

Using genomic sequencing to explore vaccinia virus diversity, recombination and evolution

by

Li Qin

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Virology

Department of Medical Microbiology and Immunology  
University of Alberta

© Li Qin, 2014

## ABSTRACT

Smallpox was eradicated using vaccinia viruses (VACV) as vaccines, including Dryvax, a calf-lymph vaccine derived from the New York City Board of Health (NYCBH) strain, and TianTan, a chicken egg cultured vaccine used exclusively in China. To take advantage of the next generation sequencing technology, this thesis examined the genetic diversity of the population of viruses present in a sample of Dryvax and TianTan stocks. In Dryvax stock, any two clones differ by approximately 570 SNPs (single nucleotide polymorphism), exhibiting a patchy pattern of polymorphic sites across the whole genome due to recombination. In addition, 110 small indels (insertion and deletion) were observed in the Dryvax stock. Over 85% of indels are associated with repeats and a rare naturally attenuated virus bearing a large deletion in the right telomere (DPP17) was also identified. In contrast, there are barely any SNPs and indels detected in the TianTan clones, suggesting that this stock has been cloned previously. Two different subclones were detected; TP03 encoding large deletions in the terminal repeats that extend into both VEGF (vaccinia epithelial growth factor) genes and create a small plaque variant, and TP05 having the longest genome in all TianTan clones.

To further study the mechanism of poxvirus recombination, I coinfecting two of my sequenced viruses (TP05 and DPP17) and used the different SNPs to track the origin of progeny recombinants. My studies showed that

recombinants contain a patchwork of DNA fragments, with the number of exchanges increasing with passage. Further passage also selected for TianTan DNA and correlated with increased plaque size. The recombinants produced through a single round of co-infection exhibited a bias towards the short conversion tracks (<1 kbp) and exhibited 1 exchange per 12 kbp, close to the ~1 per 8 kbp in the literature.

Finally, I explored the possible origin of VACVs and evolutionary relationship among extant VACVs. My study showed that VACV is probably derived from a horsepox-like virus by reductive evolution. An intermediate virus has been suggested to originate from horsepox virus and serve as an ancestral strain for all other extant VACVs. A model of illegitimate recombination is proposed to help explain this evolutionary process.

## PREFACE

This thesis is my original and independent work. Chapters 2, 3 and 4 have been published in *Journal of Virology*, *Virology* and *J. Virology* respectively. I did all the experiments, data collection and data analysis. In Chapter 1, Dr. Upton contributed to the last figure 2.12 and Dr. Hazes helped with the bioinformatics analysis. In chapter 2, Dr. Liang provided the TianTan stock and contributed to the history of the TianTan vaccine. Dr. Evans is my supervisor and was responsible for the project design and manuscript composition. Chapter 5 now is under preparation. I contributed to the project proposal, experimental design, data collection and draft writing. Dr. Famulski helped with the manuscript edits of chapter 5. Ms. Nicole Favis performed the animal virulence study.

The animal protocols were approved by the University of Alberta Research Ethics Office, under the project titled “Vaccinia virus virulence determination” (No. AUP00000506, June 5, 2013).

## **DEDICATION**

To my parents, Ping'an Qin and Suiping Lu, and my husband, Xin Zhang for all their love and support, giving me strength and courage to follow my passions.

## ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. David Evans for his guidance and support during my studies. He has always encouraged independent thinking, and given me the opportunity to try anything I am interested in as well as develop my own ideas and approaches for a scientific research. Thank you to my committee members, Dr. Michael Houghton and Dr. Bart Hazes for your helpful suggestions and advice. I would also like to extend my sincere appreciation to Dr. Craig Brunetti and Dr. Andrew Mason for serving on my examining committee.

I would also like to thank the members of the Evans lab, past and present; especially Dr. Wendy Magee for teaching me experimental skills when I first started in the lab, helpful discussions and later on the protocol for the Roche 454 Junior sequencing; Dr. Jakub Famulski for assistance with my paper and thesis editing; Ms. Nicole Favis for excellent technical assistance, animal studies and lab organization; Megan, Wondim, Chad, Melissa, Carmit and Branawan for accompanying me in the past five years. I also thank Dr. Mary Hitt, Dr. Maya Shmulevitz and their groups for helpful advice during lab meetings; MMI office administrative staff (Anne, Debbie, Tabitha and Alliston) and washup staff (San, Yadviga) for all your hard work to keep the department running.

I would like to thank the staff at the Genome Québec Innovation Centre for their assistance with DNA sequencing and Mr. Rob Maranchuk for his technical support with the Roche 454 Junior sequencer. I also thank Dr. Chris Upton for his advice on my first paper and help with bioinformatics; Dr. Min Liang for sending me the TianTan stock, Profs. Zhao Kai, Isao Arita and Dr. Yasuo Ichihashi for providing insights into the history of Asian vaccinia virus strains; Dr. Jingxin Cao and Shaun Tyler at the National Microbiology

Laboratory, Public Health Agency of Canada for providing me the viral DNA of WR stock and the sequences of all their WR clones.

This work was supported by operating grants from the Canadian Institutes for Health Research and the Alberta Cancer Foundation, by an infrastructure award from the Canada Foundation for Innovation and by the Queen Elizabeth II scholarship.

Finally, I would like to thank my family and friends for all their love, support, tolerance and encouragement for me to finish this thesis.

## TABLE OF CONTENTS

ABSTRACT .....	ii
PREFACE .....	iv
DEDICATION .....	v
ACKNOWLEDGEMENTS .....	vi
TABLE OF CONTENTS .....	viii
LIST OF TABLES .....	xii
LIST OF FIGURES .....	xiii
LIST OF ABBREVIATIONS .....	xv
CHAPTER ONE – GENERAL INTRODUCTION .....	1
1.1 SMALLPOX.....	1
1.1.1 Variolation, vaccination and eradication history .....	2
1.1.2 Origins of old non-clonal vaccinia virus vaccines .....	4
1.1.3 Modern smallpox vaccines .....	5
1.2 POXVIRUSES .....	10
1.2.1 Classification .....	10
1.2.2 General features of the genome .....	13
1.2.3 Poxvirus life cycle .....	15
1.2.4 Mechanism of replication .....	18
1.2.5 Recombination .....	20
1.3 POXVIRUS GENOMICS.....	23
1.3.1 Early genome sequencing.....	25
1.3.2 Next generation sequencing technology .....	26
1.3.3 Current status .....	30
1.4 VACCINIA GENOMICS.....	32
1.4.1 Genomes sequenced prior to my work.....	32
1.4.2 Major virulence determinants .....	32
1.4.3 Special features of vaccinia virus genomes .....	36
1.5 PURPOSE OF THIS STUDY.....	38
CHAPTER TWO – GENOMIC ANALYSIS OF THE VACCINIA VIRUS STRAIN VARIANTS FOUND IN DRYVAX VACCINE .....	40
2.1 INTRODUCTION .....	40
2.2 MATERIALS AND METHODS .....	42

2.2.1 Viruses and cells .....	42
2.2.2 Virus DNA isolation and sequencing .....	42
2.2.3 Sequence assembly, analysis, and annotation .....	43
2.2.4 Quantitative PCR .....	45
2.2.5 Southern blotting .....	45
2.3 RESULTS AND DISCUSSION .....	47
2.3.1 Virus isolation and genome assembly .....	47
2.3.2 Genome annotation .....	50
2.3.3 Genetic similarities between Dryvax clones .....	52
2.3.4 Genetic differences between Dryvax clones .....	53
2.3.5 Patterns of mutation .....	76
2.3.6 Deletions in the right telomere junction .....	79
2.3.7 Substitution mutations .....	82
2.3.8 Recombination .....	84
2.4 CONCLUSIONS .....	87
CHAPTER THREE - GENOMIC ANALYSIS OF VACCINIA VIRUS STRAIN TIANTAN PROVIDES NEW INSIGHTS INTO THE EVOLUTION AND EVOLUTIONARY RELATIONSHIPS BETWEEN ORTHOPOXVIRUSES .....	
3.1 INTRODUCTION .....	91
3.2 METHODS .....	93
3.2.1 Virus and cell culture .....	93
3.2.2 Virus sequencing and genomic analysis .....	93
3.2.3 Other methods .....	95
3.3 RESULTS AND DISCUSSION .....	95
3.3.1 Virus isolation and genome sequencing and assembly .....	95
3.3.2 The genome of TianTan clone TP05 .....	101
3.3.3 Other features of TianTan clones TP03 and TP05 .....	107
3.3.4 Large deletions help define the evolutionary trajectory of VACV strains .....	109
3.4 CONCLUSIONS .....	114
CHAPTER FOUR – GENOME SCALE PATTERNS OF RECOMBINATION BETWEEN CO-INFECTING VACCINIA VIRUSES .....	
4.1 INTRODUCTION .....	116
4.2 METHODS .....	118
4.2.1 Cells and viruses .....	118
4.2.2 Virus DNA reactivation .....	119

4.2.3 Virus sequencing and genomic analysis.....	121
4.2.4 PCR and Southern blotting.....	121
4.3 RESULTS AND DISCUSSION .....	123
4.3.1 Virus isolation and genome sequencing and assembly .....	123
4.3.2 Crossovers in DTM and DTH viruses .....	126
4.3.3 Average length of the conversion tracts .....	131
4.3.4 Biased genetic origins in progeny viruses.....	134
4.3.5 Large deletions formed through illegitimate recombination .....	136
4.3.6 Recombination in SFV-reactivated vaccinia viruses .....	142
4.4 CONCLUSIONS .....	144
CHAPTER FIVE - GENOME STRUCTURE OF VACCINIA VIRUS REVEALS THE	
EVOLUTION AND EVOLUTIONARY RELATIONSHIPS AMONG EXTANT STRAINS	
.....	147
5.1 INTRODUCTION .....	147
5.2 MATERIALS AND METHODS .....	150
5.2.1 Viruses and cells.....	150
5.2.2 Viral sequencing, data analysis and annotation.....	150
5.3 RESULTS .....	151
5.3.1 Virus isolation, genome assembly and annotation .....	151
5.3.2 IHDW.....	154
5.3.3 Genome comparison among VACVs .....	158
5.3.4 VACV phylogenetic tree .....	163
5.3.5 Three shared features in the genome structure of all VVs .....	165
5.3.6 The relationship of extant VACVs, the right TIR boundary as an	
evolutionary feature.....	168
5.4 CONCLUSIONS .....	173
5.4.1 Evolution model.....	175
5.4.2 Evolutionary relationship of vaccinia viruses.....	178
5.4.3 VACV genome boundaries .....	182
5.4.4 Comparison of cowpox to vaccinia virus.....	183
5.4.5 The driving force for VACV evolution .....	186
CHAPTER SIX - DISCUSSION AND FUTURE DIRECTIONS .....	
6.1 SNPs .....	188
6.2 SMALL INDELS.....	190
6.3 LARGE DELETION AND GENOME REARRANGEMENT.....	191

6.4 COPY NUMBER VARIATION .....	193
6.5 VIRULENCE.....	199
BIBLIOGRAPHY .....	204

## LIST OF TABLES

Table 1.1 Some vaccinia viruses .....	6
Table 1.2 Family <i>poxviridae</i> .....	11
Table 1.3 Comparison of three 454 instruments .....	28
Table 1.4 Poxvirus family and their reference genomes .....	31
Table 2.1 Identities and accession numbers of the viruses cited in this work..	44
Table 2.2 Primers used in q-PCR and PCR analysis.....	46
Table 2.3 Summary of the virus sequencing and genome assembly process...	49
Table 2.4 Complete complement of genes and gene fragments encoded by Dryvax clones .....	54
Table 2.5 Genes differences in gene length (nt) and gene complement between Dryvax clones <sup>a</sup> .....	67
Table 3.1 Genes located in a 5.7 kbp deletion in TianTan clone TP03 terminal inverted repeats .....	100
Table 3.2 Gene differences between TianTan strains TP05 and TP00.....	104
Table 4.1 PCR primers used in this study .....	120
Table 4.2 GenBank accession numbers.....	122
Table 4.3 Gene differences in the right TIR deletion: TP05 <i>versus</i> DPP17...	137
Table 4.4 Gene complement in parent and hybrid viruses .....	138
Table 5.1 Identities and accession numbers of the viruses cited in this work	152
Table 5.2 Properties of the right TIR boundaries of vaccinia viruses.....	169
Table 5.3 Cowpox genes (CPX-GRI) that are not found in VACV .....	184

## LIST OF FIGURES

Figure 1.1 The genome structure of poxvirus .....	14
Figure 1.2 The infectious cycle of vaccinia virus .....	16
Figure 1.3 Model of VACV DNA replication .....	19
Figure 1.4 SSA model for recombination .....	22
Figure 1.5 <i>Hind</i> III restriction map of all orthopoxviruses .....	24
Figure 2.1 Plaque properties of different Dryvax clones .....	48
Figure 2.2 Telomere repeat patterns shown in Southern blot.....	51
Figure 2.3 Variant forms of the DVX_204 (B11R) gene carried by Dryvax clones .....	71
Figure 2.4 Variant forms of the DVX_142 (A12L) gene carried by Dryvax clones .....	72
Figure 2.5 A unique pattern of mutations in the DPP17 DVX_084 (I4L) gene .....	74
Figure 2.6 Genetic assortment of mutations within ORFs spanning the DVX_039-041 locus .....	75
Figure 2.7 Patterns associated with insertion and deletion mutations in Dryvax clones .....	77
Figure 2.8 Patterns of gene deletion and rearrangement in the virus right telomeres .....	80
Figure 2.9 Distribution of virus deletion patterns in Dryvax stocks .....	81
Figure 2.10 Distribution of polymorphic sites in Dryvax-derived VACV genomes .....	83
Figure 2.11 Distribution of putative recombination sites.....	86
Figure 2.12 Pattern of polymorphic sites in the ribonucleotide reductase large subunit gene .....	88
Figure 3.1 Plaques formed by viruses cloned from a stock of TianTan virus..	97
Figure 3.2 Small plaque TianTan strains grow to lower titers in culture .....	98
Figure 3.3 PCR and Southern blotting methods confirm the presence of telomeric deletions .....	99
Figure 3.4 Special sequence features associated with mutations in TianTan strains .....	102
Figure 3.5 Phylogenetic relationships between different Orthopoxviruses ...	106
Figure 3.6 Analysis of the number of 59bp repeats within the TIR.....	108
Figure 3.7 Phylogenetic relationships between different Orthopoxviruses ...	110
Figure 3.8 Relationship between deletions mapping in the left telomere of common VACV strains .....	111
Figure 3.9 Relationship between deletions mapping in the right telomere of common VACV strains .....	113

Figure 4.1 Patterns of DNA exchange in recombinant vaccinia viruses.....	124
Figure 4.2 Rare mutations associated with replication and recombination ...	128
Figure 4.3 Numbers of exchanges in DTM and DTH clones .....	130
Figure 4.4 Length of the DNA segments exchanged in DTH clones.....	132
Figure 4.5 Biased selection for sequences associated with the TianTan parent .....	135
Figure 4.6 Illegitimate recombination detected during the cloning and sequencing of hybrid DTM28 .....	141
Figure 4.7 Patterns of DNA exchange in recombinant vaccinia viruses produced using Leporipoxvirus-mediated reactivation reactions. ....	143
Figure 5.1 Gene mutations due to the insertion of repeats found in IHDW clones .....	155
Figure 5.2 The whole genome alignment by dotplots.....	157
Figure 5.3 The whole genome alignment of VACVs by dotplot analysis.....	160
Figure 5.4 Phylogenetic relationships between all vaccinia virus clones .....	164
Figure 5.5 BLAST alignments comparing the HPXV left and right telomere sequences to VACV telomeres .....	166
Figure 5.6 A 2kbp deletion shared by VACV strains DPP13, IHD-W, IHDW1 and WR .....	172
Figure 5.7 A model showing the proposed mechanism for genome rearrangement in VACVs .....	176
Figure 5.8 A special example for the evolutionary model part III applied to WR evolution .....	179
Figure 5.9 Proposed evolutionary relationships between extant and hypothetical ancestral VACV strains .....	181
Figure 6.1 Diagram showing the telomeric repeat structures in different VACV ....	194
Figure 6.2 Comparison of plaque size of CPXV and other VACVs .....	197
Figure 6.3 The predicted RNA secondary structure of NR3 (191nt) .....	200
Figure 6.4 Virulence study on mice to compare DPP15, 17 and 25 .....	201

## LIST OF ABBREVIATIONS

2k	Acam2000- a Dryvax-derived clone
a.a.	amino acid
ADP	adenosine diphosphate
AIM2	Absent in melanoma 2
AMP	adenosine monophosphate
AP-1	activator protein 1
APS	adenosine 5' phosphosulfate
ATP	adenosine triphosphate
AU	arbitrary units
Bcl-2	B-cell lymphoma 2
bp	base pair(s)
BSA	bovine serum albumin
BTB	Bric-a-brac, Tramtrack and Broad complex (a conserved domain)
CAM	chorioallantoic membrane
CCD	charge coupled device
CEV	cell-associated enveloped virus
CL3	Acambis clone 3- a Dryvax-derived clone
CMLV	Camelpox virus
CMP	cytidine monophosphate
CMV	cytomegalovirus
CHO	Chinese hamster ovary
Cop	Copenhagen- a vaccinia strain
CPE	cytopathic effect
CPX-GRI	Cowpox virus strain GRI-90
CPXV	cowpox virus
CrmC	cytokine response modifier C
C-terminal	carboxy terminal
CTL	cytotoxic T lymphocyte
CTP	cytidine triphosphate
CVA	Chorioallantois vaccinia virus Ankara
dA	deoxyadenosine
DAI	DNA-dependent activator of IFN-regulatory factors
dATP	deoxyadenosine triphosphate
dC	deoxycytosine
DC	dendritic cell
dCTPs	deoxycytidine triphosphates
ddATPs	dideoxyadenosine triphosphates
ddCTPs	dideoxycytidine triphosphates
ddGTPs	dideoxyguanosine triphosphates
ddNTPs	2',3'-dideoxynucleotide triphosphates

ddTTPs	dideoxythymidine triphosphates
dG	deoxyguanosine
dGTP	deoxyguanosine triphosphate
DI	Dairen I - Vaccinia virus produced in Japan
DNA	deoxyribonucleic acid
DNA-PK	DNA protein kinase
dNTPs	2'-deoxynucleotide triphosphates
ds	double stranded
DSB	double-strand break
dT	deoxythymidine
dTTP	deoxythymidine triphosphate
dUTP	deoxyuridine triphosphate
ECTV	Ectromelia virus
EDTA	ethylenediamine tetraacetic acid
EEV	extracellular enveloped virus
EGF	epithelial growth factor
eIF2 $\alpha$	eukaryotic translation initiation factor 2 $\alpha$
ERK2	extracellular signal-regulated kinase 2
ETF	early transcription factor
FBS	fetal bovine serum
GF	growth factor
GM-CSF	granulocyte macrophage colony stimulating factor
GMP	guanosine monophosphate
GPT	xanthine-guanine phosphoribosyltransferase
hGAAP	human Golgi anti-apoptotic protein
hpi	hours post infection
HPXV	horsepox virus
hr	host range
HU	hydroxyurea
IEV	intracellular enveloped virus
IFN	interferon
IHD	International Health Division-vaccinia strain
IHD-J	International Health Division-Japan
IHD-W	International Health Division-White
IKK $\epsilon$	inhibitor of kappa B kinase $\epsilon$
IL	interleukin
IL-18BP	IL-18 binding protein
IL-1R	IL-1 receptor
IMV	intracellular mature virus
indel	insertion and deletion
IRAK	interleukin 1 receptor-associated kinase
IRF	IFN regulatory factor
ISG	IFN-stimulated gene
ISGF3	IFN-stimulated gene factor 3

ISRE	IFN-stimulated response element
IV	immature virions
I $\kappa$ B $\alpha$	inhibitor of kappa B $\alpha$
JAK	Janus kinase
kb	kilobase
kbp	kilobase pair
kDa	kilo Dalton
MAP	mitogen-activated protein
MAPK	mitogen-activated protein kinase
MCS	multiple cloning site
MCV	molluscum contagiosum
MOI	multiplicity of infection
MPA	mycophenolic acid
MPV	Monkeypox virus
mRNA	messenger RNA
MVA	modified vaccinia virus Ankara
MYXV	myxoma virus
NDP	nucleoside diphosphate
NF- $\kappa$ B	nuclear factor kappa-light-chain-enhancer of activated B cell
NGS	next generation sequencing
NK	natural killer
NR1	nonrepeating I (including hairpin loop and concatemer resolution sequences)
NR2	nonrepeating II
NR3	nonrepeating III
nt(s)	nucleotide(s)
N-terminal	amino terminal
NTP	nucleoside triphosphate
NYCBH	New York City Board of Health
OAS	2'-5' oligoadenylate synthase
OPV	orthopoxvirus
ORF	open reading frame
PAMP	pathogen-associated molecular pattern
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PFU	plaque forming unit
PKR	protein kinase R
PNK	polynucleotide kinase
PPi	pyrophosphate
PRR	pattern recognition receptor
R1	repeat 1
R2	repeat 2
RAP	RNA polymerase-associated protein
RDP	Recombination Detection Program

RIG-I	retinoic acid inducible gene I
RLR	RIG-I-like receptor
RNA	ribonucleic acid
RNase	ribonuclease L
RDP	Recombination Detection Program
RPXV	Rabbitpox virus
RR	ribonucleotide reductase
RT	room temperature
SB	Southern Blot
SFV	Shope fibroma virus
siRNA	small interfering RNA
SNP	single nucleotide polymorphism
SOLiD	Supported Oligo Ligation Detection
SPI-2	serine proteinase inhibitor 2
SSA	single strand annealing
SSB	single strand DNA-binding
ssDNA	single-stranded DNA
STAT	signal transducer and activator of transcription
T4	T4 bacteriophage
TANK	TRAF family member-associated NF-kappa-B activator
TATV	Tatera poxvirus
TBK1	TANK-binding kinase 1
TF	transcription factor
TIR	terminal inverted repeats
TK	thymidine kinase
TLR	toll-like receptor
TMK	thymidylate kinase
TNF	tumour necrosis factor
TNFR	tumour necrosis factor Receptor
TRAF	tumour necrosis factor receptor-associated factor
TRAM	TRIF-related adaptor molecule
TRIF	TIR-domain-containing adapter-inducing interferon- $\beta$
<i>ts</i>	temperature-sensitive
TT	Tian Tan
VACV	Vaccinia virus
VARV	Variola virus
vCCI	VACV CC chemokine inhibitor
vCKBP	VACV chemokine binding protein
VCP	vaccinia complement protein
VEGF	viral epithelial growth factor
vGAAP	viral Golgi anti-apoptotic protein
WB	western blot
WR	western reserve
YFP	yellow fluorescent protein

## CHAPTER ONE – GENERAL INTRODUCTION

### 1.1 SMALLPOX

Smallpox is an acute infectious disease caused by variola virus (VARV). Smallpox obtained its name from the Latin word for spotted and refers to the raised bumps on the face and body of an infected person. Since humans are the only known natural hosts for variola, it can be transmitted only from an infected symptomatic person to another via aerosols and droplets. Infection occurs most frequently through the respiratory tract and rarely via the conjunctiva or placenta. Interestingly, inoculation through the skin, a process called variolation, results in a protective infection [1].

Two clinical forms of smallpox are present, variola major and variola minor. The more common form is variola major, producing a severe disease, with an extensive rash and high fever. Four clinical types of variola major are identified: ordinary (more than 90% of smallpox cases); modified (mild with an accelerated course, and occurring mostly in vaccinated persons); flat; and hemorrhagic (both rare and greatly severe). Generally, variola major has an overall mortality rate of 30-35%; however, flat and hemorrhagic smallpox are usually fatal. In contrast, variola minor is less common, causing a milder disease, with death rates of about 1% [1].

The incubation period of smallpox is 7–19 days after infection, usually 12 days. The initial symptoms are not specific, similar to other virus infection, including fever, malaise, headache, severe back pain, and vomiting. After 2–3 days, the body temperature falls and a rash appears, first on the mouth, throat, then face, hands and forearms, and later on the trunk. VARV mainly infects skin cells, rapidly producing the characteristic pimples or macules spread to the whole body in only 2 to 3 days. The macules are then developed quickly to form pustules, eventually leaving characteristic scars commonly on the face.

In addition to the scars, blindness and limb deformities are less common complications, occurring in less than 5% of smallpox survivors [1].

The last naturally occurring case of smallpox in the world was in Somalia in 1977. In 1979, the World Health Organization declared the worldwide eradication of smallpox by a collaborative global vaccination programme. Consequently, routine vaccination against smallpox in the public ended. However, after the events of September 11<sup>th</sup> 2001, a concern raised that the VARV might be used as an agent of bioterrorism. For this reason, precautions have to be taken to prevent a smallpox outbreak, including the study and storage of new smallpox vaccines, since any specific treatment for smallpox disease is not available; the only prevention is vaccination.

### **1.1.1 Variolation, vaccination and eradication history**

The earliest physical evidence of smallpox comes from the pockmarked rash on the mummified body of ancient Egypt 1157 BCE. Since there is no animal vector, an endemic outbreak of smallpox could occur only when the human population reached a high level (around 3000 BC). From 4<sup>th</sup> to 19<sup>th</sup> centuries, continuous epidemics occurred across the whole world, reducing human population and profoundly changing human history. The fight to eradicate smallpox was progressive and also accompanied by the development of knowledge in infectious disease and immunity [1].

The first medical practice to prevent smallpox was called variolation, by deliberately inoculating a person with materials taken from a patient in the hope that a milder disease can occur which would produce immunity against later natural infection of smallpox. The method was first used in China in 15<sup>th</sup> century by the intranasal insufflation and the Middle East in 16<sup>th</sup> century by the cutaneous inoculation (scarification) before it was introduced into England in the 1720s. The method is no longer used today because of a much safer alternative, vaccination [1].

Vaccination was introduced by the British physician Edward Jenner in 1796, by the immunization of a small boy with materials taken from a cow suffering from cowpox disease. The boy was immune to a later smallpox challenge due to the cross-protection raised by the cowpox virus. Vaccination was accepted very quickly since it exhibited milder symptoms than variolation. To distribute vaccines, the materials from a cow pustule were maintained by the arm-to-arm vaccination of children in order to keep the potency of vaccines, especially for long-distance shipments. Ivory points, glass slides or capillary tubes were also used to air-dry and transport the vaccines [1].

However, many problems arose. One of the problems was the contamination of the cowpox material. Arm-to-arm transfer was a possible cause of VARV contamination, especially in materials distributed in patients. Also, other human diseases such as syphilis were transmitted. Another problem was the shortages of cowpox virus. Cowpox is a rare and sporadic disease occurring unpredictably. Therefore, Dr. Jenner recommended one alternative source, materials from the lesions of horses suffering from a disease called grease, caused rarely by horsepox virus [1].

To overcome these problems, the production of vaccine on calves started to be developed in Italy in 1805. However, calf vaccines were confined to Italy until 1864. In the late half of 19<sup>th</sup> century, the use of calf vaccines were gradually extended and adopted in Europe and arm-to-arm transfer was eventually banned. Except calf, other animals were also used for vaccine productions, such as sheep, introduced by the Lister Institute, and water buffaloes used in India. In addition, to keep the potency of vaccines, vaccine providers periodically passaged the seed virus through animals, usually rabbits, sometimes monkeys, donkeys or even human subjects, although these manipulations were regarded as unnecessary later. Surprisingly, at some time during the 19<sup>th</sup> century, the nature of viruses used for vaccination changed from cowpox virus to vaccinia virus (VACV). Downie first reported in 1939 [2,

3] that biological properties of smallpox vaccines were different from those of cowpox virus. The origin of VACV remains unknown, although my investigations provide some support for the hypothesis that it derives from a (likely) now extinct horsepox-like strain (Chapter 5). However, since cowpox virus (CPXV), VACV and VARV all belong to the *orthopoxvirus* genus and induce closely related immunity; both VACV and CPXV are able to provide effective protections against smallpox. Due to the intensified smallpox eradication programme initiated in 1967 using VACVs, smallpox was eradicated world wide in 1979.

### **1.1.2 Origins of old non-clonal vaccinia virus vaccines**

Numerous VACV strains have been used throughout history, and were seldom subjected to clonal purification. During vaccine production on a large scale, high passage through alternative hosts made the viruses more genetically heterogeneous, normally named a quasispecies, a collective of viruses, comprising many different sequence variants, no one of which defines the virus present in a given stock. Generally, a vaccine was named according to the country or health agency involved in its propagation [4], for example, the New York City Board of Health (NYCBH) strain and Copenhagen (Cop) strain. Today, it is hard to identify documentation recording the origin and passage history of these vaccine strains across nearly 200 years. What we know mainly comes from the investigation of the origin of 35 strains by Wokatsch In 1972 [5]. Although many of these strains were used only for laboratory studies and not for vaccination, it provides valuable information for the history and general methods used in that time to passage and keep the viral stocks. Seven of those 35 strains were reported to derive from VARV: Dairen, Ikeda, Lister, LMC, Tashkent, Tian Tan (or Temple of Heaven) and Williamsport. Since it is well known that VARV could not convert into VACV, these strains must have been contaminated by VACVs, whose origins we are unable to track. Table 1.1

summarizes some of them related to this study. Four strains were most widely used in the last stages of the smallpox eradication campaign: Lister, New York City Board of Health (NYCBH), EM-63 and Tian Tan [1]. These four viruses had a better safety record than other VACVs, including Bern, Cop and Tashkent, which were also more reactogenic in humans [1, 6].

### **1.1.3 Modern smallpox vaccines**

The first generation of smallpox vaccines, i.e. vaccines produced in live animals do not meet the requirement for the processing, safety, sterility and quality control of modern vaccines, since the vaccines were easily contaminated with adventitious agents such as bacteria. This has led to the production of the second generation smallpox vaccines using tissue culture systems. ACAM2000<sup>TM</sup> is a cloned VACV strain isolated from Dryvax and manufactured in Vero cells, a cell line derived from kidney epithelial cells of an African green monkey. This vaccine has been licensed for use in the USA as of August 2007. However, there are concerns over the incidence of post-vaccination myopericarditis with ACAM2000 [7]. Therefore, studies are ongoing to develop safer and immunogenic vaccines. This so called third generation vaccine is generated by sequential passage to attenuate the virus but still retain their immunity against smallpox. Four candidates are available: MVA, Lister (LC16m8), NYVAC and DIs. MVA (modified Vaccinia Ankara) was derived from Ankara by 572 passages in chick embryo fibroblasts, causing 6 large deletions and  $\sim 30$  kb of the genome DNA to be lost compared to of parental virus [8]. MVA is unable to replicate in most mammalian cells [9].

Table 1.1 Some vaccinia viruses

Strain	Origin	Comments
Ankara	Ankara, Turkey, 1954	Passed in equines and then in chorioallantoic membrane. also called CVA (chorioallantois vaccinia Ankara)
Bern	Swiss Serum and Vaccine Institute, Berne Switzerland, 1898	Used for vaccine production in Berne 1898-1962. Virulent strain, use in man discontinued. Passaged in calf skin/n
Buffalopox	Northern India, 1967	Circulating in buffalos in Northern India. Also found in Bangladesh, Pakistan, Egypt, Indonesia
Cop	Copenhagen, Denmark	First VACV strain sequenced
Dairen I	University of Tokyo, Japan, 1934	Isolated from a smallpox patient in Tokyo with vesicular rash. Passed in rabbit skin/30, CAM/27. Mainly for experimental studies.
DIs	National Institute of Health, Tokyo, Japan, 1959	After 13 passages on 1-day-old eggs, a minute pock mutant appeared which was twice clone-purified on CAM and additionally passed 26 times in CAM. Not pathogenic for mice, guinea-pigs or Rabbits.
Dryvax	NYCBH	Freeze-dried calf lymph vaccine produced by Wyeth Laboratories.
Ecuador	Institute of National Hygiene, Guayaquil, Ecuador, 1940	Derived from Massachusetts Department of Health, Boston, USA. Low virulence
EM-63	Russia, 1963	Derived from strain Ecuador. Induced mild reactions in children. Used widely in smallpox eradication campaign
International Health Division (IHD)	New York City Department of Health, USA, 1954	Derived from the NYCBH strain. Passaged by mice (intracerebral)/51, CAM/4. (ATCC VR 156)
IHD-J (Japan)	New York City Department of Health	Dr. Yasuo Ichihashi isolated this strain from IHD in Japan and first reported in 1971.
IHD-White (IHD-W)	New York City Department of Health	Derived from IHD-J strain, hemagglutinin negative, isolated by Dr. S. Dales.

Ikeda	Osaka University, Japan, 1939	Isolated from a smallpox patient. Passaged by rabbits/5, calves/11 alternatively. Used for vaccine production in Japan for many years.
LC16m8	Japan, 1972	Attenuated strain derived from Lister. Used for smallpox vaccination in Japan since 1974
Lister	Lister Institute, Elstree, London, UK, 1892	Isolated from Prussian soldier in Franco-Prussian war in 1870. Used in UK since 1892 at the Lister Institute. Also known as Nigeria strain and Liverpool strain. Passaged originally by man/n, rabbit and sheep skin alternatively.
LMC	The Lister Institute, Elstree, England, 1936	Isolated from a patient with alastrim. Passaged in Monkey/1, rabbit (intratesticular)/6, rabbit skin/6. Since 1937, the strain was used from smallpox vaccine production at the Stock Medical Research Laboratories, Khartoum, Sudan.
Modified virus Ankara (MVA)	Munich, Germany, 1971	Attenuated strain derived from Ankara strain by 572 passages in chick embryo fibroblasts
New York City Board of Health (NYCBH)	Taken from England to New York, USA, 1856	Loines brought strain from England in 1856. Developed at New York City Board of Health Laboratories. Passage history: rabbits and calves with humanization 1-3 times yearly. Wyeth-calf adapted line was passed 21 times in calves since 1929 (ATCC: VR325).
NYVAC	Albany, New York, 1992	Derived from Copenhagen by deletion or inactivation of 18 genes. Attenuated and host range restricted.
Rabbitpox	Utrecht, Holland, 1941	Isolated from rabbits during outbreak of poxvirus infection in rabbit colony
Tashkent	Origin unknown	Virulent strain, use in man discontinued. Low neuropathogenicity for rabbits and mice. Passed by CAM/n.
Tian Tan	China, 1926	Isolated from a smallpox patient in 1926, passed through monkey skin, rabbit skin and testis, calf skin, human babies/8-9, rabbit and calf skin alternatively. Vaccine production was made on calf skins (before 1965) or chicken eggs (after 1965).

Western Reserve	Western Reserve University, Ohio, 1941	Derived from NYCBH by passage in mouse brain/24 (ATCC: VR119). Widely used laboratory strain after adapted to L-cells by 200-347 passages (ATCC:VR-1354)
Williamsp ort	Indiana University Medical Center, USA, 1951	Isolated from a patient with mild clinical smallpox symptoms. Passaged by rabbit cornea/1, rabbit (intratesticular)/n, also rabbit skin/1, rabbit (intratesticular)/1, CAM/1. Lethal for mice after intracerebral inoculation.

adapted from [5, 10] and other sources.

CAM: chorioallantoic membrane.

MVA was safely used to vaccinate over 120,000 individuals in Germany in the smallpox eradication campaign without adverse reactions [11, 12], but its effectiveness against smallpox remains untested since these vaccinees were not exposed to VARV. However, studies have shown MVA is still immunogenic even though it replicates poorly in mammals, providing protection against other orthopoxvirus (OPV) challenge [13-15].

LC16m8 was developed in Japan by repeatedly passaging the Lister virus through primary rabbit kidney epithelial cells (PRK) at a low temperature (30°C) and has a small plaque phenotype [16]. It was safely used in Japan for smallpox vaccination from 1974 with milder reactions [17] and also induced protection against OPV challenge in animal models [18, 19]. The small plaque phenotype was due to a frame shift mutation leading to the disruption of the B5R gene [20].

NYVAC is replication-deficient in most human cells and was produced by deletion of 18 nonessential genes implicated in virulence, host range, or pathogenicity from parental Cop strain [21]. It has been shown that NYVAC is immunogenic against several infectious diseases [21] and worked effectively as a recombinant vaccine delivery system [22, 23].

Dairen I (DI) strain was produced by serial passages of Dairen vaccine in chicken eggs with the selection for a tiny plaque phenotype on CAM [24]. DI contains a 15.4kbp deletion in the left terminal region, from Cop C9L to K5L [25]. Interestingly, this deletion is similar to the region deleted in the left terminus of NYVAC [26].

In addition to development of new smallpox vaccines, VACV has continued to be studied intensively. VACV strains are now genetically modified to carry on foreign genes and shown to have potential application as new live vaccines [27-30]. The capacity of VACV to encode and express foreign DNA is at least 25 kbp [31] and a polyvalent vaccine constructed by introduction of multiple foreign genes produced antibodies to all foreign

antigens [32]. Apart from the use of recombinant VACVs as live vaccines, VACV has been used to construct oncolytic viruses to selectively replicate in and destroy cancer cells. For example, JX-594, a thymidine kinase-deleted vaccinia virus armed with granulocytemacrophage colony stimulating factor (GM-CSF) [33-35], is currently in phase II clinical trial for hepatocellular carcinoma [36] and refractory metastatic cancers [33]. Furthermore, basic studies on elucidating the cycle of virus replication and discoveries of virus-encoded proteins that affect cell growth and modulate immune defense continue to provide valuable insights into pathogen-host relationships. All these efforts continue to make significant improvements in VACV safety and efficacy for human use.

## **1.2 POXVIRUSES**

Poxviruses are a group of viruses mainly associated with diseases that produced pocks in the skin, which also exclusively replicate in the cytoplasm of infected cells [37]. Most poxvirus virions are typically brick-shaped with dimensions of about 360x270x250nm. The mature virion is composed of an outer membrane enclosing a dumbbell-shaped core with two lateral bodies underlying the membranes. The core contains proteins necessary for early mRNA biosynthesis and core morphogenesis, and a genome composed of a single linear ds DNA molecule of 130-380 kbp with covalently closed ends.

### **1.2.1 Classification**

The family *poxviridae* is divided into two subfamilies *Chordopoxvirinae* and *Entomopoxvirinae*, based on vertebrate and insect host range, respectively (Table 1.2). The subfamilies are further divided into eight genera for *Chordopoxvirinae* and three genera for *Entomopoxvirinae*. Members of the same genus are genetically closely related due to the similar morphologies, molecular properties and host organisms.

Table 1.2 Family *poxviridae*

Genus	Virus	Major Hosts	Host Range	Geographic Distribution
chordopoxvirinae (subfamily)				
Orthopox virus	Variola	Humans	Narrow	Eradicated globally
	Vaccinia virus	humans, cattle, buffalo, swine, rabbits	Broad	Worldwide
	Cowpox virus	rodents, domestic cats and large felids, cattle, humans, elephants, rhinoceros, okapi, mongoose	Broad	Europe, Asia
	Camelpox virus	Camels	Narrow	Asia, Africa
	Ectromelia virus	Mice, voles	Narrow	Europe
	Monkeypox virus	Numerous: squirrels, monkeys, anteaters, great apes, humans	Broad	Western and central Africa
	Tatera poxvirus	Gerbils ( <i>Tatera kempi</i> )	?	Western Africa
	Raccoon poxvirus	Raccoons	Broad	North America
	Volepox virus	Voles ( <i>Microtus californicus</i> )	?	California
	Skunkpox virus	Skunks ( <i>Mephitis mephitis</i> )	?	North America
	Yoka poxvirus	Mouse?	?	Africa
Capripox virus	Sheeppox virus	Sheep, goats	Narrow	Africa, Asia
	Goatpox virus	Goats, sheep	Narrow	Africa, Asia
	Lumpy skin disease virus	Cattle, Cape buffalo	Narrow	Africa
Suipoxvirus	Swinepox virus	Swine	Narrow	Worldwide
Leporipox virus	Myxoma virus, rabbit fibroma virus	Rabbits ( <i>Oryctolagus</i> and <i>Sylvilagus</i> spp.)	Narrow	Americas, Europe, Australia
	Hare fibroma virus	European hare ( <i>Lepus europaeus</i> )	Narrow	Europe

	Squirrel fibroma virus	Eastern and western gray, red squirrels	Narrow	North America
Molluscipox virus	Molluscum contagiosum virus	Humans	Narrow	Worldwide
Yatapoxvirus	Yabapox virus and tanapox virus	Monkeys, humans	Narrow	West Africa
Avipoxvirus	Fowlpox, crowpox, juncopox, (etc.)	Chickens, turkeys, many other bird species	Narrow	Worldwide
Parapoxvirus	Orf virus	Sheep, goats, humans (related viruses of camels and chamois)	Narrow	Worldwide
	Pseudocow pox	Cattle, humans	Narrow	Worldwide
	Bovine papular stomatitis virus	Cattle, humans	Narrow	Worldwide
	Ausdyk virus	Camels	Narrow	Africa, Asia
	Sealpox virus	Seals, humans	Narrow	Worldwide
	Parapoxvirus of red deer	Red deer	Narrow	New Zealand
	Squirrel Poxvirus	Red and gray squirrels	Narrow	Europe and North America
Entomopoxvirinae (subfamily)				
AlphaEntomopoxvirus	Melontha melontha	beetles	Narrow	Worldwide
BetaEntomopoxvirus	Amsacta moorei, Melanoplus sanguinipes	butterflies, moths, grasshoppers and locusts	Narrow	Worldwide
Gamma Entomopoxvirus	Chrionomus luridus	flies and mosquitoes	Narrow	Worldwide

Four genera of *Chordopoxvirinae* may infect humans: orthopox, parapox, yatapox, molluscipox. However, only variola virus (VARV) and molluscum contagiosum virus (MOCV) are sole human pathogens; all the other human pathogenic poxvirus infections are zoonoses.

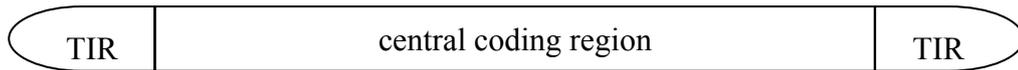
### **1.2.2 General features of the genome**

Poxvirus genomes consist of a single molecule of linear double-stranded DNA ranging in size from 130 kbp (parapoxviruses) up to 380 kbp (entomopoxviruses). The G+C content of its genome is also variable from 18% in Betaentomopoxvirus to 64% in orf virus.

The genome is flanked by terminal inverted repeats (TIR), which are identical but having the opposite orientation. The length of TIRs varies because of deletions/insertions or transpositions. Each TIR contains a hairpin loop at very end which connects the two DNA strands; a conserved concatemer resolution sequence which is required to resolve the replicating DNA concatemeric molecules into single length of genome DNA; tandem repeats array (R1 and R2); and a few of open reading frames (ORFs ) (variola virus encodes zero ORFs in its TIRs). (Figure 1.1)

Genome sequences of poxvirus show that genes are largely non-overlapping with little space between genes. The coding sequence of each gene is intronless and contiguous. Highly conserved genes are usually located in the central part of genome, and are essential for viral replication, transcription or morphogenesis; in contrast, variable genes are located in the terminal regions and usually involved in host interactions. Nearly 300 gene families are represented and 49 of them are conserved in all poxviruses. An additional 41 families are also conserved in chordopoxvirus [38].

A: genome



B: Terminal inverted repeats (TIR)

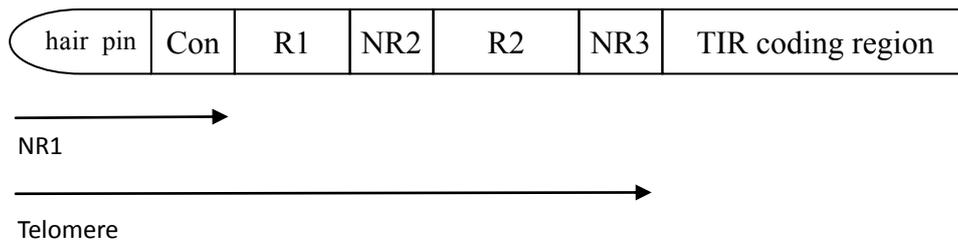


Figure 1.1 The genome structure of poxvirus

Panel A shows the whole genome structure. TIR: terminal inverted repeats. Panel B shows the composition of the TIR (according to VACV WR strain). Con: concatemer resolution sequence; NR: nonrepeating region; NR1: including hairpin and concatemer resolution sequence; R: repeats; Telomere: including NR1, NR2, NR3 and R1, R2. (Figures are not drawn to scale.)

### 1.2.3 Poxvirus life cycle

Poxvirus displays an unusual degree of autonomy in host cells. The prototype, vaccinia virus, encodes most, if not all, of the gene products required for three temporally regulated phases of viral replication; transcription, viral DNA repair, recombination and viral packaging [39, 40].

There are two distinct infectious virus particles: the intracellular mature virus (IMV), bearing a single outer membrane, and the extracellular enveloped virus (EEV), which has an additional lipid envelope (Figure 1.2). The extra membrane of the EEV is derived from the Golgi or endosomes during morphogenesis and contains several proteins which are absent from the IMV. Only the IMV membrane fuses with the cellular plasma membrane and the additional EEV membrane is disrupted and removed [41]. Infection is initiated by the attachment of vaccinia virion to an unknown receptor on the cell surface. Then the virion fuses directly with the cell membrane and releases the viral core into the cytoplasm, where it remains intact for several hours, while the early genes are transcribed within the core.

Poxvirus gene expression is temporally regulated at the level of transcription initiation, resulting in three classes of sequential expression consisting of early, intermediate and late genes [42]. A complete early transcriptional system is present within the infectious virus particles, including a large number of virus-encoded enzymes and factors, such as a DNA-dependent RNA polymerase, VACV early transcription factor (vETF), RNA polymerase-associated protein of 94-kDa (RAP94), capping and methylating enzymes, poly (A) polymerase and an early transcription termination factor. Early mRNA synthesis is detected within 20 minutes of infection [43], clearly independent of host protein synthesis. Half of the VACV genome is transcribed prior to DNA replication by RNA/DNA hybridization [44], including genes encoding proteins for host interaction, viral DNA synthesis and intermediate gene expression.

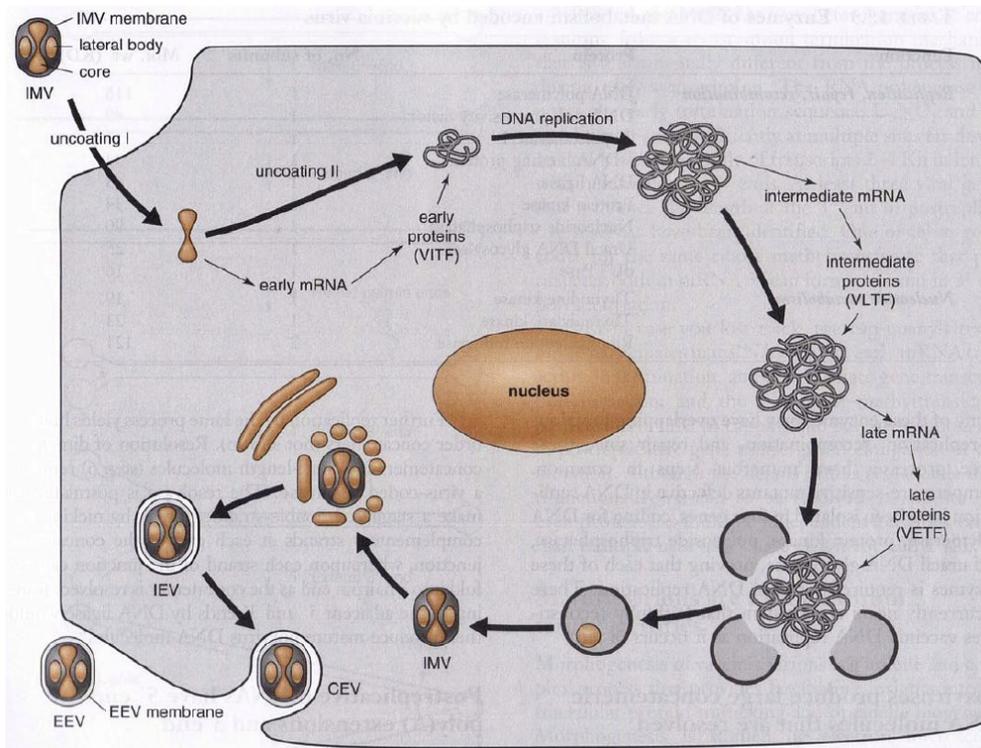


Figure 1.2 The infectious cycle of vaccinia virus

IMV: intracellular mature virus; IEV: intracellular enveloped virions; CEV: cell-associated virions; EEV: extracellular enveloped virus.

(adapted from NH Acheson, Fundamentals of molecular virology, Wiley, 2007)

Early mRNA transcripts are subsequently extruded through the pores in the core surface[45] and found to be aligned on microtubules, where they are translated by host protein translation machinery [46] . After early proteins are synthesized, core disassembly occurs, which releases the viral genome into cytoplasm for genome replication [47, 48]. DNA replication generates concatemeric molecules, necessary for the switch to transcription of intermediate genes. Several viral proteins are involved in this process, including virus intermediate transcription factors (VITF) VITF-1 (a component of the viral RNA polymerase) [49], VITF-2 (a host factor) [50], VITF-3 (A8R and A23R heterodimer) [51] and the capping enzyme [52]. VACV intermediate transcripts encode proteins needed for late gene expression. The products of the late genes include structural proteins, encapsidation enzymes and early transcription factors to be packaged for the next round of replication.

After resolved into unit length, the viral DNA is packaged in immature virions, a process requiring viral proteins of A32, I6, I1 and A13 [53-56]. Immature virions are wrapped by a membrane derived from the intermediate compartment between the ER and the Golgi apparatus [57]. Viral proteins required for immature virions formation contain F10 kinase [58], H5[59], G5[60] and A11[61]. Next, the spherical immature virion is converted to a barrel-shaped particle intracellular mature virions (IMV) with internal reorganization [62]. Conversion of immature virions to IMVs is impaired when expression of the A9[63], L1[64] or H3[65, 66] membrane protein is blocked. The IMV is the most abundant infectious form of VACV, although it can only be released by cell lysis. A subset of virions acquire an additional wrapping from the trans-Golgi network (intracellular enveloped virions, IEV) [67], fuse with the plasma membrane, and are released by exocytosis [68-70] or by direct budding depending on the virus species and strain[71-73]. Some of IEV remain associated to the cell surface (cell-associated virions, CEV)

[74], and mediate viral spread to adjacent cells. Others are released fully, forming infectious EEV (extracellular enveloped virions) and mediate long-range spread.

#### **1.2.4 Mechanism of replication**

DNA replication takes place in the cytoplasm of the infected cell in discrete cytoplasmic foci, termed viral factories or virosomes, which are sites of virion morphogenesis and DNA synthesis [75]. At high multiplicities of infection, the onset of DNA replication is detectable at approximately 2 hours post infection (hpi), lasts until 10-12 hpi, and results in the generation of about 10,000 genome copies per cell, half of which are thought to be encapsidated into progeny virions [76].

The current working model, the self-priming model, for poxvirus replication (Figure 1.3) proposes that a nick in a site near one of genomic termini is introduced, forming a 3' hydroxyl group to serve as a primer for the viral DNA polymerase [77, 78]. Strand displacement synthesis proceeds toward the hairpin telomere, elongating to the end of the template, which then folds back on itself to generate hairpin structure, leading the synthesis of entire genome. Although the primary product would be a dimeric tail-to-tail or head-head concatemer, if the newly synthesized strand again folded back on itself, a tetramer could be formed, and so on, forming high-molecular-weight concatemers after several rounds of replication [79-81]. Late gene products are necessary for the resolution of concatemers into unit-length molecules [79, 82-84]. When late gene expression is impaired, these concatemers accumulate and cannot be resolved into mature genomes. Three virus proteins are necessary for the resolution process: DNA topoisomerase I (Cop H7R)[85, 86], which relaxes supercoiled DNA; a DNase with nicking joining activity (Cop K4L) [87]; and a Holliday junction resolvase (Cop A22R) [88, 89].

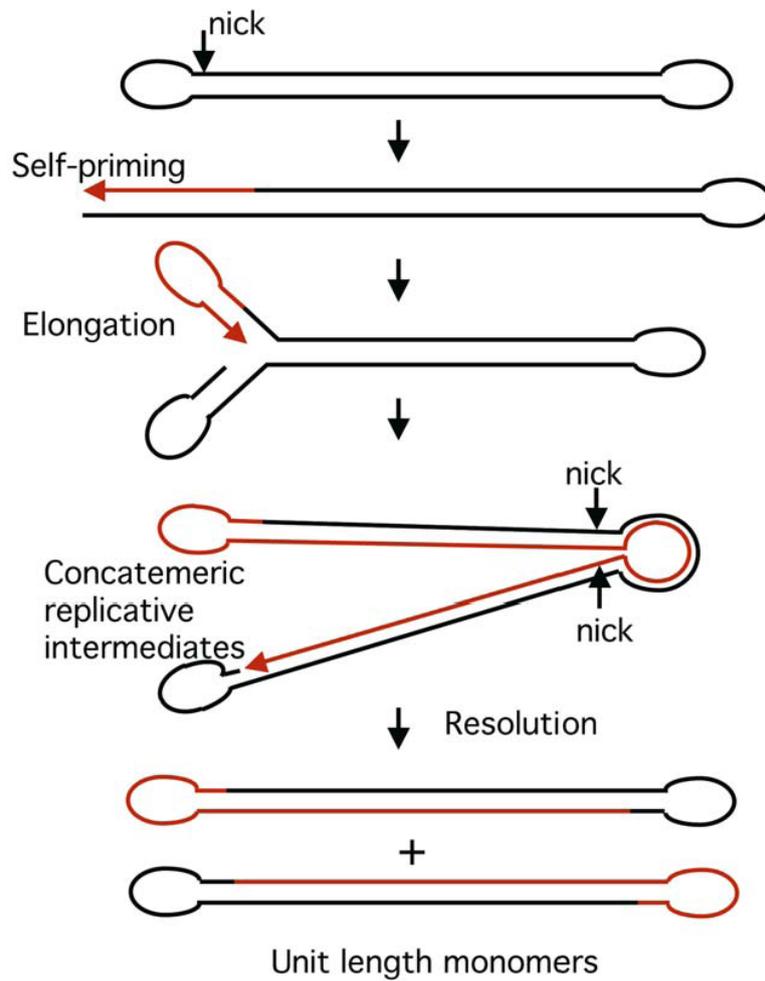


Figure 1.3 Model of VACV DNA replication

A nick is introduced close to genome end, which enable the DNA chain extension, resulting in a concatemeric DNA intermediate. After resolution, unique length monomer is created. Parental DNA is indicated in blank and daughter DNA in red.

Adapted from [10].

The self-priming model is proposed based on early experiments. However, recent studies showed that poxvirus encodes a functional primase [90] and a predicted flap-like nuclease G5R [91], supporting another replication model of leading-lagging strand synthesis. Further evidence is needed to illustrate the replication mechanism for poxviruses.

### **1.2.5 Recombination**

Recombination occurs within poxvirus-infected cells [92] and produces hybrids between related poxviruses, for example: malignant rabbit virus derived from recombination between Shope fibroma virus and myxoma virus [93]; and one type of capripoxvirus, Yemen goat-1 virus, is a recombinant of the other two types, Iraq goat-1 and Kenya cattle-1 virus [94]; nearly full length reticuloendotheliosis virus (REV) has recombined into the genome of field and vaccine strains of fowlpox virus [95].

Recombination also can explain some interesting phenomenon observed in poxvirus genomes. Mirror image deletions within TIRs were described first by Grant McFadden, suggesting that poxviruses tend to maintain two identical TIRs [96]. In addition, the sequences at both extremes seem to be interchangeable [97, 98]. Two mutants of rabbitpox virus (RPXV) showed two fragments from the left-most of genome inserted into the right end of genome, simultaneously causing deletions in the right terminus and extensions of TIRs; another mutant showed a right end fragment was transposed to the left [97]. The similar result was found in Monkeypox virus [98], cowpox [99] and VACVs [100]. This kind of translocation was suggested to be responsible for reduced infectivity in cells [101] or altered pathogenicity due to the host range change [97]. Furthermore, Moss and colleagues showed variations in the copy numbers of tandem repeats in VACV TIRs [102]. Except tandem repeats in TIRs, unique genes can also be duplicated to generate multiple copies in the genomes under hydroxyurea selection [103] or Protein Kinase R (PKR)

immune pressure [104]. This suggests that poxviruses may adapt highly specific gene amplification to express an excess of viral proteins as the primary and rapid response against host defenses. Later on, point mutation may happen [104].

Due to the effectiveness of poxvirus driven recombination, recombination has been used to genetically modify viral genomes. Transfection of DNA fragments flanked by poxvirus homology sequence [105] is routinely used to produce knock out or knock in mutants for studying gene function, construction of vaccines, generation of therapeutic viruses, or large scale expression of proteins in mammalian cells.

Poxvirus recombination is catalyzed by the VACV E9 viral DNA polymerase and I3 single-strand DNA binding (SSB) protein, which both primarily reside within virus factories, where recombination occurs. VACV uses a single-strand annealing (SSA) mechanism (Figure 1.4) to produce recombinants [106-109]. SSA occurs upon exposure of complementary regions between two ends of a broken DNA or separate DNAs. This process involves producing 3' ssDNA tails and is conserved from phage to human [110-112]. The 3' ssDNA tail allows two complementary ssDNAs to join together by base pairing. The region between these two homologous sequences is displaced as flaps and subsequently removed by nucleases [110]. The two molecules are then ligated by DNA ligases. Two features are observed in this model: little or no requirement of DNA synthesis; and the loss of the sequences between the two homologies in the final recombinant [113, 114].

Recombination is also necessary for repair of DNA damage. DNA damage occurs due to environmental factors or normal metabolic processes inside the cells, such as UV radiation, mutagenic chemicals, reactive oxygen species or replication errors.

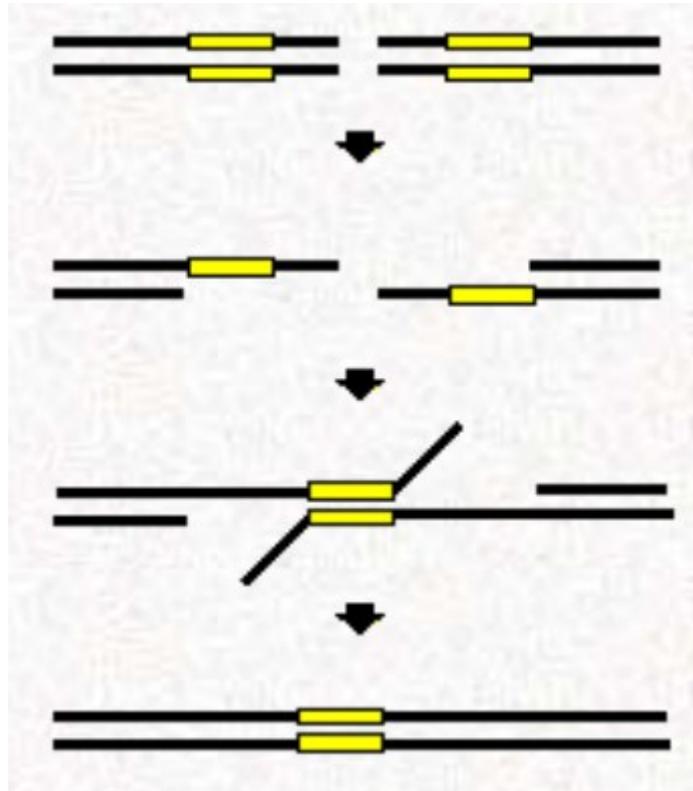


Figure 1.4 SSA model for recombination

Broken double strand molecules sharing homology (yellow boxes) are processed in a 5'-to-3' direction by nucleases to yield 3' ssDNA tails and expose complementary regions. Annealing of the two molecules at complementary regions may create 3' flaps of non-homologous sequences that must be removed by exo- or endonucleases. The processed strands are then sealed by ligase generating a shorter recombinant product than the parental strain. Adapted from [114].

Apparently, some of the damage can inhibit viral replication unless repaired. Genetic recombination is the process whereby two broken DNA molecules are joined together as shown in the above SSA mechanism. Apparently, VACV DNA polymerase and SSB play important roles to link the replication, recombination and DNA repair all together.

### 1.3 POXVIRUS GENOMICS

Before the advent of modern nucleic acid technologies, genetic analysis of viruses consisted of the random isolation of large numbers of mutants, generated naturally or induced (by chemicals and high temperature). The mutants were assigned to a particular complementation group while the physical order of genes on the virus genome was determined by recombination analysis. In the 1970s, advances in nucleic acid technology brought some new techniques, including marker rescue and restriction enzyme mapping analysis.

Marker rescue is a physical mapping technique that is used to localize the site of a mutation within a specific sub fragment of a virus genome. A polynucleotide fragment, the 'marker', usually encoding a wild-type sequence, is used to recombine with the mutant genome. If the fragment includes sequences spanning the mutation site, recombination can 'rescue' the mutant by causing it to revert to the wild-type phenotype.

Restriction endonuclease digestion analysis can distinguish clearly between all species of orthopoxvirus [115] (Figure 1.5). *HindIII* restriction mapping exhibits a conserved pattern of the central part of all orthopoxvirus genomes and a more variable pattern at the terminal regions. Among orthopoxviruses, VARV-major and -minor strains are similar but different from VACV, including HPXV, WR, Cop and CL3 (Acam clone 3). Interestingly, HPXV shows a similar pattern to other VACVs, suggesting all VACVs share the same origin.

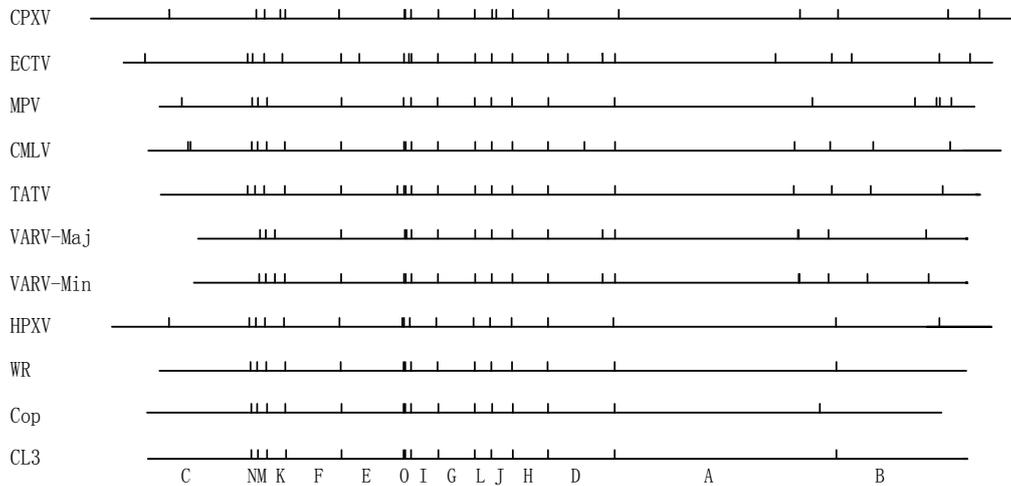


Figure 1.5 *HindIII* restriction map of all orthopoxviruses

Reference genomes were obtained from Genbank and *HindIII* digestion map was drawn to scale. CPXV: cowpox virus (NC\_003663); ECTV: ectromelia virus (NC\_004105); MPV: monkeypox virus (NC\_003310); CMLV: camelpox virus (NC\_003391); TATV: taterapox virus (NC\_008291); VARV-Maj: variola virus major (NC\_001611); VARV-Min: variola virus minor (Y16780); HPXV: horsepox virus (DQ792504); WR: (NC\_006998); cop: (VACCG); CL3: Acambis clone 3, a Dryvax derived clone (AY313848). The sixteen *HindIII* restriction fragments from the Cop strain were named alphabetically starting from the longest fragment as shown in A, B, C, ending at O (bottom line).

The VACV Cop strain was the first orthopoxvirus to have its genome sequenced and this explains the historical origins of the naming convention used for many VACV genes. *Hind*III digestion of Cop genome generates sixteen fragments which were named in alphabetical order from the longest to the shortest (Figure 1.5). The genes within each fragments were named in increasing order from left to right, for example, the first gene in fragment D was named D1 (L/R). The third letter L/R indicates the orientation of whether the gene is transcribed in a left or rightwards direction. If the digestion site is within a gene, the gene would be named according to the location of its start codon. Due to the similar digestion pattern of VACVs, the gene name is closely conserved between different strains.

### **1.3.1 Early genome sequencing**

New methods for DNA sequencing were developed during the 1970s. In 1977, Walter Gilbert and Allan Maxam at Harvard developed a DNA sequencing method known as Maxam-Gilbert sequencing or chemical sequencing [116]. This method involved radioactive labelling the end of DNA followed by chemical degradation of 4 nucleotide bases. Polyacrylamide gel electrophoresis resolved the labelled fragments and the autoradiograph showed a pattern of bands where the sequence can be inferred. Interestingly, the first variola genomic sequence was sequenced by this method [117, 118]. However, due to its high error rates, the sequence was revised using Sanger sequencing and deposited into Genbank as the reference of variola virus (NC\_001611). In fact, the Sanger method is the basis of the first generation of automated DNA sequencers and led Frederick Sanger to win the Nobel Prize in Chemistry in 1980.

The Sanger method, also named the chain-termination or dideoxy method, was introduced by Sanger in 1977 [119]. The principle of this method is that 2',3'-dideoxynucleotide triphosphates (ddNTPs) would inhibit the activation

of DNA polymerase in a DNA synthesis reaction because ddNTPs contains no 3'-hydroxyl group required to form the phosphodiester bond between one nucleotide and the next, preventing the DNA chain from extension further. As a result, chain termination occurs specifically at positions where ddNTPs are incorporated. An important feature in this DNA synthesis reaction is to contain a lower amount of ddNTPs compared to normal 2'-deoxynucleotides (dNTPs) (1:100) so that only partial incorporation of ddNTPs takes place. For example, when 4 normal dNTPs (dATP, dCTP, dGTP and dTTP) and one ddATP are added into the reaction, DNA fragments with varying lengths are produced since the DNA chain can be randomly terminated in every possible position of As. Then, the different fragments are separated according to the size using electrophoresis on denatured polyacrylamide gels, which would produce a pattern of bands showing the distribution of ddAs in the newly synthesized DNA strand. Four different reactions were needed for each template, because each reaction contains a different ddNTP terminator. Collectively, the sequence of the template would be inferred.

Using the Sanger method, the genome sequence of vaccinia virus Cop strain [120] was sequenced in 1990, a variola major strain Bangladesh-1975 (L22579) was sequenced in 1994 [121, 122], and myxoma virus [123] and Shope fibroma virus [124] were sequenced in 1999.

### **1.3.2 Next generation sequencing technology**

Sanger sequencing prevailed from the time it was invented to the mid-2000s when next generation sequencing (NGS) technologies emerged. A limitation of first generation sequencing was that one had to amplify the DNA fragments in vivo by cloning into bacterial hosts, which is both labor intensive and prone to bacterial-induced bias. In contrast NGS omits the cloning step by using more efficient PCR amplification methods in vitro, which reduced

dramatically the cost to sequence one human-size genome from \$100 million in 2001 to \$ 6,000 in 2013 [125].

The first NGS machine was the 454 Life Science system launched in 2005. The Solexa Genome Analyzer was introduced in 2006 and SOLiD (Supported Oligo Ligation Detection) from Agencourt also in 2006. These three platforms are the major, and massively parallel, sequencing systems for NGS commercially available. In 2008, the 454 GS FLX Titanium system was released, which was updated to GS FLX+ in 2011, promising the long read length up to 1000bp. Also, in 2009, Roche launched the GS Junior, a bench top version of the 454 sequencer. Table 1.3 compares the performance of the three systems.

454 systems require four steps to complete one sequencing run, including the preparation of a single stranded DNA library, emulsion PCR for clonal amplification of the library, data acquisition through sequencing-by-synthesis, and finally data analysis by provided software.

The starting materials can be a genomic DNA, PCR products (amplicon), cDNA or BACs (bacterial artificial chromosome). About 500ng of high quality sample DNA is needed for a shotgun library preparation. The genomic DNA is first nebulized with nitrogen, randomly shearing the template DNA into fragments of 400-900bp. The sheared DNA fragments are repaired by T4 DNA polymerase to create blunt ends, and PNK (polynucleotide kinase) to phosphorylate the 5' ends. The two ends of sheared DNA are then ligated with 2 short adaptors, A- and B-adaptor to each, which provide the binding sites for amplification and sequencing. B-adaptor is tagged with 5'-biotin, allowing the library DNA to be immobilized to streptavidin-coated magnetic beads, with one fragment binding to one capture bead by limited dilution.

Table 1.3 Comparison of three 454 instruments

Instrument	Junior system	GS FLX system	GS FLX+ system
System Type	Benchtop instrument	Instrument with floor-standing base	
Throughput	~35Mb	~400 Mb	~700 Mb
HQ Reads per Run	~100,000 shotgun, 70,000 amplicon	~1,000,000 shotgun, 700,000 amplicon	
Average Read Length	~400 bases	~400 bases	~700 bases, up to 1000bp
Accuracy	>99%	>99%	99.997%
Run Time	10 hours	10 hours	23 hours
Sample Input	gDNA, PCR products (amplicons) or cDNA,		
Multiplexing	Multiplex Identifiers (MIDs): 132 Gasket: 1 region	Multiplex Identifiers (MIDs): 132 Gaskets: 2, 4, 8, or 16 regions	
Computing	High-performance desktop PC	Cluster recommended	
Dimensions (w x l x h)	15.8" x 23.6" x 15.8" (40 cm x 60 cm x 40 cm)	29.5" x 35.4" x 51.2" (75 cm x 90 cm x 130 cm)	
Weight	55 lb (25 kg)	532 lb (242 kg)	

Note: adapted from the 454 online brochure.

Then, emulsion PCR is performed to amplify the fragments, which creates numerous micro reactors, for each containing only one bead with PCR reagents in a water-in-oil emulsion, making each template independently amplified on the surface of one bead. The emulsion PCR would produce about  $10^7$  clonal copies for each template [126], required to generate sufficient light signal intensity for the next sequencing step.

The beads are recovered by multiple wash steps to get rid of the emulsion using propanol and ethanol solution. The beads still contain a mixture of null beads with no fragments amplified on them, or mixed beads with more than one library fragments amplified, or DNA beads with products amplified from correctly captured one library fragment. Therefore, streptavidin-coated enrichment beads are used to capture all the beads (mixed beads and correctly assembled beads) carrying products containing biotinylated primers during emulsion PCR. Then the PCR beads are released and counted. A threshold parameter of 20% of the enrichment yield (2 million if starting with 10 million beads) is used to evaluate whether the mixed beads are too abundant, which cannot be sequenced correctly and should be excluded. If it is, the emulsion PCR has to be repeated with lower amount of input library DNA. On the other hand, if the number of enriched DNA beads is lower than 2 million, half a million of the DNA beads are loaded onto the PicoTiterPlate (PTP) device, which contains half a million of wells (one well for one bead). A single dNTP (dATP, dGTP, dCTP, dTTP) is supplied to the reaction mixture to all beads on the PTP at one time with the help of ATP sulfurylase, luciferase, luciferin, DNA polymerase and adenosine 5' phosphosulfate (APS). The incorporation of one nucleotide by the polymerase complementary to the template in the growing chain releases a pyrophosphate (PPi) group, which is used by the enzyme ATP sulphurylase to transform APS to ATP, a chemical then driving the luciferin into oxyluciferin and generating a visible chemiluminescent light signal. The light intensity is proportional to the amount of ATP generated from

incorporated nucleotides. The PTP is an optical fibre chip, allowing light emitted from the well to be captured by a sensitive CCD (charge coupled device) camera. The unmatched nucleotides are washed out and also degraded by apyrase in the wash solution, an enzyme used to decrease the background signal by converting free dNTP into dNMP. Then another dNTP is added into the reaction and the pyrosequencing reaction is repeated for 200 cycles of each four dNTPs in order.

One of the advantages of 454 systems are their running speed, Manpower can be reduced with automation of library preparation and semi-automation of emulsion PCR. A disadvantage of the 454 system is that it is prone to errors when estimating the number of bases in a long string of identical nucleotides. This is referred to as a homopolymer error and occurs when there are 6 or more identical bases in row [127]. Another disadvantages is that the price of reagents is relatively more expensive per nucleotide sequenced compared with other next-generation sequencers.

In this thesis, I used both the GS FLX system and the GS Junior benchtop to sequence more than 100 vaccinia virus strains.

### **1.3.3 Current status**

Due to the availability of advanced sequencing technology, more and more virus sequences are available. Now, almost every species in poxvirus family has at least one genome sequenced. Table 1.4 summarize all the species in poxvirus and their representative virus strains. Among them, some of the species, such as cowpox, vaccinia and variola, have dozens of sequences available. All of the sequencing information will greatly facilitate the genetic analysis of pox and other viruses.

Table 1.4 Poxvirus family and their reference genomes

Genus	Virus Species	Genome	accession no.
Chordopoxviridnae (subfamily)			
Avipoxvirus	Canarypox fowlpox,	~290-360kbp, G+C ~30%	NC_005309 NC_002188
Capripoxvirus	goatpox	~150kbp, G+C ~25%	NC_004003
Leporipoxvirus	rabbit fibroma, and myxoma	~160kbp, G+C ~40%	NC_001266 NC_001132
Molluscipoxvirus	molluscum contagiosum	~190kbp, G+C ~63%	NC_001731
Orthopoxvirus	camelpox, cowpox, ectromelia, monkeypox, taterapox, vaccinia, variola	~200kbp, G+C ~36%	NC_003391; NC_003663; NC_004105; NC_003310; NC_008291; NC_006998; NC_001611
Parapoxvirus	orf virus Bovine papular stomatitis virus	~140kbp, G+C ~64%	NC_005336  NC_005337
Suipoxvirus	swinepox	~147kbp, G+C ~27%	NC_003389
Yatapoxvirus	Tanapox Yaba-like disease virus	~145kbp, G+C ~33%	NC_009888 NC_002642
Crocodilepox virus	crocodilepox	190kbp, G+C 62%	NC_008030
Entomopoxvirinae (subfamily)			
Alphaentomo poxvirus	anomala cuprea	~246kbp, G+C ~20%	NC_00023426
Betaentomo poxvirus	Amsacta moorei	~232kbp, G+C ~18%	NC_002520

## **1.4 VACCINIA GENOMICS**

The approximately 200kb VACV genome is an AT-rich (66.7%), linear duplex, flanked by  $\sim$  10kb terminal inverted repeat (TIR) regions with covalently closed hairpin termini [120]. As the prototype for poxvirus, most studies have been conducted using vaccinia strains, such as WR, Cop, IHD, MVA and Lister etc.

### **1.4.1 Genomes sequenced prior to my work**

Due to the importance of VACV in history, a great majority of strains had been sequenced before my thesis, including HPXV (DQ792504), Western Reserve (NC\_006998), RPXV (AY484669), Copenhagen (M35027), four Dryvax-derived clones (Acambis clone 2000 (AY313847), Acambis clone 3 (AY313848), Duke (DQ439815) and 3737 (DQ377945)), Ankara strain CVA (AM501482), TianTan (AF095689), and five Lister strains (Lister-LO (AY678276), VACV107 (DQ121394), LC16m8 (AY678275), LC16m0 (AY678277) and GLV-1h68 (EU410304)).

### **1.4.2 Major virulence determinants**

The concept of using VACV as an oncolytic virus or vaccine vector requires some understanding of which virus genes cause undesirable levels of virulence. This section will give a general review of virulence genes in VACV. For a detail description, please refer to a recent review by Geoffrey Smith [128]. Virulence genes are located throughout the entire VACV genome, although a majority of cluster in terminal regions [129]. Proteins encoded by these genes can be either secreted, usually as glycoproteins, or cell-associated, affecting virulence, immunomodulation and host range. Secreted proteins can bind to and neutralize host complement factors, cytokines, chemokines or interferons (IFNs). In contrast, the intracellular viral proteins can inhibit

apoptosis or signaling cascades leading to induction of pro-inflammatory host responses.

#### **1.4.2.1 Secreted immunomodulators**

VACV infection triggers a series of host responses to rid the host of foreign pathogens, including an early nonspecific innate and later adaptive immune response [130]. Upon infection, the cells of the host innate immune system, including neutrophils, macrophages, dendritic cells (DCs) and natural killer (NK) cells, start to produce antiviral cytokines, such as tumor necrosis factors (TNFs), IFNs, interleukin (IL)-1 $\beta$ , IL-18, and chemokines. Cytokines function to recruit migratory leukocytes to sites of virus infection and induce a T helper type 1 (Th1) response, which is especially critical for poxvirus clearance [130]. This is a very complex response pathway and involves multiple cell types and proteins responding to many overlapping signals. However, VACVs are able to inhibit key aspects of the host immune response by the cooperative actions of several virulence factors [129, 131].

VACV WR B18R encodes a secreted protein that contains an immunoglobulin domain that binds and inhibits IFN- $\alpha/\beta/\delta/\omega$  (type I) from various mammalian species [132, 133]. Both non-infected and infected cells can be bound by B18 proteins, a mechanism used to inhibit the host antiviral response [134]. Except B18 (type I IFN receptor homolog), WR also encode a soluble IFN gamma (type II) receptors, B8, binding human, bovine, rat, rabbit IFN gamma with similar affinity, but having a significantly lower relative affinity to murine IFN gamma [135, 136]. Furthermore, a B8R deletion virus was clearly attenuated in a rabbit infection model and surprisingly in a mouse model since B8 binds to murine IFN gamma with low affinity [137, 138].

In addition to IFN receptor homologs, VACV also encodes some other soluble immunomodulators, such as WR B15R, an IL-1 $\beta$  receptor [139], C12L, an IL-18 binding protein[140], C3L, a complement control protein, binding to

complement components C3b and C4b and blocking activation of the complement cascade [141], TNF receptors [142] and two chemokine binding proteins, vCCI (VACV CC chemokine inhibitor) [143] and A41 [144]. All these soluble viral proteins play important roles in viral virulence.

#### **1.4.2.2 Inhibition of intracellular signaling**

VACV is a large DNA virus and one half of its genes contribute to immunomodulation. Most of these genes are expressed early during infection for rapid inhibition of the innate immune system. Multiple viral proteins are known to function on the same cellular pathway, although they target different sites and generally serve non-redundant roles. Interestingly, some these immunomodulator proteins may have multiple functions.

The most important innate immune response to viral infection is the production of IFNs, triggered by the sensing of viral pathogen-associated molecular patterns (PAMPs) by host pattern recognition receptors (PRRs). Stimulated PRRs induce signaling pathways to activate IFN regulatory factors 3 (IRF3), NF- $\kappa$ B and activator protein 1 (AP-1). These three proteins translocate into the nucleus and induce expression of genes encoding IFNs, cytokines and chemokines.

VACV uses several strategies to interfere with IFNs production. In fact, many viral proteins are expressed to inhibit the IFN signal pathway at multiple levels. The primary way to block IFNs is to prevent the production of IFNs. Upon infection, host mRNAs are targeted for degradation by two virus encoded mRNA-decapping enzymes, D9 and D10 proteins [145-147], a process leading to shut down the host protein synthesis, including host IFNs and other pro-inflammatory molecules, within a few hours of infection [148]. In addition, VACV encodes proteins to prevent binding of PRRs, such as E3 which sequesters dsRNA via a C-terminal dsRNA binding domain and thereby prevents activation of dsRNA-binding PRRs [149]. Cytosolic DNA can be

recognized by DNA-dependent protein kinase (DNA-PK) and trigger IFN production [150] in an IRF3-dependent manner. However, WR C16 binds DNA-PK and thereby inhibits the interaction of DNA-PK with DNA therefore blocking IFNs production [151]. Finally, the virus also express proteins, acting downstream of PRR and PAMPs to directly block the activation of IRF3 and NF- $\kappa$ B. Numerous viral proteins inhibit NF- $\kappa$ B activation, including A46, A49, A52, B14, C4, E3, K1, K7, M2 and N1. These proteins are expressed early and block NF- $\kappa$ B signaling pathway at differing sites. Similar for the IRF3 pathway, many viral proteins (E3, K7, C6, A46, N2, C16) are involved. In summary, abundant inhibitors in IRF3 and NF- $\kappa$ B pathway provide collective inhibition to IFN production.

Although VACV has numerous proteins that function to inhibit IFN production, IFNs are still present in infected hosts. This leads to the binding of cellular receptors on the surface of the same or neighbor cells and triggering activation of the JAK/STAT (Janus kinase/signal transducer and activator of transcription) pathway. Active JAK/STAT leads to the induction of IFN stimulated gene factor 3 (ISGF3) complex and transcription of hundreds of IFN stimulated genes (ISGs). To oppose the actions of the secreted IFN, VACV employs strategies to counter its downstream effects. First, the virus blocks IFN from reaching its receptors by encoding soluble IFN receptors as described above. Second, the virus can inhibit IFN-induced signal transduction. VACV VH1 dephosphorylates STAT1 and STAT2, thus inhibiting the signaling pathway generated by IFN receptors [152, 153]. Finally, VACV can inhibit ISGs. For example, VACV E3 binds ISG15 and prevents its antiviral activity [154]. K3, a viral protein similar to translation initiation factor 2 (eIF2 $\alpha$ ), acts as a pseudo-substrate for PKR, competitively blocking the phosphorylation of eIF2 $\alpha$  by PKR and shutting down host protein synthesis [155, 156].

In addition to the IFN pathway, apoptosis is another powerful mechanism to eliminate virus-infected cells [157]. Pro-apoptotic and anti-apoptotic

proteins of the Bcl-2 family are involved in apoptotic pathways. Apoptotic stimuli, for example, growth factor (GF) binding to its receptor, can activate pro-apoptotic proteins Bad or Bid to bind the effector proteins Bax and Bak. This interaction leads to the release of mitochondrial cytochrome c into the cytoplasm resulting in caspase-3 activation and subsequent apoptosis. Two viral Bcl-2 proteins are encoded to counter this pathway; protein F1, inhibiting Bak [158, 159], and N1, blocking Bid and Bad signaling [160, 161]. In addition, TNF or Fas can also trigger apoptosis. After binding to their receptors, a set of caspase proteases are activated, eventually leading to apoptosis. However, two viral proteins are found to be involved to block this pathway. VACV WR B13 (SPI-2) inhibits caspase-1 activity [162]. Another inhibitor is a viral Golgi anti-apoptotic protein (vGAAP), encoded by Lister strains, which is a homologue to human GAAP protein. Both of GAAPs modulate calcium in intracellular stores, leading to protection of cells from apoptosis [163, 164].

In summary, a lot of virulence genes have been found to participate in host-virus interactions. Further studies of the remaining genes will help us understand more about mechanisms of host immune defenses.

### **1.4.3 Special features of vaccinia virus genomes**

All the poxviruses have the same general genomic structure as show in Figure 1.1. However, VACVs show some interesting features that cannot be seen in other poxviruses. Firstly, all VACVs bear different length of TIRs, which are determined by the length of right TIR and its upstream fragments [18, 120, 165-170]. Two of them, Lister and Dryvax, show at least two groups of viruses with different length of TIRs in their original stocks [165, 169]. This finding will be discussed further in this thesis.

The second feature concerns the indels (insertion and deletions). Coulson et al. [171] examined the available genomes of variola virus and VACVs and

found that most indels are associated with repeats (of 3-25 nucleotides). It had originally been suggested that indels are the result of strand slippage errors caused by misalignments during DNA replication in bacteria [172, 173]. Strand slippage was suggested to take place preferentially during syntheses of both leading and lagging strands [174, 175]. However, it is not clear whether lagging strand synthesis is involved since Okazaki fragments have not been detected in poxvirus replication and the current model for poxvirus replication is the self-priming model as discussed above. Coulson et al. also examined the genome of MVA, compared to its parental CVA strain. In all 112 indels, only 13 indels are insertions. In contrast, 50% of indels in variola virus genomes are insertions. Moreover, MVA also showed 6 large deletions, which are likely responsible for its reduced virulence. Only 1 of these 6 deletions is related to repeats (7 nt). Since the other large deletions are not relevant to repeats, it is interesting to understand the mechanism for the generation of large deletions. Examining the mechanism for genome rearrangement, primarily indel formation, is the general topic of this thesis.

The third feature is the telomeric repeats. The VACV WR genome (Figure 1.1) consists of 13 copies of a 70 bp repeat (R1), a 325 bp spacer (also named nonrepeating II sequence, NR2), an additional 18 copies of the 70 bp repeat, 2 copies of a 125 bp repeat and 8 copies of a 54bp repeat (R2), following nonrepeating III (NR3) elements. The nonrepetitive sequence elements, NR1 and NR2 are conserved in the TIRs of all orthopoxviruses [176, 177]. Interestingly, VACV late promoter sequences are found not only in NR1 but also NR2 [178, 179]. Some of these late transcripts from NR2 extend through the hairpin region, indicating such transcription may be necessary for opening up the duplex region to facilitate entry of resolution proteins or topological rearrangement [179]. Although the general structure of TIRs is conserved among poxviruses, the precise number and organization of the repeats do vary. 37 of 38 Orthopoxvirus strains show cross-hybridization with the vaccinia

virus 70 bp repeat [180], indicating that these repeats are also conserved. Furthermore, based on the 54bp, 125bp and 70bp repeats bearing extensive sequence homologies and redundancies, evolution by unequal crossing over is proposed to explain the evolution of one type of repeat from another. For example, a recombinant element 86% homologous to the 125bp repeat was formed by a crossover between 54bp repeats [181].

More interestingly, restriction endonuclease analysis of terminal fragments demonstrated an array of eight or more fragments differing in size by 1650 bp increments even after the virus was repeatedly plaque purified. The transition between the longer telomere and the original one may be mediated by recombination events [102]. This 1.6 kb fragment contains NR2 and part of R2 (18 copies of 70 bp repeats). The authors suggested that this difference in length of TIRs is generated by recombination. Diffuse or submolar terminal restriction endonuclease fragments after serial passages of VACV stocks could also be observed in other vaccinia strains [182-185].

There is little evidence available concerning the possible function of these repeats, however, they are believed to participate in frequent inter- and intragenomic recombination events during viral life cycle.

## **1.5 PURPOSE OF THIS STUDY**

The purpose of this study was to take advantage of next generation sequencing technologies to explore viral diversity, recombination and evolution in VACVs. Since all the vaccines used in the smallpox global eradication programme were non-clonal and expected to comprise a “swarm” of viruses called a quasispecies, my first study was to examine the diversity of VACV genomes, specifically Dryvax and TianTan strains. Unexpectedly, naturally attenuated clones were isolated and whether these viruses might be useful to construct oncolytic viruses or serve as vectors for other vaccines such as HBV or HCV will be interesting to know. This study was next broadened to

investigate the mechanism of recombination since it is well established that poxviruses are subjected to recombination reactions, which might affect the evolution of viral genome structures. Finally, my study concluded with an exploration of the evolutionary relationships between different vaccinia strains from genome structure analysis.

## CHAPTER TWO – GENOMIC ANALYSIS OF THE VACCINIA VIRUS STRAIN VARIANTS FOUND IN DRYVAX VACCINE

Li Qin, Chris Upton, Bart Hazes, and David H. Evans<sup>1</sup>

### 2.1 INTRODUCTION

Smallpox was eradicated in the 1970s through the use of intensive vaccination in combination with campaigns designed to discover and isolate residual pockets of disease [1]. The vaccines used in many of these campaigns were composed of live vaccinia virus (VACV) cultured in large quantities on the skin of animals, usually cows. Many different vaccinia virus strains were used as vaccines towards the end of this era, including a strain that was distributed in a lyophilized formulation called Dryvax (DVX) and produced by Wyeth laboratories [165, 186]. This calf-lymph vaccine derives from the “New York City Board of Health” VACV strain and shares this origin with the most commonly studied VACV research strain called “Western Reserve” (WR). However, the two viruses have long been propagated in isolation. The last stocks of Dryvax were produced after passaging the virus 22-28 times in cows [165] while the sequenced strain of WR (NC\_006998) has a complex 70 year history of passage first in rabbits and mice followed, in more recent decades, by extensive passage in cell culture [187], Condit, personal communication].

These old smallpox vaccines were rarely subjected to clonal purification; in fact the methods used to propagate them would have readily produced mixtures of viruses that are commonly called a quasispecies. They were also contaminated with adventitious agents including bacteria and bacterial debris [188]. This situation is considered intolerable for modern licensure requirements and created problems when the need arose to produce new

---

<sup>1</sup> A version of this chapter has been published in *Journal of Virology* (2011), 85 (24), 13049-13060. Copyright 2011. American Society for Microbiology.

smallpox vaccine supplies in the late 1990s. This led to the development of ACAM-2K (Acambis clone 2000), a licensed vaccine comprising a VACV strain cloned from Dryvax and cultured on VERO cells [7, 186, 189, 190]. ACAM-2K was one of six viruses originally cloned from a pool of Dryvax production lots, and shown to replicate the immunogenicity of Dryvax in humans while exhibiting a seemingly comparable (albeit still not ideal) safety profile [190].

Genome sequencing has suggested that vaccines like Dryvax are comprised of a complex mixture of viruses. The Esposito laboratory reported that there are 573 single-nucleotide polymorphisms (SNPs) and 53 insertion-deletions (indels) of varying sizes that differentiate ACAM-2K (originally called Clone-2 or CL2) from a more neurovirulent sister clone, CL3 (Clone-3) [165]. Similar degrees of sequence difference are observed when these viruses are compared in a pairwise manner with other independently isolated Dryvax subclones including VACV-3737 and VACV-Duke. VACV-Duke is of special interest because it was isolated from a patient who developed progressive vaccinia after being vaccinated with Dryvax [191]. It thus may represent a more virulent component of the original inoculum. The fact that old smallpox vaccines comprise a quasispecies is not restricted to Dryvax of course. Garcel *et al.* have documented a diversity of phenotypes exhibited by clones isolated from a stock of the VACV strain Lister [18] and shotgun sequencing of unpurified stocks identified >1200 polymorphic sites distributed across a mix of Lister genomes [192].

These observations raise intriguing questions about the degree of genome diversity that can be found in old smallpox vaccines. In this communication we have taken advantage of recent advances in DNA sequencing technologies to explore this question in greater detail. Our results illustrate the remarkable complexity of the quasispecies that characterize stocks of old unpurified smallpox vaccines and suggest that the viruses that have been isolated to date

represent only a small fraction of the diversity of viruses in these preparations. These genomic studies also provide insights into the origin of viruses like VACV-Duke and of Orthopoxvirus evolution under the selection processes associated with classical VACV propagation methods.

## **2.2 MATERIALS AND METHODS**

### **2.2.1 Viruses and cells**

Viruses were isolated from a stock vial of Dryvax (lot #1556-14) and propagated on mycoplasma-free monkey kidney epithelial (BSC-40) cells in modified Eagle's medium (MEM) supplemented with 5% fetal bovine serum, 1% nonessential amino acids, 1% L-glutamine, and 1% antibiotic at 37°C in a 5% CO<sub>2</sub> atmosphere. BSC-40 cells were grown to 80% confluence in 24-well plates and then infected with virus at a multiplicity of infection of approximately one PFU/well in 100 µl of phosphate-buffered saline (PBS) for 1 h at 37°C. The viruses were cultured for three days and then harvested from wells containing only one plaque. These virus were then cloned twice more by limiting dilution as described above. Plaque images were preprocessed with ImageJ [193].

### **2.2.2 Virus DNA isolation and sequencing**

Each stock of plaque-purified virus was bulked up using sequential passages on BSC-40 cells and then purified by centrifugation through sucrose gradients as described previously [194]. The DNA was extracted from each purified virus using proteinase K digestion followed by phenol/chloroform extraction. The amount of DNA was determined by spectrophotometry and then 5 µg of each specimen was sequenced at the Genome Québec Innovation Centre (Montréal, Québec) using a high throughput pyrosequencing approach on a Roche 454 GS FLX Titanium sequencer platform. A total of 12 viruses

were sequenced using this approach, 11 were successfully assembled into complete viruses.

### **2.2.3 Sequence assembly, analysis, and annotation**

Different sized contigs were assembled from the raw sequencing data using Newbler and then CLC Genomics Workbench 4 software was used to inspect the trace data and complete the assembly of nearly full-length genomes. Conflicts between the reference sequences and our assemblies were resolved by using the PCR to amplify the region of interest followed by Sanger sequencing of the amplicons. Bioinformatic analyses were performed using Viral Genome Organizer [195, 196] and Viral Orthologous Clusters [38, 197] ([www.virology.ca](http://www.virology.ca)). The program LAGAN [198] ([genome.lbl.gov](http://genome.lbl.gov)) was used to produce alignments of multiple genomic sequences and Base-by-Base software [199] was used to fine-tune the alignments and to produce a visual summary of the whole genome alignments. To explore the phylogenetic arrangements, 98.8kb of conserved DNA sequences (spanning genes DVX\_058 to DVX\_155) were extracted from the multiple genome alignment and analyzed using a maximum likelihood analysis with the Recombination Detection Program (RDP) [200] using 1000 bootstrap replicates. Phylogenetic trees were plotted using TreeView [201]. The plot of putative recombination sites (Figure 2.11) was produced using the program Simplot/ Bootscan [202] with a 200 nt window, 20 nt steps, gap stripping “on”, 100 replicates, and employed a neighbor joining method of tree calculation. The Genome Annotation Transfer Utility (GATU) [203] was used to initially transfer a reference annotation to our Dryvax-derived viral genome sequences. Artemis [204] was used to visualize and edit the annotation. Table 2.1 lists the accession numbers for the VACV genomes cited in this communication.

Table 2.1 Identities and accession numbers of the viruses cited in this work

Virus ID	Virus	Source or strain	Accession number
CMLV-CMS	Camelpox virus	CMS	AY009089.1
CPXV-BR	Cowpox virus	Brighton Red	AF482758
CPXV-GRI	Cowpox virus	GRI-90	X94355
ECTV	Ectromelia virus	Moscow	NC_004105
HSPV	Horsepox virus	MNR-76	DQ792504.1
MPXV-ZAI	Monkeypox virus	Zaire-96-I-16	AF380138
RPXV	Rabbitpox virus		AY484669
TATV	Taterapox virus	Dahomey 1968	NC_008291
ACAM-2K	Vaccinia virus	Dryvax (Acambis clone 2000)	AY313847
3737	Vaccinia virus	Dryvax	DQ377945
CL3	Vaccinia virus	Dryvax (Acambis clone 3)	AY313848
Duke	Vaccinia virus	Dryvax (human isolate)	DQ439815
VACV-ANK	Vaccinia virus	Ankara	AM501482
VACV-COP	Vaccinia virus	Copenhagen	M35027
VACV-LO	Vaccinia virus	Lister-LO	AY678276
VACV-TP01	Vaccinia virus	TianTan (Temple of Heaven)	LQ/DE unpublished
VACV-TT	Vaccinia virus	TianTan (Temple of Heaven)	AF095689
VACV-WR	Vaccinia virus	Western Reserve	NC_006998
VARV-BAN	Variola major virus	Bangladesh-1975	L22579
VARV-GAR	Variola minor virus	GAR66	Y16780

To facilitate gene annotation, all of the complete experimental genomes were aligned and used to create a synthetic genome encoding a collective of all the open reading frames (ORFs). This synthetic, or “master”, DVX genome was used as a reference for annotation purposes.

#### **2.2.4 Quantitative PCR**

Quantitative PCR (q-PCR) was used to determine the relative abundance of the virus types discovered through genome sequencing. The primers used in this experiment are shown in Table 2.2 A and are named according to the genes they target. A pool of virus DNA was prepared by boiling the Dryvax vaccine in 5% (v/v) ion-exchange resin (Sigma, C7901) for 30 min followed by centrifugation for 20 min at 10,000 ×g. The supernatant was transferred to a clean tube and used as a source of DNA for the q-PCR reactions. The gene designated as DVX-209 was encoded by all four of the virus variants and was therefore used as standard to normalize the amount of virus DNA.

The q-PCR reactions were assembled using a SYBR green “supermix” (BioRad, 170-8882) and processed in a Bio-Rad Min-Opticon cycler according to manufacturer’s direction. Cloned virus DNAs were prepared as described above for use in ordinary PCR reactions.

#### **2.2.5 Southern blotting**

Virus DNA was digested with *SalI* (Fermentas) and size fractionated by electrophoresis through 0.7% agarose gels. The DNA was fragmented *in situ* with 0.2 M HCl, denatured with 0.4 M NaOH and 1 M NaCl, transferred to a nylon membrane (Pall Corporation, B60207), washed, and then UV cross-linked. A 445 bp biotin-labeled probe was prepared using the PCR in reactions containing biotin-16-dUTP (Roche, 1093070), two oligonucleotide primers (5'-GACTTAAACAACGGACAC-3' and 5'-GGCATAAAACACG AAGAGAA -3'), and *Taq* DNA polymerase (Fermentas).

Table 2.2 Primers used in q-PCR and PCR analysis

A

Group or virus class	Amplicon size (bp)	Reference genome	Primer ID	Primers (5'-to-3')
CL3-like virus	145	DPP25	DVX-215F	AGAACTCCCACCCATAAT
			DVX-215R	CATCTTCCACTTATCATCAC
2K-like virus	166	DPP20	DVX-212F2	GCGGAAGATACGACTGTT
			DVX-217R	GCATGTCCGTACCATTTATT
DUKE-like virus	147	DPP21	DVX-213F2	CGTACACCACTTCATTGC
			DVX-216R	TTGTATCCTCCTCCATATCT
DPP17-like virus	163	DPP23	DVX-209F	CGAAGAAGATGATGGGGAC
			DVX-226R2	GGCATAAAACACGAAGAGAA
Generic primer	148	DPP20	DVX-209F2	GACTTAAACAACGGACAC
			DVX-209R	ATTCTATCCCGTACCTCT

B

Group or virus class	Amplicon size (kb)	Primer ID	Primers (5'-to-3')
CL3-like virus	1.7	DVX-214F	CTGGACCCATCCTTTTATTCT
		DVX-215R	CATCTTCCACTTATCATCAC
2K-like virus	1.4	DVX-212F	AACCTCCTTCATGCATTC
		DVX-220R	GTTCTACCAACACCTTTATC
DUKE-like virus	0.9	DVX-213F	CGTTGGATGGATTCGATA
		DVX-229R	CCAGCTGCTCCATGATTT
DPP17-like virus	1.1	DVX-209F	CGAAGAAGATGATGGGGAC
		DVX-226R	ATAAGAGGAAAGAGGACAC

After hybridizing the probe to the DNA, the membrane was stained with IRDye 800CW-coupled streptavidin (LI-COR, 926-32230), and imaged using a LI-COR infrared imager as recommended by manufacturer.

## **2.3 RESULTS AND DISCUSSION**

### **2.3.1 Virus isolation and genome assembly**

A single round of passage by limiting dilution on BSC-40 cells was used to isolate >50 different randomly selected Dryvax clones. A total of 25 viruses were then chosen for further passage by limiting dilution two more times on BSC-40 cells. These viruses were separately “bulked up” and the high titer stocks were purified using sucrose gradients and the virus DNA isolated using phenol-chloroform extractions. This method produced varying yields of virus DNA, and we arbitrarily elected to sequence the 12 viruses that yielded the greatest amounts of DNA for 454 sequencing. Although this may have biased the selection in favor of viruses that replicate most efficiently in BSC-40 cells, the 11 viruses that were eventually sequenced and assembled into complete contigs produced plaques that were not obviously any different from the range of plaque sizes produced by the original pool of 25 viruses (Figure 2.1). The choice of viruses thus represents a modestly representative sample of the diversity of viruses in Dryvax, although one would probably not have sequenced any virus exhibiting a profound replication defect in BSC-40 cells.

The viruses were sequenced using a multiplex approach and a Roche 454 GS FLX Titanium sequencer. Table 2.3 summarizes the sequencing statistics. The sequence reads were automatically assembled into initial contigs and then manually assembled into nearly complete final contigs using several different sequencing tools.

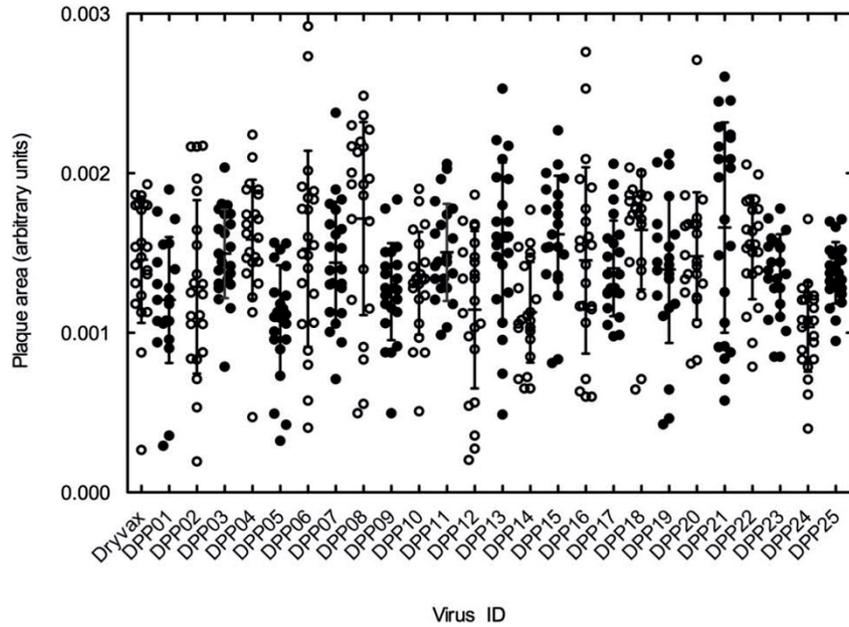


Figure 2.1 Plaque properties of different Dryvax clones

Each stock of virus was plaque purified three times and then ~50 pfu of each isolate was plated on a monolayer of BSC-40 cells. The viruses were cultured for three days, stained with crystal violet, imaged, and the plaque sizes determined using the program ImageJ [193]. No single clone produced uniquely smaller or large plaques, although the smallest plaques (e.g. DPP05, DPP24) are significantly ( $p < 0.05$ ) different from the largest ones (e.g. DPP04, DPP18).

Table 2.3 Summary of the virus sequencing and genome assembly process

Virus ID	Number of reads	Number of sequenced bases	Average read length (nt)	Average quality	Largest initial contig (bp)	Matched reads	Total nucleotides	Average read redundancy	Final contig length (bp)	Intact genes	Accession number
DPP09	25,579	8,074,625	315	32	166,052	16,435	5,429,541	27.4	198,518	174	JN654976
DPP10	27,379	8,745,307	319	32	166,052	8,546	2,918,944	14.7	198,464	172	JN654977
DPP11	24,102	7,693,966	319	32	166,073	15,484	5,178,206	26.1	198,554	175	JN654978
DPP12	19,219	6,238,917	324	31	166,165	10,322	3,529,518	17.8	198,741	173	JN654979
DPP13	11,786	3,818,832	324	32	39,700	4,580	1,631,481	8.4	194,800	172	JN654980
DPP15	19,281	6,639,255	344	32	166,263	9,965	3,656,426	18.4	198,547	174	JN654981
DPP16	24,302	7,943,833	326	31	166,158	10,597	3,730,887	18.8	198,820	172	JN654982
DPP17	30,391	8,800,008	289	32	165,633	9,059	2,831,001	14.8	191,709	170	JN654983
DPP19	26,422	8,304,653	314	32	143,909	8,101	2,777,440	14.0	198,609	175	JN654984
DPP20	23,453	7,523,115	320	32	108,090	6,793	2,332,401	11.7	198,699	173	JN654985
DPP21	17,503	5,395,284	308	32	30,921	4,834	1,593,050	8.2	194,916	174	JN654986

Because the average read length was only ~320 nt, these sequencing methods do not provide accurate insights into the structure of the highly repeated elements located in the virus telomeres [181, 205]. We therefore elected to define the left and right ends of each genome as each comprising four copies of the 54 bp repeats located proximal to the boundaries of the terminal inverted repeats (TIR). DNA sequencing data and Southern blots showed that these viruses exhibit the variable numbers of (54, 69, and 125 nt) telomeric repeat elements that have previously been reported [181, 191] to characterize vaccinia virus strains (Figure 2.2).

Pyrosequencing methods are prone to producing indel-type sequencing errors within homopolymeric base runs. In many cases these mistakes were easily spotted wherever the consensus produced frameshift errors within normally intact VACV ORFs, and the true sequence could be deduced from visual inspection of the sequence of one or more of the aligned high quality replicate reads. We also PCR amplified six of the sites where it was not possible to deduce the true sequence from the replicate reads and resequenced them. In all of the six instances, the correct sequence was the one that supported the original reference sequence and an intact ORF. This led us to assume that “mutations” in homopolymeric runs of more than 4 bases were artifacts of the sequencing technology and were thus edited to maintain previously described open-reading frames. It is possible that a few true mutations were missed due to this assumption.

### **2.3.2 Genome annotation**

Four other Dryvax-derived whole genome sequences had been assembled and annotated previously to starting this project. To facilitate the direct comparison between these viruses we first produced a master genome, which contains all of the ORFs that have been identified as being encoded by one or more Dryvax derivatives, and then transferred the earlier annotations to our sequences. The genes were numbered according to the system used to annotate strains ACAM-2K and CL3 [165], although we added four additional genes (DVX 063.5, DVX080.5, DVX 164.5 and DVX192.5), which are widely conserved amongst VACV strains including VACV strain WR.

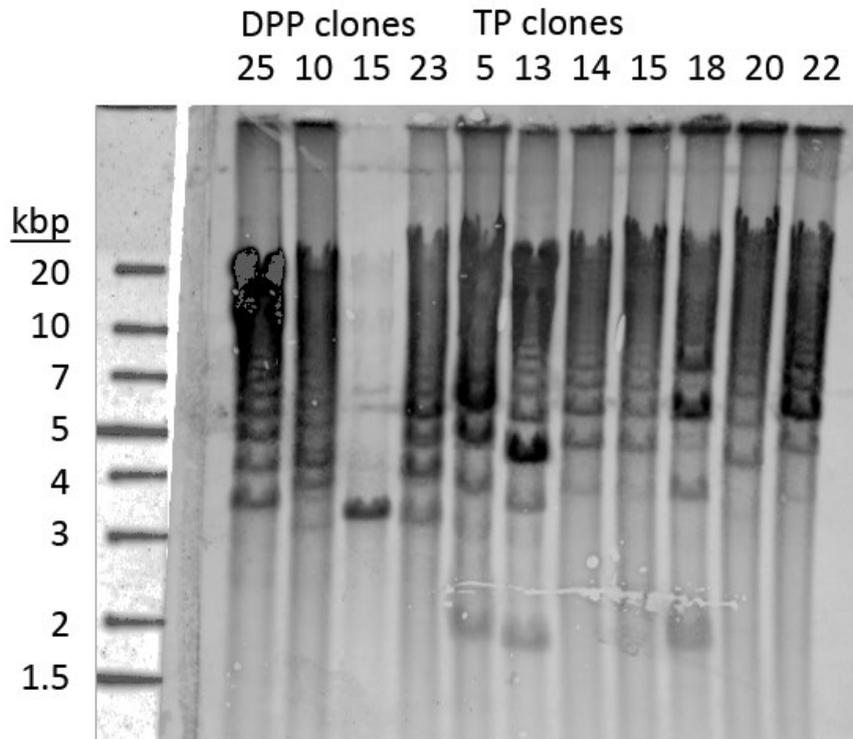


Figure 2.2 Telomere repeat patterns shown in Southern blot

DNA was extracted from Dryvax (DPP25, 10, 15, 23) and TianTan clones (TP5, 13, 14, 15, 18, 20, 22), then digested with *SalI*, and detected using a probe targeting the telomere repeats. *SalI* cuts the very end of ORF001 (open reading frame), resulting in telomere fragmentation. Consistent with previous study [102], all clones, except the DPP15, show variable telomere lengths.

This system generally defines an ORF as comprising at least 50 amino acids and where a gene was identified as being fragmented it refers to the fact that a much larger contiguous ORF is seen in either one of the Dryvax clones (e.g. M1L or DVX\_041), or it has been truncated by at least one-third of its original length, or split into two or more pieces, in another vaccinia strain (e.g. strain Copenhagen). In some cases it is not clear which of several ATG codons encoded the initiating MET; thus the different annotations can create the false appearance of variably sized ORFs. Where possible, we used the known transcription start site to identify the most likely start codon [206]. This was not possible with intermediate and late genes, so where these discrepancies were noted we identified the most highly conserved, consensus ATG as the probable start codon. Table 2.4 summarizes all of the genes and other large ORFs encoded by these cloned viruses, along with the reported gene complements for strains ACAM-2K, CL3, Duke, and 3737.

We should note that our analysis was limited mainly to examine coding regions of these frames. We recognize that mutations in other genetic elements (e.g. promoters) could play an important role in viral biology. However, there are currently no tools available to examine the effects of SNPs on particular promoter activities.

### **2.3.3 Genetic similarities between Dryvax clones**

An important feature of the data summarized in Table 2.4 is the large number of genes that seem to be conserved between the different viruses, insofar as they comprise nearly identical ORFs. This is not a surprising feature of the genes encoded within the conserved central core of these viruses (see below), but it does suggest that many genes in the relatively unstable telomere region also provide some selective advantage under the conditions in which these viruses have been propagated.

Many of the conserved genes appear to regulate inflammatory (and other) host antiviral processes such as DVX\_001 (a chemokine-binding protein), DVX\_015 (an IL-1 receptor antagonist), DVX\_013 (SPI-1 serpin), DVX\_034 (a secreted complement binding protein), and DVX\_042 (an NF $\kappa$ B inhibitor). Kretzschmar et al. [207] have noted that vaccines derived from the New York City Board of Health strain produced ~10-fold less postvaccinial encephalitis

compared with other once widely used smallpox vaccines (e.g. Bern, Lister, and Copenhagen). It is possible that the conservation of such anti-inflammatory genes is a contributing factor behind this phenotype, although this is clearly not a sufficient explanation as most of these genes are still also encoded by strains Lister and Copenhagen (Bern has not been sequenced). It may also confer a replicative advantage for viruses that can delay the induction of a sterilizing immune response in the animals where these vaccines were propagated. Several genes of still unknown function (e.g. DVX\_009, 010, and 012) have also been maintained intact within the telomeric regions and thus may be deserving of further investigation.

### **2.3.4 Genetic differences between Dryvax clones**

Table 2.5 summarizes the major differences in gene complement between the 15 sequenced Dryvax clones. A great many genes differ only slightly in length and this is commonly due to the acquisition of one or more in-frame deletions.

One of the more interesting examples of these genes is DVX\_204 (B11R), which encodes anywhere from 1-to-12 copies of a six-nucleotide repeat (5' ACAGAT 3') in the 5'-end of the gene (Figure 2.3). The in-frame deletion (or insertion) of this 6 bp sequence creates ORFs ranging in length from 219 to 285 bp amongst these VACV clones (Table 2.5). This variable repeat is conserved amongst Orthopoxviruses and the longest reported set of these repeats (23 copies) is encoded by the B11R homolog of monkeypox virus, strain Zaire. A more typical example of this pattern of mutagenesis is seen in DVX\_142 (A12L). This gene ranges in length from 570 to 579 bp, due to combinations of 0, 3 and/or 6 bp in-frame indels near the middle of the gene (Figure 2.4). A notable feature of the generation of the deletion variants is that over long periods of time it could rationalize how poxviruses have evolved to encode some of the smallest known examples of several different enzymes.

Besides an accumulation of non-frameshifting indels, several genes have been disrupted by a varying pattern of frameshift mutations. For example, most Dryvax clones encode intact homologs of DVX\_084 (I4L) and DVX\_177 (A41L), but these genes are disrupted by single frameshifts in DPP17 and DPP13, respectively.

Table 2.4 Complete complement of genes and gene fragments encoded by Dryvax clones

Gene or gene fragment	DPP clone number											Other Dryvax clones			
	09	10	11	12	13	15	16	17	19	20	21	CL3	3737	2k	Duke
Chemokine-binding protein(Cop-C23L)_VAC_DVX_001	726	726	726	726	726	726	726	726	726	726	726	726	726	726	726
TNF-alpha-receptor (CrmB)VAC_DVX_002/3		441	441	441	441		441		432		441	441	441	441	441
	285	285	285	285	285	285	285	285	285	285	285	369	369	369	369
Ankyrin-like VAC_DVX_004	336	240	240	336	336	336	336	336	336	336	336	336	342	336	336
Ankyrin-like VAC_DVX_005	387	387	387	387	387	387	387	234	387	387	387	234	387	387	234
Ankyrin-like VAC_DVX_006	414	435	435	348	414	414	435	765	435	444	339	450	450	414	441
Ankyrin (CPXV-008)_VAC_DVX_007/8	453	597	597	597	453	597	453	453	597	597	453	597	597	597	597
	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149
Unknown (Cop-B22R)_VAC_DVX_009	546	546	546	546	354	546	546	546	546	546	546	546	546	546	546
Unknown(Cop-C15L)_VAC_DVX_010	264	264	264	264	264	276	270	276	270	264	270	276	276	270	276
Surface glycoprotein fragment VAC_DVX_011	156	156	156	156	156	156	156	156	156	156	156	156	156	156	156
Unknown (Cop-C14L) VAC_DVX_012	573	573	573	573	573	573	573	573	573	573	573	573	573	573	573
Serpin (SPI-1)Cop-C12L)_VAC_DVX_013	1074	1074	1074	1074	1074	1074	1074	1074	1074	1074	1074	1005	1074	1074	930
EGF growth factor (Cop-C11R)_VAC_DVX_014	423	423	423	423	423	426	426	426	426	426	426	423	420	420	426
IL-1 receptor antagonist(Cop-C10L)_VAC_DVX_015	996	996	996	996	996	996	996	996	996	996	996	996	996	996	996
Ubiquitin LigaseVAC_DVX_016/017/018	135	135	135	135	135	135	135	135	135	135	135	135	135	135	135
	252	201	201	201	252	252	252	273	201	201	201	252	252	252	252
	189	189	189	189	189	189	189	189	189	189	189	189	189	189	189

Gene or gene fragment	DPP clone number											Other Dryvax clones			
	09	10	11	12	13	15	16	17	19	20	21	CL3	3737	2k	Duke
IL-18 BP(Bsh-D7L)(CopC12L)_VAC_DVX_019	381	381	375	375	381	381	381	381	381	381	381	375	381	375	375
Ankyrin(Bang-D8L)_VAC_DVX_020/1/2/3/4	273	273	273	273	273	273	273	273	273	273	273	273	273	273	273
	522		522	522	522		540	522	522	522	522	429	429	429	429
	513	894	513	513	513	894	408	513	498	513	513	408	408	408	408
	252	252	252	252	252	252	234	252	252	252	252	234	234	234	234
	216	216	216	216	216	216	216	216	216	216	216	216	216	216	216
Unknown(Tan-TC10L)_VAC_DVX_025	168	168	168	168	168	171	171	171	168	168	168	171	171	168	168
Ankyrin(Cop-C9L)_VAC_DVX_026.1				726	726	726	726				573				
Ankyrin(Cop-C9L)_VAC_DVX_026	1905	1905	1905	1191	1191	1191	1191	1905	1905	1905	894	1905	1905	1905	1905
Unknown(Cop-C8L)_VAC_DVX_027	534	534	534	534	534	534	534	534	534	534	534	534	534	534	534
Host range virulence factor(Cop-C7L)_VAC_DVX_028	453	453		453	453	453	453	453	453	453	453	453	453	453	453
Unknown(Cop-C6L)_VAC_DVX_029	456	456	456	456	456	456	456	456	456	456	456	456	456	456	456
Unknown(Cop-C5L)_VAC_DVX_030	606	606	606	606	606	606	606	606	606	606	606	615	618	615	606
IL-1 receptor antagonist (Cop-C4L)_VAC_DVX_031/32/33	189	189	189	189	189	189	189	189	189	189	189	189	189	189	189
	387	387	231	231	231	231	231	387	231	387	231	411	231	411	231
	180	174		180	180	180	180	180	180	180	180	180	180	180	180
Complement binding (secreted)(Cop-C3L)_VAC_DVX_034	786	792	786	786	786	786	792	792	786	786	792	786	786	792	792
Kelch-like(Cop-C2L)_VAC_DVX_035	1521	1521	1521	1521	1521	1521	1521	1521	1521	1521	1521	1521	1521	1521	1521
Putative TLR signalling inhibitor(Cop-C1L)_VAC_DVX_036	672	675	675	675	675	675	672	675	672	675	672	675	675	675	675
Virokine(Cop-N1L)_VAC_DVX_037	354	354	354	354	354	354	354	354	354	354	354	354	354	354	354

Gene or gene fragment	DPP clone number											Other Dryvax clones			
	09	10	11	12	13	15	16	17	19	20	21	CL3	3737	2k	Duke
Alpha-amanitin sensitivity(Cop-N2L)_VAC_DVX_038	528	528	528	528	528	528	528	528	528	528	528	528	528	528	528
Ankyrin(Cop-M1L)_VAC_DVX_039/40/41	219	219	219	219	219		219		219	219	219	219			
	645	645		645			645	855	645	645		645			
	480	483	1200	486	1200	1383	486	483	486	486	1197	480	1419	1410	1410
NFκB inhibitor(Cop-M2L)_VAC_DVX_042	663	663	663	663	663	663	663	663	663	663	663	663	663	663	663
Ankyrin/ NFκB inhibitor(Cop-K1L)_VAC_DVX_043	855	855	855	855	855	855	855	855	855	855	855	855	855	855	855
Serpin (SPI-3)(Cop-K2L)_VAC_DVX_044	1110	1110	1110	1110	1110	1110	1110	1095	1110	1095	1095	1110	1110	1110	1110
IFN resistance/elf2 alpha-like PKR inhib (Cop-K3L) _VAC_DVX_045	267	267	267	267	267	267	267	267	267	267	267	267	267	267	267
Nicking-joining enzyme(Cop-K4L) _VAC_DVX_046	1275	1275	1275	1275	1275	1275	1275	1275	1275	1275	1275	1275	1275	1275	1275
Putative monoglyceride lipase (Cop-K5L) _VAC_DVX_047 (Cop-K6L)_VAC_DVX_048	513	516	513	513	516	516	366	366	366	516	516	366	513	411	516
	246	195	195	333	246	195	333	333	246	246	195	333	246	246	246
Host immune response repressor(Cop-K7R) _VAC_DVX_049	450	450	450	450	450	450	450	450	450	450	450	450	450	450	450
Caspase-9 (apoptosis) inhibitor(Cop-F1L) _VAC_DVX_050	681	681	669	681	681	681	681	681	681	681	681	681	681	681	681
dUTPase(Cop-F2L)_VAC_DVX_051	444	444	444	444	444	444	444	444	444	444	444	444	444	444	444
Kelch-like, innate immune response modifier, virulence factor Cop-F3L) _VAC_DVX_052.1/052								678							
	1434	1434	1434	1443	1434	1434	1434	678	1443	1443	1434	1443	1434	1443	1443
Ribonucleotide reductase small subunit(Cop-F4L) _VAC_DVX_053	960	960	960	960	960	960	960	960	960	960	960	960	960	960	960
36kDa major membrane protein(Cop-F5L) _VAC_DVX_054	966	966	966	966	966	966	966	966	966	966	966	966	966	966	966
Unknown(Cop-F6L)_VAC_DVX_055	225	225	225	225	225	225	225	225	225	225	225	225	225	225	225
Unknown(Cop-F7L)_VAC_DVX_056	243	243	243	243	243	243	243	243	243	243	243	243	243	243	243

Gene or gene fragment	DPP clone number											Other Dryvax clones			
	09	10	11	12	13	15	16	17	19	20	21	CL3	3737	2k	Duke
Cytoplasmic protein(Cop-F8L)_VAC_DVX_057	198	198	198	198	198	198	198	198	198	198	198	198	198	198	198
S-S bond formation pathway protein(Cop-F9L)_VAC_DVX_058	639	639	639	639	639	639	639	639	639	639	639	639	639	639	639
Ser/Thr kinase Morph(Cop-F10L)_VAC_DVX_059	1320	1320	1320	1320	1320	1320	1320	1320	1320	1320	1320	1320	1320	1320	1320
RhoA signalling inhibitor, virus release protein(Cop-F11L)_VAC_DVX_060	1047	1047	1047	1047	1047	1047	1047	1047	1065	1047	1047	1047	1047	1047	1047
IEV associated (Cop-F12L)_VAC_DVX_061	1908	1908	1908	1908	1908	1908	1908	1908	1908	1908	1908	1908	1908	1908	1908
Palmitoylprotein; major IEV antigen (Cop-F13L)_VAC_DVX_062	1119	1119	1119	1119	1119	1119	1119	1119	1119	1119	1119	1119	1119	1119	1119
Unknown(Cop-F14L)_VAC_DVX_063	222	222	222	222	222	222	222	222	222	222	222	222	222	222	222
IMV protein (YMTV-28.5L)(WR53.5)_VAC_DVX_063.5	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150
Unknown(Cop-F15L)_VAC_DVX_064	477	477	477	477	477	477	477	477	477	477	477	477	477	477	477
Unknown(Cop-F16L)_VAC_DVX_065	696	696	696	696	696	696	696	696	696	696	696	696	696	696	696
DNA-binding phosphoprotein,VP11(Cop-F17R)_VAC_DVX_066	306	306	306	306	306	306	306	306	306	306	306	306	306	306	306
Poly-A polymerase-large sub. (VP55)(Cop-E1L)_VAC_DVX_067	1440	1440	1440	1440	1440	1440	1440	1440	1440	1440	1440	1440	1440	1440	1440
Required for IEV morphogenesis(Cop-E2L)_VAC_DVX_068	2214	2214	2214	2214	2214	2214	2214	2214	2214	2214	2214	2214	2214	2214	2214
IFN resistance/PKR inhibitor(Cop-E3L)_VAC_DVX_069	573	573	573	573	573	573	573	573	573	573	573	573	573	573	573
RNA polymerase (RPO30) (Cop-E4L)_VAC_DVX_070	780	780	780	780	780	780	780	780	780	780	780	780	780	780	780
Virosome component (Cop-E5R)_VAC_DVX_071/072	570	570	996	570	507	570	570	570	570	570	570	570	570	570	996
	450	450		450	450	450	450	450	450	450	450	450	450	450	
Virion protein; required for assembly(Cop-E6R)_VAC_DVX_073	1704	1704	1704	1704	1704	1704	1704	1704	1704	1704	1704	1704	1704	1704	1704
Soluble/myristyl EEV(Cop-E7R)_VAC_DVX_074	501	501	501	501	501	501	501	501	501	501	501	501	501	501	501
ER-localized MP; virion core protein(Cop-E8R)_VAC_DVX_075	822	822	822	822	822	822	822	822	822	822	822	822	822	822	822

Gene or gene fragment	DPP clone number											Other Dryvax clones			
	09	10	11	12	13	15	16	17	19	20	21	CL3	3737	2k	Duke
DNA polymerase(Cop-E9L)_VAC_DVX_076	3018	3021	3021	3021	3018	3018	3021	3018	3021	3018	3018	3018	3018	3018	3018
Sulfhydryl oxidase(FAD-linked)(Cop-E10R)_VAC_DVX_077	288	288	288	288	288	288	288	288	288	288	288	288	288	288	288
Virion core protein(Cop-E11L)_VAC_DVX_078	390	390	390	390	390	390	390	390	390	390	390	390	390	390	390
Unknown(Cop-O1L)_VAC_DVX_079	2001	2001	2001	2001	2001	2001	2001	2001	2001	2001	2001	2001	2001	2001	2001
Glutaredoxin 1(Cop-O2L)_VAC_DVX_080	324	324	324	324	324	324	324	324	324	327	324	327	327	327	327
Entry-fusion complex(Cop-O3L, WR069.5)_VAC_DVX_080.5	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108
DNA-binding core protein(Cop-I1L)_VAC_DVX_081	939	939	939	939	939	939	939	939	939	939	939	939	939	939	939
IMV membrane protein(Cop-I2L)_VAC_DVX_082	222	222	222	222	222	222	222	222	222	222	222	222	222	222	222
ssDNA-binding phosphoprotein(Cop-I3L)_VAC_DVX_083	810	810	810	810	810	810	810	810	810	810	810	810	810	810	810
Ribonucleotide reductase large subunit (Cop-I4L)_VAC_DVX_084/084.1	2316	2292	2316	2310	2316	2316	2316	1251	2292	2316	2316	2316	2316	2316	2316
								1152							
IMV protein VP13(Cop-I5L)_VAC_DVX_085	240	240	240	240	240	240	240	240	240	240	240	240	240	240	240
Telomere-binding protein(Cop-I6L)_VAC_DVX_086	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149
Virion core cysteine protease(Cop-I7L)_VAC_DVX_087	1272	1272	1272	1272	1272	1272	1272	1272	1272	1272	1272	1272	1272	1272	1272
RNA-helicase/NPH-II(Cop-I8R)_VAC_DVX_088	2031	2031	2031	2031	2031	2031	2031	2031	2031	2031	2031	2031	2031	2031	2031
Predicted metalloprotease(Cop-G1L)_VAC_DVX_089	1776	1776	1776	1776	1776	1776	1776	1776	1776	1776	1776	1776	1776	1776	1776
Entry/fusion complex component(Cop-G3L)_VAC_DVX_090	336	336	336	336	336	336	336	336	336	336	336	336	336	336	336
VLTF (Late trans. elongation factor)(Cop-G2R)_VAC_DVX_091	663	663	663	663	663	663	663	663	663	663	663	663	663	663	663
Disulfide oxidoreductase(Cop-G4L)_VAC_DVX_092	375	375	375	375	375	375	375	375	375	375	375	375	375	375	375
FEN1-like nuclease(Cop-G5R)_VAC_DVX_093	1305	1305	1305	1305	1305	1305	1305	1305	1305	1305	1305	1305	1305	1305	1305

Gene or gene fragment	DPP clone number											Other Dryvax clones			
	09	10	11	12	13	15	16	17	19	20	21	CL3	3737	2k	Duke
RNA polymerase (RPO7)(Cop-G5.5R) _VAC_DVX_094	192	192	192	192	192	192	192	192	192	192	192	192	192	192	192
Virulence factor/NlpC/P60 superfamily protein(Cop-G6R) _VAC_DVX_095	498	498	498	498	498	498	498	498	498	498	498	498	498	498	498
Virion phosphoprotein; early morphogenesis(Cop-G7L) _VAC_DVX_096	1116	1116	1116	1116	1116	1116	1116	1116	1116	1116	1116	1116	1116	1116	1116
VLTF-1 (Late gene transcription factor)(Cop-G8R) _VAC_DVX_097	783	783	783	783	783	783	783	783	783	783	783	783	783	783	783
Entry-fusion complex protein(Cop-G9R) _VAC_DVX_098	1023	1023	1023	1023	1023	1023	1023	1023	1023	1023	1023	1023	1023	1023	1023
Myristylated MP IMV(Cop-L1R) _VAC_DVX_099	753	753	753	753	753	753	753	753	753	753	753	753	753	753	753
Crescent membrane and immature virion formation protein(Cop-L2R) _VAC_DVX_100	258	258	258	258	258	258	258	258	258	258	258	258	258	258	258
Internal virion protein(Cop-L3L) _VAC_DVX_101	1053	1053	1053	1053	1053	1053	1053	1053	1053	1053	1053	1053	1053	1053	1053
Core package/transcription(Cop-L4R) _VAC_DVX_102	756	756	756	756	756	756	756	756	756	756	756	756	756	756	756
Entry and Fusion IMV protein(Cop-L5R) _VAC_DVX_103	387	387	387	387	387	387	387	387	387	387	387	387	387	387	387
Virion morphogenesis(Cop-J1R) _VAC_DVX_104	462	462	462	462	462	462	462	462	462	462	462	462	462	462	462
Thymidine kinase(Cop-J2R) _VAC_DVX_105	534	534	534	534	534	534	534	534	534	534	534	534	534	534	534
Poly-A polymerase-small sub. (VP39)(Cop-J3R) _VAC_DVX_106	1002	1002	1002	1002	1002	1002	1002	1002	1002	1002	1002	1002	1002	1002	1002
RNA polymerase (RPO22)(Cop-J4R) _VAC_DVX_107	558	558	558	558	558	558	558	558	558	558	558	558	558	558	558
Membrane-associated proteins(Cop-J5L) _VAC_DVX_108	402	402	402	402	402	402	402	402	402	402	402	402	402	402	402
RNA polymerase (RPO147)(Cop-J6R) _VAC_DVX_109	3861	3861	3861	3861	3861	3861	3861	3861	3861	3861	3861	3861	3861	3861	3861
Tyr/Ser phosphatase,INF-gamma inhibitor(Cop-H1L) _VAC_DVX_110	516	516	516	516	516	516	516	516	516	516	516	516	516	516	516
Entry and cell-cell Fusion(Cop-H2R) _VAC_DVX_111	570	570	570	570	570	570	570	570	570	570	570	570	570	570	570
IMV heparin binding surface protein(Cop-H3L) _VAC_DVX_112	975	975	975	975	975	975	975	975	975	975	975	975	975	975	975
RAP94(RNA pol assoc protein)(Cop-H4L) _VAC_DVX_113	2388	2388	2388	2388	2388	2388	2388	2388	2388	2388	2388	2388	2388	2388	2388

Gene or gene fragment	DPP clone number											Other Dryvax clones			
	09	10	11	12	13	15	16	17	19	20	21	CL3	3737	2k	Duke
VLTF-4 (Late transcription factor)(Cop-H5R)_VAC_DVX_114	612	612	612	612	612	612	612	612	612	612	612	612	612	612	612
Topoisomerase type I(Cop-H6R)_VAC_DVX_115	945	945	945	945	945	945	945	945	945	945	945	945	945	945	945
Involved in crescent membrane and immature virions formation (Cop-H7R)_VAC_DVX_116	441	441	441	441	441	441	441	441	441	441	441	441	441	441	441
mRNA capping enzyme large subunit(Cop-D1R)_VAC_DVX_117	2535	2535	2535	2535	2535	2535	2535	2535	2535	2535	2535	2535	2535	2535	2535
Virion core protein(Cop-D2L)_VAC_DVX_118	441	441	441	441	441	441	441	441	441	441	441	441	441	441	441
Virion core protein(Cop-D3R)_VAC_DVX_119	714	714	714	714	714	714	714	714	714	714	714	714	714	714	714
Uracil-DNA glycosylase(Cop-D4R)_VAC_DVX_120	657	657	657	657	657	657	657	657	657	657	657	657	657	657	657
NTPase(Cop-D5R)_VAC_DVX_121	2358	2358	2358	2358	2358	2358	2358	2358	2358	2358	2358	2358	2358	2358	2358
VETF-s (Early transcription factor small)(Cop-D6R)_VAC_DVX_122	1914	1914	1914	1914	1914	1914	1914	1914	1914	1914	1914	1914	1914	1914	1914
RNA polymerase (RPO18)(Cop-D7R)_VAC_DVX_123	486	486	486	486	486	486	486	486	486	486	486	486	486	486	486
Carbonic anhydrase; GAG-binding IMV(Cop-D8L)_VAC_DVX_124	915	915	915	915	915	915	915	915	915	915	915	915	915	915	915
mutT motif/decapping enzyme(Cop-D9R)_VAC_DVX_125	642	642	642	642	642	642	642	642	642	642	642	642	642	642	642
mutT motif/decapping enzyme(Cop-D10R)_VAC_DVX_126	747	747	747	747	747	747	747	747	747	747	747	747	747	747	747
NPH-I/Helicase, virion(Cop-D11L)_VAC_DVX_127	1896	1896	1896	1896	1896	1896	1896	1896	1896	1896	1896	1896	1896	1896	1896
mRNA capping enzyme small subunit(Cop-D12L)_VAC_DVX_128	864	864	864	864	864	864	864	864	864	864	864	864	864	864	864
Trimeric virion coat protein (rifampicin res.)(Cop-D13L)_VAC_DVX_129	1656	1656	1656	1656	1656	1656	1656	1656	1656	1656	1656	1656	1656	1656	1656
VLTF-2 (late transcription factor2)(Cop-A1L)_VAC_DVX_130	453	453	453	453	453	453	453	453	453	453	453	453	453	453	453
VLTF-3 (late transcription factor3)(Cop-A2L)_VAC_DVX_131	675	675	675	675	675	675	675	675	675	675	675	675	675	675	675
S-S bond formation pathway protein (Cop-A2.5L)_VAC_DVX_132	231	231	231	231	231	231	231	231	231	231	231	231	231	231	231
P4b precursor(Cop-A3L)_VAC_DVX_133	1935	1935	1935	1935	1935	1935	1935	1935	1935	1935	1935	1935	1935	1935	1935

Gene or gene fragment	DPP clone number											Other Dryvax clones			
	09	10	11	12	13	15	16	17	19	20	21	CL3	3737	2k	Duke
Core protein(Cop-A4L)_VAC_DVX_134	840	840	822	846	846	840	840	840	846	840	840	846	846	846	846
RNA polymerase (RPO19)(Cop-A5R) _VAC_DVX_135	495	495	495	495	495	495	495	495	495	495	495	495	495	495	495
Virion morphogenesis, virion core protein(Cop-A6L)_VAC_DVX_136	1119	1119	1119	1119	1119	1119	1119	1119	1119	1119	1119	1119	1119	1119	1119
VETF-L (early trans. factor large)(Cop-A7L) _VAC_DVX_137	2133	2133	2133	2133	2133	2133	2133	2133	2133	2133	2133	2133	2133	2133	2133
VITF-3 (intermediate trans. factor 3)(Cop-A8R) _VAC_DVX_138	867	867	867	867	867	867	867	867	867	867	867	867	867	867	867
Membrane protein, early morph(Cop-A9L)_VAC_DVX_139	285	327	327	327	327	327	327	327	327	327	327	327	327	327	327
P4a precursor(Cop-A10L)_VAC_DVX_140	2676	2676	2676	2676	2676	2676	2676	2676	2676	2676	2676	2676	2676	2676	2676
viral membrane formation(Cop-A11R) _VAC_DVX_141	957	957	957	957	957	957	957	957	957	957	957	957	957	957	957
Structural protein(Cop-A12L)_VAC_DVX_142	576	576	576	576	570	576	570	576	576	576	576	576	576	579	576
Virion maturation; IMV membrane protein(Cop-A13L)_VAC_DVX_143	213	213	213	213	207	207	213	213	213	213	213	213	213	213	207
IMV PO4 MP(Cop-A14L)_VAC_DVX_144	273	273	273	273	273	273	273	273	273	273	273	273	273	273	273
IMV-MP/virulence factor(Cop-A14.5L) _VAC_DVX_145	162	162	162	162	162	162	162	162	162	162	162	162	162	162	162
Core protein, morphogenesis(Cop-A15L)_VAC_DVX_146	285	285	285	285	285	285	285	285	285	285	285	285	285	285	285
Myristylated entry/cell fusion protein(Cop-A16L) _VAC_DVX_147	1137	1137	1137	1137	1134	1137	1137	1137	1137	1137	1137	1137	1134	1137	1134
IMV MP PO4(Cop-A17L)_VAC_DVX_148	612	612	612	612	612	612	612	612	612	612	612	612	612	612	612
DNA helicase; transcript release factor(Cop-A18R)_VAC_DVX_149	1482	1482	1482	1482	1482	1482	1482	1482	1482	1482	1482	1482	1482	1482	1482
Zinc finger-like protein(Cop-A19L) _VAC_DVX_150	234	234	234	234	234	234	234	234	234	234	234	234	234	234	234
Entry and cell-cell Fusion(Cop-A21L) _VAC_DVX_151	354	354	354	354	354	354	354	354	354	354	354	354	354	354	354
DNA polymerase processivity factor(Cop-A20R)_VAC_DVX_152	1281	1281	1281	1281	1281	1281	1281	1281	1281	1281	1281	1281	1281	1281	1281
Holliday junction resolvase(Cop-A22R)_VAC_DVX_153	531	531	531	531	531	531	531	531	531	531	531	531	564	564	531

Gene or gene fragment	DPP clone number											Other Dryvax clones			
	09	10	11	12	13	15	16	17	19	20	21	CL3	3737	2k	Duke
VITF-3 (intermediate trans. factor 3)(Cop-A23R)_VAC_DVX_154	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149
RNA polymerase132(RPO132)(Cop-A24R)_VAC_DVX_155	3495	3495	3495	3495	3495	3495	3495	3495	3495	3495	3495	3495	3495	3495	3495
Cowpox A-type inclusion protein VAC_DVX_156/7/8/9	198	198	198	198	198	198	198	198	198	198	198	198	198	198	198
	465	465	465	465	465	465	465	465	465	465	465	465	465	465	465
	684	684	684	684	684	684	684	684	684	684	684	684	684	684	684
	2166	2166	2166	2166	2166	2166	2166	2166	2166	2166	2166	2166	2166	2166	2166
P4c precursor(Cop-A26L)_VAC_DVX_160	1503	1503	1503	1503	1503	1503	1503	1503	1503	1503	1503	1503	1503	1503	1503
Fusion protein, IMV surface protein(Cop-A27L)_VAC_DVX_161	333	333	333	333	333	333	333	333	333	333	333	333	333	333	333
IMV MP/Virus entry(Cop-A28L)_VAC_DVX_162	441	441	441	441	441	441	441	441	441	441	441	441	441	441	441
RNA polymerase35(RPO35)(Cop-A29L)_VAC_DVX_163	918	918	918	918	918	918	918	918	918	918	918	918	918	918	918
Virion morphogenesis, IMV protein(Cop-A30L)_VAC_DVX_164	234	234	234	234	234	234	234	234	234	234	234	234	234	234	234
Unknown (YMTV-120.5L) (WR153.5) _VAC_DVX_164.5	129	129	129	129	129	129	129	129	129	129	129	129	129	129	129
Unknown(Cop-A31R)_VAC_DVX_165	375	375	375	375	375	375	375	375	375	375	375	375	375	375	375
ATPase/DNA packaging protein(Cop-A32L) _VAC_DVX_166	813	813	813	813	813	813	813	813	813	813	813	813	813	813	813
EEV membrane phosphoglycoprotein; C-type lectin-like domain(Cop-A33R) _VAC_DVX_167	558	558	558	558	558	558	558	558	558	558	558	558	558	558	558
C-type lectin-like EEV/IEV protein(Cop-A34R)_VAC_DVX_168	507	507	507	507	507	507	507	507	507	507	507	507	507	507	507
MHC-II antigen presentation inhibitor(Cop-A35R)_VAC_DVX_169	531	531	531	531	531	531	531	531	531	531	531	531	531	531	531
IEV transmembrane phosphoprotein(Cop-A36R)_VAC_DVX_170	666	666	666	666	666	666	666	666	666	666	666	666	666	666	666
Unknown(Cop-A37R)_VAC_DVX_171	801	801	792	801	801	801	801	792	792	792	792	792	792	792	792
Unknown(Gar-A43R)_VAC_DVX_172	189	189	189	189	189	189	189	189	189	189	189	189	189	189	189

Gene or gene fragment	DPP clone number											Other Dryvax clones			
	09	10	11	12	13	15	16	17	19	20	21	CL3	3737	2k	Duke
CD47-like, integral membrane protein(Cop-A38L)_VAC_DVX_173	834	834	834	834	834	834	834	834	834	834	834	834	834	834	834
Semaphorin(Cop-A39R)_VAC_DVX_174/175	681	768	399	399	399	774	687	405	681	768	681	405	681	399	687
	429	429	429	429	429	429	429	429		429	207	429		429	429
Lectin homolog(Cop-A40R)_VAC_DVX_176	507	507	507	507	492	507	507	507	507	507	492	507	507	507	507
Chemokine binding protein (Cop-A41L)_VAC_DVX_177/177.1	660	660	660	660	327	660	660	660	660	660	660	660	660	660	660
					207										
Profilin-like; ATI-localized(Cop-A42R)_VAC_DVX_178	402	402	402	402	402	402	402	402	402	402	402	402	402	402	402
Membrane glycoprotein-class I(Cop-A43R)_VAC_DVX_179	582	582	582	582	582	582	582	582	582	582	582	585	582	585	591
Unknown(MVA-156R)_VAC_DVX_180									237			237	237		237
Hydroxysteroid dehydrogenase(Cop-A44L)_VAC_DVX_181	1041	1041	1041	1041	1041	1041	1041	1041	1041	1041	1041	1041	1041	1041	1041
Superoxide dismutase-like(Cop-A45R)_VAC_DVX_182	378	378	378	378	378	378	378	378	378	378	378	378	378	378	378
IL-1/TLR signaling inhibitor(Cop-A46R)_VAC_DVX_183	723	723	723	723	723	723	723	723	723	723	723	723	723	723	723
Unknown; immunoprevalent protein(Cop-A47L)_VAC_DVX_184	735	735	735	735	735	735	735	735	735	735	735	735	735	735	735
Thymidylate kinase(Cop-A48R)_VAC_DVX_185	615	615	615	615	615	615	615	615	615	615	615	684	615	615	684
Putative phosphotransferase(Cop-A49R)_VAC_DVX_186	432	432	432	432	432	432	432	432	432	432	432	489	489	489	489
ATP-dependent DNA ligase(Cop-A50R)_VAC_DVX_187	1659	1659	1659	1659	1659	1659	1659	1659	1659	1659	1659	1659	1659	1659	1659
Unknown(Cop-A51R)_VAC_DVX_188/188.1	987	210	210	993	210	210	210	210	237	210	216	993	210	210	210
		801	801		801	801	801	801	759	801	807		801	801	801
Intracellular TLR and IL-1,NFκB sig.inhibitor(Cop-A52R)_VAC_DVX_189	573	573	573	573	573	573	573	573	573	573	573	573	573	573	573
TNF receptor (CrmC)(Cop-A53R)_VAC_DVX_190/190.1	312	312	561	396	246	399	399	561	561	312	399	561	561	399	558
					270										

Gene or gene fragment	DPP clone number											Other Dryvax clones			
	09	10	11	12	13	15	16	17	19	20	21	CL3	3737	2k	Duke
Kelch-like(Cop-A55R)_VAC_DVX_191	1695	1695	1695	1695	1695	1695	1695	1695	1695	1695	1695	1695	1695	1695	1695
Hemagglutinin(Cop-A56R)_VAC_DVX_192	930	930	933	930	933	930	930	930	930	930	930	930	933	930	930
Guanylate kinase (WR181.5)VAC_DVX_192.5	114	114	114	114	114	114	114	114	114	114	114	114	114	114	114
Guanylate kinase(Cop-A57R)_VAC_DVX_193	456	456	456	456	456	456	456	456	456	456	456	456	456	456	456
Ser/Thr kinase(Cop-B1R)_VAC_DVX_194	903	903	903	903	903	903	903	903	903	903	903	903	903	903	903
Schlafen(Cop-B2R)_VAC_DVX_195	660	660	660	660	660	660	660	660	660	660	660	660	660	660	660
Schlafen(Cop-B3R)_VAC_DVX_196	795	795	795	795	795	795	795	795	786	795	786	795	801	375	801
Ankyrin(Cop-B4R)_VAC_DVX_197	1677	1677	1677	1677	1677	1677	1677	1677	1677	1677	1677	1677	1677	1677	1677
EEV type-1 membrane glycoprotein; protective antigen; virulence protein (Cop-B5R)_VAC_DVX_198	954	954	954	954	954	954	954	954	954	954	954	954	954	954	954
Ankyrin-like protein(Cop-B6R)_VAC_DVX_199	522	522	504	522	522	504	504	522	522	522	522	522	522	522	522
Virulence, ER resident(Cop-B7R)_VAC_DVX_200	549	549	549	549	549	549	549	549	549	549	549	549	549	549	549
Interferon-gamma receptor-like(Cop-B8R)_VAC_DVX_201	819	819	819	819	819	819	819	819	819	819	819	819	819	819	819
Virulence factor(Cop-B9R)_VAC_DVX_202	234	234	234	234	234	234	234	234	234	234	234	234	234	234	234
Kelch-like (CPV-GRI-B9R)(Cop-B10R)_VAC_DVX_203	501	501	501	501	501	501	501	501	501	480	501	501	501	501	501
Unknown(Cop-B11R)_VAC_DVX_204	219		255	255	255	255	255	255	255	273	255	279	261	219	285
Ser/Thr kinase(Cop-B12R)_VAC_DVX_205	852	852	852	852	852	852	852	852	852	852	852	852	852	852	852
Serpin (SPI-2)(Cop-B13R)_VAC_DVX_206	381	381	381	381	381	381	381	381	297	381	381	381	381	381	381
Serpin (SPI-2)(Cop-B14R)_VAC_DVX_207	669	669	669	669	669	669	669	669	669	669	669	669	669	669	669
Unknown(Cop-B15R)_VAC_DVX_208	450	450	450	450	450	450	450	450	450	450	450	450	450	450	450
IL-1-beta-receptor(Cop-B16R)_VAC_DVX_209	981	981	981	981	981	981	981	1002	981	981	981	981	981	981	981

Gene or gene fragment	DPP clone number											Other Dryvax clones			
	09	10	11	12	13	15	16	17	19	20	21	CL3	3737	2k	Duke
Unknown(Cop-B17L)_VAC_DVX_210	1023	1023	1023	1023	1023	1023	1023		1023	1023	1023	1023	1023	1023	1023
Ankyrin(Cop-B18R)_VAC_DVX_211	1725	1725	1725	1725	1725	1725	1725		1725	1725	1725	1725	1725	1725	1725
IFN-alpha/beta-receptor(Cop-B19R)_VAC_DVX_212	798	798	798	798	1062	798	798		798	798	1062	1062	798	798	1056
Ankyrin (Bang-B18R) (Cop-B20R)-VAC_DVX_213					1650						1650	2376			1650
Kelch-like(EV-M-167) -VAC_DVX_214												243			
Kelch-like(EV-M-167) -VAC_DVX_215												945			
Kelch-like(EV-M-167) -VAC_DVX_216					405						405	405			405
TIR paralog DVX_217/026	276	276	276	276		276	276		276	276				276	
TIR paralog DVX_218/025	168	168	168	168		171	171		168	168			171	168	
TIR paralog DVX_219/024	216	216	216	216		216	216		216	216				216	
TIR paralog DVX_220/023	252	252	252	252		252	234		252	252			234	234	
TIR paralog DVX_221/022	513	894	513	513		894	408		498	513			408	408	
TIR paralog DVX_222/021	522		522	522			540		522	522			429	429	
TIR paralog DVX_223/020	273	273	273	273		273	273		273	273			273	273	
TIR paralog DVX_224/019	381	381	375	375		381	381		381	381			381	375	
TIR paralog DVX_225/018	189	189	189	189		189	189		189	189			189	189	
TIR paralog DVX_226/017	252	201	201	201		252	252	273	201	201			252	252	
TIR paralog DVX_227/015	996	996	996	996		996	996	996	996	996			996	996	
TIR paralog DVX_228/014	423	423	423	423		426	426	426	426	426			423	420	
TIR paralog DVX_229/013	1074	1074	1074	1074	1074	1074	1074	1074	1074	1074	1074	1062	1074	1074	1074

Gene or gene fragment	DPP clone number											Other Dryvax clones			
	09	10	11	12	13	15	16	17	19	20	21	CL3	3737	2k	Duke
TIR paralog DVX_230/012	573	573	573	573	573	573	573	573	573	573	573	573	573	573	573
TIR paralog DVX_231/011	156	156	156	156	156	156	156	156	156	156	156	156	156	156	156
TIR paralog DVX_232/010	264	264	264	264	264	276	270	276	270	264	270	276	276	270	276
TIR paralog DVX_233/009	546	546	546	546	354	546	546	546	546	546	546	546	546	546	546
TIR paralog DVX_234/008	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149
TIR paralog DVX_235/007	453	597	597	597	453	597	453	453	597	597	453	597	597	597	597
TIR paralog DVX_236/006	414	435	435	348	414	414	435	765	435	444	339	450	450	414	441
TIR paralog DVX_237/005	387	387	387	387	387	387	387	234	387	387	387	234	387	387	234
TIR paralog DVX_238/004	336	240	240	336	336	336	336	336	336	336	336	336	342	336	336
TIR paralog DVX_239/003	285	285	285	285	285	285	285	285	285	285	285	369	#	369	369
TIR paralog DVX_240/002		441	441	441	441		441		432		441	441	#	441	441
TIR paralog DVX_241/001	726	726	726	726	726	726	726	726	726	726	726	726	#	726	726

# Sequence is incomplete

Table 2.5 Genes differences in gene length (nt) and gene complement between Dryvax clones<sup>a</sup>

Gene ID	DPP clone number											Dryvax			
	09	10	11	12	13	15	16	17	19	20	21	CL3	3737	2k	Duke
TNF-alpha-receptor (CrmB) (Cop-C22L)_VAC_DVX_002/003		441	441	441	441		441		432		441	441	441	441	441
	285	285	285	285	285	285	285	285	285	285	285	369	369	369	369
Ankyrin-likeVAC_DVX_004/005/006	336	240	240	336	336	336	336	336	336	336	336	336	342	336	336
	387	387	387	387	387	387	387	234	387	387	387	234	387	387	234
	414	435	435	348	414	414	435	765	435	444	339	450	450	414	441
Ankyrin (CPXV-008)_VAC_DVX_007/008	453	597	597	597	453	597	453	453	597	597	453	597	597	597	597
	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149	1149
Unknown (Cop-B22R)_VAC_DVX_009	546	546	546	546	354	546	546	546	546	546	546	546	546	546	546
Unknown (Cop-C15L)_VAC_DVX_010	264	264	264	264	264	276	270	276	270	264	270	276	276	270	276
Serpin (SPI-1) (Cop-C12L)_VAC_DVX_013	1074	1074	1074	1074	1074	1074	1074	1074	1074	1074	1074	1005	1074	1074	930
EGF growth factor (Cop-C11R)_VAC_DVX_014	423	423	423	423	423	426	426	426	426	426	426	423	420	420	426
Ubiquitin ligase VAC_DVX_016/017/018	135	135	135	135	135	135	135	135	135	135	135	135	135	135	135
	252	201	201	201	252	252	252	273	201	201	201	252	252	252	252
	189	189	189	189	189	189	189	189	189	189	189	189	189	189	189
IL-18 BP (Bsh-D7L) (Cop-C12L)_VAC_DVX_019	381	381	375	375	381	381	381	381	381	381	381	375	381	375	375
Ankyrin (Bang-D8L)_VAC_DVX_020/021/22/23/24	273	273	273	273	273	273	273	273	273	273	273	273	273	273	273
	522		522	522	522		540	522	522	522	522	429	429	429	429
	513	894	513	513	513	894	408	513	498	513	513	408	408	408	408
	252	252	252	252	252	252	234	252	252	252	252	234	234	234	234
	216	216	216	216	216	216	216	216	216	216	216	216	216	216	216
Unknown (Tan-TC10L)_VAC_DVX_025	168	168	168	168	168	171	171	171	168	168	168	171	171	168	168
Ankyrin (Cop-C9L)_VAC_DVX_026.1/026				726	726	726	726				573				
	1905	1905	1905	1191	1191	1191	1191	1905	1905	1905	894	1905	1905	1905	1905

Gene ID	DPP clone number											Dryvax			
	09	10	11	12	13	15	16	17	19	20	21	CL3	3737	2k	Duke
Host range virulence factor (Cop-C7L)_VAC_DVX_028	453	453		453	453	453	453	453	453	453	453	453	453	453	453
Unknown(Cop-C5L)_VAC_DVX_030	606	606	606	606	606	606	606	606	606	606	606	615 <sup>b</sup>	618 <sup>b</sup>	615 <sup>b</sup>	606
IL-1 receptor antagonist (Cop-C4L)_VAC_DVX_032/033	387	387	231	231	231	231	231	387	231	387	231	411	231	411	231
	180	174		180	180	180	180	180	180	180	180	180	180	180	180
Putative TLR <sup>c</sup> signalling inhibitor (Cop-C1L)_VAC_DVX_036	672	675	675	675	675	675	672	675	672	675	672	675	675	675	675
Ankyrin (Cop-M1L)_VAC_DVX_039/040/041	219	219	219	219	219		219		219	219	219	219			
	645	645		645			645	855	645	645		645			
	480	483	1200	486	1200	1383	486	483	486	486	1197	480	1419	1410	1410
Serpin (SPI-3)(Cop-K2L)_VAC_DVX_044	1110	1110	1110	1110	1110	1110	1110	1095	1110	1095	1095	1110	1110	1110	1110
Putative monoglyceride lipase (Cop-K5L)_VAC_DVX_047	513	516	513	513	516	516	366	366	366	516	516	366	513	411	516
Putative monoglyceride lipase (Cop-K6L)_VAC_DVX_048	246	195	195	333	246	195	333	333	246	246	195	333	246	246	246
Apoptosis inhibitor (Cop-F1L)_VAC_DVX_050	681	681	669	681	681	681	681	681	681	681	681	681	681	681	681
Kelch-like, innate immune response modifier, virulence factor (Cop-F3L)_VAC_DVX_052.1/052								678							
	1434	1434	1434	1443	1434	1434	1434	678	1443	1443	1434	1443	1434	1443	1443
RhoA signalling inhibitor, virus release protein (Cop-F11L)_VAC_DVX_060	1047	1047	1047	1047	1047	1047	1047	1047	1065	1047	1047	1047	1047	1047	1047
Virosome component (Cop-E5R)_VAC_DVX_071/072	570	570	996	570	507	570	570	570	570	570	570	570	570	570	996
	450	450		450	450	450	450	450	450	450	450	450	450	450	
DNA polymerase (Cop-E9L)_VAC_DVX_076	3018	3021	3021	3021	3018	3018	3021	3018	3021	3018	3018	3018	3018	3018	3018
Glutaredoxin 1 (Cop-O2L)_VAC_DVX_080	324	324	324	324	324	324	324	324	324	327	324	327	327	327	327
Ribonucleotide red. large sub. (Cop-I4L)_VAC_DVX_084 <sup>d</sup> /084.1	2316	2292	2316	2310	2316	2316	2316	1251	2292	2316	2316	2316	2316	2316	2316
								1152							
IMV heparin binding surface protein (Cop-H3L)_VAC_DVX_112	975	957	957	975	957	975	975	957	957	975	957	957	957	975	975
Core protein (Cop-A4L)_VAC_DVX_134	840	840	822	846	846	840	840	840	846	840	840	846	846	846	846
Membrane protein, early morph (Cop-A9L)_VAC_DVX_139	285	327	327	327	327	327	327	327	327	327	327	327	327	327	327

Gene ID	DPP clone number											Dryvax			
	09	10	11	12	13	15	16	17	19	20	21	CL3	3737	2k	Duke
Structural protein (Cop-A12L)_VAC_DVX_142	576	576	576	576	570	576	570	576	576	576	576	576	576	579	576
Virion maturation; IMV membrane protein (Cop-A13L)_VAC_DVX_143	213	213	213	213	207	207	213	213	213	213	213	213	213	213	207
Myristylated entry/cell fusion protein (Cop-A16L)_VAC_DVX_147	1137	1137	1137	1137	1134	1137	1137	1137	1137	1137	1137	1137	1134	1137	1134
Holliday junction resolvase (Cop-A22R)_VAC_DVX_153	531	531	531	531	531	531	531	531	531	531	531	531	564 <sup>b</sup>	564 <sup>b</sup>	531
Unknown (Cop-A37R)_VAC_DVX_171	801	801	792	801	801	801	801	792	792	792	792	792	792	792	792
Semaphorin (Cop-A39R)_VAC_DVX_174/175	681	768	399	399	399	774	687	405	681	768	681	405	681	399	687
	429	429	429	429	429	429	429	429		429	207	429		429	429
Lectin homolog (Cop-A40R)_VAC_DVX_176	507	507	507	507	492	507	507	507	507	507	492	507	507	507	507
Chemokine binding protein (Cop-A41L)_VAC_DVX_177/177.1	660	660	660	660	327	660	660	660	660	660	660	660	660	660	660
					207										
Membrane glycoprotein-class I (Cop-A43R)_VAC_DVX_179	582	582	582	582	582	582	582	582	582	582	582	585 <sup>b</sup>	582	585 <sup>b</sup>	591
Unknown (MVA-156R)_VAC_DVX_180									237			237	237		237
Thymidylate kinase (Cop-A48R)_VAC_DVX_185	615	615	615	615	615	615	615	615	615	615	615	684 <sup>b</sup>	615	615	684 <sup>b</sup>
Putative phosphotransferase (Cop-A49R)_VAC_DVX_186	432	432	432	432	432	432	432	432	432	432	432	489 <sup>b</sup>	489 <sup>b</sup>	489 <sup>b</sup>	489 <sup>b</sup>
Unknown (Cop-A51R)_VAC_DVX_188/188.1	987	210	210	993	210	210	210	210	237	210	216	993	210	210	210
		801	801		801	801	801	801	759	801	807		801	801	801
TNF receptor (CrmC) (Cop-A53R)_VAC_DVX_190/190.1	312	312	561	396	246	399	399	561	561	312	399	561	561	399	558
					270										
Hemagglutinin (Cop-A56R)_VAC_DVX_192	930	930	933	930	933	930	930	930	930	930	930	930	933	930	930
Schlafen (Cop-B2R)_VAC_DVX_196	795	795	795	795	795	795	795	795	786	795	786	795	801	375	801
Ankyrin-like protein (Cop-B6R)_VAC_DVX_199	522	522	504	522	522	504	504	522	522	522	522	522	522	522	522
Kelch-like (CPV-GRI-B9R) (Cop-B10R)_VAC_DVX_203	501	501	501	501	501	501	501	501	501	480	501	501	501	501	501
Unknown (Cop-B11R)_VAC_DVX_204	219		255	255	255	255	255	255	255	273	255	279	261	219	285
Serpin (SPI-2) (Cop-B13R)_VAC_DVX_206	381	381	381	381	381	381	381	381	297	381	381	381	381	381	381

Gene ID	DPP clone number											Dryvax			
	09	10	11	12	13	15	16	17	19	20	21	CL3	3737	2k	Duke
IL-1-beta-receptor (Cop-B16R)_VAC_DVX_209	981	981	981	981	981	981	981	1002	981	981	981	981	981	981	981
Unknown (Cop-B17L)_VAC_DVX_210	1023	1023	1023	1023	1023	1023	1023		1023	1023	1023	1023	1023	1023	1023
Ankyrin (Cop-B18R)_VAC_DVX_211	1725	1725	1725	1725	1725	1725	1725		1725	1725	1725	1725	1725	1725	1725
IFN-alpha/beta-receptor (Cop-B19R)_VAC_DVX_212	798	798	798	798	1062	798	798		798	798	1062	1062	798	798	1056
Ankyrin (Bang-B18R) (Cop-B20R)_VAC_DVX_213					1650						1650	2376			1650
Kelch-like (EV-M-167)_VAC_DVX_214/215/216												243			
												945			
					405						405	405			405
TIR paralog DVX_217/026	276	276	276	276		276	276		276	276				276	
TIR paralog DVX_218/025	168	168	168	168		171	171		168	168			171	168	
TIR paralog DVX_219/024 TIR paralog DVX_220/023 TIR paralog DVX_221/022	216	216	216	216		216	216		216	216				216	
	252	252	252	252		252	234		252	252			234	234	
	513	894	513	513		894	408		498	513			408	408	
TIR paralog DVX_222/021 TIR paralog DVX_223/020	522		522	522			540		522	522			429	429	
	273	273	273	273		273	273		273	273			273	273	
TIR paralog DVX_224/019	381	381	375	375		381	381		381	381			381	375	
TIR paralog DVX_225/018	189	189	189	189		189	189		189	189			189	189	
TIR paralog DVX_226/017	252	201	201	201		252	252	273	201	201			252	252	
TIR paralog DVX_227/015	996	996	996	996		996	996	996	996	996			996	996	
TIR paralog DVX_228/014	423	423	423	423		426	426	426	426	426			423	420	
TIR paralog DVX_229/013	1074	1074	1074	1074	1074	1074	1074	1074	1074	1074	1074	1062	1074	1074	1074

<sup>a</sup> The table shows only genes that differ in some manner between different clones and ORFs of  $\geq 50$  amino acids. TIR paralogs DVX\_230/012, DVX\_231/011, DVX\_232/010, DVX\_233/009, DVX\_234/008, DVX\_235/007, DVX\_236/006, DVX\_237/005, DVX\_238/004, DVX\_239/003, DVX\_240/002 and DVX\_241/001 have the same arrangement as the respective left TIR.

<sup>b</sup> The start codon is probably assigned incorrectly. The real gene length is probably identical to that of other “shorter” genes.

<sup>c</sup> TLR, Toll-like receptor. <sup>d</sup> Ribonucleotide red. Large sub., large subunit of ribonucleotide reductase.



```

|126,117
DPP16 ATGGCGGATAAAAAAATTAGCCGTTAGAAGCAGTTACGATGATTATATCGAAACAGTTAATAAGATTACACCACAGCTTAAAAATCTACTAGCGCAAATCGGTGGAGATGCAGCCGTC 60
DPP09 ATGGCGGATAAAAAAATTAGCCGTTAGAAGCAGTTACGATGATTATATCGAAACAGTTAATAAGATTACACCACAGCTTAAAAATCTACTAGCGCAAATCGGTGGAGATGCAGCCGTC 60
Acam_2k ATGGCGGATAAAAAAATTAGCCGTTAGAAGCAGTTACGATGATTATATCGAAACAGTTAATAAGATTACACCACAGCTTAAAAATCTACTAGCGCAAATCGGTGGAGATGCAGCCGTC 60
*****

DPP16 AAAGGAGCAACAATAATCTTAATTCTCAAACAGATGTGACTGCCGGCGCATGTGATACAAAATCAAAGAGTACAAAATGTATTACATGT-----AAATCAAATCCTCGTCTTCTTCT 180
DPP09 AAAGGAGCAACAATAATCTTAATTCTCAAACAGATGTGACTGCCGGCGCATGTGATACAAAATCAAAGAGTACAAAATGTATTACATGTAAACCAAATCAAATCCTCGTCTTCTTCT 180
Acam_2k AAAGGAGCAACAATAATCTTAATTCTCAAACAGATGTGACTGCCGGCGCATGTGATACAAAATCAAAGAGTACAAAATGTATTACATGTAAACCAAATCAAATCCTCGTCTTCTTCT 180
*****

DPP16 ACATCAACATCCAAGGCTCCAAAAATCTTCTGGTCTCTAAACGTAGAACAACAG---CACTACATCAATCAACGCAATGGATGGTCAGATTGTCCAGCTGTACTAATCGTGT 292
DPP09 ACATCAACATCCAAGGCTCCAAAAATCTTCTGGTCTCTAAACGTAGAACAACAG---CACTACATCAATCAACGCAATGGATGGTCAGATTGTCCAGCTGTACTAATCGTGT 298
Acam_2k ACATCAAGCTCCAAGGCTCCAAAAATCTTCTGGTCTCTAGACGTAGAACAACAGTACTACTACATCGTCAATGGATGGTCAGATTGTCCAGCTGTACTAATCGTGT 300
*****

DPP16 AAAATAGTTTATGTACCCTCAGAGACGGCCAAATTAGAAGTTCGTGGAATGGTCGGAGAGATCAATCAGATCTTCTAGGTATCGAATCAGTTAATGCTGGGAAAAAGAACCATCTAAA 411
DPP09 AAAATAGTTTATGTACCCTCAGAGACGGCCAAATTAGAAGTTCGTGGAATGGTCGGAGAGATCAATCAGATCTTCTAGGTATCGAATCAGTTAATGCTGGGAAAAAGAACCATCTAAA 417
Acam_2k AAAATAGTTTATGTACCCTTAGAGACGGCCAAATTAGAAGTTCGTGGAATGGTCGGAGAGATCAATCAGATCTTCTAGGTATCGACTCAGTTAATGCTGGGAAAAAGAACCATCTAAA 420
*****

DPP16 AAGATGCCTACTAATAAAAAGATTAATATATGTCGTCGGGTATGAGACGACAGGAACAGATTAATCCAGACGATTTGTCTGGATATGGGAATGTATTA 570
DPP09 AAGATGCCTACTAATAAAAAGATTAATATATGTCGTCGGGTATGAGACGACAGGAACAGATTAATCCAGACGATTTGTCTGGATATGGGAATGTATTA 576
Acam_2k AAGATGCCTACTAATAAAAAGATTAATATATGTCGTCGGGTATGAGACGACAGGAACAGATTAATCCAGACGATTTGTCTGGATATGGGAATGTATTA 579
*****

```

Figure 2.4 Variant forms of the DVX\_142 (A12L) gene carried by Dryvax clones

Three different isoforms of the gene are illustrated, two of which bear one or two in-frame deletions of amino acids within the coding sequence.

The DVX\_084 gene is of some interest as it encodes the large subunit of the ribonucleotide reductase. The gene is not essential [208] and, although most poxviruses encode a small subunit for the ribonucleotide reductase subunit, only a subset of Orthopoxviruses, and the Suipoxvirus swinepox, encode a gene for the large subunit. An interesting feature of the frameshift mutation in I4L is that it is linked to a number of nearby point mutations that are unique to DPP17 (Figure 2.5). Clusters of DNA damage and mutations are a hallmark of exposure to ionizing radiation [209] although this pattern of mutations could also be consequence of “patchy” recombination. Interestingly, horsepox virus encodes several frameshift mutations in the I4L homolog, one of which is identical to the  $\Delta T$  mutation in DPP17 [166]. This peculiar feature of the horsepox genome is discussed more in the sections that follow.

Several genes also show a classic pattern of accumulating frameshift and point mutations that seem to be progressively degrading the residual ORFs. This is presumably a consequence of the fact that, once a gene has been inactivated by an initial mutation, there is no longer any further selection for the maintenance of gene function and sequence drift can occur. For example the DVX\_039-041 ORFs derive from a single larger gene (M1L) that is intact in several of the strains including ACAM-2K and WR (DVX\_041). However, the length of the gene varies due to a combination of in-frame indels in some viruses, and the gene is disrupted by a 2 bp frameshift and/or C-to-A nonsense mutations in other viruses (Figure 2.6). These mutations appear to be assorting independently between different viruses and create at least eight different alleles of the one original gene.

```

DUKE      CTTATAAAGAATGAATGGAGTACCAGTTTCAATCTGAGATTCTATAATCGCTTCCAGACGACTCGAGCCTTATTATAGATTTGTATCTCCTTTCTCTT
3737      CTTATAAAGAATGAATGGAGTACCAGTTTCAATCTGAGATTCTATAATCGCTTCCAGACGACTCGAGCCTTATTATAGATTTGTATCTCCTTTCTCTT
Acam_2k   CTTATAAAGAATGAATGGAGTACCAGTTTCAATCTGAGATTCTATAATCGCTTCCAGACGACTCGAGCCTTATTATAGATTTGTATCTCCTTTCTCTT
CL3       CTTATAAAGAATGAATGGAGTACCAGTTTCAATCTGAGATTCTATAATCGCTTCCAGACGACTCGAGCCTTATTATAGATTTGTATCTCCTTTCTCTT
DPP09     CTTATAAAGAATGAATGGAGTACCAGTTTCAATCTGAGATTCTATAATCGCTTCCAGACGACTCGAGCCTTATTATAGATTTGTATCTCCTTTCTCTT
DPP10     CTTATAAAGAATGAATGGAGTACCAGTTTCAATCTGAGATTCTATAATCGCTTCCAGACGACTCGAGCCTTATTATAGATTTGTATCTCCTTTCTCTT
DPP11     CTTATAAAGAATGAATGGAGTACCAGTTTCAATCTGAGATTCTATAATCGCTTCCAGACGACTCGAGCCTTATTATAGATTTGTATCTCCTTTCTCTT
DPP12     CTTATAAAGAATGAATGGAGTACCAGTTTCAATCTGAGATTCTATAATCGCTTCCAGACGACTCGAGCCTTATTATAGATTTGTATCTCCTTTCTCTT
DPP13     CTTATAAAGAATGAATGGAGTACCAGTTTCAATCTGAGATTCTATAATCGCTTCCAGACGACTCGAGCCTTATTATAGATTTGTATCTCCTTTCTCTT
DPP15     CTTATAAAGAATGAATGGAGTACCAGTTTCAATCTGAGATTCTATAATCGCTTCCAGACGACTCGAGCCTTATTATAGATTTGTATCTCCTTTCTCTT
DPP16     CTTATAAAGAATGAATGGAGTACCAGTTTCAATCTGAGATTCTATAATCGCTTCCAGACGACTCGAGCCTTATTATAGATTTGTATCTCCTTTCTCTT
DPP19     CTTATAAAGAATGAATGGAGTACCAGTTTCAATCTGAGATTCTATAATCGCTTCCAGACGACTCGAGCCTTATTATAGATTTGTATCTCCTTTCTCTT
DPP20     CTTATAAAGAATGAATGGAGTACCAGTTTCAATCTGAGATTCTATAATCGCTTCCAGACGACTCGAGCCTTATTATAGATTTGTATCTCCTTTCTCTT
DPP21     CTTATAAAGAATGAATGGAGTACCAGTTTCAATCTGAGATTCTATAATCGCTTCCAGACGACTCGAGCCTTATTATAGATTTGTATCTCCTTTCTCTT
*****
DPP17     CTTATAAAGAATGAacGGAGTACCAGTTTCAATCTGAGATTCTATAATCGCTTCCAGACGACTCGAGCCTTcA-TATAGtctTGTATCTctTTCTCTT
          o                o o                o o                o

DUKE      TCGTATAGTGATATAAATCGTTTGAACCTCGTCTCCCAAAACATTGTCCAATCCAGGACATTCATCCGGACACATCAACGACCCTCTCCGTCATCCTTCA
3737      TCGTATAGTGATATAAATCGTTTGAACCTCGTCTCCCAAAACATTGTCCAATCCAGGACATTCATCCGGACACATCAACGACCCTCTCCGTCATCCTTCA
Acam_2k   TCGTATAGTGATATAAATCGTTTGAACCTCGTCTCCCAAAACATTGTCCAATCCAGGACATTCATCCGGACACATCAACGACCCTCTCCGTCATCCTTCA
CL3       TCGTATAGTGATATAAATCGTTTGAACCTCGTCTCCCAAAACATTGTCCAATCCAGGACATTCATCCGGACACATCAACGACCCTCTCCGTCATCCTTCA
DPP09     TCGTATAGTGATATAAATCGTTTGAACCTCGTCTCCCAAAACATTGTCCAATCCAGGACATTCATCCGGACACATCAACGACCCTCTCCGTCATCCTTCA
DPP10     TCGTATAGTGATATAAATCGTTTGAACCTCGTCTCCCAAAACATTGTCCAATCCAGGACATTCATCCGGACACATCAACGACCCTCTCCGTCATCCTTCA
DPP11     TCGTATAGTGATATAAATCGTTTGAACCTCGTCTCCCAAAACATTGTCCAATCCAGGACATTCATCCGGACACATCAACGACCCTCTCCGTCATCCTTCA
DPP12     TCGTATAGTGATATAAATCGTTTGAACCTCGTCTCCCAAAACATTGTCCAATCCAGGACATTCATCCGGACACATCAACGACCCTCTCCGTCATCCTTCA
DPP13     TCGTATAGTGATATAAATCGTTTGAACCTCGTCTCCCAAAACATTGTCCAATCCAGGACATTCATCCGGACACATCAACGACCCTCTCCGTCATCCTTCA
DPP16     TCGTATAGTGATATAAATCGTTTGAACCTCGTCTCCCAAAACATTGTCCAATCCAGGACATTCATCCGGACACATCAACGACCCTCTCCGTCATCCTTCA
DPP15     TCGTATAGTGATATAAATCGTTTGAACCTCGTCTCCCAAAACATTGTCCAATCCAGGACATTCATCCGGACACATCAACGACCCTCTCCGTCATCCTTCA
DPP19     TCGTATAGTGATATAAATCGTTTGAACCTCGTCTCCCAAAACATTGTCCAATCCAGGACATTCATCCGGACACATCAACGACCCTCTCCGTCATCCTTCA
DPP20     TCGTATAGTGATATAAATCGTTTGAACCTCGTCTCCCAAAACATTGTCCAATCCAGGACATTCATCCGGACACATCAACGACCCTCTCCGTCATCCTTCA
DPP21     TCGTATAGTGATATAAATCGTTTGAACCTCGTCTCCCAAAACATTGTCCAATCCAGGACATTCATCCGGACACATCAACGACCCTCTCCGTCATCCTTCA
*****
DPP17     TCGTATAGTGATATAAATCGTTTGAACCTCGTCTCCCAAAACATTGtTtAAATCCAGaACATTCATCCaaACACATCAACGACCCTCTCCGTCATCCTTCA
          oo                o                oo

```

Figure 2.5 A unique pattern of mutations in the DPP17 DVX\_084 (I4L) gene  
DPP17 is unique in that it is the only virus that encodes a frameshift mutation in the DVX\_084 gene. The  $\Delta T$  frameshift mutation is associated with many additional point substitution mutations, most of which are unique to this particular DVX\_084  $\Delta T$  allele (marked by circles).

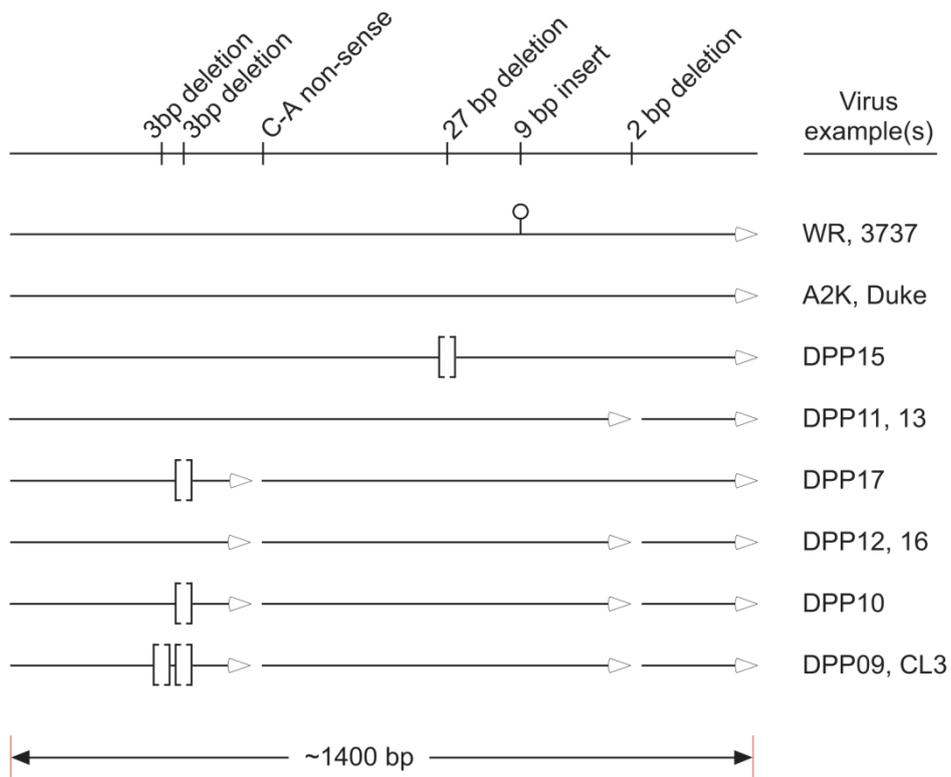


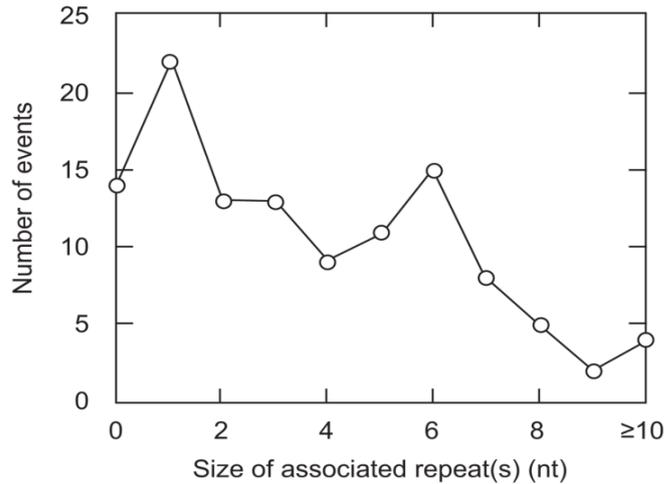
Figure 2.6 Genetic assortment of mutations within ORFs spanning the DVX\_039-041 locus

The three ORFs are derived from a single large ORF that has been designated the M1L gene in VACV strain Copenhagen. This locus seems to have been fragmented by a series of mutations, which also appear to be assorting independently within extant Dryvax stocks. Strains ACAM-2K and Duke carry intact versions of the M1L gene, but it has been broken up into two or three gene fragments in other strains. In-frame indels also modify the lengths of individual ORFs (arrows). Also shown are in-frame deletions ([]) and an in-frame insertion (o).

### 2.3.5 Patterns of mutation

It has been previously noted that many of the indels which differentiate VACV from VARV are associated with small duplications, and this led Coulson and Upton [171] to suggest that poxvirus replication is susceptible to strand slippage errors. Our sequencing of a group of much more closely related and co-cultivated VACV strains provides many further examples in support of this hypothesis. We find that >85% of the indel mutations are associated with the presence of repeated sequences in one or more of the cloned viruses. (Note that one cannot determine with certainty which sequence is ancestral from just these data, although if only one virus encodes a particular sequence it most likely derives from a virus resembling the consensus sequence.) Figure 2.7 illustrates the manner in which duplications of nucleotide sequence are associated with indels. It is important to note that many of the insertions and deletions associated with 1-to-3 nt of identity occur at sites of repeated sequence (e.g. a T from TTT or an AT from ATAT) and thus tend to skew these statistics even when we discount deletions occurring in longer patches of homopolymeric repeated sequence as being a probable artifact of the sequencing technology. A similar pattern of Streisinger frameshifting within short homopolymer repeats has been previously noted in recombinant VACV MVA strains encoding cloned HIV genes [210]. The number of events declines as the length of the duplication increases, possibly due to the greater instability of longer repeats and thus their deletion from the virus pool.

We did look closely at sites where there were no apparent sequence duplication(s) associated with particular indels, i.e. the “0 nt” class of events (Figure 2.7). One unique putative insertion mutation was located adjacent to a classic VACV topoisomerase recognition site [211], which is conserved in all of the other viruses (Figure 2.7, lower panel), but no other distinguishing features were noted regarding these sites.



Indel involving homopolymer repeats

```

|176160
...TATAAATCCGTTAAAATAATTAATAA-TTAATAACAAACAAGTATCAAAAGATTAAAGAC... DPP12 and DPP20
...TATAAATCCGTTAAAATAATTAATAATTTAATAACAAACAAGTATCAAAAGATTAAAGAC... All other clones

```

Indel involving simple repeats:

```

|131380
...TTACTACAGTTCTCCTG---GATGTACGTTACCAACTGCAGACGTAGTTCTAGTACAAT... DPP13 (and others)
...TTACTACAGTTCTCCTGGATGATGTACGTTACCAACTGCAGACGTAGTTCTAGTACAAT... DPP16 (and others)

```

Precise deletion between separated direct repeats:

```

|20040
...AAACTTATATGCTTATTAGCCCAATCTGTAA-----ATATCGTTTCTT... DPP21
...AAACTTATATGCTTATTAGCCCAATCTGTAAATATCGGATTATTAACATATCGTTTCTT... All other clones

```

Imprecise indel involving direct repeats:

```

|154700
...TAATAATTATATAACAACATCTATAAAAGTAGAGGATG-----AGGATACATTAGTATG... DPP19 (and others)
...TAATAATTATATAACAACATCTATAAAAGTAGAGGATGCGGATAAGGATACATTAGTATG... DPP17 (and others)

```

Putative topoisomerase sites:

```

|13360
...ATATCTACGTCCATGTCAATTATTGTCATCATTATCGTCATCCGGCGTATCATCGGGTGTA... CL3
...ATATCTACGTCCATGTCAATTATTGTCATCCTT-----CTTATCATCGGGTGTA... All other clones

```

Figure 2.7 Patterns associated with insertion and deletion mutations in Dryvax clones

We scanned across the aligned genomes, counting the sites of sites where one or more viruses carried an insertion or deletion mutation, and searching for repeated nucleotide sequences in the immediate vicinity of the indel. (Top) The number of times an indel at a particular site was flanked by repeats of a given length on one or more of the other viruses. For example, we detected 15 different indels associated with variably spaced 6-nt duplications. The most

frequent class of events involved the insertion or deletion of a single nucleotide from a short A or T run (1 nt), although a substantial number of mutations also occurred at sites lacking any apparent similarity between the sequences flanking the indel (0 nt). Collectively, over 85% of indels are associated with repeats of varying lengths. (Bottom) The different kinds of rearrangements that are seen at indel sites, for example, a 1-nt deletion in an A (or T) run in DPP12 and DPP20, and a 16-nt deletion associated with 6-nt repeats in DPP21. CL3 encodes a 12-nt insertion, and two point mutations, next to an otherwise highly conserved poxvirus topoisomerase recognition site (5'-YCCTT-3'), but no other special sequence features are seen at any of the other indels associated with the 0-nt class of events.

### 2.3.6 Deletions in the right telomere junction

It has previously been noted that the junction regions, where the boundaries of the terminal inverted repeats (TIRs) are located, are prone to high frequencies of mutation and rearrangement [165, 191]. This is true of these different clones, where several large rearrangements create significant alterations in the gene structure surrounding the junction with the right-hand TIR (Figure 2.8A).

Besides discovering additional examples of viruses bearing telomeres resembling Duke (DPP13 and DPP21) and ACAM-2K/3737 (DPP09-12, 15, 16, 19, and 20), we also detected an additional 11.7 kb deletion in isolate DPP17. This deletion completely excised the DVX\_212 (B19R) gene, an interferon binding protein [133, 212] that is partially deleted in ACAM-2K [165], as well as the genes of unknown function encoded by DVX\_210 and 211. We did not initially sequence an example of a virus bearing the deletion characteristic of strain CL3, although it is present amongst the clones and in the vaccine stock (clone DPP25, see below). Such significant alterations in the gene complement of these strains might be expected to alter the abundance of the strain types in the original virus stock. To test this, we designed sets of PCR primers that can differentiate the four major “deletion types” of these viruses (Table 2.2). PCR analysis showed that of the 25 cloned viruses, thirteen encoded the deletion characteristic of ACAM-2K, nine resembled Duke, one looked like CL3, and two encoded the largest deletion characteristic of DPP17 (Figure 2.9, panels A and B). Because picking plaques biases the recovery of viruses, we also used qPCR (Figure 2.9 D) to measure the abundance of the different deletion alleles in a pool of virus DNA extracted directly from the stock of vaccine. These studies confirmed that (as defined by these large deletions) the ACAM-2K-like viruses were the dominant form (~60%) followed by Duke-like viruses (~40%) and <1% of viruses resembled CL3 and DPP17 in these stocks.

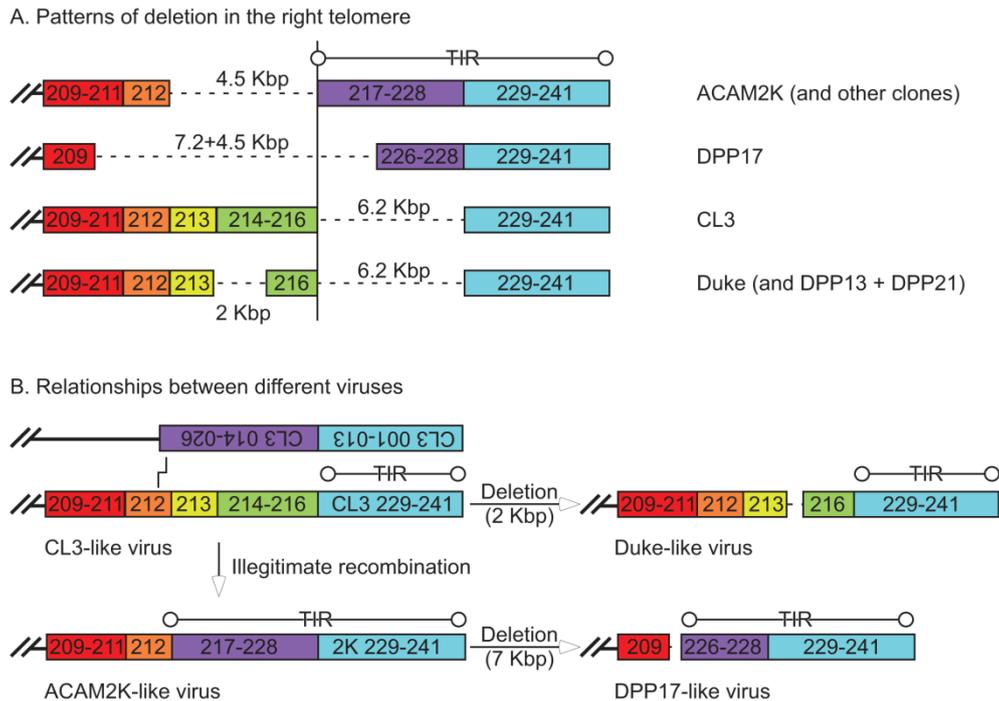


Figure 2.8 Patterns of gene deletion and rearrangement in the virus right telomeres

(A) Structures of the right-hand telomeric region with genes colour coded to facilitate the identification of homologous sequences. The sequences that comprise the terminal inverted repeats of each virus are also noted (TIR). The differences between CL3 and Duke have been described in detail elsewhere [165, 191]. CL3-like viruses encode a 6.2 kb deletion from DVX\_217 to DVX\_228, compared to 2k-like viruses. However these genes are still encoded by the CL3 left terminus. DPP17 is a new kind of Dryvax clone that includes the largest right-hand deletion (dashed lines). The deletion eliminates the DVX\_209 stop codon and extends the gene by 21 bp. (B) A simple scheme for producing these viruses. An ancestral virus resembling CL3 could, through a simple pattern of deletion, produce a virus resembling the Duke-like viruses. Alternatively, an illegitimate recombination event that transposed a copy of the left TIR onto the right side of a CL3-like ancestor, would create a virus resembling ACAM-2K. Another deletion event would produce a DPP17-like virus.

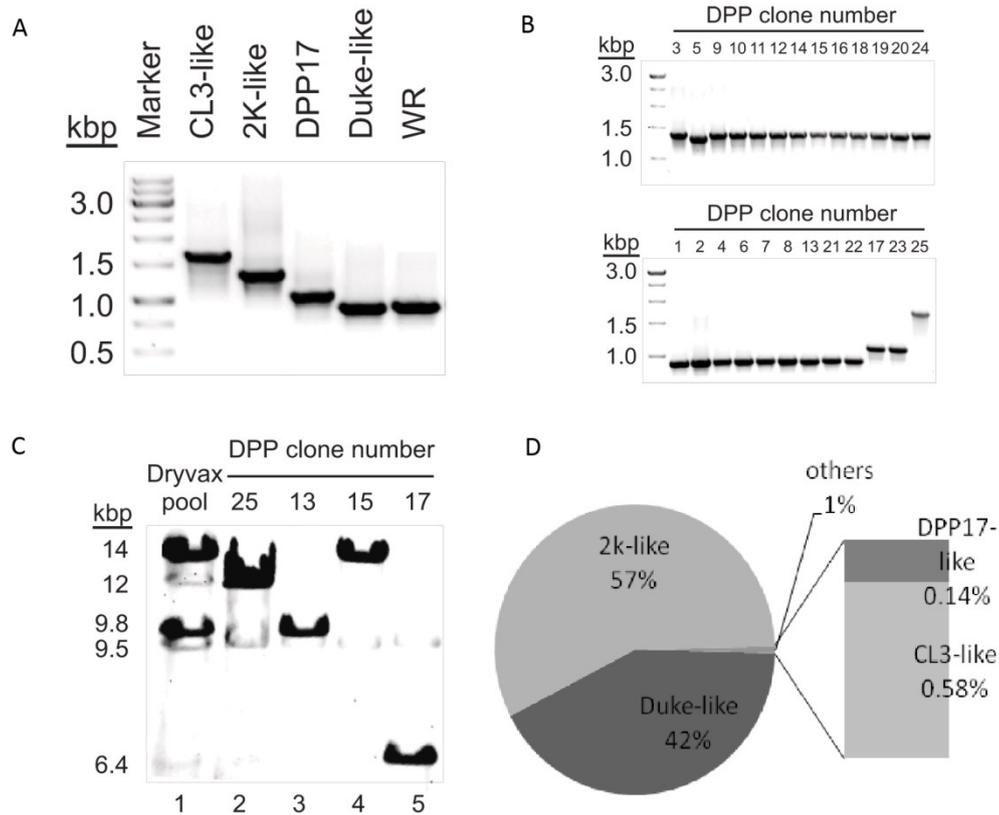


Figure 2.9 Distribution of virus deletion patterns in Dryvax stocks

(A) Control PCR products used to differentiate the four major variant viruses. The PCR amplicons span the sites of deletion in the right TIR. This method was then used to genotype the 25 different cloned viruses shown in Panel B. Most viruses bear right telomere structures resembling ACAM-2K- or Duke-like strains. Note that DPP05 may encode another small deletion that has not been further characterized. (C) DNA was extracted either from the Dryvax pool (lane 1) or from viruses representing each of the four deletion patterns (lanes 2 to 5), digested with *SalI*, and Southern blotted using a probe targeting the right-telomere region. The 14 and 9.8 kb bands that predominate in the Dryvax pool are characteristic of ACAM-2K- and Duke-like viruses (DPP15 and DPP13, respectively). One sees much weaker signals from the 12 and 6.4 kb fragments that characterize CL3- and DPP17-like viruses (DPP25 and DPP17, respectively). Note that this probe also detects a 9.5-kb fragment that is common to all of the viruses. It includes ORFs DVX\_006/226 and derives from the left TIR. (D) DNA from Dryvax stock was extracted and analyzed for clone content by quantitative PCR using primers specific to DPP13, 15, 17 and 25. To quantify the amounts of these four groups in the Dryvax pool, DNA from clones of DPP13, DPP15, DPP17 and DPP25 were used as standards representing four different groups (Duke-, 2k-, DPP17- and CL3-like, respectively).

Southern blotting of DNA extracted from the Dryvax pool produced a pattern of hybridization signals with intensities that were also consistent with these measurements (Figure 2.9, panel C). An intriguing aspect of the discovery of the different viruses is that they may reflect the past evolutionary history of the telomere junctions. The seemingly complicated structures can be explained if it is assumed that a virus originally resembling CL3 was subjected to a rearrangement that transposed a copy of the left telomere sequence into the right telomere.

This would create a family of viruses resembling ACAM-2K. A simple process of additional deletions could then produce the Duke- and DPP17-like viruses from the CL3- and ACAM-2K-like viruses, respectively (Figure 2.8 B). The relative rarity of viruses resembling CL3 and DPP17 suggests that some of the genes in this interval may have adaptive value in competitive growth environments, but it is not possible to say with certainty which ones.

### **2.3.7 Substitution mutations**

Figure 2.10 shows a plot of the distribution of sequence differences across the different genomes. Inspection of this plot suggests that the density of sequence differences varies unevenly across the different genomes, with fewer polymorphic sites in the region between nucleotides ~40,000 and ~150,000. More quantitatively, we detected a density of about 0.8 single-nucleotide polymorphisms (SNPs) per 100 bp between nucleotides 40-150,000 and 1.3 SNPs per 100 bp in the 30 kb segments on either side of this central region. Most of the SNPs (74%) comprise transition substitutions, i.e. pyrimidine for pyrimidine and purine for purine. This interval is bounded by the genes DVX\_053/F4L and DVX\_170/A36R and encompasses the F9R-A32L region that has been previously identified as encoding the highly conserved core of poxvirus genes [38].

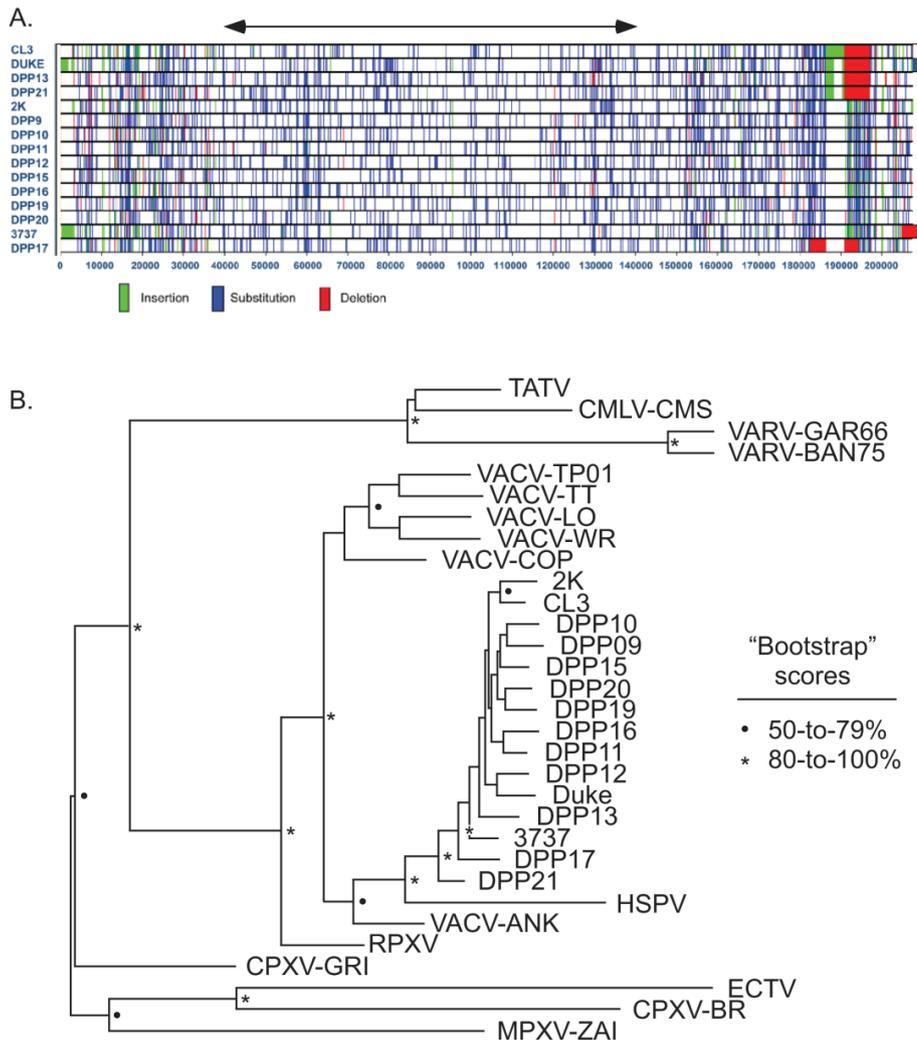


Figure 2.10 Distribution of polymorphic sites in Dryvax-derived VACV genomes

(A) Sites of sequence variation were identified as being substitutions, insertions, and/or deletions by reference to a consensus sequence. The 15 different genomes were aligned by LAGAN and edited using Base-By-Base software. A lower density of polymorphic sites is clearly seen in the central portion of the viruses. Note that because different approaches were used to define where the sequence begins and ends in the different genomes, the very ends to these viruses are not properly illustrated by this methodology. (B) Phylogenetic relationships between Dryvax clones and other Orthopoxviruses. A multiple alignment was compiled using sequences encoding the conserved core genes DVX058 (F9L) to DVX155 (A24R) and marked off with an arrow in panel A. The alignment, a maximum likelihood approach, and a thousand bootstrap replicates were used by the program RDP [200] to create the plot. All of the Dryvax-derived clones cluster as one group with good bootstrap values, but the relationship between viruses within this cluster is less certain.

The lower rate of accumulation of mutations would be consistent with the large number of essential and highly conserved proteins encoded within this region.

The SNPs located between nucleotides 42240 and 141145 (aligned position) were used to examine the relationship between the different cloned isolates. Figure 2.10 shows a phylogenetic tree that clearly demonstrates a clustering of the viruses that share a common historical origin as the Dryvax vaccine. A notable feature of this tree is that most of the viruses that we have isolated, cluster as one group and separately from the ACAM-2K and CL3 strains that were independently isolated from another stock of the same vaccine. However, the association is not absolute with DPP21 and DPP17 falling elsewhere within the “Dryvax cluster”. It is also curious that horsepox virus fall into this grouping, providing some support for the hypothesis that HSPV derives from a feral vaccine strain [166] in much the same manner as VACV appears to have established new zoonotic infections in Asian water buffalo and South American cattle [213, 214]. It has previously been noted that there is more sequence diversity amongst all the sequenced VACV strains (including Tian Tan and Copenhagen) than amongst extant Variola virus strains [165]. Judging by the branch lengths, our phylogenetic analysis shows that the VACV that can be isolated from a single vial of Dryvax stock, also exhibits more sequence diversity than is seen in strains of Variola major and Variola minor viruses.

### **2.3.8 Recombination**

Within the cluster of Dryvax clones, the branching is not securely supported by the bootstrap values. This most likely reflects the fact that these viruses are likely genomic mosaics, generated by multiple recombination events at some point in their history and this obscures any clear relationship between the different isolates [171]. This hypothesis is supported by an

analysis of how the patterns of SNPs and indels are shared between the different genomes. For example, the widely used program Bootscan [202, 215] calculates how often a bootstrapping algorithm assigns two viruses to a common branch of several possible trees, and how this relationship changes as one scans across a window encompassing different polymorphic sites. Bootscan discovered evidence that each virus shares many short patches of sequences that closely resemble portions of other “sister” viruses. For example, within an interval towards the left end of the virus containing a relatively high density of informative polymorphic sites, DPP13 encodes blocks of sequence that resemble portions of homologous loci in DPP11, 16, 20, and 21 (Figure 2.11).

Different patterns are detected in different viruses although the reciprocal signals are readily detected (i.e. DPP15 encodes a patch of sequence resembling DPP17 and *vice versa*, Figure 2.11, panel B and C).

How much recombination these viruses have been subjected to is more difficult to determine. Previous studies from our laboratory have shown that replicating poxviruses can very efficiently recombine DNA under certain special circumstances, but there are also some physical constraints operating within a cell that, in combination with issues relating to the multiplicity of infection, limit recombination between co-infecting viruses [216]. This dichotomous situation is also seen in the Dryvax stocks. On the one hand, many mutations show a complete loss of linkage, as is illustrated by the A12L and M1L genes (Figure 2.4 and Figure 2.6, respectively), and appear to be assorting randomly amongst the genomes. On the other hand, there has not been enough recombination to fully obscure the patchwork patterns of closely linked polymorphic sites that are detected by a recombination detection algorithm (Figure 2.11).

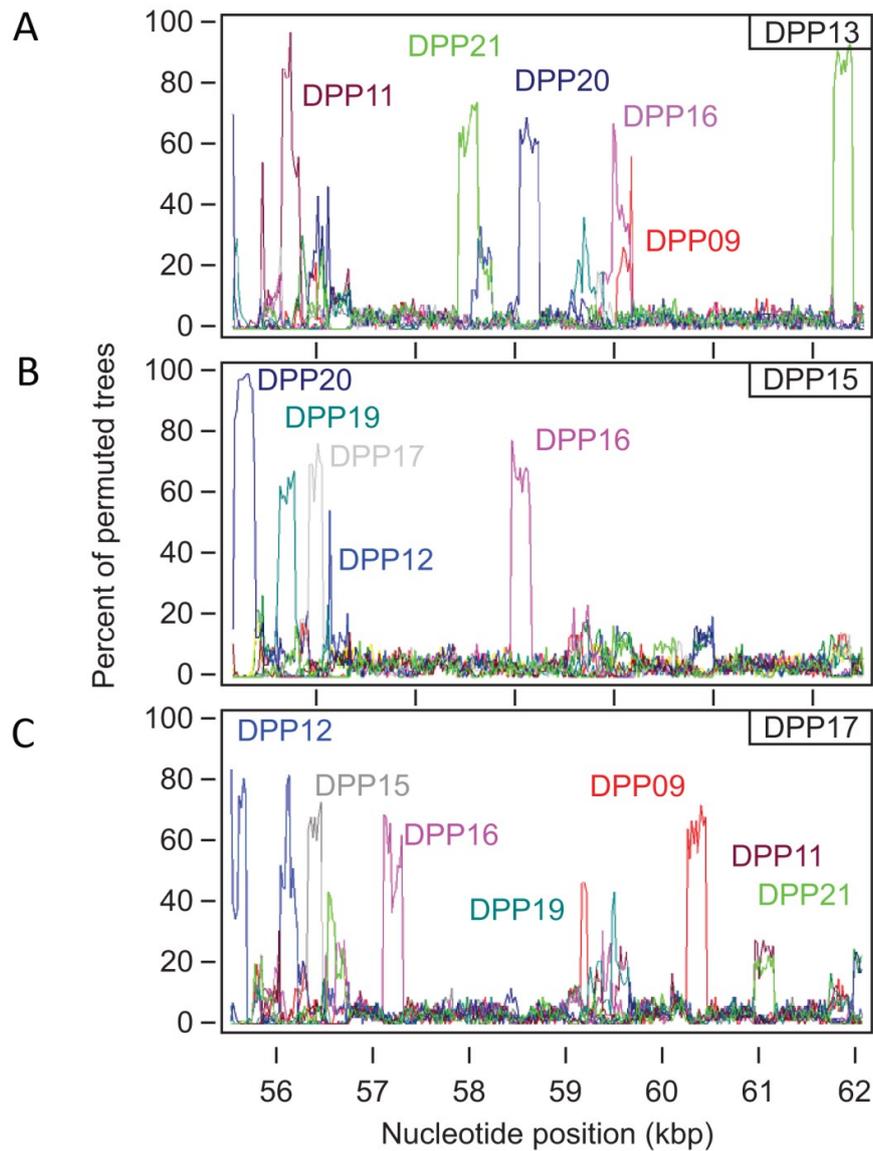


Figure 2.11 Distribution of putative recombination sites

The program Simplot/Bootscan [202] was used to search for varying phylogenetic relationships across a 7-kb region of virus encoding DVX\_075 (E8R) to DVX\_080 (O2L). The search used a 200-nt window of sequence, which was moved in 20-nt steps, and compared each of the indicated viruses (DPP13, 15, and 17) with all of the other sequenced clones. The scan encompassed nucleotides 55,534-62,079 (DPP\_09) and the colored peaks indicate sites where a disproportionate fraction of the 100 calculated neighbor-joining trees suggested a close similarity between a particular virus pair in that region.

The simplest explanation for this situation is that these stocks are composed of a mixture of recombinant viruses, but there has not been sufficient recombination to completely obscure the presence of a few still linked markers.

There is one fascinating illustration of this patchy pattern of recombination. Most of the mutations in DPP17 DVX\_084 gene are tightly linked and seemingly unique to that virus (Figure 2.5). We have found no other Dryvax strains encoding a similar pattern of markers. However, what is exceedingly curious about the DPP17 DVX\_084  $\Delta$ T frameshift mutation is that it lies in the centre of a 200 nt patch of DPP17 DNA, that aside from a new 2 bp deletion, closely resembles the homologous locus in the horsepox R1 gene. The remainder of the DPP17 DVX\_084 gene is clearly more closely related to viruses like Acambis-2K as judged by the surrounding pattern of polymorphic sites (Figure 2.12). This looks like a “molecular fossil”, and provides further evidence of a shared origin (or at least some co-cultivation) of horsepox and vaccinia viruses.

## **2.4 CONCLUSIONS**

These studies provide new insights into the population structure and evolutionary trajectories of classical smallpox vaccines. Within a single vial of Dryvax, we have identified at least four different variants of VACV as defined by the pattern of large deletions in the right hand telomere. By this crude definition, viruses resembling ACAM-2K comprise about 60% of the viruses in this population, and this observation provides additional retrospective support for the wisdom of selecting this isolate to serve as a clonal representative of the VACV in Dryvax. These stocks also contain a substantial fraction of viruses (~40%) bearing the right telomeric deletion characteristic of strain Duke as well as sufficient other genetic commonalities to cluster Duke in a phylogenetic tree with DPP12 (Figure 2.10).

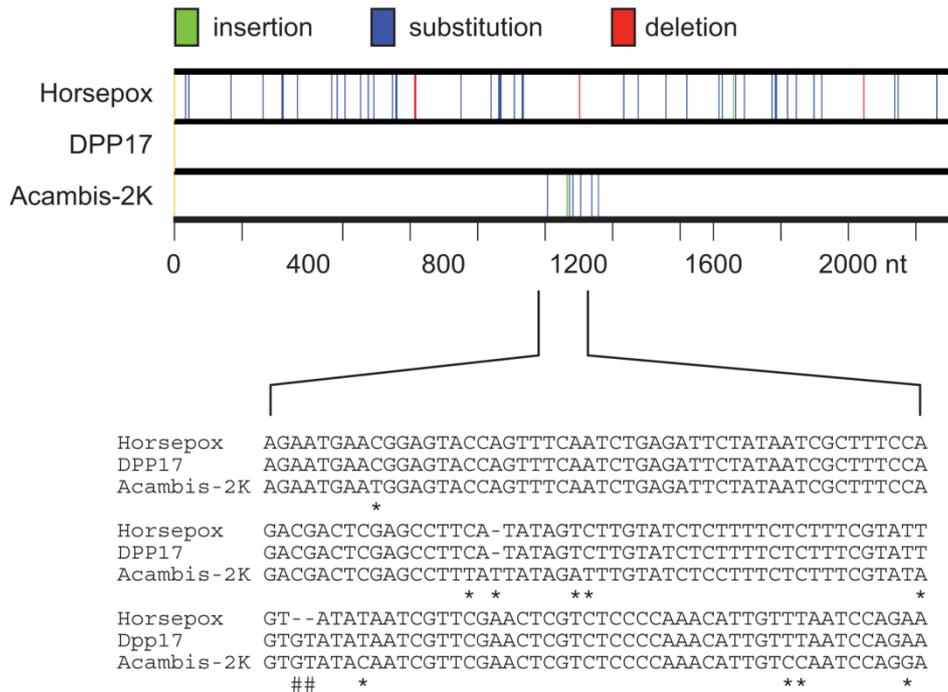


Figure 2.12 Pattern of polymorphic sites in the ribonucleotide reductase large subunit gene

The I4L genes of the three indicated viruses were aligned and the upper figure prepared using Base-by-Base software. Where the sequences differ between DPP17 and Horsepox virus, and between DPP17 and ACAM-2K, are shown at the top and bottom, respectively. The sequence differences are shown aligned and expanded below (nucleotides 1100 to 1249). Note that in this region, nearly all of the sequence differences are between DPP17 and ACAM-2K (\*), and that DPP17 is nearly identical to horsepox. The exception is a 2-nt deletion found only in horsepox virus (#).

This suggests that the Duke strain was not a novel form of spontaneously arising virulent VACV, but rather represents a pre-existing strain type that was subjected to clonal selection in a patient susceptible to *vaccinia necrosum* [191]. Rather surprisingly we could detect only small numbers of viruses resembling the virulent CL3 strain ( $\leq 4\%$ ), even though it represented one of the six viruses originally characterized during the development of ACAM-2K [165]. Whether this reflects some differences between different lots of virus, experimental protocols, or simply random chance is difficult to say. We have also identified a new variant strain of VACV, DPP17, which was probably produced by the deletion of 7 kb from the right telomere of a virus resembling ACAM-2K (Figure 2.8). Given this lineage, and the evidence suggesting that ACAM-2K represents a less virulent form of VACV (compared with CL3), it would be interesting to test the safety of a DPP17-based vaccine. It is possible we have missed viruses bearing other large deletions, because the Poisson 95% confidence interval is 0.0-to-3.7 with zero observed events and thus a screen of 25 plaques could have easily missed any viruses comprising less than  $3.7/25$  (i.e. 15%) of the population. However, the true abundance would have to be far less than 15% because no other variants were detected by Southern blotting. We found no viruses encoding previously unknown genes, although again the small sample size makes it impossible to conclude there are no other genes remaining to be discovered in rare DVX clones.

Finally, these studies also provide some interesting insights into the behaviour of VACV in the face of the selective forces imposed by classical calf-lymph culture methods. It is not surprising that the central core of conserved genes appears to resist mutation, but selection also appears to favor the retention of at least some genes in the classically “unstable” telomeres. As a number of authors have previously noted, the viruses in vaccine stocks differ greatly due to an abundance of single-nucleotide and other polymorphisms [165, 192]. However, our sequencing data also highlight the natural

instabilities associated with sites bearing small duplications. We have previously noted that only limited sequence identity is needed to support recombinational repair in VACV-infected cells [217], and this pattern of repeat instability is perhaps reflective of this process. Recombination also appears to be rearranging the different genomes, but not to such an extent as to completely unlink all of the mutations. These viruses have clearly been subjected to a long history of mutational drift, and periodic rearrangement by recombination, in the absence of severe selection pressure. This has created a much greater degree of genetic diversity in a VACV vaccine than is seen in viruses like Variola, which would have been subjected to very different evolutionary pressures, especially bottlenecks, during the natural passage of smallpox from person-to-person.

**CHAPTER THREE - GENOMIC ANALYSIS OF VACCINIA VIRUS  
STRAIN TIAN TAN PROVIDES NEW INSIGHTS INTO THE  
EVOLUTION AND EVOLUTIONARY RELATIONSHIPS  
BETWEEN ORTHOPOXVIRUSES**

Li Qin, Min Liang<sup>1</sup> and David H. Evans<sup>2</sup>

### **3.1 INTRODUCTION**

The Chinese Center for Disease Prevention was created in 1919 with a mandate to produce a smallpox vaccine using a seed stock of Japanese origin. In 1924 an employee, Mr. Qi Changqing, was sent to Japan to study the technique for vaccine production. After his return home in 1926, Mr. Qi is reported to have isolated a sample of virus from a smallpox patient and then passaged it on the skin of monkeys, rabbits and cows. His virus, which acquired the name “TianTan” after Beijing’s Temple of Heaven where the specimen was collected, became the smallpox vaccine of choice for most of China’s history [218, 219]. Its use spanned a period of manufacture in rabbits and cows from 1926 to 1954, a brief production hiatus while a Russian strain was used from 1955 to 1960, and then a return to production from 1960 to 1980, when smallpox was declared eradicated. Through the last part of this period (1965-1980) the vaccine was produced in chicken eggs [220]. Although no longer used much as a smallpox vaccine, TianTan virus is still widely used by Chinese researchers as a vaccine vector.

Mr. Qi is rightly remembered for his discovery of the virus that played such an important role in helping to eradicate smallpox in China. This is a history that spans many turbulent decades, and includes the remarkable story

---

<sup>1</sup> Dr. Liang is a scientist in TOTBIOPHARM Shanghai R&D Center, Shang, China. He provides us the TianTan virus and contributes to the history of TianTan.

<sup>2</sup> A version of this chapter has been published in *Virology* (2013), 442: 59-66.

of how Mr. Li Yanmao later preserved the TianTan strain at great personal risk during the Japanese attack on Beijing in 1937. However, this history lacks scientific credibility, as variola virus (the causative agent of smallpox) does not exhibit a host range encompassing monkeys, rabbits, or cows [122]. Based upon the DNA sequencing done to date, TianTan is clearly a vaccinia virus (VACV) and phylogenetically distinct from other Orthopoxviruses like variola, horsepox, and monkeypox. Most probably, the sample was either isolated from an individual exhibiting one of the rarer complications of smallpox vaccination [221], or contaminated with another strain being propagated at the time in the Center for Disease Prevention. Considering the difficulties associated with virus culture in the 1920s, and the fact that the origins of even modern viruses continues to generate controversy [222], the later explanation seems more probable. Dairen (a name of Chinese origin) and Ikeda (a Japanese name) are some of the oldest VACV strains once used in Japan (Dr. I.Arita, personnel communication), and whether they are related to TianTan would be interesting to determine.

Given the critical role that TianTan virus has served in eradicating smallpox in China, it is unfortunate that the only available complete genome sequence [167] comes from an era when DNA sequencing was far more difficult, and thus it contains a number of sequencing errors. These include at least nine frameshift mutations in what are generally believed to be essential genes [38]. Because of the errors, these sequence data are often avoided in comparisons with other Orthopoxviruses. To address this concern, we have taken advantage of next generation sequencing technologies to produce more accurate genome sequences for two clonal Wuhan laboratory strains of TianTan virus, which come originally from the National Vaccine and Serum Institute. Besides providing an accurate sequence for one of the most historically important smallpox vaccines, we also describe a curious pattern of deletions and rearrangements, which may account for some of the

unique properties as a vaccine. This pattern of genome rearrangements also provides further insights into the evolution of Orthopoxviruses.

## **3.2 METHODS**

### **3.2.1 Virus and cell culture**

Vaccinia virus (strain TianTan) was obtained from the China Center for Type Culture Collection (Wuhan, Hubei). The passage history is obscure, although publications and dissertations indicate that it comes originally from the Institute of Virology, Chinese Academy of Preventive Medicine, and before that the Chinese National Vaccine and Serum Institute. These records suggest it would have been previously cultured on chick embryo chorioallantoic membranes and primary chick embryo fibroblasts; possibly also Vero cells. This stock has also been plaque purified at some point in its recent history, but when is also unknown. Monkey kidney BSC-40 cells were obtained from the American Type Culture Collection (Manassas, VA). The cells and virus were cultured in modified Eagle's medium (MEM, Gibco) supplemented with 5% fetal bovine serum, 1% nonessential amino acids, 1% L-glutamine, and 1% antibiotic at 37°C in a 5% CO<sub>2</sub> atmosphere. The cells were tested periodically and shown to be free of mycoplasma.

### **3.2.2 Virus sequencing and genomic analysis**

The viruses were plaque purified, cultured, purified using sucrose gradients, and sequenced as described previously using a Roche 454 GS FLX Titanium sequencer platform and multiplex technology [223]. Twelve different clones were sequenced, to read redundancies ranging from 9- to 42-fold, but discovered to be nearly all identical apart from sequence features described below. Pyro-sequencing methods produce indel-type sequencing errors within homopolymeric base runs. The true sequence can usually be deduced from inspection of the sequence of the aligned high

quality replicate reads, but where that was not possible we assumed that the correct read was the one that maintained the open-reading frame in alignment with many other sequenced VACV strains. (In our previous study, 6/6 of these problematic sites were Sanger sequenced and all of them were found to be pyro-sequencing artifacts.) The viruses discussed in detail in this communication were designated TP03 and TP05 (i.e. TianTan plaque 3 or 5) and have been assigned GenBank accession numbers KC207810 and KC207811, respectively. For purposes of gene comparison we reference the gene numbering system initially established for Acambis clone 2000 (AY313847) and Acambis clone 3 (AY313848) viruses and subsequently used to label other Dryvax clones [223]. We did not sequence far into the repeated elements located in the virus telomeres, the first nucleotide in each of the assemblies was defined as the first nucleotide in one of the four 54 bp repeats preceding the first open reading frame.

The other VACV and accession numbers used in this study are: Ankara strain CVA (AM501482), Copenhagen (M35027), DPP9 (JN654976), DPP13 (JN654980), DPP15 (JN654981), DPP17 (JN654983), DPP21 (JN654986), Duke (DQ439815), Lister (AY678276), Western Reserve (NC\_006998), and 3737 (DQ377945). Other Orthopoxviruses included: camelpox strain CMS (AY009089), cowpox strains GRI-90 (X94355) and Brighton Red (NC\_003663), ectromelia strain Moscow (NC\_004105), horsepox (DQ792504), monkeypox strain Zaire (AF380138), rabbitpox (AY484669), taterapox (NC\_008291), and variola strains Bangladesh (L22579) and Garcia-66 (Y16780). For simplicity, we refer to the original TianTan virus sequence (AF095689) as “TT00” throughout this communication. Genome assemblies were prepared using CLC Genomics Workbench (v4.6) and annotated using GATU [203], also as described previously [223]. Bioinformatic analyses were performed using Viral Genome Organizer [195, 196] and Poxvirus Orthologous Clusters [197].

These and additional bioinformatics tools can be accessed at <http://www.virology.ca>. Phylogenies were assembled using RDP3 [200].

### **3.3.3 Other methods**

To determine the lengths of the repeat elements in a 1.2 kbp region corresponding to the genes previously designated DVX-004 to DVX-006, we used the PCR and “repeat forward” (5’ GCAGTAGGCTAGTATCTT 3’) plus “repeat reverse” (5’ TACCGGCATCATAAACAC 3’) primers. The products were then sequenced using dideoxy-sequencing methods. To confirm the size and presence of the larger deletions we used the PCR and either of two forward primers (TT010F: 5’ TTTTGTAGGAAGGAGGC 3’ and TT004F: 5’ TGGATGGCCGTATTGATT 3’) in combination with a single reverse primer (TT011R: 5’ CCGGGAGATGGGATATATGA 3’). For Southern blots virus DNAs were digested with *ScaI*, fractionated on a 0.7% agarose gel, transferred to nylon membranes, and blotted with a biotin-labeled probe prepared using primers TT010F and TT011R (above) and TP05 virus DNA as a template.

## **3.3 RESULTS AND DISCUSSION**

### **3.3.1 Virus isolation and genome sequencing and assembly**

We started by randomly selecting 24 different viruses from a stock of VACV strain TianTan and plaque purifying them on BSC-40 cells. After being cloned and plated, these viruses produced two kinds of plaques. Viruses of the more abundant TP03 type (21/24 clones) formed smaller plaques with an area about half that of viruses of the TP05 type (3/24 clones). Figure 3.1 illustrates this property of some representative viruses. A single step growth curve, prepared using BSC-40 cells infected with virus at a low multiplicity of infection (MOI=0.01), showed that the small plaque variety (TP03) also grew slightly slower and to lower titers than did the large plaque

variety (TP05) (Figure 3.2). This suggested that our stock contained at least two variant forms of viruses.

To explore this question we sequenced 12 of the viruses (clones designated TP01-to-TP06 and TP08-to-TP13) and assembled the genomes using standard methods [223]. Interestingly these TianTan clones all seemed to be essentially identical except for symmetrical 5.7 kbp deletions located in the telomeric inverted repeats (TIRs) in viruses of the small plaque TP03 type. To confirm that this was not an artifact of the assembly algorithm, we used the PCR and Southern blotting methods to examine the DNA structure in this region of the genome (Figure 3.3). These methods provided independent evidence of 5.7 kb deletions in clones TP03, TP11, and TP12 and the presence of the correspondingly larger fragment in clones TP05, TP13, and TP18. The small plaque phenotype correlated perfectly with the presence of the deletions (Figure 3.1). As would have been expected, we also noted that the unpurified stock of virus contained viruses of both types (Figure 3.3).

These mutations delete or inactivate several genes of mostly still unknown function from the small plaque variants (Table 3.1). The deletion of the homolog of the Copenhagen strain C12L gene from clone TP03 (TT\_009 in clone TP05) excises the SPI-1 serpin gene, but that is not expected to alter the growth of the virus in culture as suggested by studies using SPI-1 mutants of VACV strain WR [162]. However, it is striking that these mutations would delete the promoters and N-termini of both copies of the epidermal growth factor gene (VEGF) homolog from TP03 (TT\_011 in clone TP05). Depending upon the cell type and the growth state of the host cells, deleting the VEGF can affect the growth of VACV in culture [224], and it is likely the major reason why the TP03-like clones produce smaller plaques than the TP05-like clones.

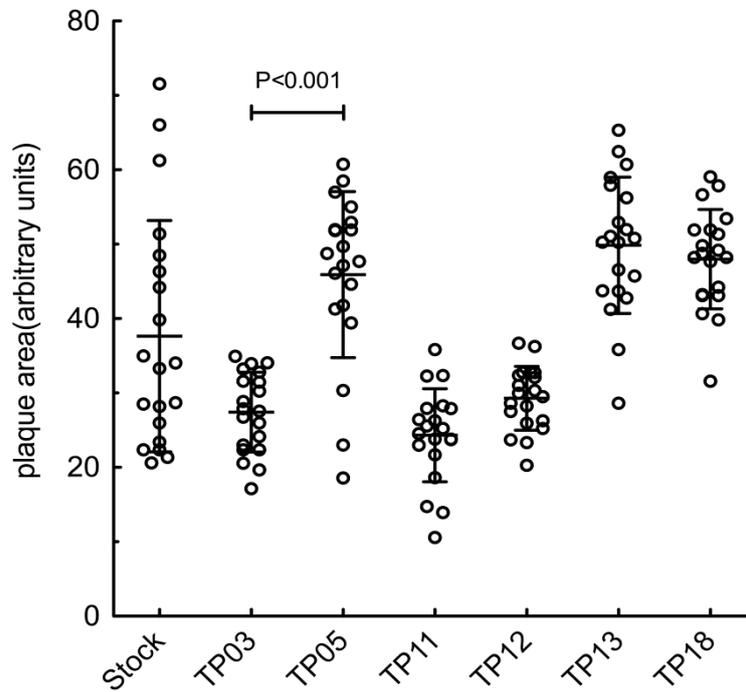


Figure 3.1 Plaques formed by viruses cloned from a stock of TianTan virus. Each of the plaque-purified viruses were separately plated on BSC-40 cells, cultured for two days under liquid overlay, fixed, stained with crystal violet, and photographed. ImageJ [225] was then used to measure the sizes of 20 plaques per virus, randomly selected from each dish. Viruses of the TP05 type form plaques that are significantly ( $P < 0.001$ ) larger than viruses of the TP03 type and the original stock appears to contain a mix of both large plaque and small plaque viruses. Neither of these two strains of TianTan virus produces high level of extracellular enveloped virus, as judged by the lack of “comets” that arise through secondary infections in liquid overlay.

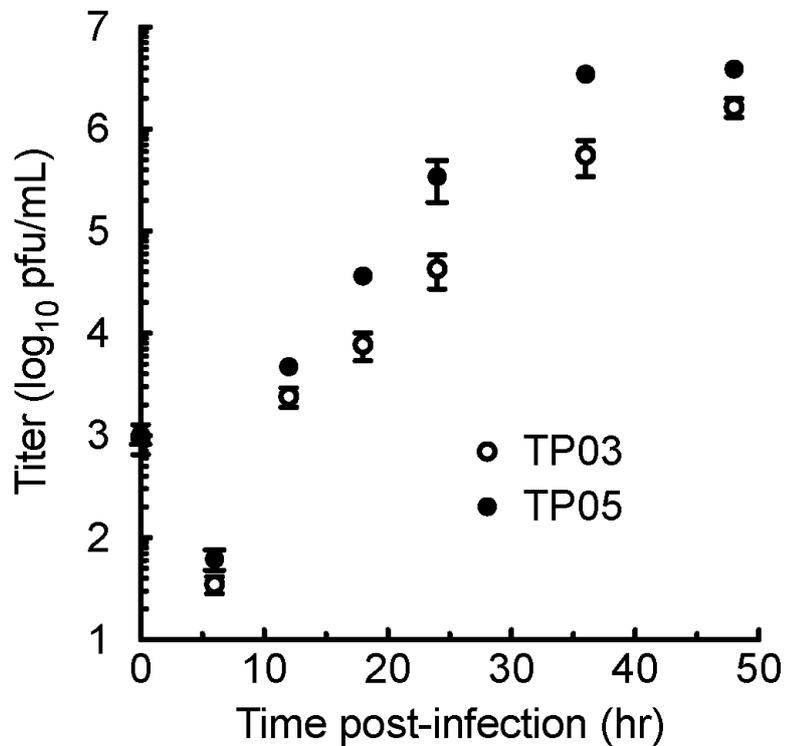


Figure 3.2 Small plaque TianTan strains grow to lower titers in culture  
 BSC-40 cells were infected with the indicated viruses at MOI=0.01, incubated at 37° C, harvested by freeze-thaw, and the yield of virus determined by titration on BSC-40 cells. The TP03 viruses grow somewhat more slowly, and to lower titers, than the larger plaque forming TP05 strain. The difference between the two data sets is statistically significant. Beyond the 10h time point, TP05 produces ~5-fold more virus than TP03 (P=0.007, 95% CI=1.9-8.6 fold, 2-tailed ratio *t* test).

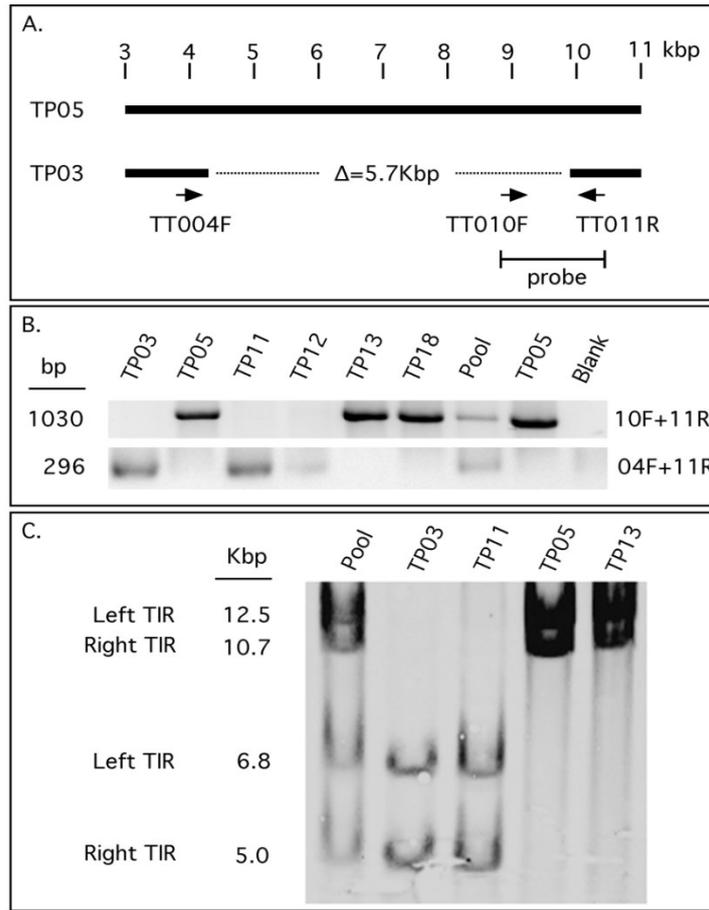


Figure 3.3 PCR and Southern blotting methods confirm the presence of telomeric deletions

Panel A shows the PCR primer binding sites relative a map of the TP03 and TP05 left telomeres. The TT004F and TT011R primer pair are located sufficiently far apart (on the TP05 genome) that they can serve as PCR primers only if the distance between the primers is reduced by the deletion in TP03. The deletion eliminates the TT010F binding site, thus preventing any DNA amplification using primers TT010F and TT011R. Panel B shows that clones TP03, TP11, and TP12 encode the 5.7 kb deletion. Panel C shows a Southern blot of *ScaI* digested VACV DNAs. The probe encodes DNAs spanning the right side of the deletion boundary (Panel A). The deletion of 5.7 kb greatly shortens the *ScaI* fragments encoding the left and right TIRs in clones TP03 and TP11.

Table 3.1 Genes located in a 5.7 kbp deletion in TianTan clone TP03 terminal inverted repeats

Gene function or feature	Copenhagen	TianTan TP05		TP03
	Gene number	Gene number	Size (bp)	Size (bp)
Ankyrin	C17L/B23R	TT_005	528	0
Unknown	C16L/B22R	TT_006	546	0
Unknown	C15L/B21R	TT_007	264	0
Unknown	-	TT_008	156	0
Unknown	C14L	TT_009	573	0
Serpin (SPI-1)	C12L	TT_010	1062	0
Epidermal growth factor	C11R	TT_011	426	253 <sup>1</sup>

<sup>1</sup>C-terminal fragment

Interestingly several different variants of a VACV strain called “vvDD”, containing a double deletion of the thymidine kinase and VEGF genes, have been tested for more cancer-specific oncolytic activity [226]. The TP03-like TianTan viruses seem to have naturally anticipated an element of this strategy (i.e. deletion of VEGF gene).

### **3.3.2 The genome of TianTan clone TP05**

The TP03-like variants are probably an artifact of passage in culture, which likely occurred sometime after the Wuhan strains were plaque purified in the 1990’s. We therefore focused our attention on characterizing the larger TP05-like isolates, which aside from the 5.7 kbp insertions relative to the TP03-like viruses, are nearly identical. (Other differences between TP03 and TP05 include four single- nucleotide substitutions and three small insertion/deletions or indels.) In our final assembly TP05 spans 196,260 bp and encodes 219 open reading frames, including the genes duplicated in the TIRs. The TIRs comprise approximately 12 kbp each and duplicate 12 genes.

The revised sequence of TianTan clone TP05 validates and extends some of the concerns expressed previously [38], in that a number of genes that are typically found intact in other VACV strains contain frame shift mutations in the original TianTan genome sequence (referred to as “TP00” herein). However, not all of the sequence differences that we have identified are necessarily a result of sequencing errors in TT00. We noted several small indels, where the longer sequence encoded two flanking and duplicated copies of a sequence found only once in the shorter sequence (Figure 3.4A). These are typically unstable sites in poxviruses [171, 223] and likely reflect real differences between different clonal isolates. Where these mutations create in-frame deletions, they characteristically alter the lengths of the encoded proteins by only a few amino acids.

A. Indels linked to duplicated sequences

```
>TP05_041
TP05:   TATGTAATAGGTTTCCAATATTTACAATATATGTAATCAT
TP00:   TATGTAATAGGTTTC_____CAATATATGTAATCAT

> TP05_151
TP05:   TCGATGTAC_____TTGGCATCGAAACACTTATTAA
TP00:   TCGATGTACTTGGCATGATTGGCATCGAAACACTTATTAA
```

B. Gene disrupting 59 bp repeats

```
69 bp   AACTTTTTTACGACTCCATCAGAAAGAGGTTTAATATTTTTGTGAGACCATCGAAGGAGAAAGAGATAA
59 bp   AACTTTTTTACGACTCCATCAGAAAGAGGTTTAATATTTTTGTGAGACCATCGAAGGAG_____
```

Figure 3.4 Special sequence features associated with mutations in TianTan strains

Panel A illustrates two of the small indels that differentiate the original TianTan genome sequence (TP00) from clone TP05. These indels in VACV strains are typically flanked by small duplications (underlined). Panel B illustrates the 59 bp repeats, which disrupts genes in the left and right TIRs of clones TP03 and TP05. The upper row shows the sequence of the 69 bp repeats found elsewhere in the telomeres of VACV strain TP00 (Genbank accession AF095689), the lower row shows the 59 bp repeats which disrupt genes in the TP03 and TP05 clones. Eight or nine copies of the 59 bp repeat are found in different TianTan clones. Note that VACV strain WR encodes an extra “A” between the two underlined “G’s” in a 70 bp repeat [205]. Many of our sequencing reads detected fragments of sequence derived from the virus telomeric repeats, and also encoding the 69 bp repeats reported previously [167].

We also noted a great many single-nucleotide polymorphisms (SNPs); far too many to be explained as sequencing errors and in numbers ( $\sim 0.5\%$  differences between TP00 and TP03/05) consistent with these viruses being different clonal isolates of an original vaccine stock. Table 3.2 summarizes the genes that exhibit significant size differences between the two sequences. It also highlights some small open-reading frames that were not previously annotated, but are now recognized as true genes due to their conservation across many different strains.

We also noted that TP05 differs significantly from TP00 in that much of the sequences comprising the TP00 TIR have been excised by a deletion spanning the homologs of the DVX002-12 genes. Unlike the 5.7 kbp deletion in the TP03 clone, the larger deletion in TP00 spares the virus copy of the VEGF gene. The TIRs of TP05 thus most closely resemble those of other extant VACV strains, suggesting that TP05 is probably more representative of the original TianTan lineage than are these other two TianTan clones.

We used these and other Orthopoxvirus genome sequences to assemble updated phylogenetic trees using an alignment of the conserved core region from gene homologs of DVX058 to DVX155 (VACV Copenhagen genes F9L-to-A24R). Different methods assign more distantly related viruses to different branches of these trees (Figure 3.5), with rabbitpox being assigned varying degrees of relatedness to other VACV strains. Horsepox virus regularly trees with viruses belonging to the Wyeth/Dryvax cluster but the sequence has drifted substantially judging by the branch lengths. However, regardless of the method used, TP03 and TP05 always cluster with TP00 and in a group that always forms a clade comprising VACV strains WR, Lister, and Copenhagen. VACV strain Copenhagen is also always assigned as a more basal root to this cluster (Figure 3.5).

Table 3.2 Gene differences between TianTan strains TP05 and TP00

Gene function or feature	ORF <sup>a</sup> number (TP05)	Size (bp)	
		TP05	TP00
Chemokine-binding protein(Cop-C23L/B29R)	TT_001	318	735
	TT_002	414	-
EGF growth factor (Cop-C11R)	TT_011	426	423
Ubiquitin ligase (WR012/207)	TT_014	189	156
Ankyrin-like protein (WR015)	TT_018	399	348
Ankyrin-like protein (WR017)	TT_020	216	231
$\alpha$ -amanitin sensitivity(Cop-N2L)	TT_033	528	291 <sup>b</sup>
Ankyrin(Cop-M1L)	TT_034	1413	1419
NF $\kappa$ B inhibitor(Cop-M2L)	TT_035	663	591
Ankyrin/ NF $\kappa$ B inhibitor(Cop-K1L)	TT_036	855	570
Putative monoglyceride lipase (Cop-K5L)	TT_040	366	405
Putative monoglyceride lipase (Cop-K6L)	TT_041	255	246
Cytoplasmic protein(Cop-F8L)	TT_050	198	153
IEV associated (Cop-F12L)	TT_054	1908	1095 <sup>b</sup>
IMV protein (WR053.5)	TT_057	150	150 <sup>c</sup>
IFN resistance/PKR inhibitor(Cop-E3L)	TT_063	573	342 <sup>b</sup>
Virosome component (Cop-E5R)	TT_065	1026	774
Putative ORF (Cop-E5R fragment)	-	-	321
F10L kinase substrate, core protein (Cop-E8R)	TT_068	822	825
DNA polymerase(Cop-E9L)	TT_069	3018	3021
Virulence factor(Cop-G6R)	TT_089	498	267 <sup>b</sup>
RNA polymerase (RPO147)(Cop-J6R)	TT_103	3861	2958
RNA polymerase co-factor (RAP94) (Cop-H4L)	TT_107	2388	1812 <sup>b</sup>
Decapping enzyme(Cop-D10R)	TT_120	747	507 <sup>b</sup>
NPH-I/Helicase (Cop-D11L)	TT_121	1896	1710 <sup>b</sup>
Disulfide bond pathway (Cop-A2.5L)	TT_126	231	- <sup>b</sup>
Core protein(Cop-A4L)	TT_128	852	510
IMV membrane protein (Cop-A14L)	TT_138	273	228
Entry and cell-cell fusion (Cop-A21L)	TT_145	354	351

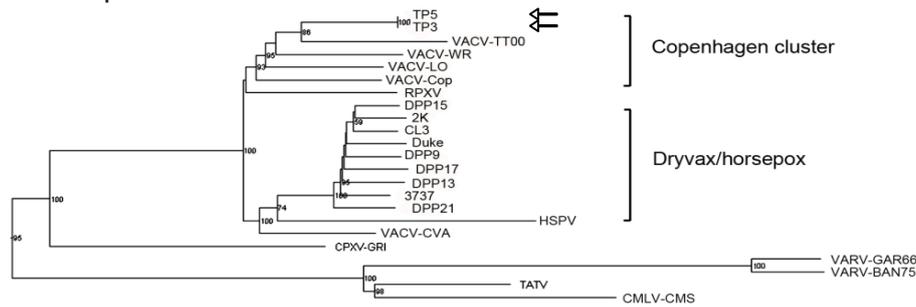
Cowpox A-type inclusion protein (Cop-A26L)	TT_151	684	693
P4c precursor (Cop-A26L)	TT_153	1509	690
Gene function or feature	ORF <sup>a</sup> number (TP05)	Size (bp)	
		TP05	TP05
Unknown(Cop-A31R)	TT_159	375	426
Unknown (YMTV-120.5L) (WR153.5)	TT_158	129	-
ATPase/DNA packaging protein (Cop-A32L)	TT_160	813	1176
EEV membrane protein (Cop-A33R)	TT_161	558	438 <sup>b</sup>
Unknown (WR161)	TT_166	195	195 <sup>c</sup>
Semaphorin (Cop-A39R)	TT_168	774	687
Semaphorin(Cop-A39R)	TT_169	933	429
Membrane glycoprotein(Cop-A43R)	TT_173	582	585
IL-1/TLR signaling inhibitor(Cop-A46R)	TT_176	711	633
TNF receptor (CrmC)(Cop-A53R)	TT_183	561	372
Unknown (Cop-B11R)	TT_198	219	231
IFN $\alpha/\beta$ receptor (Cop-B19R)	TT_206	1056	1062
Ankyrin (Cop-B20R)	TT_207	1794	1842

<sup>a</sup> Open-reading frame. We generally annotated any open-reading frame (> 150 bp) that had been previously identified as encoding a VACV gene. Where a gene was disrupted by a frame shift mutation (e.g. Cop-C23L/B29R) the open reading frames are indicated next to gene they are proposed to derive from.

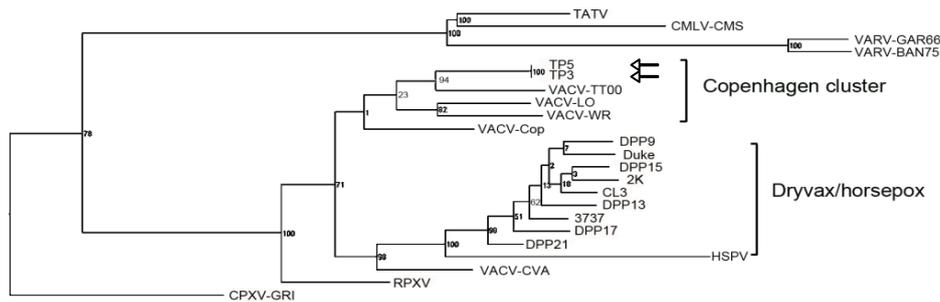
<sup>b</sup> Error noted and corrected previously [38]. The column labeled “TT00” shows the original reported gene size, the sequences we report in the column labeled “TP05” are identical to the corrected values reported by Upton et al.

<sup>c</sup> Not previously annotated.

### Least Squares method



### Maximum likelihood method



### Neighbour joining method

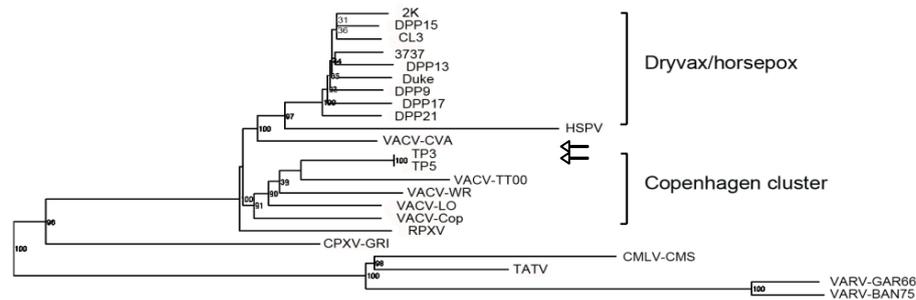


Figure 3.5 Phylogenetic relationships between different Orthopoxviruses

We prepared an alignment of virus sequences lying between the gene homologs of genes DVX058 to DVX155 (VACV Copenhagen genes F9L-to-A24R) for the indicated viruses, and then used three different phylogenetic methods to examine the relationship between the TianTan clones and other Orthopoxviruses. The three methods all produced very similar trees, with the TianTan clones always falling in a “Copenhagen” cluster. Cowpox (strain Brighton Red), ectromelia, and monkeypox viruses were included in the original alignment, but always mapped outside of these groupings and have been removed for clarity. White arrows indicate the two newly sequenced strains.

The depth of the branches separating TP03/05 from TP00 is much greater than the branches encompassing the Dryvax-derived clones (Duke, DPP9-21, Acambis 2000, CL3, etc), reflecting the fewer SNPs that differentiate the Dryvax clones.

One rather inexplicable feature of these phylogenetic trees is that the VACV WR strain persistently maps within the Lister/Copenhagen cluster, regardless of the method used to assemble the trees. VACV strain Western Reserve is reported to derive from the same New York City Board of Health virus stock as does Wyeth/Dryvax [summarized in [223]] and logically should associate with that cluster. We have not explored whether the contradictions between the reported history of WR and these phylogenetic trees relates to the particulars of the alignment we have used, or the choice of core genes, or perhaps reflects something unknown about the long and complex passage history of strain WR.

### **3.3.3 Other features of TianTan clones TP03 and TP05**

Another notable feature of both TP05 and TP03 is that the TIRs encode a 59 bp repeat sequence clearly related to the 69 or 70bp repeats [205] found in the telomeres of VACV strains (Figure 3.4 B). The acquisition of this sequence is linked to the deletion of genes that probably would have been homologs of DVX004, -5, and -6 (Copenhagen C19L-to-C21L). The length of the 59bp repeat region is too long to be determined by pyro-sequencing and so we used Sanger sequencing and the PCR and primers anchored in the flanking sequences to determine the cumulative lengths and number of repeats. This analysis showed that TP03 and TP05 encode 9 copies of the repeat whereas other sub-cloned viruses (e.g. TP04 and TP08) encoded one less copy (Figure 3.6). This feature appears to be an example of DNA having been accidentally captured through illegitimate recombination with another virus telomere, but why the repeat should be resistant to deletion is less certain.

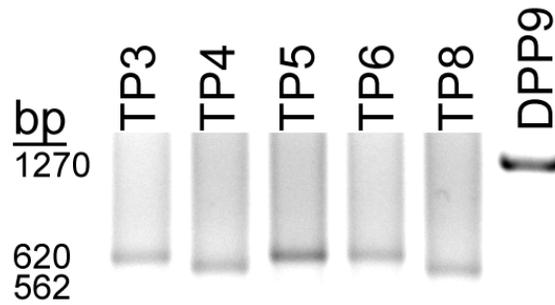


Figure 3.6 Analysis of the number of 59bp repeats within the TIR

PCR was used to detect the number of 59bp repeats and Sanger sequencing was used to confirm the result. TP3, TP5 and TP8 show a 621bp PCR product, corresponding to 9 copies of repeats. In comparison, TP4 and TP8 show bands of 562 bp in length, consistent with 8 copies of 59bp repeats. DPP9 is a dryvax clone, showing the length of original region replaced by the 59bp repeats in TianTan clones.

One possible explanation is that, once these sequences have been accidentally incorporated into the TIRs, the processes responsible for maintaining the inverted duplications also helps stabilize the inserts. The homologous region in TP00 has been excised by the large deletion, thus providing no insights into whether this mutation is a characteristic feature of all TianTan strains.

### **3.3.4 Large deletions help define the evolutionary trajectory of VACV strains**

If one compares the sequence of the TianTan strains with other sequenced VACV, one also sees specific patterns of gene deletions that can help define different lineages and identify genes essential for passage in culture. Moreover, if it is assumed that unique assemblages of genes are generally lost as virus strains evolve under culture pressure[227], such analyses can help identify the modern strains that likely still resemble ancestral VACV strains. For example, it has been suggested that VACV shares a common ancestry with a virus resembling horsepox virus [166, 228], although based upon the analysis of conserved core genes, this hypothesis is only supported by the simplest of the phylogenetic tests we have used (Figure 3.7). To gain a better understanding of the relationships between different VACV strains, we ran a BLASTN search using ~20 kbp of sequence spanning the left end of the HSPV genome (Figure 3.8). The sequenced VACV strains fall into various groupings that are defined by different deletion and TIR boundaries and show evidence of previously noted evolution through gene deletion [227]. All of the VACV strains are distinguished from HSPV by a common 10.7 kbp deletion (DNA encoding the mostly fragmented HSPV007- 015b genes [166], grey bar in Figure 3.8) that juxtaposes the SPI-1 and VEGF genes in VACV.

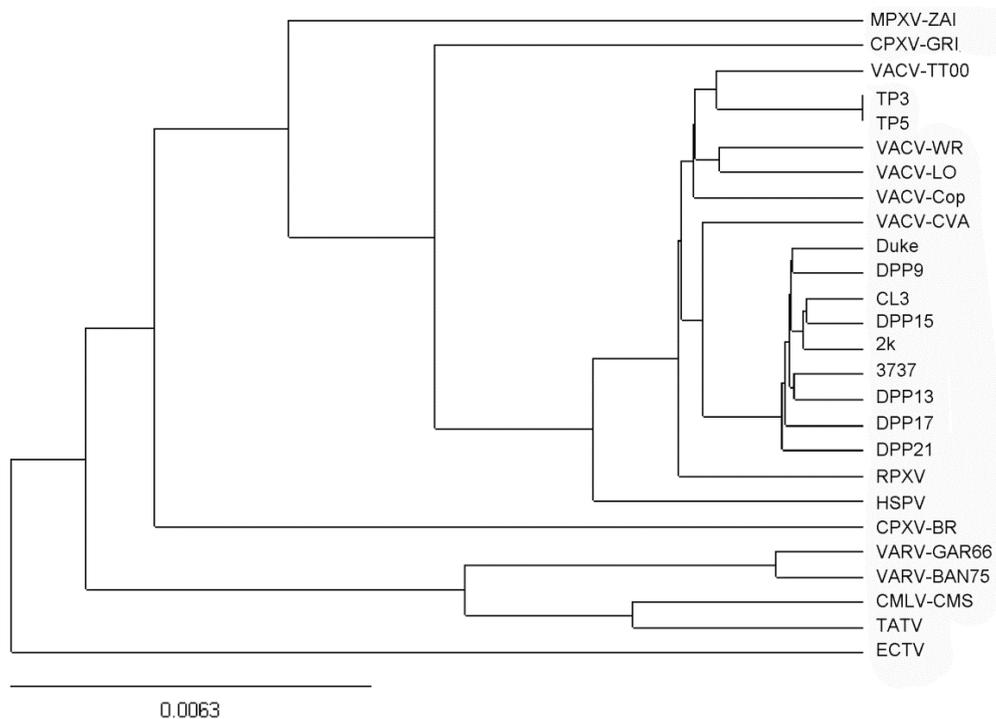


Figure 3.7 Phylogenetic relationships between different Orthopoxviruses

A simple UPGMA (unweighted pair group method with arithmetic mean) approach was also used, along with the alignment prepared for Figure 3.5, to explore the relationship between Orthopoxvirus strains. The method again groups TianTan strains with VACVs like Copenhagen, but now places horsepox virus (HSPV) on a branch distinct from other VACV-like strains. Tulman et al. similarly located HSPV in an early, but distinct branch of the VACV family [166].

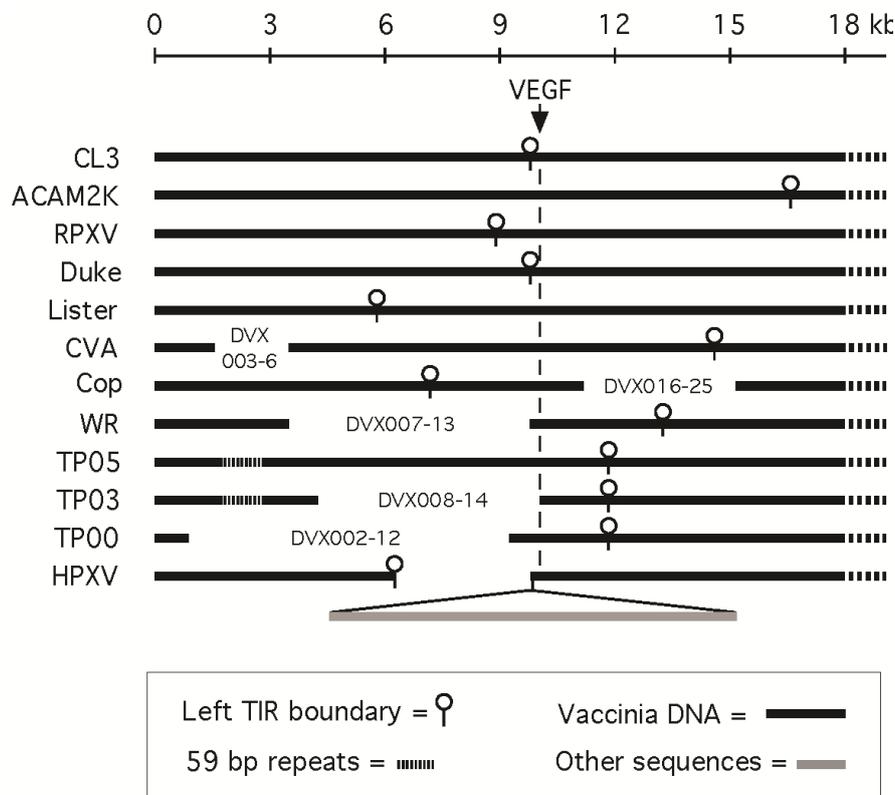


Figure 3.8 Relationship between deletions mapping in the left telomere of common VACV strains

The TP05 strain encodes the longest and most complete TIR in the TianTan lineage. Note that the VEGF gene seems to be a privileged site, which is sometimes duplicated, and nearly always retained in VACV strains. The deletion extending into the VEGF gene promoter and N-terminus in the TP03 genome is unusual in this regard. Horsepox virus (HPXV) encodes an additional 10.7 kbp insert upstream of the VEGF gene (grey bar). The labels (e.g. DVX007-13) refer to genes present in a consensus of VACV Dryvax clones, which would have been located in the region spanned by the deletion[223].

Many genomes have been cloned from vaccine stocks (e.g. Lister and ACAM2K), which encode a full complement of the remaining genes in this part of the genome, although the location of the left TIR boundary (Figure 3.8, circled vertical bars) varies between these strains due to the presence of different deletions in the region surrounding the right TIR boundary (see below). TP05 resembles these more complete VACV strains, aside from the disruption of genes homologous to DVX004-6 by the insertion of repeat elements near the left telomere. TP03 and TP00 exhibit more extensive deletions, much like those that have accumulated in VACV strain WR. An interesting feature of these viruses is that nearly all, except TP03, retain at least one functional copy of DVX014, the VEGF gene homolog discussed above. This strongly suggests that the gene is under positive selection pressure in culture.

A similar pattern of gene deletions is also seen at the right end of the different VACV genomes (Figure 3.9), although this part of the genome appears to be more dynamic and subject to rearrangement. Here the “Acambis clone 3” (CL3) isolate is an anomaly, as it is the only type of VACV to retain DNAs homologous to 1.1 kbp of HSPV sequence in the region near genes DVX214-15. Related sequences are also detected in ectromelia, camelpox, and variola viruses. The CL3-like viruses are a rare type of VACV clone found in Dryvax stocks [223], which exhibits unusually high virulence compared to strains like ACAM2K [186]. These sequences have been deleted from all other known VACV isolates including the TianTan strains and it is possible that they play some role in CL3’s unusual virulence. Otherwise, there is little to differentiate (or relate) the pattern of gene deletions in TianTan virus from/to most other VACV strains although the lack of alignment to some portions of the Lister TIR (Figure 3.9, grey bar) is not supportive of a direct relationship between the viruses.

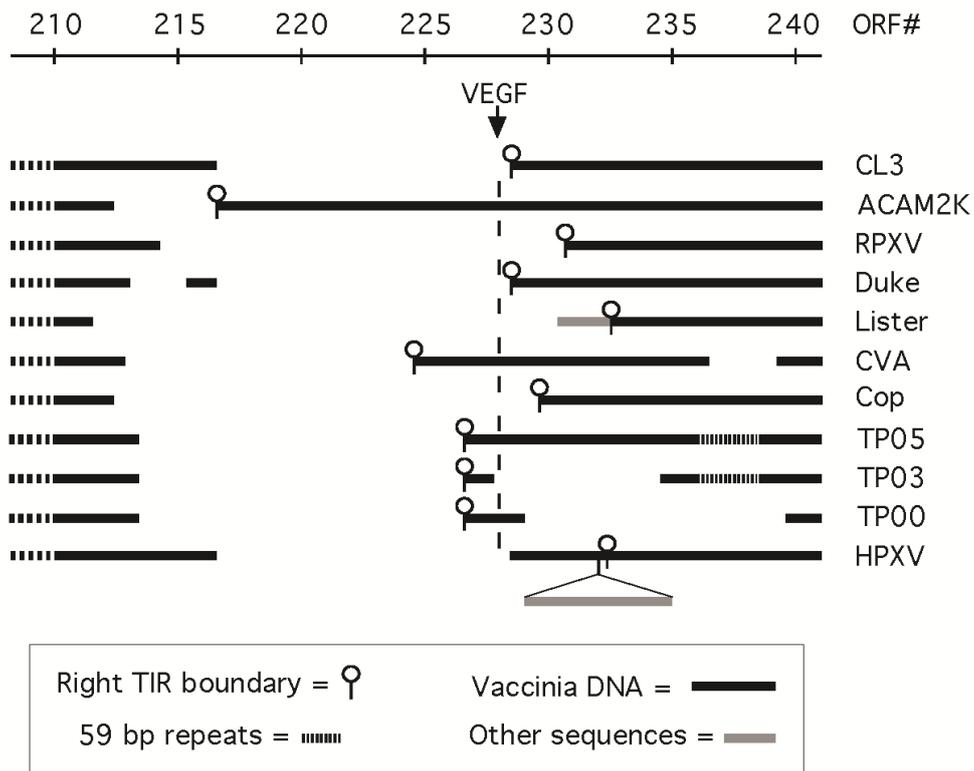


Figure 3.9 Relationship between deletions mapping in the right telomere of common VACV strains

The TP05 strain encodes a complement of genes in the right telomere similar to most other VACV, apart from the disruption of three genes in the TIRs by the 59 bp repeats. The left end of the deletion extending to the left of the TIR boundary (barred circle) does not extend into the DVX212/TT\_206 gene (a secreted IFN  $\alpha / \beta$  receptor homolog) in the TianTan strains, in contrast to strains like ACAM2K where this gene is partially truncated by a larger deletion. Note that because of the complexity of the maps in this region of the different genomes, we have used the open reading frames (ORF#) to align the different sequences instead of true map distances. We have also omitted strain WR, due to the additional complexity of the rearrangements in this portion of the WR genome.

Note that because of the complexity of the rearrangements in this region of the genome, we have used ORF numbers to simplify the comparison rather than distances shown in Figure 3.8.

Assuming that VACV strains have undergone progressive deletion of genes in culture, this analysis of the telomeres suggests that they have followed an evolutionary path involving HSPV-like and then CL3-like viruses, after which different strains have evolved along different pathways. Viruses that have been cultured for part of their history in eggs (like CVA/MVA and TianTan) seem to have suffered further deletions.

### **3.4 CONCLUSIONS**

Our sequencing of viruses belonging to the TianTan lineage, clearly documents its relationship to other widely used VACV viruses, and suggests that TianTan virus most likely shares a common origin with a strain also ancestral to VACV strain Copenhagen. This relationship is clearly supported by all available phylogenetic methods and is compatible with the patterns of deletions.

However, the genome of the virus has clearly also been affected by its separate passage history, with additional mutations and deletions that have further altered the TIRs. Although the virus we obtained must have been plaque purified at some point in its recent history, it still contained two types of viruses that exhibited significantly different growth properties due to deletions into the region encompassing the VACV epidermal growth factor gene homologs.

A comparison of our sequence with the published TP00 sequence still detected far too many single-nucleotide polymorphisms and indels for all to be a product of simple sequencing errors. Indeed, many of the indels illustrate the same pattern of repeat driven mutations (Figure 3.4) seen in Dryvax clones [223]. Our strain(s) and the TP00 isolate are likely just several examples of the viruses found in the original TianTan stocks, and the divergent sequences further illustrates the complexity of the viruses comprising the quasispecies that makes up a traditional vaccine. Although it

is not widely used anymore as a smallpox vaccine, VACV continues to be used as a vector in diverse applications. The complexity detected by DNA sequencing technologies illustrates the importance of characterizing any clones randomly “fished” out of a traditional unpurified stock of smallpox vaccine, because viruses encoding or lacking a gene as critical as the VACV growth factor gene will likely exhibit dramatically different safety and efficacy profiles.

**CHAPTER FOUR – GENOME SCALE PATTERNS OF  
RECOMBINATION BETWEEN CO-INFECTING VACCINIA  
VIRUSES**

Li Qin and David H. Evans<sup>3</sup>

**4.1 INTRODUCTION**

Recombination plays an essential role in DNA repair and, by creating new combinations of genetic traits, it averts the decline in fitness caused by the accumulation of mutations ("Muller's ratchet") while creating the genetic diversity that is the substrate for Darwinian selection. Bacteriophage were the first viruses shown to be subject to recombination [229] and the phenomenon was soon also detected in co-cultures of many other viruses including herpes simplex virus [230] and vaccinia virus [92]. In the years immediately following, research showed that *in vitro* recombination could also produce hybrids between related poxviruses such as rabbit fibroma and myxoma viruses [231] and between variola virus and cowpox and rabbitpox viruses [232, 233]. The subsequent discovery of another natural hybrid between rabbit fibroma and myxoma viruses, called malignant rabbit virus [234], suggested that poxviruses can also recombine in co-infected animals and the significance for human health is illustrated by the fact that variola minor virus may be a recombinant derived from more virulent West African and Asian variola major strains [235]. We have recently published an analysis of some of the strain variants found in an old non-clonal smallpox vaccine, Dryvax, and detected one virus bearing a small region of sequence wherein the single nucleotide polymorphisms (SNPs) were more

---

<sup>3</sup> A version of this chapter has been published in *J Virol* 88(10): 5277-86 (2014).  
*Copyright 2014. American Society for Microbiology.*

characteristic of horsepox than of vaccinia virus (VACV) [223]. This may represent a sequence relic that has been retained in the absence of counter selection and the population bottlenecks caused by periodic plaque purification.

Genetic crosses between viruses encoding different selectable markers were once also used to try and assemble recombination-based maps of VACV [236-238], although the method never proved very useful and was soon supplanted by marker rescue and DNA sequencing technologies. This was due to the limited distances over which linkage is retained relative to the spacing between many markers (<20 kbp over a 200 kbp genome) and the difficulty of reproducibly measuring recombinant frequencies [216]. Further studies showed that homologous recombination can be used to genetically modify VACV [239, 240], that this is an accurate process [241, 242], and that these processes also operate *in trans* and can be detected using transfected DNAs [243]. Poxviruses replicate in sequestered structures called factories [244], each of which derives from a single infecting particle, and it is presumed that recombinants can only form within these factories if different DNAs mix in the presence of the recombination machinery. VACV uses a single-strand annealing mechanism to produce recombinant molecules in a reaction catalyzed by the E9 viral DNA polymerase and I3 single-strand DNA binding protein [108, 109] and, since both proteins primarily reside within virus factories, that is presumably where recombination also occurs. We have suggested that random variations in the timing and degree of mixing of virus factories within co-infected cells could explain why recombinant frequencies proved difficult to measure reproducibly [216]. If two or more viruses infect any particular cell, but a portion of the factories don't mix, such a process would decrease the yield of recombinants in a stochastic manner relative to the number of non-recombinant (i.e. fully parental) viruses produced by DNA replication.

Although much has been learned concerning the mechanism of poxvirus recombination, questions remain regarding how these processes and physical constraints might affect the patterns of DNA exchange, and thus the overall genetic composition of the resulting pool of parental and recombinant viruses. How genetic linkage distances relate to the actual numbers of exchanges in recombinant viruses also remains to be established. In this study we have used the ~1400 SNPs that differentiate two strains of VACV, a Dryvax clone and a TianTan clone, as sequence tags that can be used to track the origins of the different DNA segments in recombinant progeny. Our study shows that VACV recombination reactions produce genomes exhibiting a “patchwork” of exchanges, some apparently derived from a succession of crossovers over the course of a single infection cycle. Interestingly, viruses produced using non-genetic reactivation methods [245], appear indistinguishable from recombinants produced in a more regular manner, showing that such viruses are likely subjected to similar replication and developmental pathways.

## **4.2 METHODS**

### **4.2.1 Cells and viruses**

Vaccinia virus strains DPP17 (GenBank JN654983), TP03 (GenBank KC207810), and TP05 (Genbank KC207811) were cloned from stocks of Dryvax (DPP17) and TianTan (TP03/05) viruses [223, 246]. They were cultured on BSC-40 cells in modified Eagle’s medium (MEM, Gibco) supplemented with 5% fetal bovine serum, 1% nonessential amino acids, 1% L-glutamine, and 1% antibiotic at 37°C in a 5%CO<sub>2</sub> atmosphere. Two types of recombinant virus stocks were prepared. The DTM viruses (Dryvax-TianTan mixture) were generated by co-infecting cells with DPP17 and TP05 at a multiplicity of infection (MOI) of 0.02 (each 0.01), culturing the cells for two days, harvesting the cell-virus mixture, and releasing the

virus by freeze-thaw. A sample (10  $\mu$ L, 0.5% of the lysate, or  $\sim$ 0.02 PFU/cell) was then used to infect another fresh dish of cells and this was repeated for a total of five rounds of passage. Individual DTM viruses were then isolated using three rounds of cloning by limiting dilution as described [223]. The DTH viruses (Dryvax-TianTan high MOI) were produced by co-infecting cells with DPP17 and TP05 at MOI=10 (each 5) for 24 hr, followed by three rounds of cloning. Plaque images were processed with ImageJ [193].

#### **4.2.2 Virus DNA reactivation**

A third collection of recombinant VACV were prepared using DNA reactivation reactions as described previously [245]. Briefly, BGMK cells were grown to near confluence in 60 mm dishes and infected with Shope fibroma virus (SFV, strain Kasza) at MOI=1 in PBS. After one hour at 37°C, the medium was replaced with MEM containing 10% fetal bovine serum, incubated for another hour, and the MEM replaced with Opti-MEM (Gibco). DPP17 and TP03 VACV DNAs were extracted from sucrose gradient-purified virions using phenol chloroform, mixed in 1:1 ratio, and the SFV-infected cells transfected with 5  $\mu$ g of this DNA using Lipofectamine 2000 (Gibco). The cells were incubated for 4 hr, the Opti-MEM media was replaced with MEM containing 10% serum, and the cells cultured for another 3 days. The cells were then subjected to three rounds of freeze-thaw and 0.2 mL used to infect BSC-40 cells (which do not support SFV growth).

The reactivated VACV were cloned using three rounds of limited dilution and recombinants were identified using the PCR and the primer pairs DVX-209F and DVX-226R plus DVX-004F and DVX-007R (Table 4.1). The 209F/226R primers should produce a 1.1 kbp amplicon in reactions containing DPP17 DNA whereas TP03 DNA does not serve as a substrate.

Table 4.1 PCR primers used in this study

Amplicon size (bp)		Primer ID	Primer (5'-to-3')
DPP17	TP05		
1120	-	DVX-209F	CGAAGAAGATGATGGGGAC
		DVX-226R	ATAAGAGGAAAGAGGACAC
-	653	DVX-213F	CGTTGGATGGATTTCGATA
		DVX-226R	ATAAGAGGAAAGAGGACA C
1230	665	DVX-004F	GCAGTAGGCTAGTATCTT
		DVX-007R	TACCGGCATCATAAACAC
225	260	DVX-107F	AACTGGAGTAGAGATAGC
		DVX-108R	CCGAGAATATAGCTGTCC
		201F	AATATGATGGTGATGAGCGA
		239R	TATTGCGAGATGTGAAGG
712	712	208F	TTTCTTCTCTTCTCCCTTTC
		209R	ATTCTATCCCGTACCTCT

The 004F/007R primer pair targets telomeric repeat sequences and should produce 665 bp and 1230 bp products in reactions containing TP03 and DPP17 DNAs, respectively. After cloning, the viruses were purified and sequenced as described. Four DTD clones (Dryvax-TianTan DNA reactivation) were sequenced.

#### **4.2.3 Virus sequencing and genomic analysis**

Stocks of 16 DTM clones, 15 DTH, and 4 DTD clones were prepared and purified over sucrose gradients. Viral DNAs were extracted and sequenced as described previously [223] using a Roche 454 GS Junior system. Roche GS De Novo Assembler software was used to deconvolute and assemble the raw sequencing data into contigs and nearly full-length genomes were generated using CLC Genomics Workbench 6. The average read redundancy was 15, which permitted the assembly and mapping of all of the recombinant junctions with confidence. Multiple sequence alignments were prepared using the program LAGAN (<http://genome.lbl.gov>) [198] and Base-by-Base software [199] was used to produce a visual summary of the whole genome alignments. The assembled sequence data have been assigned GenBank accession numbers KJ467582 to KJ467616, inclusive (Table 4.2).

#### **4.2.4 PCR and Southern blotting**

Southern blotting was used to confirm the rearrangement detected in clone DTM28. Virus DNA was digested with *Nde*I (Fermentas) and size fractionated by electrophoresis on 0.7% agarose gels. The DNA was fragmented with 0.2M HCl, denatured with 0.4M NaOH, transferred to a nylon membrane, and fixed with a UV cross-linker. Two primers (201F and 239R, Table 4.1) and the PCR were used to prepare a probe in a reaction containing biotin-16-dUTP (Roche, 1093070), which was subsequently hybridized to the prepared membrane and detected using IRDye 800CW-coupled streptavidin (Li-Cor; 926-32230) and a Li-Cor imager.

Table 4.2 GenBank accession numbers

The sequences were submitted without extra annotation (e.g. gene features, open reading frames, etc.) and thus will not show up in BLAST databases.

Recombinant Clone ID number	GenBank Accession Number		Recombinant Clone ID number	GenBank Accession Number
DTM03	KJ467582		DTH10	KJ467600
DTM04	KJ467583		DTH10.2	KJ467601
DTM06	KJ467584		DTH13	KJ467602
DTM08	KJ467585		DTH14	KJ467603
DTM09	KJ467586		DTH21	KJ467604
DTM10	KJ467587		DTH22	KJ467605
DTM11	KJ467588		DTH27	KJ467606
DTM19	KJ467589		DTH28	KJ467607
DTM22.1	KJ467590		DTH30	KJ467608
DTM22.2	KJ467591		DTH31	KJ467609
DTM27	KJ467592		DTH34	KJ467610
DTM28	KJ467593		DTH36	KJ467611
DTM29	KJ467594		DTH41	KJ467612
DTM30	KJ467595		DTD03	KJ467613
DTM32	KJ467596		DTD06	KJ467614
DTM33	KJ467597		DTD11	KJ467615
DTH06	KJ467598		DTD18	KJ467616
DTH08	KJ467599		-	-

The DTM28 and DTM28 $\Delta$  viruses were also differentiated using the PCR and 208F and 209R (Table 4.1) primers. This was done in combination with 201F and 239R in a PCR reaction containing all four primers.

## **4.3 RESULTS AND DISCUSSION**

### **4.3.1 Virus isolation and genome sequencing and assembly**

We used three different methods to produce VACV recombinants. The first method was designed to explore the effects of repeated passage, at low MOI, on a seed mixture initially composed of just two different genetically tagged viruses. These viruses were originally cloned from stocks of Dryvax (DPP17) and TianTan (TP05) vaccines and differ in sequence by 1 SNP per ~140 bp. Compared to TP05, DPP17 also encodes a 6 kbp deletion near the right terminal inverted repeat boundary as well as ~150 other smaller insertions and deletions (indels) distributed across the two genomes (Figure 4.1). The TP05 strain forms plaques that are approximately twice the diameter of those formed by the DPP17 strain, which provided an opportunity to explore what effect a growth bias might have on the pattern of recombinants.

For this first experiment, a 1:1 mixture of the two different VACV were used to infect BSC-40 cells at MOI=0.02, cultured 48 hr, and a portion (10  $\mu$ L or ~0.02 PFU/cell) of the resulting progeny passaged again under the same conditions. This was repeated to produce a total of 5 rounds of replication. Each time the infection produced overlapping plaques that partly cleared the entire plate. We then plated out the diluted virus in 24-well plates, and identified 36 wells each containing just a single random plaque. These 36 viruses were then cloned again, also by limiting dilution, and designated DTM (Dryvax-TianTan mixture) strains.

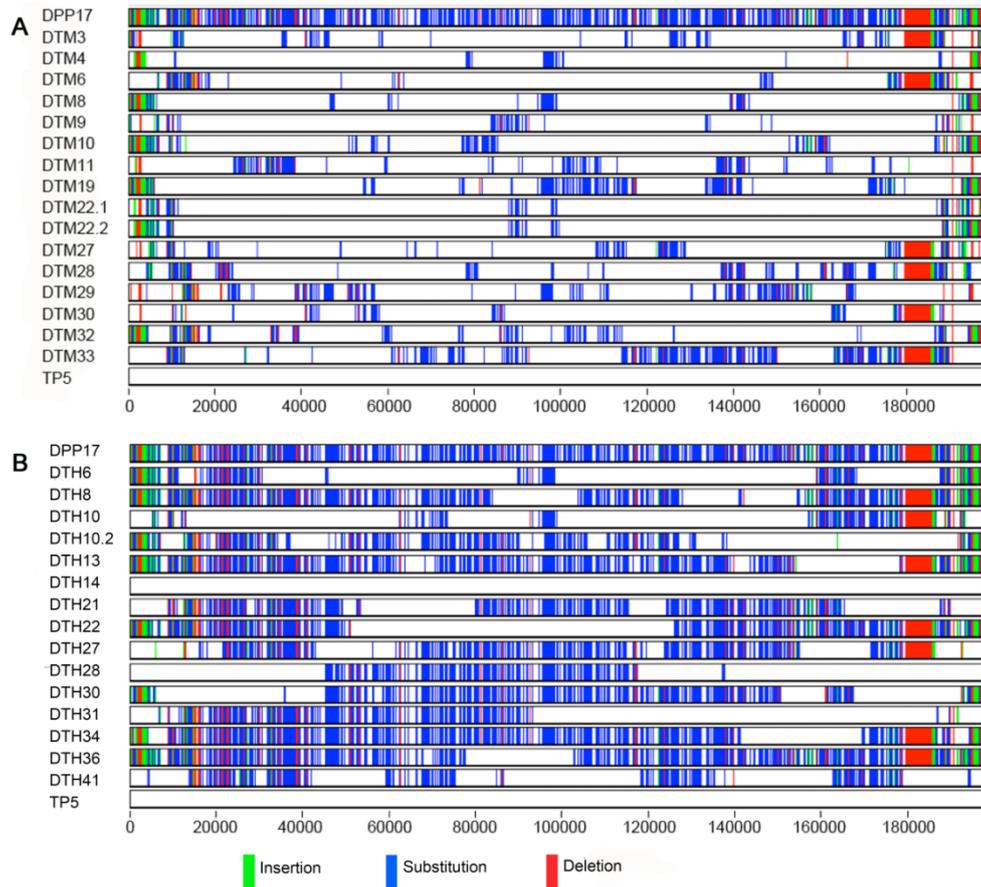


Figure 4.1 Patterns of DNA exchange in recombinant vaccinia viruses

The genome sequences of DTM (panel A) and DTH (panel B) recombinant clones were aligned against the parent genomes DPP17 and TP05 using the program “LAGAN”, and edited using the program “Base-by-Base”. TP05 was used as the reference strain and any differences between a given virus and TP05 are colour coded to indicate insertion, substitution, and deletion mutations derived from strain DPP17 (inset at bottom). Thus the blank regions represented fragments derived from TP05.

Using this method minimized the risk of picking certain plaque types, since the only criteria we used to choose a clone was that the virus had to have been diluted to the point where it was the only plaque in a well, in the first round of selection. To avoid the problem of resequencing any non-recombinant parental stains, the PCR and three different primer pairs were used to determine the genetic origin of three different sites within each genome: within the terminal inverted repeats (primers 004F/007R), in the central part of the genome (primers 107F/108R), and near the junction with the right terminal inverted repeat (213F/209F/226R) (Table 4.1). Fourteen clones were selected because at least one position was recombinant with respect to either of the other two sites. We also chose two additional viruses, which exhibited a parental arrangement of markers at these three sites, although these were subsequently determined to also be recombinants. These viruses were cloned two more times and 16 were sequenced. After sequencing and assembly, these DTM recombinants exhibited a patchy pattern of SNPs suggesting each virus was the product of approximately 30 exchanges over the course of virus replication.

One expects that when viruses are passaged five times under these conditions, it should provide an opportunity for repeated rounds of replication and recombination. We also examined what the virus progeny would look like if they were permitted just a single round of infection, although it is expected that this would still involve multiple rounds of replication. To do this, we co-infected BSC-40 cells with DPP17 and TP05 viruses at MOI=10 (5 pfu/cell of each virus) and cultured the viruses for just 24 hr. These viruses were cloned and designated DTH strains (Dryvax-TianTan high MOI). After the first round of cloning, 43 viruses were randomly selected and the PCR was used to identify putative recombinants as described above. Thirteen hybrid viruses were cloned twice more, along with two additional viruses (DTH13 and DTH14) that exhibited a parental pattern of markers at the three positions tested by the PCR.

DTH14 was subsequently identified as being identical to the TP05 parent virus, while DTH13 proved to be a recombinant. Ultimately, 15 DTH clones were sequenced and assembled as described before. These recombinants exhibited a mean of 18 exchanges per genome.

We should note one *caveat* regarding these methods, in that single plaques isolated in the first round of purification were not always pure, and this provided some limited opportunity for additional rounds of replication and recombination. For example, when a plaque initially identified as recombinant DTM22 was cloned a second time, and the subsidiary plaques reanalyzed by PCR, it was realized that the two daughter plaques (DTM22.1 and DTM22.2) were not identical. However, they are clearly “sibs”, viruses sharing a common genetic origin as judged by a shared pattern of exchanges in the center of the two genomes (Figure 4.1). We also noted one case where a single apparently recombinant starting plaque resolved into two clearly unrelated recombinant clones upon replating (clones DTH10 and DTH10.2, Figure 4.1). For simplicity, our analysis has treated these particular clones as being the same as the other recombinants isolated in the study, although they may have experienced some additional limited opportunities to undergo recombination.

#### **4.3.2 Crossovers in DTM and DTH viruses**

After assembly, the sequences were aligned with program LAGAN and the alignment corrected manually using Base-by-Base. Inspecting these sequences we could readily identify the origin of each SNP-tagged segment of DNA as belonging to either the DPP17 or TP05 parent (Figure 4.1). What was remarkable was the very low frequency of observed mutation even though numerous SNPs and small indels commonly differentiate clones isolated from a viral stock like Dryvax [223]. No mutations were detected in any of the DTH clones, compared with the two parent viruses, and just two mutations in two of the DTM clones. One was a small deletion in DTM29 at

alignment position 900, which removed two nucleotides (Figure 4.2, panel A) just 6 nt upstream of the ORF001 start codon. Although most small deletions are associated with repeats [171, 223], this event was not. It was located immediately adjacent to a SNP that differentiates DPP17 from TP05. We also discovered a point mutation in DTM27 at alignment position 70,493 (Figure 4.2, panel B). This causes a C-to-T transition mutation and an alanine-to-valine substitution in gene DVX\_088 (RNA-helicase). A crude estimate of the VACV replicative error rate in laboratory culture can be calculated from the following observations and assumptions. We note that there were only two independent mutations detected in 16 DTM viruses over the course of 5 rounds of infection, and there are ~200,000 nt copied per genome per each round of infection. Each round of infection typically expands the VACV titer ~10,000-fold (i.e. between  $2^{13}$  and  $2^{14}$  doublings of the genome) and thus the error rate is very crudely estimated as  $2 \div [16 \times 200,000 \times 5 \times 13.5]$  or  $\sim 1 \times 10^{-8}$  mutations per nucleotide copied per cycle of replication. Alternatively this is  $2 \div [16 \times 5 \times 13.5]$  or ~1 mutation per 500 genomes per cycle of replication. By a similar method, the absence of any mutations detected amongst the 15 DTH clones over the course of a single round of infection suggests an error rate of  $< 5 \times 10^{-8}$ . The VACV E9L gene encodes a typical B-family proofreading DNA polymerase of a type encoded by a variety of viruses and bacteriophage, and this error frequency resembles that reported for phage [247]. Although some drug-resistant E9L alleles cause altered spontaneous mutation rates *in vivo* [248-250], for comparison purposes these cannot be converted into absolute mutation rates given the uncertainties in the size or number of genetic target(s). We could find no other reported absolute error rates for poxviruses in the literature. Beyond these two rare mutations, the remainder of the sequences in the recombinant genomes could be ascribed to having been inherited from one or the other of the two parent viruses.

A. Deletion mutation.		
ACATGCATGCCAGGACGATATATTGTT		DPP17/6 DTM Clones
ACATGCATGCCAGTACGATATATTGTT		TP05/10 DTM Clones
ACATGCATGCCAGT__GATATATTGTT		DTM29
B. C-to-T substitution.		
AATTTAGTTACTGCTATACAGATGTAT		DPP17/TP05/16 DTM Clones
AATTTAGTTACTGTTATACAGATGTAT		DTM27

Figure 4.2 Rare mutations associated with replication and recombination  
 Panel A shows a deletion mutation in DTM29, immediately adjacent to a G-to-T SNP found at alignment position 70493. Panel B shows a C-to-T substitution detected only in clone DTM27 at position 900.

In total, 1399 SNPs (single nucleotide polymorphisms) can be used to differentiate DPP17 from TP05 and we used these SNPs to track the origins, and thus the sites of crossing over, in the hybrids. The relative abundance of these variant sites (1399 scattered across ~200 kbp) allowed us to map the site(s) of crossing over with an average resolution of ~140 nt. In general, each hybrid virus encoded variable-length blocks of DNA derived from each of the two parent strains, and no uniquely conserved block (a hallmark of a highly selected patch of DNA) was detected in all of the viruses. The lengths of these blocks of recombined sequences varied, ranging in length from one to several hundred SNPs. We detected none of the large gene duplications that have been described by other authors [103, 104], but would not expect to do so given that these structures are stable only in the presence of strong selection pressure.

To examine the pattern of crossing over in greater detail, we used “Base-by-Base” software to produce a table ascribing each of the 1399 SNPs detected in each hybrid as being of either DPP17 or TP05 origin (data not shown). This provided a tool for calculating the number of exchanges on the assumption that each time the pattern of SNPs changed from DPP17 to TP05 (or *vice versa*), that an exchange had occurred and was counted as one crossover. It is important to note that this is still an underestimate of the true physical recombination frequency, as any recombination between two parental genomes (e.g. DPP17×DPP17 or TP05×TP05), or recombination occurring over distances less than the distances between SNPs cannot be detected by this method. Figure 4.3 shows the results of this analysis. We had expected that viruses given more opportunities for recombination would exhibit a greater number of exchanges, and this was supported by these measurements. The number of crossovers in the DTM viruses ranged from 14 to 44 (mean = 30±11 [SD]).

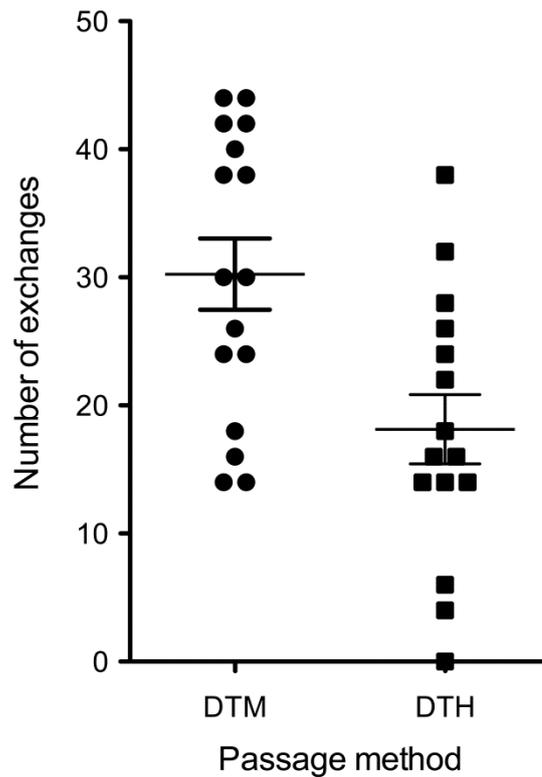


Figure 4.3 Numbers of exchanges in DTM and DTH clones

Each of the hybrid viruses was first aligned against strains TP05 and DPP17. Then, the program “Base-by-Base” was used to determine where each cross-over was located relative to the 1399 SNPs that differentiate the two strains, along with the number of such exchanges. The viruses passaged five times (DTM) exhibited more exchanges per genome than the viruses passed just once (DTH). That is  $30 \pm 11$  versus  $18 \pm 11$  exchanges/genome, respectively.

In contrast, the numbers of crossovers in the DTH viruses ranged from 0 to 38 (mean =  $18 \pm 11$  [SD]), with one virus, DTH14, identical to the TP05 parents. In a previous publication [216] we conducted a meta-analysis of all of the published VACV genetic recombination data and attempted to correlate these data with the known physical map of the virus. From these studies we derived an estimate that half-maximal recombination was detected in a single round of infection (akin to the method used to produce the DTH viruses), when the distance between VACV markers was  $\sim 8$  kbp. Although this number is difficult to estimate with precision due to a number of experimental factors [216], from our measurements of the conversion track lengths (see below) we calculated that the DTH viruses exhibited a mean of about 1 physical crossover per 12 kbp, a number compatible with the  $\sim 8$  kbp deduced from the older genetic literature. Historically, classical genetic crosses never proved a useful tool for mapping VACV genomes, due to a combination of experimental noise and short linkage distances relative to the distances between most VACV markers. When one considers that our estimate of 1 crossover per 12 kbp is associated with a standard deviation of 19 kbp (i.e.  $12 \pm 19$  kbp), the source of this problem is clearly apparent.

#### **4.3.3 Average length of the conversion tracts**

Besides measuring the numbers of exchanges suffered by each of the recombinant viruses, we also examined the lengths of the DNA segments exchanged between viruses (i.e. the conversion track length) in the DTH group. To do this, the calculation assumed that the start and end of each exchange lay midway between the SNPs flanking the two sites of exchange. The resolution of the method varies depending upon the local SNP density, but with an average of 1 SNP per 140 bp we could detect exchanges ranging in size from 55 to 92,000 bp. An interesting feature of VACV recombination is illustrated by this analysis, which showed that there were relatively more short conversion tracks than long ones (Figure 4.4).

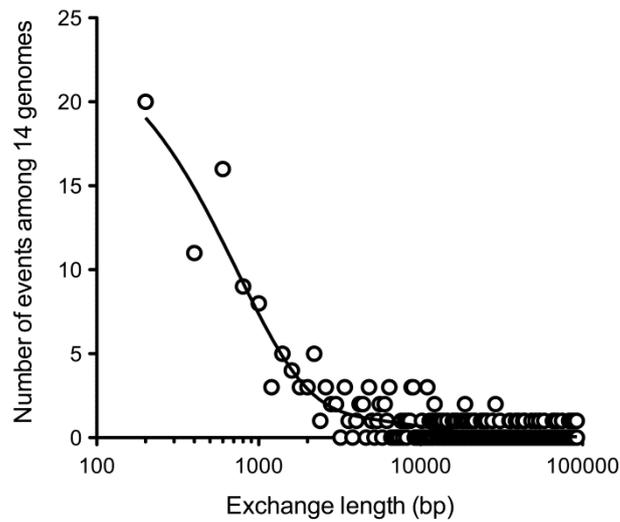


Figure 4.4 Length of the DNA segments exchanged in DTH clones

The lengths of all the conversion tracks were measured in all 14 genomes using midpoints defined by the four SNPs flanking the two bounding sites of exchange. The numbers of events, of a given exchange length, were then determined by assignment to 200 bp bins. A semi-logarithm of the bin size (i.e. exchange length) is presented because the values differ so greatly in scale across the different genomes. VACV recombination appears to be associated with a disproportionate number of very short exchanges. Because there is approximately one SNP per 140 bp, greater resolution than ~200 bp is not achievable.

Thus while the mean length of a conversion track was 12 kbp, the median was only 2.6 kbp. The abundance of short conversion tracks would help favor intragenic recombination events, which can be detected between markers spaced only 54 bp apart [251]. We should note that this estimate of the recombination frequency is lower than has been previously reported. For example we detected a loss of linkage at distances exceeding 350 bp in one study [252]. However, this earlier experiment measured the yield of recombinants when DNA was transfected into Shope fibroma virus-infected cells, and it is possible that the non-specific DNA replication that is seen under these circumstances [253] also exposes transfected DNAs to higher levels of recombination than is normally experienced by viruses.

Crossing over is not the only process that could produce this abundance of short exchanges. Poxvirus replication and recombination reactions also produce hybrid (or heteroduplex) DNA [254]. Such molecules would contain mismatched bases wherever the sequences differ and if a subset of mismatches were subjected to mismatch-specific and directionally biased short patch DNA repair prior to further replication, it could create the appearance of closely linked crossovers. However, our data provide no evidence that mismatch repair is producing artifactual exchanges. If any repair bias existed, it would most probably occur at sites where the hybrid DNA contained G·T mismatches, since G·T mismatches are generally repaired to G·C basepairs in cells by a pathway employing a thymine DNA glycolylase [255]. A G·T mismatch would be formed (along with a reciprocal A·C mismatch) at sites where substitution mutations differentiate the two viruses. Such substitutions (i.e.  $G \rightarrow A$ ,  $A \rightarrow G$ ,  $C \rightarrow T$ , or  $T \rightarrow C$ ) comprise 72% of the SNP markers in these crosses. If one makes the simplifying assumption that A·C mismatches are just replicated and not repaired, and that all G·T mismatches are converted to G·C prior to replication, then these single marker exchanges should be biased 2:1 in favour of forming (or retaining) a G or C. We detected 51 single exchanges

at sites containing base substitutions (82% of all single exchanges), in the DTM and DTH viruses. Of these, 29 retained a G or C, and 22 retained an A or T, which is not significantly different from a 1:1 split ( $\chi^2=0.96$ ,  $P=0.33$ ). Although we cannot *disprove* the hypothesis that biased mismatched repair created some of the short conversion tracts, the simplest explanation for these data is that poxvirus recombination reactions produce an abundance of short conversion tracks through a process formally akin to crossing over.

#### **4.3.4 Biased genetic origins in progeny viruses**

An interesting difference between the TianTan and Dryvax clones used in this study is that TP05 forms plaques twice the size of DPP17 plaques on BSC40 cells. We wondered how this phenotype might segregate amongst the recombinants deriving from either the DTH or DTM crosses. We used the SNPs to determine what fraction of each genome derived from TP05 or DPP17, and plated all of the cloned viruses on BSC-40 cells, at the same time, to determine the average plaque size. The DTH viruses, passaged just once, showed no particular compositional bias, comprising about equal portions ( $50\pm 27\%$ ) of each of the parental viruses (Figure 4.5A). In contrast the DTM hybrids bear a diminished ( $19\pm 11\%$ ) fraction of the genome derived from DPP17 SNPs (Figure 4.5A). Oddly, there seems to be a simple linear relationship between plaque size and the proportion of the genome derived from each parent strain, with larger plaque sizes associated with a greater proportion of TP05-derived DNA (Figure 4.5B). These data suggest that the TP05-derived DNA may confer a selective growth advantage in multiple rounds of culture (i.e. DTM viruses), but one round of growth (i.e. DTH viruses) provides insufficient time or selective pressure to bias the composition of the recombinants.

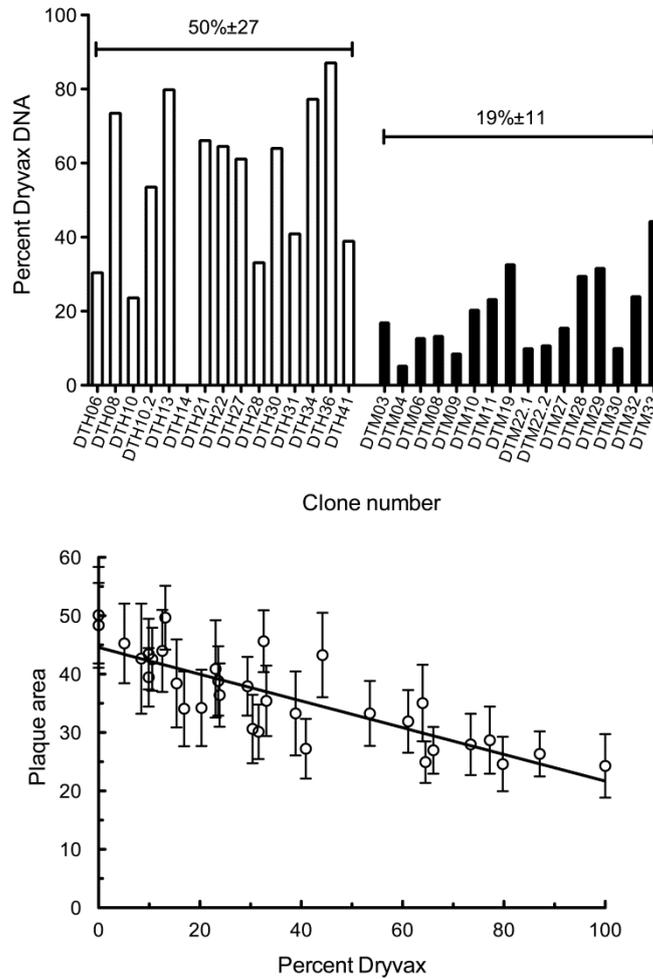


Figure 4.5 Biased selection for sequences associated with the TianTan parent. The percent of DNA derived from each of the parental viruses was determined from the fraction of SNPs derived from each parent. Panel A shows how the composition varied in viruses passaged just once (DTH hybrids) or five times (DTM hybrids) prior to cloning. Passage appeared to select for SNPs linked to the TP05 TianTan parent, as the percentage of Dryvax DNA decreased from  $50 \pm 27\%$  to  $19 \pm 11\%$  with continued passage. Panel B illustrates how the plaque size is related to the genetic origins of the hybrid. The viruses forming smaller plaques, more closely resemble the DPP17 parent. To measure the plaque size, each of the cloned hybrids was plated on BSC-40 cells (in parallel), cultured for two days, stained with crystal violet, scanned, and the plaque area determined using ImageJ. Twenty randomly selected plaques were measured for each virus.

What would produce this effect is not clear. Plaque size is likely determined by many different genetic factors, and there are many differences in the gene composition of the two parent strains. For example DPP17 contains a large deletion in the right TIR compared to TP05 and this deletion bears a number of different genes (Table 4.3). However, this deletion is not completely responsible for plaque variation since DPP25, containing all the genes in this region, forms plaques only slightly larger than DPP17 [223]. We subsequently annotated all of the hybrid genomes using GATU [203] and evaluated the differences (Table 4.4). There are many mutant genes segregating in complex ways between the different viruses including mutant forms of I4L [208], F3L [256], E5R [257], M1L, A51R, and C23L, but no obvious distribution patterns could be discerned by inspection beyond the fact that the more presumably functional genes the virus encoded (Table 4.4, grey cells), the better it grew. The non-transcribed telomeric repeats in VAC telomeres cannot be assembled into contiguous sequences, due to the redundancies in the repeats, but some of the fragments of junction sequences differ enough in TianTan and Dryvax to deduce the origins of the telomeres. This analysis detected a trend suggesting that viruses bearing TP05-like telomeres also form larger plaques. However, all that could really be concluded from this analysis is that TianTan-derived sequences generally contribute greater advantage in culture than does Dryvax DNA.

#### **4.3.5 Large deletions formed through illegitimate recombination**

Poxviruses are also known to suffer deletion mutations during passage. The most extreme example is probably modified vaccinia virus Ankara (MVA), which accumulated six large deletions, and many smaller ones, when it was passaged >570 times in chicken embryo fibroblasts [258]. Over the course of these experiments we did detect one such large deletion mutation when we sequenced clone DTM28.

Table 4.3 Gene differences in the right TIR deletion: TP05 *versus* DPP17

Gene/virus			Nucleotide length (bp)	
Gene (or feature)	Copenhagen	Dryvax	DPP17	TP05
IL-1- $\beta$ -receptor	Cop-B16R	DVX_209	1002	981
Unknown	Cop-B17L	DVX_210	-	1023
Ankyrin motifs	Cop-B18R	DVX_211	-	1725
IFN- $\alpha/\beta$ -receptor	Cop-B19R	DVX_212	-	1056
Ankyrin motifs	Cop-B20R	DVX_213	-	1794

Table 4.4 Gene complement in parent and hybrid viruses

Virus	Plaque area	Telomere	right TIR	F4L	C23L	M1L	F3L	E5R	A51R
DPP17	100%	D <sup>1</sup>	- <sup>3</sup>	-	+ <sup>4</sup>	-	-	-	-
DTH13	101	D	-	+	+	-	-	-	+
DTH22	103	D	-	+	+	-	-	+	-
DTH36	108	D	-	-	+	-	-	-	-
DTH21	111	T <sup>2</sup>	+	+	-	-	-	-	-
DTH31	112	T	+	-	-	-	-	-	+
DTH08	115	D	-	-	+	-	-	-	-
DTH34	118	D	-	-	+	-	-	-	+
DTM29	124	T	+	+	-	+	+	-	+
DTH06	126	D	+	+	-	-	+	+	-
DTH27	131	T	-	-	-	-	-	+	+
DTH10.2	137	D	+	+	+	-	+	-	+
DTH41	137	T	+	-	-	-	-	+	+
DTM03	140	D	-	+	+	+	+	+	+
DTM10	141	D	+	+	+	+	+	+	-
DTH30	144	D	+	-	-	+	+	-	-
DTH28	146	T	+	-	-	+	+	-	+
DTM32	150	D	+	+	+	+	+	+	+
DTM28	156	D	-	+	-	+	+	+	-
DTM27	158	T	-	-	-	+	+	+	+
DTH10	160	T	-	-	-	+	+	+	-
DTM30	162	T	-	+	-	+	+	-	+
DTM11	168	T	+	+	-	-	-	+	+
DTM22.2	175	D	+	+	-	+	+	+	+
DTM09	176	D	+	+	+	+	+	+	+
DTM33	178	T	-	-	-	-	+	+	+

DTM22.1	179	T	+	+	-	+	+	+	+
DTM06	181	T	-	+	-	+	+	+	+
DTM04	186	T	+	+	-	+	+	+	+
DTM19	188	T	+	+	+	+	+	+	+
DTH14	199	T	+	+	-	+	+	+	+
DTM08	204	T	+	+	-	+	+	+	+
TP05	206	T	+	+	-	+	+	+	+

1. D=DPP17-like telomere repeats
2. T=TP05-like telomere repeats
3. "-"=Gene or genes are truncated or deleted
4. "+"=Gene appears intact (gray shading)

The deletion spans 21 kbp and encompasses genes DVX\_201 to 239 (Figure 4.6). In the initial assembly, we found 11 sequence reads that started in gene DVX\_201 and terminated in gene DVX\_239 (Figure 4.6a), along with sequence reads derived from all of the intervening genes. The deletion spans the right TIR boundary, but amongst the reads were some from the unique genes DVX\_202-209 suggesting we had sequenced a stock of virus containing the DTM28 parent contaminated by a virus bearing the deletion (DTM28 $\Delta$ ). To confirm this interpretation of the data we prepared primers 201F and 239R (which are located 21 kbp apart in genes DVX\_201 and DVX\_239, respectively; Figure 4.6b) and used the PCR to detect the novel 1.2 kbp amplicon that was predicted to be formed in this process (Figure 4.6c). We also tested DNAs extracted from the virus stocks that had been archived during the process of passaging these viruses 5 times, before cloning, as well as DTM27, another independent clone that was purified in parallel. Only the purified DTM28 stock contained a virus bearing the deletion (Figure 4.6c) suggesting that DTM28 $\Delta$  arose during the expansion of the stock. Finally we used the 1.2 kbp amplicon to probe a Southern blot of *Nde*I-cut virus DNA, and showed that the DTM28 stock contains viruses contributing a 6.7kbp band characteristic of the deletion-containing fragment as well as a 5.4 kbp band, which derives from the two 5.4 kbp *Nde*I fragments that encode the boundaries of the deletion (Figure 4.6 d). We subsequently sub-cloned this stock and separately isolated the two viruses, confirming the viability of DTM28 $\Delta$  and the fact that none of the deleted genes are essential (Figure 4.6 e) in cell culture. The DNA surrounding the vaccinia virus right TIR boundary is a well-established hotspot for large deletion mutations [246]. The mechanism is probably the same as that which drives the formation of small deletion mutations, starting with the misalignment of regions containing imperfect repeats [171, 223].

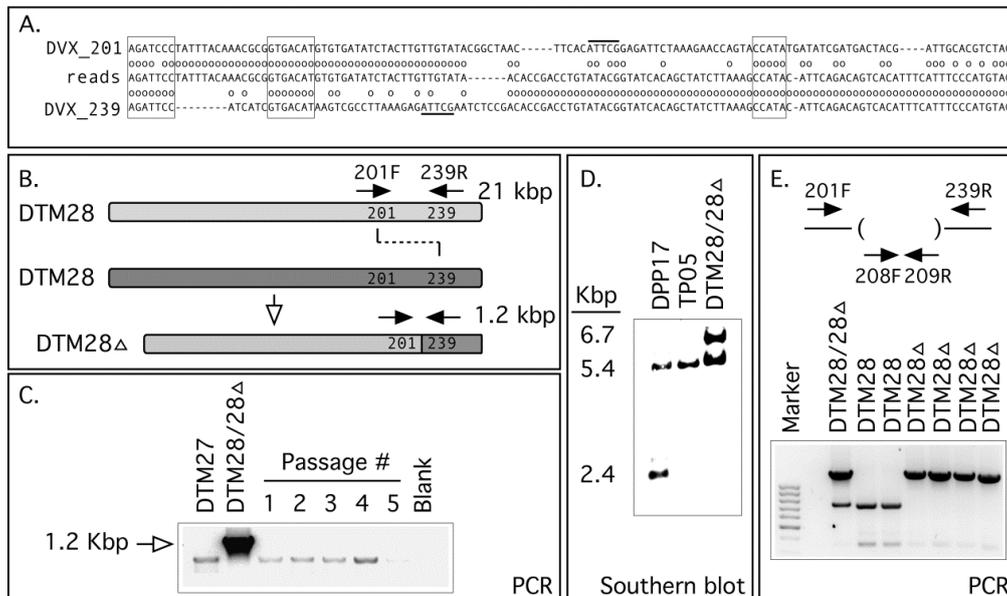


Figure 4.6 Illegitimate recombination detected during the cloning and sequencing of hybrid DTM28

During the sequencing of DTM28, 11 reads were detected linking gene DVX\_201 to gene DVX\_239. Panel A shows an alignment of these reads to sequences within the two genes, which are normally spaced 21 kbp apart. We have also identified sequence identities (circles), short patches of homology (boxed) and a simple repeat (underlined) common to sequences flanking the fusion site. The sequence in these reads transitions cleanly from one gene to the next, with no evidence of any unrelated additional sequences having been added in the process. Panel B showed a way to form this deletion. Illegitimate recombination between identical parents (DTM28) excised 21kbp and created the virus we subsequently called DTM28Δ. Panel C shows the results of a PCR analysis using primers targeting sites flanking the fusion site. These are located too far apart in the parent viruses (e.g. DTM27) to amplify 21 kpb of intervening sequence. The DTM28Δ virus was probably formed during the expansion of the clone prior to sequencing, as it is not detected in intermediary viruses during the course of passages. Panel D shows a Southern blot of *NdeI*-digested virus DNA showing that the sequenced virus stock contained two viruses. These are the DTM28 hybrid (indicated by a 5.4 kbp fragment common to both parent strains), and the DTM28Δ (indicated by a 6.7 kbp fragment containing the fusion junction). Panel E shows the DTM28Δ is independently viable. Six randomly selected viruses were separately plaque purified from the sequenced stock and the PCR was used to detect sequences found only in the deleted region in DTM28 (primers 208F+209R) or only capable of being amplified if the intervening sequences are deleted (primers 201F+239R).

If one aligns the reads spanning the junction boundary between DVX\_201 and DVX\_239, one sees several small blocks of homology (Figure 4.6a, boxed) that could have stabilized the first step in an illegitimate recombination reaction. It is difficult to establish an exact rate by which such mutations arise, but this stock was plaque purified three times, following bulk up and only one virus in 16 DTM viruses passaged in parallel suffered a deletion of this type. This creates a rate of  $\sim 0.06/4$  passages or 1 deletion per 70 passages. The six large deletions introduced into MVA over 570 passages are thus quite consistent with this estimate although, of course, the selection pressures are very different in the two experiments.

#### **4.3.6 Recombination in SFV-reactivated vaccinia viruses**

A third small collection of recombinant viruses was also produced using Shope fibroma virus mediated DNA reactivation assays and DNAs extracted from DPP17 and TianTan strain TP03 [245]. [We used TP03, instead of the TP05 used for the preceding experiments, to test whether viruses could also be produced encoding all three of the large TP03 and DP17 telomeric deletions.] This method relies upon a replicating helper virus (SFV) to rescue or “reactivate” fragments of transfected virus DNA (VACV). The SFV is subsequently eliminated by passage on a cell line that supports only VACV growth. Figure 4.7 shows the maps of the viruses that were recovered by this method. There were just four viruses obtained and two (DTD03 and DTD11) are so similar that they are probably “sibs” sharing a mostly common history. These viruses were too few in number, and the passage history too complicated, to derive much in the way of statistics about recombination patterns, but the pattern of exchanges generally resembled the lesser numbers and longer conversion tracks seen in the DTH viruses. The method did also produce clones bearing the three large telomeric deletions (DTD03 and DTD18, Figure 4.7), which left the virus with TIRs just 7.3 kbp long.

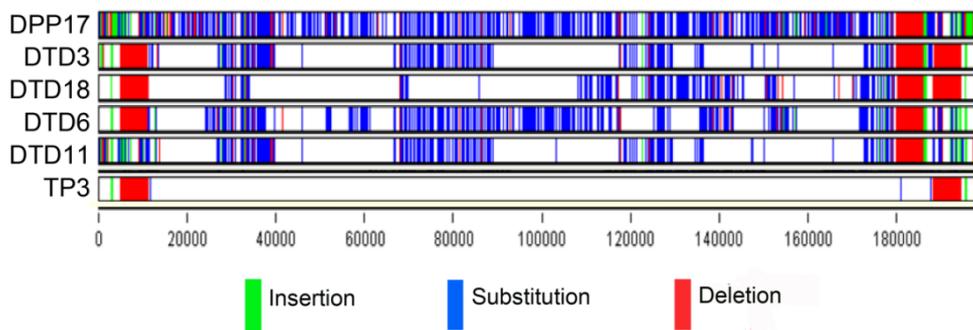


Figure 4.7 Patterns of DNA exchange in recombinant vaccinia viruses produced using Leporipoxvirus-mediated reactivation reactions.

The genome sequences of the DTD recombinant clones were aligned against the parent genomes DPP17 and TP03 using the program “LAGAN”, and edited using the program “Base-by-Base”. Because this experiment used TP03 DNA, and TP05 was always used as the reference strain in all of our analyses (Figure 4.1), the telomeric deletion mutations that differentiate TP05 from TP03 show up as additional red blocks in the TP03 alignment.

An important *caveat* is that the DTD viruses were recovered from cells that had been transfected for a few hours and then incubated for three days, so whether the recombinants were produced during the reactivation stage or during subsequent rounds of re-infection and replication is difficult to deduce.

Overall, there were no strikingly unique features of these reactivated viruses that would differentiate them from any other type of recombinant poxvirus. Perhaps the most important conclusion that could be drawn from this brief study is that this process is very accurate (no mutant viruses were recovered) and no Shope fibroma virus DNA sequences were detected in any of the reactivated VACV. The two most similar genes in SFV and VACV are S068R and J6R, respectively [124], which share only 73% nucleotide sequence identity with no blocks of perfect alignment >17 nt. There is even less similarity between VACV and fowlpox virus, another virus that has also been used to reactivate Orthopoxviruses [259]. This is probably insufficient sequence similarity to support frequent recombination between the helper and reactivated viruses. Additionally VACV hybrids may well be rare and difficult to isolate in the absence of selection, if not simply inviable. Such data support the long-standing suspicion that using a heterologous helper virus, like SFV or fowlpox virus, to reactivate Orthopoxviruses can be done without mutation and does not produce hybrid strains.

#### **4.4 CONCLUSIONS**

Next generation DNA sequencing technologies are greatly improving our understanding of the genome structures and genes encoded by large DNA viruses. Here we show that these methods can also be used to characterize the structures of recombinant poxviruses. These studies show that recombinant VACV are not surprisingly composed of a patchwork of DNA fragments derived from the parent viruses. The numbers of exchanges

varies depending upon the passage history, but if one uses methods like those classically used to produce VACV recombinants (a high multiplicity of infection [10] and one day of co-culture), one detects about 1 physical crossover per 12 kbp in the DTH viruses, a number only slightly higher than the ~8 kbp we have estimated from a review of the older genetic literature. However, there is a lot of noise observed in this number ( $12 \pm 19$  kbp), perhaps explaining why accurate classical recombination maps were never produced for VACV.

Interestingly the lengths of the recombinant patches (i.e. the conversion tracts) are heavily biased towards shorter sizes, something that would favour intragenic recombination. What mechanism would produce such an effect is difficult to identify, although we have previously used genetic methods to show that VACV replication and recombination are intimately linked processes [107, 243], probably because the VACV E9 DNA polymerase exhibits properties characteristic of a recombinase both *in vitro* [109] and *in vivo* [108]. Thus recombination may just be an indirect by-product of virus replication, conceivably associated with the DNA polymerase-catalyzed repair of broken replication structures. Regardless of the mechanism, this process could have interesting genetic consequences for virus evolution, as it would create a lot of diversity *within* recombinant genes, not just diverse combinations of different genes. This becomes of critical importance when one considers the challenge posed to viruses by rapidly evolving responses to biological features like immunodominant epitopes. Short conversion tracks offer a selective advantage for a virus, as they provide a mechanism for rearranging and eliminating peptide epitopes while still retaining gene function.

These studies also show how sequencing could be used to characterize more complex virus traits than those regulated by single genes. Continued passage of the DTM viruses selected for viruses bearing greater proportions of the TianTan genome and this was correlated in some, still unclear, manner

to plaque size. By producing recombinants, applying a selection strategy (perhaps in an iterative manner), and then sequencing clones bearing the desired traits, it should be possible to map genes that collectively regulate the phenotype of interest. This is not a novel approach of course; related methods have been used for decades to map complex genetic traits in many different organisms. However, the widespread availability of next generation sequencing technologies creates a tool that could easily be used by many more laboratories studying gene families and gene interactions in large DNA viruses.

**CHAPTER FIVE - GENOME STRUCTURE OF VACCINIA VIRUS  
REVEALS THE EVOLUTION AND EVOLUTIONARY  
RELATIONSHIPS AMONG EXTANT STRAINS**

Li Qin and David H. Evans<sup>4</sup>

**5.1 INTRODUCTION**

The *Orthopoxvirus* genus encompasses many immunologically related poxviruses, which vary greatly in their capacity to infect different hosts. Of these, cowpox virus (CPXV) probably exhibits the greatest genetic diversity and a broad host range while variola virus (VARV), the causative agent of smallpox, exhibits relatively little genetic variation and naturally infects only humans. The CPXV group comprises at least 5 subtypes [260] and these encode all of the genes present in all other *Orthopoxviruses*. This has led to the suggestion that gene loss has played an important role in *Orthopoxvirus* evolution and that ancestral CPXV-like strains have evolved into all of the modern *Orthopoxviruses* through “reductive evolution” [227, 261].

Large genome deletions with or without genome rearrangements were observed naturally [99] or under certain conditions such as multiple passage through cells [101]. These studies suggest that the regions around TIR are interchangeable; the fragment from the left-most end was inserted into the right end with a simultaneous deletion in the insertion site and causing extended TIRs, or vice versa [97]. Actually, the unique length of genome is shortened if one gene is only counted once, exhibiting examples for gene loss evolution. Furthermore, this kind of translocation was suggested to be responsible for the reduced infectivity in cells [101] or altered pathogenicity due to the passage through different kinds of animals and/or human hosts

---

<sup>4</sup> A version of this chapter is in preparation for publication.

[97]. Different animals and humans were commonly used in the 19<sup>th</sup> century for vaccinia vaccine production. How it might have led to a change from CPXV to VACV over 200 years is still interesting to know.

The relationships between *Orthopoxviruses* are important because in the late 18<sup>th</sup> century, Dr. Edward Jenner showed that humans could be vaccinated against a smallpox challenge using material extracted from a cowpox lesion, a process far safer than the existing practice of variolating with VARV [262, 263]. In the years that followed, and prior to the advent of modern *in vitro* culture methods, the various inocula used as vaccinating agents were distributed around the world while being repeatedly passaged in humans and amplified in animals. This generally involved virus culture on the skin of cows, rabbits, or sheep, but donkeys and chicken eggs were also sometimes used [264]. During this era it is now recognized that passage through different hosts would likely have attenuated these agents, and viruses causing less adverse side effects in humans were also likely to have been retained for use as future stocks [265], but how these interventions might have had affected virus evolution is unknown. Curiously, the virus that was eventually used to eradicate smallpox was the one we now call vaccinia virus (VACV), another *Orthopoxvirus*, but one that is clearly different from the known strains of CPXV. The biological origin of VACV is uncertain, although it has been suggested that a horsepox-like virus (HPXV) was an ancestor because a surviving HPXV genome encodes dozens of extra genes [166]. This hypothesis is supported by Jenner's own report that he obtained his later inocula from an infection in horses called "grease" [263].

During the period when VACV was being grown and distributed for use as a smallpox vaccine, it acquired many different names that reflect the country or health agency involved in its propagation [4]. In many cases these viruses became linked to geographical regions. The New York City Board of Health (NYCBH) strain was originally transported from England in 1856, and was later widely used in America and West Africa. A Russia strain,

EM-63, was also (probably) derived from NYCBH by a circuitous route. Europe used many different strains including Lister, Bern, Paris, Copenhagen (Cop), and Ankara, while the Tian Tan and Ikeda strains were used in China and Japan, respectively (9). Interestingly, viruses related to VACV continue to evolve in wild habitats including South American cattle [266] and Southeast Asian water buffalo [214], and this has led to speculation as to whether these might be escaped human vaccine strains [267, 268]. The historical records suggest that these different strains also varied greatly in virulence, a biological feature of “wild” smallpox vaccines that is still poorly understood.

Some of these VACV strains have been sequenced entirely, and this can provide valuable information that can be used to explore the historical relationships between viruses and virus strains. The most common differences that are detected by these methods in poxviruses are single nucleotide polymorphisms (SNPs) and small insertions and deletions (indels), although these mutations accumulate in such great numbers that they provide few insights into the deeper elements of virus phylogenies. Large deletions and genome rearrangements are also sometimes detected and, because these are less common and not readily obscured by sequence drift, provide a more useful tool for studying deeper features of the evolution and evolutionary relationships between viruses. In some rare cases one can also detect evidence of horizontal gene transfer between related Orthopoxviruses including two CPXV-like genes in VACV strain Lister [269] and a small region of sequence encoding HPXV-like SNPs in a Dryvax subclone (DPP17) [223]. It was at one time common practice to periodically co-cultivate smallpox vaccines with other poxviruses, including VARV [188], so as to “refresh” the vaccine efficacy. Such activities could have provided opportunities for the production of recombinants between related Orthopoxviruses.

Although a virus resembling HPXV is often assumed to be ancestral to VACV, the evolutionary path(s) from CPXV-like and HPXV-like viruses to VACV is unclear. So are the relationships between the many VACV strains. Using data obtained from whole genome alignments between different VACVs, we describe some probable routes by which the extant VACV strains can be related. We also provide some insights into the likely route by which VACV evolved from a hypothetical HPXV-like ancestral strain.

## **5.2 MATERIALS AND METHODS**

### **5.2.1 Viruses and cells**

IHD-W (International Health Department White) stock was obtained from Dr. S. Dales (isolated from IHD-J (International Health Department - Japan)). Lister stock was newly bought from ATCC (VR-1549). WR (Western Reserve), Cop (Copenhagen) and cowpox strains were lab stocks. Viruses were propagated on mycoplasma-free monkey kidney epithelial (BSC-40) cells in modified Eagle's medium (MEM) supplemented with 5% fetal bovine serum, 1% nonessential amino acids, 1% L-glutamine, and 1% antibiotic at 37°C in a 5% CO<sub>2</sub> atmosphere. BSC-40 cells were grown to 80% confluence in 24-well plates and then infected with viruses (Dryvax, IHD-W, WR) at a multiplicity of infection of approximately one PFU/well in 100 µl of phosphate-buffered saline (PBS) for 1 h at 37°C. Then the medium was changed and the viruses were cultured for three days, harvested from wells containing only one plaque. These virus were further cloned twice more by limiting dilution as described above. Clones of DPP25 (from Dryvax), IHDW1, 2, 3 and WR72 were obtained.

### **5.2.2 Viral sequencing, data analysis and annotation**

Each purified virus was amplified (bulked up on) in BSC-40 cells and then purified by centrifugation through sucrose gradients as described

previously [194]. The DNA was extracted and 500ng of each viral DNA was sequenced using a Roche 454 GS Junior system. DPP25, WR72 and three of IHDW clones were sequenced. Different sized contigs were assembled from the raw sequencing data using GS De Novo Assembler software and then CLC Genomics Workbench 6 software was used to complete the assembly of nearly full-length genomes. Conflicts between the reference sequences and our assemblies were resolved by using the PCR to amplify the region of interest followed by Sanger sequencing of the amplicons. Bioinformatics analyses were performed using Viral Genome Organizer [195, 196] and Viral Orthologous Clusters [38, 197] ([www.virology.ca](http://www.virology.ca)). The program LAGAN [198] ([genome.lbl.gov](http://genome.lbl.gov)) was used to produce alignments of multiple genomic sequences and Base-by-Base software [199] was used to check the alignments. To explore the phylogenetic arrangements, 98.8kb of conserved DNA sequences (spanning genes DVX\_058 to DVX\_155) were extracted from the multiple genome alignment and analyzed with the Recombination Detection Program (RDP) [200] using 1000 bootstrap replicates. JDotter was used to make dotplots [270]. The Genome Annotation Transfer Utility (GATU) [203] was used to initially transfer a reference annotation to our viral genome sequences. Artemis [204] was used to visualize and edit the annotation. Table 5.1 lists the accession numbers for the VACV genomes cited in this communication.

## **5.3 RESULTS**

### **5.3.1 Virus isolation, genome assembly and annotation**

Our previous study showed that Dryvax, an old vaccine stock for smallpox, is composed of a number of different strains, which we classified into four sub groups according to genome structure analysis [223]. During our studies of Dryvax, we isolated DPP25 based on a search for a virus resembling Acam clone 3 (CL3) (Table 5.1).

Table 5.1 Identities and accession numbers of the viruses cited in this work

Virus ID	Virus	Source or strain	Accession number
HSPV	Horsepox virus	MNR-76	DQ792504.1
RPXV	Rabbitpox virus		AY484669
CL3	Vaccinia virus	Dryvax (Acambis clone 3)	AY313848
Duke	Vaccinia virus	Dryvax (human isolate)	DQ439815
3737	Vaccinia virus	Dryvax	DQ377945
2K	Vaccinia virus	Dryvax (Acambis clone 2000)	AY313847
LC16m8		Lister	AY678275
LC16m0		Lister	AY678277
VACV107		Lister	DQ121394
GLV-1h68		Lister	EU410304
VACV-LO	Vaccinia virus	Lister-LO	AY678276
TP5		TianTan	KC207811
TP3		TianTan	KC207810
TT12		TianTan	JX489139
TT11		TianTan	JX489138
TT10		TianTan	JX489137
TT9		TianTan	JX489136
TT8		TianTan	JX489135
IHD-W		International health division- White	KC201194
CVA	Vaccinia virus	chorioallantois vaccinia virus Ankara	AM501482
COP	Vaccinia virus	Copenhagen	M35027
WR	Vaccinia virus	Western Reserve	NC_006998
DPP9		Dryvax	JN654976
DPP10		Dryvax	JN654977
DPP13		Dryvax	JN654980

DPP15		Dryvax	JN654981
DPP17		Dryvax	JN654983
DPP20		Dryvax	JN654985
DPP21		Dryvax	JN654986
CPX-GRI	Cowpox virus	Cowpox virus strain GRI-90	X94355

The DPP25-like virus is a very low abundance virus (<1%) in the Dryvax stock, bearing a 1.1kbp fragment (DVX\_214 to 216) which is absent from all other VACVs except HPXV [223]. To further study the genome structure of VACVs, we sequenced the DPP25 viral genome, assembled, annotated and deposited it into Genbank under the accession number of KJ125438. A synthetic genome, DVX (Dryvax) was created and contained the genome of DPP25 (with the longest unique fragments) and DPP15 (with the longest TIRs) in Dryvax clones [223]. This synthetic genome was used to compare DPP25 to other VACV stocks.

Since the published IHD-W sequence (GenBank accession number: KC201194) is missing the right TIR region, we also decided to sequence our IHDW virus. We cloned IHDW strains randomly by limiting dilution three times on BSC-40 cells as described above. Three IHDW clones were then sequenced. Since the sequences of these three IHDW clones are very similar, we annotated only one of them, called IHDW1, and deposited it into Genbank (accession numbers: KJ125439).

### **5.3.2 IHDW**

A notable feature of IHD-W strains is that they form comet-like plaques due to a point mutation (Lys-151-Glu) in the lectin homologue encoded by the A34R gene (DVX\_168) [271]. Indeed, our IHDW strains also show this comet phenotype and encode the causative amino acid mutation, confirming our IHDW virus is of IHD-W origin.

After assembling the IHDW sequencing data, we observed a frame-shift due to 6As instead of 5 in DVX\_176 (Cop-A40R) (Figure 5.1A). To test whether this was a sequencing error, we re-sequenced the viral DNAs using the Sanger method. Consistently, we found the 6As in our IHDW clones compared to 5As in the published IHD-W strain and the DPP25 clone.

```

A: DVX_176 (Cop-A40R) Lectin homolog
DPP25    AATGTATCCATTTATCTACTGATCGAAAAA-CCTGGGAGGAAGGACGTAATGCATGCAA
IHD-W    AATGTATCCATTTATCTACTGATCGAAAAA-CCTGGGAGGAAGGACGTAATGCATGCAA
IHDW1    AATGTATCCATTTATCTACTGATCGAAAAAACCTGGGAGGAAGGACGTAATGCATGCAA
IHDW2    AATGTATCCATTTATCTACTGATCGAAAAAACCTGGGAGGAAGGACGTAATGCATGCAA
                                                *

B: DVX_124 (Cop-D8L) Carbonicanhydrase; GAG-binding IMV
DPP25    ATGTAGTGATGATAACACATATTCATTGGGGAGAAACCCTCCACTTATATAT--CCTCC
IHD-W    ACGTAGTGATGATAACACATATTCATTGGGGAGAAACCCTCCACTTATATAT--CCTCC
IHDW1    ATGTAGTGATGATAACACATATTCATTGGGGAGAAACCCTCCACTTATATATATCCTCC
IHDW2    ATGTAGTGATGATAACACATATTCATTGGGGAGAAACCCTCCACTTATATATATCCTCC
#                                               **

C: DVX_192 (Cop-A56R) Hemagglutinin
DPP25    TTAAGTATAATGTAGAAGATCATAACAGACAC--CGTCACATACACTAGTGATAGCATT
IHD-W    TTAAGTATAATGTAGAAGATCATAACAGACAC--CGTCACATACACTAGTGATAGCATT
IHDW1    TTAAGTATAATGTAGAAGATCATAACAGACACACCGTCACATACACTAGTGATAGCATT
IHDW2    TTAAGTATAATGTAGAAGATCATAACAGACACACCGTCACATACACTAGTGATAGCATT
                                                **

```

Figure 5.1 Gene mutations due to the insertion of repeats found in IHDW clones

Three genes are found to be mutated due to frameshifts by inserting small repeats, supposed to be viral replication errors since functional genes have one copy less of the repeats. Panel A shows gene DVX\_176. DPP25 and published IHD-W contain 5A. However, IHDW1 and 2 bear 6A which would truncate the gene. Panel B show DVX\_124. DPP25 and IHD-W show 3 AT repeats but the IHDW1 and 2 show 4 copies of AT repeats. Panel C indicates DVX\_192. DPP25 and IHD-W show 2 copies of AC repeats instead of IHDW1/2 having three copies. \* indicates the position of the repeats. # shows a point mutation in IHD-W strain in DVX\_124.

A40R, a type II membrane glycoprotein with C-type lectin domain, is nonessential for viral replication and virulence [272]. We also noted another two small insertions in D8L and A56R (Figure 5.1B and C). VACV D8L encodes a virion transmembrane protein, that binds to cell surface chondroitin sulfate and mediates the adsorption of intracellular mature virions to cells [273]. Deletion of D8L attenuates the virus dramatically in mice [274]. Lastly, VACV A56R encodes the hemagglutinin (HA) protein, which is able to bind two other viral proteins, a serine protease inhibitor (K2) [275, 276] and the complement- control protein (VCP) [277] and express them at the surface of infected cells. Interestingly, the IHD-W strain is reported to be a HA-negative due to a two nucleotide (AC) insertion in A56R [278], forming syncytia on infected cells [279]. We observed this mutation in our IHDW strain, but not the published IHD-W (Figure 5.1C). Our IHDW clone is therefore different from the published IHD-W clone based on A40, A56 and D8 truncations. Whether these mutations have some collective influence on the viral growth remains to be investigated.

Previous studies have found that most of the small insertions and deletions (indels) are related to repeats [171, 223]. However, most of these indels cause in-frame changes and it is often hard to identify whether it was an insertion or a deletion [223]. It is assumed that the genes should be functional in origin, thus our IHDW strain would generate small insertions, which result in frame shifts, giving rise to the truncation of genes (Figure 5.1A, B and C). The insertions are likely due to replication errors, suggesting that the DNA polymerase of IHDW strain is very susceptible to increase repeat errors.

When comparing the genome of IHDW to DPP25, dot matrix analysis shows that DPP25 bears a 2kb insertion close to the right TIR (Figure 5.2A). Except for this 2kb difference, the genomes of these two viruses are highly similar.

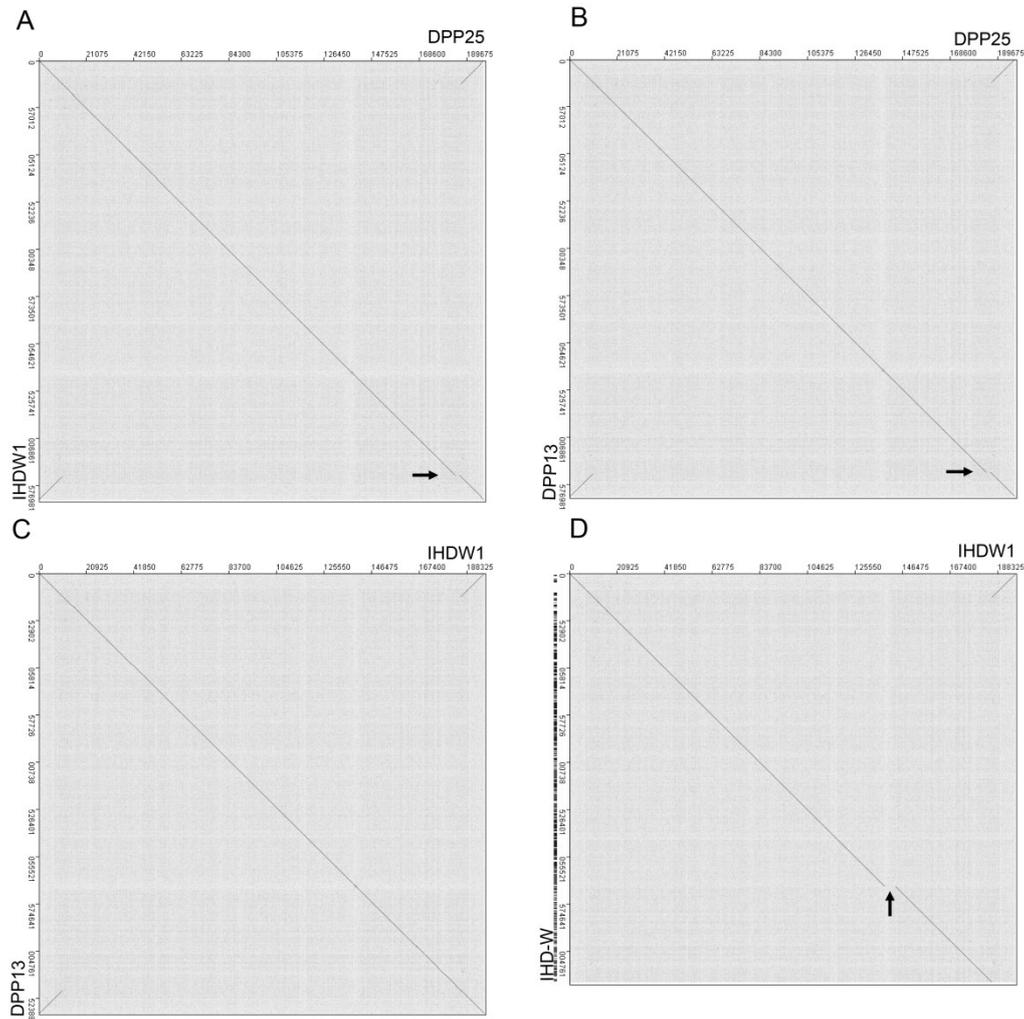


Figure 5.2 The whole genome alignment by dotplots

Panel A shows a dotplot comparing DPP25 to IHDW1. Except for a 2kbp deletion (right arrow), the genome structure is nearly identical. Panel B: DPP25 to DPP13. An exactly identical 2kbp deletion is shown by a right arrow. Panel C: IHDW1 to DPP13. The genome structure is identical. Panel D: IHDW1 to IHD-W. The vertical arrow shows a 4.9kbp deletion within the ATI gene of IHD-W, not IHDW1. Note that the published IHD-W sequence is missing the right TIR.

Interestingly, DPP13, a clone also isolated from Dryvax stock, bears this exact same 2kb deletion (Figure 5.2B). In fact, when compared by dotplot analysis, IHDW has an identical genome structure as that of DPP13 (Figure 5.2C).

The IHD-W virus derived from IHD-J by selection for a hemagglutinin (HA) negative phenotype [278], has the same genome structure as DPP13, a Dryvax clone. The IHDW strain is derived from IHD-J, which was cloned by Dr. Yasuo Ichihashi in Japan from an IHD stock and was first described in 1971 [279] (personal communication with Dr. Hisatoshi Shida). The IHD stock was originally derived by passage of the NYCBH VACV strain through 51 rounds of intracerebral infection in mice followed by 4 passages using chorioallantoic membrane (CAM) culture. Therefore, it is to be expected that IHDW has the same genome structure as DPP13 since Dryvax was also derived from the NYCBH strain. In addition, compared to our IHDW stock, the published IHD-W strain is missing the right TIR sequence and shows a 4.9kb deletion within ATI gene (A-type inclusion) (Figure 5.2D). Further PCR testing confirmed that the ATI gene fragment is present in our IHDW stock. Deletion or fragmentation of the ATI gene is commonly seen in VACVs, such as Copenhagen strain (4.1kbp deletion) [120], buffalopox [214] and Belo Horizonte virus (a Brazilian vaccinia virus-like strain, 4.0kbp deletion) [280], but the length of the deletion and the position vary, indicating that these deletions occur randomly and independently during VACV evolution.

### **5.3.3 Genome comparison among VACVs**

#### **5.3.3.1 HPXV to DPP25**

DPP25, whose genomic structure is identical to CL3, has the second longest VACV genome after HPXV. Since genome reduction theory [227] suggests that gene loss may play an important role in poxvirus evolution, we

hypothesize that a DPP25-like virus represents an early intermediate in VACV evolution, one that likely derived from a HPXV-like virus. In an effort to track the origin of VACV, we decided to analyze and compare the DPP25 genome structure to viruses found in other stocks of HPXV, Lister, WR, IHDW, CVA, TianTan and Cop strains.

HPXV and DPP25 shared the same gene content and gene order between DVX\_014 (growth factor) to DVX\_212 (Cop B19R) (Figure 5.3 A). There are differences in individual indels and/or SNPs, which may affect specific genes, but this is expected due to random sequence drift. Interestingly, compared to HPXV, DVX\_010 to 013 are present in both the right and left terminal regions of DPP25, suggesting a transposition occurred at some point in the virus history. In addition, DPP25 bears a 5.5kbp deletion close to the right TIR and a 10.7kbp deletion close to the left TIR compared to HPXV.

Taken together, our findings suggest that DPP25 has a close relationship to HPXV in terms of genome structure. Given the hypothesis that a HPXV-like virus might be ancestral to other VACVs, this suggested that a DPP25-like virus might have been an intermediate in the VACV evolutionary tree. In order to model additional steps in VACV evolution, we therefore decided to compare the genome structure of DPP25 to all of the other VACVs available today.

### **5.3.3.2 Comparing DPP25 to Cop and RPXV**

The Cop strain was the first VACV genome to be sequenced [120]. Compared to DPP25, its genome exhibits three deletions; 3.7kbp, 4.1kbp and 5.7kbp respectively (Figure 5.3 D). These findings suggest the Cop strain is a distant relative of DPP25 and has undergone several genome re-arrangements during its evolution.

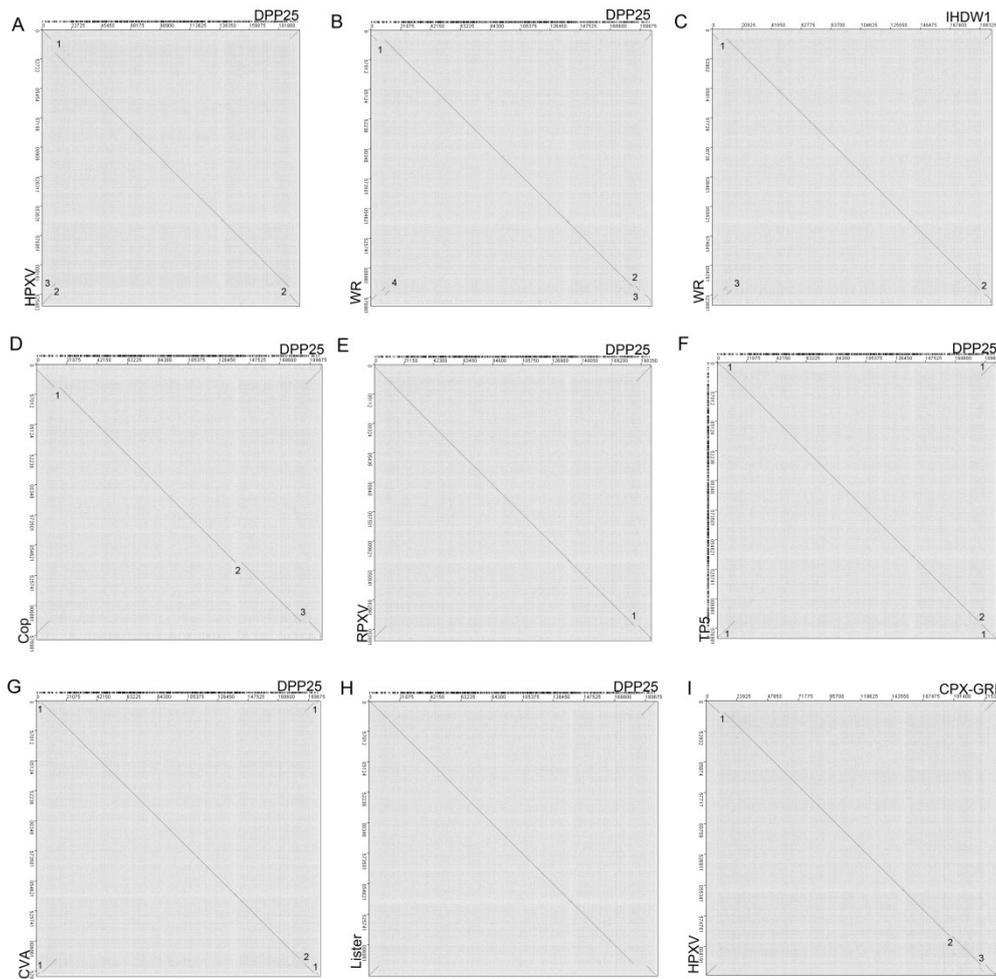


Figure 5.3 The whole genome alignment of VACVs by dotplot analysis

Panel A: A dot-matrix plot comparing DPP25 to HPXV. Please note, all the DNA fragments (except HPXV) are named according to the synthetic genome DVX gene numbering system. Number 1 indicates a 10.7kbp deletion relative to HPXV. Note that gaps are found on both viruses in this region. One comprises the 10.7kbp fragment in HPXV. Another is a fragment containing DVX\_010 to 013 and is found only in the HPXV right end (number 3). The number 2 indicates a 5.5kbp fragment absent in both DPP25 TIRs. Panel B compares DPP25 to WR. Number 1 shows a fragment absent in WR (6.1kbp, DVX\_007-013). Number 2 is the 2kbp fragment absent in WR. Number 3 shows gaps found in two viral genomes. DPP25 lacks a fragment encoding DVX\_018-014 and WR is missing a fragment containing DVX\_013 to 007. Number 4 shows a translocation; the fragment DVX\_018-014 translocates with DVX\_013-012 in WR. Panel

C compares IHDW1 to WR. Except for the 2kbp deletion shown in panel B number 2, all other changes are the same as the panel B, since except for this 2kbp fragment, the genome structures of DPP25 and IHDW are the same. Panel D compares DPP25 to Cop. Number 1: A 3.7kbp (DVX\_016-025) deletion in Cop. Number 2: A 4.1kbp deletion (DVX\_157-9, ATI gene) in Cop. Number 3: A 5.7kbp deletion (DVX\_213-6, DVX\_013-012) in Cop. Panel E compares DPP25 to RPXV. Number 1: A 3.3kbp (DVX\_214-6, DVX\_013) deletion in RPXV. Panel F compares DPP25 to TP5. Number 1: In TP5, a 59bp repeats (500bp) replaces DVX\_004-6 (1.2kbp) in DPP25. Number 2: A 2.4kbp fragment in DPP25 (DVX\_214-6) is replaced with a 2.3kbp fragment in TP5 (DVX\_015-014). Panel G compares DPP25 to CVA. Number 1: A 1.8kbp deletion (DVX\_003 -005) within both TIRs of CVA. Number 2: A 3.4kbp fragment (DVX\_213-6) in DPP25 is replaced with a 3.6kbp segment in CVA (DVX\_019-014). Panel H compares DPP25 to Lister. A 9.3kbp fragment in DPP25 (DVX\_211-6, DVX213-009) is replaced with a 1.7kbp fragment in Lister (Lister 195-6 or CPX-GRI 208-9). Panel I compares CPX-GRI to HPXV. Number 1: A 4kbp deletion (CPX-GRI D8L to D11L) in HPXV; number 2: A 1.6kbp deletion (CPX-GRI B8R to B9R) in HPXV; and number 3: A 3.6kbp deletion (CPX-GRI K1R to T1R) in HPXV.

Interestingly, when comparing RPXV to DPP25, RPXV differs from DPP25 by a single 3.3kbp deletion (Figure 5.3 E), suggesting a closer evolutionary relationship between DPP25 and RPXV.

### **5.3.3.3 Comparison of DPP25 to Lister, CVA and TianTan**

Unlike Cop and RPXV, the comparisons between TianTan (Figure 5.3 F), CVA (Figure 5.3 G) and Lister (Figure 5.3 H) to DPP25 are more complex. We found one region in the right hand terminus bearing two different sequences in each dot matrix plot. When comparing DPP25 to CVA, we observe a 3.6kbp fragment (DVX\_019-014) being replaced by a 3.4kbp (DVX\_213-6) fragment in DPP25. Similarly, when comparing DPP25 with Lister, we observe an expansion of a 1.7kbp fragment into a 9.3kbp region in DPP25. Finally, comparison of DPP25 with TianTan showed a 2.3kbp substitution with a 2.4kbp in DPP25. Our results suggest that Lister, CVA and TianTan illustrate examples of terminal re-arrangements during evolution from a DPP25-like ancestor.

### **5.3.3.4 Comparison of DPP25 to WR**

WR is the accepted reference strain for VACVs. Compared to DPP25, dot matrix analysis showed a 6.1kbp (DVX\_007-013) deletion in the TIR region and 2kbp deletion in right end of the WR genome. In addition, WR encodes a translocation that occurred between DVX\_018-014 and DVX\_013-012 (Figure 5.3 B). Interestingly, IHDW exhibits an almost identical genomic pattern compared to WR except for the 2kbp deletion (Figure 5.3 C). In fact, WR is the only VACV analyzed to date that has this translocation. As such, using WR as the VACV reference strain can complicate comparisons between VACV strains and it is unfortunate that it has been chosen as the VACV reference strain.

Based on our whole genome analysis of VACV strains we concluded that the structure of the central part of the VACV genome is conserved while

the TIR is more variable in gene organization (Figure 5.3). In fact, the terminal regions of VACVs appear to be unique to different strains of VACV. However, using genomic structure comparisons alone we were unable to get a clear answer on the evolutionary path between HPXV-like, DPP25-like and modern VACV strains.

#### **5.3.4 VACV phylogenetic tree**

Having compared the genome structure of various VACV strains, we next sought to generate a phylogenetic tree to further map VACV evolution. Our recent study detected a patchy distribution of polymorphic sites across the different genomes among Dryvax clones, showing that no single virus can represent the whole viral stock [223]. Phylogenetic tree analysis produces variable outputs according to the methods used to build the trees (least squares, neighbor joining or maximum likelihood), even if the same genome alignment is analyzed [246]. Therefore, we used as many as clones as possible from each group to assemble the trees using the same genomic alignment spanning a 100kbp conserved region from DVX\_058 to DVX\_155 (Cop F9L to A24R). GLV-1h68, a clone isolated from an old Lister stock, LIVP, had been modified to insert a Renilla luciferase-green fluorescent protein fusion cassette into F14.5L and a  $\beta$ -galactosidase into the J2R locus [169]. To build the tree, we deleted these two insertions from the reported sequence. WR72 was a clone from our laboratory stock of VACV strain WR.

Different methods produced very similar phylogenetic trees (Figure 5.4). All three phylogenetic methods correctly assigned viruses into related groups, such as Dryvax, TianTan, WR, Ankara, Lister, IHDW. HPXV maps separately from all of the other VACVs, consistent with the possibility that it may closely resemble the first generation of VACV while having the longest genome among vaccinia species. Two subgroups map downstream of HPXV.

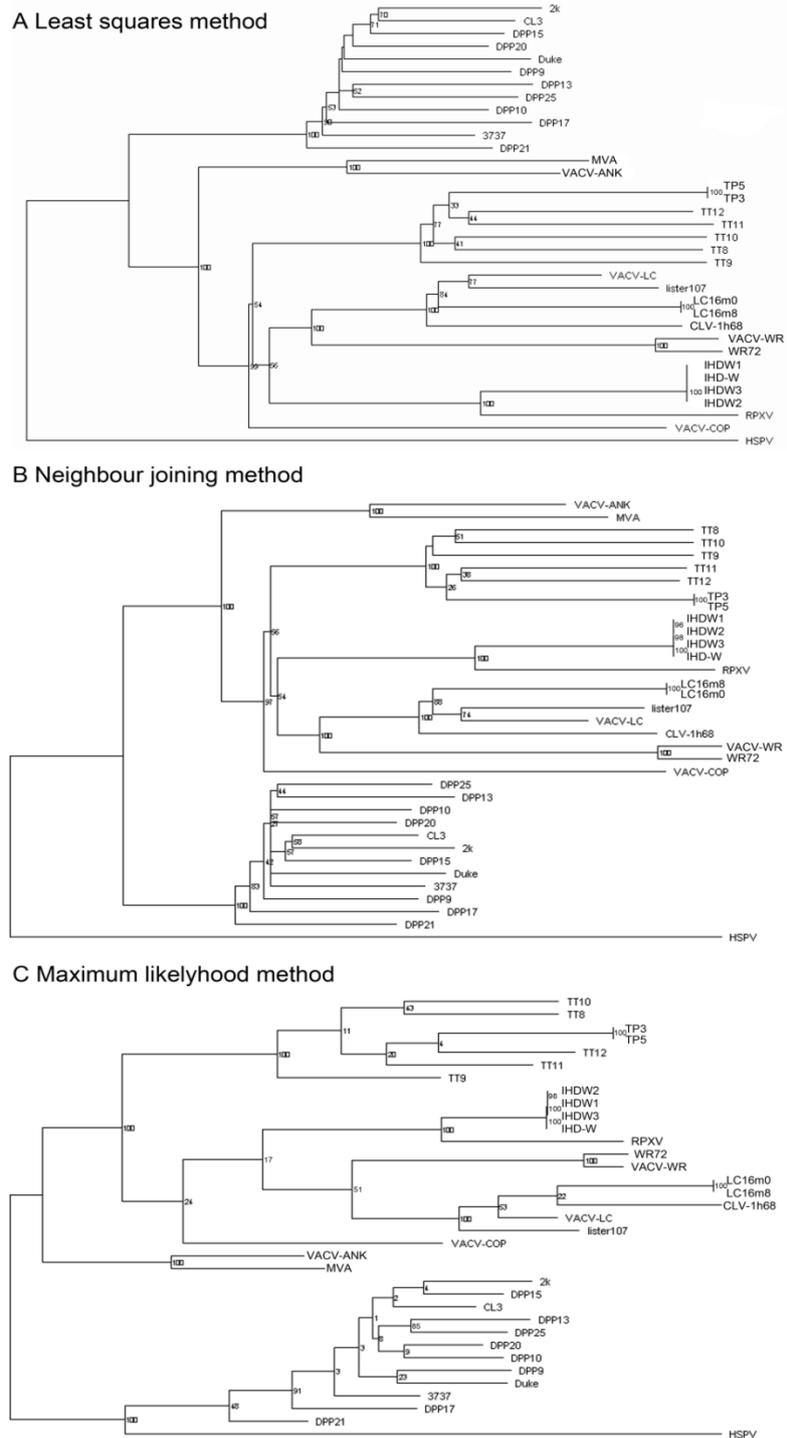


Figure 5.4 Phylogenetic relationships between all vaccinia virus clones

A multiple alignment was compiled using sequences encoding the core region of the different genomes spanning DVX\_058 (F9L) to DVX\_155 (A24R). The alignment, three approaches, and a thousand bootstrap replicates were used by the program RDP [200] to create the plots shown here.

One is an American cluster, including all the Dryvax clones; another is the European/Asian, bearing all the remaining VACV strains. This makes sense since Dryvax originated from a seed stock supplied to the New York City Health Department in 1856 and likely evolved independently from European strains for more than 100 years. The exceptions to this rule are the WR and IHDW strains, which (like Dryvax) are both derived from New York City Board of Health; however, they do not group with Dryvax in the “American” cluster.

All three phylogenetic analyses point to a HPXV-like virus as being the most likely parental strain of all other VACVs. However, none of the trees indicated that DPP25-like strains are an intermediate between HPXV and all other VACV. In fact, none of the trees identify an intermediate virus. Our phylogenetic analysis required the comparison of a highly conserved region of the VACV genome. We (and others) had previously noted that the central genome of VACVs is highly conserved while the terminal regions show the most diversity. Perhaps the central genome is under stringent constraint for functionality and survival, and is therefore a poor region for mapping evolutionary relationships. Based on these results, we decided to continue our efforts by concentrating our analysis on the terminal boundaries of VACVs as markers for evolution.

### **5.3.5 Three shared features in the genome structure of all VVs**

Since very little variation is found within the conserved central domain, but the TIRs show distinct differences, we focused our analysis of VACV evolution on a careful examination of the TIR boundaries.

Having observed distinct TIR differences between DPP25 and HPXV, we analyzed how the TIR boundaries compare amongst the various phylogenetic groups of VACVs. Performing a blast-search of HPXV 005 to 019 (16932nt) (Figure 5.5 A) revealed that all VACVs share a 10.7 kbp deletion in the left TIR.

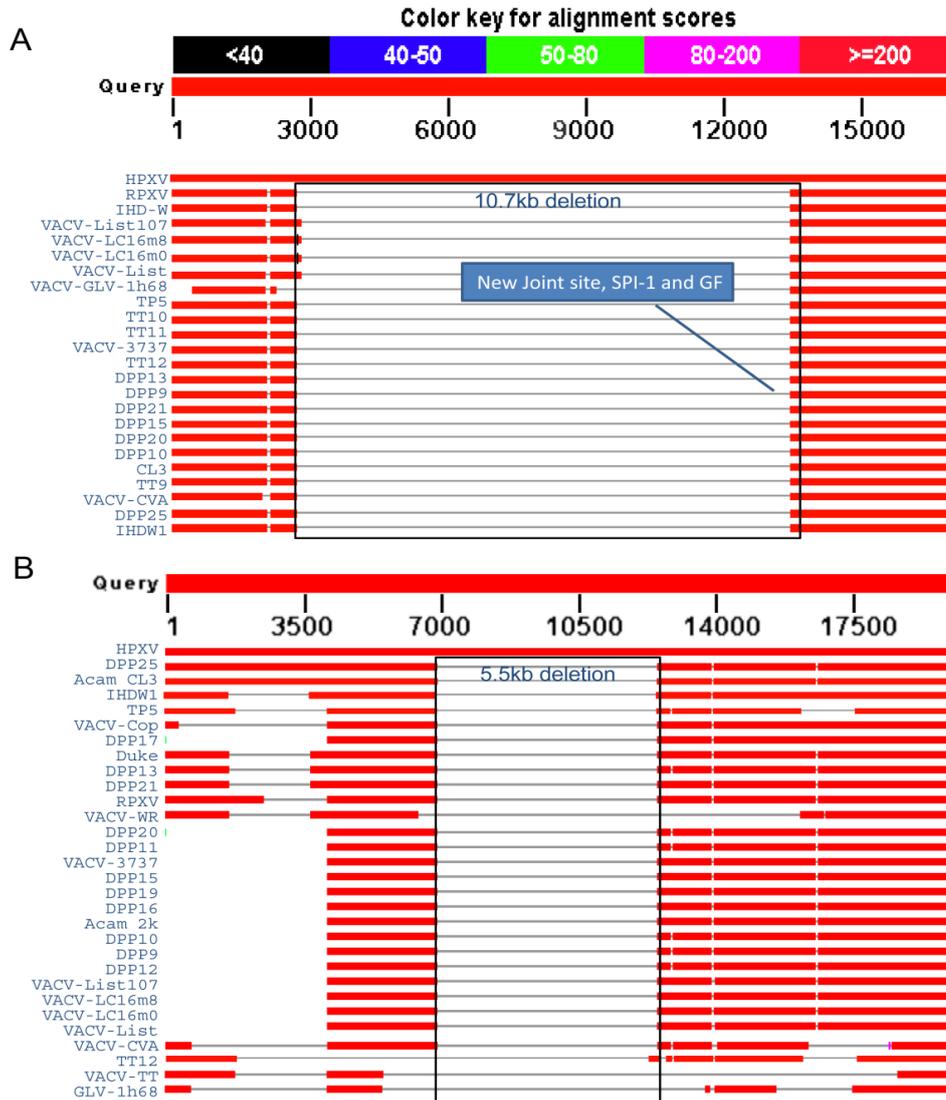


Figure 5.5 BLAST alignments comparing the HPXV left and right telomere sequences to VACV telomeres

Panel A: The left end of HPXV (orf 005-019) compared to VACV telomeres. All VACVs show a 10.7kb deletion. The right boundary of this deletion marks the position of the growth factor gene (GF). In VVs, the growth factor gene is linked to the SPI-1 (serpin-1) gene due to a translocation of DVX010 to 013 from the right end of the HPXV genome. We also note that the four Lister strains, encode a bit more DNA to the left of the deletion. Further analysis indicated that in the left TIR of Lister, the 10.7kb deletion boundary is much the same as other VVs. However, Lister appears to have captured a homolog of the CPXV CrmE gene and a bit of this insert shows up in the alignment. Panel B: The right end of HPXV (orf 196-206) compared to VACV telomeres. A 5.5kbp deletion is shared by all VACVs, except for TT12, a clone from a TianTan stock, which delete DVX\_013 to 011(3kbp) from the TIR. Note this 5.5kbp deletion truncates only one gene, HPXV200, the longest gene in HPXV (5763nt) and the left fragment forms DVX\_011.

The 10.7kbp deletion has a left boundary located next to the TIR of HPXV and CPXV (DVX\_009, Cop-B22R) and a right boundary corresponding to the growth factor gene (DVX\_014, Cop-C11R). The sequence encoding DVX\_010 to 013 is present within both of the VACV TIRs while HPXV/CPXV contains this fragment of DNA sequence at only the right end of the genome.

It has been suggested that this arrangement results from a transposition of DNA from the right to the left end of the HPXV genome [166]. However, since DVX\_001 to 009 are within the HPXV TIR, it is more likely that the entire region, from DVX\_013 to 001 and corresponding to HPX196 to the end of the genome, transposed from the right to left end of HPXV. This therefore results in a new joint site connecting the SPI-1 gene (DVX\_013) and the growth factor gene (DVX\_014) (Figure 5.5). We did note that the four Lister strains encode additional small DNA sequence downstream of the deletion. Lister has captured two genes from a CPX-like virus, and a small part of this insert (67nt) shows up in the alignment.

Next, we used blast analysis to compare the right end of the HPXV genome (HPX196 to 206 (19984nt) to VACVs (Figure 5.5 B). We observed a conserved 5.5kbp deletion (within one gene of HPX200), that is shared by all VACVs. We noticed that a specific TianTan stock, TT12, has a slightly extended deletion boundary, as TT12 contains a deletion of DVX\_013 to 011(3kbp) within its TIRs. However, another TianTan clone, TP5, which has the longest genome of all TianTan clones, contains the conserved 5.5kbp deletion. We assume that TT12 is a daughter strain that derives from a TP5-like virus within the TianTan stock and that a further deletion within the DVX\_013-011 region made this area more variable (less stable), therefore resulting in a 200bp fragment extending into the deletion. We therefore conclude that the 5.5kbp deletion is also shared by all VACVs, and this differentiates them from HPXV-like viruses.

The three conserved genomic features of VACVs (the 10.7kbp and 5.5kbp deletions, the translocation of DVX\_010 to 013, forming a new joint site connecting SPI-1 and growth factor genes) suggest that there might once have existed an intermediate strain, which derived from a HPXV-like virus, but also serves as the ancestral strain for all other VACVs. A virus with a DPP25-like genome, bearing the longest genome in all other VACVs, might have represented this hypothetical ancestral strain. Of course it is important to realize that the virus we hypothesize is one that existed hundreds of years ago and is not found any longer in modern stocks of virus.

### **5.3.6 The relationship of extant VACVs, the right TIR boundary as an evolutionary feature**

We have noted that all VACVs have the same gene content and order starting with DVX\_001 (Cop-C23L) and ending with DVX\_211 (COP-B18R) (Figure 5.3, Table 5.2). Exceptions exist in the form of individual gene deletions, for example, the Cop strain has a 3.7kbp deletion (DVX\_016 to 025) and a 4.1kb deletion (DVX\_158-159, within the ATI gene), the Ankara strain has a 1.8 kbp deletion (DVX\_003 to 005) while WR has a 6.1kbp deletion (DVX\_007 to 013). In contrast, the length of each VACV TIR is different. Therefore, we chose to focus on the right TIR boundary, which determines the length of TIR, and the whole length of unique genes, to characterize various VACV strain relationships. To simplify the comparison, we use DVX to unify the gene names. Table 5.2 summarizes the components of TIR boundaries, showing that clones from the same groupings, either from phylogeny or analysis of deletions, share TIR boundary features.

In the TianTan strains the TIR boundary terminates at 76% of DVX\_213 N-terminal to DVX\_015 (Table 5.2).

Table 5.2 Properties of the right TIR boundaries of vaccinia viruses

Strain	DVX_211 (Cop- B18R)	DVX_212 (Cop- B19R)	DVX_213	DVX_214 -6 <sup>1</sup>	TIR
DPP25	100%	100%	100%	100%(Fra g) <sup>1</sup>	DVX_013-001
RPXV	100%	100%	100%	3% N-	DVX_012-001
TP5	100%	100%	76% N-	0	DVX_015-001
TT12	100%	100%	76% N-	0	DVX015,014, 010-001 <sup>6</sup>
DPP13	100%	100%	69% N-	28% C-	DVX_013-001
IHDW1	100%	100%	69% N-	28% C-	DVX_013-001
CVA	100%	100%	28%N-	0	DVX_019-006, 002-001 <sup>5</sup>
Cop	100%	100%	16%N-	0	DVX_011-001
GLV-1h68	100%	100%	28%N-	0	DVX_019-013, 008,003-001 <sup>4</sup>
Lister	71% N-	0	0	0	CPX-GRI <u>208</u> <sup>2</sup> , <u>209</u> <sup>2</sup> , 210-214
DPP15	100%	72%N-	0	0	DVX_026-001
WR	100%	100%	69% N-	28% C-	<u>DVX_013</u> <sup>3</sup> , <u>012</u> <sup>3</sup> , <u>018-014,006-</u> 001

<sup>1</sup> Similar to CPXV-GRI B19R, but truncated into three pieces (DVX\_214-6) in DPP25. The B19R protein belongs to BTB/Kelch family associated with cullin-3 based E3 ubiquitin ligase.

<sup>2</sup> CPX-GRI 208 and 209 of Lister strain are not located in TIR.

<sup>3</sup>DVX\_013 and 012 of WR is not within TIR, but still present downstream of DVX\_216.

<sup>4</sup>GLV-1h68 has further deletions in its TIR, a 2.6kbp (DVX\_009-012) and a 1.7kbp (DVX\_004-007) fragments.

<sup>5</sup> CVA has an extra 1.8kbp (DVX\_003-005) deletion in its TIR.

<sup>6</sup>TT12 has an extra 3kbp deletion (DVX\_011-013) in its TIR.

Our previous study showed that out of 25 Dryvax clones, 13 clones have the DPP15-like (Acam 2k-like) TIR boundaries terminating at 72% of DVX\_212, while 9 clones have the DPP13-like (Duke-like) TIR boundary terminating at 69% of DVX\_213 (Table 5.2) [223].

Lister strains terminate their TIR most prematurely of all VACV at 71% of DVX\_211 (Table 5.2). Lister encodes two genes, List195 (CPXV-GRI-208, K3R, encoding the CrmE TNF receptor that contributes to virulence) [269] and List196 (CPXV-GRI-209, T1R, an apoptosis inhibitor, encoding a Golgi anti-apoptotic protein, also contributes to virulence) [163], that are both absent from all other VACVs, except the USSR and Evans strains [269].

Most Lister strains bear these two genes close to the right TIRs. However, GLV-1h68, a clone isolated from an old Lister stock, LIVP, is missing List195 and List196 [169]. Table 2 shows that GLV-1h68 strain shares its TIR boundary with the CVA strain, both terminating at 28% of DVX\_213, suggesting Lister and CVA have a shared evolutionary relationship (Table 5.2). The fact that the LIVP stock does contain viruses encoding List195 and 196 homologues, suggests that this old Lister stock still comprises a mixture of viruses [169]. Therefore, we group CVA and Lister together. We also tested another Lister stock (ATCC, VR-1549), but were unable to find any viruses with GLV-1h68 right TIR boundary characteristics, indicating the GLV-1h68-like viruses may have been lost from this Lister stock.

DPP13, IHDW and WR all share the same 2kb deletion compared to DPP25 (Table 5.2, Figure 5.6), suggesting that all these viruses might be derived from an ancestor virus. In fact, DPP13 and IHDW1 have identical genome structures (Figure 2 C). Analysis of TIR boundaries reveals that WR, IHDW and DPP13 all terminate at 69% of DVX\_213 (Table 5.2). We therefore group DPP13, IHDW and WR together. The fact that both WR and IHDW are neurovirulent in mice adds support to our conclusion [281].

```

DPP13  TATACAACATAACTTCACTCAGATTGCTAAGTACTTATTAGATCGAGGAGCTGATATA-----
IHD-W  TATACAACATAACTTCACTCAGATTGCTAAGTACTTATTAGATCGAGGAGCTGATATA-----
WR     TATACAACATAACTTCACTCAGATTGCTAAGTACTTATTAGATCGAGGAGCTGATATA-----
IHDW1  TATACAACATAACTTCACTCAGATTGCTAAGTACTTATTAGATCGAGGAGCTGATATA-----
DPP25  TATACAACATAACTTCACTCAGATTGCTAAGTACTTATTAGATCGAGGAGCTGATATATCATTAAAGACAGACGATGGTAAAAC
CL3    TATACAACATAACTTCACTCAGATTGCTAAGTACTTATTAGATCGAGGAGCTGATATATCATTAAAGACAGACGATGGTAAAAC

DPP13  -----GTCGTACCCAACACATTGATTATACATCAGTACATACAGTAAATAGCATAGATA
IHD-W  -----GTCGTACCCAACACATTGATTATACATCAGTACATACAGTAAATAGCATAGATA
WR     -----GTCGTACCCAACACATTGATTATACATCAGTACATACAGTAAATAGCATAGATA
IHDW1  -----GTCGTACCCAACACATTGATTATACATCAGTACATACAGTAAATAGCATAGATA
DPP25  .....GAATCAATACATATACATTATAGGCGGTCGTACCCAACACATTGATTATACATCAGT-----AAATAGCATAGATA
CL3    .....GAATCAATACATATACATTATAGGCGGTCGTACCCAACACATTGATTATACATCAGTACATACAGTAAATAGCATAGATA

```

Figure 5.6 A 2kbp deletion shared by VACV strains DPP13, IHD-W, IHDW1 and WR

The upper panel shows the alignment of the left boundary of this deletion. The bottom panel shows the right boundary. The conservation of this deletion in so many different strains indicates a relationship between them. We group these strains together. Note DPP25 has a 9nt small deletion. Please note the full 2kbp sequence is omitted, shown by “.....” in DPP25 and CL3.

WR is derived from the NYCBH strain, which is also the parent strain of Dryvax and IHDW. Compared to DPP25, a distinguishing feature of the WR genome is a translocation found in the right end of the genome (Figure 5.3 B). WR has a 6.1kbp deletion (DVX\_013-007) in its TIRs. It also has an altered gene order from DVX\_018 to DVX\_014, downstream of DVX\_012. For all other VACVs, DVX\_018 will follow genes 017 to 016 till 001 in a descending order. Therefore, WR exhibits a translocation between DVX\_013,012 and DVX\_018 to 014. This suggests that a genome rearrangement created a new connection of the fragment DVX\_018-014/006-001 following DVX\_012. WR has a shorter genome (further deletion of DVX\_013-007), suggesting that WR could be derived from DPP13 or IHDW-like strains.

Based on our analysis of VACV TIRs, we propose that the TIR boundaries can be used as an additional important marker of VACV evolutionary groups. This is, of course, not the only marker that should be used to differentiate VACV strain. In reality, it is a collection of different genomic features, that help define a particular strain. According to our scheme, an unknown VACV could be analyzed for characteristics of its TIR right boundary, and assigned to a particular grouping with reasonable confidence. Nevertheless, the additional sequence data from elsewhere in the genome is required to fully identify the particular strain.

## **5.4 CONCLUSIONS**

In this study, we provide a way to explore the evolution and evolutionary relationships among VACVs one that differs from the traditional method based on phylogenetic analysis. Through the comparison of whole genome sequence, we confirmed that the central genome of VACV strains is highly conserved while most of the major variation across strains occurs at the termini. When examining the two terminal regions we observed

three distinct features conserved in all VACV strains. We also found that the right TIR boundary is highly specific to different VACV strains. Therefore, we define each group of VACVs according to their right TIR boundaries. This type of analysis has led us to the discovery that the Lister strain co-evolved with CVA while WR co-evolved with Dryvax and IDHW.

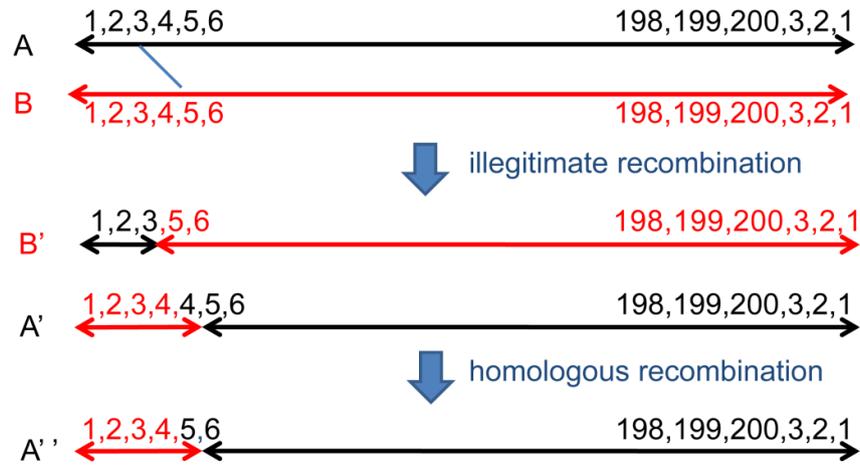
Based on our study, we hypothesize that DPP25 may be a relic example of an intermediate virus between HPXV-like and all the other extant VACV strains. DPP25 was first isolated from a stock of Dryvax vaccine. Judged by the right TIR boundary, Dryvax clones are grouped into four distinct populations [223]. The dominant strains in Dryvax are DPP15-like (60% in Dryvax stock) and DPP13-Like (40%). Both the DPP25-like virus and DPP17-like virus, a virus with a larger deletion in its genome and possibly a novel component of a Dryvax vaccine, make up less than 1% percent of all viruses in this stock [223]. Theoretically, a DPP25 virus can generate a DPP13-like virus by deleting a 2kbp fragment of DVX\_213 to 216. A DPP15-like virus can be produced by the rearrangement of the right TIR of the DPP25 or the DPP13-like virus. It seems unlikely that DPP13 or 15 could generate DPP25. We can imagine a similar process occurring in other VACVs strains. For example, when a DPP25-like virus deletes a 3.3kbp fragment, containing DVX\_214 to 216 and DVX\_013/229 in its right TIR boundary, a RPXV-like virus would be generated with the TIR ranging from DVX\_012 to 001 (Table 5.2). The same process could have taken place for the Cop strain, involving a 5.7kbp deletion containing DVX\_213 to 216 till DVX\_013-012 in a DPP25-like virus or a DPP13-like virus (3.7kbp). While it is easier to explain genome rearrangement resulting in a virus with a shortened TIR, we did find viruses with extended TIRs. For example, DPP15 has the longest TIR ranging from DVX\_026 to 001. Our previous study has suggested that to generate this virus, a fragment of DVX\_026 to 014 (most likely the end of the genome) from the left end of DPP25/DPP13-like virus was translocated to the right end at some point in

the viruses evolution, connecting to DVX\_212 by illegitimate recombination [223].

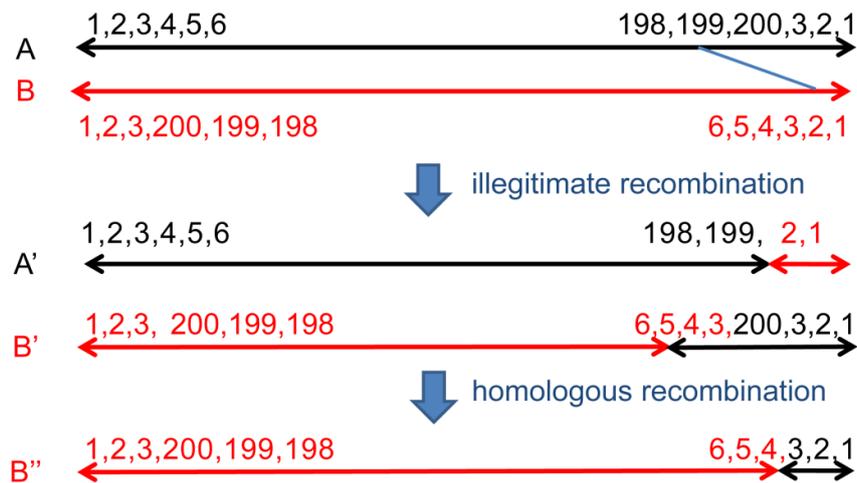
#### **5.4.1 Evolution model**

Based on all the data shown here, we propose an illegitimate recombination model to explain the origin of large VACV genomic deletions, and changes in the lengths of the TIRs (Figure 5.7). To simplify the model, all three examples begin with two sister strains A (Black) and B (Red), although it is important to note that a simple deletion could involve just a single viral genome. Part I shows how an illegitimate recombination causes genome rearrangement. Viruses A and B encode ORFs numbered 1-200 and 1-3 are within the TIRs. An illegitimate recombination event takes place between non-homologous regions, here between orf3 to orf5, generating virus B' missing gene 4 and A' having two copies of gene 4. However, such a large repeat in VACV is unstable. Therefore, the most probable outcome would be that the A' virus would delete one copy of gene 4 by homologous recombination and return back to a virus with a genome structure like wild type virus A''. Or the A' virus is simply lost from the viral population. Part II shows how an illegitimate recombination event can generate a virus with a large deletion and shortened telomere. This time, the illegitimate recombination step is located within the boundaries defined by gene 199 and gene 2. The progeny virus, A', acquires a large deletion from gene 200 to gene 3 in its right TIR, leaving a TIR containing just two genes. For viral B', gene 3 is triplicated and homologous recombination would delete extra copies and leave one copy of gene 3. The virus reverts to wild type or it is deleted from the viral population. Part III shows the mechanism for producing a large deletion with extended telomere. The illegitimate recombination reaction maps within gene 199 to gene 5. Virus A' will retain genes 1 to 5 within its TIR but lacks gene 200. Note that gene 4 and gene 5 are translocated from left end to right end of the genome.

Part I: A large deletion



Part II: A large deletion with shortened telomere



Part III: A large deletion with extended telomere

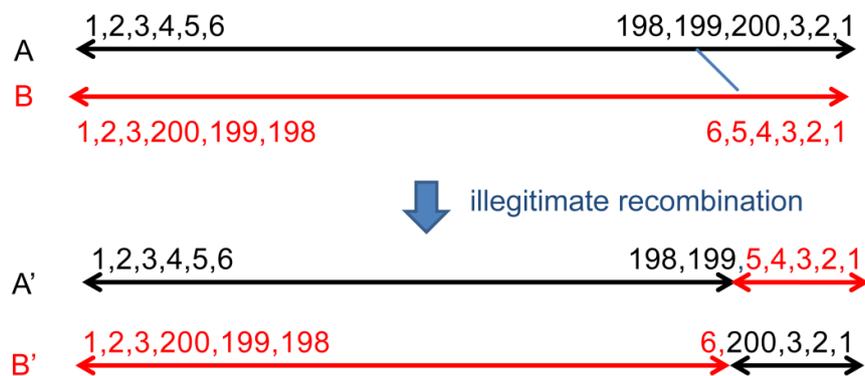


Figure 5.7 A model showing the proposed mechanism for genome rearrangement in VACVs

Virus B' also acquires a longer TIR encompassing genes from 1,2,3, to 200. Note gene 200 is also translocated from the right end to the left end.

Large deletions are commonly seen in VACVs, for example, MVA, after being passaged more than 570 times, acquired at least 6 large deletions compared to the parent strain Ankara [170]. I have also recently described another example of a virus bearing a large deletion, that was discovered during our characterization of recombinant VACV, and there was some sequence data to suggest that a small homology in nearby sequence can support illegitimate recombination [282]. Our model aims to show that illegitimate recombination can account for all three types of genome rearrangements (simple large deletions, large deletions with shortened TIRs and large deletions with extended TIRs) seen in VACVs.

The schemes outlined in Figure 5.7 can explain the structures of different VACVs. For example, in Part II we model a large deletion with a shortened telomere. Examples of such events are seen in RPXV (deletion of DVX\_214 to 216 till DVX\_013, TIR is shortened to DVX\_012) (Figure 5.3 E) and Cop (deletion of DVX\_213 to 216 till DVX\_013, 012, TIR is shortened to DVX\_011) (Figure 5.3 D).

In another example (Part III in Figure 5.7) we model a virus with an extended telomere. This is seen in DPP15 (deletion of DVX\_213 to 216 in the unique genes and extended TIR till DVX\_026); TP5 (deletion of DVX213 to 216 and extended TIR to DVX\_015) (Figure 5.3 F) and CVA (deletion of DVX213 to 216 and extended TIR to DVX\_019) (Figure 5.3 G). In scheme III virus A' has lost a unique gene, orf200, and extended its TIR (orf1 to 5). Virus B' also deletes unique genes (orf 4 and orf 5) and duplicates orf 200 in the right TIR. Interestingly, one could explain how a DPP25-like virus could derive from a HPXV-like ancestor through the deletion of HPX 007 to 015b (corresponding to orf 4,5 in this model) and the translocation of HPX 196 to 199 (DVX\_013 to 010, corresponding orf 200 in this model) from the right to the left end of the genome, making the TIR

longer. In this case, a block of genes from the right end is translocated into the left end. A recent paper [283] examining myxoma virus genome structure reported an event where two genes present in the right TIR of myxoma were translocated into the left TIR resulting in a small deletion. These findings provide real life support for our extended telomere model (Figure 5.7, Part III).

A unique example among VACVs is WR, which has an extended telomere. In order to explain the formation of WR's genome structure (DVX\_216 is present, following DVX\_013, 012, rearranging to DVX\_018), we can partially modify the model in Part III. First, an orf3 is deleted from the left but not right end (corresponding to DVX\_013 to 007). Then, we connect virus A orf3 in the right end to virus B orf5 in the left end. This process would generate a gene translocation in the right end between orf3 and orf4, 5 (corresponding to DVX\_013-12 to DVX\_018-014). In this way, a WR-like virus is produced (Figure 5.8).

Lister is a unique case because Lister appears to be a recombinant between VACV and CPXV. This event connected a gene homolog of CPX-GRI 208 to DVX\_211 (most possibly from a CVA-like virus, see below), causing an insertion comprising homologs of CPX-GRI gene 208 and 209. Interestingly, the Lister TIR is the same as that of CPX-GRI (DVX\_009/233). Compared to DPP25, a fragment, which is homologous to DVX\_211 to 216 till DVX\_013 to 010, is absent in Lister. We can explain this event using the same non-homologous recombination scheme shown in Figure 5.7, except that the two viruses must have comprised VACV and CPXV-like ancestors.

#### **5.4.2 Evolutionary relationship of vaccinia viruses**

This illegitimate recombination model can account for all the deletions and genome rearrangements observed in various vaccinia strains.

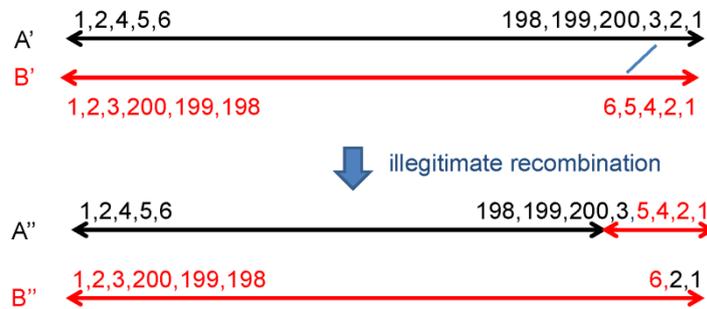


Figure 5.8 A special example for the evolutionary model part III applied to WR evolution

Both parental viruses A' and B' have deleted the left orf3. Since there is no longer any homology remaining downstream of orf3, the virus cannot revert back to wild type. Subsequently, an illegitimate recombination step happens, connecting black orf3 to red orf5. The resulting virus A'' has an extended telomere (orf1 to 5). However, orf3 and orf4/5 has been exchanged. The resulting virus B'' bears a large deletion in the left terminus.

We predict that a virus with a longer genome can generate a virus with a shorter genome no matter the length of the TIR. A corollary to our model is that a DPP25-like virus could be the ancestor strain for all other VACVs (downstream of a HPXV-like virus).

A scheme that could account for these relationships is shown in Figure 5.9. Panel A shows the right terminus of all VACVs. Since the right TIR is a duplicate of the left, I have only drawn the end of unique genes in genomes to represent each virus group. From the model, we predict that any virus with shorter genome can potentially derive from a virus with a longer genome. Consequently a DPP25-like virus could give rise to all other VACVs either directly or sequentially. Furthermore, based on the arrangement of the right TIR in VACVs, we can easily identify most (although not all) strains of viruses simply by analyzing its right TIR boundary. We suggest that the evolution of VACVs originated with some kind of HPXV, followed by the evolution of an intermediate DPP25-like virus, which subsequently evolved into the various VACV strains circulating today (Figure 5.9 B). Starting with an HPXV-like virus, there is the selection for a DPP25-like virus, then through a 2kbp deletion, one can generate an IHDW1/ DPP13-like virus. Interestingly, both viruses (DPP25 and IHDW/DPP13) are still present in Dryvax stocks, however the dominant strains in Dryvax are DPP15-like virus (60%) and DPP13-like (40%). Cop and Ankara likely derived from either DPP25-like or IHDW1/DPP13-like viruses. Theoretically, both of them can also derive from Rabbitpox or TianTan virus. Lister most likely originated from Ankara because the old lister stock, LIVP, bears an Ankara-like right TIR boundary. The presence of CrmE in Lister, USSR and Evans strains suggests that these three viruses likely share the same origin.

We did wonder if some of the older hypothetical viruses might still be present in extant virus stocks.

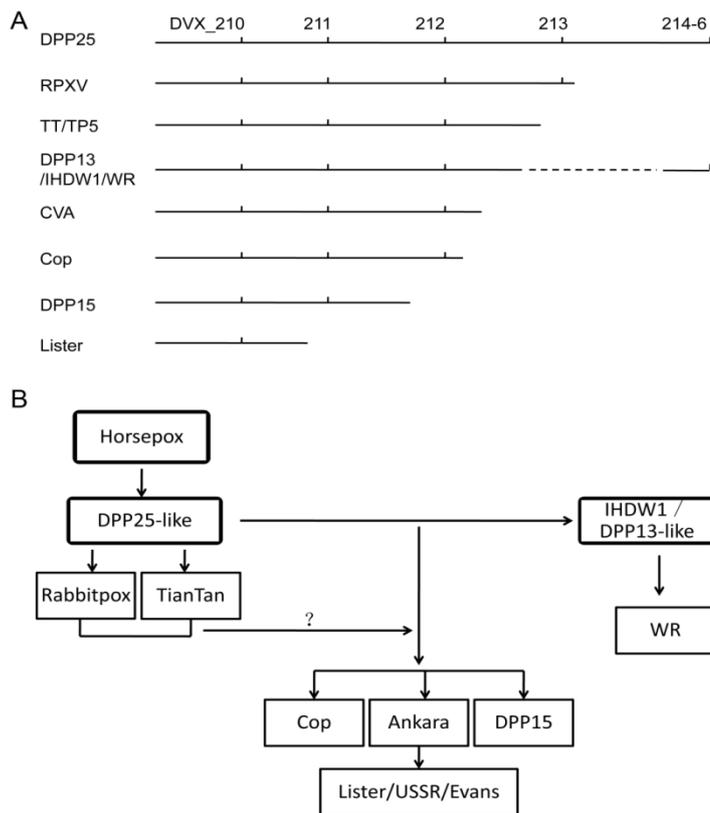


Figure 5.9 Proposed evolutionary relationships between extant and hypothetical ancestral VACV strains

Panel A shows the right genomic end of VACVs. According to the model, the virus with longer genome can evolve into any virus with a shorter genome. Panel B shows the deduced relationship between VACVs. Some ancestral HPXV evolved into a DPP25-like virus. Through a 2kbp deletion, IHDW1/DPP13-like viruses are generated, which can further evolve into WR-like virus. Rabbitpox and TianTan could only have derived from a DPP25-like virus. However, Cop and Ankara strains can come from either DPP25- or DPP13-like viruses. Due to the longer genomes of Rabbitpox and TianTan viruses, they can potentially evolve into Cop and Ankara strains. One can argue that Cop and Ankara are supposed to be very old strains and TianTan has circulated only in China for a long while (a ? mark shows here). Lister is proposed to derive from Ankara since a shared genome structure is found in both strains and Ankara has the larger genome.

However, using appropriate PCR primers we still could not find a CVA-like virus in our Lister stock (ATCC VR-1549) or a DPP13-like virus in an old WR stock (ATCC VR-119). Similarly we looked for DPP25-like viruses in stocks of Cop, TianTan, CVA, Lister and WR, and were unsuccessful. These data suggest that viruses with these hypothetical older genome structures have been lost from modern viral stocks.

### **5.4.3 VACV genome boundaries**

A previous study suggested that the left boundary of VACV genome seems to be defined by the need to retain the epithelial growth factor gene (DVX\_014) [246]. Here we wondered what gene (or genes) might define the right boundary of the genome. Aside from genes duplicated in the right TIR, the end of the genome seems to be defined by the DVX\_214-216, genes that have been truncated/deleted in all extant VACVs (Table 5.2). Interestingly, while some VACV genomes end at DVX\_214, like RPXV; most VACVs end within DVX\_213 (Table 5.2), suggesting that these two genes are not essential. Interestingly, The Lister genome has a right genome boundary extending into DVX\_212, which encodes an interferon alpha/beta receptor [134]. IFN is one of the most effective antiviral responses elicited by the host, not surprisingly poxviruses have evolved various ways of interfering with the IFN signaling pathways and thus greatly restricting the production of IFN by infected cells [128]. To prevent IFN signaling in uninfected cells, poxvirus secretes two soluble IFN receptors, DVX\_212 (IFN alpha/beta receptor) and DVX\_201 (IFN gamma receptor), which sequester IFNs and inhibit their binding to cell receptors. Given this biological role for DVX\_212, it is not surprising that most VACVs keep this gene. We therefore conclude that the right genomic boundary for most VACV is DVX\_212. Interestingly, as mentioned in chapter 3, the left boundary of VACV genomes is DVX\_014 (growth factor gene). In fact, most of VACV strains keep the same gene contents and gene orders between these two genes.

An interesting feature of the most widely used VACV vaccines is illustrated by these two genes. Four VACVs (Lister, Dryvax, EM-63 and TianTan) were selected for the smallpox eradication campaign in 1967 [188], because they had a better safety record [6] than some other VACVs. Among them, Lister, Dryvax and EM-63 (NYCBH origin) all encode DVX\_212 truncation, while some (e.g. TP03) Tian Tan viruses encode DVX\_014 deletions.

#### **5.4.4 Comparison of cowpox to vaccinia virus**

Compared to CPX-GRI, HPXV only differs by three deletions and there is no change in the TIR (Table 5.3). The first deletion is 4kbp in size bearing CPX-GRI D8L to D11L (all gene names refer to CPXV). The second is a 1.6kbp deletion disrupting B8R and B9R and the third is a 3.6kbp deletion encompassing K1R, K2R, K3R and T1R, just upstream of the TIR region (Lister inserts back the fragments containing K3R and T1R). Note that the second and third deletions retained by DPP25 and all other VACVs, further support for the hypothesis that a HPXV-like virus may be an intermediate virus between CPXV and VACV.

In fact, as human vaccines and to a first approximation, HPXV and VACV showed no obvious difference from CPXV. It wasn't until 1939 that Downie first reported [2, 3] that the biological properties on the chorioallantoic membrane of smallpox vaccines (which by that time are assumed to be VACV) were different from those of CPXV. Loss of the aforementioned genes should therefore account for the difference between CPXV and VACV on CA membranes. Interestingly, half of these genes encode ankyrin-F-box-like or BTB-kelch-like proteins, including DVX\_213 (ankyrin- F-Box) and DVX\_214 (BTB-kelch) (Table 5.2). These proteins comprise the largest viral multigene families in orthopoxviruses since CPXV encodes fourteen ankyrin- containing proteins, thirteen of which also contain an F-box domain, and six BTB-kelch-like proteins [284, 285].

Table 5.3 Cowpox genes (CPX-GRI) that are not found in VACV

CPXV-GRI	Family/gene name	Length aa	HPXV	VV-DPP25
Left telomere (15kbp CPXV to VACV) (10.7kbp HPXV to VACV)				
D6L	Putative TLR signaling inhibitor	219	HPXV 007	
D7L	BTB Kelch-like	273	HPXV 008	
D8L	Ankyrin (Bang-B18R)	661	4kbp deletion in HPXV	
D9L	C-type lectin (FPV-V-008)	75		
D10L	CPV-B-012	96		
D11L	BTB Kelch-like	521		
D12L	TNF receptor (CrmB)	202	HPXV 009	
D13L	Unknown	111	HPXV 010	
D14L	Ankyrin (Cop-B18R)	764	HPXV 011a,b,c	
C1L	Ankyrin	437	HPXV 012	
C2L	MPV-Z-N3R	178	HPXV 013	
C3L	Ankyrin (Cop-B18R)	833	HPXV 014a,b,c,d	
C4L	Unknown (Bang-D3L)	170	HPXV 015a,b	
Middle genome (1.6kbp CPXV to HPXV)				
B8R	Virulence factor (Cop-B9R)	221	HPXV 186 (Frag)	DVX_202 (Cop-B9R) (Frag)
B9R	Kelch-like protein	501	HPXV 187 (Frag)	DVX_203 (Cop-B10R) (Frag)
Right genome (5.5kbp HPXV to VACV)				
B22R	Surface glycoprotein	1933	HPXV200	DVX_011(Frag)
Right genome (3.6kbp CPXV to HPXV)				
K1R	Ankyrin (Cop-B25R)	581	HPXV 201a,b (Frag)	DVX_010 (Cop-C15L) (Frag)

K2R	TNF receptor (CrmD)	322		
K3R	TNF-alpha receptor (CrmE)	167		
T1R	Viral Golgi anti-apoptotic protein (vGAAP)	210		

Note: 20 genes in CPX-GRI are missing in DPP25.

Ankyrin and BTB proteins interact with cellular Cullin-1 and Cullin-3-containing ubiquitin-protein ligases, respectively, and are assumed to be involved in regulating a host range response [284, 285]. CPXV has a broadest host range, bearing 14 ankyrin genes, MPXV has eight and VACV-Cop has five. VARV only infects human, encodes only 5 ankyrin genes, and 3 of these are shared with VACV-Cop.

VACV-MVA has very limited replication capacity in most mammalian cells, and encodes only one ankyrin gene (Cop-B18R, DVX\_211) [170, 258]. Furthermore, deleting B18R from MVA caused a reduction of viral intermediate and late protein synthesis, suggesting that this may be the minimum ankyrin gene complement to permit VACV infection [286]. In contrast, VARV does not encode any BTB proteins, while MVA and MPXV encode only one, Cop-F3L. It should be noted that the ankyrin multigene family appears to be redundant, since the CPXV CHOhr (Chinese hamster ovary, host range) gene CP77 can partly replace the host range function of K1L [287]. CPXV also encodes five tumour necrosis factor receptors (TNFR) [288], all of which are truncated in VACV WR and Cop strains [128]. In conclusion, assuming that CPXV was first used to vaccinate humans, these numerous host ranges genes were lost due to deletion or truncation, as they were not required for human infection.

#### **5.4.5 The driving force for VACV evolution**

VACV is a large DNA virus, encoding approximately 200 genes, which can roughly be divided into two types; essential genes, which are necessary for viral growth and replication, and non-essential genes, which serve more specialized functions to regulate the host immune systems. Viruses adapt well to their environment in nature and its host. However, when people started using HPXV or CPXV as vaccines for smallpox, the situation changed. Firstly, vaccines were required in large amounts. Secondly, the hosts for growing vaccines kept changing. This process created many

opportunities for mutations, including genome rearrangements, SNPs and indels. Mutations occurred randomly and whether they were under selection pressure depended on the environment. Passage through humans and particular animals, appears to have favoured loss of many genes, perhaps due to selection for viruses that caused milder vaccination sequelae. In conclusion, VACV evolution seems to have been accompanied by loss or truncation of virulence genes perhaps due to human intervention, during which genome rearrangement plays an important role.

## CHAPTER SIX - DISCUSSION AND FUTURE DIRECTIONS

DNA sequencing technology provides a genetic foundation for our understanding of life processes, evolution, behavior and disease. However, DNA sequencing technology was laborious and time consuming, until the advent of the next generation sequencing (NGS). NGS technologies have enabled huge progress in biological and medical research in the past several years, including de novo whole genome sequencing of modern and extinct species, characterization of difference between individuals in a species or the difference between cells in an individual (cancer genomics, metagenomics and immunogenomics) and decoding of underlying cellular mechanisms (epigenetics, transcriptome, DNA methylation, active regulatory chromatin, protein-DNA interaction and so on). To take advantage of NGS technology, I used a 454 sequencer to sequence more than one hundred vaccinia virus genomes, with the purpose of exploring viral diversity, recombination and evolution. It is known that diversity and evolution are associated with various genetic changes, including recombination, point mutations (SNPs), small indels, large deletions with or without genome rearrangement, and copy number variation, all of which are the topics of this thesis.

### 6.1 SNPs

Vaccinia virus has a complex history and an unknown origin. What we know is that the vaccines used for smallpox eradication programme have never been cloned and were passaged at different times through humans, animals and/or chicken eggs. In a Lister stock, more than 1200 SNPs were discovered between clones [18]. In a Dryvax stock (chapter two), I found 570 SNPs on average between either of two clones, scattered across the whole genome. However, in our lab TianTan stock, I found only a few SNPs that differentiated two clones, however, approximately 880 SNPs were

found in either of two clones isolated from a Beijing TianTan stock [289]. We assume that our TianTan stock has been cloned previously. There are some interesting differences between the different TianTan clones. One of our viruses, TP05, encoded the longest genome among all TianTan clones, and contained two epithelial growth factor genes within its TIR. The growth factor gene is lost in most of the other TianTan clones, perhaps due to the absence of selection when cultured in chicken eggs. One possible explanation is that although viruses encoding one or two growth factor genes comprise a minor part of the virus swarm, passage through mammalian cell culture favours their growth. Another example is the p28 gene in VACV strain Lister. The p28 gene (encoding an E3 ligase in poxvirus [290] and a virulence factor [291]) is truncated in most VACVs except for the IHD strain. This gene is however found intact in LC16m8 and LC16m0, two attenuated strains that were derived from a Lister stock [18], while all the other Lister clones sequenced truncated this gene.

SNPs are generated randomly due to replication errors caused by DNA polymerase. In chapter four, I found only one SNP (and one small deletion) in my hybrid viruses and estimated the VACV E9 polymerase mutation rate, at about  $10^{-8}$ , consistent with other types of DNA polymerase. Since SNPs are produced spontaneously, what selection mechanism does the virus engage when keeping or removing a new SNP? Of course, most SNPs are neutral since they do not change any amino and most likely these mutations are simply subject to random drift without any selection.

SNPs are commonly used to determine branch length of a phylogenetic tree. The more SNPs that differ between two viruses, the longer the branch. However, the branching is not completely supported by the bootstrap scores that are used to construct the tree (chapter two), due to recombination. For example, a patchy pattern of recombination among all Dryvax- derived clones is observed, complicating SNP derived phylogenetic analysis. In chapter four, I explored recombination in more detail. DPP17 and TP05

differ by 1399 SNPs (1 SNP in 140bp fragment on average) across the whole genome, which were used to track the origin of the segments in progeny recombinants. Hybrid viruses contained 18 (DTH viruses) and 30 (DTM viruses) crossovers on average. I could detect no evidence for recombination hotspots, and the lengths and spacing of the genetic exchanges appeared to be random. According to my data, genetic material is exchanged approximately once every 12 kB (close to the 1 per 8 kbp in previous study) with a bias that favors the exchange of shorter segments. We also noticed that DTD clones, produced by non-genetic reactivation with SFV, did not show any exchange with the helper virus.

## **6.2 SMALL INDELS**

Small indels are less common than SNPs. In a Dryvax pool (chapter two), I detected only 110 small indels and more than 85% were associated with small repeats. I also noted that the longer the repeats, the less frequently they were associated with indels. We believe the reason for this is that longer repeats are unstable and are rapidly deleted from the virus pool. It is generally assumed that the indels are generated due to strand slippage errors during DNA replication, a phenomenon having also been observed in VARVs and MVA [171]. My whole genome sequencing analysis of multiple clones in a single viral stock, further supports this idea (Chapter two).

These mutations are called “indels” because one cannot always decide whether it is an insertion or deletion. In chapter five, one of my IHDW clones, IHDW1, contained indels that must have been insertions since they truncated genes. This shows that VACV DNA polymerase can cause insertion errors. In contrast, MVA, when passaged through chick embryo fibroblasts, mainly acquired deletions [8].

Indels are more likely to have more severe genetic consequences than SNPs, since they are likely to create frameshift mutations. This would

generally be disfavoured if the gene serves a useful purpose. However, if a gene is not essential, and once it has accumulated one inactivating mutation, there is generally no further selection for gene function and multiple mutations can then be observed. We see examples of this in the VACV MIL and I4L genes (Chapter two). Interestingly, the I4L gene in DPP17 seems to provide a rare example of a “molecular fossil” where a block of linked mutations can be seen also in HPXV. This provides further support for the hypothesis that DPP17 or Dryvax shares a common origin with HPXV.

### **6.3 LARGE DELETION AND GENOME REARRANGEMENT**

Compared to SNPs and indels, large deletions and genome rearrangements are generally rare events. In the Dryvax pool (Chapter two), we observed only three of these events. Assuming that a virus resembling CL3/DPP25 is the parental strain; DPP15, Duke and DPP17 groups can be derived from such a parental virus by a combination of large deletions and genome rearrangements. We also found a large deletion in one of my hybrid viruses, DTM28, and proposed that the mechanism underlying such a genome rearrangement is illegitimate recombination (Chapter four). We further noted that small repeats and some limited sequence homology around the deletion site likely supported illegitimate recombination.

Due to the large effect that large deletions and genome rearrangements make on the viral genome, we consider that this event is irreversible (as long as the deletion excises all copies of the affected sequence). That is to say DPP17 cannot revert back to a DPP15-like virus. Or that if there is another genome rearrangement in DPP15, there is little chance it will result in a DPP17-like virus.

Because of their irreversibility and rarity, genome rearrangements can be looked as features, which are likely retained by all progeny viruses. It is the basic idea behind a scheme for modeling viral evolution that I proposed

in chapter five. To explore this idea, we decided to investigate the presumed origins of VACVs. A critical assumption behind my model is that “reductive evolution” [227, 261] has been an important feature of *Orthopoxvirus*. Given that the CPXV group (which comprises at least 5 subtypes [260]) encodes all of the genes present in all other *Orthopoxviruses*, it is probable that CPXV might be the ancestral strain for all other *Orthopoxviruses*.

HPXV belongs to the vaccinia virus species as judged by phylogenetic analysis and based upon analysis of SNPs. However, it also contains several large additional fragments not present in other VACVs [166]. If one looks at the genome structure, HPXV is actually quite similar to a CPX-GRI-like virus, suggesting that genome structure and SNPs are not co-evolving. It has long been suggested that a HPXV-like virus is the likely ancestor of other VACVs although there is no evidence to prove this idea. In chapter five, I characterized three conserved genome features that are unique to all VACVs when compared to HPXV, suggesting a hypothetical intermediate virus which could have derived from HPXV and served as the ancestral strain for all other VACVs. A DPP25-like virus is proposed to be the candidate for such an intermediate virus.

Although illegitimate recombination can be proposed to explain simple deletions, our own studies do not provide direct evidence for large DNA fragments to be translocated from the left end of the genome to the right, or vice versa. However, the early literature has documented this phenomenon: large genome deletions with or without genome rearrangements occurring [99] [101] [97], suggesting that the regions around the TIRs are interchangeable. Based on these many observations, I believe this model for virus evolution (Chapter five) can easily rationalize and explain the likely evolutionary relationship between all VACVs and possibly all other orthopox virus evolution.

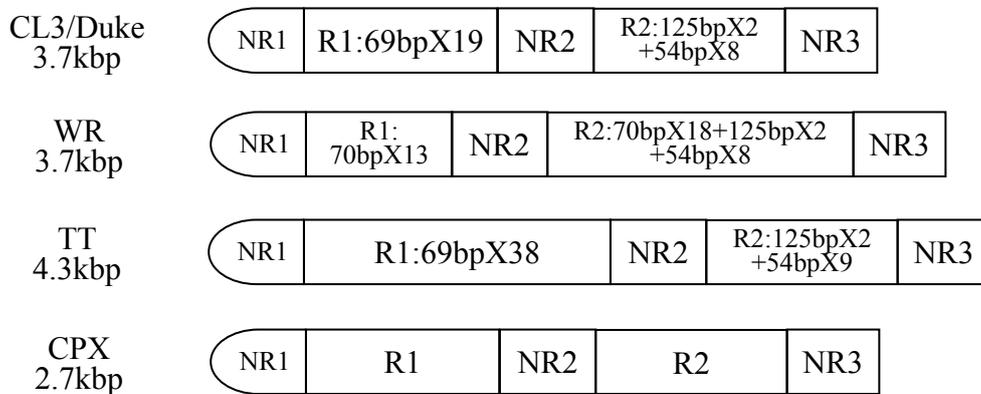
## 6.4 COPY NUMBER VARIATION

Under hydroxyurea drug selection, the VACV I4L gene is duplicated to generate more copies in the genome [103]; K3L is also amplified to produce multiple copies in response to PKR immune pressure [104]. These data suggested that poxviruses can adapt rapidly in response to their environment by increasing expression of targeted viral proteins. Interestingly, copy number variation is also seen in the number of tandem repeats in VACV TIRs [102] [182-185]. In my study, Dryvax and TianTan virus also show this variation in the number of tandem repeats (Chapter two). Why this occurs is unclear and it would be an interesting area of future investigation.

From our work as well as the work of others, the high variability in VACV TIRs is surprising. Seemingly, no two strains of VACVs have the same TIR (Figure 6.1 A). The difference is caused by changes in the lengths of repeats R1 and R2 since sequences NR1, NR2 and NR3 are highly conserved. The R1 sequence in WR consists of multiple copies of a 70bp repeats. Other VACV, like CL3 and Duke, encode a 69bp repeat derived from the 70bp repeat through deletion of one nucleotide. In contrast, R2 consists of 54bp, 125bp and 70bp repeats. Unequal crossing over is proposed to explain changes in the numbers of these repeats [181]. Interestingly, in my TianTan clones, I found that an array of 59bp repeats had replaced a region encompassing of DVX\_004 to 006. These 59bp repeats appear to be derived from telomeric 69bp repeats by accidental illegitimate recombination.

In addition to the variable lengths of telomeres, copy number variations in the number of tandem repeats in VACV WR were also first reported by Moss [102]. Restriction endonuclease analysis of terminal fragments detected an array of eight or more fragments differing in size by 1650 bp increments even after the viruses were repeatedly plaque purified (Figure 6.1 B).

A:



NR: non-repeats, R: Repeats

B:



Figure 6.1 Diagram showing the telomeric repeat structures in different VACV

A: the unique composition of TIR for each virus. NR1 includes hairpin loop and concatemer resolution sequences. CL3 and Duke have the same TIR structure, one that is also found in Drvax clones. The numbers at left indicate the length of telomere of each different virus. (Figure is not drawn to scale)

B: 1.6kbp increment in WR strain.

The transition between the longer telomere and the original one may be driven by recombination [102]. This 1.6 kbp repeated fragment contains NR2 and part of R2 (17 copies of 70 bp repeats). There was reported to be no difference between the clones having longer telomeres (unstable variants in the Moss paper) or unique length of telomere (stable variant) in terms of plaque morphology and growth rate. The paper also showed that only 20% of the unstable variants with longer telomeres reverted back to the original length during serial passages, suggesting that the longer TIR is favoured in culture. This phenomenon is seen only in vaccinia virus and confirmed by our study (chapter two, Figure 2. 2). I used *SaII* to digest the genomic DNAs of Dryvax and TianTan clones. The restriction endonuclease fragments produced telomere ladders for all clones except DPP15, suggesting most were unstable and variable in length. DPP15 exhibited a unique length and the shortest telomere in Dryvax clones. Based on my sequencing data, I estimate that the telomeric increment in Dryvax clones is about 0.9kbp, consisting of R2, NR3 and 12 nt of the end of orf1 (a gene on the negative strand).

One of the things that clearly differentiates Tian Tan and Dryvax clones is that the Tian Tan clones produce plaques that are twice the size of Dryvax clones (chapter four, TP5 to DPP17). We do not know why, although it was interesting that the hybrid viruses bearing Tian Tan telomeres seemed to produce larger plaques than hybrids with Dryvax telomere (Chapter four). These data led us to hypothesize that the length and/or composition of the TIRs may play some role in determining the rate of viral replication.

To test this idea, I isolated viruses with unique lengths of telomere (This was possible because the stable variants comprise only 20% of WR stock according to the Moss study). Four viruses were isolated, WR72, DPP15, CPX1 (used as a control), and DTM4, a clone from my hybrid collection with TianTan-like telomeres. The plaque sizes of CPX, WR,

TianTan and DPP15 are dramatically different, although curiously they do correspond roughly with the length of telomere (Figure 6.1 and Figure 6.2).

Interestingly, WR and DPP15 (CL3/Duke-like telomere) have the same unique length of telomere (3.7 kbp shown in Figure 6.1 A); however, the plaque sizes of WR and TianTan are nearly twice of that of DPP15. Compared to DPP15, WR has a longer R2 and short R1, and TianTan has a longer R1 and a similar R2, suggesting that the composition of TIR may also play a role in viral replication.

In order to study the role of TIR repeats in virus replication, the rest of the virus has to be isogenic and the most straightforward way to do this is to swap telomeres between two viruses. (Although I did not have time to complete these studies, I made some progress towards investigating this question and my preliminary studies lay the foundations for some interesting future research). The experimental design required a virus with a unique restriction site near the telomere. Since I was unable to find any unique endonuclease recognition sites in this region, I inserted a unique I-SceI digestion fragment into the junction site between NR3 and orf1, with a YFP-GPT cassette as a selection marker, and used this construct to genetically modify CPX1, WR72, DPP15 and DTM4. DPP9 was used as a control (for reasons that are explained below). To construct these viruses, two homologous flanking regions were chosen, one is the NR3 element (a 190nt conserved virus sequence, whose function is unclear) and the other is orf1. All the viruses were successfully constructed with insertions introduced into the two NR3 sites within both TIRs. DPP9 is used as a control since it contains multiple copies of NR3 in its telomere repeats (telomere ladders seen in Chapter 2).

## comparing plaque sizes of CPX1 and VACV

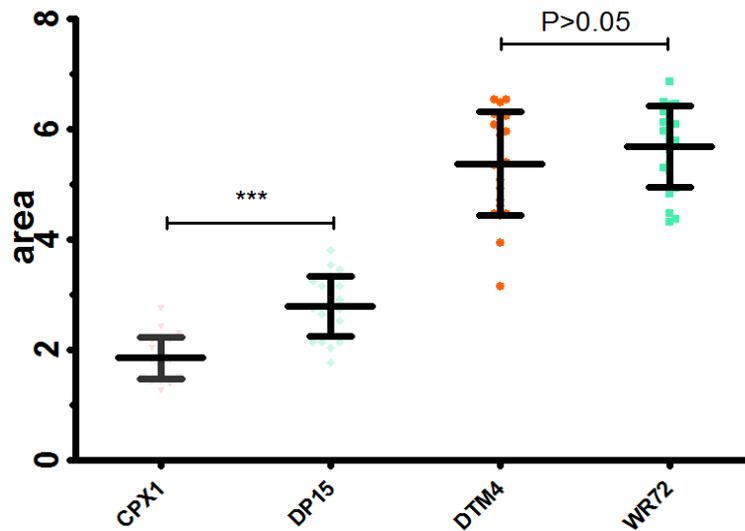


Figure 6.2 Comparison of plaque size of CPXV and other VACVs

CPX1 is a clone from our lab stock of CPXV. DP15 is a Dryvax-derived clone. DTM4 is a hybrid that was isolated as described in Chapter 4. It bears a TianTan-like telomere. WR72 is a WR clone.

NR3 is located between orf1 and the telomere repeats and is conserved in all poxviruses. As I noted above, the function of NR3 is unclear, and out of curiosity I also constructed additional plasmids that would delete either the whole of NR3 or a half of NR3 (leaving 90nt of NR3 proximal to the hairpin end of the genome). Interestingly, I could disrupt one of the two copies of NR3 in viruses encoding just two copies of NR3 (CPX1, WR72, DPP15, and DTM4) and I could disrupt one of the two NR3 copies closest to Orf1 in DPP9 (which encodes multiple copies of NR3). However, I was not able to disrupt both copies of NR3 in the first set of viruses, nor could I disrupt both of the NR3's in DPP9 that lie next to Orf1. This suggests that at least one copy of NR3 is essential for VACV, moreover it may have to be a copy adjacent to Orf1.

**Future direction:**

These preliminary investigations into the functions of poxvirus telomeres suggest a number of areas for future research.

Firstly, does the TIR length and/or composition regulate viral replication? To test this, viruses with I-SceI digestion sites near NR3 have been constructed, including WR72, DPP15, DTM4 and CPX1. To swap telomeres, it should be possible to isolate these virus DNAs, cut them with I-SceI, and then co-transfect them into SFV-infected cells (as I described in Chapter 4) along with DNA fragments encoding the desired telomeres. SFV-catalyzed reactivation reactions, along with recombination between the overlapping sequences, should permit swapping telomeres between the viruses. [245] (Chapter 4, the method was used to make DTD clones). Alternatively one could genetically modify virus "bacmids" in bacteria [292] [293] prior to rescuing them using reactivation technologies. After making a series of viruses that differ only in the composition of their telomeric repeats, the virus growth rates and yields in culture will be compared. I would predict that translocation of the TIR from one Orthopoxvirus to another would change the viral growth rate, ideally in a reciprocal manner.

The second question concerns what function (if any) the variation in telomere repeat number might serve. Although Moss has shown that the plaque morphology of the stable variants is identical to that of an unstable variant, it does not explain why most viruses in a stock retain the longer repeats. It is possible that these repeats serve some function in replication that is not yet understood; perhaps the variation is a reflection of recombination being used to initiate replication without necessarily affecting growth rates or virus yields.

The third issue concerns NR3 function. My data suggests that NR3 is necessary for viral replication. It is well known that the genes within TIR are transcribed in a direction towards the hairpin telomere and an early transcript termination signal TTTTNT [294] is found on the telomere side of NR3. Thus it seems quite possible that an early transcript initiating in the first ORF, could read through NR3 and terminate at this signal. It has been reported that a non-coding RNA in ectromelia virus is required for viral replication [295]. It would be very interesting to test if NR3 encodes a non-coding RNA that serves some possible role in viral replication.

In this regard it is noteworthy that a transcript was produced containing NR3, it would likely fold into a tRNA-like structure (Figure 6.3) and tRNAs do serve as primers in HIV replication [296]. This raises the very intriguing possibility that VACV encodes an RNA primer that serves to initiate replication in the virus telomeres.

## **6.5 VIRULENCE**

One of the purposes of this thesis was to isolate a naturally attenuated virus for use as an oncolytic virus or as a vaccine vector. We have tested some of my viruses to compare the virulence of different strains. Figure 6.4 shows a recent animal virulence study performed by Ms. Nicole Favis.

Secondary structure:  $\Delta G = -49.4\text{kcal/mol}$

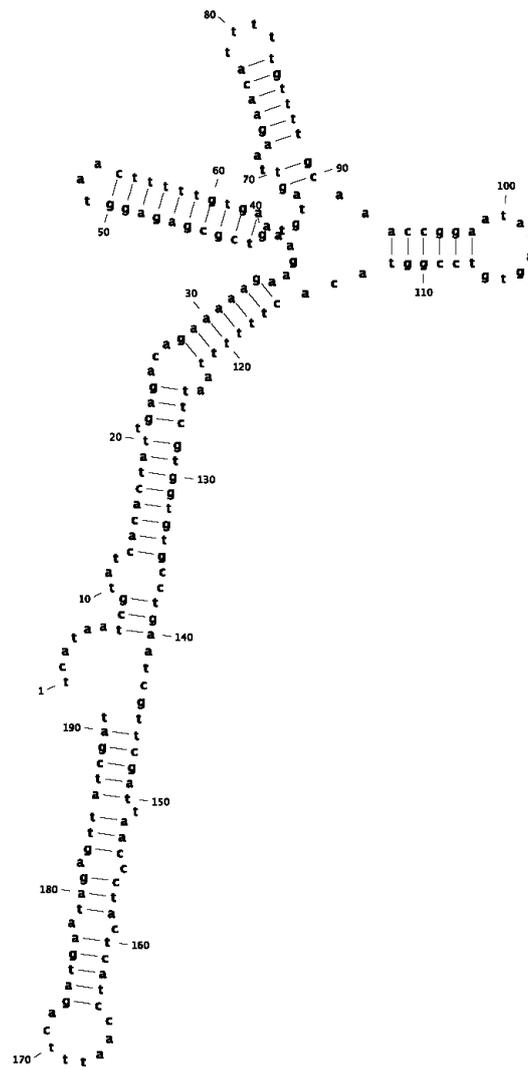


Figure 6.3 The predicted RNA secondary structure of NR3 (191nt)

Using CLC genome workbench software, this NR3 (191nt) RNA structure is predicted when the DNA sequence is transcribed into the RNA sequence. Please note, at position 30, there is the termination signal of early transcripts (TTTTTNT). In my study, a half NR3 (nt 1 to 90) is necessary for viral replication.

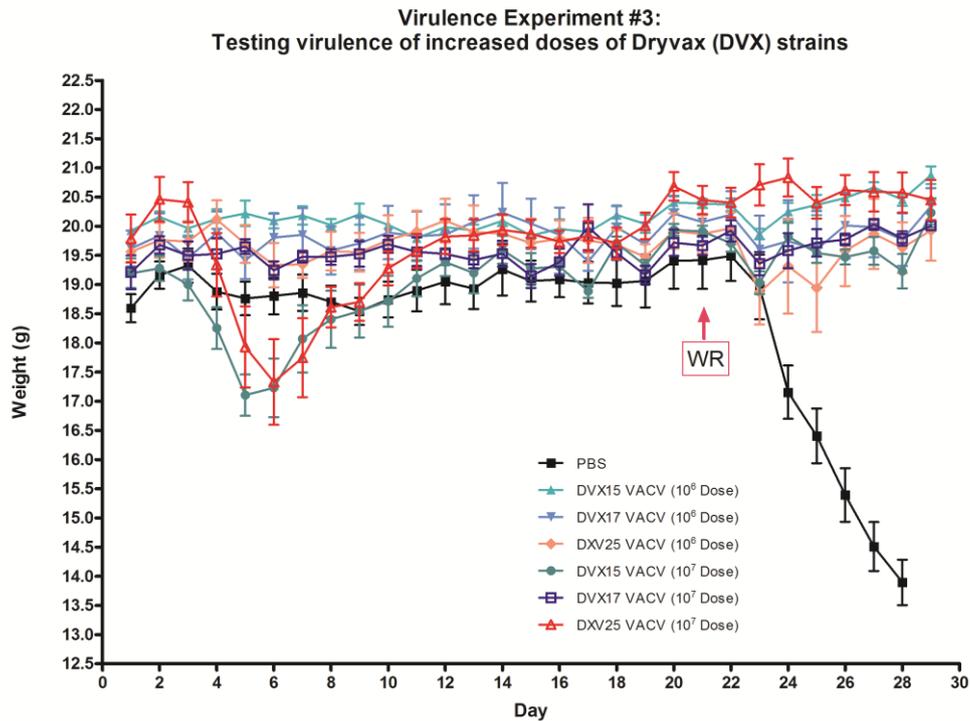


Figure 6.4 Virulence study on mice to compare DPP15, 17 and 25

Female Balb/c mice, 4 weeks of age, were obtained from Charles River Laboratories (U.S.). DPP15, DPP17, DPP25 viruses were purified using sucrose gradients. These 3 viruses and PBS control at day 1 were inoculated into mice (5 mice per group) by the intranasal route with  $10^7$  or  $10^6$  PFU/mouse of virus diluted in  $10\mu\text{l}$  of PBS. Body weights were recorded everyday over one month or until the mouse had lost 30% of its body weight and had to be euthanized. At day 21, a lethal dose of WR virus ( $10^6$  PFU/mouse) was inoculated into all of mice by the same route. At 7 days after WR inoculation, all mice in PBS control group were too sick and were euthanized. In contrast, all mice immunized with Dryvax clones survived the WR challenge. (This animal study was performed by Miss Nicole Favis)

In this study, we compared three Dryvax clones (DPP15, 17 and DPP25) and using two different doses ( $10^6$  and  $10^7$  PFU/mouse). Note that, DPP15 is very similar to ACAM2000, a vaccine licensed in US for smallpox prevention. The figure shows that at the lower dose, all three viruses have no effect on mouse weight. However, in animals inoculated with  $10^7$  PFU (the highest dose we could use), both DPP15 and DPP25 groups showed a roughly 10% of weight loss, compared to the DPP17 group, which showed no evidence of pathology. This shows that the mutations we identified in DPP17 (in particular the 7kbp deletion compared to DPP15) create an attenuated virus relative to other viruses in the stock.

After the mice had recovered from infection, they were challenged with a lethal dose of WR strain,  $10^6$  PFU/mouse at day 21. All mice in the PBS control group died, compared with 100% survival of all immunized mice, indicating that all of the Dryvax clones provide an effective protection for mice against WR challenge. Among them, DPP17 stands out as the most promising clone since it had the least virulence.

### **Future direction**

According to our data, DPP17 may be a safer VACV vector than DPP15, or possibly ACAM2000 as a smallpox vaccine. It has been shown that ACAM2000 is linked to post-vaccination myopericarditis in vaccinees [7]. Thus, a safer vaccine is still needed. DPP17 is derived from the same stock as ACAM2000 and shows reduced virulence in mice, suggesting it might be a good candidate for use either as a smallpox vaccine or as a vector for other recombinant antigens. To document this application, additional virulence and toxicity tests will be required in a greater variety of animal models such as rabbits and perhaps monkeys.

Another interesting virus is TP3, one of the two viruses we isolated from TianTan stock. We have shown that this virus has a small plaque phenotype compared to TP5, a virus with the longest known genome amongst all the TianTan viruses. A spontaneous deletion in TP3

encompasses both copies of the epidermal growth factor gene, a gene affecting virulence in mice and rabbits [224]. This is also of interest because a virus with double deletion of thymidine kinase and growth factor genes showed tumor-selective oncolytic activity [226]. Moreover, deleting the SPI-1 and SPI-2 genes from VACV produced a strain that grew preferentially in p53-null tumor cells [297]. SPI-1 and SPI-2 are host range genes and inhibit apoptosis [297]. Interestingly, TP3 has also truncated both genes naturally. Therefore, TP3 would seem to be a good candidate for use as an oncolytic virus or as a vaccine vector especially in China where it would be the only VACV strain acceptable for use by health authorities. We are currently working to license the virus to a Shanghai-based company with the hope that we will someday be able to test the clinical performance of TP3 *in vivo*.

In conclusion, NGS technology has proven to be a very valuable tool for examining many aspects of vaccinia virus biology, with applications ranging from basic science (replication, recombination, and evolution) to virus-based therapeutics.

## BIBLIOGRAPHY

1. Fenner, F., *Smallpox and its eradication*. History of international public health. 1988, Geneva: World Health Organization. xvi, 1460 p.
2. Downie, A.W., *Immunological relationship of the virus of spontaneous cowpox to vaccinia virus*. Br J Exp Pathol, 1939. **20**(158-176).
3. Downie, A.W., *A study of the lesions produced experimentally by cowpox virus*. J Pathol Bacteriol, 1939. **48**: p. 361-379.
4. Jacobs, B.L., et al., *Vaccinia virus vaccines: past, present and future*. Antiviral Res, 2009. **84**(1): p. 1-13.
5. Wokatsch, R., *vaccinia virus*, in *Strains of human viruses*, M. Majer, Editor. 1972, Karger, Basel. p. 241-257.
6. Polak, M.F., et al., *A Comparative Study of Clinical Reaction Observed after Application of Several Smallpox Vaccines in Primary Vaccination of Young Adults*. Bull World Health Organ, 1963. **29**: p. 311-22.
7. Greenberg, R.N. and J.S. Kennedy, *ACAM2000: a newly licensed cell culture-based live vaccinia smallpox vaccine*. Expert Opin Investig Drugs, 2008. **17**(4): p. 555-64.
8. Meyer, H., G. Sutter, and A. Mayr, *Mapping of deletions in the genome of the highly attenuated vaccinia virus MVA and their influence on virulence*. J Gen Virol, 1991. **72** ( Pt 5): p. 1031-8.
9. Carroll, M.W. and B. Moss, *Host range and cytopathogenicity of the highly attenuated MVA strain of vaccinia virus: propagation and generation of recombinant viruses in a nonhuman mammalian cell line*. Virology, 1997. **238**(2): p. 198-211.
10. Smith, G.L., *Genus orthopoxvirus: Vaccinia virus*, in *Poxviruses*, A.A. Mercer, Editor. 2007, Birkhauser Advances in Infectious diseases. p. 1-46.
11. Mayr, A., *Smallpox vaccination and bioterrorism with pox viruses*. Comp Immunol Microbiol Infect Dis, 2003. **26**(5-6): p. 423-30.
12. Stickl, H. and V. Hochstein-Mintzel, *[Intracutaneous smallpox vaccination with a weak pathogenic vaccinia virus ("MVA virus")]*. Munch Med Wochenschr, 1971. **113**(35): p. 1149-53.
13. Coulibaly, S., et al., *The nonreplicating smallpox candidate vaccines defective vaccinia Lister (dVV-L) and modified vaccinia Ankara (MVA) elicit robust long-term protection*. Virology, 2005. **341**(1): p. 91-101.
14. Mayr, A., *[TC marker of the attenuated vaccinia vaccine strain "MVA" in human cell cultures and protective immunization against orthopox diseases in animals]*. Zentralbl Veterinarmed B, 1976. **23**(5-6): p. 417-30.
15. Earl, P.L., et al., *Immunogenicity of a highly attenuated MVA smallpox vaccine and protection against monkeypox*. Nature, 2004. **428**(6979): p. 182-5.
16. Kenner, J., et al., *LC16m8: an attenuated smallpox vaccine*. Vaccine, 2006. **24**(47-48): p. 7009-22.
17. Friedman, D., *Safety and Immunogenicity of LC16m8, an Attenuated Smallpox Vaccine in Vaccinia-Naive Adults*. J Infect Dis, 2012. **206**(7): p. 1149; author reply 1149-50.
18. Morikawa, S., et al., *An attenuated LC16m8 smallpox vaccine: analysis of full-genome sequence and induction of immune protection*. J Virol, 2005. **79**(18): p. 11873-91.
19. Kidokoro, M., M. Tashiro, and H. Shida, *Genetically stable and fully effective smallpox vaccine strain constructed from highly attenuated vaccinia LC16m8*. Proc Natl Acad Sci U S A, 2005. **102**(11): p. 4152-7.
20. Takahashi-Nishimaki, F., et al., *Regulation of plaque size and host range by a vaccinia virus gene related to complement system proteins*. Virology, 1991. **181**(1): p. 158-64.
21. Tartaglia, J., et al., *NYVAC: a highly attenuated strain of vaccinia virus*. Virology, 1992. **188**(1): p. 217-32.
22. Pincus, S., J. Tartaglia, and E. Paoletti, *Poxvirus-based vectors as vaccine candidates*. Biologicals, 1995. **23**(2): p. 159-64.

23. Tartaglia, J., et al., *Safety and immunogenicity of recombinants based on the genetically-engineered vaccinia strain, NYVAC*. Dev Biol Stand, 1994. **82**: p. 125-9.
24. Tagaya, I., T. Kitamura, and Y. Sano, *A new mutant of dermovaccinia virus*. Nature, 1961. **192**: p. 381-2.
25. Ishii, K., et al., *Structural analysis of vaccinia virus DIs strain: application as a new replication-deficient viral vector*. Virology, 2002. **302**(2): p. 433-44.
26. Paoletti, E., J. Tartaglia, and J. Taylor, *Safe and effective poxvirus vectors--NYVAC and ALVAC*. Dev Biol Stand, 1994. **82**: p. 65-9.
27. Panicali, D., et al., *Construction of live vaccines by using genetically engineered poxviruses: biological activity of recombinant vaccinia virus expressing influenza virus hemagglutinin*. Proc Natl Acad Sci U S A, 1983. **80**(17): p. 5364-8.
28. Paoletti, E., et al., *Construction of live vaccines using genetically engineered poxviruses: biological activity of vaccinia virus recombinants expressing the hepatitis B virus surface antigen and the herpes simplex virus glycoprotein D*. Proc Natl Acad Sci U S A, 1984. **81**(1): p. 193-7.
29. Paoletti, E., et al., *Genetically engineered poxviruses: a novel approach to the construction of live vaccines*. Vaccine, 1984. **2**(3): p. 204-8.
30. Smith, G.L., M. Mackett, and B. Moss, *Infectious vaccinia virus recombinants that express hepatitis B virus surface antigen*. Nature, 1983. **302**(5908): p. 490-5.
31. Smith, G.L. and B. Moss, *Infectious poxvirus vectors have capacity for at least 25 000 base pairs of foreign DNA*. Gene, 1983. **25**(1): p. 21-8.
32. Perkus, M.E., et al., *Recombinant vaccinia virus: immunization against multiple pathogens*. Science, 1985. **229**(4717): p. 981-4.
33. Breitbach, C.J., et al., *Targeted and armed oncolytic poxviruses for cancer: the lead example of JX-594*. Curr Pharm Biotechnol, 2012. **13**(9): p. 1768-72.
34. Parato, K.A., et al., *The oncolytic poxvirus JX-594 selectively replicates in and destroys cancer cells driven by genetic pathways commonly activated in cancers*. Mol Ther, 2012. **20**(4): p. 749-58.
35. Kim, J.H., et al., *Systemic armed oncolytic and immunologic therapy for cancer with JX-594, a targeted poxvirus expressing GM-CSF*. Mol Ther, 2006. **14**(3): p. 361-70.
36. Heo, J., et al., *Randomized dose-finding clinical trial of oncolytic immunotherapeutic vaccinia JX-594 in liver cancer*. Nat Med, 2013. **19**(3): p. 329-36.
37. Moss, B., *Poxviridae: the viruses and their replication*. Fields virology, 4th ed. P.M.H.e. D. M. Knipe. 2001: Lippincott-Raven, Philadelphia, Pa. p. 2849-2884.
38. Upton, C., et al., *Poxvirus orthologous clusters: toward defining the minimum essential poxvirus genome*. J Virol, 2003. **77**(13): p. 7590-600.
39. Moss, B., *Regulation of vaccinia virus transcription*. Annu Rev Biochem, 1990. **59**: p. 661-88.
40. Beaud, G., *Vaccinia virus DNA replication: a short review*. Biochimie, 1995. **77**(10): p. 774-9.
41. Moss, B., *Poxvirus entry and membrane fusion*. Virology, 2006. **344**(1): p. 48-54.
42. Broyles, S.S., *Vaccinia virus transcription*. J Gen Virol, 2003. **84**(Pt 9): p. 2293-303.
43. Baldick, C.J., Jr. and B. Moss, *Characterization and temporal regulation of mRNAs encoded by vaccinia virus intermediate-stage genes*. J Virol, 1993. **67**(6): p. 3515-27.
44. Boone, R.F. and B. Moss, *Sequence complexity and relative abundance of vaccinia virus mRNA's synthesized in vivo and in vitro*. J Virol, 1978. **26**(3): p. 554-69.
45. Cyrklaff, M., et al., *Cryo-electron tomography of vaccinia virus*. Proc Natl Acad Sci U S A, 2005. **102**(8): p. 2772-7.
46. Mallardo, M., S. Schleich, and J. Krijnse Locker, *Microtubule-dependent organization of vaccinia virus core-derived early mRNAs into distinct cytoplasmic structures*. Mol Biol Cell, 2001. **12**(12): p. 3875-91.
47. Woodson, B., *Vaccinia mRNA synthesis under conditions which prevent uncoating*. Biochem Biophys Res Commun, 1967. **27**(2): p. 169-75.
48. Joklik, W.K., *The Intracellular Uncoating of Poxvirus DNA. II. The Molecular Basis of the Uncoating Process*. J Mol Biol, 1964. **8**: p. 277-88.
49. Rosales, R., et al., *Purification and identification of a vaccinia virus-encoded intermediate stage promoter-specific transcription factor that has homology to eukaryotic*

- transcription factor SII (TFIIS) and an additional role as a viral RNA polymerase subunit.* J Biol Chem, 1994. **269**(19): p. 14260-7.
50. Rosales, R., G. Sutter, and B. Moss, *A cellular factor is required for transcription of vaccinia viral intermediate-stage genes.* Proc Natl Acad Sci U S A, 1994. **91**(9): p. 3794-8.
  51. Sanz, P. and B. Moss, *Identification of a transcription factor, encoded by two vaccinia virus early genes, that regulates the intermediate stage of viral gene expression.* Proc Natl Acad Sci U S A, 1999. **96**(6): p. 2692-7.
  52. Harris, N., R. Rosales, and B. Moss, *Transcription initiation factor activity of vaccinia virus capping enzyme is independent of mRNA guanylation.* Proc Natl Acad Sci U S A, 1993. **90**(7): p. 2860-4.
  53. Cassetti, M.C., et al., *DNA packaging mutant: repression of the vaccinia virus A32 gene results in noninfectious, DNA-deficient, spherical, enveloped particles.* J Virol, 1998. **72**(7): p. 5769-80.
  54. Klemperer, N., et al., *The vaccinia virus I1 protein is essential for the assembly of mature virions.* J Virol, 1997. **71**(12): p. 9285-94.
  55. Grubisha, O. and P. Traktman, *Genetic analysis of the vaccinia virus I6 telomere-binding protein uncovers a key role in genome encapsidation.* J Virol, 2003. **77**(20): p. 10929-42.
  56. Unger, B. and P. Traktman, *Vaccinia virus morphogenesis:  $\alpha$ 13 phosphoprotein is required for assembly of mature virions.* J Virol, 2004. **78**(16): p. 8885-901.
  57. Sodeik, B., et al., *Assembly of vaccinia virus: role of the intermediate compartment between the endoplasmic reticulum and the Golgi stacks.* J Cell Biol, 1993. **121**(3): p. 521-41.
  58. Traktman, P., et al., *Temperature-sensitive mutants with lesions in the vaccinia virus F10 kinase undergo arrest at the earliest stage of virion morphogenesis.* J Virol, 1995. **69**(10): p. 6581-7.
  59. DeMasi, J. and P. Traktman, *Clustered charge-to-alanine mutagenesis of the vaccinia virus H5 gene: isolation of a dominant, temperature-sensitive mutant with a profound defect in morphogenesis.* J Virol, 2000. **74**(5): p. 2393-405.
  60. da Fonseca, F.G., et al., *Vaccinia virus mutants with alanine substitutions in the conserved G5R gene fail to initiate morphogenesis at the nonpermissive temperature.* J Virol, 2004. **78**(19): p. 10238-48.
  61. Resch, W., A.S. Weisberg, and B. Moss, *Vaccinia virus nonstructural protein encoded by the A11R gene is required for formation of the virion membrane.* J Virol, 2005. **79**(11): p. 6598-609.
  62. Condit, R.C., N. Moussatche, and P. Traktman, *In a nutshell: structure and assembly of the vaccinia virion.* Adv Virus Res, 2006. **66**: p. 31-124.
  63. Yeh, W.W., B. Moss, and E.J. Wolffe, *The vaccinia virus A9L gene encodes a membrane protein required for an early step in virion morphogenesis.* J Virol, 2000. **74**(20): p. 9701-11.
  64. Ravello, M.P. and D.E. Hruby, *Conditional lethal expression of the vaccinia virus L1R myristylated protein reveals a role in virion assembly.* J Virol, 1994. **68**(10): p. 6401-10.
  65. da Fonseca, F.G., et al., *Effects of deletion or stringent repression of the H3L envelope gene on vaccinia virus replication.* J Virol, 2000. **74**(16): p. 7518-28.
  66. Lin, C.L., et al., *Vaccinia virus envelope H3L protein binds to cell surface heparan sulfate and is important for intracellular mature virion morphogenesis and virus infection in vitro and in vivo.* J Virol, 2000. **74**(7): p. 3353-65.
  67. Schmelz, M., et al., *Assembly of vaccinia virus: the second wrapping cisterna is derived from the trans Golgi network.* J Virol, 1994. **68**(1): p. 130-47.
  68. Ichihashi, Y., S. Matsumoto, and S. Dales, *Biogenesis of poxviruses: role of A-type inclusions and host cell membranes in virus dissemination.* Virology, 1971. **46**(3): p. 507-32.
  69. Morgan, C., *Vaccinia virus reexamined: development and release.* Virology, 1976. **73**(1): p. 43-58.
  70. Payne, L.G. and K. Kristenson, *Mechanism of vaccinia virus release and its specific inhibition by N1-isonicotinoyl-N2-3-methyl-4-chlorobenzoylhydrazine.* J Virol, 1979. **32**(2): p. 614-22.

71. Hiramatsu, Y., et al., *Poxvirus virions: their surface ultrastructure and interaction with the surface membrane of host cells*. J Electron Microsc (Tokyo), 1999. **48**(6): p. 937-46.
72. Meiser, A., C. Sancho, and J. Krijnsse Locker, *Plasma membrane budding as an alternative release mechanism of the extracellular enveloped form of vaccinia virus from HeLa cells*. J Virol, 2003. **77**(18): p. 9931-42.
73. Tsutsui, K., *Release of vaccinia virus from FL cells infected with the IHD-W strain*. J Electron Microsc (Tokyo), 1983. **32**(2): p. 125-40.
74. Blasco, R. and B. Moss, *Role of cell-associated enveloped vaccinia virus in cell-to-cell spread*. J Virol, 1992. **66**(7): p. 4170-9.
75. Pennington, T.H. and E.A. Follett, *Vaccinia virus replication in enucleate BSC-1 cells: particle production and synthesis of viral DNA and proteins*. J Virol, 1974. **13**(2): p. 488-93.
76. Joklik, W.K. and Y. Becker, *The Replication and Coating of Vaccinia DNA*. J Mol Biol, 1964. **10**: p. 452-74.
77. Pogo, B.G., E.M. Berkowitz, and S. Dales, *Investigation of vaccinia virus DNA replication employing a conditional lethal mutant defective in DNA*. Virology, 1984. **132**(2): p. 436-44.
78. Pogo, B.G., M. O'Shea, and P. Freimuth, *Initiation and termination of vaccinia virus DNA replication*. Virology, 1981. **108**(1): p. 241-8.
79. Moyer, R.W. and R.L. Graves, *The mechanism of cytoplasmic orthopoxvirus DNA replication*. Cell, 1981. **27**(2 Pt 1): p. 391-401.
80. Merchlinsky, M., C.F. Garon, and B. Moss, *Molecular cloning and sequence of the concatemer junction from vaccinia virus replicative DNA. Viral nuclease cleavage sites in cruciform structures*. J Mol Biol, 1988. **199**(3): p. 399-413.
81. DeLange, A.M. and G. McFadden, *The role of telomeres in poxvirus DNA replication*. Curr Top Microbiol Immunol, 1990. **163**: p. 71-92.
82. DeLange, A.M., *Identification of temperature-sensitive mutants of vaccinia virus that are defective in conversion of concatemeric replicative intermediates to the mature linear DNA genome*. J Virol, 1989. **63**(6): p. 2437-44.
83. Merchlinsky, M. and B. Moss, *Resolution of vaccinia virus DNA concatemer junctions requires late-gene expression*. J Virol, 1989. **63**(4): p. 1595-603.
84. Hu, F.Q. and D.J. Pickup, *Transcription of the terminal loop region of vaccinia virus DNA is initiated from the telomere sequences directing DNA resolution*. Virology, 1991. **181**(2): p. 716-20.
85. Shuman, S. and B. Moss, *Identification of a vaccinia virus gene encoding a type I DNA topoisomerase*. Proc Natl Acad Sci U S A, 1987. **84**(21): p. 7478-82.
86. Klemperer, N. and P. Traktman, *Biochemical analysis of mutant alleles of the vaccinia virus topoisomerase I carrying targeted substitutions in a highly conserved domain*. J Biol Chem, 1993. **268**(21): p. 15887-99.
87. Eckert, D., et al., *Vaccinia virus nicking-joining enzyme is encoded by K4L (VACWR035)*. J Virol, 2005. **79**(24): p. 15084-90.
88. Garcia, A.D. and B. Moss, *Repression of vaccinia virus Holliday junction resolvase inhibits processing of viral DNA into unit-length genomes*. J Virol, 2001. **75**(14): p. 6460-71.
89. Garcia, A.D., et al., *Quaternary structure and cleavage specificity of a poxvirus holliday junction resolvase*. J Biol Chem, 2006. **281**(17): p. 11618-26.
90. De Silva, F.S., N. Paran, and B. Moss, *Products and substrate/template usage of vaccinia virus DNA primase*. Virology, 2009. **383**(1): p. 136-41.
91. Da Silva, M., et al., *Predicted function of the vaccinia virus G5R protein*. Bioinformatics, 2006. **22**(23): p. 2846-50.
92. Fenner, F. and B.M. Comben, *[Genetic studies with mammalian poxviruses. I. Demonstration of recombination between two strains of vaccinia virus]*. Virology, 1958. **5**(3): p. 530-48.
93. Block, W., C. Upton, and G. McFadden, *Tumorigenic poxviruses: genomic organization of malignant rabbit virus, a recombinant between Shope fibroma virus and myxoma virus*. Virology, 1985. **140**(1): p. 113-24.
94. Gershon, P.D., et al., *Poxvirus genetic recombination during natural virus transmission*. J Gen Virol, 1989. **70** ( Pt 2): p. 485-9.

95. Hertig, C., et al., *Field and vaccine strains of fowlpox virus carry integrated sequences from the avian retrovirus, reticuloendotheliosis virus*. *Virology*, 1997. **235**(2): p. 367-76.
96. McFadden, G. and S. Dales, *Biogenesis of poxviruses: mirror-image deletions in vaccinia virus DNA*. *Cell*, 1979. **18**(1): p. 101-8.
97. Moyer, R.W., R.L. Graves, and C.T. Rothe, *The white pock (mu) mutants of rabbit poxvirus. III. Terminal DNA sequence duplication and transposition in rabbit poxvirus*. *Cell*, 1980. **22**(2 Pt 2): p. 545-53.
98. Esposito, J.J., et al., *Intragenomic sequence transposition in monkeypox virus*. *Virology*, 1981. **109**(2): p. 231-43.
99. Archard, L.C., et al., *The genome structure of cowpox virus white pock variants*. *J Gen Virol*, 1984. **65** ( Pt 5): p. 875-86.
100. Kotwal, G.J. and B. Moss, *Analysis of a large cluster of nonessential genes deleted from a vaccinia virus terminal transposition mutant*. *Virology*, 1988. **167**(2): p. 524-37.
101. Paez, E., S. Dallo, and M. Esteban, *Generation of a dominant 8-MDa deletion at the left terminus of vaccinia virus DNA*. *Proc Natl Acad Sci U S A*, 1985. **82**(10): p. 3365-9.
102. Moss, B., E. Winters, and N. Cooper, *Instability and reiteration of DNA sequences within the vaccinia virus genome*. *Proc Natl Acad Sci U S A*, 1981. **78**(3): p. 1614-8.
103. Slabaugh, M.B., N.A. Roseman, and C.K. Mathews, *Amplification of the ribonucleotide reductase small subunit gene: analysis of novel joints and the mechanism of gene duplication in vaccinia virus*. *Nucleic Acids Res*, 1989. **17**(17): p. 7073-88.
104. Elde, N.C., et al., *Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses*. *Cell*, 2012. **150**(4): p. 831-41.
105. Moss, B., *Vaccinia virus: a tool for research and vaccine development*. *Science*, 1991. **252**(5013): p. 1662-7.
106. Merchlinsky, M., *Intramolecular homologous recombination in cells infected with temperature-sensitive mutants of vaccinia virus*. *J Virol*, 1989. **63**(5): p. 2030-5.
107. Willer, D.O., et al., *Vaccinia virus DNA polymerase promotes DNA pairing and strand-transfer reactions*. *Virology*, 1999. **257**(2): p. 511-23.
108. Gammon, D.B. and D.H. Evans, *The 3'-to-5' exonuclease activity of vaccinia virus DNA polymerase is essential and plays a role in promoting virus genetic recombination*. *J Virol*, 2009. **83**(9): p. 4236-50.
109. Willer, D.O., et al., *In vitro concatemer formation catalyzed by vaccinia virus DNA polymerase*. *Virology*, 2000. **278**(2): p. 562-569.
110. Lyndaker, A.M. and E. Alani, *A tale of tails: insights into the coordination of 3' end processing during homologous recombination*. *Bioessays*, 2009. **31**(3): p. 315-21.
111. Cromie, G.A., J.C. Connelly, and D.R. Leach, *Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans*. *Mol Cell*, 2001. **8**(6): p. 1163-74.
112. Reuven, N.B., et al., *The herpes simplex virus type 1 alkaline nuclease and single-stranded DNA binding protein mediate strand exchange in vitro*. *J Virol*, 2003. **77**(13): p. 7425-33.
113. Paques, F. and J.E. Haber, *Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae**. *Microbiol Mol Biol Rev*, 1999. **63**(2): p. 349-404.
114. Helleday, T., et al., *DNA double-strand break repair: from mechanistic understanding to cancer treatment*. *DNA Repair (Amst)*, 2007. **6**(7): p. 923-35.
115. Esposito, J.J., J.F. Objieski, and J.H. Nakano, *Orthopoxvirus DNA: strain differentiation by electrophoresis of restriction endonuclease fragmented virion DNA*. *Virology*, 1978. **89**(1): p. 53-66.
116. Maxam, A.M. and W. Gilbert, *A new method for sequencing DNA*. *Proc Natl Acad Sci U S A*, 1977. **74**(2): p. 560-4.
117. Shchelkunov, S.N., et al., *Analysis of the nucleotide sequence of a 43 kbp segment of the genome of variola virus India-1967 strain*. *Virus Res*, 1993. **30**(3): p. 239-58.
118. Shchelkunov, S.N., et al., *[Entire coding sequence of the variola virus]*. *Dokl Akad Nauk*, 1993. **328**(5): p. 629-32.
119. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors*. *Proc Natl Acad Sci U S A*, 1977. **74**(12): p. 5463-7.

120. Goebel, S.J., et al., *The complete DNA sequence of vaccinia virus*. *Virology*, 1990. **179**(1): p. 247-66, 517-63.
121. Massung, R.F., et al., *Potential virulence determinants in terminal regions of variola smallpox virus genome*. *Nature*, 1993. **366**(6457): p. 748-51.
122. Massung, R.F., et al., *Analysis of the complete genome of smallpox variola major virus strain Bangladesh-1975*. *Virology*, 1994. **201**(2): p. 215-40.
123. Cameron, C., et al., *The complete DNA sequence of myxoma virus*. *Virology*, 1999. **264**(2): p. 298-318.
124. Willer, D.O., G. McFadden, and D.H. Evans, *The complete genome sequence of Shope (rabbit) fibroma virus*. *Virology*, 1999. **264**(2): p. 319-43.
125. Wetterstrand, K., *DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)*. Available at: [www.genome.gov/sequencingcosts](http://www.genome.gov/sequencingcosts). National Human Genome Research Institute., 2013.
126. Margulies, M., et al., *Genome sequencing in microfabricated high-density picolitre reactors*. *Nature*, 2005. **437**(7057): p. 376-80.
127. Mardis, E.R., *Next-generation DNA sequencing methods*. *Annu Rev Genomics Hum Genet*, 2008. **9**: p. 387-402.
128. Smith, G.L., et al., *Vaccinia virus immune evasion: mechanisms, virulence and immunogenicity*. *J Gen Virol*, 2013. **94**(Pt 11): p. 2367-92.
129. Seet, B.T., et al., *Poxviruses and immune evasion*. *Annu Rev Immunol*, 2003. **21**: p. 377-423.
130. Smith, S.A. and G.J. Kotwal, *Immune response to poxvirus infections in various animals*. *Crit Rev Microbiol*, 2002. **28**(3): p. 149-85.
131. Alcami, A. and U.H. Koszinowski, *Viral mechanisms of immune evasion*. *Immunol Today*, 2000. **21**(9): p. 447-55.
132. Symons, J.A., A. Alcami, and G.L. Smith, *Vaccinia virus encodes a soluble type I interferon receptor of novel structure and broad species specificity*. *Cell*, 1995. **81**(4): p. 551-60.
133. Vancova, I., C. La Bonnardiére, and P. Kontsek, *Vaccinia virus protein B18R inhibits the activity and cellular binding of the novel type interferon-delta*. *J Gen Virol*, 1998. **79** ( Pt 7): p. 1647-9.
134. Alcami, A., J.A. Symons, and G.L. Smith, *The vaccinia virus soluble alpha/beta interferon (IFN) receptor binds to the cell surface and protects cells from the antiviral effects of IFN*. *J Virol*, 2000. **74**(23): p. 11230-9.
135. Mossman, K., et al., *Species specificity of ectromelia virus and vaccinia virus interferon-gamma binding proteins*. *Virology*, 1995. **208**(2): p. 762-9.
136. Alcami, A. and G.L. Smith, *Vaccinia, cowpox, and camelpox viruses encode soluble gamma interferon receptors with novel broad species specificity*. *J Virol*, 1995. **69**(8): p. 4633-9.
137. Symons, J.A., et al., *A study of the vaccinia virus interferon-gamma receptor and its contribution to virus virulence*. *J Gen Virol*, 2002. **83**(Pt 8): p. 1953-64.
138. Sroller, V., et al., *Effect of IFN-gamma receptor gene deletion on vaccinia virus virulence*. *Arch Virol*, 2001. **146**(2): p. 239-49.
139. Alcami, A. and G.L. Smith, *A soluble receptor for interleukin-1 beta encoded by vaccinia virus: a novel mechanism of virus modulation of the host response to infection*. *Cell*, 1992. **71**(1): p. 153-67.
140. Symons, J.A., et al., *The vaccinia virus C12L protein inhibits mouse IL-18 and promotes virus virulence in the murine intranasal model*. *J Gen Virol*, 2002. **83**(Pt 11): p. 2833-44.
141. McKenzie, R., et al., *Regulation of complement activity by vaccinia virus complement-control protein*. *J Infect Dis*, 1992. **166**(6): p. 1245-50.
142. Alcami, A., et al., *Vaccinia virus strains Lister, USSR and Evans express soluble and cell-surface tumour necrosis factor receptors*. *J Gen Virol*, 1999. **80** ( Pt 4): p. 949-59.
143. Alcami, A., et al., *Blockade of chemokine activity by a soluble chemokine binding protein from vaccinia virus*. *J Immunol*, 1998. **160**(2): p. 624-33.
144. Ng, A., et al., *The vaccinia virus A41L protein is a soluble 30 kDa glycoprotein that affects virus virulence*. *J Gen Virol*, 2001. **82**(Pt 9): p. 2095-105.

145. Parrish, S. and B. Moss, *Characterization of a vaccinia virus mutant with a deletion of the D10R gene encoding a putative negative regulator of gene expression*. J Virol, 2006. **80**(2): p. 553-61.
146. Parrish, S. and B. Moss, *Characterization of a second vaccinia virus mRNA-decapping enzyme conserved in poxviruses*. J Virol, 2007. **81**(23): p. 12973-8.
147. Parrish, S., W. Resch, and B. Moss, *Vaccinia virus D10 protein has mRNA decapping activity, providing a mechanism for control of host and viral gene expression*. Proc Natl Acad Sci U S A, 2007. **104**(7): p. 2139-44.
148. Moss, B. and N.P. Salzman, *Sequential protein synthesis following vaccinia virus infection*. J Virol, 1968. **2**(10): p. 1016-27.
149. Chang, H.W., J.C. Watson, and B.L. Jacobs, *The E3L gene of vaccinia virus encodes an inhibitor of the interferon-induced, double-stranded RNA-dependent protein kinase*. Proc Natl Acad Sci U S A, 1992. **89**(11): p. 4825-9.
150. Ferguson, B.J., et al., *DNA-PK is a DNA sensor for IRF-3-dependent innate immunity*. Elife, 2012. **1**: p. e00047.
151. Peters, N.E., et al., *A mechanism for the inhibition of DNA-PK-mediated DNA sensing by a virus*. PLoS Pathog, 2013. **9**(10): p. e1003649.
152. Mann, B.A., 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): p. 367-80.
153. Najarro, P., P. Traktman, and J.A. Lewis, *Vaccinia virus blocks gamma interferon signal transduction: viral VH1 phosphatase reverses Stat1 activation*. J Virol, 2001. **75**(7): p. 3185-96.
154. Guerra, S., et al., *Vaccinia virus E3 protein prevents the antiviral action of ISG15*. PLoS Pathog, 2008. **4**(7): p. e1000096.
155. Carroll, K., et al., *Recombinant vaccinia virus K3L gene product prevents activation of double-stranded RNA-dependent, initiation factor 2 alpha-specific protein kinase*. J Biol Chem, 1993. **268**(17): p. 12837-42.
156. Beattie, E., J. Tartaglia, and E. Paoletti, *Vaccinia virus-encoded eIF-2 alpha homolog abrogates the antiviral effect of interferon*. Virology, 1991. **183**(1): p. 419-22.
157. Tait, S.W. and D.R. Green, *Mitochondria and cell death: outer membrane permeabilization and beyond*. Nat Rev Mol Cell Biol, 2010. **11**(9): p. 621-32.
158. Kvsanakul, M., et al., *Vaccinia virus anti-apoptotic F1L is a novel Bcl-2-like domain-swapped dimer that binds a highly selective subset of BH3-containing death ligands*. Cell Death Differ, 2008. **15**(10): p. 1564-71.
159. Postigo, A., et al., *Interaction of F1L with the BH3 domain of Bak is responsible for inhibiting vaccinia-induced apoptosis*. Cell Death Differ, 2006. **13**(10): p. 1651-62.
160. Aoyagi, M., et al., *Vaccinia virus N1L protein resembles a B cell lymphoma-2 (Bcl-2) family protein*. Protein Sci, 2007. **16**(1): p. 118-24.
161. Cooray, S., et al., *Functional and structural studies of the vaccinia virus virulence factor N1 reveal a Bcl-2-like anti-apoptotic protein*. J Gen Virol, 2007. **88**(Pt 6): p. 1656-66.
162. Kettle, S., et al., *Vaccinia virus serpins B13R (SPI-2) and B22R (SPI-1) encode M(r) 38.5 and 40K, intracellular polypeptides that do not affect virus virulence in a murine intranasal model*. Virology, 1995. **206**(1): p. 136-47.
163. Gubser, C., et al., *A new inhibitor of apoptosis from vaccinia virus and eukaryotes*. PLoS Pathog, 2007. **3**(2): p. e17.
164. de Mattia, F., et al., *Human Golgi antiapoptotic protein modulates intracellular calcium fluxes*. Mol Biol Cell, 2009. **20**(16): p. 3638-45.
165. Osborne, J.D., et al., *Genomic differences of Vaccinia virus clones from Dryvax smallpox vaccine: the Dryvax-like ACAM2000 and the mouse neurovirulent Clone-3*. Vaccine, 2007. **25**(52): p. 8807-32.
166. Tulman, E.R., et al., *Genome of horsepox virus*. J Virol, 2006. **80**(18): p. 9244-58.
167. Jin, Q., et al., *Characterization of the complete genomic sequence of the vaccinia virus Tian Tan strain*. Sci China C, 1997. **27**: p. 562-67.
168. Li, G., et al., *Complete coding sequences of the rabbitpox virus genome*. J Gen Virol, 2005. **86**(Pt 11): p. 2969-77.

169. Zhang, Q., et al., *The highly attenuated oncolytic recombinant vaccinia virus GLV-1h68: comparative genomic features and the contribution of F14.5L inactivation*. Mol Genet Genomics, 2009. **282**(4): p. 417-35.
170. Meisinger-Henschel, C., et al., *Genomic sequence of chorioallantois vaccinia virus Ankara, the ancestor of modified vaccinia virus Ankara*. J Gen Virol, 2007. **88**(Pt 12): p. 3249-59.
171. Coulson, D. and C. Upton, *Characterization of indels in poxvirus genomes*. Virus Genes, 2011. **42**(2): p. 171-7.
172. Feschenko, V.V. and S.T. Lovett, *Slipped misalignment mechanisms of deletion formation: analysis of deletion endpoints*. J Mol Biol, 1998. **276**(3): p. 559-69.
173. Lovett, S.T., *Encoded errors: mutations and rearrangements mediated by misalignment at repetitive DNA sequences*. Mol Microbiol, 2004. **52**(5): p. 1243-53.
174. Trinh, T.Q. and R.R. Sinden, *Preferential DNA secondary structure mutagenesis in the lagging strand of replication in E. coli*. Nature, 1991. **352**(6335): p. 544-7.
175. Weston-Hafer, K. and D.E. Berg, *Deletions in plasmid pBR322: replication slippage involving leading and lagging strands*. Genetics, 1991. **127**(4): p. 649-55.
176. Chen, N., et al., *The genomic sequence of ectromelia virus, the causative agent of mousepox*. Virology, 2003. **317**(1): p. 165-86.
177. Massung, R.F., J.C. Knight, and J.J. Esposito, *Topography of variola smallpox virus inverted terminal repeats*. Virology, 1995. **211**(1): p. 350-5.
178. Stuart, D., et al., *The target DNA sequence for resolution of poxvirus replicative intermediates is an active late promoter*. J Virol, 1991. **65**(1): p. 61-70.
179. Parsons, B.L. and D.J. Pickup, *Transcription of orthopoxvirus telomeres at late times during infection*. Virology, 1990. **175**(1): p. 69-80.
180. Esposito, J.J. and J.C. Knight, *Orthopoxvirus DNA: a comparison of restriction profiles and maps*. Virology, 1985. **143**(1): p. 230-51.
181. Baroudy, B.M. and B. Moss, *Sequence homologies of diverse length tandem repetitions near ends of vaccinia virus genome suggest unequal crossing over*. Nucleic Acids Res, 1982. **10**(18): p. 5673-9.
182. Wittek, R., H.K. Muller, and R. Wyler, *Length heterogeneity in the DNA of vaccinia virus is eliminated on cloning the virus*. FEBS Lett, 1978. **90**(1): p. 41-6.
183. McCarron, R.J., et al., *Structure of vaccinia DNA: analysis of the viral genome by restriction endonucleases*. Virology, 1978. **86**(1): p. 88-101.
184. Wittek, R., et al., *Inverted terminal repeats in rabbit poxvirus and vaccinia virus DNA*. J Virol, 1978. **28**(1): p. 171-81.
185. Takahashi-Nishimaki, F., et al., *Genetic analysis of vaccinia virus Lister strain and its attenuated mutant LC16m8: production of intermediate variants by homologous recombination*. J Gen Virol, 1987. **68** ( Pt 10): p. 2705-10.
186. Weltzin, R., et al., *Clonal vaccinia virus grown in cell culture as a new smallpox vaccine*. Nat Med, 2003. **9**(9): p. 1125-30.
187. Parker, R.F., L.H. Bronson, and R.H. Green, *Further studies of the infectious unit of vaccinia*. J Exp Med, 1941. **74**: p. 263-281.
188. Fenner, F., et al., *Smallpox and its eradication*. 1988, Geneva: World Health Organization. 1460.
189. Monath, T.P., et al., *ACAM2000 clonal Vero cell culture vaccinia virus (New York City Board of Health strain)--a second-generation smallpox vaccine for biological defense*. Int J Infect Dis, 2004. **8 Suppl 2**: p. S31-44.
190. Frey, S.E., et al., *Comparison of the safety and immunogenicity of ACAM1000, ACAM2000 and Dryvax in healthy vaccinia-naive adults*. Vaccine, 2009. **27**(10): p. 1637-44.
191. Li, G., et al., *Genomic sequence and analysis of a vaccinia virus isolate from a patient with a smallpox vaccine-related complication*. Virol J, 2006. **3**: p. 88.
192. Garcel, A., et al., *Phenotypic and genetic diversity of the traditional Lister smallpox vaccine*. Vaccine, 2009. **27**(5): p. 708-17.
193. Abramoff, M.D., Magalhaes, P.J., Ram, S.J., *Image Processing with ImageJ*. Biophotonics International, 2004. **11**(7): p. 36-42.
194. Lee, H.J., K. Essani, and G.L. Smith, *The genome sequence of Yaba-like disease virus, a yatapoxvirus*. Virology, 2001. **281**(2): p. 170-92.

195. Upton, C., et al., *Viral genome organizer: a system for analyzing complete viral genomes*. Virus Res, 2000. **70**(1-2): p. 55-64.
196. Lefkowitz, E.J., et al., *Poxvirus Bioinformatics Resource Center: a comprehensive Poxviridae informational and analytical resource*. Nucleic Acids Res, 2005. **33**(Database issue): p. D311-6.
197. Ehlers, A., et al., *Poxvirus Orthologous Clusters (POCs)*. Bioinformatics, 2002. **18**(11): p. 1544-5.
198. Brudno, M., et al., *LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA*. Genome Res, 2003. **13**(4): p. 721-31.
199. Brodie, R., et al., *Base-By-Base: single nucleotide-level analysis of whole viral genome alignments*. BMC Bioinformatics, 2004. **5**: p. 96.
200. Martin, D.P., et al., *RDP3: a flexible and fast computer program for analyzing recombination*. Bioinformatics, 2010. **26**(19): p. 2462-3.
201. Page, R.D., *TreeView: an application to display phylogenetic trees on personal computers*. Comput Appl Biosci, 1996. **12**(4): p. 357-8.
202. Lole, K.S., et al., *Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination*. J Virol, 1999. **73**(1): p. 152-60.
203. Tcherepanov, V., A. Ehlers, and C. Upton, *Genome Annotation Transfer Utility (GATU): rapid annotation of viral genomes using a closely related reference genome*. BMC Genomics, 2006. **7**: p. 150.
204. Rutherford, K., et al., *Artemis: sequence visualization and annotation*. Bioinformatics, 2000. **16**(10): p. 944-5.
205. Baroudy, B.M., S. Venkatesan, and B. Moss, *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): p. 315-24.
206. Yang, Z., et al., *Simultaneous high-resolution analysis of vaccinia virus and host cell transcriptomes by deep RNA sequencing*. Proc Natl Acad Sci U S A, 2010. **107**(25): p. 11513-8.
207. Kretzschmar, M., et al., *Frequency of adverse events after vaccination with different vaccinia strains*. PLoS Med, 2006. **3**(8): p. e272.
208. Gammon, D.B., et al., *Vaccinia virus-encoded ribonucleotide reductase subunits are differentially required for replication and pathogenesis*. PLoS Pathog, 2010. **6**(7): p. e1000984.
209. Sutherland, B.M., et al., *Clustered damages and total lesions induced in DNA by ionizing radiation: oxidized bases and strand breaks*. Biochemistry, 2000. **39**(27): p. 8026-31.
210. Wyatt, L.S., et al., *Elucidating and minimizing the loss by recombinant vaccinia virus of human immunodeficiency virus gene expression resulting from spontaneous mutations and positive selection*. J Virol, 2009. **83**(14): p. 7176-84.
211. Shuman, S. and J. Prescott, *Specific DNA cleavage and binding by vaccinia virus DNA topoisomerase I*. J Biol Chem, 1990. **265**(29): p. 17826-36.
212. Colamonici, O.R., et al., *Vaccinia virus B18R gene encodes a type I interferon-binding protein that blocks interferon alpha transmembrane signaling*. J Biol Chem, 1995. **270**(27): p. 15974-8.
213. Silva, D.C., et al., *Clinical signs, diagnosis, and case reports of Vaccinia virus infections*. Braz J Infect Dis, 2010. **14**(2): p. 129-34.
214. Bhanuprakash, V., et al., *Zoonotic infections of buffalopox in India*. Zoonoses Public Health, 2010. **57**(7-8): p. e149-55.
215. Salminen, M.O., et al., *Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning*. AIDS Res Hum Retroviruses, 1995. **11**(11): p. 1423-5.
216. Lin, Y.C. and D.H. Evans, *Vaccinia virus particles mix inefficiently, and in a way that would restrict viral recombination, in coinfecting cells*. J Virol, 2010. **84**(5): p. 2432-43.
217. Yao, X.D. and D.H. Evans, *Characterization of the recombinant joints formed by single-strand annealing reactions in vaccinia virus-infected cells*. Virology, 2003. **308**(1): p. 147-56.
218. Dong, S.L., *Qi Changqing is the founder of Tiantan strain of vaccinia virus*. Progress In Microbiology and Immunology, 2009. **37**(3): p. 1-3.

219. Qing, N.S., *Qi Changqing is an eyewitness to the development of Chinese biological products*. Wei Sheng Wu Xue Bao, 2006. **46**(3): p. 339-340.
220. Zhao, K., *Medical Biopreparatics*. second ed. 2007, Beijing: People's Medical Publishing House. 966-967.
221. Bray, M., *Pathogenesis and potential antiviral therapy of complications of smallpox vaccination*. Antiviral Res, 2003. **58**(2): p. 101-14.
222. Crewdson, J., *Science fictions : a scientific mystery, a massive coverup, and the dark legacy of Robert Gallo*. 1st ed. 2002, Boston: Little, Brown. xvii, 670 p. (16 p. of plates).
223. Qin, L., et al., *Genomic analysis of the vaccinia virus strain variants found in Dryvax vaccine*. J Virol, 2011. **85**(24): p. 13049-60.
224. Buller, R.M., et al., *Deletion of the vaccinia virus growth factor gene reduces virus virulence*. J Virol, 1988. **62**(3): p. 866-74.
225. Schneider, C.A., W.S. Rasband, and K.W. Eliceiri, *NIH Image to ImageJ: 25 years of image analysis*. Nat Methods, 2012. **9**(7): p. 671-5.
226. McCart, J.A., et al., *Systemic cancer therapy with a tumor-selective vaccinia virus mutant lacking thymidine kinase and vaccinia growth factor genes*. Cancer Res, 2001. **61**(24): p. 8751-7.
227. Hendrickson, R.C., et al., *Orthopoxvirus genome evolution: the role of gene loss*. Viruses, 2010. **2**(9): p. 1933-67.
228. Baxby, D., *Jenner's smallpox vaccine : the riddle of vaccinia virus and its origin*. 1981, London: Heinemann Educational.
229. Hershey, A.D. and R. Rotman, *Linkage Among Genes Controlling Inhibition of Lysis in a Bacterial Virus*. Proc Natl Acad Sci U S A, 1948. **34**(3): p. 89-96.
230. Wildy, P., *Recombination with herpes simplex virus*. J Gen Microbiol, 1955. **13**(2): p. 346-60.
231. Woodroffe, G.M. and F. Fenner, *Genetic studies with mammalian poxviruses. IV. Hybridization between several different poxviruses*. Virology, 1960. **12**: p. 272-82.
232. Bedson, H.S. and K.R. Dumbell, *Hybrids Derived from the Viruses of Variola Major and Cowpox*. J Hyg (Lond), 1964. **62**: p. 147-58.
233. Bedson, H.S. and K.R. Dumbell, *Hybrids Derived from the Viruses of Alastrim and Rabbit Pox*. J Hyg (Lond), 1964. **62**: p. 141-6.
234. Strayer, D.S., et al., *Malignant rabbit fibroma virus causes secondary immunosuppression in rabbits*. J Immunol, 1983. **130**(1): p. 399-404.
235. Esposito, J.J., et al., *Genome sequence diversity and clues to the evolution of variola (smallpox) virus*. Science, 2006. **313**(5788): p. 807-12.
236. Ensinger, M.J. and M. Rovinsky, *Marker rescue of temperature-sensitive mutations of vaccinia virus WR: correlation of genetic and physical maps*. J Virol, 1983. **48**(2): p. 419-28.
237. Ensinger, M.J., *Isolation and genetic characterization of temperature-sensitive mutants of vaccinia virus WR*. J Virol, 1982. **43**(3): p. 778-90.
238. Condit, R.C., A. Motyczka, and G. Spizz, *Isolation, characterization, and physical mapping of temperature-sensitive mutants of vaccinia virus*. Virology, 1983. **128**(2): p. 429-43.
239. Panicali, D. and E. Paoletti, *Construction of poxviruses as cloning vectors: insertion of the thymidine kinase gene from herpes simplex virus into the DNA of infectious vaccinia virus*. Proc Natl Acad Sci U S A, 1982. **79**(16): p. 4927-31.
240. Mackett, M., G.L. Smith, and B. Moss, *Vaccinia virus: a selectable eukaryotic cloning and expression vector*. Proc Natl Acad Sci U S A, 1982. **79**(23): p. 7415-9.
241. Ball, L.A., *High-frequency homologous recombination in vaccinia virus DNA*. J Virol, 1987. **61**(6): p. 1788-95.
242. Ball, L.A., *Fidelity of homologous recombination in vaccinia virus DNA*. Virology, 1995. **209**(2): p. 688-91.
243. Evans, D.H., D. Stuart, and G. McFadden, *High levels of genetic recombination among cotransfected plasmid DNAs in poxvirus-infected mammalian cells*. J Virol, 1988. **62**(2): p. 367-75.
244. Condit, R.C., *Vaccinia, Inc.--probing the functional substructure of poxviral replication factories*. Cell Host Microbe, 2007. **2**(4): p. 205-7.

245. Yao, X.D. and D.H. Evans, *High-frequency genetic recombination and reactivation of orthopoxviruses from DNA fragments transfected into leporipoxvirus-infected cells*. J Virol, 2003. **77**(13): p. 7281-90.
246. Qin, L., M. Liang, and D.H. Evans, *Genomic analysis of vaccinia virus strain TianTan provides new insights into the evolution and evolutionary relationships between Orthopoxviruses*. Virology, 2013. **442**(1): p. 59-66.
247. Goodman, M.F. and K.D. Fygenon, *DNA polymerase fidelity: from genetics toward a biochemical understanding*. Genetics, 1998. **148**(4): p. 1475-82.
248. Taddie, J.A. and P. Traktman, *Genetic characterization of the vaccinia virus DNA polymerase: identification of point mutations conferring altered drug sensitivities and reduced fidelity*. J Virol, 1991. **65**(2): p. 869-79.
249. Taddie, J.A. and P. Traktman, *Genetic characterization of the vaccinia virus DNA polymerase: cytosine arabinoside resistance requires a variable lesion conferring phosphonoacetate resistance in conjunction with an invariant mutation localized to the 3'-5' exonuclease domain*. J Virol, 1993. **67**(7): p. 4323-36.
250. Andrei, G., et al., *Cidofovir resistance in vaccinia virus is linked to diminished virulence in mice*. J Virol, 2006. **80**(19): p. 9391-401.
251. Fathi, Z., et al., *Intragenic and intergenic recombination between temperature-sensitive mutants of vaccinia virus*. J Gen Virol, 1991. **72 ( Pt 11)**: p. 2733-7.
252. Parks, R.J. and D.H. Evans, *Effect of marker distance and orientation on recombinant formation in poxvirus-infected cells*. J Virol, 1991. **65**(3): p. 1263-72.
253. DeLange, A.M. and G. McFadden, *Sequence-nonspecific replication of transfected plasmid DNA in poxvirus- infected cells*. Proc Natl Acad Sci U S A, 1986. **83**(3): p. 614-8.
254. Fisher, C., et al., *Heteroduplex DNA formation is associated with replication and recombination in poxvirus-infected cells*. Genetics, 1991. **129**(1): p. 7-18.
255. Maiti, A. and A.C. Drohat, *Dependence of substrate binding and catalysis on pH, ionic strength, and temperature for thymine DNA glycosylase: Insights into recognition and processing of G.T mismatches*. DNA Repair (Amst), 2011. **10**(5): p. 545-53.
256. Froggatt, G.C., G.L. Smith, and P.M. Beard, *Vaccinia virus gene F3L encodes an intracellular protein that affects the innate immune response*. J Gen Virol, 2007. **88**(Pt 7): p. 1917-21.
257. Murcia-Nicolas, A., et al., *Identification by mass spectroscopy of three major early proteins associated with virosomes in vaccinia virus-infected cells*. Virus Res, 1999. **59**(1): p. 1-12.
258. Antoine, G., et al., *The complete genomic sequence of the modified vaccinia Ankara strain: comparison with other orthopoxviruses*. Virology, 1998. **244**(2): p. 365-96.
259. Scheiflinger, F., F. Dorner, and F.G. Falkner, *Construction of chimeric vaccinia viruses by molecular cloning and packaging*. Proc Natl Acad Sci U S A, 1992. **89**(21): p. 9977-81.
260. Carroll, D.S., et al., *Chasing Jenner's vaccine: revisiting cowpox virus classification*. PLoS One, 2011. **6**(8): p. e23086.
261. Shchelkunov, S.N., et al., *The genomic sequence analysis of the left and right species-specific terminal region of a cowpox virus strain reveals unique sequences and a cluster of intact ORFs for immunomodulatory and host range proteins*. Virology, 1998. **243**(2): p. 432-60.
262. Baxby, D., *Edward Jenner, William Woodville, and the origins of vaccinia virus*. J Hist Med Allied Sci, 1979. **34**(2): p. 134-62.
263. Baxby, D., *The origins of vaccinia virus*. J Infect Dis, 1977. **136**(3): p. 453-5.
264. Geddes, A.M., *The history of smallpox*. Clin Dermatol, 2006. **24**(3): p. 152-7.
265. Parrino, J. and B.S. Graham, *Smallpox vaccines: Past, present, and future*. J Allergy Clin Immunol, 2006. **118**(6): p. 1320-6.
266. Nagasse-Sugahara, T.K., et al., *Human vaccinia-like virus outbreaks in Sao Paulo and Goias States, Brazil: virus detection, isolation and identification*. Rev Inst Med Trop Sao Paulo, 2004. **46**(6): p. 315-22.
267. Singh, R.K., et al., *Buffalopox: an emerging and re-emerging zoonosis*. Anim Health Res Rev, 2007. **8**(1): p. 105-14.
268. de Assis, F.L., et al., *Reemergence of vaccinia virus during Zoonotic outbreak, Para State, Brazil*. Emerg Infect Dis, 2013. **19**(12): p. 2017-20.

269. Garcel, A., et al., *Genomic sequence of a clonal isolate of the vaccinia virus Lister strain employed for smallpox vaccination in France and its comparison to other orthopoxviruses*. J Gen Virol, 2007. **88**(Pt 7): p. 1906-16.
270. Brodie, R., R.L. Roper, and C. Upton, *JDotter: a Java interface to multiple dotplots generated by dotter*. Bioinformatics, 2004. **20**(2): p. 279-81.
271. Blasco, R., J.R. Sisler, and B. Moss, *Dissociation of progeny vaccinia virus from the cell membrane is regulated by a viral envelope glycoprotein: effect of a point mutation in the lectin homology domain of the A34R gene*. J Virol, 1993. **67**(6): p. 3319-25.
272. Wilcock, D., et al., *The vaccinia virus A4OR gene product is a nonstructural, type II membrane glycoprotein that is expressed at the cell surface*. J Gen Virol, 1999. **80** ( Pt 8): p. 2137-48.
273. Matho, M.H., et al., *Structural and biochemical characterization of the vaccinia virus envelope protein D8 and its recognition by the antibody LA5*. J Virol, 2012. **86**(15): p. 8050-8.
274. Rodriguez, J.R., D. Rodriguez, and M. Esteban, *Insertional inactivation of the vaccinia virus 32-kilodalton gene is associated with attenuation in mice and reduction of viral gene expression in polarized epithelial cells*. J Virol, 1992. **66**(1): p. 183-9.
275. Turner, P.C. and R.W. Moyer, *The cowpox virus fusion regulator proteins SPI-3 and hemagglutinin interact in infected and uninfected cells*. Virology, 2006. **347**(1): p. 88-99.
276. Turner, P.C. and R.W. Moyer, *The vaccinia virus fusion inhibitor proteins SPI-3 (K2) and HA (A56) expressed by infected cells reduce the entry of superinfecting virus*. Virology, 2008. **380**(2): p. 226-33.
277. DeHaven, B.C., et al., *Poxvirus complement control proteins are expressed on the cell surface through an intermolecular disulfide bridge with the viral A56 protein*. J Virol, 2010. **84**(21): p. 11245-54.
278. Brown, C.K., D.C. Bloom, and R.W. Moyer, *The nature of naturally occurring mutations in the hemagglutinin gene of vaccinia virus and the sequence of immediately adjacent genes*. Virus Genes, 1991. **5**(3): p. 235-42.
279. Ichihashi, Y. and S. Dales, *Biogenesis of poxviruses: interrelationship between hemagglutinin production and polykaryocytosis*. Virology, 1971. **46**(3): p. 533-43.
280. Trindade, G.S., et al., *Belo Horizonte virus: a vaccinia-like virus lacking the A-type inclusion body gene isolated from infected mice*. J Gen Virol, 2004. **85**(Pt 7): p. 2015-21.
281. Zhang, C.X., et al., *A mouse-based assay for the pre-clinical neurovirulence assessment of vaccinia virus-based smallpox vaccines*. Biologicals, 2010. **38**(2): p. 278-83.
282. Qin, L. and D.H. Evans, *Genome Scale Patterns of Recombination between Coinfecting Vaccinia Viruses*. J Virol, 2014. **88**(10): p. 5277-86.
283. Kerr, P.J., et al., *Genome scale evolution of myxoma virus reveals host-pathogen adaptation and rapid geographic spread*. J Virol, 2013. **87**(23): p. 12900-15.
284. Barry, M., et al., *Poxvirus exploitation of the ubiquitin-proteasome system*. Viruses, 2010. **2**(10): p. 2356-80.
285. Shchelkunov, S.N., *Interaction of orthopoxviruses with the cellular ubiquitin-ligase system*. Virus Genes, 2010. **41**(3): p. 309-18.
286. Sperling, K.M., et al., *The orthopoxvirus 68-kilodalton ankyrin-like protein is essential for DNA replication and complete gene expression of modified vaccinia virus Ankara in nonpermissive human and murine cells*. J Virol, 2009. **83**(12): p. 6029-38.
287. Ramsey-Ewing, A.L. and B. Moss, *Complementation of a vaccinia virus host-range K1L gene deletion by the nonhomologous CP77 gene*. Virology, 1996. **222**(1): p. 75-86.
288. Panus, J.F., et al., *Cowpox virus encodes a fifth member of the tumor necrosis factor receptor family: a soluble, secreted CD30 homologue*. Proc Natl Acad Sci U S A, 2002. **99**(12): p. 8348-53.
289. Zhang, Q., et al., *Genomic sequence and virulence of clonal isolates of vaccinia virus Tiantan, the Chinese smallpox vaccine strain*. PLoS One, 2013. **8**(4): p. e60557.
290. Huang, J., et al., *The poxvirus p28 virulence factor is an E3 ubiquitin ligase*. J Biol Chem, 2004. **279**(52): p. 54110-6.
291. Senkevich, T.G., E.V. Koonin, and R.M. Buller, *A poxvirus protein with a RING zinc finger motif is of crucial importance for virulence*. Virology, 1994. **198**(1): p. 118-28.

292. Berger, I., D.J. Fitzgerald, and T.J. Richmond, *Baculovirus expression system for heterologous multiprotein complexes*. Nat Biotechnol, 2004. **22**(12): p. 1583-7.
293. Sele, C., et al., *Low-resolution structure of vaccinia virus DNA replication machinery*. J Virol, 2013. **87**(3): p. 1679-89.
294. Shuman, S. and B. Moss, *Bromouridine triphosphate inhibits transcription termination and mRNA release by vaccinia virions*. J Biol Chem, 1989. **264**(35): p. 21356-60.
295. Esteban, D.J., et al., *Expression of a non-coding RNA in ectromelia virus is required for normal plaque formation*. Virus Genes, 2014. **48**(1): p. 38-47.
296. Das, A.T., et al., *Human immunodeficiency virus uses tRNA(Lys,3) as primer for reverse transcription in HeLa-CD4+ cells*. FEBS Lett, 1994. **341**(1): p. 49-53.
297. Guo, Z.S., et al., *The enhanced tumor selectivity of an oncolytic vaccinia lacking the host range and antiapoptosis genes SPI-1 and SPI-2*. Cancer Res, 2005. **65**(21): p. 9991-8.