# Genetic Structure of Cydia pomonella Granulovirus Isolates and their Potential in Overcoming Resistance in Codling Moth

TECHNISCHE UNIVERSITÄT DARMSTADT

## vom Fachbereich Biologie der Technischen Universität Darmstadt

zur Erlangung des Grades
Doctor rerum naturalium
(Dr. rer. nat.)

**Dissertation**
**von Jiangbin Fan**

Erstgutachter: Prof. Dr. Johannes A. Jehle
Zweitgutachter: Prof. Dr. Gerhard Thiel

Darmstadt 2019

书山有路勤为径，学海无涯苦作舟。

Diligence is the path to the mountain of knowledge, hard-working is the boat to the endless sea of learning.

For my parents, HX. Zhang and XM. Fan

## CONTENTS

## List of Abbreviations

| | |
|---|---|
| % | percent |
| et al. | and others |
| etc. | et cetera |
| e.g. | for example |
| ng | nanogram |
| bp | base pair |
| *Bt* | *Bacillus thuringiensis* |
| BV | budded virus |
| CM | codling moth |
| CpS | susceptible *Cydia pomonella* |
| CpRR1 | type I resistance *Cydia pomonella* |
| CpR5M | type II resistance *Cydia pomonella* |
| CpRGO | type III resistance *Cydia pomonella* |
| dpi | days post infection |
| EPNs | entomopathogenic nematodes |
| EC | European Commission |
| GV | granulovirus |
| ICTV | International Committee on Taxonomy of Viruses |
| indel | insertion or deletion |
| IPM | Integrated pest management |
| JKI | Julius Kühn-Institut |
| kbp | kilo base pairs |
| L3 | third instar larva |
| $LC_{50}$ | median lethal concentration |
| ml | milliliter |
| MNPV | multiple nucleopolyhedrovirus |
| NGS | next generation sequencing |

| | |
|---|---|
| nm | nanometer |
| NPV | nucleopolyhedrovirus |
| OB | occlusion body |
| ODV | occlusion derived virus |
| ORF | open reading frame |
| pH | measure of acidity of basicity of aqueous solutions |
| PM | peritrophic membrane |
| RFLP | restriction fragment length polymorphism |
| SIR | sterile insect release |
| SIT | sterile insect technique |
| SNP | single nucleotide polymorphism |
| SNPV | single nucleopolyhedrovirus |

Viruses:

| | |
|---|---|
| AcMNPV | Autographa californica multiple nucleopolyhedrovirus |
| AdhoNPV | Adoxophyes honmai nucleopolyhedrovirus |
| AgMNPV | Anticarsia gemmatalis multiple nucleopolyhedrovirus |
| ArGV | Artogeia rapae granulovirus |
| BmNPV | Bombyx mori nucleopolyhedrovirus |
| CpGV | Cydia pomonella granulovirus |
| CrleGV | Cryptophlebia leucotreta granulovirus |
| DapuNPV | Dasychira pudibunda nucleopolyhedrovirus |
| ErelGV | Erinnyis ello granulovirus |
| HaSNPV, HearNPV, HearSNPV | Helicoverpa armigera nucleopolyhedrovirus |
| HzSNPV | Helicoverpa zea single nucleopolyhedrovirus |
| LdMNPV | Lymantria dispar multiple nucleopolyhedrovirus |
| OpbuNPV | Operophtera brumata nucleopolyhedrovirus |
| PaflNPV | Panolis flammea nucleopolyhedrovirus |

| | |
|---|---|
| PlxyGV | Plutella xylostella granulovirus |
| PhopGV | Phthorimaea operculella granulovirus |
| SeMNPV | Spodoptera exigua multiple nucleopolyhedrovirus |
| SfMNPV | Spodoptera frugiperda multiple nucleopolyhedrovirus |
| SpltNPV | Spodoptera litura nucleopolyhedrovirus |
| TnSNPV | Trichoplusia ni single nucleopolyhedrovirus |

## Summary

Cydia pomonella granulovirus (CpGV) is a highly virulent pathogen of codling moth (CM, *Cydia pomonella* L.) larvae. It has been developed to one of the most successful commercial baculovirus biocontrol agents used on hundred thousands of hectares of pome fruit production worldwide. In recent years, however, three types (I to III) of field resistance to CpGV as well as the existence of resistance-breaking CpGV isolates have been discovered, providing an ideal model for studying baculovirus-host adaptation. This thesis aims to elucidate the potential of recently collected isolates of CpGV from northwest China to infect susceptible and resistant CM colonies and to study the stability and restoration of type I resistance in laboratory rearing by mass-crossing and selection. To further explore the genotypic and biological differences of CpGV, the population structure of 20 CpGV isolates was analyzed on the basis of Illumina next generation sequence (NGS) data. The isolates included seven new Chinese CpGV isolates, termed CpGV-ZY, -JQ, -ALE, -KS1, -KS2, -ZY2 and -WW, the re-sequenced isolates CpGV-M, -S, -E2, -I12 and the Iranian isolate CpGV-I0X, as well as the active ingredients of commercial virus selections from MadexPlus, MadexMAX, MadexTOP (V15), V14, V34, V45 and Carpovirusine EVO2.

First, resistance testing or full range bioassays were conducted to determine the resistance-breaking capacity or the median lethal concentration ($LC_{50}$) of the Chinese CpGV isolates against susceptible and three resistant CM strains, representing type I to III resistance (Chapter 2). The isolates were further screened for the presence of the additional 2×12 bp repeat insertion in CpGV gene *pe38* (ORF24), which had been proposed to be the target of type I resistance in the Mexican isolate CpGV-M. It was found that the isolates CpGV-JQ, -KS1 and -ZY2 could break type I resistance, though a distinct delay was observed in the infection process. The isolates followed the previously established "*pe38* model" of resistance-breaking, except CpGV-WW, which lacked the 2×12 bp repeat involved in resistance-breaking but failed to overcome type I resistance. However, CpGV-WW was able to overcome type II and type III resistance. Correlation of bioassay results and the isolates' *pe38* repeat structure were in agreement with the potential role of *pe38* as the major target for resistance in CpRR1, except for CpGV-WW.

Second, resistance tests with CpGV-M revealed a certain decline of the resistance level of CM strain CpRR1, expressing type I resistance, after it had been reared for several years without virus pressure (Chapter 3). Therefore, two newly selected lines, CpRR1_F5 and CpRR1_F7, were established by mass crossing experiments combined with virus selection on CpGV-M. Resistance level of the newly selected lines was determined by full range bioassays. The successful selection process resulted in a 15- to 160-fold increase of the $LC_{50}$ of CpRR1_F5 resistance compared to CpRR1, suggesting that the rearing in absence of virus selection was most likely the main factor involved the observed resistance decline of CpRR1. Additionally, some fitness costs of fecundity were recorded in the re-selected CpRR1_F5. Single-pair crossing of CpRR1_F5 and CpRR1-F7 with susceptible CM, followed by a resistance testing with a discriminating concentration of CpGV-M occlusion bodies, revealed a dominant but not fully sex-linked inheritance arguing for a partial change of previous genetic traits in CpRR1.

Third, in Chapter 4 the genomic difference among seven new Chinese CpGV isolates could provide some answers for the virulence difference observed in bioassays. After Illumina NGS sequencing, the genome annotation and phylogenetic analyses of these isolates indicated that the genomes were highly conserved and related to known CpGV isolates, despite a considerable geographic distance. However, two new phylogenetic lineages, termed genome groups F (CpGV-JQ and -ZY2) and G (CpGV-ALE), were proposed in addition to previous phylogenetic genome groups A to E. The genetic composition of the isolates was further quantified on the basis of previously identified genome group specific single nucleotide polymorphisms (SNPs). In addition of 223 new SNP positions out of total 563 SNPs were detected against CpGV-M reference sequence, which represented virus characteristics of Chinese isolates. Whereas CpGV-WW was proposed to be genetically highly homogeneous, belonging to genome group E, the other six isolates were mixtures of at least two genotypes. Thereof CpGV-ZY, -KS1 and -KS2 were highly similar and were composed of variable ratios of genome group A (CpGV-M) and genome group E (CpGV-WW). Detailed quantification of the 12 bp repeat unit of *pe38* corresponded to the results obtained from PCR and Sanger sequence analyses (Chapter 2).

Fourth, to achieve a fully comprehensive perspective of CpGVs of different origin, 20 CpGV isolates, including twelve natural isolates from different geographic locations and eight selected CpGV strains, were analyzed together for the distribution and frequency of single nucleotide polymorphisms (SNPs) in NGS genome data and for the abundance of the 12 bp repeat unit in *pe38* (Chapter 5). The results indicated that CpGV-M, -WW, -S and MadexPlus were genetically highly homogenous isolates with a low rate of polymorphisms, while other isolates were composed of two or more genome groups at different ratios. Based on hierarchical clustering on principal components (HCPC) analysis, six distinct clusters were proposed, which represents the previously proposed main phylogenetic lineages, though the insertions and deletions were not included in cluster analysis. Relative location of different isolates in HCPC further reflected the ratio of variable compositions of different genome groups. For the quantification of the proportions of 1-5×12 bp repeat units in the different CpGV isolates a “read counting” method was developed and showed a high diversity and less conserved characteristics in *pe38* than literature reported before. The established methods for SNP quantification and HCPC analysis provide novel tools to decipher the molecular complexity of genome mixtures in virus isolates, thus depicting the population structure of baculovirus isolates in a more adequate form than genome consensus based analyses.

In summary, the results in this thesis showed that resistance loss in CpRR1 is developing in laboratory under continuous rearing without virus pressure. Newly discovered CpGV isolates exhibited high potential for control of known types of field resistance of CM. The established methods to determine positional SNP distribution can be easily extended to other (baculo)viruses to assess isolate composition and genetic diversity and to study quality and stability of virus mixtures during propagation. It can be further applied to determine its potential for control of resistant CM on molecular level, since CpGV isolates with the similar virulence patterns were found to be grouped together considering their spatial location in factor map of HCPC. Understanding CpGV population structure and the genetic adaption between baculovirus and host insect give a crucial blueprint to improve current strategies of CpGV resistance management in the field.

## Zusammenfassung

Das Cydia pomonella granulovirus (CpGV) ist ein hoch virulentes Pathogen der Larven des Apfelwicklers (= Obstmade) (CM, *Cydia pomonella* L.) und ist eines der erfolgreichsten kommerziellen Baculovirus-basierten biologischen Pflanzenschutzmittel; es wird weltweit auf mehreren hunderttausend Hektar im Kernobstanbau eingesetzt. In den letzten Jahren wurden drei Resistenztypen (I bis III) in Freilandpopulationen des Apfelwicklers gegenüber CpGV beobachtet, andererseits wurden aber auch Resistenz-brechende CpGV-Isolate entdeckt. Das System CpGV-Apfelwickler stellt somit ein ideales Model dar, die Interaktion und genetische Adaptation von Baculoviren und ihren Wirten zu untersuchen.

Das Ziel dieser Dissertation ist die Evaluierung des Potentials kürzlich entdeckter CpGV-Isolate aus dem Nordwesten Chinas anfällige und resistente Apfelwicklerlinien zu infizieren, und die Stabilität und Wiederherstellung des Reistenzttyp I des Apfelwicklerstammes CpRR1 mittels Massenkreuzungen und viraler Selektion zu untersuchen. Um die genotypischen und biologischen Unterschiede des CpGV näher zu bestimmen, wurde die Populationsstruktur von 20 CpGV-Isolaten mittels Illumina next generation sequenzierung (NGS) charakterisiert. Diese umfassen sieben neue chinesische CpGV-Isolate, die als CpGV-ZY, -JQ, -ALE, -KS1, -KS2, -ZY2 und -WW bezeichnet wurden, die re-sequenzierten Isolate CpGV-M, -S, -E2, -I12 und das Iranische Isolate CpGV-I0X, sowie die Wirkstoffe der kommerziellen Selektionen MadexPlus, MadexMAX, MadexTOP (V15), V14, V34, V45 und Carpovirusine EVO2.

Zunächst wurden Resistenztests oder Bioassays durchgeführt, um die Resistenz-brechenden Eigenschaftenoder die mittlere letale Konzentration ($LC_{50}$) der sieben chinesischen CpGV-Isolate gegenüber anfälligen und resistenten Apfelwickler-Stämmen des Resistenztyps I-III zu bestimmen (Kapitel 2). Die Isolate wurden weiterhin auf das Vorhandensein der zusätzlichen 2×12 bp-Wiederholungsinsertion in CpGV-Gen *pe38* (ORF24), das als Zielort des Resistenztyps I in CpGV-M gilt, mittels PCR und Sanger-Sequenzierung überprüft. Es wurde festgestellt, dass die Isolate CpGV-JQ, -KS1 und -ZY2 den Resistenztyp I brechen, obgleich eine deutliche Verzögerung im Infektionsverlauf zu beobachten war. Alle Isolate folgten dem früher etablierten „*pe38*-Modell“ der Resistenzbrechung, mit Ausnahme von CpGV-WW. Dieses Isolat beherbergte zwar den genetischen Faktor zur Resistenzbrechung, zeigte aber eine geringe Virulenz gegenüber CpRR1 (Resistenztyp I), andererseits jedoch eine hohe Aktivität gegenüber Apfelwicklerstämmen des Resistenztyps II und III. Der Zusammenhang der Ergebnisse aus den Infektionsversuchen und die Kartierung der 12 bp-Wiederholungen in den einzelnen Isolaten stützen ganz überwiegend die frühere Hypothese, dass *pe38* das Hauptziel des Resistenztyps I ist.

Zweitens, Resistenztests mit CpGV-M zeigten eine gewisse Abnahme des Resistenzniveaus des Apfelwickler-Stammes CpRR1 (Resistenztyps I), nachdem dieser Stamm mehrere Jahre ohne Virusdruck gehalten worden war (Kapitel 3). Daher wurden zwei neu selektierte Apfelwicklerlinien, CpRR1_F5 und CpRR1_F7, durch Massenkreuzungen in Verbindung mit einer Selektion durch CpGV-M etabliert. Das Resistenzniveau der neu ausgewählten Linien

wurde mittels Bestimmung der mittleren letalen Konzentration ($LC_{50}$) ermittelt. Das erfolgreiche Selektionsverfahren führte zu einem 15- bis 160-fachen Anstieg des $LC_{50}$-Wertes von CpRR1_F5 gegenüber CpRR1, was darauf hindeutete, dass die mehrjährige Zucht ohne Virusselektion die Hauptursache gewesen sein könnte, die zu dem beobachteten Resistenzverlust von CpRR1 geführt hatte. Zudem wurden Fitnesskosten in Bezug aud die Fertilität der Linie CpRR1_F5 erfasst. Einzelpaarkreuzungen von CpRR1_F5 und CpRR1_F7 mit einer anfälligem Apfelwicklerlinie, gefolgt von einem Resistenztest mit einer diskriminierenden Konzentration von CpGV-M Einschlusskörpern, ergab eine dominante, aber nicht vollständig geschlechtsgebundene Vererbung der Resistenz, die auf eine partielle Veränderung des ursprünglichen genetischen Merkmals von CpRR1 hinweist.

Da die genomischen Unterschiede zwischen den sieben Chinesischen CpGV-Isolaten mit ihrer biologischen Aktivitäten korreliert sein könnten, wurden deren Genome mittels Illumina NGS sequenziert (Kapitel 4). Die Genomannotation und die phylogenetischen Untersuchungen zeigten, dass deren Genome trotz der beträchtlichen geographischen Entfernung zu anderen Herkunftsorten des CpGV stark konserviert waren und Verwandtschaften mit anderen bekannten CpGV-Isolaten aufwiesen. Zusätzlich zu den bereits früher definierten phylogenetischen CpGV-Genomgruppen A bis E wurden zwei weitere Genomgruppen F (CpGV-JQ und -ZY2) und G (CpGV-ALE) gefunden. Die genetische Zusammensetzung der sieben CpGV-Isolate wurde anhand früher identifizierter Genomgruppen-spezifischer Einzelnukleotidpolymorphismen (single nucleotide polymorphisms, SNPs) quantifiziert. Im Vergleich zum bisherigen Kenntnisstand der genetischen Diversität des CpGV wurden in den neuen CpGV-Isolaten 223 neue SNP-Positionen von insgesamt 563 SNPs gegenüber der Referenzsequenz des CpGV-M nachgewiesen, die Sequenzsignaturen chinesischer CpGV-Isolate darstellten. Basierend auf der SNP-Verteilung wurde das Isolat CpGV-WW als ein sehr homogenes Isolat der Genomgruppe E klassifiziert, während die sechs anderen Isolate Mischungen von mindestens zwei Genotypen darstellten: die Isolate CpGV-ZY, -KS1 und -KS2 waren sehr ähnlich zu einander und bestanden aus einem variablen Verhältnis von Isolaten der Genomgruppe A (CpGV-M) und der Genomgruppe E (CpGV-WW). Eine detaillierte Quantifizierung der 12-bp-Wiederholungseinheit im Gen *pe38* entsprach den mittels PCR und Sanger-Sequenzierung erhaltenen Ergebnissen aus Kapitel 2.

Um einen umfassenden Blick auf die genetische Diversität von 20 CpGV-Isolaten verschiedener Herkünfte, einschließlich zwölf geografischer Isolate und acht selektierten CpGV-Stämmen, zu werfen, wurden die Verteilung und die Häufigkeit von SNPs und das Vorkommen der 12 bp-Wiederholung im Gen *pe38* untersucht (Kapitel 5). Die Ergebnisse zeigten, dass CpGV-M, -WW, -S und MadexPlus genetisch sehr homogene Isolate mit einer sehr niedrigen Rate an Polymorphismen waren, während andere Isolate Mischungen aus zwei oder mehr Genomgruppen mit unterschiedlichen Verhältnissen zusammengeordnet werden konnten. Basierend auf einer Hauptkomponentenanalyse und hierarchischen Clusterung (HCPC) der Häufigkeitsverteilungen aller SNPs wurden die 20 CpGV-Genome in sechs verschiedene Cluster eingeordnet, die den früher vorgeschlagenen phylogenetischen Hauptlinien des CpGV entsprechen. Insertionen und Deletionen wurden bei der Clusterung nicht berücksichtigt. Die relative Lage des verschiedenen Isolate innerhalb der HCPC-Analyse spiegelte ferner das Verhältnis der variablen Zusammensetzungen verschiedener Genomgruppen wider. Die Messung des Anteils der 1-5×12 bp-Wiederholungseinheiten des *pe38* in den verschiedenen Isolaten wurde unter Verwendung einer „Read-Counting“-Methode optimiert und zeigte eine sehr hohe Variabilität dieser Wiederholungsstruktur innerhalb der

einzelnen CpGV-Isolate. Die etablierten Methoden der SNP-Quantifizierung und HCPC-Analyse bieten neuartige Werkzeuge zur Entschlüsselung der molekularen Komplexität von Genomgemischen in Virusisolaten, wodurch die Populationsstruktur von Baculovirus-Isolaten in einer angemesseneren Form als durch Analysen, die auf Konsensussequenzen der Genome beruhen, dargestellt werden kann.

Zusammenfassend haben die Ergebnisse dieser Arbeit gezeigt, dass sich der Resistenzverlust von CpRR1 im Labor durch dauerhafte Zucht ohne Virusdruck entwickelte und nur teilweise re-selektiert werden konnte. Neu entdeckte CpGV-Isolate aus China sind potenziell für die Bekämpfung von Apfelwicklerpopulationen mit CpGV-Resistenz geeignet. Die Untersuchung der genetischen Diversität und der Populationsstruktur CpGV-Isolaten mit unterschiedlicher Herkunft bieten nicht nur einen vertieften und detaillierten Einblick in die molekular Diversität des CpGV in Korrelation mit seinen Virulenzeigenschaften, sowie einen wichtigen Beitrag zum Verständnis der gegenseitigen Anpassung zwischen Baculovirus und ihren Wirtsinsekten, sondern liefern auch entsprechende genetische Informationen um die derzeitigen Strategien des Resistenzmanagements bei CpGV weiterzuentwickeln.

## Aim of this thesis

In recent years, three types of CpGV field resistance as well as the existence of resistance-breaking CpGV isolates have been discovered, applied and studied. This system provides an ideal model for studying baculovirus-host adaptation. This thesis aims to elucidate the molecular mechanism of the resistance-breaking in CpGV and codling moth by addressing the following questions:

1. What resistance-breaking capacities have recently discovered CpGV isolates from China?
2. How is the resistance level and marker location in CpRR1 after long-term laboratory rearing without virus pressure?
3. How is the genetic composition of the Chinese CpGV isolates related to their virulence in resistant CM?
4. What are the in-depth genetic composition and changes of CpGVs isolated from nature and selected in the laboratory?

In Chapter 2, resistance testing and the median lethal concentration ($LC_{50}$) assays against susceptible and resistant CM strains were conducted to determine the virulence of seven new CpGV isolates derived from northwest China. The 2×12 bp insertion in CpGV gene *pe38* (ORF24) was identified as well. The correlation between resistance-breaking and this insertion was observed in CpGV-JQ, -KS1 and -ZY2 follow the model of *pe38* type I resistance-breaking. CpGV-WW failed to break type I resistance, though the lack of 2×12 bp insertion. However CpGV-WW was able to overcome type II and type III resistance. Combining the bioassay and Sanger sequencing data, it is pointed out that *pe38* is the major target for resistance in CpRR1.

In Chapter 3, five and seven generations of selection of CpRR1 against CpGV-M were performed to establish new CpRR1_F5 and CpRR1_F7 lines. Resistance ratio of these selected lines was determined by full range bioassays and single-pair crossings with susceptible CM individuals followed by a resistance test using a discriminating concentration of CpGV-M occlusion bodies. Regain of high level of resistance in CpRR1_F5 and CpRR1_F7 indicated the absence of virus pressure for long-term under the laboratory condition could result in resistance decrease. Single-pair crossing with susceptible codling moth revealed a dominant but not fully sex-linked inheritance arguing for a partial loss of previous resistance traits in CpRR1.

In Chapter 4, the phylogeny and genetic composition of seven Chinese CpGV isolates were analyzed by consensus sequence and detected SNPs based on the NGS data. Both results concluded two new genome groups F and G represented by CpGV-JQ and -ZY2 and -ALE. The previous genome group A, D and E were also identified in Chinese CpGV populations. Combining the genome group specific SNPs and SNP distribution, it is proposed that only CpGV-WW was genetically highly homogeneous isolate, other six isolates of CpGV-ZY, -JQ, -ALE, -KS1, -KS2 and -ZY2 were mixtures of different ratios of at least two genotypes.

In Chapter 5, isolate/genome group specific SNPs obtained from *de novo* pathway were used to quantify the genetic structure of NGS data of 20 CpGV isolates, twelve geographic isolates and eight selected ones. Four homogeneous CpGV-M, -S, -WW and MadexPlus were identified with low variation in their genomes, whereas other isolates were mixtures or heterogeneous. These CpGV isolates can be classified into six clusters corresponding to the genotype composition on the basis of SNP location and frequency using HCPC method. The spatial location in factor map of HCPC also reflects the different ratio of genome group in a given isolate. A read counting method was conducted to measure the overall variants of 12 bp repeat unit in *pe38*.

# Chapter I: General introduction

Codling moth (CM, *Cydia pomonella* L.) is a cosmopolitan insect pest of commercial apple and pear production. To reduce economic damage caused by it, Cydia pomonella granulovirus (CpGV) has been extensively used as biological control agent worldwide. After at least one decade of successful CpGV application, three different types of CpGV resistance have been discovered in European CM field populations (Asser-Kaiser et al., 2007; Jehle et al., 2017; Sauer et al., 2017a; Sauer et al., 2017b), mainly in organic orchards. On the other hand, resistance-breaking CpGV isolates have been identified and are meanwhile used to control CM (Eberle et al., 2008; Berling et al., 2009a; Zingg et al., 2011; Gueli Alletti et al., 2017; Sauer et al., 2017b). Elucidating the different mechanisms of CpGV resistance, modes of inheritance of resistance as well as the capacity of resistance-breaking CpGV isolates to overcome different types of resistance is a major scientific challenge to adapt application of CpGV as well as to develop suitable resistance management strategies in practice.

In China, CM is an invasive species, which have spread from two directions of northwest along Hexi Corridor or northeast to major apple production areas, threatening the development of local apple production (Zhang et al., 2012). CpGV has therefore a great potential to be used for CM control in China. The use of CpGV isolates indigenous to China and avoiding development of CpGV resistance are two major issue of importance for further development of CM control in China.

## Occurrence and distribution of codling moth

*C. pomonella*, (Synonym: *Grapholitha pomonella* L., *Carpocapsa pomonella* L., *Laspeyresia pomonella*), is a species of family Tortricidae of the order Lepidoptera, which larvae cause severe damage to pome fruit (apple, pear, quince) production; also stone fruits (apricot, plum and peaches) as well as walnuts can be affected by CM (Barnes, 1991). CM neonates bore into premature apple and develop into third instars; after that some of them can transfer to adjacent apples or continue to grow up until the last (fifth) larval stage (Figure 1-1). Mature larvae leave the fruit to either pupate for next generation in the same year or become over-wintering diapausing larvae. All pre-pupation larvae hide inside cocoons under the bark of the tree (Barnes, 1991). For diapausing larvae temperature is the sole factor triggering begin of pupation and eclosion in spring. Once the degree-day accumulation reaches around 222.2 °C, and the lower environmental temperature is higher than 11.1 °C, overwintering larvae start to pupate and soon emerge into adults (Pickel et al., 1986). After 7-15 days, eggs can hatch and neonate larvae infest the fruit. For the development from hatching to fully grown larva, the caterpillar needs 3-4 weeks. In general, one to four generations occur in apple growing regions (Pringle et al., 2003; Odendaal et al., 2015), depending on climate conditions. A female CM moth can lay 30-60 eggs. Egg deposition is on the surface of leaves or directly on fruits; distribution of eggs is more in the north and east part of the apple trees than in other directions (Wearing, 2016). In south Germany, generally two generations occur, first of which start in early May, whereas in north Germany only one generation exists (Steineke and Jehle, 2002). In northwest China, where *C. pomonella* is

considered as an invasive species, adults emerge between late April and early May resulting up to three to four generations per year (Zhang et al., 2012).

The Caucasus region is considered to be the native place of CM (Rezapanah et al., 2002). From there it spread to other regions supported by human migration and transport, as well as via transportation of planting material, infested fruits and packaging material. Nowadays, CM is distributed in temperate zones of the north and south hemisphere, across all main apple growing areas worldwide (CABI, 2018); it is present in Europe, Mediterranean countries, Middle Asia, North America, Argentina, Eastern Australia, New Zealand, South Africa etc.; it was also localized in Brazil, India, China etc. (Figure 1-2). However, in areas where CM has not yet established, very strict quarantine regulations are in force to prevent further pest invasion.

**Figure 1-1**. Damage of apples caused by codling moth larvae.

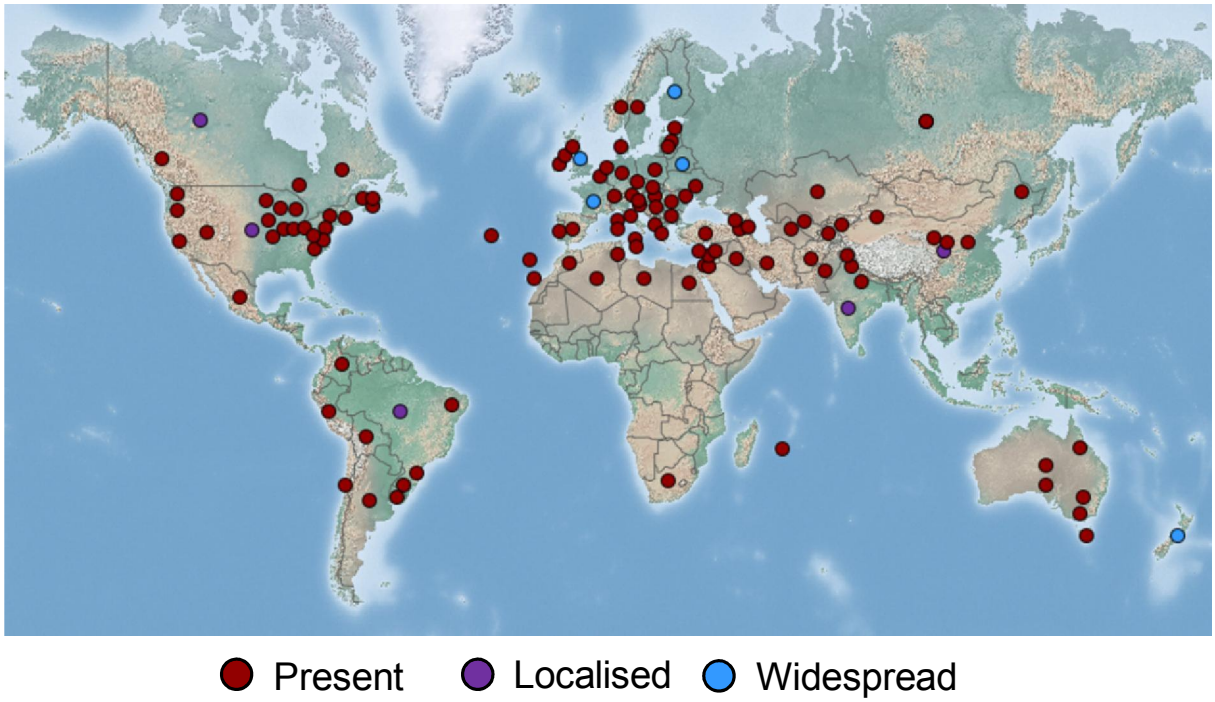



**Figure 1-2**. Cosmopolitan occurrence of codling moth (CM).

## Biological control tools for codling moth

Integrated pest management (IPM) as defined by the European Commission (EC) is the "careful consideration of all available plant protection methods and subsequent integration of appropriate measures that discourage the development of populations of harmful organisms and keep the use of plant protection products and other forms of intervention to levels that are economically and ecologically justified and reduce or minimize risks to human health and the environment" (EC, 2009). IPM emphasizes the growth of a healthy crop with the least possible disruption to agro-ecosystems and encourages natural pest control mechanisms, e.g. preferential application of biological, physical and other non-chemical methods (Ehler, 2006).

Biological control (or biocontrol) is an important component for IPM (EC, 2009); it is considered as "the use of living organisms to suppress the population density or impact of a specific pest organism, making it less abundant or less damaging than it would otherwise be" (Eilenberg et al., 2001), which underlines the use the living organisms, including viruses.

Biological control tools can be applied in organic and integrated production systems. For control of CM a series of methods has been developed, e.g. sterile insect technique (SIT), augmentative release of parasitoids and application of *Bacillus thuringiensis* (Bt), entomopathogenic fungi, entomopathogenic nematodes and CpGV.

SIT is a biological control tool that aims to eradicate targeted pest by the release of many sterile males, which have been sterilized via exposure to ionizing radiation or chemicals or genetic modification (Robinson, 2005). Therein, ionizing radiation is extensively applied in large-scale SIT programmes (Robinson, 2005). SIT was first practiced to reduce CM field populations in 1994 in Okanagan Valley and Similkameen Valley in Southern British Columbia, Canada. After implementing an area-wide Sterile Inset Release (SIR) program, local CM populations decreased. Later on, SIT control of CM has been also applied in South Africa and New Zealand, however it presented several challenges threatening the continuation of the program, such as economic costs and efficacy at high population densities (Barnes et al., 2015; Horner et al., 2016).

Parasitoid wasps of *Trichogramma* sp. and *Mastrus ridens* parasitize eggs and larvae of CM, respectively; they have been used in the field with limited success in America, Australia, New Zealand and Europe (Mills et al., 2000; Sandanayaka et al., 2011; Lefoe et al., 2013). Other parasitoids, such as *Ascogaster quadridentata*, *Pristomerus vulnerator* and *Perilampus tristis,* were investigated for possible application as well (Maalouly et al., 2015).

Although *B. thuringiensis* strains have been shown to be highly pathogenic to CM larvae in laboratory tests, control efficacy in orchards seems to be low, due to the reason that CM neonates cannot ingest a sufficient lethal dose before boring into the apple (Andermatt et al., 1988; Boncheva et al., 2006). Therefore its application in CM control is rare.

*Metarhizium robertsii* and *Beauveria bassiana* are the major entomopathogenic fungi species used in biological control (Meyling et al., 2011; Wang and Feng, 2014). Field investigations indicated that 23.6% examined CM specimens were infected by *B. bassiana* in Germany and Austria (Zimmermann et al., 2013), whereas *M. robertsii* could be isolated in 51% of soil samples in South Africa (Abaajeh and Nchu, 2015). It was reported that CM larvae exposed to commercial preparations of *B. bassiana* strains induced up to 96% mortality (Gürlek et al., 2018).

The microsporidium *Nosema carpocapsae* naturally occurs in many CM populations as reported from diagnostic data from 1972 to 1990; it was prevalent in 57.3% of German CM field populations (Zimmermann et al., 2013). Fecundity and fertility of infested CMs were significantly reduced by 64% and 46%, respectively (Zimmermann et al., 2013). In laboratory experiments it was shown that CM infected with *N. carpocapsae* had a delayed development by one week and also laid less eggs than healthy individuals (Siegel et al., 2001).

Entomopathogenic nematodes (EPNs) from the Steinernematidae and Heterorhabditidae families have been mostly studied for control of cocooned CM larvae. *Steinernema carpocapsae* has been commercialized; cold-tolerant species of *Heterorhabditis* sp. were applied in environmental temperature below 15 °C (Lacey and Unruh, 2005). Application of infective juveniles of *S. feltiae* with adjuvants (wood flour foam or Barricade II fire retardant gel) resulted in significant increase of mortality from 24% to 85% or 97% of CM larvae (Lacey et al., 2010).

The baculovirus Cydia pomonella granulovirus (CpGV) was first isolated in Mexico, thus termed Mexican isolate CpGV-M (Tanada, 1964). CpGV-M is highly specific and extraordinarily virulent to CM larvae and was therefore the first studied and tested CpGV isolates to control CM in Europe and North America in the 1970s and 1980s, (Huber, 1998). Today, commercial CpGV products are used worldwide in nearly all major apple growing regions where CM is a pest.

### Classification status and histopathology of Cydia pomonella granulovirus

CpGV is a double-strand DNA virus, belonging to the genus of *Betabaculovirus* (family *Baculoviridae)*. The family *Baculoviridae* is classified into four genera of *Alphabaculovirus*, *Betabaculovirus*, *Gammabaculovirus*, *Deltabaculovirus* (Herniou et al., 2011). All baculoviruses have rod-shaped nucleocapsids and have a double-stranded, covalently closed DNA genome. They are specific for larval stages of insects from the order Lepidoptera, Hymenoptera and Diptera. During the viral replication cycle two virion phenotypes are produced: occlusion derived virus (ODV) and budded virus (BV) (Figure 1-3) (Blissard, 1996). ODV, embedded inside of protein matrix (polyhedrin or granulin) called occlusion body (OB), is responsible for peroral infection of an individual host; BV, comprised with a single nucleocapsid and a virion envelope, is in charge of infection transmission among host tissue cells (Herniou et al., 2011). Historically, baculoviruses were classified according their OB morphology; since the OB of alphabaculoviruses, gammabaculoviruses and deltabaculoviruses have a polyhedral shape, these viruses were termed nucleopolyhedroviruses (NPV), whereas betabaculoviruses have an ovicylindrical OB and are considered as granulovirus (GV) (Jehle et al., 2006).

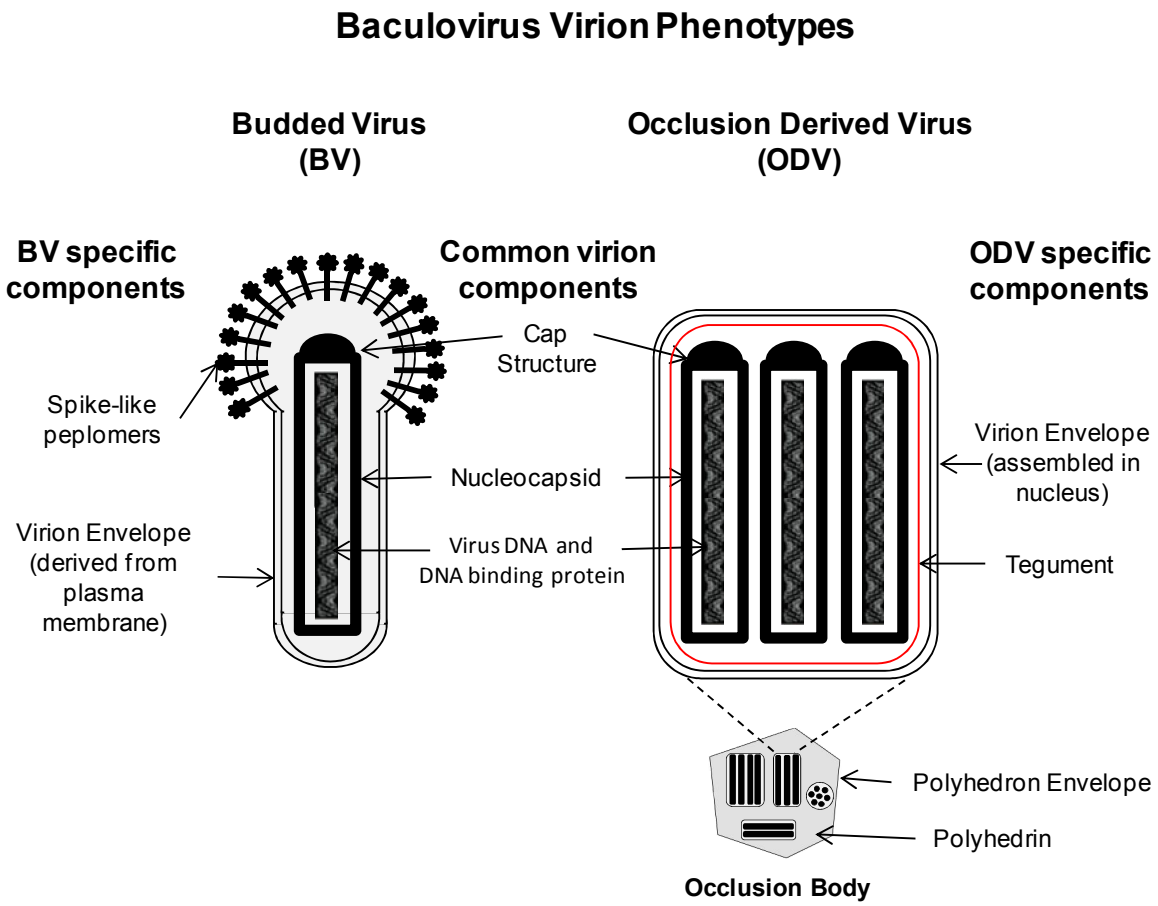



**Figure 1-3**. Baculovirus virion phenotypes of budded virus (BV) and occlusion derived virions (ODV) and occlusion body, based on shared and phenotype-specific components. Illustrations modified from Blissard (1996).

The OB of all baculoviruses including alphabaculoviruses, gammabaculoviruses, deltabaculoviruses except betabaculoviruses, contain more than one ODV (Figure 1-3 and Figure 1-4). ODV from alphabaculoviruses show two morphological types: a single virion may contain multiple nucleocapsids (MNPV) or a single nucleocapsid (SNPV) (Figure 1-3 and 1-4); ODVs from betabaculoviruses, gammabaculoviruses and deltabaculoviruses typically contain only a single nucleocapsid (Slack and Arif, 2007). OB size depends on OB shape, for example OB from NPV range between 0.5-5 $\mu$m whereas OB from GV are ovi-cylindrical and much smaller (0.3-0.5 $\mu$m). Nucleocapsids of all baculoviruses share the similar size in diameter from 30 to 60 nm and in length from 250 to 300 nm (Herniou et al., 2011).

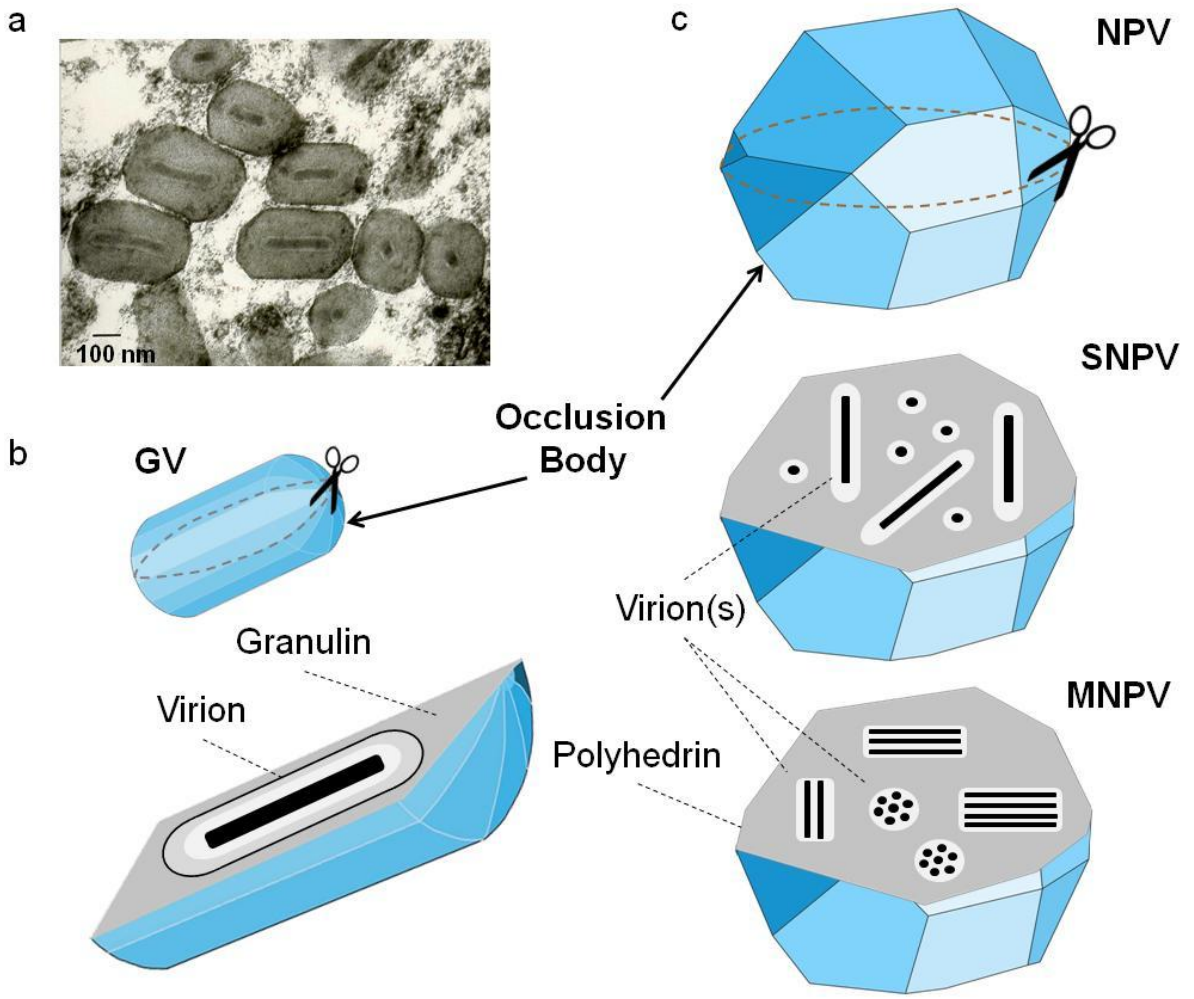



**Figure 1-4**. Morphological traits of occlusion bodies of baculoviruses. (a) Transmission electron microscope photograph of Cydia pomonella granulovirus (CpGV) (photographed by Huger, A., JKI). Graphic illustrations of (b) occlusion body (OB) of granulovirus (GV) containing a single virion and (c) OB of nucleopolyhedrovirus (NPV) contain several virions. NPV virions contain either single nucleocapsid (SNPV) or multiple nucleocapsids (MNPV). The structure and size of GV and NPV OB are described in the text. (Schematic diagrams modified from Sauer (2017) and Wennmann (2014)).

The OB of CpGV contains only a single occluded virion (Hess and Falcon, 1987; Jehle et al., 2006). The OB size is approximately 360×210 nm (Wu, 2015). Once OB is ingested by CM larva, ODV is released from OB in the alkaline midgut with a pH value of 8-12. After the virion attaches to and enters midgut epithelium cells it is transported to the nucleus. Nucleocapsids pass through the nuclear pore. Then host cell nucleus shows “clearing”, which represents the first electron microscopic sign of infection (Hess and Falcon, 1987). With initiation of viral replication a virogenic stroma is formed via recruitment of cellular components for its own proliferation. Newly formed nucleocapsids bud through plasma membrane generating BVs, which are transported via tracheae and hemolymph to other cells or tissues (Engelhard et al., 1994; Barrett et al., 1998). At the end of infection, ODV are produced and embedded into granulin matrix to form OB. Finally, OB is released from larval cadavers, generating a source for a next infection cycle for CM larvae (Figure 1-5).

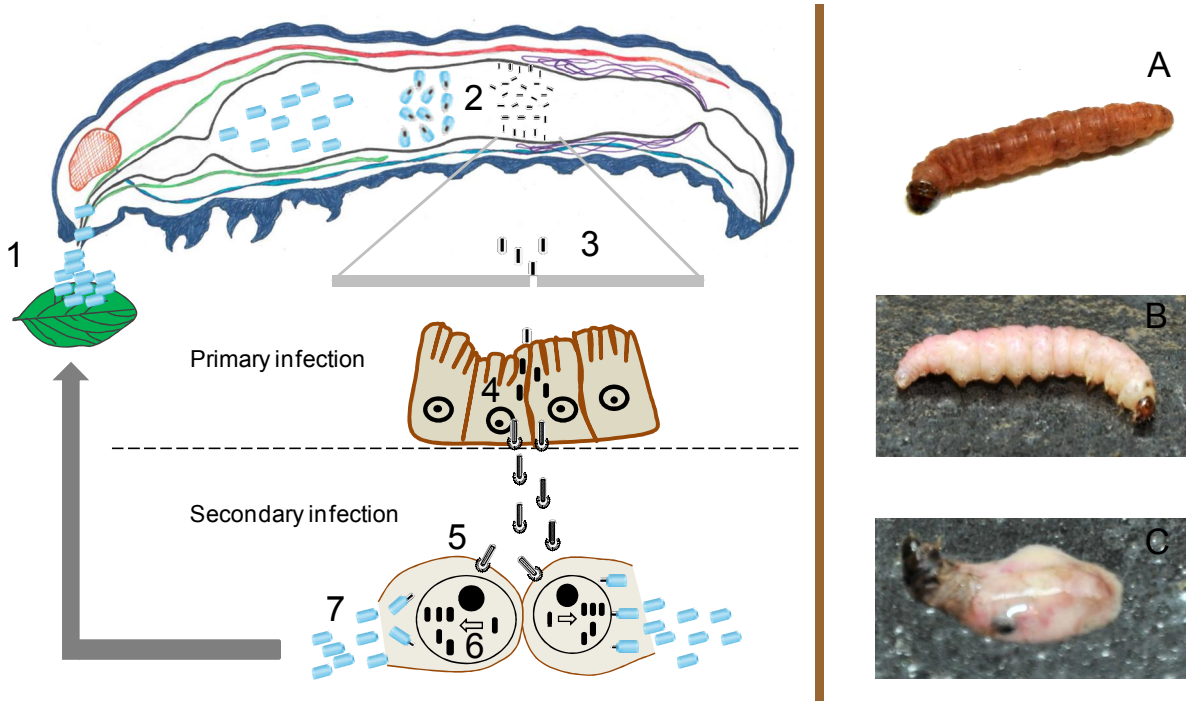



**Figure 1-5**. Graphic illustration of CpGV life cycle in codling moth (CM). On the left side the infection steps (primary and secondary infection) of CM larvae with CpGV are shown. Primary infection includes four steps: (1) Occlusion bodies (OB) are first ingested by CM larvae and transported to the midgut; (2) in the alkaline milieu of midgut (pH 8-12) (Dow, 1992) the granulin matrix protein is dissolved and occlusion derived viruses (ODV) are then released from OB; (3) ODV pass through the peritrophic membrane (PM) (dark gray) and subsequently bind to the microvilli membrane of midgut epithelial cells; (4) nucleocapsids enter into the nucleus where viral DNA is transcribed and replicated and viral nucleocapsids are produced in the nucleus; (5) newly generated nucleocapsids bud from infected cells through the plasma membrane to form budded virus (BV) that is responsible for spread of infection to different cells and tissues to initiate secondary infection; (6) BVs bind to cell membrane and enter into the cell and nucleus to initiate virus replication. At late stage of infection; (7) new ODVs are produced and embedded into OB, which will be released from dead larvae. Symptoms of infected CM larvae are given on right side. The images show healthy (A), infected (B) and CM cadaver (C). The infected larva (B and C) show representative characteristics of baculovirus infection with a whitish, milky color. The cuticle of dead larva is broken and CpGV OB is released from the cadaver. (Left CM larva painted by Dr. Xi, Y., right images photographed by Feiertag, S., JKI Darmstadt).

Different geographic CpGV isolates have been discovered worldwide (Table 1-1). Two isolates, CpGV-CJ01 and CpGV-ZY, have been recently discovered in China. They were isolated from a CM field population and from a laboratory-reared CM individual collected from the same field (Liu, 2010; Shen et al., 2012) and showed high activity to CM larvae (Zheng et al., 2011). Further isolates from China, described in this thesis, were found in field infected larvae from Xinjiang and Gansu province.

Full genomes sequences have been determined for CpGV-M1, -M, -E2, -I07, -I12, and -S (Luque et al, 2001, Gebhardt et al., 2014, Wennmann et al., 2017). The genome size of all sequenced CpGV isolates ranged from 120.8-124.3 kbp, encoding 137-142 ORFs. Transcription of baculovirus genes is classified into four temporal groups: immediate early

genes, delayed-early genes, late genes and very late gens (van Oers and Vlak, 2007). Early genes are transcribed by host RNA polymerase II before the onset of viral DNA replication whereas late and very late genes are transcribed after DNA replication using a viral encoded RNA Polymerase (Rohrmann, 2013). A set of 38 core genes is found in all baculovirus genome sequences so far and plays an important role in the virus life cycle (Table 1-2). The putative function of 76 CpGV ORFs was predicted based on sequence homology to genes of previously studied baculoviruses but little is known about gene function from direct studies with CpGV (Luque et al., 2001). A major repeat region, located in ORF25 and ORF26 was identified but varied in the six sequenced CpGVs; 13 imperfect palindromes (homologous regions, hrs), proposed to act as origins of genomic DNA replication (Hilton and Winstanley, 2007) and transcriptional enhancers (Kool et al., 1994), were found in these isolates (Wennmann et al., 2017). The palindromes *hr3* and *hr4* play an important role for CpGV-M replication (Elmenofy and Jehle, 2015). Heterogenous expression of both CpGV-GP37 and matrix metalloproteases (MMPs) were able to enhance the infectivity of NPVs (Liu et al., 2011; Ishimwe et al., 2015).

Based on sequence analysis of CpGV-M, -I12, and -S it was suggested that CpGV genome of single isolate is highly homogenous, harboring only small variants (Wennmann et al., 2017). A total of 788 single nucleotide polymorphism (SNP) sites were identified, when the genome sequences of CpGV-M, -E2, -I07, -I12, and -S were analysized (Wennmann et al., 2017, Gueli Alletti, 2017). Based on either concatenated nucleotide sequences of *granulin* and *lef-8* or concatenated amino acid sequences of 35 core genes as well as whole genome nucleotide sequences, CpGV can be classified into five distinguishable phylogenetic lineages, termed genome groups A to E (Eberle et al., 2009; Gebhardt et al., 2014; Wennmann et al., 2017), representing the known diversity and phylogeny of CpGV (Figure 1-6).

**Table 1-1**. Geographic origin of known CpGV isolates.

| Location | Isolates | Reference |
|---|---|---|
| Mexican | M | (Tanada, 1964) |
| Russia | R | (Crook et al., 1985) |
| England | E, E2 | (Crook et al., 1985) |
| Iran | I07, I12, I01, I08, I66, I68, I15, I22, I28, I30, I67 | (Rezapanah et al., 2008; Eberle, 2010) |
| Azerbaijan | AZ1, AZ2, AZ3, AZ4, AZ5, AZ6, AZ7, | (Eberle, 2010) |
| Georgia | G01, G02, G03, G04 | (Eberle et al., 2009; Eberle, 2010) |
| Canada | S | (Vincent et al., 2007) |
| China | ZY, CJ01 | (Liu, 2010; Shen et al., 2012) |
| Argentina | 2.17, 3.8, 6.9, 6.16, P118, Col19, C1, C6, M3, M10, M18, P7 | (Arneodo et al., 2015) |
| South Africa | SA | (Motsoeneng, 2016) |

**Table 1-2**. Core genes of baculovirus. Given are the ORF name, ORF number in CpGV and AcMNPV and their function

| Gene Designation | ORFs identical to | | Transcription category | Function Description | Reference |
|---|---|---|---|---|---|
| | CpGV | AcMNPV | | | |
| *odv-e18* | 14 | 143 | late | ODV envelope protein; mediateing BV production | (Braunagel et al., 1996b; McCarthy and Theilmann, 2008) |
| *p49/49k* | 15 | 142 | late | Required for BV production | (Vanarsdall et al., 2007; Miele et al., 2011) |
| *pif-5/odv-e56* | 18 | 148 | late | ODV envelope protein (PIF-5) | (Braunagel et al., 1996a; Harrison et al., 2010) |
| *pif-3* | 35 | 115 | late | Required for per os infection (PIF-3) | (Ohkawa et al., 2005) |
| *lef-2* | 41 | 6 | early | DNA replication/primase-associated factor | (Evans et al., 1997) |
| *pif-1* | 48 | 119 | late | Mediates binding of ODV to midgut (PIF-1) | (Kikhno et al., 2002; Peng et al., 2010) |
| *pif-7* | 53 | 110 | late | Impact on the oral infectivity | (Jiantao et al., 2016; Javed et al., 2017) |
| *odv-ec43* | 55 | 109 | late | Associated with ODV and BV | (Braunagel et al., 2003; Wang et al., 2010b) |
| *p74* | 60 | 138 | late | Mediates binding of ODV to midgut (PIF-0) | (Kuzio et al., 1989; Peng et al., 2010) |
| *p47* | 68 | 40 | early | RNA polymerase subunit | (Guarino and Summers, 1986; Carstens et al., 1993) |
| *lef-1* | 74 | 14 | early | DNA primase | (Evans et al., 1997) |
| *pif-2* | 75 | 22 | late | Required for per os infection (PIF-2) | (Pijlman et al., 2003) |
| *p48* | 83 | 103 | early | BV production and ODV envelopment | (Garavaglia et al., 2012) |
| *p40* | 85 | 101 | late | Subunit of protein complex in ODV and BV | (Braunagel et al., 2001; Braunagel and Summers, 2007) |
| *p6.9* | 86 | 100† | late | Nucleocapsid protein | (Wilson et al., 1987; Wang et al., 2010a) |
| *lef-5* | 87 | 99 | early | Transcription initiation factor | (Guarino et al., 2002) |
| *38k* | 88 | 98 | late | Required for nucleocapsid assembly | (Wu et al., 2008) |
| *pif-4/ac96* | 89 | 96 | early | Required for per os infection (PIF-4) | (Xu et al., 2006; Fang et al., 2016) |
| *helicase* | 90 | 95 | late | Unwinding DNA | (Bideshi et al., 1998; McDougal and Guarino, 2000) |
| *odv-e25* | 91 | 94 | late | ODV envelope protein | (Chen et al., 2012) |
| *p18* | 92 | 93† | late | Egress of nucleocapsids, involved in ODV envelop formation | (Yuan et al., 2011) |
| *p33* | 93 | 92 | late | Sulfhydryl oxidase, required for BV production | (Wu et al., 2010) |
| *lef-4* | 95 | 90 | late | RNA polymerase subunit/RNA capping enzyme | (Gross and Shuman, 1998; Guarino et al., 1998b; Jin et al., 1998) |
| *vp39* | 96 | 89 | late | Major capsid protein | (Pearson et al., 1988) |

*Table 1-2 cont.*

| | | | | | |
|---|---|---|---|---|---|
| *odv-ec27* | 97 | 144 | early | ODV envelope protein, involved in cell cycle arrest | (Braunagel et al., 1996b; Belyavskyi et al., 1998) |
| *vp91/pif8* | 101 | 83 | early | Viral capsid-associated protein | (Russell and Rohrmann, 1997; Javed et al., 2017) |
| *ac81* | 103 | 81 | late | Unknown function | (Chen et al., 2007) |
| *gp41* | 104 | 80 | late | Tegument protein | (Whitford and Faulkner, 1992; Braunagel et al., 2003) |
| *ac78* | 105 | 78 | late | Associated with virions formation and envelope | (Garavaglia et al., 2012; Tao et al., 2013) |
| *vlf-1* | 106 | 77 | late | Regulating very late gene transcripts e.g.*p10*; nucleocapsid assembly | (McLachlin and Miller, 1994; Vanarsdall et al., 2006) |
| *dnapol* | 111 | 65 | n.d. | DNA replication | (McDougal and Guarino, 1999) |
| *desmoplakin* | 112 | 66 | early | Required for egress of nucleocapsids from the nucleus | (Ke et al., 2008) |
| *pif6* | 114 | 68 | early | Required for per os infection (PIF-6) | (Li et al., 2011; Nie et al., 2012) |
| *lef-9* | 117 | 62 | late | RNA polymerase subunit | (Guarino et al., 1998a) |
| *alk-exo* | 125 | 133 | early | Involved in DNA recombination and replication | (Mikhailov et al., 2003; Okano et al., 2004) |
| *lef-8* | 131 | 50 | late | RNA polymerase subunit | (Passarelli et al., 1994; Vanarsdall et al., 2006) |
| *ac53* | 134 | 53 | late | Involved in nucleocapsid assembly | (Liu et al., 2008) |
| *vp1054* | 138 | 54 | early, late | Nucleocapsid protein | (Olszewski and Miller, 1997) |

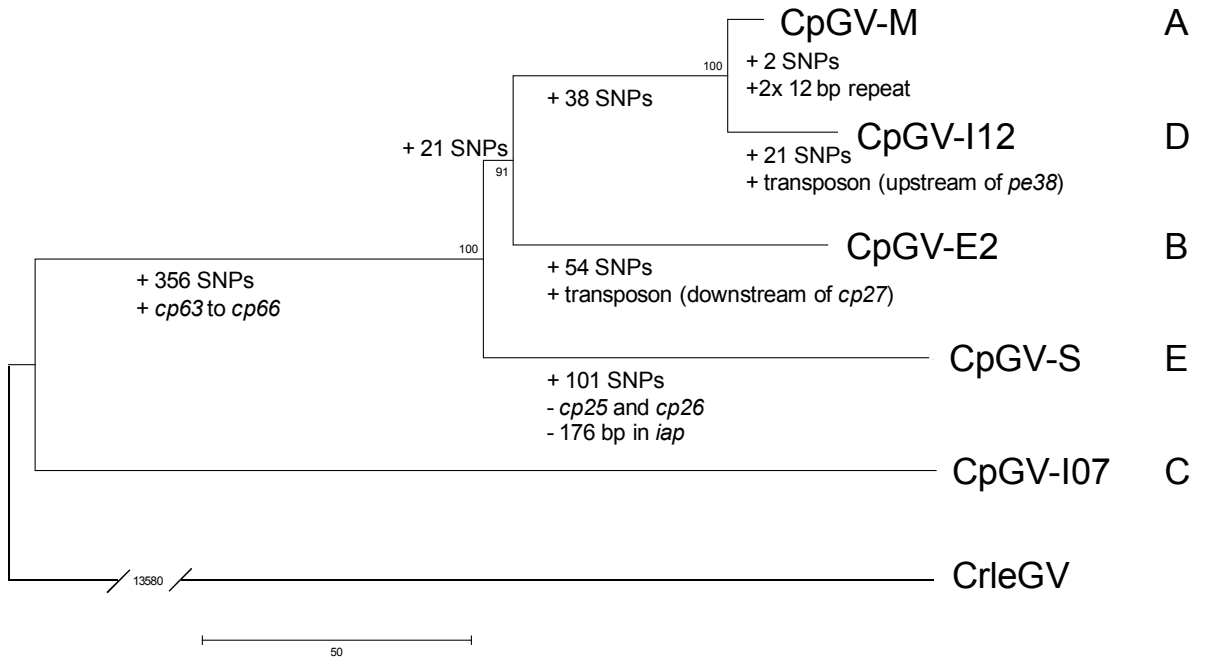



**Figure 1-6**. Minimum evolution tree of the isolates CpGV-M, -I12, -E2, -S, and -I07. The genome of Cryptophlebia leucotreta granulovirus (CrleGV) was used as an outgroup. The isolates were classified into five phylogenetic lineages, termed genome groups A to E. Specific SNPs and indels from each group are added to branch node to express the differences. Taken from Wennmann et al. (2017).

### Cydia pomonella granulovirus field application

CpGV is highly virulent and specific for CM larvae. Infected neonates start to die within 3-7 days following the typical symptoms of fragile and milky white body (Lacey et al., 2008). Four commercial CpGV products of Madex (Switzerland, 1987), Carpovirusine (France, 1998), Granupom (Germany, 1989), Cyd-X (USA, 1995) based on CpGV-M, were registered and used in apple and pear orchards (Huber, 1998; Saint Gès et al., 2014). In addition, a Canadian isolate was registered as Virosoft in the USA and Canada in 2000 (Vincent et al., 2007). Appropriate time, dosage and frequency of CpGV spraying during egg hatching period from the first generation and last generation provide considerable control of CM, ranging from 75.5% to 100% efficacy in field trials carried out in Germany, Czech Republic, Canada and USA (Huber and Dickler, 1977; Jaques et al., 1994; Stará and Kocourek, 2003; Arthurs and Lacey, 2004).

It is very important to apply CpGV products during egg hatching time, because neonate CM larvae bore into the fruit and are then difficult to be controlled. It was shown that newly hatched CM larvae can be infected by exposure to CpGV OB within minimal 3.5 min, but they move around on the leaves or fruits for up to 4.5 h before boring into the fruit (Ballard et al., 2000). Once CM neonates stay and grow inside apple, some larvae would leave the original drill hole and bore into apple again because of rainfall and other environmental factors, providing a second opportunity to control CM larvae (Steineke, 2004). Although the mortality of L1 and L3 induced by CpGV has shown no significant difference, infection of L3 larvae results in big larval cadavers and leaves useless apples (Steineke, 2004). Therefore it is very important to control CM neonates. Since at least two generations of CM

occur in almost all pome fruit regions and the development of one generation requires 36-82 days depending on the environmental conditions, approximately 8-26 sprays with CpGV are necessary to cover the whole growing season in one year (Arthurs et al., 2005)

### Occurrence of field resistance to Cydia pomonella granulovirus

The first CpGV products based on CpGV-M were registered in the late 1980s and early 1990s; they were then frequently used in organic apple production (Huber 1998, Lacey et al., 2008). Failure of CM control with commercial CpGV products was first noticed in Germany in 2003 and little later in France (Fritsch et al., 2005; Sauphanor et al., 2006). When CM field populations were reared in the laboratory and subjected to bioassays an increase of the median lethal concentration ($LC_{50}$) by a factor of 1,000 to 100,000 was observed, indicating field resistance to CpGV (Fritsch et al, 2005; Asser-Kaiser et al., 2007). A resistant field population in south Germany (Südbaden, DE-BW-FI-03, CpR) was then used to establish a genetically homogenous strain CpRR1 by single-pair crossings (Asser-Kaiser et al., 2007). By crossings and backcrossings with individuals of a susceptible CM strain (CpS) it was shown that resistance of CpRR1 followed a dominant inheritance that is linked to the Z chromosome (Asser-Kaiser et al., 2007). This type of resistance was later termed type I resistance (Jehle et al., 2017). CpR, the original genetically heterogeneous CM colony, showed a 1,000-fold resistance level against CpGV-M when its $LC_{50}$ value was compared to that of CpS. Resistance of CpR kept more or less stable for over 30 generations under virus-free conditions; once this strain was exposed to CpGV-M for two or five generations, resistance level of CpR increased to >1,000,000-fold without any fitness costs in fecundity and fertility (Undorf-Spahn et al., 2012). Meanwhile, type I resistance of CM field populations has been widely found in >40 orchards located in Germany, France, the Netherlands, Switzerland, Austria, Italy, Czech Republic (Schmitt et al., 2013; Zichová et al., 2013; Jehle et al., 2014; Sauer, 2017). When the infection process of CpGV-M in CpRR1 larvae was studied with a eGFP-expressing recombinant of CpGV-M (bacCpGV$^{hsp\text{-}eGFP}$) and by quantitative polymerase chain reaction (PCR), ultra-weak virus replication was quantified in midgut, hemolymph and fat body after both oral ingestion of OB and intra-hemocoelic infection of BV, demonstrating that type I resistance against CpGV-M has no tissue specificity (e.g. midgut-related factor) but is based on a systemic block at an early stage of virus replication (Asser-Kaiser et al., 2011). However, fluorescent GFP signals with at low intensity were detected in CpRR1 larvae, testifying the proliferation of CpGV-M and its limited exit to other cells, though without infection outburst and systemic spread (Asser-Kaiser et al., 2011).

Resistance testing on CpRR1 larvae showed that isolates from genome group B-E can break type I resistance (Eberle et al., 2008; Berling et al., 2009a; Gebhardt et al., 2014). When the genomes of CpGV isolates from genome groups A to E were sequenced and compared to each other, an insertion of 24 bp (2×12 bp repeat unit, 2×GACACAGTGGAT) in the gene *pe38* of CpGV-M was identified as the only common difference between the genomes of CpGV-M (genome group A) and the resistance-breaking isolates from group B to E. This finding suggested that *pe38* could be the target of type I resistance (Gebhardt et al., 2014). The gene, *pe38*, was proposed to be involved in virus gene replication during early stage of infection (Krappa and Knebel-Mörsdorf, 1991). A recombinant CpGV-M, carrying the *pe38* of the resistance-breaking CpGV-S instead of its own, was able to infect CpRR1 as efficiently as the susceptible strain CpS, strongly supporting the hypothesis that the lack of 2×12 bp repeat unit insertion in *pe38* is a key for overcoming type I resistance (Gebhardt et al., 2014).

Beyond CpGV isolates from genome groups B to E, newly discovered isolates and those successfully selected on resistant CM strains provide abundant resources to control CpGV resistance in the field and avoid future resistance development (Eberle et al., 2008; Rezapanah et al., 2008; Berling et al., 2009a; Graillot et al., 2014; Motsoeneng, 2016). Field trials indicated that resistance-breaking isolates CpGV-I12 and NPP-R1 can efficiently control both susceptible and resistant CM populations (Berling et al., 2009b).

A second form of resistance (type II), exhibiting cross-resistance to CpGV-M (genome group A) and CpGV-S (genome groupe E), was identified after low efficacy of CM control with novel resistance-breaking CpGV products had been observed in an orchard in northwest Germany (DE-NRW-WE-08) which had a long history of CpGV-M application and 3-year treatment of the resistance-breaking product MadexPlus (Jehle et al., 2017). From this field population, two genetically homogenous strains, namely CpR5M and CpR5S, have been established by mass-crossing and selection using CpGV-M and CpGV-S, respectively (Sauer et al., 2017a). Further crossing and backcrossing experiments revealed an autosomal dominant inheritance in CpR5M and CpR5S, as well as a systemic resistance for CpGV-M and a midgut related mechanism for CpGV-S (Sauer et al., 2017a).

A third type of resistance (type III) was verified using the same methods of crossings and backcrossings. The orchard was identified in eastern Germany (SA-GO-08) and overwintering larvae were collected in 2008 (Schmitt et al., 2013). Inbreeding using single-pair crossings resulted in strain CpRGO which showed a complex combination of autosomal and Z-linked inheritance (Sauer et al., 2017b).

### Application of Illumina Solexa sequencing in baculovirus

The genome sequence of AcMNPV clone 6 (accession No: NC_001623.1) was determined in 1994 (Ayres et al., 1994), composed of ~133.9 kbp, harboring 59% A+T content, encoding 156 ORFs, and providing the archetype for later sequencing and genome comparison of baculoviruses. In 2001, CpGV-M1 (accession No: NC_002816.1), a clone from Mexican isolate CpGV-M, was sequenced and annotated; the genome is 123.5 kbp in length and has a G+C content of 45.2% (Luque et al., 2001).

Recent progress has been made with application of next generation sequencing (NGS) techniques. Less costs and more output have driven more and more baculovirus isolates as well as virus populations to be sequenced. For example, AcMNPV strain E2, derived from infected Sf9 cells, was NGS sequenced and was compared to AcMNPV strain C6 to identify the differences between them (Maghodia et al., 2014). Isolate CpGV-M, -E2, -S and -I07 have been sequenced using 454 pyrosequencing (Wennmann et al., 2017). Operophtera brumata nucleopolyhedrovirus (OpbuNPV) and Dasychira pudibunda nucleopolyhedrovirus (DapuNPV) have been sequenced with high coverage using NGS sequencing (Krejmer et al., 2015; Harrison et al., 2017). Genomic knowledge of these NPVs has been annotated after reads assembly and consensus sequence generation. Similarly, different origins of baculovirus isolates have been sequenced to study the genetic diversity, variant and evolutionary trend. Such as, Noune and Hauxwell (2016) differentiated the variants among HaSNPV-AC53 geographic and passage-derived strains. Based on NGS data it was further proposed that AcMNPV, which has a broad range of Lepidoptera hosts, serves as vector of insect transposons between different moth species, thereby impacting the evolution of host genome (Gilbert et al., 2014). Widely distributed variations with low frequency in genome of AcMNPV derived

from natural population was evidenced by ultra deep Illumina sequencing (Chateigner et al., 2015). Genome group specific polymorphisms have been identified and applied to determine the genetic composition of three commercial CpGV products. Commercialized agents, MadexMAX, Carpovirusine EVO2, and MadexTOP (V15) were evaluated to be comprised of variable proportions of genome group A and E or genome group B and E, respectively. Utilization of genetic database in combination with efficiency of different isolates in resistance test that point out the way to optimize the genome group ratio in virus production and to acquire the optimal combination of isolates in a commercial product (Gueli Alletti et al., 2017).

## Chapter II: Novel diversity and virulence patterns found in new isolates of Cydia pomonella granulovirus from China


### Abstract

Cydia pomonella granulovirus (CpGV) is used worldwide as a powerful biological control agent of codling moth (CM, *Cydia pomonella* L.). Whereas in Europe CM populations with different modes of resistance against commercial CpGV preparations have been detected, CM is considered as an invasive pest species to China, requiring for new control options. To develop CM control strategies and to avoid emergence of CpGV resistance as it appeared in Europe, it is urgent to discover new CpGV isolates and to further utilize the naturally occurring genetic diversity of CpGV. Therefore seven new Chinese CpGVs were isolated in northwest China. Resistance testing using a discriminating CpGV concentration and median lethal concentration ($LC_{50}$) assays were conducted to determine their virulence against susceptible and resistant CM strains. The isolates were further screened for the presence of the 2×12 bp insertion in CpGV gene *pe38* (ORF24), which appeared to be the target of type I resistance by CM. It was found that the isolates CpGV-JQ, -KS1 and -ZY2 could break type I resistance, though a delayed infection pattern was observed in infection process. The isolates followed the *pe38* model of resistance-breaking, except CpGV-WW, which harboured the genetic factor but failed to overcome type I resistance. However, CpGV-WW was able to overcome type II and type III resistance. Based on the bioassay results and Sanger sequencing data, it is proposed that *pe38* is the major target for resistance in CpRR1. The new isolates show some distinct new virulence patterns when infection of different CM strains is considered.


### Introduction

Codling moth (CM), *Cydia pomonella* (L.), has spread and established local populations in temperate climate zones worldwide. Its larvae cause considerable economic damage to apple, pear, apricot, plum, peach, nectarine and walnut production (Barnes, 1991; Willett et al., 2009; Men et al., 2013). An efficient way to control CM without adverse effects to human health and the environment is the application of Cydia pomonella granulovirus (CpGV) a double-stranded DNA virus from the *Betabaculovirus* genus in *Baculoviridae* family (Tanada, 1964; Jehle and Hilton, 2011). CpGV was first discovered in Mexico in 1963 and the isolate was termed as CpGV-M (Tanada, 1964; Herniou et al., 2011). It belongs to type Ⅱ granulovirus which induces systemic infection in CM larvae, for instance of fat body, tracheal cells, as well as, Malpighian tubules and results in larval death at 3-7 days post-infection (Tanada and Leutenegger, 1968; Asser-Kaiser et al., 2007). Two virion phenotypes are produced during the infection process: the occlusion-derived virion (ODV) embedded in a viral occlusion body (OB) and the budded virus (BV) and which are responsible for primary infection of midgut epithelial cells and secondary invasion of other larval tissues, respectively (Hess and Falcon, 1987; Eberle et al., 2012). Genome size of CpGV range from 120.8-123.9 kbp in length and encode 137-142 genes (Luque et al., 2001; Gebhardt et al., 2014; Wennmann et al., 2017). Natural isolates were grouped into five phylogenetic lineages,

termed genome groups A to E, which were established by phylogenetic analysis using concatenated sequences of partial viral ORFs *granulin* and *lef-8* (Eberle et al., 2009; Arneodo et al., 2015), baculovirus core genes (Gebhardt et al., 2014), whole genomes (Wennmann et al., 2017), or via genomic fingerprints from specific single nucleotide polymorphisms (SNPs) (Gueli Alletti et al., 2017). Recently, diseased CM larva showing typical signs of baculovirus infection were field-collected in northwest China and new CpGV isolates, namely CpGV-ZY, -KS1, and -KS2 have been purified from the insect cadavers(Gan et al., 2011; Zheng et al., 2011).

In China, codling moth was first recorded as invasive pest in Xinjiang in 1953 (Zhang, 1957). In recent years, this CM population had spread eastward and so far, the pest had been found in five provinces in northwest China, including Xinjiang, Gansu, Ningxia and west part of Inner Mongolia (Zhang et al., 2012). Moreover a disjunct population, possibly from Russia has established in Jilin and Heilongjiang provinces in northeast China (Zhao et al., 2015). As the CM has started to spread in northern regions of China, where the major apple production area is located, it became a possible threat to Chinese apple production. In Europe, after more than a decade of successful application of CpGV, over 40 field CM populations resistant to commercial CpGV products have been recorded in seven European countries since 2005 (Asser-Kaiser et al., 2007; Zichová T., 2011; Schmitt et al., 2013). Three types (I-III) of CpGV field resistance with differing inheritance and susceptibility patterns have been identified (Asser-Kaiser et al., 2007; Jehle et al., 2017; Sauer et al., 2017a; Sauer et al., 2017b). Therein, type I resistance, first discovered in Germany, is the most common one (Fritsch et al., 2005). Represented by the genetically homogeneous resistant CM strain, CpRR1, inheritance of type I resistance is linked to the Z chromosome and follows a dominant mode (Asser-Kaiser et al., 2007; Asser-Kaiser et al., 2010). It was further shown that this resistance is systematic and instar-independent, and causes an early block in virus replication (Eberle et al., 2008; Asser-Kaiser et al., 2011). Only resistant female or homozygous resistant males successfully pupated and emerged into adults. Type I resistance is only directed against isolates from genome group A, such as CpGV-M, whereas isolates from other genome groups were able to break this type of resistance (Gebhardt et al., 2014). Genome sequence comparison and further molecular analysis have demonstrated that type I resistance is targeted against a 2×12 bp repeat within viral *pe38* (ORF24), which is present in CpGV-M of genome group A (Gueli Alletti et al., 2017). Other CpGV isolates from genome group B, C, D and E, which do not contain the additional repeat sequence, can break type I resistance (Berling et al., 2009b; Gebhardt et al., 2014). This 2×12 bp difference between CpGV-M and the resistance-breaking isolate CpGV-R5 was also used to distinguish the replication of both isolates in susceptible and resistant CM from France (strain $R_{GV}$) (Graillot et al., 2016). According to current knowledge, *pe38* is a suitable genetic marker to identify whether a given CpGV isolate is able to break type I resistance in CM. Type II and type III resistances are less well studied but follow different inheritance patterns and they are resistant to CpGV isolates from genome group A and E, such as CpGV-M and -S (Jehle et al., 2017; Sauer et al., 2017a; Sauer et al., 2017b).

Therefore, new isolates of CpGV would be an effective approach to control this insect pest concerning to both the Chinese CM threat and European CM resistance issues. Seven new CpGV isolates from China were tested for virulence against the susceptible and resistant CM strains at a discriminating concentration. Therein, three CpGV isolates were further used to conduct full range bioassays. Sequence analyses of the 2×12 bp repeat insertion of *pe38*

among these isolates revealed that not all isolates followed the current resistance mode, suggesting additional genetic factors involved in CpGV resistance.

## Material and methods

### Insect

Four strains of codling moth (CM) (*Cydia pomonella* L.), CpS, CpRR1, CpR5M and CpRGO, were used to perform resistance tests under laboratory conditions. In detail, CpS (susceptible CM to all CpGV isolates), CpRR1 (type I resistance), CpR5M (type II resistance) and CpRGO (type III resistance), were reared in light chambers at 26 °C under a 16/8 h light/dark photoperiod. CpRR1 was established in 2007 and was reared since 2011 without selection pressure (Asser-Kaiser et al., 2007; Undorf-Spahn et al., 2012). Larvae were reared on a semi-artificial diet at the Institute for Biological Control of the Julius Kühn-Institut (JKI), Darmstadt, Germany (Ivaldi-Sender, 1974). Fifth instar larvae were pooled from the diet to allow pupation in wrinkled cardboard strips. Adults were kept under the same rearing conditions and were allowed to mate to produce eggs. Freshly hatched neonates were directly used for bioassays.

### CpGV isolates and Propagation

Seven geographical isolates of CpGV isolates, designated CpGV-ZY, -JQ, -ALE, -KS1, -KS2, -ZY2, and -WW according to specimen location name (ZY: Zhangye, JQ: Jiuquan, ALE: Aral, KS: Kashgar, WW: Wuwei), were collected from northwest China during 2006 to 2014 (Figure 2-1). All viruses were isolated from individual dead larvae, showing typical signs of baculovirus infection in their larval or diapause phases. Infected larva normally died inside apples or hid under the cracks of bark on the strunk of apple trees. The occlusion bodies (OBs) were isolated from CM cadavers using the method described by Smith and Crook (1988). Propagation of these isolates was carried out by feeding purified OBs to fourth instar larvae of susceptible CM strain CpS. OBs were purified using glycerol gradient centrifugation (Arends and Jehle, 2002). Virus stocks of CpGV-M, -S, -E2 derived from -20 °C stocks at JKI (Eberle et al., 2009; Gebhardt et al., 2014). Enumeration of viral OBs was performed by dark field microscopy using a Petroff-Hauser counting chamber ($2.5\times10^{-3}$ $mm^2\times0.02$ mm depth) (Hausser Scientific, Horsham, PA, USA) and a light microscope (Leica DM RBE, Leica, Wetzlar, Germany).

### Virulence determination

A discriminating concentration of $5.8\times10^4$ OB/ml, inducing over 95% mortality on susceptible CpS larvae (Asser-Kaiser et al., 2007), was used to differentiate pathogenicity of CpGV-ZY, -JQ, -ALE, -KS1, -KS2, -ZY2, -WW as well as CpGV-M, -S and -E2 on CpS, CpRR1,CpR5M and CpRGO neonates. Infection experiments were conducted in autoclavable 50-well plates containing 45 ml of diet mixed with 5 ml OB suspension as described by (Eberle et al., 2008). For untreated controls, the virus suspension was substituted by 5 ml $H_2O$. Larvae that died from handling were recorded on the first day and were excluded from the experiment. Mortality of neonates was recorded at 7 and 14 days post infection (dpi). Each treatment was independently repeated three to five times. Observed virus-induced mortality was corrected for mortality of the untreated control group using the formula of Abbott (1925). The corrected mean mortality rates were analyzed using one-way ANOVA with pairwise comparison and Tukey HSD post-hoc test in R (R version 3.3.1 in RStudio 1.0).

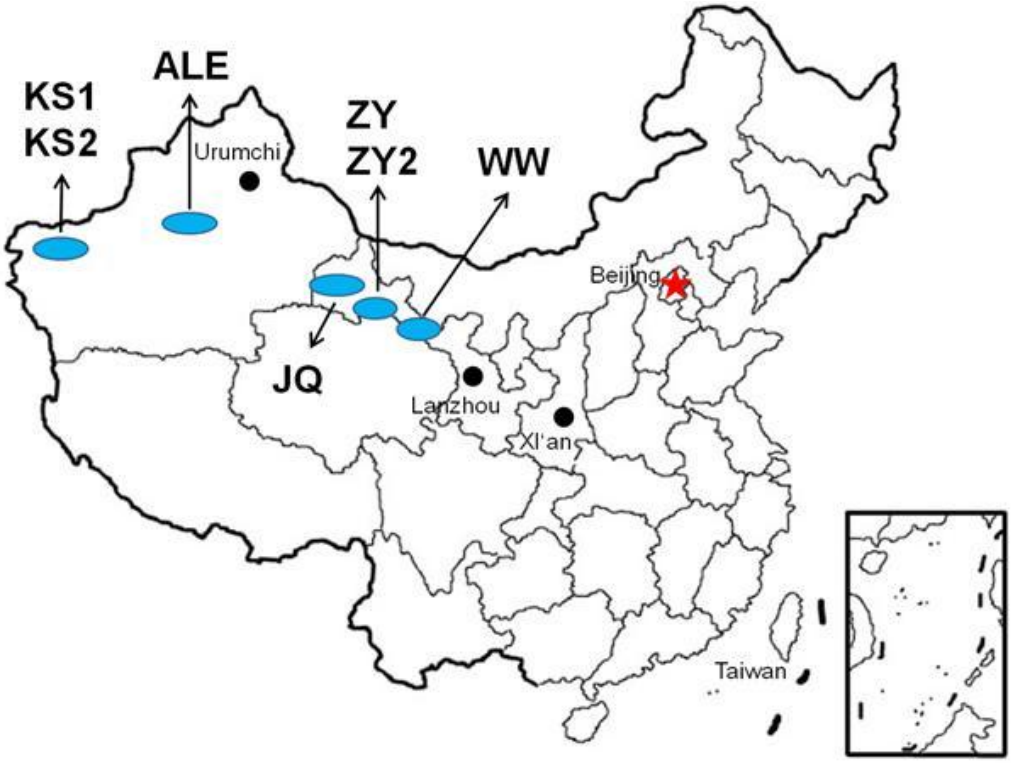



**Figure 2-1.** Location and index of new CpGV isolates collected in northwest China.The location names: KS: Kashgar (in Xinjiang); ALE: Alar (in Xinjiang); JQ: Jiuquan (in Gansu); ZY: Zhangye (in Gansu); WW: Wuwei (in Gansu).

Full-range bioassays of the isolates CpGV-ZY and -ALE were done to determine the median lethal concentration ($LC_{50}$ value) using six different virus concentrations ($3\times10^2$, $1\times10^3$, $3\times10^3$, $1\times10^4$, $3\times10^4$ and $1\times10^5$ OB/ml) (Eberle et al., 2008). Another six different virus concentrations ($5\times10^2$, $1\times10^3$, $5\times10^3$, $1\times10^4$, $5\times10^4$ and $1\times10^5$ OB/ml) were used to determine the LC50 value of CpGV-WW against CM larvae. For untreated controls, the virus suspension was substituted by 5 ml $H_2O$. Bioassays were processed as the same as for the discriminating test. Mortality data were determined at 7 and 14 dpi and were corrected for control mortality (Abbott, 1925) as well. Full range bioassay data were analyzed by Probit analysis using ToxRat Standard Software (Version 2.10) (ToxRat Solutions GmbH, 2005).

**Comparison of *pe38* among CpGV isolates**

Virus DNA extraction was performed as described by Arends and Jehle (2002). In brief, viral OBs were first dissolved in 100 mM $Na_2CO_3$. Then the solution was neutralized using 1.0 M HCl, 1% SDS was added to disrupt the virions. After two washings with a mixture of TE-saturated phenol : chloroform : isoamyl alcohol [25:24:1 (v/v/v)] the viral DNA was precipitated in 96% ethanol, dried and solubilized in double distilled water. A pair of PCR oligonucleotide primers, based on the CpGV-M genome, covering the region of 2×12 bp insertion of *pe38*, was designed using Primer Premier 5 software, *pe38*F: 5'TGGATAAGAAGGAATTGGAGG3', *pe38*R: 5'TTAATGGGTTTTTTGGTGGC3'. The fragments of each CpGV isolate were amplified in a volume of 50 $\mu$L, containing 2 $\mu$L of each primer (10 pmol/$\mu$L), 100 ng of viral DNA, 5 $\mu$L of 10× PCR reaction buffer, 4 $\mu$L of $MgCl_2$ (25 mM), 0.5 $\mu$L of dNTP (10 mM) and 0.5 $\mu$L of Taq DNA polymerase and adding distilled water to a final volume of 50 $\mu$L. The amplification conditions were set as

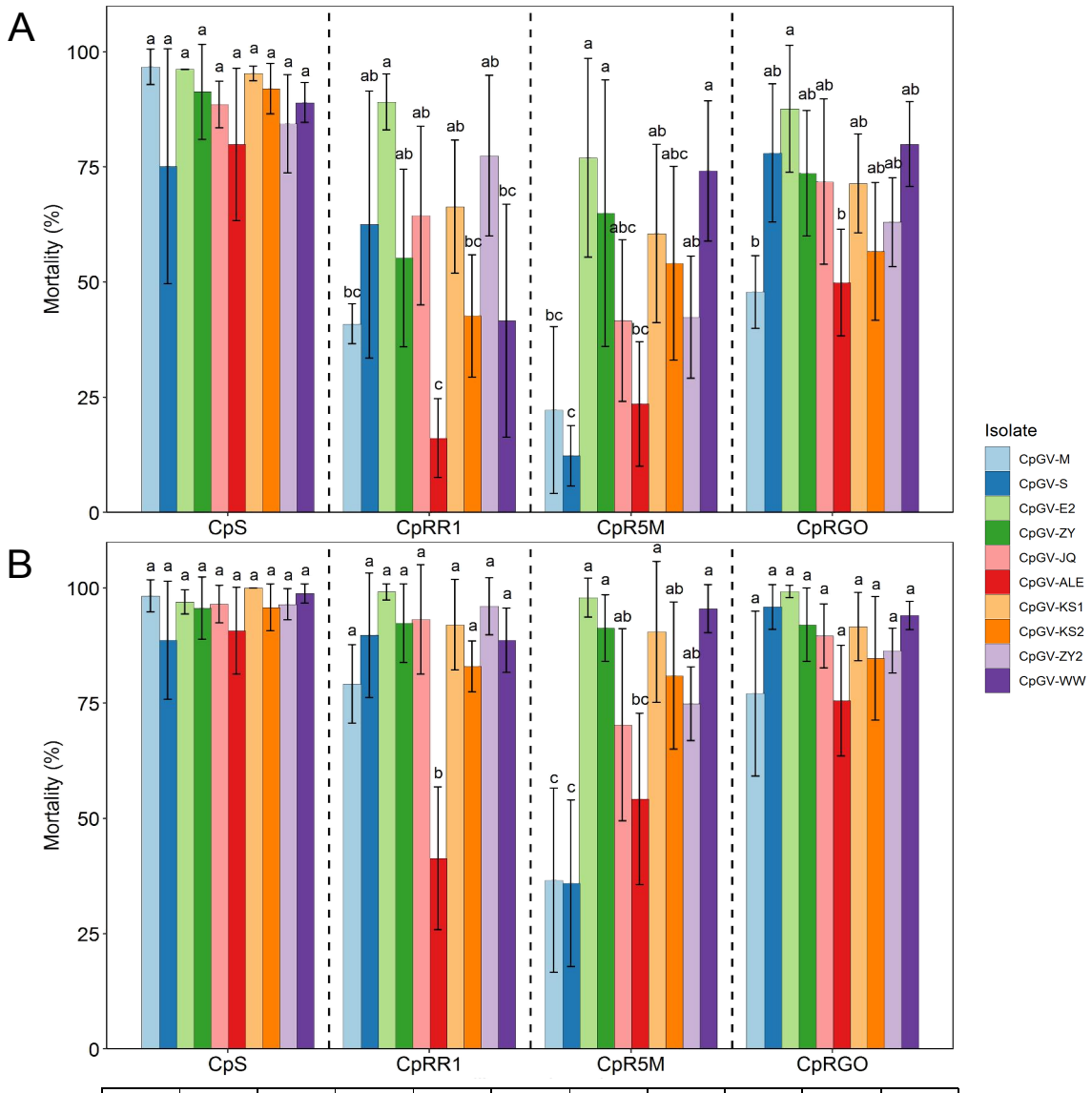
following: 3 min pre-denaturation at 94 °C, followed by 35 cycles of 94 °C for 30 s, 55 °C for 30 s and 72 °C for 45 s, a final extension period of 72 °C for 10 min. A small amount of 10 $\mu$L of PCR products was applied on 2% agarose gel electrophoresis stained with Midori Green Advance (Nippon Genetics Europe GmbH, Düren, Germany) and visualized under ChemoCam Imager ECL UV transilluminator and software (INTAS Science Imaging Instruments GmbH, Göttingen, Germany). The major PCR products were purified according to the manufacture of DNA Clean & Concentrator™ kit (Zymo Research, Freiburg, Germany). Purified DNA was cloned into pGEM®-T Easy Vector (Promega Corporation, Madion, USA) as described by the manufacturer. Two to four white colonies were picked for Sanger sequencing at StarSEQ GmbH (Mainz, Germany). All sequenced fragments were aligned using MUSCLE Alignment in Geneious version 9 (Biomatters, New Zealand) and using *pe38* of CpGV-M as reference.

## Results

### Virulence of CpGV isolates

Seven new CpGV isolates, designated CpGV-ZY, -JQ, -ALE, -KS1, -KS2, -ZY2, and -WW, all collected from northwest China (Figure 2-1), as well as the reference isolates CpGV-M, -S and -E2 were used to test their infectivity to neonate larvae of four CM strains, namely CpS (CpGV susceptible), CpRR1 (type I resistance), CpR5M (type II resistance) and CpRGO (type III resistance). A discriminating concentration of $5.8\times10^4$ OB/ml was applied. This concentration would cause >95% virus-induced mortality in susceptible CM larvae (Asser-Kaiser et al., 2007; Sauer et al., 2017b).

All isolates showed a similar activity in CpS larvae at 7 and 14 days post infection (dpi). For CpRR1, mortality caused by CpGV-JQ, -KS1 and -ZY2 at 7 dpi were 64.4%, 66.4% and 77.4%, respectively, which were higher than 40.9% and 62.5% induced by CpGV-M and the resistance-breaking isolate CpGV-S, respectively (Figure 2-2A). Therefore, these new isolates were considered as resistance-breaking isolates. In contrast, CpGV-ALE, -KS2 and -WW showed mortality of only 16.1%, 42.6% and 41.6%, respectively. CpGV-ZY showed an intermediate infection with 55.2% mortality, ranging between those induced by CpGV-M and -S. At 14 dpi all isolates except CpGV-ALE and -M could induce >80% mortality on CpRR1 larvae (Figure 2-2B). When comparing mortality rates between 7 and 14 dpi for each isolate, a delayed mortality appeared in CpRR1 larvae (Figure 2-2). Interestingly, the virus-caused mortality varied significantly in CpR5M (resistance type II). Mortality induced by CpGV-ZY and -WW on CpR5M was 65.0% and 74.1% at 7 dpi, respectively, and were higher than 55.2% and 41.6% on CpRR1 (Figure 2-2A). A similar trend could be observed after 14 dpi, though at a higher level of mortality. This finding indicated that CpGV-ZY and -WW were more virulent to CpR5M than to CpRR1. Strikingly, both CpGV-JQ and -ZY2 caused high mortality on CpRR1, but not on CpR5M, however CpGV-WW could overcome resistance in CpR5M but not in CpRR1. Both CpRR1 and CpR5M were least susceptible to CpGV-ALE, with mortality rates of less than 25% and 55% after 7 and 14 dpi, respectively. Compared to the treatment with resistance-breaking isolate CpGV-E2 the isolates CpGV-M and -ALE showed a similar low mortality for CpRGO (type III resistance) after 7 dpi, whereas after 14 dpi mortality of CpRGO neonates was statistically similar for all isolates (Tukey-HSD post-hoc test, $p > 0.05$, $\alpha = 0.05$). These comparative tests suggested that the isolates CpGV-JQ, -KS1 and -ZY2 were type I resistance-breaking, as their larvicidal activities were similar or higher than CpGV-S and CpGV-ZY and -WW were able to overcome type II resistance.




| | M | S | E2 | ZY | JQ | ALE | KS1 | KS2 | ZY2 | WW |
|---|---|---|---|---|---|---|---|---|---|---|
| CpS (n/N) | 132 / 4 | 142 / 4 | 86 / 3 | 140 / 4 | 143 / 4 | 138 / 4 | 138 / 4 | 142 / 4 | 110 / 3 | 106 / 3 |
| CpRR1 (n/N) | 152 / 4 | 148 / 4 | 117 / 3 | 142 / 4 | 147 / 4 | 157 / 4 | 161 / 4 | 147 / 4 | 117 / 3 | 117 / 3 |
| CpR5M (n/N) | 86 / 3 | 59 / 2 | 65 / 3 | 114 / 4 | 114 / 4 | 103 / 4 | 114 / 4 | 102 / 4 | 67 / 3 | 74 / 3 |

**Figure 2-2.** Determination of the activity of seven Chinese isolates (CpGV-ZY, -JQ, -ALE, -KS1, -KS2 -ZY2, and -WW) on four strains of codling moth (CpS, CpRR1, CpR5M and CpRGO). CpGV-M, -S, -E2 were included as controls to test resistance-breaking ability. A discriminating concentration of $5.8\times10^4$ OB/ml was used in all assays. Mortality at 7 dpi (A) and 14 dpi (B) was corrected for control mortality (Abbott, 1925) and plotted with the standard deviation. Differences in mean mortality induced by ten isolates were pairwise compared using one-way ANOVA and Tukey HSD post-hoc test for each strain. The same letters indicate no difference between two isolates. n and N represented the total tested larvae and replicates of four strains of CM against ten CpGV isolates, respectively.

**Table 2-1**. Median lethal concentration ($LC_{50}$) of three CpGV isolates tested on susceptible (CpS) and resistant (CpRR1) codling moth strains. Number (No.) of tested insect larvae, the $LC_{50}$ values including 95% Confidence Interval (CI) values for 7 and 14 days post infection, as well as the slope of probit line and its standard error (SE), and the $\chi^2$ value of the fit to the probit analysis and the degree of freedom (df) are given.

| Isolate | Strain | No. | 7 days post-infection | | | 14 days post-infection | | |
|---|---|---|---|---|---|---|---|---|
| | | | $LC_{50}$ [$10^3$×No. of Bs/ml] (95% CI) | Slope ± SE | $\chi^2$ (df) | $LC_{50}$ [$10^3$×No. of Bs/ml] (95% CI) | Slope ± SE | $\chi^2$ (df) |
| ZY | CpS | 914 | 2.99 (1.81-4.77) | 1.33 ± 0.12 | 11.87 (4) | 0.55 (0.14-1.17) | 1.04 ± 0.03 | 16.48 (4) |
| | CpRR1 | 971 | 34.84 (28.67-43.29) | 1.40 ± 0.01 | 2.30 (4) | 3.32 (1.58-6.53) | 1.35 ± 0.01 | 25.49 (4) |
| ALE | CpS | 1162 | 18.76 (13.53-27.14) | 1.26 ± 0.01 | 15.96 (7) | 5.15 (3.28-7.33) | 1.58 ± 0.01 | 25.31 (7) |
| | CpRR1 | 1528 | n.d. | 0.72 ± 0.04 | 33.99 (4) | 53.85 (32.65-110.6) | 1.05 ± 0.01 | 11.84 (4) |
| WW | CpS | 1255 | 14.65 (9.36-23.01) | 1.71 ± 0.01 | 19.19 (4) | 5.06 (3.67-6.92) | 1.61 ± 0.01 | 9.28 (4) |
| | CpRR1 | 1607 | 38.31 (16.79-161.86) | 0.73 ± 0.01 | 23.54 (4) | 4.57 (3.24-6.34) | 1.12 ± 0.01 | 8.52 (4) |

n.d.: not determined because the calculated value is beyond the tested concentrations by more than factor 1,000.

When median lethal concentration ($LC_{50}$) of three isolates (CpGV-ZY, -ALE, and -WW) were determined for CpS and CpRR1, it was found that CpRR1 larvae appeared to be resistant with more than 2.6-fold level against three isolates at 7 dpi (Table 2-1). This resistance phenomenon was also recorded at 14 dpi except for isolate CpGV-WW which then induced similar mortality in both CpS and CpRR1 larvae (Table 2-1).

**Analysis of *pe38* PCR products**

Amplification of *pe38* (ORF24) was carried out to identify if the new isolates harbour the 2×12 bp insertion, which was hypothesized in previous studies to be the target of the type I resistance in CpRR1. Therefore, a partial *pe38* fragment covering the 2×12 bp insertion was amplified by polymerase chain reaction (PCR) from all seven Chinese CpGV isolates with CpGV-M as a reference. Agarose gel electrophoresis of the amplified fragments revealed that PCR fragments of CpGV-ZY, -ALE, -KS1 and -M were within the same size of approximately 304 bp (Figure 2-3). However, the isolates CpGV-JQ, -ZY2 and -WW showed smaller bands with an estimated size of around 280 bp. Interestingly, CpGV-KS2 showed a slightly larger size than that found in CpGV-M. Apparently, CpGV-ZY2 and possibly also -KS1 and -ZY showed a submolar band similar in size to the fragment of -KS2 defining them as a mixture of genotypes carrying heterogeneous versions of *pe38*.

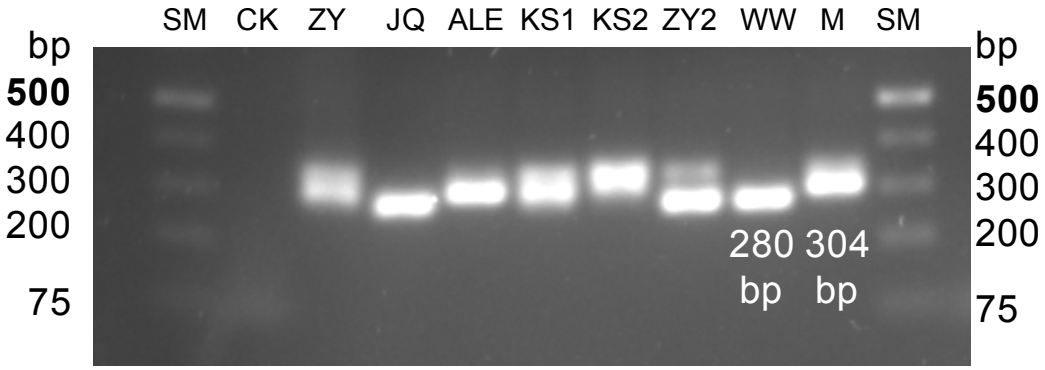



**Figure 2-3.** Agarose gel of PCR fragments of 2×12 bp insertion region of the *pe38* gene from isolates CpGV-ZY, -JQ, -ALE, -KS1, -ZY2, -WW, -M. CK, negative control; SM = Size marker GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific, Carlsbad, CA, USA). Size of bands is given in bp..

**Multiple sequence alignment of *pe38* segment**

Sanger sequencing of PCR fragments showed that only CpGV-ALE and -KS1 contained the 2×12 bp repeat insertion (eight amino acids) as is typical for CpGV-M and related isolates, whereas isolates CpGV-JQ and -WW lacked the insertion as would be typical for isolates breaking type I resistance (Figure 2-4A). Two types of *pe38*, with and without these 24 nucleotides, were detected in two isolates of CpGV-ZY and -ZY2. For CpGV-KS2 a 3×12 bp insertion was found resulting in one and three repeats of the DTVD motive (Figure 2-4B), and no 2×12 bp insertion virus population was detected as well. In summary, there were three types of *pe38* representing no insertion (CpGV-WW and -JQ), 2×12 nucleotides (-ALE) or 3×12 nucleotides insertion (-KS2).


**Figure 2-4.** Sequence analysis of *pe38* obtained from PCR products of seven CpGV isolates. Sequences of CpGV-M, -S and -E2 are given as a reference (Gebhardt et al., 2014). (A) Nucleotide sequence of randomly picked clones from each isolate; (B) the corresponding amino acid sequences. The number of the sequenced colony of each isolate was added to the end of the isolate name. Isolates CpGV-ZY and -ZY2 indicated two types of *pe38* with and without 2×12 bp insertion; isolate -KS2 showed a 3×12 bp insertion or no insertion; isolates -JQ and -WW contained no 2×12 bp insertion in *pe38.* The isolates -ALE and -KS1 showed the same pattern of *pe38* as CpGV-M.

## Discussion

The infectivity of seven new CpGV isolates was tested in terms of their capacity to infect susceptible and resistant CM strains. All isolates were highly infectious for susceptible CM and several of them appeared to be resistance-breaking. The mortality of CpRR1 induced by CpGV-M was 40.9% at 7 dpi and 79.2% at 14 dpi, respectively. Though a reduced susceptibility of CpRR1 to CpGV-M was clearly notable when compared to CpS, the mortality of CpRR1 larvae was considerably higher than mortality of 5% and 11% under similar conditions in previous studies (Sauer et al., 2017b). This finding suggests that CpRR1 may have lost some degree of its high level of resistance during the last couple of years of rearing without further resistance selection. A similar phenomenon was reported for CpR (Undorf-Spahn et al., 2012) that showed a decline but not a complete loss of resistance during 60 generations of rearing without viral selection pressure. CpRR1 strain was reared since 2011 without selection pressure. Also for CpR5M and CpRGO mortality caused by CpGV-M and CpGV-S appeared slightly higher, though the general pattern of resistance was clearly confirmed (Sauer et al., 2017b). In relation to CpRR1 (type I resistance), CpGV-M is a target of resistance, whereas CpGV-S is a resistance-breaking isolate (Gebhardt et al., 2014; Sauer et al., 2017a). Therefore the mortality rates induced by CpGV-M and CpGV-S were set as fix points to determine the resistance-breaking activity of the new isolates. The isolates CpGV-

ALE, -KS2 and -WW were not able to efficiently kill CpRR1 larvae as it caused a similar low mortality at 7 dpi as CpGV-M (genome group A) (Gebhardt et al. 2014).

Isolate CpGV-WW killed CpR5M and CpRGO as efficiently as the resistance-breaking CpGV-E2, which was shown to be the most powerful CpGV isolate breaking all three types of resistance (Gueli Alletti et al., 2017; Sauer et al., 2017b). On the other hand, its efficacy against CpRR1 was significantly lower at 7 dpi (Figure 2-2A), leaving CpGV-WW as the first isolate able to break type II and III resistance. However it was not highly effective against type I resistance though its efficacy was considerably delayed when compared to other isolates breaking type I resistance. The results of resistance testing at the discriminating concentration were also reflected in selected full range bioassays done for CpGV-ZY, -ALE, and -WW on CpS and CpRR1. The $LC_{50}$ at 7 dpi could not be determined for CpGV-ALE due to the very low efficacy and also the $LC_{50}$ values at 14 dpi were significantly higher than those of CpGV-ZY and -WW (Table 2-1).

The viral gene *pe38* was proposed as a target of type I resistance (Gebhardt et al., 2014) and this gene has also been used as a genetic marker to distinguish between CpGV isolates prone to this type of resistance or breaking it (Gebhardt et al., 2014; Graillot et al., 2016), depending on the presence of a 2×12 bp repeat insertion, which was found in CpGV-M and other CpGVs of genome group A but not in resistance-breaking isolates CpGV-E2, -S and others.

PCR and sequence analysis of the 2×12 bp repeat insertion region in *pe38* of the seven new CpGV isolates showed there were three types of *pe38*; those without the insertion (CpGV-JQ, -WW), some with a 2×12 bp repeat insertion (CpGV-ZY, -ZY2, -ALE, -KS1) and one isolate CpGV-KS2 with a 3×12 bp repeat insertion (Figure 2-3 and 2-4). Gel electrophoresis of PCR products indicated that some isolates, such as CpGV-ZY, -ZY2 and likely -KS1 contained double bands of the 2×12 bp repeat area. Sanger sequencing of cloned PCR fragments further corroborated this finding for CpGV-ZY, -ZY2, and -KS2 but not for -KS1 (Figure 2-4). Thus, based on *pe38* heterogeneity, at least some of the isolates clearly represented genotype mixtures, which may further explain the high variability in the resistance testing.

Strikingly, the efficacy of CpGV-KS2 and -WW was low in CpRR1, though the two isolates contained partly or entirely genomes without the 2×12 bp repeat insertion (Figure 2-2 and Figure 2-4A), which had been proposed as a molecular marker for resistance-breaking of CpRR1 (Gebhardt et al., 2014). Field isolates of baculoviruses often show different proportions of genotypes and distinct larvicidal activities (Cory et al., 2005; Del-Angel et al., 2018). Mixtures of genotypes may contribute to the genetic plasticity of a virus populations in the field (Herniou et al., 2003). This appears also to be the case for CpGV, for which field isolates have been identified as mixtures with potential adaptation to different ecological habitats (Fuxa, 1993; Rezapanah et al., 2008). On the other hand, laboratory selections on resistant CM populations ($R_{GV}$) promoted CpGV to adapt to the resistant host and resulted in an increased virulence against type I resistance, e.g. progeny virus of isolate NPP-R1 propagated on $R_{GV}$ for either 4 or 16 generations caused a decrease of CpGV-M-like genotypes and an increase of its efficiency on type I resistance (Berling et al., 2009a; Graillot et al., 2014).

In previous studies, CpGV-E2, a minor variant of CpGV-E, could not be obtained by *in vivo* passage experiments using CpGV-E as inoculum on susceptible CM larvae (Crook et al., 1985).

Restriction endonuclease analyses and whole genome sequencing of CpGV-E2 proposed that this isolate is a mixture of genome group A and B isolates (Crook et al., 1985; Gebhardt et al., 2014). As CpGV-E2 showed high pathogenicity of the CpRR1 strain, it is conceivable that genotypes from genome group A may positively influence virulence of resistance-breaking genotypes, though genome group A viruses alone cannot break resistance. This hypothesis is supported by the observation of a $LC_{50}$ value of isolate CpGV-ZY which was lower than that of CpGV-WW induced in full range bioassay on CpRR1 at 7 and 14 dpi (Table 2-1). A mixture of CpGV-M and -R5 (each 50%) induced mortality in resistant CM strain $R_{GV}$ (type I resistance strain) which was similar as that caused by the resistance-breaking isolate CpGV-R5 alone (Graillot et al., 2016).

Interestingly, *pe38* of CpGV-WW appeared highly homogenous, it did not contain the 2×12 bp repeat insertion, though its efficacy against CpRR1 was reduced or at least delayed. This finding contradicts the previously established model that the lack of 2×12 bp repeat insertion of *pe38* causes high viral activity on type I resistance. This inconsistent result implies that the structure of *pe38* alone may not fully explain whether an isolate breaks type I resistance or not. Berling et al. (2013) proposed that more than one gene is involved in type I resistance, the observed response of CpGV-WW provides a hint in the same direction.

In conclusion, resistance tests and the partial sequence of *pe38* of seven new CpGV isolates from northwest China revealed novel patterns of infectivity towards different CM strains. CpGV-ALE showed consistent low efficacy in all three resistant CM colonies, whereas CpGV-WW had a superior effect in CpR5M and CpRGO but not in CpRR1. Even though there is no CpGV resistance reported in China till now, new isolates could be use as biocontrol agents for CM, some of which can break the CM resistance and these genetic difference related to resistance-breaking may provide new genetic resources in future resistance management strategies.

## Chapter III: Partial loss of inheritable type I resistance of codling moth to Cydia pomonella granulovirus


### Abstract

Three types of field resistance of codling moth (CM, *Cydia pomonella*, L) against commercial products of Cydia pomonella granulovirus (CpGV) are known. Current knowledge of resistance management is founded mainly on the interaction between the Mexican isolate CpGV-M and CpRR1, a genetically homogeneous imbreed line carrying a high potency of type I resistance. Whereas previous studies indicated that CpRR1 harbored high level of resistance against CpGV-M, the resistance level of laboratory-reared CpRR1 towards CpGV-M was recently found to have decreased considerably, probably because of rearing on virus-free diet without further selection for many years. To understand the background of this phenomenon, CpRR1 larvae were exposed over several generations to $2\times10^4$ OB/ml of CpGV-M for re-selection of the original resistance level. After five and seven generations of selection, new CpRR1_F5 and CpRR1_F7 lines were established. The resistance ratio of these selected lines was determined by full range bioassays and single-pair crossings with susceptible CM individuals followed by a resistance test using a discriminating concentration of CpGV-M occlusion bodies. The CpRR1_F5 strain regained a higher level of resistance against CpGV up to $10^4$-fold based on $LC_{50}$ values compared to susceptible larvae (CpS), which indicated that the absence of virus selection had resulted in a reduction of resistance under laboratory rearing conditions. In addition, some fitness costs of fecundity were observed in the re-selected strain CpRR1_F5. Single-pair crossings with susceptible codling moth revealed a dominant but not fully sex-linked inheritance which suggests for a partial loss of previous resistance traits in CpRR1.


### Introduction

Cydia pomonella granulovirus (CpGV) is applied as one of the most important and commercially successful baculovirus-based insectcides used to control larvae of a codling moth (CM, *Cydia pomonella* L.) in pome fruit production (Huber, 1998; Eberle and Jehle, 2006; Lacey et al., 2008). In 2003, a CM field population, designated DE-BW-FI03 or CpR, was collected from an organic apple orchard in southwest Germany, where application of commercial CpGV preparations, containing the Mexican isolate CpGV-M, had failed (Fritsch et al., 2005; Eberle and Jehle, 2006). When individuals of this population were reared in the laboratory for two years on virus-free diet, the established colony still exhibited an at least 100-fold resistance to CpGV-M. Initial genetic studies based on mass crossing experiments suggested an autosomal and incompletely dominant inheritance pattern of resistance (Eberle and Jehle, 2006). This early finding was challenged after a highly resistant and genetically homogenous inbred-line of CpR, termed CpRR1, was selected. Single-pair crossing experiments between individuals of CpRR1 and of susceptible CM strain CpS, followed by backcrossing experiments unequivocally demonstrated a dominant, sex-linked resistance, which was later also confirmed for CpR (Eberle and Jehle, 2006; Asser-Kaiser et al., 2007).

Later on, more than 40 local orchards with CM populations presumed to carry type I resistance have been discovered in Austria, Czech Republic, France, Italy, the Netherlands and Switzerland (Sauphanor et al., 2006; Asser-Kaiser et al., 2010; Schmitt et al., 2013; Zichová et al., 2013). It has been suggested that this type I resistance to CpGV-M is geographically distributed in Europe, though concentrated within individual orchards, because resistance-breaking CpGV isolates were successfully applied in nearly all orchards with CpGV resistance (Jehle et al., 2017). When the heterogeneous CpR was reared for more than 60 generations without virus selection pressure, its resistant level, based on median lethal concentration values, was still 100-fold higher than susceptible CM against CpGV-M (Undorf-Spahn et al., 2012). Once the CpR strain was exposed to CpGV-M, the resistance level increased up to 1,000,000-fold which demonstrates resistant CM larvae have developed a fast and effectively adaptive mechanism against CpGV-M with no discernible fitness cost.

Two types of crossing methods, single-pair crossings and mass crossings, were used to select for genetically homogenous resistant CM colonies. CpRR1 was founded by consecutive single-pair crossings starting with CpR individuals (Asser-Kaiser et al., 2007). An alternative approach using successive mass crossings combined with selections on CpGV-M (e.g. $3\times10^3$ to $3\times10^5$ OB/ml) was applied to obtain the strain CpR-CZ from a Czech CM field population (Zichová et al., 2013). Also other genetically homogenous resistant CM strains, such as CpR5M/CpR5S (both type II resistance) and CpRGO (type III resistance) were selected by mass crossings and single-pair crossings, respectively (Sauer et al., 2017a; Sauer et al., 2017b).

In contrast, the resistant CM strain RGV-8 was generated from eight generationS of mass crossings between individuals of a resistant French CM field population and a susceptible CM strain combined with selection of the progeny on CpGV OBs (Berling et al., 2013). The resulting strain RGV-8 exhibited a strongly dominant, sex-linked resistance, which suggested that a further resistance-related allel or mode was needed to fully explain the observed results of backcrossing and resistance testing (Berling et al., 2013). Mass crossing selection of another CM population from North Rhine-Westphalia in Germany, namely DE-NRW-WE-08, resulted in the discovery of the so-called type II resistance, which was shown to be autosomally, dominantly inherited. Type II resistance is targeted against isolates other than only the CpGV-M isolate (Jehle et al., 2017)). Two lines, CpR5M and CpR5S, were selected for five generations using $10^5$ OBs/ml of either CpGV-M or CpGV-S in each generation. Both lines were cross-resistant to CpGV-M and CpGV-S suggesting a somehow interlocked mode of action (Sauer et al., 2017a). Finally, a third resistance type (III), carrying Z-chromosomal and autosomal inheritance traits was discovered, demonstrating that resistance against CpGV is a complex phenomenon, which includes several mechanisms and inheritance pathways (Sauer et al., 2017b)

Most laboratory-selected baculovirus-resistant insect colonies become virus sensitive again when virus selection pressure is removed, most likely because of fitness costs of carrying the resistance (Fuxa and Richter, 1998). On the other hand, resistance of Cabbage loopers (*Trichoplusia ni*) against Trichoplusia ni single nucleopolyhedrovirus (TnSNPV) was stable for at least five generations without selection following nine generations of TnSNPV selection (Milks and Myers, 2000). Artificially selected resistance of smaller tea tortrix strain against Adoxophyes honmai nucleopolyhedrovirus (AdhoNPV), harboring a midgut-conferring resistance, showed an over 67,000-fold increased resistance level compared to susceptible

larvae; the resistance was stable for more than 168 generations in 13 years of laboratory rearing without further AdhoNPV selection pressure (Iwata et al., 2017; Nakai et al., 2017).

Recently, resistance tests using a discriminating concentration of $5.8\times10^4$ OB/ml showed that the neonates of the CpRR1 strain were more sensitive to CpGV-M than observed in previous tests (Chapter 2; Asser-Kaiser et al., 2007). Because this type of resistance is assumed to occur in numerous CM field populations, the expression of resistance within the CpRR1 strain after several years of rearing without CpGV pressure is a matter of considerable interest. A mass-crossing selection process was conducted for seven generations to regain the high level of resistance in the CpRR1 strain. Newly selected CpRR1_F5 and CpRR1_F7 were single-pair crossed with CpS to confirm the inheritance pattern of resistance in resistance testing using a discriminating CpGV-M concentration.

## Material and methods

### Insects and virus

Larvae of laboratory CM populations were maintained individually on a semi-artificial diet (Ivaldi-Sender, 1974) in autoclavable 50-well chambers (Licefa, Bad Salzuflen, Germany) at 26 °C under a photoperiod of 16:8 hours (light:dark). Both the susceptible strain (CpS) and resistant strain (CpRR1, type I resistance) were reared at the Institute for Biological Control of the Julius Kühn-Institut, Darmstadt, Germany. CpRR1 was originally selected from the field population DE-BW-FI03 (= CpR) in 2006 (Asser-Kaiser et al., 2010). It was reared for six years without virus selection pressure after an interim single-pair crossing selection in 2011 (Pietruska, 2018). Occlusion bodies (OBs) of the Mexican strain of CpGV, termed CpGV-M (Tanada, 1964), was stored at -20 °C as a stock virus.

### Re-selection of CpRR1

Re-selection of CpRR1 was initiated by exposing 200 fifth instar (L5) larvae and then 100 larvae at the fourth instar (L4) to a concentration of $2\times10^4$ OB/ml of CpGV-M. All surviving pupae were pooled and allowed to emerge to adults, which were then mass crossed. Eggs derived from these crosses were incubated to obtain first instar (L1) larvae, which were considered as F1 (generation 1). Then, F1 larvae were reared on diet containing either $2\times10^4$ OB/ml or $2\times10^5$ OB/ml of CpGV-M until they developed into adults. After further mass-crossing the L1 larvae hatching from the offspring of F1, were used for the next round of selection. Selection was repeated for five generations resulting in CpRR1_F5. A cohort of the F5 larvae was separated for resistance testing and further single-pair crossing experiments. The other cohort of F5 larvae were continued to bereared under virus exposure of $2\times10^4$ OB/ml until the seventh generation (CpRR1_F7). There were always 30-45 larvae fed on virus-free diet to keep the currently selected generation for the next round of selection. If there were no adult survivors during selection, the untreated adults were forwarded to mass crossing and their neonates were eventually used for the subsequent selection. A summary of the complete selection process is given in Figure 3-1.

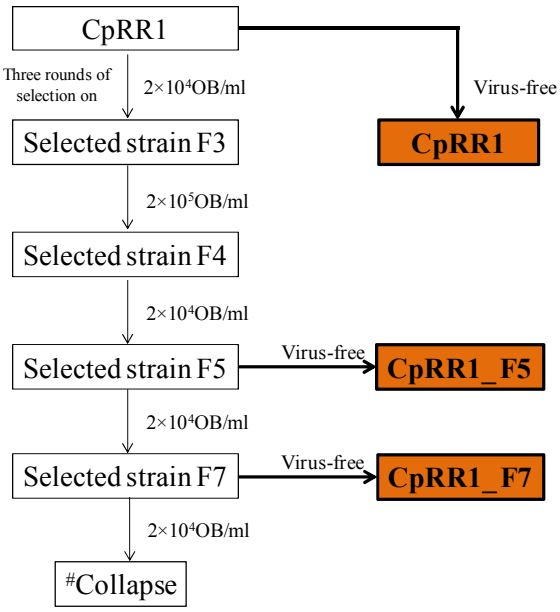



**Figure 3-1.** Selection process to regain high level of resistance of CpRR1. #Collapse of colony since only a very few male or only female adults survived during selection on CpGV-M.

**Bioassays**

Full range bioassays were conducted to determine the median lethal concentration ($LC_{50}$) of CpS, CpRR1 and CpRR1_F5. Neonate larvae of the different strains were placed into an autoclavable 50-well box containing a mixture of semi-artificial diet with a series concentration of CpGV-M involving $1\times10^3$, $3\times10^3$, $1\times10^4$, $3\times10^4$, $1\times10^5$ OB/ml for bioassay of CpRR1_F5. For CpS and CpRR1, different virus concentrations of $5\times10^2$, $1\times10^3$, $5\times10^3$, $1\times10^4$, $5\times10^4$ and $1\times10^5$ OB/ml) were used. Groups of 30-40 larvae were prepared for every concentration (for details see Eberle and Jehle, 2006). Larvae which died on the first day were assumed to be killed from handling and were therefore excluded in the following analyses. Dead larvae were recorded at 7 and 14 days post-infection (dpi). Each assay was then repeated independently two times. Bioassay data were corrected for untreated control mortality according to the formula of Abbott (1925). Median lethal concentration ($LC_{50}$) was determined by Probit analysis using ToxRat Standard Version 2.10 software (ToxRat Solutions GmbH, 2005).

**Single-pair crossings**

The selected generations CpRR1_F5 and CpRR1_F7 were reared on virus-free diet for at least two generations to obtain sufficient individuals. For single-pair crossings, pupae from susceptible (CpS) and CpRR1_F5 and CpRR1_F7 were separated by sex (Eberle and Jehle, 2006). After adults emerged, single males of CpS (CpSm) and females of CpRR1_F5 or CpRR1_F7 (CpRR1_F5f or CpRR1_F7f) were transferred to transparent plastic containers for mating and egg deposition. Crossings were also performed with single CpSf×CpRR1_F5m or CpRR1_F7m. Paired adults were fed on 10% sugar solution in a 35 mm Petri dish. Eggs were collected every two days and stored at 8-10 °C for a maximum of six days (Asser-Kaiser et al., 2010). Pooled eggs were incubated at the same conditions as mentioned above. Offspring of single-pair crossings were divided into two cohorts, one of which was used for bioassay and the other of which was set as untreated control. Neonate larvae were exposed to the discriminating concentration of $5.8\times10^4$ OB/ml, which was previously shown to cause >95%

mortality in susceptible CpS larvae at 7 dpi (Asser-Kaiser et al., 2007). Mortality was recorded at 7 and 14 dpi and corrected for untreated control mortality according to the formula of Abbott (1925).

The number of eggs produced by CpRR1_F5 and CpS in single-pair crossings was counted to evaluate the fecundity. The ratio of hatched neonates to all eggs was considered as representative of fertility. Differences in fecundity and fertility were then used to assess the fitness cost of CpRR1_F5 compared to CpS using Student's *t*-test in R software package (R version 3.4.4 in RStudio 1.1.442).

## Results

### Mass crossing selection

Re-selection of previously resistant level larvae was achieved by consecutive mass crossings of CpRR1 adults followed by exposure of L1 larvae to CpGV-M OB. For this, fourth (L4) and early fifth (L5) instars of CpRR1 larvae (CpRR1_F0) were exposed to $2\times10^4$ OB/ml of CpGV-M in diet during the rest of larval development (Table 3-1). Starting from the obtained progeny CpRR1_F1, two rounds of selection followed at the same concentration but with neonate (L1) larvae exposed to virus OB, resulting in generation of CpRR1_F3. Survival ranged between 22.0% and 7.7% for the different generations (Table 3-1). Then the OB concentration was increased to $2\times10^5$ OB/ml to select for CpRR1_F4 larvae. As the number of surviving adults decreased drastically to only six (3.1%), CpRR1_F4 larvae, the following generations CpRR1_F5 to CpRR1_F7 were further selected at the lower virus concentration of $2\times10^4$ OB/ml. Eventually, when the CpRR1_F7 population was built up, the selection could not be continued because the survival rate was only 0.6% and either only female or only male larvae developed to adults from four different crossing experiments, which prevented continuation of the selection experiment.

**Table 3-1.** Survival and adult sex ratio of CpRR1 progeny at each selected generation

| Generation | Larval instar | CpGV seletion (OBs $ml^{-1}$) | Surviving pupae (%) | Surviving adult (%) | Male (%) (m:f) |
|---|---|---|---|---|---|
| F0 | L5+L4 (300) | $2\times10^4$ | 27.3 (82) | 22.0 (67) | 53.7 (36:31) |
| F1 | L4 (100) | $2\times10^4$ | 11.0 (11) | 9.0 (9) | 66.7 (6:3) |
| F2 | L1 (326) | $2\times10^4$ | 8.6 (28) | 7.7 (25) | 64.0 (16:9) |
| F3 | L1 (196) | $2\times10^5$ | 4.1 (8) | 3.1 (6) | 16.7 (1:5) |
| F4 | L1 (363) | $2\times10^4$ | 17.6 (64) | 12.4 (45) | 53.3 (24:21) |
| F5 | L1 (95) | $2\times10^4$ | 2.1 (2) | 2.1 (2) | 0 (0:2) |
| F6[a] | L1 (92) | $2\times10^4$ | 5.4 (5) | 5.4 (5) | 60.0 (3:2) |
| F7[#] | L1 (163) | $2\times10^4$ | 0.6 (1) | 0.6 (1) | collapse (1:0) |

F6[a] was the progeny of F6♀ crossed with F5♂ population.

F7[#] population obtained from surviving larvae in this seventh selection, was used to continue with CpGV-M selection for four generations under the same concentration of $2\times10^4$ OBs/ml, but there was no-paired adult survival (only male survival adults or only female adult alive) causing collapse of the colony.

**Table 3-2.** Median lethal concentration ($LC_{50}$) observed for the resistant codling moth strains CpRR1, CpRR1_F5 and the susceptible strain of CpS infected with CpGV-M. Given are the total number (N) of tested L1 larvae, the $LC_{50}$ values at 7 and 14 days post infection (dpi), the 95% confidence interval of the $LC_{50}$s, slope and standard error (SE) of the probit line as well as $\chi^2$ value and degree of freedom (df), as well as the potency ($LC_{50\ strain}$ / $LC_{50\ CpS}$) of the different strains.

| Strain | N | 7 dpi | | | | 14 dpi | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $LC_{50}$ [$10^3$×No. of OBs/ml] (95% CI) | Slope ± SE | $\chi^2$ (df) | Potency | $LC_{50}$ [$10^3$×No. of OBs/ml] (95% CI) | Slope ± SE | $\chi^2$ (df) | Potency |
| CpS | 1075 | 1.99 (0.95-4.05) | 1.26 ± 0.01 | 30.81 (4) | 1 | 0.45 (0.21-1.19) | 1.01 ± 0.01 | 14.72 (4) | 1 |
| CpRR1 | 1550 | 127.39 (107.64-163.02) | 2.36 ± 0.09 | 2.37 (4) | 64 | 16.02 (14.16-18.12) | 1.97 ± 0.01 | 6.47 (4) | 36 |
| CpRR1_F5 | 362 | >20,600* | 0.45 ± 0.01 | 3.17 (3) | >10350 | 232.49 (98.09-1321.53) | 0.77 ± 0.02 | 4.26 (3) | 517 |

* not computable because value is beyond the tested concentrations by more than factor 1000; extrapolated estimation of $LC_{50}$ at 7 dpi for CpRR1_F5 was >$2.06\times10^7$ OB/ml based on probit line.

**Table 3-3.** Resistance testing of neonate larvae from single-pair crossings of CpRR1_F5 or CpRR1_F7 with CpS exposed to CpGV-M at a discriminating concentration of $5.8\times10^4$ OB/ml. Given are the crosses and the number of independent single pair crossings (N), total number of tested larvae (n), Progeny genotypes by hypothesis of Z-chromosomal inheritance (Z)[$], mean mortality and standard deviation (SD) after 7 and 14 days post infection (dpi), as well as the sex ratio of pupae; f, female; m, male,

| Strain | Crosses | N, n | Progeny genotypes, by hypothesis[$] | | Observed mortality at 7 dpi (%) | Observed mortality at 14 dpi (%) | Male (%) (m:f) |
|---|---|---|---|---|---|---|---|
| | | | (Z) | Exp (Z) at 7 dpi (%) | Mean ± SD | Mean ± SD | Sex ratio (m:f) |
| CpS | CpSf × CpSm | 4, 318 | $Z^SZ^S$, $Z^SW$ | 100 | 96.7 ± 3.9 | 98.3 ± 3.4 | - |
| CpRR1 | CpRR1f × CpRR1m | 5, 437 | $Z^RZ^R$, $Z^RW$ | 0 | 40.9 ± 4.3 | 79.2 ± 8.5 | - |
| CpRR1_F5 | CpRR1_F5f × CpRR1_F5m | 7, 363 | $Z^RZ^R$, $Z^RW$ | 0 | 49.2 ± 25.1 | 85.8 ± 14.5 | n.d. |
| CpRR1_F5 | CpRR1_F5f × CpSm | 8, 302 | $Z^RZ^S$, $Z^SW$ | 50 | 63.4 ± 15.9 | 87.2 ± 14.1 | n.d. |
| | CpRR1_F5m × CpSf | 11, 638 | $Z^RZ^S$, $Z^RW$ | 0 | 12.1 ± 8.9 | 57.4 ± 23.4 | 2:5 (28.6) |
| CpRR1_F7 | CpRR1_F7f × CpSm | 13, 518 | $Z^RZ^S$, $Z^SW$ | 50 | 16.0 ± 8.4 | 16.3 ± 13.1 | 7:10 (41.2) |
| | CpRR1_F7m × CpSf | 8, 452 | $Z^RZ^S$, $Z^RW$ | 0 | *3.8 ± 5.2 | *4.1 ± 7.7 | n.d. |

[$] based on hypothesis of dominant, Z-linked resistance inheritance in CpRR1 according to Asser-Kaiser et al (2007), progeny genotypes and expected mortality (%) (Exp) at 7 dpi were shown above. * indicates that negative value of Abbott (1925) corrected mortality in independent single-pair crossings (5 out of 8) were set as zero. "-" was not counted in the experiment. n.d., not determined because of fungal contamination at the pupal stage. Mortality rates of CM larvae were determined in bioassays with neonate larvae at 7 and 14 days post infection (dpi), exposed to CpGV-M at a discriminating concentration of $5.8\times10^4$ OB/ml.

### Resistance ratio

Median lethal concentration ($LC_{50}$) of CpS, CpRR1 and CpRR1_F5 against CpGV-M was determined to compare the susceptibility and the resistance level of the different CM colonies. At 7 dpi, the $LC_{50}$ values of CpGV-M in CpS and CpRR1 neonates were $1.99\times10^3$ and $1.2\times10^5$ OB/ml, respectively, In contrast, the $LC_{50}$ in CpRR1_F5 could not be determined due to low larval mortality, but it was estimated from the extrapolated probit line to be $>2.06\times10^7$ OB/ml. At 14 dpi, the $LC_{50}$ values of CpS, CpRR1 and CpRR1_F5 were $4.5\times10^2$, $1.60\times10^4$ and $2.32\times10^5$ OB/ml, respectively (Table 3-2). Thus, the resistance ratio based on $LC_{50}$ values of CpRR1 compared to CpS increased by 64-fold and 36-fold at 7 and 14 dpi, respectively. For CpRR1_F5, the resistance ratio was increased 517-fold at 14 dpi. Since most CpRR1_F5 larvae were still alive at 7 dpi, its estimated resistance ratio compared to CpS was >10,350-fold (Table 3-2).

### Single-pair crossing

Resistance testing using the discriminating concentration of $5.8\times10^4$ OB/ml resulted in 97% mortality at 7 dpi, proving the susceptibility of CpS, whereas CpRR1 and CpRR1_F5 showed a mortality of 41% and 49%, respectively (Table 3-3). To further analyze the mode of resistance

in CpRR1_F5 and CpRR1_F7 individual, single-pair backcrosses with susceptible CpS followed by resistance testing was performed. When CpRR1_F5 females were crossed with CpS males, mortality was 63% at 7 dpi, slightly higher than the expected 50% mortality (Table 3-3). In CpRR1_F5m × CpSf crosses, induced mortality was 12% at 7 dpi, also slightly exceeding the expected 0%. At 14 dpi mortality increased to 87% and 57% for CpRR1_F5f × CpSm and CpRR1_F5m × CpSf, respectively, indicating high larval mortality with increasing incubation time. Most of the infected insects died prior to pupation as only two male and seven female pupae were obtained from a total of 638 infected progenies in 11 single-pair crossings.

Induced mortality of offspring from CpRR1_F7f × CpSm was 16%, at both 7 dpi and 14 dpi, which was much lower than the predicted 50% mortality at 7 dpi. In particular, seven male and ten female larvae survived and developed into pupae in this bioassay, which was also not compatible with the model of a Z-linkage of resistance (Table 3-3). In the CpRR1_F7m × CpSf crossings mortality was only 4% at 7 dpi and 14 dpi.

**Fitness cost**

The average number of eggs of CpRR1_F5f was significantly lower than CpSf in single-pair crossings (Figure 3-2) (*T*-test, $p < 0.05$, $\alpha = 0.05$) using Student's *t*-test. However fertility of CpRR1_F5f (51.9%) and CpSf (59.8%) was similar to one aother (*T*-test, $p = 0.28$, $\alpha = 0.05$). These observations showed that there were some fitness costs in fecundity in the CpRR1_F5 strain when resistance was partially recovered.

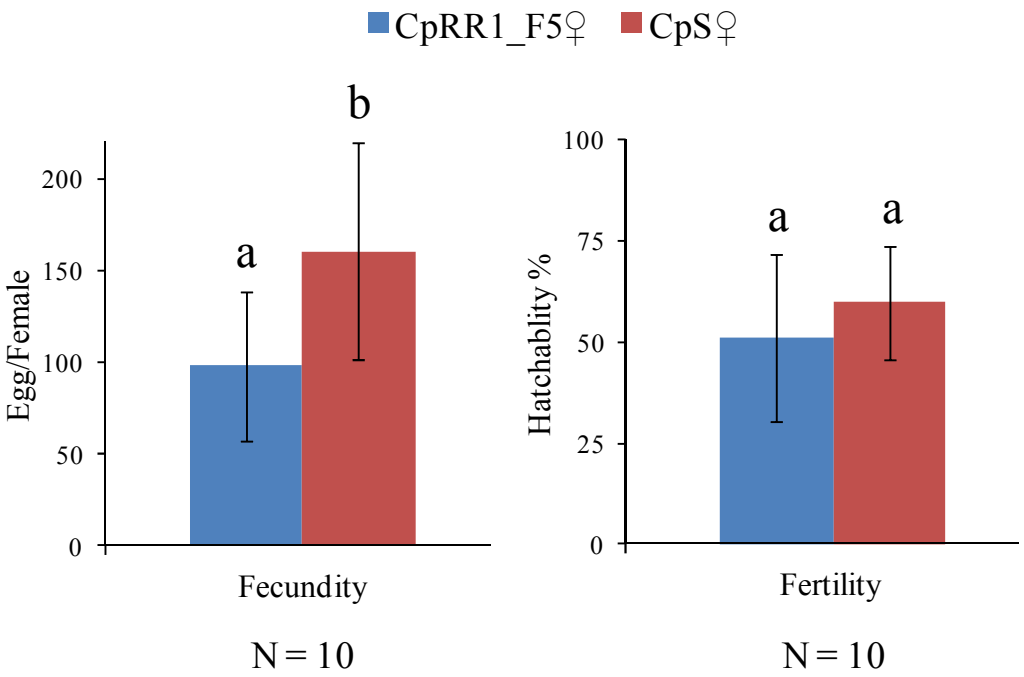



**Figure 3-2** Fecundity and fertility of CpRR1_F5 and CpS. The eggs resulting from each 10 (= N) single-pair crossings of CpRR1_F5 × CpRR1_F5 and CpS × CpS, respectively, were counted to estimate their fecundity. The fertility was assessed the percentage of successfully hatched eggs. Error bar indicate standard deviation (SD).. The different letters indicate statistically significant differences (*T*-test, $p < 0.05$, $\alpha = 0.05$).

## Discussion

The resistant CM strain CpRR1 was selected in 2006/2007 from the resistant field population CpR (Asser-Kaiser et al., 2007). In previous bioassays, CpGV-M did not cause significant mortality in CpRR1 larvae and the $LC_{50}$ was estimated to be $6.92\times10^{8}$ OB/ml and $8.53\times10^{6}$ OB/ml at 7 dpi and 14 dpi respectively, suggesting a resistance level of $10^{4}$-$10^{5}$ fold (Asser-Kaiser et al., 2007; Asser-Kaiser et al., 2011). Since then, the CpRR1 strain has been reared on semi-artificial diet under virus-free conditions in the laboratory. In 2011, a re-selection was performed with single-pair crossings (Pietruska, 2018). When tested in 2015, CpRR1 appeared to be more susceptible to CpGV-M than in previous tests (Chapter 2). This finding was corroborated by an $LC_{50}$ determination, demonstrating a considerable decline of the resistance level of the current CpRR1 rearing to only 64- and 36-fold resistance in 7- and 14-day bioassays, respectively (Table 3-2). When CpR was reared for more than 60 generations without CpGV pressure under laboratory conditions, a more or less stable resistance level and no fitness costs of resistance was observed (Undorf-Spahn et al., 2012). Other examples from literature showed that laboratory-selected resistance of *Adoxophyes honmai* against Adoxophyes honmai nucleopolyhedrovirus (AdhoNPV) can be stable at a 800-fold level after stopping selection for 168 generations (Nakai et al., 2017). Therefore, it is hypothesized that the CpRR1 strain lost full expression of its original resistance trait because of long-term laboratory rearing without CpGV pressure.

To elucidate the reasons for the decline of resistance and to re-establish a highly resistant CM colony, seven consecutive mass crossings of CpRR1 adults, each followed by selection of progeny larvae on CpGV-M were carried out. The OB concentration used ranged from $2\times10^{4}$ OB/ml to $2\times10^{5}$ OB/ml. After five generations, the selection line CpRR1_F5 was estalished with an estimated $LC_{50}$ of $2.06\times10^{7}$ OB/ml at 7 dpi, representing a $10^{4}$-fold resistance compared to CpS, which was about 35 times lower than that in the original CpRR1 (Asser-Kaiser et al., 2007). When mortality after 14 days is considered, the resistance factor of CpRR1_F5 was about 517, again about 20 times lower than the $10^{4}$-fold level observed in previous experiments. Further selection to induce higher resistance failed when CpRR1_F7 was established, because the colony was lost due to high mortality. It cannot be ruled out that inbreeding effects caused the loss of CpRR1_F7. Female adults of CpRR1_F5 produced significantly fewer eggs than CpS females, which could be an effect of inbreeding or caused by fitness costs related to resistance. As shown in Table 3-1, only a very low number of adults survived during the selection process, suggesting that inbreeding may have reduced the genetic diversity present in the selection lines. Mass-crossing selections have been successfully applied to select genetically homogenous CM strains resistant to baculoviruses: The French selection line RGV-8 was established through mass-crossings using a discriminating concentration of CpGV-M causing 98% mortality of susceptible CM larvae for eight generations (Sauphanor et al., 2006; Berling et al., 2009a) and the obtained RGV-8 colony showed a dominant, sex-linked resistance pattern with a 7,000-fold resistance level compared to susceptible CM (Berling et al., 2013). Similarly, the CM strain CpR-CZ originated from the survivors of three mass crossings of a resistant field colony, showing a dominant, Z-linked inheritance pattern (Zichová et al., 2013). It is assumed that mass-crossed RGV-8 and CpR-CZ harbor a more heterogeneous genetic background as this was the case for CpRR1, which was developped from a very few individuals. Theoretically, the CpRR1 strain originating from single resistant pairs should not contain susceptible individuals if resistance is stably inherited, even without further selection of resistance. This is apparently not the case (Asser-

Kaiser et al., 2007). Resistance to CpGV-M in CpRR1_F5 and CpRR1_F7 was regained by selection on CpGV-M, though its resistance level was about 30-50 times lower than that of the original CpRR1 strain (Table 3-2 and 3-3). The low survival of pupae and low emergence of adults observed during the experiments and finally leading to the collapse of the selection line CpRR1-F7, a further indication that the original level of resistance of CpRR1 could not be recovered.

According to the well-known mode of type I resistance of CpRR1, dominant resistance factor(s) are located on the Z chromosome and follows classical dominant Mendelian inheritance (Asser-Kaiser et al., 2007). Single-pair crossings carried out between the newly established CpRR1_F5 or CpRR1_F7 strains and susceptible CpS larvae as well as among CpRR1_F5 larvae revealed the interesting phenomenon that their resistance did not fully comply with the dominant, Z-linked inheritance mode. First, the mortality of progeny of CpRR1_F5f × CpRR1_F5m was nearly 50% after 7 days and was therefore much higher than expected from the $LC_{50}$ determination. Second, the observed mortality in single-pair crossings of CpRR1_F5f × CpSm and CpRR1_F5m × CpSf was 63% and 12%, respectively, and thus slightly higher than the expected value of 50% and 0%, respectively (Table 3-3). Though the observed mortality rates were still in favor of a dominant Z-linked inheritance, yet at a lower resistance level, the finding of two surviving males and five surviving female pupae from the CpRR1_F5m × CpSf crosses cannot be explained by the previous resistance model established by Asser-Kaiser et al. (2007), which would have predicted no survival of heterozygous $Z^RZ^S$ males and susceptible $Z^SW$ females from these crosses. Third, reciprocal single-pair crosses of CpRR1_F7 × CpS resulted in low mortality of only 16% and 4%, after 7 and 14 days, indicating that there was no strong Z-linkage of resistance in this line (Table 3-3). Strikingly, seven male (expected $Z^RZ^S$) and ten female (expected $Z^SW$) pupae survived the experiment, also contradicting the previous resistance model (Asser-Kaiser et al., 2007). Thus, these results overturned the sex-linked pattern of resistance because expected $Z^SW$ female larvae would not survive in the 7-day resistance test, nor would they pupate. Survival of male ($Z^RZ^S$) and female ($Z^SW$) pupae in the resistance tests hinted that the resistance factor in the Z chromosome was no longer fixed.

These findings are somewhat reminiscent of the observation made with the resistant CM strain CpRGO (type III resistance) (Sauer et al., 2017b). In that case, no coherent Mendelian pattern of Z-linked or autosomal inheritance was found for resistance to CpGV-M. For CpRGO, the results could be partly explained by a combination of Z-linked and autosomal traits contributing to a heterorgeneous response when progeny of different crossings were challenged with CpGV-M. A similar situation appeared again in the CpRR1_F7 after seven rounds of re-selection. Interestingly, despite a strong dominant, Z-dependent inheritence controlled by a single major gene, additional factors not subjected to the rules of Mendelian inheritance were also observed for the French colony RGV-8 (Berling et al., 2013).

In conclusion, the type I resistance of CpRR1 is not fully stable and was lost in part during many generations of virus-free rearing. Re-selection recovered resistance, yet not at its previous level. Though a dominant resistance mode could be determined after re-selection, its linkage to the Z chromosome could not be fully verified. These findings indicate a certain instability and loss of resistance in CpRR1 as well as potential autosomal factors involved in its expression.

## Chapter IV: Single nucleotide polymorphism (SNP) frequencies and distribution reveal complex genetic composition of seven novel natural isolates of Cydia pomonella granulovirus


### Abstract

As a consequence of the co-evolution between baculoviruses and their insect hosts, pathogen-host interaction prompts mutual adaptation enabling the co-existence of both the virus and the insect. This process results in selection of virus populations, which are geographically and sometimes even functionally distinct. To explore this phenomenon at the molecular level, seven recent, naturally occurring isolates of Cydia pomonella granulovirus (CpGV) derived from infected codling moth larvae from orchards in northwest China were studied. The genomes of these CpGV isolates were sequenced via Illumina next generation sequencing (NGS) and compared to those of previously characterized CpGV isolates for insertion and deletion mutations, single nucleotide polymorphisms (SNPs) and phylogenetic relationship. A total of 563 SNP positions were detected in comparison with previously known CpGV isolates. New members of previously defined phylogenetic genome groups A, D and E of CpGV, as well as two new phylogenetic lines, termed genome group F and G, were identified. Combining SNP frequency distribution with the prevalence of genome group specific SNPs, revealed that six isolates of CpGV, CpGV-ZY, -JQ, -ALE, -KS1, -KS2 and -ZY2 were mixtures of different ratios of at least two genotypes, whereas only CpGV-WW was genetically highly homogeneous. Strikingly, variable proportions of CpGV-WW were found in the other six new CpGV isolates. Considering the economic importance of CpGV, which is used worldwide as a biocontrol agent of codling moth larvae in pome fruit production, this study significantly extends our current understanding of the genetic diversity of CpGV and opens new lines of application of this virus.


### Introduction

Cydia pomonella granulovirus (CpGV) has been studied for over 50 years since the first isolate, termed CpGV-M, was discovered in infected larvae of codling moth (CM), *Cydia pomonella* L. from Mexico in 1963 (Tanada, 1964). CpGV has a covalently closed dsDNA genome and belongs to the insect virus family *Baculoviridae*, which includes a large number of viruses isolated from Lepidoptera, Hymenoptera and Diptera (Herniou et al., 2011; Rohrmann, 2013) and genus *Betabaculovirus*. In addition to CpGV-M, many geographic isolates of CpGV have been identified and characterized during the last decades, including isolates from Russia (CpGV-R) (Harvey and Volkman, 1983), England (CpGV-E) (Crook et al., 1985), Iran (CpGV-I01, -I07, -I08, -I12, -I66, -I68) (Eberle et al., 2008; Rezapanah et al., 2008; Eberle et al., 2009), Georgia (CpGV-G01, -G02) (Eberle et al., 2009), Argentina (CpGV-2.17, -3.8, -6.9, -6.16, -P118, -Col19, -C1, -C6, -M3, -M10, -M18, -P7) (Arneodo et al., 2015), and three isolates from China (CpGV-ZY, -KS1, -KS2) (Gan et al., 2011; Zheng et al., 2011). To understand the diversity of CpGV at the molecular level, the genomes of nine isolates, namely CpGV-M1, -M, -I07, -I12, -S, -E2, -0006, -R5, V15, have been previously

sequenced; the genome size of these isolates ranged from 120.8 to 124.3 kbp, encoding 137-142 ORFs (Luque et al., 2001; Gueli Alletti et al., 2017; Wennmann et al., 2017).

Whereas the CpGV genome sequences revealed high conservation in size, GC content, ORF quantity and transcription direction, as well as the distribution of insertions and deletions (indels) and single nucleotide polymorphisms (SNPs), the phylogenetic analyses based on concatenated nucleotide sequences of single genes (*granulin* and *late expression factor 8* (*lef-8*)) or concatenated amino acid sequences of the 35 baculovirus core genes or whole genome alignments revealed five CpGV lineages, which were eventually termed genome groups A to E (Eberle et al., 2009; Gebhardt et al., 2014; Arneodo et al., 2015; Wennmann et al., 2017).

Several CpGV isolates have been developed as biological control agents of CM larvae in apple and pear plantations (Huber, 1998; Lacey et al., 2008). CM is distributed to all temperate climate zones worldwide and can cause serious economic loss to apple, pear, apricot, plum, peach, nectarine and walnut (Barnes, 1991; Willett et al., 2009; Men et al., 2013). Nowadays CpGV is one of the most important commercial baculovirus biocontrol agents, being registered in nearly all apple growing countries worldwide. Since 2005, three distinguishable forms of field resistance (type I to III) have been identified in apple plantations. CM populations resistant to CpGV isolates belong to genome groups A, D and E, and only the isolates from genome group B and C were able to break all three types of resistance (Sauer et al., 2017b).

The occurrence of SNPs in viral nucleotide sequences has been widely used to identify virus genotypes (Loparev et al., 2004; Williams et al., 2010; Cornman et al., 2013; Iyer et al., 2015). SNP distributions have been further applied to identify the genetic diversity within a given virus isolate and to distinguish virulence variations among different virus populations (Hoen et al., 2013; Sangket et al., 2015). For example, Loparev et al. (2004) classified 326 varicella-zoster virus samples into E, M and J type strains using specific SNPs in ORF22 as a genotype marker. In terms of overcoming CpGV resistance, it is essential to identify specific SNPs located in resistance-breaking CpGV isolates, as a pool of potential determinants of isolate composition and virulence. In fact, Wennmann et al. (2017) recorded 788 potential nucleotide variations in CpGV on the basis of alignments of five CpGV genomes, namely CpGV-M, -E2, -I07, -I12, -S, representing the genome groups A to E. In practice, 534 specific SNPs were selected as genetic markers for the different CpGV genome groups, namely group A (2 SNPs), B (54), C (356), D (21), E (101), to investigate the population structure of commercial CpGV isolates (Gueli Alletti et al., 2017). By applying SNP-based genome grouping it was possible to correlate isolate composition with the capacity to break different types of resistance, e.g. isolate CpGV-V15 was found to be composed of similar ratios of group B and group E CpGV and can break type I-III resistance in laboratory bioassays as well as in field applications (Zingg et al., 2011; Gueli Alletti et al., 2017).

Moving forward to explore the population structure and phylogeny of geographic CpGV isolates, Illumina Solexa next generation sequencing (NGS) and bioinformatic analyses were applied to study seven CpGV isolates, recently collected from northwest China (Gan et al., 2011; Zheng et al, 2011; Chapter 2). A comprehensive SNP mapping was developed to determine the composition and frequency of different genome groups within the population of a given isolate. By applying phylogenetic analysis we were able to determine two new lineages in the CpGV members. The identified SNP patterns of CpGV can be used for isolate identification, studying their geographic distribution and sheds new light on the genetic

diversity of CpGV in particular and intra-specific genome complexity of bacauloviruses in general.

## Material and methods

### CpGV isolates and propagation

Seven CpGV isolates, designated as CpGV-ALE, -JQ, -KS1, -KS2, -WW, -ZY, and -ZY2 derived from single diseased CM larvae collected in northwest China as described in Chapter 2. The isolates were propagated in CpS (susceptible CM) to obtain an adequate amount of occlusion bodies (OBs) (Chapter 2). Genomic DNA samples were then purified from OBs using standard methods (Arends and Jehle, 2002; Gueli Alletti et al., 2017). The quantity and quality of viral DNA preparations were verified by measuring their absorption at 260 nm, 280 nm and 320 nm using a NanoDrop 2000 (Thermo Fisher Scientific, MA, USA).

### Genome sequencing and assembly

At least 100 ng purified viral DNA of each isolate was used for paired-end sequencing with Illumina NextSeq500$^{TM}$ platform at StarSEQ GmbH (Mainz, Germany). Paired-end reads were 151 nucleotides in length. For the generation of isolate specific consensus sequences, all bioinformatic steps were conducted with software Geneious R9 (Biomatters, New Zealand) in a combination of mapping reads against internal reference sequence CpGV-M (GeneBank number: KM217575) and use of reads for *de novo* assembly. For all steps, reads were adapter trimmed and quality filtered with a Phred score above 30, corresponding to a 99.9% base call accuracy, and a minimal read length of 50 nt. In a first approach, all reads of each isolate were mapped against CpGV-M by Bowtie 2 implemented in Geneious R9. A preliminary consensus sequence was extracted from the read assembly by picking bases exceeding the frequency of 60% (Day and McMorris, 1992). In a second step, all reads of each isolate were *de novo* assembled by using the Genome assembler implemented in Geneious. Congtigs longer than 1,000 nt were kept and mapped against the preliminary consensus. For all seven Chinese isolates, the contigs comprised the entire CpGV genome, indicated by the genes *granulin* and *me53*, defined as the first and last ORF of the linearized circular genome, respectively. Differences between contigs and preliminary consensus were carefully evaluated and further determined by local Sanger sequencing if needed. In a last step to confirm the consensus sequence, the individual isolate consensus sequences were used as reference genomes for mapping of isolate reads. This final step was conducted by using BWA-MEM algorithm (Li, 2011b) implemented in the Galaxy server of JKI.

### Phylogenetic analysis

For phylogenetic analysis, the consensus sequences of CpGV-ZY, -JQ, -ALE, -KS1, -KS2, -ZY2, and -WW were aligned with those of CpGV-M, -S (KM217573.1), -I07 (KM217574.1), -I12 (KM217576.1), -E2 (KM217577.1) and Cryptophlebia leucotreta granulovirus (CrleGV) (NC_005068.1), chosen as an outgroup, using progressive Mauve algorithm in Geneious (Kearse et al., 2012). A phylogenetic tree was constructed using maximum likelihood algorithm with Best-Fit substation model implemented in MEGA6 (Tamura et al., 2013).

### SNP analyses

Quality-filtered reads ($Q \geq 30$) of the seven sequenced isolates were assembled against the internal reference genome CpGV-M using BWA-MEM on the Galaxy server of JKI. All

calculated sequence alignment map (SAM) files were further processed with MPileup to detect single nucleotide variant sites. Monomorphic sites were filtered with the BCF tool (Li, 2011b) to obtain variant sites only. The variant call format (VCF) file was then loaded and processed in R (R version 3.3.1 in RStudio 1.0) using Variant Annotation package (Obenchain et al., 2014). The following criteria were set to filter significant variable sites for each isolate: (i) a variable SNP position should be supported by a minimum coverage of 100, and (ii) the detected alternative nucleotide must exceed a coverage of 10. Variant sites that did not fulfil these criteria were considered as minor variants. The entire number of detected SNPs over all seven Chinese isolates were split into three groups: (i) previously detected and specific SNP positions for genome groups A to E (Wennmann et al., 2017), (ii) newly detected SNP positions and their specificity to CpGV groups designed as genome group F, G, as well as (iii) newly detected SNPs that were shared by more than one isolate, termed SN.

The genetic diversity of each Chinese CpGV isolate was calculated based on the previously detected SNPs as well as newly detected SNP positions specific for different isolates as described by (Gueli Alletti et al., 2017). The average frequencies of major SNPs represent the genotype of seven new isolates respectively. In addition, all quantified reads containing 1×12 bp (GACACAGTGGAT, amino acid: DTVD) repeat region in *pe38* were picked up to calculate the proportion of this repeat region in seven new CpGV isolates.

## Results and Discussion

### Genome assembly and phylogeny

To analyze the genetic diversity of seven new Chinese CpGV isolates, namely CpGV-ALE, -JQ, -KS1, -KS2, -WW, -ZY, and -ZY2, DNA was purified from occlusion bodies (OBs) and submitted to Illumina Solexa paired-end sequencing. The number of reads of each isolate ranged between 607,566 (CpGV-ZY) and 1,181,528 (-ALE) resulting in a mean nucleotide coverage between 594 and 1,075, respectively (Table 4-1). A consensus sequence of each isolate was obtained by the combination of two genome assembly approaches, comprising a *de novo* read assembly as well as a mapping of reads to the reference genome of CpGV-M (Larem et al., 2019). The isolates´ consensus sequences were extracted based on the criterion of 60% nucleotide majority in the mapped reads. The consensus sequence of the different isolates was about 123,700 bp, ranging from 123,669 bp (CpGV-JQ) to 123,759 bp (-ZY), and was thus similar to previously sequenced CpGV isolates (Gebhardt et al, 2014, Wennmann et al., 2017). The consensus sequence of only CpGV-WW was considerably shorter; with 123,392 bp in length it was slightly larger than CpGV-S (123,193 bp) (Wennmann et al 2017; Gebhardt et al., 2014). The genomes of all isolates had a GC content of 45.3%, also similar to previously sequenced CpGV isolates.

The obtained consensus nucleotide sequences were aligned with those of CpGV-M, -I12, -S, -E2 and -I07 (Wennmann et al., 2017) for phylogenetic analysis. As shown in Figure 4-1, some of the new CpGV isolates fitted well into the basic lineage structure of CpGV genome group classification, but with some exceptions. Whereas CpGV-KS2, -ZY and -KS1 clustered with CpGV-M and -I12 (genome group A and D), CpGV-ALE was found as a new sister lineage to them, termed group G. Similarly, CpGV-WW appeared closely related to CpGV-S (genome group E), whereas CpGV-JQ and -ZY2 constituted a new group F. There were two reasons why the whole genome consensus sequences were used for the phylogenetic analysis. Firstly, the highly conserved genomes showed only few dissimilarities in both genome size and

arrangement (Table 4-1, Figure 4-2) allowing a very robust multiple alignment. Secondly, by including all informative nucleotide positions of the genomes, more informative sites were included than in previous phylogenies based on single gene comparisons or baculovirus core genes only (Eberle et al., 2009; Gebhardt et al., 2014; Wennmann et al., 2017). Compared to previously studied Argentinean and Iranian CpGV isolates, which comprised isolates either from genome groups A and B or from genome groups A, C and D, respectively (Eberle et al., 2009; Arneodo et al., 2015), the novel isolates from northwest China revealed a striking genetic diversity covering four phylogenetic lineages, two previously described ones (A/D and E) and two new ones (F and G).

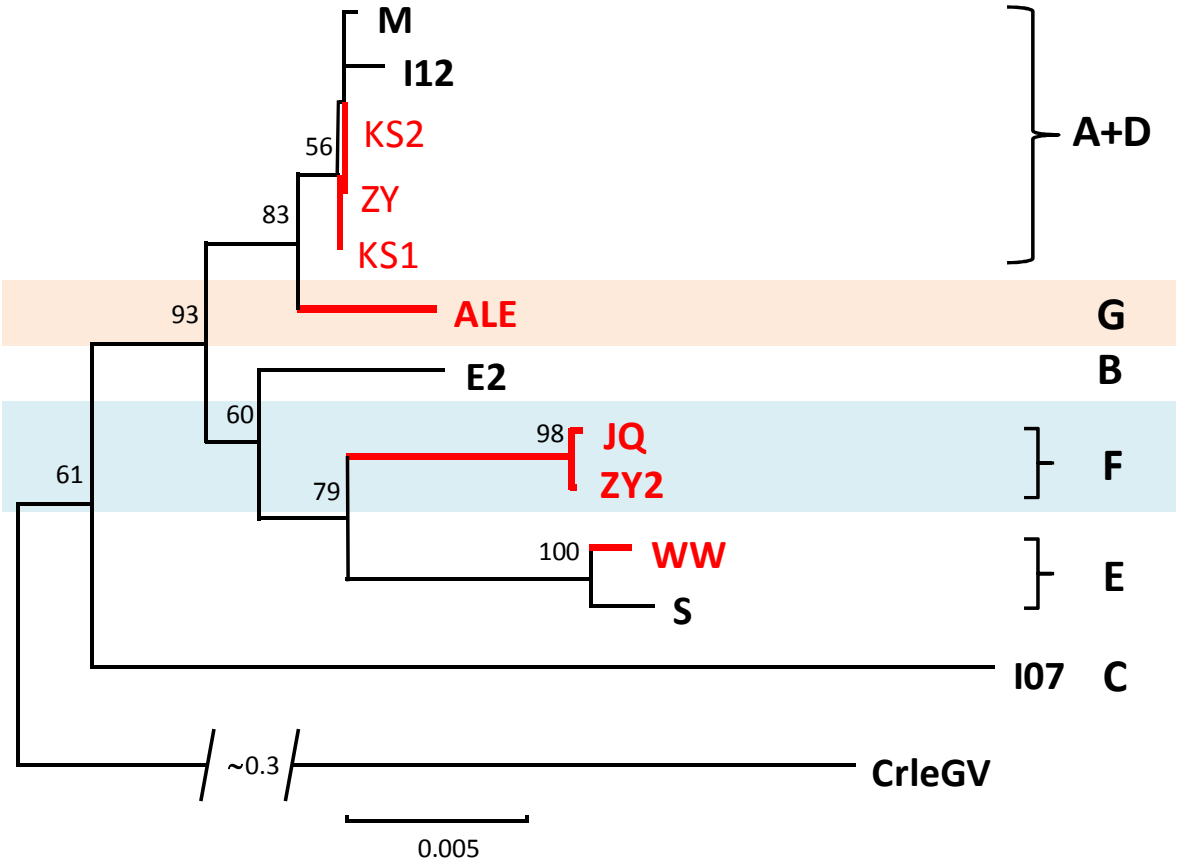



**Figure 4-1.** Phylogenetic analysis of twelve CpGV isolates. The phylogenetic tree was conducted on the basis of maximum likelihood method of the twelve whole genome nucleotide sequence alignment with 500 bootstrap replicates. Genome group A to F including two new groups (G and F) were placed behind the isolate name. Cryptophlebia leucotreta granulovirus (CrleGV) was chosen as the outgroup for rooting the CpGV tree. The seven studied CpGV isolates are indicated in red. The new genome groups G and F are highlighted with light orange and blue background, respectively.

**Table 4-1.** Statistics of assembled reads (Phred-quality score ≥ 30) against CpGV-M and genome characteristics of seven Chinese isolates.

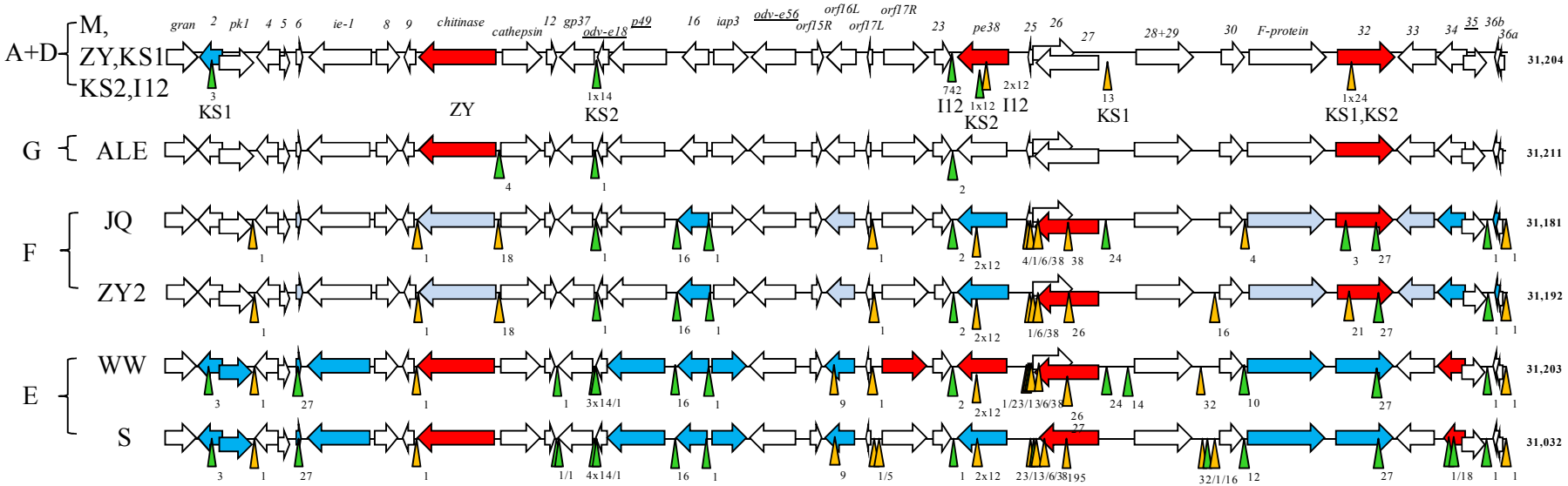
| CpGV isolates | No. and percentage (%) of assembled reads | Mean coverage ± SD | Genome consensus size (bp) | No. of ORFs | No. of SNPs | No. of Indels (1-210 bp) |
|---|---|---|---|---|---|---|
| KS2 | 795,937 (90.5) | 776 ± 148 | 123,702 | 142 | 197 | 13 |
| ZY | 607,621 (93.4) | 594 ± 121 | 123,759 | 142 | 230 | 6 |
| KS1 | 1,092,892 (96.3) | 1056 ± 214 | 123,702 | 142 | 197 | 16 |
| ALE | 1,092,435 (92.5) | 1075 ± 208 | 123,705 | 142 | 265 | 11 |
| JQ | 965,856 (94.4) | 949 ± 184 | 123,669 | 142 | 386 | 43 |
| ZY2 | 987,529 (99.1) | 967 ± 195 | 123,692 | 141 | 339 | 44 |
| WW | 953,716 (99.7) | 928 ± 203 | 123,392 | 141 | 192 | 58 |

CpGV-M was used as the reference genome for SNP calling and indels determination. GC content of each Chinese isolate was 45.3%. No. is the total number.

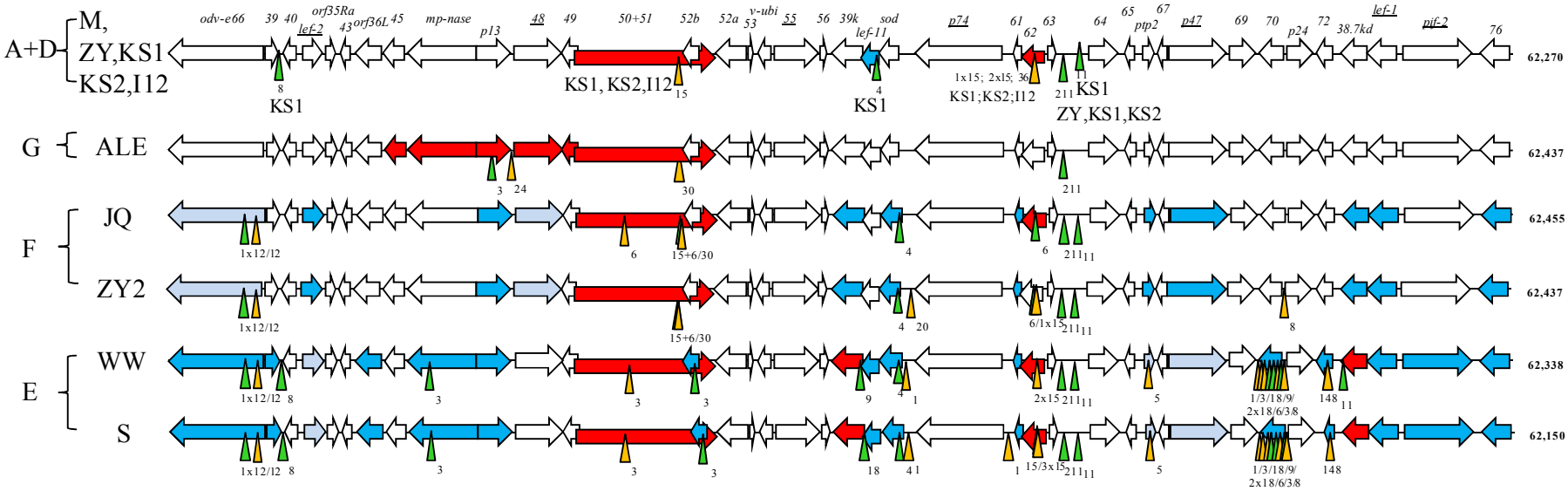
### Genome annotation and comparison

When ORFs were mapped and compared to other CpGVs, 142 ORFs were found for most isolates (Table 4-1, Figure 4-2). Only the putative ORF25 (*cp25*), located in a repeat region, was missing in the two isolates CpGV-ZY2 and -WW, reducing their number of ORFs to 141 (Figure 4-2). ORF26 (*cp26*) could be identified in all seven isolates but not in CpGV-S (Wennmann et al., 2017), which is most closely related to -WW. On the other hand, the new CpGV isolates contained between 192 and 386 single nucleotide polymorphisms (SNPs) (detailed analysis below) and between 6 and 58 indel mutations when compared to the CpGV-M genome as a reference (Table 4-1, Figure 4-2). They all shared a common 211 bp insertion with CpGV-S at the genome position 51,992 (between ORF63 and ORF64), which is lacking in other CpGV isolates (Figure 4-2).

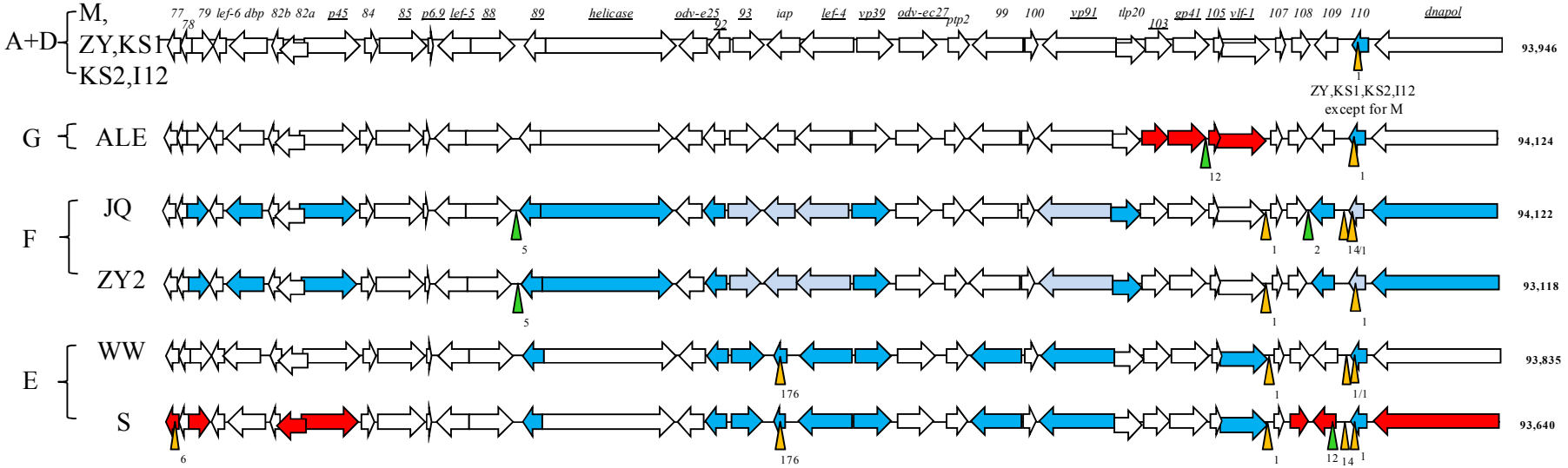
## Group Isolate




## Group Isolate



**Group Isolate**




**Group Isolate**

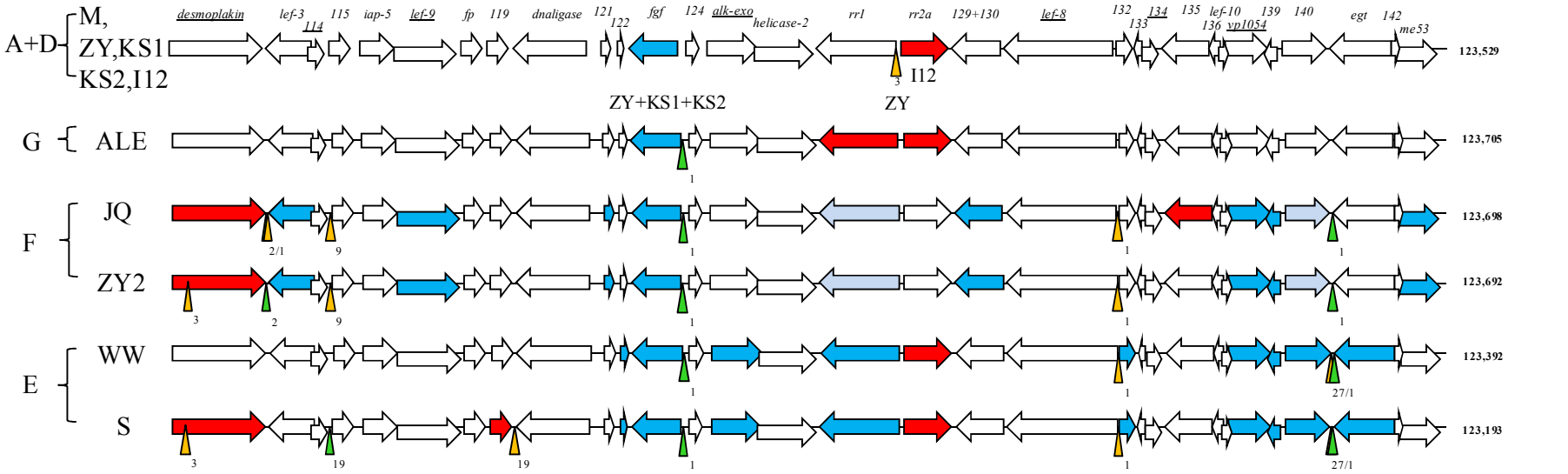


**Figure 4-2.** Graphic illustration of genome differences of seven Chinese isolates CpGV-ZY (ZY), -KS2 (KS2), -KS1 (KS1), -ALE (ALE), -JQ (JQ), -ZY2 (ZY2), -WW (WW) compared to CpGV-M (M), -I12 (I12) and -S (S) (Wennmann et al., 2017). CpGV-M was selected as the reference genome sequence. ORFs (open reading frames) were plotted with different size and orientation of arrows. Color code of ORFs: white = ORFs' amino acid (aa) sequence of the indicated genome is identical to that in CpGV-M, red = ORFs' aa sequence is unique to their own, blue and darkblue (80% lighter) = ORFs' aa shared in the same sequence, but are different from that in CpGV-M. Insertion and deletion (indel) were are represented by the arrows using orange and green triangles. Isolate specific mutations were marked below ORFs.

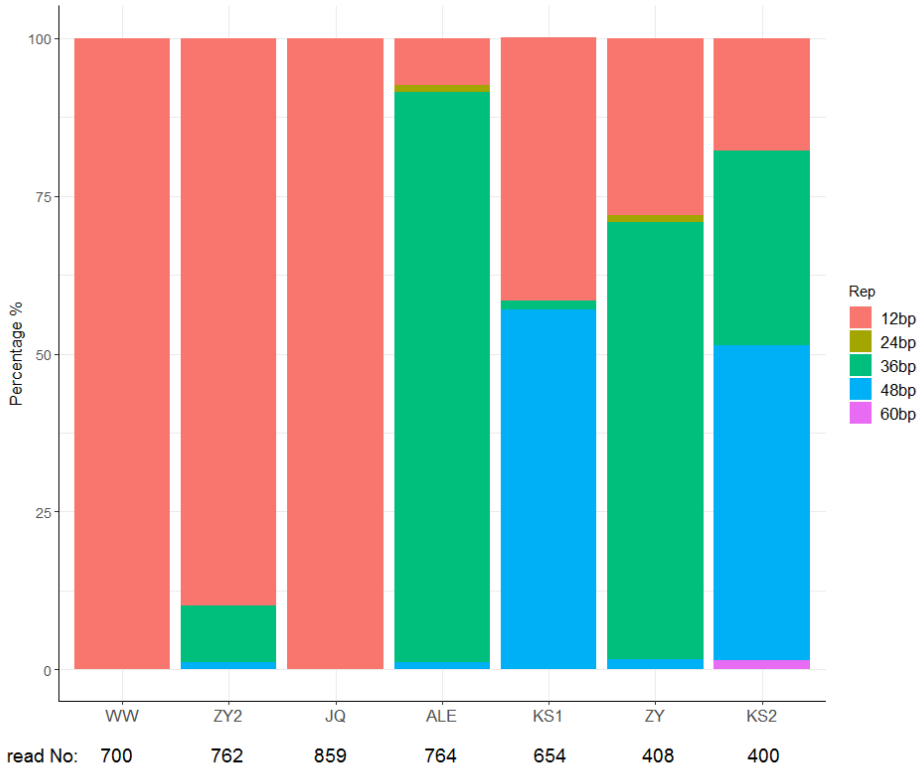



**Figure 4-3.** Proportion of 12 bp repeat unit GACACAGTGGAT (amino acid: DTVD) in *pe38* of seven CpGV isolates. CpGV-WW (WW), -ZY2 (ZY2), -JQ (JQ), -ALE (ALE), -KS1 (KS1), -ZY (ZY) and -KS2 (KS2) contain different ratios of one, two, three, four or five times repeats of 12 bp. Total number of reads covering the repeat region in each isolate are indicated below.

**Table 4-2.** Repetitive region of 12 bp in pe38 of seven new CpGV isolates

| Isolates | Read number in differently repetitive types | | | | | |
|---|---|---|---|---|---|---|
| | Total | 1×12 bp | 2×12 bp | 3×12 bp | 4×12 bp | 5×12 bp |
| WW | 700 | 700 | 0 | 0 | 0 | 0 |
| ZY2 | 762 | 692 | 0 | 69 | 1 | 0 |
| JQ | 859 | 859 | 0 | 0 | 0 | 0 |
| ALE | 764 | 57 | 1 | 705 | 1 | 0 |
| KS1 | 654 | 272 | 0 | 3 | 379 | 0 |
| ZY | 408 | 118 | 1 | 286 | 3 | 0 |
| KS2 | 400 | 72 | 0 | 128 | 200 | 2 |

### The pe38 repeat region

A most significant indel mutation related to the biological activity of CpGV is located in *pe38,* because this gene had been proposed to be involved in breaking type I resistance to CpGV (Gebhardt et al., 2014). *Pe38* is a poorly conserved baculovirus viral transcription factor, sharing about 26% amino acid identity with others (Luque et al., 2001). A deletion of *pe38* in the genome of Autographa californica multiple nucleopolyhedrovirus (AcMNPV) had an impact on virus replication and BV production, which resulted in a lower virulence in *Heliothis virescens* larvae, indicating that *pe38* plays an important role during *in vivo* infection (Milks et al., 2003). For CpGV, a 2×12 bp repeat in *pe38*, not present in resistance-breaking isolates, was identified as a target of type I resistance to CpGV-M (Gebhardt et al., 2014). The *pe38* repeat region was either 3×12 bp (CpGV-M) or 1×12 bp (CpGV-S, -E2, -I12 and -I07) in length (Gebhardt et al., 2014). Polymorphisms of *pe38* of the Chinese CpGV isolates have been already suggested by PCR and Sanger sequence analyses of the *pe38* region (Chapter 2). However, the NGS data allowed a more detailed insight of this region due to the high sequencing coverage; to analyse the heterogenity of *pe38*, all those reads encompassing the entire *pe38* repeat region, defined by non repetitive flanking sequences present on single reads within the data set of reads were counted (Figure 4-3 and Table 4-2). By this approach, between 402 (CpGV-KS2) to 859 (CpGV-JQ) reads were successfully detected in the isolate data set encompassing the entire repeat region (Figure 4-3). Within the seven sequenced isolates, CpGV-WW and -JQ were homogenous in the length of the repeat region of only 1×12 bp (Figure 4-3). For CpGV-ZY2 a 12 bp long region accounted for 90.8% of the total of 762 reads, followed by 3×12 bp (9.1%) and 4×12 bp (0.1%), represented by 69 reads and a single read, respectively. Based on these data, the isolates belonging to genome group E and F possessed dominantly a single 12 bp repeat within *pe38*. For CpGV-ALE, representing the new genome group G, the main portion (92.3%) of the repeat region was 3×12 bp, similar to CpGV-ZY of genome group A, where 70.1% of all reads contained a 3×12 bp repeat region. A new repeat length of 2×12 bp was detected only in CpGV-ALE and CpGV-ZY, although based on only a single read only (Table 4-2). For CpGV-KS1 and -KS2, the majority of reads contained a 4×12 bp long repeat region, with 58% and 49.8% frequency, respectively. The high coverage of this study allowed the detection of even very low genetic variation, such as the presence of a 5×12 bp repeat region located on two reads (Table 4-2). Such minor genetic variations were not reflected by the isolate consensus sequences since they represent the majority of the intra-population diversity. In previously studied genome populations of CpGV, such as CpGV-M and CpGV-S, the genome coverage was low compared to this study, showing an average genome coverage between 4 to 22, except for CpGV-E2 with a 243-fold average coverage (Wennmann et al., 2017), and genetic variation in the *pe38* repeat regions of a single isolate was not detected.

The lack of the 2×12 bp insertion in *pe38* relative to CpGV-M was verified not only in the new CpGV isolates (Figure 4-2 and 4-3, Table 4-2) but in the selected French isolate (NPP-R1) (Berling et al., 2009a; Graillot et al., 2014). Increasing the ratio of CpGV-R5 (without 2×12 bp insertion) from 10% to 50% in a resistance test against the resistant strain $R_{GV}$, resulted in an approximately 10-fold reduction of median lethal concentration ($LC_{50}$) (Graillot et al., 2016). Whether the length of the repeat region is correlated to resistant-breaking profiles needs to be further addressed. For example, CpGV-JQ and -ZY2 from genome group F (with no or few 2×12 bp insertions) showed higher infectivity than CpGV-M (resistant isolate containing 2×12 bp insertion) against CpRR1 strain (Chapter 2). Taken together, these data

undoubtedly indicate that a high proportion of *pe38* without 2×12 bp insertion correlates with the capacity of resistance-breaking type I resistance.

**SNP categories and isolate composition**

To analyze the intra-specific variation within each meta-population of the seven new CpGV isolates, variable SNP sites of each isolate were determined by a consensus free method. Instead of aligning all isolates' consensus sequences and checking for SNP positions, as was performed for CpGV-M, -I12, -E2, -S and -I07 (Wennmann et al., 2017), all reads of the Chinese CpGV isolates were mapped to a common reference, namely CpGV-M, in order to standardize SNP positions referring to CpGV-M. By this newly applied approach, a total of 563 SNP positions was detected throughout all CpGV isolates from China. This approach considered SNPs with a frequency below 60%, which was the threshold for calling consensus sequences, and could further identify their specificity. The nucleotide alignment of CpGV sequences used for the SNP analyses comprised 123,529 positions in total. Theoretically, the nucleotide at each SNP position can be either identical to the reference or to one of the three alternative nucleotides (Alternative #1 to #3) (Figure 4-4). The frequency of the second (#2) and third (#3) alternative nucleotides at single SNP positions was generally below 1% (Figure 4-4). Therefore they were not further included in the analyses and only alternative #1 was considered as a reliable SNP. A total of 563 SNP positions was filtered using a coverage filtering that required at least a 100-fold coverage and a minimum alternative nucleotide occurrence of a total of 10. These criteria were set to increase the reliability of SNP positions and to avoid randomly occurring SNPs based on sequencing errors. Finally, 540 SNP sites of reliable coverage equalling 0.44% of nucleotide positions of the whole genome sequences showed polymorphisms. A total of 473 and 67 SNPs were found in coding and non-coding sequences, respectively. Among SNPs in coding regions, the number at the first, second and third codon positions were 108, 80 and 285, showing that more than half SNPs were synonymous and did not alter the encoded amino acid.

Compared to other baculoviruses the number of identified SNP position was rather low, underlining the highly conserved genomic architecture of CpGV. For example, seven Erinnyis ello granulovirus (ErelGV) isolates carried close to two thousand (1,893) SNPs in coding and noncoding genome areas (Brito et al., 2018). In contrast, another genetically highly homogenous betabaculovirus is Phthorimaea operculella granuloviurs (PhopGV), for which only 503 SNP positions were identified in twelve isolates originating from different continents (Larem et al., 2019). For the alphabaculovirus AcMNPV, 3,243 SNPs were found in wild-type isolates and four clusters of SNPs were determined with frequencies of 0.3% (2,561), 9.3% (400), 19.1% (164) and 34.9% (118) (Chateigner et al., 2015). In the CpGV isolates, however, SNPs did not show any frequency limitation and ranged from 0 to 100%.

Figure 4-5 shows a comprehensive map of the seven analyzed CpGV isolates with all SNP positions and their frequencies in relation to previously identified SNPs of the genome groups A to E of CpGV (Wennmann et al., 2017; Gueli Alletti et al., 2017). Compared to the latter the new CpGV isolates contained 334 previously identified SNPs, 116 new SNP positions specific for the phylogenetic genome groups F (99 SNPs) and G (17 SNPs), as well as 113 ungrouped SNPs (SN, new SNPs shared by Chinese CpGV isolates) (Table 4-3). A striking observation of the SNP mapping in Figure 4-5 is that the SNPs in a given isolate are often present with the same frequency (e.g. CpGV-WW) or at two frequency levels (e.g. CpGV-JQ) indicating that the isolates consist of pure or mixed genotypes. By using SNPs characterized with a

similar frequency and genome group specific trait, the composition of seven new CpGV isolates could be determined and quantified as shown in Table 4-4.

CpGV-WW carried 192 SNPs, of which 152 SNPs were previously identified in CpGV-S and representing genome group E (Wennmann et al., 2017). Their frequency was 99.1%, suggesting that this isolate is genetically highly homogenous (Figure 4-4 and 4-5) and closely related to CpGV-S as found in the phylogenetic analysis (Figure 4-1). In addition, 17 out of 31 SN SNPs, not present in CpGV-S, were found with the same frequency of >99%. Re-sequencing of CpGV-S revealed that three of these SNPs located at genome position 4,952, 40,992 and 87,368 were not present in CpGV-S representative of genome group E, but other 14 SNPs were found in CpGV-S (Chapter 4.). Thus, these three SNPs can be considered as CpGV geographic fingerprints of genome group E from China.

Three isolates, namely CpGV-KS1, -ZY, and -KS2, carried SNPs virtually identical to CpGV-WW, specific for genome group E, and to CpGV-M, specific for genome group A. The SNP frequencies related to genome group E and A were 22.8% and 76.7% (for CpGV-KS1), 13.9% and 85.7% (for -ZY), and 30.4% and 69.1% (for -KS2), respectively (Figure 4-5, Table 4-4).

The group E and group A specific SNPs help explain the >99% of genome variability, which argues that all three isolates are composed of CpGV-WW and a group A virus (Table 4-4).

Another example of physical mixtures of different CpGV genotypes is that of CpGV-0006 and -R5 which are comprised of 32% and 36% genome group E and 67% and 64% genome group A, respectively (Gueli Alletti et al., 2017). DNA restriction endonuclease analyses revealed that natural CpGV isolates often are genotype mixtures (Rezapanah et al., 2008). For example, the Iranian isolates CpGV-I22 and -I28 had characteristics similar to -I07, which belongs to genome group C, but CpGV-M was the dominant genotype in these isolates; other isolates were either identical to CpGV-M (-I15, -I66), or similar to CpGV-M but combined with additional diagnostic profiles (CpGV-I70 and -I67), clearly indicating their mixed nature (Rezapanah et al., 2008).

CpGV-ALE had 17 specific SNPs, of which the frequency was 78.4% (77.1-77.8%), representing new SNP fingerprints not present in other isolates and hence considered as specific for genome group G. Further, 79 group E specific SNPs accounted for 6.5% of CpGV-ALE SNPs, suggesting that it is composed of 78.4% group G, 6.5% group E, 15.1% unknown genotype.

CpGV-JQ and -ZY2 showed fingerprints of 83 genome group E specific SNPs suggesting the presence of CpGV-WW in the population at about 16.7% and 5.5%, respectively (Figure 4-5). In addition to that they harboured 99 specific SNPs largely not present in other isolates at a frequency of 77.8% (-JQ) and 76.4% (-ZY2), respectively. These SNPs were considered a fingerprint of the phylogenetic lineage F as they are not present in other isolates.

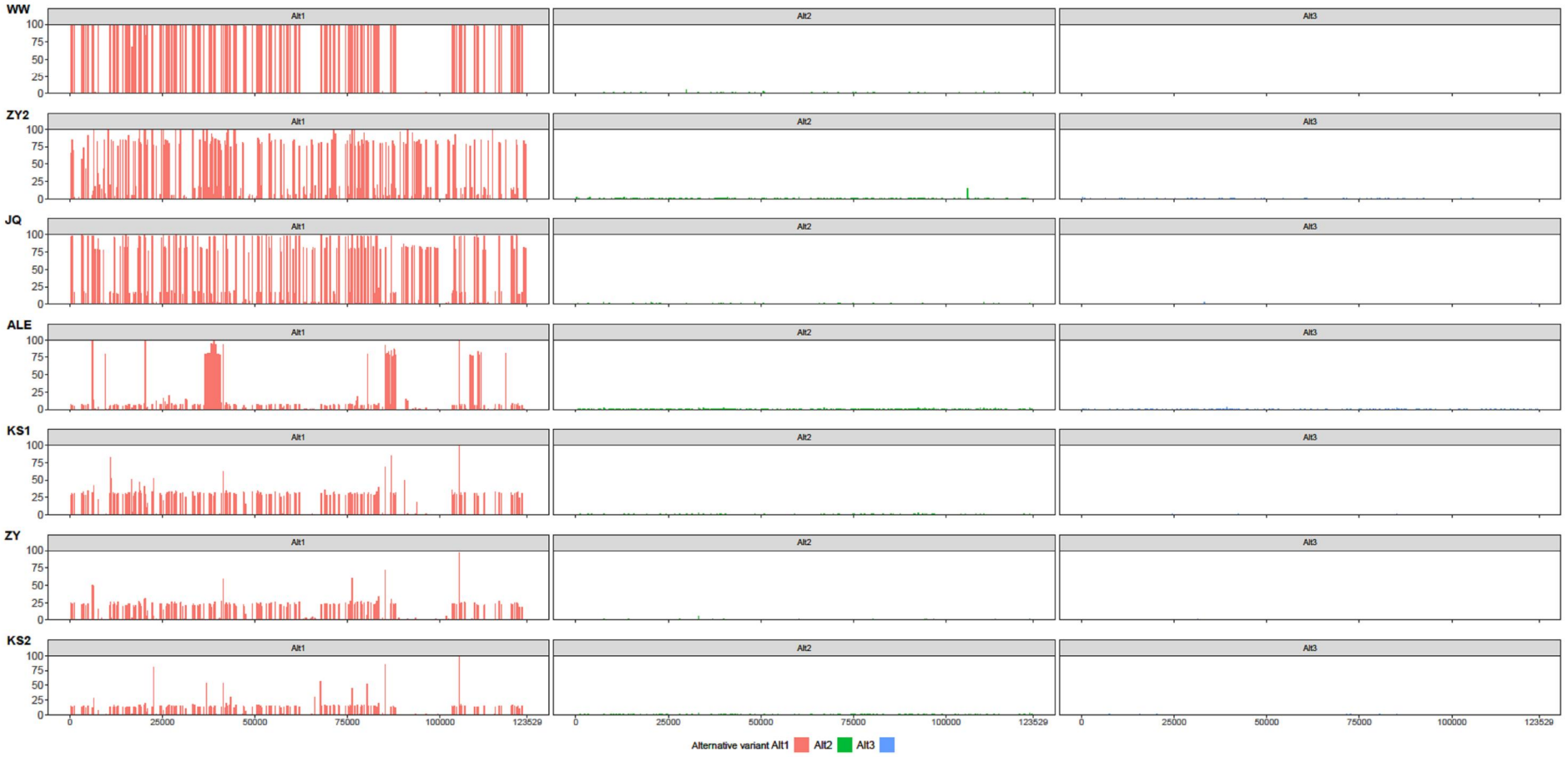



**Figure 4-4.** Position and frequency (%) of 563 SNP variants of seven CpGV isolates referencing to CpGV-M genome sequence. Three alternative nucleotide types were plotted in red, green and blue respectively.

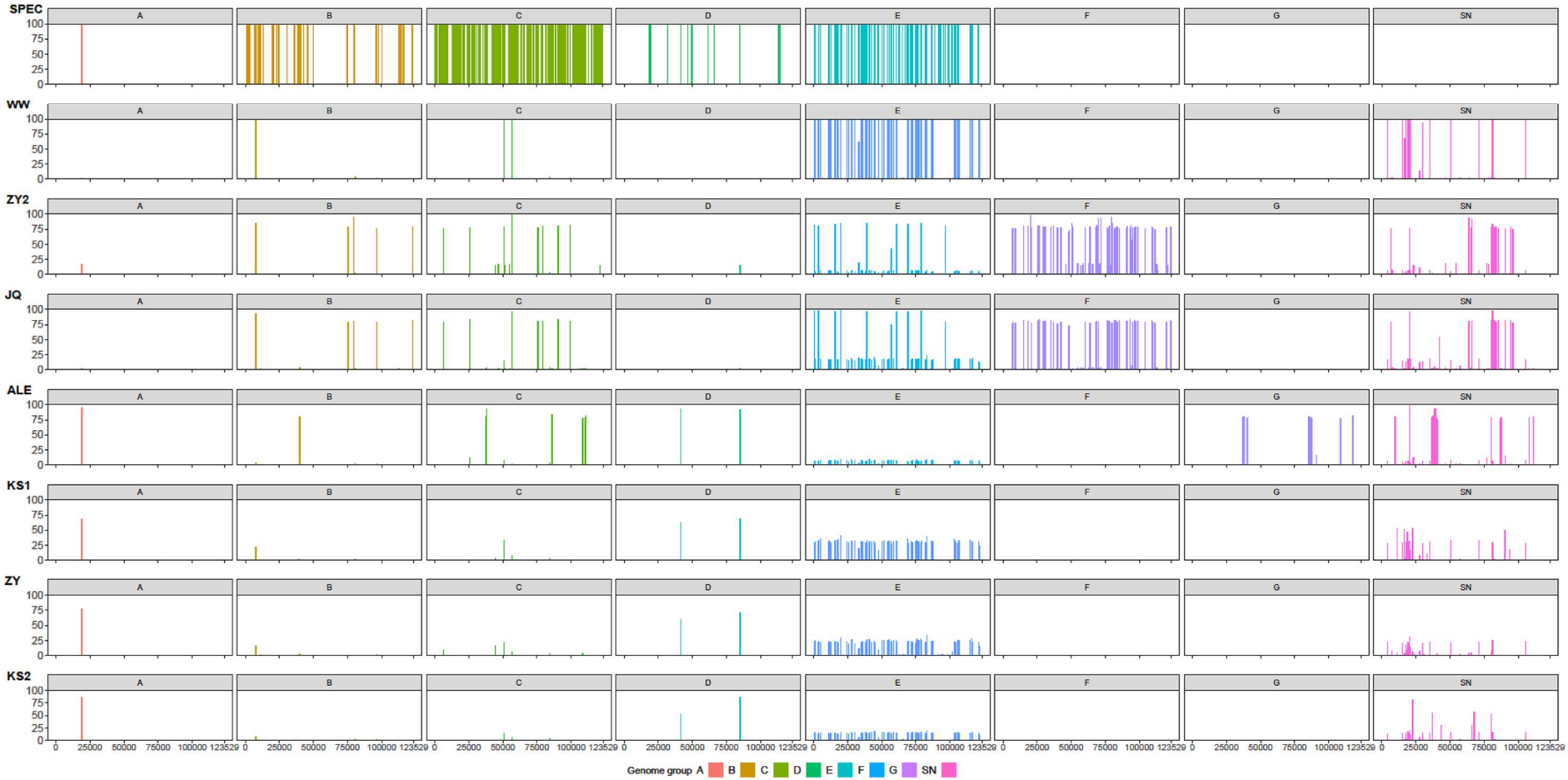



**Figure 4-5.** Frequencies (%) and positions of specific SNPs of seven Chinese CpGV isolates. Specific SNPs of genome group A-E obtained from Wennmann et al. (2017) and Gueli Alletti et al. (2017) were plotted in the top map, represented by SPEC. All SNP positions were plotted against the CpGV-M genome sequence. A total 302 specific SNPs excluding ambiguous ones, were plotted in genome group A, B, C, D, E, F and G, possessing 2, 18, 36, 2, 92, 99 and 17 specific SNPs, respectively. Additonal 113 ungrouped new SNPs (SN), discovered in the Chinese isolates, were randomly distributd in different CpGV genomes.

**Table 4-3.** Genome group SNPs numbers of seven new CpGV isolates was filtered by the criterion of total reads/coverage reads with 100/10.

| Genome group | Seven isolates | | | | | | | Total SNPs |
|---|---|---|---|---|---|---|---|---|
| | ZY | JQ | ALE | KS1 | KS2 | ZY2 | WW | |
| SN | 48 | 86 | 63 | 34 | 34 | 49 | 31 | 113 |
| G | - | - | 17 | - | - | - | - | 17 |
| F | - | 84 | - | - | - | 94 | - | 99 |
| E | 89 | 83 | 79 | 80 | 80 | 81 | 80 | 92 |
| D | 2 | 1 | 2 | 2 | 2 | 1 | - | 2 |
| C | 7 | 30 | 13 | 5 | 4 | 24 | 5 | 36 |
| B | 5 | 13 | 6 | 3 | 3 | 8 | 3 | 18 |
| A | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| Old combined SNPs | 77 | 87 | 83 | 71 | 72 | 81 | 71 | 161 |
| Total | 230 | 386 | 265 | 197 | 197 | 339 | 192 | 540 |

**Table 4-4**. Composition of genome group in seven CpGV isolates on the basis of frequency and specific SNPs.

| Iso | Grouped SNPs/ total No. | Median (95% CI of median) of genome group frequency | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | A % | B % | C % | D % | E % | F % | G % | Unkn. % |
| KS2 | 153/197 | 85.7 (85.5-86.0) | - | - | - | 13.9 (13.6-14.1) | - | - | 0.4 |
| ZY | 163/230 | 76.7 (76.4-77.1) | - | - | - | 22.8 (22.3-23.3) | - | - | 0.5 |
| KS1 | 152/197 | 69.1 (68.6-69.5) | - | - | - | 30.4 (30.1-30.7) | - | - | 0.5 |
| ALE | 168/265 | - | - | - | - | 6.5 (6.3-6.7) | - | 78.4 (77.1-79.9) | 15.1 |
| JQ | 230/386 | | - | - | - | 16.7 (16.3-17.2) | 77.8 (4.8-78.8) | - | 5.5 |
| ZY2 | 189/339 | 16.1 (15.3-16.9) | - | - | - | 5.5 (5.2-5.8) | 76.4 (75.4-77.8) | - | 2 |
| WW | 152/192 | - | - | - | - | 99.1 (99.0-99.2) | - | - | 0.9 |

Unkn., the frequency of unknown SNPs distributed in seven CpGV isolates. CI is the Confidence Interval of median. Iso, isolate name.

Strikingly, specific SNPs of CpGV-WW, related to CpGV-S (genome group E), were observed in all seven new CpGV isolates with variable ratios from 5.5% to 99.1% (Figure 4-5, Table 4-4), indicating that CpGV-WW is widely distributed at least in Chinese codling moth populations. The other isolate of genome group E, CpGV-S, was originally isolated from diseased CM larvae in Canada and is the active ingredient of the commercial product Virussoft® (Vincent et al., 2007; Gebhardt et al., 2014). Isolates belonging to genome group E are also ingredients of three other commercial isolates, namely CpGV-0006, -R5, and V15 with different proportion of 68%, 64% and 50% of genome group E genotypes, respectively (Gueli Alletti et al., 2017). Considering their virulence patterns against CM with type I and type II resistance (Chapter 2), these new CpGV isolates provide novel genetic diversity of CpGV and probably functional differences which should be further evaluated for use as agents for CM control.

Genomic diversity was also observed for natural isolates of other betabaculoviruses, e.g. ErleGV (Brito et al., 2018) and PhopGV (Larem et al., 2019), as well as in alphabaculoviruses, e.g. in Lymantria dispar multiple nucleopolyhedrovirus (LdMNPV) (Harrison et al., 2016), Anticarsia gemmatalis nucleopolyhedrovirus (AgMNPV) (Del-Angel et al., 2018), Bombyx mori nucleopolyhedrovirus (BmNPV) (Xu et al., 2013), Helicoverpa armigera nucleopolyhedrovirus (HearNPV) (Arrizubieta et al., 2015; Raghavendra et al., 2017).

Since the consensus sequences in general refer only to the most frequent nucleotides in sequence data assembly (Day and McMorris, 1992), minor mutations as well as submolar genotypes are often ignored in such analyses. On the other hand, specific SNPs of respective genome groups could be used as genetic markers to quantify the genotype composition of new isolates. The graphical output of SNP mapping, as presented in this study, allows an easy visual notion of the genomic diversity of complex genotype mixtures, as well as the identification of geographic footprints of a given virus isolate.

Seven new CpGV isolates did not fully follow previous SNP patterns representative genome group A-E, not only in total SNP quantity but location in genome sequence. New and already known SNPs can be detected when new isolates are included in SNP analyses, which may result in a reduction of genome group-specific SNPs, as they may be found also in other isolates or lineages. Nearly all SNPs (92 of 101) specific for genome group E were still group specific and followed the CpGV lineages defined based on single isolates, other SNPs were not lineage/group specific anymore.

In conclusion, the analyses of the genomes of seven novel CpGV isolates from China revealed some remarkable genetic heterogeneity, including two new phylogenetic lineages of CpGV, termed genome group F and G. All isolates, except for CpGV-WW, could be clearly identified as mixtures of at least two or more isolates belonging to different genome groups. It appears that in natural isolates of CpGV and many other baculoviruses, genotype mixtures are more prevalent than pure genotypes. It can be therefore hypothesized that there could be a functional cooperation of genomes and genotypes during the infection process. SNP based transcriptome analyses, which allow identifying the contribution of the different genotypes of such mixed isolates to the infection process may help inform how such genomes interact.

## Chapter V: In-depths population structure of Cydia pomonella granulovirus isolates revealed by multi component analyses of their SNP distribution


### Abstract

Genetic diversity of viruses is driven by genomic mutations and selection through its host, resulting in differences of virulence as well as host response. For baculoviruses, which are naturally occurring pathogens of insects and which are every year repeatedly sprayed as biocontrol agents on hundred thousand to millions of hectares, the phenomenon of virus-host co-evolution is of particular scientific interest and economic importance because high virulence of baculovirus products is essential and emergence of host resistance needs to be avoided. In the present study, the population structure of 20 Cydia pomonella granulovirus (CpGV) isolates, including twelve isolates from different geographic origins and eight laboratory selections, were studied by Illumina next generation sequencing (NGS) and by analyzing the distribution of single nucleotide polymorphisms (SNPs). On average, 91.9% of 2,124,330 QC-passed reads ($Q \geq 30$) of each isolate were mapped separately against CpGV-M reference sequence. Based on the position and frequency of the detected 753 (0.61% of whole genome) SNP sites as well as deletions in 25 ORFs (19.7% of 142 ORFs) and insertions in 23 ORFs (18.3% of 142 ORFs), especially five types of 12 bp repeat insertions in ORF 24 (*pe38*), a genetic fingerprint database was constructed using a consensus-free method. The results showed that the isolates CpGV-M, -WW, -S and MadexPlus were genetically highly homogenous exhibiting a low rate of polymorphisms, while other isolates were composed of two or more genotypes at different ratios. Based on hierarchical clustering on principal components (HCPC) six distinct isolate clusters were identified, representing the previously proposed main phylogenetic lineages of majority-based consensus sequences but comprising full genome information from virus mixtures. Relative location of different isolates in HCPC reflected the ratio of variable compositions of different genotypes. The established methods provide novel analysis tools to decipher the molecular complexity of genotype mixtures in virus isolates, thus depicting the population structure of baculovirus isolates in a more adequate form than consensus based analyses.


### Introduction

Most of our knowledge of virus functions derives from laboratory studies of purified virus isolates containing a single genotype. To obtain such pure genotypes, isolates had to be purified, e.g. by plaque purification (Cooper, 1962; Brown and Faulkner, 1978; Durantel et al., 1998; Kariuki and McIntosh, 1999; Harrison, 2009; Gueli Alletti et al., 2018) or *in vivo* cloning (Smith and Crook, 1988; Winstanley and Crook, 1993; Luque et al., 2001). Naturally occurring viruses, however, are often mixed populations of different genotypes, exhibiting genetic variation caused by insertions/deletions (indels) mutations (Crook et al., 1985; Thézé et al., 2014) and single nucleotide polymorphisms (SNPs) (Chateigner et al., 2015; Wennmann et al., 2017; Larem et al., 2019). Such populations may have different functions than pure genotypes (Ferrelli et al., 2012; Gueli Alletti et al., 2017). With advent of next generation sequencing (NGS) based deep sequencing, studying virus populations has been

taken to a new level. Population structure can be described and analyzed, allowing insight into diversity as well as evolutionary constraints.

In the following the population structure of Cydia pomonella granulovirus (CpGV), which is highly pathogenic for the codling moth (CM, *Cydia pomonella* L) was studied. CpGV is the type species of genus *Betabaculovirus* (family *Baculoviridae*) (Herniou et al., 2011). Its genome varies between 120.8 and 124.3 kbp in length and encodes 137 to 142 open reading frames (ORFs) (Wennmann et al., 2017). Because of the high virulence of CpGV against larval stages of CM and its potential application in biological control of this pest insect in pome fruit production (Huber, 1998), considerable efforts were undertaken to study this virus since it was first isolated from diseased CM larvae in Mexico in 1963 (Tanada, 1964). This isolate, termed CpGV-M, was also the first commercial CpGV isolate registered as biocontrol agent (Huber, 1998). As a consequence of these efforts, a large number of further natural isolates have been discovered from different geographic regions worldwide (Crook et al., 1985; Rezapanah et al., 2008; Gan et al., 2011; Arneodo et al., 2015). Based on phylogenetic studies CpGV isolates are grouped into 7 phylogenetic lineages, termed genome group A to G (Gebhardt et al., 2014; Wennmann et al., 2017; Chapter 4). With the development of CM populations resistant to CpGV products, scientific and economic interest in exploiting the genetic diversity of CpGV became even more significant (Asser-Kaiser et al., 2007; Jehle et al., 2017; Sauer et al., 2017b). Nowadays several natural isolates and laboratory selections are used in different commercial products worldwide, including the isolates from genome groups A, B and E (Huber, 1998; Lacey et al., 2008, Vincent et al., 2007; Graillot et al., 2014; Gueli Alletti et al., 2017). To further characterize the genomic functions of CpGV and to identify the molecular nature of resistance-breaking CpGV isolates, genomes of different isolates were previously studied by Sanger and 454 pyrosequencing. SNP pattern analysis was successfully applied to determine the identity and composition of natural and commercial CpGV isolates and correlating their composition with different CpGV genotypes and their activity towards CpGV resistance (Gueli Alletti et al., 2017).

The picture of CpGV diversity and phylogeny was significantly extended when seven new Chinese isolates were characterized by NGS (Wennmann et al., 2017; Chapter 4): (i) in addition to previously defined genome groups A-E, two new phylogenetic lines (groups F and G) were found; (ii) new SNP positions were identified; (iii) based on limited dataset used in previous studies, some recently identified group specific SNPs were not group-specific anymore; (iv) highly homogenous isolates and highly complex genotype mixtures could be identified in single isolates.

As genotype mixtures are common in baculovirus field isolates (Xu et al., 2013; Thézé et al., 2014; Harrison et al., 2016; Brito et al., 2018; Larem et al., 2019), a consensus sequence generated from an ultra deep sequenced baculovirus isolate can only reflect the major frequency that was chosen to extract it for every base from the assembly data (Day and McMorris, 1992), resulting in loss of genomic information. In conventional phylogenetic analyses based on consensus sequences of CpGV isolates, genetic composition were not taken into account (Wennmann et al., 2017). This study also revealed that one limitation of the analyses laid in the lack of NGS data of CpGV isolates founding the phylogenetic genome groups A to E. Because of the poor sequencing depth below 250 (Wennmann et al., 2017) and the use of sequence consensus data for SNP analyses, no information was available whether these isolates were also mixtures and which SNPs were indeed genome group specific.

It addresses this problem by re-sequencing the previously sequenced isolates CpGV-M, -I12, -S and E2 as well as some new natural isolates and new commercial isolates. Indels and SNP detection for these newly and re-sequenced isolates was performed *de novo* in a consensus free approach, independently from previously identified indels and SNPs (Wennmann et al. 2017). Data were compiled and analyzed together with the NGS data of Chinese isolates (Chapter 4) and commercial isolates (Gueli Alletti et al., 2017), extending the data set to 20 CpGV genomes. Analysis of SNP variation and frequency based on multiple component analyses (MCA) was developed. Such MCA allows an improved representation of the diversity and composition of virus populations than the use of conventional phylogenetic analysis tools.

## Material and method

### Geographic CpGV isolates and Illumina sequencing

Twenty different CpGV isolates were analyzed in this study. Therein, twelve field isolates, derived from different geographic locations of Mexico (1), England (1), Canada (1), Iran (2) and China (7), whereas eight isolates were commercial selections or were derived from commercial CpGV products. CpGV-M (Mexico), -E2 (England), -I12 (Iran), -S (Canada) were re-sequenced from virus samples previously used for Sanger or 454 pyrosequencing (Gebhardt et al., 2014, Wennmann et al., 2017), -I0X was a unknown characterized CpGV isolate from Iran (Rezapanah et al., 2008). The isolates CpGV-ALE, -JQ, KS1, -KS2, -ZY, -ZY2, -WW (China) had been described previously (Chapter 4). The following commercial isolates were obtained from Andermatt Biocontrol, Stahlermatten, Switzerland: MPlus (MadexPlus, V003, selection from CpGV-M, resistance-breaking to type I resistance); MMaxF (batch number 48 of MadexMAX); V14 (batch number 11, commercial propagation of CpGV-E2); V34 and V45 (commercial selections of natural CpGV isolates in resistant *C. pomonella* larvae). Sequence data of Carpovirusine EVO2 (CEVO2,) (EVO2, isolate R5, Arysta Lifescience, Noguères, France), MMaxP (CpGV-0006, active ingredient of MadexMAX, resistance-breaking to type I resistance) and V15 (batch number 002, active ingredient of MadexTOP) were obtained from previous study (Gueli Alletti et al., 2017).

For newly sequenced isolates, viral DNA was extracted using the standard protocol as previously described (Arends and Jehle, 2002; Gueli Alletti et al., 2017). At least 100 ng genomic DNA was sequenced using Illumina NextSeq500$^{TM}$ platform at StarSEQ GmbH (Mainz, Germany), generating paired-end reads with 151 nucleotides in length.

### Sequence datasets assembly

All obtained reads of each isolate were processed in same workflow. First, adapter trimmed reads were quality-controlled using the parameters of Phred-quality score ≥30 and read length ≥ 50. Second, high quality reads of every isolate were assembled separately with the BWA-MEM mapper using CpGV-M (KM217575) as reference sequence (Li, 2011a, 2013) on the Galaxy server of the Julius Kühn-Institut. Third, assembled data sets of paired-reads and unpaired-reads were processed using SAMtools MPileup and followed by BCFTools command to call variant sites only (For details see Chapter 4, Larem et al., 2019).

The average coverage for a whole genome can be calculated from the length of the original genome (G), the number of reads (N), and the average read length (L) as N×L/G. Mapped read depth refers to the total number of bases sequenced and aligned at a given reference base position.

### SNP mapping and genotype composition

After read assembly, detected variant sites were visualized using R (R version 3.4.4 in RStudio 1.1.442) programming with "VariantAnnotation" package (Obenchain et al., 2014) and developed scripts by Wennmann (unpublished). For each position and isolate, the frequencies of the reference nucleotide, as well as all three possible alternative nucleotides was calculated resulting in a SNP frequency table with 753 variant sites for all 20 isolates. Minimum 100-fold coverage of total reads and minimum 10-fold variant reads at each position were set as variant criteria. If a variant position of an isolate did not exceed this threshold, the variant base count was set to zero. The SNP distribution was plotted along the CpGV-M reference genome to identify whether an isolate was a highly homogenous genotype, a mixture of homogenous genotypes, highly heterogeneous, or in combination of these possibilities. For the twelve field collected/non-commercial CpGV isolates, namely CpGV-M, -E2, -S, -I12, -I0X, -ALE, -KS1, -KS2, -JQ, -ZY, -ZY2, and -WW, excluding low quality and commercial product specific SNPs, the specificity of 723 variant sites, was performed on the SNP frequency table with the following criteria: (i) a variant site was isolate specific (= specific for one isolate only), when all other isolates had no alternative variant base at the same position; (ii) a variant site was group specific (= specific for at least two isolates), when isolates had a common alternative base at the same position. SNP specificities assigned to each position were plotted as heatmap in R using CpGV-M as reference. The SNP frequency table was further used to determine the genetic composition of each isolate. Isolate/genome group specific SNPs in every isolate were extracted, median values of these specific SNPs were used to evaluate the ratio of genome groups A, B, D, E, F and G in a given isolate (Chapter 4; Wennmann et al., 2017).

### Indels screening

Assembled paired-reads of each isolate was input into Geneious 9 separately and further used to identify insertion or deletion (indel) mutations; the following parameters were applied: minimum coverage for 100, minimum variant frequency for 0.01, maximum variant P-value for $10^{-6}$, minimum strand-bias P-value for $10^{-5}$ when exceeding 65% bias. Indels were picked up from output and listed in table. Especially for reads containing multiples of the 12 bp motif repeat in *pe38* (GACACAGTGGAT) (Chapter 4; Gebhardt et al., 2014) were picked up using the following parameters: a read contains unique fragment of 12 bp and 10 bp upstream and downstream of the 12 bp repeat region, respectively. These reads were used to calculate the different repeat numbers and their frequency by applying the software "ShortRead" (Morgan et al., 2009) and "stringr" (Wickham, 2018) R package. Finally the output of 12 bp ratio in both methods were plotted using ggplot2 (Wickham, 2016).

### Hierarchical clustering on principal components (HCPC)

For the clustering of 20 CpGV isolates, the SNP frequency table with the entity of 753 found SNP positions were used by applying the HCPC method implemented in "FactoMineR" R package (Lê et al., 2008). The HCPC process was included four steps of factorial analysis (principal components analysis), hierarchical clustering, cutting clustering tree and consolidation using K mean with the cluster centers. The last factor was removed from analysis to make the clustering more robust. Six clusters were set to correspond to six genome group A, B, D, E, F, G (Chapter 4). Hierarchical cluster tree and factor map were generated representing the lineages of highly homogenous, mixture and heterogeneous CpGV isolates.

Three dimension cluster result was plotted as well. Assembled dataset of each isolate was used to generate a consensus sequence with the threshold of 50% (Bases matching at least 50% of the sequence). These consensuses were applied for phylogeny construction using minimum evolution method with the parameters: 500 of bootstrap replications, Tajima-Nei model, Gamma distributed for rates among sites, gaps/missing data treated by partial deletion in MEGA7 (Kumar et al., 2016).

## Results

### Illumina sequence data of 20 CpGV isolates

In total 20 CpGV isolates, the previously sequenced CpGV-ALE, -JQ, -KS1, -KS2, -ZY, -ZY2, -WW and CEVO2 (EVO2), MMaxP, V15 (Chapter 4; Gueli Alletti et al., 2017) and newly re-sequenced isolates CpGV-M, -E2, -I12, -S, I0X and MPlus, MMaxF, V14, V15, V34, V45, were presently analyzed together (Table 5-1). On average, 91.9% of 2,124,330 high quality reads of each isolate were mapped against CpGV-M as reference genome (Table 5-2). Mean coverage ranged from 540- to 4345-fold (Table 5-2). On the other hand, read depth, referring to the total number of bases sequenced and aligned at a given reference base position, was from 406 to 1163-fold when assembled reads were aligned together to call variants (Figure 5-1). Median reads depth of re-sequenced isolates CpGV-M, -S, -E2, -I12 ranged from 1109 to 1138-fold.

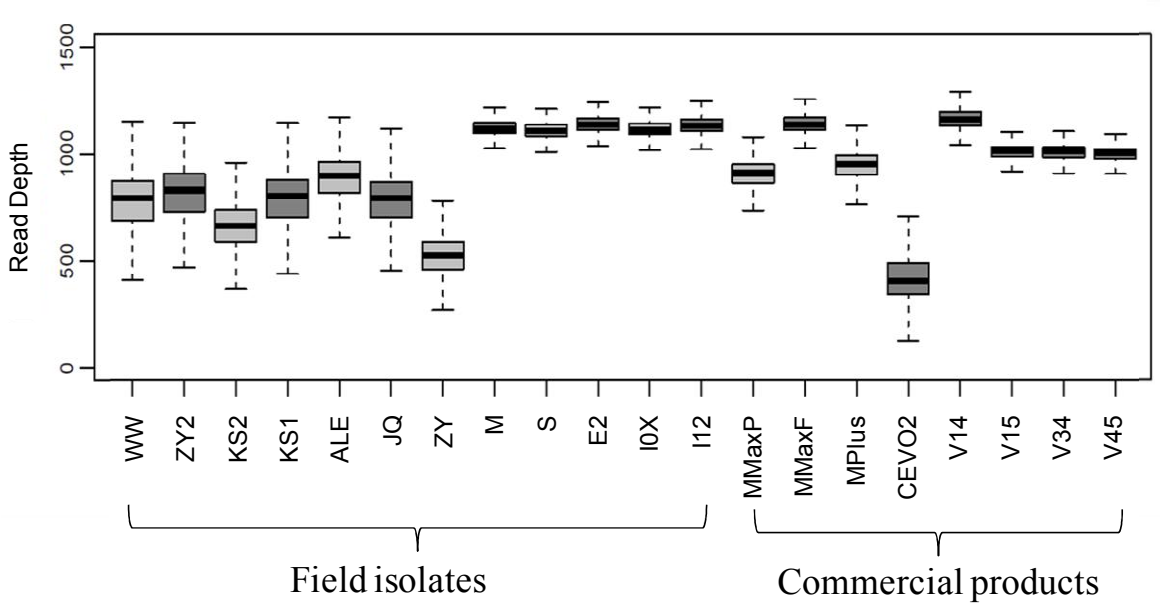



**Figure 5-1**. Average coverage of 20 CpGV isolates against CpGV-M (KM217575) reference genome. The field isolates CpGV-WW, -ZY2, -KS2, -KS1, -ALE, -JQ, -ZY derive from China had been described previously (Chapter 4); CpGV-I0X was isolated from Iran; CpGV-M, -S, -E2, -I12 were re-sequenced isolates. The following commercial isolates were MadexMAX (MMaxP and MMaxF), MadexPlus (MPlus), V14, V15, V34 and V45, of which all were obtained from Andermatt Biocontrol, Stahlermatten, Switzerland; Carpovirusine EVO2 (CEOV2, isolate R5) was obtained from Arysta Lifescience, Noguères, France.

**Table 5-1.** Genetic composition of twenty field, commercial and commercially formulated isolates of CpGV. Median frequencies of SNP variants with correspondent 5 to 95% percentiles were measured based on their isolate specificity. Given are the previously named genome group A, BDEFG based on the phylogenetic reconstruction. F[#], CpGV-ZY2 specific SNP contributing on genome group F.

| | Median (5-95%) (%) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Genome group | A | B | D | F[#] | F | G | FG | BDEFG | DEFG |
| Isolate (no. SNPs) | M (58) | E2 (68) | I12, I0X (24) | ZY2 (30) | ZY2, JQ (89) | ALE (21) | ALE, JQ (22) | WW, ZY2, KS2, KS1, ALE, JQ, ZY, S, I12, E2 (75) | WW, ZY2, KS2, KS1, ALE, JQ, ZY, S, I12 (45) |
| CpGV-M | 100 (100-100) | | | | | | | | |
| CpGV-MPlus | 100 (100-100) | | 0 (0-78) | | | | | | |
| CpGV-S | 0 (0-1) | | | | | | | 100 (100-100) | 100 (93-100) |
| CpGV-WW | 0 (0-1) | | | | | | | 100 (100-100) | 100 (98-100) |
| CpGV-KS1 | 69 (65-73) | | | | | | | 29 (26-33) | 30 (23-35) |
| CpGV-KS2 | 86 (84-90) | | | | | | | 14 (9-17) | 14 (10-17) |
| CpGV-ZY | 77 (73-81) | | | | | | | 22 (20-28) | 23 (17-27) |
| CpGV-ALE | 94 (70-95) | | | | | 80 (75-83) | 80 (76-94) | 6 (4-88) | 6 (4-8) |
| CpGV-ZY2 | 15 (0-95) | | | 16 (13-18) | 77 (14-81) | | | 42 (4-100) | 5 (4-84) |
| CpGV-JQ | 0 (0-84) | | | | 80 (2-84) | | 2 (1-3) | 79 (15-100) | 17 (13-100) |
| CpGV-I12 | 56 (40-82) | | 2 (1-3) | | | | | 46 (31-57) | 42 (14-57) |
| CpGV-I0X | 93 (50-97) | | 35 (4-100) | | | | | | |
| CpGV-E2 | 20 (0-98) | 84 (28-100) | | | | | | 67 (3-100) | |
| V14 | 27 (0-100) | 81 (29-100) | | | | | | 60 (2-100) | |
| V15 | 7 (3-53) | 42 (0-48) | | | | | | 93 (44-97) | 49 (42-55) |
| MMaxF | 26 (23-29) | 0 (0-1) | 0 (0-4) | | | | 0 (0-1) | 75 (64-78) | 74 (62-77) |
| MMaxP | 32 (29-35) | | 0 (0-4) | | | | 0 (0-2) | 68 (63-72) | 68 (59-70) |
| CEVO2 | 33 (24-64) | | 0 (0-8) | | | | | 67 (29-76) | 64 (30-72) |
| V34 | 38 (25-75) | 0 (0-39) | 0 (0-19) | | | | | 67 (23-76) | 28 (23-67) |
| V45 | 61 (0-98) | 2 (0-72) | 0 (0-67) | | | | | 30 (9-99) | 19 (5-71) |

**Table 5-2.** Statistics of paired and unpaired reads with an average Phred-quality score ≥ 30 obtained from Illumina NextSeq500 sequencing and assembled against CpGV-M genome using BWA-MEM method.

| CpGV isolate | Phred-quality score ≥ 30 | reads (%) assembled to CpGV consensus | reads (%) not assembled to CpGV consensus | Mean coverage ± SD |
|---|---|---|---|---|
| M | 3,650,570 | 99.4% | 0.6% | 3,995 ± 622 |
| MPlus | 1,573,038 | 98.7% | 1.3% | 1,301 ± 234 |
| S | 3,359,199 | 88.8% | 11.2% | 3,320 ± 614 |
| WW | 962,278 | 99.7% | 0.3% | 928 ± 204 |
| KS1 | 1,134,546 | 96.3% | 3.7% | 1,057 ± 214 |
| KS2 | 879,222 | 90.5% | 9.5% | 776 ± 149 |
| ZY | 650,566 | 93.4% | 6.6% | 595 ± 121 |
| ALE | 1,181,528 | 92.5% | 7.5% | 1,076 ± 209 |
| ZY2 | 996,638 | 99.1% | 0.9% | 968 ± 195 |
| JQ | 1,022,825 | 94.4% | 5.6% | 950 ± 184 |
| I12 | 3,317,723 | 98.0% | 2.0% | 3,603 ± 593 |
| I0X | 3,141,095 | 96.9% | 3.1% | 3,406 ± 523 |
| E2 | 3,658,231 | 98.8% | 1.2% | 4,045 ± 725 |
| V14 | 4,039,489 | 97.0% | 3% | 4,345 ± 818 |
| V15 | 2,388,975 | 99.7% | 0.3% | 2,654 ± 448 |
| MMaxF | 3,485,678 | 96.5% | 3.5% | 3,734 ± 657 |
| MMaxP | 1,428,117 | 98.9% | 1.1% | 1,584 ± 438 |
| CEV02 | 3,102,552 | 15.7% | 84.3% | 540 ± 162 |
| V34 | 2,140,576 | 99.5% | 0.5% | 2,389 ± 404 |
| V45 | 2,110,218 | 99.6% | 0.4% | 2,361 ± 371 |

### Variants detection and category

A global detection of CpGV polymorphisms was achieved by mapping reads of the 20 analyzed isolates to the common reference CpGV-M. The main focus was laid on the detection of SNPs that were used in further steps for the identification and quantification of CpGV isolates and mixtures. In total 753 SNP sites corresponding to 0.61% of the entire CpGV-M genome (123,529 bp) were detected in all aligned reads. The SNPs included 544 transitions and 209 transversions. 665 SNPs (88.3%) were distributed over 124 open reading frames (ORFs), whereas 88 SNPs (11.7%) were found in non-coding regions (Figure 5-2). Only 18 ORFs were without any SNP and thus fully conserved in all isolates. Out of 665 SNPs in coding regions, 163 (eleven SNPs in overlapping ORFs) and 128 SNPs (four SNPs in overlapping ORFs) were located in first and second codon position, respectively, whereas 398 SNPs (ten SNPs in overlapping ORFs) were found in the third codon position; 309 SNPs were non-synonymous nucleotide changes, resulting in amino acid changes (Figure 5-2). Annotated CpGV genes were sorted into five groups, i.e. biological regulation, DNA replication, metabolic process, transcription, and structural proteins (Table 5-3) (Luque et al., 2001; Rohrmann, 2013). The average SNP number per ORF with known function ranged from 2.75 (DNA replication) to 3.59 (biological regulation) per kbp, whereas the SNP density in unknown ORF is slightly higher, with 5.61 per kbp (Figure 5-3).

**Table 5-3.** Annotation of 103 ORFs harbored SNPs. Based on the annotation reviewed by (Luque et al., 2001; Rohrmann, 2013) identified SNPs were classified into five groups of biological regulation, DNA replication, metabolic process, structure protein, virus transcription.

| ORF | gene | function | Annotation |
|---|---|---|---|
| 24 | *pe38* | delay DNA replication and synthesis | biological regulation |
| 94 | *iap* | inhibitor of apoptosis | biological regulation |
| 140 | *fgf-3* | vFGF initiate a cascade of events to accelate systemic infection | biological regulation |
| 17 | *iap-3* | inhibitor of apoptosis | biological regulation |
| 102 | *tlp20* | BV production and DNA replication delay | biological regulation |
| 76 | *fgf-1* | vFGF initiate a cascade of events to accelate systemic infection | biological regulation |
| 13 | *gp37* | enhance per os infection | biological regulation |
| 123 | *fgf* | accelerate the establishment of systemic infections | biological regulation |
| 143 | *me53* | BV and ODV associated | biological regulation |
| 41 | *lef-2* | DNA primase accessory factor | DNA replication |
| 58 | *lef-11* | DNA repliacate | DNA replication |
| 113 | *lef-3* | single-stranded DNA binding protein | DNA replication |
| 81 | *dbp* | DNAbinding protein, production of nucleocapsids and virogenic stroma | DNA replication |
| 125 | *alk-exo* | an exonuclease activity, involved in DNA recombination | DNA replication |
| 111 | *dnapol* | DNA polymerases | DNA replication |
| 90 | *helicase* | DNA replication | DNA replication |
| 74 | *lef-1* | DNA primase | DNA replication |
| 128 | *rr2a* | nucleotide metabolism_ribonucleotide reductase subunits | metabolic process |
| 10 | *chitinase* | degrade chitins | metabolic process |
| 117 | *rr1* | nucleotide metabolism_ribonucleotide reductase subunits | metabolic process |
| 141 | *egt* | inactivating these insect molting hormones | metabolic process |
| 11 | *cathepsin* | a metalloprotease along with chitinase | metabolic process |
| 66 | *ptp-2* | pro-apoptotic protein | structure protein |
| 35 | *pif-3* | per os infection factor | structure protein |
| 46 | *mp-nase* | assisting in their viral transmission | structure protein |
| 22 | orf17R | baculovirus PEP N domain | structure protein |
| 73 | *38.7K* | nonessential structure and encode a late gene | structure protein |
| 23 | *pe/pp34* | polyhedron envelop protein | structure protein |
| 47 | *p13* | killing associated gene | structure protein |
| 56 | ac108 | associated with the PIF complex | structure protein |
| 48 | *pif-1* | oral infection factor | structure protein |
| 89 | *pif-4* | per os infectivity factor | structure protein |
| 14 | *odv-e18* | ODV envelop fraction | structure protein |
| 101 | *vp91* | structure protein | structure protein |
| 96 | *vp39* | structure protein | structure protein |
| 112 | *desmoplakin* | associated with BV and ODV (egress of virions from nuclei) | structure protein |
| 83 | *p45(p48)* | essential for generate perfect ODV and BV | structure protein |
| 31 | *f-protein* | glycosylated and associated with the envelope of BV | structure protein |
| 93 | *p33* | associated with BV and ODV | structure protein |
| 37 | *odv-e66* | ODV envelopes | structure protein |
| 108 | *ac75* | required for the nuclear egress of nucleocapsids and intranuclear microvesicle formation | structure protein |
| 15 | *p49* | affect DNA synthesis | structure protein |
| 92 | *p18* | core gene, egress of nucleocapsids | structure protein |
| 138 | *vp1054* | capsid protein | structure protein |
| 106 | | structure protein in ODV and BV | structure protein |
| 8 | ac146 | BV production | structural protein |
| 71 | *p24capsid* | ODV component and associated with BV | structure protein |
| 1 | *granulin* | structural protein | structural protein |
| 75 | | per os onfectivity factors | structure protein |

*Cont. Table 5-3*

| | | | |
|---|---|---|---|
| 3 | *pk1* | catalytic domain activity | structural protein |
| 97 | *odv-ec27* | ODV and BV associated and affect nucleocapsid formation | structure protein |
| 104 | *gp41* | located between the virion envelope and capsid | structure protein |
| 60 | *p74* | first identified per os infectivity factors | structure protein |
| 85 | *bv/odv-c42 (p40)* | encode a capsid-associated protein of BV and ODV | structure protein |
| 57 | *pp31/39K* | capable of binding to DNA not a virion structure protein | virus transcription |
| 68 | *p47* | transient late gene expression and component of late gene polymerase complex | virus transcription |
| 95 | *lef-4* | component of the late baculovirus RNA polymerase | virus transcription |
| 131 | *lef-8* | baculovirus RNA polymerase subunit | virus transcription |
| 117 | *lef-9* | baculovirus RNA polymerase subunit | virus transcription |
| 7 | *ie-1* | transient DNA replication | virus transcription |
| 26 | | | |
| 49 | | | |
| 61 | | | |
| 27 | | | |
| 6 | | | |
| 36b | | | |
| 62 | | | |
| 122 | | | |
| 39 | | | |
| 70 | | | |
| 79 | | | |
| 25 | | | |
| 36a | | | |
| 32 | | | |
| 34 | | | |
| 72 | | | |
| 109 | | | |
| 121 | | | |
| 82a | | | |
| 45 | | | |
| 12 | | | |
| 20 | orf16L | | |
| 129/130 | | | |
| 42 | orf35a | | |
| 105 | ac78 | | |
| 2 | | | |
| 50/51 | | | |
| 77 | | | |
| 139 | | | |
| 100 | | | |
| 43 | | | |
| 135 | | | |
| 110 | | | |
| 132 | | | |
| 33 | | | |
| 119 | | | |
| 115 | | | |
| 103 | ac81 | | |
| 16 | | | |
| 44 | orf36L | | |
| 52a | | | |
| 99 | | | |
| 28/29 | | | |

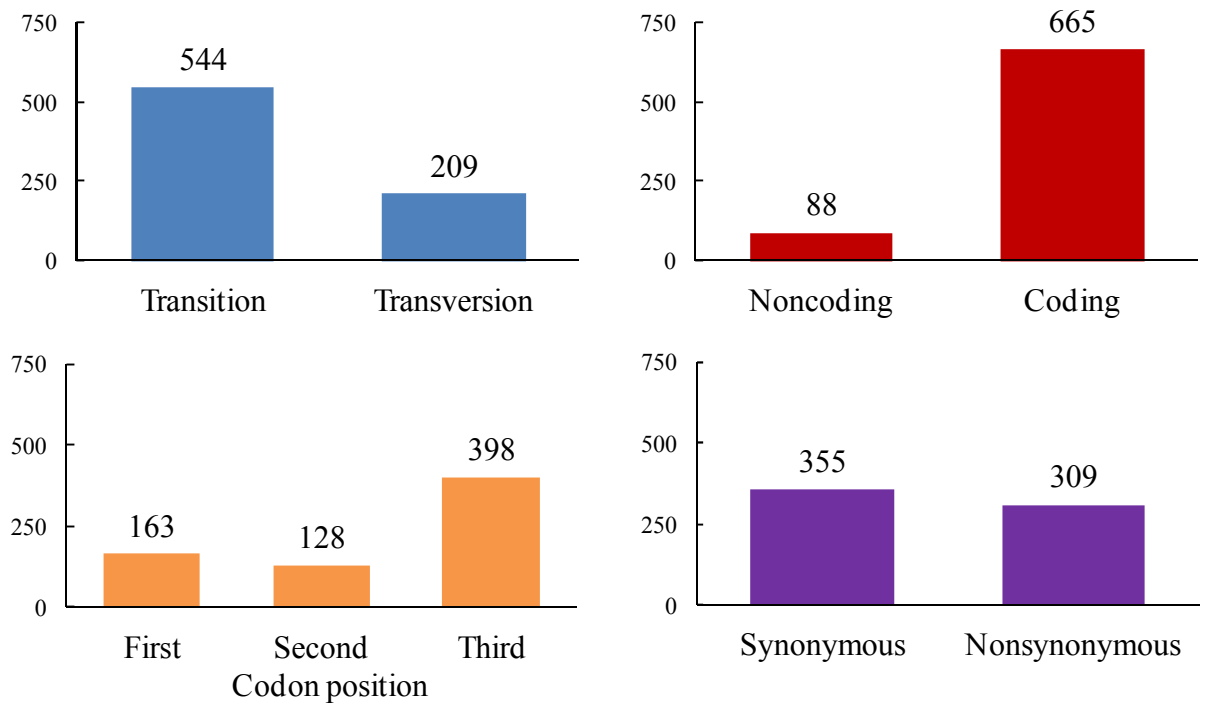



**Figure 5-2.** SNP categories. Total 753 SNPs were classified into different groups based on the polymorphisms (transition or transversion); or position in genome (noncoding and coding) and codon position; as well as synonymous or nonsynonymous. Thereof eleven and four SNPs in first and second position were overlapping with adjacent ORF.

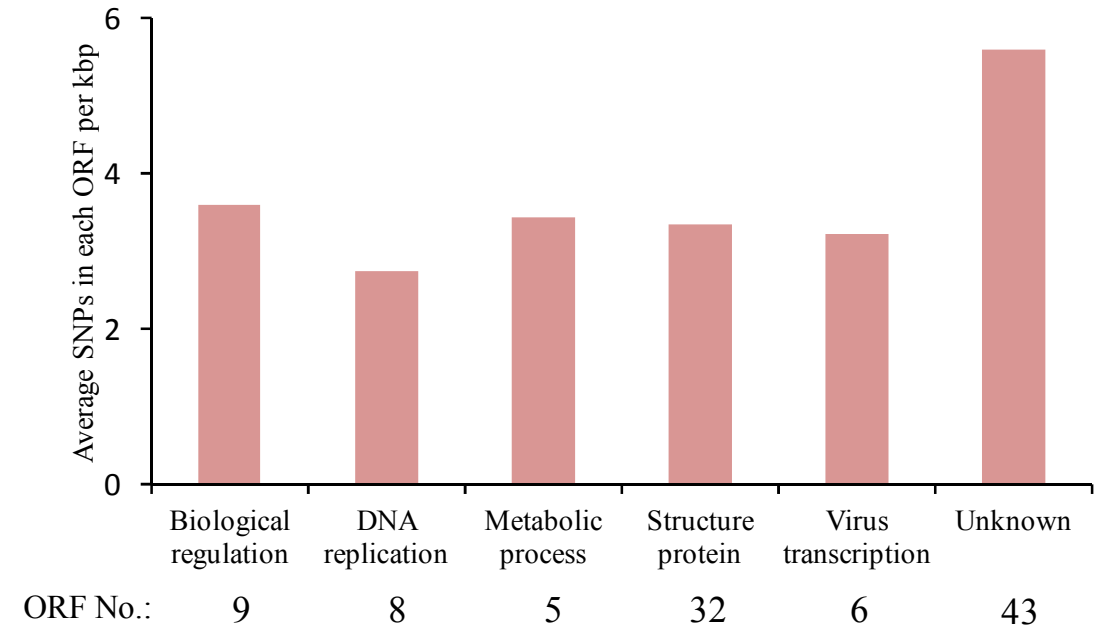



**Figure 5-3.** Density of non-synonymous SNPs in 103 ORFs. ORFs containing SNPs were grouped into biological regulation, DNA replication, metabolic process, structure protein, virus transcription and unknown function. 309 non-synonymous SNPs in each ORFs were calculated for respective SNP number per kbp.

**SNP mapping and genotype composition**

The SNP frequencies at all 753 SNP positions of each CpGV isolate were plotted against the CpGV-M reference (Figure 5-4). Because not every position was a variant within each isolate, the number of variant SNP positions of each isolate was lower than 753. The lowest number of variant SNPs in field collected isolates was detected in CpGV-M (51 SNPs) (Figure 5-4). In field isolates CpGV-I0X (244), -WW (246), -KS2 (250), -KS1 (258), -S (268), -ZY (278) and -E2 (282) variant SNP positions varied from 200 to 300, whereas -ZY2 and -JQ had the highest number of detected SNPs with 445 and 475 positions, respectively (Figure 5-4). The total number of SNP positions for these twelve field isolates was 723. Because isolates from commercial products were propagated and selected in CM, the SNPs of twelve field isolates were chosen to reflect the natural polymorphisms in CpGV. These SNP positions excluding the low quantity of SNPs and specific SNP from commercial products were used for quantification of 20 isolates as documented below. Frequency of specific SNP in genome group A (CpGV-M), genome group B (CpGV-E2), genome group E (CpGV-S), genome group F (CpGV-JQ and -ZY2), genome group G (CpGV-ALE) as well as combined genome group BDEFG (CpGV-WW, -ZY2, -KS2, -KS1, -ALE, -JQ, -ZY, -S, -I12, -E2) and DEFG (CpGV-WW, -ZY2, -KS2, -KS1, -ALE, -JQ, -ZY, -S, -I12) were applied to quantify the genotype composition of all isolates (Table 5-4). For a mixtures of two genotypes a and b, often three different SNP frequencies can be noted: SNP frequency for genotype a, frequency for genotype b and isolate frequency a+b (=often 1) when genotype a and b share the same SNP (for Example Figure 5-4, isolate CpGV-JQ). Based on specific SNP and frequency, most isolates were determined to be a mixture of two or three viral genome groups, only CpGV-M, -S, -WW, MPlus appeared to be highly homogenous (Figure 5-4, Table 5-1).

CpGV-M

The assembly of the re-sequenced isolate CpGV-M against its own reference revealed a low number of variable sites. Only 51 SNPs were identified of which only five exceeded an alternative frequency of 20%. In particular, only one SNP had a frequency above 95% (position no. 105.178, ORF123, 96.7%). In the remaining 19 CpGV isolates, the alternative frequency of this position ranged from 92% to 100%, strongly proposing that this position is most likely an annotation error in the consensus of the CpGV-M reference genome. On other hand, the high accuracy of the reference sequence was confirmed by the absence of highly variable sites in this analysis (Figure 5-4). It is important to mention that these observed 51 SNPs positions were not specific for CpGV-M only, but most likely reflected the natural and internal variation of CpGV-M. For the detection of only CpGV-M specific SNPs all other isolate sequencing results needed to be considered. In total 58 SNP positions that were variable for all isolates except CpGV-M, were counted as CpGV-M specific and were used for the quantification of CpGV-M (Figure 5-5, Table 5-1). These SNP positions were not visible in the alternative SNP frequency plot since these were identical to the reference CpGV-M itself. Instead, these 58 SNP positions were variable for all other 11 field CpGV isolates (Figure 5-5). These SNPs were considered as CpGV-M specific SNP positions and were used for the quantification of CpGV-M in all other analyzed isolates. SNPs being specific for other isolates and genome groups were not detected underlining the purity of CpGV-M (Table 5-1). Occurring SNP locations in CpGV-M were concentrated in ORF1 (*granulin*), ORF2, ORF6, ORF7 (*ie1*), ORF10 (*chitinase*) and ORF139, ORF140 (FGF-3), ORF141 (*egt*). In this region a single SNP island I/II with a SNP frequency <5%, covering 8.6 kpb from map unit at position

119,522 to 6149 was noted in the circular genome of CpGV-M. These SNPs were also identified with variable frequency across all CpGV isolates except for MPlus. Non-synonymous SNPs were found in *granulin* (1 SNP), ORF2 (2), *pk-1* (1), ORF6 (2), *ie1* (1), ORF8 (1), *chitinase* (1), ORF139 (1), *fgf-3* (4), *egt* (3).

**Table 5-4.** isolate and genome group specific SNPs. Genome group A specific SNP was represented by all combined genome group but not appeared in group A. Different genome specific group A (CpGV-M), B (CpGV-E2), D (CpGV-I12), E (CpGV-S and -WW), F (CpGV-ZY2 and -JQ) and G (CpGV-ALE) were obtained from Chapter 4, Wennmann et al, 2014 and Gebhardt et al., 2014. Genome specific group A* indicated that SNP positions that were variable for all isolates except CpGV-M, were counted as CpGV-M specific.

| Isolate group | Group count | Genome specific group |
|---|---|---|
| ZY2_JQ | 89 | F |
| WW_ZY2_KS2_KS1_ALE_JQ_ZY_S_I12 | 75 | DEFG |
| E2 | 68 | B |
| WW_ZY2_KS2_KS1_ALE_JQ_ZY_S_I12_I0X_E2 | 58 | A* |
| WW_ZY2_KS2_KS1_ALE_JQ_ZY_S_I12_E2 | 45 | BDEFG |
| ZY2 | 30 | F |
| I12_I0X | 24 | D |
| ALE_JQ | 22 | FG |
| ALE | 21 | G |
| WW_ZY2_KS2_KS1_ALE_JQ_ZY_S_M_I12_I0X_E2 | 19 | ABDEFG |
| ZY2_JQ_I12_I0X | 17 | DF |
| I0X | 16 | |
| S_I12 | 15 | DE |
| WW_ZY2_KS2_KS1_ALE_JQ_ZY | 12 | EFG |
| ZY2_JQ_E2 | 12 | BF |
| ALE_I12_I0X_E2 | 9 | BDG |
| JQ_I12_I0X | 9 | DF |
| ZY2_JQ_I12_I0X_E2 | 9 | BDF |
| I12 | 8 | D |
| JQ_I12_I0X_E2 | 6 | BDF |
| ZY2_I12_I0X | 6 | DF |
| ALE_JQ_I12_I0X | 5 | DFG |
| ALE_JQ_ZY | 5 | FG |
| JQ_I0X | 5 | F |
| WW_ZY2_KS2_KS1_ALE_JQ_ZY_S_I12_I0X | 5 | DEFG |
| WW_ZY2_KS2_KS1_ALE_JQ_ZY_S_M_I12_I0X | 5 | ADEFG |
| ZY2_JQ_S_I12_I0X_E2 | 5 | BDCEF |
| JQ | 4 | F |
| S_I12_E2 | 4 | BDE |
| WW_ZY2_KS2_KS1_ALE_JQ_ZY_S_M_I12_E2 | 4 | ABDEFG |
| ZY2_ALE_JQ | 4 | FG |
| ZY2_JQ_ZY | 4 | F |
| ALE_JQ_E2 | 3 | BFG |
| I12_I0X_E2 | 3 | BD |
| ZY2_JQ_I0X | 3 | F |
| ZY2_JQ_ZY_S_I12_I0X_E2 | 3 | |
| ALE_JQ_M | 2 | AFG |
| KS1 | 2 | - |
| KS1_ZY_M_I12 | 2 | AD |
| KS2 | 2 | - |
| KS2_KS1_ALE_JQ_ZY | 2 | FG |
| M_I12_I0X_E2 | 2 | ABD |
| S_I12_I0X_E2 | 2 | BDE |
| WW_ZY2_JQ | 2 | |
| WW_ZY2_KS2_KS1_ALE_JQ_ZY_E2 | 2 | |
| WW_ZY2_KS2_KS1_JQ_ZY_S_I12_I0X_E2 | 2 | BDEF |
| ZY2_JQ_M_E2 | 2 | |
| ZY2_ZY_I12_I0X | 2 | |
| ALE_I0X | 1 | G |
| ALE_I12 | 1 | DG |
| ALE_I12_I0X | 1 | DG |
| ALE_JQ_I0X | 1 | FG |
| ALE_JQ_I12_I0X_E2 | 1 | BDFG |
| ALE_JQ_ZY_E2 | 1 | BFG |

| | | |
|---|---|---|
| ALE_JQ_ZY_I12_I0X | 1 | DFG |
| ALE_JQ_ZY_I12_I0X_E2 | 1 | BDFG |
| JQ_S_I12 | 1 | DEF |
| JQ_S_I12_I0X | 1 | DEF |
| JQ_ZY_I12_I0X_E2 | 1 | BDF |
| JQ_ZY_S_M_I12 | 1 | ADEF |
| KS1_E2 | 1 | B |
| KS1_I12 | 1 | D |
| KS1_ZY | 1 | - |
| KS1_ZY_M | 1 | A |
| KS2_ALE_ZY_M_I12_I0X | 1 | ADG |
| KS2_I12 | 1 | D |
| KS2_I12_I0X | 1 | D |
| KS2_KS1_ALE_ZY_M_I12 | 1 | AD |
| KS2_KS1_ALE_ZY_M_I12_I0X | 1 | ADG |
| KS2_KS1_ZY | 1 | - |
| KS2_ZY_I12 | 1 | D |
| M | 1 | A |
| M_I12 | 1 | AD |
| M_I12_I0X | 1 | AD |
| S | 1 | E |
| S_I12_I0X | 1 | DE |
| WW_KS1_ALE_JQ_ZY | 1 | |
| WW_KS1_S | 1 | |
| WW_KS2_ALE_E2 | 1 | |
| WW_KS2_KS1_ALE_JQ_ZY | 1 | |
| WW_KS2_KS1_JQ_ZY_S_I12 | 1 | |
| WW_ZY_S_I12_E2 | 1 | |
| WW_ZY2_JQ_I12_I0X | 1 | |
| WW_ZY2_JQ_ZY | 1 | |
| WW_ZY2_KS1_JQ_S_I12_I0X_E2 | 1 | |
| WW_ZY2_KS1_JQ_ZY_S_I12_E2 | 1 | |
| WW_ZY2_KS2_KS1_ALE_JQ_S_I12_E2 | 1 | |
| WW_ZY2_KS2_KS1_ALE_JQ_ZY_S_M_I12 | 1 | ADEFG |
| WW_ZY2_KS2_KS1_E2 | 1 | |
| WW_ZY2_KS2_KS1_JQ_I12_E2 | 1 | |
| WW_ZY2_KS2_KS1_JQ_I12_I0X_E2 | 1 | |
| WW_ZY2_KS2_KS1_JQ_ZY_S_I12 | 1 | DEF |
| WW_ZY2_KS2_KS1_JQ_ZY_S_I12_E2 | 1 | BDEF |
| ZY_I12 | 1 | |
| ZY_M | 1 | |
| ZY_S_I12_I0X_E2 | 1 | |
| ZY2_ALE_JQ_I12_I0X | 1 | |
| ZY2_ALE_JQ_I12_I0X_E2 | 1 | |
| ZY2_ALE_JQ_S_I12_I0X_E2 | 1 | |
| ZY2_JQ_S_I12 | 1 | DEF |
| ZY2_JQ_S_I12_E2 | 1 | BDEF |
| ZY2_JQ_S_I12_I0X | 1 | |
| ZY2_JQ_S_M_I12_I0X_E2 | 1 | |
| ZY2_JQ_ZY_I12_I0X | 1 | |
| ZY2_JQ_ZY_S_I12 | 1 | |
| ZY2_JQ_ZY_S_I12_I0X | 1 | |
| ZY2_KS1_ALE_JQ | 1 | |
| ZY2_KS1_ALE_ZY_S_I12_I0X_E2 | 1 | |
| ZY2_KS1_JQ_ZY_S_I12_E2 | 1 | |
| ZY2_KS1_ZY_M | 1 | |
| ZY2_KS2_JQ_ZY_S_I12_I0X_E2 | 1 | |
| ZY2_KS2_KS1_ALE_ZY_M_I12_I0X | 1 | ADFG |
| ZY2_M_I12_I0X | 1 | |
| ZY2_S_I12_E2 | 1 | |
| ZY2_ZY_M_I12_I0X_E2 | 1 | |

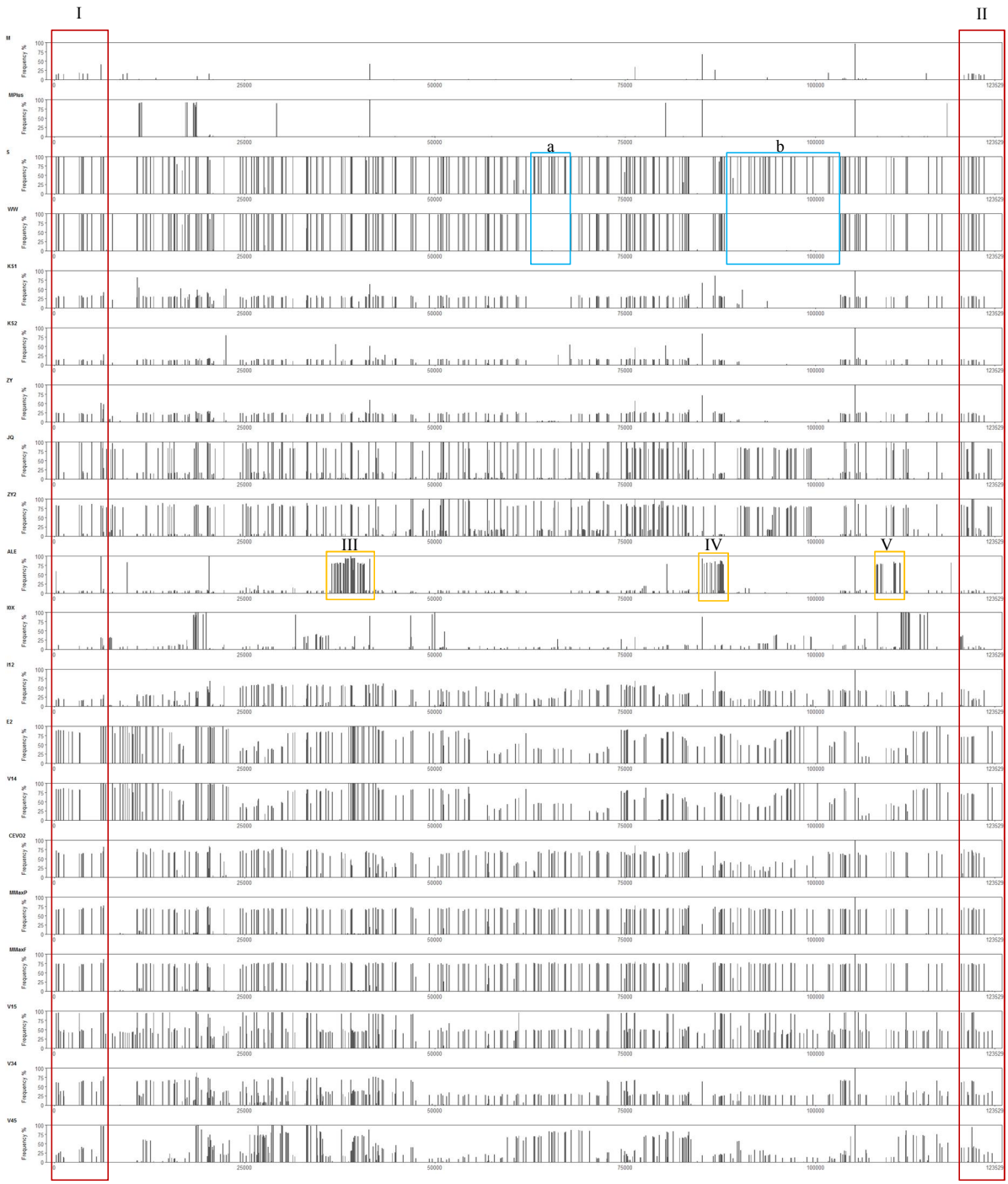
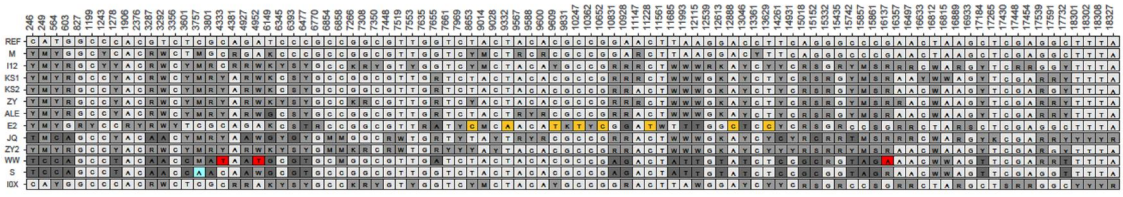
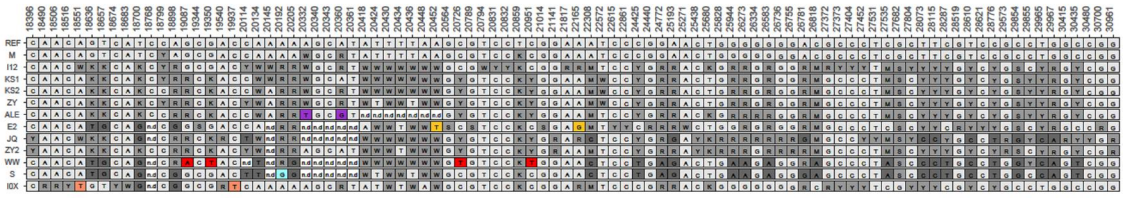
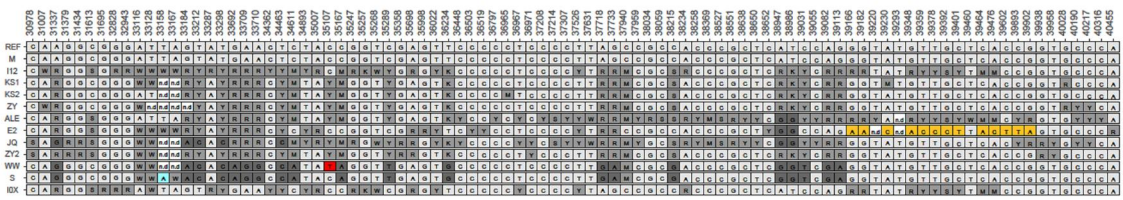
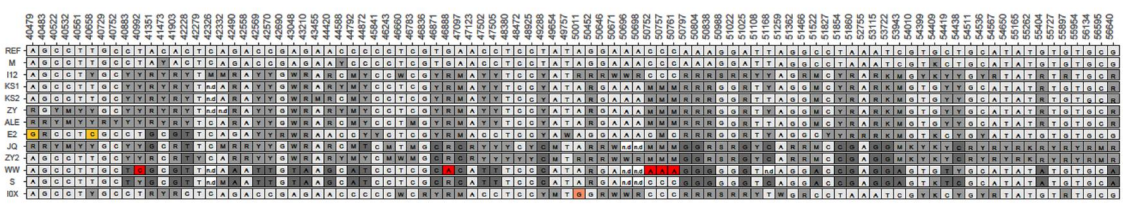


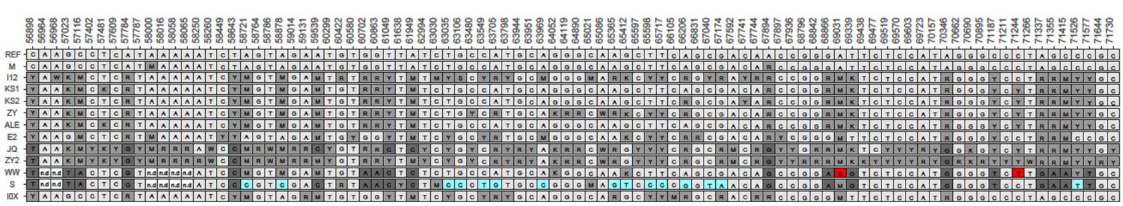

**Figure 5-4**. SNP distribution against CpGV-M reference sequence. Re-sequencing of CpGV-M, -S, -E2 and -I12 revealed 51, 268, 282, 393 SNPs, respectively, as well as 244 SNPs in -I0X. ALE, JQ, KS1, KS2, ZY, ZY2 and WW representing CpGV-ALE, -JQ, -KS1, -KS2, -ZY, -ZY2 and WW obtained from China (Chapter 4) contain 320, 475, 258, 250, 278, 445 and 246 SNPs, respectively. V15, CEVO2 (R5), MMaxP (0006) (Gueli Alletti et al., 2017) as well as MPlus, MMaxF, V14, V34 and V45, obtained from commercial products, contain 331, 281, 377 as well as 51, 364, 273, 399 and 384 SNPs, respectively. SNP island I/II across all CpGV isolates were marked with red box. SNP hotspot regions III, IV and V in CpGV-ALE and two stretches a and b in CpGV-WW were indicated by orange and blue box, respectively.

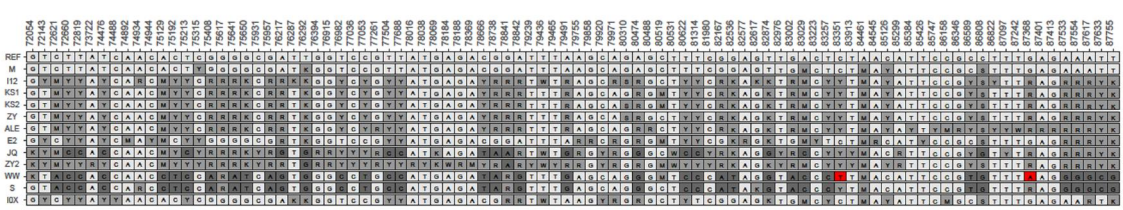


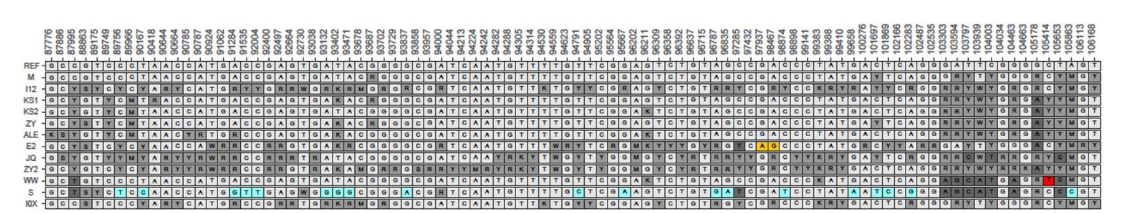












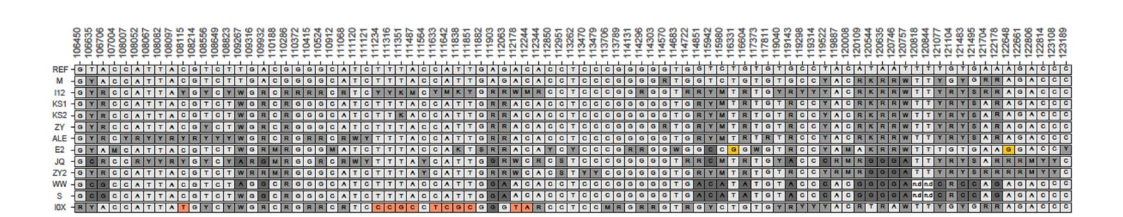


**Figure 5-5.** Position of previous CpGV-M genome sequence was set as reference (REF). Specific SNPs that is isolate specific variant site (= unique for one isolate only), where all other isolates had no alternative variant at the same position were plotted with purple (ALE), goldenrod1 (E2), red (WW), and sky blue (S), and dark orange (I0X). Ambiguous SNPs were filled by gray40 or gray60. (See page 68).

MPlus

MPlus, a resistance-breaking isolate selected from CpGV-M, contained also 51 SNPs but only ten SNPs (position no. 1649, 11147, 18799, 20332, 20360, 2859, 41473, 76292, 85126 and 105178) were identical to those positions found in CpGV-M (Figure 5-4). The overall SNP pattern revealed 26 and 25 SNPs with an alternative SNP frequency of <6% and >73%, respectively. With its unique SNP pattern it appeared very homogenous. The 58 CpGV-M specific SNP positions identified MPlus as highly similar to CpGV-M (Table 5-1) with a different SNP pattern. SNPs were mainly concentrated in ORF15 and ORF22-24, whereas the SNP island I/II was missing.

CpGV-WW and -S

The overall SNP patterns of CpGV-WW and -S were highly identical in their location and frequency (Figure 5-4). The majority of alternative SNP frequencies of CpGV-WW and -S in relation to CpGV-M were more or less equally close to 100% identifying them as homogenously different to CpGV-M. Only 18 of 246 SNP positions in CpGV-WW and 18 of 268 SNP positions in CpGV-S were <100%. Two major regions with significant difference between -S and -WW were found in genome region at position 63030 to 67744 and 88863 to 102535; here CpGV-WW lacked specific SNPs present in -S, and the -WW sequence is identical to that of the CpGV-M reference. These sequence stretches were termed CpGV-WW fingerprint regions (a) and (b) (Figure 5-4). From all 246 and 268 SNP positions of CpGV-WW and -S, respectively, no SNP positions were detected to be only specific for one or both isolates. The majority of CpGV-WW and -S specific SNPs was further specific for isolates CpGV-KS1, -KS2, -ZY, -ZY2, -JQ, -ALE and -I12 (45 SNP positions, Table 5-1) as well as, -KS1, -KS2, -ZY, -ZY2, -JQ, -ALE, -I12 and -E2 (75 SNP positions, Table 5-1). These two groups of SNPs were characterized as suitable for CpGV-WW and -S quantification for three reasons: (i) The SNP frequencies were 100% for these two groups in CpGV-WW and -S only; (ii) the genome-wide characteristic SNP pattern of CpGV-WW (including the -WW fingerprint regions a and b) was found in -KS1, -KS2 and -ZY, but at lower frequencies, identifying these isolates as mixtures with -WW (see analysis below); (iii) the difference between combined genome group of BDEFG and DEFG is the presence of shared CpGV-E2 (group B), which allowed to quantify the composition of combined genome group BE and specific genome group E, respectively.

CpGV-KS1, -KS2 and -ZY

The three isolates CpGV-KS1, -KS2 and -ZY exhibited a similar SNP pattern in their position and frequency (Figure 5-4 and 5-5). No SNP position solely specific for one of these three isolates was found (Figure 5-5). The similarity of these three isolates was further reflected by the total number of SNPs found: 258, 250 and 278 for CpGV-KS1, -KS2 and -ZY, respectively (Figure 5-4). When assembled against the CpGV-M reference, the genome-wide SNP density and abundance resembled the picture of -WW (including fingerprint region a and b) indicating a common composition. Based on the 58 CpGV-M specific SNP positions a median

ratio of 69%, 86% and 77% of CpGV-M within these Chinese field isolates was calculated (Table 5-1). Beside CpGV-M SNPs, there were only the CpGV-WW and -S specific SNPs in two combined genome groups BDEFG and DEFG, which could be used for quantification of -WW at a ratio of about 29% to 30% for -KS1, 14% for -KS2 and 22% to 23% for -ZY (Table 5-1). Based on the quantification data, these field isolates were mixtures of CpGV-M and -WW.

CpGV-I12 and -I0X

The two Iranian isolates CpGV-I12 and CpGV-I0X differed in their SNP abundance and density. Whereas CpGV-I12 was identified as a mixture of 56% CpGV-M and 46 or 42% CpGV-S according to CpGV-M and -S specific group SNP and lack -WW signature a+b, though frequencies of CpGV-M specific SNP showed uneven distribution. CpGV-I0X was only measured to consist of mainly 93% CpGV-M (Table 5-1). Besides the quantification based on CpGV-M and -S, additional 24 SNPs specific for -I12 and I07 were detected with frequencies ranging from 2% to 35% for CpGV-I12 and -I07 (Figure 5-4, Table 5-1). Especially for CpGV-I0X these specific SNP frequencies ranged between 1% to 100% representing an Iranian geographic fingerprint and the main difference to CpGV-M (Table 5-1).

CpGV-ALE

A total number of 320 variant SNP positions were detected for CpGV-ALE, the reference isolate of genome group F as defined in Chapter 4 (Figure 5-4). Ninety-six SNPs had a frequency above 50%, represented by three SNP clusters around genome position 36,000, 86,000 and 108,000, and termed SNP clusters III, IV and V in (Figure 5-4). For CpGV-ALE, 21 SNP positions were detected to be solely specific for CpGV-ALE and 22 being group specific for CpGV-ALE and -JQ (Table 5-1). Both, the 21 and 22 specific SNPs were located within the three SNP clusters III, IV, and V with SNP frequencies above 50%, representing fingerprint SNPs of CpGV-ALE (Figure 5-4). Based on the CpGV-M specific SNP frequencies the CpGV-ALE is represented by CpGV-M by about 93% and a smaller amount of about 6% CpGV-S (Table 5-1). According to the analysis, CpGV-ALE appeared to be similar to CpGV-M but with a characteristic unique fingerprint represented by the specific clusters III to V, which were solely present in CpGV-ALE specific SNPs.

CpGV-JQ and -ZY2

CpGV-JQ and -ZY2 shared 239 SNPs with -WW, echoing that genome group E was inside of these isolates. A total number of 89 SNP positions were identified to be specific for CpGV-JQ and CpGV-ZY2 only, representing their own genetic marker for genome group F defined in Chapter 4. According to these group F specific SNPs, these isolates were not homogenous but a mixture, which were also visible by two major SNP frequency groups in Figure 5-4. Group B, D and G could be excluded to be part of CpGV-JQ and CpGV-ZY2 since frequencies for their isolate specific SNPs were not measured (Table 5-1). In conclusion, the SNPs of the two combined genome groups BDEFG and DEFG could be reduced to the specificity for E and F (EF) (Table 5-1). For CpGV-JQ, the SNP frequency in group F was similar with that in group BDEFG (75 specific SNPs) but different from DEFG (45 specific SNPs), indicating that 79% genome group F was represented by genome group BDEFG, whereas 17% genome group E was indicated by genome group DEFG. 15% CpGV-M (group A) and 77% genome group F was also identified in -ZY2 as quantification in Table 5-1. It was difficult to determine the exact genomic proportion of CpGV-WW, but ranging from 5 to 42%.

CpGV-E2

CpGV-E2 and V14 were unique as they showed a highly uneven SNP frequency distribution, resembling a “wave”-like pattern, clearly visible for the alternative SNP frequencies between genome positions from 10,000 to 113,000. Another characteristic was the presence of two cluster regions with SNP frequencies of 100% at genome positions (i) 10,000 to 13,000 and (ii) 39,000 to 43,000. Despite the heterogeneous SNP pattern, a total number of 68 only for CpGV-E2 specific SNPs positions was detected (Table 5-1). These specific SNPs were used to calculate the median presence of CpGV-E2. Its quantification is hampered by its genotype heterogeneity, which is reflected by the self-quantification with median ratio of 84% (28-100%) by the CpGV-E2 specific SNPs only. When the 75 SNPs specific for CpGV-WW/-S and -E2 were used for quantification, a median ratio of 67% (3-100%) was measured. However, the 45 group specific SNP for CpGV-WW/S without -E2 did not detect the presence of any CpGV-WW or -S (Table 5-1). Except for the 58 CpGV-M specific SNPs with a frequency of up to 20% no other specific SNP frequencies could be calculated (Table 5-1), hinting that it is a mixture of CpGV-M, but difficult to quantify due to the range of the CpGV-M specific SNP frequencies.

**CpGV commercial isolates**

The isolate V14 was a commercial *in vivo* propagation of CpGV-E2. SNP positions of both isolates (269/282 in CpGV-E2 and 269/273 in V14) were identical.

CEVO2, MMaxF and MMaxP showed the similar SNP pattern as CpGV-I12, of which all were mixtures of CpGV-M and -S. Since 9 and 14 SNPs specific for CpGV-S were identified in CpGV-WW fingerprint regions a and b, respectively, indicating that these isolates contain genome group E virus similar to CpGV-S but not to CpGV-WW. SNP frequency of identical SNP positions in MMaxP (322/331) and MMaxF (322/364) were used to evaluate the genotype ratio from different production batches during propagation.

V15 was comprised of 42% CpGV-E2 based on the -E2 specific SNPs (Table 5-1) and 49% CpGV-S based on the WW/S group specific SNPs that did not include -E2 (Table 5-1). Consequently, the 75 WW/S group specific SNPs that were further specific for -E2 showed a ratio of 93% indicating a mixture of CpGV-E2 and CpGV-S (Table 5-1). Besides the 58 SNP positions specific for CpGV-M with a frequency of 7% (3-53%), no other isolate was detected (Table 5-1). The difficulty in the quantification of CpGV-E2 due its high heterogeneity was reflected by 5 to 95 percentiles ranging from 0 to 48% in V15 as well as -E2 with 28 to 100% (Table 5-1).

V34 and V45 also appeared to have an uneven frequency distribution of SNPs. Since no specific SNPs from genome groups D, F and G were identified in these isolates, the specific genome group combination of BDEFG and DEFG were thus shorted into group BE and E respectively. Both isolates contained SNP signals of genome group E were then 28% and 19% in V34 and V45 respectively. Genome group A was presented in V34 and V45 with 38% and 61% as well. In consideration of above of group E ratio, the group B ratio of V34 and V45 was thus 39% and 11% respectively, derived from the specific genome group combination BE (Table 5-1).

## Indels location and frequency

In reference to the genome of CpGV-M, 58 deletions in 25 ORFs and 40 insertions in 23 ORFs were detected with frequencies >10% (Table 5-5). Insertions and deletions with fewer frequencies were considered as minor mutations and were excluded from the analysis. Major indels located in ORF 25-26 were excluded in this analysis for the abundance of repeated sequence located in region of ORF25, ORF26 and ORF27 (Wennmann et al., 2017). Twelve geographic CpGVs harbored insertions and deletions in *pe38*, *odv-e66*, *mp-nase*, *p13*, *39K*, *lef11*, *38.7kd*, *p33*, *desmoplakin*, *fgf3*, *egt* and other unannotated genes evidenced by 10-95.6% of aligned reads. A deletion of a thymidine (T) residue (position no. 90,356, ORF110) was found in all read assemblies of all 20 CpGV isolates, where its frequency ranged from 90.4-97.7% indicating a possible error might exist in reference CpGV-M.

Because of the putative role of *pe38* in breaking type I resistance (Gebhardt et al., 2014), its 12 bp repeat motif of (GACACAGTGGAT) was analyzed with two different methods (Figure 5-6). First, with a read counting method, which considered reads only that comprised the entire repeat motif including 12 and 10 bp unique sequences downstream and upstream, respectively, of the motif (Figure 5-6A). Reads that were found to contain the entire repeat region were used to measure the length distribution of the repeat motif for each isolate. The second, the length of the motif was analyzed by the Geneious software implemented insertion/deletion detection algorithm, and was considered as the standard motif length measurement in this thesis (Figure 5-6B). Between one to five copies of the 12 bp repeat motifs (1-5×12 bp) were identified in the *pe38* of the 20 CpGV isolates (Figure 5-6A), which was consistent to the previously detected repeat motifs (Chapter 4). However only 1×12 bp (all isolates except CpGV-M and -ALE), 3×12 bp (all except CpGV-E2 and -I0X) and 4×12 bp (only CpGV-KS2) were detected (Figure 5-6B), when the indels detection workflow was applied to measure the 12 bp motif. Hence, the result of read counting method was used to describe the ratio of different 12 bp repeat motif in 20 CpGV isolates. In the re-sequenced CpGV-M a 1×12 bp repeat motif was present with 0.1% (3 reads), however its abundance increased into 95% (1293 reads) in MPlus, which is a selection from CpGV-M, whereas the rest of the genomes of CpGV-M and MPlus showed highly similar SNP patterns. The 3×12 bp repeat motif was found in CpGV-ZY2, -ALE, -KS1, -KS2, -M, -I12, MMaxP, MMaxF, CEVO2, V14, V15, V34 and V45 with different ratios. The 5×12 bp repeat type of *pe38* was found only in CpGV-KS1, -KS2 and CEVO2 with 58% (379 reads), 50% (200) and 6.7% (22). According to read counting method an additional 2×12 bp motif (CpGV-ALE, -ZY, and MMaxP, MMaxF) was found but at low proportions. A 5×12 bp motif was found within reads of CpGV-KS2 (Figure 5-6A).

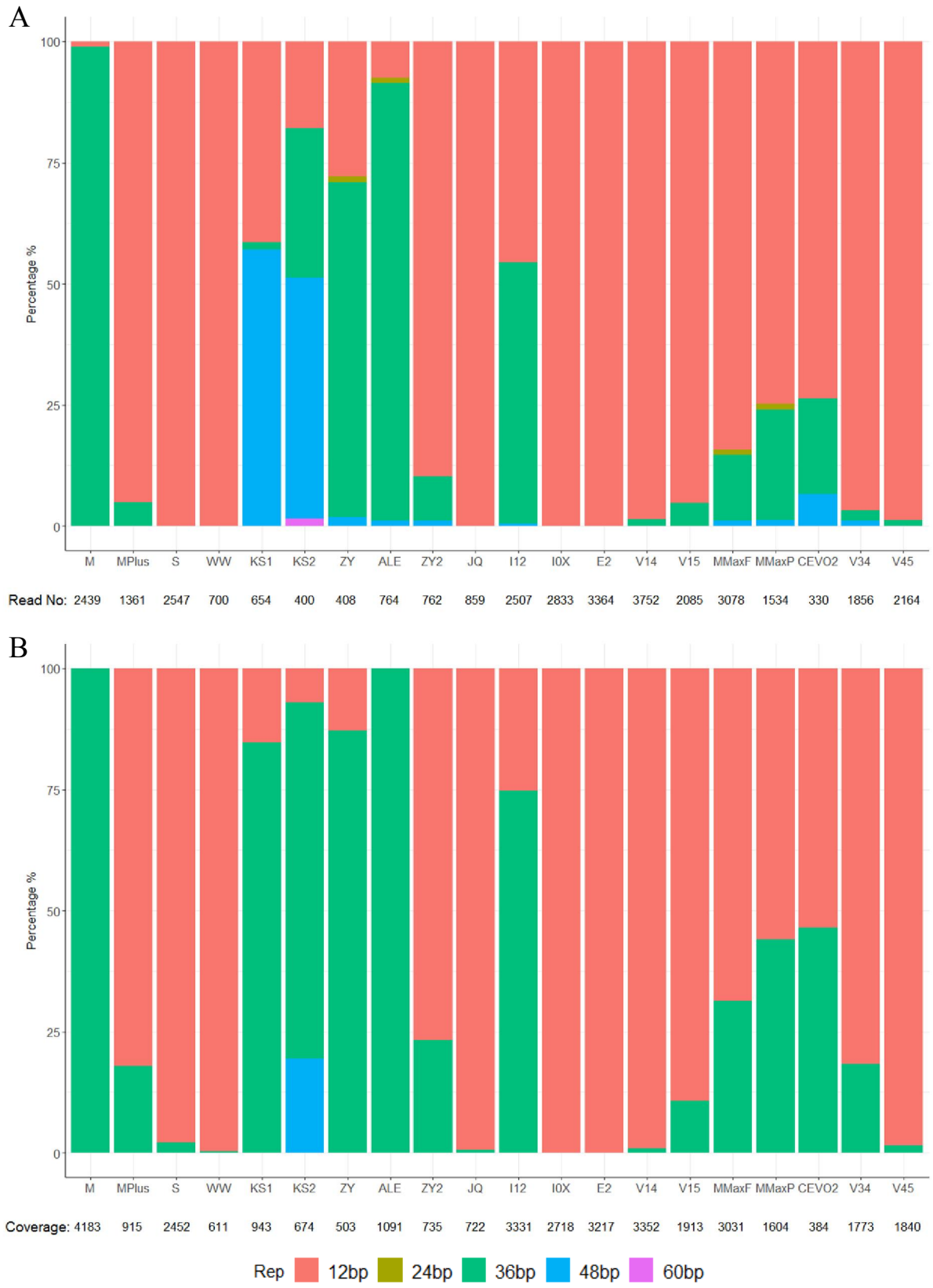



**Figure 5-6.** Percentage of five types of 12 bp repeat variants in *pe38* in 20 CpGV isolates. A: percentage and amount of reads obtained from read counting methods; B: percentage and amount derived from genome assembly. Total reads and coverage number was plotted below.

### Hierarchical clustering on principal components (HCPC)

Consensus sequence based phylogenies can properly reflect the relationship of isolates, only if the isolates are highly homogenous. In case of high sequence heterogeneity and/or mixtures of different genomes as observed for most of the studied isolates, calculation of a consensus sequence ignores existing variants in a virus population. To overcome this limitation, a hierarchical clustering on principal components (HCPC) based on the SNP frequency and position was applied to identify the dissimilarity and similarity among 20 CpGV isolates. HCPC program initiated from factorial analysis (principal component analysis, PCA), followed by hierarchical clustering. The clustering tree was chosen and consolidated by K means with the cluster centers. Since the first seven principal components (PC) covered 95.2% information (variances) existing in the SNP data set and excluded the noise from data (Figure 5-7A), individual (isolate) factor map was drawn on a plot that corresponded to the spatial relative position of each isolate (Figure 5-7B). The hierarchical tree suggested three clusters because the inertia gain to more clusters was minor. For the further analyses, however, it was decided to use six clusters, in consideration of two new genome group F (CpGV-ZY2 and -JQ) and group G (CpGV-ALE) based on the CpGV phylogeny in Chapter 4 (Figure 5-7C). Highly homogenous isolates CpGV-M and MPlus were distributed in bottom left of quadrant and CpGV-I0X, -KS1, -KS2, -ZY and -ALE containing the major genotype from genome group A (CpGV-M) were closer to them (Figure 5-7B and 7D); on the contrary the homogenous isolates CpGV-WW and -S were located at the right bottom in the coordinate diagram and MMaxP, MMaxF, CEOV2 containing the major genome group E (CpGV-S); highly heterogenous CpGV-E2, -I12 and V14, V45 were located nearly on Dim1 axis (first principal component) except V34 which was composed of almost one third of genome group A (38%), genome group B (28%) and genome group E (39%) respectively and located in the middle of these groups; V15 was also located in the middle between genome group E and genome group B as it was composed of nearly 50% CpGV-E2 and 50% CpGV-S (Figure 5-7B and 7D). Although CpGV-ZY2 and -JQ seemed closer to CpGV-E2 and -S in the factor map (Figure 5-7B), these isolates were actually far away from CpGV-E2 and -S in three dimensional plot (Figure 5-8). The ratio of CpGV-M to -WW or -S in isolate -KS2, -ZY, -KS1, -I12, CEVO2, MMaxP, MMaxF determined the location of each isolate on the factor map (Figure 5-7B and 5-7D) in the three dimensions (Figure 5-8). Phylogenetic analysis corresponded somehow to the clustering results, but it was not able to distinguish whether a given isolate was genetically homogeneous, a mixture of homogenous viruses or heterogeneous. For example, phylogenetic analyses grouped the homogeneous CpGV-M and -S together with mixed MMaxP and MMaxF, which is different from HCPC analysis where they have been clustered together based on genetic composition (Figure 5-9).

**Table 5-5.** Insertions and deletions (indels) located in open reading frame and their position and frequency, of which all were obtained from genome assembly.

Deletions frequency (%) in 20 CpGVs

| Postion | ORF gene | M | KS2 | MMplus | I0X | ALE | KS1 | ZY | I12 | V34 | V45 | E2 | V14 | ZY2 | JQ | MMaxP | MMaxF | CEVO2 | V15 | WW | S |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2,965 | ORF5 | 13.7 | | | | | | | | | | | | | | | | | | | |
| 9,086 | ORF12 | 19.7 | | | | 74.3 | | | | | | | | 11.9 | | | | | | | |
| 15,663 | ORF20 orf16L | | | 68.9 | | | 22.1 | 16.5 | 34.3 | 60.6 | | | | | | 58.6 | 69.5 | 59.6 | 46.2 | 98.3 | 99.9 |
| 15,671 | ORF20 orf16L | | | | | | 20.9 | | | | | | | | 12.0 | | | 58.8 | 44.8 | 93.4 | 98.9 |
| 17,349 | ORF22 orf17R | | | | | | | | | 15.3 | | | | | | | | | | | |
| 18,745 | ORF24 pe-38 | | 7.1 | 82.1 | 100.0 | | 15.3 | 12.8 | 25.2 | 81.6 | 98.4 | 100.0 | 99.1 | 76.8 | 99.4 | 56.0 | 68.7 | 53.5 | 89.2 | 99.7 | 97.9 |
| 27,384 | ORF32 | | | | | | | | | | | | | | 11.5 | | | | | | |
| 27,507 | ORF32 | | | | | | | | | | | | | | | | | 11.5 | | | |
| 27,693 | ORF32 | | | | | | | | | | 27.2 | | | | | | | | | | |
| 27,728 | ORF32 | | | | | | | | | | 65.6 | | | | | | | | | | |
| 27,730 | ORF32 | | | | | | | | | | 64.4 | | | | | | | | | | |
| 29,998 | ORF34 | | | | | | | | | | 76.4 | | | | | | | | | | |
| 33,156 | ORF37 odv-e66 | | | | | | 20.7 | | | 37.8 | 79.8 | 28.0 | 28.5 | 62.4 | 71.3 | 57.2 | | 54.2 | 59.3 | 75.2 | |
| 33,158 | ORF37 odv-e66 | | | | | | 21.8 | | | 37.4 | 82.3 | | | 62.6 | | 58.4 | | 55.5 | 61.2 | | |
| 33,161 | ORF37 odv-e66 | | | | | | | | | | | 27.6 | | | | | | | | | |
| 33,162 | ORF37 odv-e66 | | | | | | | | | | | | | | 58.1 | | | | | 64.7 | |
| 33,164 | ORF37 odv-e66 | | | | | | | | | 35.4 | 76.7 | | 25.8 | 57.6 | 62.6 | | | | 55.9 | 66.4 | |
| 33,167 | ORF37 odv-e66 | | | | | | 22.2 | | | 34.9 | 73.0 | 24.5 | 24.9 | 55.4 | 57.1 | 55.4 | | 53.8 | 53.8 | 63.0 | |
| 36,557 | ORF45 | | 49.0 | | | | | | | | | | | | | | | | | | |
| 39,289 | ORF47 p13 | | | | | 69.3 | | | | 26.8 | 54.7 | 77.4 | 74.4 | | | | | | 34.1 | | |
| 42,315 | ORF50/51 | | | | 11.2 | | | | | | | | | | | | | | | | |
| 42,321 | ORF50/51 | | | | | | | | | | | | | 27.8 | | | | | | | |
| 42,324 | ORF50/51 | | | | | | | | | 27.2 | 33.6 | | | 29.2 | | 46.1 | | | | 41.6 | |
| 42,327 | ORF50/51 | | | | | | | | | | | | | 24.7 | | | | | | | |
| 43,491 | ORF50/51 | | 16.1 | | 31.3 | 24.0 | 15.5 | | 33.3 | 24.6 | 31.3 | | | | | 37.6 | 31.5 | 27.9 | 21.1 | | 33.7 |
| 43,506 | ORF50/51 | | | | | 28.6 | | | | | | | | 24.3 | 25.0 | | | | | | |
| 43,519 | ORF50/51 | | | | | | | | | | | | | 26.3 | 28.5 | | | | | | |
| 47,488 | ORF57 pp31/39K; ORF58 lef-11 | | | | | | | | | | | | | 12.2 | | | | | | | |
| 47,490 | ORF57 pp31/39K | | | | | | | | | | | | | 12.1 | | | | | | | |
| 47,491 | ORF58 lef-11 | | | | | | | | | | | | | 12.3 | | | | | | | |
| 47,493 | ORF57 pp31/39K | | | | | | | | | | | | | 12.5 | | | | | | | |
| 47,494 | ORF58 lef-11 | | | | | | | | | | | | | 12.6 | | | | | | | |

*Cont. Table 5-5*

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 47,496 | ORF57 pp31/39K | | | | | | | | | | | | | 12.6 | | | | | | |
| 47,497 | ORF58 lef-11 | | | | | | | | | | | | | 12.7 | | | | | | |
| 47,499 | ORF57 pp31/39K | | | | | | | | | | | | | 12.3 | | | | | | |
| 47,500 | ORF58 lef-11 | | | | | | | | | | | | | 12.3 | | | | | | |
| 47,502 | ORF57 pp31/39K | | | | | | | | | | | | | 11.7 | | | | | | |
| 50,901 | ORF61 | | | | | | | | | | | | | 13.0 | | | | | | |
| 51,137 | ORF62 | | | | | | 15.4 | 12.8 | 23.5 | 17.3 | | 55.5 | 54.7 | | | 45.9 | 44.9 | 37.1 | 32.0 | | 67.8 |
| 51,145 | ORF62 | | | | | | | | | | | | | | 10.0 | | | | | | |
| 51,394 | ORF62 | | | | | | | | | | | | | 10.4 | | | | | | |
| 51,397 | ORF62 | | | | | | | | | | | | | 10.6 | | | | | | |
| 53,000 | ORF64 | | | | | | | | 12.3 | | | | | | | 16.7 | 18.1 | 13.2 | 17.6 | | 29.6 |
| 53,987 | ORF66 | | 12.8 | | | | 22.7 | 20.7 | 38.1 | 21.0 | 9.4 | | | | 14.5 | 58.5 | 73.0 | 71.7 | 40.6 | 99.3 | 99.9 |
| 56,838 | ORF70 | | 12.6 | | | | 27.4 | 19.9 | | 22.3 | 17.7 | | | | 16.2 | 60.4 | | 59.0 | 40.1 | | |
| 56,881 | ORF70 | | | | | | 16.7 | 12.8 | | 14.7 | 12.6 | | | | 10.4 | 42.1 | | 46.7 | 28.1 | 79.0 | |
| 56,898 | ORF70 | | | | | | 15.0 | 11.4 | | 13.7 | 11.6 | | | | | 39.6 | | 42.3 | 26.1 | 69.1 | |
| 56,963 | ORF70 | | 11.3 | | | | 26.4 | 18.0 | 30.7 | 24.9 | 19.0 | | | | 13.0 | 45.6 | 49.6 | 45.9 | 40.1 | 61.7 | 61.9 |
| 56,964 | ORF70 | | | | | | | | 35.9 | | | | | | | 60.8 | 66.4 | 59.2 | | 90.2 | 91.0 |
| 58,338 | ORF73 38.7kd | | | | | | | | | | | | | | | | | | | | |
| 62,508 | ORF77 | | | | | | | | 29.5 | 21.1 | 57.4 | | | | | 54.5 | 64.6 | 14.9 | 37.7 | | 85.0 |
| 76,278 | ORF93 P33 | | | | | | | | | | | | | 11.3 | | | | | | | |
| 90,356 | ORF110 | 92.9 | 96.2 | 94.5 | 93.1 | 96.3 | 95.4 | 96.6 | 93.0 | 97.7 | 97.2 | 93.5 | 93.1 | 95.9 | 95.9 | 97.7 | 93.2 | 90.4 | 97.4 | 96.0 | 93.6 |
| 94,304 | ORF112 desmoplakin | | | | 28.9 | | | | 34.9 | 22.5 | 17.9 | 47.4 | 55.4 | 68.8 | 70.2 | 55.5 | 58.7 | 25.8 | 82.6 | | 76.3 |
| 94,306 | ORF112 desmoplakin | | | | | | | | | | | | | 68.6 | | | | | | | |
| 94,318 | ORF112 desmoplakin | | | | | | | | | | | | | 12.0 | | | | | | | |
| 112,319 | ORF129/130 | | | | 87.2 | | | | | | | | | | | | | | | | |
| 119,851 | ORF140 FGF-3 | 14.4 | | | | | | | | | | | | | | | | | | | |

Insertions frequency in 20 CpGVs

| Position | ORF gene | M | KS2 | MMplus | I0X | ALE | KS1 | ZY | I12 | V34 | V45 | E2 | V14 | ZY2 | JQ | MMaxP | MMaxF | CEVO2 | V15 | WW | S |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 919 | ORF2 | 13.5 | | | | | | | | | | | | | | | | | | | |
| 924 | ORF2 | | | | | | 14.1 | 10.2 | | 16.4 | | | | | | 43.9 | 43.9 | 38.5 | 32.8 | 54.1 | |
| 3,188 | ORF6 | | | | | | | | | | | | | | | | | | | | |
| 18,745 | ORF24 pe-38 | | 19.4 | | | | | | | | | | | | | | | | | | |
| 27,380 | ORF32 | | | | | | | | | | | | | 36.7 | | | | | | | |
| 27,381 | ORF32 | | | | | | | | | | | | | | 38.4 | | | | | | |
| 27,405 | ORF32 | | | | | | | | | | | | | 15.4 | 12.8 | | | | | | |

*Cont. Table 5-5*

| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 27,572 | ORF32 | | | | | | | | | | 60.0 | | | | | | | | | | |
| 27,683 | ORF32 | | | | | | | | 17.4 | 19.2 | | 21.1 | 22.5 | 17.1 | 20.5 | 21.3 | 21.3 | 22.9 | 27.4 | 20.0 | 26.9 |
| 27,691 | ORF32 | | | | | | | | | | 22.6 | | | | | | | | | | |
| 33,099 | ORF37 odv-e66 | | | | | | | | | | 17.4 | | | | | 14.3 | 14.3 | 24.0 | 10.4 | | |
| 33,162 | ORF37 odv-e66 | | | | | | | | | | | | | | | | | | 14.9 | | |
| 35,227 | ORF43 | | | | 28.2 | | | | | | 62.6 | | | | | | | | | | |
| 37,049 | ORF46 mp-nase | | | | | | 14.6 | 12.1 | 42.0 | 24.2 | 22.0 | | | | 10.0 | 49.1 | 49.1 | 41.2 | 37.1 | 60.2 | 76.5 |
| 39,219 | ORF47 p13 | | | | | 84.8 | | | | 33.7 | 50.2 | 95.6 | 95.1 | | | | | | 41.2 | | |
| 40,469 | ORF49 | | | | | | | | | 22.3 | 45.8 | 53.9 | 53.4 | | | | | | 27.2 | | |
| 43,506 | ORF50/51 | | | | | | | | | | | | | 38.9 | 45.1 | | | | | | |
| 43,541 | ORF50/51 | | | | | | | | | 16.5 | | 40.4 | 42.7 | | | | | | 15.5 | | |
| 43,995 | ORF50/51 | | 15.5 | | | | 25.9 | 23.1 | 53.2 | 32.0 | 27.0 | 66.0 | 66.2 | | 14.6 | 60.4 | 60.4 | 33.7 | 46.7 | | 87.7 |
| 43,995 | ORF52b | | 15.5 | | | | 25.9 | 23.1 | 53.2 | 32.0 | 27.0 | 66.0 | 66.2 | | 14.6 | 60.4 | 60.4 | 33.7 | 46.7 | | 87.7 |
| 47,482 | ORF57 pp31/39K | | | | | | | | 12.8 | | | | | | | | | | | 43.2 | 24.5 |
| 47,482 | ORF58 lef-11 | | | | | | | | 12.8 | | | | | | | | | | | 43.2 | 24.5 |
| 50,953 | ORF61 | | | | | | | | | | | | | 11.9 | | | | | | | |
| 50,976 | ORF61 | | | | | | | | | | | | | 13.3 | | | | | | | |
| 51,047 | ORF61 | | | | | | | | | | | | | 15.7 | | | | | | | |
| 51,152 | ORF62 | | | | | | | | | | | 12.1 | 11.0 | 58.6 | 57.9 | | | | | | 14.1 |
| 52,662 | ORF64 | | | | | | | | | | | 36.0 | 37.6 | | | | | | | | |
| 53,292 | ORF64 | | | | | | | | | | | 59.2 | 61.6 | | | | | | | | |
| 56,939 | ORF70 | | | | | | 14.3 | | 25.4 | 16.5 | 12.4 | | | 11.8 | | 44.3 | 44.3 | 47.4 | 28.7 | 64.8 | 74.6 |
| 57,003 | ORF70 | | | | | | 20.6 | 15.2 | 35.0 | 20.6 | 14.9 | | | | 10.5 | 52.1 | 52.1 | 56.0 | 33.6 | 63.8 | 83.1 |
| 58,361 | ORF73 38.7kd | | | | | | 10.3 | | | | | | | | | | | | | 33.3 | |
| 58,587 | ORF73 38.7kd | | | | | | | | | | | | | 11.5 | | | | | | | |
| 58,647 | ORF73 38.7kd | | | | | | | | | | | | | 10.2 | | | | | | | |
| 65,131 | ORF82a | | | | | | | | | | | | | 12.5 | | | | | | | |
| 89,965 | ORF109 | | | | | | | | 11.4 | 13.4 | 17.7 | | | | | 43.1 | 43.1 | 28.6 | 28.2 | | 55.3 |
| 96,030 | ORF112 desmoplakin | | | | | | | | | | | 44.1 | 47.5 | | | | | | 23.7 | | |
| 119,914 | ORF140 FGF-3 | 13.1 | | | | | | | | | | | | | | | | | | | |
| 121,512 | ORF141 egt | | | | | | 17.6 | | | | | | | | | | | | | 78.4 | |
| 122,309 | ORF141 egt | 12.9 | | | | | | | | | | | | | | | | | | | |
| 122,439 | ORF142 | | | | | | | | | | | | | | | | | 10.5 | | | |

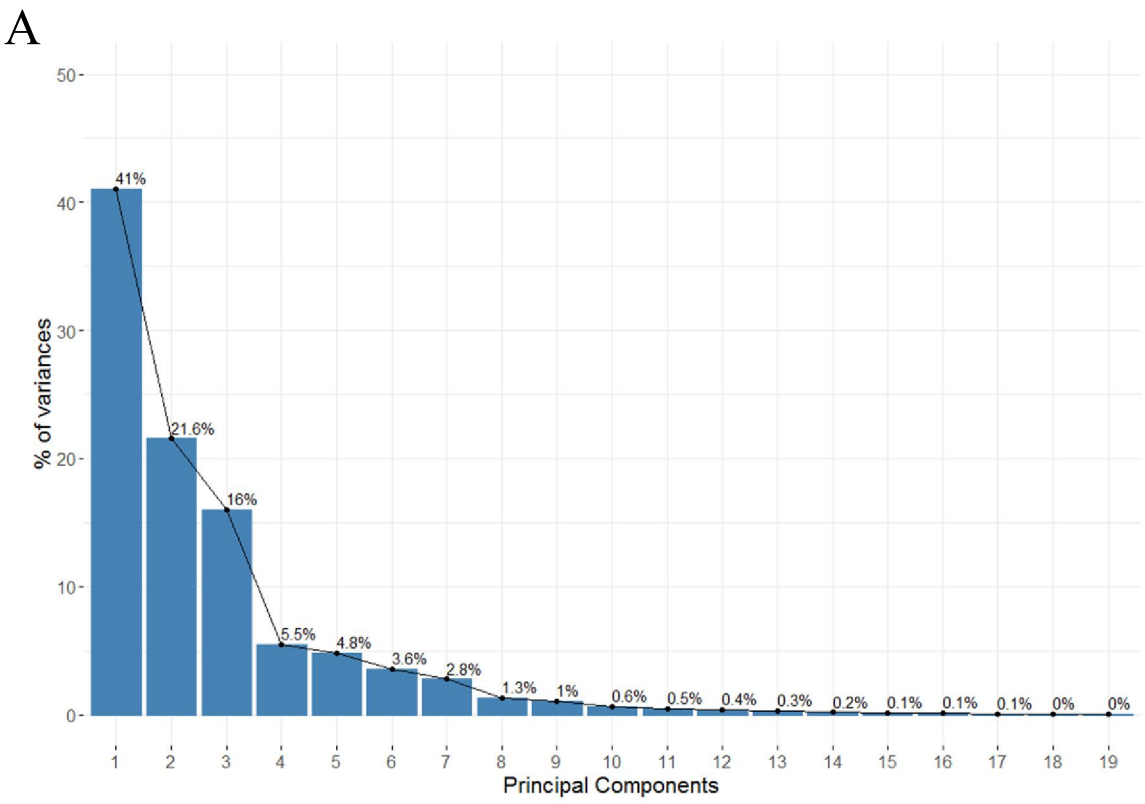



**Figure 5-7.** Hierarchical Clustering on Principal Component analysis of 20 CpGV isolates based on identified SNPs dataset. A: percentage of variance of each component; B: relative position of 20 CpGV isolates in two dimension factor map evaluated by principal component analysis; C: the first seven components was used for hierarchical clustering instead of three suggested by inertia gain and six clusters based on the previously proposed genome group A, B, D, E, F and G; D: the hierarchical clustering was plotted on factor map in two dimension to elucidate the isolate clustering in spatial position.

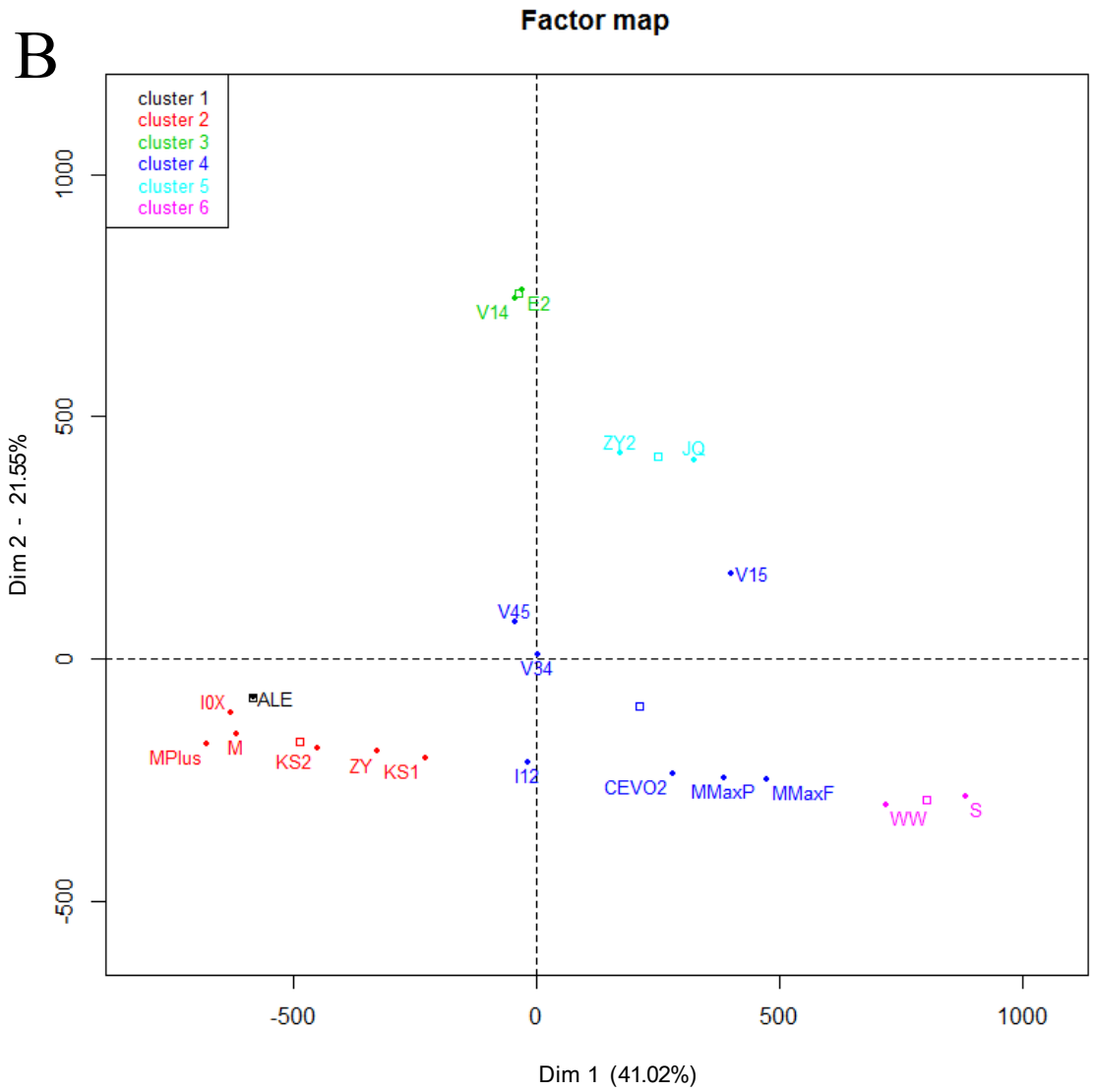



Figure 5-7 continued legend see page 78

C

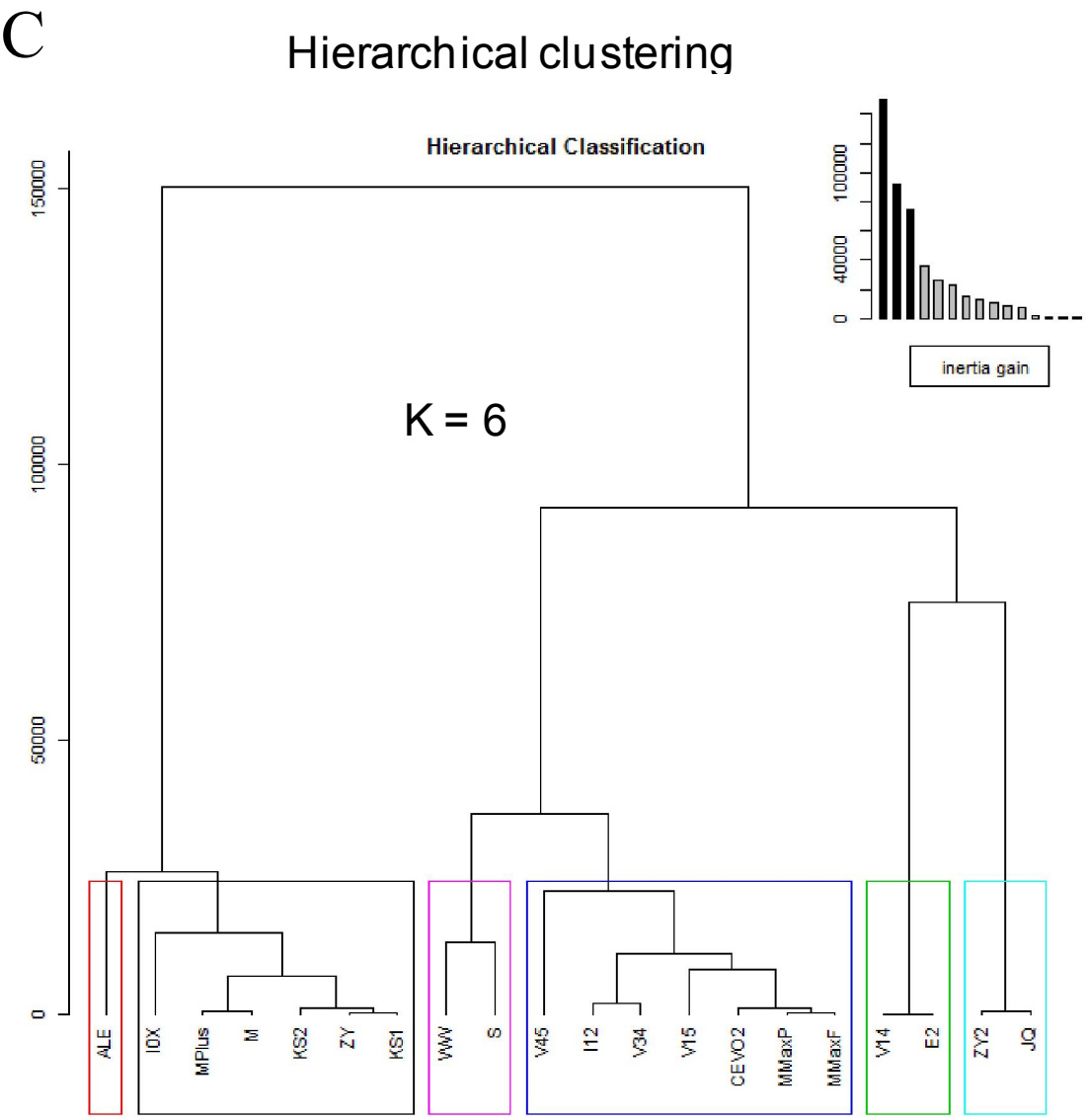



Figure 5-7 continued legend see page 78

D

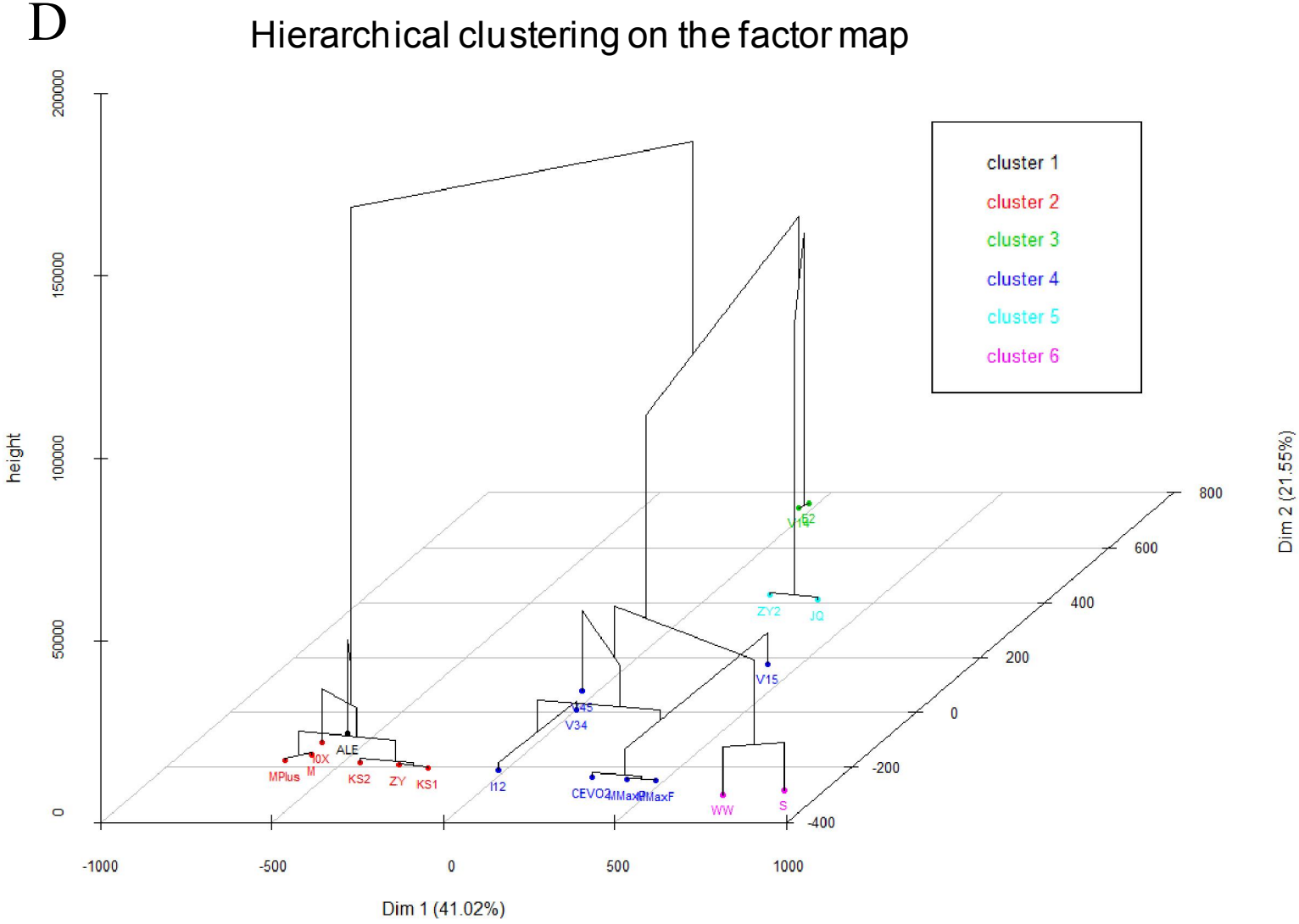



Figure 5-7 continued legend see page 78

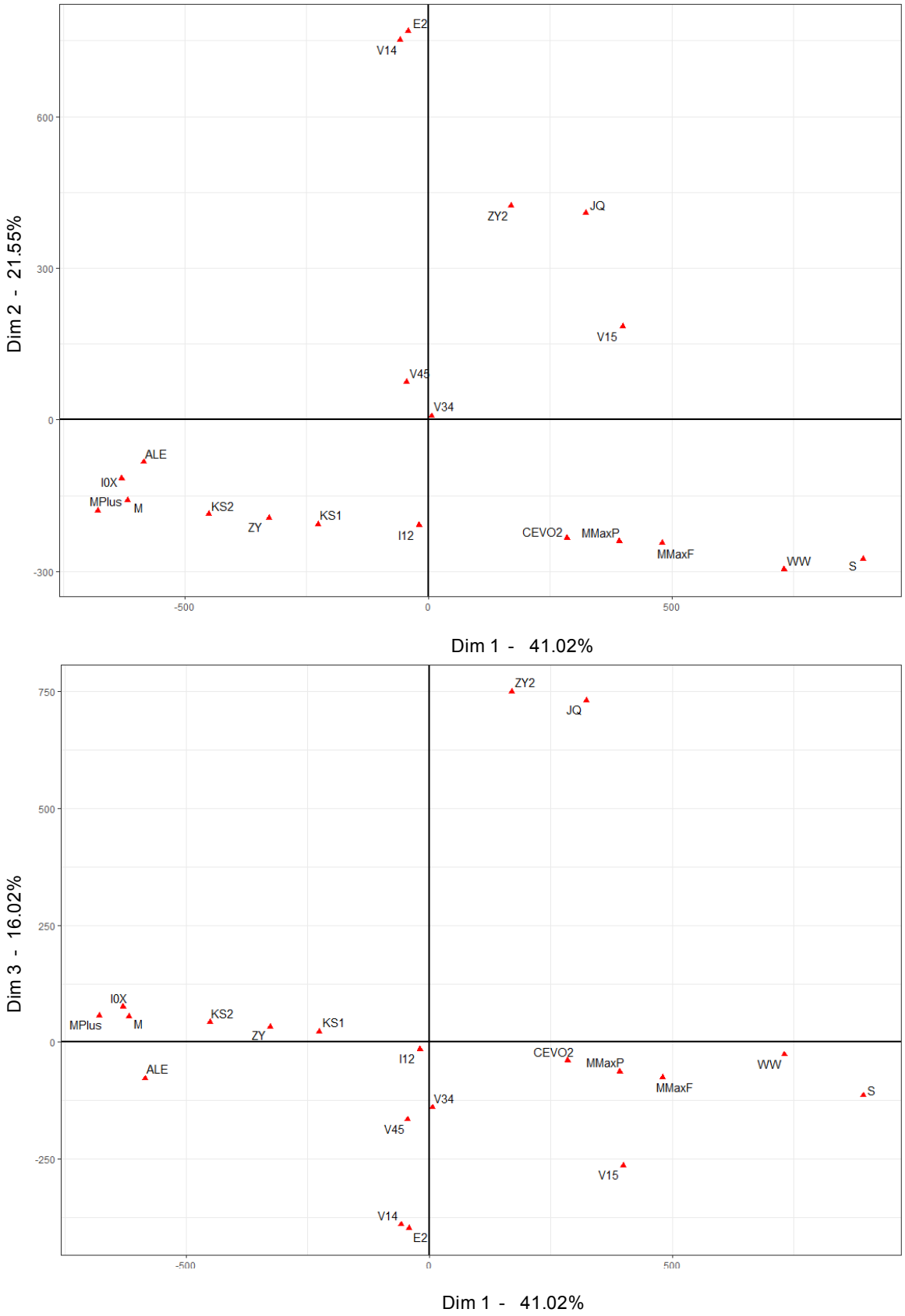



**Figure 5-8.** The spatial clustering position of 20 CpGV isolates was plotted in three dimesion (Dim 1 – Dim 3) in factorial analysis using the first three principal components, of which sum up to 78.6% is able to represent the whole data information and remove noise.

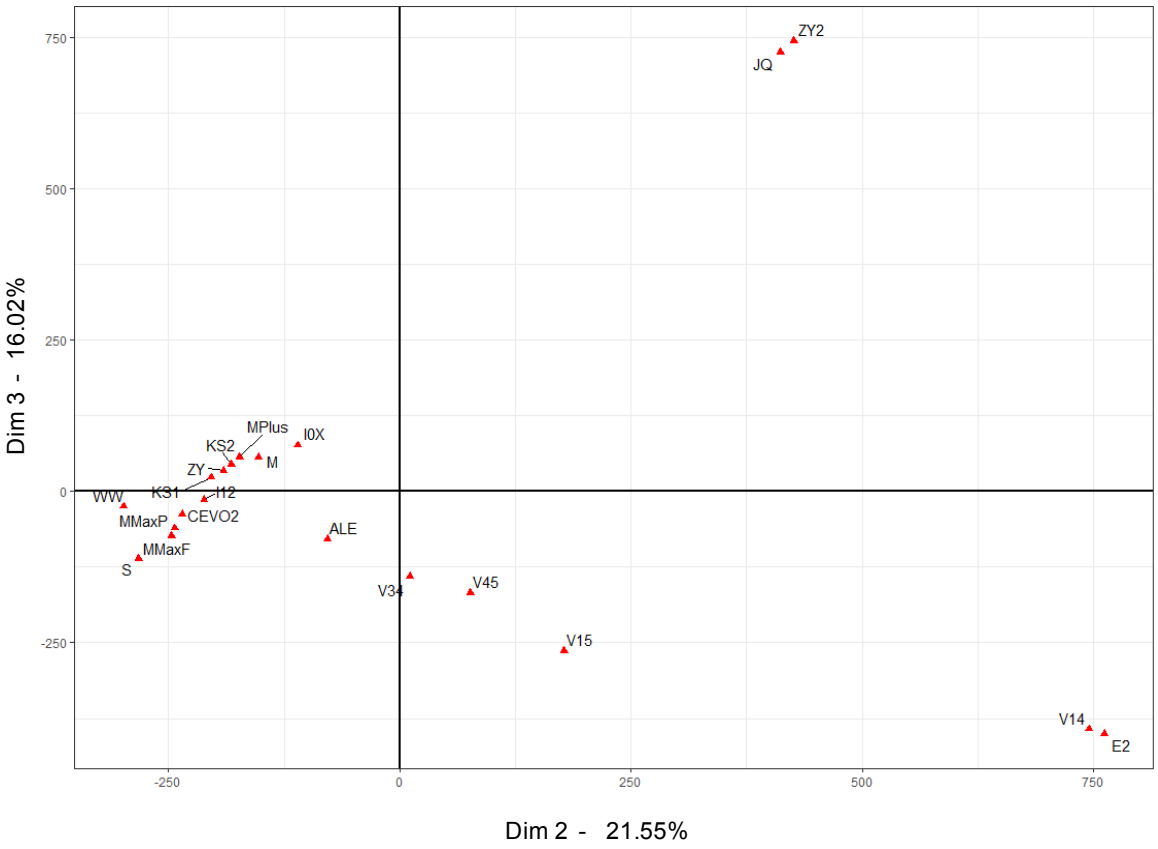



Figure 5-8 continued legend see page 82

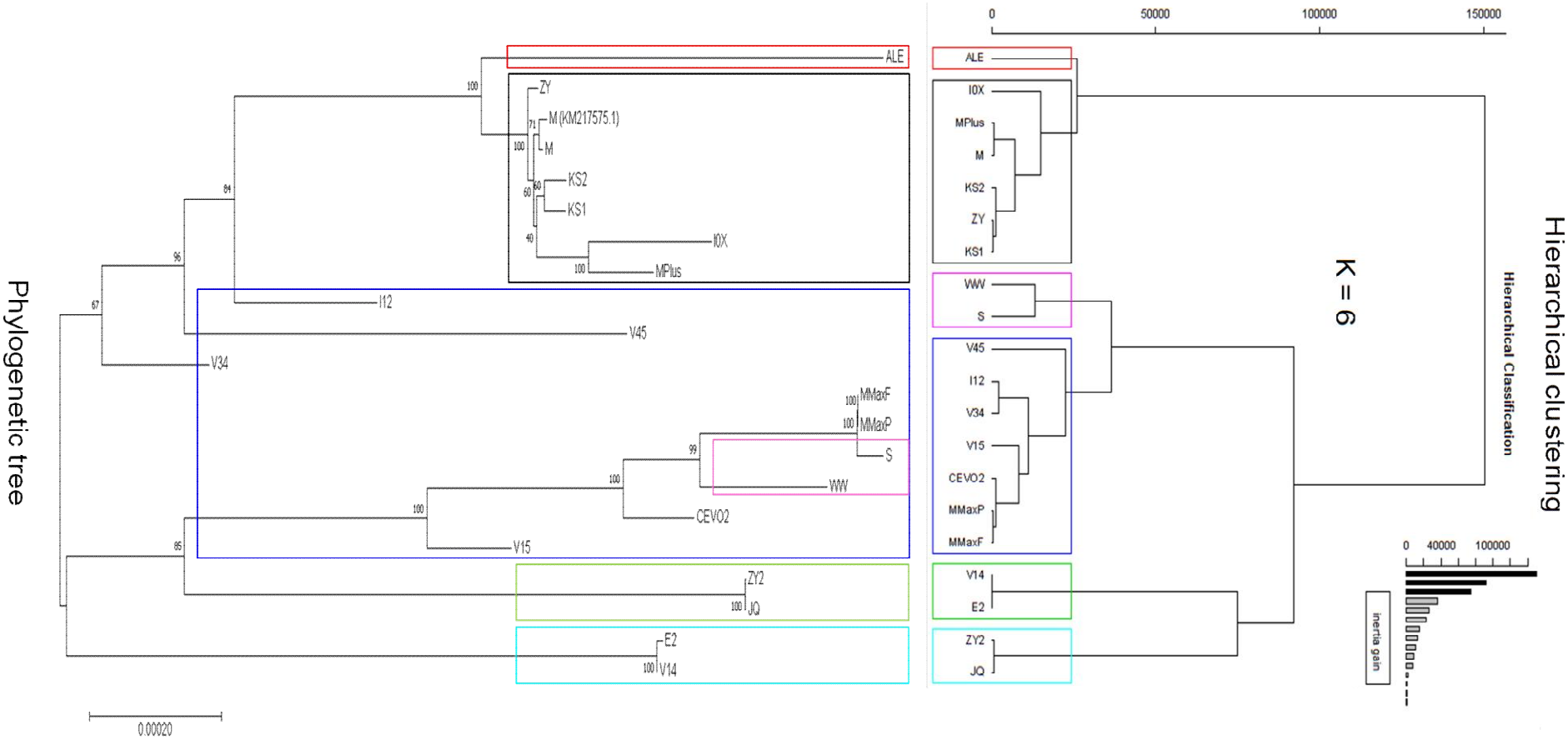



**Figure 5-9.** Phylogenetic reconstruction and hierarchical clustering. Phylogenetic tree was constructed on the basis of 20 CpGV consensus sequences using minimum evolution method with 500 of bootstrap replications, Tajima-Nei model, Gamma distributed for rates among sites, gaps/missing data treated by partial deletion in MEGA7 (Kumar et al., 2016). Hierarchical clustering on principal component (HCPC) containing six clusters was clustered using 753 SNP dataset and consolidated by K means clustering with the cluster centers as well.

## Discussion

NGS methods with high genome coverage provide new tools to study the population structure of virus isolates. As a consequence, more comprehensive genetic information therefore can be unveiled than in previous studies. The data sets of re-sequenced genomes of CpGV-M, -S, -E2 and -I12, with an average nucleotide coverage of 3320-4045 was up to 100- to 1,000-fold higher than previous data sets of these viruses, ranging from 3.9-fold to 243-fold coverage (Wennmann et al., 2017). This high sequencing coverage allows determining the location of SNPs as well as their frequency distribution. Previous SNP analyses of CpGV-M, -S, -E2, -I12, -I07, representing the genome groups A to E, rendered between 2 to 356 group or isolate specific SNPs (Wennmann et al., 2017). The extension of analyses to seven new Chinese isolates and application of the consensus-free method extended the pictures of CpGV diversity to seven groups (A to G) (Chapter 4). It also revealed, however, that a number of SNPs, previously considered as genome group specific, were either present in isolates classified to different groups or not present in closely related isolates belonging to the same phylogenetic lineage, suggesting that these SNPs were rather isolate specific (Wennmann et al., 2017; Chapter 4). By using NGS data sets of 20 CpGV isolates, different phylogenetic lineages of CpGV could be re-evaluated on the basis of the distribution of isolate and group specific SNPs (Table 5-1). In consequence, the re-examination and identification of group specific SNPs from 20 CpGV genomes is much more robust than in previous studies with five genomes, when each phylogenetic genome group was based on only one viral genome. In addition, SNP detection obtained directly from aligned reads from NGS data rather than consensus sequence alignments can generate a much more complete picture of nucleotide variations, since minor SNP frequencies can be noted rather than neglected. Therefore, it is suggested that SNP screening in baculovirus sequence analysis needs to be done before consensus sequence is generated.

Another resource of mutations in CpGV genomes are indels. The motif of the 12 bp repeat unit in *pe38* corresponding to type I resistance was screened in Chapter 4 and in this Chapter. When correlating the results of Chapter 2 with the distribution of 1×12 bp repeats in different CpGV isolates (Chapter 4), isolates with fraction of 1×12 bp repeats larger than 46% were able to overcome type I resistance (CpRR1), otherwise slow or no resistance-breaking activity is observed (Chapter 2; Gueli Alletti et al., 2017; Jehle et al., 2017; Sauer et al., 2017b). In addition, CpGV-M and -ALE are the only isolates with little or extremely low percentage of 1×12 bp repeats, both isolates were the least active isolates in CpRR1 (Chapter 2). The other isolates showed higher virulence towards CpRR1 supporting the *pe38* model of resistance-breaking isolates (Gebhardt et al., 2014). Also in other species of the *Baculoviridae* family indels are also widely present as infective factors, e.g. Plutella xylostella granulovirus (PlxyGV), Phthorimaea operculella granulovirus (PhopGV) (Jukes et al., 2016; Larem et al., 2019). *Egt* and *sf122* deletion were found in some of three geographic SfMNPV isolates, SfMNPV-3AP2 of which lacked *egt* gene was considered as a fast killing phenotype (Harrison et al., 2008; Simón et al., 2011).

Four highly homogenous isolates, namely CpGV-M, -S, -WW and MPlus, were identified by nucleotide frequency of the majority of SNPs of close to 100%. These isolates represent highly pure genotypes. Subpolar SNPs at a frequency of less than 5% were concentrated in the genome of CpGV-M in a single SNP island I/II. The majority of these SNP positions are also

present in viruses containing genome group E isolates, suggesting that the highly homogenous isolate CpGV-M contained a small sub-population with this genetic signature. MPlus is a laboratory selection from CpGV-M on resistant CM larvae (Daniel Zingg, pers. communication). As demonstrated by the sequencing data and SNP analysis, it is highly identical to CpGV-M, though a 95% portion of the reads covering the repeat region of *pe38* contained only a 1×12 bp repeat motif without the additional 2×12 bp repeat that is characteristic for CpGV-M, where 1×12 bp repeat was only found at a 5% ratio (Figure 5-6A). Lack of the 2×12 bp repeat was previously identified to be essential for breaking type I resistance (Gebhardt et al., 2014). Interestingly, 3 out of 2439 reads of *pe38* of CpGV-M contained the 1×12 bp motif, clearly demonstrating that the capacity of resistance-breaking is present in CpGV population, yet only at 0.1% frequency (Figure 5-6). The minor presence of a *pe38* with a 1×12 bp repeat in CpGV-M may explain how MPlus (major presence of 1×12 bp repeat) could be selected from CpGV-M in type I resistant larvae. On the other hand, only 10 SNP positions were shared between CpGV-M and MPlus, also excluding those from SNP island I/II, proposing that they were lost during the selection process, while nucleotide frequency at other SNP positions were enriched or new SNPs were selected, e.g. in the MPlus specific SNP hotspot covering orf17R (ORF22), PEP (ORF23) and *pe38* (ORF24). As most SNPs frequencies of MPlus were close to 100%, it can be concluded that the virus selection successfully resulted in a notably "pure" genotype which was even more homogenous than the original CpGV-M.

The SNP pattern of CpGV-WW was highly similar to CpGV-S. The most obvious differences of both viruses were the WW fingerprint regions (a) and (b), covering a total number of 30 SNPs present in CpGV-S but not in -WW. As these regions are apparently shared between CpGV-WW and -M, it is plausible to assume that they are the result of a recent recombination event between CpGV-WW and CpGV-M or its ancestors. Despite the overall similarity of CpGV-S and -WW, there were also significant differences related to the WW fingerprint regions (a and b), which are located in *p45*, *dnapol*, *desmoplakin*, *lef3*, *iap5*, *lef9*, *dna-ligase*, genes which are supposed to be involved in viral DNA replication and BV production. SNPs in these regions result in amino acid sequence changes, which might be the possible reason for the virulence differences between the two viruses in infection experiments of larvae with type I (CpRR1) and type II (CpR5M) resistance (Chapter 2). When the genotype attribution of the seven Chinese CpGV isolates (Chapter 4) is considered, a similar genetic composition of CpGV was determined, irrespectively, whether the genome group specific SNP positions from consensus sequences or from new consensus-independent SNP positions were applied for quantification. Thus, the previous approach also provided reliable results, although at lower level of accuracy.

Both CpGV-E2 and V14, a commercial production batch of -E2, were different propagations of the same virus with almost identical SNP distribution and frequencies (Figure 5-4) and a similar indel frequencies at the same positions (Table 5-5, Figure 5-6) The frequencies of their SNPs could not be grouped into one, two or three majority classes as it was observed for other pure genotype isolates or mixtures of them. Their SNP frequencies occurred in a highly uneven distribution suggesting that CpGV-E2 consists of unusually manifold genotypes, adding to the "wave"-like distribution patterns of SNPs as shown in Figure 5-4. Similar "wave"-patterns, though to a lower degree and at different SNP positions were also noted for CpGV-I12, CEVO2, and V45. There is no obvious reason which could explain such patterns. But it is striking that CpGV-E2 and V14 showed virtually the identical SNP frequency pattern, although they were independently propagated in different CM strains and laboratories,

starting from the same virus inoculums. This observation can only be explained by the existence of selection constraints in CM larvae and/or in CpGV-E2 resulting in the stabilization of the complex composition of this virus.

Interestingly, CpGV-E2 was shown to be one of the most virulent CpGV isolates, being infective for all types of CpGV resistance I to III (Gueli Alletti et al., 2017; Sauer et al., 2017b). It can be speculated that CpGV isolates with such an internal heterogeneity (= "heterosis") are the most potent ones, suggesting that stable and heritable "heterosis" is essential for their high virulence. It was previously noticed that genotypes of CpGV-E2 could not be plaque purified (Winstanley and Crook, 1993). This finding might be explained by its highly complex composition and potentially co-acting genotypes. Studying replication of CpGV-E2 as well as the population genetics of this isolate may provide further answers.

Similarly, MMaxP and MMaxF, derived from different production batches using the same virus inoculums, which were mixtures of two viruses (CpGV-M and -S). As the SNP distributions as well as the resulting quantification of founding virus types were each highly similar (Figure 5-4, Table 5-1), it can be concluded that the population structure of genotype mixtures appears to be highly stable during *in vivo* propagation when a stable composition of virus in a given host is reached. Only new genetic host background, e.g. resistant individuals, may change this stable composition. Inheritable composition of virus mixture was highly prevalent in Spodoptera frugiperda multiple nucleopolyhedrovirus (SfMNPV), ensuring highly genetic diversity and infective activity in its field geographic populations (Escribano et al., 1999; Simón et al., 2004; Barrera et al., 2011). These first findings that even complex compositions of virus mixtures can be stably propagated are an important issue for quality control of commercial baculovirus production, since it demonstrates the identity of such product compositions can be stably produced. NGS techniques as developed here and applied on production batches will allow easy and straight forward tools for quality control of isolates consisting of mixed genotypes.

It is apparent that wild-type CpGV isolates originating from natural CM populations are in their majority genotype mixtures. Field-obtained CpGV-E and NPP-R1 were comprised of at least two genotypes, one of which is most likely identical to CpGV-M (Crook et al., 1985; Berling et al., 2009a). *Artogeia brassicae* larvae infected either with Artogeia rapae granulovirus 1 (ArGV1) or other ArGV strains, revealed in most cases the presence of ArGV1 or recombinants between inoculums and ArGV1, suggesting ArGV1 is present as latent genotype in the host population (Smith and Crook, 1993). For wild-type ArGV3 and Lymantria dispar multiple nucleopolyhedrovirus (LdMNPV) it was shown by physical mapping that they were comprised of three and eight distinct genotypes, respectively (Smith and Crook, 1988; Harrison et al., 2016). Either the same or different inoculum doses of Helicoverpa armigera nucleopolyhedrovirus (HearNPV) could induce the genetic diversity in HearNPV populations (Baillie and Bouwer, 2013; Kitchin and Bouwer, 2018). Besides that, host and environmental stress were correlated with genetic diversity of HearNPV, which is presumably caused by activation of latent viruses (Moscardi, 1999). A similar observation, validated by NGS analyses, was recently noted for PhopGV, when infection of *Phthorimaea operculella* larvae with the isolate PhopGV-GR3 resulted in the activation of a latent virus PhopGV-R (Larem, 2019).

Genotype mixtures of wild-type viruses may benefit to the virulence of a given virus. An experimental insect population of SfMNPV consisted of a mixture of two genotypes, 78% of a complete genotype B and 22% of a deletion genotype C. In infection experiments, a mutualistic interaction between these genotypes with increased virulence compared to pure genotypes was noted (Lopez-Ferber et al., 2003; Simon et al., 2006). Five genotypes of Anticarsia gemmatalis multiple nucleopolyhedrovirus (AgMNPV) have been identified from field-collected samples whit a high natural diversity (Del-Angel et al., 2018). In the practice of *Anticarsia gemmatalis* control, field AgMNPV isolates were collected and used to produce commercially biological control agent (Moscardi, 1999; Cory and Franklin, 2012). The natural trait of baculovirus genetic diversity has been exploited to select more virulent genotypes for overcoming resistance. This strategy, collecting virus inoculum from infected insects in the field or from virus collections and further selecting them in resistant insect colonies was practiced to prevent emergence of host resistance against AgMNPV and to overcome CpGV resistance (Moscardi, 1999; Zingg et al., 2011).

The gene *pe38* is less conserved but characterized by a leucine zipper and a Zinc finger domain (Krappa and Knebel-Mörsdorf, 1991). It was identified to be the key gene for breaking type I resistance because the lack of 2×12 bp repeat units, present in CpGV-M but not in other resistance-breaking isolates, rendered isolates as resistance-breaking (Luque et al., 2001; Gebhardt et al., 2014). To further analyze the distribution of the 12 bp repeat units in the different isolates only those reads which harbored a unique border sequence of *pe38* and could therefore be nested to the genome sequences were considered for quantitative analyses. Statistical ratio of 1×12 bp unit was generated from genome assembly where all reads containing complete or partial of 1×12 bp is used for calculation. This could increase the absolute number of reads having 1×12 bp repeat, but of which part might contain multiple 1×12 bp repeat unit. As a result, the read counting method is more confident than genome mapping for indels identification. Except for CpGV-E2, -JQ, -I0X and V14, V34, V45, the ratio 1×12 bp repeat unit was somehow correlated with the ratio of genome group A (e.g.CpGV-M) in a given isolate, as the presence of multiples of the 12 bp repeat unit is the significant difference between genome group A and other genome groups (Chapter 2 and Chapter 4) (Gebhardt et al., 2014; Wennmann et al., 2017). From this point, it is proposed that independent indels can be used to sort genome groups, though not accurately and completely.

The principal component analysis (PCA) is a mathematical procedure to reduce a multiple dimension problem to a lower number of dimensions with highest influence on the data, aiming to summarize the overall divergence in individuals, as it was done in this study. Here, the first two and three components explained 60% to 80% of the data distribution, respectively. PCA is a standard procedure in analyzing RNA sequencing, where reliability of biological replicates as well as the effect of treatments can be visualized and explained (Xue et al., 2012; Treutlein et al., 2016; Xing et al., 2017). The analysis can be performed on the frequency of SNP position without any previous knowledge about isolates genetic composition and allows a reliable grouping. This stays in contrast with the common analysis based on consensus sequences that are drawn from majorities of every single base in the read assembly. In addition, the clustering of isolates, based on the same SNP frequency table, allowed an assignment of isolates into groups that were almost followed by the consensus based phylogeny approach, except that mixtures of isolates were not directly visible. In the HCPC an imaginary line can be drawn between homogenous genome group A (CpGV-M) and genome

group E (CpGV-WW and -S), between which isolates were arranged according to their corresponding mixtures of group A and E (Figure 5-7 B and C). HCPC cluster reflects genetic compositions better than consensus based phylogenetic reconstructions. Hence, the method applied in this paper might be an important tool for future analysis of field and selected baculovirus isolates to identify their hetero- or homogeneity.

In conclusion, analyses of Illumina Solexa NGS data of 20 CpGV isolates revealed high conservation of CpGV genomes with only 753 SNP sites. The analyses of frequency patterns revealed three classes of genetic composition: (i) highly homogeneous isolates, such as CpGV-M, -WW,-S and MPlus; (ii) mixtures of at least two homogenous viruses with different ratios, often containing either genotypes from genome group A (CpGV-M) or genome group E (CpGV-S or -WW); and (iii) isolates with heterogenous composition as demonstrated by highly uneven and “wave”-like frequency distribution of SNPs, as found in CpGV-E2 and V14 but also in CpGV-I12, CEVO2 and V45. Based on SNP location and frequency using HCPC method, CpGV isolates can be classified into six clusters, which clearly correspond to the genotype composition and phylogenetic lineage of the major components.

## Chapter VI: General discussion

CpGV is one of the most extensively used viral control agents. It has been first registered in Switzerland, Germany and some other countries in the late 1980s and is nowadays commercialized in nearly all countries where CM occurs and causes economic loss in pome and stone fruit as well as walnut production (reviewed by Huber, 1998; Lacey et al., 2008). Over more than one decade of extensive application of CpGV products, several CM field populations appeared to be less susceptible (resistant) to CpGV, first noted in Germany and soon after in France (Fritsch et al., 2005; Sauphanor et al., 2006), and then also in other European countries (Schmitt et al., 2013; Sauer, 2017). Until now >30 orchards with resistant CM populations have been identified in Germany, of which most were representative for type I resistance (Asser-Kaiser et al., 2007; Schmitt et al., 2013). The most likely reason of resistance occurrence was that all commercial products contained the Mexican isolate CpGV-M. The problem of CpGV resistance could be initially solved by using resistance-breaking CpGV isolates which were eventually registered (Berling et al., 2009b; Eberle et al., 2009; Zingg et al., 2011; Jehle et al., 2016).

Via discovering new CpGV isolates in the field and selection on resistant CM individuals, additional CpGV variants, e.g. CpGV-SA, -I12, -I07, -NPP-R1, -R5 have been identified and successfully produced (Graillot et al., 2014; Motsoeneng, 2016; Sauer, 2017). After several years of successful application of resistance-breaking CpGV products, however, additional resistant populations not following the original type I resistance was recorded (Jehle et al., 2017; Sauer et al., 2017a). Some of these populations were already present in 2008, even before resistant-breaking isolates were used. To date three types of CpGV resistance are known: type I resistance (Z-chromosomal, dominant), type II resistance (autosomal dominant) and type III resistance (both autosomal and Z-chromosomal and proposed to be a mixture resistance of type I and type II resistance).

For the management of resistance it has been reported that CpGV isolates belonging to genome groups B to E, represented by CpGV-E2, -S, -I07, -I12 but not viruses from group A (represented by CpGV-M) can break type I resistance, which has been evidenced in numerous discriminating bioassays (Eberle et al., 2008; Sauer, 2017). The presence of an early promoter motif located in *pe38* of CpGV and cDNA microarray data indicated that *pe38* is an early transcribed gene (Pietruska, 2018). *Pe38* plays an important role during CpGV infection in CpS larvae; bacCpGV *pe38*M (knockout of *pe38*) did not induce mortality and generate occlusion bodies as well as budded viruses (Gebhardt et al., 2014). Two types of 12 bp repeat units, 1×12 bp and 3×12 bp in *pe38* were identified (Eberle, 2010). The lack of 2×12 bp repeat insertion in *pe38* have been further identified as the key factor for overcoming type I resistance in CpRR1 (Gebhardt et al., 2014). If 2×12 bp repeat unit insertion was present, virus replication was blocked in early virus infection phase in type I resistant CM larvae, resulting in a loss of virus infectivity (Asser-Kaiser et al., 2011). Resistance testing of CpRR1 larvae with mixed CpGV isolates which *pe38* partially (CpGV-KS1, -ZY and -ZY2) or fully (CpGV-JQ) lacked the 2×12 bp repeat insertion showed their type I resistance-breaking ability (Chapter 2, 4). It is therefore postulated that the early gene *pe38* without the 2×12 bp repeat unit insertion is in favor of CpGV infection and propagation by transcriptionally

transactivating viral genes and thus results in a continuation of initiated infextion which can then become systemic. The above hypothesis is supported by functional studies of AcMNPV PE38, the homologue of PE38 of CpGV. AcMNPV PE38 can enhance IE1-induced apoptosis about twofold, but PE38 alone had no function in Sf-21 cells which proposed that PE38 is one of viral genes together with others to augment cellular apoptosis at baculovirus early infection (Prikhod'ko and Miller, 1999). Interestingly, CpGV-ALE caused low mortality on strain CpRR1, which was even significantly lower than the mortality induced by CpGV-M at 14 dpi (Chapter 2), but showed the same infectivity as other isolates against CpS, indicating that all seven Chinese CpGV isolates can be used to control susceptible CM populations in China. Comparison of mortality induced by the seven Chinese isolates of CpGV and CpGV-M, -S and -E2 showed a slightly delayed infection pattern observed in CpRR1 as noted before by Asser-Kaiser et al. (2011). Quantitative analysis of virus gene expression in CpRR1 infected with CpGV-M using microarray demonstrated temporal viral gene expression was reduced and delayed (Pietruska, 2018). Two functional domains of PE38, N-terminal RING-finger motif and C-terminal leucine zipper motif, existing in AcMNPV and CpGV, providing a plausible hypothesis that 2×12 bp mutation in *pe38* may interact with two functional domains resulting in alteration of PE38 structure and a delayed and suppressed infection process in CpRR1 larvae (Gebhardt et al., 2014). A mutant *pe38* knockout virus of AcMNPV showed a delayed and reduced BV production, no impact on the initiation of DNA synthesis compared to wild-type of AcMNPV (Milks et al., 2003).

The highly homogenous isolate CpGV-WW (genome group E) which lacked the 2×12 bp repeat unit in *pe38*, was able to overcome type II and type III resistance but not type I resistance (Chapter 2, 4 and 5). In other words, CpRR1 is resistant to CpGV-WW but the "*pe38* model" fails to explain the virulence pattern of CpGV-WW in the three resistant CM colonies. In this virus-host combination, additional resistance mechanisms and virus target gene(s) of resistance must be involved, not only *pe38*. Or PE38 functions as an intermediate protein or enhancer protein in successful development of CpGV in CpRR1.

Consider that type I and type II resistance were discovered in the same orchard (Jehle et al., 2017). There was somehow a heritable relation between these two resistances. Sauer et al. (2017a) have verified that type II resistance is not caused of resistant gene transfer from Z chromosome to autosome, but induced by other genetic mutation or inter-chromosomal rearrangement (Sauer et al., 2017a). The genomic and pathogenic differentiations between CpGV-S and -WW might answer the above question that the type II resistance-breaking factor(s) is located in CpGV-WW genome. V15, consisting of resistance-breaking genome group B and E isolates, can overcome type I to III resistance (Gueli Alletti et al., 2017). Two resistance-breaking markers, the 2×12 bp repeat unit in *pe38* (responsible for type I) and currently unknown unit(s) (responsible for breaking type II resistance), are located in genome group B CpGV. These results demonstrated that autosomal resistance in CM targets at least one more CpGV gene than Z-chromosomal resistance and may carry a more complicated resistance mechanism.

In the last two years, V14, highly similar to CpGV-E2, has been tested in organic orchards in German where type II resistance was recorded (BVL 2017, 2018). Recently, six newly discovered CpGV isolates from Argentina were grouped into the same branch as CpGV-E2, which however showed 57% difference in contrast to CpGV-E2 in phylogeny (Arneodo et al., 2015). As genome groups A to E are corresponding to their virulence, the Argentinean

CpGV isolates show potential market in resistance management. Although emergence of CpGV-M resistance in CM has been regarded as the result of applying the single genotype of CpGV-M for more than a decade, application of CrleGV-SA in South Africa against *Thaumatotibia leucotreta* for more than ten years has not resulted in any loss of efficacy (van der Merwe et al., 2017). Up to five generations of *T. leucotreta* occur during one year because there is diapause stage (Daiber, 1980), which implies that at least 75 applications have been done without any evidence of resistance selection . This story is quite different from CpGV resistance cases and proposes the question why the mutual adaption between CM and CpGV develops apparently faster than resistance evolution in *T. leucotreta* and CrleGV. The answer to this question will provide an additional clue to develop strategies for CM resistance management in field.

A decline of the resistance level of strain CpRR1, expressing type I resistance, was recorded in aforementioned resistance testing and bioassays. The subjects' CpRR1 strain had been selected in 2006 and was re-selected by single pair crossings in 2011 (Asser-Kaiser et al., 2007; Pietruska, 2018), since then it has been reared on CpGV-free diet. Newly selected CpRR1_F5 and CpRR1_F7 were established to figure out the reason for the decline of resistance in CpRR1 and to regain original resistance levels (Chapter 3). Though the backcrosses CpRR1_F5 and CpRR1_F7 with susceptible CpS demonstrated dominant inheritance the survival patterns did not fully favor single Z-chromosomal linkage anymore, as female larvae could develop into pupae in single-pair crossings and discriminating resistance tests, although they would have been expected to be fully susceptible according to a sex-linked model of resistance. Indeed CpRR1_F5 could regain a high resistance level but it was estimated to be 10- to 100-fold below the original value of CpRR1. Also some fitness costs of fecundity were noted, though no fitness costs were recorded before in CpR, the origin of CpRR1 (Undorf-Spahn et al., 2012). However, in consideration of the double inbreeding in CpRR1 in 2006 and 2011 to achieve high resistance, a low genetic diversity of CpRR1 needs to be taken into account, which may also have contributed to the reduced fecundity. Compared to other baculovirus-host systems, resistance level of *Adoxophyes honmai* incurred by AdhoNPV was extremely stable, no obvious resistance reduction has been measured after 128 generations of rearing without any virus pressure (Nakai et al., 2017). It was proposed that there was no fitness cost in resistance strain of *A. honmai*. Also resistance of *Trichoplusia ni* to AcMNPV-AalT did not reveal any no fitness costs in the selection process (Milks and Theilmann, 2000). For BmNPV resistance in *Bombyx mori* a pair of autosomal dominant genes have been demonstrated by backcrossing and self-crossed progeny from second generation in resistance tests and resistance marker screening, whether these two genes are same or not that need further study (Feng et al., 2012). A further study posed the hypothesis that genes for sugar transporter located on chromosome 21 might participate and that this protein serves as receptor in favor of virus entry (Govindaraj et al., 2016). The present day Z chromosome of CM and other Olethreutinae is a fusion product of an ancestral Z chromosome and homologue of chromosome 15 of *Bombyx mori* (Nguyen et al., 2013), hinting that newly rearranged CM Z chromosome harboring one copy (haploid) of chromosome 15 somehow inherit the function from Z chromosome and the autosome. The obtained results in this thesis support that the long-term rearing type I resistance under virus-free condition and multiple inbreedings of CpRR1 were two potential reasons which had induced the decline of resistance over the years.

NGS sequencing provides the opportunity to determine the genetic composition of diverse geographic and manufacturer produced isolates. Genomic analyses of twelve naturally

occurring CpGV isolates in this thesis pinpointed that only three isolates, namely CpGV-M, -S, -WW were highly homogenous, termed as “pure” ones. The other field CpGV isolates were mixtures or compositions of highly heterogeneous genomes. This finding echoes former studies that naturally occurring baculoviruses usually contain mixed genomic components rather than a single genotype. For example, two Mexican isolates of SeMNPV, SeSLP6 and SeSIN6, collected from long-distance geographic areas, show a major portion of ODVs which DNA generates additional bands in the variable region *V01,* as illuminated by PCR and RFLP analyses (Zamora-Avilés et al., 2017). These isolates have been considered to be further developed into potential pesticides to reduce *S. exigua* in the field. It is assumed that high viral diversity offers an advantage of high virulence against host larvae. In addition of different origins of HearNPV, $LC_{50}$ values of HearSP2 is 2.8- and 2.6-fold higher than HearPT2 and HearSP7, respectively (Figueiredo et al., 2009). HearSP4 and HearPT2 can kill larvae significantly faster than HearSP8. Although polymorphisms of Pakistan SpltNPV isolates are originated from a single biotype; the speed of kill of three genotypes was different from each other, which implies that geographic distance and variable agricultural systems may contribute to viral pathogenic differences (Ali et al., 2018). More than one Panolis flammea nucleopolyhedrovirus (PaflNPV) genotype, co-infected in *Panolis flammea*, were isolated even from a single larva and there was no dominant genotype in the virus population, demonstrating that genetic diversity in baculoviral populations is maintained during *in vivo* infection (Cory et al., 2005). Studying different origins of LdMNPV (isolated from three continents) and CrleGV (isolated from South Africa) did not only expand the genetic diversity of baculoviruses but were suggested to provide a virus pool available for pest control and resistance management, before resistance emerges in these virus host systems (Opoku-Debrah et al., 2013; Harrison et al., 2014). Genotype heterogeneity generally exists in baculovirus populations, demonstrating that heterogeneous isolates can better adapt in the co-evolution between baculovirus and host or under variable environmental conditions. Naturally collected 30 Mexican AgMNPVs were comprised of genotypes 1-5, a mixture which appeared to be more pathogenic than other genotype composition with variable ratios (Del-Angel et al., 2018). Genotype 3 commonly exists in AgMNPV populations, similar to naturally occurring Chinese and other CpGV isolates where CpGV-M (genome group A) and/or CpGV-WW (genome group E) were the most prevalent genotypes (Chapter 4 and Chapter 5).

Genome analyses of baculoviruses revealed that two types of genomic variation, SNPs and indels, are not evenly distributed along the genome but that certain parts carry high ratios of mutations, marked as “hotspot” of hypervariable regions, such as *hr* regions and *bro* genes (Cory et al., 2005; Erlandson, 2009). This hotspot phenomenon was identified in CpGV-ALE and -I0X, though in other isolates, leaving it open whether such “hotspots” are a general genomic feature of all baculoviruses (Chateigner et al., 2015). SNPs and indels in 20 CpGVs were distributed with different amounts and frequencies. Four isolates CpGV-M, -S, -WW and MadexPlus possess less variation than others, proposing that they are highly homogenous. Vice versa other isolates are mixtures or highly heterogeneous ones.

Genome group specific SNPs obtained from CpGV consensus sequence alignments were initially used to determine the composition of different isolates and to allocate the viruses to different genome groups (Gueli Alletti et al., 2017; Wennmann et al., 2017). However, these SNPs from consensus sequence were not able to reflect the real genetic information, as the consensus sequence was processed and generated by choosing the major frequency nucleotide to compose a genome (Day and McMorris, 1992). In this thesis, a *de novo* SNP detection

method was applied to discover the SNP from original virus population that provide initial genetic information of a given isolate. Based on the isolate/genome group specific SNPs, genetic composition of most CpGVs were distinguishable except highly heterogeneous isolates, such as CpGV-E2 and V14.

When a given isolate consists of a mixture of two or more homogenous genotypes or of highly heterogenous genotypes, it is not possible to assess the population structure by phylogenetic analysis tools since no proper consensus genome is generated (Day and McMorris, 1992). To solve this problem, Hierarchical Clustering on Principal Components (HCPC) based on identified SNP frequency and position was conducted, which allowed distinguishing the (dis)similarity between isolates. Compared to the results from previous resistance testing and this thesis, it is found that CpGV isolates with similar infectivity were clustered together (Chapter 2 and Chapter 5).

In summary, virulence to different CM strains, genetic composition and diversity of CpGV isolates were studied in this thesis, providing a pathway to understand resistance-breaking mechanism from CpGV polymorphisms. The established methods of identification and quantification of genetic diversity can be extended to other baculovirus agents, as major natural isolates or commercial products of baculoviruses may consist of more than one genotype. Identity of SNP frequency and position in given baculoviruses can be used to assess the genome stability and the evolutionary direction on species level. Thus, this kind of analyses is not only useful for field isolates but is also important for quality control of baculovirus productions and preparations. Connecting the isolates´ genetic composition, as expressed in position in the HCPC factor map and their virulence parameters will allow establishing a blueprint for current and future CM resistance management.

## Reference


Abaajeh, A.R., Nchu, F., 2015. Isolation and pathogenicity of some South African entomopathogenic fungi (Ascomycota) against eggs and larvae of *Cydia pomonella* (Lepidoptera: Tortricidae). Biocontrol Sci Technol 25, 828-842.

Abbott, W.S., 1925. A method of computing the effectiveness of an insecticide. J Econ Entomol 18, 265-267.

Ali, G., Abma-Henkens, M.H.C., van der Werf, W., Hemerik, L., Vlak, J.M., 2018. Genotype assembly, biological activity and adaptation of spatially separated isolates of *Spodoptera litura* nucleopolyhedrovirus. J Invertebr Pathol 153, 20-29.

Andermatt, M., Mani, E., Wildbolz, T., Lüthy, P., 1988. Susceptibility of *Cydia pomonella* to *Bacillus thuringiensis* under laboratory and field conditions. Entomologia Experimentalis et Applicata 49, 291-295.

Arends, H.M., Jehle, J.A., 2002. Homologous recombination between the inverted terminal repeats of defective transposon TCp3.2 causes an inversion in the genome of *Cydia pomonella* granulovirus. J Gen Virol 83, 1573-1578.

Arneodo, J.D., De Anna, J., Salvador, R., Farinon, M., Quintana, G., Sciocco-Cap, A., 2015. Prospection and molecular analysis of CpGV isolates infecting *Cydia pomonella* at different geographical locations in Argentina. Ann Appl Biol 166, 67-74.

Arrizubieta, M., Simón, O., Williams, T., Caballero, P., 2015. Genomic sequences of five *Helicoverpa armigera* nucleopolyhedrovirus genotypes from Spain that differ in their insecticidal properties. Genome Announcements 3, e00548-00515.

Arthurs, S.P., Lacey, L.A., 2004. Field evaluation of commercial formulations of the codling moth granulovirus: persistence of activity and success of seasonal applications against natural infestations of codling moth in Pacific Northwest apple orchards. Biol Control 31, 388-397.

Arthurs, S.P., Lacey, L.A., Fritts, R., Jr., 2005. Optimizing use of codling moth granulovirus: effects of application rate and spraying frequency on control of codling moth larvae in Pacific Northwest apple orchards. J Econ Entomol 98, 1459-1468.

Asser-Kaiser, S., Fritsch, E., Undorf-Spahn, K., Kienzle, J., Eberle, K.E., Gund, N.A., Reineke, A., Zebitz, C.P., Heckel, D.G., Huber, J., Jehle, J.A., 2007. Rapid emergence of baculovirus resistance in codling moth due to dominant, sex-linked inheritance. Science 317, 1916-1918.

Asser-Kaiser, S., Heckel, D.G., Jehle, J.A., 2010. Sex linkage of CpGV resistance in a heterogeneous field strain of the codling moth *Cydia pomonella* (L.). J Invertebr Pathol 103, 59-64.

Asser-Kaiser, S., Radtke, P., El-Salamouny, S., Winstanley, D., Jehle, J.A., 2011. Baculovirus resistance in codling moth (*Cydia pomonella* L.) caused by early block of virus replication. Virology 410, 360-367.

Ayres, M.D., Howard, S.C., Kuzio, J., Lopez-Ferber, M., Possee, R.D., 1994. The complete DNA sequence of *Autographa californica* nuclear polyhedrosis virus. Virology 202, 586-605.

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.

Ballard, J., Ellis, D.J., Payne, C.C., 2000. Uptake of granulovirus from the surface of apples and leaves by first instar larvae of the codling moth *Cydia pomonella* L. (Lepidoptera: Olethreutidae). Biocontrol Sci Technol 10, 617-625.

Barnes, B.N., Hofmeyr, J.H., Groenewald, S., Conlong, D.E., Wohlfarter, M., 2015. The Sterile Insect Technique in Agricultural Crops in South Africa: A Metamorphosis …. but Will it Fly? African Entomology 23, 1-18.

Barnes, M.M., 1991. Codling moth occurrence, host race formation and damage, in: Van der Guest LPS, E.H. (Ed.), Tortricid pests: their biology, natural enemies and control. Elsevier, Amsterdam, pp. 313-328.

Barrera, G., Simón, O., Villamizar, L., Williams, T., Caballero, P., 2011. *Spodoptera frugiperda* multiple nucleopolyhedrovirus as a potential biological insecticide: Genetic and phenotypic comparison of field isolates from Colombia. Biol Control 58, 113-120.

Barrett, J.W., Brownwright, A.J., Primavera, M.J., Retnakaran, A., Palli, S.R., 1998. Concomitant primary infection of the midgut epithelial cells and the hemocytes of *Trichoplusia ni* by *Autographa californica* nucleopolyhedrovirus. Tissue and Cell 30, 602-616.

Belyavskyi, M., Braunagel, S.C., Summers, M.D., 1998. The structural protein ODV-EC27 of *Autographa californica* nucleopolyhedrovirus is a multifunctional viral cyclin. Proc Natl Acad Sci U S A 95, 11205-11210.

Berling, M., Blachere-Lopez, C., Soubabere, O., Lery, X., Bonhomme, A., Sauphanor, B., Lopez-Ferber, M., 2009a. *Cydia pomonella granulovirus* genotypes overcome virus resistance in the codling moth and improve virus efficiency by selection against resistant hosts. Appl Environ Microbiol 75, 925-930.

Berling, M., Rey, J.B., Ondet, S.J., Tallot, Y., Soubabère, O., Bonhomme, A., Sauphanor, B., Lopez-Ferber, M., 2009b. Field trials of CpGV virus isolates overcoming resistance to CpGV-M. Virol Sin 24, 470-477.

Berling, M., Sauphanor, B., Bonhomme, A., Siegwart, M., Lopez-Ferber, M., 2013. A single sex-linked dominant gene does not fully explain the codling moth's resistance to granulovirus. Pest Manag Sci 69, 1261-1266.

Bideshi, D.K., Hice, R.H., Ge, B., Federici, B.A., 1998. Molecular characterization and expression of the *Trichoplusia ni* granulovirus *helicase* gene. J Gen Virol 79 ( Pt 5), 1309-1319.

Blissard, G.W., 1996. Baculovirus-insect cell interactions. Cytotechnology 20, 73-93.

Boncheva, R., Dukiandjiev, S., Minkov, I., de Maagd, R.A., Naimov, S., 2006. Activity of *Bacillus thuringiensis* δ-endotoxins against codling moth (*Cydia pomonella* L.) larvae. J Invertebr Pathol 92, 96-99.

Braunagel, S.C., Elton, D.M., Ma, H., Summers, M.D., 1996a. Identification and analysis of an *Autographa californica* nuclear polyhedrosis virus structural protein of the occlusion-derived virus envelope: ODV-E56. Virology 217, 97-110.

Braunagel, S.C., Guidry, P.A., Rosas-Acosta, G., Engelking, L., Summers, M.D., 2001. Identification of BV/ODV-C42, an *Autographa californica* nucleopolyhedrovirus orf101-encoded structural protein detected in infected-cell complexes with ODV-EC27 and p78/83. J Virol 75, 12331-12338.

Braunagel, S.C., He, H., Ramamurthy, P., Summers, M.D., 1996b. Transcription, translation, and cellular localization of three *Autographa californica* nuclear polyhedrosis virus structural proteins: ODV-E18, ODV-E35, and ODV-EC27. Virology 222, 100-114.

Braunagel, S.C., Russell, W.K., Rosas-Acosta, G., Russell, D.H., Summers, M.D., 2003. Determination of the protein composition of the occlusion-derived virus of *Autographa californica* nucleopolyhedrovirus. Proc Natl Acad Sci U S A 100, 9797-9802.

Braunagel, S.C., Summers, M.D., 2007. Molecular biology of the baculovirus occlusion-derived virus envelope. Curr Drug Targets 8, 1084-1095.

Brito, A.F., Melo, F.L., Ardisson-Araújo, D.M.P., Sihler, W., Souza, M.L., Ribeiro, B.M., 2018. Genome-wide diversity in temporal and regional populations of the betabaculovirus *Erinnyis ello granulovirus* (ErelGV). BMC Genet 19, 698.

Brown, M., Faulkner, P., 1978. Plaque assay of nuclear polyhedrosis viruses in cell culture. Appl Environ Microbiol 36, 31-35.

Carstens, E.B., Lu, A.L., Chan, H.L., 1993. Sequence, transcriptional mapping, and overexpression of *p47*, a baculovirus gene regulating late gene expression. J Virol 67, 2513-2520.

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

Chen, H., Chen, K., Yao, Q., Guo, Z., Wang, L., 2007. Characterization of a late gene, ORF67 from *Bombyx mori* nucleopolyhedrovirus. FEBS letters 581, 5836-5842.

Chen, L., Hu, X., Xiang, X., Yu, S., Yang, R., Wu, X., 2012. *Autographa californica* multiple nucleopolyhedrovirus *odv-e25* (Ac94) is required for budded virus infectivity and occlusion-derived virus formation. Arch Virol 157, 617-625.

Cooper, P.D., 1962. The plaque assay of animal viruses, in: Smith, K.M., Lauffer, M.A. (Eds.), Advances in virus research. Academic Press, pp. 319-378.

Cornman, R.S., Boncristiani, H., Dainat, B., Chen, Y., vanEngelsdorp, D., Weaver, D., Evans, J.D., 2013. Population-genomic variation within RNA viruses of the Western honey bee, Apis mellifera, inferred from deep sequencing. BMC Genet 14, 154.

Cory, J.S., Franklin, M.T., 2012. Evolution and the microbial control of insects. Evolutionary Applications 5, 455-469.

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.

Crook, N.E., Spencer, R.A., Payne, C.C., Leisy, D.J., 1985. Variation in *Cydia pomonella* granulosis virus isolaes and physical maps of the DNA from three variants. J Gen Virol 66, 2423-2430.

Daiber, C.C., 1980. A study of the biology of the false codling moth *Cryptophlebia leucotreta* (Meyr.): the adult and generations during the year. Phytophylactica 12, 187-194.

Day, W.H.E., McMorris, F.R., 1992. Consensus sequences based on plurality rule. Bull Math Biol 54, 1057-1068.

Del-Angel, C., Lasa, R., Rodríguez-del-Bosque, L.A., Mercado, G., Beperet, I., Caballero, P., Williams, T., 2018. *Anticarsia gemmatalis* nucleopolyhedrovirus from soybean crops in Tamaulipas, Mexico: diversity and insecticidal characteristics of individual variants and their co-occluded mixtures. Florida Entomologist 101, 404-410.

Dow, J.A., 1992. pH gradients in Lepodopteran midgut. J. exp. Biol. 172, 355-375.

Durantel, D., Croizier, L., Ayres, M.D., Croizier, G., Possee, R.D., Lopez-Ferber, M., 1998. The *pnk/pnl* gene (ORF 86) of *Autographa californica* nucleopolyhedrovirus is a non-essential, immediate early gene. J Gen Virol 79, 629-637.

Eberle, K.E., 2010. Novel isolates of Cydia pomonella granulovirus (CpGV): deciphering the molecular mechanism for overcoming CpGV resistance in codling moth (*Cydia pomonella*), Johannes Gutenberg-Universität Mainz, Mainz, p.171.

Eberle, K.E., Asser-Kaiser, S., Sayed, S.M., Nguyen, H.T., Jehle, J.A., 2008. Overcoming the resistance of codling moth against conventional *Cydia pomonella* granulovirus (CpGV-M) by a new isolate CpGV-I12. J Invertebr Pathol 98, 293-298.

Eberle, K.E., Jehle, J.A., 2006. Field resistance of codling moth against *Cydia pomonella* granulovirus (CpGV) is autosomal and incompletely dominant inherited. J Invertebr Pathol 93, 201-206.

Eberle, K.E., Jehle, J.A., Huber, J., 2012. Microbial control of crop pests using insect viruses, in: Abrol, D.P., Shankar, U. (Eds.), Integrated pest management: principles and practice. CABI, Croydon, pp. 281-298.

Eberle, K.E., Sayed, S., Rezapanah, M., Shojai-Estabragh, S., Jehle, J.A., 2009. Diversity and evolution of the *Cydia pomonella* granulovirus. J Gen Virol 90, 662-671.

Ehler, L.E., 2006. Integrated pest management (IPM): definition, historical development and implementation, and the other IPM. Pest Manag Sci 62, 787-789.

Eilenberg, J., Hajek, A., Lomer, C., 2001. Suggestions for unifying the terminology in biological control. Biocontrol 46, 387-400.

Elmenofy, W.H., Jehle, J.A., 2015. Possible functional co-operation of palindromes *hr3* and *hr4* in the genome of Cydia pomonella granulovirus affects viral replication capacity. J Gen Virol 96, 2888-2897.

Engelhard, E.K., Kam-Morgan, L.N., Washburn, J.O., Volkman, L.E., 1994. The insect tracheal system: a conduit for the systemic spread of *Autographa californica* M nuclear polyhedrosis virus. Proc Natl Acad Sci U S A 91, 3224-3227.

Erlandson, M.A., 2009. Genetic variation in field populations of baculoviruses: Mechanisms for generating variation and its potential role in baculovirus epizootiology. Virol Sin 24, 458-469.

Escribano, A., Williams, T., Goulson, D., Cave, R.D., Chapman, J.W., Caballero, P., 1999. Selection of a nucleopolyhedrovirus for control of *Spodoptera frugiperda* (Lepidoptera: Noctuidae): structural, genetic, and biological comparison of four isolates from the Americas. J Econ Entomol 92, 1079-1085.

Evans, J.T., Leisy, D.J., Rohrmann, G.F., 1997. Characterization of the interaction between the baculovirus replication factors LEF-1 and LEF-2. J Virol 71, 3114-3119.

Fang, Z., Shao, J., Weng, Q., 2016. *De novo* transcriptome analysis of *Spodoptera exigua* multiple nucleopolyhedrovirus (SeMNPV) genes in latently infected Se301 cells. Virol Sin 31, 425-436.

Feng, F., Fu, J., Hu, P., Zhang, X., Yao, Q., Chen, K., 2012. Genetic analysis of baculovirus resistance in lepidopteran model insect *Bombyx mori* L. African Journal of Biotechnology 11, 14417-14421.

Ferrelli, M.L., Salvador, R., Biedma, M.E., Berretta, M.F., Haase, S., Sciocco-Cap, A., Ghiringhelli, P.D., Romanowski, V., 2012. Genome of *Epinotia aporema* granulovirus (EpapGV), a polyorganotropic fast killing betabaculovirus with a novel thymidylate kinase gene. BMC Genet 13, 548.

Figueiredo, E., Muñoz, D., Murillo, R., Mexia, A., Caballero, P., 2009. Diversity of Iberian nucleopolyhedrovirus wild-type isolates infecting *Helicoverpa armigera* (Lepidoptera: Noctuidae). Biol Control 50, 43-49.

Fritsch, E., Undorf-Spahn, K., Kienzle, J., Zebitz, C.P.W., Huber, J., 2005. Apfelwickler-Granulovirus: Erste Hinweise auf Unterschiede in der Empfindlichkeit lokaler Apfelwickler-Populationen [Codling moth granulovirus: First indication of variations in the susceptibility of local codling moth populations]. Nachrbl Deut Pflanzenschutzd 57, 29-34.

Fuxa, 1993. Insect resistance of viruses, in: Beckage, N.E., Thompson, S.A., Federici, B.A. (Eds.), Parasites and Pathogens of Insects, Pathogens. Academic Press, Baton Rouge, Louisiana pp. 197-209.

Fuxa, J.R., Richter, A.R., 1998. Repeated reversion of resistance to nucleopolyhedrovirus by *Anticarsia gemmatalis*. J Invertebr Pathol 71, 159-164.

Gan, E., Li, X., Yu, H., Wu, Z., Li, X., Zhang, Y., Wang, D., 2011. Virulence determination for three native *Cydia pomonella* granulovirus strains and their control effect in field. Journal of Northwest A&F University (Nat. Sci. Ed) 39, 119-122.

Garavaglia, M.J., Miele, S.A., 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.

Gebhardt, M.M., Eberle, K.E., Radtke, P., Jehle, J.A., 2014. Baculovirus resistance in codling moth is virus isolate-dependent and the consequence of a mutation in viral gene *pe38*. Proc Natl Acad Sci U S A 111, 15711-15716.

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

Govindaraj, L., Gupta, T., Esvaran, V.G., Awasthi, A.K., Ponnuvel, K.M., 2016. Genome-wide identification, characterization of sugar transporter genes in the silkworm *Bombyx mori* and role in *Bombyx mori* nucleopolyhedrovirus (BmNPV) infection. Gene 579, 162-171.

Graillot, B., Bayle, S., Blachere-Lopez, C., Besse, S., Siegwart, M., Lopez-Ferber, M., 2016. Biological characteristics of experimental genotype mixtures of Cydia pomonella granulovirus (CpGV): ability to control susceptible and resistant pest populations. Viruses 8, 147.

Graillot, B., Berling, M., Blachere-Lopez, C., Siegwart, M., Besse, S., Lopez-Ferber, M., 2014. Progressive adaptation of a CpGV isolate to codling moth populations resistant to CpGV-M. Viruses 6, 5135-5144.

Gross, C.H., Shuman, S., 1998. RNA 5'-triphosphatase, nucleoside triphosphatase, and guanylyltransferase activities of baculovirus LEF-4 protein. J Virol 72, 10020-10028.

Guarino, L.A., Dong, W., Jin, J., 2002. In vitro activity of the baculovirus late expression factor LEF-5. J Virol 76, 12663-12675.

Guarino, L.A., Summers, M.D., 1986. Functional mapping of a trans-activating gene required for expression of a baculovirus delayed-early gene. J Virol 57, 563-571.

Guarino, L.A., Xu, B., Jin, J., Dong, W., 1998a. A virus-encoded RNA polymerase purified from baculovirus-infected cells. J Virol 72, 7985-7991.

Guarino, L.A., Xu, B., Jin, J., Dong, W., 1998b. A virus-encoded RNA polymerase purified from baculovirus-infected cells. J Virol 72, 7985-7991.

Gueli Alletti, G., Carstens, E.B., Weihrauch, B., Jehle, J.A., 2018. Agrotis segetum nucleopolyhedrovirus but not Agrotis segetum granulovirus replicate in AiE1611T cell line of *Agrotis ipsilon*. J Invertebr Pathol 151, 7-13.

Gueli Alletti, G., Sauer, A.J., Weihrauch, B., Fritsch, E., Undorf-Spahn, K., Wennmann, J.T., Jehle, J.A., 2017. Using next generation sequencing to identify and quantify the genetic composition of resistance-breaking commercial isolates of Cydia pomonella granulovirus. Viruses 9, 250.

Gürlek, S., Sevim, A., Sezgin, F.M., Sevim, E., 2018. Isolation and characterization of *Beauveria* and *Metarhizium* spp. from walnut fields and their pathogenicity against the codling moth, *Cydia pomonella* (L.) (Lepidoptera: Tortricidae). Egyptian Journal of Biological Pest Control 28, 50.

Harrison, R., Rowley, D., Mowery, J., Bauchan, G., Burand, J., 2017. The *Operophtera brumata* nucleopolyhedrovirus (OpbuNPV) represents an early, divergent lineage within Genus *Alphabaculovirus*. Viruses 9, 307.

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.

Harrison, R.L., Puttler, B., Popham, H.J.R., 2008. Genomic sequence analysis of a fast-killing isolate of *Spodoptera frugiperda* multiple nucleopolyhedrovirus. J Gen Virol 89, 775-790.

Harrison, R.L., Rowley, D.L., Keena, M.A., 2016. Geographic isolates of *Lymantria dispar* multiple nucleopolyhedrovirus: genome sequence analysis and pathogenicity against European and Asian gypsy moth strains. J Invertebr Pathol 137, 10-22.

Harrison, R.L., Sparks, W.O., Bonning, B.C., 2010. *Autographa californica* multiple nucleopolyhedrovirus ODV-E56 envelope protein is required for oral infectivity and can be substituted functionally by *Rachiplusia ou* multiple nucleopolyhedrovirus ODV-E56. J Gen Virol 91, 1173-1182.

Harvey, J.P., Volkman, L.E., 1983. Biochemical and biological variation of Cydia pomonella (codling Moth) granulosis virus. Virology 124, 21-34.

Herniou, E.A., Arif, B.M., Bonning, B.C., Theilmann, D.A., Blissard, G.W., Becnel, J.J., Jehle, J.A., Harrison, R., 2011. Family-Baculoviridae, in: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier, San Diego, pp. 163-174.

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.

Hess, R.T., Falcon, L.A., 1987. Temporal events in the invasion of the codling moth, *Cydia pomonella*, by a granulosis virus: An electron microscope study. J Invertebr Pathol 50, 85-105.

Hilton, S., Winstanley, D., 2007. Identification and functional analysis of the origins of DNA replication in the *Cydia pomonella* granulovirus genome. J Gen Virol 88, 1496-1504.

Hoen, A.G., Gardner, S.N., Moore, J.H., 2013. Identification of SNPs associated with variola virus virulence. BioData Min 6, 3-3.

Horner, R., Walker, J., Rogers, D., Lo, P., Suckling, D., 2016. Use of the sterile insect technique in New Zealand: benefits and constraints. NZ Plant Prot 68, 296-304.

Huber, J., 1998. Western Europe, in: Hunter-Fujita F.R., Ph.F. Entwistle, Evans H.F., N.E., C. (Eds.), Insect Viruses and Pest Management. Wiley & Sons pp. 201-215.

Huber, J., Dickler, E., 1977. Codling moth granulosis virus: its efficiency in the field in comparison with organophosphorus insecticides. J Econ Entomol 70, 557-561.

Ishimwe, E., Hodgson, J.J., Passarelli, A.L., 2015. Expression of the Cydia pomonella granulovirus matrix metalloprotease enhances Autographa californica multiple nucleopolyhedrovirus virulence and can partially substitute for viral cathepsin. Virology 481, 166-178.

Ivaldi-Sender, C., 1974. Techniques simples pour un elevage permanent de la Tordeuse orientale, *Grapholita Molesta* (Lepidoptera, Tortricidae) sur milieu artificiel. Ann Zool Ecol Anim 6, 337-343.

Iwata, K., Haas-Stapleton, E., Kunimi, Y., Inoue, M.N., Nakai, M., 2017. Midgut-based resistance to oral infection by a nucleopolyhedrovirus in the laboratory-selected strain of the smaller tea tortrix, *Adoxophyes honmai* (Lepidoptera: Tortricidae). J Gen Virol 98, 296-304.

Iyer, S., Casey, E., Bouzek, H., Kim, M., Deng, W., Larsen, B.B., Zhao, H., Bumgarner, R.E., Rolland, M., Mullins, J.I., 2015. Comparison of major and minor viral SNPs identified through single template sequencing and Pyrosequencing in acute HIV-1 infection. PLoS One 10, e0135903.

Jaques, R.P., Hardman, J.M., Laing, J.E., Smith, R.F., Bent, E., 1994. Orchard trials in Canada on control of *Cydia pomonella* (Lep: Tortricidae) by granulosis virus. Entomophaga 39, 281-292.

Javed, M.A., Biswas, S., Willis, L.G., Harris, S., Pritchard, C., van Oers, M.M., Donly, B.C., Erlandson, M.A., Hegedus, D.D., Theilmann, D.A., 2017. Autographa californica multiple nucleopolyhedrovirus AC83 is a *per os* infectivity factor (PIF) protein required for occlusion-derived virus (ODV) and budded virus nucleocapsid assembly as well as assembly of the PIF complex in ODV envelopes. J Virol 91, e02115-02116.

Jehle, J.A., Hilton, S., 2011. Betabaculovirus, in: Tidona, C., Darai, G. (Eds.), The Springer Index of Viruses. Springer, New York, NY, pp. 119-127.

Jehle, J.A., Sauer, A., Gueli Alletti, G., Undorf-Spahn, K., Fritscha, E., 2016. Granulovirusresistenz beim Apfelwickler und neue resistenzbrechende Granulovirus-Isolate. Öko-Obstbau 2, 16-18.

Jehle, J.A., Lange, M., Wang, H., Hu, Z., Wang, Y., Hauschild, R., 2006. Molecular identification and phylogenetic analysis of baculoviruses from Lepidoptera. Virology 346, 180-193.

Jehle, J.A., Sauer, A., Fritsch, E., Undorf-Spahn, K., 2014. Resistance to Cydia pomonella granulovirus: novel findings on its distribution and diversity. Fördergemeinschaft Ökologischer Obstbau e.V. (FÖKO), Weinsberg, pp. 244-246.

Jehle, J.A., Schulze-Bopp, S., Undorf-Spahn, K., Fritsch, E., 2017. Evidence for a second type of resistance against Cydia pomonella granulovirus in field populations of codling moths. Appl Environ Microbiol 83, e02330-02316.

Jiantao, L., Zhu, L., Zhang, S., Deng, Z., Huang, Z., Yuan, M., Wu, W., Yang, K., 2016. The *Autographa californica* multiple nucleopolyhedrovirus *ac110* gene encodes a new *per os* infectivity factor. Virus Res 221, 30-37.

Jin, J., Dong, W., Guarino, L.A., 1998. The LEF-4 subunit of baculovirus RNA polymerase has RNA 5'-triphosphatase and ATPase activities. J Virol 72, 10011-10019.

Jukes, M.D., Motsoeneng, B.M., Knox, C.M., Hill, M.P., Moore, S.D., 2016. The comparative analysis of complete genome sequences from two South African betabaculoviruses: *Phthorimaea operculella* granulovirus and *Plutella xylostella* granulovirus. Arch Virol 161, 2917-2920.

Kariuki, C.W., McIntosh, A.H., 1999. Infectivity studies of a new baculovirus isolate for the control of the diamondback moth (Plutellidae: Lepidoptera). J Econ Entomol 92, 1093-1098.

Ke, J., Wang, J., Deng, R., Wang, X., 2008. *Autographa californica* multiple nucleopolyhedrovirus ac66 is required for the efficient egress of nucleocapsids from the nucleus, general synthesis of preoccluded virions and occlusion body formation. Virology 374, 421-431.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647-1649.

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

Kitchin, D., Bouwer, G., 2018. Significant differences in the intra-host genetic diversity of *Helicoverpa armigera* nucleopolyhedrovirus dnapol after serial *in vivo* passages in the same insect population. Arch Virol 163, 713-718.

Kool, M., Ahrens, C.H., Goldbach, R.W., Rohrmann, G.F., Vlak, J.M., 1994. Identification of genes involved in DNA replication of the *Autographa californica* baculovirus. Proc Natl Acad Sci U S A 91, 11212-11216.

Krappa, R., Knebel-Mörsdorf, D., 1991. Identification of the very early transcribed baculovirus gene PE-38. J Virol 65, 805-812.

Krejmer, M., Skrzecz, I., Wasag, B., Szewczyk, B., Rabalski, L., 2015. The genome of *Dasychira pudibunda* nucleopolyhedrovirus (DapuNPV) reveals novel genetic connection between baculoviruses infecting moths of the Lymantriidae family. BMC Genet 16, 759.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33, 1870-1874.

Kuzio, J., Jaques, R., Faulkner, P., 1989. Identification of *p74*, a gene essential for virulence of baculovirus occlusion bodies. Virology 173, 759-763.

Lacey, L.A., Shapiro-Ilan, D.I., Glenn, G.M., 2010. Post-application of anti-desiccant agents improves efficacy of entomopathogenic nematodes in formulated host cadavers or aqueous suspension against diapausing codling moth larvae (Lepidoptera: Tortricidae). Biocontrol Sci Technol 20, 909-921.

Lacey, L.A., Thomson, D., Vincent, C., Arthurs, S.P., 2008. Codling moth granulovirus: a comprehensive review. Biocontrol Sci Technol 18, 639-663.

Lacey, L.A., Unruh, T.R., 2005. Biological control of codling moth (*Cydia pomonella*, Lepidoptera: Tortricidae) and its role in integrated pest management, with emphasis on entomopathogens. Vedalia 12, 33-60.

Larem, A., 2019. Diversity and Host Interaction of Phthorimaea operculella granulovirus, Biology. Technische Universität Darmstadt, Darmstadt, p. 133.

Larem, A., Ben-Tiba, S., Wennmann, J.T., Gueli Alletti, G., Jehle, J.A., 2019. Elucidating the genetic diversity of Phthorimaea operculella granulovirus (PhopGV). J Gen Virol, doi: 10.1099/jgv.0.001215.

Lê, S., Josse, J., Husson, F., 2008. FactoMineR: An R Package for Multivariate Analysis. J Stat Softw 25, 1-18.

Lefoe, G., Hossain, M., Hossain, M., Williams, D., 2013. Application for the release of *Mastrus ridens* (Hymenoptera: Ichneumonidae) for the biological control of codling moth *Cydia*

*pomonella* (Lepidoptera: Tortricidae). Final Risk Analysis Report for the release of *Mastrus ridens* for the biological control of codling moth (*Cydia pomonella*), 1.

Li, G., Chen, H., Tang, Q., Huang, G., Deng, R., Wang, J., Wang, X., 2011. Effect of *ac68* knockout and *lef3* leading sequence disruption on viral propagation. Current Microbiology 62, 191-197.

Li, H., 2011a. Improving SNP discovery by base alignment quality. Bioinformatics 27, 1157-1158.

Li, H., 2011b. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27, 2987-2993.

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, arXiv e-prints.

Liu, C., Li, Z., Wu, W., Li, L., Yuan, M., Pan, L., Yang, K., Pang, Y., 2008. *Autographa californica* multiple nucleopolyhedrovirus *ac53* plays a role in nucleocapsid assembly. Virology 382, 59-68.

Liu, Q., 2010. Isolation and inentification of *Cydia pomonella* granulovirus Zhangye (CypoGV-ZY), Forrest Protection. Northwest A&F University Yangling, p. 34.

Liu, X., Ma, X., Lei, C., Xiao, Y., Zhang, Z., Sun, X., 2011. Synergistic effects of *Cydia pomonella* granulovirus GP37 on the infectivity of nucleopolyhedroviruses and the lethality of *Bacillus thuringiensis*. Arch Virol 156, 1707-1715.

Loparev, V.N., Gonzalez, A., Deleon-Carnes, M., Tipples, G., Fickenscher, H., Torfason, E.G., Schmid, D.S., 2004. Global identification of three major genotypes of varicella-zoster virus: longitudinal clustering and strategies for genotyping. J Virol 78, 8349-8358.

Lopez-Ferber, M., Simon, O., Williams, T., Caballero, P., 2003. Defective or effective? Mutualistic interactions between virus genotypes. Proc Biol Sci 270, 2249-2255.

Luque, T., Finch, R., Crook, N., O'Reilly, D.R., Winstanley, D., 2001. The complete sequence of the *Cydia pomonella* granulovirus genome. J Gen Virol 82, 2531-2547.

Maalouly, M., Franck, P., Lavigne, C., 2015. Temporal dynamics of parasitoid assemblages parasitizing the codling moth. Biol Control 82, 31-39.

Maghodia, A.B., Jarvis, D.L., Geisler, C., 2014. Complete genome sequence of the *Autographa californica* multiple nucleopolyhedrovirus strain E2. Genome announcements 2, e01202-01214.

McCarthy, C.B., Theilmann, D.A., 2008. AcMNPV *ac143* (*odv-e18*) is essential for mediating budded virus production and is the 30th baculovirus core gene. Virology 375, 277-291.

McDougal, V.V., Guarino, L.A., 1999. *Autographa californica* nuclear polyhedrosis virus DNA polymerase: measurements of processivity and strand displacement. J Virol 73, 4908-4918.

McDougal, V.V., Guarino, L.A., 2000. The *Autographa californica* nuclear polyhedrosis virus *p143* gene encodes a DNA *helicase*. J Virol 74, 5273-5279.

McLachlin, J.R., Miller, L.K., 1994. Identification and characterization of *vlf-1*, a baculovirus gene involved in very late gene expression. J Virol 68, 7746-7756.

Men, Q.-L., Chen, M.-H., Zhang, Y.-L., Feng, J.-N., 2013. Genetic structure and diversity of a newly invasive species, the codling moth, *Cydia pomonella* (L.) (Lepidoptera: Tortricidae) in China. Biol Invasions 15, 447-458.

Meyling, N.V., Thorup-Kristensen, K., Eilenberg, J., 2011. Below- and aboveground abundance and distribution of fungal entomopathogens in experimental conventional and organic cropping systems. Biol Control 59, 180-186.

Miele, S.A.B., Garavaglia, M.J., Belaich, M.N., Ghiringhelli, P.D., 2011. Baculovirus: molecular insights on their diversity and conservation. International journal of evolutionary biology 2011, 379424-379424.

Mikhailov, V.S., Okano, K., Rohrmann, G.F., 2003. Baculovirus alkaline nuclease possesses a 5'-->3' exonuclease activity and associates with the DNA-binding protein LEF-3. J Virol 77, 2436-2444.

Milks, M.L., Myers, J.H., 2000. The development of larval resistance to a nucleopolyhedrovirus is not accompanied by an increased virulence in the virus. Evolutionary Ecology 14, 645-664.

Milks, M.L., Theilmann, D.A., 2000. Serial selection for resistance to a wild-type and to a genetically modified nucleopolyhedrovirus in *Trichoplusia ni*. Biol Control 19, 283-289.

Milks, M.L., Washburn, J.O., Willis, L.G., Volkman, L.E., Theilmann, D.A., 2003. Deletion of *pe38* attenuates AcMNPV genome replication, budded virus production, and virulence in heliothis virescens. Virology 310, 224-234.

Mills, N., Pickel, C., Mansfield, S., McDougall, S., Buchner, R., Caprile, J., Edstrom, J., Elkins, R.B., Hasey, J.K., Kelley, K., Krueger, B., Olson, B., Stocker, R., 2000. Mass releases of wasps can reduce damage from codling moth. California Agriculture 54, 22-25.

Morgan, M., Anders, S., Lawrence, M., Aboyoun, P., Pages, H., Gentleman, R., 2009. ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. Bioinformatics 25, 2607-2608.

Moscardi, F., 1999. Assessment of the application of baculoviruses for control of Lepidoptera. Annu Rev Entomol 44, 257-289.

Motsoeneng, B.M., 2016. Genetic and biological characterisation of a novel South African *Cydia pomonella* granulovirus (CpGV-SA) isolate. Rhodes University.

Nakai, M., Takahashi, K., Iwata, K., Tanaka, K., Koyanagi, J., Ookuma, A., Takatsuka, J., Okuno, S., Kunimi, Y., 2017. Acquired resistance to a nucleopolyhedrovirus in the smaller tea tortrix *Adoxophyes honmai* (Lepidoptera: Tortricidae) after selection by serial viral administration. J Invertebr Pathol 145, 23-30.

Nguyen, P., Sýkorová, M., Šíchová, J., Kůta, V., Dalíková, M., Čapková Frydrychová, R., Neven, L.G., Sahara, K., Marec, F., 2013. Neo-sex chromosomes and adaptive potential in tortricid pests. Proceedings of the National Academy of Sciences 110, 6931-6936.

Nie, Y., Fang, M., Erlandson, M.A., Theilmann, D.A., 2012. Analysis of the *Autographa californica* multiple nucleopolyhedrovirus overlapping gene pair *lef3* and *ac68* reveals that AC68 is a *per os* infectivity factor and that LEF3 is critical, but not essential, for virus replication. J Virol 86, 3985-3994.

Noune, C., Hauxwell, C., 2016. Comparative analysis of HaSNPV-AC53 and derived strains. Viruses 8. 280, doi:10.3390/v8110280

Obenchain, V., Lawrence, M., Carey, V., Gogarten, S., Shannon, P., Morgan, M., 2014. VariantAnnotation : a Bioconductor package for exploration and annotation of genetic variants. Bioinformatics 30, 2076-2078.

Odendaal, D., Addison, M.F., Malan, A.P., 2015. Control of codling moth (*Cydia pomonella*) (Lepidoptera: Tortricidae) in South Africa with special emphasis on using entomopathogenic nematodes. African Entomology 23, 259-274.

Ohkawa, T., Washburn, J.O., Sitapara, R., Sid, E., Volkman, L.E., 2005. Specific binding of *Autographa californica* M nucleopolyhedrovirus occlusion-derived virus to midgut cells of *Heliothis virescens* larvae is mediated by products of *pif* genes *ac119* and *ac022* but not by *ac115*. J Virol 79, 15258-15264.

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

Olszewski, J., Miller, L.K., 1997. Identification and characterization of a baculovirus structural protein, VP1054, required for nucleocapsid formation. J Virol 71, 5040-5050.

Opoku-Debrah, J.K., Hill, M.P., Knox, C., Moore, S.D., 2013. Overcrowding of false codling moth, *Thaumatotibia leucotreta* (Meyrick) leads to the isolation of five new *Cryptophlebia leucotreta* granulovirus (CrleGV-SA) isolates. J Invertebr Pathol 112, 219-228.

Passarelli, A.L., Todd, J.W., Miller, L.K., 1994. A baculovirus gene involved in late gene expression predicts a large polypeptide with a conserved motif of RNA polymerases. J Virol 68, 4673-4678.

Pearson, M.N., Russell, R.L., Rohrmann, G.F., Beaudreau, G.S., 1988. *p39*, a major baculovirus structural protein: immunocytochemical characterization and genetic location. Virology 167, 407-413.

Peng, K., van Oers, M.M., Hu, Z., van Lent, J.W., Vlak, J.M., 2010. Baculovirus per os infectivity factors form a complex on the surface of occlusion-derived virus. J Virol 84, 9497-9504.

Pickel, C.P., Bethell, R.S., Coates, W.W., 1986. Codling moth management using degree days, Publication 4, University of California Statewide IPM Project, Berkeley, California, USA.

Pietruska, D., 2018. Vergleichende Transkriptomanalysen des Cydia pomonella granulovirus (CpGV) in sensiblen und resistenten Apfelwicklern (*Cydia pomonella*). Julius Kühn-Institut, Bundesforschungsinstitut für Kulturpflanzen, Quedlinburg, Germany.

Pijlman, G.P., Pruijssers, A.J., Vlak, J.M., 2003. Identification of *pif-2*, a third conserved baculovirus gene required for per os infection of insects. J Gen Virol 84, 2041-2049.

Prikhod'ko, E.A., Miller, L.K., 1999. The baculovirus PE38 protein augments apoptosis induced by transactivator IE1. J Virol 73, 6691-6699.

Pringle, K.L., Eyles, D.K., Brown, L., 2003. Trends in codling moth activity in apple orchards under mating disruption using pheromones in the Elgin area, Western Cape Province, South Africa. African Entomology 11, 65-75.

Raghavendra, A.T., Jalali, S.K., Ojha, R., Shivalingaswamy, T.M., Bhatnagar, R., 2017. Whole genome sequence and comparative genomic sequence analysis of *Helicoverpa armigera* nucleopolyhedrovirus (HearNPV-L1) isolated from India. VirusDisease 28, 61-68.

Rezapanah, M., Kharazi, P.A., Kamali, K., Huber, J., 2002. Survey on natural occurrence of Cydia pomonella granulovirus in apple orchards of Iran. Journal of Applied Entomology and Phytopathology, 69, 49-56.

Rezapanah, M., Shojai-Estabragh, S., Huber, J., Jehle, J.A., 2008. Molecular and biological characterization of new isolates of Cydia pomonella granulovirus from Iran. J Pest Sci 81, 187-191.

Robinson, A.S., 2005. Genetic Basis of the Sterile Insect Technique, in: Dyck, V.A., Hendrichs, J., Robinson, A.S. (Eds.), Sterile Insect Technique: Principles and Practice in Area-Wide Integrated Pest Management. Springer Netherlands, Dordrecht, pp. 95-114.

Rohrmann, G.F., 2013. in: 3rd (Ed.), Baculovirus Molecular Biology. National Center for Biotechnology Information (US) George Rohrmann., Bethesda (MD).

Russell, R.L.Q., Rohrmann, G.F., 1997. Characterization of P91, a protein associated with virions of an *Orgyia pseudotsugata* baculovirus. Virology 233, 210-223.

Sandanayaka, W., Chhagan, A., Page-Weir, N., Charles, J., 2011. Colony optimisation of *Mastrus ridens* (Hymenoptera: Ichneumonidae), a potential biological control agent of codling moth in New Zealand. New Zealand Plant Protection 64, 227-234.

Sangket, U., Vijasika, S., Noh, H., Chantratita, W., Klungthong, C., Yoon, I.K., Fernandez, S., Rutvisuttinunt, W., 2015. SNPer: An R library for quantitative variant analysis on single nucleotide polymorphisms among influenza virus populations. PLoS One 10, e0122812.

Sauer, A.J., 2017. Novel types of resistance of codling moth to Cydia pomonella granulovirus, Biology. Technische Universität Darmstadt, Darmstadt, p. 130.

Sauer, A.J., Schulze-Bopp, S., Fritsch, E., Undorf-Spahn, K., Jehle, J.A., 2017b. A third type of resistance to Cydia pomonella granulovirus in codling moths shows a mixed Z-Linked and autosomal inheritance pattern. Appl Environ Microbiol 83, e01036-01017.

Sauphanor, B., Berling, M., Toubon, J.-F., Reyes, M., Delnatte, J., Allemoz, P., 2006. Carpocapse des pommes cas de résistance au virus de la granulose en vergers biologiques: fruits et légumes. Phytoma-La défense des végétaux 590, 24-27.

Schmitt, A., Bisutti, I.L., Ladurner, E., Benuzzi, M., Sauphanor, B., Kienzle, J., Zingg, D., Undorf-Spahn, K., Fritsch, E., Huber, J., Jehle, J.A., 2013. The occurrence and distribution of resistance of codling moth to *Cydia pomonella* granulovirus in Europe. J Appl Entomol 137, 641-649.

Shen, J., Liu, W., Wan, F., Zhang, F., 2012. Characterization of the *Cydia pomonella* granulovirus CpGV-CJ01 from northwest China. Chin J Appl Ecol 49, 96-103.

Siegel, J.P., Lacey, L.A., Vossbrinck, C.R., 2001. Impact of a north American isolate of the Microsporidium Nosema carpocapsae on a laboratory population of the codling moth, *Cydia pomonella*. J Invertebr Pathol 78, 244-250.

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.

Simon, O., Williams, T., Caballero, P., Lopez-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., Caballero, P., 2004. Genetic structure of a *Spodoptera frugiperda* nucleopolyhedrovirus population: high prevalence of deletion genotypes. Appl Environ Microbiol 70, 5579-5588.

Slack, J., Arif, B.M., 2007. The baculoviruses occlusion-derived virus: virion structure and function. Advances in virus research 69, 99-165.

Smith, I.R., Crook, N.E., 1993. Characterization of new baculovirus genotypes arising from inoculation of *Pieris brassicae* with granulosis viruses. J Gen Virol 74, 415-424.

Smith, I.R.L., Crook, N.E., 1988. *In vivo* isolation of baculovirus genotypes. Virology 166, 240-244.

Stará, J., Kocourek, F., 2003. Evaluation of efficacy of *Cydia pomonella* granulovirus (CpGV) to control the codling moth (*Cydia pomonella*, Lep.: Tortricidae) in field trials. Plant Protect Sci 39, 117-125.

Steineke, S.B., 2004. Populationsdynamik des Cydia pomonella Granulovirus. Johannes Gutenberg-Universität Maiz, Mainz, p. 135.

Steineke, S.B., Jehle, J.A., 2002. Mathematische Modellierung der Populationsdynamik von genetisch veränderten Mikroorganismen am Beispiel von Baculoviren: Forschungsbericht 29889418. Umweltbundesamt.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol 30, 2725-2729.

Tanada, Y., Leutenegger, R., 1968. Histopathology of a granulosis-virus disease of the codling moth, *Carpocapsa pomonella*. J Invertebr Pathol 10, 39-47.

Tanada, Y.A., 1964. Granulosis virus of codling moth, *Carpocapsa pomenella* (Linnaeus) (Olethreutidae, Lepidoptera). J Insect Pathol 6, 39-47.

Tao, X.Y., Choi, J.Y., Kim, W.J., Lee, J.H., Liu, Q., Kim, S.E., An, S.B., Lee, S.H., Woo, S.D., Jin, B.R., Je, Y.H., 2013. The *Autographa californica* multiple nucleopolyhedrovirus ORF78 is essential for budded virus production and general occlusion body formation. J Virol 87, 8441-8450.

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.

Treutlein, B., Lee, Q.Y., Camp, J.G., Mall, M., Koh, W., Shariati, S.A.M., Sim, S., Neff, N.F., Skotheim, J.M., Wernig, M., Quake, S.R., 2016. Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. Nature 534, 391.

Undorf-Spahn, K., Fritsch, E., Huber, J., Kienzle, J., Zebitz, C.P., Jehle, J.A., 2012. High stability and no fitness costs of the resistance of codling moth to *Cydia pomonella* granulovirus (CpGV-M). J Invertebr Pathol 111, 136-142.

van der Merwe, M., Jukes, D.M., Rabalski, L., Knox, C., Opoku-Debrah, K.J., Moore, D.S., Krejmer-Rabalska, M., Szewczyk, B., Hill, P.M., 2017. Genome analysis and genetic stability of the *Cryptophlebia leucotreta* Granulovirus (CrleGV-SA) after 15 years of commercial use as a biopesticide. Int J Mol Sci 18, doi:10.3390/ijms18112327.

van Oers, M.M., Vlak, J.M., 2007. Baculovirus genomics. Curr Drug Targets 8, 1051-1068.

Vanarsdall, A.L., Okano, K., Rohrmann, G.F., 2006. Characterization of the role of very late expression factor 1 in baculovirus capsid structure and DNA processing. J Virol 80, 1724-1733.

Vanarsdall, A.L., Pearson, M.N., Rohrmann, G.F., 2007. Characterization of baculovirus constructs lacking either the Ac101, Ac142, or the Ac144 open reading frame. Virology 367, 187-195.

Vincent, C., Andermatt, M., Valéro, J., 2007. Madex® and VirosoftCP4®, viral biopesticides for codling moth control, Biological control: a global perspective. CABI, Wallingford, UK, pp. 336-343.

Wang, C., Feng, M., 2014. Advances in fundamental and applied studies in China of fungal biocontrol agents for use against arthropod pests. Biol Control 68, 129-135.

Wang, M., Tuladhar, E., Shen, S., Wang, H., van Oers, M.M., Vlak, J.M., Westenberg, M., 2010a. Specificity of baculovirus P6.9 basic DNA-binding proteins and critical role of the C terminus in virion formation. J Virol 84, 8821-8828.

Wang, R., Deng, F., Hou, D., Zhao, Y., Guo, L., Wang, H., Hu, Z., 2010b. Proteomics of the *Autographa californica* nucleopolyhedrovirus budded virions. J Virol 84, 7233-7242.

Wearing, C.H., 2016. Distribution characteristics of eggs and neonate larvae of codling moth, *Cydia pomonella* (L.) (Lepidoptera: Tortricidae). Int J Insect Sci 8, 33-53.

Wennmann, J.T., Radtke, P., Eberle, K.E., Gueli Alletti, G., Jehle, J.A., 2017. Deciphering single nucleotide polymorphisms and evolutionary trends in isolates of the Cydia pomonella granulovirus. Viruses 9, 227, doi:10.3390/v9080227.

Whitford, M., Faulkner, P., 1992. Nucleotide sequence and transcriptional analysis of a gene encoding *gp41*, a structural glycoprotein of the baculovirus *Autographa californica* nuclear polyhedrosis virus. J Virol 66, 4763-4768.

Wickham, H., 2016. ggplot2: elegant graphics for data analysis. Springer.

Wickham, H., 2018. R: A language and environment for statistical computing, R Foundation for Statistical Computing, Houston, USA.

Willett, M.J., Neven, L., Miller, C.E., 2009. The occurrence of codling moth in low latitude countries: validation of pest distribution reports. Horttechnology 19, 633-637.

Williams, L.M., Ma, X., Boyko, A.R., Bustamante, C.D., Oleksiak, M.F., 2010. SNP identification, verification, and utility for population genetics in a non-model genus. BMC Genet. doi: 10.1186/1471-2156-11-32.

Wilson, M.E., Mainprize, T.H., Friesen, P.D., Miller, L.K., 1987. Location, transcription, and sequence of a baculovirus gene encoding a small arginine-rich polypeptide. J Virol 61, 661-666.

Winstanley, D., Crook, N.E., 1993. Replication of *Cydia pomonella* granulosis virus in cell cultures. J Gen Virol 74, 1599-1609.

Wu, C.P., Huang, Y.J., Wang, J.Y., Wu, Y.L., Lo, H.R., Wang, J.C., Chao, Y.C., 2010. *Autographa californica* multiple nucleopolyhedrovirus LEF-2 is a capsid protein required for amplification but not initiation of viral DNA replication. J Virol 84, 5015-5024.

Wu, W., Liang, H., Kan, J., Liu, C., Yuan, M., Liang, C., Yang, K., Pang, Y., 2008. *Autographa californica* multiple nucleopolyhedrovirus *38K* is a novel nucleocapsid protein that interacts with VP1054, VP39, VP80, and itself. J Virol 82, 12356-12364.

Wu, Z., 2015. The adaptability of *Cydia pomonella* granulovirus response to strong UV index sunlight in Xinjiang relative to the biological control of the codling moth. Northwest A&F University, Yangling, p. 130.

Xing, L., Yuan, C., Wang, M., Lin, Z., Shen, B., Hu, Z., Zou, Z., 2017. Dynamics of the interaction between cotton bollworm *Helicoverpa armigera* and nucleopolyhedrovirus as revealed by integrated transcriptomic and proteomic analyses. Mol Cell Proteomics 16, 1009-1028.

Xu, H.J., Yang, Z.N., Wang, F., Zhang, C.X., 2006. *Bombyx mori* nucleopolyhedrovirus ORF79 encodes a 28-kDa structural protein of the ODV envelope. Arch Virol 151, 681-695.

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

Xue, J., Qiao, N., Zhang, W., Cheng, R.-L., Zhang, X.-Q., Bao, Y.-Y., Xu, Y.-P., Gu, L.-Z., Han, J.-D.J., Zhang, C.-X., 2012. Dynamic interactions between *Bombyx mori* nucleopolyhedrovirus and its host cells revealed by transcriptome analysis. J Virol 86, 7345-7359.

Yuan, M., Huang, Z., Wei, D., Hu, Z., Yang, K., Pang, Y., 2011. Identification of *Autographa californica* nucleopolyhedrovirus *ac93* as a core gene and its requirement for intranuclear microvesicle formation and nuclear egress of nucleocapsids. J Virol 85, 11664-11674.

Zamora-Avilés, N., Murillo, R., Lasa, R., Pineda, S., Figueroa, J.I., Bravo-Patiño, A., Díaz, O., Corrales, J.L., Martínez, A.M., 2017. Genetic and biological characterization of four nucleopolyhedrovirus isolates collected in Mexico for the control of *Spodoptera exigua* (Lepidoptera: Noctuidae). J Econ Entomol 110, 1465-1475.

Zhang, R.Z., Wang, F.X., Zhang, Y.L., Chen, H.J., Luo, J.C., Wang, Q.Y., Liu, W.X., Ainiwaer, M., Pu, C.J., Yan, Y.G., Guo, J.M., Liu, X.Y., Chen, J.G., Zhang, Z.F., Yang, S., Xu, J.J., Cui, G.Z., Xu, J., 2012. Progress on monitoring and control of the codling moth, *Cydia pomonella* (L.). Chin J Appl Ecol 49, 37-42.

Zhang, X.Z., 1957. Taxonomic notes on the codling moth, *Carpocapsa pomonella* L. in Sinkiang. Acta Entomol Sin 7, 467-472.

Zhao, L., Hou, P., Zhu, G., Li, M., Xie, T., Liu, Q., 2015. Mapping the disjunct distribution of introduced codling moth *Cydia pomonella* in China. Agric Forest Entomol 17, 214-222.

Zheng, C., Liu, Q., Li, J., Dong, K., Feng, J., Zhang, Y., Wang, D., 2011. Virulence determination for *Cydia pomonella* granulovirus Zhangye strain ( CypoGV-zy). Journal of Northwest Forestry University 26, 121-123.

Zichová, T., Stará, J., Kundu, J.K., Eberle, K.E., Jehle, J.A., 2013. Resistance to *Cydia pomonella* granulovirus follows a geographically widely distributed inheritance type within Europe. Biocontrol 58, 525-534.

Zichová T., F.V., Kocourek F., Stará J., 2011. Differences in the susceptibility of codling moth populations to *Cydia pomonella* granulovirus in the Czech Republic. Horticulture 38, 21-26.

Zimmermann, G., Huger, A.M., Kleespies, R.G., 2013. Occurrence and prevalence of insect pathogens in populations of the codling moth, *Cydia pomonella* L.: a long-term diagnostic survey. Insects 4, 425-446.

Zingg, D., Züger, M., Bollhalder, F., Andermatt, M., 2011. Use of resistance overcoming CpGV isolates and CpGV resistance situation of the codling moth in Europe seven years after the first discovery of resistance to CpGV-M. IOBC-WPRS Bull 66, 401-404.

## Acknowledgments

First of all, I would like to convey my Herzlichen Dank to my supervisor Prof. Dr. Johannes A. Jehle for his inspirations, suggestions and helps during my doctoral study at Institute for Biological Control, Julius Kühn-Institut (JKI) in Darmstadt, Germany. From the beginning of English learning to final doctoral thesis, he teaches me how to design, perform and finalize research in a proper way, as well as how to present my results to different audiences. The most important and lifelong benefit thing is to cultivate my rigorous scientific way of thinking. With his help and effort, I could attend three international conferences (France, USA and Australia) and three national meetings in Germany (Quedlinburg, Siebeldingen and Stuttgart). These scientific communications offered me an opportunity to connect other researchers and to better understand the science in a global perspective. I also would like to thank Prof. Dr. Gerhard Thiel for his willingness to be my second supervisor and to read my doctoral thesis.

I highly appreciated the support from Prof. Dr. Dun Wang (Northwest A&F University). With his favor and advices, I decided to apply the funding from China Scholarship Council (CSC) to study abroad. His personality gives me optimistic impact on my master and one year of doctoral study in China.

Many thanks to the “Gemeinschaft der Förderer und Freunde des Julius Kühn-Instituts e.V (GFF)”, and Society for Invertebrate Pathology (SIP) and JKI, these organizations funded me to join the conferences. With the support from JKI I could complete my doctoral study.

My two office roommates of Dr. Jörg Wennmann and Dr. Annette Sauer are especially acknowledged for their super assistance of R programming and great favor of understanding German style, respectively. Thank Dr. Eckhard Koch for daily conversation with about the history of the world in English and sometimes in German.

With the help of our working group, I could smoothly conduct all bioassays and resistance selections. I owe a big thank to my group members of Birgit Ruoff, Dr. Eva Fritsch and Dr. Karin Undorf-Spahn, Doris El Mazouar and Eckhard Gabrys, Dr. Yu Xi, Dr. Gianpiero Gueli Alletti, Dr. Andreas Larem, Sarah Biganski, saoussen Bentiba, Marina Eigenbrod and Maximilian Amberger. I appreciate Dr. Doreen Gabriel and her introduction to the statistics program "R" that help me a lot to analyze my data. Additional appreciation is given to Simon Feiertag for excellent photography.

Other PhD students Katharina Saar, Mokhtar Abonaem, Helen Pfitzner, Silvia Mátray, Carina Ehrlich, Camilla Englert and Shili Yang always offered their hand on me, once I need a favor from study or daily life. This support and understanding have allowed me to get this far.

During the last four years and now, I was enjoying a great time at JKI and had lots of wonderful conversations with Susanne Ganzer, Juliana Pelz, Dietrich Stephan, Annette Herz.

Finally, I thank my parents! Without your support, patience, motivation and love, this doctoral thesis would never be finished!

## Curriculum Vitae

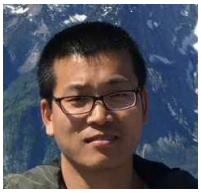
March 2019

Jiangbin Fan

16.07.1987, Shaanxi
Nationality: Chinese

Julius Kühn-Institut (JKI), Institute for Biological Control
Federal Research Center for Cultivated Plants
Heinrichstr. 243, 64287 Darmstadt, Germany

(0049)6151 407 248, Fan.Jiangbin@julius-kuehn.de
fanjiangbin@outlook.com



### PROFESSIONAL RECORDS

Since 2015 — PhD student – Work at the Institute for Biological Control, Julius Kühn-Institut in Darmstadt, Germany. Enroll as PhD student at theTechnische Universität Darmstadt, Darmstadt, Germany. Study on codling mothresistance to Cydia pomonella granulovirus (CpGV).

### EDUCATIONAL RECORDS

2014 – 2015 — PhD student – study at the Department of Plant Protection in Northwest A&F University, Yangling, China. Topic: Prepare RNA samples from codling moth for RNA-seq and confirm the RNA-seq results using qRT-PCR; CpGV gene expression in vitro using Bac-to-Bac Baculovirus Expression System.

2011 – 2014 — MSc.– study Northwest A&F University, Yangling, China. Thesis title: The mechanisms of CpGV resistance to UV radiation and rearing the codling moth on semi-artificial diet.

2006 – 2010 — BSc. – study at the Department of Plant Protection – Qingdao Agricultural University, Qingdao, China.

### SCHOLARSHIPS AND AWARDS

2015 – 2017 — Funded by China Scholarship Council (CSC) for study onresistance in field codling moth against CpGV at the Institute for Biological Control, Julius Kühn-Institut in Darmstadt, Germany.

2017 — Travel Award of the International Congress for Invertebrate Pathology and Microbial Control (Microbial Control Division) in San Diego, USA. 50th Meeting of the SIP.

2018 — Travel Award of the International Congress for Invertebrate Pathology and Microbial Control (Virus Division) in Gold Coast, Australia. 51th Meeting of the SIP.

### OTHER ACTIVITIES

2016 – 2019 — Member of the Society for Invertebrate Pathology (SIP). Conference participation from 2016 to 2018.

2017 – 2019 — Student representative of Microbial Control Division (SIP).

### CONFERENCE CONTRIBUTIONS

Annual meeting of the Society for Invertebrate Pathology (SIP):

2018 — Deciphering the population structure of genotype mixtures of CpGV field isolates by Next Generation Sequencing techniques and improved sequence analyses methods (*oral presentation*) (Gold Coast, Australia).

2017 — New Chinese isolates of Cydia pomonella granulovirus provide novel genetic diversity in the arms-race of resistance of codling moth. (*oral presentation*) (San Diego, USA).

2016 Novel Cydia pomonella granulovirus isolates break virus resistance in codling moth (*poster*) (Tours, France)

German Conference on Plant Protection (national conference):

2018 The genetic diversity of Cydia pomonella granulovirus in natural isolates and commercially applied products (Stuttgart, Germany).

Annual Young Scientists Meeting (YSM) at Julius Kühn-Institut

2016 Novel Cydia pomonella granulovirus isolates break virus resistance in codling moth. (*oral presentation*) (Quedlinburg, Germany).

2017 Genetic diversity of new Chinese Cydia pomonella granulovirus isolates (*oral presentation*) (Siebeldingen, Germany).

**PUBLICATIONS**

*Peer reviewed journals*:

Wu, Z., **Fan**, **J**., Yu, H., Wang, D., Zhang, Y. (2015). Ultraviolet protection of the Cydia pomonella granulovirus using zinc oxide and titanium dioxide. Biocontrol Science and Technology 25(1): 97-107.

**Fan**, **J**., Wu, Z., Liu, G., Wang, D. (2015). Synergistic enhancement of Cydia pomonella granulovirus virulence using an adjuvant. Journal of Safety 24(4): 310-314 (in Chinese).

**Fan**, **J**., Wu, Z., Shang, S., Mijiti, A., Zhang, Y., Wang, D. (2015). Optimization of semi-artificial and temperature in rearing the codling moth *Cydia pomonella* (L.). Journal of Plant Protection 42(1):45-50 (in Chinese).

Darmstadt, March 2019

Ehrenwörtliche Erklärung:

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit entsprechend den Regeln guter wissenschaftlicher Praxis selbstständig und ohne unzulässige Hilfe Dritter angefertigt habe.

Sämtliche aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie sämtliche von Anderen direkt oder indirekt übernommenen Daten, Techniken und Materialien sind als solche kenntlich gemacht. Die Arbeit wurde bisher bei keiner anderen Hochschule zu Prüfungszwecken eingereicht.

Darmstadt, den ……

.....