

https://dspace.org/

# The effects of in vivo passaging, solar radiation exposure and inoculum dose on the genetic diversity of Helicoverpa armigera nucleopolyhedrovirus

**Kitchin, Dale John**

**2017**

https://hdl.handle.net/10539/26508

Kitchin, Dale John, (2017) The effects of in vivo passaging, solar radiation exposure and inoculum dose on the genetic diversity of Helicoverpa armigera nucleopolyhedrovirus, University of the Witwatersrand, Johannesburg, https://hdl.handle.net/10539/26508

Downloaded from a DSpace repository. https://dspace.org/

# The effects of *in vivo* passaging, solar radiation exposure and inoculum dose on the genetic diversity of *Helicoverpa armigera* nucleopolyhedrovirus

Dale John Kitchin

School of Molecular and Cell Biology
University of the Witwatersrand
Johannesburg, South Africa

A thesis submitted to the Faculty of Science, University of the Witwatersrand, in fulfilment of the requirements for the degree of Doctor of Philosophy.

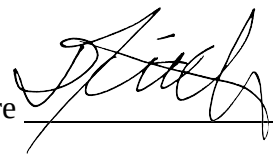
Johannesburg, June 2017

# DECLARATION

I, Dale John Kitchin (0611622Y), am a student registered for the degree of Doctor of Philosophy in the academic year 2017.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise and acknowledged.
- I have not submitted this work before for any other degree or examination at this or any other University.
- The information used in the Thesis/Dissertation/Research Report has not been obtained by me while employed by or working under the aegis of, any person or organisation other than the University.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signature  5th day of June 2017

# ABSTRACT

The nucleopolyhedroviruses (NPVs) are a group of entomopathogenic viruses, with large double-stranded DNA genomes, which belong to the *Baculoviridae* family. Considerable genetic diversity is known to exist within NPV populations, with a number of studies having shown that NPV populations consist of multiple genotypes. Studies have also shown that the genotypic composition of NPV populations can vary between different host individuals and geographical locations. Yet, how this genetic diversity is generated and maintained in NPV populations over successive infections is not fully understood. In this study, a denaturing gradient gel electrophoresis (DGGE) assay and deep amplicon sequencing were used to assess the effects of serial *in vivo* passaging, pre-infection exposure to solar radiation and occlusion derived virion (ODV) inoculum dose on the genetic diversity of *Helicoverpa armigera* nucleopolyhedrovirus (HearNPV) populations within regions of the *DNA polymerase* and *me53* genes. The serial *in vivo* passaging of a HearNPV population, using a high OB inoculum dose ($LD_{90} \times 5$), did not result in a common or similar genotypic composition in the 125 third passage HearNPV populations isolated from individual host insects. Shifts in the relative abundance of inoculum genetic variants and the generation of novel genetic variants caused the HearNPV populations isolated from individual insects to diverge in genetic variant composition over the three serial *in vivo* passages. Although pre-infection exposure of HearNPV OBs to solar radiation did not increase the frequency at which novel genetic variants were generated over the course of an infection, it did appear to contribute to the divergence in genetic variant composition of HearNPV populations over *in vivo* passages. It was also observed that the higher the ODV inoculum dose the greater the probability that the genetic variant composition of the post-passage HearNPV population would resemble that of the inoculum population. Overall it was shown that *in vivo* passaging, pre-infection exposure of OBs to solar radiation and ODV inoculum dose all affect the genotypic composition of HearNPV populations. The findings of this study contribute to the understanding of the factors that affect the generation and maintenance of HearNPV genetic diversity over the course of an infection and provide insights into how the genetic diversity between distant geographical HearNPV isolates may develop in the field.

# TABLE OF CONTENTS

DECLARATION ..... II

ABSTRACT ..... III

TABLE OF CONTENTS ..... IV

LIST OF TABLES ..... IX

LIST OF FIGURES ..... X

LIST OF ABBREVIATIONS ..... XIII

ACKNOWLEDGMENTS ..... XV

Chapter 1 ..... 1

General Introduction ..... 1

1.1. The *Baculoviridae* ..... 1

1.1.1. Defining the *Baculoviridae* ..... 1
1.1.2. Baculovirus taxonomy and phylogenetics ..... 2
1.1.3. Core baculovirus genes ..... 3
1.1.4. Nucleopolyhedrovirus replication cycle ..... 4
1.1.5. *Helicoverpa armigera* nucleopolyhedrovirus (HearNPV) ..... 5

1.2. Baculovirus genetic diversity ..... 7

1.2.1. The genotypic variation of baculovirus populations ..... 7
1.2.2. Baculovirus genetic diversity: atypical for dsDNA viruses with large genomes? ..... 8
1.2.3. Factors that affect the level of baculovirus genetic diversity ..... 9

1.3. Techniques used to study baculovirus genetic diversity ..... 15

1.3.1. Well-established techniques for detecting baculovirus genotypic variants .....15
1.3.2. Newer methods for studying baculovirus genetic diversity .....18

1.4. Study justification .....22

1.5. Objectives and aims .....24

1.5.1. Overall project objective .....24
1.5.2. Specific research aims .....24

Chapter 2 ..... 27

Evaluation of the use and reproducibility of PCR-DGGE to assess genetic diversity in *Helicoverpa armigera* nucleopolyhedrovirus populations. 27

2.1. Abstract .....27

2.2. Introduction .....28

2.3. Methods .....32

2.3.1. Insect rearing .....32
2.3.2. Virus propagation and purification .....32
2.3.3. Extraction of viral DNA .....33
2.3.4. PCR amplification for DGGE .....33
2.3.5. DGGE conditions .....34
2.3.6. Analysis of DGGE profiles .....35
2.3.7. DGGE band extraction and reamplification .....36
2.3.8. Cloning of *me53* and *DNA polymerase* amplicons reamplified from DGGE bands...36
2.3.9. "Reconditioning" PCR .....37
2.3.10. Sanger dideoxy sequencing .....37
2.3.11. Deep amplicon sequencing .....38
2.3.12. Deep sequencing data analysis .....39

2.4. Results .....41

2.4.1. Reproducibility of DGGE profiles .....41
2.4.2. Multiple sequences associated with individual bands .....41
2.4.3. Origins of DGGE profile artefacts .....43
2.4.4. Deep *me53* amplicon sequencing .....46

2.5. Discussion .....48

2.6. Conclusions .....52

Chapter 3 ..... 64

The effect of serial *in vivo* passages on the genetic diversity of two core *Helicoverpa armigera* nucleopolyhedrovirus genes ..... 64

3.1. Abstract ..... 64

3.2. Introduction ..... 65

3.3. Materials and methods ..... 68

3.3.1. Insect rearing ..... 68
3.3.2. Virus propagation and purification ..... 68
3.3.3. Dose-mortality assay ..... 68
3.3.4. Serial *in vivo* passaging ..... 68
3.3.5. DNA extraction ..... 69
3.3.6. PCR-DGGE ..... 70
3.3.7. Deep amplicon sequencing ..... 70

3.4. Results ..... 72

3.4.1. DGGE profile analysis ..... 72
3.4.2. Deep *me53* amplicon sequencing ..... 74

3.5. Discussion ..... 78

3.6. Conclusions ..... 82

Chapter 4 ..... 92

Evaluation of the effect of solar radiation on the genetic diversity of *Helicoverpa armigera* nucleopolyhedrovirus populations ..... 92

4.1. Abstract ..... 92

4.2. Introduction ..... 93

4.3. Materials and methods ..... 96

4.3.1. Insect rearing ..... 96
4.3.2. Virus propagation ..... 96

4.3.3. Exposure of HearNPV to solar radiation....96
4.3.4. Post-exposure HearNPV *in vivo* passage....97
4.3.5. DNA extraction....98
4.3.6. PCR-DGGE....98
4.3.7. Deep amplicon sequencing....98

4.4. Results....100

4.4.1. DGGE profile analysis....100
4.4.2. Deep amplicon sequencing....102

4.5. Discussion....105

4.6. Conclusions....109

Chapter 5 .... 116

The effect of ODV inoculum dose on genetic variation within *Helicoverpa armigera* nucleopolyhedrovirus populations .... 116

5.1. Abstract....116

5.2. Introduction....117

5.3. Materials and methods....119

5.3.1. Insect rearing....119
5.3.2. Virus propagation and purification....119
5.3.3. Preparation of ODVs....119
5.3.4. Inoculation of neonate larvae....120
5.3.5. DNA extraction....120
5.3.6. PCR-DGGE....121
5.3.7. Deep amplicon sequencing....121

5.4. Results....122

5.4.1. Structural integrity and infectivity of ODVs....122
5.4.2. DGGE profile analysis....122
5.4.3. Deep amplicon sequencing....124

5.5. Discussion....126

5.6. Conclusions....129

Chapter 6 ..... 136

General discussion and conclusion ..... 136

6.1. General discussion .....136

6.2. Conclusion .....143

References ..... 144

# LIST OF TABLES

**Table 1.1.** Core genes present within all baculovirus and lepidopteran-specific baculovirus genomes sequenced to date. ..... 25

**Table 2.1.** Sequences and characteristics of the primers specific to the HearNPV *me53* and *DNA polymerase* genes, used for PCR-DGGE, Sanger sequencing and MiSeq library preparation. . 53

**Table 2.2.** The true relative abundance of genetic variants mV1, mV2 and mV5 in the two mock populations and their observed relative abundance in the corresponding deep amplicon sequencing data. ..... 54

**Table 3.1.** The results of the PERMANOVA and PERMDIST analyses of *DNA polymerase* DGGE profiles from the different passage groups, after the second and third passages, respectively. . 83

**Table 3.2.** The results of the PERMANOVA and PERMDIST analyses of *me53* DGGE profiles from the different passage groups, after the second and third passages, respectively. ..... 83

**Table 4.1.** The results of the PERMANOVA and PERMDIST analyses of *DNA polymerase* and *me53* DGGE profiles from the three solar UVR exposure groups..... 110

**Table 5.1.** The empirically determined ODV lethal doses and the corresponding dilutions of the neat ODV preparation. ..... 131

**Table 5.2.** The results of the PERMANOVA and PERMDIST analyses of *me53* DGGE profiles from the different ODV inoculum groups ($LD_5$-$LD_{95}$). ..... 131

# LIST OF FIGURES

**Figure 1.1.** The generic biphasic replication cycle of the nucleopolyhedroviruses.. .....26

**Figure 2.1.** Methods flowchart showing the steps taken to assess the reproducibility and the prevalence of artefacts in PCR-DGGE gel profiles, when the technique is used to assess the genetic diversity of the *me53* and *DNA polymerase* genes in HearNPV populations. .....55

**Figure 2.2.** Similarity of *me53* and *DNA polymerase* DGGE profiles across PCR and gel replicates. .....56

**Figure 2.3.** DGGE profiles for *me53* amplicons originating from the reamplification of DNA extracted from six bands in the *me53* profile of HearNPV population P4. .....57

**Figure 2.4.** Nucleotide sequences of *me53* variants. .....58

**Figure 2.5.** The DGGE profiles for the *me53* clones originally derived from the six bands (mB1-mB6) in the P4 *me53* DGGE profile. .....59

**Figure 2.6.** DGGE profiles and nucleotide sequences of the *DNA polymerase* clone variants. .....60

**Figure 2.7.** Artefactual bands and reconditioning reactions for *me53* DGGE profiles. .....61

**Figure 2.8.** Artefactual bands and reconditioning reactions for *DNA polymerase* DGGE profiles. .62

**Figure 2.9.** Differences in the relative intensities of bands mB1-mB4 between *me53* DGGE profiles for multi-template reamplification reactions, where different molar ratios of two *me53* clone variants were used as template. .....63

**Figure 3.1.** The overall experimental design for the serial *in vivo* passages of HearNPV in third instar *H. armigera* larvae. .....84

**Figure 3.2.** Cluster analysis of the *DNA polymerase* DGGE profiles of HearNPV populations isolated from third instar *H. armigera* larvae inoculated with an $LD_{90} \times 5$ over three serial *in vivo* passages. .....86

**Figure 3.3.** Cluster analysis of the *me53* DGGE profiles of HearNPV populations isolated from third instar *H. armigera* larvae inoculated with an $LD_{90} \times 5$ over three serial *in vivo* passages. .....88

**Figure 3.4.** Examples of shifts in the band composition of DGGE profiles that were observed after successive *in vivo* passages .....89

**Figure 3.5.** The *me53* variants detected using deep amplicon sequencing in HearNPV populations where PCR-DGGE detected significant shifts in *me53* variant composition over a passage lineage. .....90

**Figure 3.6.** Highlighter plot showing the nucleotide sequences of the *me53* variants relative to a reference HearNPV *me53* gene sequence (GenBank accession number: AF271059).....91

**Figure 4.1.** Cluster analysis of the *DNA polymerase* DGGE profiles of HearNPV populations isolated from third instar *H. armigera* larvae inoculated with HearNPV exposed to either 0 h, 12 h or 21 h of solar radiation. .....111

**Figure 4.2.** Cluster analysis of the *me53* DGGE profiles of HearNPV populations isolated from third instar *H. armigera* larvae inoculated with HearNPV exposed to either 0 h, 12 h or 21 h of solar radiation.....112

**Figure 4.3.** Non-metric multidimensional scaling (NMDS) analyses of *DNA polymerase* and *me53* DGGE profiles for HearNPV populations from the three solar UVR exposure groups. .....113

**Figure 4.5.** Highlighter plot showing the nucleotide sequences of the *me53* variants relative to a reference HearNPV *me53* gene sequence (GenBank accession number: AF271059).....115

**Figure 5.1.** TEM micrographs showing ODVs in the HearNPV sample after the alkaline-dissolution process.....132

**Figure 5.2.** Cluster analysis of the *me53* DGGE profiles for the HearNPV inoculum population and populations isolated from *H. armigera* neonate larvae inoculated with one of the five HearNPV ODV doses ($LD_5$, $LD_{25}$, $LD_{50}$, $LD_{75}$ and $LD_{95}$). .....133

**Figure 5.3.** The *me53* variants detected using deep amplicon sequencing in a subset of HearNPV populations isolated from larvae inoculated with the different ODV doses. .....134

**Figure 5.4.** Highlighter plot showing the nucleotide sequences of the *me53* variants relative to a reference.....135

# LIST OF ABBREVIATIONS

| | |
|---|---|
| ANOVA | analysis of variance |
| BV | budded virus |
| CP | common progenitor |
| dB | *DNA polymerase* DGGE band |
| DGGE | denaturing gradient gel electrophoresis |
| DNA | deoxyribonucleic acid |
| dP | *DNA polymerase* DGGE profile type |
| dsDNA | double stranded DNA |
| dV | *DNA polymerase* genetic variant |
| GV | granulovirus |
| HearNPV | *Helicoverpa armigera* nucleopolyhedrovirus |
| HTS | high-throughput sequencing |
| IAH | independent action hypothesis |
| LD | lethal dose |
| mB | *me53* DGGE band |
| MNPV | multiple nucleopolyhedrovirus |
| MOI | multiplicity of infection |
| mP | *me53* DGGE profile type |
| mV | *me53* genetic variant |

| | |
|---|---|
| MWCO | molecular weight cut off |
| NGS | next generation sequencing |
| NMDS | non-metric multidimensional scaling |
| NPV | nucleopolyhedrovirus |
| OB | occlusion body |
| ODV | occlusion derived virion |
| PCR | polymerase chain reaction |
| PG | passage group |
| PIF | *per os* infectivity factor |
| REN | restriction endonuclease |
| RFLP | restriction fragment length polymorphism |
| RNA | ribonucleic acid |
| SDS | sodium dodecyl sulphate |
| SNP | single nucleotide polymorphism |
| SNPV | single nucleopolyhedrovirus |
| TEM | transmission electron microscope |
| UPGMA | unweighted pair group method with arithmetic mean |
| UVR | ultraviolet radiation |

# ACKNOWLEDGMENTS

I would like to thank the following people and institutions for their support and assistance, which allowed me the opportunity to start and ultimately finish this study and thesis:

1. My supervisor, Prof. Gustav Bouwer, for the many hours of discussion, during which he provided me with invaluable guidance on both technical and theoretical aspects related to this study. I'd also like to thank him for his detailed and thorough critique of my draft chapters which I feel greatly improved the quality of this thesis and my scientific writing in general.
2. Dr Arshad Ismael and Dr Cathrine Scheepers at the National Institute for Communicable Diseases for their guidance and advice with regards to the deep amplicon sequencing component of this study.
3. Prof. Penny Moore for initially encouraging me to pursue a PhD and continuing to provide words of advice and encouragement throughout the duration of my studies.
4. The University of the Witwatersrand's postgraduate merit award (Wits PGMA) as well as the National Research Foundation (NRF) and German Academic Exchange Service (DAAD) for generous financial assistance during my studies.
5. My parents, Alan and Louise, and sister, Ruth, for endless amounts of encouragement and support, especially when the "bugs" were getting me down.
6. Natasha Abraham, for her belief in me that I'd get it done…eventually (And sticking it out with me till then!).
7. Alice Martin, Mark Mthombeni, Kelvin Tam, Jonathan Linde, Warren Baas, Premal Bhima and Michael Weston, for forgiving me for declining all their invitations to go climbing/sailing/cycling and a number of other exciting things, so that I could spend more time with my "butterflies".
8. All the people in Room 207 over the years, especially Maabo Moralo and Imah Mwaba, for telling me to "JBS" and, of course, all the Friday afternoons spent at the PIG.

# Chapter 1

# General Introduction

## 1.1. The *Baculoviridae*

### 1.1.1. Defining the *Baculoviridae*

The *Baculoviridae* family consists of a diverse group of arthropod-specific viruses which infect members of a number of invertebrate genera, predominantly in the insect orders Lepidoptera, Hymenoptera and Diptera (Herniou *et al.*, 2003; Rohrmann, 2013). All baculoviruses have circular, supercoiled, double-stranded DNA (dsDNA) genomes. These range from around 80 to over 180 kbp in size and are packaged in enveloped, rod-shaped nucleocapsids (Jehle *et al.*, 2006b; Rohrmann, 2013). Another defining characteristic is that they encode their own RNA polymerase enzyme (Fuchs *et al.*, 1983; Guarino *et al.*, 1998; Herniou *et al.*, 2003). This is especially significant for certain baculoviruses, known as nucleopolyhedroviruses, which replicate in the nucleus, as it is unusual for a DNA virus which replicates in the nucleus to not rely entirely on the host cell's transcriptional apparatus (Rohrmann, 2013).

In addition to these features, the characteristics which are most associated with baculoviruses are their unusual biphasic replication cycle and the occlusion of virions within a proteinaceous crystal matrix, known as an occlusion body (OB) (Jehle *et al.*, 2006b; Rohrmann, 2013). In the biphasic replication cycle, two genetically identical virion phenotypes, occlusion derived virions (ODVs) and budded virions (BVs), are responsible for initiating an infection in a susceptible insect's midgut epithelial cells and spreading the infection from the midgut throughout the host's tissues, respectively (Jehle *et al.*, 2006b; Okano *et al.*, 2006; Rohrmann, 2013). The importance of the study of baculoviruses is a result of their use as protein expression vectors in both insect and mammalian cell culture systems (reviewed in Kost *et al.*, 2005), and as biological control agents against a number of agricultural insect pests (reviewed in Haase *et al.*, 2015; Moscardi *et al.*, 2011; Szewczyk *et al.*, 2005).

### 1.1.2. Baculovirus taxonomy and phylogenetics

Viruses belonging to the *Baculoviridae* family were originally divided into two main groups which subsequently became the *Nucleopolyhedrovirus* (NPV) and *Granulovirus* (GV) genera (Herniou *et al.*, 2001; Jehle *et al.*, 2006a). This classification system was largely based on the morphology and structure of OBs and ODVs (Herniou *et al.*, 2003). The OBs of NPVs (0.6-2 µm in diameter) are polyhedral-shaped, contain multiple virions and tend to be larger than those of GVs (0.2-0.4 µm in diameter) which only contain a single virion and are ellipsoidal-shaped (Jehle *et al.*, 2006b; Pearson & Rohrmann, 2002). NPV ODVs consist of either a single nucleocapsid (SNPVs) or one or more nucleocapsid aggregates (MNPVs) within an envelope, with all NPV isolates having a distinctive SNPV or MNPV phenotype (Rohrmann, 2013, 2014); while GVs typically have singly enveloped nucleocapsids (Pearson & Rohrmann, 2002). The major protein component of OBs also differs between the two genera; the crystalline matrix is composed of a protein called polyhedrin in NPVs and granulin in GVs (Bideshi *et al.*, 2000; Herniou *et al.*, 2003; Zanotto *et al.*, 1993).

Phylogenetic studies using data generated from the sequencing of 29 baculovirus genomes, and the molecular characterisation of some of these viruses, led to a review of the older two-genera classification of baculoviruses (Herniou *et al.*, 2011; Jehle *et al.*, 2006b). These studies indicated that the phylogenetic distance between lepidoptera NPVs and GVs is closer than between lepidopteran, dipteran and hymenoptera NPVs. These findings were supported by a later study by Wu *et al.* (2009), who constructed the whole-proteome phylogeny for 142 large double-stranded DNA (dsDNA) eukaryote viruses from 11 virus families, including members of the *Baculoviridae.* These findings support a host-dependent process for baculoviruses evolution where a virus becomes genetically isolated and speciates along with its associated host (Jehle *et al.*, 2006b). Phylogenetic studies have also shown that the SNPV and MNPV nucleocapsid phenotype does not follow the phylogeny of the *Baculoviridae* (Herniou *et al.*, 2003; Zanotto *et al.*, 1993). However lepidoptera NPVs can be divided into two phylogenetic groups based on the envelope fusion protein utilised during the cell to cell spread of viruses. Group I NPVs utilise a GP46 envelope fusion protein while group II NPVs make use of a furin-cleaved envelope protein (F protein) for the same purpose (Pearson *et al.*, 2000; Pearson & Rohrmann, 2002).

Based on the phylogenetic and molecular evidence a new classification system was proposed and later adopted for the *Baculoviridae,* where the family was split into four genera; *Alphabaculovirus, Betabaculovirus, Gammabaculovirus* and *Deltabaculovirus* (Herniou *et al.,* 2011; Jehle *et al.,* 2006a). The genus *Alphabaculovirus* includes all lepidopteran specific NPVs, from both group I and II, while the *Betabaculovirus* genus consists of the existing lepidopteran-specific *Granulovirus* genus (Herniou *et al.,* 2011; Jehle *et al.,* 2006b). The *Gammabaculovirus* and *Deltabaculovirus* genera consist of hymenopteran- and dipteran-specific NPVs, respectively (Herniou *et al.,* 2011; Jehle *et al.,* 2006b).

### 1.1.3. Core baculovirus genes

There are a group of 37 core genes which have been found in all baculovirus genomes sequenced to date (Garavaglia *et al.,* 2012; Herniou & Jehle, 2007; Jehle *et al.,* 2006b). These genes are considered orthologous as they are thought to have been present in the last common ancestor of the baculoviruses and to have diverged through speciation events (Garavaglia *et al.,* 2012). Phylogenetic analysis of these core genes was used to infer the phylogeny of members of the *Baculoviridae* family, and this was the basis of the new taxonomic organisation of baculoviruses into the *Alphabaculovirus, Betabaculovirus, Gammabaculovirus* and *Deltabaculovirus* genera (Herniou *et al.,* 2011; Herniou & Jehle, 2007; Jehle *et al.,* 2006b). These genes are believed to encode products which are crucial for baculovirus DNA replication, gene transcription, virion structure, *per os* infectivity and several auxiliary processes required for successful baculovirus replication (Garavaglia *et al.,* 2012; Herniou & Jehle, 2007). However the functions of several core gene products remain unknown (Garavaglia *et al.,* 2012). A group of 26 additional core genes have also been identified, which occur in all lepidopteran-specific baculovirus genomes sequenced thus far (Garavaglia *et al.,* 2012; van Oers & Vlak, 2007). These additional core genes are thought to be essential for the replication and infectivity of members of the *Alphabaculovirus* and *Betabaculovirus* genera in their lepidopteran hosts (Garavaglia *et al.,* 2012; van Oers & Vlak, 2007). The 37 core baculovirus genes and the 26 additional core lepidopteran-specific baculovirus genes are listed in Table 1.1.

Baillie & Bouwer (2012a) have previously shown that in *Helicoverpa armigera NPV* (HearNPV) populations there is a significant amount of genetic variation within regions of a core baculovirus gene, *DNA polymerase*, and two core lepidopteran-specific baculovirus

genes, *me53* and *dbp1*. In the present study the effects of successive *in vivo* passaging, solar radiation and variable inoculum dose on the level of genetic variation within regions of the *DNA polymerase* and *me53* genes were assessed. The *DNA polymerase* gene encodes a DNA dependent DNA polymerase which has been shown to be essential for the replication of baculovirus DNA during an infection (Vanarsdall *et al.*, 2005). The *me53* gene is thought to be involved in the regulation of viral DNA replication, as it encodes a protein with a putative zinc finger DNA-binding motif (Knebel-Mörsdorf *et al.*, 1993; Xi *et al.*, 2007). In addition the *me53* gene product has been implicated in the efficient budding of BVs from infected cell membranes (de Jong *et al.*, 2011).

### 1.1.4. Nucleopolyhedrovirus replication cycle

Lepidopteran nucleopolyhedroviruses (members of the *Alphabaculovirus* genus) follow a biphasic replication cycle which is a characteristic of the *Baculoviridae*. OBs, also called polyhedra, transmit virions between insect hosts. OBs consist of a number of virions occluded within a crystalline polyhedrin protein matrix, and are surrounded by an outer envelope that forms a smooth, seamless outer surface (Rohrmann, 2013; Theilmann & Blissard, 2008). OBs allow virions to remain stable and persist in the environment for several years, in the absence of suitable host insects (Theilmann & Blissard, 2008). The two genetically identical virion phenotypes, ODVs and BVs, are responsible for different stages of the biphasic replication cycle (Okano *et al.*, 2006; Theilmann & Blissard, 2008). ODVs are released upon the dissolution of OBS in the midgut of a potential insect host and initiate infection in the midgut epithelial cells (Rohrmann, 2013; Theilmann & Blissard, 2008). BVs are responsible for the spread of an infection from the midgut to the rest of a host's tissues through cell-to-cell transmission (Rohrmann, 2013; Theilmann & Blissard, 2008). The biphasic replication cycle of the NPVs is summarised in Figure 1.1.

Both replication and transcription of the NPV genome occurs in an infected cell's nucleus. Genes are expressed in four transcriptional phases: immediate early, delayed early, late and very late (Okano *et al.*, 2006). The early gene products include viral transcription factors which are necessary for the transcription of delayed early genes (van Oers & Vlak, 2007; Szewczyk *et al.*, 2005). Many of the products of the delayed early genes are required for viral DNA replication (Okano *et al.*, 2006). Late genes are those expressed after the initiation of DNA replication and their products include nucleocapsid structural proteins and the envelope

fusion proteins GP64 and F (Rohrmann, 2013). The very late gene, encoding the polyhedrin protein, is only expressed at very high levels during the final stages of an infection, just prior to the occlusion of virions within the crystalline protein matrix (Okano *et al.*, 2006).

The mechanism of baculovirus genome replication is complex and has yet to be fully elucidated (reviewed in Carstens, 2009 & Vanarsdall *et al.*, 2007). There appear to be many regions in baculovirus genomes that serve as origins (*ori*) of replication including homologous repeated sequences (*hrs*) and early promoter sequences (Okano *et al.*, 2006). DNA replication may be initiated at a number of these sites in the genome simultaneously and most likely proceeds by a rolling-circle mechanism (Leisy & Rohrmann, 1993; Mikhaĭlov, 2003). Replication appears to result in the production of intermediate sub-genomic DNA fragments (Okano *et al.*, 2004, 2006). Viral proteins, including an alkaline nuclease, are involved in processing these fragments to produce mature closed circular dsDNA genomes (Okano *et al.*, 2004, 2006). Evidence suggests that this process of genome maturation is dependent on extensive homologous recombination occurring between sub-genomic fragments, ultimately resulting in the production of full size NPV genomes (Okano *et al.*, 2004). This recombination-dependent mechanism of DNA replication has been implicated in the generation of genotypic variation within baculovirus populations (Hajós *et al.*, 2000; Okano *et al.*, 2006).

### 1.1.5. *Helicoverpa armigera* nucleopolyhedrovirus (HearNPV)

*Helicoverpa armigera* nucleopolyhedrovirus (HearNPV) is a single-nucleocapsid group II NPV, and a member of the *Alphabaculovirus* genus (Herniou & Jehle, 2007; Jehle *et al.*, 2006b; Okano *et al.*, 2006). HearNPV selectively infects the African Bollworm, *Helicoverpa armigera* Hübner (Lepidoptera: Noctuidae) and closely related lepidopteran species (Chen *et al.*, 2002). *H. armigera* is a highly polyphagous and economically important agricultural pest internationally. Both wild-type and recombinant forms of HearNPV have been used as biological control agents for *H. armigera* in several countries (Chen *et al.*, 2000; Cherry *et al.*, 2000; Sun *et al.*, 2004).

HearNPV is a relatively well characterised NPV, as the whole genome of a number of geographically disparate isolates have been sequenced (Baillie & Bouwer, 2012a; Chen *et al.*, 2001, 2002, Noune & Hauxwell, 2015, 2016) and other studies have focused on various

aspects of its biology, genomics and development as an effective biopesticide (Huang *et al.*, 2012; Long *et al.*, 2006; Peng *et al.*, 2010; Zhang *et al.*, 2008). Of particular relevance for the current study is the previous development and application of a novel denaturing gradient gel electrophoresis (DGGE) assay by Baillie & Bouwer (2011, 2012a, b, 2013), to assess the genetic variation within regions of core genes in HearNPV populations. This existing knowledge base makes HearNPV an attractive model system for the purposes of evaluating the effects of host and environmental factors on the genetic diversity of NPV populations.

## 1.2. Baculovirus genetic diversity

### 1.2.1. The genotypic variation of baculovirus populations

It has been well documented that a high level of genetic diversity exists within baculovirus populations (Baillie & Bouwer, 2012b; Chateigner *et al.*, 2015; Cory *et al.*, 2005; Gettig & Mccarthy, 1982; Patel *et al.*, 2009b; Thézé *et al.*, 2014). The use of several molecular techniques have shown that this genotypic variation exists within baculovirus populations at a number of spatial levels; between populations from the same host species at different geographic locations (Cooper *et al.*, 2003; Graham *et al.*, 2004), between populations from different host individuals at the same site (Shapiro *et al.*, 1991) and even within populations isolated from individual hosts (Baillie & Bouwer, 2012b; Cory *et al.*, 2005; Hodgson *et al.*, 2001). It is thought that the level of genetic diversity in baculovirus populations may vary over time in response to changes in the environment, or phenotypic shifts in the host population (Cory *et al.*, 2003).

The high levels of genetic diversity in baculovirus populations would largely be irrelevant if it did not result in different baculovirus phenotypes. However it has been shown that different genetic variants may vary in a number of phenotypic traits, including virulence, speed of kill, OB size and morphology, and progeny yield (Cory *et al.*, 2005; Hamm & Styer, 1985; Hodgson *et al.*, 2001; Simón *et al.*, 2008; Thézé *et al.*, 2014). Overall the generation and maintenance of genetic diversity is thought to benefit baculovirus populations by providing a source of selectable variation which allows them to adapt to and survive changes in their environment (Chateigner *et al.*, 2015; Harrison, 2009). In addition it has been shown that mixed genotype co-infections result in greater host mortality and higher yields of progeny virus, relative to single genotype infections (Bernal *et al.*, 2013; Hodgson *et al.*, 2004). Therefore there are clear advantages for baculoviruses in maintaining population genetic diversity; however the exact mechanisms responsible for the high levels of genotypic diversity and its maintenance in baculovirus populations are not yet fully understood (Baillie & Bouwer, 2013; Cory & Myers, 2003; Simón *et al.*, 2008).

### 1.2.2. Baculovirus genetic diversity: atypical for dsDNA viruses with large genomes?

Many dsDNA viruses, including the baculoviruses, have followed a "gene gatherer" evolutionary strategy, which has led to the large size of their genomes (Duffy *et al.*, 2008). These viruses are thought to have collected new genes from hosts and co-infecting pathogens, through horizontal gene transfer events over the course of their evolutionary history (Gilbert *et al.*, 2014, 2016; Iyer *et al.*, 2001; Shackelton & Holmes, 2004). These collections of genes act as a "molecular toolbox" which allows these dsDNA viruses to manipulate their hosts by causing changes to host morphology, physiology and behaviour (Belshaw *et al.*, 2008; Lucas *et al.*, 2008; Shackelton & Holmes, 2004). A number of conserved baculovirus genes are thought to have been obtained by ancestral viruses through lateral gene transfers. Two examples are the conserved lepidopteran-specific baculovirus *egt* and *chitinase* genes, which phylogenetic analyses have shown were most likely acquired from an ancestral host insect and γ-proteobacteria, respectively (Hughes & Friedman, 2003; Katsuma *et al.*, 2008; Wang *et al.*, 2004). Both of these genes are classed as auxiliary genes, since they are not essential for baculovirus replication, but do increase the probability of successful baculovirus transmission (Daimon *et al.*, 2003; Katsuma *et al.*, 2008).

Most RNA viruses are thought to have followed an alternative evolutionary strategy to the "gene gatherer" strategy of dsDNA viruses, which has resulted in their characteristically small genome sizes (Duffy *et al.*, 2008). Due to the inherent chemical instability of RNA, it has been suggested that natural selection favours viruses with high replication rates that produce large numbers of progeny, and that this allows RNA virus populations to remain viable (Belshaw *et al.*, 2008; Drake, 2006). However a high replication rate, by virus encoded RNA polymerases lacking exonuclease proofreading activity, is thought to come with the concomitant fitness cost of low replication fidelity (Belshaw *et al.*, 2008; Drake, 2006; Furio *et al.*, 2005). The term "fitness" is used here to describe the relative ability of an individual virion to initiate infection and replicate within a host, and be successfully transmitted to a new host, under a set of prevailing environmental conditions (Duffy *et al.*, 2008; Lynch, 2010; Manrubia & Lazaro, 2006). Most mutations are deleterious, and the low level of replication fidelity is thought to constrain RNA viruses to small, compact genomes. RNA viruses with larger genomes would potentially risk crossing an "error threshold", where the rate at which deleterious mutations would accumulate in viral genomes would soon lead to a virus population producing non-viable progeny (Belshaw *et al.*, 2008; Duffy *et al.*, 2008).

However a beneficial spin-off of the high mutation rate of RNA viruses is that it produces a great degree of genetic diversity within virus populations. This should in theory allow these virus populations to adapt quickly to changing environments, for example, escape mutations which would allow viruses to evade a host's immune response (Belshaw *et al.*, 2008; Elena & Sanjua, 2005; Vignuzzi *et al.*, 2006). The opposite is true for many large dsDNA viruses, which have acquired DNA polymerases with exonuclease proofreading activity. These proofreading polymerases allow large dsDNA viruses to maintain the coding integrity of their genomes, and prevent them from crossing the "error threshold" (Drake, 1991; Sanjuán *et al.*, 2010). However this concomitant reduction in the genome-wide mutation rate with the evolution of large dsDNA genomes is believed to be largely responsible for the relatively low levels of genetic diversity typically seen in these virus populations (Belshaw *et al.*, 2008; Drake, 2006; Drake *et al.*, 1998; Duffy *et al.*, 2008). Therefore baculoviruses represent an anomaly amongst the dsDNA viruses with large genomes as, although it appears they have followed the "gene gatherer" evolutionary strategy, considerable genetic diversity exists within baculovirus populations (Baillie & Bouwer, 2012b; Cory *et al.*, 2005; Hodgson *et al.*, 2001).

### 1.2.3. Factors that affect the level of baculovirus genetic diversity

Factors which affect the level of genetic diversity in baculovirus populations can be divided into two broad groups: (1) those which affect the frequency at which new genetic variation is generated within a population and, (2) factors which act on and shape the genetic variation that already exists within a population.

#### 1.2.3.1. Generation of genetic variation

Mutations are the ultimate source of genetic variation in any population and largely arise as a consequence of polymerase errors during DNA replication (Duffy *et al.*, 2008). The large amount of genetic diversity within baculovirus populations could be seen as an indication that new mutations frequently arise during the replication of baculovirus DNA. However this is thought unlikely, as 3′ − 5′ exonuclease proofreading activity has been associated with several baculovirus DNA polymerases (Hang & Guarino, 1999; Huang & Levin, 2001; Mikhailov *et al.*, 1986). In addition, a high degree of replication fidelity, associated with strong 3′ − 5′ exonuclease proofreading activity, is a conserved trait of the B-family DNA

polymerases, of which baculovirus DNA polymerases are members (Darmawan *et al.*, 2015). Therefore the origin of the extensive genetic heterogeneity in baculovirus populations has largely been attributed to homologous recombination occurring between co-infecting heterogeneous virus genotypes, during the recombination-dependent replication of baculovirus DNA (Hajós *et al.*, 2000; Okano *et al.*, 2006). These recombination events may be responsible for the sequence deletions and insertions in baculovirus genomes observed in some studies (Chateigner *et al.*, 2015; Thézé *et al.*, 2014). Although homologous recombination between genotypes occurs at a high frequency within cell culture, it is not known to what extent it occurs *in vivo* (Hajós *et al.*, 2000). However it is believed to occur at significant levels, as the co-infection of cells by two or more genomes during the systemic phase of infection appears to be a fairly common event (Bull *et al.*, 2001; Simón *et al.*, 2006). Two recent studies on *Autographa californica* multiple nucleopolyhedrovirus (AcMNPV) have also shown using ultra deep whole genome sequencing that there appears to be a relatively high frequency of horizontal gene transfer (HT) between host and virus genomes over the course of an *in vivo* infection, with an estimated 4.8% of AcMNPV genomes harbouring at least one sequence of host origin (Gilbert *et al.*, 2014, 2016). The majority of the horizontally transferred host sequences belonged to known groups of transposable elements, while a minority of them were hypothesised to result from recombination events between host and virus genomes (Gilbert *et al.*, 2016). The horizontally transferred sequences were not maintained in AcMNPV genomes over multiple *in vivo* passages and there appeared to be a constant turnover of host DNA inserted into the viral genome with each new infection (Gilbert *et al.*, 2016). However, it is not known whether similarly high levels of host-to-virus HT occur with other baculovirus species (Gilbert *et al.*, 2014).

A number of studies have shown using whole genome next generation sequencing, that single nucleotide polymorphisms (SNPs) are a major contributor to the genetic diversity in populations of several nucleopolyhedrovirus species (Baillie & Bouwer, 2012a; Chateigner *et al.*, 2015; Thézé *et al.*, 2014; Xu *et al.*, 2013). These studies indicate that SNPs occur across the whole baculovirus genomes in all functional gene classes (Chateigner *et al.*, 2015; Thézé *et al.*, 2014). Although the majority of SNPs detected in these studies were synonymous, a significant number were non-synonymous and would result in changes to the amino acid sequence of expressed proteins. Interestingly, both Theze *et al.* (2014) and Baillie & Bouwer (2012a) detected numerous SNPs within baculovirus core genes in AcMNPV and HearNPV

populations, respectively, while Xu *et al.* (2013) found that the core baculovirus genes in a *Bombyx mori* nucleopolyhedrovirus (BmNPV) population were conserved and contained no SNPs. Single nucleotide substitution events would not be generated as a consequence of homologous recombination, and baculovirus DNA replication by a high fidelity enzyme would be expected to greatly reduce their frequency. Therefore this may be indicative of additional, and as yet unknown, factors which increase the frequency at which new genetic variation is generated within baculovirus populations. Solar ultraviolet radiation (UVR) is known to cause damage to DNA that can potentially lead to the generation of new mutations (reviewed in Sinha & Häder, 2002). As OBs may spend prolonged periods in the environment, exposure to solar UVR may be one such factor that contributes to the generation of novel baculovirus genetic variants.

In another study, Baillie & Bouwer (2013) showed that the generation of novel genetic variants over the course of an *in vivo* passage contributed to the greater detectable genetic diversity in HearNPV populations isolated from larvae inoculated with a low dose ($LD_5$), relative to the amount of genetic diversity detected in the inoculum population and populations isolated from larvae inoculated with a high dose ($LD_{95}$). The authors suggested that at low inoculum doses any novel genetic variants arising early in an infection may be amplified to detectable levels due to an increased number of replication cycles, which ultimately results in a higher virion yield, relative to high inoculum doses (Baillie & Bouwer, 2013). This inference was supported by an empirically based mathematical model which showed that increases in dose are associated with a reduction in the yield of progeny virions at the time of host death (White *et al.*, 2012). These findings indicate that inoculum dose may play an important role in determining whether novel genetic variants generated over the course of an infection contribute to the genetic diversity of post-passage NPV populations. In theory, at very low doses a host may be infected by a single virion, and in the case of SNPVs this would mean the host would be infected by a single virus genome (Rohrmann, 2013; Theilmann & Blissard, 2008). It is not known in these instances, where there is no potential for recombination between different genotypes, whether novel genetic variants would still be generated over the course of an *in vivo* passage resulting in a genotypically heterogeneous post-passage population. Overall it appears that the genetic diversity observed in baculovirus populations is generated by a number of different mechanisms, including mutations,

recombination between heterogeneous virus genomes and host-to-virus horizontal gene transfer events.

#### 1.2.3.2. Factors which shape existing genetic diversity

The two main forces which act on existing genetic variation in populations, and drive evolution in all organisms, are differential selection and random genetic drift (Gutiérrez *et al.*, 2012; Rambaut *et al.*, 2004). It is thought that the genotypic composition of a baculovirus population is largely shaped by the set of selective pressures imposed by the prevailing environmental conditions and the local host population (Hitchman *et al.*, 2007; Hodgson *et al.*, 2004; Hodgson & Vanbergen, 2002). Host species and host food-plant species are two factors which have been shown to differentially select genetic variants in baculovirus populations (Hitchman *et al.*, 2007; Hodgson & Vanbergen, 2002).

Many baculoviruses are capable of infecting several closely related host species, often members of the same insect genus (Allaway & Payne, 1984; Moscardi, 1999). It has been shown that when a baculovirus population, containing a mixture of genotypes, is passaged in two different host species, the genotype which will ultimately become predominant in the baculovirus population is host species-dependent (Hitchman *et al.*, 2007; Weitzman *et al.*, 1992). This suggests that certain genetic variants within a baculovirus population may be better adapted to a particular host species than other variants within the population, and that each host species places a distinct set of selective pressures on the variant pool of a baculovirus population (Hitchman *et al.*, 2007; Weitzman *et al.*, 1992). Similarly it has been found that host food-plant species can differentially affect both the pathogenicity and the progeny virus yield of different genotypes in a baculovirus population during an infection (Hodgson *et al.*, 2004; Hodgson & Vanbergen, 2002).

However it has been argued that it is unlikely that the directional selective pressures on a baculovirus population will ever be stringent enough to result in the complete loss of minor genotypes from the population over multiple *in vivo* passages (Hitchman *et al.*, 2007). This is because no natural environment is static, and shifts in the selective pressures will occur. This could potentially be one mechanism of genetic diversity maintenance in baculovirus populations, as shifts in the selective pressures would result in a shift in the relative fitness of genetic variants within the baculovirus population (Hitchman *et al.*, 2007). It has also been

suggested that the mixing of genotypically heterogeneous populations from different geographical regions may occasionally occur, resulting in novel populations of genetic variants within individual hosts (Hodgson & Vanbergen, 2002).

Fitness trade-offs associated with certain phenotypic traits of genetic variants in a population may further prevent selective pressures from progressively reducing the genetic diversity of baculovirus populations over a number of *in vivo* passages. Trade-offs occur when there are both beneficial and disadvantageous phenotypic traits, which carry a fitness cost, associated with a genotype (Cory & Myers, 2003). The most well studied trade-off in baculoviruses is between the time taken for an infection to cause host mortality, also known as speed-of-kill, and the yield of progeny OBs upon host death (Cory & Myers, 2003). Baculoviruses must kill their host for horizontal transmission to occur; therefore fast-killing genotypes may have a transmission advantage over those that take longer to cause mortality (Simón *et al.*, 2008; White *et al.*, 2012). However, studies have shown that the yield of progeny OBs is higher for genotypes that take longer to kill their hosts (Hodgson *et al.*, 2004; Simón *et al.*, 2008). Therefore in trade-off situations opposing traits are unlikely to lead to the superior fitness and selection of one genotype over the other. Both Hodgson *et al.* (2004) and Simón *et al.* (2008) have also shown that genotypes may actually gain increased yields in mixed infections relative to single genotype infections. This suggests that complementation rather than competition between co-infecting genetic variants may occur. This could indicate that collections of genotypes are the unit of natural selection in baculovirus populations, rather than single genotypes (Hodgson *et al.*, 2004).

Whereas natural selection would drive the gradual adaptation of a baculovirus population to a set of prevailing host or environmental conditions, genetic drift is an evolutionarily neutral, stochastic process where the frequency of genetic variants within a population is determined by a random sampling event (Duffy *et al.*, 2008; Gutiérrez *et al.*, 2012). The horizontal transmission step for many viruses is characterised by a population bottleneck, where only a single or limited pool of genotypes establishes an infection (Bull *et al.*, 2011; Carlson *et al.*, 2014; Gutiérrez *et al.*, 2012). It is still unclear to what extent transmission is a neutral or selective process, but there is evidence that at least under certain circumstances the process determining which virus genotypes establish an infection is largely stochastic and not dependent on virus fitness (Edwards *et al.*, 2006). The relative impact of bottleneck events,

and potentially genetic drift, on baculovirus genetic diversity has not been assessed. However it is possible that in environments where the concentration of infectious OBs is low that transmission bottlenecks may cause significant reductions in the genetic diversity of baculovirus populations. If these processes are selectively neutral then genetic drift may also significantly impact on the genotype composition of populations with regards to which baculovirus genotypes are dominant within a host.

Overall, both differential selection and random genetic drift may act to shape the genotypic composition of a NPV population over the course of an *in vivo* passage. What is not known is how infection and environmental variables affect the relative contributions of these two forces in determining the genetic diversity of NPV populations. It may be that under different sets of conditions either differential selection or random genetic drift plays a greater role, relative to the other, in shaping the post-passage genotypic composition of a NPV population. In particular, the inoculum dose may play an important role in determining the potential for multiple genotypes to co-infect a host (Zwart *et al.*, 2009). It is thought that in many virus-host interactions that at high inoculum doses there is a higher probability of the host being exposed to and subsequently being infected by a greater proportion of the genotypes in the inoculum population (Zwart *et al.*, 2011). Essentially there may be greater potential for differential selection by host factors to shape the genetic diversity of a virus population at higher doses, whereas at low doses random genetic drift may be largely responsible for the post-passage genetic diversity (Gutiérrez *et al.*, 2012; Zwart *et al.*, 2011). The co-occlusion of several genotypes within a single OB, a phenomenon which has been shown to occur in NPV infections, is thought to contribute to the maintenance of genetic diversity by increasing the probability of multiple genotypes co-infecting a host, thus reducing the effect of random genetic drift (Clavijo *et al.*, 2010). However, no studies have yet shown how the use of non-occluded ODVs as the inoculum will affect the maintenance of diversity at lower doses.

## 1.3. Techniques used to study baculovirus genetic diversity

Over the years a number of techniques have been used to study the heterogeneity of natural baculovirus populations. One of the first methods relied on differences in plaque morphology to detect and isolate different variants from a baculovirus isolate, after *in vitro* passage in insect cell culture (Potter *et al.*, 1976). However, this early method for studying baculovirus population diversity was made redundant with the advent of a variety of far more sensitive molecular techniques, which are able to discriminate between different genetic variants within a population. Restriction fragment length polymorphism (RFLP) analyses, denaturing gradient gel electrophoresis (DGGE), Sanger-dideoxy DNA sequencing and, most recently, next generation DNA sequencing (NGS) methods are four molecular techniques that have been used to detect and study genetic variation in a number of microbial species. Of these RFLP analyses have been used extensively (Hodgson *et al.*, 2001; Lee & Miller, 1978; Redman *et al.*, 2010) and, to a lesser extent, Sanger-dideoxy DNA sequencing to detect and study genetic variation in baculovirus populations. DGGE, although not a new technique, has only recently been used to study variation in baculoviruses (Baillie & Bouwer, 2011). NGS technologies are a comparatively recent development, which have been widely used to assess genetic diversity in other groups of viruses (reviewed by Radford *et al.*, 2012), and are seeing greater application in studies of baculovirus genetic diversity (Baillie & Bouwer, 2012a; Chateigner *et al.*, 2015; Gilbert *et al.*, 2014, 2016; Thézé *et al.*, 2014; Xu *et al.*, 2013). All of these molecular techniques have different strengths and levels of sensitivity and are suited to different applications in the study of baculovirus genetic diversity.

### 1.3.1. Well-established techniques for detecting baculovirus genotypic variants

#### 1.3.1.1. Restriction fragment length polymorphism (RFLPs)

RFLP, also known as restriction endonuclease (REN) profile analysis, has been the most extensively used technique to detect genotypic variants in baculovirus isolates (Hodgson *et al.*, 2001; Lee & Miller, 1978; Redman *et al.*, 2010). Restriction endonucleases digest dsDNA at enzyme-specific nucleotide sequence recognition sites, and the digestion of baculovirus genomic DNA results in the production of fragments of variable length which are subsequently separated by gel electrophoresis (Busch & Nitschko, 1999; Olive & Bean, 1999; Smith & Summers, 1978). This results in the production of a highly reproducible restriction

fragment pattern for each genotype (Olive & Bean, 1999; Smith & Summers, 1978). The basis of this technique is that the position of the endonuclease recognition sites in the genomes of different variants can be polymorphic, as sites may be lost, added or shuffled around in a genome by mutation or recombination events (Lee & Miller, 1978; Smith & Summers, 1978). A variation of this technique is PCR-based locus-specific RFLPs, where a specific genetic locus is first amplified and then subjected to RFLP analysis (Busch & Nitschko, 1999; Murillo *et al.*, 2006).

RFLPs allow variation between baculovirus genotypes to be detected at a genome-wide scale; however their discriminatory ability is low as sequence variation between variants must change the position or number of endonuclease recognition sites in the genome in order for it to be detected (Busch & Nitschko, 1999; Olive & Bean, 1999). Restriction fragment patterns are also difficult to interpret and do not provide any information on what differences exist between variants at the nucleotide sequence level and where these variations occur within the genome (Busch & Nitschko, 1999; Olive & Bean, 1999). Although the restriction patterns generated with locus-specific RFLPs are simpler and easier to interpret, their discriminatory ability is even lower due to the fact that they do not detect variation between genotypes at a genome-wide scale (Murillo *et al.*, 2006; Olive & Bean, 1999). However despite these limitations, RFLPs have been successfully used to detect genetic variation between baculovirus populations from different geographical areas, and between different hosts from a single agricultural field (Murillo *et al.*, 2006; Patel *et al.*, 2009a; Shapiro *et al.*, 1991). They have also been used to show that there are multiple genetic variants within baculovirus populations isolated from individual insect hosts (Cory *et al.*, 2005; Hodgson *et al.*, 2001).

#### 1.3.1.2. Sanger-dideoxy sequencing

DNA sequencing has been described as the "gold standard" technique for distinguishing between viral genetic variants, as it allows the nucleotide sequences of polymorphic genomic regions to be determined. In contrast, RFLPs and DGGE only allow relative differences in the underlying genetic variant composition of populations to be detected (Muyzer *et al.*, 1993; Olive & Bean, 1999). Prior to the advent of NGS sequencing technologies, sequence analyses were most commonly carried out using some variation of the Sanger-dideoxy chain-terminating method (Mardis, 2013; Sanger *et al.*, 1977; Shendure & Ji, 2008). In the most widely used, semi-automated, capillary-based version of this method, dideoxy analogues of

the four deoxyribonucleotides are each labelled with a different fluorescent tag (Prober *et al.*, 1987; Smith *et al.*, 1986). This allows the different extension products of the cycle sequencing reaction to be detected upon their electrophoretic separation in a capillary based polymer gel (Shendure & Ji, 2008). Software is then able to generate the nucleotide sequence from the trace file containing the emission spectra of the four different fluorescence signals (Shendure & Ji, 2008).

There are two main drawbacks to the Sanger-dideoxy sequencing method, which make it problematic with regards to the study of genetic diversity within natural virus populations. The first is that the resultant sequence is a consensus sequence which is generated from the dominant signal at each nucleotide position for the collection of template sequences in the sample being sequenced (Olive & Bean, 1999; Shendure & Ji, 2008). This hinders its ability to detect and sequence individual genetic variants present within a viral isolate, unless sequences from the different variants are first isolated, for example, by including a single genome amplification (SGA) step prior to sequencing (Gong *et al.*, 2013; Gregori *et al.*, 2014). The second drawback, which follows from the first, is that to sequence a good representative sample of genetic variants from a natural virus population (also referred to as good sequencing coverage or "depth"), requires a time intensive, multi-step process. This involves the isolation of single genomes, followed by cycle sequencing reactions for each of the genomes, and finally asynchronous analysis of all sequencing reactions on a capillary electrophoresis system (Mardis, 2013; Shendure & Ji, 2008).

Although Sanger-dideoxy sequencing has been used in this manner to study the within-host sequence diversity of human immunodeficiency virus (HIV) (Novitsky *et al.*, 2013) and hepatitis B virus (HBV) (Gong *et al.*, 2013) populations, it has not been used to study the genetic diversity within baculovirus populations isolated from individual hosts. Rather Sanger-dideoxy sequencing in baculovirology has largely been used to characterise and infer the phylogenetic relationships between baculovirus species (Jehle *et al.*, 2006a; Lange *et al.*, 2004), and to differentiate between baculovirus populations isolated from different hosts and geographical regions, based on variation between the consensus sequences of conserved baculovirus genes (Ferreira *et al.*, 2014; Harrison *et al.*, 2014; Rowley *et al.*, 2010).

### 1.3.2. Newer methods for studying baculovirus genetic diversity

#### 1.3.2.1. Next generation sequencing (NGS)

Over the last decade a number of NGS technologies, also known as high-throughput sequencing (HTS) systems, have emerged as a viable and attractive alternative to Sanger-dideoxy sequencing, for studying genetic diversity within natural virus populations (reviewed in Beerenwinkel *et al.*, 2012; Nelson & Hughes, 2015; Radford *et al.*, 2012). Although there are a variety of different technologies that fall under the HTS label, most commercially available HTS platforms are of the cyclic-array sequencing variety (Shendure & Ji, 2008). These include Roche Diagnostics' 454, Ilumina's MiSeq and Life Technologies' SOLiD and Ion Torrent sequencing platforms (reviewed by Ansorge, 2009; Goodwin *et al.*, 2016; Mardis, 2013; Shendure & Ji, 2008). With these platforms, a DNA sample (either genomic DNA or PCR amplicons) is first used to construct a library of short DNA fragments with platform-specific adaptors ligated to each end (Goodwin *et al.*, 2016; Mardis, 2013; Shendure *et al.*, 2011). A platform specific protocol is then used to create a two-dimensional spatial array of immobilised PCR colonies, also known as sequencing foci (Mardis, 2013). Each sequencing focus consists of a cluster of amplicons, attached to the array surface, which originated from a single library molecule (Shendure & Ji, 2008). The cyclical sequencing process is either polymerase or ligase driven, and consists of a nucleotide addition step followed by a nucleotide detection and wash step (Mardis, 2013). During the nucleotide detection step, data are acquired by imaging the full sequencing array, allowing for the simultaneous detection of the platform-specific signal emitted from each sequencing focus (Shendure & Ji, 2008). Although there are differences between these platforms with regards to the method used to generate the array and the underlying sequencing biochemistry, the overall sequencing strategy is similar across all cyclic-array sequencing platforms.

HTS systems have several advantages over Sanger-dideoxy sequencing, including a greater degree of automation and the massively parallel nature of the sequencing process, in which potentially millions of individual template molecules can be sequenced simultaneously (Mardis, 2013). These properties of HTS ultimately mean that far more sequencing data can be generated in a much shorter time period relative to Sanger dideoxy sequencing. However the ability to amplify and sequence single templates in a massively parallel manner is particularly advantageous for the study of genetic diversity in virus populations, as it allows

the full sequence variant spectrum in a virus population to be determined (Beerenwinkel *et al.*, 2012). For this reason HTS is potentially the most powerful technique for the study of within-host baculovirus diversity. In one of the first baculovirology studies to use HTS technology, Baillie & Bouwer (2012a) used the Roche 454 HTS platform to detect multiple SNPs within a core baculovirus gene (*DNA polymerase*) and two core lepidopteran-specific baculovirus genes (*dbp1* and *me53*), in a pooled HearNPV sample made up of different geographical HearNPV isolates. Subsequently, several other studies have used a variety of HTS platforms to conduct whole genome sequencing for nucleopolyhedrovirus isolates of several different species, including *Spodoptera exigua* multiple nucleopolyhedrovirus (SeMNPV), BmNPV, AcMNPV and *Catopsilia pomona* nucleopolyhedrovirus (CapoNPV) (Gilbert *et al.*, 2014, 2016; Thézé *et al.*, 2014; Wang *et al.*, 2016; Xu *et al.*, 2013). These whole genome sequencing studies have been used to assess the overall level and distribution of genetic variation across viral genomes within populations of different nucleopolyhedrovirus species (Baillie & Bouwer, 2012a; Chateigner *et al.*, 2015; Thézé *et al.*, 2014; Xu *et al.*, 2013).

There remains an opportunity to use HTS to study the effect of infection, host and environmental variables on the genetic variant composition of baculovirus populations at the nucleotide sequence level. Deep amplicon sequencing of selected genomic loci has been used extensively to characterise the microbial composition of bacterial and fungal communities as well as some virus populations (Gonçalves Rossi *et al.*, 2016; Görzer *et al.*, 2010; Ihrmark *et al.*, 2012; Quartieri *et al.*, 2016; Sinclair *et al.*, 2015). Although deep amplicon sequencing does not provide genome-wide sequence information it has several advantages for studies where the aim is not to assess the genome-wide genetic diversity within a population, but to evaluate the effect of different variables on the variant composition of virus populations (Goodwin *et al.*, 2016). These include maintaining linkage between contiguous sequence regions and individual virus variants, simpler and less time consuming data analysis, and being considerably more affordable when multiple populations need to be assessed (Goodwin *et al.*, 2016; Grad *et al.*, 2014). However, all experimental applications utilising the currently available HTS platforms suffer from number of disadvantages. These include relatively short sequencing reads and higher average error rates relative to Sanger-dideoxy sequencing (Kircher & Kelso, 2010; Shendure & Ji, 2008), the initial capital outlay expense for acquiring a HTS system, which remains prohibitively high for many research groups, and the relatively

complex downstream data management and analysis processes for the typically large HTS datasets (Beerenwinkel *et al.*, 2012; Kircher & Kelso, 2010).

#### 1.3.2.2. Denaturing gradient gel electrophoresis (DGGE)

DGGE has been extensively used to study genetic diversity in a variety of microbial communities (reviewed in Muyzer & Smalla, 1998 & Muyzer, 1999). The technique has also been used to evaluate the level of genetic variation within natural virus populations (Carmona *et al.*, 2012; Harris & Teo, 2001; Motta *et al.*, 2002, 2006), although it has only relatively recently been used to study baculovirus genetic diversity (Baillie & Bouwer, 2011). DGGE allows DNA fragments of the same length to be separated according to differences in their melting temperatures ($T_m$). As the $T_m$ of a fragment is dependent on its relative complement of A-T and G-C base pairs, fragments are essentially separated due to differences in their nucleotide sequences (Fischer & Lerman, 1980, 1983).

A PCR amplification step is usually included to amplify a variable genomic region from a heterogeneous genotype pool (Short & Suttle, 1999). During DGGE, PCR fragments migrate through a gel containing a linear denaturing gradient of formamide and urea (Fischer & Lerman, 1980, 1983). Electrophoretic separation of the amplicons is brought about by partially denatured DNA having a far lower mobility in the gel than dsDNA (Fischer & Lerman, 1980, 1983). Therefore the distance migrated through a gel by a fragment is dependent on its $T_m$ and denaturation characteristics. When a synthetic, highly stable GC clamp is added to the 5´ end of the PCR primers the technique is sensitive enough to detect single nucleotide polymorphisms (SNPs) between closely related amplicons (Myers *et al.*, 1985; Sheffield *et al.*, 1989).

Although the discriminatory ability of PCR-DGGE is much higher than that of locus-specific RFLPs (Enwall & Hallin, 2009; Matsuyama *et al.*, 2007), it can only be used to detect and assess the level of variation in specific genomic regions and not at a genome-wide level (Muyzer & Smalla, 1998). In addition this technique will only detect polymorphisms in a region if they affect the denaturation characteristics of the amplified fragments (Muyzer & Smalla, 1998). Like RFLPs, DGGE has a further limitation in that it does not provide any information on the differences between variants at the nucleotide sequence level. However, DGGE does make it easier to isolate polymorphic amplicons, which can then be excised from

the gel and for further analysis by Sanger-dideoxy sequencing (Fromin *et al.*, 2002). Some previous studies have also calculated quantitative measures of genetic diversity from DGGE profiles, such as the Shannon diversity index, based on the premise that each band represents a distinct sequence variant and its relative intensity indicates the variant's relative abundance in a population (Fromin *et al.*, 2002; Gafan *et al.*, 2005). However, more recently an extensive study by Neilson *et al.* (2013) showed that several artefacts that commonly occur in multi-template PCR-DGGE affect the accuracy of these quantitative measures of population genetic diversity. Their results suggest that although PCR-DGGE remains an excellent technique to assess the relative structure and composition of microbial communities, it provides a biased quantification of variant richness and relative variant abundance in communities (Neilson *et al.*, 2013).

HTS could potentially generate far more extensive and comprehensive data on the genetic diversity present within baculovirus populations, relative to DGGE. However, the cost of HTS platforms and individual sequencing "runs", and the relative complexity of HTS data analysis, make it an unfeasible option for many research groups (Kircher & Kelso, 2010). This is where the DGGE analysis of baculovirus genetic diversity is advantageous, as it has greater sensitivity than the commonly used RFLP analysis technique (Enwall & Hallin, 2009; Matsuyama *et al.*, 2007), while having a much faster turnaround time and being significantly cheaper than HTS, as it can be conducted in-house over a 24-hour period (Green *et al.*, 2010). DGGE data are also comparatively quick and simple to analyse with the use of genetic fingerprint analysis software, relative to the analysis of HTS data (Fromin *et al.*, 2002). Baillie & Bouwer (2011) developed a DGGE assay to detect genetic variation in a core baculovirus gene (*DNA polymerase*) and two core lepidopteran-specific baculovirus genes (*me53* and *dbp1*). They then showed that this DGGE assay could be used to distinguish between seven different HearNPV geographical isolates based on differences in their *DNA polymerase, me53* and *dbp1* DGGE profiles (Baillie & Bouwer, 2011). The DGGE assay they developed is used in this study to evaluate the effects of successive *in vivo* passaging, solar radiation and variable inoculum dose on the variant composition of HearNPV populations.

## 1.4. Study justification

Many previous studies have shown that considerable genetic diversity exists within NPV populations (Baillie & Bouwer, 2012b; Chateigner *et al.*, 2015; Cory *et al.*, 2005; Gettig & Mccarthy, 1982; Patel *et al.*, 2009b). This high level of genetic diversity is unusual for viruses with large dsDNA genomes, especially as the replication of NPV genomes is conducted by virus-encoded DNA polymerases with proofreading activity (Duffy *et al.*, 2008; Huang & Levin, 2001). This would be expected to decrease the frequency at which mutations are generated and subsequently fixed within NPV populations. Although several studies have looked at how the genetic diversity of NPV populations change under different host and environmental conditions (Hitchman *et al.*, 2007; Hodgson & Vanbergen, 2002), a number of key questions remain with regards to which factors contribute to the maintenance and potential generation of genetic variants in NPV populations over the course of an *in vivo* passage.

These questions include whether a stable consensus genotypic composition will emerge over several successive *in vivo* passages as a result of selective pressures, or whether genetic drift and/or the generation of novel genetic variants during each transmission and infection cycle will result in constant shifts in the post-passage genetic diversity. It is also not known if the pre-infection exposure of OBs to solar radiation, which is known to cause significant amounts of NPV inactivation in the field (Petrik *et al.*, 2003; Witt, 1984), affects the genetic diversity of post-passage NPV populations. It is possible that in addition to decreasing the transmission of low frequency genotypes through virion inactivation, the UV component of solar radiation, a potent DNA mutagen, may increase the frequency at which novel NPV genetic variants are generated *in vivo*.

There is also a need to further evaluate the effect of inoculum dose on the genetic diversity of post-passage NPV populations. A previous study by Baillie & Bouwer (2013) found that the level of genetic diversity was higher in low OB dose ($LD_5$) infections, relative to high OB dose ($LD_{95}$) infections. However, it is not known whether the co-occlusion of virions contributes to the maintenance and generation of genetic variants in low dose OB infections, and whether a similar effect would be observed over a range of doses if ODVs rather than OBs were used as the inoculum. At low ODV doses there is a high probability that a host will

only be infected by the dominant genotype in the inoculum population. In these instances, where host infection may be initiated by a single genotype, it is unclear whether the post-passage population will be genotypically homogeneous or whether the generation of novel genetic variants over the course of the infection will result in a genotypically heterogeneous population.

Studies where techniques are used that are able to detect close to the full spectrum of genetic variants in a NPV population, either over the whole genome or within a gene region, are required to answer these questions. In this study a PCR-DGGE assay developed by Baillie & Bouwer (2012a) was used to assess the genetic variation within regions of a baculovirus core gene, *DNA polymerase* and a core lepidopteran-specific baculovirus gene, *me53*, in HearNPV populations. This assay was chosen as in theory it has greater sensitivity than the routinely used RFLP techniques (Enwall & Hallin, 2009; Matsuyama *et al.*, 2007). However, several studies have reported the co-migration of genetic variants and the presence of additional artefactual bands in PCR-DGGE gel profiles, and these would affect the accuracy and reproducibility of this technique (Hong *et al.*, 2007; Neilson *et al.*, 2013; Samarajeewa *et al.*, 2015). This meant that the ability to use PCR-DGGE to assess the genetic diversity of HearNPV populations first needed to be established. The PCR-DGGE assay was subsequently used to detect changes in the genetic diversity of a HearNPV population over several *in vivo* passages, as well as to evaluate whether pre-infection exposure of OBs to solar radiation contributed to changes in the genetic variant composition of HearNPV populations over the course of a passage. The PCR-DGGE assay was also used to detect any differences in the post-passage genetic diversity of populations isolated from *H. armigera* larvae inoculated with different ODV doses. Overall, it was hoped that this study would further our understanding of the factors that affect both the level of genetic diversity in post-passage NPV populations and the degree to which the genetic variant composition of a post-passage population differs from that of an inoculum population.

## 1.5. Objectives and aims

### 1.5.1. Overall project objective

The main objective of this study was to evaluate the effects of successive *in vivo* passaging, the exposure of OBs to solar radiation and variable ODV inoculum doses on the genetic diversity of HearNPV populations.

### 1.5.2. Specific research aims

1. To evaluate the accuracy and reproducibility of PCR-DGGE when used to assess the genetic diversity of HearNPV populations.

2. To assess the effect of successive *in vivo* passaging on the genetic diversity of HearNPV populations.

3. To evaluate if the pre-infection exposure of OBs to solar radiation affects the genetic diversity of HearNPV populations after a post-exposure *in vivo* passage.

4. To assess and compare the effects of different ODV inoculum doses on the genetic diversity of HearNPV populations after an *in vivo* passage.

**Table 1.1.** Core genes present within all baculovirus and lepidopteran-specific baculovirus genomes sequenced to date (based on Garavaglia et al., 2012).

| Gene product function | Core baculovirus genes | Core lepidopteran-specific genes |
|---|---|---|
| Replication | *lef-2, primase/lef-1, DNA polymerase, DNA helicase, alkaline nuclease/alk-exo* | *dbp1, lef-3, me53* |
| Transcription | *p47, lef-8, lef-9, vlf1, lef-4, lef-5, p74/pif-0* | *lef-6, ac35, pp31, lef-11, ac38, ie-1* |
| Structural | *vp1054, ac66, gp41,p95/vp91, vp39, p6.9, odv-e25, odv-e43, 49k, odv-e18, odv-e27* | *polh, ac23, fp/25k, pe/pp34* |
| *Per os* infectivity factors | *pif-2, pif-6, pif-4, pif-3, pif-1, odv-e56/pif-5* | *p11, p24, ac145* |
| Auxiliary function | *ubiquitin ligase, sulphydryl oxidase/p33, phosphatase/38k* | *pk1, gp37, ac75, p12, exon-0* |
| Unknown function | *ac78, ac81, p18, p40, p48* | *ac13, tlp, ac106/107, ac110, ac144* |

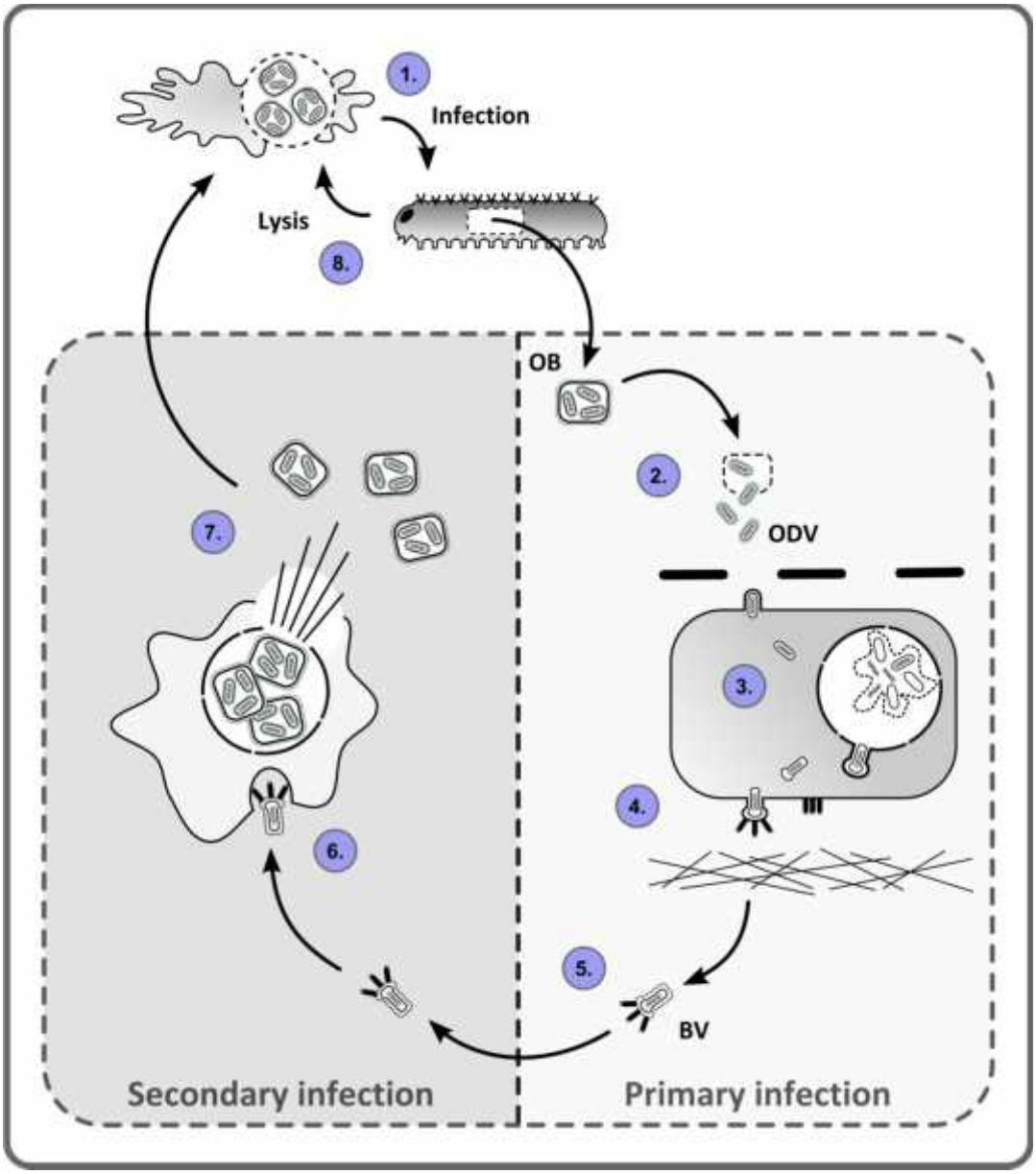



**Figure 1.1.** The generic biphasic replication cycle of the nucleopolyhedroviruses. **1.** OBs are ingested by insect larvae during feeding. **2.** The polyhedrin protein matrix of OBs is dissolved in the alkaline midgut of larvae, releasing ODVs, which pass through the peritrophic membrane. **3.** ODVs initiate the primary infection phase when their envelopes fuse with the membrane of a midgut epithelial cell, allowing their nucleocapsids to enter the cytoplasm. Viral DNA replication and nucleocapsid assembly occurs in the virogenic stroma, within the host cell's nucleus. **4.** Systemic infection occurs when progeny BVs bud out of infected midgut epithelial cells, through areas of the plasma membrane which have been modified with virus envelope fusion proteins. **5.** After transiting through the basal lamina surrounding the midgut epithelium, BVs initiate the secondary infection phase by spreading the infection to the host's different tissues. **6.** Virus fusion proteins mediate the entry of BVs into somatic cells by triggering endocytosis. **7.** In the terminal stages of an infection, multiple progeny OBs accumulate in the nuclei of infected cells, and are released upon virus-induced cellular disintegration. **8.** Viral chitinase and cathepsin enzymes are responsible for the lysis or "melting" of a host's body, leading to the release of progeny OBs into the environment.

# Chapter 2

# Evaluation of the use and reproducibility of PCR-DGGE to assess genetic diversity in *Helicoverpa armigera* nucleopolyhedrovirus populations

## 2.1. Abstract

PCR-DGGE has been widely used to assess and compare the genetic variant composition of microbial populations. However, previous studies have identified a number of artefacts that arise during multi-template PCR-DGGE that can prevent the accurate measure of genetic variant richness and abundance from DGGE profiles. The aim of this study was to assess the effect of any artefacts on the use and reproducibility of PCR-DGGE to assess genetic diversity in populations of a dsDNA insect virus, *Helicoverpa armigera* nucleopolyhedrovirus (HearNPV). Heteroduplex bands were found to be a major artefact in HearNPV profiles which would result in artificially inflated measures of genetic variant richness for HearNPV populations. However HearNPV DGGE profiles were highly reproducible across PCR and gel replicates and could be reliably used as genetic fingerprints to discriminate between populations that differ in variant composition. The number of heteroduplex bands also increased with the underlying number of variants in a population and the profiles could be used to compare the relative level of genetic diversity in HearNPV populations. An additional finding was that deep amplicon sequencing on the MiSeq platform was superior to the traditional method of Sanger sequencing of cloned fragments extracted from DGGE bands, as it was able to detect a low frequency sequence variant in a HearNPV population that was not detected by the latter approach. PCR-DGGE and deep amplicon sequencing could be used as complementary techniques to assess and compare the genetic diversity of HearNPV populations in future studies. In this approach, the relative variant composition of HearNPV populations would first be assessed using PCR-DGGE. Subsequently deep amplicon sequencing would be used to evaluate the underlying differences in the number and identity of sequence variants between populations where PCR-DGGE had indicated differences in variant composition existed.

## 2.2. Introduction

Denaturing gradient gel electrophoresis (DGGE) has been used extensively to assess the genetic diversity of a wide range of microbial communities since its use in the field of microbial ecology was first pioneered by Muyzer *et al.* (1993). Most applications of PCR-DGGE have focused on using regions of the 16S rRNA gene as phylogenetic markers to assess the taxonomic composition of bacterial communities in a wide variety of environments and to evaluate the impact that environmental perturbations have on the structure of these communities (Dowd *et al.*, 2008; Ferris *et al.*, 1996; Legatzki *et al.*, 2012; Muyzer *et al.*, 1993; Rosario *et al.*, 2007). However, PCR-DGGE has also been used extensively to assess genetic diversity in fungal communities (Gomes *et al.*, 2003; Kim *et al.*, 2010; Wang *et al.*, 2015) and virus populations (Baillie & Bouwer, 2011; Harris & Teo, 2001; Motta *et al.*, 2002; Short & Suttle, 1999).

One of the strengths of PCR-DGGE is that it produces reproducible genetic fingerprints that can be used to discriminate between communities that differ in their microbial composition (Hong *et al.*, 2007; Kowalchuk *et al.*, 1997; Nicolaisen & Ramsing, 2002). In addition to using PCR-DGGE purely as a fingerprinting technique, many researchers have applied quantitative methods to DGGE band profiles in order to evaluate the α-diversity of microbial communities (Ampe & Miambi, 2000; Fromin *et al.*, 2002; Gafan *et al.*, 2005; Loisel *et al.*, 2006). When used in the context of DGGE studies, α-diversity is a measure of the richness and relative abundance of sequence variants in a community at a particular time (Jost, 2007; Neilson *et al.*, 2013). In these studies each band in a DGGE profile is taken to represent a unique genetic variant, while the band's relative intensity indicates the relative abundance of the genetic variant in the community (Fromin *et al.*, 2002). This assumption allows ecological diversity and evenness indices, such as the Shannon diversity index (Shannon, 1948), to be calculated from DGGE band profiles, enabling the objective comparison of α-diversity between microbial communities (Fromin *et al.*, 2002; Gafan *et al.*, 2005). It is also common practice to extract DNA fragments from DGGE bands for Sanger sequencing in order to link individual band positions to known taxonomic groups or virus variants and to assess the underlying phylogenetic diversity of microbial communities (Dowd *et al.*, 2008; Samarajeewa *et al.*, 2015; Schroeder *et al.*, 2003).

However, numerous studies have reported that several types of artefact can occur in DGGE profiles, which could potentially make the technique unsuitable for assessing the α-diversity of certain microbial communities. These artefacts include: 1) artefactual bands associated with heteroduplex or chimeric amplicons formed as a consequence of multi-template PCR (Hong *et al.*, 2007; Neilson *et al.*, 2013; Samarajeewa *et al.*, 2015), 2) sequence variants that exist in a number of stable conformational states and are associated with different bands in a gel profile (Neilson *et al.*, 2013), 3) co-migration of sequence variants, resulting in a single band being associated with more than one sequence variant (Legatzki *et al.*, 2011; Neilson *et al.*, 2013; Samarajeewa *et al.*, 2015; Sekiguchi *et al.*, 2001), and 4) amplification bias during the PCR which results in relative band intensities that do not necessarily reflect the relative abundance of sequence variants in a community (Ferris & Ward, 1997; Neilson *et al.*, 2013). These artefacts skew the quantitative association between band number and variant richness or band intensity and variant abundance, and would result in inaccurate measures of α-diversity in microbial communities (Neilson *et al.*, 2013; Pylro *et al.*, 2016; Reeson *et al.*, 2003). As many of the artefacts associated with PCR-DGGE have their origins in the multi-template PCR step, attempts have been made to develop PCR protocols that reduce their formation (Janse *et al.*, 2004; Lenz & Becker, 2008; Smyth *et al.*, 2010; Thompson *et al.*, 2002). However, many studies that have continued to use PCR-DGGE to assess the genetic diversity of microbial communities have not reported any attempts to first evaluate the degree to which DGGE band or sequence artefacts contribute to their estimates of microbial diversity, nor have they used modified PCR protocols which could potentially reduce their occurrence (Dong & Reddy, 2010; Xu *et al.*, 2015; Zhong *et al.*, 2015).

In recent years deep amplicon sequencing on high-throughput sequencing platforms has become the favoured alternative to PCR-DGGE for assessing the diversity of microbial communities (Quartieri *et al.*, 2016; Sinclair *et al.*, 2015; Yu *et al.*, 2015). The advantages of amplicon deep sequencing relative to PCR-DGGE include the greater detection of low frequency genetic variants and the ability to assess the underlying nucleotide sequence variation within a community without having to conduct additional band extraction and cloning steps (Kraková *et al.*, 2016; Samarajeewa *et al.*, 2015). However, PCR-DGGE remains significantly cheaper than deep amplicon sequencing and has a much faster turnaround time as it can be conducted in-house over a 24-hour period (Green *et al.*, 2010). This makes PCR-DGGE a suitable technique for studies where the genetic diversity of a large

number of microbial populations needs to be assessed. Several studies have compared the use of PCR-DGGE and deep amplicon sequencing for assessing the diversity of various microbial communities (David *et al.*, 2014; Kraková *et al.*, 2016; Samarajeewa *et al.*, 2015; Yu *et al.*, 2015), and some have suggested that the two techniques could complement each other and be used together to detect and study differences in microbial community diversity (Kraková *et al.*, 2016; Samarajeewa *et al.*, 2015).

*Helicoverpa armigera* nucleopolyhedrovirus (HearNPV) is a dsDNA insect virus and member of the *Baculoviridae* family. Previous studies have used PCR-DGGE as a fingerprinting method to discriminate between different HearNPV populations (Baillie & Bouwer, 2011), as well to conduct quantitative assays of the genetic diversity in HearNPV populations (Baillie & Bouwer, 2012b, 2013). Although studies using 16S rRNA genes as phylogenetic markers have consistently reported PCR-DGGE artefacts that result in biased or inaccurate measures of genetic diversity in bacterial communities (Hong *et al.*, 2007; Neilson *et al.*, 2013; Samarajeewa *et al.*, 2015), no study has looked at the potential effects of PCR-DGGE artefacts on estimates of genetic diversity in populations of a dsDNA virus species. Members of an environmental bacterial community can potentially belong to different genera, families or orders (Ciccazzo *et al.*, 2014; Legatzki *et al.*, 2011; de Smidt, 2016; Zhong *et al.*, 2015), whereas members of virus populations under study often consist of the circulating genetic variants of a virus species isolated from a single host individual (Harris & Teo, 2001; Motta *et al.*, 2006). Consequently the amount of genetic variation between members of bacterial communities may be significantly higher than for dsDNA virus populations. It is known that the propensity for PCR artefacts to form that may result in artefactual DGGE bands, such as chimeric and heteroduplex amplicons, is higher when there is less variation between heterogeneous sequences in multi-template PCR (Qiu *et al.*, 2001; Wang & Wang, 1996). This could result in differences in the frequency of artefacts that are observed in PCR-DGGE profiles between studies on bacterial communities and those on dsDNA virus populations. In short, the findings of previous studies on the use of PCR-DGGE to assess α-diversity in bacterial communities may not necessarily apply to dsDNA virus populations.

The primary objective of this study was to establish the prevalence and type of artefacts that occur when PCR-DGGE is used to assess the genetic diversity of HearNPV populations and

the degree to which they affect the estimates of richness and relative abundance of sequence variants in these virus populations. DGGE profiles for regions of *me53* and *DNA polymerase*, two HearNPV genes essential for virus replication, were used to answer four key questions: 1) are DGGE profiles for different HearNPV populations reproducible across different PCRs and gels, 2) does each band in a DGGE profile represent a single genetic variant, 3) are there cases where a single genetic variant is associated with more than one band, and 4) can adjustments to the PCR protocol be used to reduce the formation of any artefacts that may occur?

## 2.3. Methods

A methods flowchart showing the steps taken to assess the reproducibility and the prevalence of artefacts in PCR-DGGE gel profiles, when the technique is used to assess genetic diversity of the *DNA polymerase* and *me53* genes in HearNPV populations, is shown in Figure 2.1.

### 2.3.1. Insect rearing

*H. armigera* eggs, which were originally obtained from the Agricultural Research Council (Pretoria, South Africa), were used to set up a laboratory population. Larvae were reared in individual containers on an artificial diet consisting of a mixture of wheat germ, inactivated brewer's yeast, skim milk powder, agar, ascorbic acid, myo-inositol, choline chloride, L-cysteine hydrochloride, nipagin and cholestrol in water (Bot, 1966). Adults were reared on a 5% (w/v) sucrose solution and kept in glass cages covered with a net on which oviposition took place. Eggs and pupae were sterilised in sodium hypochlorite solutions of 0.2% (w/v) and 0.25% (w/v), respectively, in order to prevent microbial contamination of the population. Both larvae and adults were kept in an insect rearing room under constant temperature and relative humidity conditions of 25 ˚C (± 1) and 60% (± 5), respectively, with a 12:12 h light:dark photocycle.

### 2.3.2. Virus propagation and purification

HearNPV was propagated *in vivo* in third instar *H. armigera* larvae. Larvae were considered to be in the third instar when they were between four and five days old and on the basis of their head capsule widths. Larvae were inoculated with a lethal dose of a HearNPV population using the diet contamination method (Evans & Shapiro, 1997). Thirty larvae displaying overt symptoms of HearNPV infection (cessation of feeding and movement, lack of growth and and development relative to healthy larvae and a shiny integument) were harvested after death, prior to lysing, and stored at -20 ˚C. HearNPV OBs were isolated from larvae cadavers using a modified version of the protocol described by Crook & Payne (1980). Briefly, cadavers were homogenised in 0.1% (w/v) SDS, filtered through a microfibre cloth and subjected to low speed centrifugation to pellet any remaining insect tissue debris. OBs were concentrated using centrifugation (19 000 × *g* for 30 min) and purified by centrifugation through a single discontinuous sucrose gradient of 40-65% (w/w) (49 000 × *g* for 90 min). The OB band at the interface between the 55% and 60% zones was removed and washed

twice in sterile distilled water using differential centrifugation (21 000 × *g* for 30 min). OBs were concentrated by a final centrifugation step (10 500 × *g* for 30 min) and the resulting pellet was re-suspended in a small volume of sterile distilled water, before being stored at 4 ˚C for further use. This was repeated for each of the four HearNPV populations (P1, P2, P3 and P4) used in this study.

### 2.3.3. Extraction of viral DNA

Viral genomic DNA was extracted separately from four HearNPV populations using the NucleoSpin® Tissue XS kit (Macherey-Nagel), using a modified version of the manufacturer's protocol for the purification of DNA from cultured cells. OBs were pelleted by centrifugation (10 000 × *g* for 1 min) from 250 µl of a gradient-purified HearNPV suspension. The supernatant was discarded and the OBs were re-suspended in 60 µl of sterile distilled water. Sodium carbonate was added to a final concentration of 50 mM and the suspension was incubated at 37 ˚C for 30 min, after which 8 µl of a 10% (w/v) SDS solution was added and it was incubated at 37 ˚C for a further 30 min. The addition of sodium carbonate and SDS are necessary to dissolve the polyhedron protein matrix of the OBs and disrupt the envelopes and nucleocapsids of the occlusion derived virions (ODVs), respectively. The pH of the suspension was adjusted to pH 7 by the addition of 20 µl of Tris-HCl buffer (pH 7.0, 1 M). Viral DNA was then extracted from the neutralised suspension as per the manufacturer's protocol for extracting genomic DNA from tissue, forgoing the initial pre-lysis step. The concentration of DNA in each preparation was determined using a NanoDrop™ ND-1000 spectrophotometer before storage at -20 ˚C.

### 2.3.4. PCR amplification for DGGE

A 355 bp region of the *me53* gene (1080 bp) and a 323 bp region of the *DNA polymerase gene* (1750 bp) were amplified from viral genomic DNA using the me53(1) and DNA-pol primers sets, respectively (primer sequences are listed in Table 2.1). To facilitate DGGE analysis of the amplicons GC clamps were attached to the 5′ end of the me53(1)-R primer and to both the DNA-pol primers (GC clamp sequences listed in Table 2.1). These primers were designed by Baillie & Bouwer (2011) using the MELTingeny 1.0 software (Ingeny International), which identifies amplicons which have melting profiles suitable for DGGE analysis and allows the addition of GC clamps to the primers. In addition the primers were

designed to anneal to regions in the two genes where no single nucleotide polymorphisms (SNPs) were detected following the analysis of whole genome resequencing data for the HearNPV genome (Baillie & Bouwer, 2012b).

PCR amplification was performed using Kapa Hifi Hotstart DNA polymerase kits (Kapa Biosystems) as per the manufacturer's instructions, with 50 ng of viral DNA added as template to each 25 µl reaction volume. The Kapa Hifi Hotstart DNA polymerase was chosen as it is a robust, high fidelity enzyme which has been shown to have a low amplification bias relative to other commonly used high fidelity DNA polymerases (Brandariz-Fontes *et al.*, 2015; Oyola *et al.*, 2012; Quail *et al.*, 2012). A "touchdown" PCR protocol (Don *et al.*, 1991) was used to reduce the amplification of non-target sites in the HearNPV genome. The PCR protocol was as follows: an initial denaturation step at 95 ˚C for 5 min, followed by 10 amplification cycles of 95 ˚C for 30 s, an annealing step for 30 s and 72 ˚C for 30 s. The annealing temperature was initially 10 ˚C above the "touchdown" annealing temperature for the primer set, and was reduced by 1 ˚C with each cycle. Once the primer-specific "touchdown" annealing temperature was reached, a further 15 amplification cycles were conducted at this annealing temperature. A final extension step at 72 ˚C for 5 min was included to ensure the complete polymerisation of all amplicons. For the reamplification of cloned gene fragments, the PCR conditions were as described above except that 10 ng of plasmid DNA was added as template to these reactions and the PCR protocol consisted of 20 amplification cycles using the primer-specific annealing temperature, without any preceding "touchdown" cycles.

The quality of amplicons was assessed by electrophoresis on 1% (w/v) agarose gels, stained with 1 × GelRed™ nucleic acid stain (Biotium). Gels were viewed using a Chemidoc™ MP system (Biorad) at an excitation wavelength of 302 nm. The concentration of amplicons in each reaction was determined using the gel quantification function of the Image Lab™ software (Biorad), with selected bands of the MassRuler™ DNA ladder (Thermo Scientific) used as reference bands.

### 2.3.5. DGGE conditions

The electrophoretic conditions and gel composition for the DGGE assay used in this study for the targeted HearNPV *me53* and *DNA polymerase* gene regions were previously optimised by

Baillie & Bouwer (2011). The optimised conditions result in DGGE profiles with clearly resolved bands. DGGE was performed with an Ingeny PhorU-2 apparatus (Ingeny International) and the linear denaturation gradient was formed using a conventional gradient mixer (Ingeny International). The gels were 8% (w/v) polyacrylamide, prepared in 0.5 × TAE (20 mM Tris-HCl, 10 mM sodium acetate, 0.5 M EDTA, pH 8.0), with a linear 20-50% urea-formamide denaturation gradient (100% denaturants is 7 M urea and 40% (v/v) formamide). For each sample, a standardised amount of amplicons (30 ng), mixed with an equal volume of the loading dye (5 mM Tris-HCl (pH 7.6), 0.015% (w/v) bromophenol blue, 30 mM EDTA and 30% (v/v) glycerol) was loaded onto gels. Electrophoresis was performed in 0.5 × TAE for 16 h at a constant temperature and voltage of 60 ˚C and 100 V, respectively. Gels were stained with 0.1 × SYBR Green I nucleic acid stain (Sigma-Aldrich), freshly prepared in 0.5 × TAE, for 45 min post-electrophoresis. Stained gels were viewed using a Chemidoc™ MP system (Biorad) at an excitation wavelength of 300 nm.

### 2.3.6. Analysis of DGGE profiles

GelCompar II V6.0 software (Applied Maths) was used to analyse DGGE profiles. The software was first used to correct for any gel-associated variation in profiles across the width of a gel as well as between different gels. During the gel correction process the positions of bands in a standard profile were used as reference positions. The standard profile was developed in-house, with five distinct bands distributed along its length, and three of these standard profiles were included at equal intervals across the width of each gel. GelCompar was then used to fit Gaussian curves to the densitometric profile of each DGGE profile at the positions of software detected bands (band detection settings: 3% minimum profiling). The area under each Gaussian curve reflected the relative intensity of the corresponding band (Fromin *et al.*, 2002). Band assignments were manually corrected in areas of any profile where the software had incorrectly assigned a band to an artefactual gel spot or failed to detect a faint but clearly visible band that correlated with a peak in the densitometric profile. Relative band intensity values were transformed using the arcsine square root transformation, before calculating the mean and standard error of the mean (SE) for replicate DGGE profiles. This was followed by a back-transformation to yield the reported mean and SE values for the relative intensities of bands in replicate profiles.

In order to assess the similarity of DGGE profiles, pairwise similarity matrices of the profiles for the *DNA polymerase* and *me53* gene regions were constructed using the cosine correlation coefficient algorithm of the software (algorithm settings: 1% profile alignment optimisation, 0% curve smoothing). The cosine correlation coefficient was chosen as it takes into account both the relative position and intensity of bands when comparing two profiles. The similarity matrices were used to create dendrograms showing the mean similarity between profiles, using the UPGMA (Unweighted Pair Group Method with Arithmetic Mean) cluster analysis function of the software.

### 2.3.7. DGGE band extraction and reamplification

The gel composition and electrophoresis conditions for DGGE gels used for band extraction and reamplification were the same as described above. However, these gels were stained post-electrophoresis for 30 minutes with a 5 µg/ml ethidium bromide (Sigma-Aldrich) solution, and destained in distilled water for 15 minutes, to allow bands to be viewed on a standard UV transillumination table. A sterile needle was used to extract amplicons from a DGGE band of interest by puncturing the gel repeatedly across the width of the band. The needle was immersed in a PCR tube containing a PCR master mix, providing template for the reamplification reaction. Kapa Hifi Hotstart DNA polymerase kits were used with the me53(1) and DNA-pol primer sets, as described previously. A "touchdown" PCR protocol was not used as the template material for these reactions consisted of previously amplified gene fragments and the risk of amplifying non-target HearNPV genomic sequences was considered low. The PCR protocol consisted of an initial denaturation step at 95 ˚C for 5 min, followed by 30 amplification cycles comprising of 95 ˚C for 30 s, 30 s at the primer-specific annealing temperature and 72 ˚C for 30 s. A final extension step at 72 ˚C for 5 min was included to ensure the complete polymerisation of all amplicons. The quality of amplicons was assessed by agarose electrophoresis, as described above.

### 2.3.8. Cloning of *me53* and *DNA polymerase* amplicons reamplified from DGGE bands

Amplicons reamplified from DGGE bands were purified using Agencourt® AMPure® XP beads (Beckman Coulter), as per the manufacturer's protocol. The purified amplicons were cloned into pJET1.2/blunt cloning vectors (Thermo Scientific), as per the manufacturer's instructions. NEB® 5-alpha competent *E. coli* (New England Biolabs) were transformed with

the recombinant vectors using the high efficiency transformation protocol provided by the manufacturer. Transformants were selected by plating cells on LB agar plates supplemented with ampicillin (100 µg/ml). Non-recombinant pJET1.2/blunt cloning vectors express a lethal restriction enzyme following transformation, resulting in only those transformants containing recombinant vectors being able to grow on the LB-ampicillin plates. Four colonies were randomly selected for each cloned DGGE band and used to inoculate overnight cultures of LB-media supplemented with ampicillin (100 µg/ml). Plasmid DNA was isolated from each overnight broth culture with the Zyppy™ plasmid miniprep kit (Zymo Research), as per the manufacturer's protocol. Isolated plasmids were stored at -20 ˚C until their inserts were sequenced and reamplified for DGGE analysis.

### 2.3.9. "Reconditioning" PCR

The effectiveness of a "reconditioning" PCR protocol (Thompson *et al.*, 2002) at reducing the formation of DGGE artefactual bands was assessed. The "reconditioning" PCR protocol consisted of an initial 20 cycle reamplification of cloned *me53* or *DNA polymerase* inserts, as described in Section 2.3.4, after which the reaction products were "reconditioned" by diluting them 10 × in a 25 µl fresh reaction mixture and conducting a further four amplification cycles. For the *me53* reconditioning PCR, equimolar clone mixtures containing either two or five different sequence variants, previously isolated from DGGE bands, were used as template in the initial 20 cycle reamplification, with four replicate reactions for each clone mixture. Similarly, for the *DNA polymerase* PCRs an equimolar clone mixture of the two *DNA polymerase* sequence variants isolated from DGGE bands were used as template. The me53(1) and DNA-pol primer sets were used with Kapa Hifi Hotstart DNA polymerase kits for all amplification reactions.

### 2.3.10. Sanger dideoxy sequencing

Both the forward and reverse reads of cloned *me53* and *DNA polymerase* inserts were sequenced. Sequencing was performed by Inqaba Biotec (Pretoria, South Africa) using the pJET1.2 forward and reverse sequencing primers, supplied with the pJET1.2/blunt cloning vector, and the BigDye® Terminator v3.1 cycle sequencing kit (Applied Biosystems). Sequences were resolved on an ABI 3500xL automated genetic analyser (Applied Biosystems). Direct sequencing of *me53* amplicons was conducted as described for the

sequencing of cloned *me53* inserts but the me53(S1)-F and me53(S1)-R primers (primer sequences are listed in Table 2.1) were used to sequence the forward and reverse reads, respectively.

Sequence chromatograms were viewed in Finch TV V1.4 (Geospiza) and poor quality sequence regions were trimmed from the 5' and 3' ends of each read. The forward and reverse reads were aligned in BioEdit V7.2.5, checked for consensus and consolidated into a single sequence. The forward and reverse primer sequences were trimmed from each end of the sequences, following which the sequences were manually aligned in BioEdit. Nucleotide polymorphisms between the different *me53* and *DNA polymerase* clone sequences were shown using the Highlighter tool (http://www.hiv.lanl.gov) (Keele *et al.*, 2008).

### 2.3.11. Deep amplicon sequencing

Deep amplicon sequencing on the MiSeq (Illumina) sequencing platform was used to identify and characterise *me53* sequence variants in a HearNPV population (Population P4). The *me53* region evaluated in the DGGE assay was amplified from viral genomic DNA using the me53(MS1)-F and me53(MS1)-R primers (primer sequences are listed in Table 2.1). The primers were designed to amplify a 307 bp region of the *me53* gene, within the same region amplified by the DGGE me53(1) primer set. Baillie & Bouwer (2012a) previously detected multiple SNPs in this region of the *me53* gene and the significant overlap of forward and reverse MiSeq reads produced using the me53(MS1) primer set allowed for greater confidence in assessing nucleotide sequence variation in this region (Edgar & Flyvbjerg, 2015). Adaptor sequences were added to the 5′ end of the target-specific sequences of the forward and reverse primers (adaptor sequences are listed in Table 2.1), which allowed Nextera XT® (Illumina) indexing tags to be added to amplicons during the preparation of the sequencing library. The indexing tags allow for multiple samples to be pooled in a single MiSeq sequencing run and for reads to be subsequently demultiplexed prior to analysis.

Amplification was performed using Kapa Hifi Hotstart DNA polymerase kits in duplicate 25 µl reactions, with 25 ng of viral genomic DNA added to each reaction as template. The "touchdown" PCR protocol described in Section 2.3.4 was used. The duplicate reactions were subsequently pooled to reduce the effect of any potential PCR drift on the relative abundance of sequence variants in either amplicon population (Polz & Cavanaugh, 1998). PCR products

were purified using Agencourt® AMPure® XP beads, as per the manufacturer's protocol. Nextera XT® indexing tags were added to the forward and reverse amplicon strands during a limited cycle PCR using the FailSafe™ PCR system (Epicentre). The 25 µl reaction consisted of 5 µl of each Nextera XT® index, 0.5 µl of FailSafe™ PCR enzyme, 25 µl of FailSafe™ PCR Premix E, 13.5 µl of sterile distilled water and 1 µl of the purified amplicons. The PCR conditions were 72 ˚C for 3 min, 95 ˚C for 30 s, 12 cycles of 95 ˚C for 30 s, 55 ˚C for 30 s and 72 ˚C for 30 s, followed by a final extension step of 72 ˚C for 5 min. The indexed amplicons were purified using Agencourt® AMPure® XP beads and the addition of the indexing tags to the amplicons was confirmed by assessing the size of the amplicons using the Agilent high sensitivity DNA assay on an Agilent 2100 Bioanalyzer system (Agilent Technologies). The indexed sample was quantified on a Qubit™ 3.0 (Invitrogen) fluorometer using the Qubit™ dsDNA HS Assay (Invitrogen), as per the manufacturer's protocol. The sample was subsequently diluted to 8 nM and denatured by adding an equal volume of freshly prepared 0.2 N NaOH. The denatured sample, at a final concentration of 12 pM with 15% Phix control (Illumina), was submitted for sequencing on a MiSeq instrument using the 2 × 300 paired-end chemistry. Sequencing was conducted at the sequencing core facility of the National Institute for Communicable Diseases (Johannesburg, South Africa).

### 2.3.12. Deep sequencing data analysis

FASTQ files for the forward and reverse reads were automatically generated by the MiSeq platform's software. The command-line tool cutadapt (Martin, 2011) was used to remove any reads containing uncalled bases and to trim both partial and complete *me53* primer sequences from the 5' end of paired-end reads. Any read pairs where primer sequences were not trimmed from both the forward and reverse read were discarded from further analysis. The USEARCH version 9 software suite (Edgar & Flyvbjerg, 2015) was used for all subsequent read processing steps. Paired-end reads were merged using the fastq_mergepairs comman and the fastq_filter algorithm was subsequently used to discard any merged reads with greater than 0.5 total expected errors (Edgar, 2013; Edgar & Flyvbjerg, 2015). The remaining merged reads were collapsed to unique sequences using the fastx_uniques algorithm. The unique sequencing reads were screened for potentially chimeric sequences using the uchime2_denovo algorithm (Edgar, 2016; Edgar *et al.*, 2011). To enable the automated processing of MiSeq reads, the above read processing steps were incorporated into a single Z shell (zsh) script.

Only sequence read variants with a relative abundance ≥ 0.1% were included in the following analyses. Although this cut-off limit may lead to a conservative assessment of the amount of sequence diversity in the HearNPV population, there is a high probability that low frequency sequence variants represent PCR or sequencing artefacts (Bokulich *et al.*, 2012; Edgar, 2013). The Highlighter tool was used to show nucleotide polymorphisms between the different sequence variants. The relative abundances for the sequence variants were plotted using the ggplot2 package (Wickham, 2009) in R version 3.3.1 (R Core Team, 2016).

## 2.4. Results

### 2.4.1. Reproducibility of DGGE profiles

In order to assess the reproducibility of *me53* and *DNA polymerase* DGGE profiles for HearNPV populations, three replicate PCRs were conducted with each primer set for three different HearNPV populations (P1, P2 and P3). The PCR products for each reaction were loaded on two replicate DGGE gels. HearNPV populations P1, P2 and P3 were chosen after an initial PCR-DGGE screening of multiple populations showed that these populations had different numbers of bands in their *me53* and *DNA polymerase* DGGE profiles (data not shown). This was in order to assess whether DGGE profiles with low, intermediate or high numbers of bands would consistently prove to be reproducible across different PCRs and gels.

Replicate profiles for HearNPV populations P1, P2 and P3 had very similar banding patterns with mean similarities of 96.5%, 95.7% and 92.8% for *me53* profiles, and 98.6%, 98.3% and 99.0% for *DNA polymerase* profiles, respectively (Figure 2.2a-b). In both the *me53* and *DNA polymerase* dendrograms, replicate profiles for each HearNPV population which originated from the same DGGE gel clustered together, while profiles for each PCR replicate did not appear to have a greater degree of similarity (Figure 2.2a-b). This shows that while PCR-DGGE generates reproducible *me53* and *DNA polymerase* profiles for the three HearNPV populations across replicate PCRs and gels, gel-to-gel variation does have a small but detectable effect on the degree of similarity between profile replicates. This effect was more pronounced for *me53* profiles where bands could occur across a greater proportion of a gel's length. This could make the position of bands in *me53* profiles more sensitive to any variation in the denaturing gradient between gels that may be produced during the gel pouring step.

### 2.4.2. Multiple sequences associated with individual bands

In this study bands in *DNA polymerase* DGGE profiles were abbreviated "dB" while those in *me53* profiles were abbreviated "mB". Similarly, *DNA polymerase* and *me53* genetic variants were referred to as "dV", while *me53* variants were referred to as "mV". During the preliminary stages of this study, in order to assess whether each band in a DGGE profile represented a single genetic variant, DNA was extracted from six bands (Figure 2.3a, mB1-

mB6) in the *me53* DGGE profile of a fourth HearNPV population (P4), reamplified and sequenced. The sequencing chromatograms for the amplicons from bands mB2-mB6 contained multiple double peaks indicating the amplicon populations contained heterogeneous sequences (sequencing data not shown). When amplicons from the six band reamplifications were loaded onto a second DGGE gel the resulting profiles contained multiple bands and did not only consist of a single band at the position of the original reamplified band (Figure 2.3a). The DGGE profile for reamplified band mB1 was unique in that it only contained a single dominant band at the mB1 position and a number of additional faint, minor bands often referred to as "ghost bands" in the literature (Figure 2.3b). The profiles for the reamplified mB2-mB6 bands in addition to containing multiple minor bands also contained multiple dominant bands, some of which occurred at the mB1-mB6 positions and others that were not present in the original P4 *me53* profile (Figure 2.3b).

These preliminary results suggested that the reamplified amplicon populations for bands mB1-mB6 may have contained multiple genetic variants. In order to further assess the number and identity of genetic variants associated with individual bands in *me53* and *DNA polymerase* profiles, amplicons from band reamplifications were cloned and several clones from each band were sequenced. In addition, each sequenced *me53* and *DNA polymerase* clone insert was reamplified and loaded onto a new DGGE gel in order to establish the DGGE profile associated with each sequence variant. Five *me53* sequence variants (mV1-mV5) were identified from twenty four clones sequenced (Figure 2.4a). Whereas the variants mV3, mV4 and mV5 were each isolated from a single band, variants mV2 and mV1 were isolated from four and all six of the P4 bands, respectively. However, upon reamplification the five *me53* clone variants were only associated with three distinct DGGE profiles (Figure 2.3 and Figure 2.4, mP1-mP3). Each of these profiles contained a single dominant band and multiple low frequency "ghost" bands (Figure 2.5). The variants mV1 and mV4 both produced the mP1 profile with a dominant band at the mB1 position, while variants mV2 and mV5 produced the mP2 profile with the dominant band at the mB2 position. The mV3 variant was the only variant that produced the mP3 profile, which had a dominant band, mB7, between the mB2 and mB3 positions.

From the 21 *DNA polymerase* clones sequenced only two sequence variants were isolated from the four dominant bands in the P4 *DNA polymerase* profile (Figure 2.6b). The dV1

variant, which accounted for twenty of the twenty one sequenced clones, was isolated from all four dominant bands (dB1-dB4) in the P4 *DNA polymerase* profile, while the second variant, dV2, was only isolated from the dB2 band (Figure 2.6a-b). The DGGE profiles for the reamplification reactions of these two clone variants produced two distinct profiles, each containing a single dominant band and multiple low frequency "ghost" bands (Figure 2.6a, dP1 and dP2). The dV1 variant was associated with a single dominant band at the dB1 position, while dV2 was associated with a single dominant band at the dB2 position (Figure 2.6a). These results indicate that for the P4 HearNPV population the mB3, mB4, mB5 and mB6 bands of the *me53* profile and the dB3 and dB4 bands of the *DNA polymerase* profile were artefactual bands that formed due to an interaction between the different sequence variants, either during the amplification reactions or on the gel during electrophoresis.

### 2.4.3. Origins of DGGE profile artefacts

PCR template mixtures and post-PCR amplicon mixtures of the different clone variants were used to determine whether the artefactual bands detected in the *me53* and *DNA polymerase* DGGE profiles were a consequence of multi-template PCR or due to interactions between heterogeneous amplicon sequences on the gel during electrophoresis. Artefactual bands were generated in both *me53* and *DNA polymerase* profiles when equimolar mixtures of *me53* or *DNA polymerase* clone variants were used as template for reamplification reactions. When equimolar mixtures of the mV1 and mV2 *me53* variants were used as template in four replicate reamplification reactions, the artefactual bands mB3 and mB4 occurred in all four DGGE profiles (Figure 2.7a, A1). When all five of the *me53* clone variants were included as template in four replicate reamplification reactions, all four of the artefactual bands observed in the P4 HearNPV population's *me53* DGGE profile (mB3-mB6), as well as multiple additional artefactual bands, occurred in all four profiles (Figure 2.7a, A2). The DGGE profiles for post-PCR amplicon mixtures of *me53* clone variants, where each variant was reamplified in a separate PCR, did not contain any of the artefactual bands and only consisted of combinations of the three dominant bands from the mP1-mP3 profiles associated with the five variants (Figure 2.7a, 3A and 4A). The two *DNA polymerase* artefactual bands, dB3 and dB4, were also only produced when a mixture of the dV1 and dV2 *DNA polymerase* variants were used as template for replicate reamplification reactions, and were not present in profiles

where post-PCR amplicon mixtures of the two variants were loaded onto the DGGE gel (Figure 2.8a, A1 and A2).

These results indicate that it was not an interaction between amplicons of different sequence variants on the gel during electrophoresis that resulted in artefactual bands on the *me53* and *DNA polymerase* gels. Rather, artefactual bands appear to have their origins in the multi-template PCR step, as they were only produced when two or more *me53* or *DNA polymerase* sequence variants were co-amplified in the same PCR. The results for the *me53* clone mixture reamplifications also suggest that the number of artefactual bands in a profile will increase with the number of different sequence variants present in the PCR template population. For clarity, the mB1, mB2 and mB7 *me53* bands and dB1 and dB2 *DNA polymerase* bands, which were not produced as a consequence of the multi-template PCR, are referred to as "true" bands for the remainder of the study. All artefactual bands occurred at positions above the true bands in *me53* and *DNA polymerase* profiles. This suggested that the DNA fragments in the artefactual bands had lower melting temperatures than the amplicons in the true bands.

Given that the artefactual bands were a consequence of the multi-template PCR step and that the DNA fragments in these bands had lower melting temperatures than the amplicons in the true bands, it seems likely that the artefactual bands were caused by heteroduplex amplicons (Ruano & Kidd, 1992; Thompson *et al.*, 2002). Heteroduplex amplicons are thought to be formed by the re-annealing of denatured heterogeneous PCR products during the last annealing step of a PCR when conditions for primer-template annealing are suboptimal (Muyzer & Smalla, 1998; Ruano & Kidd, 1992). Due to nucleotide base mismatches between the DNA strands in a heteroduplex molecule its melting temperature will be lower than for the homoduplex molecules of each constituent strand. Thompson *et al.* (2002) developed a simple modified PCR protocol, called "reconditioning" PCR, which they showed could reduce the formation of heteroduplexes in multi-template PCRs. The efficacy of the "reconditioning" PCR protocol was assessed in the current study in an attempt to reduce the formation of the heteroduplexes suspected of causing the artefactual bands. However, it was found that the modified PCR protocol did not prevent or significantly reduce the formation of artefactual bands in either the *me53* or *DNA polymerase* DGGE profiles (Figure 2.7b and Figure 2.8b). Although it is also possible that chimeric amplicons caused some or all of the *me53* and *DNA polymerase* artefactual bands, this was considered unlikely as none of the

cloned sequence variants isolated from *me53* bands were identified as chimeric by the UCHIME algorithm during the deep sequencing data analysis, and only two *DNA polymerase* sequence variants were isolated from DGGE bands. Chimeric amplicons would also not necessarily have a lower melting temperature than their two parent sequences, and as a result chimeric artefactual bands would not always occur above the true bands in a DGGE gel profile.

The degree of similarity between two or more microbial communities is often calculated based on a comparison of the densitometric curves of their DGGE profiles, where both the number and relative intensity of bands in each profile are used to calculate pairwise similarity or distance values such as the cosine correlation coefficient (Fromin *et al.*, 2002). This approach can only be justified if it is known that changes in the underlying abundance of sequence variants in a population affect the relative intensity of bands on the gel in a predictable way. However, the formation of artefactual heteroduplex bands has in some instances been observed to cause the relative intensity of bands in a profile to no longer reflect the underlying relative abundance of their associated variants in a community (Ruano & Kidd, 1992). Seven template mixtures of the mV1 and mV2 variants at different ratios were used to evaluate whether the formation of artefactual heteroduplex bands, mB3 and mB4, affected the relative intensities of the mB1 and mB2 bands (Figure 2.9a-c). It was observed that the relative intensities of the two artefactual bands remained approximately equal and constant over the ten-fold change in the ratio of mV1:mV2 in the PCR template mixture (Figure 2.9c). The relative intensity of the mB1 and mB2 bands increased as the clone variant associated with each band became more abundant in the PCR template mixture (Figure 2.9c). However, it should be noted that the ratio of the relative band intensities for mB1 and mB2 did not exactly match the ratio of the clone variants mV1 and mV2 in the template mixtures of each of the reamplification reactions. These results show that although the relative intensity of heteroduplex artefactual bands stayed relatively constant over a broad range of clone variant ratios, the relative intensity of the mB1 and mB2 bands reflected the underlying relative abundance in the PCR template mixture of the clone variant associated with each band. However, the ratio of the relative intensities of the true bands should not be used to infer the proportional variant composition of a HearNPV population.

### 2.4.4. Deep *me53* amplicon sequencing

Sequencing of cloned *me53* amplicons isolated from the P4 *me53* DGGE profile identified five sequence variants, whereas only two sequence variants were isolated from the P4 *DNA polymerase* profile. From an initial screening of different HearNPV populations, the *me53* DGGE profiles also showed greater complexity than the *DNA polymerase* profiles, with multiple dominant bands distributed along the length of the profiles of many HearNPV populations. These results suggested that there were a greater number of *me53* sequence variants in the HearNPV populations under consideration than *DNA polymerase* variants. In order to further evaluate the ability of PCR-DGGE to detect the full *me53* genetic variant complement of a HearNPV population, a second technique, deep amplicon sequencing, was used to assess the *me53* genetic diversity of the P4 HearNPV population.

Deep *me53* amplicon sequencing initially identified seventeen sequence variants (S1-S17) in the P4 population which each accounted for ≥ 0.1% of the total 197942 processed sequencing reads (Figure 2.4b, S1-S17). However eleven of these were identified as sequence chimeras by the UCHIME2 algorithm of the USEARCH software. All five of the *me53* clone variants were detected and classified as non-chimeric sequences by UCHIME2 (Figure 2.4a-b, mV1-mV5), with only one sequence variant identified by deep sequencing not being detected using the DGGE band extraction and cloning approach (Figure 2.4b, S7). The *me53* amplicon deep sequencing data for the P4 HearNPV population suggests that even though a limited number of clones for amplicons isolated from each dominant DGGE band were sequenced, the majority of genetic variants in the P4 HearNPV population were detected using the band extraction and cloning approach. The relative abundance of sequence variants in the deep sequencing data also corresponded to the number of clones sequenced for each of the five clone variants (mV1-mV5). The mV1and mV2 variants were the most abundant and second most abundant sequence variants in both the deep sequencing and clone sequencing data, respectively, while the mV3, mV4 and mV5 variants, which were each only represented by one or two sequenced clones, together accounted for less than 5% of the processed deep sequencing reads (Figure 2.4a-b).

However, a number of studies have shown that amplification bias during multi-template PCR can result in the observed relative abundance of sequence variants in the deep sequencing data not being an accurate reflection of the true relative abundance of sequence variants in the

microbial community (Amend *et al.*, 2010; Brooks *et al.*, 2015; Pinto & Raskin, 2012; Taylor *et al.*, 2016). In order to assess whether amplification bias could be responsible for the relative abundance of the six detected sequence variants, two mock populations consisting of different template ratios of mV1, mV2 and mV5 clone variants were amplified and sequenced using the same protocols as used for the P4 population. These three sequence variants were chosen as they covered the complete relative abundance range of variants in the P4 deep amplicon sequencing data, with mV1 having the greatest abundance (68%), mV2 an intermediate abundance (21.8%) and mV5 the lowest abundance (0.7%). The observed relative abundance of each sequence variant in the sequencing data was compared to the true relative abundance of each variant in the mock populations, in order to detect the effects of any potential PCR bias. PCR amplification was observed to have a negligible effect on the observed composition of both mock populations, with the observed relative abundance of each sequence variant closely matching their true relative abundance in both mock populations (Table 2.2). Even when the relative abundance of the three variants was reversed in the second mock population relative to their abundance in the P4 deep sequencing data, with mV5 having the greatest abundance (73%) and mV1 having the lowest abundance (9%) in the template mixture, the observed and true relative abundance values closely matched each other (Table 2.2). These results indicate that amplification bias was not responsible for the relative abundance of the sequence variants in the P4 deep amplicon sequencing data.

## 2.5. Discussion

The *me53* and *DNA polymerase* DGGE profiles for HearNPV populations were highly reproducible across replicate PCRs and different gels. This supports the findings of previous studies that have consistently shown PCR-DGGE is a reliable and reproducible fingerprinting technique for discriminating between communities which differ in their microbial composition (Hong *et al.*, 2007; Neilson *et al.*, 2013; Nicolaisen & Ramsing, 2002; Simpson *et al.*, 1999). Hong *et al.* (2007) in particular showed a similar level of similarity between DGGE profiles for replicate PCRs (> 95% mean similarity); although they noted that gel-to-gel variation posed a significant hurdle for accurate profile comparisons, which was not found in the current study. It is clear that PCR-DGGE can be used to discriminate between HearNPV populations on the basis of their *me53* and *DNA polymerase* DGGE profiles and, that on the basis of the existing evidence for PCR-DGGE (Neilson *et al.*, 2013), this approach could be expanded to additional HearNPV genes and probably to other dsDNA virus species with success. However, the results suggest that samples should be loaded on the same gel for the greatest degree of confidence when comparing profiles, as gel-to-gel variation did reduce the similarity of DGGE profile replicates to a small but detectable degree. This also reiterates the importance of adhering to an optimised protocol during the gel-pouring, electrophoresis and staining steps, with the aim of reducing the effects of gel-to-gel variation, even though it appears unlikely that this source of variation could ever be completely eliminated. Steps aimed at reducing variation between gels appear to be of particular importance for profiles with many bands as the effects of gel-to-gel variation on the similarity of replicate profiles was greater the more bands a profile contained.

In order to use DGGE profiles not only as genetic fingerprints but to measure the α-diversity of a microbial community, it is necessary to first establish whether each band in a profile represents a single genetic variant. Several studies have shown that the presence of artefactual bands (Ferris & Ward, 1997; Hong *et al.*, 2007; Neilson *et al.*, 2013; Samarajeewa *et al.*, 2015; Satokari *et al.*, 2001) and bands that are associated with more than a single genetic variant (Hong *et al.*, 2007; Samarajeewa *et al.*, 2015) can make quantitative measures of α-diversity calculated from DGGE profiles inaccurate. When calculating quantitative measures of α-diversity from PCR-DGGE profiles, such as the Shannon diversity index, the number of bands in a profile is taken as representing the richness of a community. Artefactual

bands artificially increase the apparent richness of a community whereas bands associated with multiple genetic variants result in the observed richness being lower than the true richness. Four of the seven bands assessed in the P4 *me53* profile were artefactual bands (mB3, mB4, mB5 & mB6), while two of the four bands assessed in the P4 *DNA polymerase* profile were artefactual bands (dB3 & dB4). Multiple sequence variants were isolated from all of the artefactual *me53* bands. However these were not examples of variants that had similar denaturation dynamics co-migrating during electrophoresis, as these artefactual bands only occurred in the profiles of multi-template PCRs and did not occur in the profiles of the single clone variant reamplifications. The bands at the mB1 and mB2 positions did appear to be examples of variant co-migration, as the mV1 and mV4 variants were both associated with the mB1 band, while the mV2 and mV5 variants were associated with the mB2 band, in the single clone variant reamplification profiles.

The artefactual bands detected in both the *me53* and *DNA polymerase* profiles appeared to consist of heteroduplex amplicons that originated as a result of the multi-template PCRs. Heteroduplexes are a major source of artefactual bands in PCR-DGGE, as evidenced by the large number of studies that report their occurrence (Ferris & Ward, 1997; Hong *et al.*, 2007; Neilson *et al.*, 2013; Samarajeewa *et al.*, 2015; Satokari *et al.*, 2001; Speksnijder *et al.*, 2001). In these studies heteroduplex bands have been found to occur in the upper section of profiles near the sample loading wells. In the *me53* and *DNA polymerase* profile for the P4 population the heteroduplex bands occurred at positions much lower down in the profiles than observed in previous studies. However this is not unexpected due to the high nucleotide sequence similarity between variants in the HearNPV population, and given that heteroduplexes with fewer nucleotide mismatches would migrate further in a gel before denaturation. It is also clear from the profiles for the template mixtures of clone variants that the formation of heteroduplex bands occurs in a highly reproducible manner and that the number of heteroduplex bands increases with number of sequence variants in the template mixture. The potential for heteroduplexes to form also appears higher for HearNPV populations relative to the number of heteroduplex bands observed in previous studies on bacterial communities (Ferris & Ward, 1997; Hong *et al.*, 2007; Neilson *et al.*, 2013), where the degree of sequence similarity between members of the community is most likely lower than between members of a HearNPV population. This is supported by the findings of Ferris & Ward (1997) who observed that heteroduplexes were only detected among closely related

sequences in their study. The results also suggest that it is not possible to prevent or reduce the formation of these artefactual bands associated with the HearNPV profiles using a simple reconditioning PCR protocol, as proposed by Thompson *et al.* (2002).

Another common group of artefacts produced by multi-template PCRs are chimeric sequences, with some studies estimating that up to 30% of amplicons from a typical PCR can be chimeric products (Smyth *et al.*, 2010). Although chimeric sequences did account for a significant number of the MiSeq *me53* reads, surprisingly no chimeric sequences were isolated from either the P4 *me53* or *DNA polymerase* profiles. Although chimeric artefactual bands have been identified in previous PCR-DGGE studies (Chen & LaPara, 2006; Reeson *et al.*, 2003), their occurrence does appear to be less frequently reported in the PCR-DGGE literature than heteroduplex bands. It was clear from the number of artefactual bands observed in the P4 DGGE profiles, and from the propensity for artefactual heteroduplex bands to form in the mixed clone variant template profiles, that each band in a HearNPV *me53* or *DNA polymerase* DGGE profile could not be taken to represent a single sequence variant without extensive prior assessment. This meant that quantitative measures of α-diversity calculated from *me53* or *DNA polymerase* profiles would greatly overestimate the amount of genetic diversity in a HearNPV population.

Although previous studies, in particular Neilson *et al.* (2013), have extensively documented the effect of amplification bias in multi-template PCR on the relative intensity of DGGE bands, the degree of amplification bias was not assessed in the current study. Since the heteroduplex artefactual bands already prevent the accurate calculation of quantitative measures of α-diversity for HearNPV populations from DGGE profiles, it was felt it was unnecessary to perform an in-depth study of the potential effects of amplification bias on the relative intensity of bands in HearNPV profiles. However, what was shown was that the relative intensity of two *me53* artefactual heteroduplex bands remained relatively constant and equal over a ten-fold change in the ratios of their two constituent variants in the template mixture, while the relative intensity of the mB1 and mB2 true bands associated with these variants reflected their relative abundance in the template mixture. Due to the fact that both the number of true and artefactual bands and their relative intensities appear reproducible for a HearNPV population with a set variant composition, the similarity of a population's variant composition can be assessed on the basis of the similarity of the densitometric curves for

their respective profiles. In addition, although the number of bands in a DGGE profile greatly overestimates the number of variants in a HearNPV population, the number of bands can still be used as a relative gauge of the number of genetic variants in a HearNPV population, with a greater number of bands being indicative of greater variant richness.

Although DGGE profiles provide a gauge to the relative level of genetic diversity within a HearNPV population they do not provide any information on the nucleotide sequences of the variants in a population. Traditionally this was obtained through the labour intensive process of extracting, cloning and sequencing DNA fragments from DGGE bands, however, more recently deep amplicon sequencing has been used, which is able to detect low frequency genetic variants without having to sample a large number of clones (Kraková *et al.*, 2016; Yu *et al.*, 2015). Deep amplicon sequencing on the MiSeq platform proved superior to the traditional clone sequencing approach in this study, and was able to detect all sequence variants isolated from DGGE bands using the cloning approach, as well as an additional low frequency variant which had not been detected. A number of studies have previously compared the use of PCR-DGGE and deep amplicon sequencing for assessing the genetic diversity of various microbial communities and all have found that deep amplicon sequencing provides greater sensitivity for the detection of low frequency genetic variants relative to the clone sequencing approach (Koo *et al.*, 2016; Kraková *et al.*, 2016; Wang *et al.*, 2015; Yu *et al.*, 2015). There appear to be no advantages to the DGGE band extraction and clone sequencing approach of assessing the nucleotide sequence diversity of a microbial community in conjunction with PCR-DGGE, unless the aim is to establish which sequence variants are associated with certain DGGE bands for validation purposes. However, the cost of deep amplicon sequencing can be prohibitive if the aim of a study is to screen a large number of microbial communities and assess what effect certain variables have on their variant composition and genetic diversity (Goodwin *et al.*, 2016). A practical approach may be to use PCR-DGGE as an initial screening method to assess and compare the relative variant composition of HearNPV populations. Subsequently deep amplicon sequencing could be used to evaluate the underlying differences in the number and identity of sequence variants between populations where PCR-DGGE indicated differences in variant composition existed.

## 2.6. Conclusions

PCR-DGGE can clearly be used as a reproducible and reliable genetic fingerprinting method to discriminate between different HearNPV and potentially other dsDNA virus populations. The propensity for artefactual heteroduplex bands to occur in HearNPV profiles prevents the calculation of accurate quantitative measures of genetic diversity for HearNPV populations from DGGE profiles. However it is still possible to use the number of bands within a DGGE profile as a gauge of the relative level of genetic diversity at targeted gene loci within a HearNPV population. In addition, the relative similarity in variant composition of HearNPV populations can be assessed by comparing the densitometric profiles of their DGGE profiles. Future studies on the factors that affect the genetic diversity of HearNPV populations could use PCR-DGGE and deep amplicon sequencing as part of an integrated analysis pipeline, where HearNPV populations are initially screened using PCR-DGGE, and deep amplicon sequencing is subsequently used to gain in depth information on both the number and identity of sequence variants within selected populations.

**Table 2.1.** Sequences and characteristics of the primers specific to the HearNPV *me53* and *DNA polymerase* genes, used for PCR-DGGE, Sanger sequencing and MiSeq library preparation.

| Primer name | Sequence | Annealing temperature (˚C) | Nucleotide position[a] | Reference |
|---|---|---|---|---|
| me53(1)-F | 5′-GACGGCAGCGTCGCTAGT-3′ | 53 | 14 634 – 14 651 | Baillie & Bouwer 2012b |
| me53(1)-R | 5′-CTGGATGTAATATGTAAGT-3′[b] | | 14 971 – 14 989 | |
| me53(S1)-F | 5′-CAGCGTCGCTAGTTAACC-3′ | 60 | 14 639 – 14 656 | This study |
| me53(S1)-R | 5′-TGGATGTAATATGTAAGTTTGAGC-3′ | | 14 972 – 14 995 | |
| me53(MS1)-F | 5′-CAGCGTCGCTAGTTAACC-3′[c] | 60 | 14 639 – 14 656 | This study |
| me53(MS1)-R | 5′-TGGATGTAATATGTAAGTTTGAGC-3′[d] | | 14 972 – 14 995 | |
| DNA-pol-F | 5′-GTATGACTTATCACGACAATTGC-3′[b] | 61 | 59 080–59 102 | Baillie & Bouwer 2012b |
| DNA-pol-R | 5′-CGGTTTGCATATGTACTCTG-3′[e] | | 59 384–59 403 | |

[a] The nucleotide position of the primer binding region is given relative to the reference HearNPV genome (GenBank accession number: AF271059)

[b] GC clamp (5′-CGCCCGCCGCGCCCCGCGCCCGGCCCGCCGCCCCCGCCCG-3′)

[c] Illumina adaptor sequence (5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3′)

[d] Illumina adaptor sequence (5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3′)

[e] GC clamp (5′-CGCCCGCC-3′)–this short GC clamp was added to the reverse DNA-pol primer in order to give the amplicons ideal melting characteristics.

**Table 2.2.** The true relative abundance of genetic variants mV1, mV2 and mV5 in the two mock populations and their observed relative abundance in the corresponding deep amplicon sequencing data.

| Mock population | Variant | Relative abundance (%) | |
|---|---|---|---|
| | | True | Observed[a] |
| 1 | **mV1** | 33.33 | 36.28 (27673) |
| | **mV2** | 33.33 | 34.70 (26468) |
| | **mV5** | 33.33 | 29.03 (22143) |
| 2 | **mV5** | 72.73 | 71.09 (59152) |
| | **mV2** | 18.18 | 17.98 (14959) |
| | **mV1** | 9.09 | 10.93 (9098) |

[a] The actual number of sequencing reads for each sequence variant in the sequencing data for the two mock populations is given in brackets.

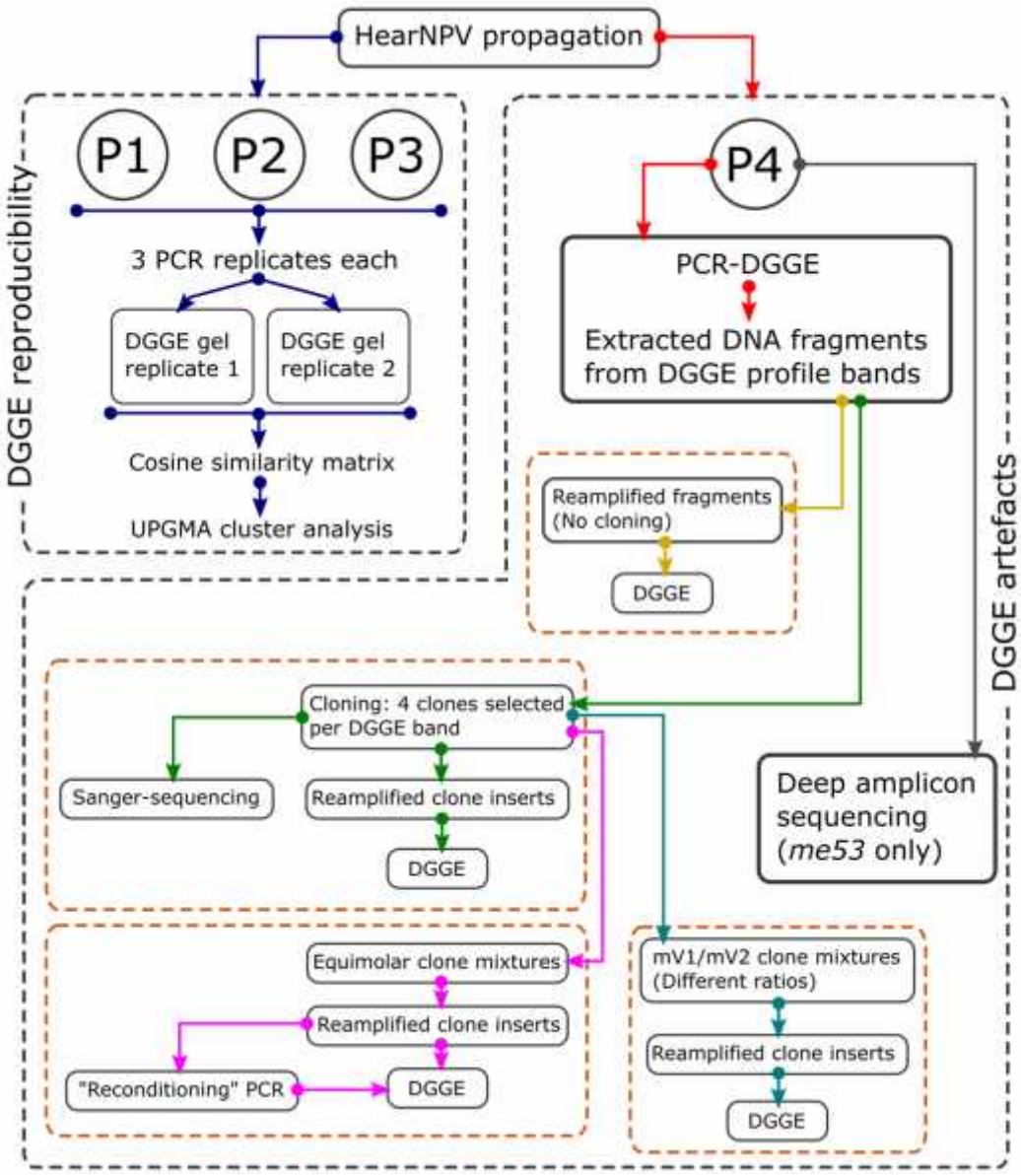



**Figure 2.1.** Methods flowchart showing the steps taken to assess the reproducibility and the prevalence of artefacts in PCR-DGGE gel profiles, when the technique is used to assess the genetic diversity of the *me53* and *DNA polymerase* genes in HearNPV populations. HearNPV populations P1, P2, P3 and P4 were propagated *in vivo* in third instar *H. armigera* larvae. (1) HearNPV populations P1, P2 and P3 were used to evaluate the effects of PCR and gel variation on the reproducibility of *me53* and *DNA polymerase* PCR-DGGE gel profiles. (2) HearNPV population P4 was used to assess the prevalence of artefacts in PCR-DGGE gel profiles and the impact of any artefacts on the ability to use PCR-DGGE to assess the genetic diversity of HearNPV populations. (a) DGGE profiles of P4 band reamplifications were assessed. (b) Amplicons from P4 band reamplifications were cloned into pJET1.2/blunt cloning vectors. The P4 insert of four clones per band were sequenced and their DGGE profiles assessed. (c) The generation of artefactual bands associated with multi-template PCR was assessed. The effectiveness of a "reconditioning" PCR protocol at reducing the formation of DGGE artefactual bands was evaluated. (d) The genetic variant complement of the P4 population was determined using deep amplicon sequencing. (e) The effect of artefactual heteroduplex bands on the relative intensity of true bands was assessed.

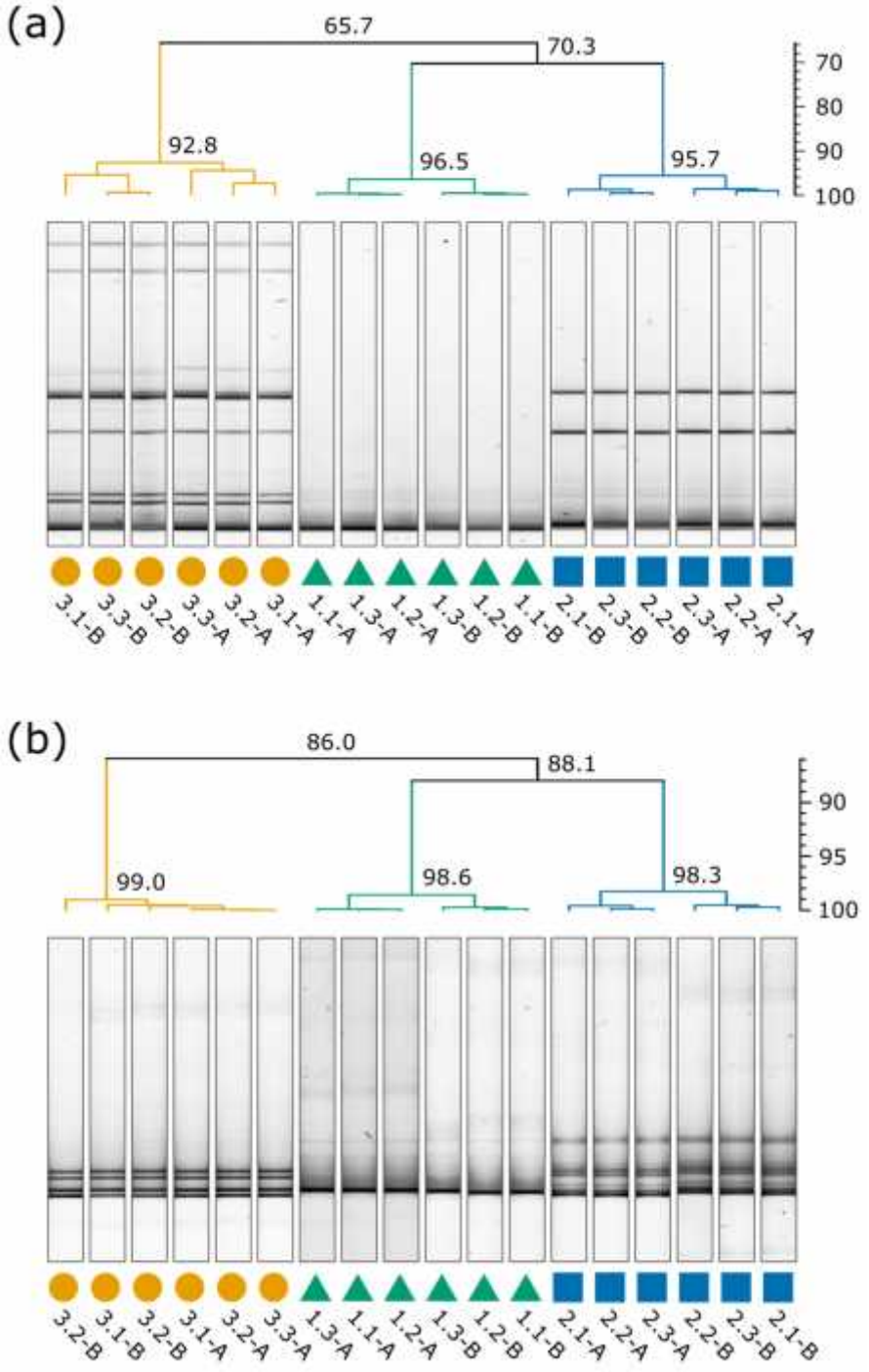



**Figure 2.2.** Similarity of *me53* and *DNA polymerase* DGGE profiles across PCR and gel replicates. **(a)** *me53* and **(b)** *DNA polymerase* replicate DGGE profiles for three HearNPV populations, P1, P2 and P3. For each HearNPV population three replicate PCRs were performed with the me53(1) and DNA-pol primer sets and an equal amount of each PCR product was loaded onto two replicate DGGE gels. Profiles derived from HearNPV populations P1, P2 and P3 are indicated by a triangle, square and circle beneath each profile, respectively. PCR replicates for the three populations are labelled 1.1-1.3, 2.1-2.3 and 3.1-3.3, respectively, with the letter A or B indicating to which gel the profile belongs. The dendrogram shows the mean percentage similarity between the different profiles.

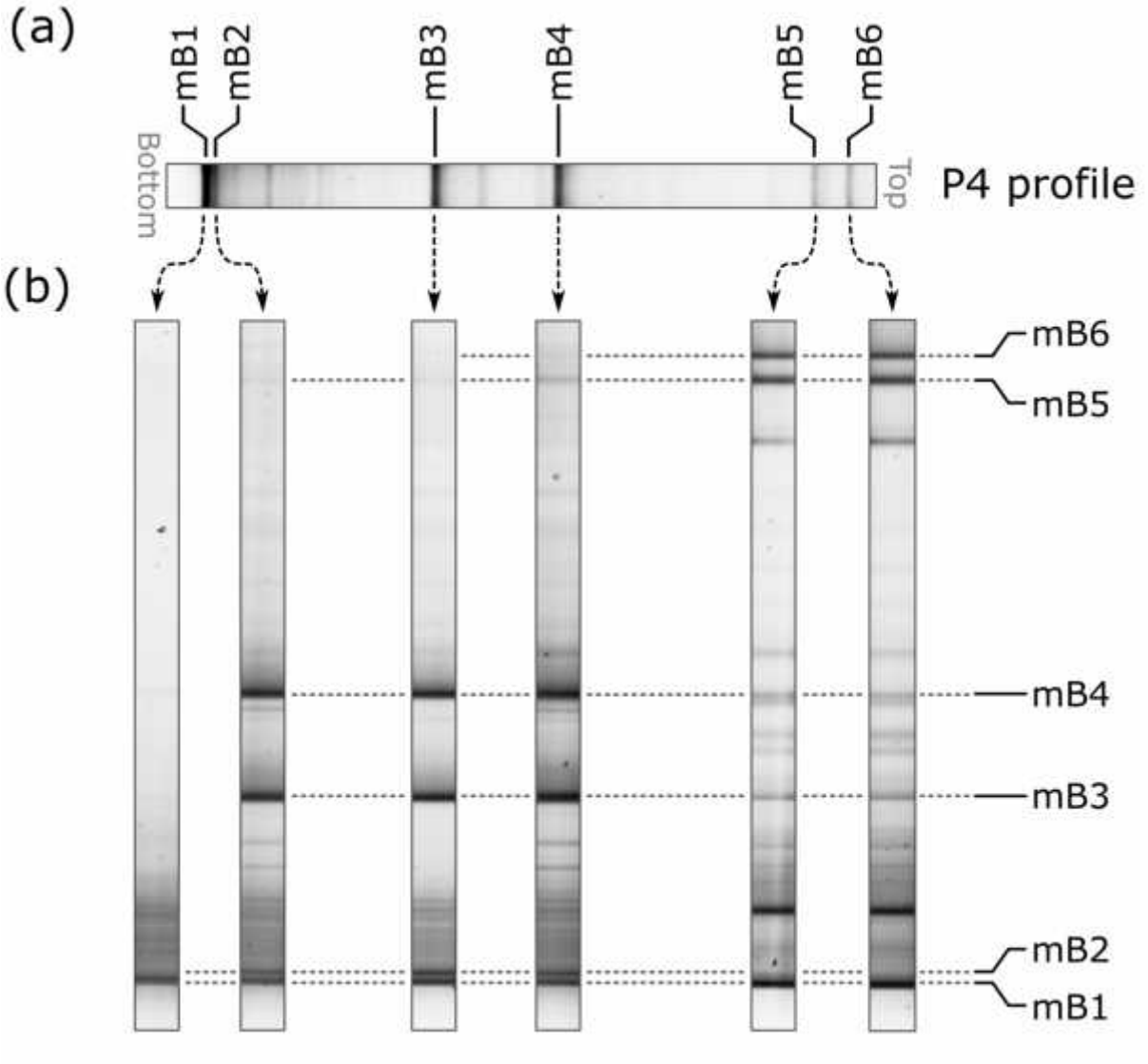



**Figure 2.3.** DGGE profiles for *me53* amplicons originating from the reamplification of DNA extracted from six bands in the *me53* profile of HearNPV population P4. The six bands in the P4 *me53* profile **(a)** and their corresponding bands in the six reamplification profiles **(b)** are labelled mB1-mB6.

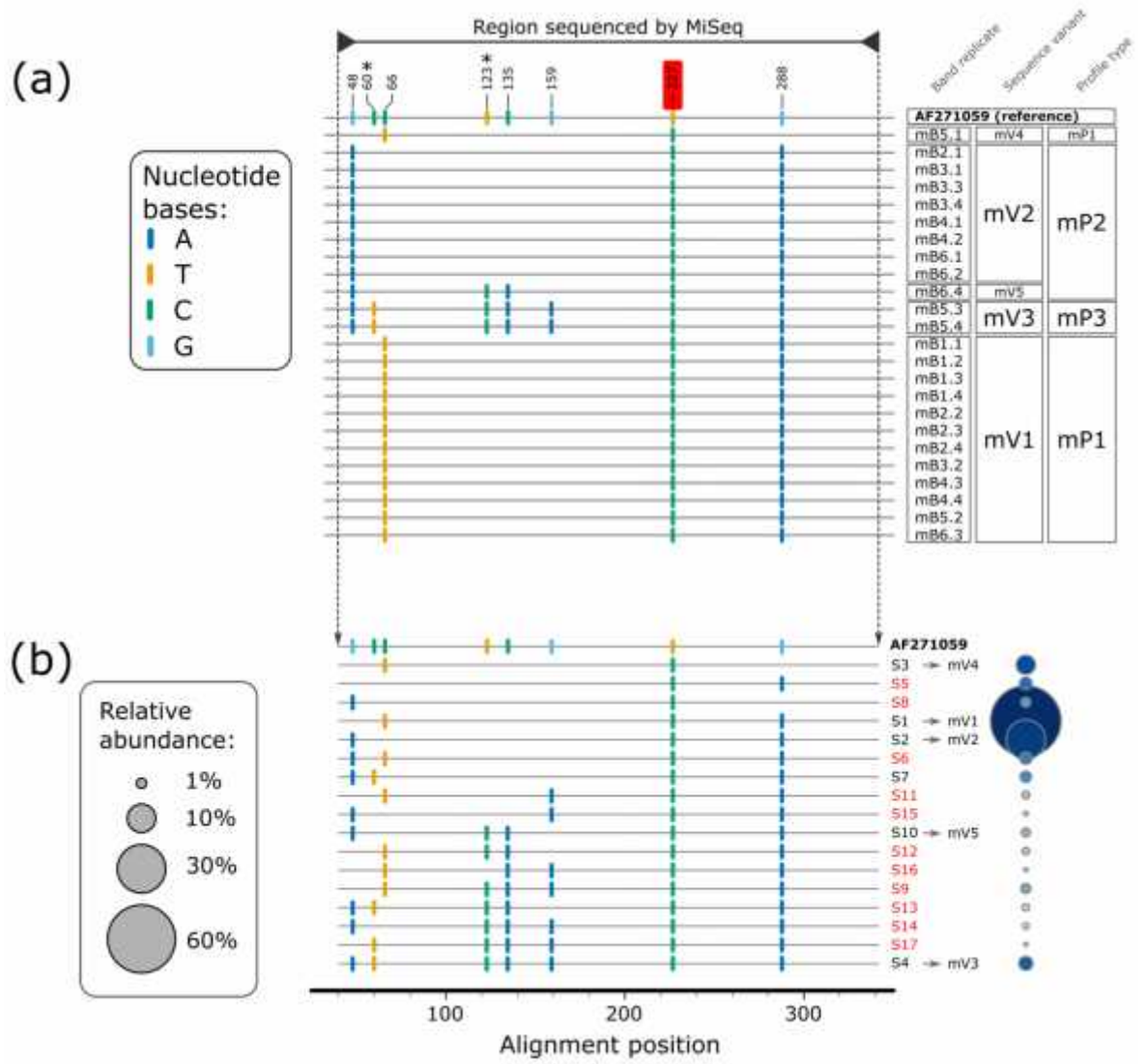



**Figure 2.4.** Nucleotide sequences of *me53* variants. Highlighter plots showing the nucleotide variation between **(a)** Sequences of *me53* clones and **(b)** MiSeq read sequences for the P4 HearNPV population. The clone labels in the first column on the right of panel (a) denote from which band in the original P4 *me53* DGGE profile each clone was originally derived and whether they were clone replicate 1-4 selected for that band. Overall there were five clone sequence variants (mV1-mV5) but these only gave rise to three distinct DGGE profile types (mP1-mP3). All five of the *me53* clone sequences (mV1-mV5) matched sequences identified by MiSeq. MiSeq read sequences which were flagged as being potential artefactual chimeric sequences by the UCHIME algorithm are shown in red (e.g. S5). Only one MiSeq read sequence (S7), which was not flagged as being chimeric, was not detected using the band extraction and cloning approach. The relative abundance of MiSeq reads for each sequence variant is shown by the size of the corresponding circle to the right of panel (b). The nucleotide positions in the highlighter plots are given relative to a reference *me53* gene sequence (GenBank accession number: AF271059). The positions of SNPs relative to the reference *me53* gene sequence are shown as coloured bars, with each colour denoting one of the four DNA bases. SNPs at positions marked with an asterisk (*) were not present in any of the previously published HearNPV genomes (Baillie & Bouwer, 2012a; Chen *et al.*, 2001; Noune & Hauxwell, 2015, 2016), while those highlighted in red (nucleotide position 227) are non-synonymous polymorphisms.

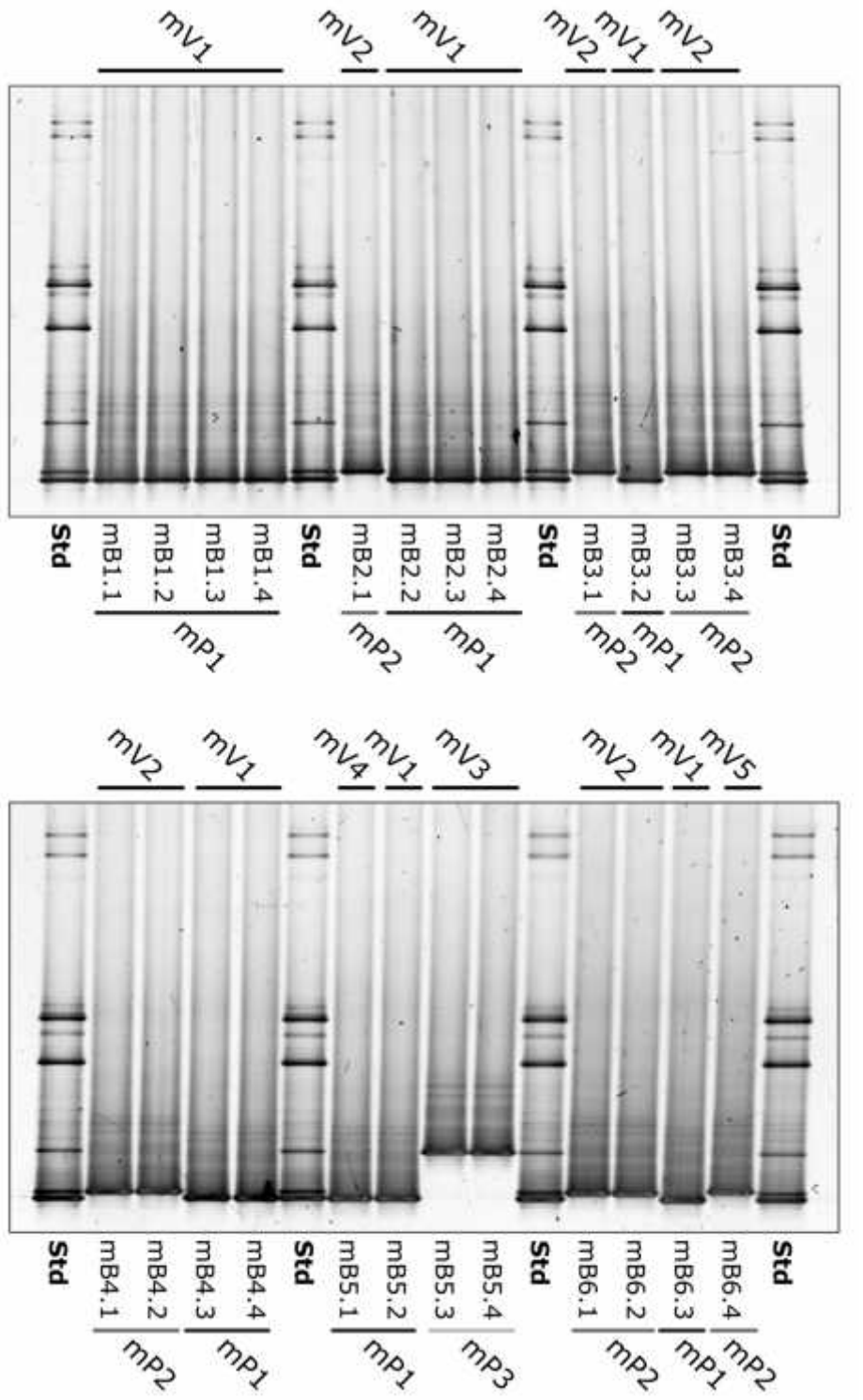



**Figure 2.5.** The DGGE profiles for the *me53* clones originally derived from the six bands (mB1-mB6) in the P4 *me53* DGGE profile. Although five clone sequence variants (mV1-mV5) were identified from the twenty four sequenced *me53* clone inserts, there were only three distinct DGGE profile types (mP1-mP3). Samples were loaded on the gels in groups of four, with profiles for replicate clones derived from the same band in the original P4 *me53* DGGE profile occurring in the same group (mB1-mB6). For example, all *me53* sequence variants isolated from the mB1 band in the P4 DGGE profile occur in the first group of four profiles (mB1.1-mB1.4). Groups are interspaced by a standard profile (Std).

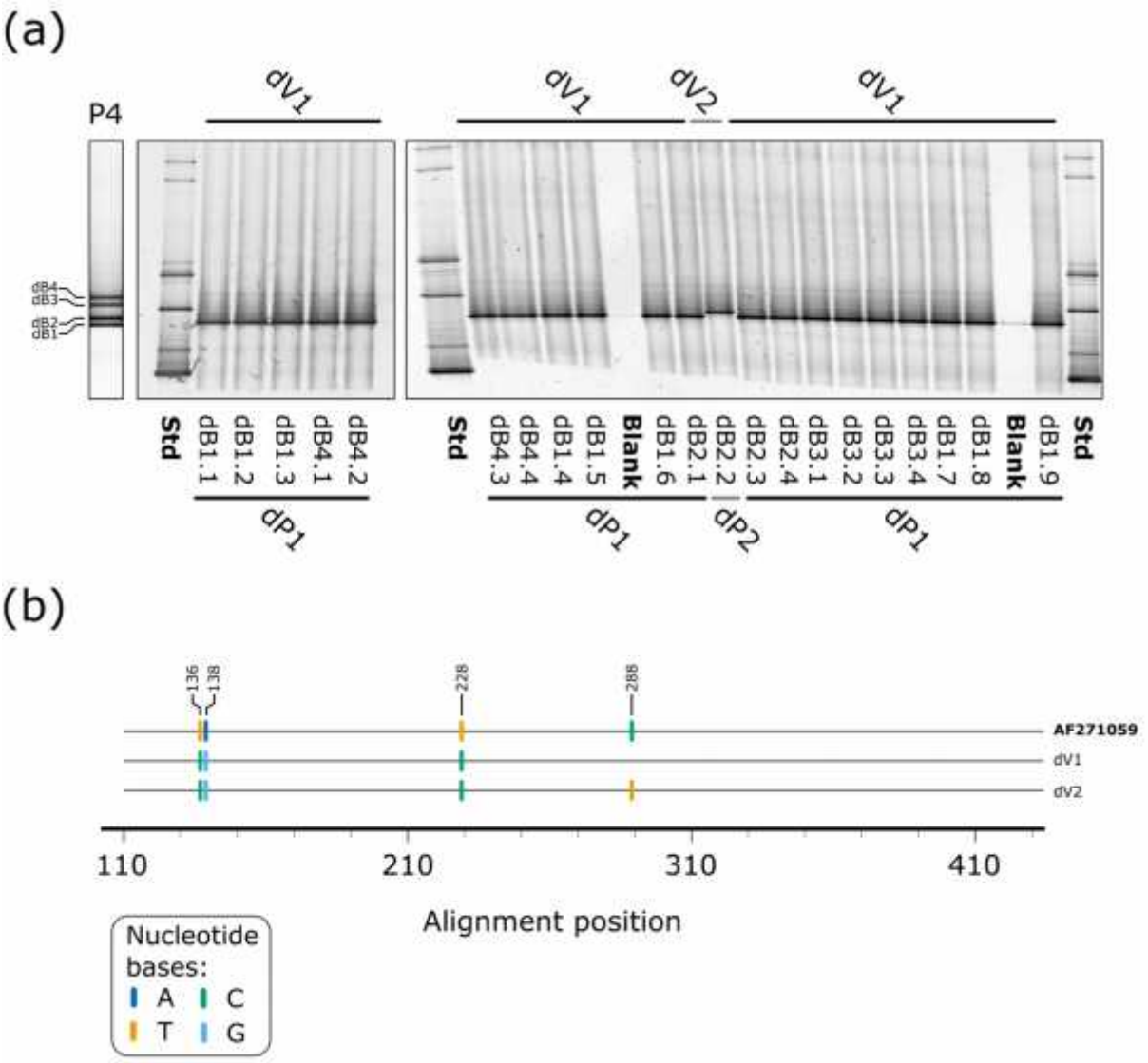



**Figure 2.6.** DGGE profiles and nucleotide sequences of the *DNA polymerase* clone variants. **(a)** The DGGE profiles for the *DNA polymerase* clones originally derived from the four bands (dB1-dB4) in the P4 *DNA polymerase* DGGE profile. Two clone sequence variants (dV1 & dV2) were identified from the twenty one sequenced *DNA polymerase* clone inserts, and each variant was associated with one of two distinct DGGE profile types (dP1 or dP2). Profile labels denote from which band in the original P4 *DNA polymerase* DGGE profile each clone was originally derived and the clone replicate number for that band. **(b)** A highlighter plot showing the nucleotide variation between the two *DNA polymerase* clone sequence variants (dV1 & dV2) and a *DNA polymerase* reference sequence (GenBank accession number: AF271059). The positions of SNPs relative to the reference *DNA polymerase* gene sequence are shown as coloured bars, with each colour denoting one of the four DNA bases. All SNPs were present in previously published HearNPV genomes (Baillie & Bouwer, 2012a; Chen *et al.*, 2001; Noune & Hauxwell, 2015, 2016) and all were synonymous polymorphisms.

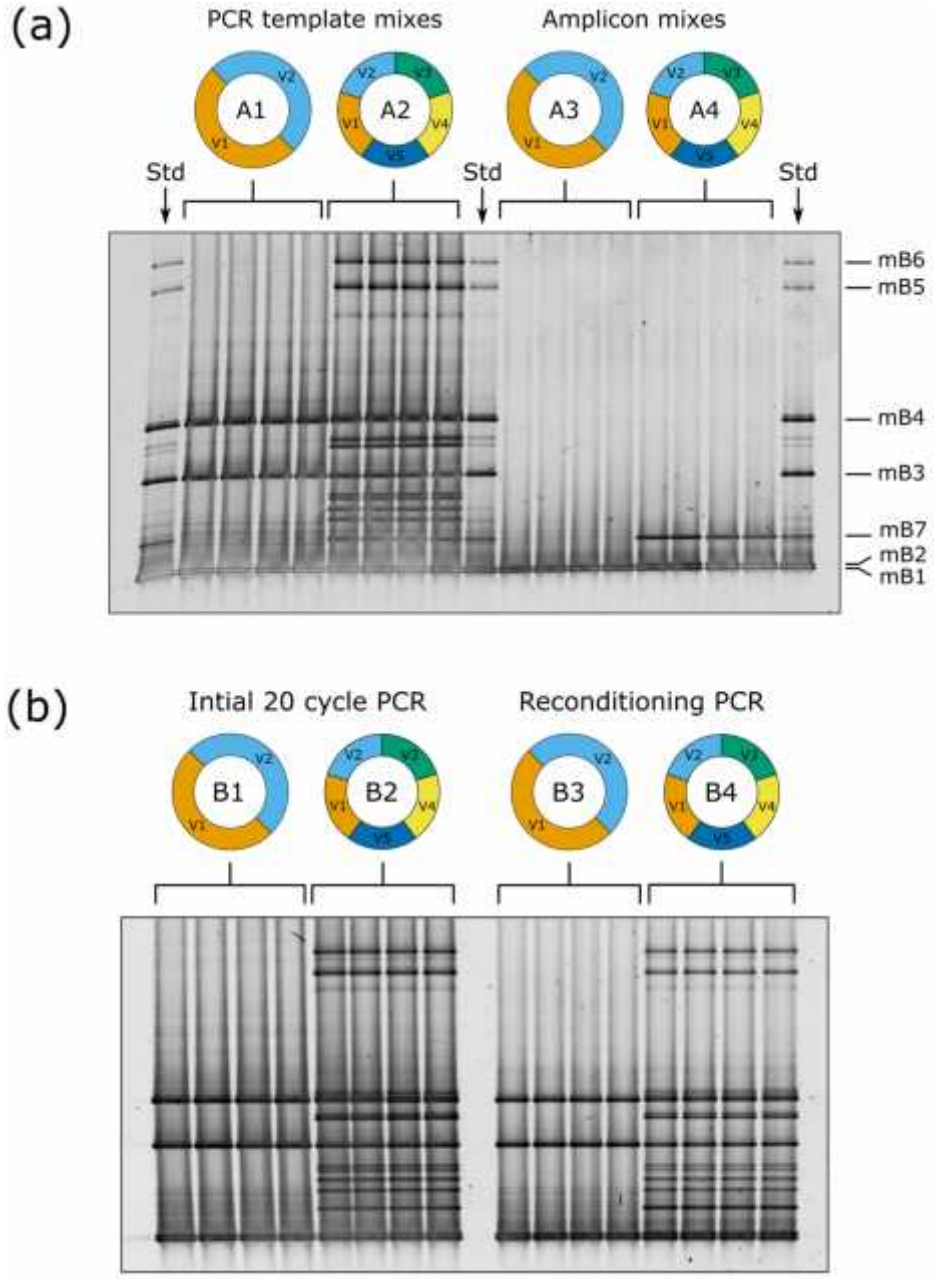



**Figure 2.7.** Artefactual bands and reconditioning reactions for *me53* DGGE profiles. **(a)** DGGE profiles of replicate reamplification reactions, where an equimolar mixture of two (A1) or five (A2) *me53* clone variants (mV1 & mV2 and mV1-mV5, respectively) were used as template, contained multiple artefactual bands. Replicate DGGE profiles for post-amplification equimolar amplicon mixtures of two (A3) or five (A4) *me53* clone variants did not contain these artefactual bands and only contained the three true bands. The positions of the three true bands (mB1, mB2 and mB7) and four artefactual bands (mB3-mB6) identified in the P4 *me53* profile are labelled on the right of the figure. **(b)** DGGE profiles for multi-template reamplification reactions (B1 & B2) and "reconditioning" reactions (B3 & B4), where PCR products from each of the initial reamplification reactions was used as template in a further four-cycle PCR. An equimolar mixture of two (B1 & B3) or five (B2 & B4) *me53* clones were used as template for the initial twenty-cycle PCRs. The "reconditioning" PCRs were not successful at reducing the formation of artefactual bands in the *me53* DGGE profiles (B3 & B4).

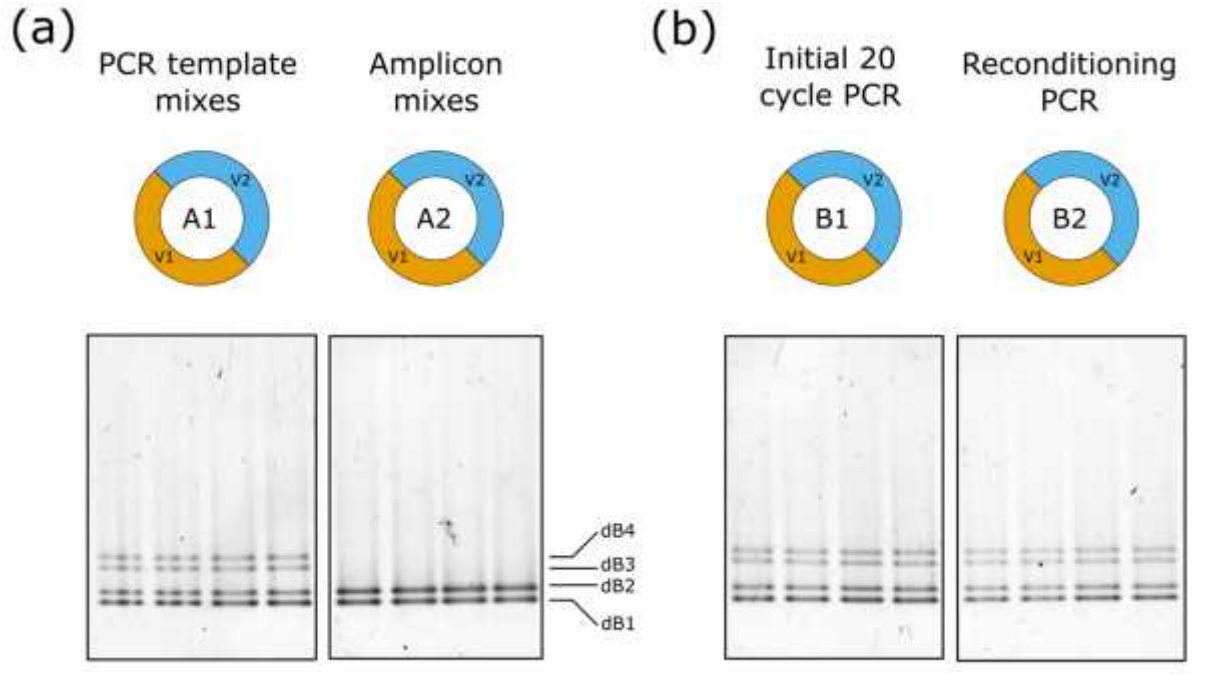



**Figure 2.8.** Artefactual bands and reconditioning reactions for *DNA polymerase* DGGE profiles. **(a)** DGGE profiles of replicate reamplification reactions, where an equimolar mixture of the two *DNA polymerase* clone variants (dV1 & dV2) were used as template (A1) contained two artefactual bands. DGGE profiles for post-amplification equimolar amplicon mixtures of the two *DNA polymerase* clone variants did not contain these artefactual bands (A2). The positions of the two true bands (dB1 and dB2) and two artefactual bands (dB3 and dB4) identified in the P4 *DNA polymerase* profile are labelled on the right of the figure. **(b)** DGGE profiles for multi-template reamplification reaction replicates (B1) and "reconditioning" reactions (B2), where PCR products from each of the initial reamplification reactions were used as template in a further four-cycle PCR. An equimolar mixture of the two *DNA polymerase* clone variants was used as template for the initial twenty-cycle PCRs (B1). The "reconditioning" PCR was not successful at reducing the formation of the two artefactual bands in the *DNA polymerase* DGGE profiles (B2).

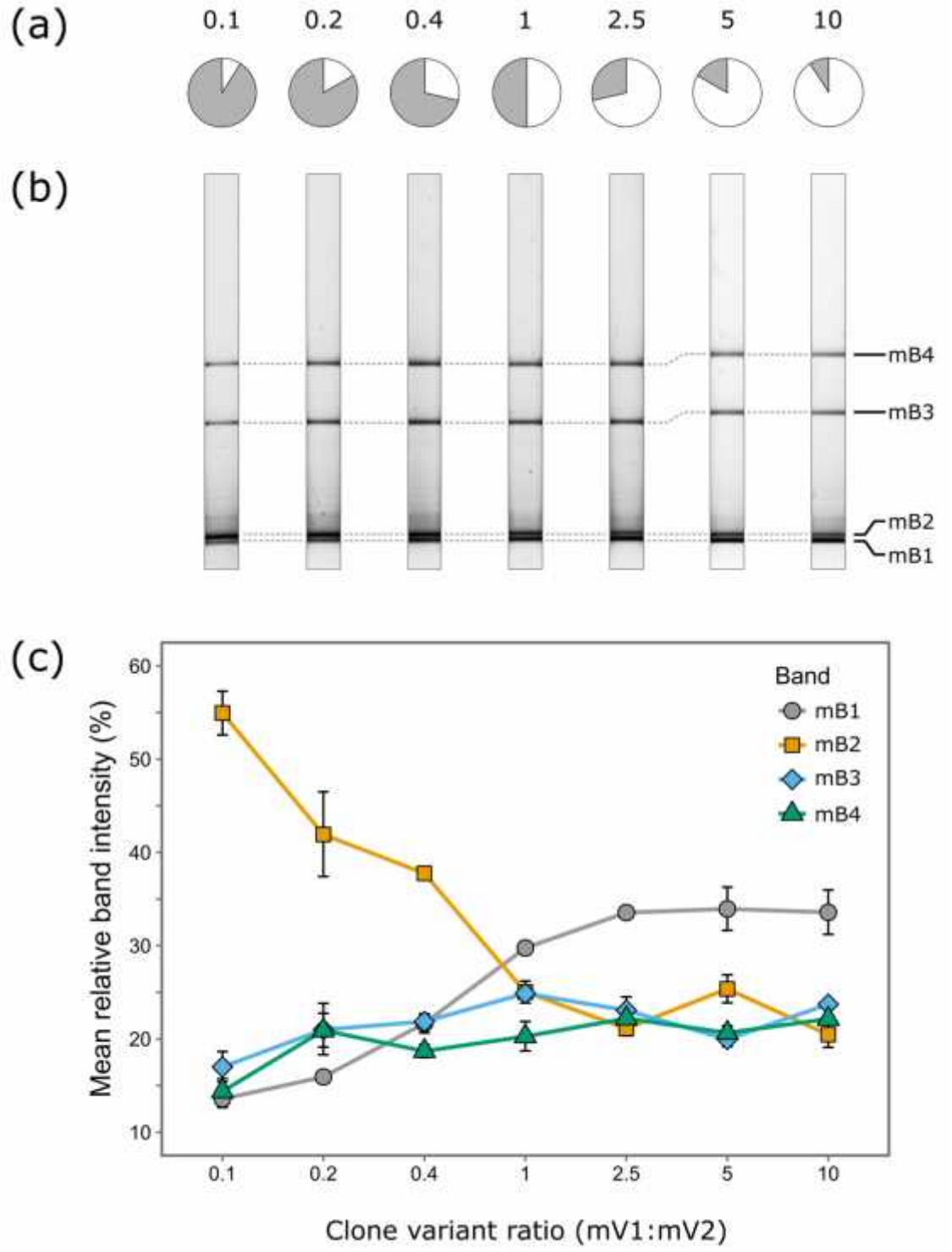



**Figure 2.9.** Differences in the relative intensities of bands mB1-mB4 between *me53* DGGE profiles for multi-template reamplification reactions, where different molar ratios of two *me53* clone variants were used as template. The two *me53* clones contained a mV1 and mV2 *me53* sequence variant insert, respectively. **(a)** The seven molar ratios of mV1:mV2 clones used as template in the reamplification reactions. **(b)** A representative DGGE profile for each group of three replicate reamplification reactions. **(c)** The relative band intensities from three replicate DGGE profiles were used to calculate the mean relative intensity and SE for each band (mB1-mB4) and these values were plotted for each clone ratio.

# Chapter 3

# The effect of serial *in vivo* passages on the genetic diversity of two core *Helicoverpa armigera* nucleopolyhedrovirus genes

## 3.1. Abstract

Considerable genetic diversity is known to exist within nucleopolyhedrovirus (NPV) populations isolated from individual host insects. A number of studies have found that the serial passaging of a NPV population in a host species results in an increase in the pathogenicity of the virus for that host species. It has been suggested that the change in pathogenicity is due to the progressive positive selection of NPV genotypes which are best adapted to exploit that particular host species. However only a limited number of studies have actually looked at how serial *in vivo* passaging affects the genetic diversity of a NPV population. In this study a denaturing gradient gel electrophoresis (DGGE) assay and deep amplicon sequencing were used to assess the genetic diversity in *Helicoverpa armigera* nucleopolyhedrovirus (HearNPV) populations of two genes essential for virus replication, *DNA polymerase* and *me53*, over three serial *in vivo* passages. It was found that there were multiple outcomes for the genetic diversity of the *DNA polymerase* and *me53* genes following an *in vivo* passage. These outcomes included the relative maintenance of the variant composition of the inoculum population, the dramatic loss of most inoculum variants from a post-passage population and the generation of novel genetic variants. These different outcomes led to the divergence of the *DNA polymerase* and *me53* variant composition of post-passage populations and a stable, common variant composition was not reached across the different passage lineages following the three *in vivo* passages.

## 3.2. Introduction

The nucleopolyhedroviruses (NPVs) are members of the *Baculoviridae* family; a group of dsDNA, arthropod-specific viruses (van Oers & Vlak, 2007; Okano *et al.*, 2006; Rohrmann, 2013; Theilmann & Blissard, 2008). It is generally accepted that dsDNA virus populations are considerably less genotypically diverse than RNA virus populations (Drake *et al.*, 1998; Sanjuán *et al.*, 2010). However several studies have shown that considerable genotypic heterogeneity exists within NPV populations isolated from individual host insects (Baillie & Bouwer, 2012b; Cory *et al.*, 2005; Hodgson *et al.*, 2001). Most RNA virus populations exist within infected hosts as "swarms" of diverse variants, also known as quasispecies (reviewed in Domingo *et al.*, 2012; Lauring & Andino, 2010). This is largely a consequence of their rapid, error-prone genome replication, which results in the continual production of novel genetic variants (Duffy *et al.*, 2008; Holmes & Moya, 2002; Smith *et al.*, 1997). In contrast the proofreading action of most host and virus DNA polymerases, together with host DNA repair mechanisms, greatly reduce the rate at which dsDNA viruses accumulate mutations, relative to RNA viruses (Duffy *et al.*, 2008; Firth *et al.*, 2010; Morse & Schluederberg, 1990).

This makes the considerable amount of genetic diversity detected in NPV populations unusual for dsDNA viruses, especially as NPV DNA polymerases demonstrate proofreading 3′ − 5′ exonuclease activity (Hang & Guarino, 1999; Huang & Levin, 2001; Mikhailov *et al.*, 1986). The genetic diversity of NPV populations is biologically relevant as different genotypes isolated from a single host can show significant differences in their pathogenicity, virulence and occlusion body (OB) yield (Cory *et al.*, 2005; Hodgson *et al.*, 2001; Rowley *et al.*, 2010; Simón *et al.*, 2005). It has also been shown that mixtures of NPV genotypic variants can be more pathogenic than the individual component genotypes alone (Hodgson *et al.*, 2004; Kikhno *et al.*, 2002; López Ferber *et al.*, 2003; Simón *et al.*, 2005, 2008). Ultimately the genetic variation may benefit NPV populations by allowing them to survive and adapt to changes in their environment (Harrison, 2009).

The *in vivo* passage of a NPV population through an insect host in theory provides multiple opportunities, during the transmission and replication stages of the infection, for the genotypic composition of the population to be altered. New genetic variants may be

continually generated as a result of DNA polymerase errors, recombination between virus genotypes during the replication of NPV genomes, or host-to-virus horizontal gene transfer (Croizier & Ribeiro, 1992; Fraser *et al.*, 1985; Gilbert *et al.*, 2014, 2016; Kamita *et al.*, 2003; van Oers & Vlak, 2007). Both selective and selectively neutral evolutionary processes could also affect the relative abundance of genetic variants within a population (Duffy *et al.*, 2008). Host and environmental factors may positively select for variants in a population, which are best adapted to initiate infection, replicate and be successfully transmitted to new hosts under the prevailing conditions (Hitchman *et al.*, 2007; Hodgson *et al.*, 2004). This may ultimately result in the complete loss of less "fit" genotypes from a NPV population over successive *in vivo* passages (Hitchman *et al.*, 2007).

In contrast selectively neutral events would result in random, non-directional changes in the relative abundance of genetic variants in a NPV population, regardless of their relative "fitness" (Manrubia & Lazaro, 2006). This is also known as genetic drift; when random sampling events determine the genotypic composition of a population (Manrubia & Lazaro, 2006). Genetic drift could be expected to occur during the horizontal transmission of NPVs between hosts, if transmission is a selectively neutral event, and infection is initiated by a single or limited pool of genotypes from the original population (Edwards *et al.*, 2006; Zwart *et al.*, 2009). The genetic diversity of other virus species is greatly reduced by population bottlenecks occurring during horizontal transmission events (Edwards *et al.*, 2006; Zwart *et al.*, 2011). However, there is limited experimental evidence showing how these processes acting over multiple *in vivo* passages would affect the nucleotide sequence variation of genes essential for virus replication in NPV populations.

In this study *Helicoverpa armigera* nucleopolyhedrovirus (HearNPV), a member of the *Alphabaculovirus* genus, was used as a model system to further elucidate the effect of serial *in vivo* passaging on the genetic diversity in NPV populations of two genes essential for virus replication. HearNPV is a SNPV which infects the African cotton bollworm, *Helicoverpa armigera* (Lepidoptera: Noctuidae), as well as several related species (Chen *et al.*, 2002). HearNPV was chosen as it is a well characterised NPV where the whole genome of a number of different HearNPV isolates have been sequenced (Baillie & Bouwer, 2012a; Chen *et al.*, 2001; Noune & Hauxwell, 2015, 2016) and for which a rapid and sensitive denaturing gradient gel electrophoresis (DGGE) assay has been developed to assess the amount of

genetic diversity in several core HearNPV genes (Baillie & Bouwer, 2011). DGGE is a genetic fingerprinting technique, which is able to discriminate between populations which differ in their underlying variant composition and, as shown in Chapter 2, can be used to assess the relative genetic diversity of HearNPV populations (Baillie & Bouwer, 2011; Muyzer *et al.*, 1993; Neilson *et al.*, 2013; Sheffield *et al.*, 1989).

The DGGE assay was used to monitor genetic variation within a core baculovirus gene, *DNA polymerase*, and a core lepidopteran-specific baculovirus gene, *me53*, in HearNPV populations over three serial *in vivo* passages. As the *DNA polymerase* and *me53* gene products have been shown to be essential for HearNPV replication (de Jong *et al.*, 2011; Vanarsdall *et al.*, 2005; Xi *et al.*, 2007), any non-synonymous sequence variations in these genes would potentially be subject to significant differential selective pressures. Populations where DGGE profiles indicated a prominent shift in the underlying variant composition of HearNPV populations over an *in vivo* passage were selected for further analysis by deep amplicon sequencing. Deep amplicon sequencing was used to assess the number and identity of nucleotide sequence variants in these populations. This allowed us to evaluate whether the *DNA polymerase* and *me53* variant composition of HearNPV populations across a number of replicate passage lineages moved towards a stable, consensus variant composition. In addition we were able assess the degree to which the generation of novel sequence variants contribute to changes in HearNPV genetic diversity over the course of an *in vivo* passage.

## 3.3. Materials and methods

### 3.3.1. Insect rearing

*H. armigera* eggs were originally obtained from the Agricultural Research Council (Pretoria, South Africa) and used to set up a laboratory culture. *H. armigera* culture maintenance was conducted as described in Chapter 2, section 2.3.1 (p. 32).

### 3.3.2. Virus propagation and purification

HearNPV was propagated *in vivo* in third-instar (± 5 d old) *H. armigera* larvae as described in Chapter 2, section 2.3.2 (p. 32). The final gradient-purified, pooled HearNPV suspension was used as the initial inoculum for all passage lineages.

### 3.3.3. Dose-mortality assay

An assay was carried out for the gradient-purified HearNPV suspension, in order to ascertain the dose which caused 90% mortality (i.e. the $LD_{90}$) in third instar *H. armigera* larvae. The assay was conducted under constant temperature and relative humidity conditions of 25 ˚C (± 1) and 60% (± 5), respectively, with a 12:12 h light:dark photocycle. Five day old, third instar larvae were infected using the diet cube contamination method (Evans & Shapiro, 1997). Larvae were presented with a HearNPV contaminated artificial diet cube; those that consumed the entire cube within 24 h were transferred back to vials containing artificial diet, while those that did not were excluded from the assay. Larvae presented with diet cubes treated with sterile distilled water acted as the uninoculated negative controls for the assay. Three replicates of 25 larvae ($n$ = 75) were used for each of the five HearNPV doses and the uninoculated control. The highest dose was $1.448 \times 10^4$ OB/larva, and a three-fold dilution series of this OB suspension was used to prepare the four lower doses. Mortality was recorded 7 d post infection and the $LD_{90}$ ($4.340 \times 10^3$ OBs/larva) was estimated by probit analysis (Bliss, 1934).

### 3.3.4. Serial *in vivo* passaging

For the primary *in vivo* passage, larvae were inoculated with a $LD_{90} \times 5$ of the gradient purified, pooled HearNPV isolate, using the diet contamination method described above. A high lethal dose was used to increase the probability that low frequency genetic variants in

the inoculum population would infect the larvae. This was intended to prevent transmission bottlenecks from playing a disproportionate role in determining the post-passage variant composition of NPV populations. Five larvae showing overt signs of NPV infection were randomly harvested just prior to the liquification of the cadavers. Each larva cadaver was homogenised in 200 µl of sterile distilled water in a 1.5 ml microcentrifuge tube, using a micro-pestle. OBs were then purified from each cadaver homogenate using a differential centrifugation protocol. The volume of each homogenate was made up to 1 ml with sterile distilled water, thoroughly vortexed and subjected to low speed centrifugation (3000 *g* for 15 s) to remove insect tissue debris. OBs were pelleted from the supernatant and washed twice in sterile distilled water during three subsequent centrifugation steps (12 000 *g* for 30 s). The final pellet was re-suspended in sterile distilled water and the concentration of OBs determined by haemocytometry. Viral DNA was extracted from half the volume of each OB suspension for DGGE analysis and deep amplicon sequencing (DNA extraction procedure described in Chapter 2, section 2.3.3, p. 33), while the remaining suspension was used as inoculum for the subsequent passage.

The process described above for the first *in vivo* passage was repeated for the second and third passages. For each passage a HearNPV isolate, derived from a single infected larva from the previous passage, was used as inoculum ($LD_{90} \times 5$) to infect a further 5 larvae. This produced a primary passage (*n = 5*), secondary passage (*n = 25*) and tertiary passage (*n = 125*) group of HearNPV isolates. The *in vivo* passaging process and overall experimental design is summarised in Figure 3.1. Serial passaging was limited to three passages as a fourth passage would have resulted in a further 625 samples which would have been impractical to analyse using PCR-DGGE. It was also thought that the effect of serial passaging on HearNPV genetic diversity would be apparent after three passages.

### 3.3.5. DNA extraction

Viral DNA was extracted separately from each of the HearNPV passage isolates as well as from the initial pooled inoculum sample. DNA extraction was performed as described in Chapter 2, section 2.3.3 (p. 33). Purified DNA was stored at -20 ˚C prior to amplification of the *DNA polymerase* and *me53* gene regions for the DGGE and deep amplicon sequencing analyses.

### 3.3.6. PCR-DGGE

A PCR-DGGE assay developed by Baillie & Bouwer (2011) was used to assess the variant composition and relative level of genetic diversity in regions of the HearNPV *DNA polymerase* and *me53* genes for all passage isolates and the initial pooled inoculum. PCR amplification and DGGE analysis of the *DNA polymerase* and *me53* gene regions was conducted as described in Chapter 2, section 2.3.4 (p. 33) and section 2.3.5 (p. 34), respectively. DGGE profiles were analysed using the GelCompar II V6.0 software (Applied Maths) as described in Chapter 2, section 2.3.6 (p. 35). The software was used to normalise gel profiles relative to a standard profile, construct cosine similarity matrices of *DNA polymerase* and *me53* profiles across all three passages and construct dendrograms showing the mean similarity of DGGE profiles. PERMANOVA analyses (Anderson, 2001) were used to assess whether there were any statistically significant differences in the band composition of *DNA polymerase* and *me53* DGGE profiles between the five passage groups after the second and third passage. PERMANOVA is robust to heterogeneity of multivariate dispersions for balanced designs (Anderson & Walsh, 2013), and a complimentary test, PERMDIST (Anderson, 2006), was used to test whether there were statistically significant differences in the amount of variation between DGGE profiles within each of the five passage groups. Both the PERMANOVA and PERMDIST analyses used the cosine similarity matrices of the DGGE profiles, after first converting them into dissimilarity matrices, and were conducted with the vegan package (Oksanen et al., 2016) in R version 3.3.1 (R Core Team, 2016).

### 3.3.7. Deep amplicon sequencing

Deep amplicon sequencing on the MiSeq (Illumina) HTS platform was used to assess the number and identity of *me53* sequence variants in the initial pooled inoculum and selected post-passage HearNPV populations. The sample preparation and sequencing steps were conducted as described in Chapter 2, section 2.3.11 (p. 38). The raw deep sequencing data was processed and analysed as described in Chapter 2, section 2.3.12. (p. 39). Samples which had <10 000 merged paired-end sequencing reads after the read processing and filtering steps were discarded. All further data analysis steps were conducted in R version 3.3.1 (R Core Team, 2016). Levenshtein distances between sequence variants were calculated using the stringDist function of the Biostrings package (Pages *et al.*, 2016) from the Bioconductor

project. The relative abundances and Levenshtein distances for the sequence variants were plotted using the ggplot2 package (Wickham, 2009).

## 3.4. Results

### 3.4.1. DGGE profile analysis

UPGMA cluster analysis was used to prepare dendrograms from the cosine coefficient pairwise similarity matrices of the *DNA polymerase* and *me53* DGGE profiles. The cosine coefficient was chosen as it takes into account both the relative position and intensity of bands when assessing the degree of similarity between profiles. It is clear from both the *DNA polymerase* and *me53* cluster analyses that passaging group, rather than passage number, largely determined which DGGE profiles would cluster together with a high degree of similarity (Figure 3.2 and Figure 3.3). Profiles for HearNPV populations isolated after the first or second passage appeared to be randomly distributed amongst profiles for populations isolated after the third passage, within clades largely consisting of profiles belonging to the same passage group (Figure 3.2, clades PG1-5 and Figure 3.3, clades PG1-4). Profiles within these passage group clades were very similar (> 93% mean similarity). These results indicate that the first *in vivo* passage of a pooled HearNPV population had a considerable effect on the variant composition of descendant populations over the subsequent passages.

This was most apparent with the *DNA polymerase* profiles, where 68% (PG1), 97% (PG2), 94% (PG3), 90% (PG4) and 61% (PG5) of profiles from the same passaging group clustered together in a clade (Figure 3.2). With the *me53* profiles there was a greater degree of similarity between profiles across different passaging groups. This was clearly shown by the clustering of 61% of PG1 and 97% of PG3 profiles together in a single clade (Figure 3.3, PG 1 & 3 clade, 97% mean similarity). However even within this clade, profiles were present in several distinct PG1 and PG3 sub-clades, and were not randomly distributed amongst each other. Relative to the PG1 and PG3 profiles, there was a lower degree of similarity between *me53* profiles in the other passaging groups, with 65% of PG2 and 74% of PG4 profiles clustering in the PG2 and PG4 clades, respectively. PG5 *me53* profiles showed the lowest degree of similarity, for profiles from the same passage group. These profiles did not form a clear PG5 clade, and were scattered amongst PG1, PG3, PG2 and PG4 profiles. A subset of six PG5 *me53* profiles belonging to the same passage lineage also formed an outlying clade (Figure 3.3, PG5 clade). The profiles in this clade had the lowest similarity (70% mean similarity) to *me53* profiles across all of the passage groups.

Both major and minor shifts in the band composition of *DNA polymerase* and *me53* profiles were observed following an *in vivo* passage. These shifts included both the increase and decrease in the relative intensity of individual bands over a single passage (Figure 3.4). In a few passage lineages these shifts were more pronounced, with the relative intensity of several bands within a profile decreasing dramatically over a single *in vivo* passage (Figure. 3.4a-c). These "drop-off" events occurred to a limited extent in both *DNA polymerase* and *me53* profiles across all passage groups after the first, second and third passages. "Drop-off" events often resulted in a single dominant band remaining in the DGGE profile (Figure 3.4a-c) and these profiles closely resemble the *DNA polymerase* and *me53* profiles in Chapter 2 for a single reamplified clone variant. This suggests that "drop-off" events may represent a significant reduction in the number of *DNA polymerase* and *me53* sequence variants in a HearNPV population that occurs over a single *in vivo* passage. The most significant of these events occurred in the PG3 lineage, where the apparent dramatic loss of *me53* variants occurred over the first passage, and the number of bands in PG3 *me53* profiles did not increase over the subsequent passages (Figure 3.4c). This suggests that the number of *me53* sequence variants remained low in HearNPV populations in PG3 lineages and that these populations potentially only contained a single variant.

Although in most instances bands in the *DNA polymerase* and *me53* inoculum profiles were maintained over successive passages, significant shifts in their relative intensities were observed. These bands generally occurred higher up in the profiles, and in some cases occurred at the same profile positions as bands that were shown to have a strong likelihood of being heteroduplex bands in Chapter 2. This was most apparent in the PG1 and PG4 *DNA polymerase* first passage profiles and a subset of PG5 *me53* third passage profiles (Figure 3.4a-b and Figure 3.5a, PG5.3.A to PG5.3.D). The significant shifts in the relative intensities of bands present in the *me53* inoculum profile over the serial passages appear to be responsible for the low similarity of the PG5 *me53* outlier profiles to all the other *me53* profiles across the five passage groups (Figure 3.3 and Figure 3.5a, PG5.3.A to PG5.3.D). These shifts in the relative intensities of both true and suspected heteroduplex bands suggest that although variants were not lost from populations, as is suspected with the observed "drop-off" events, there were considerable changes in the underlying *DNA polymerase* and *me53* variant composition of HearNPV populations over the three *in vivo* passages.

The cluster analyses of DGGE profiles indicated that the differences in the underlying *DNA polymerase* and *me53* variant composition of HearNPV populations between the five passage groups increased over the three *in vivo* passages. PERMANOVA and PERMDIST tests were performed in order to assess whether differences in the *DNA polymerase* and *me53* profiles across the five passage groups were statistically significant. The results of the cluster analyses were largely supported by the PERMANOVA and PERMDIST analyses. The PERMANOVA analyses showed that following the second and third passages there were statistically significant differences ($\alpha = 0.05$) in the band composition of *DNA polymerase* and *me53* DGGE profiles between the different passage groups (Table 3.1 and Table 3.2). The PERMDIST analyses showed that following the second passage the amount of variation between both *DNA polymerase* and *me53* profiles within the five passage groups was not statistically significantly different ($\alpha = 0.05$) (Table 3.1 and Table 3.2). However, following the third passage the variation between DGGE profiles in the five passage groups were significantly different ($\alpha = 0.05$) (Table 3.1 and Table 3.2). Overall, the cluster analyses and PERMANOVA and PERMDIST significance tests indicate that the serial *in vivo* passaging of a pooled HearNPV population using a high dose ($LD_{90} \times 5$) did not lead to a consensus population with a stable *DNA polymerase* and *me53* variant composition, which was common across all passaging groups after the third passage.

### 3.4.2. Deep *me53* amplicon sequencing

DGGE profiles indicated that the level of *me53* variant richness in HearNPV populations was generally greater than that of *DNA polymerase* variants. The shifts in *me53* variant composition of populations over successive passages also appeared to be greater than the associated shifts in *DNA polymerase* variant composition. Deep targeted sequencing of the *me53* region assayed using DGGE was conducted to assess the underlying nucleotide sequence variation and variant richness of selected HearNPV populations, and to further evaluate the effect of successive *in vivo* passages on the diversity of a gene essential for virus replication. HearNPV populations were selected for deep amplicon sequencing analysis if they were part of a passage lineage where PCR-DGGE analysis suggested significant shifts had occurred in the underlying variant composition of populations over successive *in vivo* passages. These included both gradual “drop-off” events, where the reduction in the number and intensity of bands occurred over successive passages, and dramatic “drop-offs” events where the reduction in bands occurred over a single passage. The populations belonging to a

PG5 passage lineage that made up the PG5 outlier clade of the *me53* dendrogram were also selected for deep sequencing analysis.

Using the deep amplicon sequencing approach, six *me53* sequence variants were detected in the inoculum population and a total of eleven variants were detected across the fourteen HearNPV populations selected for analysis. Of the five novel genetic variants that were not detected in the inoculum population, two were detected in PG2 populations, and the other three were detected in the PG5 outlier populations. The DGGE profiles for the PG2 passage lineage selected for deep sequencing indicated a gradual "drop-off" in *me53* variants occurred over the three passages. The deep sequencing data showed that the most significant loss of *me53* variants in this PG2 passage lineage occurred over the first passage, with only two inoculum variants, mV1 and mV2, and a low frequency novel variant, mV9, being detected in the HearNPV population following this passage (Figure 3.5b, PG2.1). These three variants were maintained in the HearNPV population following the second passage; however, only variant mV2 was detected following the third passage (Figure 3.5b, PG2.2 and PG2.3). Following the third passage variant mV2 appeared to be the dominant variant in the population, together with another novel variant, mV10, which was previously undetected in this lineage (Figure 3.5, PG2.3). The mV10 and mV2 sequences were identical except for a novel non-synonymous SNP (nucleotide position 98), which has not been reported in any previously sequenced HearNPV genomes (Figure 3.6).

In contrast to the loss and replacement of *me53* variants observed over the PG2 lineage, the deep sequencing data confirmed the dramatic loss of inoculum variants over the first passage in the PG3 lineage, as suggested by the reduction in bands in the PG3 DGGE profiles (Figure 3.5b, PG3.1). After the first passage, variant mV4 appeared to be the dominant variant in the population, accounting for 65% of the processed sequencing reads, while variant mV1 accounted for the remaining 35% (Figure 3.5b, PG3.1). The deep sequencing data also appears to confirm what was inferred from the DGGE profiles for PG3; that in this passage group the first passage was largely responsible for the *me53* variant composition of populations isolated after subsequent passages. After the reduction in variant richness observed over the first passage, the *me53* variant composition of populations in the PG3 lineage remained essentially unchanged over the second and third passages (Figure 3.5b, PG3.1, PG3.2, PG3.3.A and PG3.3.B). It is also important to note that the PG3 DGGE

profiles are clear examples of how interpretations of DGGE profiles can lead to an underestimation of the genetic diversity of HearNPV populations. In the case of PG3 profiles the co-migration of variants mV1 and mV4, as shown in Chapter 2 (Figure 2.5), resulted in profiles with a single dominant band instead of two.

The PG5 outlier populations differed from the sequenced PG2 and PG3 lineages, as *me53* inoculum variants were largely not lost over successive passages. In addition to the maintenance of inoculum variants, three novel variants were detected in the PG5 outlier populations following *in vivo* passages. Following the first passage none of the inoculum variants were lost and two novel variants, mV7 and mV8, were detected (Figure 3.5b, PG5.1). However, the inoculum variant, mV6, and both mV7 and mV8 were not detected following the second passage (Figure 3.5b, PG5.2). These three variants subsequently reappeared in three of the four populations sequenced following the third passage (Figure 3.5b, PG5.3.B, PG5.3.C and PG5.3.D). This suggests that these variants were not lost from the PG5 lineage over the second passage but were present in the population at a very low frequency, below the level necessary for detection by deep amplicon sequencing. The PG5.3.C population also contained an additional novel genetic variant, mV11. However, the variant composition of one of the four PG5 populations sequenced following the third passage differed significantly from the other three populations. In this population variants mV2, mV4 and mV5 appeared to be lost over the third passage and were not detected in the post-passage population (Figure 3.5b, PG5.3.A). This is further evidence that "drop-off" events could occur over any single *in vivo* passage. Generally, the maintenance of inoculum variants together with the presence of additional novel variants resulted in a high level of variant richness in the PG5 outlier populations. This most likely gave rise to additional true and heteroduplex bands in their DGGE profiles (Figure 3.5a) which resulted in these profiles having a low similarity to other profiles across the five passage groups.

Three SNPs, which have not been detected in previously sequenced HearNPV genomes, were identified across the eleven distinct *me53* sequence variants. However, only a single *de novo* non-synonymous SNP was detected that would have altered the amino acid sequence of the expressed protein (Figure 3.6). This SNP occurred at nucleotide position 98 in the reference *me53* gene sequence, and was only detected in the sequence of variant mV10. The fact that the majority of SNPs detected were synonymous, suggests that the *me53* nucleotide sequence

is conserved and has a low tolerance of nucleotide changes which would alter the amino acid sequence of the gene product. This suggested that the changes observed over *in vivo* passages in the *me53* variant composition of HearNPV populations were not the result of natural selection acting directly on the *me53* region assayed in this study, as there was a lack of selectable phenotypic differences between the variants detected. What does seem clear is that there were multiple outcomes for the variant composition of a HearNPV population following an *in vivo* passage, which could include the maintenance or loss of existing variants as well as the potential generation of novel sequence variants.

## 3.5. Discussion

A number of studies have found that the serial passaging of a NPV population in a host species results in an increase in the pathogenicity of the virus for that species (Kolodny-hirsch & Van Beek, 1997; Shapiro *et al.*, 1982; Shapiro & Ignoffo, 1970; Tompkins *et al.*, 1981, 1988). The pathogenicity of a baculovirus isolate is generally defined as the capacity of the virus to infect and cause the death of a host and is measured using dose-response bioassays (Bernal *et al.*, 2013; Hodgson *et al.*, 2004). It has been suggested that the increase in pathogenicity is a result of the positive selection of genotypic variants that are better adapted to a particular host species or population (Arrizubieta *et al.*, 2015; Bernal *et al.*, 2013; Hitchman *et al.*, 2007) and a number of previous studies, using RFLP or PCR-based analyses, have looked at what effect serial *in vivo* passages have on the genotypic composition of NPV populations (e.g. Arrizubieta *et al.*, 2015; Clavijo *et al.*, 2010; Hitchman *et al.*, 2007; Simón *et al.*, 2006; Weitzman *et al.*, 1992). Hitchman *et al.* (2007) found that different genotypes within a wild-type *Panolis flammea* nucleopolyhedrovirus (PaflNPV) population appeared to be selected during serial *in vivo* passages in four different host species. This suggests NPV populations may contain a predominant genotype or group of genotypes which are best adapted to the local host species, but which may be displaced following passage in an alternate host species. More recently Arrizubieta *et al.* (2015) showed that the relative frequencies of two HearNPV genotypes in a co-occluded, genotypic mixture (1:1 proportion) changed significantly over successive *in vivo* passages, with one genotype progressively increasing in prevalence with each passage and accounting for 84.6% of the virus population after five passages.

There is also clear evidence that mixed genotype infections can have greater pathogenicity and higher OB yields than infections caused by the constituent genotypes on their own (Arrizubieta *et al.*, 2015; Bernal *et al.*, 2013; Hodgson *et al.*, 2004). The studies by Hodgson *et al.* (2004), Bernal *et al.* (2013) and Arrizubieta *et al.* (2015) suggested that facilitation and complementation between genotypes may be responsible for the observed gains in fitness for genotypes in a mixed genotype infection. Therefore groups of genotypes, rather than individual genotypes, may be the unit of selection in NPV-host interactions. This would make it unlikely that only a single genotype will eventually exist in a NPV population after multiple passages in a host population (Hitchman *et al.*, 2007; Hodgson *et al.*, 2004). This is

supported by evidence that after successive *in vivo* passages, experimental *Spodoptera frugiperda* MNPV (SfMNPV) populations, with different starting ratios of complete and deletion genotypes, converged on a common equilibrium genotypic composition (Clavijo *et al.*, 2009; Simón *et al.*, 2006). On the basis of these previous passage studies, it may be predicted that the sequence variants in NPV populations for genes essential for NPV replication and transmission may converge on a stable and host-optimal composition following successive *in vivo* passages.

The results from the present study show that at least for the gene regions assayed there was no evidence of the HearNPV populations in the different passage lineages tending towards a stable consensus variant composition over three *in vivo* passages. Instead it appeared that there were multiple outcomes for the *DNA polymerase* and *me53* variant composition of HearNPV populations from different passage lineages where the same virus population was used to inoculate larvae for the first passage. The genetic variant composition of the HearNPV populations diverged from that of the original pooled inoculum population, and no convergence was observed between the 125 different passage lineages for either the *DNA polymerase* or *me53* gene region. This was shown by the absence of DGGE profiles from the same passage clustering together in distinct clades, as well as the PERMANOVA test showing that *DNA polymerase* and *me53* profiles in the different passaging groups were statistically significantly different following the third passage. If there had been convergence in the genetic variant composition of the different HearNPV populations over the three *in vivo* passages, we would have expected to observe a greater degree of similarity between populations isolated after the same passage number, especially amongst those isolated after the second and third passages. It was also clear from the deep *me53* amplicon sequencing results that each *in vivo* passage was an event where there were multiple outcomes for the variant richness and composition of the post-passage population. However, sequence analysis of the *me53* region also showed that ten of the eleven variants detected would encode the same amino acid sequence, and the non-synonymous variant was only detected in a single population. It is worth noting that the two *DNA polymerase* variants detected in Chapter 2 were also synonymous.

The *me53* deep sequencing results show that novel genetic variants were generated over the course of *in vivo* passages and contributed to the genetic diversity of HearNPV populations

across all five passage lineages. Of the five novel *me53* variants detected only variant mV10 differed from other variants by a unique and previously undetected *de novo* SNP. This *de novo* SNP also resulted in variant mV10 being the only non-synonymous variant out of the eleven variants detected across the fourteen sequenced populations. This suggests that there may be a low tolerance of any non-synonymous mutations which may be generated over the course of a passage, resulting in the conservation of the amino acid sequence encoded by the *me53* gene. Alternatively, the lack of both new non-synonymous as well as synonymous mutations detected in the *me53* sequence may indicate that the HearNPV *DNA polymerase* enzyme exhibits effective proofreading activity during the replication of the virus genome. The DNA polymerases of several other baculovirus species have been shown to possess proofreading activity (Hang & Guarino, 1999; Huang & Levin, 2001; Mikhailov *et al.*, 1986). The other four novel *me53* variants did not contain unique SNPs and appeared to be sequence mosaics of the six variants detected in the original HearNPV inoculum population. The sequences for the five novel *me53* variants detected in the current study suggest that novel sequence variants for core HearNPV genes may be predominantly generated over an *in vivo* passage through the recombination of existing variant sequences. The recombination-dependent replication of baculovirus genomes may contribute to the generation of these novel recombinant variants by facilitating the exchange of gene sequences between genomes when different genotypes co-infect a host cell (Martin & Weber, 1997).

As noted, there were multiple outcomes for the *DNA polymerase* and *me53* variant composition of a HearNPV population following an *in vivo* passage. This led to the divergence in variant composition of HearNPV populations observed between the common initial inoculum population prior to the first passage and the 125 populations following the third passage. Deep amplicon sequencing confirmed that the changes observed in the *me53* DGGE profiles for HearNPV populations over successive passages resulted from changes in the underlying richness and relative abundance of variants in these populations. As observed over the first passage, the multiple outcomes of an *in vivo* passage included post-passage populations where the variant composition of the inoculum population was largely maintained and others where a “drop-off’ event occurred and there was a dramatic reduction in genetic diversity. Of particular significance was the observation that once *me53* genetic variants were lost from a population over a “drop-off” event, the genetic diversity remained low in subsequent populations and did not recover over the following passages. These

findings hold particular significance for natural HearNPV populations in the field as it suggests that the genotypic composition of a HearNPV population can be significantly altered by a single *in vivo* passage in a new host insect. Previous studies have shown that there are greater differences in genotypic composition between NPV populations isolated from geographically distant regions than between populations isolated from sites in close proximity to each other (Cooper *et al.*, 2003; Shapiro *et al.*, 1991). Baillie & Bouwer (2012a) also detected extensive genetic diversity between seven geographically distant HearNPV populations. The shifts in variant composition that can occur over a single *in vivo* passage may be one mechanism that contributes to HearNPV genetic diversity between *H. armigera* populations that have separated from a precursor host population and become geographically isolated from each other.

The lack of non-synonymous nucleotide sequence variation in the detected *me53* variants suggests that natural selection would not act directly on this gene region and was not responsible for the shifts in *me53* variant composition observed over the *in vivo* passages. The divergence observed in both the *DNA polymerase* and *me53* DGGE profiles over the three passages also appears to suggest that a subset of *DNA polymerase* and *me53* genetic variants were not positively selected over successive passages. However, previous studies have identified a number of genes under positive selection, in several different NPV species (Chateigner *et al.*, 2015; Simón *et al.*, 2011; Xu *et al.*, 2013). In particular, Chateigner *et al.* (2015) identified genes under positive selection across all baculovirus functional gene classes, including genes essential for the initial infection of migut epithelial cells by ODVs and virus gene transcription. This suggests that natural selection could still have contributed to the shifts in genetic variant composition of HearNPV populations observed over the three *in vivo* passages, if it acted on gene sequences distant from the assayed *DNA polymerase* and *me53* gene sequences in the HearNPV genome. Selectively neutral processes may also have contributed to the changes observed in the *DNA polymerase* and *me53* genetic variant composition of HearNPV populations over successive passages. For example, genetic drift in a HearNPV population could occur during the transmission of genetic variants to a new host if the primary infection of midgut epithelial cells is a largely stochastic process. It is possible that both selective and selectively neutral processes shape the *DNA polymerase* and *me53* genetic variant composition of a HearNPV population over an *in vivo* passage, potentially acting at different stages of an infection.

## 3.6. Conclusions

PCR-DGGE and amplicon deep sequencing results showed that there were multiple outcomes for the genetic diversity of *DNA polymerase* and *me53*, two genes essential for HearNPV replication, following an *in vivo* passage in *H. armigera* larvae. These outcomes included: (i) the maintenance of genetic variants present in the inoculum population, (ii) the dramatic loss of most inoculum variants from a population over a single *in vivo* passage, and (iii) the appearance of novel genetic variants in post-passage populations. The novel *me53* genetic variants appeared to be largely generated through the recombination of existing variants and only a single instance of a mutation that led to a *de novo* SNP was detected. The variant that contained the *de novo* SNP was also the only non-synonymous variant, which encoded an altered amino acid sequence. This may suggest that the *me53* gene sequence is conserved and that there exists a low tolerance for any non-synonymous mutations that may arise over an *in vivo* passage, although it could confirm that the HearNPV DNA polymerase enzyme has effective proofreading activity. Further studies should assess whether similar changes in variant composition are observed for other HearNPV genes, such as those essential for ODV infection of midgut epithelial cells and virus gene transcription. Overall the results showed that the genetic variant composition of two genes essential for virus replication diverged from the variant composition of an inoculum HearNPV population over three *in vivo* passages and did not reach a stable, consensus variant composition across the 125 passage lineages.

**Table 3.1.** The results of the PERMANOVA and PERMDIST analyses of *DNA polymerase* DGGE profiles from the different passage groups, after the second and third passages, respectively. Both analyses were implemented using the cosine dissimilarity matrices for the DGGE profiles and *P* values were calculated following 9999 permutations according to the methods of Anderson (2001, 2006).

| | **PERMANOVA** | | **PERMDIST** | |
|---|---|---|---|---|
| | ***F*** | ***P*** | ***F*** | ***P*** |
| **2nd Passage** | $F_{4, 20} = 15.86$ | < 0.001 | $F_{4, 20} = 2.63$ | > 0.05 |
| **3rd Passage** | $F_{4, 120} = 50.51$ | < 0.001 | $F_{4, 120} = 6.51$ | < 0.001 |

**Table 3.2.** The results of the PERMANOVA and PERMDIST analyses of *me53* DGGE profiles from the different passage groups, after the second and third passages, respectively. Both analyses were implemented using the cosine dissimilarity matrices for the DGGE profiles and *P* values were calculated following 9999 permutations according to the methods of Anderson (2001, 2006).

| **Analysis** | **PERMANOVA** | | **PERMDIST** | |
|---|---|---|---|---|
| | ***F*** | ***P*** | ***F*** | ***P*** |
| **2nd Passage** | $F_{4, 20} = 7.98$ | < 0.001 | $F_{4, 20} = 1.14$ | > 0.05 |
| **3rd Passage** | $F_{4, 120} = 32.44$ | < 0.001 | $F_{4, 120} = 9.51$ | < 0.001 |

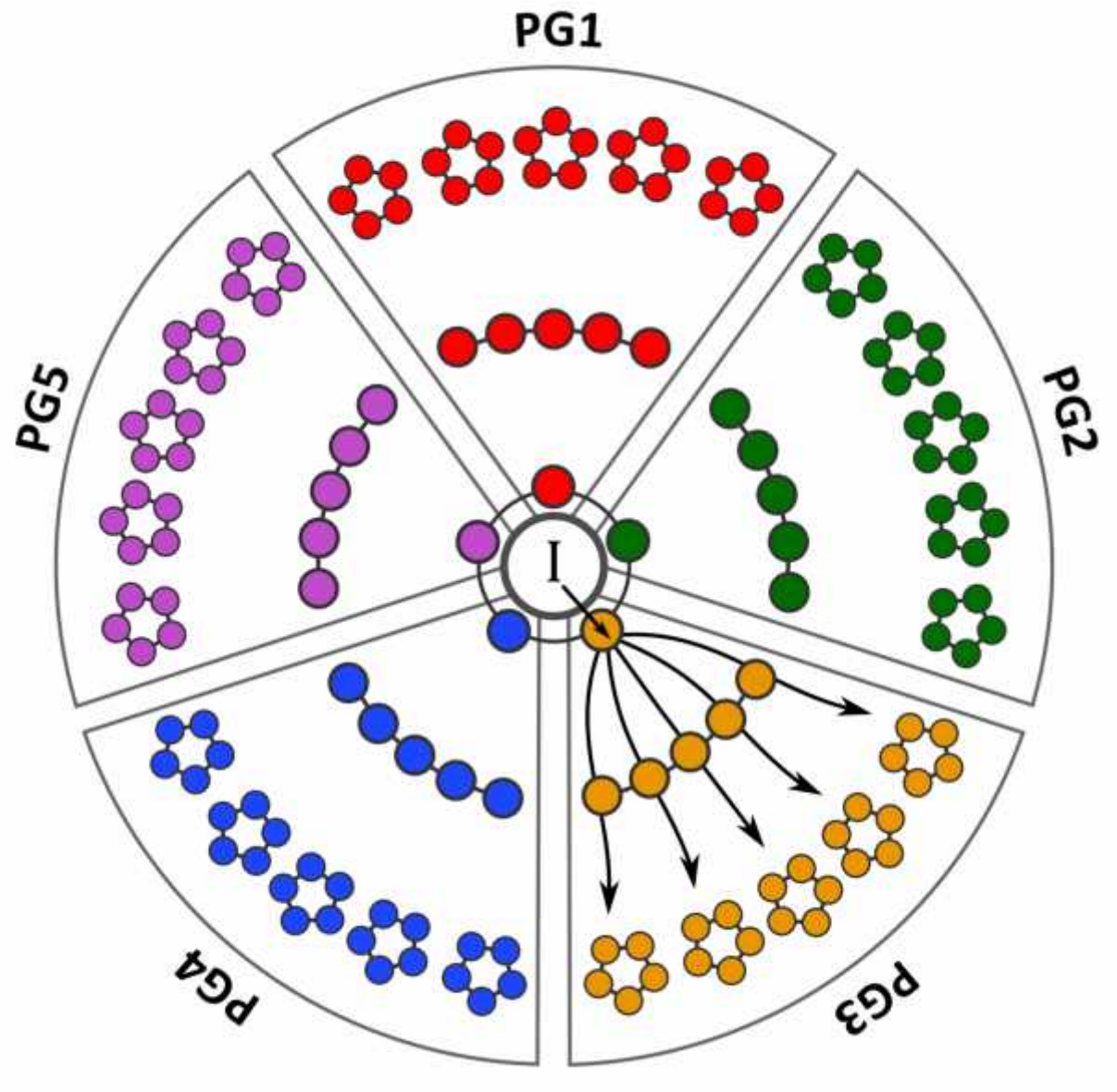
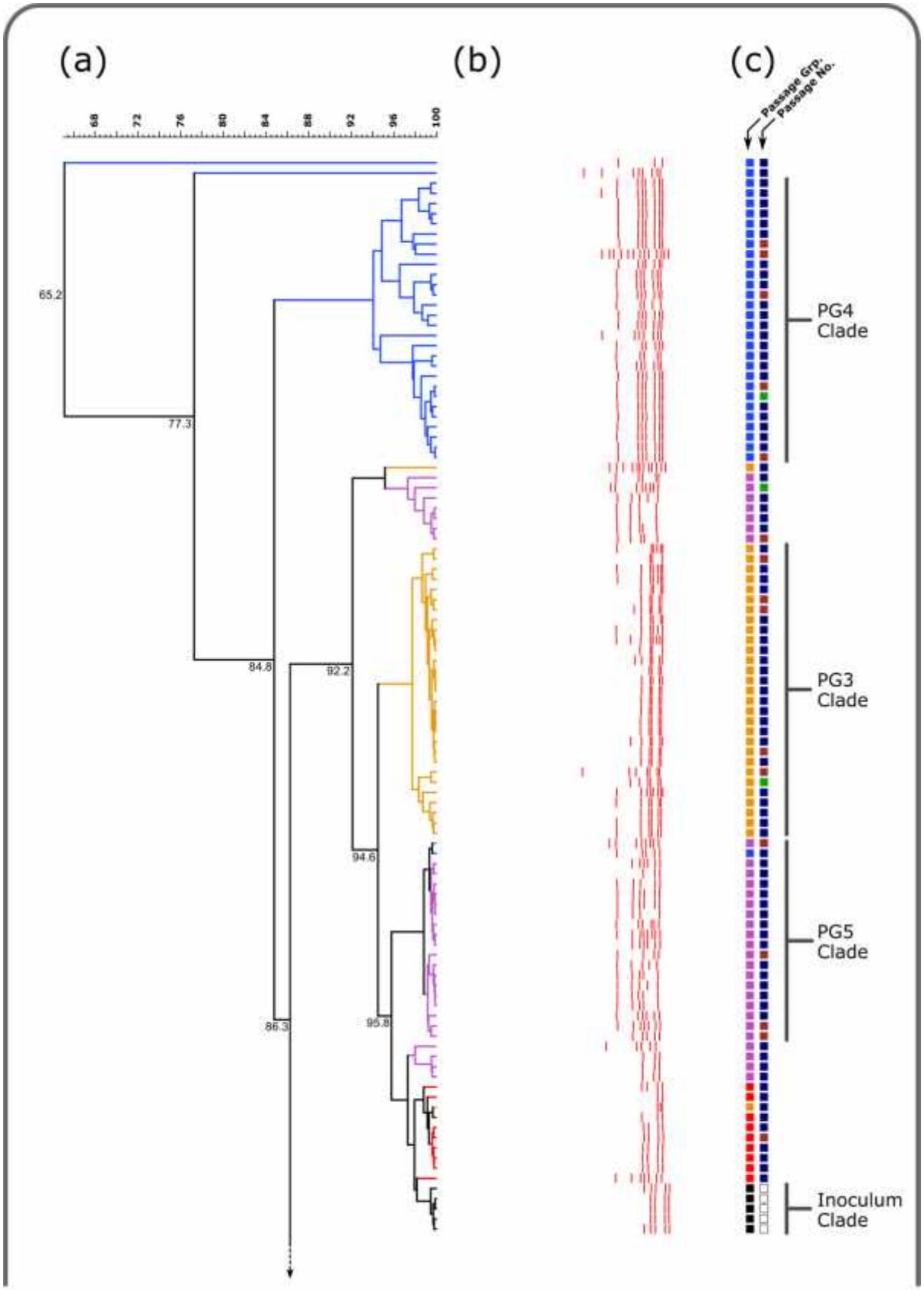



**Figure 3.1.** The overall experimental design for the serial *in vivo* passages of HearNPV in third instar *H. armigera* larvae. Each red, green, orange, blue or purple circle represents a distinct HearNPV population, isolated after a single *in vivo* passage. The five colours indicate the five different passage groups (PG1-PG5). HearNPV isolates within a passage group descended from a HearNPV population isolated from one of the five initially infected larvae. These five initially infected larvae were all inoculated with the same pooled HearNPV isolate (I). The arrows in PG3 are used as an example to show how each first and second passage HearNPV isolate was used to inoculate a further five larvae for the subsequent *in vivo* passage. This also applies to the other four passage groups. Each group of five larvae inoculated with the same HearNPV isolate was referred to as a common progenitor (CP) group.




* Figure continued on following page

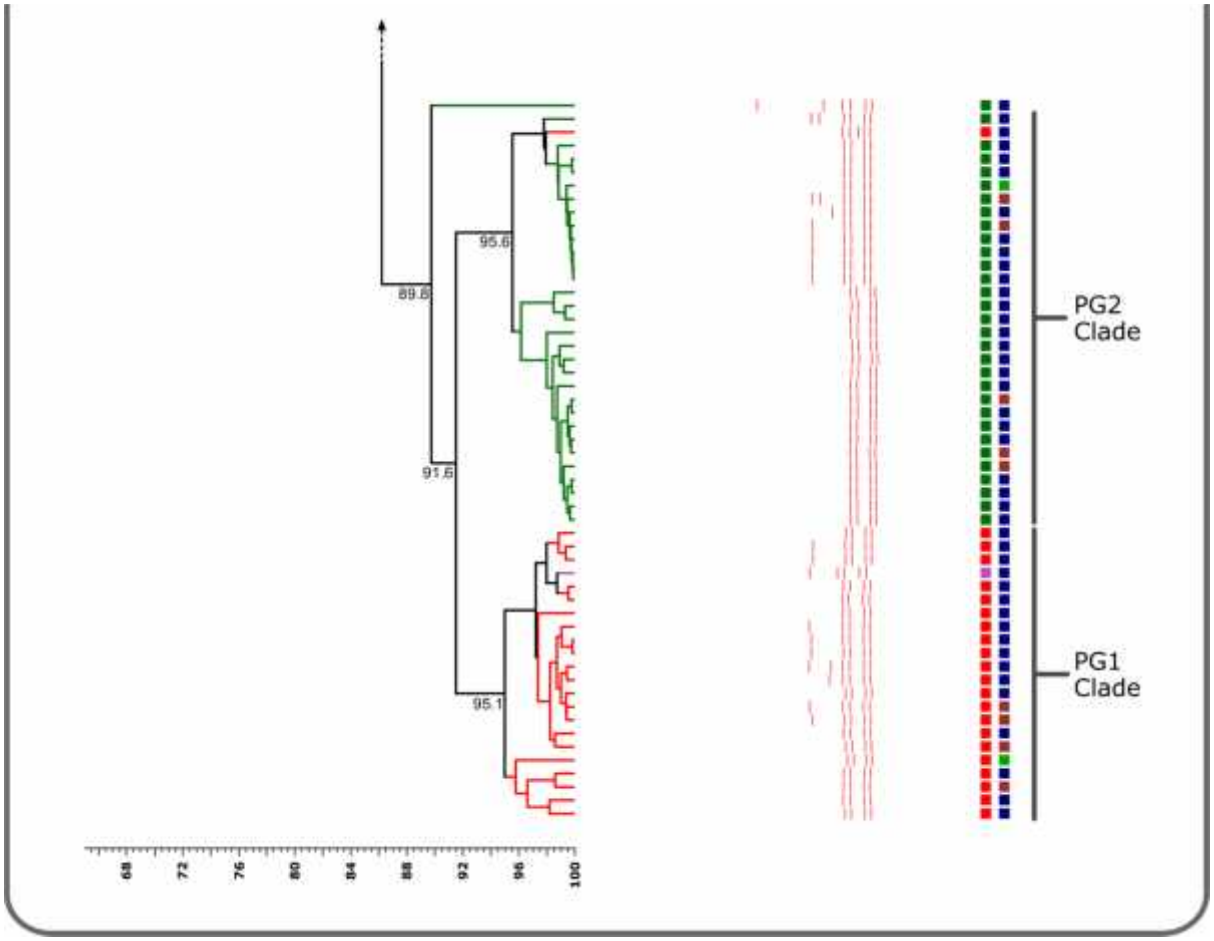
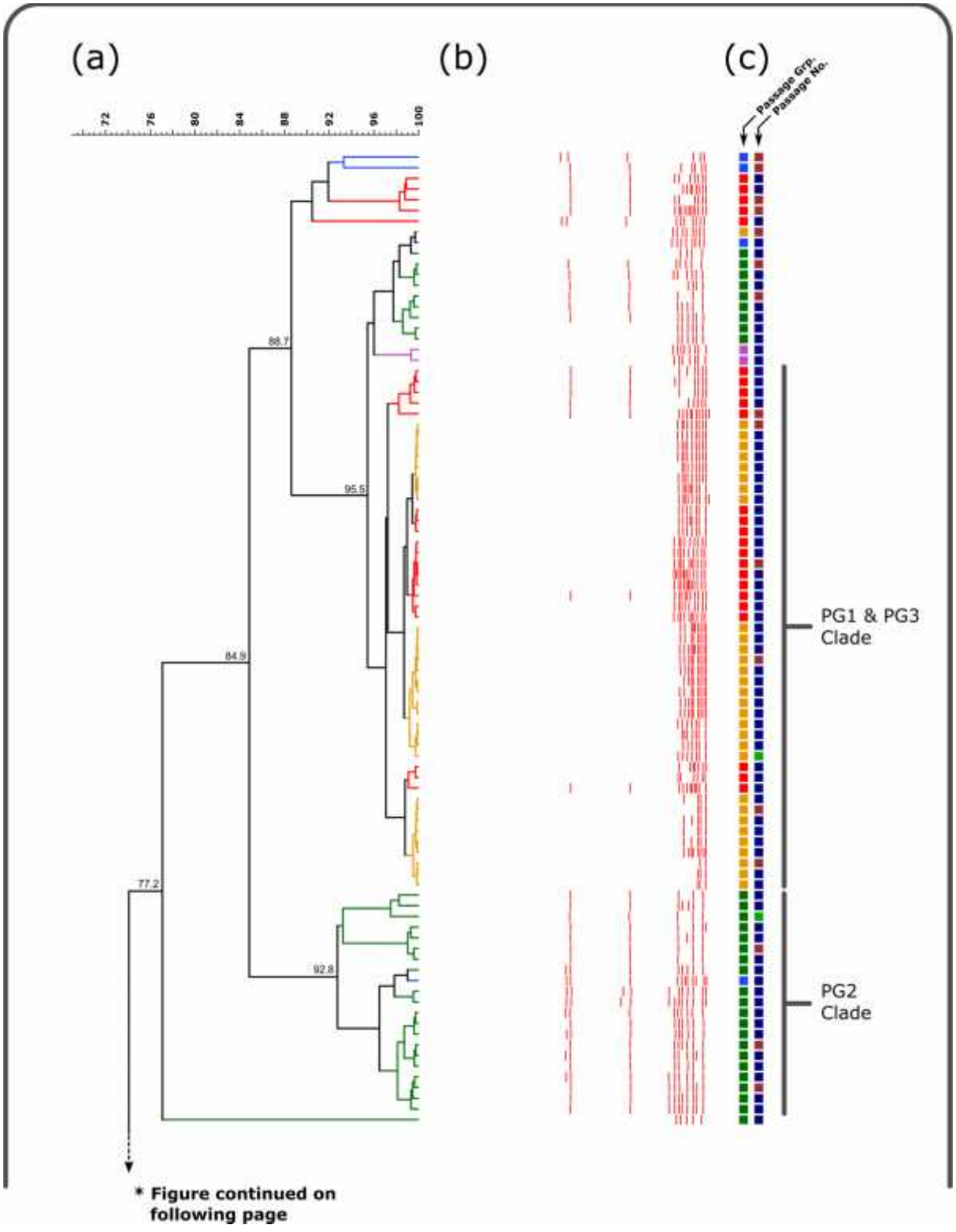



**Figure 3.2.** Cluster analysis of the *DNA polymerase* DGGE profiles of HearNPV populations isolated from third instar *H. armigera* larvae inoculated with an $LD_{90} \times 5$ over three serial *in vivo* passages. **(a)** Dendrogram showing the mean percentage similarity of the profiles. **(b)** A graphical representation of the positions of software-detected bands in each DGGE profile. **(c)** Colour coded strips indicating the passage group and passage number to which each HearNPV population belongs. In the first strip: ■ indicates inoculum samples, ■ indicates PG1, ■ indicates PG2, ■ indicates PG3, ■ indicates PG4 and ■ indicates PG5. In the second strip: □ indicates inoculum sample, ■ indicates passage 1, ■ indicates passage 2 and ■ indicates passage 3.


(a)
(b)
(c)
72
76
80
84
88
92
96
100
88.7
95.5
84.9
77.2
92.8
Passage Grp.
Passage No.
PG1 & PG3
Clade
PG2
Clade
* Figure continued on
following page

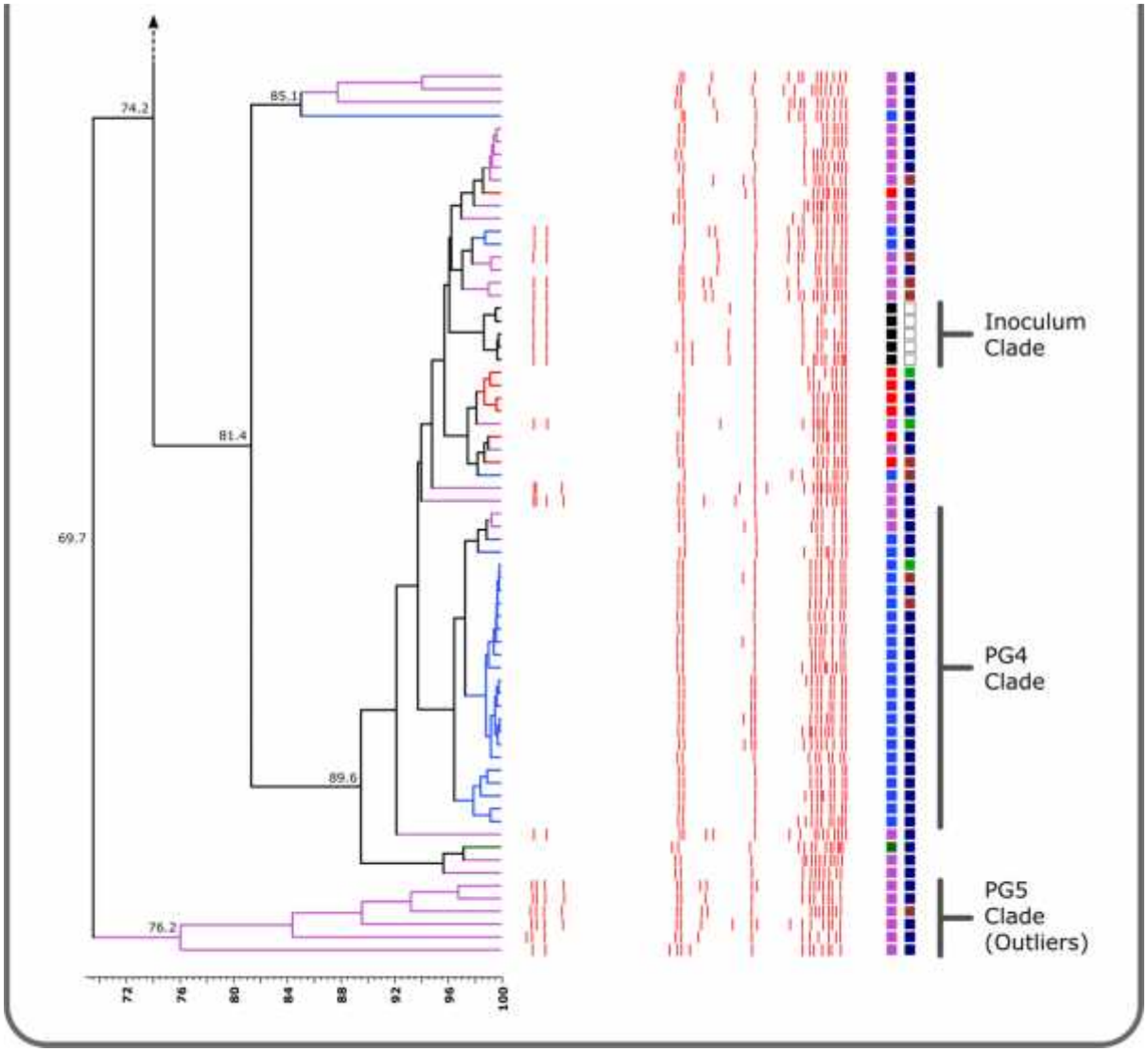



**Figure 3.3.** Cluster analysis of the *me53* DGGE profiles of HearNPV populations isolated from third instar *H. armigera* larvae inoculated with an $LD_{90}$ × 5 over three serial *in vivo* passages. **(a)** Dendrogram showing the percentage similarity of the profiles. **(b)** A graphical representation of the positions of software-detected bands in each DGGE profile. **(c)** Colour coded strips indicating the passage group and passage number to which each HearNPV population belongs. In the first strip: ■ indicates inoculum samples, ■ indicates PG1, ■ indicates PG2, ■ indicates PG3, ■ indicates PG4 and ■ indicates PG5. In the second strip: □ indicates inoculum sample, ■ indicates passage 1, ■ indicates passage 2 and ■ indicates passage 3.

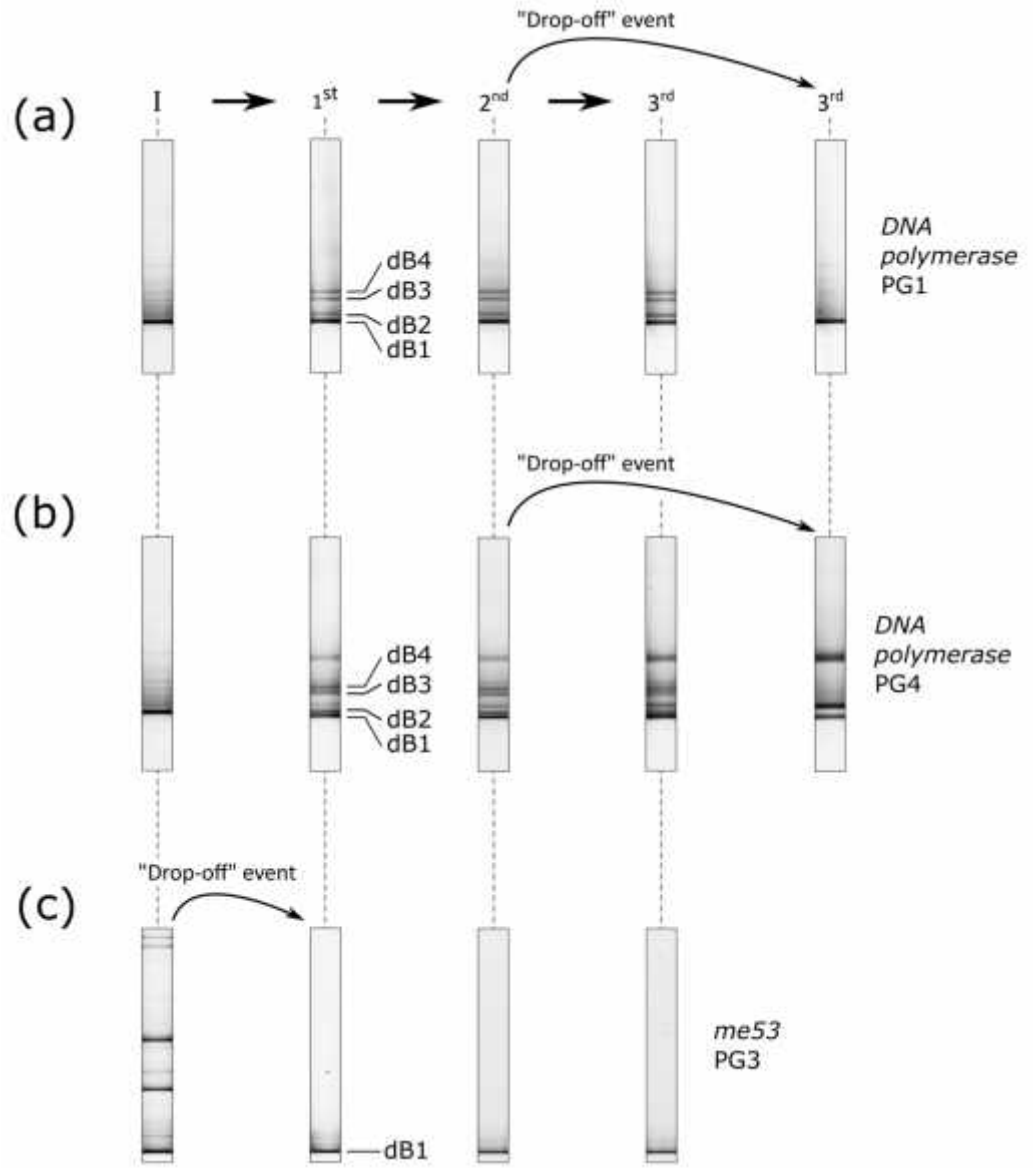



**Figure 3.4.** Examples of shifts in the band composition of DGGE profiles that were observed after successive *in vivo* passages. **(a)** A PG1 passage lineage, where the band composition of the HearNPV *DNA polymerase* inoculum profile (profile I) changed after the first passage and was then largely maintained following the second passage. Two alternative outcomes of the third passage are shown: in the first, the profile of the two preceding HearNPV populations was maintained, whereas in the second a "drop-off" event was observed, with only a single dominant band remaining. **(b)** A PG4 passage lineage is shown where there was a progressive shift in the band composition of *DNA polymerase*, with low intensity inoculum bands increasing in intensity over the three successive *in vivo* passages. An alternative outcome of the third passage is shown, where a "drop off" event occurred. **(c)** A "drop-off" event which occurred between the *me53* inoculum (I) and PG3 profile over the first *in vivo* passage. The number of bands in the *me53* profiles for the PG3 lineage did not increase over the second and third passages and only contained the single dominant band that remained after the "drop-off" event that occurred over the first passage. Positions associated with previously identified true bands (dB1 and dB2) and heteroduplex bands (dB3 and dB4) are labelled on the side of the first passage profiles.

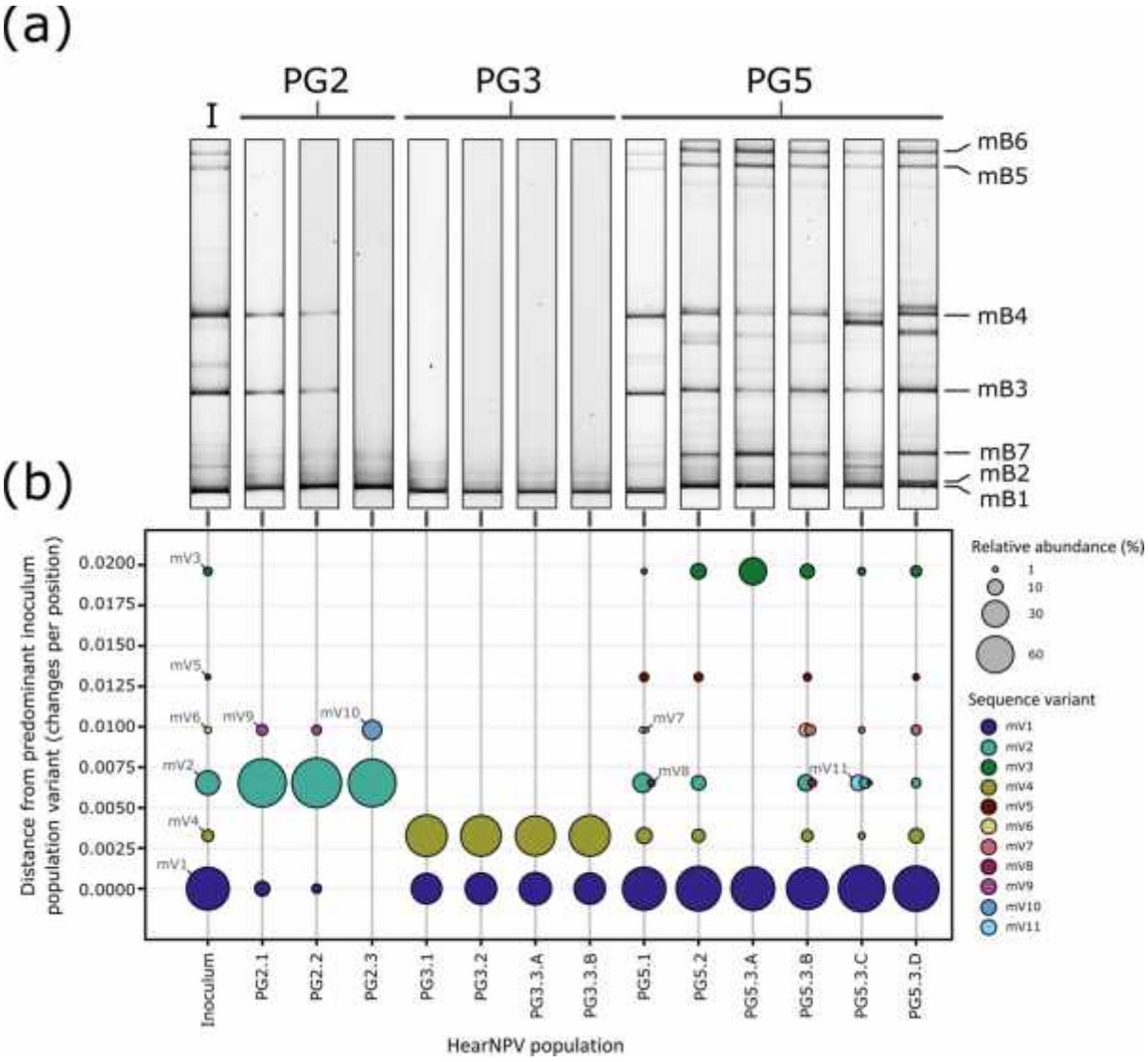



**Figure 3.5.** The *me53* variants detected using deep amplicon sequencing in HearNPV populations where PCR-DGGE detected significant shifts in *me53* variant composition over a passage lineage. Populations selected for sequencing included the original inoculum, three PG2, four PG3 and six PG5 populations. The PG2, PG3 and PG5 populations studied consisted of first, second and third passage populations from the same passage lineage (e.g. for PG2: 1$^{st}$ passage population = PG2.1, 2$^{nd}$ passage population = PG2.2, etc.). **(a)** DGGE profiles for each of the sequenced HearNPV populations. Positions associated with previously identified true bands (mB1, mB2 and mB7) and heteroduplex bands (mB3-mB6) are labelled. **(b)** The *me53* variants detected using deep amplicon sequencing in each HearNPV population. Each bubble in a column represents a *me53* variant, while the size of the bubble shows the relative abundance of the sequence variant in the processed sequencing reads. The position of each bubble on the y-axis indicates the sequence distance of each variant as changes per position from the predominant sequence variant in the inoculum population (mV1). For clarity, the position of each variant is labelled in the first population it occurs.

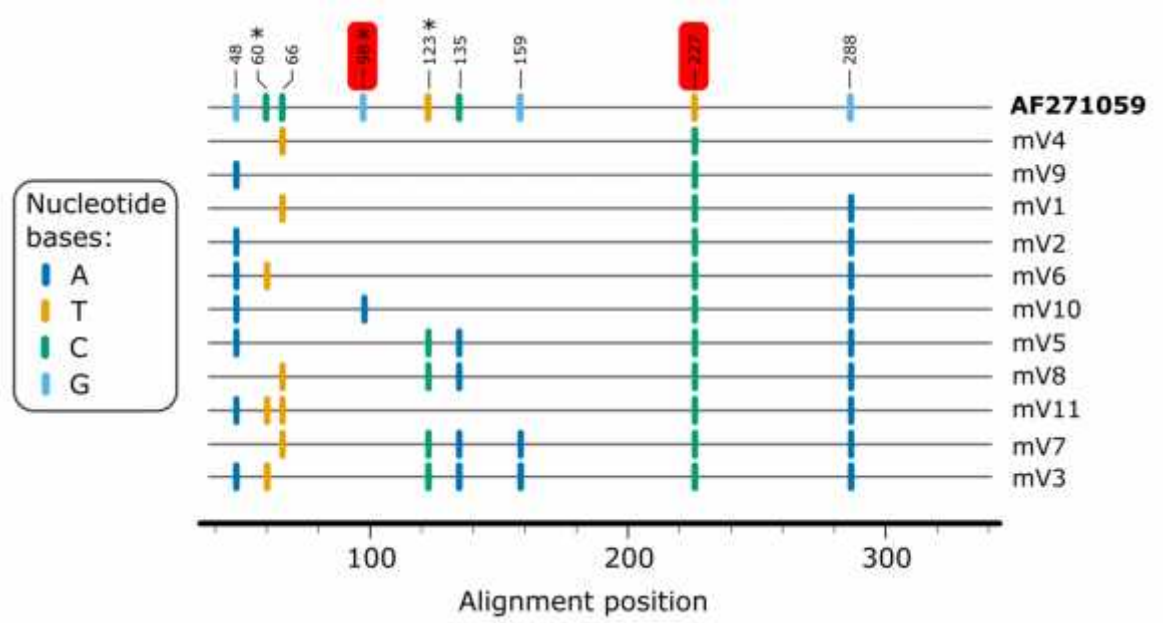



**Figure 3.6.** Highlighter plot showing the nucleotide sequences of the *me53* variants relative to a reference HearNPV *me53* gene sequence (GenBank accession number: AF271059). The positions of SNPs relative to the reference *me53* gene sequence are shown as coloured bars, with each colour denoting one of the four DNA bases. SNPs at positions marked with an asterisk (*) were not present in any of the previously published HearNPV genomes (Baillie & Bouwer, 2012a; Chen et al., 2001; Noune & Hauxwell, 2015, 2016), while those highlighted in red are non-synonymous polymorphisms.

# Chapter 4

# Evaluation of the effect of solar radiation on the genetic diversity of *Helicoverpa armigera* nucleopolyhedrovirus populations

## 4.1. Abstract

Exposure to solar ultraviolet radiation (UVR) has been shown to cause significant levels of nucleopolyhedrovirus (NPV) inactivation in the field. NPV inactivation by UVR is largely a consequence of damage to virus DNA, which either prevents virus genome replication and gene transcription, or results in non-viable genetic mutations. However it is not known whether solar UVR may also contribute to, or affect in some way, the considerable genetic diversity which exists in NPV populations isolated from host insects, which have been shown to consist of assemblies of multiple genotypes. In this study *Helicoverpa armigera* nucleopolyhedrovirus (HearNPV) was used as a model virus to evaluate whether solar radiation potentially affects the genetic diversity of NPV populations. HearNPV homogenates were exposed to solar radiation for 0, 12 or 21 h, after which the exposed viral occlusion bodies were used as inoculum for an *in vivo* passage. The genetic diversity of *DNA polymerase* and *me53* genes in post-passage HearNPV populations was evaluated using a denaturing gradient gel electrophoresis (DGGE) assay and deep amplicon sequencing. Overall solar radiation exposure did not increase the frequency of novel mutations or result in the loss of low frequency variants from post-passage HearNPV populations, even though significant levels of virus inactivation were observed in exposed populations. However, solar radiation exposure was associated with the increased variation in *DNA polymerase* and *me53* variant composition between populations in the 12 h and 21 h exposure groups relative to the 0 h exposure group. This has implications for HearNPV populations in the field where exposure to solar radiation may further add to the divergence in variant composition that occurs between populations over successive infections.

## 4.2. Introduction

Many viruses spend prolonged periods in the environment, where they are exposed to solar radiation, before coming into contact with a potential host (Srinivasan *et al.*, 2001). These include the entomopathogenic nucleopolyhedroviruses (NPVs), which are members of the *Baculoviridae* family. These viruses can persist in the environment for several years before infecting a susceptible insect host (Thompson & Scott, 1979). The characteristic occlusion of multiple NPV virions within a protein matrix, called a polyhedron or occlusion body (OB), is thought to contribute to their stability in the environment, and it has been documented that NPVs in the soil may remain viable for several years (England *et al.*, 1998; Thompson & Scott, 1979). However it has been found that when NPVs are used as biopesticides to control various lepidopteran agricultural pests, NPVs applied to foliage exposed to solar radiation are almost completely inactivated within 24 h (Bullock, 1967; Ignoffo *et al.*, 1972; Yearian & Young, 1974; Young & Yearian, 1974).

It is generally thought that the solar radiation-induced inactivation of NPVs is largely due to damage to their circular dsDNA genomes, which inhibits the transcription of NPV genes and genome replication in an infected host (Petrik *et al.*, 2003; Witt, 1984). The genotoxic property of the ultraviolet (UV) component of solar radiation has been extensively documented, and it has been shown to be a significant cause of DNA damage in a diverse range of life forms (Ikehata & Ono, 2011; Ravanat *et al.*, 2001; Sinha & Häder, 2002). The UV range of the solar radiation spectrum has historically been divided into three wavelength components: UVA (320-400 nm), UVB (280-320 nm) and UVC (200-280 nm). Several laboratory studies using artificial sources of ultraviolet radiation (UVR) have shown that the UVB component of the solar spectrum is largely responsible for NPV inactivation, and that the degree of NPV inactivation is directly proportional to the length of exposure (Griego *et al.*, 1985; Mcleod *et al.*, 1977; Witt, 1984).

UVB radiation exposure induces the formation of dimeric photoproducts, most commonly cyclobutane-pyrimidine dimers (CPDs) and 6-4 photoproducts (6-4 PPs), between adjacent bases in DNA (Ravanat *et al.*, 2001; Sinha & Häder, 2002). UVB-induced DNA damage could lead to three possible outcomes for the genomes of NPVs: (1) DNA lesions are removed and successfully repaired by one of the number of host DNA repair mechanisms that

have evolved in eukaryotic cells; (2) DNA damage causes a localised distortion of the native DNA structure that inhibits the progress of DNA and RNA polymerases, thus preventing DNA replication and transcription, which ultimately leads to virus inactivation; or (3) DNA damage is either not repaired or unsuccessfully repaired and translesional synthesis during DNA replication leads to the incorrect bases being incorporated opposite the lesion, generating a genetic mutation (Pfeifer *et al.*, 2005; Sinha & Häder, 2002). New mutations may either cause synonymous or non-synonymous changes to the nucleotide sequence, and non-synonymous changes may either be biologically viable or non-viable (Duffy *et al.*, 2008; Holmes, 2008). Non-viable mutations are those which completely prevent a gene product from performing its biological function. These may lead to that particular genetic variant being lost from a virus population through purifying selection, if the ability of the variant to successfully replicate in the host is restricted or lost completely (Duffy *et al.*, 2008).

High levels of genetic diversity have been found to exist between NPV populations isolated from different host individuals in the field (Cooper *et al.*, 2003; Shapiro *et al.*, 1991). In addition, NPV populations isolated from individual host insects have been found to consist of multiple genotypes (Baillie & Bouwer, 2012b; Cory *et al.*, 2005). Although other environmental factors, such as the species of plant on which an insect host feeds (Hodgson & Vanbergen, 2002), have been shown to have an effect on the genetic variation within NPV populations, no studies have looked at whether exposure to solar radiation has an effect on NPV genetic diversity. In this study a denaturing gradient gel electrophoresis (DGGE) assay, developed by Baillie & Bouwer (2011), was used together with deep amplicon sequencing to determine whether the exposure of a *Helicoverpa armigera* nucleopolyhedrovirus (HearNPV) inoculum sample to solar radiation, for either 12 h or 21 h, had any effect on the detectable genetic diversity of post-passage HearNPV populations. The DGGE and deep amplicon sequencing assays were used to assess the number and relative abundance of sequence variants in the different HearNPV populations, at two target regions in the *DNA polymerase* and *me53* genes, a core baculovirus and core lepidopteran-specific baculovirus gene, respectively.

Prior to this study it was hypothesised that solar radiation may have one of two effects on the genetic diversity of HearNPV populations. (1) Solar radiation induces the generation of novel, viable mutations in the genomes of HearNPV populations, by causing DNA lesions

which are neither repaired nor result in the significant inhibition of virus genome replication and gene transcription. These new genetic variants increase in abundance over the course of a single *in vivo* passage, to a level where they can be detected. (2) Solar radiation causes the inactivation of a fraction of the HearNPV inoculum population by inducing extensive DNA damage within virus genomes. Low frequency variants are more affected by the inactivation of a fraction of the NPV population than dominant variants, as due to their low frequency they are less likely to be able to sustain major losses. Therefore the solar radiation induced inactivation of virions results in the increased dominance of the existing dominant genetic variants within a HearNPV population following a single post-exposure *in vivo* passage. By comparing the number and relative abundance of different HearNPV genetic variants in each of the exposure groups to a control group, which was not exposed to solar radiation, it was possible to evaluate both hypotheses.

## 4.3. Materials and methods

### 4.3.1. Insect rearing

Eggs from the Agricultural Research Council (Pretoria, South Africa) were used to set up a permanent *H. armigera* laboratory culture. The population was maintained as described in Chapter 2, section 2.3.1 (p. 32).

### 4.3.2. Virus propagation

HearNPV was propagated *in vivo* in third instar (± 5 d old) *H. armigera* larvae. Larvae were inoculated with a lethal HearNPV dose using the diet contamination method described by Evans & Shapiro (1997). Larvae which consumed the entire contaminated diet cube within 24 h were transferred to an artificial wheatgerm diet and monitored twice daily. Cadavers of larvae which had shown overt symptoms of a NPV infection were harvested prior to liquification. Harvested cadavers were pooled and 200 µl of sterile distilled water was added per larva. The pooled larvae were then homogenised using a pestle and mortar and the homogenate stored at 4 ˚C.

### 4.3.3. Exposure of HearNPV to solar radiation

An equal volume (200 µl) of the pooled cadaver homogenate was spread evenly over the bottom of nine 35 mm Petri dishes. The homogenate was allowed to completely air-dry at 25 ˚C. Three of the Petri dishes were not exposed to solar radiation and were stored at −20 ˚C. The remaining six Petri dishes were exposed to natural sunlight on the roof of the Oppenheimer Life Sciences building at the University of the Witwatersrand, Johannesburg, South Africa (GPS coordinates: 26˚11'28.96" S, 28˚1'54.13" E), between the hours of 10h00 and 14h00, over seven consecutive clear days, during May 2013. The Petri dishes were placed on a matt light brown background (Daler-Rowney Murano paper, stone colour) and positioned on an angled platform so that the bases of the Petri dishes were approximately at a 90˚ angle to the sun's rays. The lids of the Petri dishes were removed for the duration of the solar radiation exposure. Three of the six Petri dishes were removed from the solar radiation exposure after a total of 12 h while the remaining three Petri dishes were removed after a total of 21 h. Although the initial intention was to expose homogenate samples to solar radiation for 12 h and 24 h respectively, it was found that a 24 h exposure period resulted in

such extensive HearNPV inactivation (data not shown) that it was not possible to recover sufficient quantities of viable OBs for the subsequent *in vivo* passages at an $LD_{90}$ inoculum. The effect of less than 12 h solar radiation exposure on viral genetic diversity was not assessed, as less than 12 h of exposure did not have any apparent effect on the biological viability of the HearNPV inoculum. After their respective treatments all Petri dishes were stored at −20 ˚C.

### 4.3.4. Post-exposure HearNPV *in vivo* passage

The dried homogenate in each Petri dish was re-suspended in 1 ml of sterile distilled water. OBs were purified from each homogenate using a differential centrifugation protocol. Briefly, the homogenate was centrifuged at low speed (3000 *g*, 30 s) in order to pellet insect tissue debris. OBs were pelleted from the resulting supernatant and washed twice in sterile distilled water during three further rounds of centrifugation (12 000 *g*, 30 s). The three final OB pellets for each exposure group were re-suspended in 250 µl and pooled. The concentration of OBs in each of the three pooled exposure group suspensions was determined using haemocytometry.

For the post-exposure *in vivo* passage, each exposure group HearNPV sample was used to inoculate third instar *H. armigera* larvae with a $LD_{90}$, using the diet cube contamination method described in section 4.3.2. Larvae were harvested after death, prior to the liquification of the cadavers, and stored at −20 ˚C. Previous studies have shown that solar UVR causes the inactivation of HearNPV virions (Bullock, 1967; Ignoffo *et al.*, 1972; Yearian & Young, 1974; Young & Yearian, 1974) and Baillie & Bouwer (2013) showed that inoculum dose had an effect on the amount of detectable genetic diversity in HearNPV populations. Therefore the $LD_{90}$ for each exposure group was calculated in order to prevent a variable number of viable virions in the inoculum doses from having a confounding effect on the amount of genetic diversity detected in the post-*in vivo* passage HearNPV populations. Assays were conducted to determine the dose-mortality relationship of each HearNPV exposure group in third instar *H. armigera* larvae and probit analysis (Bliss, 1934) was used to estimate the group-specific $LD_{90}$. The $LD_{90}$ values were calculated to be $4.54 \times 10^4$, $4.81 \times 10^4$ and $1.23 \times 10^5$ OBs for the 0 h, 12 h and 21 h exposure groups, respectively.

### 4.3.5. DNA extraction

DNA was extracted from the OBs purified from each larva cadaver after the post-exposure *in vivo* passage. DNA extraction was performed using NucleoSpin® Tissue XS kits (Macherey-Nagel) as described in Chapter 2, section 2.3.3 (p. 33). The concentration of the purified DNA samples was determined using a NanoDrop™ ND-1000 spectrophotometer, before storage at −20 °C.

### 4.3.6. PCR-DGGE

A PCR-DGGE assay, developed by Baillie & Bouwer (2011), was used to assess the variant composition and genetic diversity of regions of the *DNA polymerase* and *me53* genes in HearNPV populations across the three solar UVR exposure groups. PCR amplification and DGGE analysis of the targeted gene regions was conducted as described in Chapter 2, section 2.3.4 (p. 33) and section 2.3.5 (p. 34), respectively. The GelCompar II V6.0 software (Applied Maths) was used to analyse DGGE gels as described in Chapter 2, section 2.3.6 (p. 35). GelCompar was used to calculate cosine similarity matrices for the *DNA polymerase* and *me53* DGGE profiles from the three solar UVR exposure groups, and to conduct UPGMA (Unweighted Pair Group Method with Arithmetic Mean) cluster analysis to construct dendrograms showing the mean similarity of DGGE profiles. In addition to the cluster analyses conducted in GelCompar, the similarity between *DNA polymerase* and *me53* DGGE profiles in the three exposure groups were displayed using non-metric multidimensional scaling (NMDS) plots. The NMDS plots were constructed using the cosine similarity matrix (after conversion to a dissimilarity matrix, by subtracting each similarity value from 1) and the vegan package (Oksanen et al., 2016) in R version 3.3.1 (R Core Team, 2016). PERMANOVA and PERMDIST tests were conducted as described in Chapter 3, section 3.3.6 (p. 69) to determine if the DGGE profiles from the three solar UVR exposure groups were statistically significantly different.

### 4.3.7. Deep amplicon sequencing

Deep amplicon sequencing on the MiSeq (Illumina) HTS platform was used to assess the number and identity of *me53* sequence variants in the initial pooled inoculum and selected post-passage HearNPV populations from the three solar UVR exposure groups. The sample preparation and sequencing steps were conducted as described in Chapter 2, section 2.3.11

(p. 38). The raw deep sequencing data was processed and analysed as described in Chapter 2, section 2.3.12. (p. 39). Samples which had <10 000 merged paired-end sequencing reads after the read processing and filtering steps were discarded. All further data analysis steps were conducted in R version 3.3.1 (R Core Team, 2016). Levenshtein distances between sequence variants were calculated using the stringDist function of the Biostrings package (Pages *et al.*, 2016) from the Bioconductor project. The relative abundances and Levenshtein distances for the sequence variants were plotted using the ggplot2 package (Wickham, 2009).

## 4.4. Results

### 4.4.1. DGGE profile analysis

From the dose-mortality assays conducted with the post-solar radiation exposure pooled HearNPV samples, it was clear that both the 12 h and 21 h solar radiation exposure had an effect on the biological viability of the virions within the OBs of these samples. The $LD_{50}$ for HearNPV samples exposed to solar radiation for 12 h and 21 h was higher than that of the non-exposed sample by a factor of 2 and 3, respectively. This indicates that exposure to solar radiation, for 12 h or 21 h, caused damage to HearNPV virions in the cadaver homogenates at a level which was biologically relevant, as it resulted in a reduction in the ability of virus populations to cause disease and mortality in host insects. The DGGE assay was used to determine whether these biologically relevant levels of solar radiation exposure had a detectable effect on the genetic diversity of the HearNPV populations.

The *DNA polymerase* and *me53* dendrograms both contained three distinct solar radiation exposure group clades (Figure 4.1 and Figure 4.2). The dEG0 clade (dEG0 clade = clade containing *DNA polymerase* profiles from 0 h exposure group) was the largest of three *DNA polymerase* clades and contained 70% of all 0 h exposure profiles, while the dEG12 and dEG21 clades contained 55% and 62% of the 12 h and 21 h exposure profiles, respectively (Figure 4.1). The similarity of *DNA polymerase* profiles in each of these clades was high, with each having an intra-clade mean similarity of 98.9%. For the *me53* gene region, 68% of 0 h exposure profiles clustered in mEG0 (98.2% mean similarity), 72% of 12 h exposure profiles in mEG12 (91.2% mean similarity) and 48% of 21 h exposure profiles in mEG21 (96.9% mean similarity) (Figure 4.2). Overall there was less variation between the 0 h exposure *me53* profiles, with profiles for these HearNPV populations showing a greater degree of similarity relative to the 12 h and 21 h exposure profiles. There was one 12 h and three 21 h *me53* profiles which were outliers and showed very low similarity to other *me53* profiles (Figure 4.2). These four profiles differed from most other *me53* profiles in that bands were either absent at the mB3 and mB4 positions, or were present but at a very low intensity. Although there were other *me53* profiles where bands at the mB3 and mB4 positions were lost, these typically were “drop-off” events characterised by a dramatic loss multiple bands after an *in vivo* passage (Figure 4.2). Two of the outlier profiles were also significant as they each contained two *de novo* bands which were not detected in the inoculum profile

(Figure 4.4a, EG12-A and EG21-B). *De novo* bands were defined as any bands which did not appear to be present in either the pre-solar radiation exposure pooled inoculum profiles or any of the other *me53* profiles across the three exposure groups. The presence of *de novo* bands may indicate the presence of novel genetic variants in these two HearNPV populations.

PERMANOVA tests showed that there were statistically significant differences ($\alpha = 0.05$) between the *DNA polymerase* and *me53* profiles in the three exposure groups (Table 4.1). Subsequently *post-hoc* pairwise PERMANOVA comparisons with the Holm-Bonferroni correction were used to assess which exposure groups were statistically significantly different. Pairwise comparisons showed that the *me53* and *DNA polymerase* profiles in all three exposure groups were statistically different ($\alpha = 0.05$) from each other (Table 4.1). However, although profiles were generally more similar to profiles within their own exposure group, there was no obvious definitive difference in the band composition of either *DNA polymerase* or *me53* profiles between the three exposure groups. The majority of *DNA polymerase* profiles had bands at positions that in Chapter 2, section 2.4.3 were shown to be associated with two true bands (dB1 and dB2) and two heteroduplex bands (db3 and dB4), while most *me53* profiles had bands at positions associated with three true bands (mB1, mB2 and mB7) and four heteroduplex bands (dB3-dB6) (Figure 4.1 and Figure 4.2). This suggests that the variant composition of post-passage HearNPV populations was largely maintained across the three exposure groups.

Small shifts in the relative abundance of existing variants as well as the generation of novel variants may have resulted in the greater variation and lower similarity of *me53* profiles observed in the 12 h and 21 h exposure groups. NMDS plots were used as an alternative method to cluster analyses to view differences between profiles in the three exposure groups, as they showed the amount of variation between profiles more clearly. In these plots it can be seen that although *me53* profiles within each group generally showed greater similarity to each other, profiles in the 12 h and 21 h exposure groups showed a greater degree of dispersion relative to profiles in the 0 h exposure group (Figure 4.3b). This was supported by the outcomes of the PERMDIST test, which showed that there were statistically significant differences ($\alpha = 0.05$) in the amount of variation between *me53* profiles in the three exposure groups (Table 4.1). Subsequent *post-hoc* pairwise PERMDIST comparisons with the Holm-Bonferroni correction detected a statistically significant difference in the degree of dispersion

between the 0 h *me53* profiles and the *me53* profiles of the 12 h and 21 h exposure groups. Overall, the cluster analysis and NMDS plot show that there was a lower degree of similarity and greater variation between the *me53* profiles in the 12 h and 21 h exposure groups relative to those in the 0 h group, while the PERMANOVA and PERMDIST tests showed that these differences were statistically significant. For the *DNA polymerase* profiles PERMANOVA tests showed that there were statistically significant ($\alpha = 0.05$) differences between profiles in all three exposure groups (Table 4.1). However, the NMDS plot showed that *DNA polymerase* profiles within the 12 h and 21 h exposure groups overlapped in the ordination space, indicating that profiles in these groups were more similar to each other than they were to profiles in the 0 h exposure group (Figure 4.3a). This supports the results of the cluster analysis where profiles in dEG12 and dEG21 clades had a higher mean similarity, relative to profiles in the dEG0 clade (Figure 4.1). The PERMDIST tests also showed that there were no statistically significant differences ($\alpha = 0.05$) in the amount of variation between profiles in the three exposure groups (Table 4.1). Overall it appears that solar radiation exposure may have had a small but statistically significant effect on the variant composition of HearNPV populations.

### 4.4.2. Deep amplicon sequencing

The cluster analyses and NMDS plots indicated that there were greater differences between *me53* profiles across the three exposure groups, than between *DNA polymerase* profiles (Figure 4.1, Figure 4.2 and Figure 4.3). In addition, the presence of *de novo* bands in the profiles of two *me53* outlier profiles suggests that novel genetic variants were generated over the post-exposure passage in these two populations. Deep amplicon sequencing was used to further assess the *me53* variant composition and underlying nucleotide sequence diversity of selected HearNPV populations across the three solar UVR exposure groups. The pre-exposure HearNPV inoculum population and populations from the three exposure groups, whose DGGE profiles contained all seven of the characteristic *me53* bands, mB1-mB7, were selected for sequencing. These were representative of the majority of populations across the three exposure groups, with most *me53* DGGE profiles containing bands at the mB1-mB7 positions (Figure 4.2). The inoculum population also provided a baseline reference of the presence and relative abundance of *me53* variants prior to the solar UVR exposure and allowed novel genetic variants to be identified in post-exposure populations. Three of the four populations that were outliers in the *me53* profile cluster analysis, where *me53* profiles lacked

prominent bands at the mB3 and mB4 positions, were also selected for sequencing. There was insufficient viral DNA to conduct deep sequencing with the fourth outlier population.

The DGGE profiles of the inoculum population and populations EG0-A, EG0-B, EG12-B and EG21-C all contained the seven characteristic *me53* bands (mB1-mB7). Deep amplicon sequencing showed that these five populations had similar, although not identical, *me53* variant compositions (Figure 4.4a-b). All six of the inoculum variants, mV1-mV6, were detected in the 0 h exposure population EG0-B, and the variant composition of this population most closely resembled that of the inoculum population (Figure 4.4b). Populations EG0-A and EG12-B both contained five of the six inoculum variants (mV1-mV4 and mV6), while population EG21-C contained only four of the six inoculum variants (mV1-mV3 and mV6) (Figure 4.4b). In addition, a novel genetic variant, mV7, was detected in population EG12-B (Figure 4.4b). However, mV7 was the lowest frequency variant in this population and only accounted for 0.83% of the sequencing reads. The three sequenced outlier populations (EG12-A, EG21-A & EG21-B) shared several attributes and differed in composition from the above five populations in that the second most abundant inoculum *me53* variant, mV2, as well as the low frequency variants mV5 and mV6 were absent (Figure 4.4b). Another defining characteristic of the outlier populations was the increased abundance of the low frequency inoculum variant, mV3, which typically accounted for 1-2% of sequencing reads in the other sequenced populations but accounted for 20-40% of the reads in the outlier populations. Despite the differences in variant composition between the outlier populations and the other sequenced populations, the most abundant inoculum variant, mV1, remained the most abundant variant in all of the sequenced post-exposure populations (Figure 4.4b).

Two different novel genetic variants, mV12 and mV13, were detected in the two outlier populations, EG12-A and EG21-B, respectively (Figure 4.4b and Figure 4.5). These are the two outlier populations whose DGGE profiles contained the *de novo* bands (Figure 4.4a). Unlike the novel genetic variant, mV7, detected in EG12-B, the mV12 and mV13 variants accounted for a significant proportion of the sequencing reads in the EG12-A and EG21-B populations (28% and 25%, respectively). It appears likely that all three of the novel genetic variants were generated through recombination occurring between existing inoculum genetic variants. These variants did not contain any novel SNPs and given that the HearNPV DNA

polymerase is thought to have proofreading activity (Darmawan *et al.*, 2015; Hang & Guarino, 1999; Huang & Levin, 2001; Mikhailov *et al.*, 1986) it was thought less likely that new SNPs would spontaneously arise through polymerase errors at the same sites as the SNPs in existing variants (Figure 4.5). All of the SNPs detected were synonymous and the nine nucleotide sequence variants would encode the same amino acid sequence (Figure 4.5). The lack of novel synonymous and non-synonymous SNPs suggests that solar radiation exposure did not greatly increase the frequency at which new mutations were generated in HearNPV populations.

## 4.5. Discussion

Prior to this study it was hypothesised that solar radiation may affect the genetic diversity of HearNPV populations in one of two ways; either by inducing genetic mutations, which may lead to the generation of new viable genetic variants, or by reducing the representation of low frequency genetic variants through the inactivation of a proportion of a HearNPV population. A number of whole genome deep sequencing studies have shown that there are a significant number of synonymous and non-synonymous SNPs distributed throughout NPV genomes, which contribute to the considerable genetic diversity observed in NPV populations (Baillie & Bouwer, 2012a; Chateigner *et al.*, 2015; Thézé *et al.*, 2014; Xu *et al.*, 2013). Of particular relevance was the study by Baillie & Bouwer (2012a) which detected 60 SNPs within the *DNA polymerase* gene and 25 SNPs within the *me53* gene, across seven geographically distant HearNPV isolates. Although 3′ − 5′ exonuclease proofreading activity has not been conclusively demonstrated for the HearNPV DNA polymerase, it is a core, conserved baculovirus gene and proofreading activity has been shown to be associated with the enzymes of two other NPVs, *Bombyx mori* NPV and *Autographa californica* MNPV (Hang & Guarino, 1999; Mikhailov *et al.*, 1986). DNA polymerase proofreading activity should preclude HearNPV and other NPV species from having a high rate of spontaneous mutation. It is tempting to contemplate the possibility that a significant number of the SNPs detected by whole genome studies could have their origins in UVR-induced DNA damage events, especially as these viruses may spend prolonged periods in the environment prior to infecting a new host. Many studies have documented the link between UVR exposure in humans and the accumulation of mutations in cancer suppressor genes, where UVR signature mutations are found in 80% of precancerous lesions (Narayanan *et al.*, 2010; Rass & Reichrath, 2008). However no studies could be found in the literature that looked at the potential beneficial contribution of UVR induced mutations to the evolution and genetic diversity of prokaryotic or eukaryotic organisms and viruses, although it has been suggested that its role as a mutagen may have contributed to the evolution of early eukaryotes (Rothschild, 1999).

In this study *de novo* bands were detected in the *me53* DGGE profiles of two isolates exposed to solar radiation for 12 h and 21 h (4.8% of profiles from populations exposed to solar radiation); while no *de novo* bands were detected in any of the *DNA polymerase* profiles. The two *de novo* bands were not detected by the GelCompar software or visible with the naked

eye, in either the *me53* profile of the initial HearNPV inoculum population or any of the other 0 h, 12 h or 21 h profiles. Subsequent deep sequencing confirmed that both of these populations contained a different novel genetic variant which accounted for 28% and 25% of the sequencing reads of each population. In addition, a third low frequency novel genetic variant, which only accounted for 0.83% of sequencing reads, was detected in a second 12 h exposure population. However all three of these novel genetic variants were likely to have been generated through genomic recombination between existing variants present in the inoculum population. The paucity of *de novo* bands in the *DNA polymerase* and *me53* profiles and of novel synonymous and non-synonymous mutations in the *me53* variants of populations exposed to solar radiation indicates that the exposure of HearNPV OBs to solar radiation does not greatly increase the frequency of novel mutations in HearNPV populations. The generation of novel genetic variants over an *in vivo* passage was described in Chapter 3. Similar to the novel genetic variants detected in the 12 h and 21 h exposure populations, the novel genetic variants described in Chapter 3 were most likely generated through genomic recombination occurring between existing variants during the recombination-dependent replication of virus genomes over the course of an infection (Hajós *et al.*, 2000; Okano *et al.*, 2006). This suggests that the detection of three novel genetic variants in the post-exposure populations was not exceptional and that solar radiation exposure potentially played no role in their generation, as similar numbers of novel genetic variants were observed in post-passage populations in the absence of pre-passage solar radiation exposure, as reported in Chapter 3.

The first hypothesis was largely concerned with the effect of solar radiation on the genetic diversity of HearNPV populations from the perspective of UVR-induced genetic mutations in individual viruses. However the second hypothesis considered the effect of solar radiation on the relative abundance of the different variants in a HearNPV population and whether shifts occurred in the variant composition of HearNPV populations after exposure to solar radiation. The inactivation of viruses by solar radiation, which causes extensive damage to viral DNA and RNA genomes, has been extensively documented, and is in no way limited to the *Baculoviridae* family (Jacquet & Bratbak, 2003; Sagripanti & Lytle, 2007; Weber & Stilianakis, 2008). The UV component of solar radiation has been implicated in the loss of infectivity of marine cyanophages, and has even been suggested to contribute to the seasonal nature of influenza pandemics; where flu season coincides with the lower level of solar UVR

induced virus inactivation during the winter months (Noble & Fuhrman, 1997; Sagripanti & Lytle, 2007). The large degree of solar UVR induced HearNPV inactivation observed in this study, as well as in previous studies (Bullock, 1967; Ignoffo *et al.*, 1972; Yearian & Young, 1974; Young & Yearian, 1974) suggests that host-mediated DNA repair mechanisms do not act to effectively restore the activity of solar-damaged HearNPV virions. Even the ODVs of a recombinant AcMNPV, which expressed an algal virus DNA repair enzyme, did not show reduced susceptibility to inactivation by solar UVR (Petrik *et al.*, 2003). This suggests that the supportive cellular apparatus necessary for DNA repair may not be active in lepidopteran midgut epithelial cells, or that there is just not enough time for damaged viral DNA to be repaired in the midgut epithelium before these cells are lost through midgut sloughing (Petrik *et al.*, 2003).

It was thought that the inactivation of HearNPV virions in a larva homogenate by solar radiation would be a random process. Whether a particular virion would be inactivated or not would be dependent on the length of its exposure and the degree of shading its OB experienced in the larva homogenate. It was hypothesised that under this scenario, the representation of lower frequency genetic variants within a HearNPV population would decline after a post-exposure *in vivo* passage, as lower frequency variants would be less likely to be able to sustain the inactivation losses. Increases in the length of the solar radiation exposure would lead to the increased dominance of already dominant variants and the potential elimination of certain low frequency variants from post-exposure *in vivo* passage populations. Although the PERMANOVA tests showed that statistically significant differences existed between DGGE profiles of the three exposure groups, there appeared to be no obvious definitive difference in the band composition of their *DNA polymerase* and *me53* profiles. This suggests that although exposure to solar UVR may have had a statistically significant but small effect on the variant composition of HearNPV populations.

Across the three exposure groups most *DNA polymerase* profiles contained the four bands dB1-dB4 while the majority of *me53* profiles contained the seven bands mB1-mB7. Although the disappearance of low frequency *me53* variants from the 12 h and 21 h exposure outlier populations could be attributed to the effects of solar radiation exposure, similar “drop-off” events were observed to occur in a small proportion of *in vivo* passages in the absence of pre-passage solar radiation exposure, as described in Chapter 3. In addition, with the exception of

the four outlier populations, the sequenced post-exposure populations had similar *me53* variant compositions to the pre-exposure inoculum and 0 h exposure populations, and contained two to three of the five low frequency *me53* variants which together accounted for less than 10% of inoculum population sequencing reads. It is clear from these results that low frequency *DNA polymerase* and *me53* genetic variants were not progressively lost from HearNPV populations as the length of solar radiation exposure was increased. A potential explanation for the conservation of low frequency variants may be that OBs do not contain a homologous group of genotypes, but rather a "mixed-bag" of genotypic variants. This is known as the co-occlusion of variants, and has been shown to occur during HearNPV infections (Arrizubieta *et al.*, 2015). If the co-occlusion of variants frequently occurs during HearNPV infections this would decrease the likelihood of solar radiation-induced inactivation reducing the probability of low frequency variants infecting a host, as they will be distributed over a greater number of OBs in the population.

Although solar radiation exposure did not increase the frequency of novel genetic mutations or result in the loss of low frequency genetic variants from HearNPV populations, there was greater variation in the *me53* variant composition of HearNPV populations in the 12 h and 21 h exposure groups relative to the 0 h exposure group. This was shown by the lower degree of similarity between *me53* DGGE profiles in the 12 h and 21 h exposure groups relative to the 0 h exposure group, as shown by the cluster analysis and the greater dispersion of these two groups in the NMDS plot. The PERMDIST tests also showed that the variation of *me53* profiles in the 12 h and 21 h exposure groups was statistically significantly different to the amount of variation in the 0 h exposure group. From the NMDS plot of *me53* profiles it is clear that there was greater variation between profiles in the 12 h and 21 h exposure groups relative to the 0 h exposure group. This may indicate that exposure to solar radiation and the associated inactivation of virions results in perturbations to the variant composition of HearNPV populations. It appears that these perturbations are not characterised by major shifts in the variant richness but rather by smaller differences in the relative abundance of both existing and novel variants between HearNPV populations.

## 4.6. Conclusions

This study has shown that the pre-infection exposure of HearNPV to solar radiation appears to have a limited effect on the variant composition of HearNPV populations, especially when compared to the changes in composition that can occur as consequence of an *in vivo* passage in the absence of solar radiation exposure (as described in Chapter 3). The pre-infection exposure of virions to solar radiation did not increase the frequency of novel mutations in post-passage HearNPV populations. Although increased exposure to solar radiation did lead to higher rates of virus inactivation, as reported in previous studies (Bullock, 1967; Ignoffo *et al.*, 1972; Yearian & Young, 1974; Young & Yearian, 1974), this was not clearly associated with the loss of low frequency genetic variants from post-passage populations. The only clear effect associated with solar radiation exposure detected in this study was the greater variation in *me53* variant composition between populations in the 12 h and 21 h exposure groups relative to the 0 h exposure group. This has implications for HearNPV populations in the field where exposure to solar radiation may further add to the divergence in variant composition that occurs between populations over successive infections.

**Table 4.1.** The results of the PERMANOVA and PERMDIST analyses of *DNA polymerase* and *me53* DGGE profiles from the three solar UVR exposure groups. Both analyses were implemented using the cosine dissimilarity matrices for the DGGE profiles and *P* values were calculated following 9999 permutations according to the methods of Anderson (2001, 2006).

| Analysis | PERMANOVA | | PERMDIST | |
|---|---|---|---|---|
| | ***F*** | ***P*** | ***F*** | ***P*** |
| *DNA polymerase* | $F_{2, 58} = 11.05$ | < 0.001 | $F_{2, 58} = 4.39$ | < 0.05 |
| Pairwise* | 0 12 21 | | 12 0 21 | |
| *me53* | $F_{2, 60} = 10.88$ | < 0.001 | $F_{2, 60} = 5.60$ | < 0.05 |
| Pairwise* | 0 12 21 | | 0 12 21 | |

* Where there was a statistically significant overall *F*-ratio ($P < 0.05$, 9999 permutations) for DGGE profiles in the different passage groups pairwise comparisons were done. The alpha value for these pairwise tests was corrected for multiple comparisons using the Holm-Bonferroni correction (Holm, 1979). The numbers represent each of the three solar UVR exposure groups, and the underlining bars indicate groups which were not statistically significantly different ($P > 0.05$).

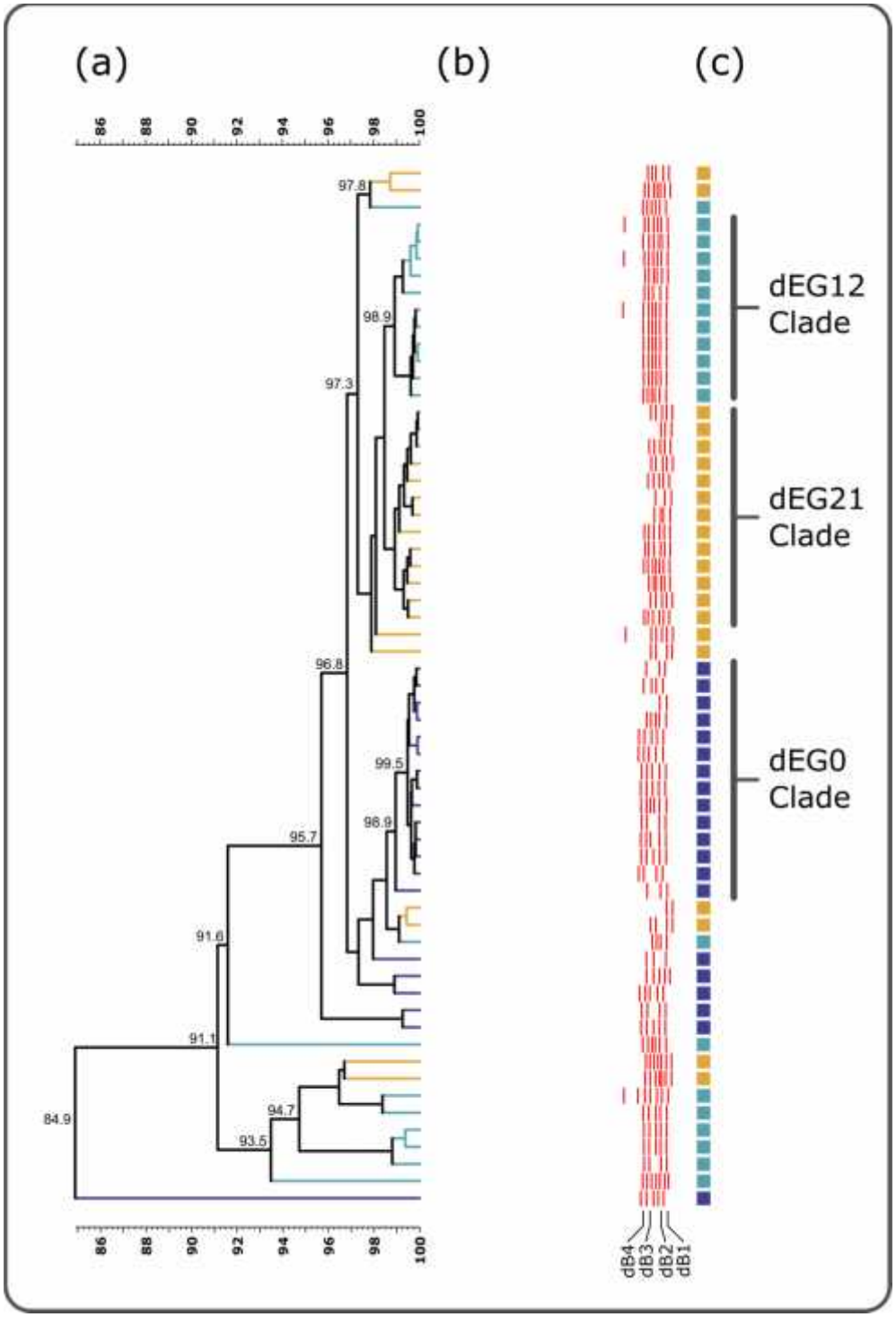



**Figure 4.1.** Cluster analysis of the *DNA polymerase* DGGE profiles of HearNPV populations isolated from third instar *H. armigera* larvae inoculated with HearNPV exposed to either 0 h, 12 h or 21 h of solar radiation. **(a)** Dendrogram showing the percentage similarity of the profiles. **(b)** A graphical representation of the relative positions of software-detected bands in each DGGE profile. Positions associated with previously identified true bands (dB1 and dB2) and heteroduplex bands (dB3 and dB4) are labelled. **(c)** Colour coded strip indicating the solar radiation treatment group to which each HearNPV population belongs. ■ indicates 0 h solar radiation exposure, ■ indicates 12 h solar radiation exposure and ■ indicates 21 h solar radiation exposure.

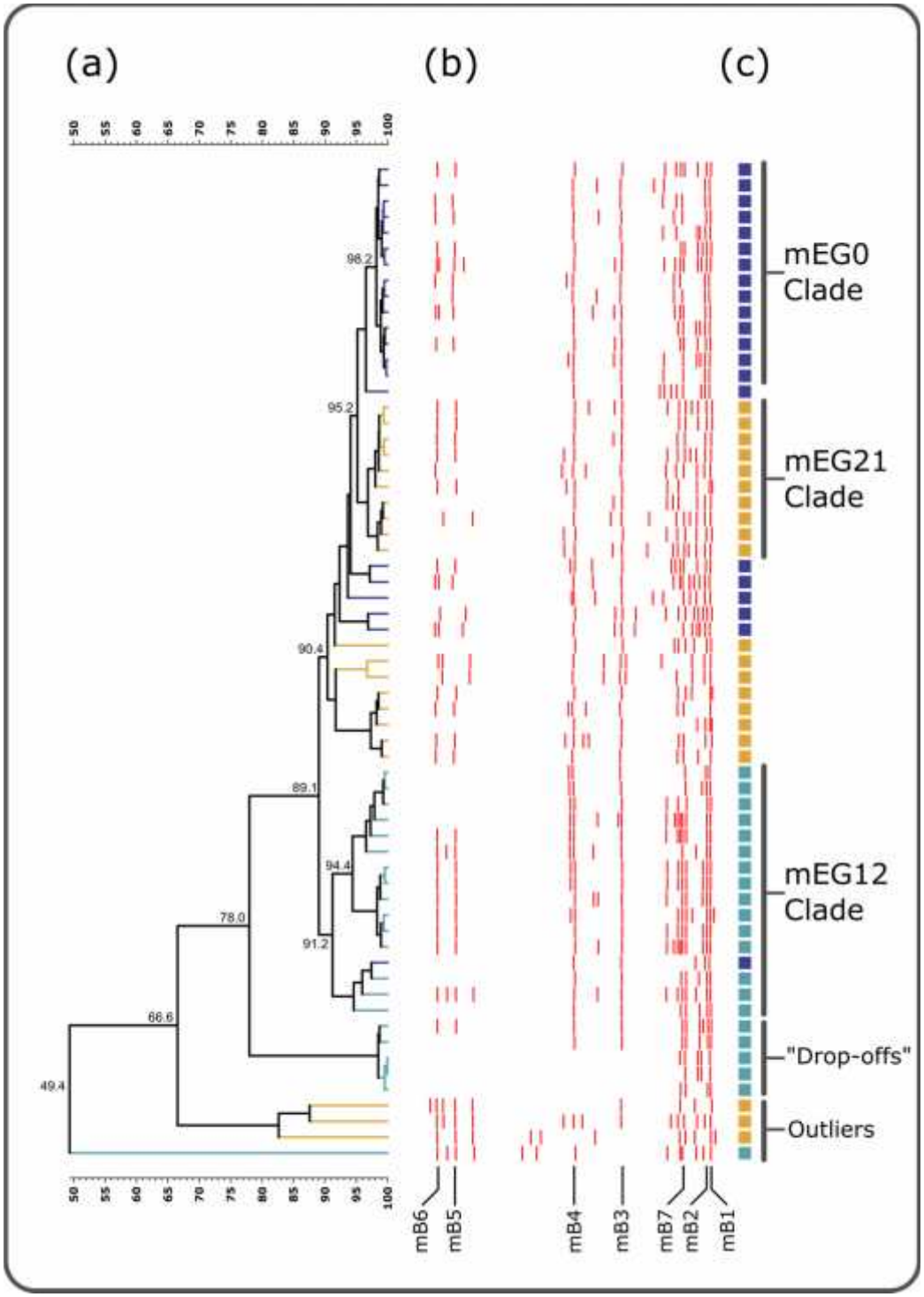



**Figure 4.2.** Cluster analysis of the *me53* DGGE profiles of HearNPV populations isolated from third instar *H. armigera* larvae inoculated with HearNPV exposed to either 0 h, 12 h or 21 h of solar radiation. **(a)** Dendrogram showing the percentage similarity of the profiles. **(b)** A graphical representation of the relative positions of software-detected bands in each DGGE profile. Positions associated with previously identified true bands (mB1, mB2 and mB7) and heteroduplex bands (mB3-mB6) are labelled. **(c) C**olour coded strip indicating the solar radiation treatment group to which each HearNPV population belongs. ■ indicates 0 h solar radiation exposure, ■ indicates 12 h solar radiation exposure and ■ indicates 21 h solar radiation exposure.

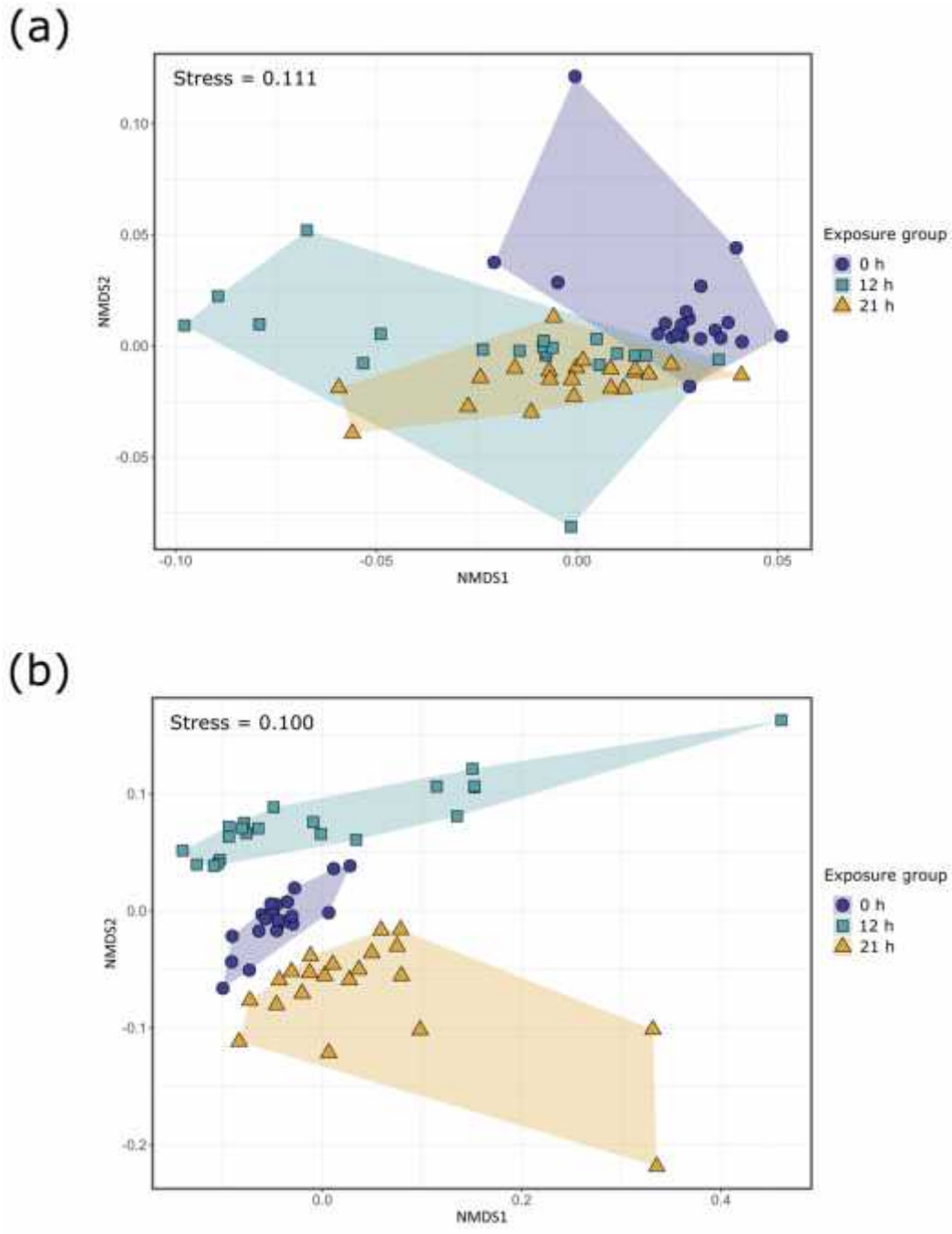



**Figure 4.3.** Non-metric multidimensional scaling (NMDS) analyses of *DNA polymerase* and *me53* DGGE profiles for HearNPV populations from the three solar UVR exposure groups. **(a)** NMDS plot for *DNA polymerase* DGGE profiles. **(b)** NMDS plot for *me53* DGGE profiles. NMDS was performed using the cosine dissimilarity matrix for each gene region. Each point in an NMDS plot represents a single DGGE profile, with the distance between any two points representing the relative dissimilarity of those two profiles. The highlighted areas encompass all points representing the DGGE profiles for each solar UVR exposure group. The "stress" parameter measures the lack of fit between DGGE profile distances in the NMDS ordination space and the calculated dissimilarity values in the dissimilarity matrix (Paliy & Shankar, 2016). Stress values ≤ 0.15 are considered generally acceptable (Clarke, 1993).

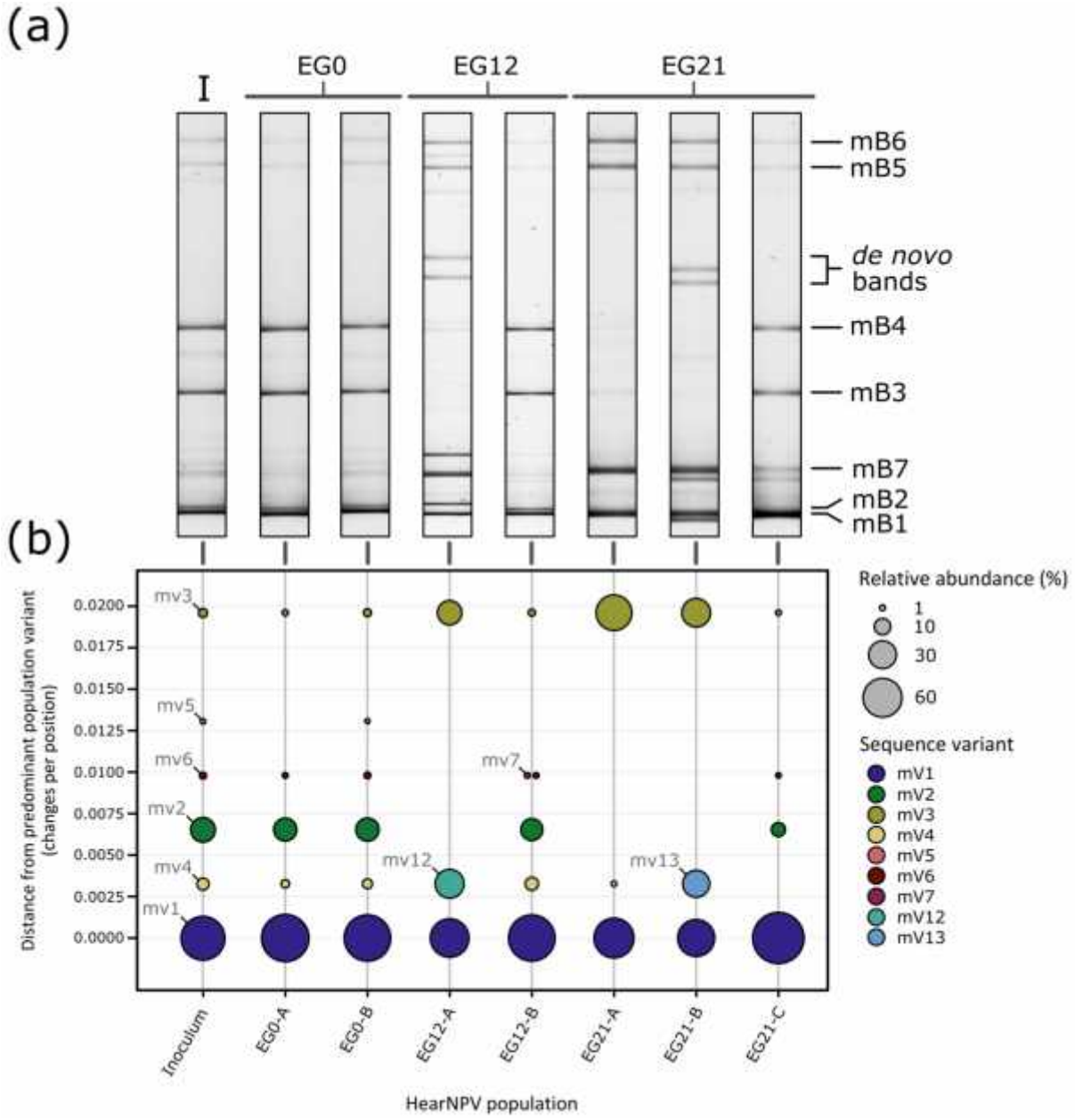



**Figure 4.4.** The *me53* variants detected using deep amplicon sequencing in selected HearNPV populations across the three solar UVR exposure groups. Populations selected for sequencing included the original HearNPV inoculum, two 0 h, two 12 h and three 21 h populations. **(a)** DGGE profiles for each of the HearNPV populations selected for sequencing. Positions in profiles previously identified as true bands (mB1, mB2 and mB7) and heteroduplex bands (mB3-mB6), as well as the *de novo* bands detected in EG12-A and EG21-B, are labelled. **(b)** The *me53* variants detected in each HearNPV population. Each bubble in a column represents a *me53* variant, while the size of the bubble shows the relative abundance of the sequence variant in the processed sequencing reads. The position of each bubble on the y-axis indicates the sequence distance of each variant as changes per position from the predominant sequence variant in the inoculum population (mV1). For clarity, the position of each variant is labelled in the first population it occurs.

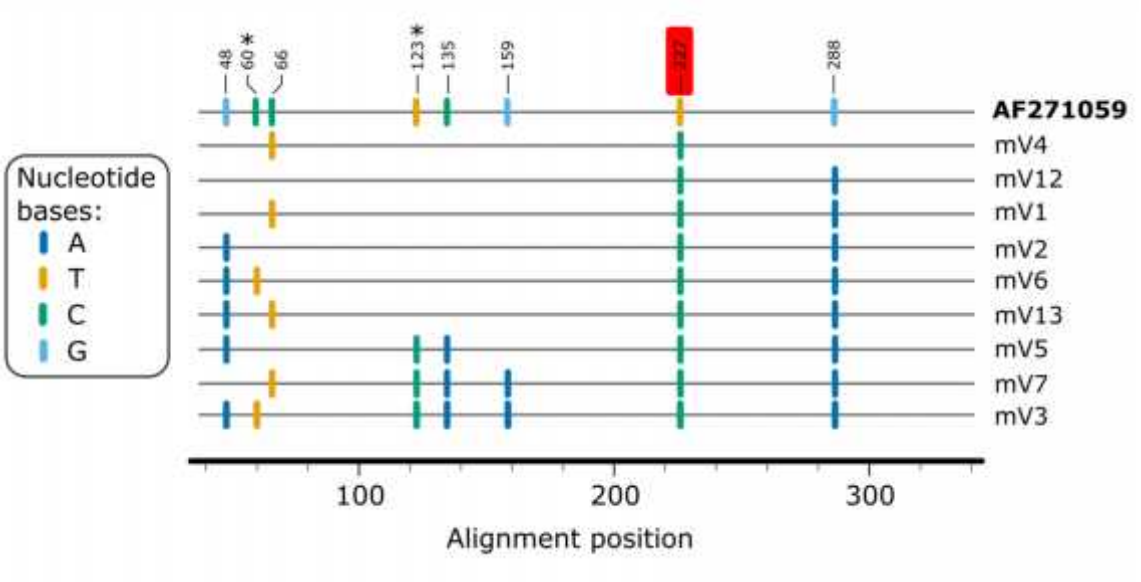



**Figure 4.5.** Highlighter plot showing the nucleotide sequences of the *me53* variants relative to a reference HearNPV *me53* gene sequence (GenBank accession number: AF271059). The positions of SNPs relative to the reference *me53* gene sequence are shown as coloured bars, with each colour denoting one of the four DNA bases. SNPs at positions marked with an asterisk (*) were not present in any of the previously published HearNPV genomes (Baillie & Bouwer, 2012a; Chen *et al.*, 2001; Noune & Hauxwell, 2015, 2016), while those highlighted in red are non-synonymous polymorphisms.

# Chapter 5

# The effect of ODV inoculum dose on genetic variation within *Helicoverpa armigera* nucleopolyhedrovirus populations

## 5.1. Abstract

It is not clear how inoculum dose, a key infection variable, affects the level of genetic diversity in *Helicoverpa armigera* nucleopolyhedrovirus (HearNPV) populations after an *in vivo* passage. The HearNPV inoculum dose may affect the number of ODVs that initiate infection in the host midgut, a factor which is thought to be a key determinant of genetic diversity in HearNPV populations isolated from an infected insect. Of particular interest is whether a single occlusion derived virion (ODV) is capable of causing a lethal infection in a host and what effect this would have on the post-passage genetic diversity of a HearNPV population. In cases where a single ODV initiates infection, a host would be infected with a single genotype and it is unclear whether the post-passage population would be genotypically homogeneous or whether the generation of novel genetic variants over the course of the infection would result in a genotypically heterogeneous population. In this study the variant composition and genetic diversity of HearNPV populations isolated from neonate *H. armigera* larvae inoculated with a range of ODV doses ($LD_5$, $LD_{25}$, $LD_{50}$, $LD_{75}$ or $LD_{95}$) was assessed using a *me53*-specific denaturing gradient gel electrophoresis assay and deep amplicon sequencing. It was found that at the lowest ODV dose ($LD_5$) most larvae appeared to have been infected by one HearNPV genotype, suggesting that infection by a single ODV is possible. In addition it was found that the higher the ODV inoculum dose, the more likely the genetic variant composition of the inoculum population would be maintained after a single *in vivo* passage. Across the different ODV doses the most abundant inoculum variant also appeared to have the highest probability of infecting a new host. The findings suggest that ODVs in the midgut infect a *H. armigera* larva in a frequency-dependent manner, where genotypic variants that have the greatest abundance in the ODV pool have the highest probability of infecting the new host.

## 5.2. Introduction

Members of the *Baculoviridae* family are a group of dsDNA entomopathogenic viruses which are characterised by their unique biphasic replication cycles and the occlusion of virions within crystalline protein matrices called occlusion bodies (OBs) (van Oers & Vlak, 2007; Okano *et al.*, 2006; Rohrmann, 2013; Theilmann & Blissard, 2008). Historically the baculoviruses were divided into two groups, the nucleopolyhedroviruses (NPVs) and the granuloviruses (GVs) (Jehle *et al.*, 2006b). The OBs of NPVs are called polyhedra and contain multiple virions whereas those of the GVs are called granules and only contain a single virion (Jehle *et al.*, 2006b; van Oers & Vlak, 2007). Furthermore occlusion derived virions (ODVs) of NPVs belonging to the same species generally exhibit one of two phenotypes, with the viral envelope enclosing either a single rod-shaped nucleocapsid (SNPV) or multiple nucleocapsids (MNPV), which encapsulate a circular dsDNA genome (van Oers & Vlak, 2007; Rohrmann, 2014; Theilmann & Blissard, 2008).

It has been extensively documented that there exists a significant amount of genotypic variation between field-collected NPV isolates belonging to the same virus species (Cory & Myers, 2003; van Oers & Vlak, 2007). Genetic variation does not only exist between isolates from different host individuals, but has also been shown to exist within virus populations isolated from a single insect (Baillie & Bouwer, 2012; Cory *et al.*, 2005; Hodgson *et al.*, 2001). A number of studies have shown that the maintenance of multiple genotypes in a NPV population is often beneficial, as mixed genotype infections can result in greater pathogenicity and higher OB yields, relative to infections caused by the constituent genotypes on their own (Hodgson *et al.*, 2004; López Ferber *et al.*, 2003; Simón *et al.*, 2005, 2008). It has been found that mixed genotype infections *in vivo* can produce OBs where genotypically distinct virions are co-occluded within the same OB (Arrizubieta *et al.*, 2015; Clavijo *et al.*, 2010; López Ferber *et al.*, 2003). The number of virions which initially infect a new host is known to be a key determinant of virus genetic diversity in a number of virus species (Carlson *et al.*, 2014; Sim *et al.*, 2015; Zwart *et al.*, 2009, 2011), as it affects the probability that virus genotypes in the inoculum population will be transmitted to the new host (Zwart *et al.*, 2011). The physical association of multiple genotypes within the baculovirus transmission body (OB) is thought to increase the probability that a new host will be exposed to and subsequently infected by a genetically diverse population of virions

(Bernal *et al.*, 2013; Clavijo *et al.*, 2010). However, the occlusion of multiple virions within a single transmission body does not always ensure hosts are infected by a large number of virions, as studies have shown that at low OB inoculum doses a host may only be infected by a very small number of virions (Smith & Crook, 1988; Zwart *et al.*, 2009).

A study by Baillie & Bouwer, (2013) found significant differences in the level of genetic diversity between *Helicoverpa armigera* nucleopolyhedrovirus (HearNPV) populations isolated from *H. armigera* larvae inoculated with low ($LD_5$) and high ($LD_{95}$) OB inoculum doses, respectively. Their study clearly showed that inoculum dose is a key infection variable which affects the genetic diversity of HearNPV populations isolated following an *in vivo* passage. However, it is still unknown whether a single HearNPV ODV is capable of causing a lethal infection in a *H. armigera* larva and what effect this would have on the post-passage genetic diversity of a virus population. HearNPV is a SNPV, and were a single ODV to initiate an infection, the new host would be infected by a single virus genotype. In these cases the HearNPV population may remain genotypically homogenous over the course of the infection, or novel genotypic variants may be generated as a consequence of mutations arising from DNA polymerase errors over successive replication cycles. The present study aimed to further evaluate the effect of inoculum dose on the genetic diversity of post-passage HearNPV populations, over a wider range of doses than previous studies. A PCR-DGGE assay and deep amplicon sequencing were used to assess the genetic diversity in a region of the *me53* gene of HearNPV populations isolated from larvae inoculated with a range of ODV doses ($LD_5$, $LD_{25}$, $LD_{50}$, $LD_{75}$ or $LD_{95}$). It was thought that the use of ODVs rather than OBs as the inoculum would increase the probability that at the lowest dose some larvae would be infected by a single ODV, allowing the effect of single ODV infections on HearNPV genetic diversity to be assessed.

## 5.3. Materials and methods

### 5.3.1. Insect rearing

*H. armigera* eggs originally obtained from the Agricultural Research Council (Pretoria, South Africa) were used to set up a *H. armigera* laboratory population. The laboratory population was maintained as described in Chapter 2, section 2.3.1 (p. 32).

### 5.3.2. Virus propagation and purification

HearNPV was propagated *in vivo* in third instar (± 5 d old) *H. armigera* larvae. Third instar larvae were inoculated with a lethal HearNPV dose using the diet contamination method (Evans & Shapiro, 1997) and cadavers were harvested prior to liquification and stored at −20 ˚C. HearNPV OBs were purified from the cadavers using a modified version of Crook & Payne's (1980) protocol as described in Chapter 2, section 2.3.2 (p. 32). The final OB suspension was stored at 4 ˚C prior to the preparation of ODVs.

### 5.3.3. Preparation of ODVs

ODVs were prepared using a modified version of Harrap & Longworth's (1974) protocol for the alkaline dissolution of OBs; the key difference being the use of dialysis during the OB dissolution steps. The sucrose gradient purified OB sample was dialysed against a 0.05 M $Na_2CO_3$ solution at 37 ˚C for 2 h. Dialysis was conducted in cassettes constructed from the cut-off lids of 1.5 ml micro-centrifuge tubes with a single layer of SnakeSkin®, 10 K MWCO, dialysis tubing (Thermo Scientific) as the dialysis membrane. The ODV preparations were then dialysed against the storage solution ($Na_2CO_3$, 0.005 M) three times, for 5 h each at 4 ˚C. Harrap & Longworth (1974) found that this storage solution minimised the aggregation of ODVs when samples were stored at 4 ˚C. The ODV preparations were centrifuged in a bench top centrifuge (13 000 *g* for 3 min) in order to pellet any OBs which may have been incompletely dissolved during the alkaline dissolution process. The ODV containing supernatant was stored at 4 ˚C.

In order to confirm the absence of OBs in the ODV preparation, ODV samples were examined using phase contrast microscopy. In addition the structural integrity of the prepared ODVs was assessed using transmission electron microscopy (TEM). The TEM also served to

confirm the absence of OBs in the ODV preparations. For TEM, micro-droplets (< 0.5 µl) of the ODV suspension were applied to formvar coated, 200 mesh, copper TEM grids, which were stabilised with carbon (Ted Pella Inc.). The micro droplets were allowed to completely air-dry before being stained for 2 min with a uranyl acetate solution (2% w/v in 30% v/v ethanol). The stained grids were then rinsed with sterile distilled water for 10 s before being allowed to completely air-dry. The prepared grids were stored in a desiccator prior to being viewed using a FEI Spirit 120 kV TEM.

### 5.3.4. Inoculation of neonate larvae

Groups of neonate larvae (≤ 12 h old) were inoculated using the droplet feeding method (Bouwer & Avyidi, 2006; Hughes & Wood, 1981) with a mixture containing 0.5 mg/ml erioglaucine, a non-toxic dye, and the ODV preparation diluted with sterile distilled water. Each group of larvae was inoculated with a different dilution of the ODV preparation. A group of 100 larvae was inoculated for each ODV dilution where the expected mortality was ≥ 50%, while a group of 250 larvae was inoculated for each of the higher dilutions. Inoculated larvae were checked after 24 h and any dead larvae discarded. The remaining larvae were checked twice a day and mortality was scored five days post inoculation. Dead larvae were harvested, prior to the liquification of the cadavers, and stored at −20 ˚C in 200 µl PCR tubes. The genetic diversity of HearNPV populations isolated from larvae inoculated with the ODV doses which resulted in approximately 5%, 25%, 50%, 75% or 95% host mortality was assessed using the PCR-DGGE assay and deep-amplicon sequencing. Table 5.1 shows these five empirically determined ODV lethal doses and the corresponding dilutions of the neat ODV preparation, as well as the number of larvae inoculated for each dose.

### 5.3.5. DNA extraction

HearNPV DNA was extracted from the individual homogenates of twelve larvae from each dose group, as well as from the ODV inoculum sample. Larvae cadavers were homogenised in 57 µl of sterile distilled water with a micropestle made from a 100 µl pipette tip where the tip aperture had been melted closed. $Na_2CO_3$ was added to a final concentration of 50 mM and the homogenate was incubated at 37 ˚C for 30 min. Thereafter 20 µl of the neutralisation buffer (1M Tris-HCl, pH 7.0) and 8 µl of a 10% (w/v) SDS solution were added and the

homogenate was incubated at 37˚C for a further 30 min. Viral DNA was then extracted from the homogenate using the NucleoSpin® Tissue XS kit (Macherey-Nagel) as per the manual's instructions for the purification of DNA from cultured cells. A NanoDrop™ ND-1000 spectrophotometer was used to determine the concentration of the purified viral DNA, which was then stored at −20 ˚C.

### 5.3.6. PCR-DGGE

A PCR-DGGE assay, developed by Baillie & Bouwer (2011), was used to assess the variant composition and genetic diversity of HearNPV populations for a region of the *me53* gene. PCR amplification and DGGE analysis of the *me53* gene region was conducted as described in Chapter 2, section 2.3.4 (p. 33) and section 2.3.5 (p. 34), respectively. DGGE profiles were analysed using the GelCompar II V6.0 software (Applied Maths) as described in Chapter 2, section 2.3.6 (p. 35). PERMANOVA and PERMDIST tests were conducted as described in Chapter 3, section 3.3.6 (p. 69) to determine if the DGGE profiles from the different inoculum groups were statistically significantly different.

### 5.3.7. Deep amplicon sequencing

Deep amplicon sequencing on the MiSeq (Illumina) HTS platform was used to assess the number and identity of *me53* sequence variants in a subset of HearNPV populations. Due to the use of neonate larvae, it was generally only possible to isolate a limited amount of viral DNA from larvae cadavers. Consequently, deep amplicon sequencing could only be used to assess the genetic diversity of HearNPV populations where there was sufficient DNA remaining following PCR-DGGE analysis. The sample preparation and sequencing steps were conducted as described in Chapter 2, section 2.3.11 (p. 38). The raw deep sequencing data was processed and analysed as described in Chapter 2, section 2.3.12. (p. 39). Samples which had <10 000 merged paired-end sequencing reads after the read processing and filtering steps were discarded. The relative abundances and Levenshtein distances for the sequence variants were plotted using the ggplot2 package (Wickham, 2009) in R version 3.3.1 (R Core Team, 2016), as described in Chapter 3, section 3.3.7 (p. 70).

## 5.4. Results

### 5.4.1. Structural integrity and infectivity of ODVs

Particles resembling ODVs were observed, using TEM, in the post alkaline-dissolution HearNPV sample. These particles, examples of which can be seen in Figure 5.1, had the characteristic rod shape of ODVs and were approximately 350 nm in length which is within the size range previously reported for ODVs of a number of baculoviruses (Ackermann & Smirnoff, 1983). No structures resembling OBs or partially dissolved OBs were observed using either phase-contrast light microscopy or TEM. This indicated that the alkaline-dissolution process was effective at dissolving OBs, and that any OBs remaining in the HearNPV sample were removed by the subsequent centrifugation step.

The majority of the observed ODVs appeared to have fully intact envelopes surrounding their nucleocapsids. This indicated that the alkaline-dissolution process did not impact on the structural integrity of the ODVs. The viral envelope and a group of envelope proteins called the *per os* infectivity factors (PIFs) are essential for facilitating the entry of the nucleocapsid into the cytosol of a midgut epithelial cell (Granados & Lawler, 1981; Peng *et al.*, 2012; Song *et al.*, 2008). The observed structural integrity of the viral envelopes therefore suggested that ODVs prepared using the dialysis version of the *in vitro* alkaline-dissolution process would be able to infect larvae *per os*. This was confirmed upon conducting a bioassay where 100% mortality was recorded for neonate larvae inoculated using the droplet feeding method with an undiluted ODV preparation. After confirming that infectious ODVs were successfully prepared using the modified alkaline-dissolution process, groups of neonate *H. armigera* larvae could be inoculated with a range of ODV doses.

### 5.4.2. DGGE profile analysis

The cluster analysis and a graphical representation of the banding patterns of the DGGE profiles are shown in Figure 5.2a and 5.2b, respectively. The cosine correlation coefficient which was used to prepare the similarity matrix, from which the dendrogram in Figure 5.2a was constructed through UPGMA cluster analysis, takes into account both the positions and relative intensities of bands when calculating the degree of similarity between two DGGE profiles. It can be seen in Figure 5.2a that the DGGE profiles clustered in two distinct major

clades (M1 and M2) which had 80.8% mean similarity. The M1 clade was predominantly made up of profiles for populations isolated from larvae inoculated with one of the three highest doses, while the M2 clade consisted largely of profiles for the two lowest doses. Clade M1 contained the ODV inoculum DGGE profile along with 75% of the $LD_{95}$, 67% of the $LD_{75}$, 44% of the $LD_{50}$, 10% of the $LD_{25}$ and 11% of the $LD_5$ DGGE profiles (profiles in the M1 clade had 85.4% mean similarity). The M2 clade (M2 profiles had 92.8% mean similarity) contained 89% of the $LD_5$, 90% of the $LD_{25}$, 56% of the $LD_{50}$, 33% of the $LD_{75}$ and 25% of the $LD_{95}$ DGGE profiles.

The DGGE profiles in the M1 clade generally contained more bands than the profiles in the M2 clade, which suggested that the *me53* variant richness was greater in the M1 populations (Figure 5.2b). Bands at positions mB1, mB2 and mB7 have previously been identified (Chapter 2, section 2.4.3) as true bands associated with five different *me53* clone sequence variants isolated from a HearNPV population, while bands at positions mB3-mB6 were identified as artefactual heteroduplex bands. Most M1 profiles contained bands at the mB1 and mB2 positions as well as bands at the mB3 and mB4 positions, while some M1 profiles also had additional bands at the mB5, mB6 and MB7 positions (Figure 5.2b and Figure 5.3a). Profiles in the M2 clade typically only had a single dominant band at either the mB1 or mB2 position, regardless of the ODV dose group to which they belonged (Figure 5.2b and Figure 5.3a). These M2 profiles closely resembled the profiles previously observed for the mV1, mV2, mV4 and mV5 *me53* clone variants (Chapter 2, Figure 2.5 ), which contained a single dominant band at either the mB1 or mB2 position and a number of low intensity “ghost” bands (Figure 5.3a).

A notable feature of the M2 clade was the sub-clade that contained eight $LD_5$ and three $LD_{25}$ profiles which had a high degree of similarity (99% mean similarity). These $LD_5$ and $LD_{25}$ profiles all had a single dominant band at the mB1 position. The only $LD_5$ profile which did not fall within this M2 sub-clade, but rather in the M1 clade, contained both the mB1 and mB2 true bands and the mB3 and mB4 heteroduplex bands. Pairwise PERMANOVA tests also showed that there were statistically significant differences ($\alpha = 0.05$) between profiles of the $LD_5$ group and profiles of all the other ODV dose groups (Table 5.2). The PERMDIST test showed that there were no statistically significant differences ($\alpha = 0.05$) in the amount of variation between DGGE profiles within each of the ODV dose groups (Table 5.2). The

presence of only a single dominant band at the mB1 position in the $LD_5$ and $LD_{25}$ profiles of M2 sub-clade suggests that at very low ODV doses *H. armigera* neonate larvae may only be infected by a single *me53* variant.

### 5.4.3. Deep amplicon sequencing

The *me53* variant richness and nucleotide sequence diversity of the ODV inoculum population and a subset of post-passage HearNPV populations from the five ODV dose groups were assessed using deep amplicon sequencing on the MiSeq platform. Due to the small size of *H. armigera* neonate cadavers the extraction of HearNPV DNA from the larvae homogenates resulted in very low DNA yields. Consequently, all HearNPV populations were sequenced where there was sufficient viral DNA remaining for the deep sequencing protocol following the initial PCR-DGGE analysis. Six *me53* sequence variants, mV1-mV6, were detected in the ODV inoculum population (Figure 5.3b and Figure 5.4). HearNPV populations isolated from larvae that had been inoculated with a higher ODV dose ($LD_{50}$-$LD_{95}$) generally appeared to contain a greater number of the six inoculum variants than populations isolated from larvae inoculated with the lowest ODV dose ($LD_5$). The $LD_{95}$ (1) population contained five of the six inoculum variants (variants mV1-mV5) and was the only population sequenced where more than two of the inoculum variants were detected (Figure 5.3b). The variant composition of the $LD_{95}$ (1) population closely resembled that of the inoculum population, and variant mV6, which accounted for the least number of sequencing reads in the inoculum population, was the only variant which was not detected in this population. Conversely, in the two $LD_5$ populations and the $LD_{50}$ (1) population only a single *me53* sequence variant was detected, while the $LD_{50}$ (2), $LD_{75}$ and $LD_{95}$ (2) populations all contained the two variants, mV1 and mV2 (Figure 5.3b). The mV1 variant was the most abundant variant in the inoculum population and following the *in vivo* passage remained the most abundant variant in six of the seven sequenced populations (Figure 5.3b). Population $LD_5$ (1) differed from the other six sequenced populations as it did not contain mV1, and variant mV4 was the only variant detected in this population (Figure 5.3b). No novel *me53* sequence variants were detected in any of the sequenced HearNPV populations isolated from ODV inoculated neonate larvae, as all six *me53* variants were detected in the HearNPV inoculum population (Figure 5.3 and Figure 5.4). The deep sequencing results suggest that the higher the ODV inoculum dose the greater the probability that multiple variants within the inoculum population will be successfully transmitted to a new host. However, even at

$LD_{95,}$ the highest inoculum dose, there was no guarantee that the full variant complement of the inoculum population would be successfully transmitted to a new host.

## 5.5. Discussion

In a viral infection the number of horizontally transmitted virions is thought to be a key determinant of viral genetic diversity, as it affects the probability that a host will be co-infected by multiple genotypes (Zwart *et al.*, 2009). The number of co-infecting genotypes affects the potential for recombination, competition and complementation to occur between genotypes during an infection (Zwart *et al.*, 2011). It has been documented, in both *in vitro* and *in vivo* systems, that the occlusion of multiple genotypes within a single OB is possible for a number of NPV species, including HearNPV (Arrizubieta *et al.*, 2015; Bull *et al.*, 2001; Clavijo *et al.*, 2010; Hamblin *et al.*, 1990). If genotypes are co-occluded within a single OB there is a higher probability that even at low OB inoculum doses larvae will be exposed to and subsequently co-infected by more than one genotype (Bernal *et al.*, 2013; Bull *et al.*, 2001; Clavijo *et al.*, 2010). A key aspect of the present study was the use of ODVs as the inoculum, instead of OBs. At low ODV doses larvae may be exposed to and subsequently infected by a less diverse pool of virions, consisting largely of the dominant genotypes in the inoculum population (Zwart *et al.*, 2009, 2011). The use of ODVs as the inoculum allowed us to evaluate the post-passage genetic diversity and detect any novel sequence variants in HearNPV populations over a broad range of inoculum doses.

The cluster analysis of DGGE profiles indicated that the *me53* variant composition of HearNPV populations isolated from larvae inoculated with one of the higher doses generally had greater similarity to that of the inoculum population than populations isolated from larvae inoculated with one of the lower ODV doses. This was clearly shown by the clustering of most $LD_{95}$ and $LD_{75}$ profiles, the two highest doses, into the M1 clade along with the inoculum population profile, while the majority of the $LD_{25}$ and $LD_5$ profiles, the two lowest doses, clustered in the M2 clade. Interestingly the profiles for virus populations isolated from larvae inoculated with a $LD_{50}$, the middle dose, were almost equally distributed between the M1 and M2 clades, with a 4:5 profile split. The number of bands in profiles suggested that there was greater *me53* variant richness in the HearNPV populations which fell in the M1 clade. All M1 profiles contained bands at positions previously shown (Chapter 2) to be associated with artefactual heteroduplexes (mB3-mB6) as well as at three true band positions associated with cloned *me53* sequence variants (mB1, mB2 and mB7). However, the majority of the $LD_5$ and $LD_{25}$ profiles in the M2 clade closely resembled the profiles in Chapter 2 of

*me53* clone variants, as they only contained a single dominant band at either the mB1 or mB2 position and did not contain any bands at the known heteroduplex positions. Artefactual heteroduplex bands reliably form when there are more than two *me53* genetic variants in a population, which suggests that populations that fell in the M2 clade contained at least two genetic variants while most of the $LD_5$ and $LD_{25}$ populations in the M2 clade contained a single genetic variant.

Subsequent deep amplicon sequencing analysis supported this interpretation of the band composition of DGGE profiles in the M1 and M2 clades. Deep sequencing detected a single *me53* variant in the two $LD_5$ populations and one $LD_{50}$ population which fell in the M2 clade and whose profiles contained a single dominant band at the mB1 position, while two or more *me53* variants were detected in the $LD_{50}$, $LD_{75}$ and $LD_{95}$ populations, whose profiles contained bands at both the true and heteroduplex positions. Together, the cluster analysis of the DGGE profiles and the deep amplicon sequencing data support one of the findings of the study by Baillie & Bouwer (2013); that the level of genetic diversity and the variant composition of a HearNPV population is largely maintained at higher inoculum doses over an *in vivo* passage. This suggests that with increases in the ODV inoculum dose, larvae were exposed to and subsequently infected by a greater proportion of the genotypes in the HearNPV inoculum population. The sequencing data also showed that the most abundant inoculum variant (mV1) appeared to have the highest probability of infecting a new host, as this variant was the most abundant variant in six of the seven sequenced HearNPV populations. These findings suggest that ODVs in the midgut infect a *H. armigera* larva in a frequency-dependent manner, where genotypic variants that have the greatest abundance in the ODV pool have the highest probability of infecting the new host.

The evidence from the current study showed that there was a high probability that populations isolated from hosts inoculated with a very low ODV dose would only contain a single *me53* variant. This finding suggests that synergism between virus genotypes is not required for HearNPV to cause a lethal infection in a *H. armigera* larva. However it should be noted that mixed genotype infections may still be beneficial to HearNPV populations, as has been suggested by previous studies (Arrizubieta *et al.*, 2015; Bernal *et al.*, 2013; Hodgson *et al.*, 2004). This finding also appears to support one of the main tenets of the independent action hypothesis (IAH) for the HearNPV-host interaction, that the average number of

horizontally transmitted virions is dose dependent (Lafforgue *et al.*, 2012; Zwart *et al.*, 2011). The IAH states that all virions have a non-zero probability of infecting a host and that virions act independently during the infection process (Druett, 1952; Lafforgue *et al.*, 2012; Zwart *et al.*, 2009, 2011). Under the IAH a single virion is capable of causing infection in a host. A previous study by Zwart *et al.* (2009) showed the first experimental evidence of infections caused by a single ODV, however it is not clear whether the IAH applies to all NPV infections, as the study showed that the IAH was only supported in an AcMNPV model system where hosts were highly susceptible to AcMNPV infection. The populations where only a single *me53* variant was detected may represent infections caused by a single ODV, although there is the possibility that the same results would have been observed had these larvae been infected by multiple ODVs of the same genotype. However, whereas the probability that multiple ODVs infected a host would be higher for the $LD_5$ (2) and $LD_{50}$ (1) populations, where only the most abundant inoculum variant (mV1) was detected, this scenario was considered less likely for the $LD_5$ (1) population where the only variant detected (mV4) was present at a low frequency in the inoculum population. This suggests that in some instances a single HearNPV ODV can successfully infect and cause host mortality in neonate *H. armigera* larvae, although further studies are needed to show this conclusively.

In the Baillie & Bouwer (2013) study it was found that the level of genetic diversity actually increased above the level observed in the inoculum population at their lowest OB inoculum dose ($LD_5$). This appeared to be due both to the maintenance of genotypes present in the inoculum population and the generation of novel sequence variants during a single *in vivo* passage. This was not the case in the present study where in most instances there was a decline in the level of HearNPV genetic diversity at the two lower ODV doses ($LD_5$ and $LD_{25}$). In addition there was no evidence of novel sequence variants being generated during infections at any of the ODV inoculum doses. Deep sequencing did not detect novel sequence variants in any of the sequenced populations and although some DGGE bands appeared at first to be "*de novo*" bands, on closer inspection these bands were found to be present in the inoculum profile, albeit below the software detection threshold. Baillie & Bouwer (2013) suggested that the novel genetic variants they detected were generated either by DNA polymerase errors or recombination between genotypes occurring during the replication of virus genomes. Although there was no evidence of novel sequence variants being generated over a single *in vivo* passage in the present study, the results do not necessarily conflict with

those of Baillie & Bouwer (2013) if the co-occlusion hypothesis is considered. Clavijo *et al.* (2010) showed that co-occlusion is likely to form part of a natural adaptive mechanism by NPVs which favours the maintenance of genetic diversity in NPV populations during insect-to-insect transmission. One would expect SNPs to be generated at a similar rate in both low dose ODV and OB infections, as this is dependent on the error rate of the viral DNA polymerase (Hang & Guarino, 1999; Huang & Levin, 2001). However the degree of recombination between genotypes may be different for the two, given that the diversity of transmitted genotypes may be higher in a low dose OB infection if genotypes are co-occluded (Bull *et al.*, 2001; Clavijo *et al.*, 2010).

For recombination between genotypes to routinely occur, the average number of viral genomes infecting host cells, also known as the multiplicity of infection (MOI), must be greater than one (Bull *et al.*, 2001). In low dose ODV infections, where a limited number of genotypes are expected to initiate infection, there is a low probability that viral genomes co-infecting a single host cell would be heterogeneous. This suggests that homologous recombination between virus genotypes may have been largely responsible for the generation of new sequence variants in the $LD_5$ OB infections in the Baillie & Bouwer (2013) study, and could explain why these were absent from the low ODV dose infections. As DNA replication has been implicated in creating opportunities for homologous recombination to occur between genotypes in co-infected cells (Kamita *et al.*, 2003; Martin & Weber, 1997), this theory does not conflict with Baillie & Bouwer's (2013) postulate that the higher number of replication cycles in larvae prior to death in low dose OB infections would favour both the generation and amplification of novel sequence variants to detectable levels.

## 5.6. Conclusions

There were two major findings in this study: (1) the higher the ODV inoculum dose the greater the propensity for the genetic variant composition of the post-passage population to resemble that of the inoculum population, and (2) at low ODV doses a single HearNPV genotype appears capable of causing a lethal infection in *H. armigera* larva, suggesting that infection by a single ODV occurred. It was also found that across the different ODV doses the most abundant variant in the inoculum population appeared to have the highest probability of infecting a new host. This suggests that primary infection in the host midgut

may occur in a frequency-dependent manner where the most abundant genotypes in the ODV pool have the highest probability of infecting the new host.

**Table 5.1.** The empirically determined ODV lethal doses and the corresponding dilutions of the neat ODV preparation.

| Dilution | No. of larvae inoculated | Empirical LD[a] |
|---|---|---|
| $10^{-1}$ | 100 | $LD_{96}$ ($LD_{95}$) |
| $10^{-2}$ | 100 | $LD_{73}$ ($LD_{75}$) |
| $20^{-3}$ | 100 | $LD_{50}$ ($LD_{50}$) |
| $10^{-4}$ | 250 | $LD_{25}$ ($LD_{25}$) |
| $50^{-4}$ | 250 | $LD_5$ ($LD_5$) |

[a] The lethal dose rounded to the nearest multiple of 5 is given in brackets.

**Table 5.2.** The results of the PERMANOVA and PERMDIST analyses of *me53* DGGE profiles from the different ODV inoculum groups ($LD_5$-$LD_{95}$). Both analyses were implemented using the cosine dissimilarity matrices for the DGGE profiles and *P* values were calculated following 9999 permutations according to the methods of Anderson (2001, 2006).

| Analysis | PERMANOVA | | PERMDIST | |
|---|---|---|---|---|
| | ***F*** | ***P*** | ***F*** | ***P*** |
| *me53* | $F_{4, 43} = 5.05$ | < 0.001 | $F_{4, 43} = 1.05$ | > 0.05 |
| Pairwise* | $LD_5$ $\underline{LD_{25}\ LD_{50}\ LD_{75}\ LD_{95}}$ | | | |

* Where there was a statistically significant overall *F*-ratio ($P < 0.05$, 9999 permutations) for DGGE profiles in the different passage groups pairwise comparisons were done. The alpha value for these pairwise tests was corrected for multiple comparisons using the Holm-Bonferroni correction (Holm, 1979). The five ODV inoculum groups are shown, and the underlining bars indicate groups which were not statistically significantly different ($P > 0.05$).

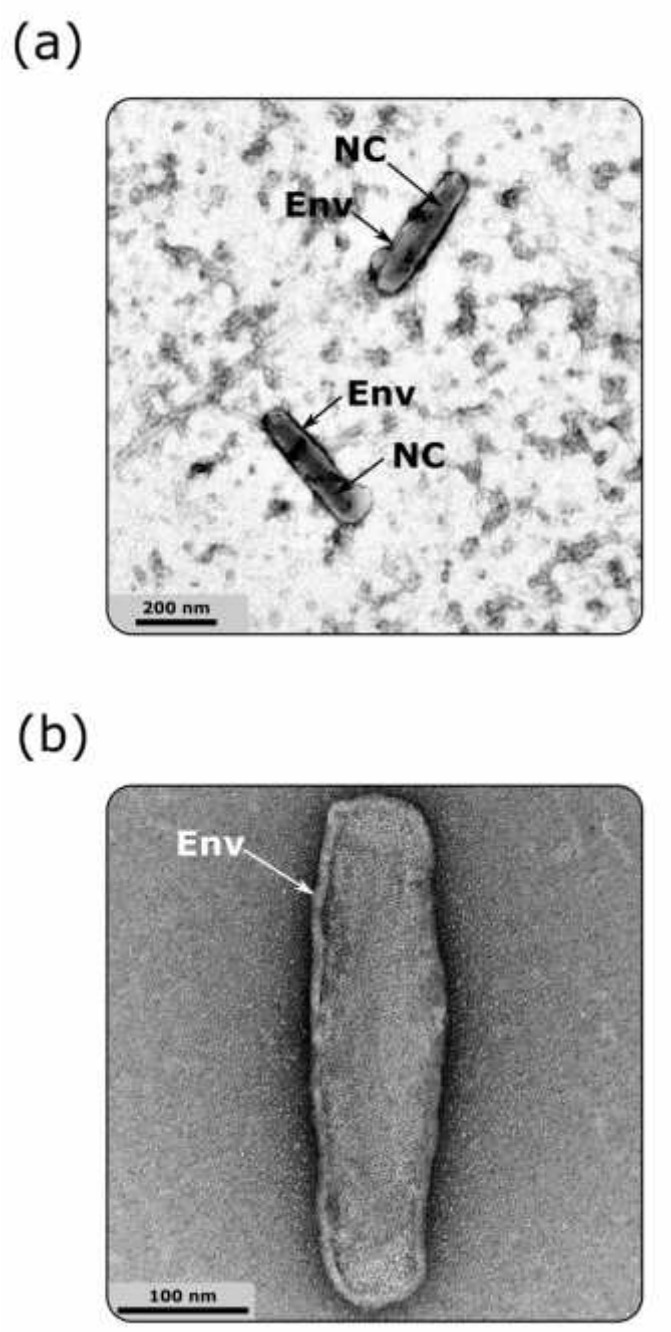



**Figure 5.1.** TEM micrographs showing ODVs in the HearNPV sample after the alkaline-dissolution process. The ODVs shown in panels **(a)** and **(b)** are representative of the majority of ODVs viewed, with good structural integrity and intact envelopes (Env) surrounding their nucleocapsids (NC).

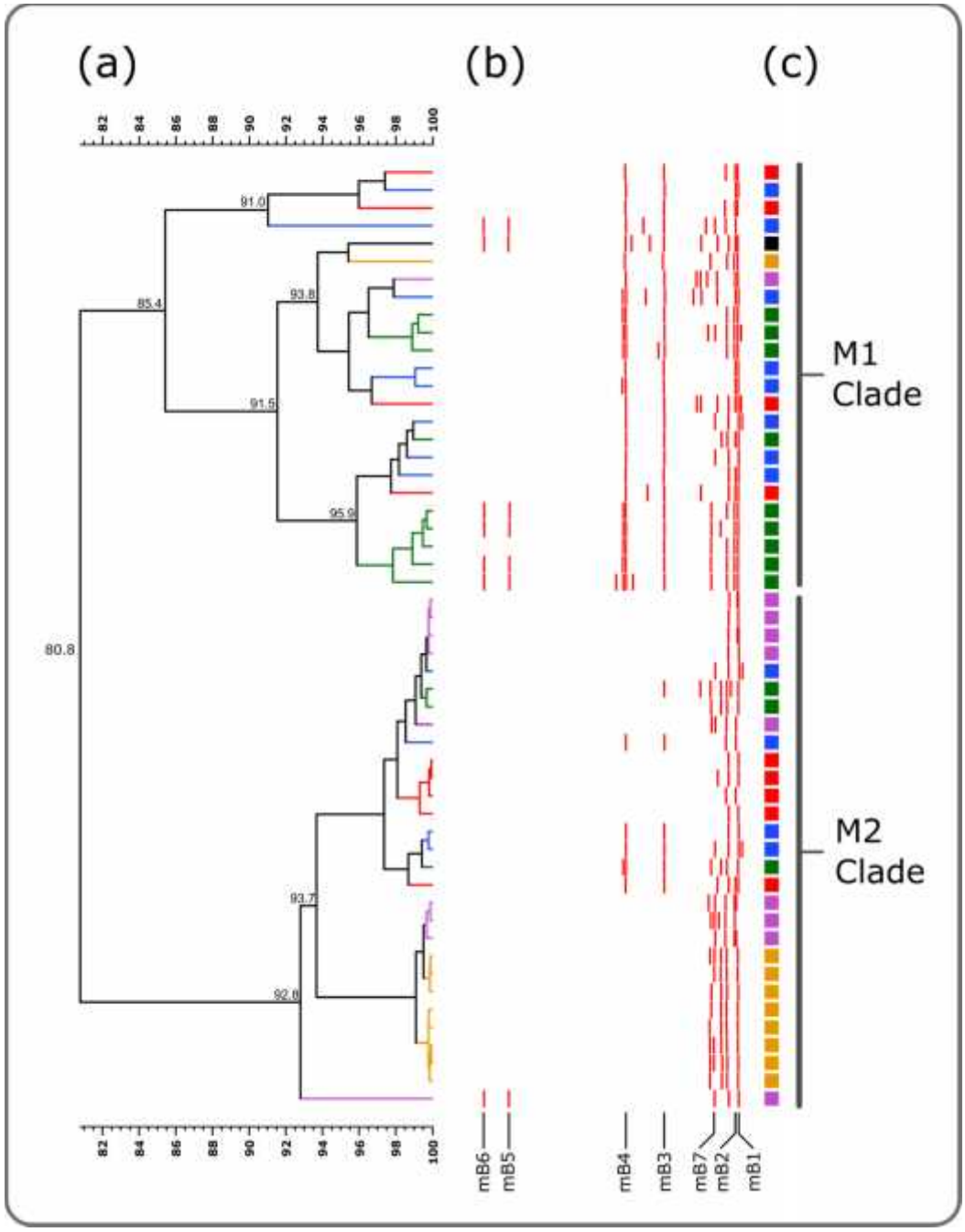



**Figure 5.2.** Cluster analysis of the *me53* DGGE profiles for the HearNPV inoculum population and populations isolated from *H. armigera* neonate larvae inoculated with one of the five HearNPV ODV doses ($LD_5$, $LD_{25}$, $LD_{50}$, $LD_{75}$ and $LD_{95}$). **(a)** A dendrogram showing percentage similarity between DGGE profiles. **(b)** A graphical representation of the relative positions of software-detected bands in each DGGE profile. Positions associated with previously identified true bands (mB1, mB2 and mB7) and heteroduplex bands (mB3-mB6) are labelled. **(c)** The HearNPV isolate from which DNA was extracted for each PCR-DGGE analysis. ■ indicates the DGGE profile for the HearNPV inoculum, ■ indicates isolates from $LD_5$-inoculated larvae, ■ indicates isolates from $LD_{25}$-inoculated larvae, ■ indicates isolates from $LD_{50}$-inoculated larvae, ■ indicates isolates from $LD_{75}$-inoculated larvae and ■ indicates isolates from $LD_{95}$-inoculated larvae.

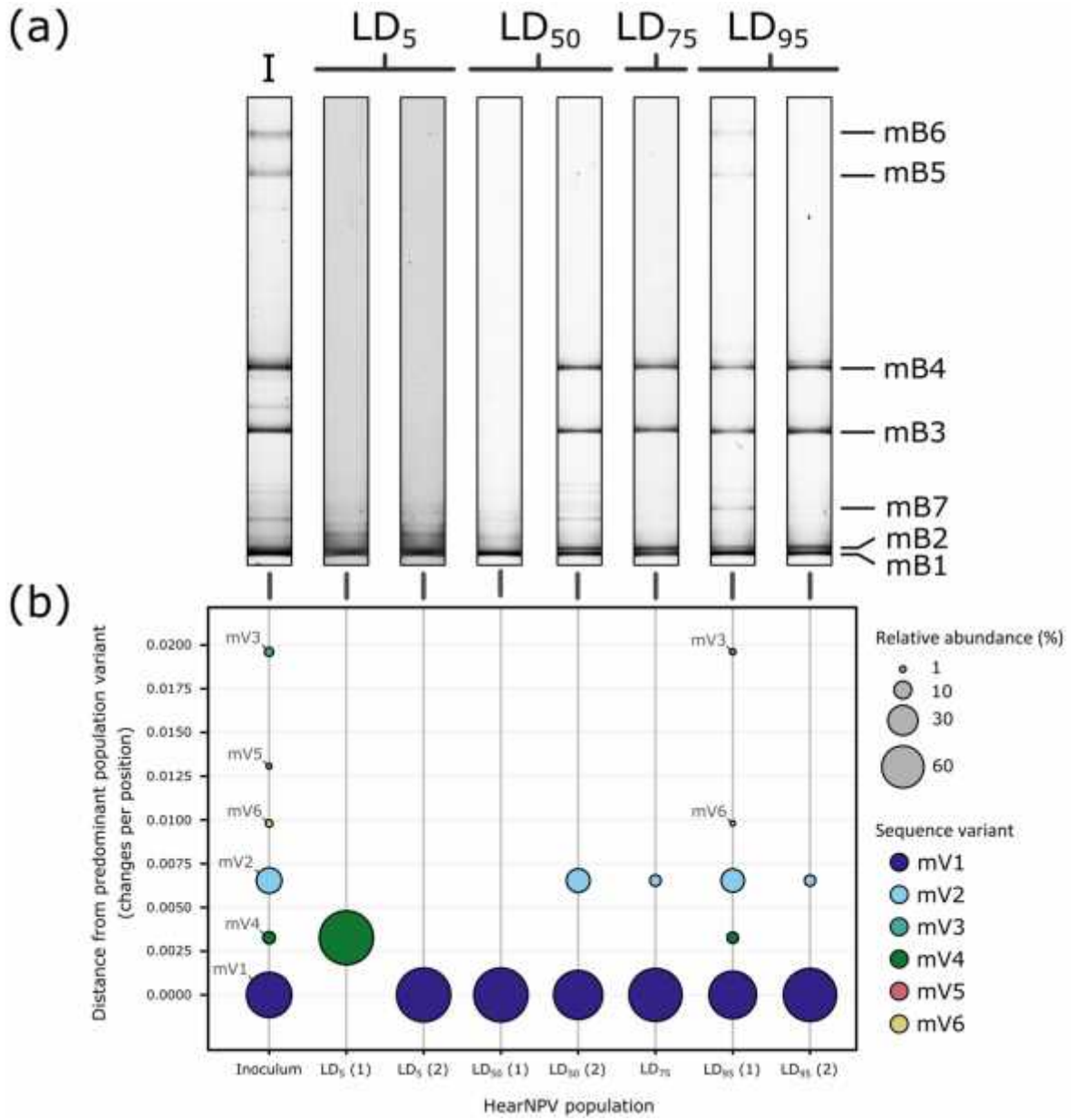



**Figure 5.3.** The *me53* variants detected using deep amplicon sequencing in a subset of HearNPV populations isolated from larvae inoculated with the different ODV doses. Populations sequenced included the original ODV inoculum, two $LD_5$, two $LD_{50}$, one $LD_{75}$ and two $LD_{95}$ populations. **(a)** DGGE profiles for each of the sequenced HearNPV populations. Positions associated with previously identified true bands (mB1, mB2 and mB7) and heteroduplex bands (mB3-mB6) are labelled. **(b)** The *me53* variants detected in each HearNPV population. Each bubble in a column represents a *me53* variant, while the size of the bubble shows the relative abundance of the sequence variant in the processed sequencing reads. The position of each bubble on the y-axis indicates the sequence distance of each variant as changes per position from the predominant sequence variant in the inoculum population (mV1). For clarity, the position of each variant is labelled in the inoculum population and in cases where the size of the bubble makes its identity difficult to discern.

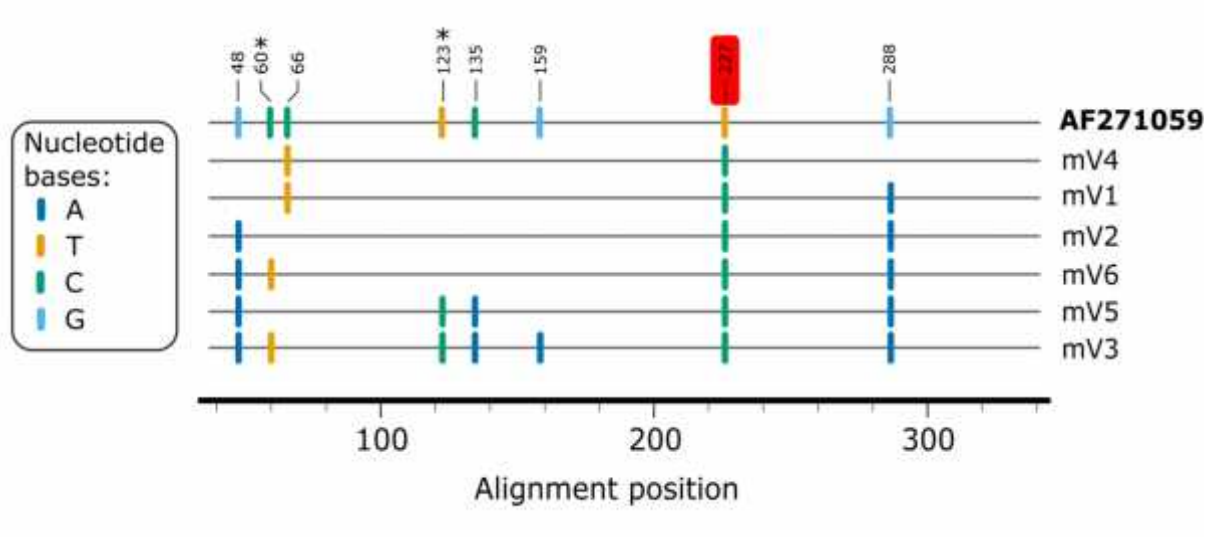



**Figure 5.4.** Highlighter plot showing the nucleotide sequences of the *me53* variants relative to a reference HearNPV *me53* gene sequence (GenBank accession number: AF271059). The positions of SNPs relative to the reference *me53* gene sequence are shown as coloured bars, with each colour denoting one of the four DNA bases. SNPs at positions marked with an asterisk (*) were not present in any of the previously published HearNPV genomes (Baillie & Bouwer, 2012a; Chen *et al.*, 2001; Noune & Hauxwell, 2015, 2016), while those highlighted in red are non-synonymous polymorphisms.

# Chapter 6

# General discussion and conclusion

## 6.1. General discussion

It is well documented that considerable genetic diversity exists within baculovirus populations in the field (Baillie & Bouwer, 2012a; Cory *et al.*, 2005; Gettig & Mccarthy, 1982; Patel *et al.*, 2009b). A number of studies have focused on evaluating and comparing the level of genetic diversity of baculovirus populations in different host species, different host populations and from different geographical areas (Baillie & Bouwer, 2012b; Cooper *et al.*, 2003; Cory *et al.*, 2005; Graham *et al.*, 2004; Shapiro *et al.*, 1991). These studies have shown that baculovirus populations consist of multiple genotypes and that the genotypic composition of populations can vary between different host individuals and geographical locations (Baillie & Bouwer, 2012b; Cooper *et al.*, 2003; Cory *et al.*, 2005; Shapiro *et al.*, 1991). In recent years, whole genome deep sequencing studies have shown that a significant number of synonymous and non-synonymous SNPs occur across NPV genomes in all functional gene classes and together with large insertions and deletions contribute to the considerable genetic diversity observed in NPV populations (Baillie & Bouwer, 2012a; Chateigner *et al.*, 2015; Thézé *et al.*, 2014; Xu *et al.*, 2013). The maintenance of genetic diversity within populations is thought to be beneficial to baculoviruses as it could in theory provide a source of selectable variation which would allow baculovirus populations to adapt to changes in their environment or host population (Harrison, 2009). However it is still unclear whether the genotypic composition of populations is largely maintained over successive baculovirus generations or whether it constantly changes with the generation of new genetic variants and/or shifts in the relative abundance of existing variants in the population.

Previous studies have made extensive use of RFLP analyses to discriminate between different baculovirus populations on the basis of their genotypic composition (Hodgson *et al.*, 2001; Lee & Miller, 1978; Redman *et al.*, 2010), while Sanger-dideoxy sequencing has been used to

characterise the nucleotide sequence diversity of baculovirus isolates (Ferreira *et al.*, 2014; Harrison *et al.*, 2014; Rowley *et al.*, 2010). However, both techniques suffer from limitations that make them unsuitable for assessing the genetic variant composition of a large number of baculovirus populations. RFLP analysis has a relatively low sensitivity and is only able to discriminate between genotypes if nucleotide sequence variations alter the number or position of discrete endonuclease recognition sites (Busch & Nitschko, 1999; Olive & Bean, 1999), while Sanger-dideoxy sequencing would require a time and reagent intensive process to sequence the full variant complement of a virus population (Gregori *et al.*, 2014; Mardis, 2013).

The use and reliability of an alternative molecular technique, PCR-DGGE, to assess the genetic diversity of HearNPV populations was evaluated in the current study (addressed in Chapter 2). It was found that PCR-DGGE generated reproducible genetic fingerprints for the targeted *DNA polymerase* and *me53* gene regions, which could be used to discriminate between HearNPV populations that differed in their underlying genetic variant composition. However, the detection of artefactual heteroduplex bands and the co-migration of sequence variants in *DNA polymerase* and *me53* profiles made PCR-DGGE an unsuitable method for the calculation of quantitative measures of genetic diversity in HearNPV populations. These findings agreed with those of previous studies where it has been found that bands in a DGGE profile do not always represent distinct genetic variants (Ferris & Ward, 1997; Hong *et al.*, 2007; Neilson *et al.*, 2013; Samarajeewa *et al.*, 2015; Satokari *et al.*, 2001). In spite of this limitation, the number of bands in DGGE profiles provided a relative gauge of the amount of genetic diversity at the targeted genetic regions in HearNPV populations, while the relative similarity in variant composition of HearNPV populations could be assessed by comparing the densitometric profiles of their DGGE profiles. Furthermore, deep amplicon sequencing on the MiSeq platform proved to be a more time efficient method of assessing the underlying nucleotide sequence diversity in HearNPV populations and enabled greater detection of low frequency genetic variants, relative to the traditional method of sequencing multiple cloned DNA fragments isolated from DGGE bands.

An integrated analysis pipeline, combining the use of both PCR-DGGE and deep amplicon sequencing, was used to assess the effect of serial *in vivo* passages on the genetic diversity of two core genes, *DNA polymerase* and *me53*, in HearNPV populations (addressed in

Chapter 3). It has been suggested that host selection may lead to NPV populations that consist of genotypic variants that are best adapted to the local host population (Arrizubieta *et al.*, 2015; Hitchman *et al.*, 2007). Studies by Hitchman *et al.* (2007) and Weitzman *et al.* (1992) have shown that different baculovirus genotypes emerge as the dominant genotypes in a population after several *in vivo* passages in different host species, while Simón *et al.* (2008) and Clavijo *et al.* (2009) have shown that the genotypic composition of NPV populations appear to be maintained to maximise survival and transmission within a host population. However, in the current study it was found that there were multiple outcomes for the *DNA polymerase* and *me53* variant composition of HearNPV populations following an *in vivo* passage in a host individual. These included: (i) the maintenance of genetic variants present in the inoculum population, (ii) "drop-off" events, where a significant number of inoculum variants were lost over a single *in vivo* passage, and (iii) the generation of novel genetic variants over the course of an infection. This led to a lack of convergence in the *DNA polymerase* and *me53* variant composition of HearNPV populations over the three successive *in vivo* passages and meant a stable, consensus variant composition was not reached for these two gene regions across the 125 passage lineages following the third passage.

Across the 125 passage lineages only one novel, low frequency *me53* genetic variant (mV10) contained a non-synonymous SNP, while the remaining ten sequence variants detected were synonymous and encoded the same amino acid sequence. This indicates that natural selection acting on the assayed *me53* region was not responsible for the changes in *me53* variant composition observed over the three passages. Instead, the divergence observed in *me53* variant composition over the passages may have resulted from natural selection acting on genomic regions outside of the assayed sequence region. This is a reasonable assumption given that several studies have identified a number of genes under positive selection in different NPV species (Chateigner *et al.*, 2015; Simón *et al.*, 2011; Xu *et al.*, 2013). In particular, Chateigner *et al.* (2015) detected genes under positive selection across all gene functional classes in AcMNPV, including a number of baculovirus core genes which encode components of the *per os* infectivity complex and factors essential for virus gene transcription. Alternatively, the shifts in variant composition could have resulted from stochastic, selectively neutral events occurring over the course of the *in vivo* passages, leading to genetic drift amongst post-passage HearNPV populations. Such events could occur during the primary infection of midgut epithelial cells if the transmission of genotypes to a

new host is largely a random process. It is likely that the selective and selectively neutral processes are not mutually exclusive and that both acted at different stages of an infection to shape the *DNA polymerase* and *me53* variant composition of post-passage HearNPV populations.

The differences in *DNA polymerase* and *me53* variant composition across the 125 passage lineages hold particular significance for natural HearNPV populations in the field as it suggests that the genotypic composition of HearNPV populations may continue to diverge with each passage in a new host individual. Previous studies have shown that NPV populations isolated from the same geographical region are more likely to have a similar genotypic composition than if they are isolated from two distant regions (Cooper *et al.*, 2003; Shapiro *et al.*, 1991), while Baillie & Bouwer (2012a) detected extensive genetic diversity between seven geographically distant HearNPV populations. If *H. armigera* host populations separate and become geographically isolated, the divergence in the genotypic composition of HearNPV populations over successive passages may be one mechanism that contributes to increasing HearNPV diversity between geographically distant host populations. In addition it was observed in the current study that there was greater variation in the genetic variant composition of HearNPV populations exposed to solar radiation relative to non-exposed populations (addressed in Chapter 4). This suggests the pre-infection exposure of HearNPV OBs to solar radiation, which leads to extensive virus inactivation (Mehrvar *et al.*, 2007; Sun *et al.*, 2004a), may be an additional factor which contributes to the divergence in genotypic composition of HearNPV populations in the field, over successive passages.

Of particular significance were the "drop-off" events observed in a small proportion of the 125 passage lineages, where there was a dramatic loss of multiple genetic variants from a HearNPV population over a single *in vivo* passage. These events had a lasting impact on the genetic variant composition of descendant populations and the genetic diversity remained low in subsequent post-passage populations. However, these "drop-off" events occurred infrequently at the high inoculum dose used and may have been caused by random events in the insect midgut which led to only a small subset of genotypes infecting a host. When the effect of ODV inoculum dose on the genetic diversity of post-passage HearNPV populations was assessed, "drop-off" events also occurred less frequently at the higher inoculum doses and there was a greater propensity for the genotypic composition of post-passage populations

to be similar to that of the inoculum population (addressed in Chapter 5). A similar observation was made by Baillie & Bouwer (2013) for high OB dose infections, where it was found that there was no significant difference between the genetic diversity of HearNPV populations isolated from hosts inoculated with an $LD_{95}$ and that of the inoculum population. This suggests that higher HearNPV inoculum doses may increase the probability of a host being exposed to and subsequently infected by the full variant spectrum of an inoculum population.

Although there were shifts in the relative abundance of certain variants in HearNPV populations over the course of three *in vivo* passages (reported in Chapter 3), a feature of the majority of the post-passage populations was the maintenance of a pool of low frequency variants present in the inoculum population. These findings appear to support those of Hitchman *et al.* (2007) who also found that low frequency variants were not progressively lost from a NPV population over several *in vivo* passages when a high inoculum dose was used. The maintenance of low frequency genetic variants also occurred in post-passage populations where the HearNPV OB inoculum was exposed to solar radiation, for 12 h or 21 h, prior to an *in vivo* passage (addressed in Chapter 4). Prior to this study it was hypothesised that low frequency variants would be less likely to be able to sustain losses caused by solar radiation-induced virus inactivation and that this would lead to the increased dominance in the post-passage population of variants already dominant in the inoculum population. However this was not observed as the pool of low frequency variants was largely maintained in post-passage HearNPV populations in both the 12 h and 21 h exposure groups. The co-occlusion of genotypes in OBs has been observed in a number of NPV species including HearNPV (Arrizubieta *et al.*, 2015; Bull *et al.*, 2001; Clavijo *et al.*, 2010), and may have contributed to the maintenance of low frequency genetic variants in post-passage HearNPV populations in the two UV exposure groups. If the co-occlusion of variants frequently occurs during HearNPV infections this would decrease the likelihood of solar radiation-induced inactivation reducing the probability of low frequency variants infecting a host, as they will be distributed over a greater number of OBs in the population.

When hosts were inoculated with low ODV doses there was a decline in the genetic diversity of the post-passage populations, relative to that of the inoculum population and populations isolated from hosts inoculated with high ODV doses. This loss of genetic diversity at low

ODV doses was due to a decline in the number of detectable variants in the post-passage populations. At the lowest ODV dose ($LD_5$) the data suggested that insect hosts were infected by a single HearNPV genotype. There was also evidence that in at least one of these hosts infection was initiated by a single ODV which resulted in host mortality (addressed in Chapter 5). If this was the case it would mean that one of the central tenets of the independent action hypothesis (IAH), that all virions have a non-zero probability of infecting a host and that virions act independently during the infection process (Druett, 1952; Lafforgue *et al.*, 2012; Zwart *et al.*, 2009, 2011), may apply to HearNPV infections in neonate *H. armigera* larvae. This would support the findings of a previous study by Zwart *et al.* (2009), which showed that a single AcMNPV ODV could initiate infections that led to host mortality in *A. californica* neonate larvae. This suggests the IAH may be more broadly applicable to NPV infections in neonate larvae, although further studies are needed to assess additional NPV-host interactions to confirm this. The results from the ODV infections also suggest that HearNPV genotypes in the midgut will infect the new host in a frequency-dependent manner, where genotypes with the greatest abundance in the ODV pool have the highest probability of infecting the host.

Whereas in the present study a drop in genetic diversity was observed for low ODV infections, a study by Baillie & Bouwer (2013) found that for hosts inoculated with a low OB dose ($LD_5$) the genetic diversity of post-passage populations was actually greater than that of the inoculum population. The increased diversity in these populations appeared to be due to the appearance of multiple novel genetic variants after the *in vivo* passage, as well as the maintenance of existing genetic variants that had been present in the inoculum population. The differences in the level of genetic diversity observed in post-passage populations isolated from hosts inoculated with low HearNPV doses, between this study and that of Baillie & Bouwer (2013), suggests that the occlusion of virions contributes to the maintenance and generation of HearNPV genetic diversity over the course of an *in vivo* passage. Baillie & Bouwer (2013) suggested that the origin of the novel *DNA polymerase* genetic variants they detected, in post-passage populations at low OB doses, were mutations generated by DNA polymerase errors. This was supported by an earlier study where they detected 60 SNPs, in the *DNA polymerase* region under study, in a pooled HearNPV population (Baillie & Bouwer, 2012a). However the rate at which mutations accumulate within a HearNPV population should not differ between low dose OB and ODV infections, as this is dependent

on the error rate of the viral DNA polymerase (Hang & Guarino, 1999; Huang & Levin, 2001).

The co-occlusion of genotypes could potentially account for the differences observed between the low dose OB infections in the study by Baillie & Bouwer (2013) and the low dose ODV infections in the current study. Co-occlusion would be expected to increase the frequency of hosts being exposed to and co-infected by multiple genotypes at low OB doses, whereas at low ODV doses, in the absence of virion occlusion, there is likely a higher probability that a host would only be exposed to and infected by the dominant genotypes in the inoculum population (Bull *et al.*, 2001; Clavijo *et al.*, 2010). Co-infection of a host by multiple genotypes would be expected to favour homologous recombination between heterologous virus genomes (Bull *et al.*, 2001; Clavijo *et al.*, 2010). This means homologous recombination between genotypes could have been higher in the low dose OB infections relative to the low dose ODV infections, and may have been largely responsible for the generation of the novel genetic variants in the Baillie & Bouwer (2013) study.

In this study a total of seven novel *me53* genetic variants which were not present in the inoculum populations were detected in post-passage populations, and appeared to have been generated over the course of an infection (addressed in Chapters 3-5). However, it appeared that six of these novel genetic variants were generated through recombination between existing inoculum variants, as they did not contain novel SNPs and were mosaics of existing genetic variant sequences. The replication of baculovirus genomes occurs through a recombination dependent mechanism (Hajós *et al.*, 2000; Okano *et al.*, 2006), which may facilitate the exchange of genetic regions between genotypes when two or more genotypes co-infect a single cell, which has been shown to occur relatively frequently *in vivo* (Bull *et al.*, 2001; Simón *et al.*, 2006). Only one of the novel genetic variants (mV10) was not a mosaic of existing sequence variants, and was more likely to have been generated through a mutation event rather than genomic recombination. The mV10 variant was also the only genetic variant detected which contained a non-synonymous SNP and encoded an amino acid sequence which differed to the other twelve genetic variants. This suggested that there was a low tolerance for non-synonymous SNPs in the *me53* gene that would lead to changes to the encoded amino acid sequence. Previous studies also found that synonymous SNPs outnumbered non-synonymous SNPs for several NPV species when the genomes of different

isolates were compared (Baillie & Bouwer, 2012a; Chateigner *et al.*, 2015; Thézé *et al.*, 2014), which suggests that non-synonymous SNPs are largely deleterious and are not maintained in virus populations over an infection. The pre-infection exposure of HearNPV OBs to solar radiation also did not appear to increase the frequency of novel synonymous or non-synonymous SNPs in post-passage populations (addressed in Chapter 5).

## 6.2. Conclusion

Overall, this study has contributed to the understanding of how HearNPV genetic diversity is maintained and generated over successive infections. PCR-DGGE and deep amplicon sequencing allowed the genetic variant composition and underlying nucleotide sequence diversity of HearNPV populations to be assessed. Using these methods it was shown that the genetic diversity of a HearNPV population can change dramatically over an *in vivo* passage. The shifts in genetic diversity were largely due to changes in the relative abundance of existing genotypes, although the generation of novel genetic variants also contributed to changes in the genotypic composition of HearNPV populations in this study. The results suggest that although the pre-infection exposure of OBs to solar radiation does not increase the frequency at which novel genetic variants are generated during an infection, it does contribute to the divergence in genotypic composition of HearNPV populations over successive passages. The occlusion of virions also appears to contribute to the increased maintenance of existing genotypes and the generation of novel genetic variants at low inoculum doses, as the level of genetic diversity detected after ODV $LD_5$ infections was much lower than that reported by Baillie & Bouwer (2013) for OB $LD_5$ infections. Further studies are needed to determine whether similar trends in genetic diversity would be observed for other gene regions within the HearNPV genome as well for the populations of different baculovirus species. Overall this study has shown that successive *in vivo* passages, together with solar radiation exposure most likely contribute to the differences in genetic diversity observed between geographical HearNPV isolates, while ODVs in the midgut appear to infect a *H. armigera* larva in a frequency-dependent manner, where genotypic variants that have the greatest abundance in the ODV pool have the highest probability of infecting the new host.

# References

**Ackermann, H. W. & Smirnoff, W. A. (1983).** A morphological Investigation of 23 Baculoviruses. *J Invertebr Pathol* **41**, 269–280.

**Allaway, G. P. & Payne, C. C. (1984).** Host range and virulence of five baculoviruses from lepidopterous hosts. *Ann Appl Biol* **105**, 29–37.

**Amend, A. S., Seifert, K. A. & Bruns, T. D. (2010).** Quantifying microbial communities with 454 pyrosequencing: Does read abundance count? *Mol Ecol* **19**, 5555–5565.

**Ampe, F. & Miambi, E. (2000).** Cluster analysis, richness and biodiversity indexes derived from denaturing gradient gel electrophoresis fingerprints of bacterial communities demonstrate that traditional maize fermentations are driven by the transformation process. *Int J Food Microbiol* **60**, 91–97.

**Anderson, M. J. (2001).** A new method for non-parametric multivariate analysis of variance. *Austral Ecol* **26**, 32–46.

**Anderson, M. J. (2006).** Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* **62**, 245–253.

**Anderson, M. J. & Walsh, D. C. I. (2013).** PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing? *Ecol Monogr* **83**, 557–574.

**Ansorge, W. J. (2009).** Next-generation DNA sequencing techniques. *N Biotechnol* **25**, 195–203.

**Arrizubieta, M., Simón, O., Williams, T. & Caballeroa, P. (2015).** A novel binary mixture of *Helicoverpa armigera* single nucleopolyhedrovirus genotypic variants has improved insecticidal characteristics for control of cotton bollworms. *Appl Environ Microbiol* **81**, 3984–3993.

**Baillie, V. L. & Bouwer, G. (2011).** Development of highly sensitive assays for detection of genetic variation in key *Helicoverpa armigera* nucleopolyhedrovirus genes. *J Virol Methods* **178**, 179–185.

**Baillie, V. L. & Bouwer, G. (2012a).** High levels of genetic variation within core *Helicoverpa armigera* nucleopolyhedrovirus genes. *Virus Genes* **44**, 149–162.

**Baillie, V. L. & Bouwer, G. (2012b).** High levels of genetic variation within *Helicoverpa armigera* nucleopolyhedrovirus populations in individual host insects. *Arch Virol* **157**, 2281–2289.

**Baillie, V. L. & Bouwer, G. (2013).** The effect of inoculum dose on the genetic diversity detected within *Helicoverpa armigera* nucleopolyhedrovirus populations. *J Gen Virol* **94**, 2524–2529.

**Beerenwinkel, N., Günthard, H. F., Roth, V. & Metzner, K. J. (2012).** Challenges and opportunities in estimating viral genetic diversity from next-generation sequencing data. *Front Microbiol* **3**, 1–16.

**Belshaw, R., Gardner, A., Rambaut, A. & Pybus, O. G. (2008).** Pacing a small cage: mutation and RNA viruses. *Trends Ecol Evol* **23**, 188–193.

**Bernal, A., Simón, O., Williams, T., Muñoz, D. & Caballero, P. (2013).** A *Chrysodeixis chalcites* single-nucleocapsid nucleopolyhedrovirus population from the Canary Islands is genotypically structured to maximize survival. *Appl Environ Microbiol* **79**, 7709–7718.

**Bideshi, D. K., Bigot, Y. & Federici, B. A. (2000).** Molecular characterization and phylogenetic analysis of the *Harrisina brillians* granulovirus granulin gene. *Arch Virol* **145**, 1933–1945.

**Bliss, C. I. (1934).** The method of probits. *Science* **79**, 38–39.

**Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., Mills, D. A. & Caporaso, J. G. (2012).** Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat Methods* **10**, 57–59.

**Bot, J. (1966).** Rearing *Heliothis armigera* (Hübner) and *Prodemia litura F.* on an artificial diet. *South African J Agric Sci* **9**, 353–538.

**Bouwer, G. & Avyidi, D. (2006).** Application of the droplet feeding assay method to two economically important African lepidopteran pests. *African Entomol* **14**, 195–198.

**Brandariz-Fontes, C., Camacho-Sanchez, M., Vilà, C., Vega-Pla, J. L., Rico, C. & Leonard, J. A. (2015).** Effect of the enzyme and PCR conditions on the quality of high-throughput DNA sequencing results. *Sci Rep* **5**, 8056.

**Brooks, J. P., Edwards, D. J., Harwich, M. D., Rivera, M. C., Fettweis, J. M., Serrano, M. G., Reris, R. A., Sheth, N. U., Huang, B. & other authors**. **(2015).** The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiol* **15**, 66.

**Bull, J. C., Godfray, H. C. J. & Reilly, D. R. O. (2001).** Persistence of an occlusion-negative recombinant nucleopolyhedrovirus in *Trichoplusia ni* indicates high multiplicity of cellular infection. *Appl Environ Microbiol* **67**, 5204–5209.

**Bull, R. A., Luciani, F., McElroy, K., Gaudieri, S., Pham, S. T., Chopra, A., Cameron, B., Maher, L., Dore, G. J. & other authors**. **(2011).** Sequential bottlenecks drive viral evolution in early acute hepatitis C virus infection. *PLoS Pathog* **7**, e1002243.

**Bullock, H. R. (1967).** Persistence of Heliothis nuclear-polyhedrosis virus on cotton foliage. *J Invertebr Pathol* **9**, 434–436.

**Busch, U. & Nitschko, H. (1999).** Methods for the differentiation of microorganisms. *J Chromatogr B Biomed Sci Appl* **722**, 263–278.

**Carlson, J. M., Schaefer, M., Monaco, D. C., Batorsky, R., Claiborne, D. T., Prince, J., Deymier, M. J., Ende, Z. S., Klatt, N. R. & other authors**. **(2014).** Selection bias at the heterosexual HIV-1 transmission bottleneck. *Science (80- )* **345**, 1254031–1254031.

**Carmona, M., Sepúlveda, D., Cárdenas, C., Nilo, L. & Marshall, S. H. (2012).** Denaturing gradient gel electrophoresis (DGGE) as a powerful novel alternative for differentiation of epizootic ISA virus variants. *PLoS One* **7**, e37353.

**Carstens, E. B. (2009).** AcMNPV as a model for baculovirus DNA replication. *Virol Sin* **24**, 243–267.

**Chateigner, A., Bézier, A., Labrousse, C., Jiolle, D., Barbe, V. & Herniou, E. (2015).** Ultra deep sequencing of a baculovirus population reveals widespread genomic variations. *Viruses* **7**, 3625–3646.

**Chen, R. & LaPara, T. M. (2006).** Aerobic biological treatment of low-strength synthetic wastewater in membrane-coupled bioreactors: the structure and function of bacterial enrichment cultures as the net growth rate approaches zero. *Microb Ecol* **51**, 99–108.

**Chen, X., Sun, X., Hu, Z., Li, M., O'Reilly, D. R., Zuidema, D. & Vlak, J. M. (2000).** Genetic engineering of *Helicoverpa armigera* single-nucleocapsid nucleopolyhedrovirus as an improved pesticide. *J Invertebr Pathol* **76**, 140–146.

**Chen, X., IJkel, W. F., Tarchini, R., Sun, X., Sandbrink, H., Wang, H., Peters, S., Zuidema, D., Lankhorst, R. K. & other authors**. **(2001).** The sequence of the *Helicoverpa armigera* single nucleocapsid nucleopolyhedrovirus genome. *J Gen Virol* **82**, 241–257.

**Chen, X., Zhang, W.-J., Wong, J., Chun, G., Lu, A., McCutchen, B. F., Presnail, J. K., Herrmann, R., Dolan, M. & other authors**. **(2002).** Comparative analysis of the complete genome sequences of *Helicoverpa zea* and *Helicoverpa armigera* single-nucleocapsid nucleopolyhedroviruses. *J Gen Virol* **83**, 673–684.

**Cherry, A. J., Rabindra, R. J., Parnell, M. A., Geetha, N., Kennedy, J. S. & Grzywacz, D. (2000).** Field evaluation of *Helicoverpa armigera* nucleopolyhedrovirus formulations for the control of chickpea pod-borer, *H. armigera* (Hubn.), on chickpea (*Cicer arietinum* var. Shoba) in southern India. *Crop Prot* **19**, 51–60.

**Ciccazzo, S., Esposito, A., Rolli, E., Zerbe, S., Daffonchio, D. & Brusetti, L. (2014).** Different pioneer plant species select specific rhizosphere bacterial communities in a high mountain environment. *Springerplus* **3**, 391.

**Clarke, K. R. (1993).** Non-parametric multivariate analyses in community structure. *Aust J Ecol* **18**, 117–143.

**Clavijo, G., Williams, T., Simón, O., Muñoz, D., Cerutti, M., López Ferber, M. & Caballero, P. (2009).** Mixtures of complete and *pif1*- and *pif2*-deficient genotypes are required for increased potency of an insect nucleopolyhedrovirus. *J Virol* **83**, 5127–5136.

**Clavijo, G., Williams, T., Muñoz, D., Caballero, P. & López Ferber, M. (2010).** Mixed genotype transmission bodies and virions contribute to the maintenance of diversity in an insect virus. *Proc Biol Sci* **277**, 943–951.

**Cooper, D., Cory, J. S. & Myers, J. H. (2003).** Hierarchical spatial structure of genetically variable nucleopolyhedroviruses infecting cyclic populations of western tent caterpillars. *Mol Ecol* **12**, 881–890.

**Cory, J. S., Green, B. M., Paul, R. K. & Hunter-Fujita, F. (2005).** Genotypic and phenotypic diversity of a baculovirus population within an individual insect host. *J Invertebr Pathol* **89**, 101–111.

**Cory, J. S. & Myers, J. H. (2003).** The ecology and evolution of insect baculoviruses. *Annu Rev Ecol Evol Syst* **34**, 239–272.

**Croizier, G. & Ribeiro, H. C. T. (1992).** Recombination as a possible major cause of genetic heterogeneity in *Anticarsia gemmatalis* nuclear polyhedrosis virus wild populations. *Virus Res* **26**, 183–196.

**Crook, N. E. & Payne, C. C. (1980).** Comparison of three methods of ELISA for baculoviruses. *J Gen Virol* **46**, 29–37.

**Daimon, T., Hamada, K., Mita, K., Okano, K., Suzuki, M. G., Kobayashi, M. & Shimada, T. (2003).** A *Bombyx mori* gene, *BmChi-h*, encodes a protein homologous to bacterial and baculovirus chitinases. *Insect Biochem Mol Biol* **33**, 749–759.

**Darmawan, H., Harrison, M. & Reha-Krantz, L. J. (2015).** DNA polymerase $3' \rightarrow 5'$ exonuclease activity: Different roles of the beta hairpin structure in family-B DNA polymerases. *DNA Repair (Amst)* **29**, 36–46.

**David, V., Terrat, S., Herzine, K., Claisse, O., Rousseaux, S., Tourdot-Maréchal, R., Masneuf-Pomarede, I., Ranjard, L. & Alexandre, H. (2014).** High-throughput sequencing of amplicons for monitoring yeast biodiversity in must and during alcoholic fermentation. *J Ind Microbiol Biotechnol* **41**, 811–821.

**Domingo, E., Sheldon, J. & Perales, C. (2012).** Viral quasispecies evolution. *Microbiol Mol Biol Rev* **76**, 159–216.

**Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K. & Mattick, J. S. (1991).** 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res* **19**, 4008.

**Dong, X. & Reddy, G. B. (2010).** Soil bacterial communities in constructed wetlands treated with swine wastewater using PCR-DGGE technique. *Bioresour Technol* **101**, 1175–1182.

**Dowd, S. E., Sun, Y., Secor, P. R., Rhoads, D. D., Wolcott, B. M., James, G. A. & Wolcott, R. D. (2008).** Survey of bacterial diversity in chronic wounds using pyrosequencing, DGGE, and full ribosome shotgun sequencing. *BMC Microbiol* **8**, 43.

**Drake, J. W. (1991).** A constant rate of spontaneous mutation in DNA-based microbes. *Proc Natl Acad Sci* **88**, 7160–7164.

**Drake, J. W. (2006).** Chaos and Order in Spontaneous Mutation. *Genetics* **173**, 1–8.

**Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. F. (1998).** Rates of spontaneous mutation. *Genetics* 1667–1686.

**Druett, H. (1952).** Bacterial invasion. *Nature* **170**, 288.

**Duffy, S., Shackelton, L. A. & Holmes, E. C. (2008).** Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet* **9**, 267–276.

**Edgar, R. (2016).** UCHIME2: improved chimera prediction for amplicon sequencing. *bioRxiv* 74252.

**Edgar, R. C. (2013).** UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* **10**, 996–998.

**Edgar, R. C. & Flyvbjerg, H. (2015).** Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics* **31**, 3476–3482.

**Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. (2011).** UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200.

**Edwards, C. T., Holmes, E. C., Wilson, D. J., Viscidi, R. P., Abrams, E. J., Phillips, R. E. & Drummond, A. J. (2006).** Population genetic estimation of the loss of genetic diversity during horizontal transmission of HIV-1. *BMC Evol Biol* **6**, 28.

**Elena, S. F. & Sanjua, R. (2005).** Adaptive Value of High Mutation Rates of RNA Viruses : Separating Causes from Consequences. *J Virol* **79**, 11555–11558.

**England, L. S., Holmes, S. B. & Trevors, J. T. (1998).** Review : Persistence of viruses and DNA in soil. *World J Microbiol Biotechnol* **14**, 163–170.

**Enwall, K. & Hallin, S. (2009).** Comparison of T-RFLP and DGGE techniques to assess denitrifier community composition in soil. *Lett Appl Microbiol* **48**, 145–148.

**Evans, H. & Shapiro, M. (1997).** Viruses. In *Man Tech Insect Pathol*, pp. 17–53. Edited by L. A. Lacey. Oxford: Elsevier Academic Press.

**Ferreira, B. C., Melo, F. L., Souza, M. L., Moscardi, F. F., Ribeiro, B. M. & Báo, S. N. (2014).** High genetic stability of peroral infection factors from *Anticarsia gemmatalis* MNPV over 20 years of sampling. *J Invertebr Pathol* **118**, 66–70.

**Ferris, M. J. & Ward, D. M. (1997).** Seasonal distributions of dominant 16S rRNA-defined populations in a hot spring microbial mat examined by denaturing gradient gel electrophoresis. *Appl Enviromental Microbiol* **63**, 1375–1381.

**Ferris, M. J., Muyzer, G. & Ward, D. M. (1996).** Denaturing gradient gel electrophoresis profiles of 16S rRNA-defined populations inhabiting a hot spring microbial mat community. *Appl Environ Microbiol* **62**, 340–346.

**Firth, C., Kitchen, A., Shapiro, B., Suchard, M. A., Holmes, E. C. & Rambaut, A. (2010).** Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. *Mol Biol Evol* **27**, 2038–2051.

**Fischer, S. G. & Lerman, L. S. (1980).** Separation of random fragments of DNA according to properties of their sequences. *Proc Natl Acad Sci U S A* **77**, 4420–4424.

**Fischer, S. G. & Lerman, L. S. (1983).** DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: Correspondence with melting theory. *Proc Natl Acad Sci U S A* **80**, 1579–1583.

**Fraser, M. J., Brusca, J. S., Smith, G. E. & Summers, M. D. (1985).** Transposon-mediated mutagenesis of a baculovirus. *Virology* **145**, 356–361.

**Fromin, N., Hamelin, J., Tarnawski, S., Roesti, D., Forestier, N., Gillet, F., Aragno, M. & Rossi, P. (2002).** Statistical analysis of denaturing gel electrophoresis (DGE) fingerprinting patterns. *Environ Microbiol* **4**, 634–643.

**Fuchs, L. Y., Woods, M. S. & Weaver, R. F. (1983).** Viral transcription during *Autographa californica* nuclear polyhedrosis virus infection: a novel RNA polymerase induced in infected *Spodoptera frugiperda* cells. *J Virol* **48**, 641–646.

**Furio, V., Moya, A. & Sanjua, R. (2005).** The cost of replication fidelity in an RNA virus. *Proc Natl Acad Sci* **102**, 10233–10237.

**Gafan, G. P., Lucas, V. S., Roberts, G. J., Petrie, A., Wilson, M., Spratt, D. A. & Petrie, A. (2005).** Statistical Analyses of Complex Denaturing Gradient Gel Electrophoresis Profiles. *J Clin Microbiol* **43**, 3971–3978.

**Garavaglia, M. J., Miele, S. A. B., Iserte, J. A., Belaich, M. N. & Ghiringhelli, P. D. (2012).** The *ac53, ac78, ac101,* and *ac103* genes are newly discovered core genes in the family *Baculoviridae*. *J Virol* **86**, 12069–12079.

**Gettig, R. R. & Mccarthy, W. J. (1982).** Genotypic variation among wild isolates of *Heliothis* spp nuclear polyhedrosis viruses from different geographical regions. *Virology* **117**, 245–252.

**Gilbert, C., Chateigner, A., Ernenwein, L., Barbe, V., Bézier, A., Herniou, E. A. & Cordaux, R. (2014).** Population genomics supports baculoviruses as vectors of horizontal transfer of insect transposons. *Nat Commun* **5**, 3348.

**Gilbert, C., Peccoud, J., Chateigner, A., Moumen, B., Cordaux, R. & Herniou, E. A. (2016).** Continuous influx of genetic material from host to virus populations. *PLOS Genet* **12**, e1005838.

**Gomes, N. C. M., Fagbola, O., Costa, R., Rumjanek, N. G., Buchner, A., Mendona-hagler, L. & Smalla, K. (2003).** Dynamics of fungal communities in bulk and maize rhizosphere soil in the tropics. *Appl Environ Microbiol* **69**, 3758–3766.

**Gonçalves Rossi, L. M., Escobar-Gutierrez, A. & Rahal, P. (2016).** Multiregion deep sequencing of hepatitis C virus: An improved approach for genetic relatedness studies. *Infect Genet Evol* **38**, 138–145.

**Gong, L., Han, Y., Chen, L., Liu, F., Hao, P., Sheng, J., Li, X. H., Yu, D. M., Gong, Q. M. & other authors**. **(2013).** Comparison of next-generation sequencing and clone-based sequencing in analysis of hepatitis b virus reverse transcriptase Quasispecies heterogeneity. *J Clin Microbiol* **51**, 4087–4094.

**Goodwin, S., McPherson, J. D. & McCombie, W. R. (2016).** Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* **17**, 333–351.

**Görzer, I., Guelly, C., Trajanoski, S. & Puchhammer-Stockl, E. (2010).** Deep sequencing reveals highly complex dynamics of human cytomegalovirus genotypes in transplant patients over time. *J Virol* **84**, 7195–7203.

**Grad, Y. H., Newman, R., Zody, M., Yang, X., Murphy, R., Qu, J., Malboeuf, C. M., Levin, J. Z., Lipsitch, M. & DeVincenzo, J. (2014).** Within-host whole-genome deep sequencing and diversity analysis of human respiratory syncytial virus infection reveals dynamics of genomic diversity in the absence and presence of immune pressure. *J Virol* **88**, 7286–7293.

**Graham, R. I., Tyne, W. I., Possee, R. D., Sait, S. M. & Hails, R. S. (2004).** Genetically variable nucleopolyhedroviruses isolated from spatially separate populations of the winter moth *Operophtera brumata* (Lepidoptera: Geometridae) in Orkney. *J Invertebr Pathol* **87**, 29–38.

**Granados, R. R. & Lawler, K. A. (1981).** *In Vivo* Pathway of *Autographa californica* Baculovirus Invasion and Infection. *Virology* **108**, 297–308.

**Green, S. J., Leigh, M. B. & Neufeld, J. D. (2010).** Denaturing gradient gel electrophoresis (DGGE) for microbial community analysis. In *Handb Hydrocarb Lipid Microbiol*, pp. 4137–4158. Edited by K. N. Timmis. Berlin: Springer.

**Gregori, J., Salicrú, M., Domingo, E., Sanchez, A., Esteban, J. I., Rodríguez-Frías, F. & Quer, J. (2014).** Inference with viral quasispecies diversity indices: Clonal and NGS approaches. *Bioinformatics* **30**, 1104–1111.

**Griego, V. M., Martignoni, M. E. & Claycomb, A. E. (1985).** Inactivation of nuclear polyhedrosis virus (Baculovirus subgroup A) by monochromatic UV radiation. *Appl Environ Microbiol* **49**, 709–10.

**Guarino, L. A., Xu, B., Jin, J., Dong, W. & Xu, B. I. N. (1998).** A Virus-Encoded RNA Polymerase Purified from Baculovirus-Infected Cells. *J Virol* **72**, 7989–7991.

**Gutiérrez, S., Michalakis, Y. & Blanc, S. (2012).** Virus population bottlenecks during within-host progression and host-to-host transmission. *Curr Opin Virol* **2**, 546–555.

**Haase, S., Sciocco-Cap, A. & Romanowski, V. (2015).** Baculovirus insecticides in Latin America: historical overview, current status and future perspectives. *Viruses* **7**, 2230–2267.

**Hajós, J., Pijnenburg, J. & Usmany, M. (2000).** High frequency recombination between homologous baculoviruses in cell culture. *Arch Virol* **145**, 159–164.

**Hamblin, M., Van Beek, N. A. M., Hughes, P. R. & Wood, H. A. (1990).** Co-occlusion and persistence of a baculovirus mutant lacking the polyhedrin gene. *Appl Environ Microbiol* **56**, 3057–3062.

**Hamm, J. J. & Styer, E. L. (1985).** Comparative pathology of isolates of *Spodoptera frugiperda* nuclear polyhedrosis virus in *S. frugiperda* and *S. exigua*. *J Gen Virol* **66**, 1249–1261.

**Hang, X. & Guarino, L. A. (1999).** Purification of *Autographa californica* nucleopolyhedrovirus DNA polymerase from infected insect cells. *J Gen Virol* **80**, 2519–2526.

**Harrap, K. & Longworth, J. (1974).** An evaluation of purification methods for baculoviruses. *J Invertebr Pathol* **24**, 55–62.

**Harris, K. A. & Teo, C. G. (2001).** Diversity of hepatitis C virus quasispecies evaluated by denaturing gradient gel electrophoresis. *Clin Vaccine Immunol* **8**, 62–73.

**Harrison, R. L. (2009).** Structural divergence among genomes of closely related baculoviruses and its implications for baculovirus evolution. *J Invertebr Pathol* **101**, 181–186.

**Harrison, R. L., Keena, M. a & Rowley, D. L. (2014).** Classification, genetic variation and pathogenicity of *Lymantria dispar* nucleopolyhedrovirus isolates from Asia, Europe, and North America. *J Invertebr Pathol* **116**, 27–35.

**Herniou, E. A., Arif, B. M., Becnel, J. J., Blissard, G. W., Bonning, B. C., Harrison, R. L., Jehle, J. A., Theilmann, D. A. & Vlak, J. M. (2011).** Family *Baculoviridae*. In *Virus Taxon ninth Rep Int Comm Taxon Viruses*, pp. 163–173.

**Herniou, E. A. & Jehle, J. A. (2007).** Baculovirus phylogeny and evolution. *Curr Drug Targets* **8**, 1043–1050.

**Herniou, E. A., Luque, T., Chen, X., Vlak, J. M., Winstanley, D., Cory, J. S. & Reilly, D. R. O. (2001).** Use of Whole Genome Sequence Data To Infer Baculovirus Phylogeny. *J Virol* **75**, 8117–8126.

**Herniou, E. A., Olszewski, J. A., Cory, J. S. & O'Reilly, D. R. (2003).** The genome sequence and evolution of baculoviruses. *Annu Rev Entomol* **48**, 211–234.

**Hitchman, R. B., Hodgson, D. J., King, L. A., Hails, R. S., Cory, J. S. & Possee, R. D. (2007).** Host mediated selection of pathogen genotypes as a mechanism for the maintenance of baculovirus diversity in the field. *J Invertebr Pathol* **94**, 153–162.

**Hodgson, D. J., Hitchman, R. B., Vanbergen, A. J., Hails, R. S., Possee, R. D. & Cory, J. S. (2004).** Host ecology determines the relative fitness of virus genotypes in mixed-genotype nucleopolyhedrovirus infections. *J Evol Biol* **17**, 1018–1025.

**Hodgson, D. J., Vanbergen, A. J., Watt, A. D., Hails, R. S. & Cory, J. S. (2001).** Phenotypic variation between naturally co-existing genotypes of a Lepidopteran baculovirus. *Evol Ecol Res* **3**, 687–701.

**Hodgson, D. & Vanbergen, A. (2002).** Differential selection of baculovirus genotypes mediated by different species of host food plant. *Ecol Lett* **5**, 512–518.

**Holm, S. (1979).** A simple sequentially rejective multiple test procedure. *Scand J Stat* **6**, 65–70.

**Holmes, E. C. (2008).** Comparative Studies of RNA Virus Evolution. In *Orig Evol viruses*, 2nd edn., pp. 119–134. Elsevier.

**Holmes, E. C. & Moya, A. (2002).** Is the Quasispecies Concept Relevant to RNA Viruses ? *J Virol* **76**, 460–462.

**Hong, H., Pruden, A. & Reardon, K. F. (2007).** Comparison of CE-SSCP and DGGE for monitoring a complex microbial community remediating mine drainage. *J Microbiol Methods* **69**, 52–64.

**Huang, H., Wang, M., Deng, F., Wang, H. & Hu, Z. (2012).** ORF85 of HearNPV encodes the *per os* infectivity factor 4 (PIF4) and is essential for the formation of the PIF complex. *Virology* **427**, 217–223.

**Huang, J. & Levin, D. B. (2001).** Expression, purification and characterization of the *Spodoptera littoralis* nucleopolyhedrovirus (SpliNPV) DNA polymerase and interaction with the SpliNPV non-*hr* origin of DNA replication. *J Gen Virol* **82**, 1767–1776.

**Hughes, A. L. & Friedman, R. (2003).** Genome-wide survey for genes horizontally transferred from cellular organisms to baculoviruses. *Mol Biol Evol* **20**, 979–987.

**Hughes, P. & Wood, H. A. (1981).** A synchronous peroral technique for the bioassay of insect viruses. *J Invertebr Pathol* **37**, 154–159.

**Ignoffo, C. M., Bradley, J. R., Gilliland, F. R., Harris, F. A., Falcon, L. A., Larson, L. & Yearian, W. C. (1972).** Field studies on stability of the Heliothis nucleopolyhedrosis virus at various sites throughout the cotton belt. *Environ Entomol* **1**, 388–390.

**Ihrmark, K., Bödeker, I. T. M., Cruz-Martinez, K., Friberg, H., Kubartova, A., Schenck, J., Strid, Y., Stenlid, J., Brandström-Durling, M. & other authors**. **(2012).** New primers to amplify the fungal ITS2 region - evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiol Ecol* **82**, 666–677.

**Ikehata, H. & Ono, T. (2011).** The Mechanisms of UV Mutagenesis. *J Radiat Res* **52**, 115–125.

**Iyer, L. M., Aravind, L. & Koonin, E. V**. **(2001).** Common origin of four diverse families of large eukaryotic DNA viruses. *J Virol* **75**, 11720–11734.

**Jacquet, S. & Bratbak, G. (2003).** Effects of ultraviolet radiation on marine virus-phytoplankton interactions. *FEMS Microbiol Ecol* **44**, 279–89.

**Janse, I., Bok, J. & Zwart, G. (2004).** A simple remedy against artifactual double bands in denaturing gradient gel electrophoresis. *J Microbiol Methods* **57**, 279–281.

**Jehle, J. a, Lange, M., Wang, H., Hu, Z., Wang, Y. & Hauschild, R. (2006a).** Molecular identification and phylogenetic analysis of baculoviruses from Lepidoptera. *Virology* **346**, 180–193.

**Jehle, J. A., Blissard, G. W., Bonning, B. C., Cory, J. S., Herniou, E. A., Rohrmann, G. F., Theilmann, D. A., Thiem, S. M. & Vlak, J. M. (2006b).** On the classification and nomenclature of baculoviruses: a proposal for revision. *Arch Virol* **151**, 1257–1266.

**de Jong, J., Theilmann, D. a, Arif, B. M. & Krell, P. J. (2011).** Immediate-early protein ME53 forms foci and colocalizes with GP64 and the major capsid protein VP39 at the cell membranes of *Autographa californica* multiple nucleopolyhedrovirus-infected cells. *J Virol* **85**, 9696–9707.

**Jost, L. (2007).** Partitioning diversity into alpha and beta components. *Ecology* **89**, 2427–2439.

**Kamita, S. G., Maeda, S. & Bruce, D. (2003).** High-Frequency homologous recombination between baculoviruses involves DNA replication. *J Virol* **77**, 13053–13061.

**Katsuma, S., Kawaoka, S., Mita, K. & Shimada, T. (2008).** Genome-wide survey for baculoviral host homologs using the *Bombyx* genome sequence. *Insect Biochem Mol Biol* **38**, 1080–1086.

**Keele, B. F., Giorgi, E. E., Salazar-Gonzalez, J. F., Decker, J. M., Pham, K. T., Salazar, M. G., Sun, C., Grayson, T., Wang, S. & other authors**. **(2008).** Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci* **105**, 7552–7557.

**Kikhno, I., Gutiérrez, S., Croizier, L., Croizier, G. & López Ferber, M. (2002).** Characterization of *pif*, a gene required for the *per os* infectivity of *Spodoptera littoralis* nucleopolyhedrovirus. *J Gen Virol* **83**, 3013–3022.

**Kim, T. W., Lee, J.-H., Park, M.-H. & Kim, H.-Y. (2010).** Analysis of bacterial and fungal communities in Japanese- and Chinese-fermented soybean pastes using nested PCR–DGGE. *Curr Microbiol* **60**, 315–320.

**Kircher, M. & Kelso, J. (2010).** High-throughput DNA sequencing - Concepts and limitations. *BioEssays* **32**, 524–536.

**Knebel-Mörsdorf, D., Kremer, A. & Jahnel, F. (1993).** Baculovirus gene ME53, which contains a putative zinc finger motif, is one of the major early-transcribed genes. *J Virol* **67**, 753–758.

**Kolodny-hirsch, D. M. & Van Beek, N. A. M. (1997).** Selection of a morphological variant of *Autographa californica* nuclear polyhedrosis virus with increased virulence following serial passage in *Plutella xylostella*. *J Invertebr Pathol* **69**, 205–211.

**Koo, O.-K., Kim, H.-J., Baker, C. A., Park, S. H. & Ricke, S. C. (2016).** Microbial diversity of ground beef products in South Korean retail market analyzed by PCR-DGGE and 454 pyrosequencing. *Food Biotechnol* **30**, 63–77.

**Kost, T. a, Condreay, J. P. & Jarvis, D. L. (2005).** Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nat Biotechnol* **23**, 567–575.

**Kowalchuk, G. a, Stephen, J. R., De Boer, W., Prosser, J. I., Embley, T. M. & Woldendorp, J. W. (1997).** Analysis of ammonia-oxidizing bacteria of the beta subdivision of the class *Proteobacteria* in coastal sand dunes by denaturing gradient gel electrophoresis and sequencing of PCR-amplified 16S ribosomal DNA fragments. *Appl Environ Microbiol* **63**, 1489–1497.

**Kraková, L., Šoltys, K., Budiš, J., Grivalský, T., Ďuriš, F., Pangallo, D. & Szemes, T. (2016).** Investigation of bacterial and archaeal communities: novel protocols using modern sequencing by Illumina MiSeq and traditional DGGE-cloning. *Extremophiles* **20**, 795–808.

**Lafforgue, G., Tromas, N., Elena, S. F. & Zwart, M. P. (2012).** Dynamics of the Establishment of Systemic Potyvirus Infection: Independent yet Cumulative Action of Primary Infection Sites. *J Virol* **86**, 12912–12922.

**Lange, M., Wang, H., Zhihong, H. & Jehle, J. a. (2004).** Towards a molecular identification and classification system of lepidopteran-specific baculoviruses. *Virology* **325**, 36–47.

**Lauring, A. S. & Andino, R. (2010).** Quasispecies theory and the behavior of RNA viruses. *PLoS Pathog* **6**, e1001005.

**Lee, H. H. & Miller, L. K. (1978).** Isolation of genotypic variants of *Autographa californica* nuclear polyhedrosis virus. *J Virol* **27**, 754–767.

**Legatzki, A., Ortiz, M., Neilson, J. W., Casavant, R. R., Palmer, M. W., Rasmussen, C., Pryor, B. M., Pierson, L. S. & Maier, R. M. (2012).** Factors influencing observed variations in the structure of bacterial communities on calcite formations in Kartchner Caverns, AZ, USA. *Geomicrobiol J* **29**, 422–434.

**Legatzki, A., Ortiz, M., Neilson, J. W., Dominguez, S., Andersen, G. L., Toomey, R. S., Pryor, B. M., Pierson, L. S. & Maier, R. M. (2011).** Bacterial and archaeal community structure of two adjacent calcite speleothems in Kartchner Caverns, Arizona, USA. *Geomicrobiol J* **28**, 99–117.

**Leisy, D. J. & Rohrmann, G. F. (1993).** Characterization of the replication of plasmids containing *hr* sequences in baculovirus-infected *Spodoptera frugiperda* cells. *Virology* **196**, 722–730.

**Lenz, T. L. & Becker, S. (2008).** Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci - Implications for evolutionary analysis. *Gene* **427**, 117–123.

**Loisel, P., Harmand, J., Zemb, O., Latrille, E., Lobry, C., Delgenes, J.-P. & Godon, J.-J. (2006).** Denaturing gradient electrophoresis (DGE) and single-strand conformation polymorphism (SSCP) molecular fingerprintings revisited by simulation and used as a tool to measure microbial diversity. *Environ Microbiol* **8**, 720–731.

**Long, G., Pan, X., Westenberg, M. & Vlak, J. M. (2006).** Functional role of the cytoplasmic tail domain of the major envelope fusion protein of group II baculoviruses. *J Virol* **80**, 11226–11234.

**López Ferber, M., Simón, O., Williams, T. & Caballero, P. (2003).** Defective or effective? Mutualistic interactions between virus genotypes. *Proc Biol Sci* **270**, 2249–2255.

**Lucas, M., Karrer, U., Lucas, A. & Klenerman, P. (2008).** Viral escape mechanisms - escapology taught by viruses. *Int J Exp Pathol* **82**, 269–286.

**Lynch, M. (2010).** Evolution of the mutation rate. *Trends Genet* **26**, 345–52.

**Manrubia, S. & Lazaro, E. (2006).** Viral evolution. *Phys Life Rev* **3**, 65–92.

**Mardis, E. R. (2013).** Next-generation sequencing platforms. *Annu Rev Anal Chem (Palo Alto Calif)* **6**, 287–303.

**Martin, D. W. & Weber, P. C. (1997).** DNA replication promotes high-frequency homologous recombination during Autographa californica multiple nuclear polyhedrosis virus infection. *Virology* **232**, 300–309.

**Martin, M. (2011).** Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10.

**Matsuyama, T., Nakajima, Y., Matsuya, K., Ikenaga, M., Asakawa, S. & Kimura, M. (2007).** Bacterial community in plant residues in a Japanese paddy field estimated by RFLP and DGGE analyses. *Soil Biol Biochem* **39**, 463–472.

**Mcleod, P. J., Yearian, W. C. & Young, S. (1977).** Inactivation of Baculovirus heliothis by Ultraviolet Irradiation, Dew, and Temperature. *J Invertebr Pathol* **30**, 237–241.

**Mehrvar, A., Rabindra, R. J., Veenakumari, K. & Narabenchi, G. B. (2007).** Effect of natural sunlight on the activity of different geographic isolates of nucleopolyhedrovirus of *Helicoverpa armigera* (Hübner). *J Biol Control* **21**, 235–239.

**Mikhailov, V. S., Marlyev, K. A., Ataeva, J. O., Kullyev, P. K. & Atrazhev, A. M. (1986).** Characterisation of 3'-5' exonuclease associated with DNA polymerase of silkworm nuclear polyhedrosis virus. *Nucleic Acids Res* **14**, 3841–3857.

**Mikhaĭlov, V. S. (2003).** Replication of the baculovirus genome. *Mol Biol* **37**, 250–259.

**Morse, S. S. & Schluederberg, A. (1990).** Emerging Viruses: The evolution of viruses and viral diseases. *J Infect Dis* **162**, 1–7.

**Moscardi, F. (1999).** Assessment of the application of baculoviruses for control of Lepidoptera. *Annu Rev Entomol* **44**, 257–89.

**Moscardi, F., Lobo de Souza, M., Batista de Castro, M. E., Moscardi, M. L. & Szewczyk, B. (2011).** Baculovirus pesticides: present state and future perspectives. In *Microbes Microb Technol Agric Environ Appl*, pp. 415–445. Edited by I. Ahmad, F. Ahmad & J. Pichtel. Springer New York.

**Motta, F. C., Rosado, a S. & Couceiro, J. N. S. S. (2002).** Standardization of denaturing gradient gel electrophoresis for mutant screening of influenza A (H3N2) virus samples. *J Virol Methods* **101**, 105–115.

**Motta, F. C., Rosado, A. S. & Siqueira, M. M. (2006).** Comparison between denaturing gradient gel electrophoresis and phylogenetic analysis for characterization of A/H3N2 influenza samples detected during the 1999-2004 epidemics in Brazil. *J Virol Methods* **135**, 76–82.

**Murillo, R., Muñoz, D., Williams, T., Mugeta, N. & Caballero, P. (2006).** Application of the PCR-RFLP method for the rapid differentiation of *Spodoptera exigua* nucleopolyhedrovirus genotypes. *J Virol Methods* **135**, 1–8.

**Muyzer, G. (1999).** DGGE/TGGE a method for identifying genes from natural ecosystems. *Curr Opin Microbiol* **2**, 317–322.

**Muyzer, G. & Smalla, K. (1998).** Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. *Antonie Van Leeuwenhoek* **73**, 127–141.

**Muyzer, G., de Waal, E. C. & Uitterlinden, A. G. (1993).** Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* **59**, 695–700.

**Myers, R. M., Fischer, S. G., Lermani, L. S. & Maniatis, T. (1985).** Nearly all single base substitutions in DNA fragments joined to a GC-clamp can be detected by denaturing gradient gel electrophoresis. *Nucleic Acids Res* **13**, 3131–3145.

**Narayanan, D. L., Saladi, R. N. & Fox, J. L. (2010).** Ultraviolet radiation and skin cancer. *Int J Dermatol* **49**, 978–86.

**Neilson, J. W., Jordan, F. L. & Maier, R. M. (2013).** Analysis of artifacts suggests DGGE should not be used for quantitative diversity analysis. *J Microbiol Methods* **92**, 256–263.

**Nelson, C. W. & Hughes, A. L. (2015).** Within-host nucleotide diversity of virus populations: Insights from next-generation sequencing. *Infect Genet Evol* **30**, 1–7.

**Nicolaisen, M. H. & Ramsing, N. B. (2002).** Denaturing gradient gel electrophoresis (DGGE) approaches to study the diversity of ammonia-oxidizing bacteria. *J Microbiol Methods* **50**, 189–203.

**Noble, R. T. & Fuhrman, J. A. (1997).** Virus decay and its causes in coastal waters. *Appl Environ Microbiol* **63**, 77–83.

**Noune, C. & Hauxwell, C. (2015).** Complete genome sequences of *Helicoverpa armigera* single nucleopolyhedrovirus strains AC53 and H25EA1 from Australia. *Genome Announc* **3**, 1–2.

**Noune, C. & Hauxwell, C. (2016).** Complete genome sequences of seven *Helicoverpa armigera* SNPV- AC53-derived strains. *Genome Announc* **4**, 3–4.

**Novitsky, V., Wang, R., Rossenkhan, R., Moyo, S. & Essex, M. (2013).** Intra-host evolutionary rates in HIV-1C *env* and *gag* during primary infection. *Infect Genet Evol* **19**, 361–368.

**van Oers, M. M. & Vlak, J. M. (2007).** Baculovirus genomics. *Curr Drug Targets* **8**, 1051–1068.

**Okano, K., Vanarsdall, A. L. & Rohrmann, G. F. (2004).** Characterization of a baculovirus lacking the alkaline nuclease gene. *J Virol* **78**, 10650–10656.

**Okano, K., Vanarsdall, A. L., Mikhailov, V. S. & Rohrmann, G. F. (2006).** Conserved molecular systems of the *Baculoviridae*. *Virology* **344**, 77–87.

**Oksanen, A. J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., Hara, R. B. O., Simpson, G. L. & other authors**. **(2016).** Vegan: community ecology package. https://CRAN.R-project.org/package=vegan.

**Olive, D. M. & Bean, P. (1999).** Principles and applications of methods for DNA-based typing of microbial organisms. *J Clin Microbiol* **37**, 1661–1669.

**Oyola, S. O., Otto, T. D., Gu, Y., Maslen, G., Manske, M., Campino, S., Turner, D. J., MacInnis, B., Kwiatkowski, D. P. & other authors**. **(2012).** Optimizing illumina next-generation sequencing library preparation for extremely at-biased genomes. *BMC Genomics* **13**, 1.

**Pages, H., Aboyoun, P., Gentleman, R. & DebRoy, S. (2016).** Biostrings: String objects representing biological sequences, and matching algorithms.

**Paliy, O. & Shankar, V. (2016).** Application of multivariate statistical techniques in microbial ecology. *Mol Ecol* **25**, 1032–1057.

**Patel, C. S., Jani, J. J., Parekh, V. B., Darji, V. B. & Vaishnav, P. R. (2009a).** Geographic variations and their impact on bioefficacy amongst *Helicoverpa armigera* Nuclear Polyhedrosis Virus isolates from India. *World J Microbiol Biotechnol* **26**, 783–794.

**Patel, C. S., Jani, J. J., Parekh, V. B., Darji, V. B. & Vaishnav, P. R. (2009b).** Genetic diversity and differentiation of *Helicoverpa armigera* nuclear polyhedrosis virus isolates from India. *Phytoparasitica* **37**, 407–413.

**Pearson, M. N., Groten, C. & Rohrmann, G. F. (2000).** Identification of the *Lymantria dispar* nucleopolyhedrovirus envelope fusion protein provides evidence for a phylogenetic division of the *Baculoviridae*. *J Virol* **74**, 6126–6131.

**Pearson, M. N. & Rohrmann, G. F. (2002).** Transfer, incorporation, and substitution of envelope fusion proteins among members of the *Baculoviridae*, *Orthomyxoviridae*, and *Metaviridae* (insect retrovirus) families. *J Virol* **76**, 5301–5304.

**Peng, K., Wu, M., Deng, F., Song, J., Dong, C., Wang, H. & Hu, Z. (2010).** Identification of protein-protein interactions of the occlusion-derived virus-associated proteins of *Helicoverpa armigera* nucleopolyhedrovirus. *J Gen Virol* **91**, 659–670.

**Peng, K., van Lent, J. W. M., Boeren, S., Fang, M., Theilmann, D. A., Erlandson, M. A., Vlak, J. M. & van Oers, M. M. (2012).** Characterization of novel components of the baculovirus per os infectivity factor complex. *J Virol* **86**, 4981–4988.

**Petrik, D. T., Iseli, A., Montelone, B. A., Van Etten, J. L. & Clem, R. J. (2003).** Improving baculovirus resistance to UV inactivation: increased virulence resulting from expression of a DNA repair enzyme. *J Invertebr Pathol* **82**, 50–56.

**Pfeifer, G. P., You, Y.-H. & Besaratinia, A. (2005).** Mutations induced by ultraviolet light. *Mutat Res* **571**, 19–31.

**Pinto, A. J. & Raskin, L. (2012).** PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. *PLoS One* **7**, e43093.

**Polz, M. F. & Cavanaugh, C. M. (1998).** Bias in template-to-product ratios in multitemplate PCR. *Appl Environ Microbiol* **64**, 3724–3730.

**Potter, K. N., Faulkner, P. & MacKinnon, E. A. (1976).** Strain selection during serial passage of *Trichoplusia ni* nuclear polyhedrosis virus. *J Virol* **18**, 1040–1050.

**Prober, J. M., Trainor, G. L., Dam, R. J., Hobbs, F. W., Robertson, C. W., Zagursky, R. J., Cocuzza, A. J., Jensen, M. A. & Baumeister, K. (1987).** A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science (80- )* **238**, 336–341.

**Pylro, V. S., Morais, D. K., Kalks, K. H. M., Roesch, L. F. W., Hirsch, P. R., Tótola, M. R. & Yotoko, K. (2016).** Misguided phylogenetic comparisons using DGGE excised bands may contaminate public sequence databases. *J Microbiol Methods* **126**, 18–23.

**Qiu, X., Wu, L., Huang, H., McDonel, P. E., Palumbo, A. V., Tiedje, J. M. & Zhou, J. (2001).** Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. *Appl Environ Microbiol* **67**, 880–887.

**Quail, M. A., Otto, T. D., Gu, Y., Harris, S. R., Skelly, T. F., McQuillan, J. A., Swerdlow, H. P. & Oyola, S. O. (2012).** Optimal enzymes for amplifying sequencing libraries. *Nat Methods* **9**, 10–11.

**Quartieri, A., Simone, M., Gozzoli, C., Popovic, M., D'Auria, G., Amaretti, A., Raimondi, S. & Rossi, M. (2016).** Comparison of culture-dependent and independent approaches to characterize fecal bifidobacteria and lactobacilli. *Anaerobe* **38**, 130–137.

**R Core Team**. **(2016).** *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

**Radford, A. D., Chapman, D., Dixon, L., Chantrey, J., Darby, A. C. & Hall, N. (2012).** Application of next-generation sequencing technologies in virology. *J Gen Virol* **93**, 1853–1868.

**Rambaut, A., Posada, D., Crandall, K. a & Holmes, E. C. (2004).** The causes and consequences of HIV evolution. *Nat Rev Genet* **5**, 52–61.

**Rass, K. & Reichrath, J. (2008).** UV damage and DNA repair in malignant melanoma and nonmelanoma skin cancer. In *Sunlight, Vitam D Ski cancer*, pp. 162–78. Springer New York.

**Ravanat, J. L., Douki, T. & Cadet, J. (2001).** Direct and indirect effects of UV radiation on DNA and its components. *J Photochem Photobiol B* **63**, 88–102.

**Redman, E. M., Wilson, K., Grzywacz, D. & Cory, J. S. (2010).** High levels of genetic diversity in *Spodoptera exempta* NPV from Tanzania. *J Invertebr Pathol* **105**, 190–193.

**Reeson, A. F., Jankovic, T., Kasper, M. L., Rogers, S. & Austin, A. D. (2003).** Application of 16S rDNA-DGGE to examine the microbial ecology associated with a social wasp *Vespula germanica*. *Insect Mol Biol* **12**, 85–91.

**Rohrmann, G. F. (2013).** *Baculovirus Molecular Biology*, 3rd edn. Bethesda MD: National Library of Medicine (US), National Centre for Biotechnology Information.

**Rohrmann, G. F. (2014).** Baculovirus nucleocapsid aggregation (MNPV vs SNPV): an evolutionary strategy, or a product of replication conditions? *Virus Genes* **49**, 351–357.

**Rosario, K., Iverson, S. L., Henderson, D. a, Chartrand, S., McKeon, C., Glenn, E. P. & Maier, R. M. (2007).** Bacterial community changes during plant establishment at the San Pedro river mine tailings site. *J Environ Qual* **36**, 1249–1259.

**Rothschild, L. J. (1999).** The influence of UV radiation on protistan evolution. *J Eukaryot Microbiol* **46**, 548–555.

**Rowley, D. L., Farrar, R. R., Blackburn, M. B. & Harrison, R. L. (2010).** Genetic and biological variation among nucleopolyhedrovirus isolates from the fall armyworm, *Spodoptera frugiperda* (Lepidoptera: Noctuidae). *Virus Genes* **40**, 458–468.

**Ruano, G. & Kidd, K. K. (1992).** Modeling of heteroduplex formation during PCR from mixtures of DNA templates. *Genome Res* **2**, 112–116.

**Sagripanti, J. & Lytle, C. D. (2007).** Inactivation of influenza virus by solar radiation. *Photochem Photobiol* 1278–1282.

**Samarajeewa, A. D., Hammad, A., Masson, L., Khan, I. U. H., Scroggins, R. & Beaudette, L. A. (2015).** Comparative assessment of next-generation sequencing, denaturing gradient gel electrophoresis, clonal restriction fragment length polymorphism and cloning-sequencing as methods for characterizing commercial microbial consortia. *J Microbiol Methods* **108**, 103–111.

**Sanger, F., Nicklen, S. & Coulson, A. R. (1977).** DNA sequencing with chain - terminating inhibitors. *Proc Natl Acad Sci* **74**, 5463–5467.

**Sanjuán, R., Nebot, M. R., Chirico, N., Mansky, L. M. & Belshaw, R. (2010).** Viral mutation rates. *J Virol* **84**, 9733–9748.

**Satokari, R. M., Vaughan, E. E., Akkermans, A. D. L. & Saarela, M. (2001).** Bifidobacterial diversity in human feces detected by genus- specific PCR and denaturing gradient gel electrophoresis. *Appl Enviromental Microbiol* **67**, 504–513.

**Schroeder, D. C., Oke, J., Hall, M., Wilson, W. H. & Malin, G. (2003).** Virus succession observed during an *Emiliania huxleyi* bloom. *Appl Environ Microbiol* **69**, 2484–2490.

**Sekiguchi, H., Tomioka, N., Nakahara, T. & Uchiyama, H. (2001).** A single band does not always represent single bacterial strains in denaturing gradient gel electrophoresis analysis. *Biotechnol Lett* **23**, 1205–1208.

**Shackelton, L. a & Holmes, E. C. (2004).** The evolution of large DNA viruses: combining genomic information of viruses and their hosts. *Trends Microbiol* **12**, 458–465.

**Shannon, C. E. (1948).** A Mathematical Theory of Communication. *Bell Syst Tech J* **27**, 379–423.

**Shapiro, D. I., Fuxa, J. R., Braymer, H. D. & Pashley, D. P. (1991).** DNA restriction polymorphism in wild isolates of *Spodoptera frugiperda* nuclear polyhedrosis virus. *J Invertebr Pathol* **58**, 96–105.

**Shapiro, M. & Ignoffo, C. M. (1970).** Nucleopolyhedrosis of Heliothis: Activity of isolates from *Heliothis zea*. *J Invertebr Pathol* **16**, 107–111.

**Shapiro, M., Martignoni, M. E., Cunningham, J. C. & Goodwin, R. H. (1982).** Potential use of the saltmarsh caterpillar as a production host for nucleopolyhedrosis viruses. *J Econ Entomol* **75**, 69–71.

**Sheffield, V. C., Cox, D. R., Lerman, L. S. & Myers, R. M. (1989).** Attachment of a 40-base-pair G + C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proc Natl Acad Sci U S A* **86**, 232–236.

**Shendure, J. & Ji, H. (2008).** Next-generation DNA sequencing. *Nat Biotechnol* **26**, 1135–1145.

**Shendure, J. a., Porreca, G. J., Church, G. M., Gardner, A. F., Hendrickson, C. L., Kieleczawa, J. & Slatko, B. E. (2011).** Overview of DNA sequencing strategies. *Curr Protoc Mol Biol* 1–23.

**Short, S. M. & Suttle, C. A. (1999).** Use of the polymerase chain reaction and denaturing gradient gel electrophoresis to study diversity in natural virus communities. *Hydrobiologia* **401**, 19–32.

**Sim, S., Aw, P. P. K., Wilm, A., Teoh, G., Hue, K. D. T., Nguyen, N. M., Nagarajan, N., Simmons, C. P. & Hibberd, M. L. (2015).** Tracking dengue virus intra-host genetic diversity during human-to-mosquito transmission. *PLoS Negl Trop Dis* **9**, e0004052.

**Simón, O., Williams, T. & López Ferber, M. (2005).** Functional importance of deletion mutant genotypes in an insect nucleopolyhedrovirus population. *Appl Environ Microbiol* **71**, 4254–4262.

**Simón, O., Williams, T., Caballero, P. & López Ferber, M. (2006).** Dynamics of deletion genotypes in an experimental insect virus population. *Proc Biol Sci* **273**, 783–790.

**Simón, O., Williams, T., López Ferber, M., Taulemesse, J.-M. & Caballero, P. (2008).** Population genetic structure determines speed of kill and occlusion body production in *Spodoptera frugiperda* multiple nucleopolyhedrovirus. *Biol Control* **44**, 321–330.

**Simón, O., Palma, L., Beperet, I., Muñoz, D., López-Ferber, M., Caballero, P. & Williams, T. (2011).** Sequence comparison between three geographically distinct *Spodoptera frugiperda* multiple nucleopolyhedrovirus isolates: Detecting positively selected genes. *J Invertebr Pathol* **107**, 33–42.

**Simpson, J. M., McCracken, V. J., White, B. A., Gaskins, H. R. & Mackie, R. I. (1999).** Application of denaturant gradient gel electrophoresis for the analysis of the porcine gastrointestinal microbiota. *J Microbiol Methods* **36**, 167–179.

**Sinclair, L., Osman, O. A., Bertilsson, S. & Eiler, A. (2015).** Microbial community composition and diversity via 16S rRNA gene amplicons: Evaluating the Illumina platform. *PLoS One* **10**, e0116955.

**Sinha, R. P. & Häder, D.-P. (2002).** UV-induced DNA damage and repair: a review. *Photochem Photobiol Sci* **1**, 225–236.

**de Smidt, O. (2016).** The use of PCR-DGGE to determine bacterial fingerprints for poultry and red meat abattoir effluent. *Lett Appl Microbiol* **62**, 1–8.

**Smith, D. B., McAllister, J., Casino, C. & Simmonds, P. (1997).** Virus 'quasispecies': making a mountain out of a molehill? *J Gen Virol* **78 ( Pt 7)**, 1511–1519.

**Smith, G. E. & Summers, M. D. (1978).** Analysis of baculovirus genomes with restriction endonucleases. *Virology* **89**, 517–27.

**Smith, I. R. L. & Crook, N. E. (1988).** *In vivo* isolation of baculovirus genotpyes. *Virology* **166**, 240–244.

**Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughs, P., Dodd, C., Connell, C. R., Heines, C., Kent, S. B. H. & Hood, L. E. (1986).** Fluorescence detection in automated DNA sequence analysis. *Nature* **321**, 681.

**Smyth, R. P., Schlub, T. E., Grimm, A., Venturi, V., Chopra, A., Mallal, S., Davenport, M. P. & Mak, J. (2010).** Reducing chimera formation during PCR amplification to ensure accurate genotyping. *Gene* **469**, 45–51.

**Song, J., Wang, R., Deng, F., Wang, H. & Hu, Z. (2008).** Functional studies of *per os* infectivity factors of *Helicoverpa armigera* single nucleocapsid nucleopolyhedrovirus. *J Gen Virol* **89**, 2331–2338.

**Speksnijder, A. G. C. L., Kowalchuk, G. A., Jong, S. D. E., Kline, E., Stephen, J. R., Laanbroek, H. J., De Jong, S., Kline, E., Stephen, J. R. & Laanbroek, H. J. (2001).** Microvariation artificats introduced by PCR and cloning of closely related 16S rRNA gene sequences. *Appl Environ Microbiol* **67**, 469–472.

**Srinivasan, V., Schnitzlein, W. M. & Tripathy, N. (2001).** Fowlpox virus encodes a novel DNA repair enzyme , CPD-photolyase , that restores infectivity of UV light-damaged virus. *J Virol* **75**, 1681–1688.

**Sun, X. X., Van Der Werf, W., Vlak, J. M. & Hu, Z. (2004a).** Field inactivation of wild-type and genetically modified *Helicoverpa armigera* single nucleocapsid nucleopolyhedrovirus in cotton. *Biocontrol Sci Technol* **14**, 185–192.

**Sun, X., Wang, H., Sun, X., Chen, X., Peng, C., Pan, D., Jehle, J. a, van der Werf, W., Vlak, J. M. & Hu, Z. (2004b).** Biological activity and field efficacy of a genetically modified *Helicoverpa armigera* single-nucleocapsid nucleopolyhedrovirus expressing an insect-selective toxin from a chimeric promoter. *Biol Control* **29**, 124–137.

**Szewczyk, B., Hoyos-Carvajal, L., Paluszek, M., Skrzecz, I. & Lobo de Souza, M. (2005).** Baculoviruses - re-emerging biopesticides. *Biotechnol Adv* **24**, 143–160.

**Taylor, D. L., Walters, W. A., Lennon, N. J., Bochicchio, J., Krohn, A., Caporaso, J. G. & Pennanen, T. (2016).** Accurate estimation of fungal diversity and abundance through improved lineage-specific primers optimized for Illumina amplicon sequencing. *Appl Environ Microbiol* **82**, 7217–7226.

**Theilmann, D. A. & Blissard, G. W. (2008).** Baculoviruses: Molecular biology of nucleopolyhedroviruses. In *Encylopedia Virol*, pp. 254–265. Oxford: Elsevier Academic Press.

**Thézé, J., Cabodevilla, O., Palma, L., Williams, T., Caballero, P. & Herniou, E. A. (2014).** Genomic diversity in European *Spodoptera exigua* multiple nucleopolyhedrovirus isolates. *J Gen Virol* **95**, 2297–2309.

**Thompson, C. G. & Scott, D. W. (1979).** Production and persistence of the nuclear polyhedrosis virus of the Douglas-fir tussock moth, *Orgyia pseudotsugata* (Lepidoptera: Lymantriidae), in the forest ecosystem. *J Invertebr Pathol* **33**, 57–65.

**Thompson, J. R., Marcelino, L. A. & Polz, M. F. (2002).** Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. *Nucleic Acids Res* **30**, 2083–2088.

**Tompkins, G. J., Vaughn, J. L., Adams, J. R. & Reichelderfer, C. F. (1981).** Effects of propogating *Autographa californica* nuclear polyhedrosis virus and its *Trichoplusia ni* variant in divergent hosts. *J Econ Entomol* **10**, 801–806.

**Tompkins, G. J., Dougherty, E. M., Adams, J. R. & Diggs, D. (1988).** Changes in the virulence of nuclear polyhedrosis viruses when propagated in alternate noctuid (Lepidoptera: Noctuidae) cell lines and hosts. *J Econ Entomol* **81**, 1027–1032.

**Vanarsdall, A. L., Okano, K. & Rohrmann, G. F. (2005).** Characterization of the replication of a baculovirus mutant lacking the DNA polymerase gene. *Virology* **331**, 175–180.

**Vanarsdall, A. L., Mikhailov, V. S. & Rohrmann, G. F. (2007).** Baculovirus DNA replication and processing. *Curr Drug Targets* **8**, 1096–1102.

**Vignuzzi, M., Stone, J. K., Arnold, J. J., Cameron, C. E. & Andino, R. (2006).** Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* **439**, 344–348.

**Wang, C., García-Fernández, D., Mas, A. & Esteve-Zarzoso, B. (2015).** Fungal diversity in grape must and wine fermentation assessed by massive sequencing, quantitative PCR and DGGE. *Front Microbiol* **6**, 1–8.

**Wang, G. C. Y. & Wang, Y. (1996).** The frequency of chimeric molecules as a consequence of PCR co- amplification of 16S rRNA genes from different bacterial species. *Appl Environ Microbiol* **142**, 1107–1114.

**Wang, H., Wu, D., Deng, F., Peng, H., Chen, X., Lauzon, H., Arif, B. M., Jehle, J. A. & Hu, Z. (2004).** Characterization and phylogenetic analysis of the chitinase gene from the *Helicoverpa armigera* single nucleocapsid nucleopolyhedrovirus. *Virus Res* **100**, 179–189.

**Wang, J., Zhu, Z., Zhang, L., Hou, D., Wang, M., Arif, B., Kou, Z., Wang, H., Deng, F. & Hu, Z. (2016).** Genome sequencing and analysis of *Catopsilia pomona* nucleopolyhedrovirus: a distinct species in group I Alphabaculovirus. *PLoS One* **11**, e0155134.

**Weber, T. P. & Stilianakis, N. I. (2008).** Inactivation of influenza A viruses in the environment and modes of transmission: a critical review. *J Infect* **57**, 361–73. Elsevier Ltd.

**Weitzman, M. D., Possee, R. D. & King, L. A. (1992).** Characterization of two variants of *Panolis flammea* multiple nucleocapsid nuclear polyhedrosis virus. *J Gen Virol* **73**, 1881–1886.

**White, S. M., Burden, J. P., Maini, P. K. & Hails, R. S. (2012).** Modelling the within-host growth of viral infections in insects. *J Theor Biol* **312**, 34–43.

**Wickham, H. (2009).** *ggplot2: elegant graphics for data analysis*. Springer Science & Business Media.

**Witt, D. J. (1984).** Photoreactivation and ultraviolet-enhanced reactivation of ultraviolet-irradiated nuclear polyhedrosis virus by insect cells. *Arch Virol* **79**, 95–107.

**Wu, G. A., Jun, S.-R., Sims, G. E. & Kim, S.-H. (2009).** Whole-proteome phylogeny of large dsDNA virus families by an alignment-free method. *Proc Natl Acad Sci* **106**, 12826–12831.

**Xi, Q., Wang, J., Deng, R. & Wang, X. (2007).** Characterization of AcMNPV with a deletion of *me53* gene. *Virus Genes* **34**, 223–232.

**Xu, X.-L., Cheng, T.-Y., Yang, H. & Yan, F. (2015).** Identification of intestinal bacterial flora in *Rhipicephalus microplus* ticks by conventional methods and PCR–DGGE analysis. *Exp Appl Acarol* **66**, 257–268.

**Xu, Y.-P., Cheng, R.-L., Xi, Y. & Zhang, C.-X. (2013).** Genomic diversity of *Bombyx mori* nucleopolyhedrovirus strains. *Genomics* **102**, 63–71.

**Yearian, W. C. & Young, S. (1974).** Persistance of Heliothis nucleopolyhedrosis virus on cotton plant parts. *Environ Entomol* **3**, 1035–1036.

**Young, S. & Yearian, W. C. (1974).** Persistance of Heliothis NPV on the foliage of cotton, soybean and tomato. *Environ Entomol* **3**, 253–255.

**Yu, L., Zhang, W., Liu, L. & Yang, J. (2015).** Determining microeukaryotic plankton community around Xiamen Island, Southeast China, using Illumina MiSeq and PCR-DGGE techniques. *PLoS One* **10**, e0127721.

**Zanotto, P. M. A., Kessing, B. D. & Maruniak, J. E. (1993).** Phylogenetic Interrelationships among Baculoviruses: Evolutionary Rates and Host Associations. *J Invertebr Pathol* **62**, 147–164.

**Zhang, Y., Ma, F., Wang, Y., Yang, B. & Chen, S. (2008).** Expression of v-cath gene from HearNPV in tobacco confers an antifeedant effect against *Helicoverpa armigera*. *J Biotechnol* **138**, 52–55.

**Zhong, F., Wu, J., Dai, Y., Yang, L., Zhang, Z., Cheng, S. & Zhang, Q. (2015).** Bacterial community analysis by PCR-DGGE and 454-pyrosequencing of horizontal subsurface flow constructed wetlands with front aeration. *Appl Microbiol Biotechnol* **99**, 1499–1512.

**Zwart, M. P., Hemerik, L., Cory, J. S., de Visser, J. A. G. M., Bianchi, F. J. J. A., Van Oers, M. M., Vlak, J. M., Hoekstra, R. F. & Van der Werf, W. (2009).** An experimental test of the independent action hypothesis in virus-insect pathosystems. *Proc R Soc B Biol Sci* **276**, 2233–2242.

**Zwart, M. P., Daròs, J.-A. & Elena, S. F. (2011).** One is enough: *in vivo* effective population size is dose-dependent for a plant RNA virus. *PLoS Pathog* **7**, e1002122.