Vaccinia Virus Gene Acquisition Through Non-Homologous Recombination

by

Gregory Thomas Vallee

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

Master of Science

in

Virology

Department of Medical Microbiology and Immunology
University of Alberta

Abstract

Poxviruses encode many genes that are orthologs of cellular genes. These orthologs serve many functions, but those that are of most interest are ones that have evolved further and now serve an immune-evasion function. Such genes were likely first acquired by poxviruses through some form of non-homologous recombination (NHR). Using an infection-transfection assay, I investigated vaccinia virus's (VAC) ability to recombine with ("capture") a non-homologous (NH) substrate that encoded a fluorescent and drug selectable protein. This study showed that VAC can capture duplex DNA and cDNA·RNA hybrid molecules, but the recombinant frequencies were much lower in the absence of VAC homology. VAC recombined most efficiently with DNA duplexes encoding homology to the N2L locus. The recombinant frequency (RF) that was measured in these control reactions was 670×10^{-6}. Forty-fold fewer recombinants were recovered from VAC-infected cells transfected with homologous cDNA·RNA hybrid molecules (RF=21×10^{-6}). However, when these experiments were repeated using substrates lacking any VAC homology, the linear duplex DNA yielded even fewer recombinants (RF=1.6×10^{-6}) and cDNA·RNA hybrid substrates yielded hardly any recombinants at all (RF=3×10^{-8}).

NHR generated a variety of viruses with genome rearrangements ranging from insertions with flanking duplications to large-scale indels. Most of the viruses that captured the NH substrate generated defective genomes. These viruses suffered large deletions of essential sequence and their replication and packaging were dependent on a wild-type co-infecting helper virus. In many cases, we also

observed partial duplications of the sequences encoding the selectable substrate and these would rapidly rearrange through homologous recombination. A review of the sequences surrounding the insert junctions suggests that VAC NHR is likely catalyzed by a microhomology-independent mechanism. The junctions exhibited little to no pre-existing microhomologies, and we did not observe any disproportionate increase in sequence identity at the junctions or surrounding sequences (~30 nt on either side). In two cases we also found examples of the VAC topoisomerase I recognition sequence [(C/T)CCTT] overlapping the site of the VAC genome that had been targeted by recombination. These studies show that VAC can catalyze NHR through a process that may reflect a form of aberrant replication fork repair supplemented by topoisomerase I action.

Preface

A portion of the data presented in Chapters 3, 4 and 5 of this thesis has been published as “Vallee G, Norris P, Paszkowski P, Noyce RS, Evans DH. 2021. Vaccinia Virus Gene Acquisition through Non-homologous Recombination, Journal of Virology, 95(14):e0031821.”

The DNA sequencing performed for this study was completed by the Applied Genomics Core at the University of Alberta. The pPN_YFP and pPN_YFP+N2H used in these experiments were made by Mr. Peter Norris. He was also responsible for the purification of ten of the VAC strains that I mapped in my study. The clones which he produced are labelled in Table 4.2. All other experiments and analyses presented in this Thesis were performed by me.

Acknowledgements

First and foremost, I would like to thank Dr. David Evans and Dr. Ryan Noyce for their assistance with my project. Thank you for the time you have taken from your busy schedules to help with troubleshooting, data analysis and guidance with my project. Your support is truly appreciated. I would also like to thank Dr. Smiley and Dr. Mason for serving on my graduate committee. Your time and support are truly appreciated.

I would like to Thank Mr. Peter Norris for all the work he accomplished with the pilot experiments for this project. Thank you for producing the plasmids that I used for my studies and creating 10 recombinant VAC strains that I analyzed in this thesis.

I enjoyed my time in the Evans laboratory and was lucky to receive assistance from and to meet some exceptional people. Thank you, Brittany, James Les, Mira, Nicole, Quinten, Sarah, and Sid. I am grateful for your kindness and intellectual support over the past 2.5 years.

Lastly, I would like to thank my significant other, Amanda and my family for their unconditional love and support throughout my Master's degree. I could not have done this without you.

Table of contents

List of tables

List of figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of abbreviations

cDNA·RNA	Complementary DNA-RNA duplex
CDV	Cidofovir
CPE	Cytopathic effect
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
DSB	Double-strand breaks
dsDNA	Double-stranded DNA
EFC	entry fusion complex
EV	Enveloped virus
FEN-1	Flap endonuclease-1
HR	Homologous recombination
HSV-1	Herpes simplex virus 1
ITR	Inverted Terminal Repeat
IV	Immature virion
LINE-1	Long interspersed nuclear elements-1
MMEJ	Microhomology-mediated end joining
MOI	Multiplicity of infection
MR	Multiplicity reactivation
MRN	MRE11–RAD50–NBS1
mRNA	Messenger Ribonucleic acid
MV	Mature virions
NCLDV	Nucleo-cytoplasmic larg DNA viruses
NHEJ	Non-homologous end joining
NHR	Non-homologous recombination
nt	Nucleotide
PARP1	Poly (ADP-ribose) polymerase 1
PFGE	Pulse-field gel electrophoresis
PFU	Plaque forming unit
pol θ	Polymerase θ
pol λ	Polymerase λ

pol μ	Polymerase μ
poly-A	Poly-adenosine
poly-dT	Poly-deoxythymidine
RF	Recombination frequency
RNA	Ribonucleic acid
sgRNA	Synthetic guide RNA
SSA	Single-strand annealing
SSB	Single-strand binding
ssDNA	Single-strand DNA
TBE	Tris-Borate- EDTA buffer
TNF	Tumor necrosis Factor
UV	Ultraviolet
VAC	Vaccinia virus
VARV	Variola virus
VMAP	Viral membrane assembly proteins
VVpol	Vaccinia virus polymerase
WR	Western Reserve
YFP/gpt	Yellow fluorescent protein fused to a guanosine phosphoribosyl transferase

Chapter 1 - Introduction

1.1 A brief history of poxviruses

Poxviruses are a family of large double-stranded DNA viruses (130-300 kb) that replicate in the cytoplasm of cells (2) This family of viruses is best known for the devastating disease of Smallpox. Smallpox is caused by Variola virus (VARV) and it is estimated to be responsible for nearly 300 million deaths throughout the 20th century (3). Historically, VARV is one of the deadliest human viruses and is the pathogen around which the concept of vaccinations was developed (3). Initial methods used to try and protect people from Smallpox date back before 1000 AD in China and India termed variolation (4). This method involved inoculating an individual with smallpox using a smallpox scab from an individual with mild disease or that was recently variolated (4). It was not until Edwards Jenner's work attempting to immunize a young boy against Smallpox by inoculating him with cowpox (3), that lead to the world's first vaccine. This ultimately led to the use of Vaccinia virus (VAC) to vaccinate individuals against VARV. VARV is a human-specific pathogen, which was ultimately eradicated in 1979 through global vaccination efforts (5). Today, this virus remains the only human pathogen to be globally eradicated. Although VARV has been eradicated, some poxviruses still pose a potential risk to human health by zoonotic transmission (6).

Poxviruses are still extensively studied as there are still many questions surrounding these viruses. Most of these studies have been performed using VAC as the prototypic poxvirus. For many years, VAC has been investigated as a tool for medicine. It is currently being studied for its potential use as an oncolytic virus (7-9) and has been explored as a vaccine vector for emerging diseases (10-12). VAC has been explored as it is relatively easy to introduce novel genetic sequences into the virus genome via homologous recombination (HR). The virus also has good clinical safety and efficacy.

1.2 Nucleo-cytoplasmic large DNA viruses and poxviruses

1.2.1 Nucleo-cytoplasmic large DNA viruses (NCLDV)

As large double-strand DNA (dsDNA) eukaryotic viruses, Poxviruses are part of a group of NCLDV which have recently been proposed to form a new Order called the *Megavirales* (13). This order is composed of the *Ascoviridae, Asfarviridae, Iridoviridae, Marseilleviridae, Mimiviridae, Pithoviridae, Phycodnaviridae* and *Poxviridae.* These encompass the largest DNA viruses ranging from 100 kb to 2.5 Mb in size (13). The largest of the NCLDVs are Pandoraviruses which encode nearly 2,000 genes with most of these being ORFans (Taxonomically-restricted genes) (14). The phylogenetic analyses of these viruses suggest there are about 40 core genes that are found in most NCLDV genus that can be traced back to what may be considered the last common ancestor of these viruses (13). Interestingly, only 3 of the core genes are found in all NCLDVs [family B DNA polymerases, D5-like primase-helicase and homologs of the poxvirus late transcription factor (VLTF3-like)] (13). It worth noting that what is considered a core protein is highly debated in the study of NCLVD. Estimates from different research groups range from 5 to more than 40 core genes.

NCLDV have quite dynamic genomes with a pangenome of over 17,000 genes (13). Interestingly there is little overlap of the pangenomic genes found in the different NCLDV families which suggests that the acquisition of these genes is lineage-specific rather than from a common ancestor. These viruses are believed to have evolved through multiple cycles of gene gain and losses over evolutionary time (13). This fluctuation in genome size may best be explained by the genome accordion hypothesis (15). This hypothesis suggests that these large virus genomes expanded and contracted by horizontal gene transfers from mobile genetic elements (15-18) or single-strand DNA invasions of novel DNA sequences at the terminal ends of the viral genomes (15, 16). Gene duplication and deletions also contributed to the expansion and contraction of these large viral genomes (15). This is evident by the presence of many paralogs encoded in some NCLVD families. For example, 35% of the genes of Mimiviruses have at least one paralog in the virus genome [up

to 398 paralogous genes in 86 gene families (19)]. The ability of virus genomes to expand and contract genes was recently demonstrated in poxviruses under modulated selective pressure (20). These viruses expanded beneficial loci under increasing selective pressure and rapidly contracted them under decreasing selective pressure.

1.2.2 Poxviruses

Poxviruses are believed to have evolved from the family of NCLDVs. The family of *Poxviridae* can be divided into two subfamilies called the *Chordopoxvirinae* and the *Entomopoxvirinae* which infect vertebrates and insects respectively. The investigation of evolutionarily conserved class II cyclobutene pyrimidine dimer-photolyases suggests that the origin of poxviruses may date as far back as the divergence of eukaryotes into fungi, animals and plants (21). This is believed to be the case as the amino acid sequences of these proteins are highly conserved when compared to those of animals and plants (>33% conserved) and are found in both subfamilies of poxviruses (21, 22).

The most relevant genus of poxviruses to human health is the *Orthopoxvirusviridae* which have a broad host range of mammals including humans (23). These viruses are quite interesting as all the core and pangenomic genes of this genus can be traced back to a single species of the group, the Cowpox virus. Of the 214 Cowpox virus genes, 174 homologs are found in all Orthopoxviruses. Different Orthopoxviruses virus species will encode a different subset of the remaining 40 pangenomic Cowpox genes. These determine their pathogenesis (23). On a broader scale, There are only 49 genes shared by all Poxviruses and around 90 genes shared by all Chordopoxviruses (23). Most of the knowledge that we have today about Poxviruses has been derived from research on VAC. For my thesis, all experiments will be performed using the Western Reserve strain of VAC.

1.2.2.1 Genome organization

Poxviruses genomes range from 130-360 kb in size and encode at least 150 genes (24). Their genomes are unique as their terminal ends are covalently closed by hairpin loops (25, 26). These hairpins are part of the poxvirus Inverted Terminal Repeats (ITRs) which are found at both the 5' and the 3' ends of the viral genome (26). These sequences are identical but inverted relative to each other. The ITRs encode the concatemer resolution sites (27, 28) and the origin of replication (29). The ITRs range in size depending on the virus but are approximately 10 kb in size in VAC (Fig 1.1). The terminal ends of the virus genomes tend to encode less evolutionarily conserved poxviruses genes (Fig 1.1, "variable') (30). These often encode host-specific virulence factors or immunomodulatory proteins which affect virulence but are non-essential to the viability of the virus. Meanwhile, most conserved genes are found in the central region of the virus genome and encode proteins that are essential for virus replication, assembly and egress (30). The presence of fewer essential genes in terminal ends of the genomes is likely why these regions of the genome appear to be evolutionary hotspots.

1.2.2.2 Evidence of poxvirus genes acquired from their hosts

Many genes encoded by poxviruses are orthologs of host genes. A curious feature of these genes is that many resemble intronless fragments of host genes. Perhaps some of the most intriguing genes are those that have been derived from host immune response genes. Poxviruses have many virokines and viroreceptor involved in the mimicry of cytokine [ex: vIL-10, (31)], and cytokine receptors [ex: CrmE, soluble Tumor Necrosis Factor (TNF) receptor (32)] (33). Such genes were more than likely acquired from the virus's host and repurposed to benefit the virus.

Poxvirus proteins that do not trace back to their NCLVD ancestry, are sometimes traced back to their hosts due to sequence similarities (34). It is reasonable to assume that a gene ortholog (or gene fragment) may have been acquired from a host if a novel gene emerges in a taxonomical class of viruses and

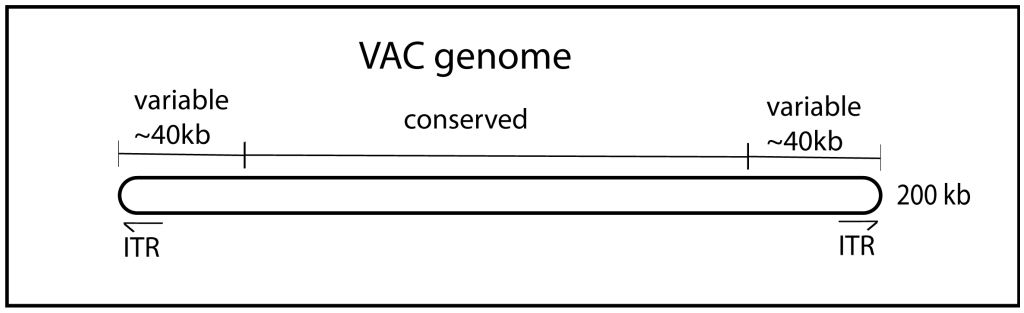


Fig 1.1 Genome diagram of VAC WR. The VAC genome is 195kb in size with inverted terminal repeats (ITR, arrowed) at the terminal ends of the genome. The size of the ITRs vary between poxviruses, but are ~10kb in size in VAC WR. Poxvirus genes are organized in the genome in a way that most genes essential to viral replication are found in the central portion of the genome ("conserved'). "Variable" regions are found at the terminal ends of the genome and can vary between strains of viruses. These variable regions encode genes typically involved in host-specific interaction or viral pathogenesis. These regions are often considered evolutionary hotspots.

that the amino acid sequence of the host and viral protein is significantly conserved. An example of such a gene would be the soluble TNF receptor-like protein encoded by many Orthopoxviruses (32). This protein resembles the N-terminal domain of the mammalian TNF-2 receptor while lacking the C-terminal portion involved in anchoring the protein to the membrane and cytoplasmic signalling (35) (Fig 1.2). The VAC protein is secreted where it binds TNF-2 and prevents signaling (32). An alignment of the VAC CrmE gene to the human TNF receptor is shown in Fig 1.2.

1.2.2.3 Retrotransposon and poxvirus horizontal gene transfers

The fact that poxviruses encode many genes that are homologs of host genes, raises questions concerning how they might have acquired such genes. Currently, the only method by which poxviruses have been shown to acquire novel host-encoded genetic sequences is by retrotransposon-mediated horizontal gene transfers (17, 18). In fact, the remnants of sequences that are hallmarks of retrotransposon-mediated gene transfers can still be seen surrounding genes that were naturally acquired by poxviruses. The Cowpox Golgi anti-apoptotic gene (vGAAP) is an example of one of these genes. It encodes remnants of a poly-adenosine (poly-A) tail, and is flanked by target site duplications that are the hallmarks of long interspersed nuclear elements-1 (LINE-1) mediate transposition (17, 18). Short-interspersed elements (SINES) can also be observed in the Taterapox virus genome (36).

1.3 Poxvirus life cycle

1.3.1 Virus entry and genome uncoating

VAC has two infectious forms, mature virions (MV) and the extracellular virions (EV). These differ in the number of surrounding lipid bilayers and the way they enter susceptible cells. MVs have one lipid bilayer (37) and enter a host cell by first binding to heparin sulfate (38-42), chondroitin (42, 43) or laminin (42, 44). These events lead to the fusion of the viral membrane with either the plasma membrane (45-47) or an endosomal membrane (42, 48, 49). The membrane fusion is mediated by the entry fusion complex (EFC) made up of 11 VAC proteins (42, 50). EVs have two lipid bilayers (37) and currently have no confirmed attachment

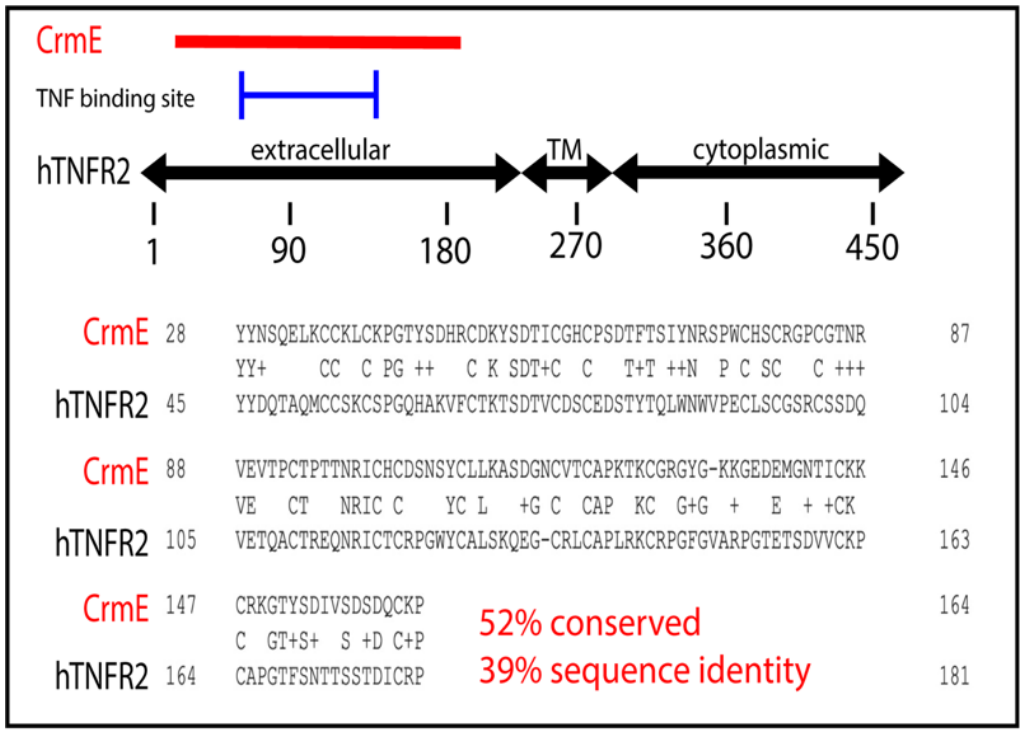


Fig 1.2 VAC CrmE protein sequence alignment. The figure presents an alignment of a portion of the VAC CrmE amino acid sequence with a portion of the sequence of the human TNF receptor Family 2 (hTNFR2) protein. The extracellular, transmembrane (TM) and cytoplasmic domains of the full-length hTNFR2 proteins are indicated. The alignment was generated using BLASTP. The TNF binding site conserved between both CrmE and hTNF2 is outlined in blue. The sequence alignment is 52% conserved and shares a 39% sequence identity. The figure was adapted from a figure created by Dr. David Evans.

factor on the cell surface. It is believed that their interaction with glycosaminoglycans leads to the disruption of the extra external membrane allowing for the interaction of the EFC with the cellular membrane (42).

Membrane fusion leads to the release of the DNA-containing core into the cytoplasm where transcription of ~100 early virus genes begins immediately within the virus core (51). This is possible because Poxviruses package their RNA polymerase, early transcription factors and mRNA modifying enzymes within the core (51). As early genes are transcribed, the transcripts are then released into the cytoplasm through the core particle pores and translated by the host cell translational machinery (52). A key early gene that is transcribed is the D5 helicase/primase. This enzyme plus the host proteosome uncoat the virus genome (53-55). Other early genes, including M2 and C5, have also been shown to be involved in DNA uncoating, although it appears that these proteins may have redundant roles (56).

1.3.2 Viral genomic DNA replication

Upon uncoating of the virus genomic DNA, replication begins in viral factories which are associated with the rough endoplasmic reticulum (57). Each of these viral factories arise from a single infectious viral particle (58). Replication can be detected as early as 2 hours post-infection (59). Poxvirus DNA replication is a complex process that requires many proteins transcribed from the early genes. Genes required for replication include the E9 DNA polymerase, D5 helicase/primase, A20 processivity factor, D4 uracil glycosylase, and the I3 single-stranded DNA-binding protein (59). Other proteins are known to be involved in DNA replication however, they may be dispensable like VAC DNA ligase (59). In the case of the ligase, it is due to the ability of VAC to recruit the cellular DNA ligase I which can compensate for the deletion of the VAC ligase (60).

The most recent research around the method by which VAC replicates its genome supports a leading/lagging strand synthesis model (61). Replication initiation is thought to mostly start near the terminal ends of the viral genome near what will eventually be converted into concatemer junctions (61). Such a method

of viral replication resembles the method used by host cells. Interestingly poxviruses encode nearly all the machinery required to replicate their genome. The enzyme they lack is an endonuclease that would produce a 3' hydroxyl that would allow for the initiation of viral replication (61). Replication of poxvirus genomes leads to the formation of concatemeric genomes with head-to-head and tail-to-tail rearrangements (62). The VAC genomes concatemers are processed into their monomeric forms, by the A22 Holliday junction resolvase (63, 64). Once cut into monomers, the virus genomes can be packaged into viral particles although precisely how resolution relates to the timing of packaging is not entirely clear.

1.3.3 Virus assembly and egress

Once DNA replication has ceased, transcription and translation of the late viral genes required for the packaging begin (65). DNA packaging and egress begins with the formation of membrane crescents derived from the ER membrane (66). The formation of the crescents is controlled by 9 viral proteins. Six regulatory proteins are complexed together in a structure collectively known as the VMAP (A6, A11, A30.5, F10, H7, L2) plus 3 structural proteins (D13, A14, A17) (67). The VMAPS stabilize naturally occurring ruptures in the endoplasmic reticulum membrane by associating with the free ends of the membrane and can coordinate the further rupture of the membrane to initiate crescent formation. These ruptures of the membrane expose the A17 transmembrane proteins to D13 trimers, which act as a scaffolding protein to promote crescent membrane formation (67). These crescents accumulate, ultimately forming immature virions (IV) (68, 69). Before sealing the crescent membranes into a spherical IV particle, the core proteins and viral genomic DNA are packaged into the growing IV (68, 70, 71). These protein-DNA complexes in IVs are called nucleoids.

The following steps transform the IV-containing nucleoids into MV. This transition begins with the removal of the D13 proteins, proteolytic processing of the A17 proteins, condensing of the core and rearrangement of the IV membrane into an inner and outer membrane (67). These changes create transcriptionally active and infectious virus known as MVs. MVs are a single lipid bilayer enveloped

virus particles that are released from cells by cell lysis (65). Infectious virus may also be trafficked to the extracellular membrane via microtubules (72) where they will gain additional membranes and be released by membrane-membrane fusion. These viruses are known as EVs and have two lipid bilayers wrapping the virus particle (73). Two intracellular membranes are gained by the MV as the virus particles pass through organelles [endosomes (74) or the Golgi Apparatus (75)]. The outer-most membrane of these trafficked viruses is ultimately lost during the membrane-membrane fusion process, releasing the EV. The second lipid bilayer wrapping the EVs have unique proteins that are not found on MVs. A number of these proteins play a role in wrapping EV [A27, B5 and F13 (65)] but also serve other functions that affect plaquing (76, 77) and actin-tail formation (78).

1.4 Genetic recombination

The ability to repair DNA damage is important as it enables an organism to protect itself from UV, oxidative compounds, and other causes of DNA damage. Many mechanisms have evolved to protect organisms against these different kinds of DNA damage and many of these are evolutionarily conserved from prokaryotes to high-level eukaryotes (79). Double-strand breaks (DSB) are considered the most lethal of all DNA damage as they can lead to mitotic catastrophe in eukaryotic cells (80). There are two main repair pathways used to repair these breaks. There is HR, a template or homology-dependent pathway (81), and there is illegitimate recombination, which is usually subdivided into homology-independent or microhomology-dependent [≤20 nucleotides (nt)] pathways (82). Both HR and illegitimate recombination can enhance the survival of bacteria and eukaryotes in DNA-damaging environments, so it is not a surprise that viruses can also benefit from being able to catalyze such reactions.

An example of such a recombination process that can be observed in many viruses is known as multiplicity reactivation (MR) (83-90). MR is a recombination mechanism by which viruses repair genomes that have inactivating damage caused by UV or ionizing radiation. The MR repair process requires two or more

coinfecting virus with damaged genomes. The viral genomes can recombine to produce a viable virus genome through genetic recombination. This was first observed with T4 bacteriophage where viable phage could be recovered from cells coinfected with UV-inactivated viruses (83). There is an extensive literature concerning bacteriophage and their ability to repair DNA damage using HR (91-93). These classical studies have identified at least two different kinds of bacteriophage recombination systems used by phage like T4 and lambda. These are initiated using either strand invasion reactions (T4 UvsX) or through single-strand annealing reactions (lambda Red recombinases) (94, 95). The SSA mechanism bears many similarities to the mechanism used by Poxviruses to catalyze HR although the polarities of the reactions are different.

1.4.1 VAC homologous recombination

VAC is a well-studied poxvirus known to catalyze HR. The initial interest in poxvirus-catalyzed recombination reactions arose from observing recombination between different Orthopoxviruses (96, 97) or between strains of VAC (98). These studies highlight how poxvirus genetic recombination might contribute to the evolution of the virus through inter-genomic and cross-species recombination. VAC is hypothesized to originate from a virus that underwent intergenomic recombination between several different Orthopoxviruses (99).

Poxvirus recombination is believed to be catalyzed by a single-strand annealing (SSA) process (100-102). A schematic of the SSA mechanism is shown in Fig 1.3. These reactions begin with the resection of the DNA strands by a 3'-to-5' exonuclease (ii), exposing homologous sequence between DNA fragments (blue). The 5' single-stranded DNA (ssDNA) overhangs can subsequently anneal together (iii). In some cases, there are non-homologous sequences at the outermost end of the 5' ssDNA (not shown). These ends are removed by an endonuclease and the remaining nicks in the joined DNA molecules are subsequently repaired (iv). These single-strand annealing repair reactions are a common HR mechanism used by viruses and bacteriophages [Herpes Simplex Virus-1 (HSV-1) (103) T4 and λ

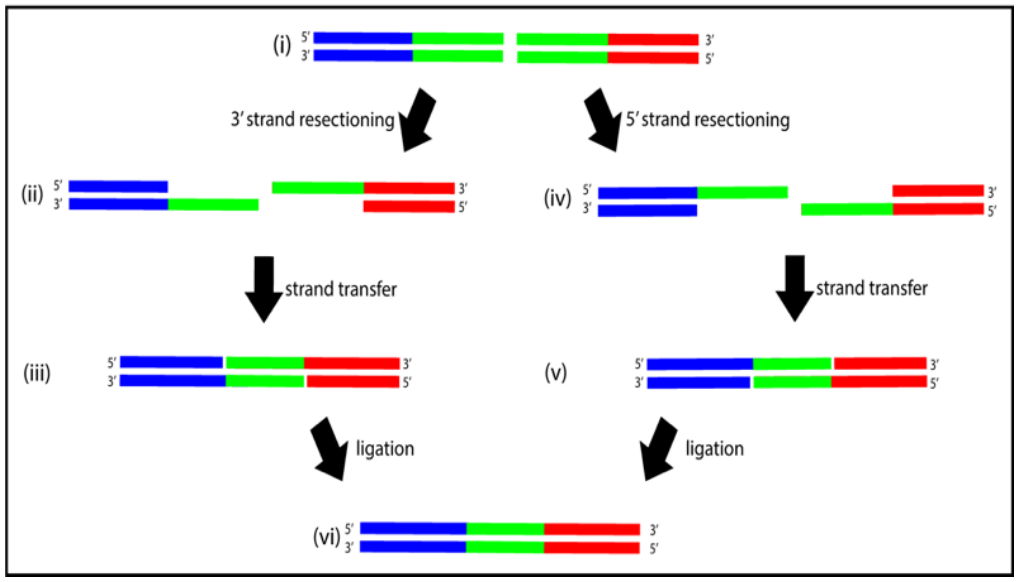


Fig 1.3 SSA recombination. (i) DNA molecules sharing homologous sequences (green) are processed through strand reseactioning to expose homologous sequences. 3' resectioning (ii) or 5' resectioning (iv) generating 5'-ssDNA tails or 3'ssDNA tails respectively. Homologous ssDNA tails can subsequently anneal together (iii and v) and nicks in the DNA backbone are ligated (vi).

(104, 105)]. However, the poxvirus single-strand annealing process differs from other described systems in that most other organisms use a 5'-to-3' exonuclease to produce 3'-ssDNA presynaptic ends (Fig 1.3, iv) [e.g., herpes viruses (106), bacteriophage, and humans (107)] whereas poxviruses use primarily use a 3'-to-5' exonuclease to produce 5'-ssDNA presynaptic ends (Fig 1.3, ii).

VAC catalyzes HR in a very efficient and accurate manner (108). HR is probably regulated by the concentration of deoxynucleotide triphosphates (dNTPs) in the microenvironment (100, 109, 110). The efficiency of VAC HR depends on the length of the homologies and the dsDNA substrates that are targeted for recombination. VAC HR can be catalyzed using as little as 12-16 bp of shared homology *in vivo* (101, 102), however, the efficiency of these reactions greatly increases when the length of the homologous sequence increased [>50 nt (101)]. VAC HR can catalyze HR between linear or circular molecules, or a combination of linear and circular molecules although linear DNA molecules appear to be the best substrates (101). VAC HR between circular molecules or a combination of linear and circular molecules are 15- to 50-fold less efficient, respectively, than between two linear substrates. Although VAC recombination is quite efficient, the presence of non-homologous DNA sequences flanking the homologous sequence decreases the efficiency of recombination (101). The presence of insertion or base substitution in homologous sequences also does not inhibit recombination but again reduces its efficiency *in vivo* (102). The presence of 1-2 nt mismatches interrupting an otherwise homologous DNA sequence has little effect on recombination efficiency (10-20% reduction) while a 1-2 nt insertion reduces recombination by nearly 75% (102).

1.4.2 Proteins involved in VAC homologous recombination

Currently, all the proteins known to be involved in VAC homologous recombination are encoded by the virus. Four key proteins have been identified to date. These proteins include the VAC E9 polymerase, the I3 single-strand DNA

binding protein, the G5 Flap endonuclease-1 (FEN-1) like protein and a DNA ligase. Details about the function and role of these proteins follows.

1.4.2.1 E9 polymerase

VAC's E9 gene encodes a family B polymerase which is conserved in all NCLDVs (111). This enzyme encodes a 5'-to-3' polymerase and 3'-to-5' proofreading exonuclease activity (112). Before the discovery of the role of the E9 polymerase (VVpol) in VAC recombination, several experiments had linked recombination and replication through the use of temperature-sensitive VAC E9L mutants (113) and DNA replication inhibitor-based experiments (114, 115). Although these experiments demonstrated a link between replication and recombination, they did not tell us what specific role of E9 played in recombination. It was not until VVpol could be purified (116) that its direct role in catalyzing recombination was demonstrated.

In vitro experiments, showed that VVpol catalyzes strand transfers between DNA molecules (117) and catalyzes the formation of concatemers made up of DNA molecules that share homology (100). The polarity of the reaction was established by sequencing VVpol-catalyzed recombination junctions. These studies showed that linear dsDNA molecules encoding bases substitutions within the hybridization region, preferentially retained the base substitutions encoded on 5'-ended strands over those found on the 3'-ended *in vitro* (101). These and other studies showed that VVpol-mediated recombination requires a 3'-to-5' proofreading exonuclease activity *in vitro*.

The role played by the VVpol 3'-to-5' exonuclease in recombination *in vivo* was difficult to confirm as the proofreading activity is essential to virus viability (109). Its role in recombination was investigated by the Evans lab using cidofovir (CDV) a dCMP analog. CDV hinders primer extension during replication when incorporated in the second last position of the primer. It also inhibits the catalytic function of wild-type VVpol 3'-to-5' exonuclease (118). However, one can isolate CDV-resistant VAC strains where the mutations map to either the exonuclease domain (promoting drug excision) or the polymerase domain (promoting enhanced

extension). Using these two cidofovir-resistant VAC strains, and DNA substrates bearing CDV incorporated into the ends of the linear duplexes, Gammon and Evans showed that both mutants could actively replicate a transfected substrate in the presence of CDV, however, only the VVpol with a mutation that mapped to the exonuclease domain, could catalyze recombination *in vivo* and *in vitro* (109). Overall, VVpol is required to resection the DNA strands and promote strand transfer between DNA molecules.

1.4.2.2 I3 single-stranded DNA binding protein (SSB)

SSB proteins are evolutionarily conserved proteins that are found in all domains of life (119). As the name indicates, these proteins are involved in non-sequence-specific binding to single-stranded DNA (ssDNA) with high affinity. Single-strand DNA binding proteins prevent the formation of secondary structures and protect the DNA from nuclease degradation (119). These functions enable organisms to maintain genomic stability. Mutant SSBs in eukaryotic cells are associated with the development of cancer as SSBs are required in all DNA repair pathways (120). The inability to appropriately repair DNA damage leads to the accumulation of mutations leading to genomic instability and cancer (120). SSBs also play an important role in DNA repair, replication, and recombination. Both in prokaryotes and eukaryotes, SSBs mediate protein-protein interactions and recruit different cellular factors and enzymes to the ssDNA that are involved in DNA metabolism (119).

The VAC I3 gene encodes a 34-kDa protein that is evolutionarily conserved in Chordopoxviruses. This protein was originally predicted to be a SSB protein because it encodes a pattern of charged and aromatic amino acids that are characteristically found in SSBs and it was later shown to bind ssDNA in vitro (121-123). *In vivo* studies showed that I3 colocalizes in the viral factories (123, 124). I3 is essential for virus replication as interfering with the capacity to bind to DNA hinders virus replication, and disrupting the gene prevents recovery of viable viruses (121, 123, 125, 126). The I3 C-terminal tail plays an important function by promoting cooperative binding of other I3 proteins to ssDNA molecules (126). This cooperative binding is thought to be important as it may stabilize DNA replication

forks in an open conformation and prevent interstrand multimerization of I3 proteins (126). These characteristics of poxviruses SSBs are similar to what has been seen in some bacteriophages and *E. coli* (126).

Much like for VAC replication, I3 plays an important role in recombination. The SSB enhances VAC recombination and E9's ability to form DNA concatemers *via* single strand annealing *in vitro* (100, 109). This enhancement is proposed to be due to I3's capacity to aggregate ssDNA (127) which would increase the likelihood that homologous ssDNA molecules are brought close together. The importance of I3 in recombination is also observed *in vivo* as reducing I3 expression reduces both viral replication and recombination (109).

1.4.2.3 G5 flap endonuclease-1 like (FEN-1 like) protein

The G5 FEN-1 like endonuclease was originally identified using protein sequence alignments and structural modelling comparisons to the human FLAP-1 endonuclease structure (128). FEN-1 nucleases are metalloenzymes that target specific DNA structures. This family of enzymes exhibit 5'-flap endonuclease activity, 5'-exonuclease activity, and gap endonuclease activity (129). These activities are used to remove the RNA primers (Okazaki fragments) that prime lagging strand DNA replication and can also cleave the single-stranded nucleotide flaps that are displaced by an enzyme such as a polymerase (129). For this type of cleavage, FEN-1 recognizes the 5'-end of ssDNA and cleaves ssDNA flaps (130).

The role of the computationally predicted FEN-1 like nuclease (G5) proved to be important for VAC recombination and replication (131). Mutant viruses lacking the G5 gene suffer a 3- to 5-fold decrease in homologous recombination efficiency and a 7- to 8-fold decrease in the ability to repair DSB. The decrease in both processes might be attributed to a G5 FEN-1 like nuclease's 5'-to-3' exonuclease activity that enables VAC to produce 3'-ssDNA tails that can be used for recombination (Fig 1.3, iv). However, the deletion of this gene did not eliminate VAC's ability to catalyze HR. This suggests that G5 may simply enhance the efficiency of HR reactions that are primarily catalyzed by the E9 polymerase (131). Another role for the FEN-1 like nuclease may be to remove the non-homologous

flaps of DNA that are formed during the annealing of complementary 5'-ended strands (Fig 1.3, ii).

1.4.2.4 DNA ligases

DNA ligases are enzymes that catalyze the joining of a 5' phosphate to a 3' hydroxyl in DNA. These enzymes are essential for DNA replication, recombination, and DNA repair (132). The first study of VAC's A50R DNA ligase, which was discovered in 1990, showed that the viral ligase was not essential to viral replication (133). It wasn't until nearly 20 years later that DNA ligases were found to be essential for viral replication (134). The lack of a phenotype in the original study of VAC ligase knockouts, can be explained by complementation by cellular DNA ligase I which is recruited to the viral factories during virus replication (134). Knocking out the viral A50R gene and knocking down DNA ligase I expression reduced viral replication. However, knocking down cellular DNA ligase I expression in cells infected with wild-type VAC had no effect on virus replication (134).

Although this study showed the importance of DNA ligase in regards to viral replication, there are no studies that have investigated whether DNA ligase I can complement VAC DNA ligase in the case of HR. It is conceivable that either ligase could be employed *in vivo* as either could theoretically catalyze the final ligation step required by a SSA mechanism (Fig 1.3, vi). In that regard it has been shown that the nicks in the DNA backbone, in the metastable molecules that have been joined together by VAC DNA polymerase, can be repaired by another bacteriophage T4 DNA ligase (135).

1.4.3 Non-homologous recombination (NHR)

NHR is sometimes also called illegitimate recombination, and it is an important process that is encountered in many organisms. NHR enables an organism to repair double-stranded breaks when a homologous sequence is not available to mediate homology-dependent repair. NHR is commonly divided into two major pathways: non-homologous end joining (NHEJ) and microhomology-mediated end joining (MMEJ). A description of both pathways follows (Sec. 1.4.3.1 and 1.4.3.2).

1.4.3.1 Non-homologous end joining (NHEJ)

NHEJ is a versatile mechanism that enables an organism to ligate blunt ended DNA molecules or DNA molecules with 3' or 5' overhangs of various lengths (<20nt) (82). I have highlighted the mechanism of eukaryotic NHEJ in Fig 1.4. NHEJ begins with the binding of the Ku complex to the broken ends of DNA molecules [Ku70-K80 (82, 136)]. These proteins bring the broken strands together and act as a dock for other proteins to bind and catalyze the joining reactions. If blunt ends are present, the DNA is ligated by DNA ligase IV (II). These ends may also undergo alternative forms of enzymatic processing, adding untemplated nucleotides to the junction before ligation (not shown). If overhangs are present (iii, vi), these strands may be degraded by nucleases (iv and ix) or filled in by a DNA polymerase (vii). In humans, the Artemis nuclease is the enzyme primarily responsible for degrading 5'- or 3'- DNA overhangs through its endo- or exonuclease activities. Polymerase μ (pol μ) and polymerase λ (pol λ) are the polymerases most commonly used to repair overhangs in NHEJ. Both polymerases can synthesize DNA in a template-dependent or template-independent manner (136). The activity of these polymerases appears to be skewed with pol λ primarily being involved in the template-dependent synthesis and Pol μ promoting template-independent synthesis (137). The template-independent synthesis is responsible for adding the non-templated nucleotides occasionally seen at the DSB site. Once the overhangs are repaired, the DNA fragments can be ligated by DNA ligase IV (82, 136) (v, viii and x).

NHEJ is a well-studied mechanism found in bacteria and eukaryotes(136). This process is important for survival. In humans, mutations in genes involved in NHEJ are associated with developmental defects, growth delays, and an increased incidence of cancer (138, 139). In addition to playing a role in promoting survival, NHR may also contribute some increase in genetic diversity. Although it hasn't been demonstrated in a laboratory setting, bacterial NHEJ has likely contributed to horizontal gene transfer events (140) and NHEJ also plays an important role in generating immune diversity during V(D)J recombination in vertebrates (141). For VAC, recent studies have demonstrated that DSB in the VAC genome can be

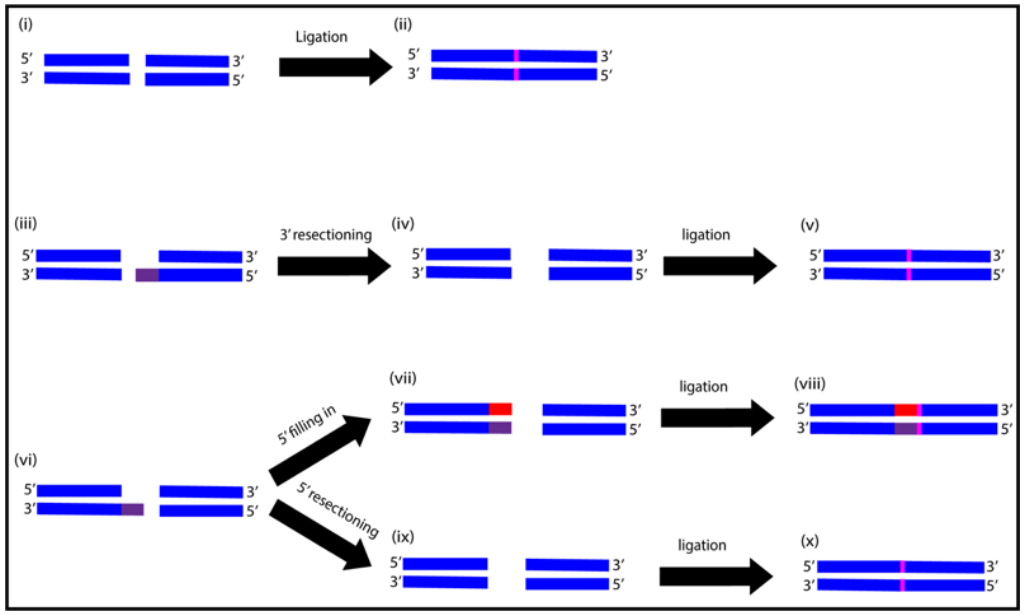


Fig 1.4 Non-homologous end-joining of a DSB. DSBs that have no overhangs (i) are preferentially ligated without additional enzymatic processing (ii, pink). If a broken molecule has a 3' overhang (iii, purple), these molecules are subject to exonuclease degradation (iv) before ligation (v, pink). in the case of 5' overhangs (vi, purple), these molecules are preferentially subject to the filling in of the overhang by a polymerase (vii, red) before ligation (viii). Alternatively, the overhangs are attacked by a nuclease (ix) before ligation (x, pink). In all cases, non-templated nucleotides may be added to the broken DNA molecules by a polymerase, before the strands are joined by ligation (not shown).

mutagenically repaired by DNA ligase IV, one of the enzymes responsible for NHEJ repair (142).

1.4.3.2 Micro-homology mediated end joining (MMEJ)

MMEJ, also known as alternative end-joining, is a repair pathway primarily used when the NHEJ pathway is compromised (143). This pathway involves extensive resection of broken DNA molecules so as to expose short stretches of sequence sufficient to support recombination [<20 nt but most commonly between 4-6 nt (82)]. These reactions are similar to single-strand annealing reactions, but require less homology and involve different proteins. The eukaryotic proteins involved in alternate end-joining are the polymerase θ (pol θ), Poly (ADP-ribose) polymerase 1 (PARP1), and MRE11–RAD50–NBS1 (MRN) (82). Mechanistically, these reactions begin with the detection of DNA damage by PARP1 (Fig 1.5, i), 3'-overhangs, 15-100 nt long, are produced by the MRN exonuclease (ii), and these are joined together by Pol θ through the microhomologies. Pol θ catalyzes these reactions by first binding and stabilizing the homologous sequences (≥2bp) shared between the 3'-ssDNA tails (iii). After stabilizing the bound microhomologies, Pol θ fills in the gapped strands (iv). The nicks left in the newly formed DNA duplex can be repaired by DNA ligase I or III (82) (v). Depending on where the microhomologies are located, 3'-overhangs may be present (vi) which require nuclease processing before pol θ can repair the DSB (vii). The identity of the nuclease responsible for this activity is currently unknown.

As previously mentioned, MMEJ is a pathway that is primarily used when NHEJ is compromised (143). How this is regulated in humans is not completely understood. MMEJ likely acts as a secondary mechanism that enables DNA repair in the absence of both HR and NHEJ. Limiting the use of this pathway would be advantageous as it can promote deletion of long segments of DNA. This is more likely to cause mutations than the small deletion or insertions created by NHEJ.

Despite the mutagenic properties of the pathway, the fact that it requires very little homology (2-25bp) make MMEJ an intriguing candidate for catalyzing horizontal gene transfers. It is known that bacteria can acquire antibiotic resistance genes from non-homologous exogenous DNA sequences using MMEJ (144). This

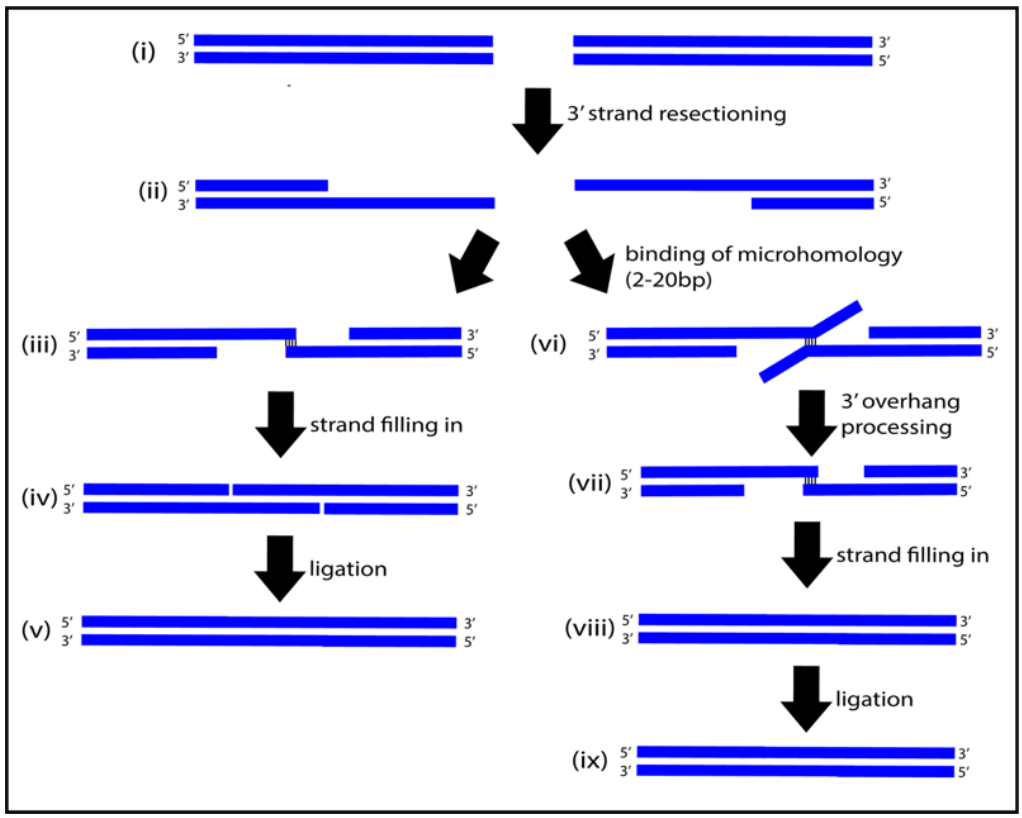


Fig 1.5 Microhomology mediated end-joining of a DSB. DSB (i) are first detected by cellular machinery and the 3'-ends are resected by an exonuclease (15-100 nt) (ii). Any exposed microhomologies within the overhangs can anneal and the joints are stabilized by host proteins (iii). If microhomologies are located at the ends of the 3'-single strand DNA tails (ssDNA), the gapped strands are filled in by a polymerase (v) and the nicks are repaired by DNA ligase (v). If If 3'-overhangs are present (vi), they are degraded by a nuclease (vii) and gapped strands are filled in (viii) and nicks are ligated (ix).

pathway also bears many resemblances to single-strand annealing reactions except insofar that it requires very little homology (≥2bp) to promote recombination between DNA molecules. This makes one wonder whether MMEJ-like reactions might contribute to horizontal gene transfers in viruses that catalyze single-strand annealing reactions [T4 and λ bacteriophages (104, 105)], HSV-1 (103) and VAC (100-102)].

1.5 Thesis goals

The overall objective of my thesis is to investigate whether VAC can capture novel genes by NHR. Currently, the only mechanism by which poxviruses are known to acquire novel genes is through retrotransposon-mediated horizontal gene transfer (17, 18). The thesis is based upon the hypothesis that there may be other mechanisms, more like NHEJ or MMEJ, that can also promote gene capture by replicating poxviruses. The research was proposed to comprise two elements.

The first part of the study was planned to test what kinds of molecules are substrates for VAC non-homologous recombination. Because poxviruses are cytoplasmic viruses, and encode intronless genes, we had to consider what kinds of nucleic acid substrates a replicating virus might encounter. These molecules presumably include mRNA, cDNA·RNA hybrid molecules, and DNA. The cDNA·RNA hybrid molecules could perhaps be produced by a reverse-transcriptase copying an mRNA template. The goal of this first aim was to develop a drug selectable gene, encoding a fluorescent marker and composed of one of the three types of aforementioned molecules. Using these substrates, the plan was to measure the frequency which VAC recombines with these different substrates and how that is affected by the presence or absence of VAC homology flanking the gene.

The second part of this thesis was planned to analyze the genome structure of the rare recombinants that were formed in these reactions. This necessitates purifying the recombinant viruses and sequencing, mapping and classifying any novel genomes. During the course of this study, I discovered that these processes produce some complex and unstable genomes that proved difficult to map using genomics alone. Therefore, much effort was expended using PCR, pulsed-field gel

electrophoresis (PFGE), and Southern blots to demonstrate that the computer-generated assemblies were correct. Finally, we also wanted to gain some understanding of what process(es) VAC might be using to assemble these rare products of NHR. Therefore, I conducted an investigation of the sequences surrounding the captured genes and compared them to the sequence of the starting VAC genome and the transfected substrates. This provided insights into whether these recombination events were mediated by MMEJ- or NHEJ-like processes.

Chapter 2 – Materials and methods

2.1 Cell lines, culturing, plasmids, oligonucleotides

2.1.1 Cell lines and general tissue culturing

BSC-40 cells were obtained from the American Type Culture Collection (ATCC) and propagated in Minimal Essential Media (MEM) substituted with 1% non-essential amino acids, L-glutamine, sodium pyruvate and antibiotic/antimycotic, and 5% FetalGro ® (RMBIO).

For passaging cells, the cells are initially washed with PBS followed by incubation with prewarmed 0.25% Trypsin-EDTA at 37 °C. Once cells have lifted off from the plate, the trypsin was deactivated by adding media, and cells were passaged. Cells were regularly passaged in 15 cm dishes. One 15 cm dish of cells was sufficient to passage in a 1:4 to 1:9 split. When passaged for experiments requiring 10 cm dishes, 6-well plates, or 12-well plates, cells were split to achieve 85-100% confluency on the following day. This required the splitting of one 15 cm dish into four 10 cm dishes, six 6-well plates, or six 12-well plates.

2.1.2 Plasmids and recombination substrates.

The plasmids used for the generation of homologous recombinant and non-homologous recombinant viruses were constructed by Peter Norris. The non-homologous plasmid (pPN_YFP) and homologous plasmid (pPN_YFP+N2H) encoded a yellow fluorescent protein fused to guanosine phosphoribosyl transferase (YFP/gpt) and regulated by a synthetic early/late poxvirus promoter (145). The pPN_YFP+N2H plasmid additionally encoded two sequences (313 bp and 463 bp) flanking the YFP/gpt marker that are homologous to the VAC N2L gene. The YFP/gpt sequence used for the pPN_YFP plasmids was originally PCR amplified from a pDGloxP plasmid (145) and the N2L-flanked YFP/gpt gene was originally amplified from a plasmid used to disrupt the VAC N2L gene through homologous recombination (146). These DNAs were amplified using Phusion® High-fidelity DNA Polymerase (Thermo Fischer) and inserted into pCR4-TOPO vectors using a TOPO Cloning Kit (Invitrogen). The plasmids also encoded a T7 promoter upstream and *Mss*I site downstream of the YFP/gpt marker and its flanking N2L homologies. This permitted transcription and linearization of the plasmid (Fig 1A).

The pPN_YFP and pPN_YFP+N2H plasmids were used to generate the six substrates used for the recombination studies. Linear DNAs were created by digesting the plasmids with *Mss*I (Thermo Fischer). Following digestion, the DNA was purified using the QIAquick PCR Purification Kit (Qiagen). These substrates were intended to represent the broken DNAs that might be captured by recombination reactions. The mRNA substrates were produced by transcribing linearized plasmid DNA with a Maxiscript™ T7 Transcription Kit and subsequently poly-adenylating the product using a Poly(A) Tailing Kit (Invitrogen). Much like the linear DNA products, the mRNA substrates were column purified, but using a RNeasy Mini Kit (Qiagen). To generate cDNA·RNA hybrid substrates, the mRNA was reverse transcribed using an oligo (dT_{12-18}) primer (Thermo Fischer), SuperScript III or SuperScript IV (Invitrogen), and the first-strand synthesis protocol provided with the enzymes. The reverse transcriptase was heat-inactivated and the reactions were used without further purification (Fig 1B).

2.1.2.1 Verification of selectable substrate production

Gel electrophoresis was used to characterize the different selectable substrates. A one percent formaldehyde-agarose denaturing gels and the controls supplied with the kits were used to confirm the success of the transcription and polyadenylation reactions. Ordinary agarose gels were used to characterize the restriction digests and cDNA·RNA hybrid structures. To confirm the formation of cDNA·RNA hybrids, 500 ng of reverse transcribed RNA was treated with 50-100 U of RNAse A (Thermo Fischer) or 2-4 U of RNAse H (Invitrogen). These products were compared to an untreated control. These RNAse treatments were used a cDNA·RNA hybrid should be sensitive to RNAse H digestion but not RNAse A. Complete linearization of the DNA substrates was verified by agarose gel electrophoresis.

2.1.3 Primers and guide RNAs

All of the primers and guide RNAs used in these experiments were produced by Integrated DNA Technologies (IDT). The synthetic guide RNAs (sgRNA) that target sequences to the N2L gene were computationally determined

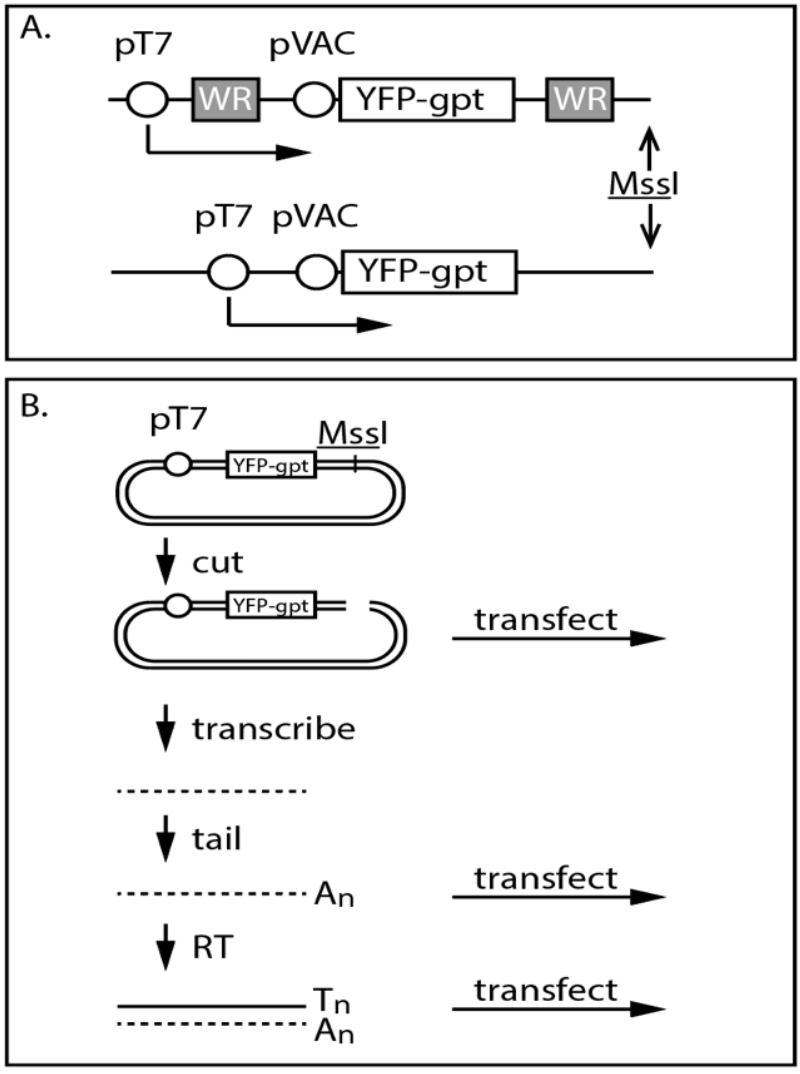

A.
pT7
pVAC
WR
YFP-gpt
WR
MssI
pT7 pVAC
YFP-gpt
B.
pT7
MssI
YFP-gpt
cut
YFP-gpt
transfect
transcribe
tail
An
transfect
RT
Tn
An
transfect

Fig 2.1 Substrates used in this study. (A) Homologous (pPN-YFP+N2H) and non-homologous (pPN-YFP) plasmid templates. Two different plasmids were constructed encoding a YFP/gpt selection marker regulated by a VAC early/late promoter and a T7 RNA polymerase promoter. One of the two plasmids also encoded 463 bp and 313 bp of sequences inserted flanking the YFP/gpt gene that are homologous to portions of the VAC WR N2L gene ("WR"). (B) Construction of different nucleic acid substrates. A duplex DNA substrate was prepared by cutting the two plasmids (in A) with *Mss*I. The RNA substrate was prepared using T7 RNA polymerase and Tailing of the RNA using *E. coli* Poly(A) Polymerase I. DNA templates were removed by Turbo DNase treatment. cDNA·RNA hybrid molecules were prepared using the poly-A tailed RNA, an oligo dT primer, and reverse transcriptase. Figure from G. Vallee et al. (1).

using IDT's custom Alt-R® CRISPR-Cas9 guide RNA design tool. The mCherry sgRNA is a guide RNA based on one described by Gowripalan *et al*. (147) with the first base changed from guanosine to adenosine. The synthetic guide RNAs were ordered as whole synthetic guides that include the target binding sequence and the sgRNA scaffold. The sequence of the scaffold and the target binding sequences used can be found in Table 2.1. The sequences of primers used in these experiments are also listed in Table 2.1.

Table 2.1: Sequence of oligonucleotides used for these experiments

sgRNA target binding sequences and scaffold	
Oligonucleotide name	Sequence
N2L-2 sgRNA	5′-UCAAUGGUGUGAUGUUGAUU-3′
mCherry sgRNA	5′-AGAUAACAUGGCCAUCAUCA-3′ (3)
A23R sgRNA	5′-GAAAGAACGCAUUUCCUCAG-3′ (142)
sgRNA scaffold sequence	5′-Target binding sequence-GUUUUAGAGCUAG AAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUG AAAAAGUGGCACCGAGUCGGUGCUUUU-3′
Biotinylated probe primers	
Oligonucleotide name	Sequence
Probe-1F	5′-GAATTGAGTGAAGGCCGTCA-3′
Probe-2R	5′-TTTTGCTGGCCTTTTGCT-3′
Non-homologous recombinant virus junction PCR primers	
Oligonucleotide name	Sequence
1F	5′-CATTATCGTGGCGCCTTATATAG-3′
2R	5′-ATCTCCGGTCGCTAATGATAG-3′
3F	5′-CGACCGGTAGCGCTAGATATA-3′
4R	5′-TGTAGAACCAACTGATGCATCTC-3′
1'F	5′-GGTCTTGCTTTGTGACTTTGAT-3′
4'R	5′-TTGTATCGCATCTCTCGGGTA-3′
e1F	5′-TCCAGTTGAATGGATTCGTCC-3′
e4R	5′-GGTGGATCCATTCAGAATACAAAC-3′
i1F	5′-CAATACCAACCCCAACAACC-3′
i4R	5′-GTACGCATTCCTGTAGATCTTG-3′
Cas9 recombinant virus primer sequences	
Oligonucleotide name	Sequence
N2L-1Fwd	5′-GAGCTCCAATGTACATACATCGCC-3′
N2L-2Rev	5′-ACTAGTGAAGTGGGATCCATAAGATG-3′
YFP-3F	5′-TGACGAATACCCAGTTC-3′
YFP-4R	5′-GATCCATGCACGTAAAC-3′

2.2 Virus culture

2.2.1 Large scale virus preparations and titering

All recombination experiments were performed using VAC strain Western Reserve (WR) originally obtained from ATCC. Stocks of WR were prepared by infecting twenty 15 cm dishes of BSC-40 cells at a multiplicity of infection (MOI) of 0.03. Cells were harvested 48-72 hours after infection [>90% cytopathic effect (CPE)] using cell scrapers and centrifuged at 2,000 ×g for 10 minutes in a JA-10 rotor (Beckman).

To culture recombinant viruses, 1 mL of plaque-purified virus was used to infect a 60 mm dish under MPA selection (25 µg/mL xanthine, 15 µg/mL hypoxanthine and 25 µg/mL mycophenolic acid) until >90% CPE was observed (48-72 hours). The cells and media were harvested and frozen and thawed three times to release the virus. Next, 500 µL of lysate was used to infect 10 cm dishes, again under MPA selection, until >90% CPE was achieved (48-72 hours). The cells and media were harvested and again frozen and thawed. Finally, 2-3 mL of these lysates was used to infect twenty-to-twenty-five 15 cm dishes for 48-72 hours. The cells were harvested using cell scrapers and recovered by centrifugation at 2,000 ×g for 10 minutes using a JA-10 rotor (Beckman).

Pelleted infected cells were resuspended in ice-cold 10 mM Tris pH 9.0 with 2 mM $MgCl_2$. Virus was released by Dounce homogenizing the cells. The inoculum was then centrifuged at 2,000 ×g for 10 min. The supernatant was put aside on ice while the pellet was resuspended in 10 mM Tris pH 9.0 with 2 mM $MgCl_2$, Dounce homogenized, and centrifuged again. The supernatant was pooled with the previously collected supernatant and the pellet was discarded. The pooled supernatants (11-12 mL) were incubated at 37 °C with 50 U/mL of Benzonase for 30 minutes, centrifuged again at 2,000 ×g for 10 minutes to remove residual cell debris and overlayed on 19 mL of 10 mM Tris pH 9.0 containing 36% sucrose. The sample was centrifuged at 26,500 ×g for 90 minutes in a JS 13.1 swinging bucket rotor (Beckman). The pellet was resuspended in 100 µL/15 cm plate of 10 mM Tris pH 9.0 to create a final virus stock. Sucrose purified virus was used as the stock for

all of the experiments reported in this thesis. All of the recombinant viruses underwent further virion purification and DNA extraction (2.4.1).

Virus titers were determined using 10-fold serial dilutions of virus in serum-free media. BSC-40 cells were infected in 6-well plates in duplicate wells, or 12-well plates in triplicate, at 37 °C for 1 hour. The inoculum was replaced with media supplemented with 1% carboxymethyl cellulose (CMC). The cells were incubated at 37 °C for 48 hours then fixed and stained with a solution of 30% formaldehyde, 5% ethanol, and 0.13% crystal violet for at least 1 hour. The solution was removed, and plaques counted to determine the titer. Virus titers were calculated using the equation below:

$$Titer = \frac{Average\ plaque\ forming\ unit\ (PFU)\ count}{\big(Volume\ plated\ (mL)\big) \times (dilution\ factor)}$$

2.2.2 Producing, isolating, and quantifying recombinant viruses

2.2.2.1 Producing recombinant viruses

Most experiments were performed on BSC-40 cells in 6 cm dishes at 90-100% confluency. The cells were infected with VAC at an MOI of 0.05-to-1. The virus inoculum was prepared in serum-free media assuming that a 6 cm dish contained 1.4×10^6 cells. The cells were infected for an hour and the inoculum replaced with complete media. Two hours later, the cells were transfected with 1-2 µg of DNA, RNA, or cDNA·RNA using Lipofectamine 2000 (Thermo Fischer). When the cells were transfected with cDNA·RNA hybrid molecules, the quantities refer to the amount mRNA that was reverse transcribed. The nucleic acids were mixed with the Lipofectamine 2000 at a ratio of 3 µL per 1 µg of nucleic acid in 500 µL of Opti-MEM and incubated on ice for 20 minutes before transfecting the infected cells. Two days later the cells were harvested using cell scrapers, freeze-and-thawed three times to release the virus, and screened for recombinant viruses. For these experiments, I found that an MOI = 0.05 and 2 µg of transfected nucleic acid was most efficient.

2.2.2.2 Quantification and isolation of recombinant viruses

The cell lysates (sec. 2.2.2.1 and 2.2.2.2) were titered as described (sec. 2.2.1) to determine the total virus yield. To count recombinant viruses, diluted virus (typically 10^{-2} to 10^{-4}) was used to infect 10 or 15 cm dishes of BSC-40 cells for 1 hour then overlaid with MPA-containing media plus 0.75% Noble agar (Difco). From 5 to 30 plates of BSC-40 cells were infected and screened. A fluorescence microscope was used to find fluorescent plaques 3-5 days later. The number of fluorescent plaques were counted and then used to estimate the recombinant frequency (RF) using the equation:

$$RF = \frac{Number\ of\ recombinant\ plaques}{Total\ virus}$$

These numbers are dependent on the transfection efficiency, so we also calculated an "RF ratio" by dividing the RF measured in the absence of any homology between substrate and virus, and the RF measured in a parallel experiment on the same day using DNA substrates bearing N2L homology:

$$RF\ ratio = \frac{RF\ (no\ homology)}{RF\ (DNA\ with\ N2L\ homology)}$$

To purify recombinant viruses, fluorescent plaques were picked from under the agar overlay, resuspended in 1 mL of 10 mM Tris·HCl pH 8, freeze/thawed, and then replated three more times under agar and with drug selection. Further rounds of plaque purification used 6-well plates, instead of 10 cm dishes, plating 10-fold dilutions of the recombinant virus in duplicate wells (1:2, 1:20 and 1:200). Plaques were picked from wells infected with the most diluted virus to avoid picking siblings. The recombinant VAC were plaque purified at least three times, the viruses were bulked up (sec. 2.2.1) and the DNA was extracted from sucrose gradient purified virus (sec. 2.4.1) for sequencing and Southern blotting (sec. 2.4.4).

2.2.2.3 Producing recombinant viruses in Cas9-transfected cells

These experiments were performed with the *Mss*I linearized pPN_YFP and pPN_YFP+N2H plasmids and cells cultured in 6-well plates. BSC-40 cells were

seeded to sub-confluency (<85%) and infected at MOI = 0.05 for 1 hour. Each well was assumed to contain $6.3x10^5$ cells. Following infection, the inocula were replaced with 2 mL of Opti-MEM and immediately transfected with 2 µg of DNA plus complexed Cas9-sgRNA.

DNA transfections were prepared in 200 µL reactions containing 2 µg of DNA and 6 µL of Lipofectamine 2000 (Thermo Scientific). The complexes were left to form for least 20 minutes before the transfection step. The Cas9-sgRNA complexes were prepared by mixing 2 µg of Cas9 protein (*S. pyogenes*, New England BioLabs, 20 µM) with 2 µg of sgRNA and diluting it to 25 µL with Opti-MEM. This was incubated at 25 °C for 10 minutes and then mixed with 4 µL of Lipofectamine 2000 (Thermo Scientific) plus 221 µL of Opti-MEM. The Cas9-sgRNA-Lipofectamine mix was allowed to complex for 30 minutes before adding it to the cells.

The DNA transfection mix (200 µL) and the Cas9-sgRNA transfection mix (250 µL) were added dropwise to each well. The cells were left to sit for 4 hours at 37 °C and the transfection mix replaced with fresh complete MEM. The cells were harvested 2 days later. Recombinant viruses were recovered and purified as described in Sec. 2.2.2.3. After 3 rounds of plaque picks, the viruses were bulked up on BSC-40 cells in the presence of MPA using 6-well plates. The infections were allowed to proceed for 3-5 days where ~ 85% CPE was observed. Viral DNA was extracted from the infected cells as described in sec. 2.5.2.

2.2.2.4 Assaying double-strand breaks in Cas9-transfected cells

To determine if VAC genomes were being cut *in vivo*, the approach described in sec. 2.2.2.2 was used with some modifications. The cells were infected at MOI = 3 (rather than 0.05) and the transfection of the plasmid substrate was omitted. Viral DNA was harvested from the infected and transfected cells at 6-, 8-, 12- and 24-hours post-infection as described (sec. 2.5.2). Ten micrograms of the extracted DNA was digested overnight with the *Hin*dIII (ThermoFischer) to ensure a complete digest and fractionated on a 1% agarose gel. The DNA was transferred to a Biodyn B membrane and Southern blotted (2.4.4) using a biotinylated probe complementary to the N2L gene.

2.3 Bioinformatics

2.3.1 Virus sequencing

Sequencing was performed by the Applied Genomics Core at the University of Alberta. For Illumina sequencing, the DNA was processed using a Nextera XT DNA library kit and sequenced with the Illumina MiSeq. For Nanopore sequencing, DNA was processed using the SQK-LSK109 kit and sequenced using a MinION Mk1B device equipped with R9 chemistry flow cells (FLO-MINSP6). Base-calling of the Nanopore sequencing data was performed using the MinKNOW v1.14.1 and v18.03.1 (Oxford Nanopore Technologies Ltd.). Porechop (v0.2.4; https://github.com/rrwick/Porechop.git) was then used to demultiplex the reads and trim the adaptors.

2.3.2 CLC genomics assemblies

CLC Genomics Workbench 12 (Qiagen) was used to assemble some of the genomes using a "map-to reference" method and the Illumina reads. Reads were mapped to the sequence of both the YFP/gpt encoding selectable marker and the VAC genome. The CLC Genomics workflow that was used is shown in Fig 2.2. Default settings were used. Alignments were then manually searched for reads that encoded both selectable marker and VAC sequences. Hypothetical genomes could subsequently be assembled and used as templates for further rounds of map-to-reference alignments. This was labour intensive and rarely worked due to extensive rearrangements of the recombinant VAC genomes and, in some cases, where recombinant virus subpopulations were present. This is why we used Nanopore methods to sequence many of our other viruses as well as a different bioinformatics platform. All of the genomes that could be assembled using the map-to-reference method were additionally assembled *de novo* (sec. 2.3.3).

2.3.3 *De novo* genome assembly, read trimming, and contig joining

We used Unicycler (v0.4.8) (148) for *de novo* assembly of the majority of recombinant genomes. Before attempting the assembly, Illumina sequencing reads were trimmed using Trimmomatic (v0.4.8b) (149). Bases at each end of the reads were trimmed until a quality score of 28 was achieved. After this process any reads shorter than 75 nucleotides were also removed.

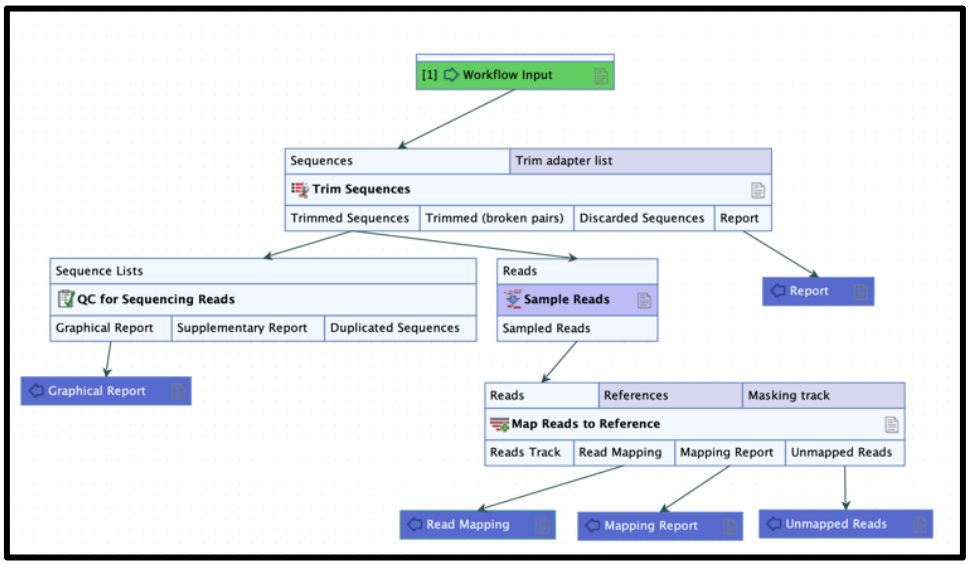


Fig 2.2 Map-to-reference workflow. This workflow shows how Illumina reads from recombinant viruses were processed and mapped to reference sequences using CLC Genomics Workbench 12. Default settings were used for each step.

The assembly of full length virus genomes was performed manually by visually inspecting the contigs and paths between them that were suggested by Unicycler . Visualization of the assemblies was performed with Bandage (150). The relative depth of coverage within the assemblies (i.e., contigs) was also used to detect the possible presence of multiple virus populations. Most assemblies were built using the "001_best_spades_graph.gfa" files generated by Unicycler. In a few cases, the "assembly_graph_with_scaffolds.gfa" found in the "spades_assembly" folder had to be used as the Unicycler algorithm removed the contigs encoding the selectable marker from the polished assemblies. Such viruses often had lower coverage of the selectable marker encoding contig relative to the VAC encoding contig. To verify the correctness of the assembly, the Illumina and Nanopore reads used for the Unicycler assembly were mapped to assembled genome using CLC Genomics Workbench 12 (Qiagen) to check for continuity.

In all cases, Unicycler could not completely assemble the sequences of the VAC inverted terminal repeats. To estimate the length of the recombinant virus genomes, the unresolved ITR sequences were assumed to be identical to the parental VAC sequences and were added to the recombinant genome assemblies. These assumptions were made to estimate the expected band sizes for the PFGE (sec. 2.4.2) and Southern blotting (sec. 2.4.4). The amount of coverage of the genomes was determined as described in sec. 2.3.2.2. An outline of the entire bioinformatics workflow is found in Fig 2.3.

2.3.3.1 Unicycler dependencies

Unicycler uses a number of additional programs to compute its genome assemblies. The versions of these programs that I used are the following: spades.py (v. 3.13.0+) (151), Racon (v1.3.2; https://github.com/lbcb-sci/racon), Pilon (v1.23) (152), Bowtie2 (v2.4.1+) (153), Samtools (v1.7+) (154) and Java (v.11.0.2+).

2.3.4 Sequencing depth of coverage analysis

The depth of coverage of the recombinant viruses was determined for both Illumina and Nanopore sequencing data. For recombinant VAC that contained marker or VAC sequence duplications, the genome assemblies hypothesized to contain the selectable marker were used to determine the depth of coverage. As for

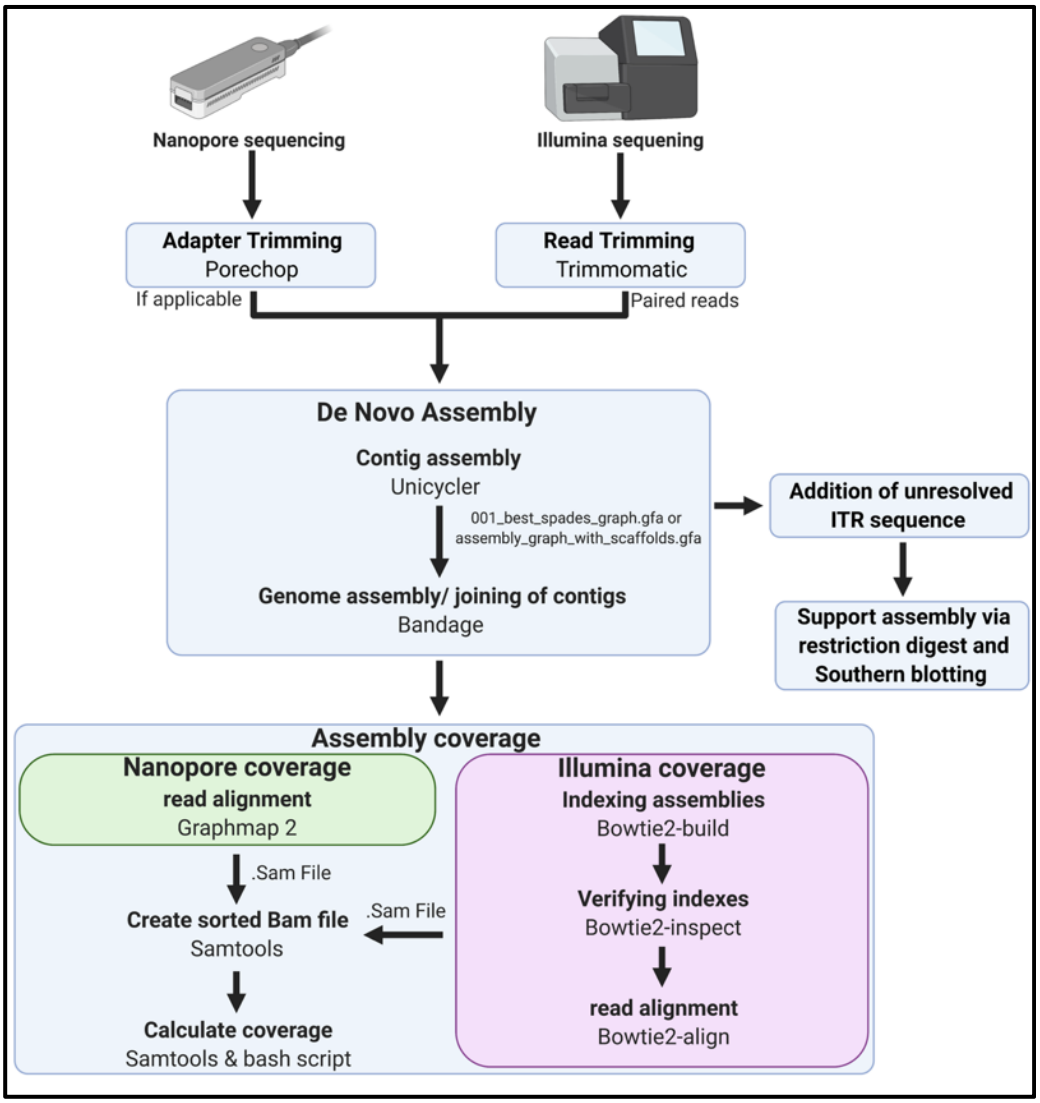


Fig 2.3 Bioinformatics workflow. The figure shows how the sequencing data were processed in a stepwise manner to build the *de novo* assemblies and calculate the genome coverage. The individual steps are listed in bold font along with the programs used at each step. The commands that were used in each step can be found in Table 2.2. Figure was created using BioRender.com.

the complementation-dependent recombinant viruses, the wild-type VAC genome and the marker-encoding defective genome were analyzed separately. VAC coverage was determined using all of the VAC encoding contigs that could easily be assembled (excluding unresolvable ITR sequence). For the complementation-dependent genomes, the depth of coverage was determined using the contigs unique to the recombinant viruses.

The read alignments for the depth of coverage calculations were performed using Bowtie2 (v2.4.1) (153) for Illumina reads and Graphmap (v0.5.2) (155) for Nanopore reads. The Nanopore reads were aligned directly to FASTA formatted assemblies but the Illumina read mapping required additional steps. The assembled viral genomic sequences were indexed using Bowtie2-build. Bowtie2-inspect was used to verify successful indexing of genomic sequences. This was confirmed by comparing the total length in base pairs of the original reference map to that of the indexed reference. Having confirmed successful indexing, paired reads were aligned using Bowtie2-aligner. The output of both Graphmap2 and Bowtie2 alignments comprised SAM formatted output files.

To calculate the depth of coverages of genome assemblies for both the Nanopore and Illumina sequencing data, the SAM formatted alignments were first converted into sorted BAM files using Samtools (v1.10). The average depth of coverage and standard deviations were subsequently calculated from the sorted BAM files using Samtools and a bash script. An outline of all the commands used for these processes can be found in Table 2.2.

2.3.5 Recombination junction analysis

DNA sequences overlapping the junction between the selectable substrate and virus sequence were aligned by hand. About 30nt on either of the junctions were aligned and inspected for matching bases between the original VAC sequence and pPN_YFP plasmid sequences. Total bases matching between the VAC and pPN_YFP sequences were noted. The number of bases expected to match randomly between VAC and plasmid sequences was calculated to be 0.24 including a correction for GC content. The topoisomerase I site motif that was used to investigate the junctions were 5’-YCCTT-3’ and 5’-AAGGR-3’ (156).

Table 2.2 Command lines used for the *de novo* assembly and genome coverage calculations

Step	Command
Trimming of Illumina reads (Trimmomatic)	java -jar trimmomatic-0.39.jar PE -phred33 input_forward.fq.gz input_reverse.fq.gz output_forward_paired.fq.gz output_forward_unpaired.fq.gz output_reverse_paired.fq.gz output_reverse_unpaired.fq.gz LEADING:28 TRAILING:28 MINLEN:75
Indexing and adapter trimming of Nanopore reads (Porechop)	porechop -i albacore_dir -b output_dir
Illumina read *de novo* assembly (Unicycler)	Unicycler -1 Fwd_Illumina_reads_1.fastq.gz -2 Rev_Illumina_reads_2.fastq.gz -o output_dir --keep 3 --no_rotate --pilon_path Path_to_pilon-1.23.jar
Illumina and Nanopore read *de novo* hybrid assembly (Uniycler)	Unicycler -1 Fwd_Illumina_reads_1.fastq.gz -2 Fwd_Illumina_reads_2.fastq.gz -l Nanopore_reads.fastq -o output_dir --keep 3 --no_rotate --pilon_path Path_to_pilon-1.23.jar
Mapping nanopore reads to reference and outputting a sam file (Graphmap)	graphmap align -r reference.fa -d reads.fasta -o output_nanopore.sam
Indexing reference sequence for Illumina mapping (bowtie2)	bowtie2-build reference_input.fasta bt2_base_index_name
Verifying bowtie index (bowtie 2)	bowtie2-inspect -s bt2_base_index_name
Mapping Illumina reads to indexed reference outputting sam file (bowtie 2)	Bowtie2 -x bt2_base_index_name -1 Fwd_Illumina_reads_1.fastq.gz -2 Rev_Illumina_reads_2.fastq.gz -S output_Illumina.sam
Converting sam files to sorted bam files (Samtools)	samtools view -u alignment.sam | samtools sort -o output_prefix.bam
Calculating average depth and standard deviation from sorted bam file (Samtools & bash script)	samtools depth -a sorted_bamfile.bam | awk '{sum+=$3; sumsq+=$3*$3} END { print "Average = ",sum/NR; print "Stdev = ",sqrt(sumsq/NR - (sum/NR)**2)}'

2.4 DNA extraction, restriction analyses, and Southern blotting

2.4.1 Gradient purification of virus stocks and virus DNA extraction

Gradient purifications were performed as described in X. D. Yao and D. H. Evans (157) using 24-40% sucrose gradient containing 1 mM Tris pH 9.0. Sucrose

cushioned virus (2.2.1) was sonicated 3 times for 30 seconds at 20 kHz, overlaid on the preformed gradient, and centrifuged for 50 minutes at 26,000 ×g using an SW28 rotor and Ultra-Clear centrifuge tubes (Beckman). The milky band that formed between the 32 and 40% sucrose layers was put aside while the pellet at the bottom of the tube was resuspended in 1 mM Tris pH 9.0, reapplied to another sucrose gradient, and centrifuged again. The two milky bands were combined, diluted at least 2-fold in 1 mM Tris pH 9.0, and pelleted at 33,000 ×g using a SW40Ti rotor. The virus pellets were resuspended in 10 mM Tris pH 8.0 and the DNA extracted as described by G. J. Kotwal and M. R. Abrahams (158). This involved lysing the virus using 2 x lysis buffer (50 mM Tris-HCl, pH 9.5, 0.7 M NaCl, 10 mM EDTA, 1% SDS) for 10 minutes at room temperature. Proteinase K was added to the sample to a final concentration of 0.5 µg/mL and digested overnight at 37 °C. Next day, the DNA was extracted twice with phenol:chloroform:isoamyl alcohol (25:24:1) followed by a chloroform:isoamyl (24:1) alcohol extraction. The DNA was precipitated by adding 2.5 volumes of 95% ethanol, and 0.1 volumes of 3 M sodium acetate pH 5.2, and frozen at -80°C for at least one hour. The DNA was then pelleted by centrifugation, washed with 70% ethanol, dried and resuspended in 10 mM Tris pH 8.0.

2.4.2 Pulsed-field gel electrophoresis (PFGE) and DNA ladders

Many of the recombinant viruses were subject to additional analysis to confirm the correctness of the assemblies. Virus DNA (500-1000 ng) was cut and mapped using restriction enzymes. Complementation-dependent recombinant viruses were cut with *Asc*I plus *Mre*I or *Not*I. All other recombinant viruses were cut with *Hind*III. All the reactions used Fastdigest enzymes and were performed using 10x Fast Digest Green Buffer (Thermo Scientific). Wild-type VAC was also included on any gel for comparison. The size markers included a lambda phage PFG ladder (NEB Labs) and a 2:3 mix of GeneRuler High Range DNA and GeneRuler 1Kb plus DNA ladders (Thermo Scientific). The lambda ladders were embedded in the wells of the gel using 1% SeaPlaque^{TM} low melt agarose dissolved in 1x Tris-Borate- EDTA buffer (TBE). PFGE was performed using 1% Seakem® LE agarose gels (Lonza) in 1x TBE. The gels were cast using the cassette supplied

with the CHEF-DRII apparatus. The DNA was separated using 0.5x TBE buffer, 5.2 V/cm, a linear switch time gradient of 1-to-10 seconds, and a 120° pulse angle. A run time of 13 hours was used for complementation-dependent viruses and 13.5 hours for all other recombinant viruses. The DNA was stained with SYBR gold and visualized using a Gel Doc™ XR+ system. The DNA was transferred from the agarose gels to Biodyn B membranes as described below (Sec. 2.4.4).

2.4.3 Biotinylated probes

Two different biotinylated probes were used for the Southern of recombinant viruses. The "PCR probe" was produced using the pDGloxP plasmid (145) as a template. The probe was amplified using Probe-1F and Probe-2R primers (Table 2.1). The PCR reactions (25-50 µL) contained the following reagents: 1x Taq buffer with $(NH_4)_2SO_4$ (Thermofisher), 3 mM $MgCl_2$, 0.25 µM dATP/dGTP/dCTP, 0.15 µM dTTP, 0.1 µM Biotin-16-dUTP (Roche), 0.4 µM of each of the forward and reverse primer, 0.4 ng/µL pDGloxP and 0.05 U/µL Taq polymerase. The reactions employed 30 amplification cycles and an annealing temperature of 51 °C.

The nick-translated probe was prepared using the pPN_YFP plasmid as a template and a protocol supplied by Fermentas. The reactions (25-50 µL) contained the following reagents: 1x DNA polI buffer, 0.05 mM dATP/dGTP/dCTP, 0.03 mM dTTP, 0.02 mM Biotin-16-dUTP (Roche), 10 ng/µL pPN_YFP 0.9 U/µL *E. coli* polI, $8x10^{-5}$ U/µL DNAse I. The reactions were incubated at 15 °C for 2 hours and then 1µL of 0.5 M EDTA was added to halt the reaction.

The N2L gene probe that was used to detect Cas9-induced double-strand breaks of the N2L locus. The probe was prepared using PCR, VAC genomic DNA, and N2L-1F and N2L-2R primers (Table 2.1). A PCR-amplified template was purified using a PCR purification column (Qiagen) prior to being used to prepare the N2L probe. The same PCR reagents and concentrations were used to make this probe as for the other PCR-generated probe. The PCR reactions were used 30 amplification cycles and an annealing temperature of 54°C.

Gel electrophoresis was used to monitor probe production. The probes were purified using PCR purification columns (Qiagen) and quantified using a spectrophotometer.

2.4.4 Southern blotting

The DNA was transferred to a Biodyn B membrane using capillary transfer. All the gel incubation steps that follow were done with sufficient volume to completely submerge the gel and placed on a rocker. The DNA was initially depurated by soaking the gel in 0.25 M HCl for 10 minutes. The gels were then rinsed twice with 300-500 mL of distilled water then treated twice with a denaturing solution (1.5 M NaCl, 0.5 M NaOH, pH 13) for 15 minutes each. The gels were washed again twice with 300-500 mL of distilled water and treated with a neutralizing solution (1.5 M HCl, 1 M Tris-HCl, pH 7.5) twice for 15 minutes. The gel was transferred to the membrane using 10x SSC (1.5 M NaCl, 150 mM sodium citrate, pH 7.0) via upward capillary transfer for 18-22 hours. Following the DNA transfer, the membrane was washed in 2x SSC (0.3 M NaCl, 30 nM sodium citrate, pH 7.0) and cross-linked with a UVP Ultraviolet Crosslinker, CL-1000 series at the "maximum" energy (Max: 9999).

Southern blot membranes were blotted using the North2South™ Chemiluminescent Hybridization and Detection Kit (Pierce). Hybridization was performed as described by the kit. For most blots, just a PCR probe was used for hybridization. In some cases, a mixture of the nick-translation and PCR probes were used. The blots were imaged using an ImageQuant LAS-4000 Imager and software (GE Lifesciences) on the chemiluminescence setting, precision exposure, Epi-luminescence imaging plate, and exposure times ranging from 0.5 seconds to 2 minutes.

2.5 SgRNA targeting verification & recombinant virus PCR

2.5.1 Cas9 *in silico* sgRNA testing

SgRNA-cleavage of VAC DNA was tested *in silico* before performing *in vitro* experiments. This used the protocol supplied with the Cas9 protein (New England BioLabs) and served to test whether the sgRNA-Cas9 complex could cut VAC DNA. To perform this experiment, 1 μmol of Cas9 was complexed with 900

nmol of sgRNA (Table 2.1) for 10 minutes at 25 °C in NEBuffer 3.1. VAC DNA (500 ng) was then added to the mix and the sample was incubated at 37 °C for 15 minutes. Half of each sample were then separated on a pulse-field gel (Sec. 2.4.2) to verify if VAC DNA cleavage occurred.

2.5.2 Isolating virus DNA from infected cells

The media was removed from virus-infected cells, in 6-well plates and replaced with 0.5 mL of virion lysis buffer (1.2% SDS, 50 mM Tris·HCl pH 8, 4 mM EDTA pH 8, 4 mM $CaCl_2$ and 0.2 mg/mL Proteinase K). The cells were digested overnight at 37 °C. The released DNA was sheered by pipetting the samples up and down multiple times, and transferred to 1.5 mL microfuge tubes. Phenol-saturated buffer (0.5 mL) was added to the lysates, vortexed, and then centrifuged at maximum speed for 15 minutes in a Beckman, 22R microfuge. About 300 µL of the aqueous phase was transferred to a new tube and the DNA was precipitated by adding 1 mL of ice-cold 95% ethanol and 50 µL of 3 M sodium acetate. The samples were incubated at -80 °C for 1 hour and centrifuged again at max speed for 30 minutes. The DNA pellet was washed with 500 µL of 70% Ethanol, centrifuged at maximum speed for 15 minutes, and the supernatant was discarded. The DNA pellets were left to dry and subsequently resuspended in 10 mM Tris pH 8.

2.5.3 PCR amplification of recombination VAC junctions

The PCR reactions containing 0.1 U/µL *Taq* DNA polymerase (Thermo-Fisher), 1x *Taq* buffer, 6 ng/µL of gradient-purified genomic DNA, 30 µM $MgCl_2$, 200 nM dNTPs, and 0.2 µM each of the forward and reverse primer were used to assay the VAC RN3 and RN5 strains. The primers used in these reactions are listed in Table 2.1(RN3 strain:1F-2R, 3F-4R, 1F-4R; RN5 strain: 1'F-3F, 2R-4'R, 1'R-4'R). All reactions used 30 thermal cycles and an annealing temperature of 48 ˚C.

The VAC RN4 strain was assayed using the following primer pairs: e1F+2R, i1F+2R, 3F+e4R, 3F+i4R, e1F+e4R, i1F+i4R (Table 2.1). PCR reactions contained 0.02 U/µL Phusion DNA polymerase (Thermo Fischer), 2 ng/µL of gradient-purified genomic DNA, 1x GC Phusion buffer, 200 nM dNTPs, and 500 nM of both the forward and reverse primer. All reactions used 30 thermal cycles.

An annealing temperature of 48 ˚C was used for the reaction containing DNA from the RN3 or RN5 strain and 60 ˚C for the RN4 strain. All reactions used 30 thermal cycles. An annealing temperature of 60 ˚C.

The recombinant VAC recovered from Cas9-sgRNA transfected cells were assayed for N2L recombinants using N2L-1F and N2L-2R primers (Table 2.1). The PCR reactions contained 0.02 U/µL Phusion DNA polymerase (Thermo Fischer), 2 ng/µL of DNA lysate (Sec. 2.5.2), 1x GC Phusion buffer, 200 nM dNTPs, and 500 nM of both the forward and reverse primer. Any recombinant viruses were further characterized by PCR using the same reaction mix listed above and the following primer pairs: N2L.1F-YFP.3F, N2L.1F-YFP.4R, N2L.2R-YFP.3F, N2L.2R-YFP.4R (Table 2.1). The PCR products were fractionated on 1% agarose gels. A GeneRuler 1Kb plus DNA ladder (Thermo Fischer) was used to determine the size of the products.

Chapter 3 – Recombination substrates and frequency of events

3.1 Introduction

To investigate the substrates that may be targeted for VAC recombination, I needed to consider the virus life cycle. These viruses are cytoplasmic viruses that encode intronless genes. This suggests that any gene-bearing sequence, acquired by a poxvirus, would likely have been spliced before being recombined into the viral genome. As a cytoplasmic virus, VAC could encounter three kinds of intronless nucleic acid targets including mRNA, cDNA·RNA hybrid molecules, or duplex DNA. Some of these molecules would have to have been reverse transcribed, this would require a reverse transcriptase encoded by a retrotransposon or endogenous or exogenous retrovirus.

Poxviruses are known to catalyze high-frequency homologous recombination between DNA substrates (102). Such reactions require as little as 12-15 bp of homology (101, 102) and in VAC-infected cells they are catalyzed by the E9 DNA polymerase (100, 109). Although this method is efficient it seems unlikely that it would have evolved to catalyze recombination between cellular and virus DNAs. Moreover, the probability that the ends of two duplex DNA molecules share homologous sequences >12bp long is statistically improbable. This suggests that capturing a fragment of a host gene would be catalyzed by a homology-independent event. Such events include retrotransposition or NHR reactions.

Retrotransposition has the potential of being an important mechanism that mediates horizontal gene transfer from the host to a virus. Recent studies have shown this by detecting rare horizontal gene transfers into VAC which bore hallmarks of LINE-1 mediated retrotransposition (17, 18). Other studies have shown that, in nature, poxviruses are large enough to transmit infectious retroviruses encoded within their genome (159, 160). This could provide a source of the reverse transcriptase activity that is needed to convert spliced mRNA into DNA substrates. These observations led us to speculate that replicating poxviruses are likely exposed to pools of different cellular nucleic acid fragments (RNA,

cDNA·RNA, and/or fragments of dsDNA). This then led us to test whether simple-transfection assays could be used to see what happens to such transfected molecules when they are encountered by the VAC replication machinery.

In this chapter, I discuss how I generated these three different kinds of nucleic acid substrates and used them to test whether such molecules are substrates for NHR in VAC-infected cells. I also describe how some of the simplest NHR VAC genomes could be sequenced and characterized using a "Map-to-Reference" method.

3.2 Results and discussion

3.2.1 Recombination substrates verification

Before preparing any of the substrates for use in recombination studies, the identity of the plasmids encoding the marker sequences were verified. The presence of the YFP/gpt marker in the pPN_YFP+N2H and pPN_YFP plasmids were verified via *EcoR*I and *MssI* digests (Fig 3.1). The *EcoR*I digests were expected to completely excise the selectable marker sequence (plus any flanking N2L homologies) from the PCR4 plasmids and the *MssI* digests were expected to linearize the plasmids. The pPN_YFP+N2H digests produced the expected restriction digest patterns. The pPN_YFP+N2H plasmid was cut into 2.1 kb and 3.9 kb fragments by *EcoR*I. It was also cut into a 6.1 kb band by *Mss*I. The pPN_YFP plasmid digest also produced the expected pattern of fragments. The plasmid was linearized into a 5.5 kb band by *Mss*I and yielded 1.5 kb and 3.9 kb bands with the *EcoR*I digest. Besides these restriction enzyme digests, the orientation of the marker sequences were further verified through Sanger sequencing. These results were consistent with the pilot studies completed by a former project student (Mr. Peter Norris), so I proceeded to make the substrates for our recombination experiments

A schematic showing how the recombination substrates were prepared can be found in Fig 2.1B. The plasmids were first linearized with *Mss*I and the completeness verified by gel electrophoresis (Fig 3.2A). Next the two linearized DNA templates were transcribed with a T7 Maxiscript Kit and poly-adenylated. with a Poly(A) Tailing Kit and aliquots were analyzed using a 1% formaldehyde

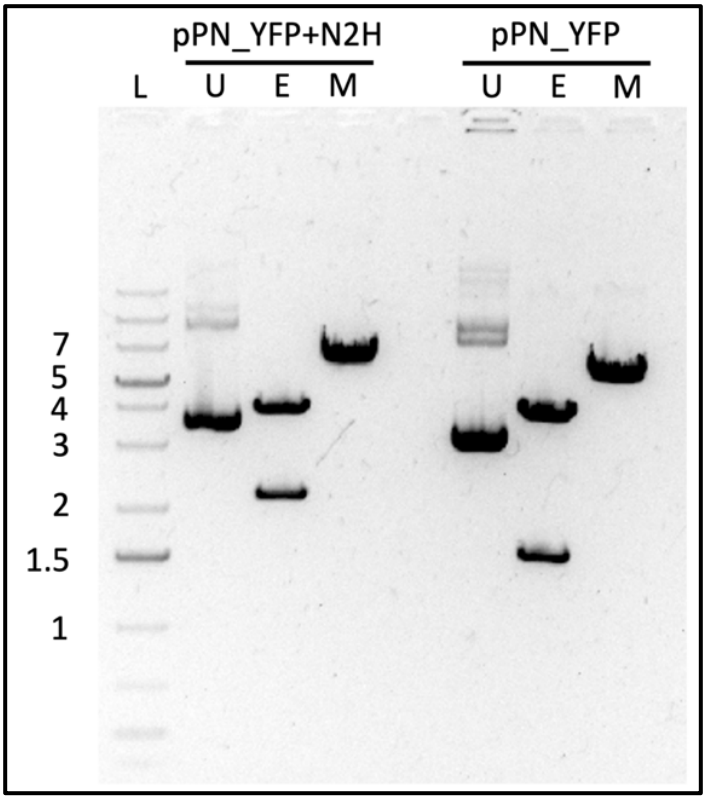


Figure 3.1 Verification of the plasmids used for these experiments. The two indicated plasmid constructs were digested with EcoRI (E) or MssI (M), or left untreated (U), and then size fractionated on 1% agarose gels and strained with *Sybr Safe*. The gel pattern confirmed the presence of the YFP/gptmarker that had been TOPO cloned into the pCR4 vector and into the plasmids created by Peter Norris (pPN_YFP+N2H and pPN_YFP). The samples were run in parallel with a 1 kb plus GeneRuler DNA ladder (L).

denaturing gels (Fig 3.2B). Both the T7 transcription (T) and poly-adenylation reactions (A) were run next to each other to confirm that the reactions were successful. The transcription reactions yielded the expected 2.2 kb and 1.6 kb bands with the pPN_YFP+N2H and the pPN_YFP templates, respectively. The poly-adenylation reactions produced "smearing" of the bands as was also expected.

Finally, the poly(A)-tailed molecules were reverse transcribed with Supercript III or IV and using oligo·$dT_{12\text{-}18}$ primers, producing the desired cDNA·RNA substrates. To document the production of these molecules, the reverse-transcribed products were treated with RNAse A and RNAse H (Fig 3.2C). These are a single-stranded RNA-specific nuclease and a cDNA·RNA specific RNA nuclease, respectively. When comparing the untreated control (C) to the RNAse treated samples we observed that the reverse-transcribed products were highly sensitive to RNAse H but not to RNAse A. This pattern supports the hypothesis that the desired product was successfully reverse-transcribed. The presence of a faint smear in the RNAse H treated lanes is likely due to the (expected) presence of ssDNA and possibly some residual ssRNA. The similarity in intensity between the control and the RNAse A treated sample further suggests that most of the RNA the reverse-transcription reactions were successfully converted into cDNA·RNA hybrid molecules.

3.2.2 Substrates used for recombination and recombination frequencies

The three different substrates were tested to see which (if any) of these molecules could be captured in VAC recombination reactions. These experiments were performed by transfecting VAC-infected cells with a nucleic acid substrate made up of linearized dsDNA, cDNA·RNA, or single-stranded RNA encoding a YFP/gpt cassette and flanked, or not, by VAC sequences homologous to the non-essential N2L gene (Fig 2.1A). The infected and transfected cells were screened for fluorescence under MPA drug selection, to look for recombinant viruses encoding the YFP/gpt selectable gene marker. As most recombination events were quite rare, large numbers of the progeny virus had to be screened. Any fluorescent VAC plaques were subsequently plaque purified, and the genomes were sequenced. To

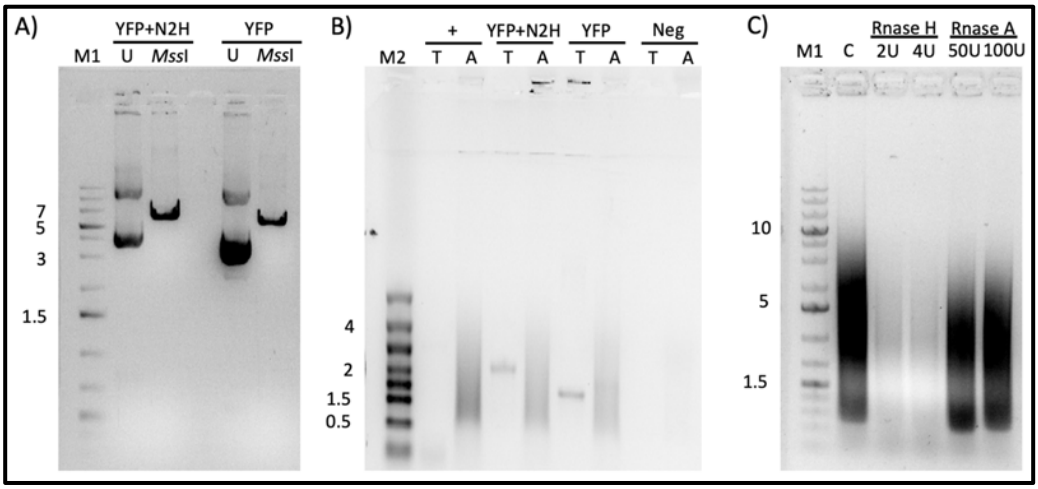


Figure 3.2 Characterization of the recombination substrates. A) Linearization of the pPN_YFP+N2H (YFP+N2H) and pPN_YFP (YFP) plasmids by *Mss*I restriction digest. Undigested (U) and linearized (*MssI*) samples were run on a 1% agarose gel and stained with SYBR-gold. B) The T7 transcribed (T) and poly-adenylated transcripts (A) were fractionated using a 1% formaldehyde denaturing gel then stained with SYBR-gold. The pPN_YFP+N2H and pPN_YFP plasmids were loaded along with a no-template negative control (Neg) and a positive control (+) supplied with the poly-A tailing kit. C) Reverse-transcription of the poly-adenylated pPN_YFP RNA transcript on a 1% agarose gel. About 500 ng of reverse-transcribed pPN_YFP plasmid transcripts (C) were run in parallel with samples that had been treated with 50-100U of RNAse A and 2-4U of RNAse H for 30 minutes before loading the gel. M1 and M2 are marker sequences used to determine the sizes of fragments.

avoid picking sibling plaques (duplicate progeny derived from a single initial event), I limited the number of plaques that were picked to just 1-4 recombinants from each transfection. The lysates were additionally titered to determine the frequency of recombinant viruses that were formed relative to the total virus population present in the infected and transfected lysates.

Recombinant viruses were successfully recovered using several different substrates. Recombinant viruses were recovered when transfecting VAC-infected cells with duplex DNA or cDNA·RNA substrates, whether or not the molecules encoded N2L homology. However, recombinants were never recovered from cells transfected with either kind of pure RNA substrate. As expected, the DNA bearing homology to the N2L gene provided the best substrates with an average recombinant frequency (RF) of about 6.7×10^{-4} over three independent trials (Table 3.1). This equates to the generation of about 670 recombinant viruses for every million viruses recovered. In comparison, about 30-fold fewer viruses were recovered from cells transfected with cDNA·RNA substrates encoding homology to N2L, about 21 recombinants for every million viruses recovered ($RF = 21 \times 10^{-6}$). The lack of N2L homology did not prevent recovery of recombinant virus from cells transfected with duplex DNA substrates. However, the frequency at which these viruses were recovered was at least 400-fold lower than the homologous DNA substrate yielding on average 1.6 viruses for every million viruses recovered ($RF = 1.6 \times 10^{-6}$). The cDNA·RNA substrate lacking any VAC homology yielded very few recombinant viruses. As a rough estimate these reactions are ~40,000 times less efficient than with duplex DNAs bearing N2L homology ($RF \leq 0.03 \times 10^{-6}$). To try and compensate for differences in transfection efficiency between different experimental replicates, I also calculated the ratio of the RF values relative to that of the RF measured for the homologous duplex DNA within that particular experimental replicate (Table 3.1). Despite this, the accuracy of our calculated RFs is difficult to state with absolute certainty. The screening process required screening of large quantities of cell lysates on multiple plates, which increased the possibility of missing recombinant plaques. However,

we can say with certainty that VAC can recombine substrates lacking homology and these substrates may be encoded by DNA or DNA·RNA.

Table 3.1 Vaccinia virus recombination frequencies (1)

Nucleic acid substrate	Type of Nucleic acid	Recombinant frequency ($\times 10^6$)	Ratio (RF/6.7x10^{-4})
Homologous (463 and 313 bp)	DNA	670	1
	cDNA・RNA	21	3×10^{-2}
	RNA	Not detected	–
Non-homologous	DNA	1.6	2×10^{-3}
	cDNA・RNA	≤0.03†	4×10^{-5}

†All of the experiments were performed 3 times but 2 of the 3 replicates produced no recombinants. Plaque counts are assumed to be Poisson distributed, so RF was estimated based on the assumption we might have missed 2 plaques in each of the negative replicates.

3.2.3 Assembling the simpler recombinant virus genomes and supporting PCR data

A number of the recombinant viruses were picked and further subjected to three rounds of plaque purification under agar and drug selection. The recombinant viruses were plated in 10-fold dilutions and plaques picked from the most diluted plating. This was done to avoid picking siblings and to purify the recombinant virus free from any wild-type parental virus that might have been present. In all, we plaque purified 24 recombinant VAC produced using non-homologous substrates. Five of these viruses were recombinants that were produced using non-homologous cDNA·RNA substrates and 19 with duplex DNA. We also included one virus that was produced using a cDNA·RNA substrate encoding N2L homology. The plaque-purified viruses were expanded in quantity (“bulked up”) and the genomic DNA was isolated for next-generation sequencing and further analysis.

Initial attempts to assemble the recombinant virus genomes were performed using a “Map-to-reference method” using Illumina sequencing reads and CLC

Genomics Workbench (Qiagen). This was accomplished by mapping sequencing reads to the selectable substrate and to the parental VAC genomic sequence. Alignments were visually inspected for clusters of reads that did not align to the substrate sequence (Fig 3.3A). Unaligned (faded) segments of reads were cross-referenced between substrate sequence alignments and the YFP/gpt alignments to ensure that they coincide (Fig 3.3B). Subsequently, the genomes were reassembled based on the alignments and the reads mapped to a new reference sequence to ensure that the alignment was correct (Fig 3.3C).

This method was labour intensive but worked well for viruses that encoded simple genome rearrangements and had high coverage of the recombination site. A couple of the viruses that were assembled this are shown mapped in Fig 3.4. In one case I observed a deletion of the VAC sequence at the site of recombination (Fig 3.4B and D). The deletion extended from the ITR sequence to the C1L gene (Fig 3.4). A second virus encoded a perfect 1.9 kb duplication of the VAC sequences flanking the substrate sequence (Fig 3.4C and E). The duplicated sequence extended from the I1L gene to the I4L gene. This method was also used to confirm the structure of a recombinant virus produced using a cDNA·RNA hybrid substrate encoding N2L homology. The sequencing data showed that the sequence encoding the N2L homology had recombined with the N2L gene via HR and lost the poly-deoxythymidine (poly-dT) tail used in the preparation of the cDNA·RNA substrate (Fig 3.5).

The map-to-reference method worked for only a handful of the virus genomes that I tried to assemble. Problems arose when large deletions, inverted translocations, or duplications appeared to be present. Some problems might also have been due to there being genetically-unstable subpopulations of viruses. This led us to change the approach from using a map-to-reference method to *de novo* assembly. This made it possible to fully automate the genome assembly process, and assemble entire virus genomes. This reduced the likelihood that we would miss something when visually inspecting mapped reads and stitching together unaligned read segments.

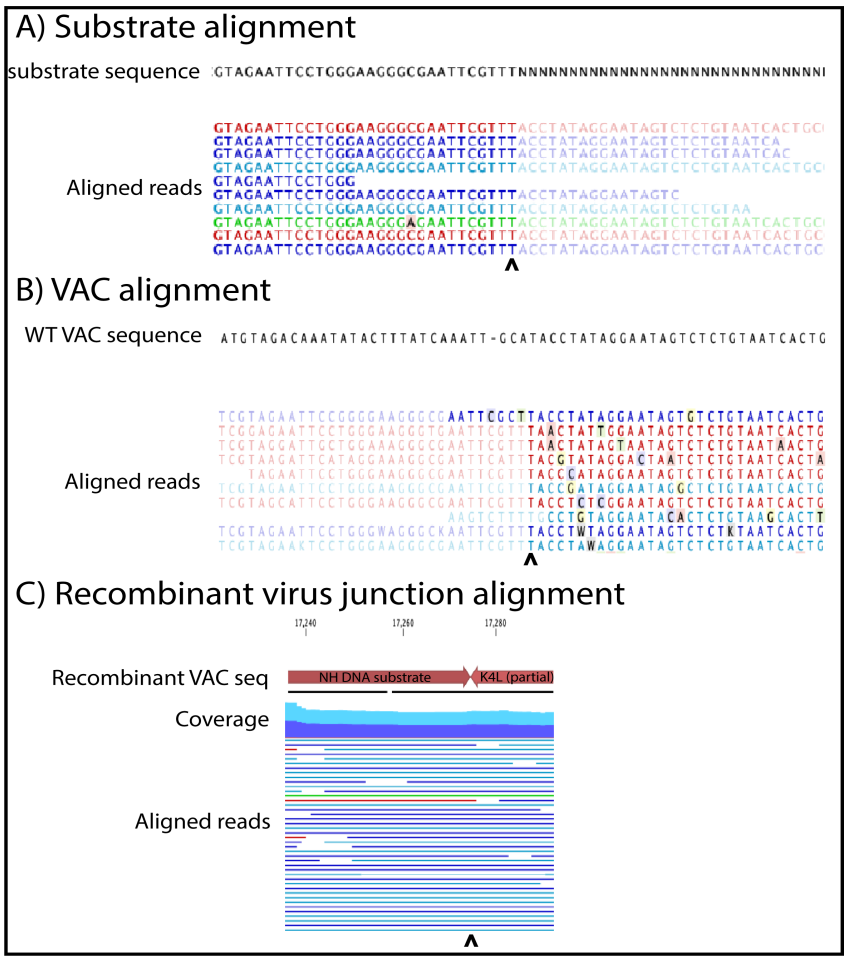


Fig 3.3 Illumina map to reference alignment and recombination junction analysis. A) CLC Genomics Workbench sequence alignment showing the Illumina reads located on the left-side of the right-hand junction sequence. Miscellaneous bases (N) were added to the end of the reference sequence to show the reads that aligned to the captured (i.e., transfected) DNA. Mapped read segments (solid colours) and unmapped read segments (translucent colours) are coloured based on the type of read that was mapped to the sequence. These include the paired reads (blues), forward reads (green) and reverse reads (red). B) Illumina reads encoding the VAC DNA sequence found on the right-side of the right-hand junction sequence. C) Final alignment of the reads to the deduced recombinant virus junction sequence. The recombination junction in all three alignments is denoted by a "^". Only a small subset of the aligned reads are shown.

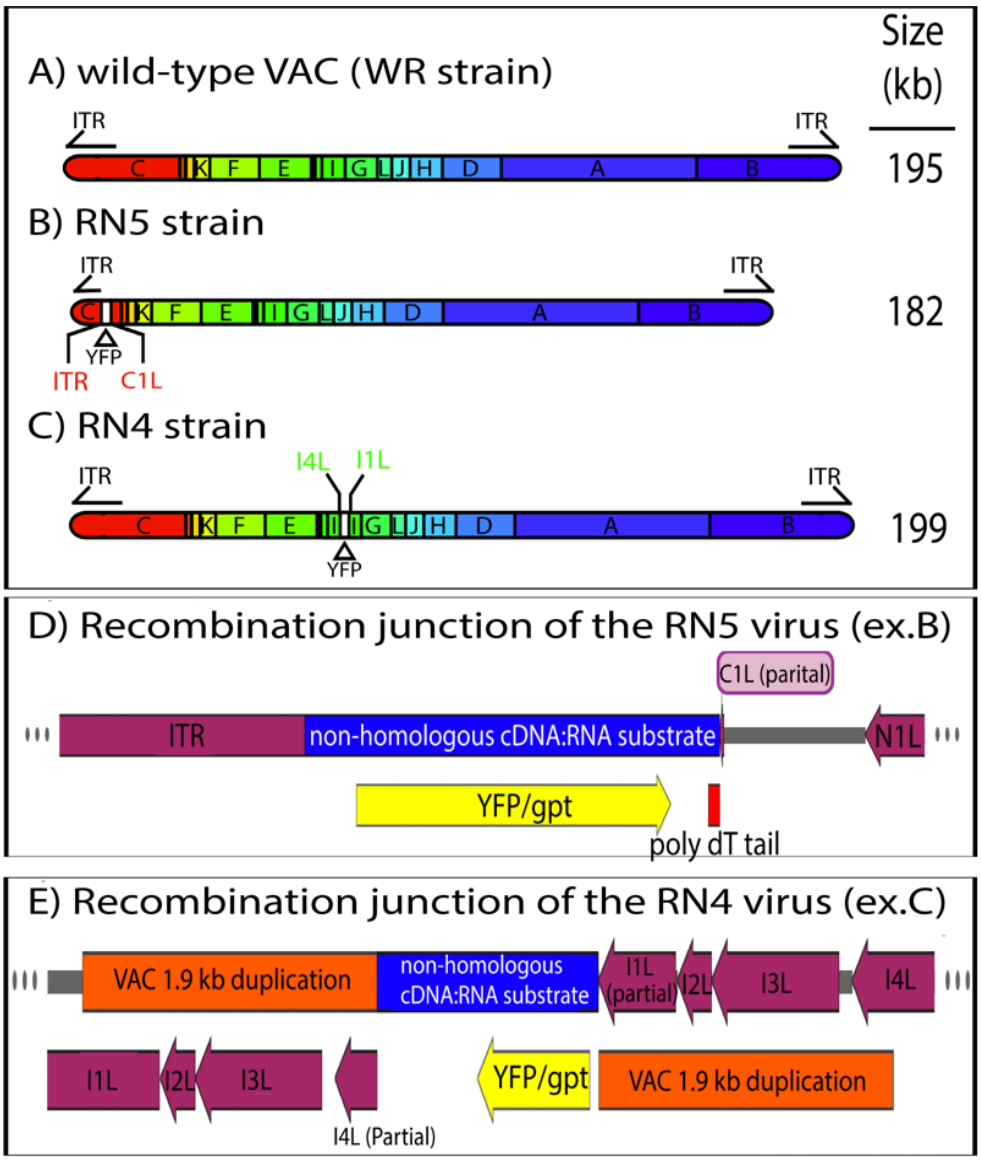


Figure 3.4 Example of simple recombinants analyzed using the map-to-reference method. (A) Schematic map of the parental VAC WR genome. Coloured and lettered segments represent the large VAC *Hin*dIII fragments for reference purposes. The right and left end of the virus genome encode inverted terminal repeats ("ITR", arrowed). (B and D) The non-homologous recombinant RN5 exhibited a 15 kb deletion next to where the transfected substrate ("YFP") had integrated. The deletion spans a segment of the left ITR and extended into the C1L gene. (C and E) The non-homologous recombinant RN4 encoded a 1.9 kb duplication flanking the inserted marker sequences. The duplicated sequence extends from I1L to the I4L.

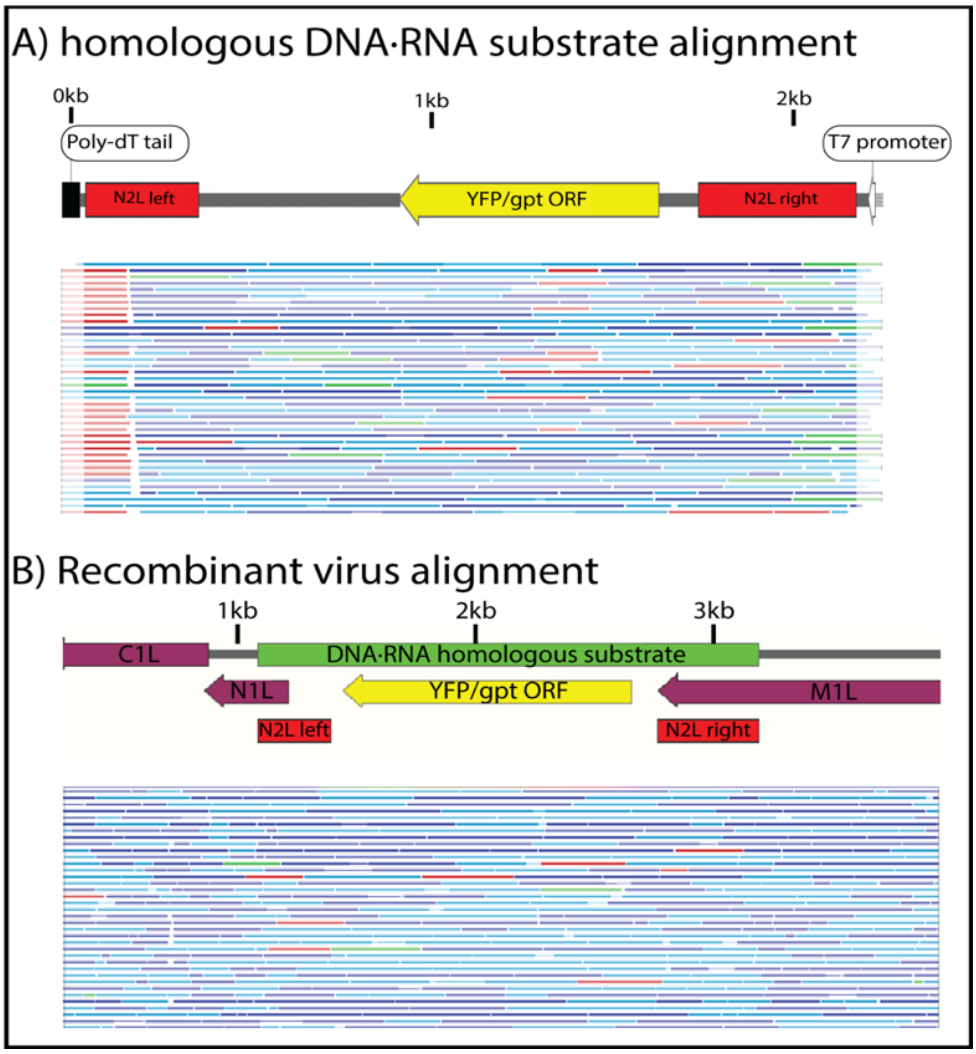


Fig 3.5 Illumina read alignments showing the site of integration of a cDNA·RNA hybrid substrate bearing N2L homology A) Virus-derived Illumina sequencing reads aligned to the original cDNA·RNA substrate. The sequences are coloured based on the type of reads that are mapped to the reference sequence. These include paired reads (blues), forward reads (green) and reverse reads (red). Note that none of the virus-derived reads aligned to the poly-dT tail on the left end of the hybrid substrate. B) Mapping the Illumina sequencing reads to the integration site and the recombination junctions. The virus encoded a clean insertion of the YFP/gpt marker accompanied by a deletion of the N2L gene.

3.2.3.1 Supporting PCR data

To support the map-to-reference assemblies, I used PCR to amplify the sequences overlapping the recombinant junctions in both the RN4 and RN5 recombinant viruses. Three pairs of primers were designed to interrogate the location and the orientation of the inserts in each virus. Two reactions served to amplify each of the two flanking junctions while a third PCR reaction was used to amplify the entire insert using VAC-specific primers.

I easily amplified the recombination junctions in the RN5 virus. The three different combinations of primers amplified the expected 2.4, 3.8, and 7.4 kb DNA fragments using primers (1'+3), (2+4'), and (1'+4'), respectively (Fig 3.6A).

The RN4 virus proved more difficult to document in this manner. In my first attempt I easily amplified the left (2.1 kb) and right (2 kb) junctions using primers (e1+2) and (3+e4), respectively (Fig 3.7C). However, the (e1+e4) primer pair yielded a 2 kb band instead of the 5.3 kb amplicon that I had expected (Fig 3.7C). This amplicon corresponded to one that would be expected if we had been using the parental VAC DNA (Fig 3.6B). This result was likely due to homologous recombination between the duplicated VAC sequence in a subset of viruses in the virus stock. These types of long tandem duplication have long been known to be unstable (161, 162). I purchased some new primers (i1 and i4) that were designed to hybridize within the VAC duplication (Fig 3.7A). These primers were designed in a way that would allow for amplification of a PCR product from a genome encoding the marker sequence and not from the subpopulation that had lost the marker through homologous recombination. Using these primers, I was able to amplify both the individual recombination junction (0.39 kb and 1.4 kb) and a 3 kb amplicon that encompassed both junctions (Fig 3.7C).

3.3 Conclusion

Overall, I was able to produce the homologous and non-homologous DNA, RNA and cDNA・RNA substrates that I required for my study of poxvirus non-homologous recombination (Fig 3.2). I demonstrated that VAC is capable of capturing a novel genetic sequence that shares no homology with the virus genome. The substrates that could be recombined into VAC's genome included DNA and

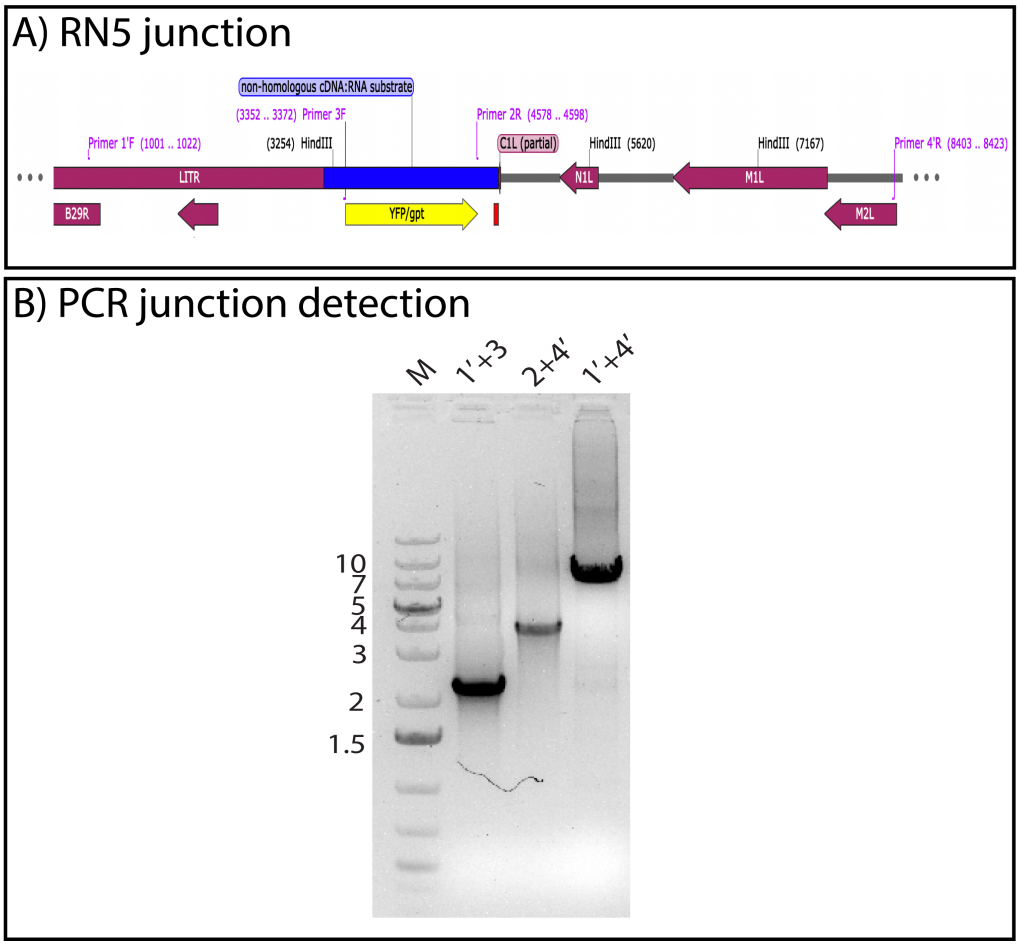


Fig 3.6 PCR amplification of recombination junctions in the RN5 virus. A) Hypothesized structure of the RN5 virus. RN5 bears a 15kb deletion next to the insertion site. The primers used for PCR are shown in purple text. B) PCR amplification of the two recombination junctions (1'F+3F and 2R+'4R) and the entire insert (1'F+4'R). The virus stock was grown under drug selection and purified before extracting the virus DNA.

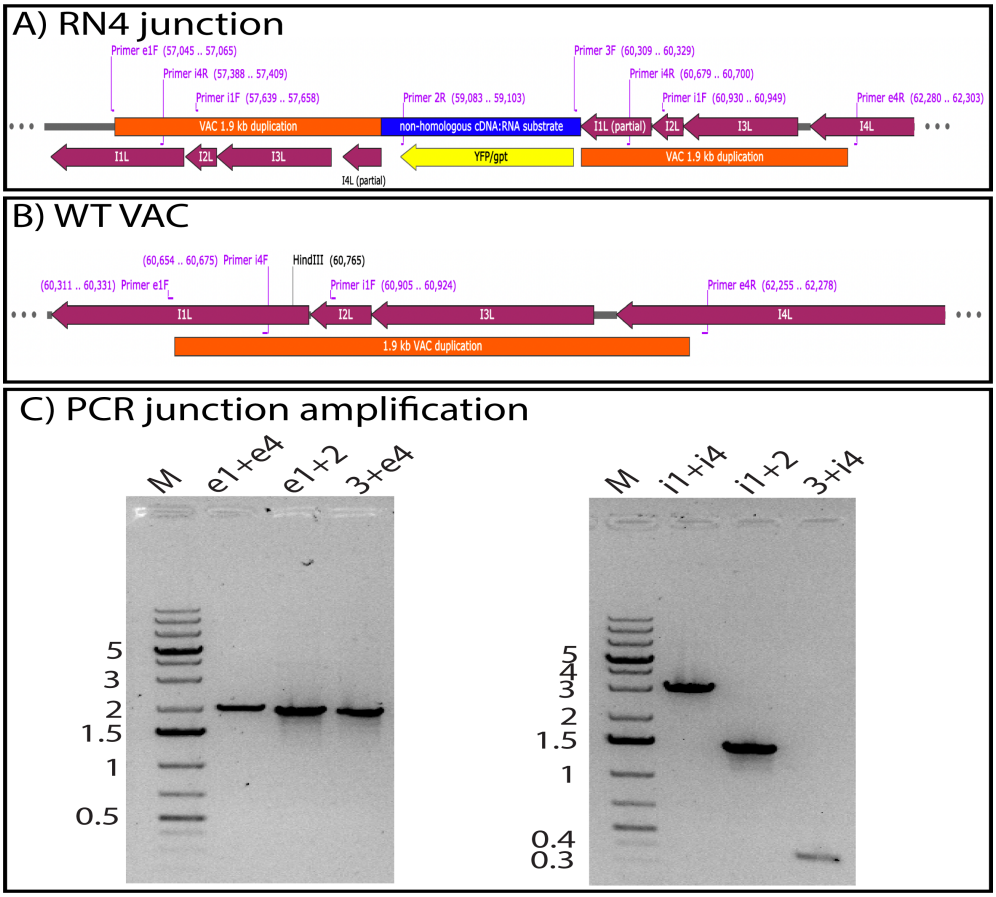


Fig 3.7 PCR verification of a recombinant strain RN4 encoding a 1.9 kb VAC duplication flanking the YFP/gpt insert. A) Map showing the sequences surrounding the insert in virus RN4. The PCR primers are labelled in purple. B) The virus that would be produced through homologous recombination between the two 1.9 kb flanking duplications. The PCR primers are also shown mapped onto this model. C) PCR amplification of the recombination junctions using genomic DNA extracted from a purified stock of the RN4 virus grown under drug selection. Primers (e1F+2R) and (3F+e4R) amplified the left and right junctions but (e1F+e4R) yielded only a 2 kb amplicon derived from virus that reverted to wild-type VAC through homologous recombination of the unstable tandem duplication. Two new primers that targeted regions within the duplicated VAC sequence were then used to amplify the individual junctions (i1F+2R, 3F+i4R) and 3 kb segments encompassing both junctions (i1F+i4R) and small segments of the duplicated VAC sequence (right panel).

cDNA・RNA. Additionally, I calculated the frequency at which I could recover recombinant virus in the presence of the homologous and non-homologous substrates (Table 3.1). The frequency at which I recovered recombinant VACs, was quite low, but may still be sufficient to be biologically relevant. In a natural poxvirus infection viremia has been recorded as high as 10^{10} genomes·ml^{-1} (163) . Lastly, I assembled a few recombinant genomes using a map-to-reference method which I supported with PCR data (Fig 3.6 and 3.7). The map-to-reference method worked for some genomes but proved unsuccessful for many of the genomes that were initially recovered. Because of this, I transitioned to using a *de novo* assembly method which I successfully used to assemble most of the virus genomes that were recovered (Chapter 4).

Chapter 4 – Mapping the VAC genomes that are products of non-homologous recombination

4.1 Introduction

As described on Chapter 3, I managed to recover and purify 14 different recombinant VAC that had been produced by transfecting non-homologous substrates into virus-infected cells. Ten additional recombinant viruses recovered by Mr. Peter Norris during the pilot studies were also included in this study (24 viruses total). Initial attempts to assemble maps of the genomes of these viruses using "map-to-reference" methods met with limited success (Chapter 3). In this Chapter I show how one can use a *de novo* assembly strategy using Illumina and some Nanopore sequencing data to build maps of most of the remaining viruses. In some cases, both sequencing technologies were used as Illumina sequencing generates very accurate short reads, while Nanopore sequencing produces long but less accurate reads. These long reads can provide a scaffold or framework that addresses the problems encountered when trying to assemble genome maps composed of duplicated and/or inverted elements.

Once I had determined the structures of most of these new recombinant virus genomes, we realized that they exhibited some shared features that could be used to categorize the different types of recombination events. As a consequence, the data in Chapter 4 are organized in a way that reflects these different classes of events. As was also noted previously, NHEJ and MMEJ can be differentiated depending upon whether or not short patches of homology (2-20 bp) are found between the recombining DNA molecules (82). In this chapter, I also examine the VAC and transfected DNA sequences that abut the recombination junctions in a search for evidence of these homologies. Collectively, by characterizing and categorizing the different classes of recombinant viruses, and documenting the junction sequences, these studies provide insights into the molecular reactions that drive these gene capture events.

4.2 Results and discussion

4.2.1 Characterization of the recombinant viruses

We recovered 24 recombinant viruses that had captured a YFP/gpt selectable marker through NHR. Five of these recombinants were produced using cDNA·RNA hybrid molecules and the remainder with duplex DNA substrates. The sequences of 21 of these viruses (including those that were assembled using a map-to-reference method) were successfully assembled, *de novo,* using Unicycler (148). This is a hybrid assembly pipeline that can be used to assemble small bacterial and viral genomes using Illumina sequence data, or a combination of Illumina and Nanopore sequencing data. The contigs and the hypothesized paths between them were stitched together using Bandage (150), a program that helps to visualize the connectivity between *de novo* assemblies of contigs. This made it much easier to assemble recombinant virus genomes without needing to performed multiple rounds of reference mapping and visual inspection of long genomic sequences. Once assembled, I grouped the viruses according to the kind(s) of rearrangement that were observed.

As will be demonstrated, many of the recombinant viruses I recovered encoded complex rearrangements, and many were genetically unstable. One consequence is that it proved impossible to build genome maps, or even discern with certainty where the YFP/gpt marker had been inserted, in three of the recombinant genomes. Of the 21 remaining recombinant viruses, I classified each of these viruses as belonging to one of five rearrangement categories (Table 4.1). Fig 4.1 shows an example of each of the five different kinds of recombinant viruses. For these assemblies, and in my subsequent analyses, we combined all of the data acquired using duplex DNA or cDNA·RNA hybrid substrates to generate the

Table 4.1 Classification of NHR events (1)

Substrate	Deletion at the insert site	Mirror inserts in the ITRs	Deletion with translocation duplication	Duplication flanking the insert	Complementation dependent
DNA	2	1	5	0	8
cDNA·RNA	1	0	0	2	2
Total	3/21 (14%)	1/21 (5%)	5/21 (24%)	2/21 (10%)	10/21 (48%)

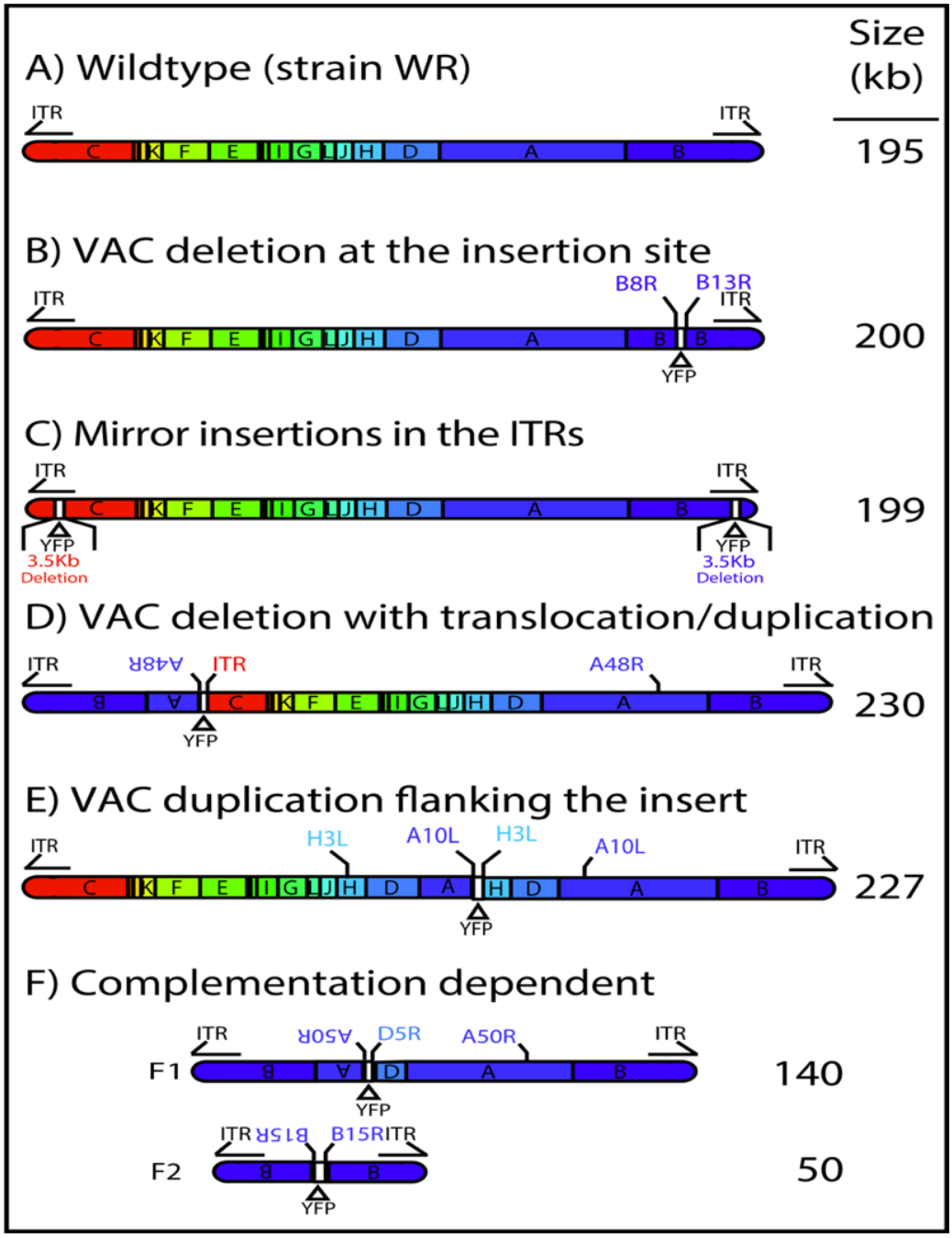

Size (kb)
A) Wildtype (strain WR)
ITR
ITR
C K F E I G L J H D A B
195
B) VAC deletion at the insertion site
B8R
B13R
ITR
ITR
C K F E I G L J H D A B B
YFP
200
C) Mirror insertions in the ITRs
ITR
ITR
C K F E I G L J H D A B
YFP
3.5Kb
Deletion
YFP
3.5Kb
Deletion
199
D) VAC deletion with translocation/duplication
ITR
A48R
ITR
A48R
ITR
B A C K F E I G L J H D A B
YFP
230
E) VAC duplication flanking the insert
ITR
H3L
A10L
H3L
A10L
ITR
C K F E I G L J H D A H D A B
YFP
227
F) Complementation dependent
F1
ITR
A50R
D5R
A50R
ITR
B A D A B
YFP
140
F2
ITR
B15R
B15R
ITR
B B
YFP
50

Fig. 4.1. Recombinant VAC genomes. (A) Schematic map of the 195 kb wildtype VAC WR genome. The lettering shows the larger *Hin*dIII fragments for reference purposes. The left and right ends are encoded by the C and B fragments and consist of two inverted terminal repeats (ITRs, arrowed). (B) Example showing a simple insertion of transfected sequences ("YFP") along with deletion of adjacent VAC sequences (G8, Table 4.2). (C) Mirror image insertions and deletions (indels) within the ITRs (5, Table 4.2). (D) Deletion of non-essential sequences and their replacement by a large, inverted duplication (G5, Table 4.2). Generally unstable. (E) Insertion associated with the duplication of flanking sequences, also unstable in the absence of selection (RN3, Table 4.2) . (F) Defective genomes, some quite small, propagation depends upon complementation by wildtype virus (R6 ("F1"), RN7 ("F2"), Table 4.2). Figure from G. Vallee et al. (1).

recombinants. The reason for doing this was that we recovered relatively few examples of each class of recombinant virus, and I did not observe any obvious differences between the kinds of viruses that were produced using duplex DNA *versus* cDNA·RNA hybrids. I have also included a table which describes all of the assembled recombinant viruses (Table 4.2).

Table 4.3 shows the depth of coverage of all the genomic assemblies. The depth of coverage was calculated differently for the simple recombinant viruses *versus* the group we classified as "complementation-dependent" viruses. This was because the ten preparations of viruses called complementation-dependent viruses each consisted of two types of viruses, a helper virus and a defective helper-dependent virus. In these cases, I calculated the coverage separately for both the mutant YFP-encoding virus and the helper VAC virus by calculating the coverage using contigs that were unique to each of the two viruses. Overall, I achieved fairly high coverage for all of the virus genomes with the lowest depth of coverage being 95 Illumina reads·bp^{-1}. Only 12 of the recombinant viruses needed to be additionally Nanopore sequenced, with the lowest Nanopore coverage being 11 reads·bp^{-1}. Although the read coverage from the Nanopore sequencing was low in some cases, this did not matter much as the Nanopore reads were always used to construct genome assemblies in conjunction with Illumina sequences. Overall, the average coverage of our recombinant viruses was good and well exceeded the typical recommendations for investigating indels (164), genotype calling (165), or copy number variations (166, 167). The generally accepted recommendations for read coverage for these applications ranges from 8-to-60 reads·bp^{-1}.

Long Nanopore sequencing reads proved very helpful as these sequences helped join together the Illumina-based contigs that otherwise appeared to be disconnected. Fig 4.2 shows an example where we revisited and reviewed the region around the YFP-insertion site and recombination junctions in a virus that was described in the previous Chapter. The RN4 virus encoded a YFP-bearing insert flanked by two unstable 1.9 kb duplications. This virus originally presented a bit of a challenge to assemble, but structure around the junctions could be easily established by mapping the long Nanopore reads across the long duplications (Fig

Table 4.2: Recombinant viruses recovered in this study

Transfected substrate	Virus isolate ("ID")	Left junction		Right junction		Approx. genome size (bp)	Rearrangement	Classification
		Gene or HindIII fragment	Recombination site (nt)	Gene or HindIII fragment	Recombination site (nt)			
NH cDNA·RNA	RN2**	C1L	21,553	A55R	160,511	60,733	138,958 bp VAC deletion	Complementation-dependent
	RN3**	A10L	119,318	H3L	88,918	226,658	30,400 bp VAC duplication	Large flanking sequence duplication
	RN4**	I4L	62,213	I1L	60,337	198,405	1,876 bp VAC duplication	Flanking duplication
	RN5**	Fragment C	6,313	C1L	21,275	181,501	14,962 bp VAC deletion	WR deletion
	RN13**	I7L	66,184	H4L	90,627	171,778	24,443 bp deletion	Complementation-dependent
NH DNA	RN7**	B15R*	174,660	B15R	174,660	49,754	174,659 bp deletion with a 20,095 bp duplication/ translocation	Complementation-dependent
	RN11**	B19R*	179,726	K4L	28,870	186,414	28,869 bp deletion with a 15,028 bp duplication/ translocation	Deletion with translocation/duplication
	RN15**	D10L*	105,383	Fragment C*	11,809	274,536	11808 bp deletion with a 108646 bp duplication/ translocation	Deletion with translocation/duplication
	RN16**	Fragment B*	180,447	A44R	149,149	62,350	149,146 bp deletion with a 14,307 bp duplication/ translocation	Complementation-dependent
	RN8**	A44L*	98,511	A44L	98,511	95,030	152,413 bp deletion with a 42,341 bp duplication/ translocation	Complementation-dependent
	1.1.1.1	C2L	19,608	Fragment A	146,690	66,058	127,081 bp deletion	Complementation-dependent

(Continued next page)

Transfected substrate	Virus isolate ("ID")	Left junction		Right junction		Approx. genome size (bp)	Rearrangement	Classification
		Gene or HindIII fragment	Recombination site (nt)	Gene or HindIII fragment	Recombination site (nt)			
NH DNA	4 (or NH4)	A4L	113,385	Fragment C*	9,161	126,682	81,369 bp deletion with a 9,161 bp duplication/ translocation	Complementation-dependent
	5 (or NH5)	LITR/RITR 70 bp repeat	Exact site undetermined	LITR/RITR	6,258/188,494 (mirrored)	~198,500	Two ~3,562 bp deletions	Insertion into the ITRs
	7 (or NH7)	E9L	54,986	J3R	81,357	170,841	26,370 bp deletion	Complementation-dependent
	R3	A34R	144,070	F3L	39,619	210.201	39,619 bp deletion with a 50,678 bp duplication/ translocation	Deletion with translocation/duplication
	R5	Fragment B	172,087	B12R	173,244	198,618	1,157 bp deletion	WR deletion
	R6	A50R*	156,490	D5R	98,511	139,549	98,510 bp deletion with a 38,264 bp duplication/ translocation	Complementation-dependent
	R8	F3L	32,717	E9L	54,723	178,237	22,005 bp deletion	Complementation-dependent
	G5	A48R*	155,158	Fragment C*	12,558	229,753	12,557 bp deletion with a 39,596 bp duplication/ translocation	Deletion with translocation/duplication
	G8	B8R	171,126	B13R	174,210	199,533	3,084 bp deletion	WR deletion
	s1-1	B8R*	171,272	K2L	26,653	197,074	26,652 bp deletion with a 23,482 bp duplication/ translocation	Deletion with translocation/duplication
Homol. DNA	RN6**	N2L	22,376	N2L	22,898	195,617	433 bp deletion	WR deletion
N/A	VAC WR	N/A		N/A		194,754	N/A	Wild-type VAC

*VAC sequence that has been duplicated/translocated from the opposite end of the viral genome

** Strains that were created by Peter Norris

Table 4.3 Depth of coverage of the recombinant viruses

Coverage (reads·bp^{-1})	Recombinant Viruses		Complementation-dependent viruses			
			Complementation-dependent viruses		Complementing viruses (wt VAC)	
	Illumina	Nanopore*	Illumina	Nanopore**	Illumina	Nanopore**
Average	1600	192	485	93	2035	316
Minimum	670	55	95	11	944	75
Maximum	2645	363	901	382	3249	1380

*Only 5 of the recombinant viruses were Nanopore sequenced

**Only 7 of the complementation-dependent viruses were Nanopore sequenced

4.2). These long reads generally proved very useful in helping to establish the structures of other unstable genome assemblies. More specifically, the translucent segments seen in Fig. 4.2 that mapped to sequences contributed by the Nanopore method, corresponded to places where a subset of the sequences reads had come from viruses that had lost the YFP/gpt marker through homologous recombination between the duplicated VAC sequences. It is not so easy to detect such unstable regions using Illumina read alignments (Fig 4.2).

4.2.2 Assembly of genome maps

These NHR reactions generated several different kinds of recombination products (Fig 4.1) with many viruses exhibiting large-scale rearrangements and, in some cases, specimens composing mixtures of viruses. Because our computer-generated assemblies pointed to the existence of so many complex genome rearrangements, we wanted to test the accuracy of these assemblies using PCR, or PFGE plus Southern blotting. In this section I will provide details that illustrate the correctness of the recombinant virus genome assemblies shown in Figure 4.1.

Of the 21 recombinant viruses that we assembled, three of them acquired deletions of VAC sequence at the YFP/gpt insertion site (Table 4.2). These deletions ranged from 1.2 kb to 15 kb in size. Two of the viruses encode deletions near the terminal ends of the virus genome, but outside the ITR, while the third encoded a deletion of non-essential sequences. Clear deletions of viral sequences

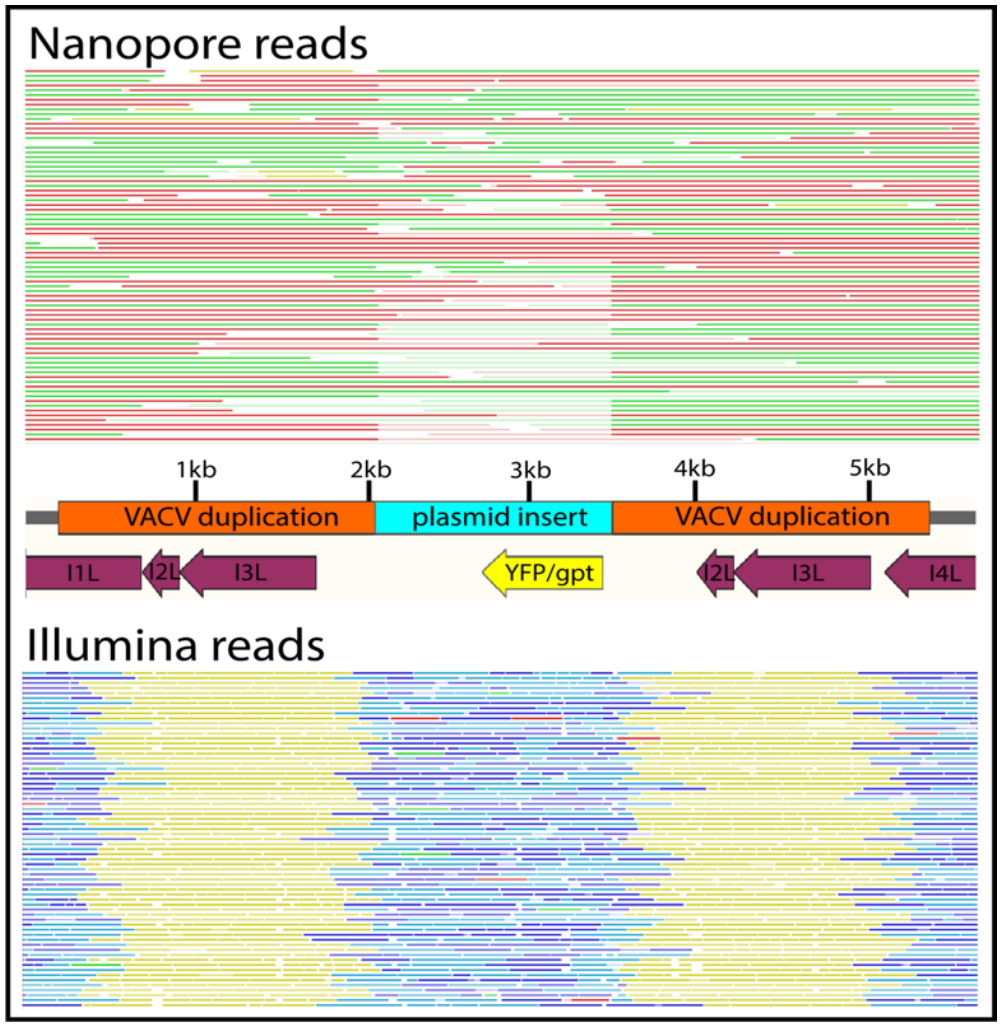


Fig 4.2 Genome assembly using Nanopore and Illumina sequencing technologies. The example shows a large duplication that was created through the insertion of non-homologous sequences into the I1L-I4L locus. These viruses are unstable, even under selection, and so the long forward (red) and reverse (green) Nanopore reads reflect the presence of a mix of mutant (solid colour) and revertant (translucent) genomes. The assembly algorithm mapped most of the shorter Illumina reads to unique places in the sequence assembly (blue) but was forced to arbitrarily allocated the duplicated elements (yellow) to map positions on either side of the unique insertion sequence. Figure from G. Vallee et al. (1).

could be observed when viewing the Illumina read alignments to the parental VAC genomic sequence (Fig. 4.3). The example virus that I have illustrated bore a 3 kb deletion of sequence that extended from the B8R to B13R genes (Fig 4.1B, Fig 4.3B and Fig 4.4B). This virus captured a 7.8 kb DNA element that encoded the YFP/gpt marker. The integrated substrate was predicted to encode a partial duplication of the marker and two *Hin*dIII sites (Fig 4.4B). This recombination event disrupts the *Hin*dIII B fragment and is visible when comparing the restriction fragment pattern of the recombinant virus with wild-type VAC. The combination of the deletion plus inserted marker sequence, within what would have been the 31 kb B fragment, generates three smaller *Hin*dIII restriction fragments (22.1, 8.4 and 5.5 kb). The three fragments encoding the inserted marker sequence were detected by Southern blotting using a probe that spanned the entire vector (Fig 4.4B). When comparing the intensities of the blotted bands the smaller 5.5kb band is underrepresented relative to the two larger bands (Fig 4.4A, "o"). This underrepresentation is due to the instability of the partial tandem duplication of vector sequences, much like what we observed for a virus that encoded a VAC sequence duplication that flanked the marker (Fig 4.2). This conclusion was additionally supported by Unicycler analysis of the read depths of contigs containing YFP/gpt marker sequences (Fig 4.5). If every virus in the virus stock encoded a partially duplicated insert (Fig 4.5B), we would expect the unique portion of the sequence insert to have about half the read depth of the duplicated sequences within the insert. The Illumina read depth of the contig encoding the duplicated marker sequence (n=1,809) was 5 times greater than the contig encoding the unique marker sequence (n=367). This suggests that our plaque-purified stock contained some viruses that encoded the tandem duplication (Fig 4.5CI) and others that did not (Fig 4.5CII). The persistence of the recombinant virus encoding the unstable duplication of marker sequences is likely due to the location of the YFP/gpt gene within the captured sequence. The loss of the duplication creates a virus that encodes only a fragment of the YFP/gpt gene (Fig 4.4C, 4.5). When viruses are grown and purified under drug selection, only those viruses that have retained the selection marker can still grow in the presence of drug. If the drug

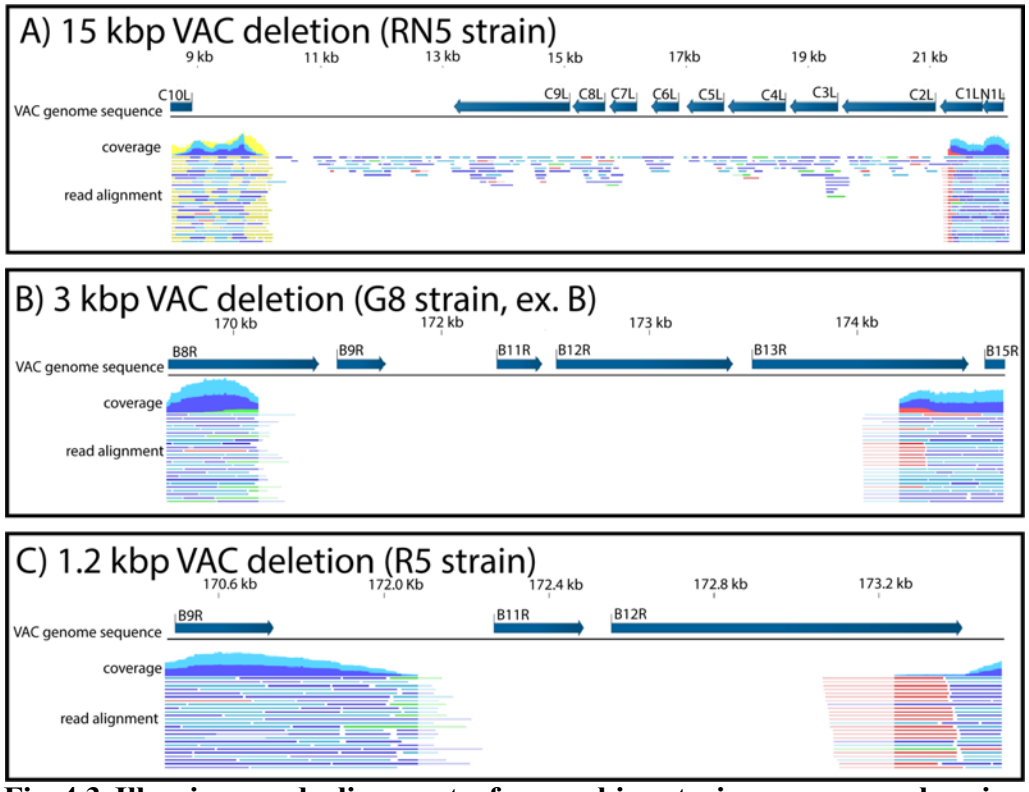
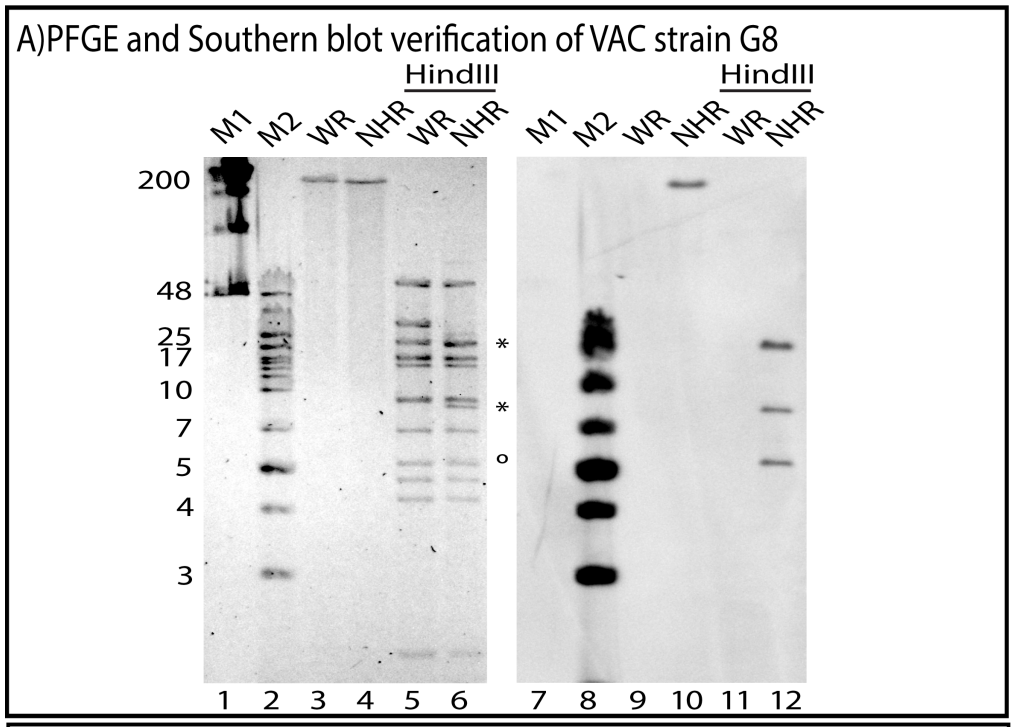
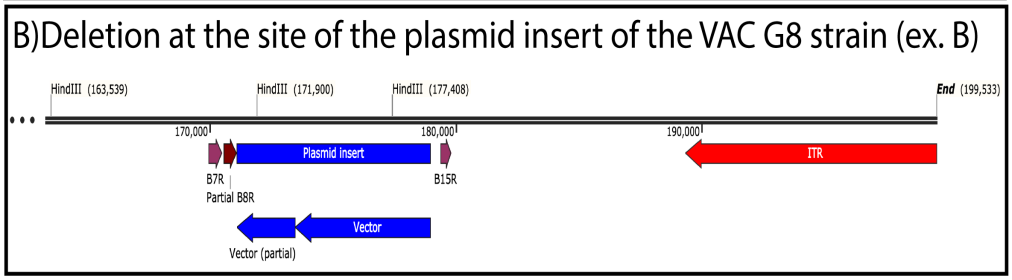
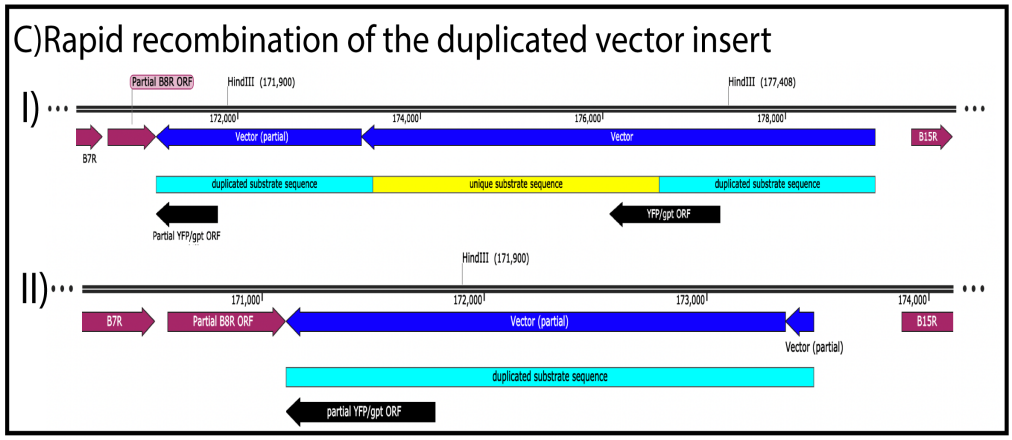


Fig 4.3 Illumina read alignment of recombinant virus sequence bearing VAC sequence deletions at the site of NHR. Recombinant VAC reads were mapped to wild-type VAC genomic DNA sequence using CLC genomics. A) recombinant virus RN5 bearing a 15kb deletion at the site of recombination B) Recombinant virus G8 (Fig 4.1B) bearing a 3 kb deletion at the site of recombination. C) Recombinant virus R5 bearing a 1.2 kb deletion at the site of recombination. Mapped read segments (solid colours) and unmapped read segments (translucent colours) are coloured based on the type of read that is mapped to the sequence. Mapped reads include the Paired reads (blues), Forward reads (green) and reverse reads (red).







(Continued next page)

Fig 4.4 Southern blot analysis of a recombinant virus genome with a deletion at the site of the insertion (Fig 4.1 B). The recombinant virus was plaque purified, grown up under drug selection, purified and the DNA extracted for sequencing, restriction mapping and Southern blotting A) *Hin*dIII PFGE restriction map (left) and Southern blot (right) of a recombinant virus genome (G8, Table 4.2) bearing a 3 kb deletion at the site of insertion (NHR). The parent VAC strain (WR) is included for comparison. A lambda phage ladder (M1) and a mix of GeneRuler standards (M2) were used as size markers. Southern blots were conducted using a probe spanning the entire plasmid vector/marker. B) Map of the recombination site illustrating the location near the right end, a 3kb deletion of VAC sequence, and the partial duplication of the marker-encoding insert. Only a subset of VAC genes are shown in the panels to orient the reader C) The two types of viruses present in these stocks. One of the recombinant viruses (CI) encodes an unstable duplication that is lost through homologous recombination (CII). This causes the loss of the 5.5 kb fragment ("o") on the SYBR-gold-stained gel (Panel A, left). Figure adapted from G. Vallee et al. (1).

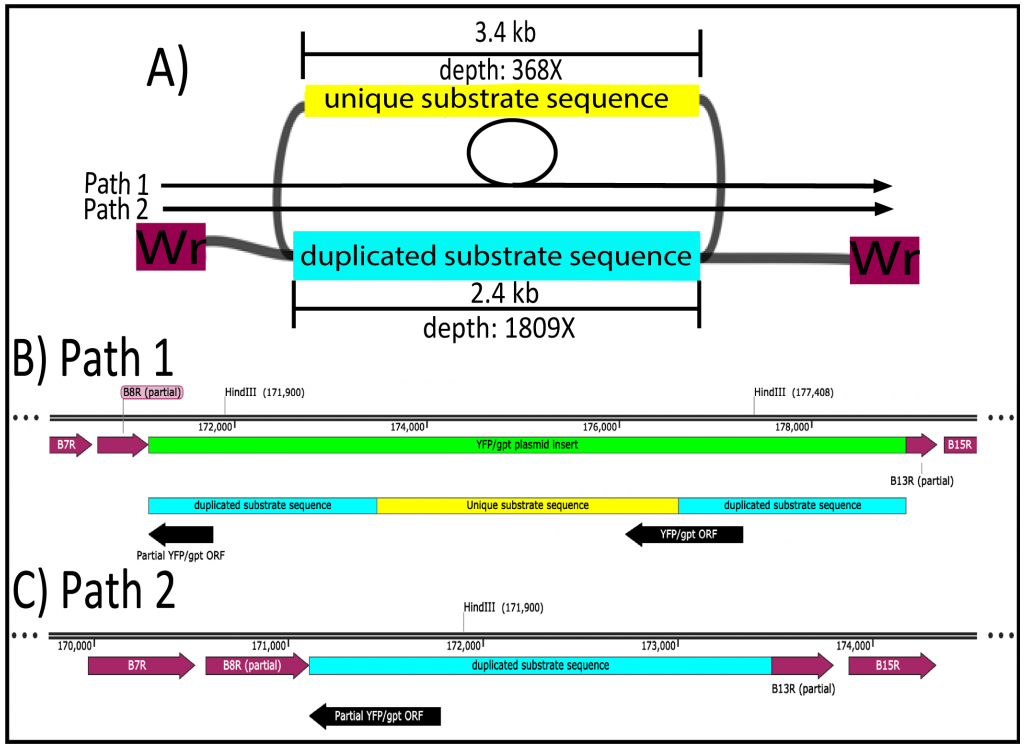


Fig. 4.5 Unicycler map showing how the sequencing data detects the presence of an unstable tandem duplication. (A) Unicycler analysis identified two different assembly paths (arrowed grey lines). Path #1 incorporated a unique sequence element (in yellow) while Path #2 did not (in blue). Moreover, reads derived from the unique sequences (in yellow) are 5-fold less abundant than reads derived from the surrounding sequences. These results can be explained if this genome encoded the YFP/gpt selectable marker embedded within duplicated sequence elements. Recombination between the duplicated DNAs in the virus described by Path #1 (B) deletes the intervening sequences and produces the virus described by Path #2 (C). Figure adapted from G. Vallee et al. (1).

selection is removed, the viruses encoding the duplicated insert would be rapidly lost.

A similar virus with deletions of VAC sequences was generated (Strain 5, Table 4.2) but differed in that the 3.5 kb deletion of VAC sequences and 5.5 kb inserts were mirrored in the two ITRs (Fig 4.1C). Such mirrored deletions are a phenomenon that was first described by G. McFadden and S. Dales (168). In this case the mirrored insertion of the marker likely required two steps. First, NHR inserted the YFP/gpt marker into one of the ITRs, perhaps the left ITR. Second, homologous recombination with the right ITR of another "sibling" virus generates the final virus (Fig 4.6). My PFGE restriction maps and Southern blot demonstrated the presence of the selectable substrates in both ITRs (Fig 4.7). The inserted sequence encodes a *Hin*dIII site which, when cut, generates two novel 6 kb restriction fragments encompassing the hairpin ends of the B and C *Hin*d*III* fragments. The exact size of this fragment could not be predicted from the sequencing data as the marker was embedded in the 70 bp telomeric repeats. The presence of this indel also had a noticeable effect on the size of the *Hin*dIII B and C fragments of the recombinant virus relative to the wild-type. These bands exhibited the altered sizes predicted from our genome assembly (26.5 and 17.5 kb, respectively). These maps were further supported by Southern blotting the restriction digest using a probe that spanned the entire selectable marker. Besides the 6 kb fragment the probe also detected the 17.5 kb and 26.5 kb fragments also expected to encode portions of the digested marker (Fig 4.7C).

The third class of viruses were those that had suffered deletions along with inversions, translocations, and duplication of VAC sequence (Fig 4.1D and Fig 4.8). These viruses were the second most common kind recovered and often encoded odd junctions within novel genomes that were often quite large. These viruses generally encoded deletions of up to 40 kb of the original sequence (Table 4.2), which were replaced by an inverted and duplicated copy of virus sequences located on the other end of the genome and next to the selectable marker. These inverted duplications varied from 15-to-109 kb in length and always extended from

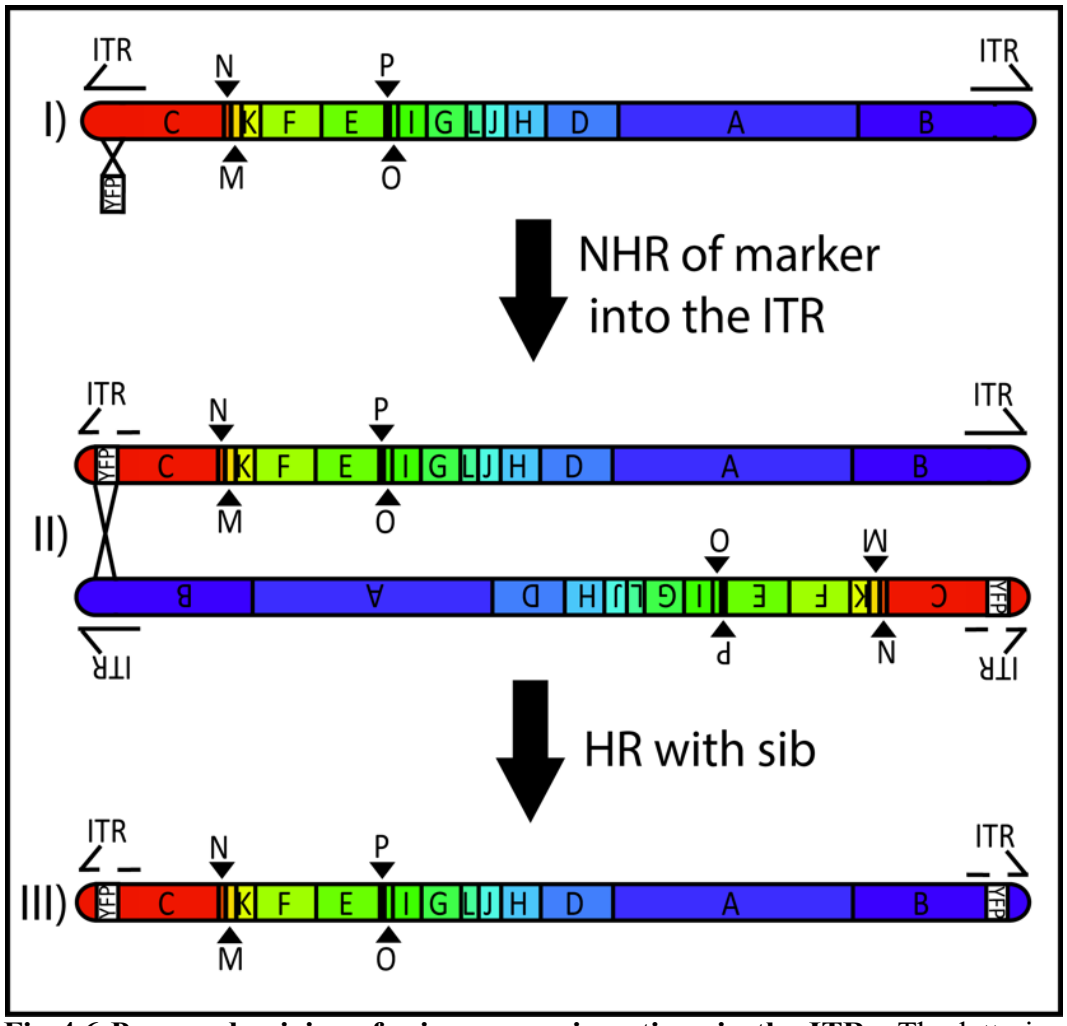


Fig 4.6 Proposed origins of mirror gene insertions in the ITRs. The lettering shows the larger *Hin*dIII fragments for reference purposes. The left and right ends encoded by the C and B fragments consist of two inverted terminal repeats (ITRs, arrowed). (I) The first NHR step where the YFP/gpt marker is inserted into one of the ITR sequences. (II) A second homologous recombination step, between sibling virus genomes, inserts the selectable marker into the second ITR. (III) The resulting recombinant virus encodes a duplicated and inverted (mirrored) insertion of the marker sequence in both ITRs. Which of the ITR (left or right) that were targeted by the initial NHR event cannot be determined.

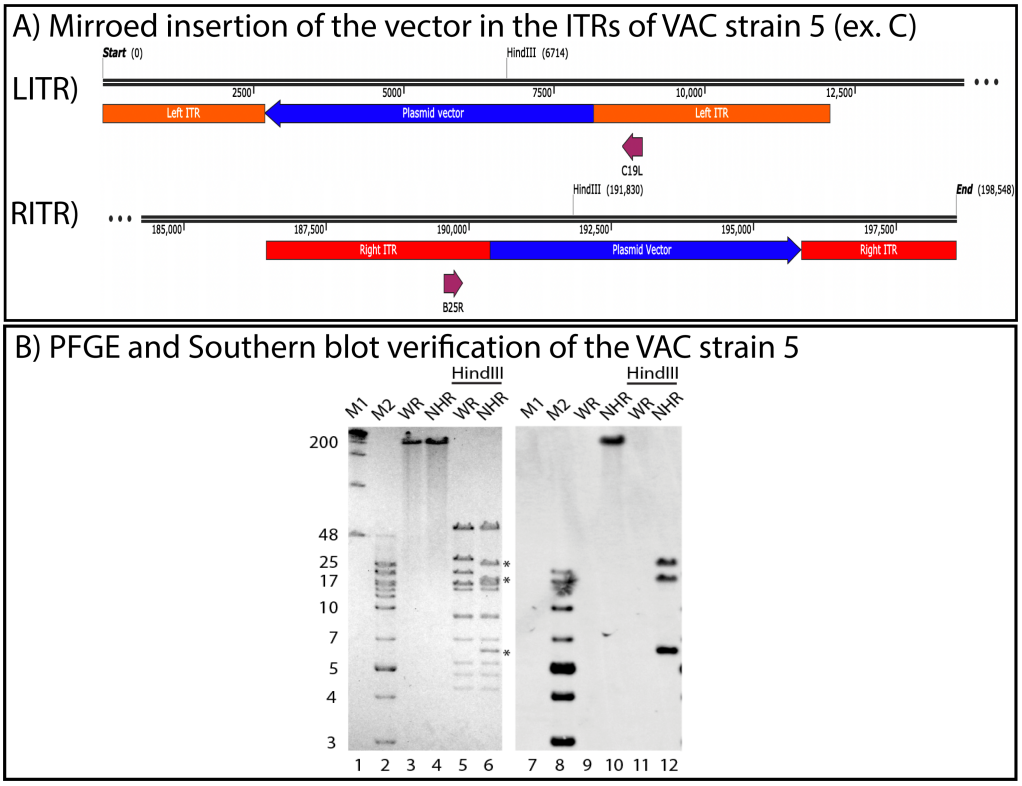
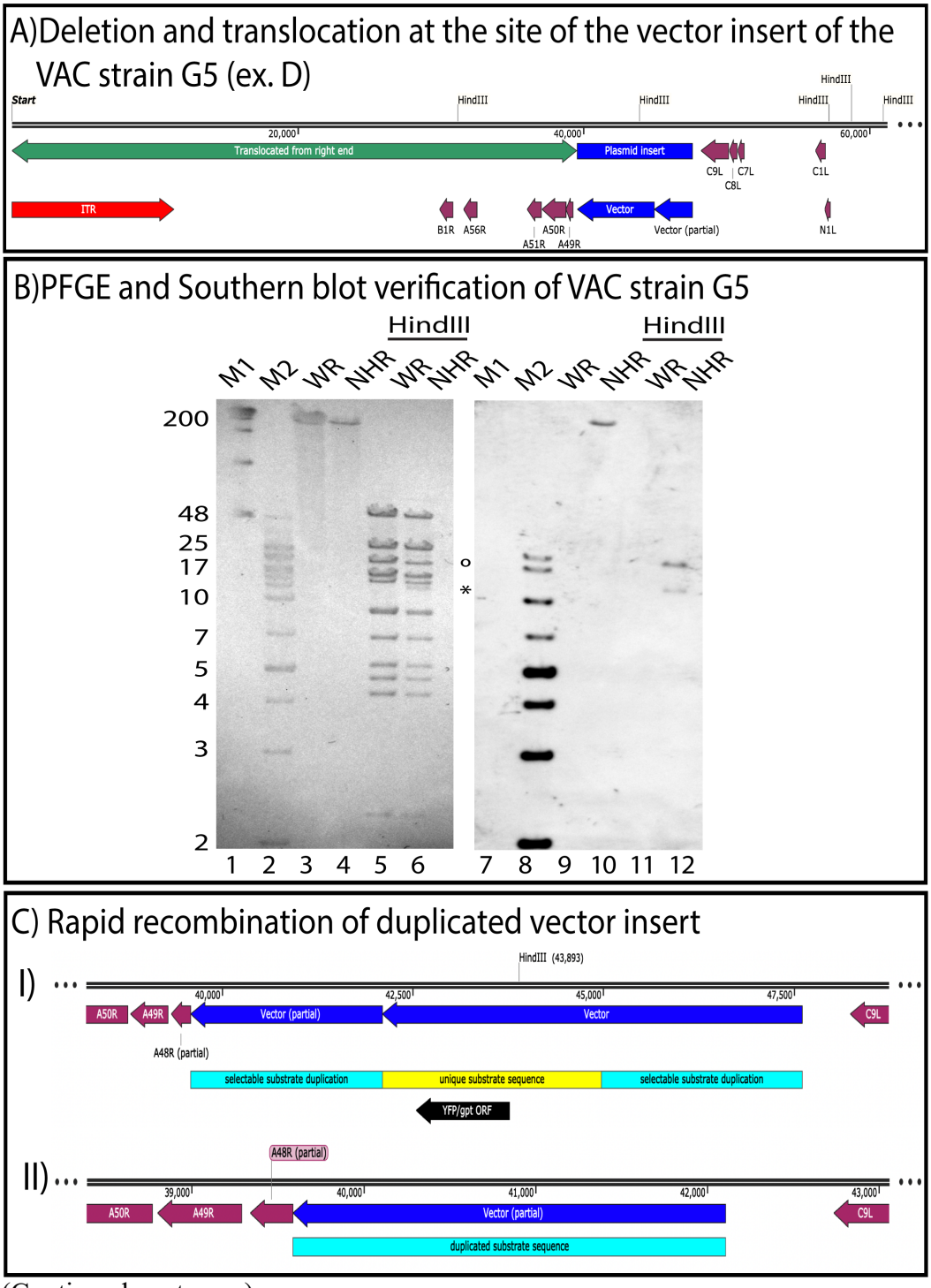


Fig 4.7 Southern blot analysis of a recombinant virus with mirror insertions in the ITRs (Fig 4.1C). Strain 5 (Table 4.2) was grown under drug selection, purified and virus DNA extracted for sequencing, restriction mapping and Southern blotting A) Maps of the left and right ITRs. The virus encodes a 3.5 kb deletion from each ITR next to the sites of insertion of the 5.5 kb YFP/gpt markers. Only a subset of VAC genes are shown to orient the reader. B) PFGE showing the *Hin*dIII restriction fragments (left) and Southern blot (right) of DNA extracted from a recombinant virus (5) and the parent VAC WR strain. A lambda phage ladder (M1) and a mix of GeneRuler standards (M2) were used as size markers. The Southern blots used a probe spanning the entire transfected plasmid substrate. Figure adapted from G. Vallee et al. (1).

a central part of the genome to an original ITR and telomeric end. The virus I have illustrated (G5, Table 4.2) bears a 12.6 kb deletion extending from the left ITR near the C9L gene. The selectable marker is inserted to the left of C9L. The deleted sequences located to the left of the marker were replaced by a 40kb inverted duplication of sequences normally found on the right end of the genome extending from A48R to the right ITR (Fig 4.8A). This generated a large 230 kb viral genome with some of the viruses encoding an unstable partial tandem duplication of the YFP/gpt sequence. Unicycler analysis detected only $1/6^{th}$ the number of reads derived from the predicted single-copy central portion of the insert compared to the duplicated portions on either side (n=299 and n=1710, respectively). This is much like what is shown in Fig 4.5 for virus isolate (G5, Table 4.2). This suggests that despite being plaque purified, this stock still contained a mix of viruses, some encoded the tandem duplication and some did not (Fig 4.8B). As with virus the virus shown in Fig 4.5, the persistence of the virus encoding the tandem duplication is due to the fact that recombination deletes the YFP/gpt gene that is under selection. This computational genome model was confirmed by PFGE and Southern blotting. The *Hin*dIII restriction map of the recombinant virus did not differ much from wild-type VAC. The only difference between the two is the appearance of a faint ~13 kb fragment (Fig 4.8C, "o"). I detected both the 13 kb fragments and a ~21 kb fragment by Southern blotting. The 21 kb fragment derives from a virus that has lost the duplication through homologous recombination, which also deletes a *Hin*dIII site located in the central portion of the partly-duplicated insert (Fig. 4.8CII). The ~13 kb fragment derives from the rarer portion of viruses that had still retained the partial duplication of the inserted sequences. These viruses encode a *Hind*III site (Fig 4.8CI) which is expected to produce two unresolvable 12.7 kb and 13.3 kb *Hin*dIII restriction fragments. The weaker intensity is consistent with the Unicycler analysis where I saw reduced depth of coverage of different portions of the inserted sequences.

The fourth kind of virus that was recovered from these experiments encoded selectable markers that were flanked by duplicated VAC sequence. Just two of these



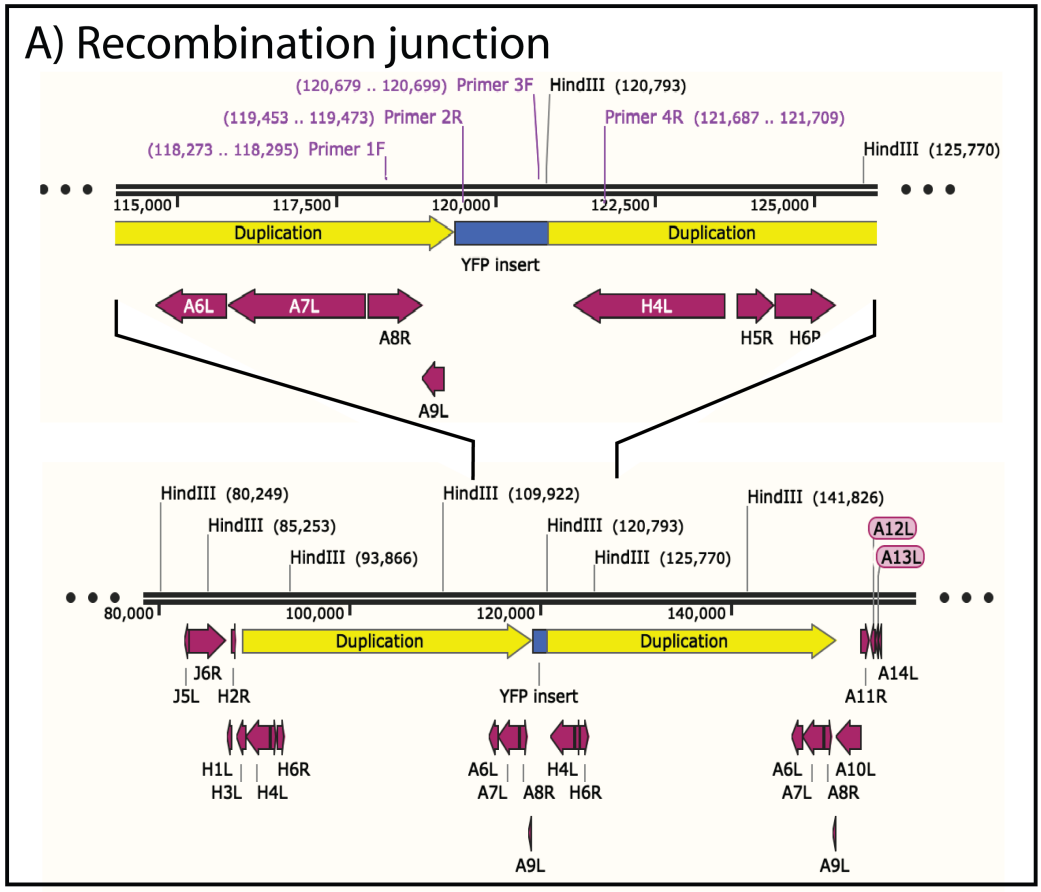
(Continued next page)

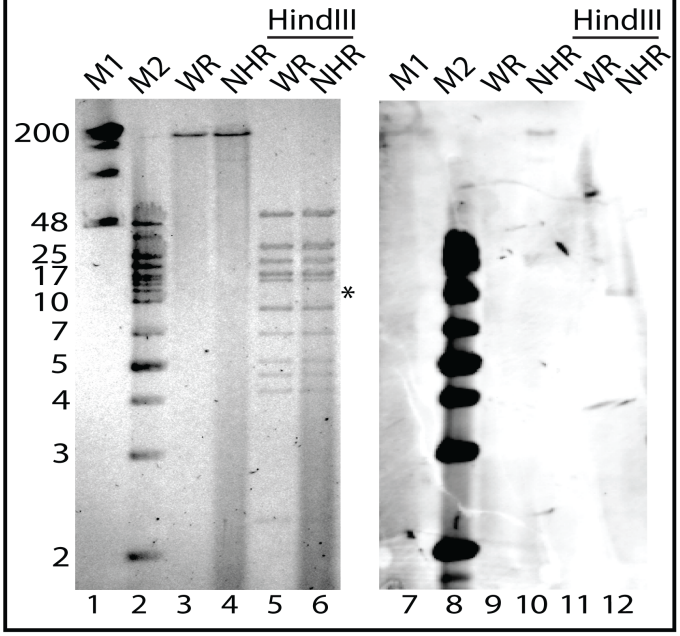
Fig 4.8 Southern blot of virus isolate (G5, Table 4.2) showing a deletion and inverted translocation/duplication at the site of recombination (Fig 4.1D). A) Map showing the left end of the genome. The virus lost 12.6 kb from the left end of the genome which was replaced with a 40 kb inverted duplication derived from the right end of the genome (Fig 4.1D). Only a subset of VAC genes are shown in the lower panels to orient the reader. B) PFGE showing the *Hin*dIII restriction fragments (left) and Southern blot (right) of DNA extracted from a recombinant virus (G5) and the parent VAC strain. A lambda phage ladder (M1) and a mix of GeneRuler standards (M2) were used as size markers. The Southern blots used a probe encompassing the entire plasmid vector/marker. C) Map showing the two recombinant viruses found in these virus stocks. The marker was partially duplicated (CI) and subject to homologous recombination (CII). Cutting the *Hin*dIII site seen in CI yields two ~13 kb pieces of DNA detectible in a Southern blot ("*"), its loss through recombination generates a 20 kb fragment ("o"). Figure adapted from G. Vallee et al. (1).

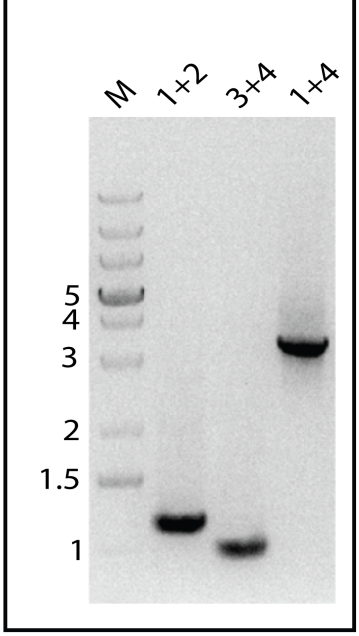
viruses were recovered. One was previously described in Chapter 3 (Fig 3.4 C, (RN4, Table 4.2)) and the second (RN3, Table 4.2), encoding a much larger VAC duplication, is shown in Fig 4.1E. The sequencing data suggested that the YFP/gpt marker was flanked by two 30.5 kb duplications encompassing a region from the H3L gene to the A10L gene (Fig 4.1E and Fig 4.9A). This large 30.5 kb virus was exceedingly unstable. The Illumina coverage of the contigs encompassing the inserted YFP/gpt element was ~40-fold lower than the average contig encoding non-duplicated VAC genomic sequence. We tried to confirm the correctness of the assembly by PFGE and Southern blotting. However, the *Hin*dIII digest resembled that wild-type VAC and our Southern blots generated only exceedingly faint 11 and 200 kb signals (Fig 4.7B). As an alternative approach I designed some PCR primers (shown in Fig 4.9A) and used to PCR amplify the different junction sequences expected to be found in this virus stock (Fig 4.9C). The region encompassing the entirety of the duplicated VAC sequence (62 kb) was too large to amplify, but these primers were able detect all of the other predicted sequence junctions (Fig. 4.9C).

4.2.3 Complementation-dependent viruses

These were the most abundant viruses that we recovered in these experiments and perhaps the most interesting. Despite three or more rounds of plaque purification under agar and drug selection, we were never able to separate the YFP/gpt-bearing recombinants from what looked like wild-type parental viruses. Upon close inspection of the sequencing data and the viral stocks, we determined that these stocks each contained at least two genomes: a wild-type VAC and a smaller complementation-dependent virus. These smaller viruses all encoded a YFP/gpt marker as well as large deletions of viral sequences from regions of the genome encoding essential genes (30). The deletions ranged from 22-to-175 kb and 5 of the 10 viruses also encoded inverted duplications of the VAC sequences (Table 4.2). Despite the large deletions, these complementation-dependent viruses all encoded complete ITR sequences, including the 200 bp terminal elements that support mini-chromosome replication in VAC-infected cells (169). Fig 4.1F illustrates one of the larger and the smallest of these viruses. The smallest complementation-dependent virus (RN7) that I recovered was 49.7 kb in size and







(Continued next page)

Fig 4.9 Southern blot of virus isolate (RN3, Table 4.2) showing a large 30.5 kb VAC duplication flanking the inserted marker (Fig 4.1E). A) Predicted genomic sequence assembly with 30.5 kb VAC duplication flanking the g/gpt encoding insert. Only a subset of VAC genes are shown to orient the reader. B) PFGE Restriction analysis (left) and Southern blotting (right) of the purified viral genomic DNA. The Southern blot used a biotinylated probe that spanned the entire marker sequence. A lambda phage ladder (M1) and a mix of GeneRuler standards (M2) were used as size markers. These viruses are unstable and duplicated sequences are rapidly lost from the virus stock through homologous recombination. C) PCR detection of the recombination junctions. Three PCR reactions were used to detect the two unique junctions found. flanking the single-copy insert (primers 1+2 and 3+4) and the 3.4 kb product encompassing both junctions (Primer 1+4). The primer locations are shown in panel A. Figure adapted from G. Vallee et al. (1).

encoded two incomplete, duplications of the marker sequence that are inverted relative to one another (Fig 4.10C). These are flanked by inverted and duplicated VAC sequences extending from the B18R gene to the end of the original right ITR. The DNA extracted from this virus stock contained several virus genomes that were visible by PFGE (Fig 4.10A, RN7 "U"). This stock contained a full-length VAC genome plus two smaller genomes approximately 45 and 50 kb in length. It was not entirely clear how one might rearrange one to form the other from the genome maps. However, upon further examination of the assemblies, we discovered four 55 bp repeats orientated in a way that would permit deleting a 5.9 kb segment from the 49.7 kb genome. This deletion would remove one of the YFP/gpt encoding loci through homologous recombination (Fig 4.10C "55 bp repeat", green arrows). The two 55 bp repeats inserted between the marker and VAC sequence appear to have been copied at these locations from its original locus in the ß-lactamase gene, although what process would do that is unclear. The second virus stock that I investigated appeared to contain a 139.5 kb subgenome (Fig 4.10, R6) with the YFP/gpt element inserted between the D5R gene and the A50R gene. This causes a deletion of 38.3 kb of essential VAC sequences (Fig 4.10D). When these viruses were Southern blotted, the YFP/gpt probe hybridized to only the smaller virus genomes that were predicted to encode the YFP/gpt inserts (Fig 4.10B). To our knowledge, this is the first description of something resembling "defective" particles in stocks of VAC. Nor are any naturally-occurring poxviruses known to employ a 50 kb genome. Our finding raises some interesting questions in terms of the encapsidation and transmission of these genomes: Are these severely truncated genomes being encapsidated alone, or are multiple copies packaged into a single capsid, or with the wild-type complementing VAC genome?

4.2.4 Junction sequences analyses

For the viruses that were recovered and successfully assembled in this study, gene capture was associated with the acquisition of some or all of the transfected nucleic acid marker sequence. To determine if this process was dependent in any way on homology, we compared the sequences of the transfected plasmid DNA and the parental virus in a region surrounded the recombination

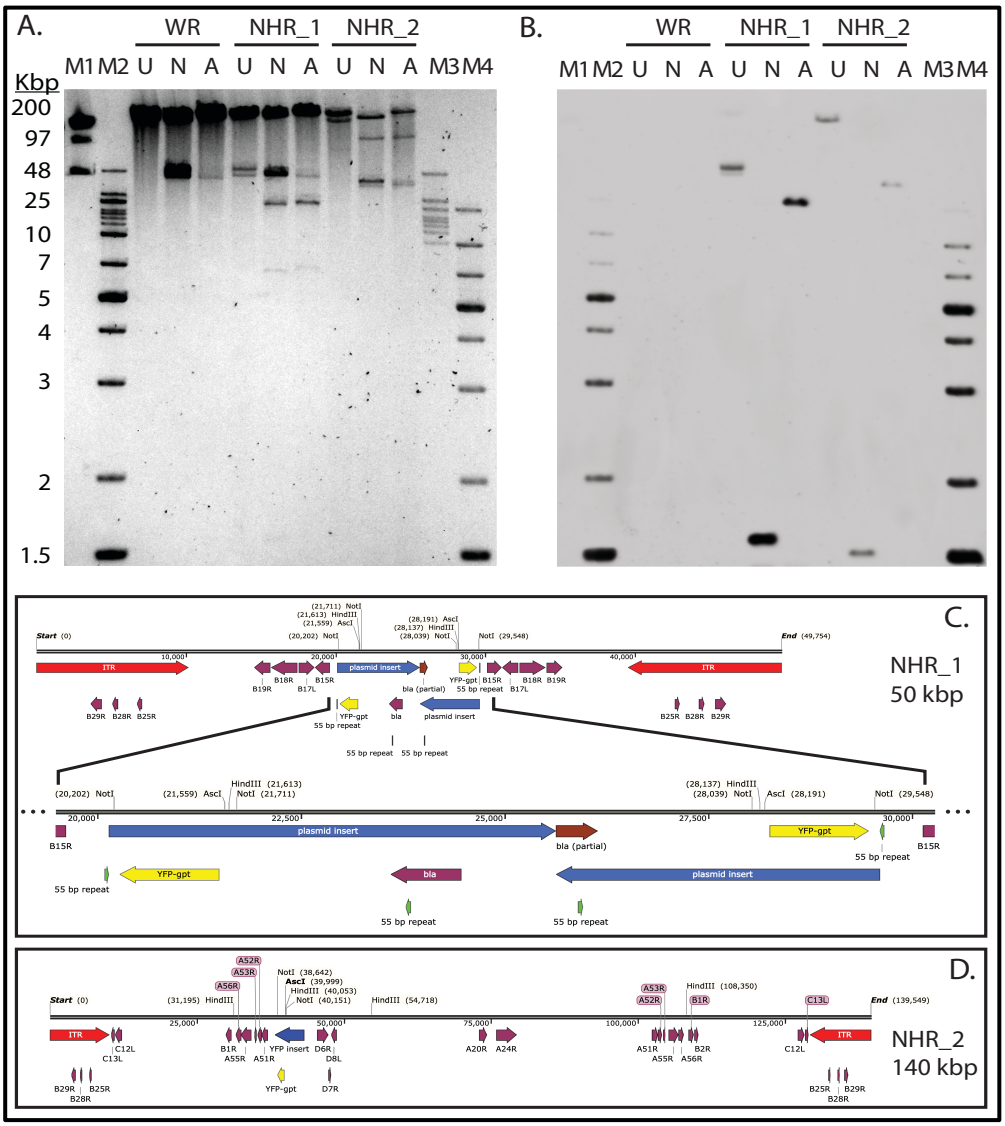


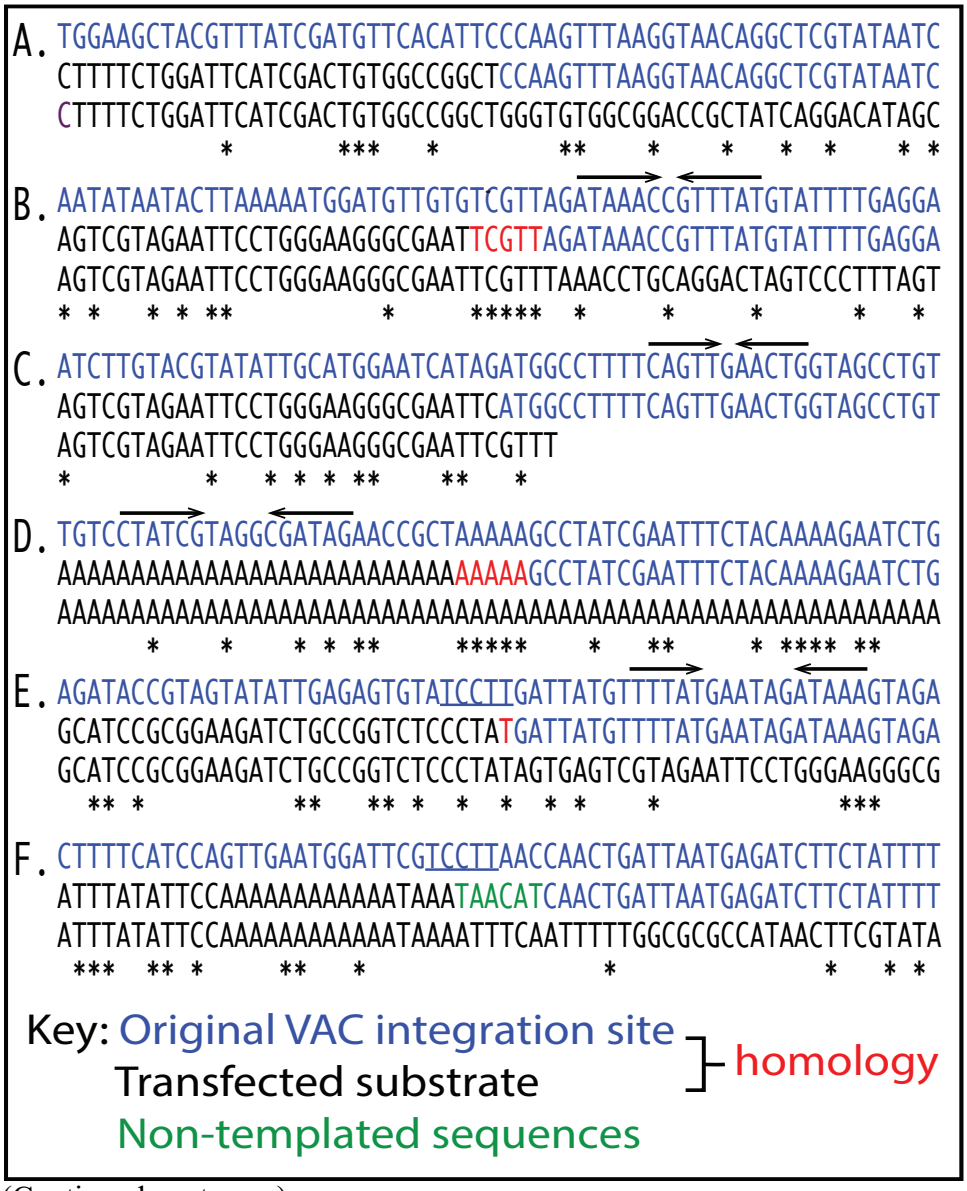
Fig 4.10 Characterization of the VAC R6 and RN7 complementation-dependent strains (Fig 4.1F). (A.) PFGE showing the parental (WR) and two recombinant (R6, RN7) viruses. Stained with SYBR-gold. Recombinant viruses were left uncut (U) or digested with *Not*I (N) or *Asc*I (A). (B.) Southern blot of the gel in A. A PCR amplified biotinylated probe encoding the YFP/gpt locus was used to probe the membrane. (C.) Map of the RN7. The orientation and location of the 55 bp repeats are illustrated in the expanded map of the virus. (D.) Map of the R6. A lambda phage PFGE ladder (M1) and a mix of GeneRuler Standards (M2) were used as size markers. M3 and M4 are High Range and 1 kb Plus GeneRuler markers, respectively. These were used to produce the M2 ladder. Figure from G. Vallee et al. (1).

junctions. Representative sequence alignments are shown in Fig 4.11 with all sequence alignments present in Table 1 of the annexe. For these alignments, we aligned ~30nt on either side of the original VAC and plasmid sequences that were targeted for recombination. Looking at the 42 recombination junctions (from the 21 recombinant viruses), most of these joints had little to no sequence similarity at the joint. Sixteen of 42 joints shared no homology (Fig 4.11A) and another 16/42 shared only one base at the joint (Fig 4.11E). The 10 remaining joints shared between 2-5 bases of homology at the joint (Fig 4.11B and E). For two junctions, we also observed the presence of non-templated sequences at the joint. The lack of homology directly overlapping the joints and the presence of non-templated nucleotides in rare cases are characteristic features of mammalian NHEJ (82, 136). When examining all of the sequences surrounding these recombination events, we did not observe any disproportionate excess of aligned bases than we'd expect to see at random. The ratio of aligned bases to the number of expected base matches to a randomly distributed nucleotide sequence was 0.89 ± 0.30. Few gene capture events involved the original ends of the transfected substrates (Fig 4.11C), most mapped at other sites in the plasmid substrate. For the viruses that were produced using transfected cDNA·RNA substrates, we were concerned that there might have been some residual contaminating dsDNA in our hybrid substrate. However, in two cases, I did observe remnants of a poly-A tail at the junction (Fig 4.11D). I additionally did not observe the presence of any sequences that are located upstream of the T7 promoter in the recombinant viruses that captured cDNA·RNA molecules. Only sequences downstream of the T7 promoter should be present as a T7 transcription kit was used to produce the RNA that was used make the cDNA·RNA duplexes (Fig 2.1B). In two junctions, we also detected examples of the VAC topoisomerase I pentameric recognition and cleavage motif [5'-(C/T)CCTT-3'] (156) that directly overlapped the recombination junction (Fig 4.11D and F, underlined). What is most interesting about these motifs is that VACs topoisomerase has been shown to catalyze NHR when expressing in *E. coli* (170) and can catalyze single-strand breaks and rejoining *in vitro* (171). Lastly, we searched the alignments for short (5-6nt) and closely spaced (≤10nt) inverted

repeats (Fig 4.11, underlined) as these have been associated with VAC gene duplication (172). We did not detect any more of these sequences than are generally distributed in the VAC genome (~70% of 60 nt sequences are expected to encode at least one repeat). Overall, our review of the sequence features associated with the recombination junctions suggests that these NHR events do not require pre-existing homology. This can be best described as a form of NHEJ.

4.3 Conclusion

In these experiments, I successfully assembled the genome of 21 recombinant viruses using Unicycler (Table 4.2). This includes the strains of virus that were also assembled using the map-to-reference method in Chapter 3 (Fig 3.4). I characterized the recombinant viruses into 5 categories (Table 4.1) for which I highlighted examples of each in Fig 4.1. The genomic assemblies of many of the viruses that were recovered suggested the presence of unstable duplication and most of the viruses we recovered appeared to be complementation-dependent due to the presence of large deletions of essential virus sequence (22-to-175 kb). For this reason, I supported the genomic assemblies from Fig 4.1 using PFGE, Southern blotting and in one case, PCR (Fig 4.4 and Fig 4.7-4.10). Lastly, I investigated all the recombination junction of the assembled recombinant viruses (Fig 4.11 and Annexe, Table 1). This investigation yielded that the joining of the substrate and VAC sequences were homology independent. These joining reactions could be best described as a form of NHEJ that may be supplemented by a topoisomerase-mediated aberrant strand end-joining reaction.



(Continued next page)

Fig 4.11 Representative sequence alignments at the recombination junctions. Parental VAC sequences are shown in blue (top line) and the transfected, non-homologous marker sequence is shown in black (bottom line). The origins of any ambiguous bases at the joints are in red while non-templated bases are shown in green. Bases that align between the parental VAC and transfected marker sequence were denoted with asterisks. Most junctions had little (E) to no homology (A and C) at the junction site. Few of the sequence alignments had up to 5 aligned nucleotides at the recombination junction (B and D). The junction aligned to various sites of the marker sequence with a mapping near the end of the substrate (C). Other junctions mapped next to the VAC type I topoisomerase recognition site 5'-YCCTT-3' (E and F) In a couple of recombinant viruses that recombined with the cDNA·RNA hybrid substrates, recombination occurred at poly-dT-primed, poly rA-tail (D). Two recombination junctions from recovered viruses encoded non-templated nucleotides. In the VAC sequence, closely spaced inverted repeats were also detected near the junctions (arrows) in many of the alignment. Figure from G. Vallee et al. (1).

Chapter 5 – Conclusions and future directions

5.1 Investigating VAC's ability to catalyze non-homologous recombination

Many genes encoded by poxviruses are orthologs of cellular genes. How poxviruses have acquired these novel genes over time is not certain, what is for certain is that such genes were captured by some form of non-homologous recombination in the broadest sense of the term. In this study, I aimed to determine if VAC could mediate the capture of a novel gene through NHR. I wanted to test the intronless substrates that these viruses may encounter in the cytoplasm and the frequency at which we can recover these recombinant viruses. I also aimed to determine what these recombinant viruses look like and examine the recombination junctions to determine if these recombination events are dependent on microhomologies.

5.1.1 Substrates targeted for VAC recombination

Recombinant VAC was recovered from cells transfected with both homologous and non-homologous duplex DNA and cDNA·RNA substrates. Overall, the transfected DNA substrates were more efficient than were cDNA·RNA hybrid molecules and the frequency at which recombinant viruses were recovered using non-homologous substrates was strikingly lower (≤400-fold) than the recombinant viruses produced using homologous substrates (Table 3.1). The fact that cDNA·RNA hybrid molecules are less suitable substrates for these reactions may be attributed to a requirement for converting the cDNA·RNA heteroduplex to a DNA duplex before the recombination event. However, we can't fully rule out the targeting of the RNA strand for recombination. VAC ligase can ligate RNA-containing duplexes to DNA with low efficiency (173). Despite the lower recombination efficiency of the cDNA·RNA substrates, these would better mimic the reverse transcribed products that a poxvirus might encounter in the cytoplasm. The frequency of recombinant viruses recovered using non-homologous transfected substrates (as few as ~3 recombinants for every 10^8 PFU), may still be enough to be biologically relevant in a natural infection. The viremia in some poxvirus infections has been reported to be as high as 10^{10} genomes·ml^{-1} (163).

In the context of the recombination frequencies I've measured, the ratios (Table 3.1) between the different substrates are likely more relevant than the absolute recombination frequencies themselves. Our experiments used a "stacked deck" approach by transfecting cells with an abundant quantity of the marker sequence that is transcriptionally regulated by a poxvirus promoter. In a natural setting, the presence of a reverse transcribed gene fragment, or the presence of a dsDNA gene fragments in the cytosol would be quite rare. Such fragments would also need to be recombined downstream of a poxvirus promoter to be expressed by a virus that has captured the novel gene fragment. These are only a few of the factors that would make these events exceedingly rare in nature. Despite these shortcomings, these results give us an idea of the substrates and the relative efficiency with which they can be recombined into the viral genome.

5.1.2 The role of homology and sequence motifs in VAC NHR

A close examination of all the recombination junctions (Fig 4.11), yielded several interesting conclusions. The joining reactions could be best described as reflecting a form of NHEJ as the junctions resemble those of eukaryotic NHEJ reactions (82, 136) and not MMEJ (82). We also observed the presence of the VAC topoisomerase pentameric motifs (156) overlapping a few of the recombination junction (Fig 4.11E and F, underlined), which suggests that this viral protein may contribute some recombinant products through an aberrant strand end-joining reaction (170). No short duplications characteristic of LINE-1 mediated horizontal gene transfers were observed at the junctions (17, 18), although our experiments were not designed to detect these kinds of events. We also did not observe any overrepresentation of short inverted repeats near the junction site which have been associated with VAC gene duplications (172) (Fig 4.11, arrows).

Although topoisomerase is likely contributing to some of the illegitimate recombination events, other enzymes are catalyzing most of these events. The viral proteins that could be contributing to these processes include the VAC E9 polymerase or A50R ATP-dependent DNA ligase. Cellular DNA repair enzymes may also be involved in such processes. DNA ligase IV is a good candidate as it can catalyze mutagenic repairs of DSB in the VAC genome (142). This enzyme is

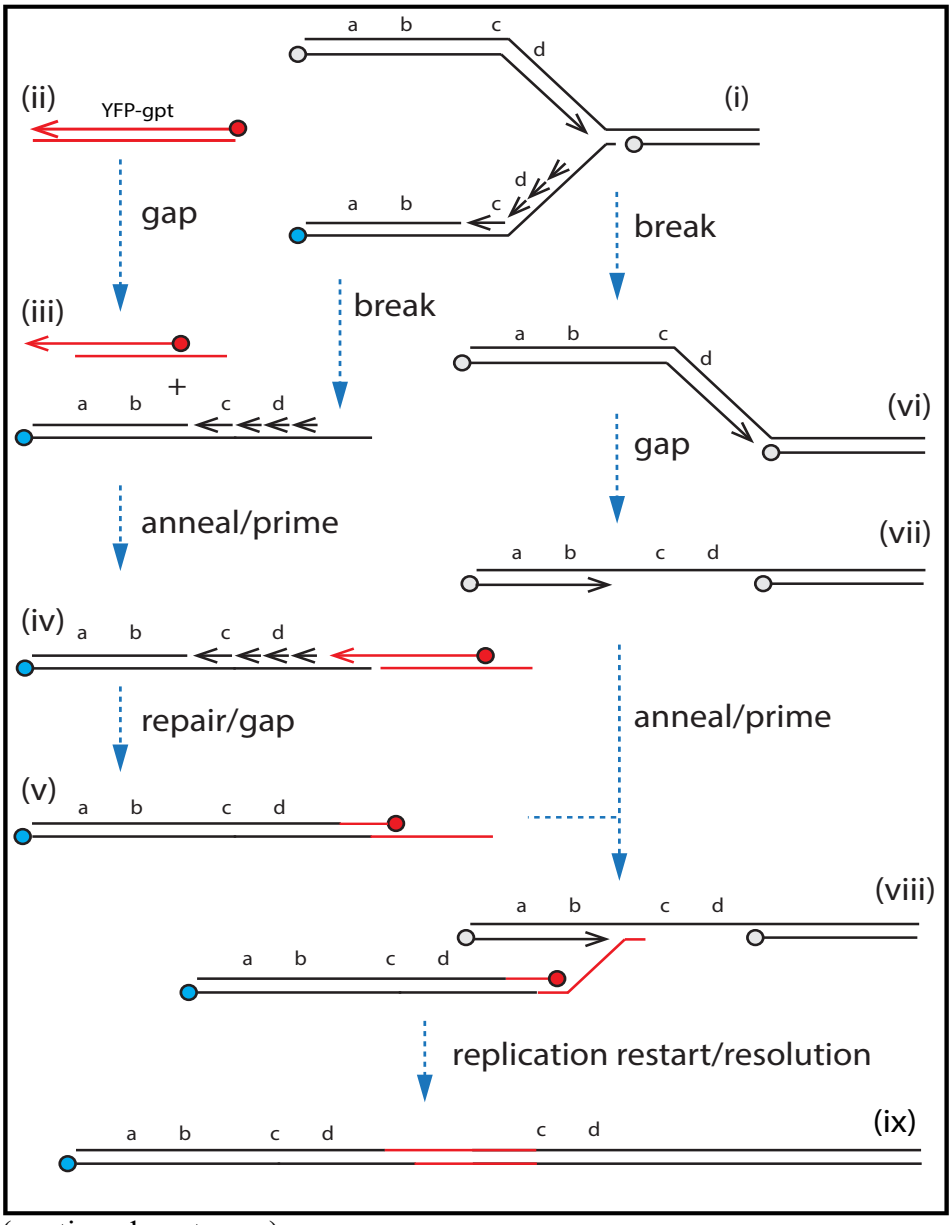
also well-known to catalyze NHEJ reactions (174) and end-processing choices (175) in eukaryotic cells. Cellular DNA ligase I is another candidate as it gets recruited to the viral factories (175). However, this enzyme does not appear to be involved in catalyzing NHEJ, but rather supports microhomology mediated repair (175).

5.1.2.1 A potential mechanisms underlying VAC NHR

The simplest way to explain the origins of VAC NHEJ-like recombination events would be a process reflecting some form of aberrant replication fork repair. This involves breaking the replication fork followed by its miss repair. I outline the mechanism in Figure 5.1 and show how a tandem VAC duplication might be formed flanking the captured marker (Fig 3.4C and Fig 4.1D). In most cases, when a replication fork (i) is broken, it is quickly repaired through the gapping of the leading strand and the annealing of the lagging strand to the newly exposed homology (not shown). Such a mechanism is referred to as homology-driven repair and it is genetically silent, that is no mutations are generated. However, if we have a non-homologous DNA molecule present (ii, red), it may be accidentally used to prime repair of any broken replication fork (iv and v), and subsequently used to prime (viii) and rebuild the broken structures (ix). Depending upon which broken replication intermediate were targeted in the "anneal/prime step", such a process would account for most of the recombination events we've seen and could also account for accordion-like tandem gene duplications (20, 172). Although our illustration shows priming as the necessary step for the joining of non-homologous molecules, random ligations between molecules could accomplish the same function (142). Random ligation of two broken molecules would better explain the lack of homology at the recombination junctions, as priming would require some degree of homology between the two molecules for it to happen.

5.1.3 Recovered recombinant viruses

Chordopoxviruses all share a highly-conserved gene order (176, 177) with the many variations between these viruses arising more from the simple gain or loss of genes that have a host-specific effect. Consequently, many of the highly rearranged genomes that I recovered are most likely evolutionary dead-ends. This



(continued next page)

Fig 5.1 The Aberrant replication fork repair mechanism. The mechanism underlying most of the creation of most of the recombinant VAC can be best explained by this process. In most cases, a broken replication fork (i) can be easily be repaired by the gapping of the leading strand and annealing of the lagging strand to the exposed homology. However, if an unrelated DNA fragment is present (ii), this intermediated may be gapped (iii), then anneal and prime the end of the broken molecule for repair (iv-v). This molecule would then be used to repair the replication fork (viii) creating an insert flanked by tandem duplication (ix). The size of the duplication is dependent on the location that the 3' end of the non-homologous substrates anneals in the gapped molecule (vii). Such broken molecules (v) may also target other sites of the fork leading to deletion or more complex rearrangements. These may also create virus genome rearrangement without incorporating the insert (ii) (not shown). Figure from G. Vallee et al. (1).

would certainly be true for the complementation-dependent recombinant viruses that were recovered (Fig 4.1F). Maintaining these sub-genomic viruses require constant selective pressure and a "helper" or complementing virus. These small poxviruses still pose some interesting questions regarding virus packaging, but are more laboratory curiosities than being biologically relevant.

Poxvirus ITRs appear to be evolutionary hotspots for recombination competent viruses. Most of the replication-competent viruses that were recovered had indels near or within the ITRs. The virus with mirror insertions in the ITRs (Fig 4.1C) is an example of a virus with the indel in the ITRs. Fig 4.1B and Fig 3.4B are examples of viruses with indels extending outside of the ITRs. Viruses encoding deletions near the ITR sequences are replication competent because many of the genes encoded in or near the ITRs are non-essential and sometimes poorly conserved Chordopoxvirus genes (30). The deletions observed in the viruses with deletions and translocation/duplications were found in similar locations as in the aforementioned viruses. These viruses differed in that they encoded inverted duplications of the virus sequences found the opposite end of the genome. This gain compensated for the loss of one of the ITRs (Fig 4.1D). These translocations conserved the prototypic poxvirus genome arrangement and ensured the virus genome retained covalently closed hairpins at both ends of the genome. The ITRs are essential to viral replication (29, 178).

In two of the recombinant viruses, we observed a large tandem duplication of essential VAC sequences flanking the integrated substrate (Fig 3.4C and Fig 4.1E). Such tandem duplications are beneficial in that they ensure at least one copy of an essential gene remains fully functional (Fig 3.4E). Tandem duplication may also act as a reservoir for genetic diversity under selective pressure (20, 179) and were also frequently observed next to sites of LINE-1 mediated horizontal gene transfer (17, 18). The presence of such transient duplications next to a gene insert would increase the likelihood of the insert stably integrating itself into the virus genome if it confers a fitness advantage.

5.2 Future directions

5.2.1 Investigating the role of double-strand breaks

Based on the results of our studies, we hypothesized that VAC NHR is catalyzed by a NHEJ process which may be a form of aberrant fork repair. If this is true then a DSB may act as a precursor intermediate that's used for recombination. As a preliminary test of this hypothesis, I designed a pilot study based on using CRISPR/Cas9 technology. To do this, I adapted a protocol from A. Gowripalan et al. (147) to see if putting a DSB in the VAC genome would promote NHR at the DSB site. As a "readout" we expected to see if a DSB would increase the number of VAC that had undergone NHR much like the previous study observed for VAC homologous recombination (180).

I began by testing the N2L sgRNA (guide RNA) that I had designed, to confirm that it would target a DSB into the VAC genome and that our methods cause cutting of the VAC genome in BSC-40 cells (Fig 5.2). VAC DNA was treated with Cas9+N2L sgRNA and PFGE showed that this produced a DSB at the expected site in the VAC genome (Fig 5.2A). The treatment generated two fragments of about 24 and ~170kb and the cut site is located at a position near nt 22,500. I also extracted DNA from cells infected with VAC, cut it with *Hin*dIII, and performed a Southern blot of the fragments encompassing the N2L locus. This detected 0.45 and 1.1kb fragments that matched the fragments produced when purified VAC DNA was cut with both N2L-sgRNA and *Hin*dIII (Fig 5.2B). These bands were most prominent at 6 hr post-infection, and while Cas9 technology doesn't appear to cut 100% of the target fragments, it is clear that the method works and targets the sequence of interest.

I next conducted some infection and transfection experiments incorporating the N2L sgRNA and either the pPN_YFP+N2H or pPN_YFP plasmids. After 48 hr the total PFU, the number of recombinant viruses, and the recombinant frequencies were calculated (Table 5.1). When comparing the total PFU and number of recombinants recovered, I observed a 3-fold decrease in the number of plaques recovered from VAC-infected cells transfected with the homologous plasmid plus N2L-sgRNA relative to cells transfected with the homologous plasmid and the

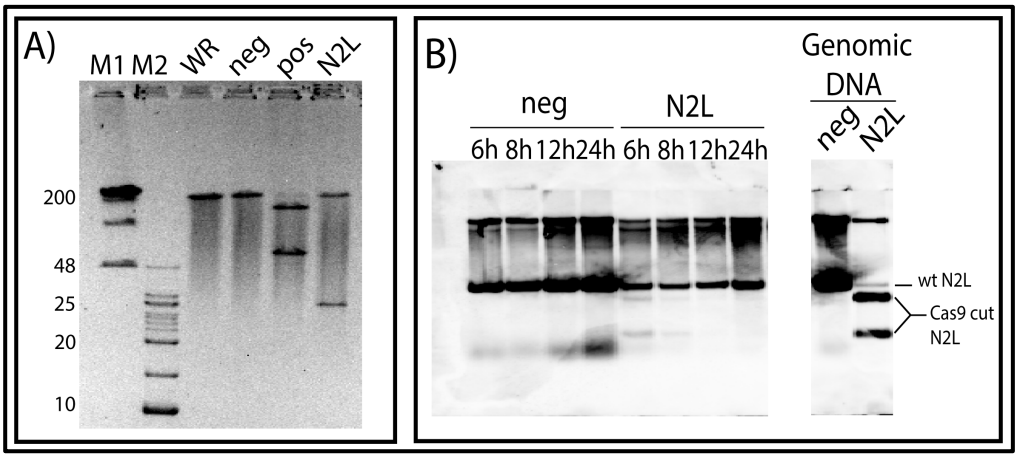


Fig 5.2 Cleavage of VAC N2L by CRISPR/Cas9. A) Purified VAC DNA was treated with Cas9 complexed with either a non-targeting guide RNA (neg), a positive control RNA targeting the A53R locus (pos) or an RNA targeting the N2L locus (N2L). The DNA fragments were separated by PFGE and stained with Sybr Safe dye. Markers included a lambda phage ladder (M1), a mix of the High Range and 1Kb plus GeneRuler ladders (M2), and uncut VAC DNA (WR). B) Southern blot analysis. VAC-infected cells were transfected with Cas9 complexed to a non-targeting sgRNA (neg) or a sgRNA complementary to the N2L locus. DNA was isolated at the indicated time points, cut with *Hind*III and size fractionated on a 1% agarose gel. The Southern blot used a biotinylated PCR probe spanning the VAC N2L locus. As a marker, VAC genomic DNA was cut with the indicated Cas9·RNA complexes and *Hind*III.

Table 5.1 Effect of DSB's on HR and NHR in VAC-infected cells

Transfected plasmid	sgRNA treatment	total PFU ($\times 10^{-6}$)	Total recombinant	Recombination frequency ($\times 10^{6}$)	Ratio (RF/145)
Homologous (463 and 313 bp)	Neg Control	53	9667	145	1.00
	N2L	16	221533	13850	101
Non-homologous	Neg Control	40	81	1.7	9×10^{-3}
	N2L	3.7	7†	1.8	9×10^{-3}

†All of the experiments were performed 3 times but 1 of the 3 replicates produced no recombinants. Plaque counts are assumed to be Poisson distributed, so the RF was estimated based on the assumption we might have missed 2 plaques in negative replicates.

control (negative) sgRNA. Despite the decrease in overall titer, I saw a 100-fold increase in the frequency of recombinant viruses that were recovered from the cells transfected with the homologous plasmid and the N2L-sgRNA. This was expected based on earlier literature (147, 180). However, when cells were transfected with the non-homologous plasmid, we saw a 10-fold decrease in titer when the cells were co-transfected with the N2L targeted sgRNA relative to the control sgRNA. Most importantly, we saw no difference in proportion of recombinants recovered from cells transfected with the negative control sgRNA *versus* the N2L sgRNA treated samples.

In addition to calculating the frequency, all of the recombinant viruses that were produced using non-homologous plasmid and the N2L sgRNA were recovered and plaque purified under MPA drug selection. Six recombinants were plaque purified and the PCR was used to amplify regions flanking the N2L locus (Fig 5.3). Five of the six recombinants still seemed to have retained the wild-type N2L locus, judging by the detection of a 1.25 kb PCR product (Fig 5.3). Presumably these had incorporated the transfected marker elsewhere in the genome. However,

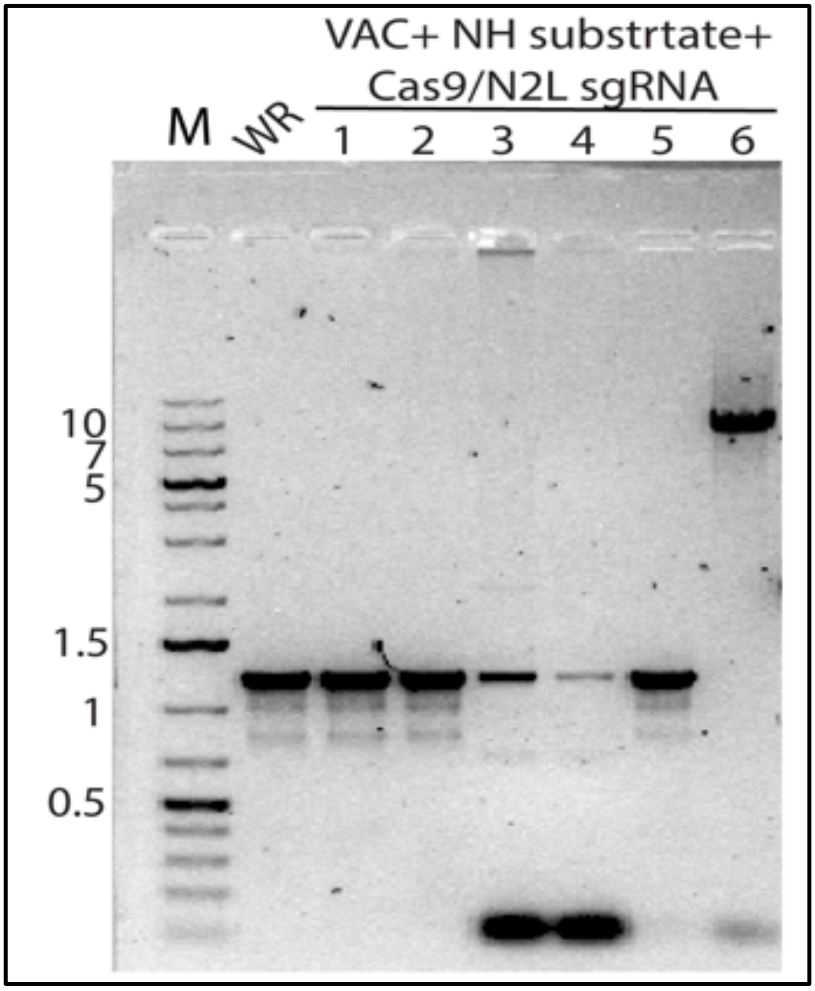


Fig 5.3 PCR amplification of the N2L locus of recombinant VAC strains 1-6. Fluorescent recombinant VAC plaques were plaque purified under drug selection. The virus DNA was extracted and the N2L locus was PCR amplified using Phusion polymerase and primers located in the flanking N1L and M1L genes. The markers included a GeneRuler 1 Kb plus ladder (M) and an N2L PCR amplicon derived from VAC strain WR.

strain 6 produced a PCR product that was ~9 kb in size. To investigate the structure of this product further, I used PCR and additional primers that bound within the YFP/gpt ORF (Fig 5.4A). These PCR reactions suggested the insert encoded at least 1 full-length copy and 2 partial copies of the YFP/gpt marker (Fig 5.4B). Further digestion of the different PCR products with *Not*I, *Blg*II, or *Xho*I (Fig 5.4B) led to the development of the map shown the Fig 5.4C. Although further studies are needed to confirm the correctness of the draft map of this particular virus, it is clear that a DSB can be a target of non-homologous recombination reactions. However, more recombinants will have to be collected and mapped to enable a comparison with the viruses described elsewhere in this thesis.

5.2.2 Further investigation of the complementation-dependent viruses

A second area of research that deserves further in investigation relates to the properties of the viruses I recovered from cells transfected with non-homologous substrates, that could only be maintained in the presence of a complementing virus. These viruses actually comprised the majority of viruses recovered in this study (Table 4.1) The recovery of such viruses introduces interesting questions about the packaging of viral genomes. Some were quite small, and one might ask if there is a lower limit to the size of a genome that poxviruses can package into a viral particle? It is conceivable that multiple small genomes could be packaged in a single viral particle, so as to collectively contribute enough DNA to fill a capsid. Such studies could perhaps be investigated with flow virometry (181, 182). A DNA dye could be used to examine if there is less DNA being encapsidated into some of the virus particles relative to wild-type VAC. It would also prove interesting to serial passage these viruses under drug selection, to see if recombination between the complementing viral genomes will eventually produce a virus encoding the YFP/gpt marker that is no longer complementation dependent. For such events to occur it would, of course, require genetic mixing and recombination between co-infecting poxviruses (162, 183).

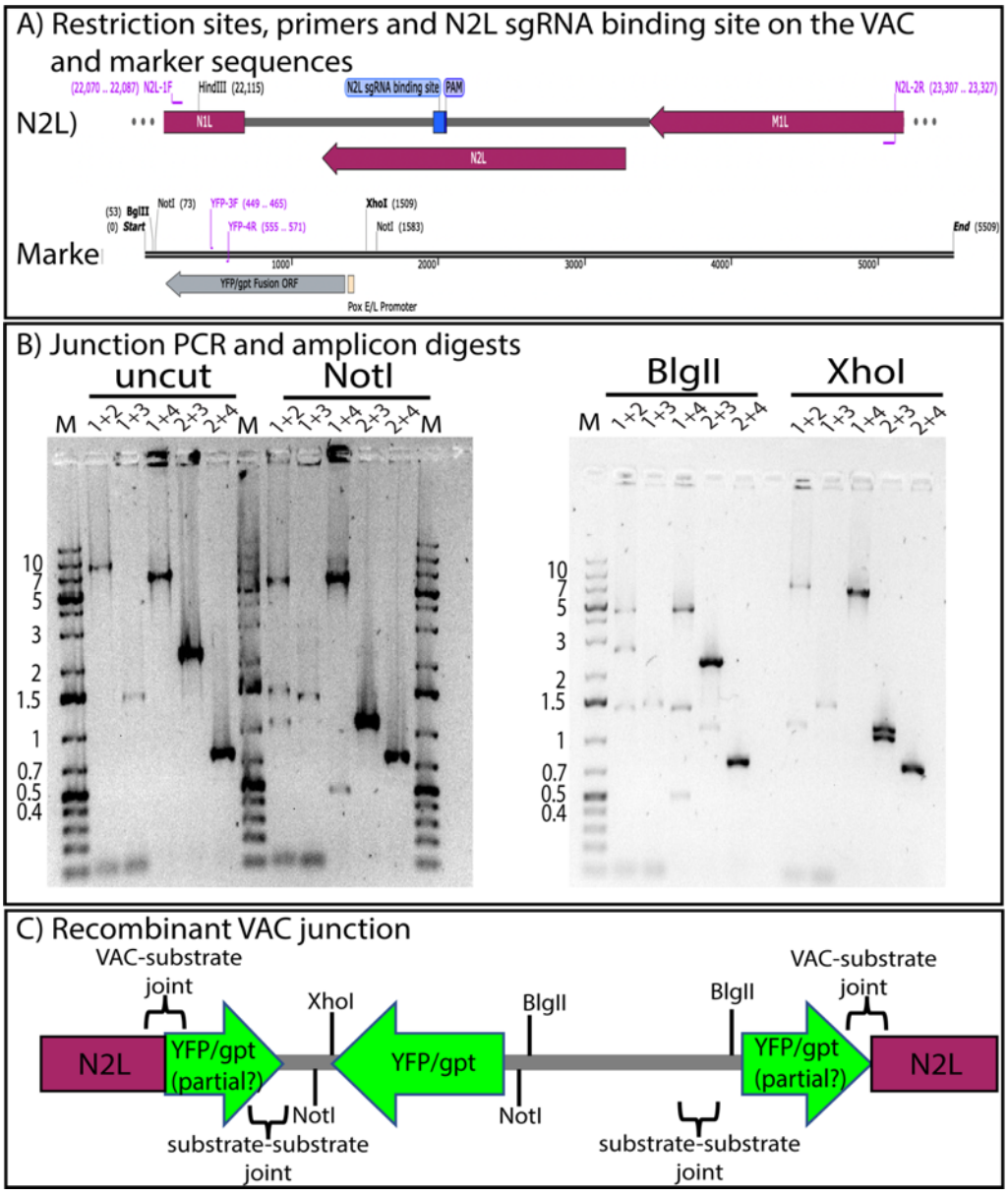


Fig 5.4 Proposed map of a NH recombinant VAC formed in association with a DSB. A) Map of the VAC N2L locus (N2L) and the transfected substrate (Marker). The location of the N2L sgRNA binding site and protospacer-adjacent motif (PAM) site are shown in blue and the PCR primer binding sites #1-4 in purple. B. PCR amplification of the recombinant N2L locus and restriction digests of those products. C) Deduced map of the region surrounding the site of integration.

References

1. Vallee G, Norris P, Paszkowski P, Noyce RS, Evans DH. 2021. Vaccinia Virus Gene Acquisition through Nonhomologous Recombination. J Virol 95:e0031821.
2. Payne S. 2017. Chapter 35 - Family Poxviridae, p 279-285. *In* Payne S (ed), Viruses doi:https://doi.org/10.1016/B978-0-12-803109-4.00035-0. Academic Press.
3. Henderson DA. 2011. The eradication of smallpox--an overview of the past, present, and future. Vaccine 29 Suppl 4:D7-9.
4. Pauli G, Blumel J, Burger R, Drosten C, Groner A, Gurtler L, Heiden M, Hildebrandt M, Jansen B, Montag-Lessing T, Offergeld R, Seitz R, Schlenkrich U, Schottstedt V, Strobel J, Willkommen H, von Konig CH. 2010. Orthopox Viruses: Infections in Humans. Transfus Med Hemother 37:351-364.
5. Fenner F. 1982. A successful eradication campaign. Global eradication of smallpox. Rev Infect Dis 4:916-30.
6. Lewis-Jones S. 2004. Zoonotic poxvirus infections in humans. Current Opinion in Infectious Diseases 17.
7. Umer BA, Noyce RS, Franczak BC, Shenouda MM, Kelly RG, Favis NA, Desaulniers M, Baldwin TA, Hitt MM, Evans DH. 2020. Deciphering the Immunomodulatory Capacity of Oncolytic Vaccinia Virus to Enhance the Immune Response to Breast Cancer. Cancer Immunology Research 8:618.
8. Guo ZS, Lu B, Guo Z, Giehl E, Feist M, Dai E, Liu W, Storkus WJ, He Y, Liu Z, Bartlett DL. 2019. Vaccinia virus-mediated cancer immunotherapy: cancer vaccines and oncolytics. Journal for ImmunoTherapy of Cancer 7:6.
9. Béguin J, Gantzer M, Farine I, Foloppe J, Klonjkowski B, Maurey C, Quéméneur É, Erbs P. 2021. Safety, biodistribution and viral shedding of oncolytic vaccinia virus TG6002 administered intravenously in healthy beagle dogs. Scientific Reports 11:2209.
10. Schweneker M, Laimbacher AS, Zimmer G, Wagner S, Schraner EM, Wolferstätter M, Klingenberg M, Dirmeier U, Steigerwald R, Lauterbach H, Hochrein H, Chaplin P, Suter M, Hausmann J. 2017. Recombinant Modified Vaccinia Virus Ankara Generating Ebola Virus-Like Particles. J Virol 91.
11. Prow NA, Liu L, Nakayama E, Cooper TH, Yan K, Eldi P, Hazlewood JE, Tang B, Le TT, Setoh YX, Khromykh AA, Hobson-Peters J, Diener KR, Howley PM, Hayball JD, Suhrbier A. 2018. A vaccinia-based single vector construct multi-pathogen vaccine protects against both Zika and chikungunya viruses. Nature Communications 9:1230.
12. García-Arriaza J, Garaigorta U, Pérez P, Lázaro-Frías A, Zamora C, Gastaminza P, del Fresno C, Casasnovas José M, Óscar SSC, Sancho D, Esteban M, Heise Mark T. COVID-19 Vaccine Candidates Based on Modified Vaccinia Virus Ankara Expressing the SARS-CoV-2 Spike Protein

Induce Robust T- and B-Cell Immune Responses and Full Efficacy in Mice. Journal of Virology 95:e02260-20.
13. Koonin EV, Yutin N. 2019. Chapter Five - Evolution of the Large Nucleocytoplasmic DNA Viruses of Eukaryotes and Convergent Origins of Viral Gigantism, p 167-202. *In* Kielian M, Mettenleiter TC, Roossinck MJ (ed), Advances in Virus Research, vol 103. Academic Press.
14. Philippe N, Legendre M, Doutre G, Couté Y, Poirot O, Lescot M, Arslan D, Seltzer V, Bertaux L, Bruley C, Garin J, Claverie J-M, Abergel C. 2013. Pandoraviruses: Amoeba Viruses with Genomes Up to 2.5 Mb Reaching That of Parasitic Eukaryotes. Science 341:281-286.
15. Filée J. 2013. Route of NCLDV evolution: the genomic accordion. Current Opinion in Virology 3:595-599.
16. Filée J, Pouget N, Chandler M. 2008. Phylogenetic evidence for extensive lateral acquisition of cellular genes by Nucleocytoplasmic large DNA viruses. BMC Evolutionary Biology 8:320.
17. Rahman MJ, Haller SL, Li J, Brennan G, Rothenburg S. 2020. Cascade evolution in poxviruses: retrotransposon-mediated host gene capture, complementation and recombination. bioRxiv doi:10.1101/2020.10.26.355610:2020.10.26.355610.
18. Fixsen SM, Cone KR, Goldstein SA, Sasani TA, Quinlan AR, Rothenburg S, Elde NC. 2020. Poxviruses capture host genes by LINE-1 retrotransposition. bioRxiv doi:10.1101/2020.10.26.356402:2020.10.26.356402.
19. Suhre K. 2005. Gene and genome duplication in Acanthamoeba polyphaga Mimivirus. Journal of virology 79:14095-14101.
20. Elde NC, Child SJ, Eickbush MT, Kitzman JO, Rogers KS, Shendure J, Geballe AP, Malik HS. 2012. Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses. Cell 150:831-41.
21. Bennett CJ, Webb M, Willer DO, Evans DH. 2003. Genetic and phylogenetic characterization of the type II cyclobutane pyrimidine dimer photolyases encoded by Leporipoxviruses. Virology 315:10-19.
22. Willer DO, McFadden G, Evans DH. 1999. The Complete Genome Sequence of Shope (Rabbit) Fibroma Virus. Virology 264:319-343.
23. Hendrickson RC, Wang C, Hatcher EL, Lefkowitz EJ. 2010. Orthopoxvirus genome evolution: the role of gene loss. Viruses 2:1933-67.
24. Lefkowitz EJ, Wang C, Upton C. 2006. Poxviruses: past, present and future. Virus Res 117:105-18.
25. Geshelin P, Berns KI. 1974. Characterization and localization of the naturally occurring cross-links in vaccinia virus DNA. J Mol Biol 88:785-96.
26. Baroudy BM, Venkatesan S, Moss B. 1982. Incompletely base-paired flip-flop terminal loops link the two DNA strands of the vaccinia virus genome into one uninterrupted polynucleotide chain. Cell 28:315-24.
27. Merchlinsky M. 1990. Mutational analysis of the resolution sequence of vaccinia virus DNA: essential sequence consists of two separate AT-rich

regions highly conserved among poxviruses. Journal of Virology 64:5029-5035.
28. DeLange AM, McFadden G. 1987. Efficient resolution of replicated poxvirus telomeres to native hairpin structures requires two inverted symmetrical copies of a core target DNA sequence. Journal of Virology 61:1957-1963.
29. Senkevich TG, Bruno D, Martens C, Porcella SF, Wolf YI, Moss B. 2015. Mapping vaccinia virus DNA replication origins at nucleotide level by deep sequencing. Proceedings of the National Academy of Sciences 112:10908-10913.
30. Upton C, Slack S, Hunter AL, Ehlers A, Roper RL. 2003. Poxvirus orthologous clusters: toward defining the minimum essential poxvirus genome. J Virol 77:7590-600.
31. Fleming SB, McCaughan CA, Andrews AE, Nash AD, Mercer AA. 1997. A homolog of interleukin-10 is encoded by the poxvirus orf virus. J Virol 71:4857-61.
32. Saraiva M, Alcami A. 2001. CrmE, a novel soluble tumor necrosis factor receptor encoded by poxviruses. J Virol 75:226-33.
33. Johnston JB, McFadden G. 2003. Poxvirus Immunomodulatory Strategies: Current Perspectives. Journal of Virology 77:6093-6100.
34. Odom MR, Curtis Hendrickson R, Lefkowitz EJ. 2009. Poxvirus protein evolution: Family wide assessment of possible horizontal gene transfer events. Virus Research 144:233-249.
35. Graham SC, Bahar MW, Abrescia NGA, Smith GL, Stuart DI, Grimes JM. 2007. Structure of CrmE, a Virus-encoded Tumour Necrosis Factor Receptor. Journal of Molecular Biology 372:660-671.
36. Piskurek O, Okada N. 2007. Poxviruses as possible vectors for horizontal transfer of retroposons from reptiles to mammals. Proceedings of the National Academy of Sciences 104:12046.
37. Payne LG, Norrby E. 1978. Adsorption and penetration of enveloped and naked vaccinia virus particles. J Virol 27:19-27.
38. Chung CS, Hsiao JC, Chang YS, Chang W. 1998. A27L protein mediates vaccinia virus interaction with cell surface heparan sulfate. J Virol 72:1577-85.
39. Hsiao JC, Chung CS, Chang W. 1998. Cell surface proteoglycans are necessary for A27L protein-mediated cell fusion: identification of the N-terminal region of A27L protein as the glycosaminoglycan-binding domain. J Virol 72:8374-9.
40. Lin CL, Chung CS, Heine HG, Chang W. 2000. 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 74:3353-65.

41. Vazquez MI, Esteban M. 1999. Identification of functional domains in the 14-kilodalton envelope protein (A27L) of vaccinia virus. J Virol 73:9098-109.
42. Schmidt FI, Bleck CK, Mercer J. 2012. Poxvirus host cell entry. Curr Opin Virol 2:20-7.
43. Hsiao JC, Chung CS, Chang W. 1999. Vaccinia virus envelope D8L protein binds to cell surface chondroitin sulfate and mediates the adsorption of intracellular mature virions to cells. J Virol 73:8750-61.
44. Rodriguez D, Rodriguez JR, Esteban M. 1993. The vaccinia virus 14-kilodalton fusion protein forms a stable complex with the processed protein encoded by the vaccinia virus A17L gene. J Virol 67:3435-40.
45. Carter GC, Law M, Hollinshead M, Smith GL. 2005. Entry of the vaccinia virus intracellular mature virion and its interactions with glycosaminoglycans. J Gen Virol 86:1279-1290.
46. Law M, Carter GC, Roberts KL, Hollinshead M, Smith GL. 2006. Ligand-induced and nonfusogenic dissolution of a viral membrane. Proc Natl Acad Sci U S A 103:5989-94.
47. Armstrong JA, Metz DH, Young MR. 1973. The mode of entry of vaccinia virus into L cells. J Gen Virol 21:533-7.
48. Dales S, Kajioka R. 1964. The Cycle of Multiplication of Vaccinia Virus in Earle's Strain L Cells. I. Uptake and Penetration. Virology 24:278-94.
49. Dales S, Siminovitch L. 1961. The development of vaccinia virus in Earle's L strain cells as examined by electron microscopy. J Biophys Biochem Cytol 10:475-503.
50. Moss B. 2012. Poxvirus cell entry: how many proteins does it take? Viruses 4:688-707.
51. Schramm B, Locker JK. 2005. Cytoplasmic organization of POXvirus DNA replication. Traffic 6:839-46.
52. Broyles SS. 2003. Vaccinia virus transcription. J Gen Virol 84:2293-2303.
53. Kilcher S, Mercer J. 2015. DNA virus uncoating. Virology 479-480:578-90.
54. Kilcher S, Schmidt FI, Schneider C, Kopf M, Helenius A, Mercer J. 2014. siRNA screen of early poxvirus genes identifies the AAA+ ATPase D5 as the virus genome-uncoating factor. Cell Host Microbe 15:103-12.
55. Mercer J, Snijder B, Sacher R, Burkard C, Bleck CK, Stahlberg H, Pelkmans L, Helenius A. 2012. RNAi screening reveals proteasome- and Cullin3-dependent stages in vaccinia virus infection. Cell Rep 2:1036-47.
56. Liu B, Panda D, Mendez-Rios JD, Ganesan S, Wyatt LS, Moss B. 2018. Identification of Poxvirus Genome Uncoating and DNA Replication Factors with Mutually Redundant Roles. J Virol 92.
57. Tolonen N, Doglio L, Schleich S, Krijnse Locker J. 2001. Vaccinia virus DNA replication occurs in endoplasmic reticulum-enclosed cytoplasmic mini-nuclei. Mol Biol Cell 12:2031-46.

58. Katsafanas GC, Moss B. 2007. Colocalization of transcription and translation within cytoplasmic poxvirus factories coordinates viral expression and subjugates host functions. Cell Host Microbe 2:221-8.
59. Moss B. 2013. Poxvirus DNA replication. Cold Spring Harb Perspect Biol 5.
60. Paran N, De Silva FS, Senkevich TG, Moss B. 2009. Cellular DNA ligase I is recruited to cytoplasmic vaccinia virus factories and masks the role of the vaccinia ligase in viral DNA replication. Cell Host Microbe 6:563-9.
61. Senkevich TG, Bruno D, Martens C, Porcella SF, Wolf YI, Moss B. 2015. Mapping vaccinia virus DNA replication origins at nucleotide level by deep sequencing. Proc Natl Acad Sci U S A 112:10908-13.
62. Moyer RW, Graves RL. 1981. The mechanism of cytoplasmic orthopoxvirus DNA replication. Cell 27:391-401.
63. Garcia AD, Aravind L, Koonin EV, Moss B. 2000. Bacterial-type DNA holliday junction resolvases in eukaryotic viruses. Proc Natl Acad Sci U S A 97:8926-31.
64. Garcia AD, Moss B. 2001. Repression of vaccinia virus Holliday junction resolvase inhibits processing of viral DNA into unit-length genomes. J Virol 75:6460-71.
65. Roberts KL, Smith GL. 2008. Vaccinia virus morphogenesis and dissemination. Trends Microbiol 16:472-9.
66. Moss B. 2018. Origin of the poxviral membrane: A 50-year-old riddle. PLoS Pathog 14:e1007002.
67. Liu L, Cooper T, Howley PM, Hayball JD. 2014. From crescent to mature virion: vaccinia virus assembly and maturation. Viruses 6:3787-808.
68. Chlanda P, Carbajal MA, Cyrklaff M, Griffiths G, Krijnse-Locker J. 2009. Membrane rupture generates single open membrane sheets during vaccinia virus assembly. Cell Host Microbe 6:81-90.
69. Suarez C, Welsch S, Chlanda P, Hagen W, Hoppe S, Kolovou A, Pagnier I, Raoult D, Krijnse Locker J. 2013. Open membranes are the precursors for assembly of large DNA viruses. Cell Microbiol 15:1883-95.
70. Morgan C. 1976. The insertion of DNA into vaccinia virus. Science 193:591-2.
71. Szajner P, Weisberg AS, Wolffe EJ, Moss B. 2001. Vaccinia virus A30L protein is required for association of viral membranes with dense viroplasm to form immature virions. J Virol 75:5752-61.
72. Hollinshead M, Rodger G, Van Eijl H, Law M, Hollinshead R, Vaux DJ, Smith GL. 2001. Vaccinia virus utilizes microtubules for movement to the cell surface. J Cell Biol 154:389-402.
73. Dodding MP, Way M. 2011. Coupling viruses to dynein and kinesin-1. EMBO J 30:3527-39.
74. Tooze J, Hollinshead M, Reis B, Radsak K, Kern H. 1993. Progeny vaccinia and human cytomegalovirus particles utilize early endosomal cisternae for their envelopes. Eur J Cell Biol 60:163-78.

75. Schmelz M, Sodeik B, Ericsson M, Wolffe EJ, Shida H, Hiller G, Griffiths G. 1994. Assembly of vaccinia virus: the second wrapping cisterna is derived from the trans Golgi network. J Virol 68:130-47.
76. Meseda CA, Campbell J, Kumar A, Garcia AD, Merchlinsky M, Weir JP. 2013. Effect of the deletion of genes encoding proteins of the extracellular virion form of vaccinia virus on vaccine immunogenicity and protective effectiveness in the mouse model. PLoS One 8:e67984.
77. Chen Y, Honeychurch KM, Yang G, Byrd CM, Harver C, Hruby DE, Jordan R. 2009. Vaccinia virus p37 interacts with host proteins associated with LE-derived transport vesicle biogenesis. Virol J 6:44.
78. Mathew E, Sanderson CM, Hollinshead M, Smith GL. 1998. The extracellular domain of vaccinia virus protein B5R affects plaque phenotype, extracellular enveloped virus release, and intracellular actin tail formation. J Virol 72:2429-38.
79. Morita R, Nakane S, Shimada A, Inoue M, Iino H, Wakamatsu T, Fukui K, Nakagawa N, Masui R, Kuramitsu S. 2010. Molecular mechanisms of the whole DNA repair system: a comparison of bacterial and eukaryotic systems. J Nucleic Acids 2010:179594.
80. Castedo M, Perfettini JL, Roumier T, Andreau K, Medema R, Kroemer G. 2004. Cell death by mitotic catastrophe: a molecular definition. Oncogene 23:2825-37.
81. Jasin M, Rothstein R. 2013. Repair of strand breaks by homologous recombination. Cold Spring Harbor perspectives in biology 5:a012740-a012740.
82. Chang HHY, Pannunzio NR, Adachi N, Lieber MR. 2017. Non-homologous DNA end joining and alternative pathways to double-strand break repair. Nat Rev Mol Cell Biol 18:495-506.
83. Luria SE, Dulbecco R. 1949. Genetic Recombinations Leading to Production of Active Bacteriophage from Ultraviolet Inactivated Bacteriophage Particles. Genetics 34:93-125.
84. McClain ME, Spendlove RS. 1966. Multiplicity reactivation of reovirus particles after exposure to ultraviolet light. Journal of bacteriology 92:1422-1429.
85. Karska-Wysocki B, Mamet-Bratley MD. 1984. Multiplicity reactivation of bacteriophage T7 inactivated by methyl methanesulfonate. Journal of virology 52:1009-1010.
86. Drake JW. 1962. Multiplicity reactivation of Newcastle disease virus. Journal of bacteriology 84:352-356.
87. Barry RD. 1961. The multiplication of influenza virus: II. Multiplicity reactivation of ultraviolet irradiated virus. Virology 14:398-405.
88. Yamamoto H, Shimojo H. 1971. Multiplicity reactivation of human adenovirus type 12 and simian virus 40 irradiated by ultraviolet light. Virology 45 2:529-31.

89. Abel P. 1962. Multiplicity reactivation and marker rescue with vaccinia virus. Virology 17:511-519.
90. Blanco M, Devoret R. 1973. Repair mechanisms involved in prophage reactivation and UV reactivation of UV-irradiated phage λ. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 17:293-305.
91. Nafissi N, Slavcev R. 2014. Bacteriophage recombination systems and biotechnical applications. Applied Microbiology and Biotechnology 98:2841-2851.
92. Murphy Kenan C, Slauch James M. 2016. λ Recombination and Recombineering. EcoSal Plus 7.
93. Mosig G. 1998. Recombination and recombination-dependent DNA replication in bacteriophage T4. Annu Rev Genet 32:379-413.
94. Hillyar CRT. 2012. Genetic recombination in bacteriophage lambda. Bioscience Horizons: The International Journal of Student Research 5.
95. Kreuzer KN. 2000. Recombination-dependent DNA replication in phage T4. Trends Biochem Sci 25:165-73.
96. Bedson HS, Dumbell KR. 1964. Hybrids Derived from the Viruses of Variola Major and Cowpox. J Hyg (Lond) 62:147-58.
97. Gershon PD, Kitching RP, Hammond JM, Black DN. 1989. Poxvirus genetic recombination during natural virus transmission. J Gen Virol 70 (Pt 2):485-9.
98. Fenner F, Comben BM. 1958. Genetic studies with mammalian poxviruses: I. Demonstration of recombination between two strains of vaccinia virus. Virology 5:530-548.
99. Baxby D. 1977. The origins of vaccinia virus. J Infect Dis 136:453-5.
100. Willer DO, Yao XD, Mann MJ, Evans DH. 2000. In vitro concatemer formation catalyzed by vaccinia virus DNA polymerase. Virology 278:562-9.
101. Yao XD, Evans DH. 2001. Effects of DNA structure and homology length on vaccinia virus recombination. J Virol 75:6923-32.
102. Yao XD, Evans DH. 2003. Characterization of the recombinant joints formed by single-strand annealing reactions in vaccinia virus-infected cells. Virology 308:147-56.
103. Schumacher AJ, Mohni KN, Kan Y, Hendrickson EA, Stark JM, Weller SK. 2012. The HSV-1 exonuclease, UL12, stimulates recombination by a single strand annealing mechanism. PLoS Pathog 8:e1002862.
104. Bleuit JS, Xu H, Ma Y, Wang T, Liu J, Morrical SW. 2001. Mediator proteins orchestrate enzyme-ssDNA assembly during T4 recombination-dependent DNA replication and repair. Proc Natl Acad Sci U S A 98:8298-305.
105. Caldwell BJ, Bell CE. 2019. Structure and mechanism of the Red recombination system of bacteriophage lambda. Prog Biophys Mol Biol 147:33-46.

106. Reuven NB, Staire AE, Myers RS, Weller SK. 2003. The herpes simplex virus type 1 alkaline nuclease and single-stranded DNA binding protein mediate strand exchange in vitro. J Virol 77:7425-33.
107. Cromie GA, Connelly JC, Leach DR. 2001. Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. Mol Cell 8:1163-74.
108. Qin L, Evans DH. 2014. Genome scale patterns of recombination between coinfecting vaccinia viruses. Journal of virology 88:5277-5286.
109. Gammon DB, Evans DH. 2009. The 3'-to-5' exonuclease activity of vaccinia virus DNA polymerase is essential and plays a role in promoting virus genetic recombination. J Virol 83:4236-50.
110. Hamilton MD, Nuara AA, Gammon DB, Buller RM, Evans DH. 2007. Duplex strand joining reactions catalyzed by vaccinia virus DNA polymerase. Nucleic acids research 35:143-151.
111. Guglielmini J, Woo AC, Krupovic M, Forterre P, Gaia M. 2019. Diversification of giant and large eukaryotic dsDNA viruses predated the origin of modern eukaryotes. Proceedings of the National Academy of Sciences 116:19585-19592.
112. Challberg MD, Englund PT. 1979. Purification and properties of the deoxyribonucleic acid polymerase induced by vaccinia virus. J Biol Chem 254:7812-9.
113. Merchlinsky M. 1989. Intramolecular homologous recombination in cells infected with temperature-sensitive mutants of vaccinia virus. Journal of virology 63:2030-2035.
114. Ball LA. 1987. High-frequency homologous recombination in vaccinia virus DNA. Journal of virology 61:1788-1795.
115. Evans DH, Stuart D, McFadden G. 1988. High levels of genetic recombination among cotransfected plasmid DNAs in poxvirus-infected mammalian cells. J Virol 62:367-75.
116. McDonald WF, Traktman P. 1994. Overexpression and purification of the vaccinia virus DNA polymerase. Protein Expr Purif 5:409-21.
117. Willer DO, Mann MJ, Zhang W, Evans DH. 1999. Vaccinia virus DNA polymerase promotes DNA pairing and strand-transfer reactions. Virology 257:511-23.
118. Magee WC, Hostetler KY, Evans DH. 2005. Mechanism of inhibition of vaccinia virus DNA polymerase by cidofovir diphosphate. Antimicrobial agents and chemotherapy 49:3153-3162.
119. Pestryakov PE, Lavrik OI. 2008. Mechanisms of single-stranded DNA-binding protein functioning in cellular DNA metabolism. Biochemistry (Mosc) 73:1388-404.
120. Wu Y, Lu J, Kang T. 2016. Human single-stranded DNA binding proteins: guardians of genome stability. Acta Biochim Biophys Sin (Shanghai) 48:671-7.

121. Rochester SC, Traktman P. 1998. Characterization of the single-stranded DNA binding protein encoded by the vaccinia virus I3 gene. Journal of virology 72:2917-2926.
122. Tseng M, Palaniyar N, Zhang W, Evans DH. 1999. DNA Binding and Aggregation Properties of the Vaccinia Virus I3L Gene Product*. Journal of Biological Chemistry 274:21637-21644.
123. Greseth MD, Boyle KA, Bluma MS, Unger B, Wiebe MS, Soares-Martins JA, Wickramasekera NT, Wahlberg J, Traktman P. 2012. Molecular genetic and biochemical characterization of the vaccinia virus I3 protein, the replicative single-stranded DNA binding protein. Journal of virology 86:6197-6209.
124. Welsch S, Doglio L, Schleich S, Krijnse Locker J. 2003. The vaccinia virus I3L gene product is localized to a complex endoplasmic reticulum-associated structure that contains the viral parental DNA. Journal of virology 77:6014-6028.
125. Gammon DB, Gowrishankar B, Duraffour S, Andrei G, Upton C, Evans DH. 2010. Vaccinia Virus–Encoded Ribonucleotide Reductase Subunits Are Differentially Required for Replication and Pathogenesis. PLOS Pathogens 6:e1000984.
126. Harrison ML, Desaulniers MA, Noyce RS, Evans DH. 2016. The acidic C-terminus of vaccinia virus I3 single-strand binding protein promotes proper assembly of DNA–protein complexes. Virology 489:212-222.
127. Tseng M, Palaniyar N, Zhang W, Evans DH. 1999. DNA binding and aggregation properties of the vaccinia virus I3L gene product. J Biol Chem 274:21637-44.
128. Da Silva M, Shen L, Tcherepanov V, Watson C, Upton C. 2006. Predicted function of the vaccinia virus G5R protein. Bioinformatics 22:2846-50.
129. Shen B, Singh P, Liu R, Qiu J, Zheng L, Finger LD, Alas S. 2005. Multiple but dissectible functions of FEN-1 nucleases in nucleic acid processing, genome stability and diseases. Bioessays 27:717-29.
130. Liu Y, Kao HI, Bambara RA. 2004. Flap endonuclease 1: a central component of DNA metabolism. Annu Rev Biochem 73:589-615.
131. Senkevich TG, Koonin EV, Moss B. 2009. Predicted poxvirus FEN1-like nuclease required for homologous recombination, double-strand break repair and full-size genome formation. Proc Natl Acad Sci U S A 106:17921-6.
132. Lehman IR. 1974. DNA ligase: structure, mechanism, and function. Science 186:790-7.
133. Colinas RJ, Goebel SJ, Davis SW, Johnson GP, Norton EK, Paoletti E. 1990. A DNA ligase gene in the Copenhagen strain of vaccinia virus is nonessential for viral replication and recombination. Virology 179:267-75.
134. Paran N, De Silva FS, Senkevich TG, Moss B. 2009. Cellular DNA ligase I is recruited to cytoplasmic vaccinia virus factories and masks the role of the vaccinia ligase in viral DNA replication. Cell host & microbe 6:563-569.

135. Hamilton MD, Evans DH. 2005. Enzymatic processing of replication and recombination intermediates by the vaccinia virus DNA polymerase. Nucleic Acids Res 33:2259-68.
136. Lieber MR. 2010. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annu Rev Biochem 79:181-211.
137. Nick McElhinny SA, Havener JM, Garcia-Diaz M, Juárez R, Bebenek K, Kee BL, Blanco L, Kunkel TA, Ramsden DA. 2005. A Gradient of Template Dependence Defines Distinct Biological Roles for Family X Polymerases in Nonhomologous End Joining. Molecular Cell 19:357-366.
138. Frank KM, Sharpless NE, Gao Y, Sekiguchi JM, Ferguson DO, Zhu C, Manis JP, Horner J, DePinho RA, Alt FW. 2000. DNA ligase IV deficiency in mice leads to defective neurogenesis and embryonic lethality via the p53 pathway. Mol Cell 5:993-1002.
139. O'Driscoll M, Gennery AR, Seidel J, Concannon P, Jeggo PA. 2004. An overview of three new disorders associated with genetic instability: LIG4 syndrome, RS-SCID and ATR-Seckel syndrome. DNA Repair (Amst) 3:1227-35.
140. Bertrand C, Thibessard A, Bruand C, Lecointe F, Leblond P. 2019. Bacterial NHEJ: a never ending story. Molecular Microbiology 111:1139-1151.
141. Malu S, Malshetty V, Francis D, Cortes P. 2012. Role of non-homologous end joining in V(D)J recombination. Immunologic Research 54:233-246.
142. Luteijn RD, Drexler I, Smith GL, Lebbink RJ, Wiertz E. 2018. Mutagenic repair of double-stranded DNA breaks in vaccinia virus genomes requires cellular DNA ligase IV activity in the cytosol. J Gen Virol 99:790-804.
143. Saito S, Kurosawa A, Adachi N. 2016. Mutations in XRCC4 cause primordial dwarfism without causing immunodeficiency. J Hum Genet 61:679-85.
144. Chayot R, Montagne B, Mazel D, Ricchetti M. 2010. An end-joining repair mechanism in Escherichia coli. Proc Natl Acad Sci U S A 107:2141-6.
145. Rintoul JL, Wang J, Gammon DB, van Buuren NJ, Garson K, Jardine K, Barry M, Evans DH, Bell JC. 2011. A selectable and excisable marker system for the rapid creation of recombinant poxviruses. PLoS One 6:e24643.
146. Teferi WM, Desaulniers MA, Noyce RS, Shenouda M, Umer B, Evans DH. 2017. The vaccinia virus K7 protein promotes histone methylation associated with heterochromatin formation. PLoS One 12:e0173056.
147. Gowripalan A, Smith S, Stefanovic T, Tscharke DC. 2020. Rapid poxvirus engineering using CRISPR/Cas9 as a selection tool. Commun Biol 3:643.
148. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595.
149. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-20.

150. Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics 31:3350-2.
151. Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, Prjibelski AD, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R, Clingenpeel SR, Woyke T, McLean JS, Lasken R, Tesler G, Alekseyev MA, Pevzner PA. 2013. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. J Comput Biol 20:714-37.
152. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963.
153. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357-359.
154. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-2079.
155. Sovic I, Sikic M, Wilm A, Fenlon SN, Chen S, Nagarajan N. 2016. Fast and sensitive mapping of nanopore sequencing reads with GraphMap. Nat Commun 7:11307.
156. Shuman S. 1991. Recombination mediated by vaccinia virus DNA topoisomerase I in Escherichia coli is sequence specific. Proceedings of the National Academy of Sciences 88:10104-10108.
157. Yao XD, Evans DH. 2004. Construction of recombinant vaccinia viruses using leporipoxvirus-catalyzed recombination and reactivation of orthopoxvirus DNA. Methods Mol Biol 269:51-64.
158. Kotwal GJ, Abrahams MR. 2004. Growing poxviruses and determining virus titer. Methods Mol Biol 269:101-12.
159. Hertig C, Coupar BE, Gould AR, Boyle DB. 1997. Field and vaccine strains of fowlpox virus carry integrated sequences from the avian retrovirus, reticuloendotheliosis virus. Virology 235:367-76.
160. Singh P, Schnitzlein WM, Tripathy DN. 2003. Reticuloendotheliosis virus sequences within the genomes of field strains of fowlpox virus display variability. J Virol 77:5855-62.
161. Ball LA. 1987. High-frequency homologous recombination in vaccinia virus DNA. J Virol 61:1788-95.
162. Paszkowski P, Noyce RS, Evans DH. 2016. Live-Cell Imaging of Vaccinia Virus Recombination. PLoS Pathog 12:e1005824.
163. Mucker EM, Chapman J, Huzella LM, Huggins JW, Shamblin J, Robinson CG, Hensley LE. 2015. Susceptibility of Marmosets (Callithrix jacchus) to Monkeypox Virus: A Low Dose Prospective Model for Monkeypox and Smallpox Disease. PLoS One 10:e0131742.
164. Fang H, Wu Y, Narzisi G, O'Rawe JA, Barron LT, Rosenbaum J, Ronemus M, Iossifov I, Schatz MC, Lyon GJ. 2014. Reducing INDEL calling errors in whole genome and exome sequencing data. Genome Med 6:89.

165. Ajay SS, Parker SC, Abaan HO, Fajardo KV, Margulies EH. 2011. Accurate and comprehensive sequencing of personal genomes. Genome Res 21:1498-505.
166. Medvedev P, Fiume M, Dzamba M, Smith T, Brudno M. 2010. Detecting copy number variation with mated short reads. Genome Res 20:1613-22.
167. Xie C, Tammi MT. 2009. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. BMC Bioinformatics 10:80.
168. McFadden G, Dales S. 1979. Biogenesis of poxviruses: mirror-image deletions in vaccinia virus DNA. Cell 18:101-8.
169. Du S, Traktman P. 1996. Vaccinia virus DNA replication: two hundred base pairs of telomeric sequence confer optimal replication efficiency on minichromosome templates. Proc Natl Acad Sci U S A 93:9693-8.
170. Shuman S. 1989. Vaccinia DNA topoisomerase I promotes illegitimate recombination in Escherichia coli. Proc Natl Acad Sci U S A 86:3489-93.
171. Shuman S. 1994. Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase. J Biol Chem 269:32678-84.
172. Slabaugh MB, Roseman NA, Mathews CK. 1989. Amplification of the ribonucleotide reductase small subunit gene: analysis of novel joints and the mechanism of gene duplication in vaccinia virus. Nucleic Acids Res 17:7073-88.
173. Sekiguchi J, Shuman S. 1997. Ligation of RNA-containing duplexes by vaccinia DNA ligase. Biochemistry 36:9073-9.
174. Wilson TE, Grawunder U, Lieber MR. 1997. Yeast DNA ligase IV mediates non-homologous DNA end joining. Nature 388:495-8.
175. Conlin MP, Reid DA, Small GW, Chang HH, Watanabe G, Lieber MR, Ramsden DA, Rothenberg E. 2017. DNA Ligase IV Guides End-Processing Choice during Nonhomologous End Joining. Cell Rep 20:2810-2819.
176. Mackett M, Archard LC. 1979. Conservation and variation in Orthopoxvirus genome structure. J Gen Virol 45:683-701.
177. McLysaght A, Baldi PF, Gaut BS. 2003. Extensive gene gain associated with adaptive evolution of poxviruses. Proc Natl Acad Sci U S A 100:15655-60.
178. Merchlinsky M. 1990. Resolution of poxvirus telomeres: processing of vaccinia virus concatemer junctions by conservative strand exchange. Journal of Virology 64:3437-3446.
179. Sasani TA, Cone KR, Quinlan AR, Elde NC. 2018. Long read sequencing reveals poxvirus evolution through rapid homogenization of gene arrays. Elife 7.
180. Yuan M, Zhang W, Wang J, Al Yaghchi C, Ahmed J, Chard L, Lemoine NR, Wang Y. 2015. Efficiently editing the vaccinia virus genome by using the CRISPR-Cas9 system. J Virol 89:5176-9.
181. Zamora JLR, Aguilar HC. 2018. Flow virometry as a tool to study viruses. Methods 134-135:87-97.

182. Lippe R. 2018. Flow Virometry: a Powerful Tool To Functionally Characterize Viruses. J Virol 92.
183. Qin L, Evans DH. 2014. Genome scale patterns of recombination between coinfecting vaccinia viruses. J Virol 88:5277-86.

Table 1. DNA sequences surrounding the recombination junctions (1)

Virus	Left junction	Overlap (bp)	Sequence alignment (60 bp)
1-1-1-1	>WR left >1-1-1-1 left >YFP right	0	ACAGTTGCTCAAAAACGATGGCAGTGACTTATGAGTTACGTTACACTTTGGAGTCTCATC ACAGTTGCTCAAAAACGATGGCAGTGACTTTGACGAGCATCACAAAAATCGACGCTCAAG GCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAG * * * * * * * * * * * ** ****
NH4	>WR left >NH4 left >YFP left	0	CCTACGCGTATACTCCCTTGCATCATATACGTTCGTCACCAGATCGTTTGTTTCGGCCTG CCTACGCGTATACTCCCTTGCATCATATACCCCCGAAGAACGTTTTCCAATGATGAGCAC CAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCAC * *** * ***** * * ** * * ** * **
NH5	>WR left >NH5 left >YFP left (30 bp)	1	CCATCAGAAAGAGGTTTAATATTTTTGTGAGACCATCGAAGAGAGAAAGAGATAAAACTT CCATCAGAAAGAGGTTTAATATTTTTGTGAACCTGCAGGACTAGTCCCTTTAGTGAGGGT AAACCTGCAGGACTAGTCCCTTTAGTGAGGGT * * * * * * * *
G5	>WR left >G5 left >YFP left	1	AGATTCCAAGAATATAACTAAGTCGGGTTTAGGCAATCCAGATTCATAACTCTTACTGAG AGATTCCAAGAATATAACTAAGTCGGGTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCT CCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCT *** * * *
G8	>WR left >G8 left >YFP left	0	TGCACGTCTACAGGTTGCAGCATAGACTTTGTCACAACAGAAAAAGTGTGCGTGACAGCA TGCACGTCTACAGGTTGCAGCATAGACTTTAACTCCAGCAGGACCATGTGATCGCGCTTC ATGCCGAGAGTGATCCCGGCGGCGGTCACGAACTCCAGCAGGACCATGTGATCGCGCTTC ** * * * * **** *
R3	>WR left >R3 left >YFP left	2	TGTTTATCGTATTGTATCCATCCATTGGCGCAAGCACTAGGCATCAGTTCTTCTTTGTAA TGTTTATCGTATTGTATCCATCCATTGGCGAAACAGGAAGGCAAAATGCCGCAAAAAAGG TTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGG * * ** * * * ** ***** * *
R5	>WR left >R5 left >YFP left	1	TCATCATTGTGTAAACTAAATAACGTCTTCTATGTATTTAGTAAGGACATTGGATATGTG TCATCATTGTGTAAACTAAATAACGTCTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTC CTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTC * * ** * * * * * * **

(continued next page)

Virus	Left junction	Overlap (bp)	Sequence alignment (60 bp)
R6	>WR left >R6 left >YFP left	0	AGTTATTAGATTTTCTAAACACATCATAGGCATTAGGACTAATAGCGTTAAGTACGTACC AGTTATTAGATTTTCTAAACACATCATAGGGCGAGTCAGTGAGCGAGGAAGCGGAAGAGC CCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGC * ** ** * * * * * *
R8	>WR left >R8 left >YFP left (30 bp)	0	GACGACGTTATGACTATCTATATATACCTTTCCAGTATACGAGTAAATAACTATAGAAGT GACGACGTTATGACTATCTATATATACCTTAAACCTGCAGGACTAGTCCCTTTAGTGAGG AAACCTGCAGGACTAGTCCCTTTAGTGAGG * * ** ** * **
RN2	>WR left >RN2 left >YFP left	0	ATGGCAAATCTAACTGCGGGTTTAGGCTTTAGTTTAGTTTTTATATCTACATCTATGTCT ATGGCAAATCTAACTGCGGGTTTAGGCTTTCTTCCCAGGAATTCTACGACTCACTATAGG ACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTCCCAGGAATTCTACGACTCACTATAGG * * * * * * * * * **
RN3	>WR left >RN3 left >YFP left	0	GTGCATCGTCTATTCTGCATCCGCTTCGTCGAATAATGTATAGCATCGCTTTGAGATTCT GTGCATCGTCTATTCTGCATCCGCTTCGTCCGGCCGCCGAGCTCCCCGGGGATCCGATAA GGAGACCGGCAGATCTTCCGCGGATGCATGCGGCCGCCGAGCTCCCCGGGGATCCGATAA * * ** * *** * * * * * * * * ** *
RN4	>WR left >RN4 left >YFP left (39 bp)	5	GCCATTTTTATGATCGTCTTTTGTGGAATTTCCCAAATAGTTTTATAAACTCGCTTAATA GCCATTTTTATGATCGTCTTTTGTGGAATTCGCCCTTCCCAGGAATTCTACGACTCACTA AAACGAATTCGCCCTTCCCAGGAATTCTACGACTCACTA ***** ** * ** * *
RN5	>WR left >RN5 left >YFP left	2	GTAGATGTTGTCCTTCTTCCTTTTGTTCGTGCCAATTGAGTAACATTATGAGAATATGAC GTAGATGTTGTCCTTCTTCCTTTTGTTCGTGAGTGAAGGCCGTCAAGGCCGCATGAATTC ATTTAGCGGCCGCGAATTCGCCCTTGAATTGAGTGAAGGCCGTCAAGGCCGCATGAATTC * * * *** * ** * * ** * * *
RN7 (WR-)	>WR YFP WR >WR YFP >YFP WR YFP	3	ATTGTCTGACGTCATCAATACTGGTCAGTCTGTCACATCCACAGTGCTGTGGTGAAAAGA ATTGTCTGACGTCATCAATACTGGTCAGTCTGGATGGAGGCGGATAAAGTTGCAGGACCA ACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCA * * * * *** * * ** *** * ** * * *

(continued next page)

Virus	Left junction	Overlap (bp)	Sequence alignment (60 bp)
RN8 (YFP-WR)	>WR YFP WR >RN8 YFP WR >YFP YFP WR (34	1	TGGTCCTATTGCTTCCATGCTACTAGTATAGATCAAATACTTGATTCCTAGGTCCACACA TAACCCTCACTAAAGGGACTAGTCCTGCAGGATCAAATACTTGATTCCTAGGTCCACACA TAACCCTCACTAAAGGGACTAGTCCTGCAGGTTT * *** * *
NH7	>WR left >NH7 left >YFP left	1	CTAACAACGACACCGACTAATGTTTCCGGAGATAGATTTCCAAAGATACACACATTAGGA CTAACAACGACACCGACTAATGTTTCCGGACCCAAGCGGCCGGAGAACCTGCGTGCAATC GTGCCCAGTCATAGCCGAATAGCCTCTCCACCCAAGCGGCCGGAGAACCTGCGTGCAATC * * * * ** * * ** * * *
RN11	>WR left >RN11 left >YFP left (30 bp)	0	GGATTATGAATAATTAATTCCTTTCCCGTTTGTGAATACTTAATGTCGTCGATATTAATT GGATTATGAATAATTAATTCCTTTCCCGTTAAACCTGCAGGACTAGTCCCTTTAGTGAGG AAACCTGCAGGACTAGTCCCTTTAGTGAGG * * * ** * *
RN13	>WR left >RN13 left >YFP left	1	GTGGAATTCAGTTGGAATATTGCCTCCGGAGTCATAAAAGGATACTAAACATTGTTTTTA GTGGAATTCAGTTGGAATATTGCCTCCGGAAGGCCGCATGAATTCATTTAGGTGACACTA ATTCGCCCTTGAATTGAGTGAAGGCCGTCAAGGCCGCATGAATTCATTTAGGTGACACTA * * * * * * * ** * ** *
RN15	>WR left >RN15 left >YFP left (35 bp)	3	TTAGATTTGGTCAAATTGTTTTTGGAAAATGGAAAGCCGCACGGAATAATGTGTAGTATT TTAGATTTGGTCAAATTGTTTTTGGAAAATAATTCGCCCTTCCCAGGAATTCTACGACTC AAACGAATTCGCCCTTCCCAGGAATTCTACGACTC *** *** * * *** * *
RN16	>WR left >RN16 left >YFP left	1	AATATTCATTATAATTTTCATAATCGCTAGATGAATAATATTGATAGTAACCGTTGTATC AATATTCATTATAATTTTCATAATCGCTAGAATGGCGTTTTTGATGTCATTTTCGCGGTG TGACATTTATATTCCCCAGAACATCAGGTTAATGGCGTTTTTGATGTCATTTTCGCGGTG * *** * *** * * ***** * *
S1-1	>WR left >s1-1 left >YFP left (30 bp)	0	TGATAACTCTGGCCATTTTGTCCTTGAAATTGGGAACATCTTCTTTGGATCTAATTGCGC TGATAACTCTGGCCATTTTGTCCTTGAAATAAACGAATTCGCCCTTCCCAGGAATTCTAC AAACGAATTCGCCCTTCCCAGGAATTCTAC * ** * ** **** *

(continued next page)

Virus	Right junction	Overlap (bp)	Sequence alignment (60 bp)
1-1-1-1	>WR right >1-1-1-1 right) >YFP right	1	AAAAAAGCTATTATCAAATACTGTACGGAATGGATTCATTCACTTCTCTTTTTATGAAAC GATAGTTTTTTTAATTAATAACTTCGTATATGGATTCATTCTCTTCTCTTTTTATGAAAC GATAGTTTTTTTAATTAATAACTTCGTATAATGTATGCTATACGAAGTTATCGGATCCCC * * * *** **** * * * * * * * * * * *
NH4	>WR right >NH4 right >YFP right (55 bp)	1	TACCGTATAATATGTAATATACAATTCCATAACCCATTGACAGTTGTTATACATCAAAAT TGCCGGTCTCCCTATAGTGAGTCGTAGAATAACCCATTGACAGTTGTTATACATCAAAAT TGCCGGTCTCCCTATAGTGAGTCGTAGAATTCCTGGGAAGGGCGAATTCGTTT *** * ** * * ** * **
NH5	>WR right >5 right >YFP right	1	GATACCGTAGTATATTGAGAGTGTATCCTTGATTATGTTTTATGAATAGATAAAGTAGAT CATCCGCGGAAGATCTGCCGGTCTCCCTATGATTATGTTTTATGAATAGATAAAGTAGAT CATCCGCGGAAGATCTGCCGGTCTCCCTATAGTGAGTCGTAGAATTCCTGGGAAGGGCGA ** * ** ** * * * * * * * ***
G5	>WR right >G5 right >YFP right	1	AGTAGTTTGACGACCTCGACATTATTACAATGCACATTAAAAAGGTATGCGTGTAAAGCA TCGCCGCAGCCGAACGACCGAGCGCAGCGATGCTCATTAAAAAGGTATGCGTGTAAAGCA TCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCC * *** * * * * * * * * * * *
G8	>WR right >G8 right >YFP right	0	TGGAAGCTACGTTTATCGATGTTCACATTCCCAAGTTTAAGGTAACAGGCTCGTATAATC CTTTTCTGGATTCATCGACTGTGGCCGGCTCCAAGTTTAAGGTAACAGGCTCGTATAATC CTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGC * *** * ** * * * * * * *
R3	>WR right >R3 right >YFP right	0	TATTGATGACACTAGTTAAGAACATGTCTTCTATTCTAGAAAACGAAAACATCCTACATG AGCTTGGCGGCGAATGGGCTGACCGCTTCCCTATTCTAGAAAACGAAAACATCCTACATG AGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATT * * ** * * ** * * * **
R5	>WR right >R5 right >YFP right	0	GAAAGTAGTATAAAAGTAATAAAACAAAAAAAAGAATATAAAAAATTTATAGCTACTTTC TTTTCCAATGATGAGCACTTTTAAAGTTCTAAAGAATATAAAAAATTTATAGCTACTTTC TTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGA * * * ** * **** * **

(continued next page)

Virus	Right junction	Overlap (bp)	Sequence alignment (60 bp)
R6	>WR right >R6 right) >YFP right	1	TCTATAGTCAGAATTTTCATGGATGTAGATTTAGACGCGTGTCTAGACGAAATAGATTAT GCGATAAGTCGTGTCTTACCGGGTTGGACTTTAGACGCGTGTCTAGACGAAATAGATTAT GCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGC *** * ** ** * * ***** ** **
R8	>WR right >R8 right >YFP right (34 bp)	0	ATCTTGTACGTATATTGCATGGAATCATAGATGGCCTTTTCAGTTGAACTGGTAGCCTGT AGTCGTAGAATTCCTGGGAAGGGCGAATTCATGGCCTTTTCAGTTGAACTGGTAGCCTGT AGTCGTAGAATTCCTGGGAAGGGCGAATTCGTTT * * * * * ** ** *
RN2	>WR right >RN2 right >YFP right	2	TTGCTGTTATTAATGGATTTAGAAATAGTGGACGATTTTGTGATATTAGTATAGTTATTA AACAGGAGAGCGCACGAGGGAGCTTCCAGGGACGATTTTGTGATATTAGTATAGTTATTA AACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGT * * ** ** ** * ** *** ****** *
RN3	>WR right >RN3 right >YFP right	3	CCTTCAAGACGGGACGAAGAAACGTAATATCCTCAATAACGTTATCGTTTTCTACAATAA TCGCGAAGGCCTACTAGTCTCGAGTTATATCCTCAATAACGTTATCGTTTTCTACAATAA TCGCGAAGGCCTACTAGTCTCGAGTTCTATAGTGTCACCTAAATGAATTCATGCGGCCTT * *** * ** *** * **
RN4	>WR right >RN4 right >YFP right	0	CTTTTCATCCAGTTGAATGGATTCGTCCTTAACCAACTGATTAATGAGATCTTCTATTTT ATTTATATTCCAAAAAAAAAAAAATAAATAACATCAACTGATTAATGAGATCTTCTATTTT ATTTATATTCCAAAAAAAAAAAATAAAATTTCAATTTTTGGCGCGCCATAACTTCGTATA *** ** * ** * * * * *
RN5	>WR right >RN5 right >YFP right (length unknown)	5	TCCTATCGTAGGCGATAGAACCGCTAAAAAGCCTATCGAATTTCTACAAAAGAATCTGTT AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCCTATCGAATTTCTACAAAAGAATCTGTT AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA * * * * ** ***** * ** * **** **
RN7 (YFP-YFP)	>YFP YFP (35 bp) >YFP YFP >YFP YFP right	1	TTAACCCTCACTAAAGGGACTAGTCCTGCAGGTTT TTAACCCTCACTAAAGGGACTAGTCCTGCAGACAGTAAGAGAATTATGCAGTGCTGCCATA AGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATA * * * * *

(continued next page)

Virus	Right junction	Overlap (bp)	Sequence alignment (60 bp)
RN8 (YFP-YFP)	>middle left >RN8 YFP YFP >YFP right (30 bp)	0	GCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGGTCTAGACGGATT GCCGGGATCACTCTCGGCATGGACGAGCTGAAACGAATTCGCCCTTCCCAGGAATTCTAC AAACGAATTCGCCCTTCCCAGGAATTCTAC * * * *
NH7	>WR right >NH7 right >YFP right	5	TATAATACTTAAAAATGGATGTTGTGTCGTTAGATAAACCGTTTATGTATTTTGAGGAAA TCGTAGAATTCCTGGGAAGGGCGAATTCGTTAGATAAACCGTTTATGTATTTTGAGGAAA TCGTAGAATTCCTGGGAAGGGCGAATTCGTTTAAACCTGCAGGACTAGTCCCTTTAGTGA * * * ** * ***** * * * * * * *
RN11	>WR right >RN11 right >YFP right (31 bp)	1	ATGTAGACAAATATACTTTATCAAATTGCATACCTATAGGAATAGTCTCTGTAATCACTG CGTAGAATTCCTGGGAAGGGCGAATTCGTTTACCTATAGGAATAGTCTCTGTAATCACTG CGTAGAATTCCTGGGAAGGGCGAATTCGTTT * * ** * * *
RN13	>WR right >RN13 right >YFP right	0	TATATTTATTCCACACACTCTACAATATGCCACACCATCTTCATAATAAATAAATTCGTT TCCCCGGGGAGCTCGGCGGCCGCATGCATGCACACCATCTTCATAATAAATAAATTCGTT TCCCCGGGGAGCTCGGCGGCCGCATGCATCCGCGGAAGATCTGCCGGTCTCCCTATAGTG * * * * ** * * * * * * **
RN15	>WR right >RN15 right >YFP right (30 bp)	2	AACTCAAATGTTATGGCCATAGAGGATGTTATTACGAATCATTAAAAAAATTAACTGAGG CCTCACTAAAGGGACTAGTCCTGCAGGTTATTACGAATCATTAAAAAAATTAACTGAGG CCTCACTAAAGGGACTAGTCCTGCAGGTTT * **
RN16	>WR right >RN16 right >YFP right (30 bp)	1	ACTGACTGGATAAGCTATAATAATAAATGTATCCATTTATCTACTGATCGAAAAACCTGG GTAGAATTCCTGGGAAGGGCGAATTCGTTTATCCATTTATCTACTGATCGAAAAACCTGG GTAGAATTCCTGGGAAGGGCGAATTCGTTT ** * * * * *** * *
S1-1	>WR right >s1-1 right >YFP right (41 bp)	1	CATATCATATTCTTCGTCATCGATTGTGATTGTATTTCCTTGCAATTTAGTAACTACGTT CTCAGAATTAACCCTCACTAAAGGGACTAGTGTATTTCCTTGCAATTTAGTAACTACGTT CTCAGAATTAACCCTCACTAAAGGGACTAGTCCTGCAGGTTT * * ** * * * **

Key: Original VAC sequence, transfected DNA, homologous junction sequence, canonical topoisomerase site, inverted repeats, non-templated DNA