bats showing no signs of clinical disease. Indeed, aside from the lyssavirus species infecting bats, few other viruses to date appear associated with clinical disease. It is therefore interesting to note that the presence of this virus was associated with typical poxvirus-like lesions, that is, vesicular to nodular skin lesions. Little red flying foxes roost in extremely large colonies, sometimes up to one million in number, which theoretically provides many opportunities for disease transmission such as deposition of the virus on body surfaces followed by infection of abrasions or auto-inoculation during grooming. Due to this, it is surprising that such a disease syndrome has not been described previously. Perhaps it is a very acute clinical disease, which resolves rapidly and does not cause the host any systemic illness, or alternatively, there may have been some immunosuppression or immunomodulation in this particular individual, allowing a virus normally carried subclinically, a now well documented situation in bat species[254], to be expressed as a clinical disease entity. Since bats live in vary large colonies, agents that infect them may be restricted to low virulence, (also perhaps chronic or persistent infections) rather than virulent acute infections which may eradicate bat colonies and the pathogen with them, such as in the current situation with the fungal disease White nose syndrome in North America[255]. The other possibility is that only flying foxes which have been captured following misadventure or illness and sent to wildlife care facilities are examined, leading to a very small sample size of the overall population, and subsequent misrepresentation of a disease of low prevalence. Current research into the virome of bat species is largely driven by the search for precursor species of known human pathogens such as MERS coronavirus[256], or scanning for species with zoonotic potential[257]. The finding of bat-associated poxvirus should therefore encourage

survey programs that screen for the prevalence of a long-harbouring poxvirus pool in the population. Unfortunately, we could not addressed the key concern with whether or not PTPV has zoonotic potential due to the inability to isolate the virus for *in vitro* testing in human derived cell lines. However, there are several indications that PTPV may have already undergone extensive host-specific adaptation, 1) relatively few non-essential genes that are usually associated with host specificity are shared with other viruses (e.g. the presence of only a single Ankyrin-like protein), 2) the relatively small genome of PTPV (gene loss in relations to narrower host range), and 3) the presence of 29 unique ORFs. Still, one cannot predict for certain whether or not PTPV might be a zoonotic pathogen.

Poxviruses have captured a variety of genes from their hosts that have subsequently evolved to encode proteins that mitigate the effectiveness of host's immune response. Some of these are receptor mimics and are secreted from infected cells to block their cognate cytokine; examples are known to block IFN-α/β, IFN-ɣ, TNF, IL-18 and chemokines[258,259]. Although fewer in number, some poxviruses also encode versions of cytokines (IL-10) and growth factors (EGF and VEGF)[260-262]. PTPV-Aus-040, which encodes a novel homolog of TRAIL, extends this repertoire of apoptosis inducing proteins further[263]. This is especially interesting because the viral TRAIL homolog, as a member of the TNF ligand-like family, could function as an "attacking cytokine" by triggering apoptosis in cells that are attacking the virus infected cells or as a "defensive cytokine receptor-mimic" by acting as a cytokine that binds TRAIL receptors on infected cells, blocking the receptors and leading to failure to trigger induction of apoptosis. In the latter example, the ligand binding domain of the viral TRAIL would need to be proteolytically cleaved to release it from the surface of the infected cell. Although TRAIL is more potent when membrane-bound, it can be released by cysteine protease cleavage[264]. Since triggering of apoptosis by TRAIL occurs through receptor

clustering[263] the trimer structure of TNF-ligands is also important. Most of the hydrophobic amino acids involved in the trimer interfaces are conserved in the PTPV-Aus-040 protein, however, it will be important to determine if the viral protein forms trimers. A further mechanism is raised by the fact that the human cytomegalovirus (HCMV) protein UL141 binds TRAIL receptors (death receptors) in the endoplasmic reticulum thereby preventing them reaching the surface of the infected cell[265]. However, the HCMV-UL141 protein has no sequence similarity to TRAIL and uses non-canonical interactions and motif mimics. Thus, PTPV-Aus-040 protein could also reduce apoptosis killing of infected cells by preventing the TRAIL receptors from transiting to the cytoplasmic membrane of infected cells.

In conclusion, we have characterised the genome of a novel, highly divergent poxvirus, which, unusually for viral infections in bats, appears to be associated with clinical disease. PTPV is unrelated to other poxviruses isolated from bats and encodes 29 novel poxvirus genes, one of which is predicted to protect infected cells from the host's immune system by a novel mechanism involving TRAIL.

3.3 Eptesipox virus

Characterization of Eptesipoxvirus, a novel poxvirus from a microchiropteran bat

Shin-Lin Tu[1]*, Yoshinori Nakazawa[2], Jinxin Gao[2], Kimberly Wilkins[2], Nadia Gallardo-Romero[2], Yu Li[2,] Ginny L. Emerson[2], Darin S. Carroll[2], Chris Upton[1].

1. Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC, Canada.
2. The National Center for Emerging and Zoonotic Infectious Diseases Centers for Disease Control and Prevention, Atlanta, GA, USA.

The work contained in this chapter constitutes the manuscript submitted to and accepted by Virus Genes. Permission to use the copyrighted material is provided in **Appendix 3**.

GLE, DSC, YL, CU conceived of the pilot study. YN, JG, KW, NGR performed pilot research. SLT, YN, JG assembled the genome; YN, JG performed Sanger sequencing check; SLT finalized the genome sequence. SLT annotated the genes and performed sequence and phylogeny analyses. SLT, GLE, CU, YN wrote the manuscript.

3.3.1 Abstract

The genome of *Eptesipoxvirus* (EPTV) is the first poxvirus genome isolated from a microbat. The 176,688 nt sequence is 67 % AT and is predicted to encode 191 genes. 11 of these genes have no counterpart in GenBank and are therefore unique to EPTV. The presence of a distantly related ortholog of *Vaccinia virus* F5L in EPTV uncovered a link with fragmented F5L orthologs in MOCV, SQPV, and Clade II viruses. Consistent with the unique position of EPTV, EPTV has 13 genes that are typical, if not exclusive, to the Clade II poxviruses, but also has 11 genes that are specific to the orthopoxviruses. This mosaic nature of EPTV blurs the distinction between the old description of the orthopoxvirus and Clade II groups. Genome annotation and characterization failed to find any common virulence genes shared with the other poxvirus isolated from bat (PTPV), however, EPTV encodes 3 genes that may have been transferred to or from DPV and SQPV viruses; 2 of these, a putative endothelin-like protein and a MHC class I-like protein are likely to have immunomodulatory roles.

3.3.2 Background

In 2013, three reports described the association of poxviruses with bats on 3 continents. Both the North American Eptesipox virus (EPTV)[219] and the West African Eidolon helvum poxvirus 1[223] were only partially sequenced, and a third virus was identified only by electron microscopy in South Australia[236]. In 2016, the first complete genome of a bat-associated poxvirus was published from a Pteropox virus found in a little red flying fox (megabat) in North Western Australia[266]. Here we revisit the sequencing data from EPTV reported from North American big brown bats (microbat) and attempt to assemble its full genome and

expand the availability of comparative data. Initial EPTV report[219] recorded that six big brown bats (*Eptesicus fuscus*) were brought into a wildlife hospital from 2009-2011 due to their inability to fly associated with necrosuppurative osteomyelitis in multiple joints of the legs and wings. Diagnostics ran for microbial and fungal infections were negative, but joint tissues were PCR-tested positive for poxvirus. Historically, osteomyelitis with arthritis was also reported in smallpox patients and smallpox vaccine recipients. However, EPTV symptoms and prevalence in microbats remains uncharacterized due to insufficient sampling/surveillance of the bat population.

Poxviruses found in bats are of special interest because bats are associated with a number of known zoonotic disease viruses, including the coronaviruses that cause SARS[224] and MERS[225], the rhabdovirus that causes rabies[267], the filoviruses ebola[229] and Marburg viruses[268], and Australian bat lyssaviruses that cause rabies-like symptoms[230]. Bats, which have been categorized into approximately 1200 species, form the second most diverse animal taxon (*Chiroptera* order) after rodents. So far, bat-isolated poxviruses have been isolated from both the megabat and microbat suborders. These suborders differ significantly; megabats are large fruit bats with big eyes, small ears (always lack tails) that are found in tropical regions, whereas the small microbats have the opposite features, eat insects and use echolocation[233,269]. It has been recently demonstrated that, statistically, bats indeed harbour a higher proportion of zoonotic viruses than all other mammalian hosts[270]. Thus, bats comprise a huge and diverse mammalian reservoir for poxviruses that may pose a risk to humans through zoonotic transmission[271]. Additionally, bat poxviruses may provide a genetic reservoir that, through recombination with other mammalian poxviruses, could generate new human pathogens[46]. It has been hypothesized that recombination events may have been

involved in the development of variola virus as a human specific pathogen[54]. We aim to expand the available genome data and advance the study of bat-associated poxviruses.

3.3.3 Results

3.3.3.1 Genome assembly and gene annotation

Taxonomer.iobio.io found at least 40% of the sequences were determined to be non-viral contamination (human, bacterial, and unknown sequences). Initially, the raw sequence reads that mapped to human genome (build 38) using the Burrows–Wheeler Alignment (BWA) tool[190] were removed, and the remaining reads were extracted using SAMtools[191]and fastqutils[192] prior to assembly. Second, using the same protocol, a separate assembly trial was performed with *Eptesicus fuscus* (bat host) sequences removed. However, both trials failed to generate a complete poxvirus genome contig, probably due to excessive read removals. Therefore, the final assembly was performed directly with the original raw read file with no decontamination process using designated ‘‘genome, denovo, accurate’’ parameters in MIRA[195].

The assembly of the unfiltered Illumina single read sequencing data by MIRA generated 2 large poxvirus specific contigs (121,151 nt and 43,742 nt), which overlapped by 54 nt. After manually joining the 2 contigs and correction/matching of the genome ITRs, a draft genome of 176,491 nt containing 2 copies of 11,999 nt ITRs was obtained. . This contig was mapped to the original read file using the Tanoti short read aligner (http://www.bioinformatics.cvr.ac.uk/tanoti.php), and subsequently visualized in Tablet[196] for coverage analyses and manual base-calling. The mapping of raw reads against this draft sequence gave an average coverage of 150x, which was consistent throughout the genome except for 1 region (134,421 - 134,514 nt). This region, which showed a dramatic increase in

coverage of 411x, was found to consist of 2 tandem repeats of about 54 nt each. Since the Illumina reads, which were approximately 71 nt on average, were unable to resolve the true number of repeats in the region, this region was amplified by PCR and re-sequenced using the Sanger method[272]. This process determined that this region consisted of six repeats rather than 2 and created a final EPTV genome sequence of 176,688 nt. A total of 191 genes were annotated (**Table 6**), of which 13 are associated with each of the ITRs. EPTV-013 and EPTV-179 span the ITR junctions, the former being a N-terminus truncated version of the EPTV-179 gene that is predicted to encode an ankyrin-like protein.

Table 6. Summary of Eptesipox virus (EPTV) genome annotations

Ortholog families are represented by VACV-Cop designations; where the ortholog is lacking in VACV-Cop, the DPV-84 and CPXV-BR gene numbers are used, respectively. Inverted Terminal Repeat regions are bolded.

Similar to Clade II
Clade II-specific orthologs
Clade II orthologs not in orthopox

Similar to orthopoxvirus
Orthopox-specific orthologs
Orthopox orthologs not in Clade II

Potential HGT genes
Novel EPTV genes
Conserved in Chordopox

Gene #	ORF position	AA #	Putative gene function	Orthologs
EPTV-001	620-141	159	Hypothetical protein	VACV-Cop-B15R
EPTV-002	1694-732	320	Serpin 2/CrmA (host range)	VACV-Cop-B13R
EPTV-003	2410-1727	227	Hypothetical protein	DPV-84-009
EPTV-004	3472-2474	332	IL-1 receptor-like protein	DPV-84-015
EPTV-005	3989-3510	159	Hypothetical protein	DPV-84-159
EPTV-006	4934-4029	301	Tyrosine protein kinase-like protein	DPV-84-158
EPTV-007	5683-4991	230	ER-localized apoptosis regulator (host range)	VACV-Cop-B9R
EPTV-008	6245-5769	158	Hypothetical protein	Unique to EPTV
EPTV-009	6804-6325	159	Hypothetical protein	VACV-Cop-B15R
EPTV-010	8696-6963	577	Ankyrin repeat-containing protein (host range)	CPXV-BR-025
EPTV-011	11040-9211	609	Ankyrin repeat-containing protein	DPV-84-019
EPTV-012	11599-11090	169	IFN resistance, eIF2a-like PKR inhibitor (host range)	VACV-Cop-K3L
EPTV-013	11992-11636	118	Ankyrin repeat protein fragment	DPV-84-019
EPTV-014	12864-12010	284	Monoglyceride lipase	VACV-Cop-K5L/K6L
EPTV-015	13183-12941	80	Secreted EGF-like protein	VACV-Cop-C11R
EPTV-016	13689-13189	166	Mitochondria anti-apoptotic factor (host range)	DPV-84-022
EPTV-017	14167-13742	141	dUTPase	VACV-Cop-F2L
EPTV-018	14598-14194	134	IFN-inducible protein (host range)	DPV-84-024
EPTV-019	15616-14642	324	Ribonucleotide reductase small subunit	VACV-Cop-F4L
EPTV-020	16786-15656	376	F5L membrane protein	VACV-Cop-F5L
EPTV-021	17026-16808	72	Hypothetical protein	Unique to EPTV
EPTV-022	17287-17063	74	Hypothetical protein	VACV-Cop-F7L
EPTV-023	17537-17346	63	Cytoplasmic protein	VACV-Cop-F8L
EPTV-024	18031-17645	128	Hypothetical protein	Unique to EPTV
EPTV-025	18714-18067	215	S-S bond formation pathway protein	VACV-Cop-F9L
EPTV-026	20017-18704	437	Ser/Thr protein kinase	VACV-Cop-F10L
EPTV-027	21337-20039	432	RhoA signalling inhibitor, virus release protein	VACV-Cop-F11L
EPTV-028	23302-21362	646	EEV maturation protein	VACV-Cop-F12L

EPTV-029	24456-23341	371	Palmitylated EEV membrane glycoprotein	VACV-Cop-F13L
EPTV-030	24703-24476	75	F14L conserved protein	VACV-Cop-F14L
EPTV-031	24950-24750	66	Hypothetical protein	Unique to EPTV
EPTV-032	25594-25148	148	F15L conserved protein	VACV-Cop-F15L
EPTV-033	26362-25682	226	Conserved non-functional serine recombinase	VACV-Cop-F16L
EPTV-034	26423-26737	104	DNA-binding phosphoprotein (VP11)	VACV-Cop-F17R
EPTV-035	28149-26731	472	Poly (A) polymerase catalytic subunit (VP55)	VACV-Cop-E1L
EPTV-036	30364-28166	732	IEV morphogenesis	VACV-Cop-E2L
EPTV-037	31138-30422	238	dsRNA-binding PKR inhibitor (host range)	VACV-Cop-E3L
EPTV-038	31903-31175	242	RNA polymerase subunit (RPO30)	VACV-Cop-E4L
EPTV-039	32072-33775	567	IMV protein, virion morphogenesis	VACV-Cop-E6R
EPTV-040	33799-34611	270	ER-localized membrane protein, virion core protein	VACV-Cop-E8R
EPTV-041	37628-34608	1006	DNA polymerase	VACV-Cop-E9L
EPTV-042	37661-37951	96	Sulfhydryl oxidase (FAD-linked)	VACV-Cop-E10R
EPTV-043	38365-37943	140	Virion core protein	VACV-Cop-E11L
EPTV-044	40424-38349	691	Virulence, modulates Raf/MEK/ERK pathway	VACV-Cop-O1L
EPTV-045	40795-40481	104	Nonessential glutaredoxin	VACV-Cop-O2L
EPTV-046	41842-40910	310	DNA-binding core protein	VACV-Cop-I1L
EPTV-047	42064-41843	73	IMV membrane protein	VACV-Cop-I2L
EPTV-048	42877-42065	270	ssDNA-binding phosphoprotein	VACV-Cop-I3L
EPTV-049	45218-42930	762	Ribonucleotide reductase large subunit	VACV-Cop-I4L
EPTV-050	45491-45255	78	IMV protein (VP13)	VACV-Cop-I5L
EPTV-051	46654-45509	381	Telomere-binding protein	VACV-Cop-I6L
EPTV-052	47933-46647	428	Viral core cysteine protease	VACV-Cop-I7L
EPTV-053	47939-49978	679	RNA-helicase, DExH-NPH-II	VACV-Cop-I8R
EPTV-054	51755-49965	596	Insulin metalloproteinase-like protein	VACV-Cop-G1L
EPTV-055	52084-51752	110	Entry/fusion complex component	VACV-Cop-G3L
EPTV-056	52078-52746	222	Late transcription elongation factor (VLTF)	VACV-Cop-G2R
EPTV-057	53090-52713	125	Thioredoxin-like protein	VACV-Cop-G4L
EPTV-058	53093-54409	438	FEN1-like nuclease	VACV-Cop-G5R
EPTV-059	54409-54600	63	RNA polymerase subunit (RPO7)	VACV-Cop-G5.5R
EPTV-060	54604-55143	179	NLPc/P60 superfamily protein	VACV-Cop-G6R
EPTV-061	56212-55109	367	Virion structural phosphoprotein, early morphogenesis	VACV-Cop-G7L
EPTV-062	56241-57023	260	Late transcription factor (VLTF-1)	VACV-Cop-G8R
EPTV-063	57039-58064	341	Myristylated entry/fusion protein	VACV-Cop-G9R
EPTV-064	58065-58814	249	Myristylated IMV envelope protein	VACV-Cop-L1R
EPTV-065	58847-59116	89	Crescent membrane/immature virion protein	VACV-Cop-L2R
EPTV-066	60073-59108	321	Internal virion protein	VACV-Cop-L3L
EPTV-067	60098-60856	252	DNA-binding virion protein (VP8)	VACV-Cop-L4R
EPTV-068	60871-61275	134	IMV protein, entry/fusion	VACV-Cop-L5R
EPTV-069	61217-61663	148	IMV membrane protein, virion morphogenesis	VACV-Cop-J1R
EPTV-070	61689-62219	176	Thymidine kinase	VACV-Cop-J2R
EPTV-071	62305-62916	203	Type I IFN inhibitor (host range)	VACV-Cop-C7L
EPTV-072	62991-63992	333	Poly (A) polymerase small subunit (VP39)	VACV-Cop-J3R
EPTV-073	63907-64464	185	RNA polymerase subunit (RPO22)	VACV-Cop-J4R
EPTV-074	64873-64466	135	IMV membrane protein, entry/fusion	VACV-Cop-J5L
EPTV-075	64979-68836	1285	RNA polymerase subunit (RPO147)	VACV-Cop-J6R
EPTV-076	69351-68833	172	Tyr/Ser kinase, virus assembly, IFN-gamma inhibitor	VACV-Cop-H1L
EPTV-077	69365-69937	190	Entry/fusion IMV protein	VACV-Cop-H2R
EPTV-078	70957-69944	337	IMV heparin-binding surface protein (p35)	VACV-Cop-H3L
EPTV-079	73348-70961	795	RNA polymerase-associated protein (RAP94)	VACV-Cop-H4L
EPTV-080	73557-74210	217	Late transcription factor (VLTF-4)	VACV-Cop-H5R
EPTV-081	74229-75170	313	DNA topoisomerase type I	VACV-Cop-H6R
EPTV-082	75202-75651	149	Crescent membrane/immature virion protein	VACV-Cop-H7R
EPTV-083	75694-78228	844	mRNA capping enzyme large subunit	VACV-Cop-D1R
EPTV-084	78633-78190	147	Virion core protein	VACV-Cop-D2L
EPTV-085	78632-79375	247	Virion core protein	VACV-Cop-D3R
EPTV-086	79372-80028	218	Uracil DNA glycosylase, DNA pol processivity factor	VACV-Cop-D4R
EPTV-087	80062-82425	787	NTPase, DNA primase	VACV-Cop-D5R
EPTV-088	82422-84329	635	Early transcription factor small subunit (VETF-s)	VACV-Cop-D6R
EPTV-089	84363-84875	170	RNA polymerase subunit (RPO18)	VACV-Cop-D7R
EPTV-090	85682-84810	290	Carbonic anhydrase, GAG-binding MV membrane protein	VACV-Cop-D8L
EPTV-091	85740-86408	222	mRNA decapping enzyme	VACV-Cop-D9R

EPTV-092	86386-87171	261	mRNA decapping enzyme	VACV-Cop-D10R
EPTV-093	89052-87145	635	ATPase, NPH1	VACV-Cop-D11L
EPTV-094	89948-89085	287	mRNA capping enzyme small subunit	VACV-Cop-D12L
EPTV-095	91661-89982	559	Trimeric virion coat protein (rifampicin resistance)	VACV-Cop-D13L
EPTV-096	92142-91687	151	Late transcription factor (VLTF-2)	VACV-Cop-A1L
EPTV-097	92847-92173	224	Late transcription factor (VLTF-3)	VACV-Cop-A2L
EPTV-098	93074-92844	76	S-S bond formation pathway protein	VACV-Cop-A2.5L
EPTV-099	95085-93094	663	P4b precursor	VACV-Cop-A3L
EPTV-100	95713-95141	190	Putative membrane-associated virion core protein (p39)	VACV-Cop-A4L
EPTV-101	95751-96260	169	RNA polymerase subunit (RPO19)	VACV-Cop-A5R
EPTV-102	97375-96257	372	Virion morphogenesis core protein	VACV-Cop-A6L
EPTV-103	99543-97399	714	Early transcription factor large subunit (VETF-L)	VACV-Cop-A7L
EPTV-104	99610-100485	291	Intermediate transcription factor (VITF-3s)	VACV-Cop-A8R
EPTV-105	100724-100482	80	IMV membrane protein, early morphogenesis	VACV-Cop-A9L
EPTV-106	103451-100725	908	P4a precursor	VACV-Cop-A10L
EPTV-107	103466-104401	311	Viral membrane formation	VACV-Cop-A11R
EPTV-108	104910-104398	170	Virion core and cleavage processing protein	VACV-Cop-A12L
EPTV-109	105224-105003	73	IMV membrane protein, virion maturation	VACV-Cop-A13L
EPTV-110	105572-105291	93	IMV membrane protein, essential	VACV-Cop-A14L
EPTV-111	105750-105589	53	IMV membrane protein, non-essential	VACV-Cop-A14.5L
EPTV-112	106033-105740	97	Core protein	VACV-Cop-A15L
EPTV-113	107159-106017	380	Myristylated protein, essential for entry/fusion	VACV-Cop-A16L
EPTV-114	107747-107160	195	IMV membrane protein	VACV-Cop-A17L
EPTV-115	107762-109201	479	DNA helicase, transcript release factor	VACV-Cop-A18R
EPTV-116	109409-109191	72	Zn finger-like protein, late morphogenesis	VACV-Cop-A19L
EPTV-117	109754-109410	114	IMV membrane protein, entry/fusion	VACV-Cop-A21L
EPTV-118	109753-111030	425	DNA polymerase processivity factor	VACV-Cop-A20R
EPTV-119	111014-111565	183	Holliday junction resolvase	VACV-Cop-A22R
EPTV-120	111562-112722	386	Intermediate transcription factor (VITF-3L)	VACV-Cop-A23R
EPTV-121	112719-116219	1166	RNA polymerase subunit (RPO132)	VACV-Cop-A24R
EPTV-122	119230-116222	1002	A-type inclusion protein	VACV-Cop-A25L
EPTV-123	121205-119274	643	P4c precursor	VACV-Cop-A26L
EPTV-124	121605-121261	114	IMV membrane protein, fusion	VACV-Cop-A27L
EPTV-125	122022-121606	138	IMV membrane protein, entry	VACV-Cop-A28L
EPTV-126	122938-122036	300	RNA polymerase subunit (RPO35)	VACV-Cop-A29L
EPTV-127	123149-122922	75	IMV protein	VACV-Cop-A30L
EPTV-128	123355-123855	166	Hypothetical protein	VACV-Cop-A31R
EPTV-129	124635-123871	254	ATPase/DNA packaging protein	VACV-Cop-A32L
EPTV-130	124771-125334	187	C-type lectin-like EEV membrane phosphoglycoprotein	VACV-Cop-A33R
EPTV-131	125386-125886	166	C-type lectin like IEV/EEV membrane glycoprotein	VACV-Cop-A34R
EPTV-132	125923-126441	172	MHC class II antigen presentation inhibitor	VACV-Cop-A35R
EPTV-133	126481-127326	281	Concanavalin-like precursor	DPV-84-135
EPTV-134	127368-128099	243	EEV glycoprotein	DPV-84-136
EPTV-135	128142-128972	276	Hypothetical protein	VACV-Cop-A37R
EPTV-136	128996-129259	87	Hypothetical protein	Unique to EPTV
EPTV-137	129840-129256	194	Truncated CD47-like protein, integral membrane protein	VACV-Cop-A38L
EPTV-138	129868-130266	132	Myristylated protein	VACV-Cop-E7R
EPTV-139	130325-131077	250	Hypothetical protein	SQPV-1300
EPTV-140	131927-131070	285	Chemokine-binding protein	VACV-Cop-A41L
EPTV-141	132054-132455	133	Profilin-like protein, ATI-localized	VACV-Cop-A42R
EPTV-142	132843-132457	128	Hypothetical protein	Unique to EPTV
EPTV-143	132917-133141	74	Hypothetical protein	Unique to EPTV
EPTV-144	133348-134403	351	3 β-hydroxysteroid dehydrogenase/δ 5→4 isomerase	VACV-Cop-A44L
EPTV-145	134449-134673	74	Hypothetical protein	Unique to EPTV
EPTV-146	135385-134732	217	Immunoprevalent protein	VACV-Cop-A47L
EPTV-147	135466-136053	195	Thymidylate kinase	VACV-Cop-A48R
EPTV-148	136085-137770	561	DNA ligase-like protein	VACV-Cop-A50R
EPTV-149	137817-139394	525	BTB kelch-domain protein	VACV-Cop-A55R
EPTV-150	139422-140048	208	A52R-like family protein	DPV-84-148
EPTV-151	140099-140596	165	Hypothetical protein	Unique to EPTV
EPTV-152	140648-141679	343	Hypothetical protein	VACV-Cop-A51R
EPTV-153	141743-142393	216	Toll/IL-1 receptor-like protein, NFkB signaling inhibitor	VACV-Cop-A52R
EPTV-154	142985-142518	155	Hypothetical protein	Unique to EPTV

EPTV-155	143074-144687	537	BTB kelch-domain protein	VACV-Cop-A55R
EPTV-156	144715-145509	264	Hemagglutinin	VACV-Cop-A56R
EPTV-157	145529-146464	311	Ser/Thr protein kinase	VACV-Cop-B1R
EPTV-158	146495-147493	332	IL-1 receptor antagonist	VACV-Cop-C10L
EPTV-159	147510-148409	299	KilA-N/RING finger protein (host range)	CPXV-BR-023
EPTV-160	148441-149031	196	Partial schlafen-like protein	VACV-Cop-B2R
EPTV-161	149126-149839	237	EEV type-1 membrane glycoprotein (host range)	VACV-Cop-B5R
EPTV-162	149954-150385	143	Anti-apoptotic Bcl-2-like protein	VACV-Cop-N1L
EPTV-163	150436-151284	282	dsRNA binding PKR inhibitor	VACV-Cop-E3L
EPTV-164	151349-152353	334	Serpin 1 (host range)	VACV-Cop-C12L
EPTV-165	152461-152919	152	Hypothetical protein	VACV-Cop-B15R
EPTV-166	152947-153822	291	Tyrosine protein kinase-like protein	DPV-84-158
EPTV-167	153877-154884	335	IL-1 beta-receptor	VACV-Cop-B16R
EPTV-168	154911-156857	648	Ankyrin repeat protein	DPV-84-019
EPTV-169	156942-157568	208	Ankyrin repeat protein	DPV-84-014
EPTV-170	157626-158333	235	Alpha-amanitin target protein	VACV-Cop-N2L
EPTV-171	158383-159060	225	NFkB inhibitor	VACV-Cop-M2L
EPTV-172	159108-159341	77	Endothelin precursor	DPV-84-006
EPTV-173	159392-160078	228	NFkB inhibitor	VACV-Cop-M2L
EPTV-174	160085-160876	263	Secreted complement binding protein (host range)	VACV-Cop-C3L
EPTV-175	160944-161372	142	IL-18 binding protein	DPV-84-021
EPTV-176	161449-161637	62	Hypothetical protein	Unique to EPTV
EPTV-177	161603-162202	199	Truncated TNFa-receptor (host range)	VACV-Cop-B28R
EPTV-178	162245-163201	318	MHC class I-like protein	SQPV-0040
EPTV-179	163239-165053	604	Ankyrin repeat protein	DPV-84-019
EPTV-180	165090-165599	169	IFN resistance, eIF2a-like PKR inhibitor (host range)	VACV-Cop-K3L
EPTV-181	165649-167478	609	Ankyrin repeat protein	DPV-84-019
EPTV-182	167993-169726	577	Ankyrin repeat protein (host range)	CPXV-BR-025
EPTV-183	169885-170364	159	Hypothetical protein	VACV-Cop-B15R
EPTV-184	170444-170920	158	Hypothetical protein	Unique to EPTV
EPTV-185	171006-171698	230	ER-localized apoptosis regulator (host range)	VACV-Cop-B9R
EPTV-186	171755-172660	301	Tyrosine protein kinase-like protein	DPV-84-158
EPTV-187	172700-173179	159	Hypothetical protein	DPV-84-159
EPTV-188	173217-174215	332	IL-1 receptor-like protein	DPV-84-015
EPTV-189	174279-174962	227	Hypothetical protein	DPV-84-009
EPTV-190	174995-175957	320	Serpin 2/CrmA (host range)	VACV-Cop-B13R
EPTV-191	176069-176548	159	Hypothetical protein	VACV-Cop-B15R

3.3.3.2 Unique EPTV genes

Most of the EPTV putative genes have >75 % aa identity with other poxviral orthologs, however, some have as little as < 30 % aa identity with the designated orthologous family. Therefore, instead of choosing an arbitrary cut-off for assigning orthologs, ORFs were evaluated by % identity, % similarity (allowing chemically similar aa to be matched), protein motifs, secondary structure prediction, synteny and gene size. Consequently, eleven EPTV genes: -008 (ITR: -184), -021, -024, -031, -136, -142, -143, -145, -151, -154, -176 could not be matched confidently with any poxvirus counterpart, and are annotated as unique hypothetical proteins. However, several of these genes are located in positions in the genome

where orthologs are missing when compared to the arrangement of genes in VACV-Cop (**Table 6**). Although it is possible that the EPTV orthologs have diverged to such an extent that no recognisable sequence patterns are discernable, it is also possible that genes have been replaced in EPTV. Interestingly, none of this set of EPTV sequences have any significant similarity to any of the sequences in the bat genome database[273], as well as the current databases when analyzed by the BLAST suite of programs[274]. A further question concerning the annotation of these ORFs is whether they are likely to be functional genes. The size of these unique ORFs range from 62-165 codons, five being in the set of twelve smallest ORFs annotated in EPTV (**Table 6**). This, together with the fact that several of these predicted proteins have isoelectric points in the extreme tails (low and high) of the distribution observed for known poxvirus proteins (not shown) suggests that some of these annotated ORFs may not represent functional genes[176]. Others may provide EPTV with functions specific to promotion of virus replication within its host.

3.3.3.3 Relationship with Clade II poxviruses

According to the ChPV phylogenetic tree, EPTV is branched off the common backbone as a sister group to the Clade II viruses (**Figure 4**). As shown in **Table 6**, EPTV possesses 7 genes that appear to only have Clade II orthologs, however, some of these genes have one or more paralogs with different distributions among the various viruses that complicate the assignment of divergent genes to particular ortholog families. The EPTV genes -004 (-188; ITR), -018, -150 and -169 are predicted to have functions associated with virulence/host range (**Table 6**). EPTV genes -003 (-189; ITR), -005 (-187; ITR) have no known function, but are conserved in DPV and are 227 and 159 codons long, respectively; this suggests that they are likely functional genes in EPTV. EPTV-134 is unusual in that it is predicted to encode a Clade II-specific structural protein that spans the outer membrane of intracellular

enveloped virus (IEV) particles[275]. This Clade II specific IEV protein is located at the same position as orthopoxvirus ortholog A36R, which also encodes an IEV protein. Experimental evidence shows these Clade II IEV orthologs (DPV-84-136) as functional orthologs of A36R, despite very low sequence identity, also induce actin tail formation, albeit via a different mechanism[276]. Using ScanSite3[277], EPTV-134 is predicted to have five Nck-binding tyrosine motifs and 1 Grb2 binding motif, thus it has the same signatures as DPV-84-136[276] and is therefore likely to represent an ortholog that is otherwise Clade II specific. Further support for the closer association of EPTV with Clade II viruses is that EPTV share a unique rearrangement (exclusive to Clade II members) of the C7L gene (EPTV-071; Type I IFN inhibitor) and E7R gene (EPTV-138; a myristylated protein), which is different from the synteny observed in the rest of the ChPVs[24].

3.3.3.4 Relationship with orthopoxviruses

EPTV also possesses several genes that are otherwise OPV specific. EPTV-010 (ITR: -182) is an ortholog of CPXV-BR-025, which encodes an ankyrin-like protein (577 aa). It is curious that the neighbouring EPTV-011 gene also encodes an ankyrin-like protein (609 aa), which is an ortholog of DPV-84-019 that is associated with Clade II, avipox-, and parapoxviruses. Despite the physical proximity of these 2 genes, the aligned predicted protein sequences are only 23 % identical. In addition, EPTV-010 and EPTV-011 have a different number of ankyrin repeats, 3 and 9, respectively. Although it is assumed that gene duplication events have created families of paralogous genes in poxviruses, at this level of sequence diversity it is very difficult to differentiate this from the possibility that some may be the result of separate gene capture events. COTV is the only other Clade II ChPV (to date) that is believed to have a mixture of ankyrin orthologs from both OPVs and Clade II viruses[24]. Together with COTV, EPTV also encodes a VACV-C3L homolog (secreted

complement binding protein) that is otherwise exclusive to the OPVs as well. Additional EPTV genes that are otherwise OPV-specific are scattered throughout the genome: EPTV-045, -141, -146 and -167 are orthologs of VACV-O2L, -A42R, -A47L and -B16R, and encode a glutaredoxin-like protein, a profilin-like protein, the immunoprevalent protein, and an IL-1 beta-receptor-like protein, respectively. EPTV also encodes a chemokine-binding protein (EPTV-140) as well as a partial schlafen-like protein (EPTV-160) that have only been found in OPVs, YKV, and the recently sequenced PTPV.

3.3.3.5 Variably present genes

Above, we tried to categorize EPTV genes by which other viruses have orthologs, however, when the presence of a particular gene is inconsistent through different genera it is also instructive to determine those genera that do not contain an ortholog. For example, the EPTV-122, -123 and -141 proteins that create or localize to an A-type inclusion body are absent from the Clade II viruses. Similarly, EPTV encodes a ribonucleotide reductase large subunit (I4L) that is missing from all Clade II viruses except swinepox (SWPV). In contrast, EPTV-006 (-185), -011 (-181), -016, -133, -166, -168, -175, and -179, are absent from the OPVs, but variably present in the Clade II viruses and other ChPVs (**Table 6**). This mosaic nature of EPTV genes makes the elucidation of evolutionary events interesting but more difficult.

3.3.3.6 A link between diverged F5L ortholog families

Overall, the genes encoded by region between F4L and F9L is also highly diverse between ChPV species from different genera. From **Table 6**, it is seen that the ChPV synteny between F4L to F9L is disrupted by 2 EPTV unique HPs. Additionally, we draw attention to EPTV-020, a putative F5L membrane protein showing global similarity with VACV-F5L

(membrane protein) and MOCV/SQPV-003 (26-29% aa ID with functional aa conserved), but peaks in local identity with Clade II DPV-84-027 ortholog group at the C-terminus. The 3 groups of protein showed no discernible similarity with each other prior to the discovery of EPTV-020 sequence (elaborated below).

Discovery of EPTV-020 revealed that MOCV003L/SQPV003 and Clade II DPV-84-027 ortholog family are divergent F5L proteins, although previously annotated as unrelated proteins with unknown function. Previously, the MOCV/SQPV orthologs had only 17% aa identity with VACV-F5L, whereas the it has up to 29% aa identity with EPTV-020 F5L ortholog. With EPTV-020, the Clade II orthologs also improved the aa identity to 39.5 % at their C-terminal region, as opposed to just 22 % aa identity with OPVs. This suggests that these F5L-positioned Clade II orthologs (DPV-84-027) are likely N-terminus truncated versions of F5L orthologs with subsequent divergence. This supports the 100% bootstrap that groups the EPTV with Clade II viruses. Furthermore, this relationship, uncovered by analysis of the EPTV-020 sequence, supports the idea that the Clade II gene at the F5L position evolved from the same ancestral F5L rather than an unknown sequence acquired by an independent gene capture event.

3.3.3.7 Other genes of interests

EPTV-172 encodes an endothelin-like polypeptide that is otherwise only found in DPV thus far. Eukaryotic endothelin is part of a multi-gene family that, along with at least 4 receptors, acts as vaso-contrictors[278]. However, as noted in the DPV genome paper[181], it has yet to be experimentally determined whether this viral ortholog acts in a manner similar to the endothelins or as an antagonist. Blocking endothelin function by binding a receptor without initiating signalling might be favorable to the virus due to a reduction of inflammation[279].

EPTV-139 is predicted to encode a cysteine-rich protein that has putative similarity to only 1 other poxvirus protein, SQPV-130. SQPV-130 is a hypothetical protein of unknown function. It is curious to see an AT-rich poxvirus, like EPTV, share a gene exclusively with a GC-rich virus like SQPV. Although these poxvirus proteins have only 23 % aa identity, the residues were conserved globally across the 250aa alignment including 10 functionally important cysteines. This is observed despite the fact that these EPTV and SQPV genes have an AT nucleotide composition of 82% and 30 %, respectively. The proteins are not predicted to have N-terminal signal sequences, but since poxviruses encode proteins that create S-S bonds within the cytoplasm[280] , these proteins may still be structurally dependent on these cysteines. No tertiary structure similarities have been predicted between these poxviral proteins and other proteins.

Similarly, the EPTV-178 protein is a member of the MHC class I family, despite the low (25%) aa identity, by virtue of the conservation of a characteristic pattern of cysteine residues. Several other poxviruses have similar genes, but it appears that at least some have been independently acquired, since EPTV-178 and SQPV-004 genes score better in BLASTP searches against microbat and squirrel MHC class I proteins, respectively, rather than with each other. COTV also has an independently acquired MHC class I protein most similar to a homolog in a Southern/Central American wild cat[24].

3.3.4 Discussion

The reconstructed phylogeny shows that, unlike what was previously suggested in the initial report[219], EPTV does not form a sister clade with COTV. Instead, it branches off the common backbone from the rest of the AT % rich chordopoxviruses and forms its own genus as a

sister clade to the Clade II poxviruses. The term "Clade II" arose to distinguish a clade of viruses (capri-, sui-, cervid-, yata-, leporipoxviruses, and COTV) from the OPVs which diverged at their last common node[17]. However, given the phylogenetic position of EPTV and its mosaic collection of orthologous Clade II and OPV genes, the term may have outlived its usefulness. EPTV along with the recently sequenced PTPV[266], demonstrate the value of sequencing further poxvirus genomes to increase the information in phylogenetic trees.

In terms of genome statistics, the 176,688 nt genome size for EPTV is considered medium length among the ChPVs. It is predicted to encode 191 genes and has one of the larger ITRs sequenced (>10,000 nt). It should be noted that EPTV and COTV genomes are the most AT % rich of the ChPVs genomes (76.4 %). The reduced synteny, including gene loss, at the right end of the EPTV genome compared to VACV-Cop reflects the fact that many VACV genes in this region may be non-essential. Three genes suspected to be the result of HGT have been found in this region (EPTV-139, -172, and -178), two of them are predicted to encode an endothelin-like gene and a MHC class I protein with potential immunomodulatory roles (**Table 6**). Further examples of genomic variability towards the terminal regions of the EPTV genome include the presence of 3 and 8 novel EPTV ORFs at the left and right ends, respectively. Some of these "unique" EPTV sequences may encode new virulence factors.

With regard to the potential for cross-species infections, EPTV possesses 11 out of 12 sets of host range factors previously reviewed (missing K1L, which is only found in OPV)[160,281]. Of these 11, unique sequence extensions and truncations are found associated with EPTV orthologs of PKR inhibitors (K3L and E3L), KilA-N/RING domain protein (p28/N1R) involved in ubiquitin or apoptosis inhibition, and tumor necrosis factor receptor family 2 (B28R). It is unknown how these altered sequences may impact the ability of the encoded

protein to avert different host immune responses. For example, SPPV with a deletion of the kelch-like protein (SPPV-019) restored 100% survival rate in infected sheep, but the same deletion failed to reduce virulence in a VACV model[282]. Therefore, the mere possession of host range proteins alone cannot serve as a marker or indicator for cross-species infection without considering the diversity of viral sequences and host immune systems, as well as any combinatorial effect of gene sets. However, we can foresee some beneficial role for EPTV in having either an increased virulence that promotes infections, or an attenuation that could contribute to the host serving as a reservoir of different viruses. The latter seems to align with the fact that majority of bat-borne viruses exhibit non-severe nor fatal symptoms in their reservoir hosts[234].

Lastly, we found no evidence of similarities between the genes shared by the bat-borne EPTV and PTPV that suggest common virulence mechanisms in bat hosts. Consistent with this theme, the clinical symptoms of EPTV in *Eptesicus fuscus* manifest in the form of joint swelling and increased lethargy[219], whereas *Pteropus scapulatus* infected with PTPV presented lesions on wing membranes[266]; neither types of symptom were directly linked to fatality. Thus, it appears that these "bat-isolated poxviruses" do not have any common genes based on their related hosts.

3.4 Kangaroopox viruses

Complete genomic characterisation of two novel poxviruses (WKPV and EKPV) from western and eastern grey kangaroos

Mark Bennett[1*], Shin-Lin Tu[2*], Chris Upton[2], Cassie McArtora[1], Amber Gillett[3], Tanya Laird[1], Mark O'Dea[1]

1. School of Veterinary and Life Sciences, Murdoch University, Perth, Western Australia, Australia

2. Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC, Canada
3. Australia Zoo Wildlife Hospital, Beerwah, Queensland, Australia

* Co-first authors: these authors contributed equally to this work.

The work contained in this chapter constitutes the manuscript submitted to and accepted by Virus Research. Permission to use the copyrighted material is provided in **Appendix 4**.

MB and MO conceived the study. MB and AG carried out necropsy, histopathology, electron microscopy and experimental work. CM carried out fieldwork, necropsy and manuscript preparation. MO and TL carried out experimental work. MO and SLT performed genome assembly; SLT finalized genome sequene. SLT annotated the genes and performed sequence, structural, and phylogeny analyses. MO and MB provided veterinary context and implications. MO, MB, SLT, and CU wrote the manuscript.

3.4.1 Abstract

Poxviruses have previously been detected in macropods with cutaneous papillomatous lesions, however to date, no comprehensive analysis of a poxvirus from kangaroos has been performed. Here we report the genome sequences of a Western grey kangaroo poxvirus (WKPV) and an Eastern grey kangaroo poxvirus (EKPV), named for the host species from which they were isolated, western grey (*Macropus fuliginosus*) and eastern grey (*Macropus giganteus*) kangaroos. Poxvirus DNA from WKPV and EKPV was isolated and entire coding genome regions determined through Roche GS Junior and Illumina Miseq sequencing, respectively. Viral genomes were assembled using MIRA and SPAdes, and annotations performed using tools available from the Viral Bioinformatics Resource Centre. Histopathology and transmission electron microscopy analysis was also performed on WKPV and its associated lesions. The WKPV and EKPV genomes show 96% identity (nt) to each other and phylogenetic analysis places them on a distinct sister branch to the established *Avipoxvirus* genera. WKPV and EKPV are 170 kbp and 167 kbp long, containing 165 and 162 putative genes, respectively. Together, their genomes encode up to 47 novel unique

hypothetical proteins, and possess virulence proteins including a MHC class II inhibitor, a semaphorin-like protein, a serpin, a 3-β-hydroxysteroid dehydrogenase/δ5→4 isomerase, and a CD200-like protein. These viruses also encode a large putative protein (WKPV-WA-039 and EKPV-SC-038) with a C-terminal domain that is structurally similar to the C-terminal domain of a cullin, suggestive of a role in the control of host ubiquitination. The relationship of these viruses to members of the *Molluscipoxvirus* and *Avipoxvirus* genera is discussed in terms of sequence similarity, gene content and nucleotide composition. A novel genus within subfamily *Chordopoxvirinae* is proposed to accommodate these two poxvirus species from kangaroos; we suggest the name, Thylacopoxvirus (thylaco-: [Gr.] *thylakos*meaning sac or pouch).

3.4.2 Background

Western grey kangaroos (*Macropus fuliginosus*) and eastern grey kangaroos (*Macropus giganteus*) are Australian macropodid marsupials classified within the 'least concern' category of the IUCN (International Union for Conservation of Nature) red list of threatened species. *M. fuliginosus* inhabits a wide swathe of southern Australia, from Western Australian coastal regions to south-western Queensland, western New South Wales and Victoria. *M. giganteus* is endemic to eastern mainland Australia (Queensland, New South Wales, Victoria, and eastern South Australia), and a subspecies, *M. g. tasmaniensis* is found in eastern Tasmania. *M. fuliginosus* and *M. giganteus* are thought to have diverged from a common macropodid ancestor during the Pliocene, approximately 2.5 ± 1.5 million years ago[220]. There have not been any documented reports of human poxvirus lesions following contact with macropod skin lesions, and currently there is no suspicion that these poxviruses infect humans or other animals. However, the enormous geographic range inhabited by these

animals would suggest that they at least come into contact with a large assortment of other animals and are therefore of interest as potential agents of cross-species infections[46].

Cutaneous poxvirus infections of *M. fuliginosus* and *M. giganteus* have been previously reported in the scientific literature[283,284]. Clinically, solitary or multiple wart-like, raised cutaneous lesions are seen on the extremities, especially the face, feet and tail. These lesions are thought to spontaneously regress after approximately three months, leaving a hairless scar. A range of other macropodid marsupials can also manifest clinically similar lesions, including *Macropus rufus*, *Macropus robustus*, *Macropus eugenii*, *Macropus agilis*, *Wallabia bicolor*, and *Setonix brachyurus*. Prior to this study, of the poxviruses found in Australian mammals, only Pteropox virus has had the entire coding region of its genome sequenced. Despite poxvirus-associated lesions being relatively common in macropods, there are no complete characterisations of the associated viruses.

In May 2013, a western grey kangaroo joey was a roadkill specimen found in the vicinity of Greenmount State Forest near Perth, Western Australia. The carcass was taken to a nearby veterinary hospital where it was stored frozen (−20 °C), until being used in a wildlife necropsy practical class for the Murdoch University Wildlife Association. The western grey kangaroo joey carcass presented with a focally extensive, well-demarcated, approximately rectangular, raised, firm area of thickened skin extending from just rostral to the medial canthi bilaterally, across the dorsal aspect of the muzzle, and rostrally to involve approximately half the length of the dorsal muzzle. Representative samples were collected from the skin mass on the dorsal muzzle.

In April 2010, an orphaned 9 kg, juvenile, female eastern grey kangaroo was found on the campus of the University of the Sunshine Coast, Sippy Downs, Queensland. The kangaroo was euthanised due to severe wounds likely caused by a dog attack, and a fresh frozen sample of papillomatous skin was sent to Murdoch University for sequencing purposes.

For the western grey kangaroo sample, an unfixed sample of lesional skin was finely minced using sterile scalpel blades and total DNA was extracted using a DNeasy tissue kit (Qiagen). A DNA library was prepared using the Roche Rapid Library kit, and sequencing was carried out on a GS Junior sequencer using GSJR titanium sequencing kit and GSJR Pico titre plate.

For the eastern grey kangaroo skin lesion, DNA was extracted using a Qiagen DNEasy Blood and Tissue kit (Qiagen). Library preparation was undertaken using a Nextera XT DNA Library Prep Kit (Illumina). The library was sequenced on an Illumina MiSeq V2 2 × 300 flowcell using standard Illumina MiSeq protocols.

3.4.3 Results

3.4.3.1 Genome organization and annotation of the KPV genomes

The genome of western grey kangaroo poxvirus WKPV-WA was assembled into a 170,141 bp sequence, including the two ITRs of 2984 bp, and the EKPV-SC genome was assembled into a 167,359 bp sequence including the 1481 bp ITRs. Sequence data for KPV species WKPV and EKPV were obtained in 2013 and 2017, respectively. Roche 454 sequence data were obtained for WKPV species and was efficiently assembled using MIRA assembler without the need to filter against host reads. Pair-end Illumina sequence data were obtained for EKPV. Different assembly trials used reads filtered against *Macropus eugenii* genomic sequence (INSDC: ABQO000000000.2; the closest available marsupial sequence to

M. giganteus), and/or human and bacterial contaminant sequences, that were additionally followed by duplicate read removal using FastUniq[193] and assembled using SPAdes[194] under the "careful" parameter. Slightly varied contigs were obtained from these different trials, and discrepancies were manually compared and validated against raw read data. PCR had been used to confirm a repeat sequence at region 50,452-50,823 in the final extended EKPV sequence. Among the ChPVs, which have genomes that range from 133 to 360 kbp, the KPVs can be considered in the low-mid size range. The KPVs have 46% AT base composition, which is currently midway between MYXV (56% AT) and CRV (38% AT) among all the sequenced ChPV genomes.

An alignment of the unique regions of the KPV genomes (ITRs excluded) showed that they are 96% identical (nucleotide) to each other. With this degree of nucleotide identity, WKPV and EKPV should be considered separate poxvirus species of the same genus. In a similar comparison, SPPV and GTPV, which belong to genus *Capripoxvirus*, are 97% identical at the nucleotide level. Annotation results are presented in **Table 7**.

Table 7. Summary of WKPV and EKPV genome annotations

Percentage aa identity between species is listed between gene numbers; fragmented regions with small ORFs are boxed in grey. Genes are designated with either the standard VACV-Cop ortholog reference, or if the ortholog is missing in VACV-Cop, then it is designated by the closest ortholog; genes that were previously identified as genus-specific orthologs are bolded; asterisks mark genes repeated at the ITR regions.

WKPV	Position	AA size	% aa ID	EKPV	Position	AA size	Putative Gene	Reference Ortholog Family
001*	508-1134	208	/	/	/	/	Hypothetical protein	Unique
002*	1165-1623	152	/	/	/	/	Hypothetical protein	Unique
003*	1832-2596	254	/	/	/	/	Hypothetical protein	Unique
004	3743-2985	252	95	001	2240-1482	252	Hypothetical protein	Unique
005	5074-3809	421	94	002	3577-2324	417	Hypothetical protein	Unique
006	5173-6033	286	97	003	3676-4536	286	Hypothetical protein	Unique
007	7211-6039	390	85	004	5764-4541	407	Hypothetical protein	Unique
008	7811-7257	184	80	005	6352-5810	180	Hypothetical protein	Unique
009	9401-7887	504	72	006	7881-6427	484	Hypothetical protein	Unique
010	11343-9526	605	93	007	9814-8006	602	Semaphorin	VACV-Cop-A39R
011	11732-11424	102	87	008	10229-9921	102	Hypothetical protein	Unique

012	12821-11841	326	93	009	11335-10340	331	Hypothetical protein	Unique
013	13514-12933	193	96	010	12032-11451	193	Hypothetical protein	Unique
014	14780-13596	394	95	011	13304-12114	396	Hypothetical protein	Unique
015	15377-15258	39	65	012	13968-13321	215	Hypothetical protein	Unique
016	16659-15451	402	97	013	15261-14053	402	Hypothetical protein	Unique
017	17207-16716	163	96	014	15838-15347	163	Hypothetical protein	Unique
018	17575-17252	107	94	015	16206-15883	107	Hypothetical protein	Unique
Fragmented & Frameshifted			/	016	16338-16475	45	Hypothetical protein	Unique
019	18520-18266	84	93	017	17138-16962	58	Hypothetical protein	Unique
020	18978-19256	92	78	018	17597-17830	77	Hypothetical protein	Unique
Fragmented & Frameshifted			/	019	18143-17931	70	Hypothetical protein	Unique
021	19877-19524	117	/	Fragmented & Frameshifted			Hypothetical protein	Unique
022	20771-20079	230	95	020	19464-18772	230	Hypothetical protein	Unique
023	20846-21445	199	82	021	19541-19945	134	Hypothetical protein	Unique
			92	022	19964-20140	58		
024	22351-22139	70	100	023	21015-20803	70	Hypothetical protein	**CNPV-VR111-140**
025	22911-22444	155	99	024	21575-21108	155	Hypothetical protein	**MOCV-st1-015L**
026	23568-22912	218	96	025	22232-21576	218	S-S bond formation pathway protein substrate	VACV-Cop-F9L
027	24868-23537	443	99	026	23532-22201	443	Essential Ser/Thr kinase	VACV-Cop-F10L
028	27458-25083	791	95	027	26077-23705	790	RhoA signalling inhibitor, virus release protein	VACV-Cop-F11L
029	29515-27563	650	98	028	28137-26185	650	EEV maturation protein	VACV-Cop-F12L
030	29868-29563	101	85	029	28490-28185	101	Hypothetical protein	Unique
031	31125-29935	396	98	030	29750-28557	397	Palmitylated EEV membrane glycoprotein	VACV-Cop-F13L
032	31859-31128	243	96	031	30484-29753	243	Hypothetical protein	**MOCV-st1-022L**
033	32324-31983	113	96	032	30954-30613	113	Hypothetical protein	Unique
034	32866-32435	143	94	033	31496-31068	142	Hypothetical protein	**CNPV-VR111-133**
035	33128-32970	52	92	034	31758-31600	52	IMV protein	VACV-Cop-F14.5L
036	33236-35236	666	96	035	31866-33866	666	Hypothetical protein	Unique
037	35695-35252	147	100	036	34325-33882	147	F15L conserved protein	VACV-Cop-F15L
038	35963-35712	73	100	037	34593-34342	73	Poxvirus APC/cyclosome regulator	MOCV-st1-026L
039	37635-35980	551	96	038	36265-34610	551	Cullin C-terminus domain (CTD)-containing protein	MOCV-st1-027L
040	38288-37632	218	95	039	36918-36262	218	Hypothetical protein	**MOCV-st1-028L**
041	39190-38396	264	96	040	37817-37023	264	Non-functional serine recombinase	VACV-Cop-F16L
042	39276-39572	98	99	041	37903-38199	98	DNA-binding phosphoprotein (VP11)	VACV-Cop-F17R
043	40974-39562	470	99	042	39601-38189	470	Poly (A) polymerase catalytic subunit (VP55)	VACV-Cop-E1L
044	43211-40977	744	97	043	41838-39604	744	IEV morphogenesis	VACV-Cop-E2L
045	43829-43251	192	98	044	42456-41878	192	RNA polymerase (RPO30)	VACV-Cop-E4L
046	43940-50086	2048	95	045	42566-48658	2030	Surface glycoprotein	CNPV-VR111-126
047	50205-56105	1966	93	046	48781-54657	1958	Surface glycoprotein	CNPV-VR111-126
048	56303-58018	571	98	047	54851-56566	571	Virion protein	VACV-Cop-E6R
049	58005-58844	279	100	048	56553-57392	279	ER-localized membrane protein, virion core protein	VACV-Cop-E8R
050	61904-58875	1009	98	049	60454-57425	1009	DNA polymerase	VACV-Cop-E9L
051	61953-62252	99	91	050	60503-60802	99	Sulfhydryl oxidase (FAD-linked)	VACV-Cop-E10R
052	62697-62260	145	97	051	61247-60810	145	Virion core protein	VACV-Cop-E11L
053	64693-62684	669	96	052	63243-61234	669	Membrane protein	VACV-Cop-O1L
054	67095-64780	771	96	053	65648-63330	772	Hypothetical protein	MOCV-st1-043L
055	67240-67130	36	100	054	65793-65683	36	Virus entry/fusion complex component	VACV-Cop-O3L
056	68220-67285	311	98	055	66773-65838	311	DNA-binding core protein	VACV-Cop-I1L

057	68430-68233	65	98	056	66983-66786	65	IMV membrane protein	VACV-Cop-I2L
058	69318-68434	294	95	057	67871-66987	294	ssDNA-binding phosphoprotein	VACV-Cop-I3L
059	69587-69348	79	97	058	68140-67901	79	IMV protein VP13	VACV-Cop-I5L
060	70809-69619	396	97	059	69362-68172	396	Telomere-binding protein	VACV-Cop-I6L
061	72089-70809	426	100	060	70642-69362	426	Virion core cysteine protease	VACV-Cop-I7L
062	72098-74146	682	98	061	70651-72699	682	RNA helicase, DExH-NPH-II domain	VACV-Cop-I8R
063	76031-74166	621	98	062	74584-72719	621	Metalloprotease	VACV-Cop-G1L
064	76351-76028	107	97	063	74904-74581	107	Entry/fusion complex component	VACV-Cop-G3L
065	76354-77076	240	98	064	74907-75629	240	VLTF (late transcription elongation factor)	VACV-Cop-G2R
066	77390-77013	125	99	065	75943-75566	125	Glutaredoxin-like protein	VACV-Cop-G4L
067	77394-78713	439	97	066	75947-77266	439	FEN1-like nuclease	VACV-Cop-G5R
068	78728-78919	63	100	067	77281-77472	63	RNA polymerase (RPO7)	VACV-Cop-G5.5R
069	78924-79517	197	98	068	77477-78070	197	NLPc/P60 superfamily protein	VACV-Cop-G6R
070	80678-79470	402	98	069	79231-78023	402	Virion phosphoprotein, early morphogenesis	VACV-Cop-G7L
071	80706-81488	260	98	070	79260-80042	260	VLTF-1 (late transcription factor 1)	VACV-Cop-G8R
072	81520-82569	349	96	071	80074-81123	349	Myristylated entry/fusion protein	VACV-Cop-G9R
073	82570-83301	243	100	072	81124-81855	243	IMV membrane protein	VACV-Cop-L1R
074	83361-83639	92	96	073	81916-82194	92	Crescent membrane/immature virion protein	VACV-Cop-L2R
075	83636-83932	98	100	074	82191-82487	98	Hypothetical protein	**MOCV-st1-071R**
076	84847-83942	301	99	075	83408-82503	301	Internal virion protein	VACV-Cop-L3L
077	84874-85629	251	99	076	83436-84191	251	ss/dsDNA binding protein (VP8)	VACV-Cop-L4R
078	85622-85981	119	97	077	84184-84543	119	Entry and fusion IMV protein	VACV-Cop-L5R
079	85971-86414	147	96	078	84533-84976	147	Virion morph	VACV-Cop-J1R
080	86477-87400	307	98	079	85039-85956	305	Poly (A) polymerase small subunit (VP39)	VACV-Cop-J3R
081	87397-87957	186	100	080	85953-86513	186	RNA polymerase (RPO22)	VACV-Cop-J4R
082	88375-87947	142	98	081	86931-86503	142	IMV membrane protein	VACV-Cop-J5L
083	88417-92286	1289	99	082	86973-90842	1289	RNA polymerase (RPO147)	VACV-Cop-J6R
084	92305-92598	97	93	083	90896-91156	86	Hypothetical protein	Unique
085	93089-92589	166	99	084	91647-91147	166	Tyr/Ser phosphatase, IFN-gamma inhibitor	VACV-Cop-H1L
086	93103-93678	191	98	085	91661-92236	191	IMV membrane protein	VACV-Cop-H2R
087	94866-93667	399	88	086	93409-92225	394	IMV heparin binding surface protein	VACV-Cop-H3L
088	97242-94867	791	99	087	95785-93410	791	RAP94 (RNA pol associated protein)	VACV-Cop-H4L
089	97394-97999	201	94	088	95942-96559	205	VLTF-4 (late transcription factor 4)	VACV-Cop-H5R
090	98049-98996	315	97	089	96609-97556	315	DNA topoisomerase type I	VACV-Cop-H6R
091	99000-99437	145	95	090	97560-97997	145	Crescent membrane/immature virion protein	VACV-Cop-H7R
092	99741-99415	108	93	091	98307-97975	110	Hypothetical protein	CNPV-VR111-191
093	99746-102340	864	99	092	98312-100906	864	mRNA capping enzyme large subunit	VACV-Cop-D1R
094	102778-102302	158	96	093	101344-100868	158	Virion core	VACV-Cop-D2L

095	102771-103529	252	94	094	101337-102095	252	Virion core	VACV-Cop-D3R
096	103543-104211	222	99	095	102110-102778	222	Uracil-DNA glycosylase, DNA polymerase processivity factor	VACV-Cop-D4R
097	104235-106610	791	99	096	102802-105177	791	NTPase, DNA primase	VACV-Cop-D5R
098	106607-108511	634	100	097	105174-107078	634	Morphogenesis, VETF-small subunit	VACV-Cop-D6R
099	108495-108980	161	98	098	107062-107547	161	RNA polymerase (RPO18)	VACV-Cop-D7R
100	109008-109655	215	99	099	107575-108222	215	mRNA decapping enzyme	VACV-Cop-D9R
101	109652-110323	223	99	100	108219-108890	223	mRNA decapping enzyme	VACV-Cop-D10R
102	112220-110316	634	99	101	110787-108883	634	ATPase, NPH1	VACV-Cop-D11L
103	113096-112224	290	99	102	111663-110791	290	mRNA capping enzyme small subunit	VACV-Cop-D12L
104	114765-113116	549	98	103	113332-111683	549	Trimeric virion coat protein (rifampicin res)	VACV-Cop-D13L
105	115256-114807	149	99	104	113823-113374	149	VLTF-2 (late transcription factor 2)	VACV-Cop-A1L
106	115970-115293	225	99	105	114537-113860	225	VLTF-3 (late transcription factor 3)	VACV-Cop-A2L
107	116239-115967	90	97	106	114806-114534	90	S-S bond formation pathway protein	VACV-Cop-A2.5L
108	118400-116271	709	98	107	116967-114838	709	P4b precursor	VACV-Cop-A3L
109	119274-118495	259	90	108	117898-117062	278	Hypothetical protein	Unique
110	119314-119799	161	100	109	117938-118423	161	RNA polymerase (RPO19)	VACV-Cop-A5R
111	120976-119771	401	98	110	119601-118432	389	Virion morphogenesis, core protein	VACV-Cop-A6L
112	123154-121037	705	99	111	121779-119662	705	VETF-L (early transcription factor large)	VACV-Cop-A7L
113	123247-124158	303	99	112	121871-122782	303	VITF-3 34kda subunit	VACV-Cop-A8R
114	124335-124105	76	100	113	122959-122729	76	IMV membrane, early morphogenesis protein	VACV-Cop-A9L
115	127017-124336	893	98	114	125641-122960	893	P4a precursor	VACV-Cop-A10L
116	127032-127862	276	100	115	125656-126483	275	Viral membrane formation	VACV-Cop-A11R
117	128370-127882	162	99	116	126995-126507	162	Virion core and cleavage processing protein	VACV-Cop-A12L
118	128387-128635	82	93	117	127012-127257	81	Hypothetical protein	Unique
119	128863-128648	71	100	118	127483-127268	71	IMV membrane protein, virion maturation	VACV-Cop-A13L
120	129151-128864	95	98	119	127771-127484	95	Essential IMV membrane protein	VACV-Cop-A14L
121	129329-129168	53	100	120	127949-127788	53	Non-essential IMV membrane protein	VACV-Cop-A14.5L
122	129667-129365	100	96	121	128289-127987	100	Core protein	VACV-Cop-A15L
123	130754-129651	367	96	122	129376-128273	367	Myristylated protein, essential for entry/fusion	VACV-Cop-A16L
124	131297-130767	176	99	123	129919-129389	176	IMV membrane protein	VACV-Cop-A17L
125	131312-132730	472	98	124	129934-131352	472	DNA helicase, transcript release factor	VACV-Cop-A18R
126	133114-132668	148	95	125	131736-131290	148	Zinc finger-like protein	VACV-Cop-A19L
127	133465-133115	116	99	126	132087-131737	116	IMV membrane protein, entry/fusion	VACV-Cop-A21L
128	133443-135272	609	94	127	132065-133915	616	DNA polymerase processivity factor	VACV-Cop-A20R
129	135273-135755	160	97	128	133916-134398	160	Holliday junction resolvase	VACV-Cop-A22R
130	135783-136934	383	99	129	134426-135577	383	VITF-3 45kda subunit	VACV-Cop-A23R
131	136968-140447	1159	100	130	135611-139090	1159	RNA polymerase (RPO132)	VACV-Cop-A24R
132	143493-140467	1008	94	131	142128-139111	1005	A type inclusion protein	VACV-Cop-A25L
133	144965-143538	475	96	132	143606-142173	477	P4c precursor	VACV-Cop-A26L

134	146659-145025	544	97	133	145300-143666	544	P4c precursor	VACV-Cop-A26L
135	146980-146699	93	77	134	145591-145340	83	Hypothetical protein	Unique
136	147412-146981	143	98	135	146023-145592	143	IMV MP/Virus entry	VACV-Cop-A28L
137	148329-147418	303	98	136	146940-146029	303	RNA polymerase (RPO35)	VACV-Cop-A29L
138	148577-148371	68	99	137	147188-146982	68	IMV protein	VACV-Cop-A30L
139	148774-148634	46	96	138	147386-147246	46	Hypothetical protein	Unique
140	148783-149142	119	97	139	147395-147754	119	A31R conserved protein	VACV-Cop-A31R
141	149150-149518	122	96	140	147762-148130	122	Hypothetical protein	Unique
142	150408-149542	288	98	141	149022-148156	288	ATPase/DNA packaging protein	VACV-Cop-A32L
143	150501-151814	437	94	142	149116-150456	446	Hypothetical protein	**MOCV-st1-141R**
144	151760-152275	171	98	143	150402-150917	171	Hypothetical protein	Unique
145	152362-152889	175	98	144	151004-151531	175	C-type lectin-like IEV/EEV glycoprotein	VACV-Cop-A34R
146	152933-153523	196	97	145	151569-152159	196	MHC class II antigen presentation inhibitor	VACV-Cop-A35R
147	153584-154471	295	99	146	152220-153107	295	Concanavalin-like precursor	MOCV-st1-144R
148	154676-155020	114	96	147	153311-153655	114	Hypothetical protein	Unique
149	155097-155735	212	88	148	153732-154370	212	Hypothetical protein	Unique
150	155777-156007	76	95	149	154413-154643	76	Hypothetical protein	Unique
151	156007-157356	449	97	150	154643-155992	449	Hypothetical protein	Unique
152	157564-157376	62	82	151	156198-155950	82	Hypothetical protein	Unique
153	157563-157748	61	83	152	156197-156415	72	Hypothetical protein	Unique
154	158034-158282	82	95	153	156644-156922	92	Hypothetical protein	Unique
Fragmented & Frameshifted			/	154	157438-157196	80	Hypothetical protein	Unique
155	158934-159032	32	100	155	157584-157682	32	Hypothetical protein	Unique
156	160045-161547	500	90	156	158698-160191	497	Serpin	VACV-Cop-K2L
157	161618-162541	307	87	157	160258-161196	312	CD200-like protein	MYXV-Lau-141R
158	162634-163797	387	98	158	161289-162458	389	3 β-hydroxysteroid dehydrogenase/δ5->4 isomerase	VACV-Cop-A44L
159	163844-164776	310	95	159	162499-163434	311	Hypothetical protein	Unique
160	164901-165605	234	91	160	163559-164263	234	Hypothetical protein	Unique
161	166134-166370	78	95	161	164897-165079	60	Hypothetical protein	Unique
162	166478-167227	249	96	162	165199-165948	249	Hypothetical protein	Unique
163*	168310-167546	254	/	/	/	/	Hypothetical protein	Unique
164*	168977-168519	152	/	/	/	/	Hypothetical protein	Unique
165*	169634-169008	208	/	/	/	/	Hypothetical protein	Unique

Although these viruses should be considered separate species, the vast majority of the KPV genomes are syntenic, with ORF differences accounting for the slight difference in gene number almost completely confined to three regions (**Figure 9**).

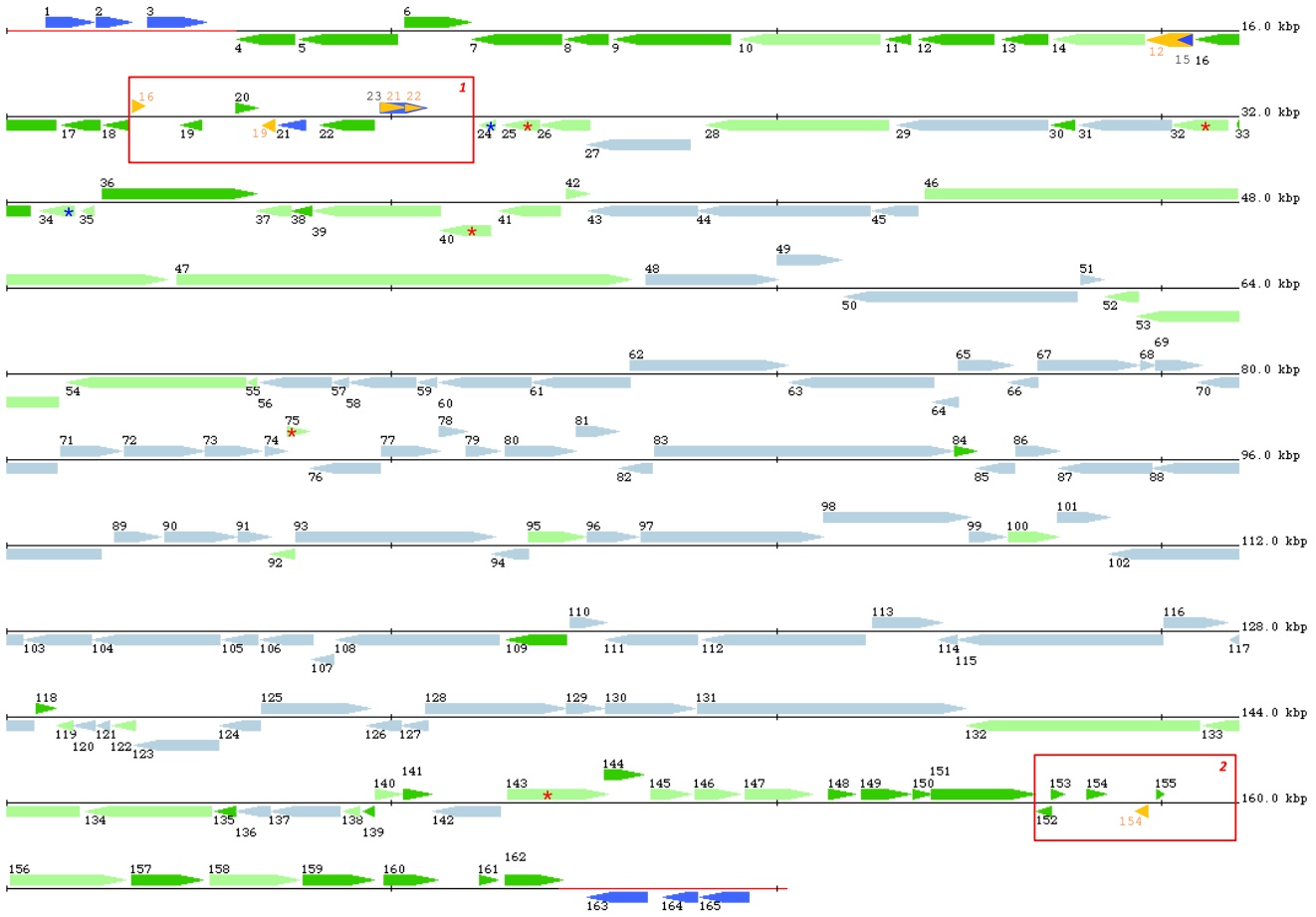

Figure 9. WKPV and EKPV genome map

Most genes are common to both viruses, numbers are for WKPV. Exceptions are noted thusly: WKPV-specific genes are shown in dark blue; EKPV-specific genes are shown and numbered in yellow. The red line indicates the WKPV ITR (the EKPV ITR is shorter). The 81 genes present in all ChPVs are shown in light blue; genes unique to the KPVs are shown in dark green; and other orthologs of ChPV genes are shown in light green. Genes only found in KPVs and MOCV or KPVs and APVs are marked with red and blue asterisks, respectively. Boxes 1 and 2 show regions of gene fragments (see text).

One of these regions is within the viral ITRs, which in WKPV contains three ORFs (hypothetical proteins) whereas none are present in the EKPV ITR (**Table 7**). The second region is flanked by WKPV-WA-018 and -024. Within this region, multiple SNPs and indels lead to variation in most of the ORFs predicted for the two viruses (**Table 7**). The third region is flanked by WKPV-WA-151 and -156 and is also likely to represent a series of fragmented ORFs. Outside of these blocks, the two KPVs contain very similar orthologs of all genes with a single exception, WKPV-WA-015 being fragmented with respect to EKPV-SC-012.

The prevalence of small ORFs in place of bigger ORFs spanning the region is out of the ordinary, and result in a coding density between 44-47% which is much lower than the KPV genome average of 92% (**Figure 10**). Analysis of the nucleotide composition of the KPV genomes revealed that these two regions are associated with higher AT content suggesting that they may represent DNA acquired relatively recently by an ancestor of these viruses. Mapping the distribution of indels from a WKPV versus EKPV comparison revealed that these regions had clusters of indels not found in known protein-coding regions (**Figure 10**). These regions lacking long ORFs, and frequented by conserved small ORFs, could be snapshots of some recent gene fragmentation event after KPV divergence in kangaroo hosts.

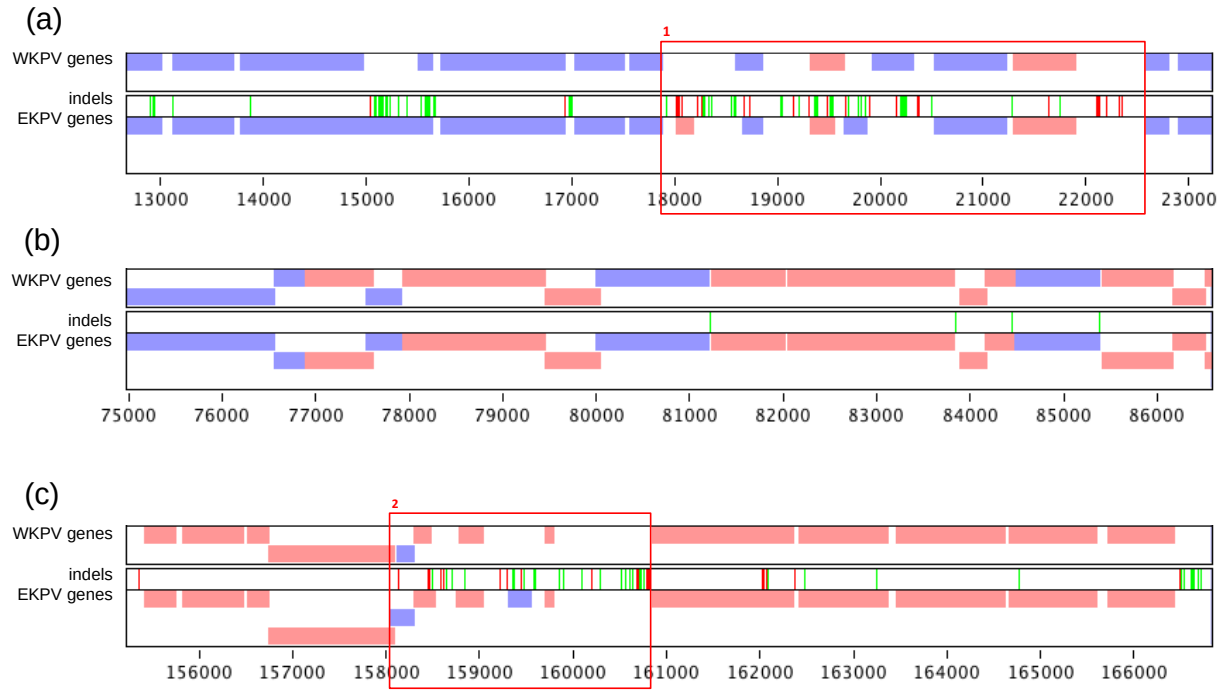

Figure 10. Visual summary of genomic regions lacking long ORFs.

A zoomed-in view of the boxed regions 1 and 2 (of **Figure 9**), depicting genomic regions lacking long ORFs, are shown in panel (a) and (c), respectively. Insertions (green) and deletions (red) within EKPV, relative to WKPV, are displayed but do not indicate the actual genetic change; e.g. a green bar representing an insertion in EKPV could be due to a loss in WKPV. Panel (b) displays insertions and deletions within a representative 10 kb conserved core region. Genes transcribed to the left and right are shown as blue and pink rectangular blocks, respectively.

As observed for other poxviruses, the regions at the termini of the KPV genomes are most varied. For example, almost all predicted genes within the first 20 kb and last 15 kb of the genomes are unique to KPVs, and conversely, there are very few genes in the central region encompassed by these boundaries that are unique to the KPVs (**Table 7**; **Figure 9**). Putatively, small ORFs (less than 50 nucleotides) were only annotated for existing orthologs or if promoter motifs (**Figure 11**) were present. Consensus motifs for early, intermediate, and late promoters of KPVs were generated from the 5′ upstream sequences to orthologs expressed at different temporal stages inferred from VACV transcriptomic data[285]. Consistent with published findings, KPV intermediate promoters have a TAAA motif that is usually

followed by a T and a G, and its late promoters are additionally characterized by a strong TAAAT consensus, and has a highly conserved T at the 10th nucleotide upstream from the motif[66]. Early promoters are more variable, have less obvious motifs, and are in general characterized as rich in AT[58].

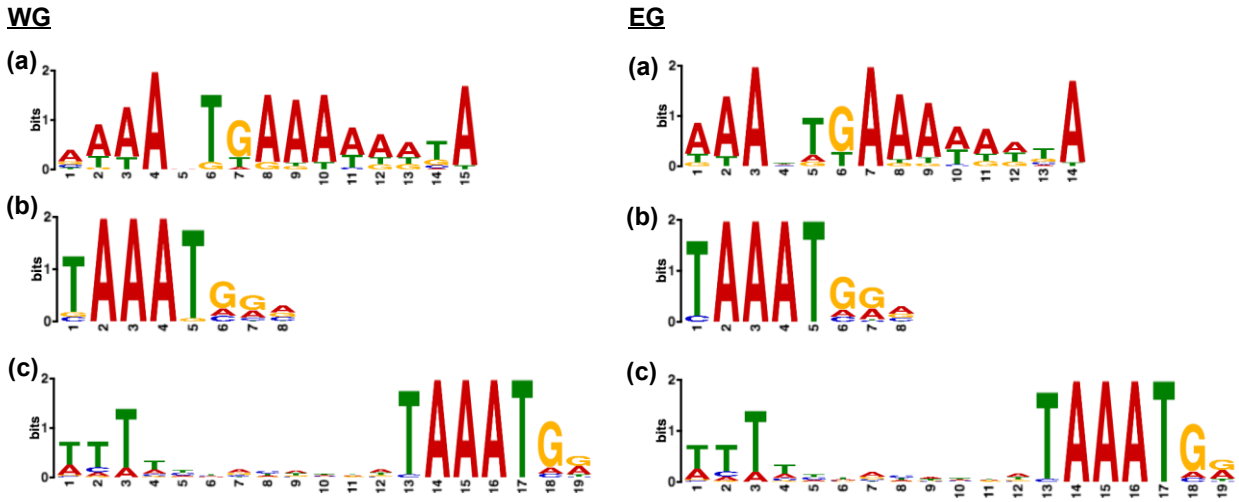

Figure 11. Promoter consensus of WKPV and EKPV

Promoter motifs are generated from the upstream sequences from 81 conserved ChPV orthologs corresponding to each of the expression stages: early (a), intermediate (b), or late (c).

The promoter results demonstrate conservation of motif sequences across the diverged ChPV species. However, without transcription analysis, it is difficult to distinguish a fragmented large gene from a series of small functional genes.

3.4.3.2 Relationship with MOCV and APV

The phylogenetic tree of ChPV placed the KPVs as a sister group to the APVs, but has sequence similarity closer to the MOCV. It is therefore interesting that WKPV-WA-024 and 034 (and EKPV orthologs) encode hypothetical proteins that were considered APV-specific, as well as having WKPV-WA-025, -032, -040, -075, and -143 (and EKPV orthologs) which

encode hypothetical proteins that were previously considered MOCV-specific (**Table 7**). MOCV and the APVs encode a predicted glutathione peroxidase, which is absent from the KPVs. These results support the idea of an ancestral virus common to MOCV, APVs and KPVs with subsequent loss of genes from the various lineages creating the patterns of gene content observed here.

3.4.3.3 Notable virulence genes

The KPVs do share several genes with other ChPVs that are believed to be involved with virus virulence in other poxviruses[160]. These include a semaphorin-like protein[286], a serpin[140], a MHC class II inhibitor[287], and a 3-β-hydroxysteroid dehydrogenase/δ5→4 isomerase[288]. In addition, WKPV-WA-157 and the EKPV ortholog show similarity to CD200 (OX2)-like proteins, which are normally involved in immune regulation. CD200-like proteins have been observed in adenoviruses, herpesviruses[289] and Clade II poxviruses[17]. Although there is very limited amino acid identity (<25%) between KPV predicted CD200-like proteins and Clade II poxvirus proteins, the genes that encode them occupy similar positions in their respective genomes, suggesting a single ancient acquisition. In MYXV, the CD200-like protein down-regulates activation of macrophages at sites of infection, and is responsible for T-cell activation *in vivo*[290]. Interestingly, the MYXV-encoded CD200-like protein does not bind to the rabbit CD200 receptor[291] and the mechanism by which immunomodulation is produced is unknown.

3.4.3.4 Virus dissemination

Consistent with the observation of inclusion bodies in the histological diagnosis of the WKPV infection, the KPVs each possess a complete gene (WKPV-WA-132 and EKPV-SC-131) encoding the A-type inclusion body protein and also the gene for the P4c precursor

protein that directs intracellular mature virus (IMV) particles to A-type inclusion bodies (ATIs)[292]. The formation of ATIs protects IMVs upon release into the extracellular environment and assists with virus spread[292]. Across the ChPVs, ATI proteins are more conserved at the N-terminus than at the C-terminus. Interestingly, the KPVs have a tandem duplication of the P4c gene, which is also found in the GC-rich ChPVs, supporting the hypothesis of a common ancestor for these viruses.

Although nine genes have been identified as specifically contributing to the formation of enveloped extracellular virus (EEV) components[293] in the OPVs (VACV-Cop nomenclature: F12L, F13L, E2L, F11L, B5R, A33R, A34R, A36R and A56R), some poxviruses have only subsets of these. For example, B5R, A33R, A36R and A56R are missing from the KPVs, APVs and most of the GC-rich poxviruses, whereas the other five genes are only absent from a few poxviruses. SGPV is unique because all nine of these genes are absent from its genome. Thus far, A36R and A56R have only been found in OPV genomes.

3.4.3.5 Putative cullin C-terminus domain (CTD)-containing protein

All hypothetical proteins were compared with the public protein sequence databases using the most sensitive parameters for BLASTP programs. However, these searches failed to predict a function for any of these potential genes. Next, the protein sequences were analyzed with HHPred tool[201], which searches protein structure databases with predicted secondary structures and is therefore less reliant on sequence similarity. Only WKPV-WA-039 and ortholog EKPV-SC-038, which are predicted to encode a 551 aa protein, gave strong hits in these searches. The predicted secondary structure of the C-terminal (291-551 aa) regions of these poxvirus proteins were matched with the C-terminal domain (CTD) of cullin proteins[294-296] that are involved in protein ubiquitination[297]. Poxvirus orthologs, MOCV and

SQPV, that have minimal aa sequence identity (approximately 24 and 28%, respectively) within this CTD region also were matched to cullin structures by HHPred. To look for further support of this relationship, we used I-TASSER[205] to determine whether the C-terminal region (291-551 aa) of the WKPV-WA-039 protein sequence could be folded to produce a cullin-CTD-like 3D structure (**Figure 12**). **Figure 12b** shows that a valid model was generated for the CTD domain of WKPV-WA-039 with a C-score of -0.71 (greater than -1.5 indicates a good model). The poxvirus structure is most closely related to the available human cullin CTD structure (PDB: 3DPL-C), over which it can be superimposed with coverage of 97.3% and a RMSD of 1.30 Å. Furthermore, this modeled structure also possesses a surface groove (**Figure 12b**) similar to that used by cullin to bind the N-terminal beta-strand of the host Rbx1 protein (red; **Figure 12b**), which serves to recruit the E2 ubiquitin-conjugating enzyme to the complex[294,298]. However, the structures differ in that KPV-WA-039 lacks the second winged helix domain (WH-B) (**Figure 12b**), which in the cullin CTD harbours a conserved lysine residue required for NEDD8 neddylation[295,299,300]. These results support the hypothesis that orthologs of WKPV-WA-039 may modify the ubiquitination pathway of the infected cell.

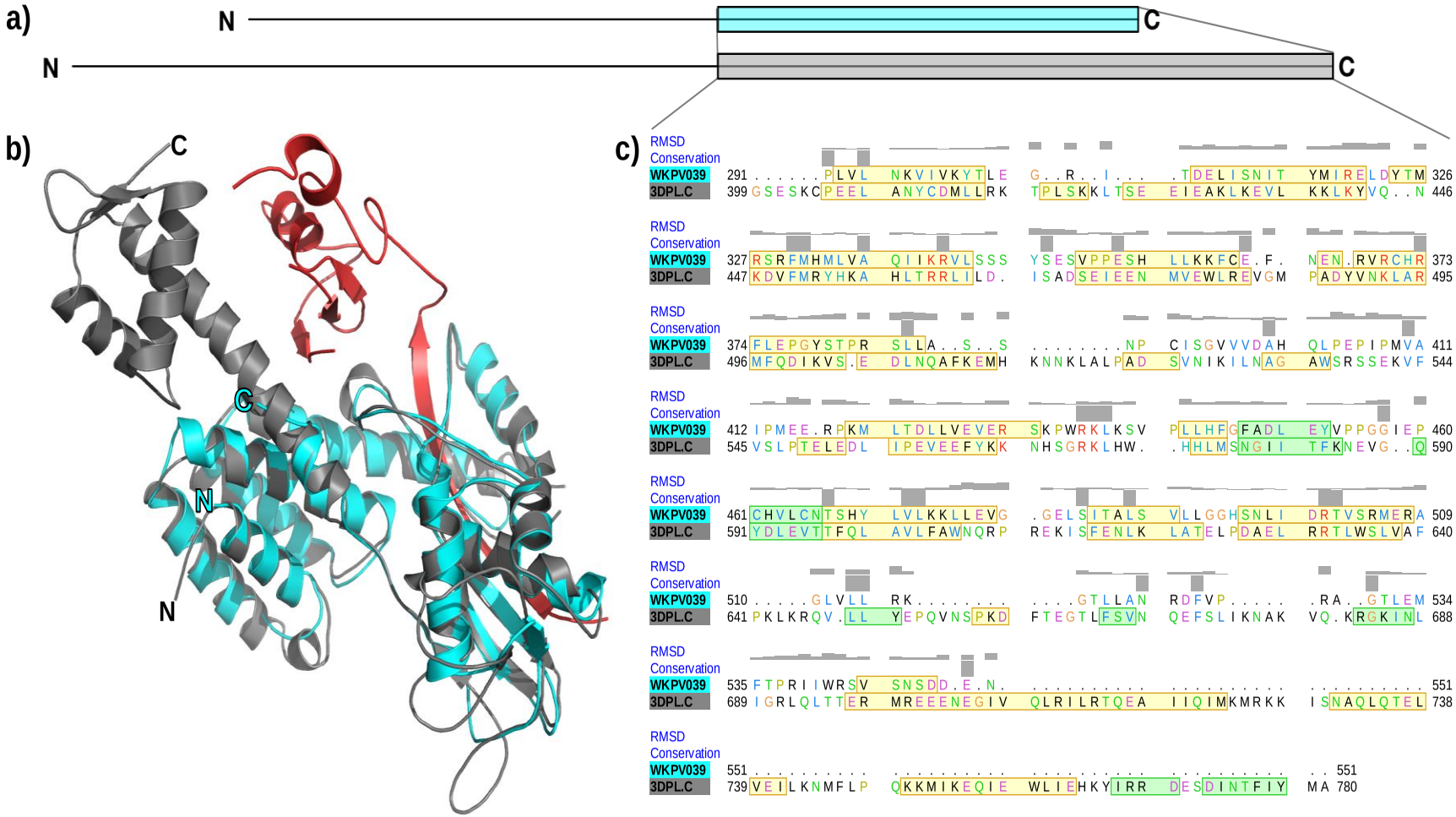

Figure 12. Predicted structure of the WKPV-WA-039 cullin CTD-containing protein.

(a) Boxes show C-terminal regions of WKPV-WA-039 (cyan) and cullin (grey) proteins used in structural modeling. (b) Predicted structure of WKPV-WA-039 (cyan) C-terminal domain superimposed on the cullin (grey) structure (PDB 3DPL.C); RMSD 1.30 Å. The cullin winged-helix domain B is at C-terminus of the structure (no KPV counterpart) and the cullin adaptor (Rbx1) is shown in red. (c) Amino acid sequence alignment generated from superimposed structures in (b); alpha-helices are yellow, beta-strands are green. RMSD values are plotted in grey

3.4.4 Discussion

The observation of presumed poxvirus infections in the Australian kangaroo population is not unusual, with naturally occurring disease being relatively minor and self-limiting according to 2012 Wildlife Health Australia fact sheet. Despite a number of published reports on the histopathology associated with poxvirus lesions in kangaroos, this report is the first that describes complete genome sequences. Discovery of KPVs suggest that kangaroo poxvirus infections have been falsely referred to as the 'molluscipoxviral diseases' of animals, when really they are caused by separate virus species. There is now compelling evidence to suggest a causal relationship between kangaroo poxvirus infection and disease, however Koch's postulates remain unfulfilled. It is possible that other as-yet unidentified factors may be important in disease pathogenesis and lesion expression.

Interestingly, the KPVs occupy a region of the poxvirus phylogenetic tree between the GC-rich mammalian ChPVs and the APVs. This is supported by gene sequence similarity and the presence or absence of particular genes among these three virus groups. Finding KPV orthologs of genes that were previously only found in MOCV or the APVs also helps produce a more accurate picture of those genes that are genus specific, thereby limiting the number of potential candidates that could be associated with specific host range restrictions. Consistent with this relationship to their most similar known relatives, the KPVs also have a relatively GC-rich nucleotide composition of approximately 46% A + T.

The two KPVs are distinct species, but are very similar overall, with only three regions accounting for most of the genomic variation between WKPV and EKPV. No function has yet been predicted for the three ORFs within the WKPV ITRs and there is no evidence except their predicted protein product sizes (152–254 aa) to suggest they might be functional genes.

The other two regions very likely contain fragmented genes in both viruses. These short ORFs have unusually high aa sequence conservation despite the number of indels associated with these regions. It would be interesting to find out if these ORFs actually translate to functional polypeptides given the retention of promoter motifs. Although, the large numbers of indels collected in these regions is out of the ordinary when compared to regions encoding functional genes. Since creation of indels is normally selected against[301], the retention of indels might suggest these sequences as non-functional. Still, it would be interesting if the ancestor virus acquired functional genes at these positions, and only lost function upon introduction into the kangaroo hosts (perhaps leading to a restriction of host range due to gene loss).

It also seems like KPVs lack many of the known virulence genes characterized in extant ChPVs, which might explained the non-acute infections seen in their natural kangaroo host populations. For example, the KPVs do not encode any ankyrin-like proteins, kelch-like proteins, chemokines, TNF-receptors, or other virokines/viroceptors[154]. Furthermore, the KPVs are missing (VACV-Cop nomenclature) F14L (unknown conserved), E3L (dsRNA-binding PKR inhibitor), E5R (virosome component), E7R (myristylated protein), O2L (non-essential glutaredoxin), I4L (ribonucleotide reductase large subunit), J2R (thymidine kinase), D8L (membrane protein), A4L (virion core protein), and A27L (IMV fusion membrane protein). None of these are conserved across all ChPVs, and most are missing from the KPVs' neighbouring viruses MOCV and the APVs as well. However, the large number of unique KPV genes distributed at the two ends of the genomes might also confer unique virulence adapted to the kangaroo hosts. It may be the case that the KPV virulence genes have not yet been recognized because of the extent of virus divergence or the lack of an available host genome sequence. Currently, only a sequence scaffold of the tammar wallaby

(*Macropus eugenii*) genome is available for analysis[302]. Thus, the sequencing of nore macropodid marsupial genomes might help confirm the host as the origin of some KPV genes.

It is very interesting to observe a putative cullin-CTD-containing protein encoded by KPV genes WKPV-039 and EKPV-038, which might participate in the host ubiquitination pathways. The conjugation of ubiquitin to proteins creates a modification that has multiple effects, though commonly ubiquitin labels proteins targeted for destruction. As part of an important sensing process in cells, it is not surprising that it is often involved as part of the signal cascade for antiviral responses. For example, the ubiquitination of IκB (inhibitor of κB), which when bound to NF-κB, would block activation of subsequent anti-viral immune responses[303]. Approximately 20% of ubiquitination involves the Cullin-RING E3 Ligase (CRL) family that interacts with various adaptors and substrate receptors to form the ubiquitination complex[294,304]. There are multiple examples of viral species hijacking CRL pathways[305]. For example, HIV encodes at least 3 proteins that interact with CRL proteins to degrade or re-localize host anti-viral factors[306]. *Epstein Barr virus* encodes viral mimics of a CRL adaptor[307] and a deneddylase that is involved in CRL inactivation[308]. In poxviruses, ectromelia virus encodes F-box proteins and BTB-kelch proteins[309] that normally interact with CRL proteins as adaptor and substrate receptor, respectively, and MYXV M-T5 protein interacts with host CRL complexes to degrade regulatory protein p27/Kip-1 to prevent anti-viral responses[310]. Our discovery of a cullin-CTD-containing protein in KPVs, with orthologs in other poxviruses is the first possible example of viral mimicry of the cullin scaffold within the CRL-based complex, adding a new mechanism by which these viruses may rewire host ubiquitination processes. Since these poxvirus proteins lack the winged helix domain B, which contains a lysine critical for neddylation and activation of the CRL-based complex,

these viral proteins might modify or inhibit ubiquitination by interfering with the assembly of a functional CRL-based complex.

Interestingly, the same subset of poxviruses that have a homolog of the cullin CTD-containing protein also encode a protein distantly related to the cullin adaptor, Rbx1. However, this protein was discovered by its similarity to anaphase-promoting complex subunit 11 (APC11)[311], which interacts with a scaffold APC2 that also has a domain with similarity to the cullin-CTD at its C-terminus. The ORFV APC11-like protein (poxvirus APC/cyclosome regulator; PACR) was shown to interact with APC2, the natural partner of APC11, and block the ubiquitination pathway through which they normally function[311]. However, the ORFV APC11-like protein was never tested with other cullin CTD-containing proteins. Although there is no evidence that the APC11-like and cullin CTD-containing proteins could function together, it is curious that not only do they appear in the same viruses, but they also are adjacent to each other in the genome.

The whole genome data of WKPV and EKPV presented here enhances the current knowledge on poxviruses, with detailed phylogenetic and genomic examination providing a potential evolutionary link between the APVs and MOCVs. Further studies utilising transcriptome analysis coupled with examination of kangaroo genomes, when they become available, could help predict the functional roles of the KPV putative genes. Future characterisation of the KPV putative cullin-CTD-containing proteins, the first examples of potential viral mimicry with the cullin scaffold of the ubiquitination complex, will add significantly to the understanding of poxviral-host immune system interactions.

3.5 Sea otterpox virus

Complete genome sequence of a novel sea otter poxvirus **(*working title; manuscript in preparation)**

Jessica M. Jacob[1], Shin-Lin Tu[2], Kuttichantran Subramaniam[1], Ole Nielsen[3], Pamela A. Tuomi[4], Chris Upton[2], Thomas B. Waltzek[1]

1. Department of Infectious Disease and Immunology, College of Veterinary Medicine, University of Florida, Gainesville, Florida, USA
2. Biochemistry and Microbiology Department, University of Victoria, Victoria, British Columbia, Canada
3. Department of Fisheries and Oceans Canada, Central and Arctic Region, 501 University Crescent, Winnipeg, MB, Canada
4. Alaska SeaLife Center, Seward, Alaska, USA

PAT provided the northern sea otter lesion sample. ON isolated virus using standard cell culture techniques. JMJ, KS, SLT completed genome assembly and annotation; SLT finalized the genome sequence and annotation. SLT performed sequence, structural, and phylogeny analyses. TBW, JMJ, CU, SLT wrote the manuscript.

3.5.1 Abstract

A novel poxvirus was discovered in two orphaned sea otter (*Enhydra* lutris) pups that developed small, superficial ulcerated skin lesions during captive care in 2009 and 2011. Histopathologic examination of the skin lesions revealed epithelial hyperplasia with affected cells displaying intracytoplasmic eosinophilic inclusions. In the northern sea otter, transmission electron microscopy revealed epithelial, brick-shaped, intracytoplasmic virions. The assembled 127,879-bp genome is 68.5% AT, encodes 132 proteins, and has 2,546-bp inverted terminal repeats on each end. Phylogenetic analysis based on core poxvirus genes showed that SOPV is divergent from other known poxviruses, and albeit most closely related to bat-isolated Pteropox virus (PTPV), should be placed into a novel genus on its own. The Sea otterpox virus (SOPV) genome is one of the first marine mammal poxvirus to be fully sequenced, and provides the first step in unraveling the position of a marine mammal poxvirus within the larger *Poxviridae* tree. Marine mammal poxviruses have been recorded for more than four decades, however the phylogenetic relationships of these poxviruses are

not well established because of the lack of complete genome sequences. This study is one of the first to fully sequence a marine mammal poxvirus genome and is the first step in unraveling the position of a marine mammal poxvirus within the larger *Poxviridae* tree. It also provides the necessary sequence to develop future molecular tools for diagnostics and epidemiological studies.

3.5.2 Background

In marine mammals, poxviruses have been detected within the families *Balaenidae, Delphinidae, Mustelidae, Otariidae, Phocidae,* and *Phocoenidae,* but little sequencing data were available[312-314]. Poxviruses infect a wide range of cetacean and pinniped species in both the Atlantic and Pacific Oceans[313,315-317]. These viruses have been seen in both free ranging and captive animals, but they do not appear to cause mass mortalities in marine mammal populations. However, poxviruses may be associated with poor health in individual animals, especially juveniles[312,316,318]. Rehabilitated pinnipeds are especially susceptible to poxviruses during their recovery process[318]. Poxviruses in marine mammals are typically diagnosed by the presence of skin lesions on the animals. In cetaceans, poxviruses are characterized by "tattoo" like lesions on the skin. These lesions are 0.5-3 cm in diameter, irregularly shaped, slightly raised, and gray, black, or yellowish in colour[315,319]. In pinnipeds, poxviruses are characterized by spherical nodules on the skin. These nodules are 1-3 cm in size, raised, firm in consistency, and grayish-white in colour[320,321].

Cetacean poxviruses are not currently classified within any known poxvirus genus, but they appear to be a sister group to the genus *Orthopoxvirus*[313]. However, there are no known transmissions of poxvirus from cetaceans to humans[322]. Most pinniped poxviruses are classified in the genus *Parapoxvirus*[323]. Several parapoxvirus infections have resulted in

humans after contact with poxvirus infected pinnipeds. PCR confirmed the presence of human sealpox in a marine mammal technician after he was bitten on the hand by a young captive grey seal (*Halichoerus grypus*) that had pox-like lesions around the muzzle[86]. In 1983, two marine mammal technicians were diagnosed with human sealpox after handling and caring for multiple grey seals (*H. grypus*) infected with poxvirus[324]. Because poxviruses are zoonotic diseases, extra caution should be taken when handling infected animals.

A novel sea otter poxvirus was discovered in two orphaned sea otter (*Enhydra lutris*) pups that developed small, superficial ulcerated skin lesions during captive care in 2009 and 2011[314]. The pups were from two different Pacific Ocean populations: the northern (*Enhydra lutris kenyoni*) population in Alaska and the southern (*Enhydra lutris nereis*) population in California. Histopathologic examination of the skin lesions in both pups revealed epithelial hyperplasia with affected cells displaying intracytoplasmic eosinophilic inclusions, which were consistent with poxvirus infection. Additionally**,** the epithelial cells in both pups showed different degrees of ballooning degeneration and necrosis. In the northern sea otter (*E. lutris kenyoni*), transmission electron microscopy revealed epithelial intracytoplasmic virions. The virions were brick shaped and had dumbbell electron-dense cores, which were also consistent with poxviruses[314].

3.5.3 Results and Discussion

3.5.3.1 Genome organization and annotations

The assembled Sea otterpox virus (SOPV) genome is 127,879 bp long, 68.7% AT, and has 2,546 bp inverted terminal repeats on each end. It has one of the smaller genomes to date. SOPV has a total of 132 gene sequences. Between SOPV and its closest relative PTPV, 118 orthologous genes are shared, with 10 exclusive to only SOPV and PTPV, and 3 shared with

GC-rich poxviruses. SOPV also has 8 unique genes: 6 hypothetical proteins (SOPV-008, 018, 034, 128, 129, 130), and 2 novel genes with predicted putative functions (SOPV-003 and 035). Phylogenetic tree (**Figure 4**) and gene annotations show that PTPV is its closest relative, but the two viruses are only about 72% identical on the 81 core gene aa level (**Table 4**), which exceeds that of the current intra-genus threshold of 10% aa difference (**Table 3**), and indicates each should be grouped into a separate novel genus. Genome annotations are summarized in **Table 8**.

Table 8. Summary of Sea otterpox virus (SOPV) genome annotations

SOPV gene	ORF position	AA #	Putative gene function	Orthologs	
				VACV-Cop	PTPV-Aus
001	990-127	287	Hypothetical protein (ITR)	-	PTPV-001
002	2226-1372	284	Hypothetical protein (ITR)	-	PTPV-001
003	2800-2354	148	IL-18 binding protein	Unique	
004	3752-3090	220	Partial schlafen-like protein	B2R	PTPV-003
005	4551-3802	249	KilA-N/RING finger protein	CPXV-BR-023	PTPV-137
006	6138-4786	450	Ankyrin repeat protein	B4R	PTPV-008
007	6869-6207	220	Hypothetical protein	-	PTPV-009
008	6979-7953	324	Hypothetical protein	Unique	
009	8603-8088	171	Hypothetical protein	-	PTPV-010
010	9054-8728	108	Hypothetical protein	-	PTPV-012
011	9814-9254	186	Hypothetical protein	-	PTPV-014
012	10404-9877	175	Hypothetical protein	-	PTPV-015
013	11616-10981	211	S-S bond formation pathway protein	F9L	PTPV-017
014	12922-11603	439	Ser/Thr kinase	F10L	PTPV-018
015	14248-13001	415	RhoA signalling inhibitor	F11L	PTPV-019
016	16191-14242	649	EEV maturation protein	F12L	PTPV-020
017	17391-16225	388	Palmytilated EEV envelope phospholipase homolog	F13L	PTPV-021
018	17723-17496	75	Hypothetical protein	Unique	
019	17968-17810	52	Hypothetical protein	F14.5L	-
020	18452-18012	146	Hypothetical protein	F15L	PTPV-024
021	18677-18471	68	RING-H2 zinc finger	-	PTPV-025
022	20281-18707	524	Hypothetical protein	-	PTPV-026
023	21006-20329	225	Conserved non-functional serine recombinase	F16L	PTPV-027
024	21060-21380	106	DNA-binding phosphoprotein (VP11)	F17R	PTPV-028
025	22799-21384	471	Poly(A) polymerase catalytic subunit (VP55)	E1L	PTPV-029
026	25039-22799	746	IEV morphogenesis	E2L	PTPV-030
027	25780-25109	223	dsRNA-binding IFN resistance/PKR inhibitor	E3L	PTPV-031
028	26574-25864	236	RNA polymerase (RPO30)	E4L	PTPV-032
029	26781-28481	566	IMV protein, virion morphogenesis	E6R	PTPV-035
030	28503-29309	268	ER-localized membrane protein, virion core protein	E8R	PTPV-036
031	29355-29798	147	dUTPase	F2L	PTPV-037
032	29862-30980	372	3 β-hydroxysteroid dehydrogenase/δ5→4 isomerase	A44L	PTPV-038
033	31001-31495	164	Hypothetical protein	-	PTPV-039
034	31539-32216	225	Hypothetical protein	Unique	
035	32335-33480	381	TNF receptor-like protein	Unique	
036	33625-34377	250	TRAIL-like protein	-	PTPV-040
037	37437-34429	1002	DNA polymerase	E9L	PTPV-042

038	37467-37757	96	Sulfhydryl oxidase (FAD-linked)	E10R	PTPV-043
039	38188-37754	144	Virion core protein	E11L	PTPV-044
040	40232-38175	685	Virulence, modulates Raf/MEK/ERK pathway	O1L	PTPV-045
041	42789-40276	837	Hypothetical protein	-	PTPV-046
042	42940-42815	41	Virus entry/fusion complex component	O3L	PTPV-047
043	43892-42960	310	DNA-binding core protein	I1L	PTPV-048
044	44108-43893	71	IMV membrane protein	I2L	PTPV-049
045	44957-44112	281	ssDNA-binding phosphoprotein	I3L	PTPV-050
046	45253-45017	78	IMV protein (VP13)	I5L	PTPV-051
047	46426-45275	383	Telomere-binding protein	I6L	PTPV-052
048	47717-46410	435	Virion core cysteine protease	I7L	PTPV-053
049	47724-49796	690	RNA helicase, DExH-NPH-II domain	I8R	PTPV-054
050	51586-49793	597	Metalloprotease	G1L	PTPV-055
051	51918-51583	111	Entry/fusion complex component	G3L	PTPV-056
052	51912-52616	234	Viral late transcription elongation factor (VLTF)	G2R	PTPV-057
053	52921-52550	123	Glutaredoxin-like protein	G4L	PTPV-058
054	52924-54237	437	FEN1-like nuclease	G5R	PTPV-059
055	54244-54435	63	RNA polymerase subunit (RPO7)	G5.5R	PTPV-060
056	54436-54996	186	NLPc/P60 superfamily protein	G6R	PTPV-061
057	56137-54983	384	Virion phosphoprotein, early morphogenesis	G7L	PTPV-062
058	56166-56948	260	VLTF-1	G8R	PTPV-063
059	56966-57973	335	Entry/fusion complex component, myristylprotein	G9R	PTPV-064
060	57974-58714	246	IMV myristylated membrane protein	L1R	PTPV-065
061	58761-59048	95	Crescent membrane/immature virion protein	L2R	PTPV-066
062	60059-59031	342	Internal virion protein	L3L	PTPV-067
063	60085-60858	257	DNA-binding virion core protein (VP8)	L4R	PTPV-068
064	60851-61252	133	Entry/fusion IMV membrane protein	L5R	PTPV-069
065	61206-61670	154	IMV membrane protein, virion morphogenesis	J1R	PTPV-070
066	61704-62705	333	Poly(A) polymerase small subunit (VP39)	J3R	PTPV-071
067	62620-63177	185	RNA polymerase (RPO22)	J4R	PTPV-072
068	63596-63174	140	Late IMV membrane protein, entry/fusion complex	J5L	PTPV-073
069	63682-67542	1286	RNA polymerase (RPO147)	J6R	PTPV-074
070	68057-67545	170	Tyr/Ser phosphatase, IFN-gamma inhibitor	H1L	PTPV-075
071	68072-68647	191	Entry/fusion IMV membrane protein	H2R	PTPV-076
072	69744-68644	366	IMV p35 heparan binding surface protein	H3L	PTPV-077
073	72123-69745	792	RNA polymerase associated protein (RAP94)	H4L	PTPV-078
074	72239-72886	215	VLTF-4	H5R	PTPV-079
075	72905-73855	316	DNA topoisomerase type I	H6R	PTPV-080
076	73856-74290	144	Crescent membrane/immature virion protein	H7R	PTPV-081
077	74329-76875	848	mRNA capping enzyme (large subunit)	D1R	PTPV-082
078	77304-76834	156	Virion core protein	D2L	PTPV-083
079	77297-78109	270	Virion core protein	D3R	PTPV-084
080	78113-78769	218	Uracil-DNA glycosylase, DNA pol processivity factor	D4R	PTPV-085
081	78828-81188	786	NTPase, DNA primase	D5R	PTPV-086
082	81185-83077	630	VETF-small subunit	D6R	PTPV-087
083	83123-83626	167	RNA polymerase (RPO18)	D7R	PTPV-088
084	83685-84302	205	mRNA decapping enzyme	D9R	PTPV-089
085	84299-85033	244	mRNA decapping enzyme	D10R	PTPV-090
086	86929-85025	634	ATPase, NPH1 transcription termination factor	D11L	PTPV-091
087	87843-86968	291	mRNA capping enzyme (small subunit)	D12L	PTPV-092
088	89522-87867	551	Trimeric virion coat protein (rifampicin resistance)	D13L	PTPV-093
089	90000-89545	151	VLTF-2	A1L	PTPV-094
090	90707-90033	224	VLTF-3	A2L	PTPV-095
091	90931-90704	75	S-S bond formation pathway protein	A2.5L	PTPV-096
092	92935-90950	661	P4b precursor	A3L	PTPV-097
093	93627-92977	216	Hypothetical protein	-	PTPV-098
094	93677-94186	169	RNA polymerase (RPO19)	A5R	PTPV-099
095	95306-94176	376	Virion morphogenesis, core protein	A6L	PTPV-100
096	97481-95343	712	VETF-large subunit	A7L	PTPV-101
097	97551-98420	289	VITF-3, small subunit	A8R	PTPV-102

098	98690-98430	86	Viral membrane protein, early morphogenesis	A9L	PTPV-103
099	101435-98691	914	P4a precursor	A10L	PTPV-104
100	101450-102481	343	Viral membrane formation	A11R	PTPV-105
101	103008-102478	176	Virion core and cleavage processing protein	A12L	PTPV-106
102	103245-103030	71	IMV membrane protein, virion maturation	A13L	PTPV-107
103	103530-103246	94	Essential IMV membrane protein	A14L	PTPV-108
104	103710-103549	53	Non-essential IMV membrane protein	A14.5L	PTPV-109
105	103986-103711	91	Core protein	A15L	PTPV-110
106	105076-103973	367	Myristylated protein, essential for entry/fusion	A16L	PTPV-111
107	105673-105098	191	IMV membrane protein	A17L	PTPV-112
108	105688-107127	479	DNA helicase, transcript release factor	A18R	PTPV-113
109	107353-107108	81	Zn finger-like protein, late virion morphogenesis	A19L	PTPV-114
110	107701-107354	115	IMV membrane protein, entry/fusion complex	A21L	PTPV-115
111	107700-108992	430	DNA polymerase processivity factor	A20R	PTPV-116
112	108985-109449	154	Holliday junction resolvase	A22R	PTPV-117
113	109516-110658	380	VITF-3, large subunit	A23R	PTPV-118
114	110698-114183	1161	RNA polymerase (RPO132)	A24R	PTPV-119
115	115384-114185	399	Truncated A-type inclusion protein	A25L	PTPV-120
116	116967-115408	519	P4c precursor	A26L	PTPV-121
117	117350-117024	108	IMV surface protein, fusion protein	A27L	PTPV-122
118	117787-117365	140	IMV MP/virus entry	A28L	PTPV-123
119	118696-117788	302	RNA polymerase (RPO35)	A29L	PTPV-124
120	118871-118665	68	IMV protein, entry/fusion complex component	A30L	PTPV-125
121	119059-119508	149	Hypothetical protein	A31R	PTPV-126
122	120271-119501	256	ATPase/DNA packaging protein	A32L	PTPV-127
123	120364-120963	199	EEV membrane phosphoglycoprotein	A33R	-
124	121014-121520	168	C-type lectin-like IEV/EEV glycoprotein	A34R	PTPV-129
125	121563-122087	174	MHC class II antigen presentation inhibitor	A35R	PTPV-130
126	122184-122849	221	Hypothetical protein	-	PTPV-131
127	122904-123803	299	Hypothetical protein	A37R	-
128	123867-124463	198	Hypothetical protein	Unique	
129	124515-124721	68	Hypothetical protein	Unique	
130	124912-125286	124	Hypothetical protein	Unique	
131	125654-126508	284	Hypothetical protein (ITR)	-	PTPV-001
132	126890-127753	287	Hypothetical protein (ITR)	-	PTPV-001

3.5.3.2 GC-rich poxviral protein orthologs (SOPV-ELK-021, -22, -41)

To date, SOPV and PTPV are the only other AT-rich poxviruses that share the GC-rich poxvirus orthologs that best hit with SQPV-019, -020, and -035 (functions unknown) at corresponding conserved syntenic positions. In contrast to the GC-rich gene sequences, which are 30-47% AT, SOPV and PTPV orthologs are 61-71% AT. This could suggest that this set of orthologs were picked up by the ancestral virus at post-divergence from the ancestral AT-rich virus but prior to the speciation into SOPV, PTPV, and GC-rich poxviruses. This would be plausible according to the ChPV phylogenetic tree, since the SOPV/PTPV clade is situated between the AT-rich and GC-rich viruses. Alternatively,

because the genes are missing in the rest of the AT-rich viruses, it could be evidence to suggest that the ancestral virus of SOPV and PTPV had picked up this set of specific poxviral orthologs from an ancient interaction with the ancestral GC-rich poxvirus while in a common host, whereby these horizontally transferred homologs provided advantage to the viruses such that they are retained to this day and have adapted an overall high AT composition.

3.5.3.3 IL-18 binding protein (SOPV-ELK-003)

IL-18, a proinflammatory cytokine, is crucial in the activation of immune responses in host cells; IL-18 can induce IFN-γ, and in turn activates natural killer (NK) and T-helper cells. IL-18 binding proteins (BPs) are found naturally in humans and mouse as a way to regulate the cytokine[325]. Currently, different IL-18 BPs have been found in VARV[326], VACV[327], ECTV[328,329], MOCV[330], YLDV[331], and avian poxvirus[121,332]. These poxviral homologs improve viral survival by binding and neutralizing IL-18 sufficiently to impair downstream immune activations. SOPV003 encodes a 148aa protein that shares only 14-26% aa identity to the other poxviral IL-18 BPs and has no detectable eukaryotic homolog. Despite the low aa identity score, SOPV003 retains several important structural features, and is suggested to be a putative novel IL-18 BP. **Figure 13** shows that the predicted SOPV003 structure adopts an immunoglobulin (Ig)-fold due to the key intra-molecular disulfide bond formed between the central beta-strands. This canonical shape and contributing cysteines (C43 and C114 in SOPV) are conserved across IL-18 BP and receptors[333], and the overall structure is found most similar to the two available poxviral IL-18 BP structures from ECTV (3f62A) and YLDV (4eeeA).

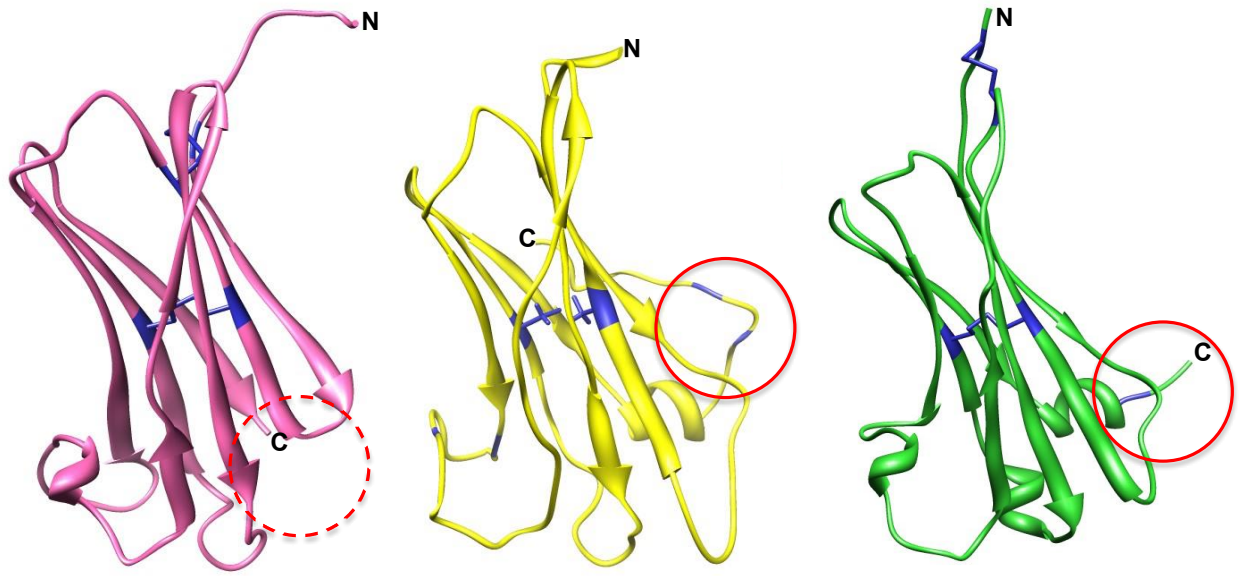

Figure 13. Structural comparison between poxviral IL-18 binding proteins (BPs).

The structure of SOPV-ELK-003 putative IL-18 BP (yellow) is predicted to fold into the canonical Ig-fold structures similar to that of the ECTV (pink; 3F62A) and YLDV (green; 4EEEA) IL-18 BP crystal structures due to conservation of the essential disulfide bond formed by the central cysteine residues (blue). At the N-terminal, ECTV and YLDV have an additional set of non-essential disulfide bond. At the C-terminal, SOPV possesses additional cysteine residues located similarly to the cysteine involved in the unique homodimerization of YLDV, which is missing in ECTV.

It should be noted that ECTV and YLDV IL-18 BPs each have an additional set of disulfide bond near the N-terminals formed between different sets of beta-strands, however, the interaction is found to be non-essential to the overall structure or ligand-binding ability of the proteins[331]. In contrast to ECTV, both SOPV and YLDV have extended C-terminus with two and one cysteines residues exposed, respectively. Interestingly, the C-terminus cysteine in YLDV forms an inter-molecular disulfide bond with that of another protomer in a back-to-back fashion, and enables the protein to form a homodimer that binds to IL-18 proteins in a 2:2 stoichiometry[331]. It is unsure whether or not the C-terminus cysteine(s) of SOPV003 also participate in such an interaction, or function as monomers like the human and ECTV IL-18 BPs (which lack the terminal cysteines).

Regardless of dimerization, the ECTV and YLDV IL-18 BPs have been shown to bind with three hydrophobic pockets on the human IL-18 ligands termed sites A, B, and C, with the most important interactions occurring at the site A vicinity with ligand residue K53[331,333]. **Appendix 5** lists the IL-18 BP residues involved at each site; essential residues are discussed.

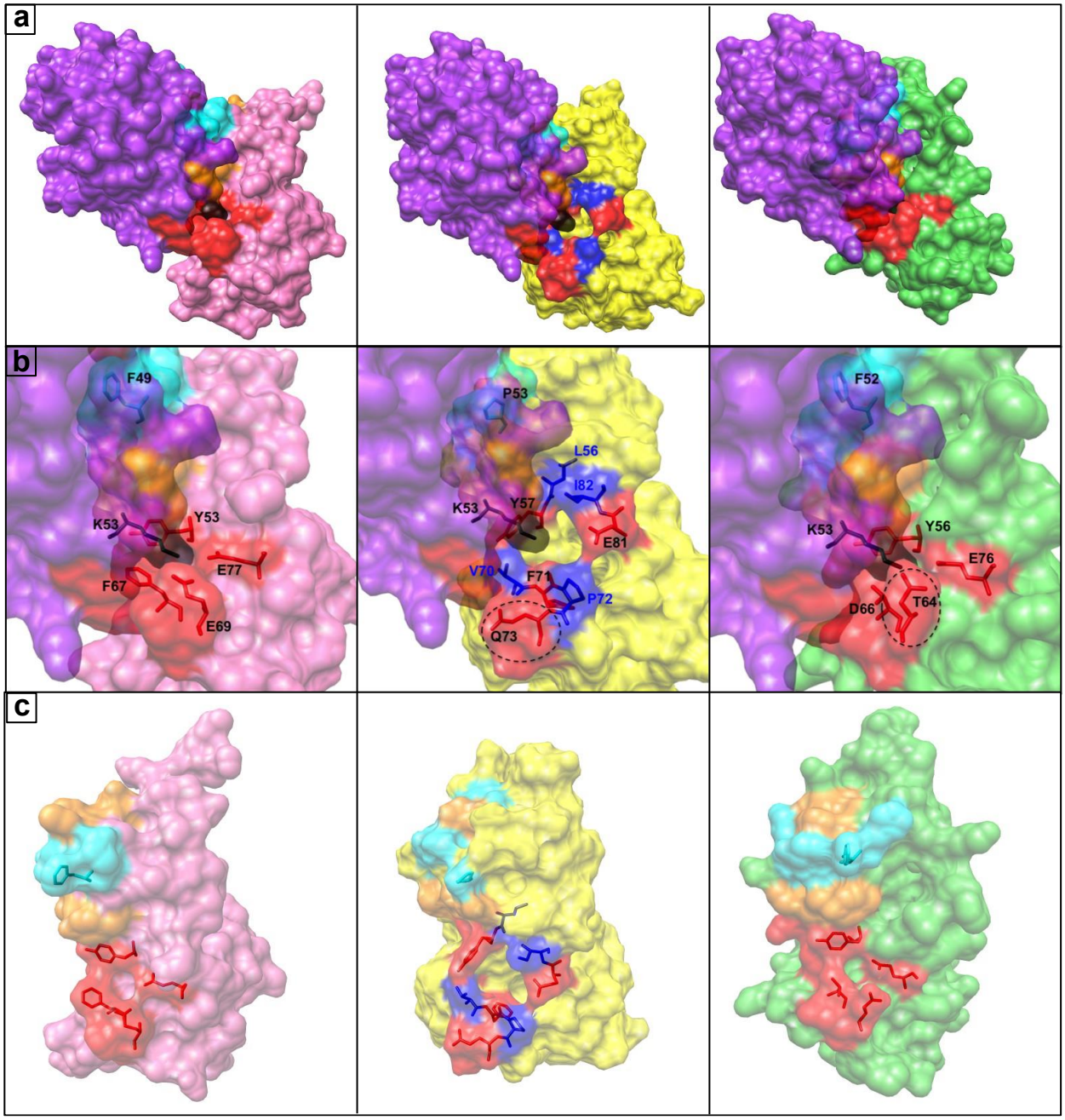

Figure 14. Binding and interaction between poxviral IL-18 BPs and human IL-18 ligands

(a) The binding interfaces and conformation between IL-18 ligands (purple) is shown between ECTV (pink) and YLDV (green) IL-18 BPs, and extrapolated for SOPV IL-18 BP

(yellow); binding interfaces at ligand site A (red), site B (orange), and site C (cyan) are shown. (b) A zoomed-in view of the conserved and novel (blue) hydrophobic residues shown to interact with IL-18 K53 at ligand site A (red); differences are circled by dashed lines. Site B (orange) residues are less essential and mainly hydrophobic. Important large hydrophobic residues (phenylalaine and proline) are shown at site C (cyan). (c) An angled view of the IL-18 BPs binding interfaces and hydrophobic pocket formed; SOPV appears to generate a larger pocket with additional hydrophobic residues.

As seen across **Figure 14a** panel, the conserved essential residues (red) form a binding interface to key residue K53 (black) on IL-18 ligand site A with multiple interactions. A zoomed view of the interaction in **Figure 14b** panel shows the aliphatic side chain of K53 is buried in the hydrophobic wall created by the conserved tyrosine and phenylalanine residues on IL-18 BPs, while the charged side chain forms strong π-cation interaction with the aromatic ring of phenylalanine and salt bridges with conserved acidic residues (glutamic acid and aspartic acid). At site A, SOPV003 is similar to other homologs in that it retains the 3 essential residues (Y57, F71, E81) involved in interacting with K53; however, SOPV lacks one H-bond forming acidic residue found in ECTV (E69) and YLDV (D66), which is replaced with a structurally similar glutamine (Q73). Unfortunately, from the SOPV003 interaction panel (**Figure 14**), the glutamine is predicted to bend away from the K53 and would be too far to form any kind of interaction. An alanine substitution at E69 in ECTV showed a decrease in binding affinity that is less dramatic than the other essential residues. Interestingly, as made apparent from an angled view of the predicted IL-18 BP in **Figure 14c**, SOPV structure is predicted with additional hydrophobic residues (blue; L56, V70, P72, I82) at the binding interface with ligand site A that are not seen in ECTV or YLDV. Alternatively, these residues increase the hydrophobic surface and interaction at the binding interface, and can likely compensate or overcome the aforementioned decrease in binding affinity. SOPV also differ from the ECTV and YLDV homologs in the conserved essential phenylalanine inserted at the hydrophobic pocket at site C. Instead, SOPV003 retains a proline (P53) at this

position, which provides the same steric hindrance and hydrophobicity as large, non-polar amino acids, and thus is predicted to have minor changes to binding affinity.

The viromimicry of a functional IL-18 BP in SOPV would significantly benefit poxvirus survival in host by blocking IFN-γ activation of T-helper and NK cells. Interestingly, because of the important activation pathways IL-18 partakes, improper regulation of IL-18 has also been shown in autoimmune diseases such as type 1 diabetes, multiple sclerosis, rheumatoid arthritis, psoriasis, and systemic lupus erythmeatosus[334]. Uncovering novel potent IL-18 BPs could additionally provide potent drug designs to combat these IL-18 induced autoimmune diseases.

3.5.3.4 TNFR-like protein (SOPV-ELK-035)

SOPV035 encodes a 381aa protein, and upon initial BLAST and CDD searches[197], shows "unusual" similarity with members of the tumor necrosis factor receptor (TNFR) superfamily. TNFR is a very diverse protein superfamily, with 29 groups reported in human alone. Members of TNFR bind to members of TNF to regulate various immune processes and pathways. For example, TNFR members such as CD40 has various regulations linked to tumorgenesis and modulation of immune responses[335,336], whereas TNFR2 suppresses inflammation and mediate anti-apoptotic responses[337,338]**.** TNFR members are difficult to categorize due to low aa identity across the family, but share similar profiles in having various tandem repeat copies of cysteine-rich domain (CRD) found on their extracellular N-terminal[339,340] (**Figure 15a**).

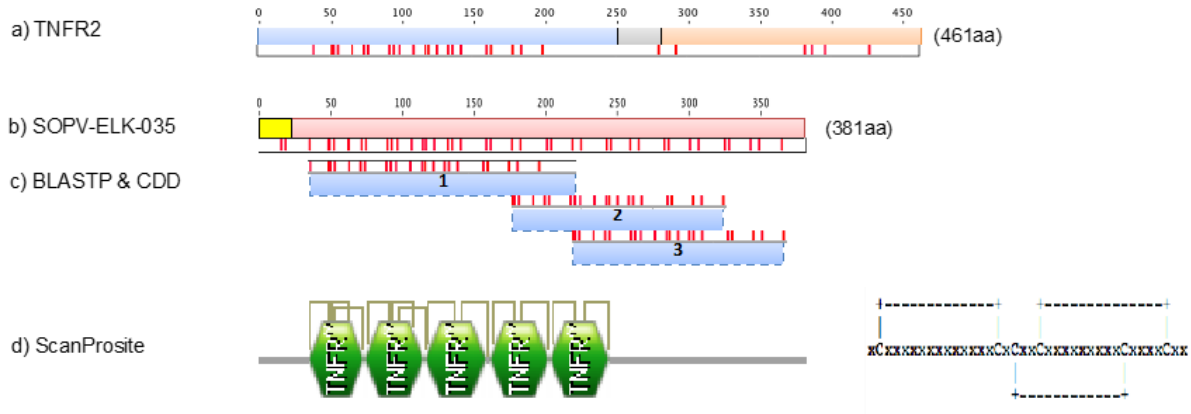

Figure 15. Distribution of cysteine-rich domains (CRDs) on putative SOPV tumor necrosis factor receptor (TNFR)-like protein.

(a) The cysteine distribution (red lines) of SOPV-ELK-035 BLASTP match: TNFR2 from gorilla (XP_004024712.1), and the overall topology: blue extracellular N-terminal, black transmembrane domain, and orange cytoplasmic are shown. (b) SOPV-035 is predicted to have a signal peptide sequence (yellow) and shows similar CRD distribution to the (c) extracellular domain of TNFR across the length of the protein (detected by BLASTP and CDD). (d) The putative SOPV TNFR-like protein is predicted by ScanProsite to have at least five TNFR CRD motifs (green) associated with forming disulfide bonds.

In this case, several motifs of these CRD are detected in the SOPV035 protein sequence (**Figure 15b**). From the consensus of BLASTP and CDD results, the N-terminal of TNFR proteins are seen aligned to the cysteine-rich SOPV035 sequence in several instances (with up to 47% aa identity with TNFR2) (**Figure 15c**). Poxviruses also have homologs of TNFR serve to modify host's TNF-mediated anti-viral responses[134,341-344], however, SOPV035 does not seem orthologous to any of them. In attempt to classify the SOPV TNFR-like protein, PrositeScan[198] predicts at least five TNFR-specific CRDs, which are important for TNF-binding (**Figure 15d**). It is thus suggested that SOPV035 sequence exhibits a unique pattern composed of CRD motifs from the TNFR superfamily members that might allow it to serve as a viroceptor in the form of a novel TNFR-like protein. Three topology prediction programs (CCTOP, Phobius, and SignalP) independently predicted a N-terminal signal peptide sequence (1-23) with no transmembrane domain (**Figure 15b**), suggesting that this putative TNFR-like protein may be targeted for release from the cell into an extracellular

environment. It is additionally observed that there is a notable drop in AT% of this gene to 59% from an average of 71% with that of the neighbouring genes (SOPV-ELK-034 is a unique HP, and -036 is a TRAIL protein). An overall shift in base composition away from genome average is at times signature of a (recent) HGT event, and could potentially encode virulence factor as in the case of host-derived pathogenicity islands seen in poxviruses[172]. Should this be the case, it could perhaps point towards a TNFR homolog of SOPV035 in a vertebrate host of low AT% genome. The genome of the sea otter, whose genome sits at 59% AT, was recently sequenced and assembled. However, the sea otter TNFR sequence is not significantly similar to the SOPV035 gene over the TNFRs from other organisms. Alternatively, this drop in AT% (or an increase in GC%) could be due to the nature of having encoded a cysteine-rich protein (cysteine codon: UGU, UGC). A quick search shows that sea otter TNFR2 has 34.4% AT, but 55.8% AT for TRAIL. Similarly, the TNFR2 of *Pteropus vamprus* is 37.1% AT, but its TRAIL protein has 59.7% AT. It is thus extrapolated that the lowered AT% of the SOPV035 gene is due to the cysteine-rich nature the encoded TNFR-like protein, which may or may not be coupled with a HGT origin from a host with a different AT%.

3.5.3.5 TRAIL-like protein (SOPV-ELK-036)

The first poxviral TRAIL (TNF-related apoptosis-inducing ligand) homolog was found in PTPV[266]. SOPV036 encodes a TRAIL protein at the same synteny position as PTPV ortholog. This evidence supports a single HGT event from their common ancestor and ancestral host rather than separate gene captures. The phylogenetic relationship between poxviral and eukaryotic TRAILs (conserved C-terminals) is demonstrated on **Figure 16**, and puts the viral TRAILs into a monophyly.

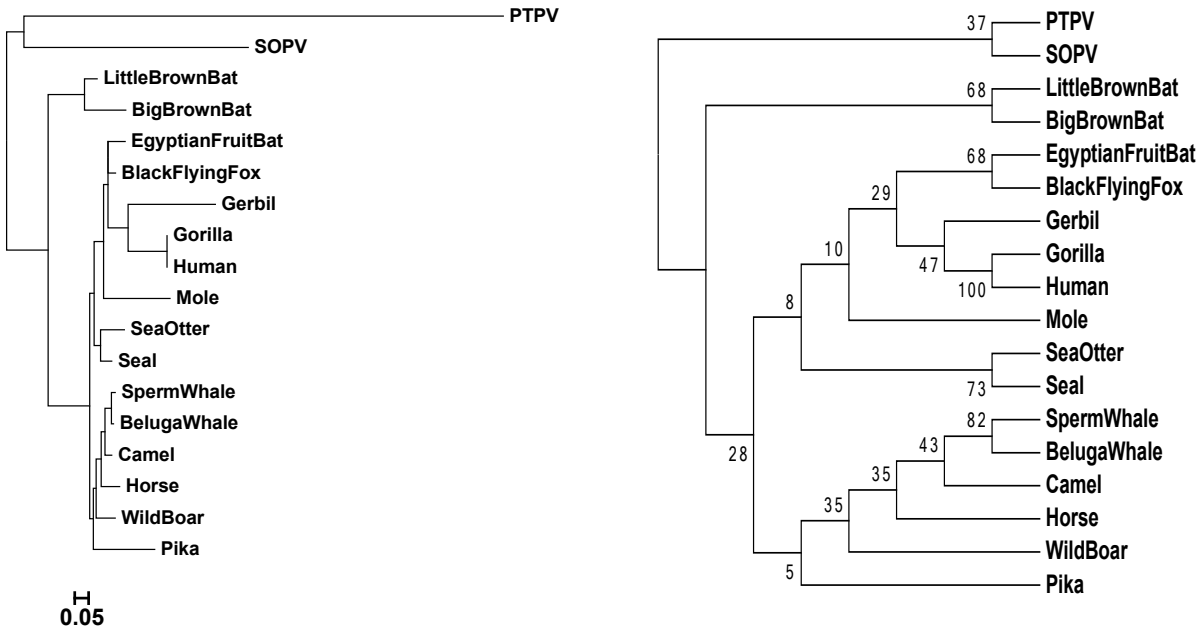

Figure 16. The phylogenetic relationship between poxviral TRAILs and other eukaryotic TRAIL representatives.

Scale bar represent the number of amino acid substitution per site.

On the amino acid level, SOPV036 is up to 49% identical with PTPV040 at the conserved C-term ligand domain (100 aa onward) on the aa level. However, SOPV homolog appears to resemble more closely to mammalian TRAILs (approximately 60% aa identity) than with PTPV homolog (49% aa identity). PTPV TRAIL on the other hand, is only about 50% identical to these eukaryotic homologs. This indicates that SOPV could potentially retain more similarity to its original ancestral HGT gene than PTPV to serve as a virokine. The fact that the difference between viral orthologs exceeds that of the 27% difference between human and microbat homologs, perhaps suggests that PTPV had undergone a strong selection pressure favouring mutations since its divergence from the last virus ancestor with SOPV; this is consistent with the long branch length observed since its divergence from the last common ancestor with SOPV (**Figure 16**). This selection pressure, perhaps due to adaptation in a new host, results in the observation of viral orthologs as distant as with the eukaryotic homologs. Since no current eukaryotic TRAIL in the database has exhibited strong similarity

to the SOPV homolog, one cannot elucidate a particular eukaryotic origin of the poxviral TRAIL gene, and attribute a potential ancestral host animal in which the TRAIL is derived from. At this stage, it is unknown if the poxviral TRAILs inhibit apoptosis of infected cells by behaving as virokines that compete with host TRAIL receptor similar to that of an adenovirus protein[345], or if they induce apoptosis in host cells by acting as attacking cytokines.

3.5.3.6 Truncated A-type inclusion body (SOPV-ELK-115)

A-type inclusion proteins (ATIs) are encoded by several chordopoxviruses (except Clade II: COTV, cervid-, capri-, lepori-, yata-, suipoxviruses) to allow for the formation of cytoplasmic bodies filled with mature virions. These ATI bodies are formed in the cytoplasm, and can protect virions after release to ensure infectivity[346]. These ATI bodies are also useful for diagnostic purposes, as they are well-defined large entities that are visible using a light microscope[347]. SOPV115 encodes a truncated version of ATI protein, with only the 399aa N-term fragment of the gene. N-terminal fragments are more likely to be expressed than C-terminal fragments due to proximity to the upstream promoter. Truncations of ATI orthologs are not uncommon. The majority of VACV orthologs have been truncated down to <100 aa, and are unable to form ATI bodies as a result. It is unsure at this stage the impact this specific truncation has on SOPV ATI body formation, however, truncation to similar extent have also been seen in MOCV, and some APVs, suggesting potentially a functional role for the N-term alone.

Chapter 4. CONCLUSIONS AND FUTURE DIRECTIONS

In the course of my work, I developed a streamlined protocol, which processed large raw sequence files, and enabled the genomic characterization of 5 novel poxviruses from 4 different genera.

The amount of poxviruses identified in novel hosts in the recent years has demonstrated the prevalence and evolutionary success and diversity of this virus group. It is interesting that none of these poxvirus infections were directly associated with the deaths of their hosts. The possibility that these viruses are well adapted in these hosts, and replicate in harmony with them (without killing off hosts) could turn the host population into a natural reservoir of the virus; this is likely the case for the KPVs. Whereas most viruses do not cause severe disease in their natural reservoir hosts, they can become highly pathogenic when introduced to a new host. The novel predicted virulence genes, such as poxviral homologs of TRAIL, IL-18 BP, TNFR, and cullin proteins, may facilitate this process. The wide geographical distributions some of these hosts covered predict a higher chance of contacts with other species, which increases the likelihood of spillover events. The concern is even greater given the behaviours of megabat poxviruses. Megabats, unlike their nocturnal insectivorous relative (microbats), are fruit bats that come out during the day that share food source with many wildlife and even human. Surveillance can be difficult, however, especially if the infection is relatively benign and goes undetected. The results generated from my research demonstrate the importance to sequence and assemble novel poxviruses which may (1) pre-emptively characterize potential zoonotic agents, (2) promote surveillance programs for poxviruses in the environment, (3) provide sequence information for diagnostic or epidemiology tools, and (4) expand the sequence repertoire for comparative analyses and/or other applications.

As more and more of the poxvirus phylogenetic tree and genomes are unravelled, future directions for poxviral bioinformatic research include the elucidation of AT- and GC-rich bias origin, promoter analyses, as well as continual similarity searches of unknown genes and wet lab characterizations of predicted novel genes. I personally find that the observed trends in poxviral ORF prediction (promoter motifs, purine skews, aa composition, and pI values mentioned in Ch. 1.5.2.3) could also be incorporated into the GATU software to improve annotation quality and performance. I hope the findings from my work have demonstrated, in part, the tremendous capabilities and potential of bioinformatic analyses from genomic data.

Moving forward, as the throughput of sequencing technologies improves, the field of bioinformatics is challenged to evolve new ways to manage and analyze these data efficiently and effectively. Data scientists today are no longer excited about “the idea of more data”, for the bottleneck now lies within the realm of analysis. There is a need for the bioinformatics field to continually improve algorithms and memory consumption in order to handle the large amount of data in a timely and efficient fashion, as well as develop user-friendly and accessible tools that answer questions in the biological context. The growth of bioinformatics will need to be sustained from 3 factors: (1) an expanded training of bioinformatics personnel, (2) the collaboration between scientists and programmers at developing bioinformatics resources, and (3) acknowledgement and valuation of computational and *in silico* analyses by the scientific communities.

The reality is that there exists a gap between the software available and the bioinformatics proficiencies of many scientists today. Bartlett *et al.* summarize the problem nicely: "the power of bioinformatics work is shaped by its dependency on life science work, which combined with the black-boxed character of bioinformatics expertise further contributes to

situating bioinformatics on the periphery of the life sciences. [Consequently], the imagined future of bioinformatics work suggest that bioinformatics will become ever more indispensable without necessarily becoming more visible, forcing bioinformaticians into difficult professional and career choices"[348]. Adding to the steep learning curve of bioinformatics research, many of the prominent tools today are command line-based (those that have a GUI may not always be free). Tools that overcome the memory consumption problems include cloud computing services (such as those provided by Compute Canada) also requires basic understanding of command lines and bash scripts. Without this, many of the essential software would not be easily accessible. Today, the bioinformatics field recognizes this crisis, and many studies are dedicated to the reform and evaluation of bioinformatics curriculum for students[349,350]. Similarly, individual research labs should also value and provide resources to train and equip scientists with skills in command-line, regular-expression, and simple scripts that immensely streamline the analyses of large amount of data between programs (as shown in the Methods section). In short, it remains crucial that the scientific communities recognizes the value computational and *in silico* analyses and promote the training of bioinformatics skills. It is important for bioinformatics tools to continually evolve and be maintained with friendly updates and features that assist with any new research needs or context.

BIBLIOGRAPHY

1. Baltimore D. Expression of Animal Virus Genomes. *Bacteriol Rev*. 1971;35(3):235–.

2. Iyer LM, Aravind L, Koonin EV. Common origin of four diverse families of large eukaryotic DNA viruses. *J Virol*. 2001;75(23):11720-11734. doi:10.1128/JVI.75.23.11720-11734.2001.

3. Koonin EV, Yutin N. Origin and Evolution of Eukaryotic Large Nucleo-Cytoplasmic DNA Viruses. *Intervirology*. 2010;53(5):284-292. doi:10.1159/000312913.

4. Yutin N, Wolf YI, Raoult D, Koonin EV. Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. *Virol J*. 2009;6(1):223. doi:10.1186/1743-422X-6-223.

5. Raoult D, Audic S, Robert C, et al. The 1.2-megabase genome sequence of mimivirus. *Science*. 2004;306(5700):1344-1350. doi:10.1126/science.1101485.

6. Philippe N. Pandoraviruses: Amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes (vol 341, pg 281, 2013). *Science*. 2013;341(6153):1452-1452.

7. Legendre M, Bartoli J, Shmakova L, et al. Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proc Natl Acad Sci USA*. 2014;111(11):4274-4279. doi:10.1073/pnas.1320670111.

8. Moreira D, Lopez-Garcia P. Ten reasons to exclude viruses from the tree of life. *Nature Reviews Microbiology*. 2009;7(4):306-311. doi:10.1038/nrmicro2108.

9. Claverie J-M, Ogata H. Ten good reasons not to exclude giruses from the evolutionary picture. *Nature Reviews Microbiology*. 2009;7(8):–615. doi:10.1038/nrmicro2108-c3.

10. Koonin EV, Krupovic M, Yutin N. Evolution of double-stranded DNA viruses of eukaryotes: from bacteriophages to transposons to giant viruses. Witzany G, ed. *Ann N Y Acad Sci*. 2015;1341(1):10-24. doi:10.1111/nyas.12728.

11. Koonin EV, Senkevich TG, Dolja VV. The ancient Virus World and evolution of cells. *Biol Direct*. 2006;1(1):29. doi:10.1186/1745-6150-1-29.

12. Holmes EC. What Does Virus Evolution Tell Us about Virus Origins? *J Virol*. 2011;85(11):5247-5251. doi:10.1128/JVI.02203-10.

13. Villarreal LP, DeFilippis VR. A hypothesis for DNA viruses as the origin of eukaryotic replication proteins. *J Virol*. 2000;74(15):7079-7084.

14. Forterre P, Gaïa M. Giant viruses and the origin of modern eukaryotes. *Current Opinion in Microbiology*. 2016;31:44-49. doi:10.1016/j.mib.2016.02.001.

15. Adams MJ, Lefkowitz EJ, King AMQ, et al. Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the

International Committee on Taxonomy of Viruses (2017). *Arch Virol*. 2017;162(8):2505-2538. doi:10.1007/s00705-017-3358-5.

16. Isaacs SN. Working safely with vaccinia virus: laboratory technique and the role of vaccinia vaccination. *Methods Mol Biol*. 2004;269(Chapter 1):1-14. doi:10.1385/1-59259-789-0:001.

17. Hughes AL, Friedman R. Poxvirus genome evolution by gene gain and loss. *Mol Phylogenet Evol*. 2005;35(1):186-195. doi:10.1016/j.ympev.2004.12.008.

18. Spiesschaert B, McFadden G, Hermans K, Nauwynck H, Van de Walle GR. The current status and future directions of myxoma virus, a master in immune evasion. *Vet Res*. 2011;42(1):76. doi:10.1186/1297-9716-42-76.

19. Kerr PJ. Myxomatosis in Australia and Europe: a model for emerging infectious diseases. *Antiviral Res*. 2012;93(3):387-415. doi:10.1016/j.antiviral.2012.01.009.

20. Kerr PJ, Ghedin E, DePasse JV, et al. Evolutionary history and attenuation of myxoma virus on two continents. *PLoS Pathogens*. 2012;8(10):e1002950. doi:10.1371/journal.ppat.1002950.

21. Kerr PJ, Liu J, Cattadori I, Ghedin E, Read AF, Holmes EC. Myxoma Virus and the Leporipoxviruses: An Evolutionary Paradigm. *Viruses 2015, Vol 7, Pages 5571-5586*. 2015;7(3):1020-1061. doi:10.3390/v7031020.

22. Burgess HM, Mohr I. Evolutionary clash between myxoma virus and rabbit PKR in Australia. *Proc Natl Acad Sci USA*. 2016;113(15):3912-3914. doi:10.1073/pnas.1602063113.

23. Kerr PJ, Cattadori IM, Liu J, et al. Next step in the ongoing arms race between myxoma virus and wild rabbits in Australia is a novel disease phenotype. *Proc Natl Acad Sci USA*. 2017;114(35):9397-9402. doi:10.1073/pnas.1710336114.

24. Afonso PP, Silva PM, Schnellrath LC, et al. Biological characterization and next-generation genome sequencing of the unclassified Cotia virus SPAn232 (Poxviridae). *J Virol*. 2012;86(9):JVI.07162–11–5054. doi:10.1128/JVI.07162-11.

25. Shackelton LA, Parrish CR, Holmes EC. Evolutionary Basis of Codon Usage and Nucleotide Composition Bias in Vertebrate DNA Viruses. *J Mol Evol*. 2006;62(5):551-563. doi:10.1007/s00239-005-0221-1.

26. Gelaye E, Mach L, Kolodziejek J, et al. A novel HRM assay for the simultaneous detection and differentiation of eight poxviruses of medical and veterinary importance. *Scientific Reports*. 2017;7:42892. doi:10.1038/srep42892.

27. Wittek R, Kuenzle CC, Wyler R. High C + G Content in Parapoxvirus DNA. *J Gen Virol*. 1979;43(1):231-234. doi:10.1099/0022-1317-43-1-231.

28. RoyChoudhury S, Pan A, Mukherjee D. Genus specific evolution of codon usage and nucleotide compositional traits of poxviruses. *Virus Genes*. 2011;42(2):189-199. doi:10.1007/s11262-010-0568-2.

29. Moss B, Shisler JL, Xiang Y, Senkevich TG. Immune-defense molecules of Molluscum contagiosum virus, a human poxvirus. *Trends Microbiol*. 2000;8(10):473-477.

30. Kuhl JT, Huerter CJ, Hashish H. A case of human orf contracted from a deer. *Cutis*. 2003;71(4):288-290.

31. MacNeil A, Lederman E, Reynolds MG, et al. Diagnosis of Bovine-Associated Parapoxvirus Infections in Humans: Molecular and Epidemiological Evidence. *Zoonoses Public Health*. 2010;57(7-8):E161-E164. doi:10.1111/j.1863-2378.2009.01317.x.

32. Centers for Disease Control and Prevention (CDC). Human Orf virus infection from household exposures - United States, 2009-2011. *MMWR Morb Mortal Wkly Rep*. 2012;61(14):245-248.

33. Oğuzoğlu TÇ, Koç BT, Kirdeci A, Tan MT. Evidence of zoonotic pseudocowpox virus infection from a cattle in Turkey. *Virusdisease*. 2014;25(3):381-384. doi:10.1007/s13337-014-0214-z.

34. Zhang K, Liu Y, Kong H, Shang Y, Liu X. Human infection with ORF virus from goats in China, 2012. *Vector Borne Zoonotic Dis*. 2014;14(5):365-367. doi:10.1089/vbz.2013.1445.

35. Huemer HP, Zobl A, Glawischnig W, Romani N, Trevisiol K, Kitchen M. Zoonotic parapoxvirus infection associated with game animals in the Tyrol region of Austria and Italy. *Journal of Clinical Virology*. 2015;70:S27-S27. doi:10.1016/j.jcv.2015.07.068.

36. Sainsbury AW, Deaville R, Lawson B, et al. Poxviral Disease in Red Squirrels Sciurus vulgaris in the UK: Spatial and Temporal Trends of an Emerging Threat. *Ecohealth*. 2008;5(3):305-316. doi:10.1007/s10393-008-0191-z.

37. Simpson VR, Hargreaves J, Butler HM, Davison NJ, Everest DJ. Causes of mortality and pathological lesions observed post-mortem in red squirrels (Sciurus vulgaris) in Great Britain. *BMC Vet Res*. 2013;9(1):229. doi:10.1186/1746-6148-9-229.

38. Chantrey J, Dale TD, Read JM, et al. European red squirrel population dynamics driven by squirrelpox at a gray squirrel invasion interface. *Ecol Evol*. 2014;4(19):3788-3799. doi:10.1002/ece3.1216.

39. Parker PG, Buckles EL, Farrington H, et al. 110 years of Avipoxvirus in the Galapagos Islands. Davis T, ed. *PLoS ONE*. 2011;6(1):e15989. doi:10.1371/journal.pone.0015989.

40. Offerman K, Carulei O, van der Walt AP, Douglass N, Williamson A-L. The complete genome sequences of poxviruses isolated from a penguin and a pigeon in South Africa and comparison to other sequenced avipoxviruses. *BMC Genomics*. 2014;15(1):1. doi:10.1186/1471-2164-15-463.

41. Ruiz-Martínez J, Ferraguti M, Figuerola J, et al. Prevalence and Genetic Diversity of Avipoxvirus in House Sparrows in Spain. Moreno-Rueda G, ed. *PLoS ONE*.

2016;11(12):e0168690. doi:10.1371/journal.pone.0168690.

42. Sarker S, Das S, Lavers JL, et al. Genomic characterization of two novel pathogenic avipoxviruses isolated from pacific shearwaters (Ardenna spp.). *BMC Genomics*. 2017;18(1):298. doi:10.1186/s12864-017-3680-z.

43. Carulei O, Douglass N, Williamson A-L. Comparative analysis of avian poxvirus genomes, including a novel poxvirus from lesser flamingos (Phoenicopterus minor), highlights the lack of conservation of the central region. *BMC Genomics*. 2017;18(1):947. doi:10.1186/s12864-017-4315-0.

44. Fenner F. Smallpox and Its Eradication. *Australian and New Zealand Journal of Medicine*. 1980;10(4):459-459.

45. Littman RJ. The Plague of Athens: Epidemiology and Paleopathology. *Mt Sinai J Med*. 2009;76(5):456-467. doi:10.1002/msj.20137.

46. Parrish CR, Holmes EC, Morens DM, et al. Cross-Species Virus Transmission and the Emergence of New Epidemic Diseases. *Microbiology and Molecular Biology Reviews*. 2008;72(3):457-470. doi:10.1128/MMBR.00004-08.

47. Babkin IV, Shchelkunov SN. Molecular evolution of poxviruses. *Russ J Genet*. 2008;44(8):895-908. doi:10.1134/S1022795408080036.

48. Shchelkunov SN. How long ago did smallpox virus emerge? *Arch Virol*. 2009;154(12):1865-1871. doi:10.1007/s00705-009-0536-0.

49. Hughes AL, Irausquin S, Friedman R. The evolutionary biology of poxviruses. *Infection, Genetics and Evolution*. 2010;10(1):50-59. doi:10.1016/j.meegid.2009.10.001.

50. Firth C, Kitchen A, Shapiro B, Suchard MA, Holmes EC, Rambaut A. Using Time-Structured Data to Estimate Evolutionary Rates of Double-Stranded DNA Viruses. *Molecular Biology and Evolution*. 2010;27(9):2038-2051. doi:10.1093/molbev/msq088.

51. Babkin IV, Babkina IN. A retrospective study of the orthopoxvirus molecular evolution. *Infect Genet Evol*. 2012;12(8):1597-1604. doi:10.1016/j.meegid.2012.07.011.

52. Babkin I, Babkina I. The Origin of the Variola Virus. *Viruses 2015, Vol 7, Pages 5571-5586*. 2015;7(3):1100-1112. doi:10.3390/v7031100.

53. Lourie B, Nakano JH, Kemp GE, Setzer HW. Isolation of Poxvirus From an African Rodent. *J Infect Dis*. 1975;132(6):677-681.

54. Smithson C, Purdy A, Verster AJ, Upton C. Prediction of steps in the evolution of variola virus host range. Meng X, ed. *PLoS ONE*. 2014;9(3):e91520. doi:10.1371/journal.pone.0091520.

55. Smith KA. Edward Jenner and the small pox vaccine. *Front Immunol*. 2011;2. doi:10.3389/fimmu.2011.00021.

56. Moss B. Vaccinia Virus - a Tool for Research and Vaccine Development. *Science*. 1991;252(5013):1662-1667.

57. Tolonen N, Doglio L, Schleich S, Locker JK. Vaccinia virus DNA replication occurs in endoplasmic reticulum-enclosed cytoplasmic mini-nuclei. *Mol Biol Cell*. 2001;12(7):2031-2046.

58. Broyles SS. Vaccinia virus transcription. *J Gen Virol*. 2003;84(Pt 9):2293-2303. doi:10.1099/vir.0.18942-0.

59. Roper RL. Characterization of the vaccinia virus A35R protein and its role in virulence. *J Virol*. 2006;80(1):306-313. doi:10.1128/JVI.80.1.306-313.2006.

60. Siqueira Ferreira JM, Abrahao JS, Drumond BP, et al. Vaccinia virus: shedding and horizontal transmission in a murine model. *J Gen Virol*. 2008;89(Pt 12):2986-2991. doi:10.1099/vir.0.2008/003947-0.

61. Senkevich TG, Wyatt LS, Weisberg AS, Koonin EV, Moss B. A conserved poxvirus NlpC/P60 superfamily protein contributes to vaccinia virus virulence in mice but not to replication in cell culture. *Virology*. 2008;374(2):506-514. doi:10.1016/j.virol.2008.01.009.

62. Unterholzner L, Sumner RP, Baran M, et al. Vaccinia Virus Protein C6 Is a Virulence Factor that Binds TBK-1 Adaptor Proteins and Inhibits Activation of IRF3 and IRF7. Buller RML, ed. *PLoS Pathogens*. 2011;7(9). doi:10.1371/journal.ppat.1002247.

63. Smith GL, Benfield CTO, de Motes CM, et al. Vaccinia virus immune evasion: mechanisms, virulence and immunogenicity. *J Gen Virol*. 2013;94(Pt 11):2367-2392. doi:10.1099/vir.0.055921-0.

64. Liu S-W, Katsafanas GC, Liu R, Wyatt LS, Moss B. Poxvirus Decapping Enzymes Enhance Virulence by Preventing the Accumulation of dsRNA and the Induction of Innate Antiviral Responses. *Cell Host & Microbe*. 2015;17(3):320-331. doi:10.1016/j.chom.2015.02.002.

65. Braunlar L, J Ostrowski. Vaccinia Virus Transcription System - Tool for Studying Mode of Action of Drugs. *Naunyn Schmiedebergs Arch Pharmacol*. 1972;272(4):363–.

66. Yang Z, Reynolds SE, Martens CA, Bruno DP, Porcella SF, Moss B. Expression Profiling of the Intermediate and Late Stages of Poxvirus Replication. *J Virol*. 2011;85(19):9899-9908. doi:10.1128/JVI.05446-11.

67. Yang Z, Cao S, Martens CA, et al. Deciphering Poxvirus Gene Expression by RNA Sequencing and Ribosome Profiling. McFadden G, ed. *J Virol*. 2015;89(13):6874-6886. doi:10.1128/JVI.00528-15.

68. Alcamí A, Smith GL. Receptors for gamma-interferon encoded by poxviruses: implications for the unknown origin of vaccinia virus. *Trends Microbiol*. 1996;4(8):321-326.

69. Tulman ER, Delhon G, Afonso CL, et al. Genome of horsepox virus. *J Virol*. 2006;80(18):9244-9258. doi:10.1128/JVI.00945-06.

70. Henderson DA, Moss B. Public Health. 1999.

71. Baxby D. Poxviruses. 1996.

72. Domingo E. Mechanisms of viral emergence. *Vet Res*. 2010;41(6):38. doi:10.1051/vetres/2010010.

73. Institute of Medicine (US) Committee on Emerging Microbial Threats to Health in the 21st Century, Smolinski MS, Hamburg MA, Lederberg J. Microbial Threats to Health: Emergence, Detection, and Response. 2003. doi:10.17226/10636.

74. Morse SS. The public health threat of emerging viral disease. *Journal of Nutrition*. 1997;127:S951-S957.

75. Wang LF, Crameri G. Emerging zoonotic viral diseases. *Rev - Off Int Epizoot*. 2014;33(2):569-581.

76. Reynolds MG, Carroll DS, Karem KL. Factors affecting the likelihood of monkeypox's emergence and spread in the post-smallpox era. *Curr Opin Virol*. 2012;2(3):335-343. doi:10.1016/j.coviro.2012.02.004.

77. Sorci G, Cornet S, Faivre B. Immunity and the emergence of virulent pathogens. *Infect Genet Evol*. 2013;16:441-446. doi:10.1016/j.meegid.2012.12.031.

78. Antia R, Regoes RR, Koella JC, Bergstrom CT. The role of evolution in the emergence of infectious diseases. *Nature*. 2003;426(6967):658-661. doi:10.1038/nature02104.

79. Vorou RM, Papavassiliou VG, Pierroutsakos IN. Cowpox virus infection: an emerging health threat. *Curr Opin Infect Dis*. 2008;21(2):153-156. doi:10.1097/QCO.0b013e3282f44c74.

80. Damaso CRA, Esposito JJ, Condit RC, Moussatché N. An Emergent Poxvirus from Humans and Cattle in Rio de Janeiro State: Cantagalo Virus May Derive from Brazilian Smallpox Vaccine. *Virology*. 2000;277(2):439-449. doi:10.1006/viro.2000.0603.

81. de Souza Trindade G, Drumond BP, Guedes MIMC, et al. Zoonotic vaccinia virus infection in Brazil: clinical description and implications for health professionals. *J Clin Microbiol*. 2007;45(4):1370-1372. doi:10.1128/JCM.00920-06.

82. Megid J, Borges IA, Abrahão JS, et al. Vaccinia Virus Zoonotic Infection, São Paulo State, Brazil. *Emerg Infect Dis*. 2011;18(1):189-191. doi:10.3201/eid1801.110692.

83. de Assis FL, Vinhote WM, Barbosa JD, et al. Reemergence of Vaccinia Virus during Zoonotic Outbreak, Pará State, Brazil. *Emerg Infect Dis*. 2013;19(12):2017-2020. doi:10.3201/eid1912.130589.

84. Abrahao JS, Campos RK, Trindade G de S, Guimarães da Fonseca F, Ferreira PCP,

Kroon EG. Outbreak of Severe Zoonotic Vaccinia Virus Infection, Southeastern Brazil. *Emerg Infect Dis*. 2015;21(4):695-698. doi:10.3201/eid2104.140351.

85. Dhar AD, Werchniak AE, Li Y, et al. Tanapox infection in a college student. *N Engl J Med*. 2004;350(4):361-366. doi:10.1056/NEJMoa031467.

86. Clark C, McIntyre PG, Evans A, McInnes CJ, Lewis-Jones S. Human sealpox resulting from a seal bite: confirmation that sealpox virus is zoonotic. *British Journal of Dermatology*. 2005;152(4):791-793. doi:10.1111/j.1365-2133.2005.06451.x.

87. Osadebe LU, Manthiram K, McCollum AM, et al. Novel poxvirus infection in 2 patients from the United States. *Clin Infect Dis*. 2015;60(2):195-202. doi:10.1093/cid/ciu790.

88. Lakis NS, Li Y, Abraham JL, et al. Novel Poxvirus Infection in an Immune Suppressed Patient. *Clin Infect Dis*. 2015;61(10):1543-1548. doi:10.1093/cid/civ643.

89. Smithson C, Meyer H, Gigante CM, et al. Two novel poxviruses with unusual genome rearrangements: NY_014 and Murmansk. *Virus Genes*. 2017;53(6):883-897. doi:10.1007/s11262-017-1501-8.

90. Henderson DA, Inglesby TV, Bartlett JG, et al. Smallpox as a Biological Weapon: Medical and Public Health Management. *JAMA*. 1999;281(22):2127-2137. doi:10.1001/jama.281.22.2127.

91. Wallin A, Luksiene Z, Zagminas K, Surkiene G. Public health and bioterrorism: renewed threat of anthrax and smallpox. *Medicina (Kaunas)*. 2007;43(4):278-284.

92. Margolis AR, Grabenstein JD. Immunizations against bioterrorism: Smallpox and anthrax. *J Am Pharm Assoc (2003)*. 2009;49(4):566-568. doi:10.1331/JAPhA.2009.09522.

93. Noyce RS, Lederman S, Evans DH. Construction of an infectious horsepox virus vaccine from chemically synthesized DNA fragments. Thiel V, ed. *PLoS ONE*. 2018;13(1):e0188453. doi:10.1371/journal.pone.0188453.

94. Wang L-F. Discovering novel zoonotic viruses. *N S W Public Health Bull*. 2011;22(5-6):113-117. doi:10.1071/NB10078.

95. Rosenberg R. Detecting the emergence of novel, zoonotic viruses pathogenic to humans. *Cell Mol Life Sci*. 2015;72(6):1115-1125. doi:10.1007/s00018-014-1785-y.

96. Molina C, Earn DJD. Game theory of pre-emptive vaccination before bioterrorism or accidental release of smallpox. *J R Soc Interface*. 2015;12(107):20141387-20141387. doi:10.1098/rsif.2014.1387.

97. Koblentz GD. The De Novo Synthesis of Horsepox Virus: Implications for Biosecurity and Recommendations for Preventing the Reemergence of Smallpox. - PubMed - NCBI. *Health Security*. 2017;15(6):620-628.

98. Li G, Chen N, Feng Z, et al. Genomic sequence and analysis of a vaccinia virus isolate from a patient with a smallpox vaccine-related complication. *Virol J*.

2006;3(1):88. doi:10.1186/1743-422X-3-88.

99. Sánchez-Sampedro L, Perdiguero B, Mejías-Pérez E, García-Arriaza J, Di Pilato M, Esteban M. The Evolution of Poxvirus Vaccines. *Viruses 2015, Vol 7, Pages 5571-5586*. 2015;7(4):1726-1803. doi:10.3390/v7041726.

100. Meyer H, Sutter G, Mayr A. Mapping of deletions in the genome of the highly attenuated vaccinia virus MVA and their influence on virulence. - PubMed - NCBI. *J Gen Virol*. 1991;72(5):1031-1038. doi:10.1099/0022-1317-72-5-1031.

101. Kaufman HL. The role of poxviruses in tumor immunotherapy. *Surgery*. 2003;134(5):731-737. doi:10.1016/S0039-6060(03)00294-0.

102. Kirn DH, Thorne SH. Targeted and armed oncolytic poxviruses: a novel multi-mechanistic therapeutic class for cancer. - PubMed - NCBI. *Nat Rev Cancer*. 2009;9(1):64-71. doi:10.1038/nrc2545.

103. Gómez CE, Nájera JL, Krupa M, Perdiguero B, Esteban M. MVA and NYVAC as vaccines against emergent infectious diseases and cancer. *Curr Gene Ther*. 2011;11(3):189-217.

104. Schlom J. Therapeutic cancer vaccines: current status and moving forward. - PubMed - NCBI. *JNCI Journal of the National Cancer Institute*. 2012;104(8):599-613. doi:10.1093/jnci/djs033.

105. Izzi V, Buler M, Masuelli L, Giganti MG, Modesti A, Bei R. Poxvirus-based vaccines for cancer immunotherapy: new insights from combined cytokines/co-stimulatory molecules delivery and “uncommon” strains. *Anticancer Agents Med Chem*. 2014;14(2):183-189.

106. Konishi E, Pincus S, Paoletti E, Shope RE, Wason PW. Avipox virus-vectored Japanese encephalitis virus vaccines: use as vaccine candidates in combination with purified subunit immunogens. *Vaccine*. 1994;12(7):633-638.

107. Pacchioni S, Volonté L, Zanotto C, Pozzi E, De Giuli Morghen C, Radaelli A. Canarypox and fowlpox viruses as recombinant vaccine vectors: an ultrastructural comparative analysis. *Arch Virol*. 2010;155(6):915-924. doi:10.1007/s00705-010-0663-7.

108. Simon C Weli MT. Avipoxviruses: infection biology and their use as vaccine vectors. *Virol J*. 2011;8(1):49. doi:10.1186/1743-422X-8-49.

109. Bissa M, Zanotto C, Pacchioni S, et al. The L1 protein of human papilloma virus 16 expressed by a fowlpox virus recombinant can assemble into virus-like particles in mammalian cell lines but elicits a non-neutralising humoral response. *Antiviral Res*. 2015;116:67-75. doi:10.1016/j.antiviral.2015.01.012.

110. Duggan AT, Perdomo MF, Piombino-Mascali D, et al. 17 th Century Variola Virus Reveals the Recent History of Smallpox. *Current Biology*. 2016;26(24):3407-3412. doi:10.1016/j.cub.2016.10.061.

111. Pajer P, Dresler J, Kabíckova H, et al. Characterization of Two Historic Smallpox

Specimens from a Czech Museum. *Viruses 2015, Vol 7, Pages 5571-5586*. 2017;9(8):200. doi:10.3390/v9080200.

112. Smithson C, Imbery J, Upton C. Re-Assembly and Analysis of an Ancient Variola Virus Genome. *Viruses*. 2017;9(9). doi:10.3390/v9090253.

113. Kumar S, Subramanian S. Mutation rates in mammalian genomes. *PNAS*. 2002;99(2):803-808. doi:10.1073/pnas.022629899.

114. Drake JW, Holland JJ. Mutation rates among RNA viruses. *PNAS*. 1999;96(24):13910-13913.

115. Babkin IV, Nepomnyashchikh TS, Maksyutov RA, Gutorov VV, Babkina IN, Shchelkunov SN. Comparative analysis of variable regions in the variola virus genome. *Mol Biol*. 2008;42(4):543-553. doi:10.1134/S0026893308040092.

116. Hendrickson RC, Wang C, Hatcher EL, Lefkowitz EJ. Orthopoxvirus Genome Evolution: The Role of Gene Loss. *Viruses 2015, Vol 7, Pages 5571-5586*. 2010;2(9):1933-1967. doi:10.3390/v2091933.

117. Reynolds MG, Guagliardo SAJ, Nakazawa YJ, Doty JB, Mauldin MR. Understanding orthopoxvirus host range and evolution: from the enigmatic to the usual suspects. *Curr Opin Virol*. 2018;28:108-115. doi:10.1016/j.coviro.2017.11.012.

118. Lefkowitz EJ, Wang C, Upton C. Poxviruses: past, present and future. 2006;117(1):105-118. doi:10.1016/j.virusres.2006.01.016.

119. McLysaght A, Baldi PF, Gaut BS. Extensive gene gain associated with adaptive evolution of poxviruses. *Proc Natl Acad Sci USA*. 2003;100(26):15655-15660. doi:10.1073/pnas.2136653100.

120. Elde NC, Child SJ, Eickbush MT, et al. Poxviruses Deploy Genomic Accordions to Adapt Rapidly against Host Antiviral Defenses. *Cell*. 2012;150(4):831-841. doi:10.1016/j.cell.2012.05.049.

121. Tulman ER, Afonso CL, Lu Z, Zsak L, Kutish GF, Rock DL. The genome of canarypox virus. *J Virol*. 2004;78(1):353-366. doi:10.1128/JVI.78.1.353-366.2004.

122. Gubser C, Hué S, Kellam P, Smith GL. Poxvirus genomes: a phylogenetic analysis. *J Gen Virol*. 2004;85(1):105-117. doi:10.1099/vir.0.19565-0.

123. Coulson D, Upton C. Characterization of indels in poxvirus genomes. *Virus Genes*. 2011;42(2):171-177. doi:10.1007/s11262-010-0560-x.

124. Camus-Bouclainville C, Gretillat M, Py R, Gelfi J, Guérin J-L, Bertagnoli S. Genome Sequence of SG33 Strain and Recombination between Wild-Type and Vaccine Myxoma Viruses. *Emerg Infect Dis*. 2011;17(4):633-638. doi:10.3201/eid1704.101146.

125. Qin L, Evans DH. Genome scale patterns of recombination between coinfecting vaccinia viruses. *J Virol*. 2014;88(10):5277-5286. doi:10.1128/JVI.00022-14.

126. Paszkowski P, Noyce RS, Evans DH. Live-Cell Imaging of Vaccinia Virus Recombination. Elde NC, ed. *PLoS Pathogens*. 2016;12(8). doi:10.1371/journal.ppat.1005824.

127. Bratke KA, McLysaght A. Identification of multiple independent horizontal gene transfers into poxviruses using a comparative genomics approach. *BMC Evolutionary Biology 2008 8:1*. 2008;8(1):1. doi:10.1186/1471-2148-8-67.

128. Odom MR, Hendrickson RC, Lefkowitz EJ. Poxvirus protein evolution: family wide assessment of possible horizontal gene transfer events. 2009;144(1-2):233-249. doi:10.1016/j.virusres.2009.05.006.

129. Campbell JA, Davis RS, Lilly LM, Fremont DH, French AR, Carayannopoulos LN. FCRL5 on Innate B Cells is Targeted by a Poxvirus MHC Class I-like Immunoevasin. *Journal of Immunology (Baltimore, Md : 1950)*. 2010;185(1):28-32. doi:10.4049/jimmunol.1000240.

130. Fleming SB, McCaughan CA, Andrews AE, Nash AD, Mercer AA. A homolog of interleukin-10 is encoded by the poxvirus orf virus. *J Virol*. 1997;71(6):4857-4861.

131. Sakala IG, Chaudhri G, Buller RM, et al. Poxvirus-Encoded Gamma Interferon Binding Protein Dampens the Host Immune Response to Infection. *J Virol*. 2007;81(7):3346-3353. doi:10.1128/JVI.01927-06.

132. Nuara AA, Walter LJ, Logsdon NJ, et al. Structure and mechanism of IFN-gamma antagonism by an orthopoxvirus IFN-gamma-binding protein. *PNAS*. 2008;105(6):1861-1866. doi:10.1073/pnas.0705753105.

133. Saraiva M, Alcamí A. CrmE, a novel soluble tumor necrosis factor receptor encoded by poxviruses. *J Virol*. 2001;75(1):226-233. doi:10.1128/JVI.75.1.226-233.2001.

134. Sedger LM, Osvath SR, Xu X-M, et al. Poxvirus tumor necrosis factor receptor (TNFR)-like T2 proteins contain a conserved preligand assembly domain that inhibits cellular TNFR1-induced cell death. *J Virol*. 2006;80(18):9300-9309. doi:10.1128/JVI.02449-05.

135. Pontejo SM, Alejo A, Alcami A. Poxvirus-encoded TNF decoy receptors inhibit the biological activity of transmembrane TNF. *J Gen Virol*. 2015;96(10):3118-3123. doi:10.1099/jgv.0.000255.

136. Gvakharia BO, Koonin EK, Mathews CK. Vaccinia virus G4L gene encodes a second glutaredoxin. *Virology*. 1996;226(2):408-411. doi:10.1006/viro.1996.0669.

137. Grossegesse M, Doellinger J, Haldemann B, Schaade L, Nitsche A. A Next-Generation Sequencing Approach Uncovers Viral Transcripts Incorporated in Poxvirus Virions. *Viruses*. 2017;9(10):296. doi:10.3390/v9100296.

138. Ching YC, Chung CS, Huang CY, Hsia Y, Tang YL, Chang W. Disulfide Bond Formation at the C Termini of Vaccinia Virus A26 and A27 Proteins Does Not Require Viral Redox Enzymes and Suppresses Glycosaminoglycan-Mediated Cell Fusion. *J Virol*. 2009;83(13):6464-6476. doi:10.1128/JVI.02295-08.

139. Chiu W-L, Lin C-L, Yang M-H, Tzou D-LM, Chang W. Vaccinia virus 4c (A26L) protein on intracellular mature virus binds to the extracellular cellular matrix laminin. *J Virol*. 2007;81(5):2149-2157. doi:10.1128/JVI.02302-06.

140. McFadden G. Poxvirus tropism. *Nature Reviews Microbiology*. 2005;3(3):201-213. doi:10.1038/nrmicro1099.

141. Moss B. Poxvirus entry and membrane fusion. *Virology*. 2006;344(1):48-54. doi:10.1016/j.virol.2005.09.037.

142. Moss B. Poxvirus Cell Entry: How Many Proteins Does it Take? *Viruses 2015, Vol 7, Pages 5571-5586*. 2012;4(5):688-707. doi:10.3390/v4050688.

143. Blasco R, Moss B. Role of cell-associated enveloped vaccinia virus in cell-to-cell spread. *J Virol*. 1992;66(7):4170-4179.

144. Doceul V, Hollinshead M, van der Linden L, Smith GL. Repulsion of Superinfecting Virions: A Mechanism for Rapid Virus Spread. *Science*. 2010;327(5967):873-876. doi:10.1126/science.1183173.

145. Condit RC. Surf and Turf: Mechanism of Enhanced Virus Spread During Poxvirus Infection. *Viruses 2015, Vol 7, Pages 5571-5586*. 2010;2(5):1050-1054. doi:10.3390/v2051050.

146. Kobiler O, Drayman N, Butin-Israeli V, Oppenheim A. Virus strategies for passing the nuclear envelope barrier. *Nucleus*. 2012;3(6):526-539. doi:10.4161/nucl.21979.

147. Moss B. Poxvirus DNA Replication. *Cold Spring Harbor Perspectives in Biology*. 2013;5(9):a010199-a010199. doi:10.1101/cshperspect.a010199.

148. Assarsson E, Greenbaum JA, Sundström M, et al. Kinetic analysis of a complete poxvirus transcriptome reveals an immediate-early class of genes. *Proc Natl Acad Sci USA*. 2008;105(6):2140-2145. doi:10.1073/pnas.0711573105.

149. Jha S, Rollins MG, Fuchs G, et al. Trans-kingdom mimicry underlies ribosome customization by a poxvirus kinase. *Nature*. 2017;546(7660):651-655. doi:10.1038/nature22814.

150. Dhungel P, Cao S, Yang Z. The 5'-poly(A) leader of poxvirus mRNA confers a translational advantage that can be achieved in cells with impaired cap-dependent translation. Evans DH, ed. *PLoS Pathogens*. 2017;13(8):e1006602. doi:10.1371/journal.ppat.1006602.

151. Smith GL, Vanderplasschen A, Law M. The formation and function of extracellular enveloped vaccinia virus. *J Gen Virol*. 2002;83(12):2915-2931. doi:10.1099/0022-1317-83-12-2915.

152. Li Y, Yuan S, Moyer RW. The non-permissive infection of insect (gypsy moth) LD-652 cells by Vaccinia virus. - PubMed - NCBI. *Virology*. 1998;248(1):74-82. doi:10.1006/viro.1998.9241.

153. Nash P, Barrett J, Cao JX, et al. Immunomodulation by viruses: the myxoma virus

story. *Immunol Rev*. 1999;168:103-120.

154. Johnston JB, McFadden G. Poxvirus Immunomodulatory Strategies: Current Perspectives. *J Virol*. 2003;77(11):6093-6100. doi:10.1128/JVI.77.11.6093-6100.2003.

155. Susanna R Bidgood JM. Cloak and Dagger: Alternative Immune Evasion and Modulation Strategies of Poxviruses. *Viruses 2015, Vol 7, Pages 5571-5586*. 2015;7(8):4800-4825. doi:10.3390/v7082844.

156. Guérin J-L, Gelfi J, Boullier S, et al. Myxoma virus leukemia-associated protein is responsible for major histocompatibility complex class I and Fas-CD95 down-regulation and defines scrapins, a new group of surface cellular receptor abductor proteins. *J Virol*. 2002;76(6):2912-2923.

157. Thome M, Schneider P, Hofmann K, et al. Viral FLICE-inhibitory proteins (FLIPs) prevent apoptosis induced by death receptors. *Nature*. 1997;386(6624):517-521. doi:10.1038/386517a0.

158. Daniel Nichols, William De Martini, Jessica Cottrell. Poxviruses Utilize Multiple Strategies to Inhibit Apoptosis. *Viruses*. 2017;9(8):215. doi:10.3390/v9080215.

159. Werden SJ, Rahman MM, McFadden G. Poxvirus host range genes. *Adv Virus Res*. 2008;71:135-171. doi:10.1016/S0065-3527(08)00003-1.

160. Haller SL, Peng C, McFadden G, Rothenburg S. Poxviruses and the evolution of host range and virulence. *Infection, Genetics and Evolution*. 2014;21:15-40. doi:10.1016/j.meegid.2013.10.014.

161. Chen W, Drillien R, Spehner D, Buller R. Restricted Replication of Ectromelia Virus in Cell-Culture Correlates with Mutations in Virus-Encoded Host Range Gene. *Virology*. 1992;187(2):433-442.

162. Ramseyewing A, Moss B. Restriction of Vaccinia Virus-Replication in Cho Cells Occurs at the Stage of Viral Intermediate Protein-Synthesis. *Virology*. 1995;206(2):984-993. doi:10.1006/viro.1995.1021.

163. Mossman K, Lee SF, Barry M, Boshkov L, McFadden G. Disruption of M-T5, a novel myxoma virus gene member of the poxvirus host range superfamily, results in dramatic attenuation of myxomatosis in infected European rabbits. *J Virol*. 1996;70(7):4394-4410.

164. Langland JO, Jacobs BL. The role of the PKR-inhibitory genes, E3L and K3L, in determining vaccinia virus host range. *Virology*. 2002;299(1):133-141. doi:10.1006/viro.2002.1479.

165. Oliveira GP, Rodrigues RAL, Lima MT, Drumond BP, Abrahao JS. Poxvirus Host Range Genes and Virus–Host Spectrum: A Critical Review. *Viruses 2015, Vol 7, Pages 5571-5586*. 2017;9(11):331. doi:10.3390/v9110331.

166. Pevsner J. *Bioinformatics and Functional Genomics*. Third edition. Hoboken, NJ: Wiley-Blackwell; 2015.

167. Kanehisa M, Bork P. Bioinformatics in the post-sequence era. *Nat Genet*. 2003;33(3s):305-310. doi:10.1038/ng1109.

168. Zhang J, Chiodini R, Badr A, Zhang G. The impact of next-generation sequencing on genomics. *J Genet Genomics*. 2011;38(3):95-109. doi:10.1016/j.jgg.2011.02.003.

169. Dolled-Filhart MP, Lee MJ, Ou-yang C-W, Haraksingh RR, Lin JC-H. Computational and Bioinformatics Frameworks for Next-Generation Whole Exome and Genome Sequencing. *ScientificWorldJournal*. 2013;2013(20):–10. doi:10.1155/2013/730210.

170. Li Z, Chen Y, Mu D, et al. Comparison of the two major classes of assembly algorithms: overlap-layout-consensus and de-bruijn-graph. *Brief Funct Genomics*. 2012;11(1):25-37. doi:10.1093/bfgp/elr035.

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

172. Da Silva M, Upton C. Host-derived pathogenicity islands in poxviruses. *Virol J*. 2005;2(1). doi:10.1186/1743-422X-2-30.

173. Bao YM, Federhen S, Leipe D, et al. National Center for Biotechnology Information Viral Genomes Project. *J Virol*. 2004;78(14):7291-7298. doi:10.1128/JVI.78.14.7291-7298.2004.

174. Sivashankari S, Shanmughavel P. Functional annotation of hypothetical proteins - A review. *Bioinformation*. 2006;1(8):335-338.

175. Davison AJ, Moss B. Structure of Vaccinia Virus Early Promoters. *J Mol Biol*. 1989;210(4):749-769.

176. Upton C. Screening predicted coding regions in poxvirus genomes. *Virus Genes*. 2000;20(2):159-164.

177. Da Silva M, Upton C. Using purine skews to predict genes in AT-rich poxviruses. *BMC Genomics*. 2005;6(1). doi:10.1186/1471-2164-6-22.

178. Rost B. Twilight zone of protein sequence alignments. *Protein Eng*. 1999;12(2):85-94.

179. Li Y, Carroll DS, Gardner SN, Walsh MC, Vitalis EA, Damon IK. On the origin of smallpox: correlating variola phylogenics with historical smallpox records. *Proc Natl Acad Sci USA*. 2007;104(40):15787-15792. doi:10.1073/pnas.0609268104.

180. Xing K, Deng RQ, Wang JW, Feng JH, Huang MS, Wang XZ. Genome-based phylogeny of poxvirus. *Intervirology*. 2006;49(4):207-214. doi:10.1159/000090790.

181. Afonso CL, Delhon G, Tulman ER, et al. Genome of deerpox virus. *J Virol*. 2005;79(2):966-977. doi:10.1128/JVI.79.2.966-977.2005.

182. Le SQ, Gascuel O. An improved general amino acid replacement matrix. *Molecular*

Biology and Evolution. 2008;25(7):1307-1320. doi:10.1093/molbev/msn067.

183. Ehlers A, Osborne J, Slack S, Roper RL, Upton C. Poxvirus Orthologous Clusters (POCs). *Bioinformatics*. 2002;18(11):1544-1545. doi:10.1093/bioinformatics/18.11.1544.

184. Upton C, Slack S, Hunter AL, Ehlers A, Roper RL. Poxvirus orthologous clusters: toward defining the minimum essential poxvirus genome. *J Virol*. 2003;77(13):7590-7600. doi:10.1128/JVI.77.13.7590-7600.2003.

185. Tcherepanov V, Ehlers A, Upton C. Genome Annotation Transfer Utility (GATU): rapid annotation of viral genomes using a closely related reference genome. *BMC Genomics*. 2006;7(1):150. doi:10.1186/1471-2164-7-150.

186. Brodie R, Roper RL, Upton C. JDotter: a Java interface to multiple dotplots generated by dotter. *Bioinformatics*. 2004;20(2):279-281. doi:10.1093/bioinformatics/btg406.

187. Hillary W, Lin S-H, Upton C. Base-By-Base version 2: single nucleotide-level analysis of whole viral genome alignments. *Microb Inform Exp*. 2011;1(1):2. doi:10.1186/2042-5783-1-2.

188. Upton C, Hogg D, Perrin D, Boone M, Harris NL. Viral genome organizer: a system for analyzing complete viral genomes. *Virus Research*. 2000;70(1-2):55-64.

189. Flygare S, Simmon K, Miller C, et al. Taxonomer: an interactive metagenomics analysis portal for universal pathogen detection and host mRNA expression profiling. *Genome Biology 2016 17:1*. 2016;17(1):111. doi:10.1186/s13059-016-0969-1.

190. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760. doi:10.1093/bioinformatics/btp324.

191. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079. doi:10.1093/bioinformatics/btp352.

192. Breese MR, Liu Y. NGSUtils: a software suite for analyzing and manipulating next-generation sequencing datasets. *Bioinformatics*. 2013;29(4):494-496. doi:10.1093/bioinformatics/bts731.

193. Xu H, Luo X, Qian J, et al. FastUniq: A Fast De Novo Duplicates Removal Tool for Paired Short Reads. Doucet D, ed. *PLoS ONE*. 2012;7(12). doi:10.1371/journal.pone.0052249.

194. Bankevich A, Nurk S, Antipov D, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol*. 2012;19(5):455-477. doi:10.1089/cmb.2012.0021.

195. Chevreux B. MIRA: an automated genome and EST assembler. 2007.

196. Milne I, Bayer M, Stephen G, Cardle L, Marshall D. Tablet: Visualizing Next-

Generation Sequence Assemblies and Mappings. *Methods Mol Biol*. 2016;1374(Chapter 14):253-268. doi:10.1007/978-1-4939-3167-5_14.

197. Marchler-Bauer A, Derbyshire MK, Gonzales NR, et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res*. 2015;43(Database issue):D222-D226. doi:10.1093/nar/gku1221.

198. de Castro E, Sigrist CJA, Gattiker A, et al. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res*. 2006;34(Web Server):W362-W365. doi:10.1093/nar/gkl124.

199. Orchard SE, Mulder N, Apweiler R. The InterPro database. *Molecular & Cellular Proteomics*. 2004;3(10):S258-S258.

200. Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res*. 2005;33(Web Server issue):W244-W248. doi:10.1093/nar/gki408.

201. Zimmermann L, Stephens A, Nam S-Z, et al. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J Mol Biol*. December 2017. doi:10.1016/j.jmb.2017.12.007.

202. Dobson L, Reményi I, Tusnády GE. CCTOP: a Consensus Constrained TOPology prediction web server. *Nucleic Acids Res*. 2015;43(W1):W408-W412. doi:10.1093/nar/gkv451.

203. Kaell L, Krogh A, Sonnhammer ELL. Advantages of combined transmembrane topology and signal peptide prediction - the Phobius web server. *Nucleic Acids Res*. 2007;35(Web Server):W429-W432. doi:10.1093/nar/gkm256.

204. Frank K, Sippl MJ. High-performance signal peptide prediction based on sequence alignment techniques. *Bioinformatics*. 2008;24(19):2172-2176. doi:10.1093/bioinformatics/btn422.

205. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*. 2008;9(1). doi:10.1186/1471-2105-9-40.

206. DeLano WL. PyMOL molecular viewer: Updates and refinements. *Abstracts of Papers of the American Chemical Society*. 2009;238.

207. Pettersen EF, Goddard TD, Huang CC, et al. UCSF chimera - A visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25(13):1605-1612. doi:10.1002/jcc.20084.

208. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002;30(14):3059-3066.

209. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. 2004;5(1):113. doi:10.1186/1471-2105-5-113.

210. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25(15):1972-1973. doi:10.1093/bioinformatics/btp348.

211. Anderson FE, Swofford DL. Should we be worried about long-branch attraction in real data sets? Investigations using metazoan 18S rDNA. *Mol Phylogenet Evol*. 2004;33(2):440-451. doi:10.1016/j.ympev.20041.06.015.

212. Shackelton LA, Holmes EC. The evolution of large DNA viruses: combining genomic information of viruses and their hosts. *Trends Microbiol*. 2004;12(10):458-465. doi:10.1016/j.tim.2004.08.005.

213. Stamatakis A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006;22(21):2688-2690. doi:10.1093/bioinformatics/btl446.

214. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*. 2013;30(12):2725-2729. doi:10.1093/molbev/mst197.

215. Afonso CL, Tulman ER, Lu Z, et al. The genome of camelpox virus. *Virology*. 2002;295(1):1-9. doi:10.1006/viro.2001.1343.

216. Gubser C, Smith GL. The sequence of camelpox virus shows it is most closely related to variola virus, the cause of smallpox. *J Gen Virol*. 2002;83(Pt 4):855-872. doi:10.1099/0022-1317-83-4-855.

217. Huelsenbeck JP, Bull JJ. A likelihood ratio test to detect conflicting phylogenetic signal. *Systematic Biology*. 1996;45(1):92-98. doi:10.2307/2413514.

218. Hillis DM, Bull JJ. An Empirical-Test of Bootstrapping as a Method for Assessing Confidence in Phylogenetic Analysis. *Systematic Biology*. 1993;42(2):182-192. doi:10.2307/2992540.

219. Emerson GL, Nordhausen R, Garner MM, et al. Novel Poxvirus in Big Brown Bats, Northwestern United States. *Emerg Infect Dis*. 2013;19(6). doi:10.3201/eid1906.121713.

220. Meredith RW, Westerman M, Springer MS. A phylogeny and timescale for the living genera of kangaroos and kin (Macropodiformes : Marsupialia) based on nuclear DNA sequences. *Australian Journal of Zoology*. 2008;56(6):395-410. doi:10.1071/ZO08044.

221. Kate S Baker PRM. Poxviruses in Bats … so What? *Viruses 2015, Vol 7, Pages 5571-5586*. 2014;6(4):1564-1577. doi:10.3390/v6041564.

222. Agnarsson I, Zambrana-Torrelio CM, Flores-Saldana NP, May-Collado LJ. A time-calibrated species-level phylogeny of bats (Chiroptera, Mammalia). *PLoS Curr*. 2011;3:RRN1212. doi:10.1371/currents.RRN1212.

223. Baker KS, Leggett RM, Bexfield NH, et al. Metagenomic study of the viruses of African straw-coloured fruit bats: Detection of a chiropteran poxvirus and isolation

of a novel adenovirus. *Virology*. 2013;441(2):95-106. doi:10.1016/j.virol.2013.03.014.

224. Ge X-Y, Li J-L, Yang X-L, et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature*. 2013;503(7477):535-538. doi:10.1038/nature12711.

225. Memish ZA, Mishra N, Olival KJ, et al. Middle East Respiratory Syndrome Coronavirus in Bats, Saudi Arabia. *Emerg Infect Dis*. 2013;19(11). doi:10.3201/eid1911.131172.

226. Young PL, Halpin K, Mackenzie JS, Field HE. Isolation of Hendra virus from pteropid bats: a natural reservoir of Hendra virus. *J Gen Virol*. 2000;81(8):1927-1932. doi:10.1099/0022-1317-81-8-1927.

227. Chua KB, Lek Koh C, Hooi PS, et al. Isolation of Nipah virus from Malaysian Island flying-foxes. *Microbes and Infection*. 2002;4(2):145-151. doi:10.1016/S1286-4579(01)01522-2.

228. Leroy EM, Kumulungui B, Pourrut X, et al. Fruit bats as reservoirs of Ebola virus. *Nature*. 2005;438(7068):575-576. doi:10.1038/438575a.

229. Leroy EM, Epelboin A, Mondonge V, et al. Human Ebola Outbreak Resulting from Direct Exposure to Fruit Bats in Luebo, Democratic Republic of Congo, 2007. *Vector-Borne and Zoonotic Diseases*. 2009;9(6):723-728. doi:10.1089/vbz.2008.0167.

230. Gould AR, Hyatt AD, Lunt R, Kattenbelt JA, Hengstberger S, Blacksell SD. Characterisation of a novel lyssavirus isolated from Pteropid bats in Australia. 1998;54(2):165-187.

231. Dacheux L, Cervantes-Gonzalez M, Guigon G, et al. A preliminary study of viral metagenomics of French bat species in contact with humans: identification of new mammalian viruses. Tse H, ed. *PLoS ONE*. 2014;9(1):e87194. doi:10.1371/journal.pone.0087194.

232. Sano K, Okazaki S, Taniguchi S, et al. Detection of a novel herpesvirus from bats in the Philippines. *Virus Genes*. 2015;51(1):136-139. doi:10.1007/s11262-015-1197-6.

233. Teeling EC, Springer MS, Madsen O, Bates P, O'Brien SJ, Murphy WJ. A Molecular Phylogeny for Bats Illuminates Biogeography and the Fossil Record. *Science*. 2005;307(5709):580-584. doi:10.1126/science.1105113.

234. Han H-J, Wen H-L, Zhou C-M, et al. Bats as reservoirs of severe emerging infectious diseases. - PubMed - NCBI. 2015;205:1-6. doi:10.1016/j.virusres.2015.05.006.

235. Zhou P, Tachedjian M, Wynne JW, et al. Contraction of the type I IFN locus and unusual constitutive expression of IFN-α in bats. *Proc Natl Acad Sci USA*. 2016;113(10):2696-2701. doi:10.1073/pnas.1518240113.

236. McLelland DJ, Reardon T, Bourne S, Dickason C, Kessell A, Boardman W.

Outbreak of Skin Nodules Associated with Riouxgolvania Beveridgei (Nematoda: Muspiceida) in the Southern Bentwing Bat (Miniopterus Schreibersii Bassanii), South Australia. Vol 49. Wildlife Disease Association; 2013:1009-1013. doi:10.7589/2012-11-288.

237. Wiley SR, Schooley K, Smolak PJ, et al. Identification and characterization of a new member of the TNF family that induces apoptosis. *Immunity*. 1995;3(6):673-682.

238. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res*. 2004;32(Web Server issue):W526-W531. doi:10.1093/nar/gkh468.

239. Mongkolsapaya J, Grimes JM, Chen N, et al. Structure of the TRAIL-DR5 complex reveals mechanisms conferring specificity in apoptotic initiation. *Nat Struct Biol*. 1999;6(11):1048-1053. doi:10.1038/14935.

240. Hymowitz SG, Christinger HW, Fuh G, et al. Triggering Cell Death. *Molecular Cell*. 1999;4(4):563-571. doi:10.1016/S1097-2765(00)80207-5.

241. Hymowitz SG, O'Connell MP, Ultsch MH, et al. A unique zinc-binding site revealed by a high-resolution X-ray structure of homotrimeric Apo2L/TRAIL. *Biochemistry*. 2000;39(4):633-640. doi:10.1021/bi992242l.

242. Mavrommatis E, Fish EN, Platanias LC. The schlafen family of proteins and their regulation by interferons. *J Interferon Cytokine Res*. 2013;33(4):206-210. doi:10.1089/jir.2012.0133.

243. Chen N, Danila MI, Feng Z, et al. The genomic sequence of ectromelia virus, the causative agent of mousepox. *Virology*. 2003;317(1):165-186.

244. Zhao G, Droit L, Tesh RB, et al. The Genome of Yoka Poxvirus. *J Virol*. 2011;85(19):10230-10238. doi:10.1128/JVI.00637-11.

245. Afonso CL, Tulman ER, Lu Z, Oma E, Kutish GF, Rock DL. The genome of Melanoplus sanguinipes entomopoxvirus. *J Virol*. 1999;73(1):533-552.

246. Herbert MH, Squire CJ, Mercer AA. Poxviral Ankyrin Proteins. *Viruses 2015, Vol 7, Pages 5571-5586*. 2015;7(2):709-738. doi:10.3390/v7020709.

247. Mercer AA, Fleming SB, Ueda N. F-box-like domains are present in most poxvirus ankyrin repeat proteins. *Virus Genes*. 2005;31(2):127-133. doi:10.1007/s11262-005-1784-z.

248. Gammon DB, Gowrishankar B, Duraffour S, Andrei G, Upton C, Evans DH. Vaccinia virus-encoded ribonucleotide reductase subunits are differentially required for replication and pathogenesis. McFadden G, ed. *PLoS Pathogens*. 2010;6(7):e1000984. doi:10.1371/journal.ppat.1000984.

249. Gjessing MC, Yutin N, Tengs T, et al. Salmon Gill Poxvirus, the Deepest Representative of the Chordopoxvirinae. *J Virol*. 2015;89(18):9348-9367. doi:10.1128/JVI.01174-15.

250. Perdiguero B, Esteban M. The interferon system and vaccinia virus evasion mechanisms. *J Interferon Cytokine Res*. 2009;29(9):581-598. doi:10.1089/jir.2009.0073.

251. Ge X-Y, Wang N, Zhang W, et al. Coexistence of multiple coronaviruses in several bat colonies in an abandoned mineshaft. *Virol Sin*. 2016;31(1):31-40. doi:10.1007/s12250-016-3713-9.

252. Yang X-L, Hu B, Wang B, et al. Isolation and Characterization of a Novel Bat Coronavirus Closely Related to the Direct Progenitor of Severe Acute Respiratory Syndrome Coronavirus. Perlman S, ed. *J Virol*. 2015;90(6):3253-3256. doi:10.1128/JVI.02582-15.

253. Yang X-L, Tan B, Wang B, et al. Isolation and identification of bat viruses closely related to human, porcine and mink orthoreoviruses. *J Gen Virol*. 2015;96(12):3525-3531. doi:10.1099/jgv.0.000314.

254. Rodhain F. [Bats and Viruses: complex relationships]. *Bull Soc Pathol Exot*. 2015;108(4):272-289. doi:10.1007/s13149-015-0448-z.

255. Frick WF, Puechmaille SJ, Hoyt JR, et al. Disease alters macroecological patterns of North American bats. *Global Ecology and Biogeography*. 2015;24(7):741-749. doi:10.1111/geb.12290.

256. Wang Q, Qi J, Yuan Y, et al. Bat origins of MERS-CoV supported by bat coronavirus HKU4 usage of human receptor CD26. *Cell Host & Microbe*. 2014;16(3):328-337. doi:10.1016/j.chom.2014.08.009.

257. Mortlock M, Kuzmin IV, Weyer J, et al. Novel Paramyxoviruses in Bats from Sub-Saharan Africa, 2007-2012. *Emerg Infect Dis*. 2015;21(10):1840-1843. doi:10.3201/eid2110.140368.

258. Upton C, Mossman K, McFadden G. Encoding of a homolog of the IFN-gamma receptor by myxoma virus. *Science*. 1992;258(5086):1369-1372.

259. Nelson CA, Epperson ML, Singh S, Elliott JI, Fremont DH. Structural Conservation and Functional Diversity of the Poxvirus Immune Evasion (PIE) Domain Superfamily. *Viruses*. 2015;7(9):4878-4898. doi:10.3390/v7092848.

260. Ouyang P, Rakus K, van Beurden SJ, et al. IL-10 encoded by viruses: a remarkable example of independent acquisition of a cellular gene by viruses and its subsequent evolution in the viral genome. *J Gen Virol*. 2014;95(Pt 2):245-262. doi:10.1099/vir.0.058966-0.

261. Buller RM, Chakrabarti S, Moss B, Fredrickson T. Cell proliferative response to vaccinia virus is mediated by VGF. *Virology*. 1988;164(1):182-192.

262. Wise LM, Inder MK, Real NC, Stuart GS, Fleming SB, Mercer AA. The vascular endothelial growth factor (VEGF)-E encoded by orf virus regulates keratinocyte proliferation and migration and promotes epidermal regeneration. *Cellular Microbiology*. 2012;14(9):1376-1390. doi:10.1111/j.1462-5822.2012.01802.x.

263. Gonzalvez F, Ashkenazi A. New insights into apoptosis signaling by Apo2L/TRAIL. *Oncogene*. 2010;29(34):4752-4765. doi:10.1038/onc.2010.221.

264. Mariani SM, Krammer PH. Differential regulation of TRAIL and CD95 ligand in transformed cells of the T and B lymphocyte lineage. *Eur J Immunol*. 1998;28(3):973-982. doi:10.1002/(SICI)1521-4141(199803)28:03<973::AID-IMMU973>3.0.CO;2-T.

265. Nemčovičová I, Benedict CA, Zajonc DM. Structure of human cytomegalovirus UL141 binding to TRAIL-R2 reveals novel, non-canonical death receptor interactions. Fremont DH, ed. *PLoS Pathogens*. 2013;9(3):e1003224. doi:10.1371/journal.ppat.1003224.

266. O'Dea MA, Tu S-L, Pang S, De Ridder T, Jackson B, Upton C. Genomic characterization of a novel poxvirus from a flying fox: evidence for a new genus? *J Gen Virol*. 2016;97(9):2363-2375. doi:10.1099/jgv.0.000538.

267. Binger T, Annan A, Drexler JF, et al. A Novel Rhabdovirus Isolated from the Straw-Colored Fruit Bat Eidolon helvum, with Signs of Antibodies in Swine and Humans. García-Sastre A, ed. *J Virol*. 2015;89(8):4588-4597. doi:10.1128/JVI.02932-14.

268. Towner JS, Amman BR, Sealy TK, et al. Isolation of Genetically Diverse Marburg Viruses from Egyptian Fruit Bats. Fouchier RAM, ed. *PLoS Pathogens*. 2009;5(7):e1000536. doi:10.1371/journal.ppat.1000536.

269. Springer MS, Teeling EC, Madsen O, Stanhope MJ, de Jong WW. Integrated fossil and molecular data reconstruct bat echolocation. *Proc Natl Acad Sci USA*. 2001;98(11):6241-6246. doi:10.1073/pnas.111551998.

270. Olival KJ, Hosseini PR, Zambrana-Torrelio C, Ross N, Bogich TL, Daszak P. Host and viral traits predict zoonotic spillover from mammals. *Nature*. 2017;546(7660):646-650. doi:10.1038/nature22975.

271. Ricardo Moratelli CHC. Bats and zoonotic viruses: can we confidently link bats with emerging deadly viruses? *Memórias do Instituto Oswaldo Cruz*. 2015;110(1):1-22. doi:10.1590/0074-02760150048.

272. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA*. 1977;74(12):5463-5467.

273. Zhang G, Cowled C, Shi Z, et al. Comparative Analysis of Bat Genomes Provides Insight into the Evolution of Flight and Immunity. *Science*. 2013;339(6118):456-460. doi:10.1126/science.1230835.

274. Madden T. The BLAST Sequence Analysis Tool. March 2013.

275. Smith GL, Law M. The exit of Vaccinia virus from infected cells. 2004;106(2):189-197. doi:10.1016/j.virusres.2004.08.015.

276. Dodding MP, Way M. Nck- and N-WASP-Dependent Actin-Based Motility Is Conserved in Divergent Vertebrate Poxviruses. *Cell Host & Microbe*. 2009;6(6):536-550. doi:10.1016/j.chom.2009.10.011.

277. Obenauer JC, Cantley LC, Yaffe MB. Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res*. 2003;31(13):3635-3641. doi:10.1093/nar/gkg584.

278. Goraca A. New views on the role of endothelin (minireview). *Endocr Regul*. 2002;36(4):161-167.

279. Hocher B, Schwarz A, Fagan KA, et al. Pulmonary Fibrosis and Chronic Lung Inflammation in ET-1 Transgenic Mice. *American Journal of Respiratory Cell and Molecular Biology*. 2012;23(1):19-26. doi:10.1165/ajrcmb.23.1.4030.

280. Su H-P, Lin DY-W, Garboczi DN. The Structure of G4, the Poxvirus Disulfide Oxidoreductase Essential for Virus Maturation and Infectivity. *J Virol*. 2006;80(15):7706-7713. doi:10.1128/JVI.00521-06.

281. Bratke KA, McLysaght A, Rothenburg S. A survey of host range genes in poxvirus genomes. *Infect Genet Evol*. 2013;14:406-425. doi:10.1016/j.meegid.2012.12.002.

282. Balinsky CA, Delhon G, Afonso CL, et al. Sheeppox virus kelch-like gene SPPV-019 affects virus virulence. *J Virol*. 2007;81(20):11392-11401. doi:10.1128/JVI.01093-07.

283. McKenzie RA, Fay FR, Prior HC. Poxvirus infection of the skin of an eastern grey kangaroo. *Aust Vet J*. 1979;55(4):188-190.

284. Rothwell T, Keep JM, Xu FN, Middleton DJ. Poxvirus in Marsupial Skin-Lesions. *Aust Vet J*. 1984;61(12):409-410.

285. Yang Z, Bruno DP, Martens CA, Porcella SF, Moss B. Simultaneous high-resolution analysis of vaccinia virus and host cell transcriptomes by deep RNA sequencing. *PNAS*. 2010;107(25):11513-11518. doi:10.1073/pnas.1006594107.

286. Walzer T, Galibert L, Smedt TD. Poxvirus semaphorin A39R inhibits phagocytosis by dendritic cells and neutrophils. *Eur J Immunol*. 2005;35(2):391-398. doi:10.1002/eji.200425669.

287. Rehm KE, Connor RF, Jones GJB, Yimbu K, Roper RL. Vaccinia virus A35R inhibits MHC class II antigen presentation. *Virology*. 2010;397(1):176-186. doi:10.1016/j.virol.2009.11.008.

288. Reading PC, Moore JB, Smith GL. Steroid hormone synthesis by vaccinia virus suppresses the inflammatory response to infection. *J Exp Med*. 2003;197(10):1269-1278. doi:10.1084/jem.20022201.

289. Foster-Cuevas M, Wright GJ, Puklavec MJ, Brown MH, Barclay AN. Human herpesvirus 8 K14 protein mimics CD200 in down-regulating macrophage activation through CD200 receptor. *J Virol*. 2004;78(14):7667-7676. doi:10.1128/JVI.78.14.7667-7676.2004.

290. Cameron CM, Barrett JW, Liu LY, Lucas AR, McFadden G. Myxoma virus M141R expresses a viral CD200 (vOX-2) that is responsible for down-regulation of macrophage and T-cell activation in vivo. *J Virol*. 2005;79(10):6052-6067.

doi:10.1128/JVI.79.10.6052-6067.2005.

291. Akkaya M, Kwong L-S, Akkaya E, Hatherley D, Barclay AN. Rabbit CD200R binds host CD200 but not CD200-like proteins from poxviruses. *Virology*. 2016;488:1-8. doi:10.1016/j.virol.2015.10.026.

292. Ulaeto D, Grosenbach D, Hruby DE. The vaccinia virus 4c and A-type inclusion proteins are specific markers for the intracellular mature virus particle. *J Virol*. 1996;70(6):3372-3377.

293. Roberts KL, Smith GL. Vaccinia virus morphogenesis and dissemination. *Trends Microbiol*. 2008;16(10):472-479. doi:10.1016/j.tim.2008.07.009.

294. Zheng N, Schulman BA, Song L, et al. Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. *Nature*. 2002;416(6882):703-709. doi:10.1038/416703a.

295. Duda DM, Scott DC, Calabrese MF, Zimmerman ES, Zheng N, Schulman BA. Structural regulation of cullin-RING ubiquitin ligase complexes. *Curr Opin Struct Biol*. 2011;21(2):257-264. doi:10.1016/j.sbi.2011.01.003.

296. Sarikas A, Hartmann T, Pan Z-Q. The cullin protein family. *Genome Biology 2016 17:1*. 2011;12(4). doi:10.1186/gb-2011-12-4-220.

297. Bosu DR, Kipreos ET. Cullin-RING ubiquitin ligases: global regulation and activation cycles. *Cell Div*. 2008;3(1). doi:10.1186/1747-1028-3-7.

298. Lydeard JR, Schulman BA, Harper JW. Building and remodelling Cullin-RING E3 ubiquitin ligases - "Ubiquitylation: mechanism and functions" review series. *EMBO Rep*. 2013;14(12):1050-1061. doi:10.1038/embor.2013.173.

299. Wada H, Yeh E, Kamitani T. Identification of NEDD8-conjugation site in human cullin-2. *Biochem Biophys Res Commun*. 1999;257(1):100-105. doi:10.1006/bbrc.1999.0339.

300. Read MA, Brownell JE, Gladysheva TB, et al. Nedd8 modification of Cul-1 activates SCF beta(TrCp)-dependent ubiquitination of I kappa B alpha. *Mol Cell Biol*. 2000;20(7):2326-2333.

301. Hatcher EL, Hendrickson RC, Lefkowitz EJ. Identification of Nucleotide-Level Changes Impacting Gene Content and Genome Evolution in Orthopoxviruses. *J Virol*. 2014;88(23):13651-13668. doi:10.1128/JVI.02015-14.

302. Renfree MB, Papenfuss AT, Deakin JE, et al. Genome sequence of an Australian kangaroo, Macropus eugenii, provides insight into the evolution of mammalian reproduction and development. *Genome Biology 2016 17:1*. 2011;12(8). doi:10.1186/gb-2011-12-8-r81.

303. Davis ME, Gack MU. Ubiquitination in the antiviral immune response. *Virology*. 2015;479:52-65. doi:10.1016/j.virol.2015.02.033.

304. Soucy TA, Smith PG, Milhollen MA, et al. An inhibitor of NEDD8-activating

enzyme as a new approach to treat cancer. *Nature*. 2009;458(7239):732-U767. doi:10.1038/nature07884.

305. Mahon C, Krogan NJ, Craik CS, Pick E. Cullin E3 ligases and their rewiring by viral factors. *Biomolecules*. 2014;4(4):897-930. doi:10.3390/biom4040897.

306. Collins DR, Collins KL. HIV-1 Accessory Proteins Adapt Cellular Adaptors to Facilitate Immune Evasion. Racaniello V, ed. *PLoS Pathogens*. 2014;10(1). doi:10.1371/journal.ppat.1003851.

307. Sato Y, Kamura T, Shirata N, et al. Degradation of phosphorylated p53 by viral protein-ECS E3 ligase complex. Sugden B, ed. *PLoS Pathogens*. 2009;5(7):e1000530. doi:10.1371/journal.ppat.1000530.

308. Gastaldello S, Hildebrand S, Faridani O, et al. A deneddylase encoded by Epstein-Barr virus promotes viral DNA replication by regulating the activity of cullin-RING ligases. *Nat Cell Biol*. 2010;12(4):351–U110. doi:10.1038/ncb2035.

309. Wilton BA, Campbell S, Van Buuren N, et al. Ectromelia virus BTB/kelch proteins, EVM150 and EVM167, interact with cullin-3-based ubiquitin ligases. *Virology*. 2008;374(1):82-99. doi:10.1016/j.virol.2007.11.036.

310. Johnston JB, Wang G, Barrett JW, et al. Myxoma virus M-T5 protects infected cells from the stress of cell cycle arrest through its interaction with host cell cullin-1. *J Virol*. 2005;79(16):10750-10763. doi:10.1128/JVI.79.16.10750-10763.2005.

311. Mo M, Fleming SB, Mercer AA. Cell cycle deregulation by a poxvirus partial mimic of anaphase-promoting complex subunit 11. *PNAS*. 2009;106(46):19527-19532. doi:10.1073/pnas.0905893106.

312. Van Bressem MF, Van Waerebeek K, Raga JA. A review of virus infections of cetaceans and the potential impact of morbilliviruses, poxviruses and papillomaviruses on host population dynamics. *Dis Aquat Org*. 1999;38(1):53-65. doi:10.3354/dao038053.

313. Bracht AJ, Brudek RL, Ewing RY, et al. Genetic identification of novel poxviruses of cetaceans and pinnipeds. *Arch Virol*. 2006;151(3):423-438. doi:10.1007/s00705-005-0679-6.

314. Tuomi PA, Murray MJ, Garner MM, et al. *Novel Poxvirus Infection in Northern and Southern Sea Otters (Enhydra Lutris Kenyoni and Enhydra Lutris Neiris), Alaska and California, Usa*. Vol 50. Wildife Disease Association Business Office, 810 East 10th St., Lawrence, Kansas 66044-8897, USA.; 2015:607-615. doi:10.7589/2013-08-217.

315. VanBressem MF, VanWaerebeek K. Epidemiology of poxvirus in small cetaceans from the Eastern South Pacific. *Marine Mammal Science*. 1996;12(3):371-382.

316. Nollens HH, Hernandez JA, Jacobson ER, Haulena M, Gulland F. Risk factors associated with development of poxvirus lesions in hospitalized California sea lions. *J Am Vet Med Assoc*. 2005;227(3):467-473.

317. Fury CA, Reif JS. Incidence of poxvirus-like lesions in two estuarine dolphin populations in Australia: Links to flood events. *Sci Total Environ*. 2012;416:536-540. doi:10.1016/j.scitotenv.2011.11.056.

318. Roess AA, Levine RS, Barth L, et al. Sealpox Virus in Marine Mammal Rehabilitation Facilities, North America, 2007-2009. *Emerg Infect Dis*. 2011;17(12):2203-2208. doi:10.3201/eid1712.101945.

319. Geraci JR, Hicks BD, Aubin DJS. Dolphin pox: a skin disease of cetaceans. *Canadian Journal of Comparative Medicine*. 1979;43(4):399.

320. Becher P, Konig M, Muller G, Siebert U, Thiel HJ. Characterization of sealpox virus, a separate member of the parapoxviruses. *Arch Virol*. 2002;147(6):1133-1140. doi:10.1007/s00705-002-0804-8.

321. Nollens HH, Gulland FMD, Jacobson ER, et al. In vitro susceptibility of sea lion poxvirus to cidofovir. *Antiviral Res*. 2008;80(1):77-80. doi:10.1016/j.antiviral.2008.03.007.

322. Waltzek TB, Cortes-Hinojosa G, Wellehan JFXJ, Gray GC. Marine Mammal Zoonoses: A Review of Disease Manifestations. *Zoonoses Public Health*. 2012;59(8):521-535. doi:10.1111/j.1863-2378.2012.01492.x.

323. Nollens HH, Gulland F, Jacobson ER, et al. Parapoxviruses of seals and sea lions make up a distinct subclade within the genus Parapoxvirus. *Virology*. 2006;349(2):316-324. doi:10.1016/j.virol.2006.01.020.

324. Hicks BD, Worthy GA. Sealpox in captive grey seals (Halichoerus grypus) and their handlers. *J Wildl Dis*. 1987;23(1):1-6.

325. Gracie JA, Robertson SE, McInnes IB. Interleukin-18. *Journal of Leukocyte Biology*. 2003;73(2):213-224. doi:10.1189/jlb.0602313.

326. Meng X, Leman M, Xiang Y. Variola virus IL-18 binding protein interacts with three human IL-18 residues that are part of a binding site for human IL-18 receptor alpha subunit. *Virology*. 2007;358(1):211-220. doi:10.1016/j.virol.2006.08.019.

327. Reading PC, Smith GL. Vaccinia virus interleukin-18-binding protein promotes virulence by reducing gamma interferon production and natural killer and T-Cell activity. *J Virol*. 2003;77(18):9960-9968. doi:10.1128/JVI.77.18.9960-9968.2003.

328. Smith VP, Bryant NA, Alcamí A. Ectromelia, vaccinia and cowpox viruses encode secreted interleukin-18-binding proteins. *J Gen Virol*. 2000;81(Pt 5):1223-1230. doi:10.1099/0022-1317-81-5-1223.

329. Born TL, Morrison LA, Esteban DJ, et al. A poxvirus protein that binds to and inactivates IL-18, and inhibits NK cell response. *J Immunol*. 2000;164(6):3246-3254.

330. Xiang Y, Moss B. IL-18 binding and inhibition of interferon gamma induction by human poxvirus-encoded proteins. *PNAS*. 1999;96(20):11537-11542.

331. Krumm B, Meng X, Wang Z, Xiang Y, Deng J. A Unique Bivalent Binding and

Inhibition Mechanism by the Yatapoxvirus Interleukin 18 Binding Protein. Buller RML, ed. *PLoS Pathogens*. 2012;8(8). doi:10.1371/journal.ppat.1002876.

332. Afonso CL, Tulman ER, Lu Z, Zsak L, Kutish GF, Rock DL. The genome of fowlpox virus. *J Virol*. 2000;74(8):3815-3831. doi:10.1128/JVI.74.8.3815-3831.2000.

333. Krumm B, Meng X, Li Y, Xiang Y, Deng J. Structural basis for antagonism of human interleukin 18 by poxvirus interleukin 18-binding protein. *PNAS*. 2008;105(52):20711-20715. doi:10.1073/pnas.0809086106.

334. Sedimbi SK, Hägglöf T, Karlsson MCI. IL-18 in inflammatory and autoimmune disease. *Cell Mol Life Sci*. 2013;70(24):4795-4808. doi:10.1007/s00018-013-1425-y.

335. Elgueta R, Benson MJ, de Vries VC, Wasiuk A, Guo Y, Noelle RJ. Molecular mechanism and function of CD40/CD40L engagement in the immune system. *Immunol Rev*. 2009;229(1):152-172. doi:10.1111/j.1600-065X.2009.00782.x.

336. Kawabe T, Matsushima M, Hashimoto N, Imaizumi K, Hasegawa Y. CD40/CD40 ligand interactions in immune responses and pulmonary immunity. *Nagoya Journal of Medical Science*. 2011;73(3-4):69-78.

337. Baud V, Karin M. Signal transduction by tumor necrosis factor and its relatives. *Trends in Cell Biology*. 2001;11(9):372-377. doi:10.1016/S0962-8924(01)02064-5.

338. Chan FK-M, Shisler J, Bixby JG, et al. A role for tumor necrosis factor receptor-2 and receptor-interacting protein in programmed necrosis and antiviral responses. *J Biol Chem*. 2003;278(51):51613-51621. doi:10.1074/jbc.M305633200.

339. MacEwan DJ. TNF ligands and receptors - a matter of life and death. *Br J Pharmacol*. 2002;135(4):855-875. doi:10.1038/sj.bjp.0704549.

340. Magis C, van der Sloot AM, Serrano L, Notredame C. An improved understanding of TNFL/TNFR interactions using structure-based classifications. *Trends Biochem Sci*. 2012;37(9):353-363. doi:10.1016/j.tibs.2012.06.002.

341. Sedger L, McFadden G. M-T2: A poxvirus TNF receptor homologue with dual activities. *Immunol Cell Biol*. 1996;74(6):538-545. doi:10.1038/icb.1996.87.

342. Xu XM, Nash P, McFadden G. Myxoma virus expresses a TNF receptor homolog with two distinct functions. *Virus Genes*. 2000;21(1-2):97-109.

343. Reading PC, Khanna A, Smith GL. Vaccinia virus CrmE encodes a soluble and cell surface tumor necrosis factor receptor that contributes to virus virulence. *Virology*. 2002;292(2):285-298. doi:10.1006/viro.2001.1236.

344. Bahar MW, Graham SC, Chen RAJ, et al. How vaccinia virus has evolved to subvert the host immune response. *J Struct Biol*. 2011;175(2):127-134. doi:10.1016/j.jsb.2011.03.010.

345. Tollefson AE, Toth K, Doronin K, et al. Inhibition of TRAIL-induced apoptosis and forced internalization of TRAIL receptor 1 by adenovirus proteins. *J Virol*.

2001;75(19):8875-8887. doi:10.1128/JVI.75.19.8875-8887.2001.

346. Howard AR, Moss B. Formation of Orthopoxvirus Cytoplasmic A-Type Inclusion Bodies and Embedding of Virions Are Dynamic Processes Requiring Microtubules. *J Virol*. 2012;86(10):5905-5914. doi:10.1128/JVI.06997-11.

347. Mann BA, Huang JH, Li P, et al. Vaccinia virus blocks Stat1-dependent and Stat1-independent gene expression induced by type I and type II Interferons. *J Interferon Cytokine Res*. 2008;28(6):367-379. doi:10.1089/jir.2007.0113.

348. Bartlett A, Penders B, Lewis J. Bioinformatics: indispensable, yet hidden in plain sight? *BMC Bioinformatics*. 2017;18(1). doi:10.1186/s12859-017-1730-9.

349. Welch L, Brooksbank C, Schwartz R, et al. Applying, Evaluating and Refining Bioinformatics Core Competencies (An Update from the Curriculum Task Force of ISCB's Education Committee). *PLoS Comput Biol*. 2016;12(5). doi:10.1371/journal.pcbi.1004943.

350. Madlung A. Assessing an effective undergraduate module teaching applied bioinformatics to biology students. Ouellette F, ed. *PLoS Comput Biol*. 2018;14(1):e1005872. doi:10.1371/journal.pcbi.1005872.

APPENDIX

Appendix 1. Bash scripts and general protocol

Quality Control

1. Check contamination distribution using: http://taxonomer.iobio.io/ or https://app.onecodex.com/ or http://kaiju.binf.ku.dk/server
2. Download the reference genome in which you want to remove mapped reads from
3. Prepare reference genome for Burrows-Wheeler Aligner

Index (db creation)	bwa index RefGenome.fasta
Mapping (slow)	bwa mem RefGenome.fasta R1.fastq R2.fastq > ref_aln-pe.sam

4. Remove all reads that mapped to your reference genome

Check if header exist	head ref_aln-pe.sam
If header exist, then convert SAM to BAM (relatively slow)	samtools view -bS ref_aln-pe.sam > ref_aln-pe.bam
Retrieve some stats about this file using flagstat in samtools	samtools flagstat ref_aln-pe.bam
Sort BAM file (relatively slow)	samtools sort ref_aln-pe.bam ref_sorted
Count how many reads are unmapped to the reference genome	samtools view -S -f0x4 ref_sorted.bam | wc -l
Extract only the UNMAPPED reads	samtools view -b -f 4 ref_sorted.bam > ref_unmapped.bam
Convert file from BAM to fastq	bamutils tofastq -unmapped -read1 ref_unmapped.bam >& ref_unmapped_R1.fastq bamutils tofastq -unmapped -read2 ref_unmapped.bam >& ref_unmapped_R2.fastq
Count extracted read number	expr $(cat ref_unmapped_R1.fastq | wc -l) / 4 expr $(cat ref_unmapped_R2.fastq | wc -l) / 4
Proper pair reads to remove singletons using fastqutils tool	fastqutils properpairs batref_unmapped_R1.fastq batref_unmapped_R2.fastq processed_R1 processed_R2
Give proper paired reads .fastq file designation before proceeding to the next step	—
Count reads in final processed reads	expr $(cat processed_R1.fastq | wc -l) / 4 expr $(cat processed_R2.fastq | wc -l) / 4
Calculate final raw read coverage	[(From step M: R1 # + R2 #) x avg read length) / (expected virus length)]

5. Rename reads

Such that there's a distinction between R1 and R2 reads (adjust renaming.pl accordingly)	`perl renaming.pl > processed_R1_ready.fastq` `perl renaming.pl > processed_R2_ready.fastq`

Assembly

6. Run assemblers

Run MIRA	mira MIRA_Manifest.conf >&log_MIRA.txt
Run SPAdes	spades.py -1 processed_R1_ready.fastq -2 processed_R2_ready.fastq -m 14 –careful -o SPAdes_output

7. If Assembly fails (due to large file size or short contigs); modify raw reads

BLASTn largest assembled contigs for removal	/Users/cindytu/Desktop/Bioinformatics_Programs/ncbi-blast-2.6.0+/bin/blastn -db nr -query contigs.fasta -remote -out BLASTall
Attempt different contaminant removal	taxonomer.iobio.io
Shorten read name	nemesis:Homo_sapiens cindytu$ sed 's/NS500766:3:HHC7FBGX2:/KPVEG_/g' FastUniqR1.fastq > FastUniq_R1.fastq
FastUniq duplicated reads removal	fastuniq -i input_list.txt -t q -o output_1.fastq -p output_2.fastq -c 1
Split reads for fast assembly (see if full contig is achievable)	split -l 3273052 filename.fastq

8. Check assembled contig

Create genome mapped against raw read	*tanoti -r draft_genome.fasta -i R1.fastq R2.fastq -p 1 -o draft_mapped.sam*
Visualize mapped file for manual examination of gaps and ITR	In Tablet and BBB
Check dotplot for unusual regions	In JDotter

Annotation

9. Extract ORFs

Annotate putative ORFs	In GATU
Extract ORFs from temporary GB file created	In text editor

10. Make DB from local ChPV proteins sequence extracted from VOCS

Extract up-to-date ChPV rep seq	In VOCs
Create local VOCs DB	makeblastdb -in all_pox_aa.fasta -dbtype prot -out pox_prot_blast_db

11. BLASTP locally against all VOCs ChPV

ChPV BLASTP best hit	blastp -db /path/chpv_prot_blast_db -query aa_seq.fasta -out BlastP_BestHit -word_size 2 -max_target_seqs 1 -outfmt "6 qseqid stitle sseqid pident qlen slen length mismatch gaps evalue bitscore" -num_threads 4
ChPV BLASTP full record on all hits	blastp -db /path/chpv_prot_blast_db -query aa_seq.fasta -out BlastP_full -word_size 2 -num_threads 4
Troubleshoot:	The query file needs to be in the same folder as the db files to serve as a index

12. Annotation of putative function

Manually examine BLASTP results	Record in Excel, check with VGO, BLASTX, as needed
Look into weak alignment and interesting genes	With HHPred, CDD, InterPro, ScanProsite, CCTOP, I-TASSER

Phylogenetic Analysis

13. Create individual gene alignments with MUSCLE then concatenate them

Concatenate gene alignments	#stable.py ensures that the alignments are outputted in the same order as the input file. This facilitates with the subsequent merging of the alignment files. muscle -in A7L.fasta -out A7L.afa && python stable.py A7L.fasta A7L.afa > A7L-aln.fasta && rm A7L.afa muscle -in A10L.fasta -out A10L.afa && python stable.py A10L.fasta A10L.afa > A10L-aln.fasta && rm A10L.afa muscle -in A24R.fasta -out A24R.afa && python stable.py A24R.fasta A24R.afa > A24R-aln.fasta && rm A24R.afa muscle -in D1R.fasta -out D1R.afa && python stable.py D1R.fasta D1R.afa > D1R-aln.fasta && rm D1R.afa muscle -in D5R.fasta -out D5R.afa && python stable.py D5R.fasta D5R.afa > D5R-aln.fasta && rm D5R.afa muscle -in H4L.fasta -out H4L.afa && python stable.py H4L.fasta H4L.afa > H4L-aln.fasta && rm H4L.afa muscle -in J6R.fasta -out J6R.afa && python stable.py J6R.fasta J6R.afa > J6R-aln.fasta && rm J6R.afa paste -d \- *aln.fasta > merged.fasta sed -e 's/->.*//g' merged.fasta > phylogeny_ready.fasta
TrimAl (for ML analysis)	trimal -in phylogeny_ready.fasta -htmlout phylogeny_ready_trimal_auto -out phylogeny_ready_trimal_auto.fasta -automated1
Manually examine alignment	In BBB

14. Run RAxML

Determine best method	raxmlHPC-PTHREADS-AVX2 -p 12345 -m PROTGAMMAAUTO -s phylogeny_ready.fasta -n AUTO
Run and time phylogeny analysis	time raxmlHPC-PTHREADS-AVX2 -f a -p 38941 -s phylogeny_ready.fasta -x 55688 -# 1000 -m PROTGAMMALG -n KPV_ML_LG -T 4
Visualize newick file	In MEGA
Calculate distance matrix	raxmlHPC-PTHREADS-AVX2 -s phylogeny_ready.fasta -t bestTree -m PROTGAMMALG -n distance -f x -T 4

Appendix 2. Copyright permission for Pteropox virus paper

We use Cookies (/about/cookies) to track your preferences I Understand ()

Microbiology Society (http://www.microbiologysociety.org) Microbe Post (http://www.microbepost.org) Contact us (/about/contact-us/)

MICROBIOLOGY SOCIETY (/)

Journals online

Share

Rights and Permissions

Copyright

Copyright in most content published in our journals is held by the Microbiology Society, and cannot be republished without permission.

Articles that are published using the OpenMicrobiology (/about/open-access-policy) option are published under a CC-BY (https://creativecommons.org/licenses/by/4.0/) licence; with the exception of *Microbial Genomics* (http://mgen.microbiologyresearch.org), the default licence is CC-BY-NC (https://creativecommons.org/licenses/by-nc/4.0/) but where the funder requires it the author can select CC-BY. These articles are clearly identified in the journals.

Authors requesting permission to reuse their own content

The Licence to Publish or Copyright Transfer Agreement signed by authors explicitly grants the authors permission to reuse their own content without seeking further permission, provided that the original source of the material is credited appropriately.

How to request permission to photocopy or republish material

The Society has partnered with the Copyright Clearance Center (https://www.copyright.com/) to meet your licensing needs.

To obtain permission to reuse content from a particular article, simply click on the 'Request Permission' link under the 'Tools' menu next to the article.

Academic reuse attracts no permission fee. However, please note that a small administration charge is associated with this type of permission request.

For questions about using the Copyright Clearance Center, please contact their Customer Care team via customercare@copyright.com (mailto:customercare@copyright.com).

Access Key

F Free content

OA Open access content

S Subscribed content

T Free Trial content

Join the Microbiology Society

Join the Microbiology Society and become part of the largest microbiology community in Europe. Members receive a range of benefits including a discount on the OpenMicrobiology fee when publishing open access with our journals.

Find out more (https://www.microbiologysociety.org/membership.html)

Microbial Genomics

Appendix 3. Copyright permission for Eptesipox virus paper

Copyright Clearance Center RightsLink®

Home | Account Info | Help

Title: Characterization of Eptesipoxvirus, a novel poxvirus from a microchiropteran bat

Author: Shin-Lin Tu, Yoshinori Nakazawa, Jinxin Gao et al

Publication: Virus Genes

Publisher: Springer

Date: Jan 1, 2017

Copyright © 2017, Springer Science+Business Media, LLC

Logged in as:
Shin-Lin Tu
Account #:
3001173877
LOGOUT

Order Completed

Thank you for your order.

This Agreement between Ms. Shin-Lin Tu ("You") and Springer ("Springer") consists of your license details and the terms and conditions provided by Springer and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

printable details

License Number	4246110385042
License date	Dec 11, 2017
Licensed Content Publisher	Springer
Licensed Content Publication	Virus Genes
Licensed Content Title	Characterization of Eptesipoxvirus, a novel poxvirus from a microchiropteran bat
Licensed Content Author	Shin-Lin Tu, Yoshinori Nakazawa, Jinxin Gao et al
Licensed Content Date	Jan 1, 2017
Licensed Content Volume	53
Licensed Content Issue	6
Type of Use	Thesis/Dissertation
Portion	Full text
Number of copies	1
Author of this Springer article	Yes and you are a contributor of the new work
Order reference number	
Title of your thesis / dissertation	The bioinformatics characterization of 5 novel poxviruses in 4 new genera
Expected completion date	Apr 2018
Estimated size(pages)	100
Requestor Location	Ms. Shin-Lin Tu 3800 Finnerty Road Victoria, BC V8P 5C2 Canada Attn: Ms. Shin-Lin Tu
Billing Type	Invoice
Billing address	Ms. Shin-Lin Tu 3800 Finnerty Road Victoria, BC V8P 5C2 Canada Attn: Ms. Shin-Lin Tu
Total	0.00 CAD

ORDER MORE CLOSE WINDOW

Copyright © 2017 Copyright Clearance Center, Inc. All Rights Reserved. Privacy statement. Terms and Conditions.
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

Appendix 4. Copyright permission for Kangaroopox viruses paper

Copyright Clearance Center RightsLink®

Home | Create Account | Help

Title: Complete genomic characterisation of two novel poxviruses (WKPV and EKPV) from western and eastern grey kangaroos

Author: Mark Bennett,Shin-Lin Tu,Chris Upton,Cassie McArtor,Amber Gillett,Tanya Laird,Mark O'Dea

Publication: Virus Research

Publisher: Elsevier

Date: 15 October 2017

© 2017 Elsevier B.V. All rights reserved.

LOGIN

If you're a copyright.com user, you can login to RightsLink using your copyright.com credentials. Already **a RightsLink user** or want to learn more?

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: https://www.elsevier.com/about/our-business/policies/copyright#Author-rights

BACK | CLOSE WINDOW

Copyright © 2017 Copyright Clearance Center, Inc. All Rights Reserved. Privacy statement. Terms and Conditions.
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

Appendix 5. A list of important residues identified in the binding of human IL-18 ligands with ECTV and YLDV IL-18 BPs; corresponding positions in SOPV are extrapolated based on structural alignment.

	SOPV	YLDV	ECTV	Function
Structural	/	C21-C87	C25-C42	Forms intra-molecular disulfide bonds (SS)
	C43-C114	C43-C111	C39-C110	Forms intra-molecular disulfide bonds (SS)
	W58	W57	W54	N/A
	L99	L95	L95	N/A
	L129	L127	L125	N/A
Binding at IL18 site A	Y57	Y56	*Y53	ECTV Y53 interact with hIl18 K53 YLDV's phenolic group interact with hIL18 D54 YLDV Y56's aromatic group chain with YLDV T64 (SOPVF71, ECTVF67)
	V59	I58	L55	N/A
	F71	T64	*F67	-Insert into hydrophobic pocket of site A and forms strong interaction with surface residues and strong pi-cation interaction with charged IL18 K53 -YLDV Y56's aromatic group chain with YLDV T64 and IL18 D54
	Q73	D66	*E69	Form hydrogen bond with positively charged IL18 K53
	H74	Q67	H70	Form H bond with IL18 Y1 phenol group
	E81	E76	*E77	Form salt bridges with positively charged IL18 K53
	S115	V112	V111	N/A
	Q125?	K121	K120	N/A
Binding at IL18 site B	Y55	Y54	Y51	Conformational change upon binding but less so than site A, and does not actually fully occupy the pocket. Mutagenesis had negligible effects.
	V117	I114	T113	
	V122	P119	V118	
	R48	Y48	E44	Interact with IL18 D110 and own K23 (YLDV) eq to T21(SOPV) and K24 (ECTV)
	Q49	F49	H45	-Form loops and interact through hydrophobic interactions -ECTV H45 Form H bond with side chain of IL18 S105
	/	/	M46	Form loops and interact through hydrophobic interactions
Binding at IL18 site C	R48	H51	E48	Weak van der Waals with IL18 K8
	P53	F52	*F49	Insert into large hydrophobic IL18 site C pocket. Form loops and interact through hydrophobic interactions
	P119	P116	L115	-ECTV L115 Form loops and interact through hydrophobic interactions -YLDV P116 form hydrophobic interaction with IL18 and ILBP F49, Y48, K23 (missing in SOPV). Human IL18bp P153 mutation doesn't affect binding. -SOPV P119 may form H-bond with IL18 S105 (chimera prediction)
	D120	D115?	D116	Forms H bind with IL18 R104
		K117		Specific H-bond interactions for YLDV; not found in ECTV or SOPV
Homodimer (SS bond)	C138 & C141	C132	/	Intermolecular SS bond with same position on the other protomer to form a homodimer
Homodimer (hydrophobic interaction)	IYAPH (27-31)	/	IHVPV (29-33)	Hydrophobic interaction; HP stacks over one another from diff protomer
	E44? R46?	/	E42 and R44	Inter and intra H bond with itself or another protomer
	Q125, V127	/	E122, V124	Aliphatic side chain
	L132	/	I129	N/A