Nudivirus Nexus

Synthesis of diversity, evolution and pathology within the family *Nudiviridae*

Jirka Manuel Petersen

Propositions

1. Nudiviruses already existed during the era of the dinosaurs (this thesis)

2. Data-driven and traditional virus discovery are two sides of the same coin (this thesis)

3. Everything becomes a science when you dig deep enough

4. Conservative taxonomy impairs scientific collaboration

5. Video games outcompete social media in terms of benefits for society

6. The benefits of artificial intelligence end when convenience eliminates the prospect for self-growth

Propositions belonging to the thesis entitled:

Nudivirus Nexus – Synthesis of diversity, evolution and pathology within the family *Nudiviridae*

Jirka Manuel Petersen

Wageningen, 20 March 2025

Nudivirus Nexus

Synthesis of diversity, evolution and pathology within the family *Nudiviridae*

Jirka Manuel Petersen

Thesis committee

Promotors

Prof. Dr. Monique van Oers

Professor of Virology

Wageningen University & Research

Co-promotors

Dr. Jean-Michel Drezen

Research director at Institut de Recherche sur la Biologie de l'Insecte

University of Tours, France

Dr. Annie Bézier

Research associate at Institut de Recherche sur la Biologie de l'Insecte

University of Tours, France

Other members

Prof. Dr. Dick de Ridder, Wageningen University & Research

Prof. Dr. Johannes Jehle, Julius Kühn Institute, Darmstadt, Germany

Dr. Anne-Nathalie Volkoff, INRAE, Paris, France

Dr. Mark Engelsma, Wageningen University & Research

This research was conducted under the auspices of the Doctoral School Santé, Sciences Biologiques et Chimie du Vivant (SSBCV), University of Tours, France, and the Graduate School of Production Ecology and Resource Conservation (PE&RC), the Netherlands, and as part of the joint PhD programme Insect Doctors

Nudivirus Nexus

Synthesis of diversity, evolution and pathology within the family *Nudiviridae*

Jirka Manuel Petersen

Thesis
submitted in fulfilment of the requirements for the joint degree of doctor between
University of Tours
by the authority of the President, Prof. Dr Philippe Roingeard,
and
Wageningen University
by the authority of the Rector Magnificus, Prof. Dr C. Kroeze,
in the presence of the
Thesis Committee appointed by the Academic Board of both universities
to be defended in public
on Thursday 20 March 2025
at 3:30 p.m. in the Omnia Auditorium

Jirka Manuel Petersen

Nudivirus Nexus – Synthesis of diversity, evolution and pathology within the family *Nudiviridae*

204 pages

Joint PhD thesis | University of Tours, France, and Wageningen University, Wageningen, The Netherlands | (2025)

With references, with summary in English

DOI: https://doi.org/10.18174/681223

Table of contents

“**Every generation got its own disease, and I’ve got mine.**”

– *Every Generation* by Fury in the Slaughterhouse (2002)

Chapter 1

General introduction

Of monsters and men: The pre-virology era

Ancient disease outbreaks were commonly interpreted by mankind as a consequence of supernatural forces or corrupted environments, such as dirty water. During these times, diseases such as cholera, malaria, and the plague, were often believed to originate from so called miasmatic foci, in accordance with the *Miasma theory* [1, 2]. The Miasma theory already existed during the ancient times of Greece, lingered through the Middle Ages, and persisted until early modern times [3]. Over time, people recognised human-to-human interactions as another driver for spreading diseases. A satirical illustration from the 1850s depicts this emerging awareness by showing a drop of water from the River Thames. Inside this droplet, bizarre animal-like and humanoid monsters are swimming around. This caricature aimed to symbolise the dual fear towards contaminated water and human contact for the transmission of illness (**Figure 1**) [4].

Only decades later, pioneering scientific achievements by scientists like Louis Pasteur and Robert Koch led to the distancing from and eventual replacement of the miasma theory. In 1878, Pasteur published the *Germ Theory of Disease*, proposing that diseases were not simply arising from foul places, but were rather caused by nonvisible agents, or microbes, responsible for specific types of sickness [5]. Koch recognised Pasteur's germ theory, which led him to further validate his colleague's findings with the commitment to develop methods capable of isolating and identifying those disease-causing microbes [6]. His endeavour resulted in the formulation of *Koch's postulates*, positing four criteria that are needed for a microorganism to establish disease in its host [7]. In the framework of those scientific advances, the bacteria responsible for tuberculosis and anthrax were identified, opening up new scopes for disease-related studies. However, despite that pursuing boom in pathogenic microorganism identifications, it was not until 1898 that even smaller pathogenic agents – namely viruses – were experimentally discovered.

Figure 1. "A Drop of London Water", *PUNCH* Magazine, January – June 1850. The reproduction of this illustration was approved by Punch Ltd., www.punch.co.uk

Origins: The first discovered virus

The pioneering work of Dmitri Ivanovsky and Martinus Beijerinck in the early 20th century was crucial for the official recognition of viruses as distinct microbial agents [8, 9]. While studying tobacco plants with disease symptoms, Ivanovsky found that the infectious agent could pass through filters, whereas bacteria were trapped. After confirming these findings, Beijerinck introduced the concept of a "*contagium vivum fluidum*," to propose that the agent was an infectious liquid. Eventually, this study led to discovery and characterisation of the first documented virus, namely the tobacco mosaic virus, a pathogen of various solanaceous crops [10]. Finally, the introduction of electron microscopy (EM) gave viruses "a face" in the mid-1930s, which allowed for the visualisation and morphological distinction of

viruses [11]. By assessing the morphological traits of viruses through EM-assisted studies, scientists forged the base for traditional virus classification and taxonomy [12].

Here and now: Modern virology

After the early breakthroughs of the 20th century, the current millennium brought technological advances and innovations that elevated the scope and depth of virus research to a new level. Sanger sequencing made it possible to decipher the code of nucleic acids from various biological specimens and paved the way for sequence-based virology [13]; however, it was the emergence of next-generation sequencing (NGS) techniques that enabled the data-driven virus discovery (DDVD) approach and in-depth characterisation of viromes and their genomic variations [14]. Today it is a known fact that viruses are incredibly abundant and diverse, making up a significant part of the biological mass; for instance, every litre of seawater contains up to 10 billion viruses [15]. According to the Master Species List (MLS) by the International Committee on Taxonomy of Viruses (ICTV) there are currently 14,690 taxonomically characterised virus species (https://ictv.global/msl), correlated to virus infections in humans, animals, plants and lower organisms.

Beyond Pathogens: The ecological and beneficial roles of viruses

Beyond their reputation as pathogens, viruses majorly shape our ecosystems by influencing microbial communities, biogeochemical cycles, or even affecting the evolution of life itself [16-18]. The environmental significance of viruses further extends to their societal importance. For instance, bacteria-infecting viruses, or bacteriophages, are used in phage therapy as an alternative to antibiotics, which can become ineffective with the rise of antibiotic resistant strains [19]. Further uses of viruses in the medical sector include the development and manufacturing of virus-like particles as vaccine delivery systems [20]. One of the most prominent virus-derived systems for vaccine development is the Baculoviruses Expression Vector System (BEVS) [21, 22], primarily derived from a member of the *Baculoviridae* family, which allows high-yield protein production [23]. More than ten commercially available BEVS-derived vaccines have been developed as protection against various human pathogenic diseases, including cervical cancer (Cervarix™), influenza (Flublok® and Flublok Quadrivalent®), and SARS-CoV-2 (NVX-CoV2373 and Weikexin) [24]. Baculoviruses are enveloped bacilliform (rod-shaped) viruses with large double-stranded DNA (dsDNA) genomes that are primarily found in insects of the order *Lepidoptera* (moths and butterflies). Next to their use as a powerful molecular tool, baculoviruses are commonly used as biocontrol agents against insect pests. The effectiveness of baculoviruses as biocontrol agents can partly be attributed to their ability to form occlusion bodies (OBs). Baculoviral OBs are protein matrices that protect viral progeny from UV radiation and other damaging factors, allowing them to stay outside their host for a longer period until they are ingested again [25]. The application of viruses for insect pest control, in place of chemical pesticides, offers an environmentally friendly and sustainable alternative for agriculture [26, 27], making it an extensively studied research area. Taken together, their use in bioprocessing and biocontrol makes baculoviruses one of the most widely studied and prominent virus families, with special attention drawn to the baculovirus Autographa californica multiple nucleopolyhedrovirus (AcMNPV). However, the focus on baculoviruses also led to mischaracterisations of virus groups with host ranges and morphologies similar to baculoviruses, especially before the advent of sequence-based virology.

A long over-shadowed sibling: The family *Nudiviridae*

In the past, researchers quickly associated rod-shaped virions found inside insect hosts with baculoviruses. In time, this extended beyond insects to other invertebrates, such as crustaceans. Eventually, more and more infections featuring baculovirus-like virions emerged and got characterised as such, including a group of rod-shaped viruses found in both insects and crustaceans. However, these viruses seemed to lack OBs (**Figure 2**) [28-31]. Little did virologist back then know that most of these unconventional "baculoviruses" would turn out to be a distinct clade of arthropod-infecting viruses, which are now known as nudiviruses of the family *Nudiviridae* [32].

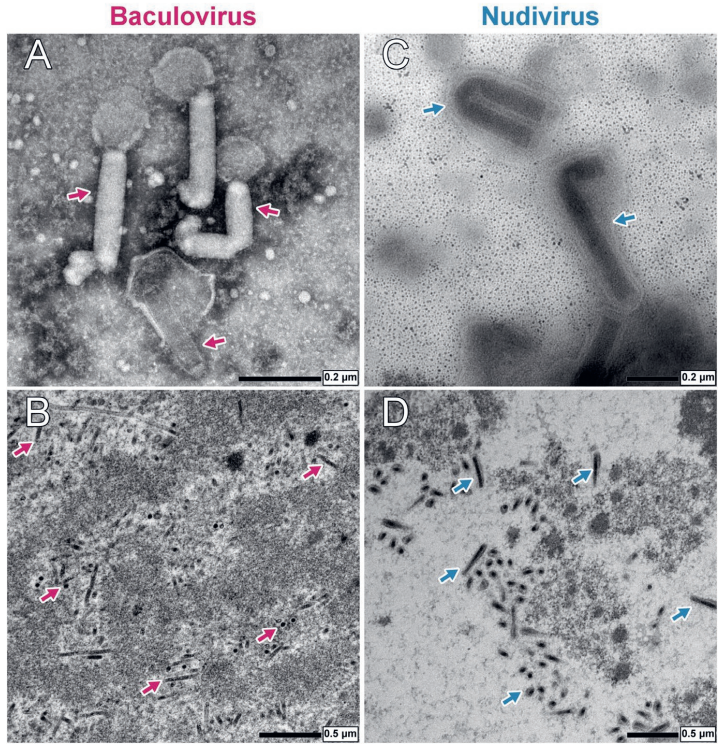


Figure 2. Electron microscopy images comparing the morphology of rod-shaped virions from a baculovirus and nudivirus in suspension or cells of their lepidopteran hosts. **(A)** Budded virions from Autographa californica multiple nucleopolyhedrovirus (AcMNPV) in suspension. Bar, 200 nm. **(B)** Intracellular virions of AcMNPV (purple arrows) 60 hours after infection of a cell line (Sf9) derived from the ovarian tissue of *Spodoptera frugiperda* (Noctuidae) [33]. Bar, 500 nm. **(C)** Virions of Heliothis zea nudivirus 1 (HzNV-1) in suspension. Bar, 100 nm. **(D)** Intracellular virions of HzNV-1 (blue arrows) 60 hours after infection of a cell line (Hz-AM1) derived from the ovarian tissue of *Helicoverpa zea* (Noctuidae) [34]. Bar, 500 nm. Using standard methods, the virion-containing suspensions (A, C) and ultrathin sections of virus-infected cells (B, D) were negatively stained with uranyl acetate. EM images courtesy Jan W. M. van Lent, Wageningen Electron Microscopy Centre (WEMC).

Despite their taxonomic distinction, nudiviruses and baculoviruses are closely related sister families. Their relatedness is based on molecular information and phylogenetic analyses of conserved viral genes, which also revealed their shared evolutionary origin with a number of other arthropod-infecting large dsDNA viruses, namely the hytrosaviruses (family *Hytrosaviridae*) and members of the recently proposed family Filamentoviridae [35]. Together, these virus families form the order *Lefavirales* [36]. In addition, the endogenised bracoviruses found in the genomes of braconid parasitoid wasps are

evolutionarily related to these large DNA viruses, and it has been proposed to place these in the new genus *Braconudivirus*. A bit more distant sits the family *Nimaviridae* that, together with the viruses classified in the order *Lefavirales*, falls under the class of *Naldaviricetes* [36] (**Figure 3**).

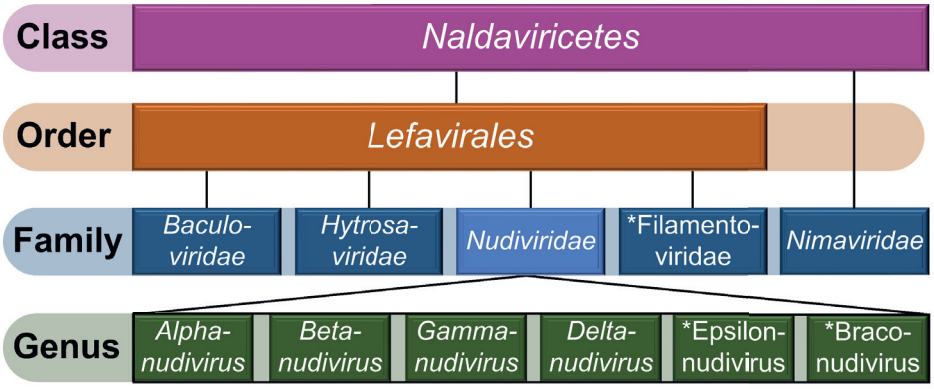


Figure 3. Taxonomic hierarchy of nuclear arthropod large DNA viruses within the class *Naldaviricetes*. This class includes five families: *Baculoviridae*, *Hytrosaviridae*, *Nudiviridae*, Filamentoviridae, and *Nimaviridae*. Within the family *Nudiviridae* sit the four officially recognised genera *Alphanudivirus*, *Betanudivirus*, *Gammanudivirus* and *Deltanudivirus*. The proposed genus Braconudivirus (commonly known as the genus *Bracovirus* from the family *Polydnaviridae*) represents endogenous viruses that have been fully integrated and domesticated within the genomes of braconid wasps. Names marked with an asterisk (*) denote proposed virus clades that are yet to be officially recognised by the International Committee on Taxonomy of Viruses (ICTV).

Since the official recognition of the family *Nudiviridae* took place quite recently in 2014 [37], only a few members have been studied in detail; for instance, the Oryctes rhinoceros nudivirus (OrNV) that is used as a biocontrol agent against infestation of the coconut rhinoceros beetle [38], or the two isolates of the single species *Betanudivirus hezeae* (namely Heliothis zea nudivirus 1, HzNV-1; Helicoverpa zea nudivirus 2, HzNV-2) of the lepidopteran pest *Helicoverpa* (formerly *Heliothis*) *zea*. A decade ago, members of the *Betanudivirus* genus were regarded as the closest exogenous relatives to the endogenous bracovirus group [39], however, just a year later, a cranefly-infecting nudivirus was identified as an even closer relative [40]. Despite this discovery, many aspects of nudivirus evolution – especially how the nudiviral ancestor became domesticated in ancestral wasps and eventually led to the domestication of bracoviruses as we know them today – remain unclear and patchy. This uncertainty also extends to their shared genetic characteristics, including the conservation of viral gene functions or replication mechanisms, which provides yet unresolved questions regarding functional genomics of the members of *Nudiviridae*. Additionally, their pathology, evolution, diversity, and ecology, remain underexplored, despite their potential as biocontrol agents or impacts as pathogenic agents in insect rearing facilities and crustacean aquacultures.

Research objectives and outline of this thesis

The aim of this thesis is to deepen and broaden our understanding of the *Nudiviridae* family and its members, including the unconventional bracoviruses. This work includes the extensive review of reposited literature and computational data, as well as experimental studies to gain more in-depth knowledge on specific members of the family, including HzNV-1 and Cotesia congregata bracovirus (CcBV).

Chapter 2 provides a detailed review of the known nudivirus species as of 2022. In addition to providing a representative phylogeny of the *Nudiviridae* family, this chapter explores the biological, pathological,

and genomic differences and similarities between nudiviruses and related viruses, including baculoviruses (*Baculoviridae*) and bracoviruses (treated at the time of writing as the genus *Bracovirus* within the family *Polydnaviridae*).

Chapter 3 features the discovery of new nudiviruses from publicly accessible sequencing read archives (SRA) through a data-driven virus discovery (DDVD) approach. This chapter emphasises the advances in nudivirus research since 2022, as discussed in Chapter 2. The full assembly of novel nudivirus genomes from ectoparasitic insect hosts prompted the proposition for two new nudivirus genera that comprise lice-associated members. Lastly, this chapter provides new insights into phylogeny-based gene synteny, evolution, geographical distribution, and potential ecological roles of the members of the family *Nudiviridae*.

The work conducted in **Chapter 4** lays the groundwork for the experimental study of the betanudivirus HzNV-1, which was specifically acquired for this research due to its initially presumed close relationship with the bracoviruses. In the Laboratory of Virology at Wageningen University, a cell culture-based production system and relevant methods were adapted to investigate the infection cycle and pathology dynamics of this virus in cell culture.

Using the established HzNV-1 system, **Chapter 5** covers in-depth gene expression profiling to examine the transcriptional dynamics in both HzNV-1 and cultured host cells over time. This analysis seeks to reveal the temporal expression patterns of HzNV-1 genes and to identify host genes that are differentially expressed, either upregulated or downregulated, throughout the infection process. Furthermore, we present a motif enriched in the upstream sequences of HzNV-1 genes that might serve as a promoter for virus gene expression.

Another experimental study is presented in **Chapter 6**, aiming to verify and identify new replication unit motifs (RUMs) within the clade of bracoviruses using long-read sequencing. To achieve this, high molecular weight DNA was extracted from the dissected ovaries of the parasitoid wasp *Cotesia congregata* and subjected to PacBio HiFi sequencing. The resulting sequencing data were analysed and compared with long-read sequencing data from another parasitoid wasp, *Toxoneuron nigriceps*. By examining the coverage patterns of the mapped reads against the wasp genomes, novel RUMs and proviral regions were uncovered, including a new proviral locus of CcBV, encoding a previously unrecognised protein tyrosine phosphatase (PTP). The work in this chapter also reveals that there is no sequence conservation between the RUMs of CcBV and Toxoneuron nigriceps bracovirus (TnBV), suggesting that TnBV's replication is likely controlled by species-specific regulatory regions.

Finally, the findings of this thesis are discussed in **Chapter 7**, where the laboratory results and the computational analyses from Chapters 4 and 5 are combined to propose a model of the HzNV-1 infection cycle. This chapter will also explore the broader implications of these findings for our understanding of viral evolution and host interactions within the *Nudiviridae* family. Potential avenues for future research are suggested, particularly in assessing the ecological role, evolutionary history and biocontrol-related potential of these viruses.

Acknowledgements

I would like to express my sincere gratitude to my promoters, Monique van Oers and Jean-Michel Drezen, for their constructive remarks and feedback on this chapter.

“**Caught in a landslide, no escape from reality,**

Open your eyes, look up to the skies and see.”

– *Bohemian Rhapsody* by Queen (1975)

Chapter 2

The naked truth: An updated review on nudiviruses and their relationship to bracoviruses and baculoviruses

Jirka M. Petersen [1,2], Annie Bézier [2], Jean-Michel Drezen [2], Monique M. van Oers [1]

[1] Laboratory of Virology, Wageningen University and Research, 6708 PB Wageningen, the Netherlands

[2] Institut de Recherche sur la Biologie de l'Insecte, UMR7261 CNRS - Université de Tours, 37200 Tours, France

Published as:

Petersen JM, Bézier A, Drezen J-M, van Oers MM. The naked truth: An updated review on nudiviruses and their relationship to bracoviruses and baculoviruses. J Invertebr Pathol. 2022; 189: 107718. DOI: 10.1016/j.jip.2022.1077

Abstract

Nudiviruses (*Nudiviridae*) are double-stranded DNA viruses with enveloped and rod-shaped virions. Several insect orders (e.g., Diptera, Lepidoptera, Coleoptera, Orthoptera) and aquatic crustaceans are susceptible to nudivirus infections, which can result in varied degrees of disease in all developmental host stages. Their pathogenicity endangers insect rearing and crustacean aquacultures, but has also proven effective in biocontrol against *Oryctes rhinoceros* infestations. This literature review aims to present all known nudivirus species and provide a comprehensive *Nudiviridae* phylogeny by including recently described nudiviral isolates, and discuss this phylogeny in comparison to current opinions and taxonomical propositions. Moreover, we aim to clarify biological, pathological and genomic differences or similarities between nudiviruses and related entomopathogenic viruses, including baculoviruses (*Baculoviridae*) and bracoviruses (*Polydnaviridae*). A phylogenetic analysis using 17 concatenated nudivirus core genes resulted in the expected structure with the genera *Alphanudivirus* and *Betanudivirus*, as well as the most recently recognised genera *Gammanudivirus* and *Deltanudivirus*. The hymenopteran Osmia cornuta nudivirus (OcNV) groups closest with the hymenopteran Fopius arisanus endogenous nudivirus (FaENV) and does not share a most common ancestor with the hymenopteran bracoviruses. Except for one node, all clades are highly supported. The proposition of a recent study to assign subgroups to the alphanudiviruses might be legitimate, but more hymenopteran and orthopteran nudiviruses, especially in bees and cricket, need to be identified to resolve this proposal. In addition, freshwater and marine nudiviruses might form taxonomic subgroups among gammanudiviruses as well, but more aquatic nudiviruses need to be identified and sequenced for better resolution. Furthermore, the search for nudiviruses in insects with (semi)aquatic life stages may aid in finding the missing link that led to the manifestation of aquatic nudiviruses.

Box 1
Nomenclature

AaBV	Astacus astacus bacilliform virus	**GrBV**	Gammarus roeselii bacilliform virus
AcMNPV	Autographa californica multiple nucleopolyhedrovirus	**HgNV**	Homarus gammarus nudivirus
AdNV	Allomyrina dichotoma nudivirus	**HzNV-1**	Heliothis zea nudivirus-1
AgENV	Aphis glycines endogenous nudivirus	**HzNV-2**	Helicoverpa zea nudivirus 2
ApBV	Austropotomobius pallipes bacilliform virus	**KV**	Kallithea virus
		MdBV	Microplitis demolitor bracovirus
BMN	Baculoviral mid-gut gland necrosis	**MdSGHV**	Musca domestica salivary gland hypertrophy virus
CcBV	Cotesia congregata bracovirus	**MNV**	Mauternbach virus
CcNV	Crangon crangon nudivirus	**MrNV**	Macrobrachium rosenbergii nudivirus
CdBV	Cherax destructor bacilliform virus		
CiBV	Chelonus inanitus bracovirus	**MsENV**	Melanaphis sacchari endogenous nudivirus
CmNV	Carcinus maenas bacilliform nudivirus	**NlENV**	Nilparvata lugens endogenous nudivirus
CpBV	Cancer pagurus bacilliform virus	**OcNV**	Osmia cornuta nudivirus
CqBV	Cherax quadricarinatus bacilliform virus	**OrNV**	Oryctes rhinoceros nudivirus
DhNV	Dikerogammarus haemobaphes nudivirus	**PlBV**	Pacifastacus leniusculus bacilliform virus
DiNV	Drosophila innubila nudivirus	**PmBV**	Pandalus montagui bacilliform virus
DuhNV	Diabrotica undecimpunctata howardi nudivirus	**PmNV**	Penaeus monodon nudivirus
DvvNV	Diabrotica virgifera virgifera nudivirus	**SsBV**	Scylla serrata bacilliform virus
EbrENV	Eurytoma brunniventris endogenous nudivirus	**TNV**	Tomelloso virus
ENV	Esparto virus	**TnBV**	Toxoneuron nigriceps bracovirus
FaENV	Fopius arisanus endogenous nudivirus	**ToNV**	Tipula oleracea nudivirus
GbNV	Gryllus bimaculatus nudivirus	**TpNV**	Tipula paludosa nudivirus
GpSGHV	Glossina pallidipes salivary gland hypertrophy virus	**VcENV**	Venturia canescens endogenous nudivirus

Nudiviruses and their societal and scientific relevance

Members of the virus family *Nudiviridae* can infect hosts from various insect orders (e.g., Diptera, Lepidoptera, Coleoptera, Orthoptera) and aquatic arthropods (e.g., lobster, shrimp, crabs). The virulence of nudiviruses poses a potential threat to a range of arthropod farms and aquacultures. For instance, PmNV causes significant mortalities in tiger shrimp (*Penaeus monodon*), endangering shrimp aquaculture [41]. GbNV causes high mortalities in four field cricket species of the genera *Gryllu*s and *Teleogryllus*, and poses a threat to cricket rearing [42, 43].

On the other hand, the coleopteran nudivirus, OrNV, has been successfully used for biocontrol in pest-infested agricultural regions. In Samoa and other southwest Pacific islands, OrNV was introduced to overcome the devastation produced by the coconut palm beetle (*Oryctes rhinoceros*), a key pest of young palms and coconuts [38]. The use of OrNV regulated and lowered the population of adult *O. rhinoceros* and their larvae [44, 45]. However, as a variety of beetle species finds use in traditional Asian medicine [46], the adverse effect of OrNV on this branch has to be considered when releasing this virus as a biocontrol agent. Japanese rhinoceros beetles (*Allomyrina dichotoma*) that are commercially bred in the Republic of Korea, have been shown to suffer from OrNV infection as well as from a distinct new nudivirus species, namely AdNV. Despite their use for biocontrol, these viruses pose a threat to Korean beetle industry [47, 48].

The abovementioned examples serve to highlight the societal relevance of nudiviruses and emphasise the scientific importance of gaining better insight into the biology and pathogenicity of this diverse virus family. Research on nudiviruses will help to assess their threat for arthropod production systems and to evaluate their potential as novel viral biopesticides to reduce the exposure of harmful chemicals and benefit the environment and agriculture. Additional data will also shed light on their mutual relationships and their evolutionary position in relation to other virus families.

The family *Nudiviridae*: general description, genomic structure and taxonomy

Before their official recognition as an own virus family (*Nudiviridae*), nudiviruses were classified as a subgroup within the family *Baculoviridae.* Baculoviruses are rod-shaped double-stranded DNA (dsDNA) viruses that infect lepidopteran, dipteran and hymenopteran species. Cells that are destined to baculovirus-induced cell lysis accumulate occlusion bodies (OBs) in their host's nuclei. OBs occlude the viral progeny responsible for primary infection, namely occlusion-derived virions (ODVs), into protective protein structures made from polyhedrin or granulin. The systemic infection of baculoviruses is coordinated by a different type of virions, called budded virus (BV) [49, 50]. Contrary to the original assumption that nudiviruses are nonoccluded baculoviruses (hence their designation from Latin "*nudus*" = naked, uncovered), occluded nudiviruses do exist, but were not classified as nudiviruses (OrNV, formerly Oryctes baculovirus) at the time [44] or only discovered and classified later on (e.g., PmNV and ToNV) [40, 41]. Today, members of the family *Nudiviridae* are assigned to the class *Naldaviricetes* together with three other families of nuclear arthropod large DNA viruses: *Baculoviridae*, *Hytrosaviridae* and *Nimaviridae*. However, nudiviruses share the same order, *Lefavirales*, only with baculoviruses and hytrosaviruses [51].

Nudiviral virions consist of single nucleocapsids surrounded by an envelope. They have an ellipsoidal or rod-shaped form of variable length and width (**Supplementary data**, Tables S1 & S2). The variety in length of rod-shaped virions correlates with a general mechanism that adapts the packaging and particle formation to match diverse viral genomes sizes [52]. The dsDNA genome of nudiviruses is a single, covalently closed circular molecule of about 96 to 232 kilobase pairs (kbp) encoding between 87

to 154 proteins depending on the virus. Open reading frame (ORF) content and order can vary greatly among different nudivirus species [53].

Genome sequencing and phylogenetic analyses strengthened the revised view that nudiviruses are closely related to baculoviruses, but form a distinct clade of viruses [53]. The phylogenetic analyses were based on a set of genes that are present in all nudiviral genomes, named core genes. The nudiviral core genes are partly conserved in members of *Baculoviridae, Hytrosaviridae* and bracoviruses (*Polydnaviridae*). A total of 32 core genes are currently assigned to all sequenced nudivirus genomes [53] and their partial conservation in baculoviruses (Table 1) and bracoviruses (Table 2) will be described in the course of this review. Phylogenies inferred from core genes of nudivirus species helped to assign clades within the family *Nudiviridae*. The first officially recognised *Nudiviridae* clades were the genera *Alphanudivirus* and *Betanudivirus* [37], and most recently the International Committee on Taxonomy of Viruses (ICTV) ratified the two additional genera: *Gammanudivirus* and *Deltanudivirus* [51]. Members of *Nudiviridae* are now officially categorised into the four genera *Alphanudivirus*, *Betanudivirus*, *Gammanudivirus* and *Deltanudivirus* (**Figure 2**). An additional genus *Epsilonnudivirus* was recently proposed based on the description of the demon shrimp infecting *Dikerogammarus haemobaphes* nudivirus (DhNV). It was shown that DhNV is most closely related to the gammanudiviruses with a low level of protein similarity at most loci (<50%) [54].

For the purpose of this review, a phylogenetic tree was inferred using the concatenated amino acid sequences of up to 17 aligned core gene products from 20 nudivirus species (**Figure 2**). Nudiviruses with too few available gene sequences (e.g., AdNV, MrNV and Charybdis crab nudivirus) were excluded from the analyses. Although the four nudiviruses ENV, KV, MNV and TNV all share the same host (*D. melanogaster*), they are classified as individual species. These viruses have been named after their different places of origin (ENV = USA: Esparto, CA; KV = Ukraine: Kharkiv; MNV = Austria: Mauternbach, TNV = Spain: Tomelloso).

As an aid for experts, a collection of well-studied nudivirus species with extra information (classification, former names, host, virion morphology, ORFs, genome size) is presented in Supplementary Table S1. Putative nudiviral agents can be found in the Supplementary Table S2.

Nudiviral biology and infection cycle compared to closely related virus clades

Nudiviruses and baculoviruses share two-thirds of their core genes (Table 1), but their biology and infection cycle differ in many aspects [55, 56]. While baculoviruses are able to efficiently infect all tissues of their host, nudiviruses are usually associated with cell type-specific pathogenesis (**Figure 3**). The localised pathology of nudiviruses is likely related to their inability to form budded virions specialised for carrying out secondary infection, as is known from baculoviruses [57].

Upon budding out of the cell, BVs lose the envelope derived from the nuclear membrane to gain a new envelope with the incorporated GP64 or F-protein. Those fusion proteins facilitate the cell entry of BVs via endocytosis during systemic infection [58]. In contrast to BVs, nudiviral virions miss homologs of the envelope fusion proteins GP64 or F-protein [59]. The virion structure of nudiviruses (**Figure 1C**) possibly resembles those of baculovirus ODVs with a single nucleocapsid and infections spread either by living cell egress or after cell lysis [57].

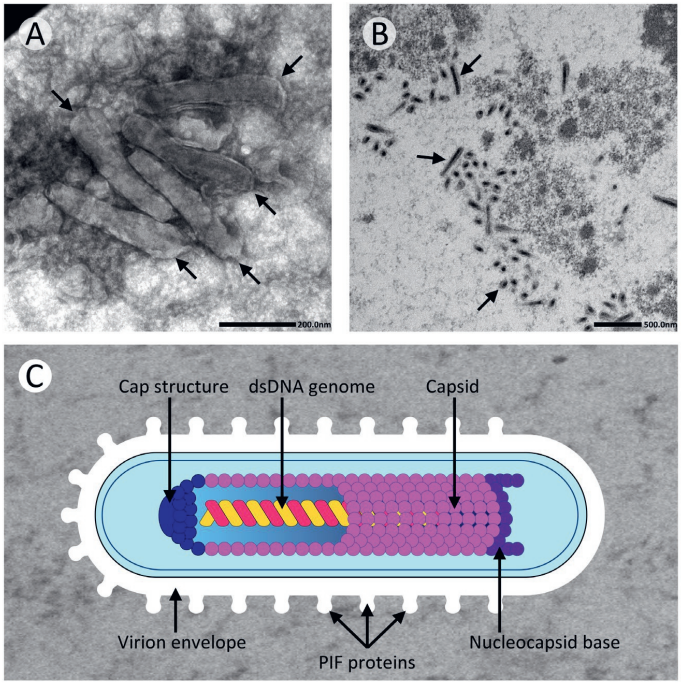


Figure 1. Virion structures of the betanudivirus HzNV-1. **(A)** Electron microscopy (EM) image of rod-shaped HzNV-1 virions negatively stained with uranyl acetate from supernatant of infected HzAM1 cells (72 hours post infection, hpi). Bar, 200 nm, **(B)** EM image of an ultrathin section made from HzNV-1 infected HzAM1 cells (60 hpi) using standard methods. Bar, 500 nm. **(C)** Exemplary illustration of nudiviral virion structure. EM image (B) courtesy Jan van Lent, Wageningen Electron Microscopy Centre (WEMC)

ODV midgut epithelial cell infection or permeation of the peritrophic membrane, or both, are facilitated by a complex of *per os* infectivity (PIF) proteins and other envelope proteins [60, 61], many of which are conserved between baculoviruses and nudiviruses (Table 1). Whether nudivirus virions enter host cells via receptor-mediated fusion, as it has been described for baculoviral ODVs [62], is yet unclear. Interestingly, an early study by Crawford and Sheehan (1985) [30] showed that single OrNV virions enter the cells via macropinocytosis. Pinocytosis is an endocytosis-related cell entry mechanisms [63]. Endocytosis-related cell entry is also known from baculoviral BVs during systemic infection, but instead of pinocytosis BVs enter the host cell via clathrin-mediated endocytosis [58].

Cell egression mechanisms can differ among nudivirus species. For instance, single OrNV virions can egress from an infected cell, but also single or multiple OrNV virions encapsulated in one or multiple membrane vesicles were observed [52]. High quantities of virion-filled vesicles have also been observed in the “waxy plug” of *Helicoverpa zea* during HzNV-2 infection, which is the main substance to spread HzNV-2 infections from one individual to another during mating [64].

A general infection or life cycle model for all nudiviruses has so far not been exemplified, but Velamoor *et al.* (2020) [52] used an electron microscopy approach to describe a unique mechanism of virion assembly and egress during OrNV infection. Although the extrapolation of this model for other members of the *Nudiviridae* should be interpreted with care, these findings highlight valuable directions for future research on other nudivirus species. The addition of genomic and proteomic data is required to extend our knowledge of the exact mechanisms of viral entry, replication, assembly and egress used by nudiviruses.

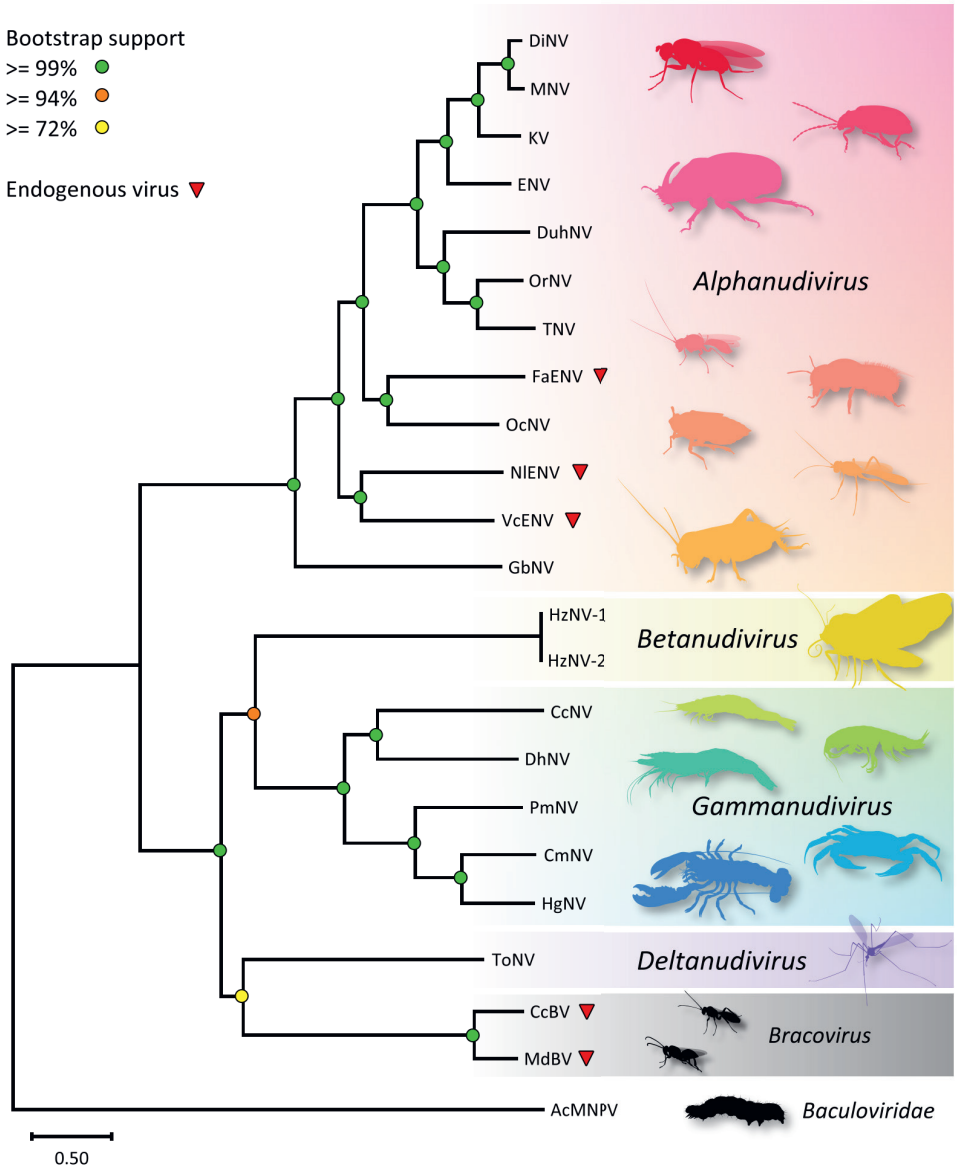


Figure 2. Phylogenetic tree of 17 concatenated nudivirus core gene products from 17 exogenous nudiviruses, three endogenous nudiviruses, two bracoviruses and the baculovirus Autographa californica nucleopolyhedrovirus (AcMNPV) as outgroup. The tree was inferred in MEGAX using maximum likelihood with the WAG+G+F+I model from an alignment of 17 concatenated amino acid sequences. The final dataset had a total of 8545 positions. Except for one node, all clades have bootstrap values of at least 94% after 1000 replicates. Percentage values of bootstrap supports are indicated as coloured circles. The evolutionary time (i.e., 0.50 substitutions *per* sequence site) between two nodes is represented by the branch length. The nudivirus species are grouped into the four officially recognised genera of *Nudiviridae*: *Alphanudivirus*, *Betanudivirus*, *Gammanudivirus* and *Deltanudivirus.* The two bracoviruses represent the *Bracovirus* genus of *Polydnaviridae*. The original phylogenetic tree can be found in the **Supplementary data** of this chapter. Accession numbers of proteins used from each virus can be found in the **Supplementary file 1** (Figure S1) and **Supplementary file 2** (online version).

Nudiviral host genome integration

Nudiviruses have the ability to durably integrate into their host's genome. For instance, genome integration was observed for HzNV-1 under laboratory conditions as part of latent infection [65] and a variety of endogenous nudiviral sequences (NlENV, VcENV, AgENV, MsENV, FaENV, EbrENV) are permanently integrated into insect genomes. Two of those (VcENV and FaENV) can produce virus-like particles (VLPs), while the others (NlENV, AgENV, MsENV and EbrENV) seem to be non-functional or at least unable to produce particles [39, 66-70]. Notwithstanding the distinction between functional and non-functional endogenous nudiviral agents, it suggests that some nudiviruses may have integration mechanisms or life cycle-related features that facilitated those endogenisation events. In addition, it was shown that an ancient nudiviral genome integration in the genetic material of parasitoid wasps gave rise to the endogenous virus clade of bracoviruses of the family *Polydnaviridae*. This family consists of the two genera, *Bracovirus* and *Ichnovirus*. In both genera, exogenous viruses were domesticated by the parasitoid wasps for the production of viral particles. Those are injected into parasitised caterpillars along with wasp eggs and are necessary for successful host parasitism. However, the two genera originated from independent viral integration events that underwent convergent evolution after endogenisation [71, 72]. It is still unknown what ancient virus laid the foundation for ichnoviruses in ichneumonid wasps [73] since conserved ichnovirus genes involved in particle production do not share similarities with currently known virus genes. On the other hand, it was shown that the domestication of an ancestral nudivirus with close relation to the currently known betanudiviruses led to the clade of bracoviruses in braconid wasps [74-76]. The endogenisation of alphanudiviruses in the genomes of braconid [66] and ichneumonid wasps [72] resulted in the manifestation of other hymenopteran endogenous nudiviral agents (FaENV and VcENV). In contrast to exogenous nudiviruses, genes required for DNA replication or particle production are not packaged into bracovirus particles [77] and therefore the production of virions in the parasitised larvae is not provided [74, 78, 79]. Thus, members of *Polydnaviridae* have been associated with the term "viriforms" to distinguish them from true viruses that have replication competent virions [80]. However, viruses are specified by the ICTV as a group of mobile genetic elements (MGEs) that packages their nucleic acid core in virions assembled from at least one virus-encoded component, or MGEs that clearly descend from a virus ancestor [80]. At least for bracoviruses that originate from nudiviruses, the second definition applies. Recent literature suggests to classify braco- and ichnoviruses as Domesticated Endogenous Viruses (DEVs) [81] to emphasise that they retained many features of their virus ancestors.

In hymenopteran ovary cells where bracovirus DNA is amplified, a viral factory is produced to assemble particles in a manner resembling nudivirus replication. Late genes coding for particle components are under the transcriptional control of the nudiviral RNA polymerase [76] as during a baculovirus infection. Once injected into the host, bracovirus particles enter its cells via PIF-mediated entry and express virulence factors to facilitate the parasitism in the lepidopteran host and the development of the wasp eggs. Thus, most of the virus cycle is conserved, but distributed over two hosts allowing the wasp species to reproduce, and the viral entity to be maintained both functionally (i.e. producing infectious particles) and as integrated sequences in the wasp chromosomes [82]. Although at first glance the biology of bracoviruses seems to resemble those of the endogenous nudiviral agents VcENV and FaENV, the latter are unable to package DNA into their particles. Instead, both hymenopteran ENVs wrap virulence proteins made by the wasp into viral envelopes containing PIF proteins (VLPs), similarly produced in a nudivirus-like viral factory in the nucleus (**Figure 3**).

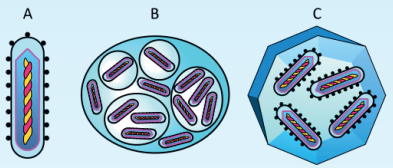
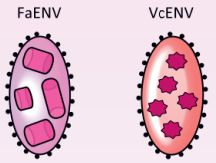
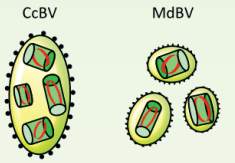
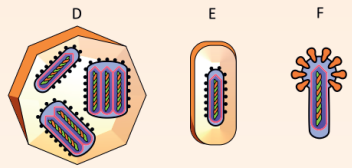
Virus	*Nudiviridae*	Nudiviral VLPs (*Alphanudivirus* ancestor)	*Bracovirus* (*Betanudivirus* ancestor)	*Baculoviridae*
Virus particle properties	 **(A)** Free virions with PIF proteins in envelope coordinate cell entry **(B)** Virion-filled vesicles accumulated in "waxy plug" of HzNV-2 and released from OrNV infected cells **(C)** OBs of ToNV with polyhedrin homolog; OBs of PmNV susceptible to acidic treatment ● No multiply enveloped nucleocapsids	 ● Viral liposomes with PIF proteins incorporated in single bilipid layer ● FaENV particles carry DNA-empty nucleocapsids with unclosed ends ● VcENV particles contain wasp-derived virulence proteins (VLP1-3)	 ● Virions form in wasp's calyx cells ● Bracoviral DNA is integrated into wasp genome ● Replication-defective virions with packaged DNA circles ● Virions can contain single (MdBV) and multiple (CcBV) nucleocapsids	 **(D)** Polyhedrin-based OBs contain numerous ODVs with single (SNPV) or multiple (MNPV) nucleocapsids per envelope **(E)** Granulovirus OBs made from granulin contain one virion with single nucleocapsid per envelope **(F)** BVs with GP64/F protein anchored to envelope coordinate systemic infection
Infection	● Peroral (excrements, cannibalism), sexual or transovarial transmission ● Ingested virions enter midgut epithelial cells via PIF mediated mechanism ● Virions transmitted during mating or transovarial transmission target reproductive tissue cells ● Non-systemic infections manifest in midgut, fat body, hepatopancreatic or reproductive tissue cells ● Occasional integration into host genome	● Parasitoid wasp injects wasp eggs and VLPs into caterpillar ● VcENV encases wasp eggs in immune-protective coating during oviposition ● VcENV proteins can bind to basal lamella of fat body cells or circulate in the hemolymph ● Function of FaENV unknown	● Braconid wasp injects wasp eggs and bracovirus virions into body cavity (hemocoel) of caterpillar ● Virions infect haemocytes, fat bodies and other tissues ● DNA from the virions integrates into host genome and virulence genes are expressed	● Lepidopteran larva ingests OBs from contaminated surface ● ODVs are released in alkaline environment of midgut and PIF proteins initiates entry into midgut epithelial cells ● Infected cells produce BVs to spread infection to all cell tissues ● At the very late stage of infection, ODVs are formed and packaged into OBs
Pathogenic effect	● Varied degrees of disease in larval and adult hosts ● Crustacean hepatopancreatic cells manifest nuclear hypertrophy and hosts can suffer from lethargy; decreased preening, feeding, growth, reproduction and longevity ● Insect hosts can suffer from lethargy and malformations; decreased growth, fertility, longevity; swelling (OrNV, GbNV); production of viscous, opalescent hemolymph (GbNV); "waxy plug" with high quantities of virus at the tip of abdomen (HzNV-2)	● VcENV believed to manipulate lepidopteran host physiology in favor of wasp development ● VLP3, a metalloprotease, was found to adhere to cells and induce the spread of haemocytes	● Expressed virulence genes suppress host immune response and alter host physiology to promote development of wasp offspring	● Viral enzymes (chitinase and cathepsin) facilitate decay of exoskeleton and cell lysis ● Liquefaction of caterpillar releases OBs into environment which eventually infect new hosts ● Baculoviruses with infections limited to the midgut do not encode degrading enzymes. Instead, infected midgut cells are sloughed off and released via excrements

Figure 3. Comparative overview of virus characteristics in the family *Nudiviridae*, nudiviral VLPs, bracoviruses and viruses in the family *Baculoviridae*. Abbreviations: SNPV, single nucleopolyhedrosis virus; MNPV, multiple nucleopolyhedrosis virus. References can be found in Supplementary file 1 (Online version).

Genetic relationship between *Nudiviridae*, *Baculoviridae* and *Hytrosaviridae*

Currently, there are 32 core genes shared between all sequenced nudivirus species , 21 are homologs to baculovirus (Table 1) and 16 to hytrosavirus genes. Due to the ongoing discovery and full sequencing of new nudivirus genomes, the number of nudivirus core genes might need adaptation in the future. For example, the recently sequenced CcNV lacks the three thymidine kinase (*tk*) genes, which would bring the total number of core genes back to 29 [83].

Table 1. The 32 core genes present in all nudivirus genomes distributed over functional groups. The collection of the core genes is based on coinciding results of different publications *[67, 70, 84]*. None of the three tk genes are present in CcNV *[83]*.*

Transcription	Infectivity	Packaging, assembly, morphogenesis	DNA replication, repair, recombination	Nucleotide metabolism	Unknown function
lef-4	***pif-0 (p74)***	***vp91 (pif-8)***	***dnapol***	*tk1*	*GbNV gp19-like*
lef-5	***pif-1***	***38K***	***helicase***	*tk2*	*GbNV gp51-like*
lef-8	***pif-2***	***p33 (ac92)***	*helicase-2*	*tk3*	*GbNV gp58-like*
lef-9	***pif-3***	***p6.9***	*integrase*		*GbNV gp67-like*
p47	***pif-4 (19 kDa)***	***vlf-1***	*fen-1*		*11K-like*
	pif-5 (odv-e56)	***vp39***			
	pif-6 (ac68)	***ac81***			

*Abbreviations: *lef*, late expression factor; *pif, per os* infectivity factor; *vlf,* very late expression factor; *dnapol*, DNA polymerase; *fen*, FLAP endonuclease; *tk*, thymidine kinase. Genes in bold indicate core genes shared with the *Baculoviridae*. Alternative gene names are in brackets.

In addition to the 32 core genes, accessory genes, such as DNA *ligase*, the ribonucleotide reductases *rr1* and *rr2*, *GbNV gp74, odv-e66* and inhibitor of apoptosis protein (*iap-3*), are commonly found in nudiviruses or only present in some species [40, 41, 55]. The entirety of the nudiviral core genes and accessory genes make up the nudiviral pangenome.

Partial conservation of genes between nudiviruses, baculoviruses and hytrosaviruses indicates similarities in their biology. For example, all five of the genes in the transcriptional group (*lef-4, lef-5, lef-8, lef-9, p47*) are conserved between the nudiviruses (including the domesticated VcENV and FaENV), baculoviruses and bracoviruses, which suggests that these viruses share a similar mode of late gene transcription [41, 72, 84, 85]. The same applies to the eight conserved PIF factors (*pif-0*/*p74*, *pif-1*, *pif-2*, *pif-3*, *pif-4*/*19kDa*, *pif-5*/*odv-e56*, *pif-6/ac68, pif-8/vp91*) in nudiviruses, baculoviruses,

bracoviruses, VcENV, but not FaENV which lacks the *pif-5* homolog. PIF proteins are all essential for the infectivity of baculoviral ODVs and believed to play a similar role for nudivirus and bracovirus virions as well as VcENV particles [76]. Neither PIF-4 nor PIF-6 have been identified as components of FaENV particles yet [66]. In baculoviral ODVs they facilitate specific binding or fusion to the midgut cells (P74, PIF-1 and PIF-2, PIF-3) [84, 86-88] stabilise the ODV-entry core-complex to resist proteolytic activity (PIF-4 and PIF-6) [89] or contribute to oral infection by a mechanism that has not been determined yet (PIF-5 and PIF-8) [90]. Despite the absence of OBs in bracoviruses and most nudiviruses, the presence of conserved *per os* infectivity factors implies a similar mechanism of cell entry. Four homologous genes with relatively high identity and similarity (*pif-0*/*p74*, *pif-1*, *pif-2*, *pif-3 and pif-5/odv-e56*) are also present in members of the family *Hytrosaviridae,* infecting Dipteran species [91]. The conservation of those proteins between baculoviruses, nudiviruses, and hytrosaviruses supports these virus families share a similar virus infectivity model [41].

The core genes *p33*, *vp91*, *vlf-1*, *38K, p6.9, vp39* and *ac81* are involved in viral particle formation. VP39 (formerly annotated as 31k-like protein in PmNV) [92] and 38K are major viral capsid proteins in baculoviruses, nudiviruses and in the bracovirus CiBV [59, 93, 94] P33 is described as an essential component of BVs and ODVs of baculoviruses [95]. VP91/PIF-8 (also) associates with the nucleocapsid and envelope, but is majorly involved in baculovirus primary infection [90]. The condensation of the viral genome into the nucleocapsid and formation of infectious virions is facilitated by the P6.9 protein [96], while the processing of the viral DNA to nucleocapsid length is coordinated by the lambda family integrase VLF-1 [97].

The protein sequences of DNA polymerase and helicase are more diverse among the nudivirus species than other core genes, which is consistent with previous findings that described *helicase* as one of the most rapidly evolving genes across baculoviruses and nudiviruses [98]. In previous studies, genetic mapping demonstrated that the sequence variation in the helicase sequence supports variations in host range and host swapping in baculoviruses [99, 100]. Phylogenetic analyses have shown greater evolutionary distances for both *dnapol* and *helicase* among the nudiviruses than among baculovirus species [41]. The *helicase-2* gene is scarcely present in baculovirus genomes [101], but is a core gene in nudiviruses and was found in both sequenced hytrosaviruses [102].

Besides the genes that closely interact with the viral DNA, there is another particularity when it comes to viral non-coding DNA sequences involved in DNA replication. Homologous regions (*hrs*) in baculoviruses have a palindromic structure and function as origins of viral DNA replication [103] and support viral transcription [104]. Despite the absence of *hrs* in nudivirus genomes, most nudiviruses contain tandem direct repeat sequences (*drs*), which are proposed of being associated with viral replication [92]. The presence of *drs* is a common feature that members of *Nudiviridae*, *Baculoviridae, Hytrosaviridae* and bracoviruses share [91, 105].

Genomic similarities between nudiviruses and bracoviruses

The bracoviral *drs*, called direct repeat junctions (DRJs) or wasp integration motifs (WIMs), are localised at the ends of the proviral segments in the wasp genome. Proviral segments make up the part of the bracoviral genome that is packaged into bracovirus virions. The proviral segments may be amplified separately or together (when contiguous in the wasp genome) within replication units (RUs terminally flanked by replication unit motifs (RUMs) [106]. The proviral segments are composed of sequences homologous to wasp genes [107] as well as transposable elements and genes of unknown origin [82, 108]. In addition to the proviral segments, the bracovirus genome consists of a second major region that presumably reflects the genome of the ancestral nudivirus that integrated into the genome

of an ancestor wasp, the nudiviral cluster [74]. The genes of the nudiviral cluster encode several of the structural proteins required for virion production, whereas the others are dispersed in the wasp genome. In contrast to the proviral segments, no RUMs were found in the nudiviral cluster, indicating a different mechanism for the amplification of genes in this genomic region [82, 109]. A study on Microplitis demolitor bracovirus (MdBV) revealed that recombination between the two flanking WIMs involves two nudiviral integrases [76] and leads to the circularisation of the DNA segments packaged into bracovirus nucleocapsids [77]. Genes of the *integrase/recombinase* superfamily that are conserved between all known bracoviruses and nudiviruses include *vlf-1* and *integrase*, whereas *HzNVorf140-like* is present in bracoviruses, but not in all nudiviruses [79]. Even though the nudiviral cluster is amplified, it is not encapsidated into the nucleocapsids like the proviral segments. Sequence analyses of Cotesia congregata bracovirus (CcBV) revealed that all RUs of the proviral segments and nudiviral cluster comprise a conserved TA-rich sequence motif. The conserved motif was shown to have variable length (up to 100 bp) and has hairpin-forming attributes. Hairpins are secondary-structures that are associated to genome replication and can function as origins of replication . Further studies on CcBV associated the hairpin-forming regions with the generation of concatemeric intermediates. They revealed unexpectedly that a CcBV locus forms head-to-head/tail-to-tail concatemers [106] as predicted by a model of linear replication as described in *Poxviridae*, while a more extensive study showed that both the later and head-to-tail concatemers characteristic of rolling circle replication as described in *Baculoviridae* are formed depending on the locus in MdBV [109]. The similarity between sequence motifs flanking RU junction sites whatever the replication intermediates, suggests a conservation of the mechanism involved to produce two concatemer types [82, 109]. How exactly those concatemers are produced has yet to be fully understood.

The wasp machinery is thought to amplify bracovirus DNA sequences, while the nudiviral machinery appears to be mostly associated with DNA processing and virion production [74, 79, 110]. This division of tasks is coherent with the absence of the nudiviral DNA polymerase in all bracovirus species (Table 2). This table indicates the nudiviral core genes present in bracovirus species (CcBV, MdBV and CiBV) [70, 82] and endogenous nudiviral agents from hymenopteran hosts (VcENV and FaENV) [66, 72]. The 14 nudiviral core genes (or pseudogenes) present in a presumably inactive nudivirus in the genome of the chalcid wasp *Eurytoma brunniventris* (EbrENV-β) [70] were also included in the comparison. Phylogenetic analysis indicated that EbrENV-β was derived from a betanudivirus and is the closest known relative to bracoviruses [70]

Table 2. Distribution of nudivirus core genes in bracovirus species, hymenopteran endogenous nudiviral agents and EbrENV. Genes present in the virus genome are indicated with filled circles "●". The presence of pseudogenes is indicated with circles in a square "◙". Open circles "○" indicate no detection of the core genes. No data on the presence or absence of genes, or pseudogenes, are indicated with empty areas. Core gene functions are labelled on the left. The table is based on the findings from [70] and was complemented with studies from [66, 68, 72, 82].
See next page.

	Gene	Bracoviruses			Endogenous nudiviruses		EbrENV
		CcBV	CiBV	MdBV	VcENV	FaENV	EbrENV-β
Transcription	*lef-4*	●	●	●	●	●	●
	lef-5	●		●	●	●	
	lef-8	●	●	●	●	●	
	lef-9	●		●	●	●	●
	p47	●		●	●	●	●
Infectivity	*pif-0 (p74)*	●	●	●	●	●	
	pif-1	●	●	●	●	●	◙
	pif-2	●	●	●	●	●	●
	pif-3	●		●	●	●	◙
	pif-4 (19kDa)	●	●	●	●	●	◙
	pif-5 (odv-e56)	●	●	●	●	○	◙
	pif-6 (ac68)	●		●	●	●	
Packaging, assembly, morphogenesis	*vp91 (pif-8)*	●	●	●	●	●	◙
	38K	●	●	●	◙	●	●
	p33 (ac92)	●		●	●	●	
	p6.9	●			◙		
	vlf-1	●	●	●	◙	○	
	vp39	●	●	●	◙	●	
	ac81	○		○	●	●	●
DNA replication, repair, recombination	*dnapol*	○		○	◙		
	fen-1	●		●	◙		●
	helicase	●		●	●	●	●
	helicase-2	○		○	○		●
	integrase	●	●	●	◙		
Nucleotide metabolism	*tk1*						
	tk2						
	tk3						
Unknown function	*GbNV gp19-like*	●		●	●	●	
	GbNV gp51-like	○		○	●	●	
	GbNV gp58-like	○		○	●	●	
	GbNV gp67-like	●			●	●	
	11K-like	●	●	●	●	○	

The presence of the above listed genes in endogenised nudiviruses does not necessarily represent their functionality or activity, as some of these genes may have obtained alternative, yet undescribed functions, or lost their functionality totally after endogenisation [111]. However, *Venturia canescens* genome analysis showed that nudiviral genes without beneficial function for the parasitism tend to become ultimately pseudogenised, such as the nudiviral DNA polymerase and genes coding for nucleocapsid components [68].

Discussion

Phylogeny and taxonomic classification

The family *Nudiviridae* comprises a rich diversity of species and their taxonomic classification is an ongoing process. Deep sequencing approaches have gradually revealed new nudiviral agents in arthropods and the full sequencing of their genomes has allowed more profound phylogenetic analyses. There are currently four genera among the family *Nudiviridae* that the ICTV officially recognised: *Alphanudivirus*, *Betanudivirus*, *Gammanudivirus*, and *Deltanudivirus* [51]. Recently discovered nudiviral agents (AgENV, MsENV, DuhNV) demonstrate a high diversity of viruses in terrestrial hosts in the genus *Alphanudivirus.* In addition, the ongoing exploration of nudiviruses in aquatic arthropods is gradually providing higher resolution of intrafamilial relationships within the *Nudiviridae.* The official recognition of two new genera, *Gammanudivirus* and *Deltanudivirus*, corresponds to the phylogenetic analyses of 15 terrestrial and 5 aquatic nudiviruses (**Figure 2**) and emphasises their evolutionary distinction. Previous phylogenetic analyses first grouped PmNV and ToNV together with the genus *Betanudivirus* [40], then later assigned ToNV and CcNV to the genus *Deltanudivirus* [112], but the inclusion of more sequences and aquatic nudivirus species in the analysis invalidated these propositions. Instead, HzNV-1 and -2 now represent the only *Betanudivirus* members and ToNV makes up the genus *Deltanudivirus*, while PmNV and CcNV group with the other aquatic nudiviruses (DhNV, CmNV, HgNV) in the genus *Gammanudivirus*.

Due to the growing diversity of alphanudiviruses, Liu *et al.* (2021) [113] proposed to upgrade this genus to the subfamily Alphanudivirinae, with three genera (Grynudivirus, Orynudivirus and Endonudivirus) due to the diverged lineages. *Grynudivirus* would currently only consist of GbNV that infects the field cricket *G. bimaculatus*, while *Orynudivirus* features nudiviruses that infect flies (*D. melanogaster*, *D. inubila*) and beetles (*O. rhinoceros*, *D. u. howardi*). The genus Endonudivirus would include the endogenous nudiviruses from aphids (AgENV and MsENV). The distinct lineage of cricket-infecting nudiviruses likely reflects the diversification of orthopteran species at the end of the Devonian period (~370 million years ago), while the common ancestor of dipteran, lepidopteran and coleopteran species manifested later between the Carboniferous and Permian period (~300 million years ago) [114] and with this ancestral host the corresponding nudiviruses of the proposed Orynudivirus genus. Therefore, the placement of cricket-infecting nudiviruses in the proposed genus Grynudivirus, next to the flies- and beetle-infecting nudiviruses in the genus Orynudivirus, appears reasonable. Nevertheless, the discovery and inclusion of more genomic data from orthopteran nudiviruses, with special regard to cricket-infecting nudiviruses, is needed to carefully evaluate this proposition.

OcNV, the only described bee (*O. cornuta;* the European orchard bee, Megachilidae)-infecting nudivirus was not included in the analysis by Liu *et al*. (2020) [69]. Our phylogenetic analysis (**Figure 2**) indicates that OcNV is an alphanudivirus that shares a most common ancestor with the endogenous nudivirus found in the hymenopteran host *Fopius arisanus*. On the other hand, OcNV is not closely related to the endogenous bracoviruses even though they both have hymenopteran hosts. Conclusively, the bee-infecting nudivirus appears to originate from an ancestral alphanudivirus, while bracoviruses share a common ancestor with the dipteran ToNV in the deltanudivirus branch, although with a rather low node support (72%) when compared to the other nodes.

ToNV, HzNV, crustacean nudiviruses and bracoviruses all seem to have diverged from a common ancestor and form a cluster distinct from the members in the proposed subfamily *Alphanudivirinae*. Therefore, Liu *et al.* (2021) [113] proposed to group the betanudiviruses in the parallel subfamily *Betanudivirinae*. This proposition is further supported by the low number of homologous genes with high similarity shared between alphanudivirins and betanudivirins [113]. An evolutionary connection between HzNV and bracoviruses is plausible, as their hosts live in a close biological relationship. A possible scenario could involve a lepidopteran host that was infected with an ancestral betanudivirus and was subsequently parasitised by an ancestor wasp. The hatched larvae of this wasp could have fed on the nudivirus-infected caterpillar and consequently ingested the betanudivirus. This event could have led to the endogenisation of the nudivirus in the genome of the wasp and ultimately in the domestication and further evolution into bracoviruses. Alternatively, an ancestral parasitoid-wasp-infecting nudivirus with tissue-specific expression in the ovaries could have been transmitted to a lepidopteran host and then manifested itself in its reproductive tissues (i.e. HzNV-2). As indicated above, the only other hymenopteran-infecting nudivirus (OcNV) known to date shows a great evolutionary distance to the betanudivirus and bracovirus branch, making the latter hypothesis less likely. However, the fact that nudiviruses of distinct genera manifested in different hymenopteran hosts, indicates that alpha- and presumably beta- and deltanudiviruses have qualities that allow infection/adaptation to hymenopteran hosts. This, in turn, leads to the question whether nudiviruses diversified along with their hosts like it was shown for baculoviruses [115], or whether an innate ability to adapt to more distantly related hosts led to the broad host range of the *Nudiviridae* as a family.

In relation to the proposed genus Endonudivirus [113], it is debatable whether all endogenous nudiviruses should be grouped into one clade. Previous phylogenetic analyses showed that endogenous nudiviral agents in hemipteran species (e.g., MsENV and AgENV) are evolutionary distinct from the ones found in hymenopteran species (FaENV and VcENV) [67, 69]. Endogenous nudiviruses from distantly related hosts are probably phylogenetically dispersed and possibly group close with their free-living relatives. This is supported by a phylogenetic tree of Cheng *et al*., (2020) [67] inferred from the P74 protein sequences of several nudiviruses and endogenous nudiviral elements found in arthropod genomes. While seven nudiviral elements of hemipteran origin share a distinct phylogenetic group, the endogenised nudivirus of the lepidopteran species *Danaus plexippus* groups with the lepidopteran virus, HzNV-1.

Among the aquatic nudiviruses, CcNV, PmNV, CmNV and PmNV infect marine crustaceans, while DhNV infects a freshwater host. It can be assumed that the nudiviruses in marine hosts are distinct to the ones in freshwater hosts and underwent separate evolutionary events at different time periods. The low average protein similarity and disparate gene synteny of DhNV compared to the described marine nudiviruses has been cited to support classification of freshwater nudiviruses as a distinct genus, proposed as Epsilonnudivirus *[54]*. The freshwater nudivirus MrNV (NCBI accession: PRJNA359633) purified from diseased giant freshwater prawn (*Macrobrachium rosenbergii*) larvae was shown to group between DhNV and PmNV when using its *iap* and *pif-2* genes for phylogenetic analysis [54]. For a more accurate phylogeny of freshwater nudiviruses, the sequences of more MrNV genes are required, as well as genomic data of other putative freshwater nudiviruses, such as AaBV, ApBV, CdBV, CqBV, GrBV and PlBV. Also, the full sequences of putative marine nudivirus genomes, including BMN, Baculo-PP, CpBV, PmBV and SsBV, will add higher resolution to the phylogeny of aquatic nudiviruses. Next to marine and freshwater habitats, it may be worthwhile to look for potential nudiviruses in brackish crustaceans as well.

Particularly interesting in this regard is the yet unknown origin of aquatic nudiviruses. It is likely that close ecological interactions between (semi)aquatic insects and crustaceans may have allowed the switch of an ancestral nudivirus to either of those host groups. A number of Dipteran families including among others, Chironomidae, Culicidae, Simuliidae and Tipulidae, are terrestrial as adults, but their larvae and pupae are aquatic [116]. Crane fly larvae (Tipulidae) have been shown to inhabit both land

and aquatic areas [117], such as marine, brackish and non-saline waters [118]. Some crane fly larvae are predatory and feed on other aquatic insects and invertebrates [119], while bigger aquatic invertebrates may naturally consume crane fly or dipteran larvae as well. Other aquatic insect larvae exist in the family of Dytiscidae, a taxonomic clade of water beetles. Members of this family are referred to as predaceous diving beetles and can be found worldwide in freshwater [120], but some species also live in brackish waters [121]. Predaceous diving beetles might form a possible bridge between insects and brackish or freshwater nudiviruses in Crustaceans, whereas crane flies could have been a driving force that led to the manifestation of nudiviruses in aquatic hosts as such. When one applies the alternative theory of nudiviral host-driven diversification, it is likely that an ancestor of all nudiviruses may have its roots in the water. The last common ancestor of hexapods and crustaceans belonged assumably to the class of Branchiopoda and lived in freshwater around 420 million years ago in the Late Silurian [122]. Therefore, it is likely that the colonisation of aquatic hexapods to terrestrial habitats might have pressured the nudiviral ancestor to adapt accordingly and laid the foundation that established the entomopathogenic group of nudiviruses that we know today. However, it remains to discover and characterise more nudiviruses in aquatic or semiaquatic invertebrates (and beyond) to refine and unravel the evolutionary history of *Nudiviridae*.

Morphological and genomic properties

The discovery and identification of nudiviruses that form OBs (ToNV, PmNV and OrNV under facultative conditions) refutes the original observation that led to the naming of this “naked” virus family. Therefore, the absence of OBs should no longer be the main criterion to discriminate between nudiviruses and baculoviruses. Instead, the localised infection and wider host range of nudiviruses as a family and the specific collection of core genes appears to be a more discriminating feature that distinguishes them from baculoviruses. The number of nudiviral core genes will most likely require revision in the future, as more nudivirus species are being discovered.

The greater evolutionary distances of the nudiviral DNA polymerase and helicase compared to their baculovirus orthologues implies that the host range within the nudivirus family is greater than the host range among baculovirus species. Indeed, this is consistent with previous findings that described the helicase gene as one of the most rapidly evolving genes across baculoviruses and nudiviruses [123]. The sequence variation of the *helicase* gene was further shown to correspond with the ability of baculoviruses to switch to new hosts [99, 100]. The corresponding broader distribution of nudiviruses over more distantly related arthropod hosts may reflect a greater evolutionary diversification that nudiviruses underwent to adjust to the manifestation of new insect orders in the course of millions of years. The nudiviral helicase is likely to facilitate a crucial function in nudivirus replication, as it has been proposed for poxviruses. It is believed that the poxviral helicase initiates DNA replication by inducing a nick in the linear dsDNA to expose a free 3’-OH group that functions as a primer for the DNA polymerase [124]. A similar role of the nudiviral helicase in DNA replication might be expected. The scarce presence of the nudiviral core gene *helicase-2* in baculoviruses might hint at a replication mechanism that further distinguishes the two virus families.

The ability of nudiviruses to maintain their replication and particle formation after host genome integration is another particularity that distinguishes them from baculoviruses. This nudivirus feature led to the formation of nudivirus-derived endogenous viral elements in genomes of arthropods and gave rise to endogenous nudiviral agents (NlENV, AgENV, MsENV, FaENV, VcENV, EbrENV) and bracoviruses. Moreover, HzNV-1 can occasionally integrate its genetic material into the host genome as part of latent infection, whereas most genome copies of HzNV-1 are suggested to persist as episomes. latent HzNV-1 infections may be reactivated for productive and lytic infection [65], but to which extent this reactivation is bound to integrated or episomal viral DNA has not been examined yet. Two differences might hint at the possibility that nudiviruses can maintain replication activity in their host genome, while baculoviruses presumably cannot: differences in the structure of nudiviral and baculoviral origins of replication (i.e. *hrs* or *drs*), and the absence of the nudivirus core gene *integrase*

in baculovirus genomes (or even the multiplicity of coding genes with integrase/recombinase functions). On the other hand, these aspects are shared between nudiviruses and bracoviruses. In bracoviruses, *drs* are loci (also called WIMs or DRJs) where recombination and excision events occur. While the nudiviral integrase probably ensures the excision of the DNA destined for circularisation from previously formed concatemeric intermediates, it can be assumed that other nudiviral genes from the integrase/recombinase superfamily (*vlf-1*, *integrase and HzNVorf140-like*) are involved in the integration of the packaged DNA circles into the genome of infected lepidopteran cells [125]. The formation of head-to-head/tail-to-tail concatemers during bracovirus replication leads to the question whether this is a trait that bracoviruses "inherited" from their nudiviral ancestor, or if this replication model newly evolved within the clade of bracoviruses. A unique sequence in the genome of HzNV-1, comparable to conserved direct repeat motifs at the RU boundary junction of bracoviruses, may play a similar role in HzNV-1 DNA replication. It is hypothesised that the *drs* originate from the duplication of a single nudivirus sequence that manifested as multiple loci over the wasp genome allowing the resolution of viral concatemers into individual genomes. The bracoviral VLF-1 and nudivirus-like integrase (INT-1) are both tyrosine recombinases that potentially bind and interact with the *drs* recombination sites. Possible *drs*-regulated mechanisms that might be associated with nudivirus replication and/or host genome integration have yet to be understood.

The conservation of *drs* and genes involved in DNA processing (*vlf-1*, *integrase and HzNVorf140-like*) between nudiviruses and bracoviruses hint at shared properties to maintain their replication mechanisms and to form virus particles after integration into their host's genome. The absence of both an integrase homolog and *drs* from baculovirus genomes may be part of the explanation why no replication competent endogenous baculovirus have been described, despite the fact that endogenised baculovirus genes were found in host genomes [126].

Infection mechanisms

Besides the uncertainties regarding the nudiviral pangenome and replication mechanisms, there are other gaps to be filled when it comes to the infection mechanisms of nudiviruses. The nuclear exit of nudiviruses involves passing a double membrane while retaining the viral envelope to keep PIF and other envelope proteins needed for the infection of subsequent cells [52]. Therefore, it can be assumed that the virion-filled vesicles observed in HzNV-2 and OrNV serve as a viral-envelope-preservation mechanism to make up for the lack of GP64 or F-protein induced endocytosis. OrNV seems to circumvent this hurdle by encapsulating its virions in multiple membrane vesicles (MMVs) that egress from the outer nuclear membrane. The utilisation of multiple vesicles may enable sequential cell-to-cell transmission of OrNV virions in exchange for a vesicle membrane each time it fuses with the plasma membrane of a different cell [30]. This may lead to dissemination into deeper tissue layers without the need for intermediate replication. The virion-filled vesicles of HzNV-2 may play a major role in horizontal transmission as infectious agents. In addition to OrNV virions that are encapsulated in MMVs, vesicle-free virions were also observed to egress from infected cells. The cell entry mechanisms of OrNV, and presumably other nudiviruses, might differ between vesicle-free virions and virion-filled vesicles. However, it has yet to be uncovered if these vesicle-free virions lose their envelopes as part of a membrane fusion and only release their nucleocapsids into the extracellular space or have a special mechanism of maintaining their viral envelope.

Moreover, the exact mechanism behind nudiviral nucleus entry still needs to be unravelled. For baculoviruses, different mechanisms have been observed. Several studies describe a docking process of nucleocapsids, followed by their entry into the nucleus through the nuclear pore complex [127, 128], while another study observed the release of the viral genome into the nucleus with empty nucleocapsids remaining on the cytoplasmic side [129]. Given the fact that nuclear pores can have a diameter of 38 – 78 nm [130] and the diameter of baculoviral nucleocapsids ranges from 30 – 60 nm [131], it is reasonably assumed that their entry through these pores is attainable [101]. The diameter of nudivirus nucleocapsids ranges from 30 nm (DiNV) to ~ 80 nm (HzNV) and it is known that the nucleocapsids

of AcMNPV are able to transit nuclear pores, though via actin polymerisation [128]. Moreover, virally-induced modification of nuclear pores can occur to allow the entry of oversized nucleocapsids into the nucleus, as it was shown for a system with AcMPNV and *Xenopus* oocytes [132]. Conclusively, most nudiviral nucleocapsids should be capable of entering the nucleus through nuclear pores as well. Whether nudiviruses are also able to reorganise the structure of nuclear pores to suit their nucleocapsids dimensions has not been described yet.

Conclusion

The intrafamilial phylogeny of nudiviruses is obtaining higher resolution as more nudivirus species are discovered and their genomes sequenced. In particular this includes aquatic bacilliform viruses that were originally classified as baculoviruses, but are now reclassified as nudiviruses. Additionally, the ever-growing diversity of alphanudiviruses may lead to a revision of the genus *Alphanudivirus* in the future. Upcoming studies should focus on discovering and sequencing the genomes of potential aquatic nudiviruses, also in regard to marine, brackish and non-saline habitats. Moreover, the search for more nudiviruses in orthopteran and hymenopteran hosts, including bee and cricket species, will provide more profound phylogenetic insight on the family of *Nudiviridae*. The same applies for yet undescribed endogenised forms in insect and crustacean genomes to retrace the evolutionary history of ancient viruses. The genome sequencing of new nudiviruses will further aid in understanding the nudiviral pangenome and eventually revise their set of core genes.

Acknowledgements

We wish to thank the workgroup of Prof. Dr. Dirck C. de Graaf for providing the transcriptomic data of the Osmia cornuta nudivirus (OcNV). We also want to thank Dr. Ronny van Aerle for his valuable remarks and support on the phylogenetic analyses. The writing of this paper has received funding from the European Union's Horizon 2020 research and innovation programme INSECT DOCTORS under the Marie Skłodowska-Curie grant agreement No 859850.

Supplementary data

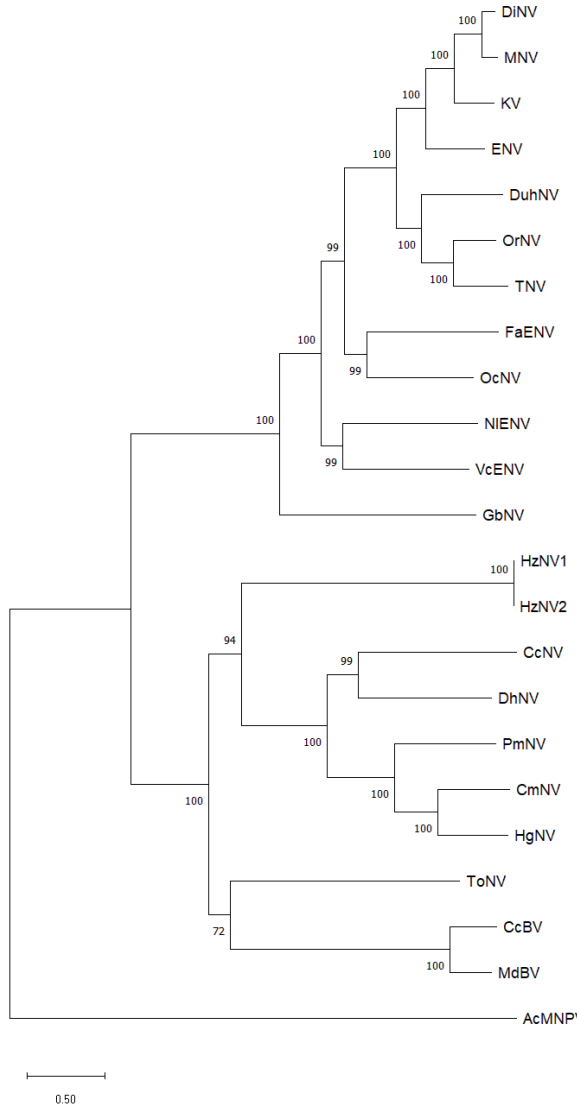


Figure S1. Phylogenetic tree of 17 concatenated nudivirus core gene products from 17 exogenous nudiviruses, 3 endogenous nudiviruses and two bracoviruses with the baculovirus Autographa californica nucleopolyhedrovirus (AcMNPV) as outgroup. The tree was inferred in MEGAX using maximum likelihood with the WAG+G+F+I model from an alignment of 17 concatenated amino acid sequences. The final dataset had a total of 8545 positions. This original tree forms the basis for the modified phylogenetic tree seen in Figure 2.

Table S1. Fully sequenced nudiviruses. Exogenous members of *Nudiviridae* that are classified and have mostly sequenced genomes.

Virus name	Former names	Genus	Host	Virion morphology	ORFs	Length in bp (# accession)	Ref.
DiNV, Drosophila innubila nudivirus	-	*Alpha-nudivirus*	Vinegar fly, *D. innubila* (Diptera: Drosophilidae)	120 x 30 nm; nonoccluded	107	155,555 (NC_040699.1)	[98, 133]
KV, Kallithea virus	-	*Alpha-nudivirus*	Common fruit fly, *D. melano-gaster* (Diptera: Drosophilidae)	200 nm x 50 nm; nonoccluded	95	152,388 (NC_033829.1)	[98, 133-135]
ENV, Esparto virus	-	*Alpha-nudivirus*	Common fruit fly, *D. melano-gaster* (Diptera: Drosophilidae)	Not described	87	183,261 (NC_040536.1)	[98, 133-137]
MNV, Mauternbach virus	-	*Alpha-nudivirus*	Common fruit fly, *D. melano-gaster* (Diptera: Drosophilidae)	Not described	95	154,465 (MG969167.1)	[98, 133, 137]
TNV, Tomelloso virus	-	*Alpha-nudivirus*	Common fruit fly, *D. melano-gaster* (Diptera: Drosophilidae)	Not described	93	112,307 (NC_040789.1)	[137]
GbNV, Gryllus bimaculatus nudivirus	Cricket baculovirus	*Alpha-nudivirus*	Two-spotted cricket, *G. bimaculatus* (Orthoptera: Gryllidae)	162 x 66 nm; nonoccluded	98	96,944 (NC_009240.1)	[28, 43, 59]
OrNV, Oryctes rhinoceros nudivirus	OrBV, Oryctes baculovirus; OrV, Oryctes virus; ROV, Rhabdionvirus oryctes	*Alpha-nudivirus*	Coconut rhinoceros beetle, *Oryctes rhinoceros* (Coleoptera: Scarabaeidae)	180 x 65 nm; facultatively occluded	139	127,615 (NC_011588.1)	[30, 43, 44, 138, 139]
DuhNV, Diabrotica undecimpunctata howardi nudivirus	-	*Alpha-nudivirus*	Spotted cucumber beetle, *Diabrotica unidecim-punctata* (Coleoptera: Chrysomelidae)	~230 nm x 52 nm, nonoccluded	109	147,179 (MW503925.1 to MW503929.1)	[113, 140]
HzNV-1, Heliothis zea nudivirus 1	Hz-1V, Hz-1 virus ; HzV-1; Hz-1 baculovirus; IMC-Hz-I-NOV; baculovirus X	*Beta-nudivirus*	Corn earworm, *Helicoverpa zea* (Lepidoptera: Noctuidae)	414 ± 30 x 80 ± 3 nm; nonoccluded	154	228,089 (AF451898.1)	[29, 141-143]

HzNV-2, Helicoverpa zea nudivirus 2	GSV, gonad-specific virus; Hz-2V	*Beta-nudivirus*	Corn earworm, *Helicoverpa zea* (Lepidoptera: Noctuidae)	382 ± 30 x 77 ± 2 nm; nonoccluded	113	231,621 (NC_004156.2)	[144, 145]
HgNV, Homarus gammarus nudivirus	-	*Gamma-nudivirus*	European lobster, *Homarus gammarus* (Decapoda: Nephropidae)	154 ± 20 x 36 ± 4 nm; nonoccluded	97	107,063 (MK439999.1)	[92]
PmNV, Penaeus monodon nudivirus	MBV, monodon baculovirus; PemoNPV, Penaeus monodon polyhedrovirus; PmSNPV, Penaeus monodon singly enveloped nuclear polyhedrosis virus	*Gamma-nudivirus*	Black tiger shrimp, *Penaeus monodon* (Decapoda: Penaidae)	246 ± 15 x 42 ± 3 nm; occluded	115	119,638 (NC_024692.1)	[41, 146-148]
CmNV, Carcinus maenas nudivirus	CmBV, Carcinus maenas bacilliform virus	*Gamma-nudivirus*	European green crab, *Carcinus maenas* (Decapoda: Portunidae)	~210 nm x 47 nm; nonoccluded	98	113,840 (MZ311577)	[83, 149, 150]
CcNV, Crangon crangon nudivirus	CcBV, Crangon crangon bacilliform virus	*Gamma-nudivirus*	Brown shrimp, *Crangon crangon* (Decapoda: Crangonidae)	215 x 72 nm; nonoccluded	105	132,068 (MZ311578)	[83, 112, 151]
DhNV, Dikerogammarus haemobaphes nudivirus	DhBV, Dikerogammarus haemobaphes bacilliform virus	*Gamma-nudivirus*	Demon shrimp, *Dikerogammarus haemobaphes* (Amphipoda: Gammaridae)	302 ± 13 x 55 ± 4 nm; nonoccluded	106	119,754 (MT488302.1)	[54, 152]
ToNV, Tipula oleracea nudivirus	-	*Delta-nudivirus*	Marsh crane fly, *Tipula oleracea* (Diptera: Tipulidae)	~160 x 40 nm; occluded	131	145,704 (NC_026242.1)	[40, 153]

Table S2. Less-studied and putative nudiviral agents in invertebrates. Nudiviral agents in insects and other arthropods without fully sequenced genomes. Putative nudiviruses that miss sequence data for phylogenetic analysis are shown in bold. Genera in brackets "()" miss sequence data for profound phylogeny and may require revision.

Virus name	Other names	Genus	Host	Morphology	Ref.
NlENV, Nilaparvata lugens endogenous nudivirus	-	*Alpha-nudivirus*	Brown planthopper, *Nilaparvata lugens* (Hemiptera: Delphacidae)	Not described	[39]
MsENV, Melanaphis sacchari endogenous nudivirus	-	*Alpha-nudivirus*	Sugarcane aphid, *Melanaphis sacchari* (Hemiptera: Aphididae)	Not described	[67, 69]
AgENV, Aphis glycines endogenous nudivirus	-	*Alpha-nudivirus*	Soybean aphid, *Aphis glycines* (Hemiptera: Aphididae)	Not described	[69]
VcVLP, Venturia canescens virus-like particles	VcENV, Venturia canescens endogenous nudivirus	*Alpha-nudivirus*	*Venturia canescens* (Hymenoptera: Ichneumonidae)	~ 110 nm diameter VLPs with no nucleocapsids	[69, 72]
FaENV, Fopius arisanus endogenous nudivirus	-	*Alpha-nudivirus*	*Fopius arisanus* (Hymenoptera: Braconidae)	37 nm diameter of cylindrical nucleocapsids in VLPs	[66]
OcNV, Osmia cornuta nudivirus	-	*Alpha-nudivirus*	European orchard bee, *Osmia cornuta* (Hymenoptera: Megachilidae)	Not described	[154]
AdNV, Allomyrina dichotoma nudivirus	-	(*Alpha-nudivirus*)	Japanese rhinoceros beetle, *Allomyrina dichotoma* (Coleoptera: Scarabaeidae)	200 – 210 x 100 – 110 nm, nonoccluded	[47]
DvvNV, Diabrotica virgifera virgifera nudivirus	-	*Alpha-nudivirus*	Western corn rootworm, *Diabrotica virgifera virgifera* (Coleoptera: Chrysomelidae)	24 – 26 nm diameter, nonoccluded	[113]
EbrENV, Eurytoma brunniventris endogenous nudivirus	-	*Beta-nudivirus*	*Eurytoma brunniventris* (Hemiptera: Eurytomidae)	Not described	[70]
Charybdis crab nudivirus 1	-	(*Gamma-nudivirus*)	*Charybdis sp.* (Decapoda: Portunidae)	Not described	[155]
Charybdis crab nudivirus 2	-	(*Gamma-nudivirus*)	*Charybdis sp.* (Decapoda: Portunidae)	Not described	[155]

MrNV, Macrobrachium rosenbergii nudivirus	-	(*Gamma-nudivirus*)	Giant freshwater prawn, *Macrobrachium rosenbergii* (Decapoda: Palaemonidae)	Not described	Accession: PRJNA359633
AaBV, Astacus astacus bacilliform virus	-	-	European crayfish, *Astacus astacus* (Decapoda: Astacidae)	343 x 71 nm; nonoccluded	[156]
ApBV, Austropotomobius pallipes bacilliform virus	-	-	European freshwater crayfish, *Austropotomobius pallipes* (Decapoda: Astacidae)	360 x 67 nm; nonoccluded	[156, 157]
BMN, Baculoviral mid-gut gland necrosis	-	-	Kuruma shrimp, *Penaeus japonicus* (Decapoda: Penaeidae)	310 x 72 nm; nonoccluded	[158]
Baculo-PP		-	Blue king crab, *Paralithodes platypus* (Decapoda: Lithodidae)	200 ± 10 x 37 – 40 nm; nonoccluded	[156, 159]
CpBV, Cancer pagurus bacilliform virus	-	-	Brown crab, *Cancer pagurus* (Decapoda: Cancridae)	210 x 60 nm; nonoccluded	[160]
CdBV, Cherax destructor bacilliform virus	-	-	Common yabby, *Cherax destructor* (Decapoda: Parastacidae)	304 x 68 nm, nonoccluded	[156, 161]
CqBV, Cherax quadricarinatus bacilliform virus	-	-	Redclaw crayfish, *Cherax quadricarinatus* (Decapoda: Parastacidae)	260 x 100; nonoccluded	[161, 162]
GrBV, Gammarus roeselii bacilliform virus	-	-	*Gammarus roeselii* (Amphipoda: Gammaridae)	224 ± 17 nm x 70 ± 13 nm; nonoccluded	[163]
PlBV, Pacifastacus leniusculus bacilliform virus	-	-	Signal crayfish, *Pacifastacus leniusculus* (Decapoda: Astacidae)	229 x 72 nm; nonoccluded	[161, 164]
PmBV, Pandalus montagui bacilliform virus	-	-	Pink shrimp, *Pandalus montagui* (Decapoda: Pandalidae)	250 x 55 nm; nonoccluded	[165]
SsBV, Scylla baculovirus	Scylla serrata bacilliform virus	-	Mud crab, *Scylla serrata* (Decapoda: Portunidae)	253 x 44 – 67 m, nonoccluded	[156, 162]
TpNV, Tipula paludosa nudivirus	TpBV, Tipula paludosa baculovirus;	(*Delta-nudivirus*)	European crane fly, *Tipula paludosa* (Diptera: Tipulidae)	235 ± 5 x 105 ± 5 nm; occluded	[166, 167]

	TpNPV, Tipula paludosa nucleopolyhedrovirus	

“**You’re just a picture, you’re an image caught in time.**”

– *Rainbow in the Dark* by Dio (1983)

Chapter 3

Nudiviruses in free-living and parasitic arthropods: evolutionary taxonomy

Jirka Manuel Petersen[1,2], Amy L. Burgess[3,4], Monique M. van Oers[1],
Elisabeth A. Herniou[2], Jamie Bojko[3,4]

[1]Laboratory of Virology, Wageningen University and Research, 6708 PB Wageningen, the Netherlands.

[2]Institut de Recherche sur la Biologie de l'Insecte, UMR7261 CNRS - Université de Tours, 37200 Tours, France.

[3]School of Health and Life Sciences, Teesside University, Middlesbrough, TS1 3BX.

[4]National Horizons Centre, Teesside University, Darlington, DL1 1HG.

Published as:

Petersen JM, Burgess AL, van Oers MM, Herniou EA, Bojko J. Nudiviruses in free-living and parasitic arthropods: evolutionary taxonomy. Trends in Parasitology. 2024. DOI: 10.1016/j.pt.2024.06.009

Abstract

The nudiviruses (family: *Nudiviridae*) are large dsDNA viruses that infect insects and crustaceans, and have most recently been identified from ectoparasitic members (fleas and lice). This virus family was created in 2014 and has since been expanded via the discovery of multiple novel viral candidates or accepted members, sparking the need for a new taxonomic and evolutionary overview. Using current information (including data from public databases), we construct a new comprehensive phylogeny, encompassing 49 different nudiviruses. We use this novel phylogeny to propose a new taxonomic structure of the *Nudiviridae* by suggesting two new viral genera (Zetanudivirus and Etanudivirus), from ectoparasitic lice. We detail novel emerging relationships between nudiviruses and their hosts, considering their evolutionary history and ecological role.

Taxonomic history of the family *Nudiviridae*

Double-stranded DNA (dsDNA) viruses are enormously diverse and are divided over separate taxonomic realms [168]. One group of invertebrate-infecting dsDNA viruses sits in the class *Naldaviricetes* [169]. This includes the family *Nimaviridae* [170] and the order *Lefavirales*, housing the *Baculoviridae*, *Hytrosaviridae*, recently proposed Filamentoviridae [35], and the *Nudiviridae* [53, 171]. The *Nudiviridae* comprise bacilliform viruses that infect hosts across the Pancrustacea [172], including insects (flies, moths, beetles, crickets, craneflies) and crustaceans (lobsters, shrimp, crabs, amphipods), with varying degrees of virulence and ecological impact [171].

Pathological and morphological data originally pointed towards the presence of baculoviruses in insect and crustacean hosts [162, 173, 174]. However, a baculovirus termed "Baculo-A", derived from the swimming crab *Callinectes sapidus,* was described in the 1980's, but later revealed to be a nudivirus [175]. Similarly, the insect nudiviruses Oryctes rhinoceros nudivirus (OrNV), Heliothis zea nudivirus (HzNV-1) and Gryllus bimaculatus nudivirus were all originally described as baculoviruses [28-30]. Subtle morphological and pathological differences, including more obvious traits (i.e. lacking occlusion bodies leading to the term 'non-occluded baculoviruses'; **Box 1**), supported the split from the other baculoviruses, leading to the suggestion of a baculoviral subfamily ("Nudivirinae") [55], which eventually became the family *Nudiviridae* [53]. Further pathological support was derived from the host specificity and tissue tropism of nudiviruses, such as those that infect the hepatopancreas of crustacean hosts [83, 160], and those that manifest across the gut, fat-bodies and reproductive tissues of insects [171]. The localised nature of nudivirus infections within an infected individual stands as another primary distinction from baculoviruses, which typically induce systemic infections [49].

The *Nudiviridae* currently comprise four officially recognised genera: *Alphanudivirus* (including endogenous nudiviral elements), *Betanudivirus*, *Gammanudivirus,* and *Deltanudivirus* [53]. A fifth genus has been suggested for a group of divergent crustacean-infecting nudiviruses ('Epsilonnudivirus') [54] and two unofficial subfamilies have been considered, which would house the *Alphanudivirus* ('Alphanudivirinae') and the *Betanudivirus*, *Gammanudivirus*, and *Deltanudivirus* genera ('Betanudivirinae') [113].

An additional genus has been proposed for inclusion in the family *Nudiviridae* – the *Bracoviriform* (formerly 'Bracovirus') genus, which consists of endogenous nudiviruses **(Box 2)** from parasitic wasps, and which are currently classified as *Polydnaviriformidae* (formerly 'Polydnaviridae') (**Figure 1**) [79, 114].

Box 1. Not all nudiviruses are "nudists": Nudiviral occlusion bodies

Occlusion bodies (OBs) are crystalline structures assembled from viral proteins that are known for enwrapping the virions of certain insect viruses [176]. In baculoviruses, these dense protein structures protect the viral progeny from environmental stresses until they are ingested by their insect host. Despite their endurance towards many abiotic factors, baculovirus OBs easily dissolve under alkaline conditions to which they are exposed in the midgut of their hosts, leading to the release of virions responsible for primary infection [49].

While it was originally assumed that nudiviruses were non-occluded baculoviruses [177], cases of OB formations in members of the *Nudiviridae* have emerged, challenging the derivation of their name from the Latin term "*nudus*" (= naked or uncovered) [53]. Occluded nudiviruses are present in both insect (ToNV) and crustacean (PvNPV, PmNV) hosts. The OBs of ToNV and PmNV are irregularly shaped [40, 41], while PvNPV occludes its virions into pyramid-shaped OBs [173] (**Figure I**). Furthermore, OBs of insect- and crustacean-infecting nudiviruses seem to dissolve under different conditions. Congruent to baculoviruses in insects, the OBs of ToNV dissolved under alkaline treatment [178], while PmNV OBs showed resistance to high pH values, but dissociated when exposed to acidic conditions [41]. The dependency of OBs on *per os* transmission signifies the importance of co-evolutionary adaptation, whereby nudiviral OBs might have been tailored to match their host's digestive environments.

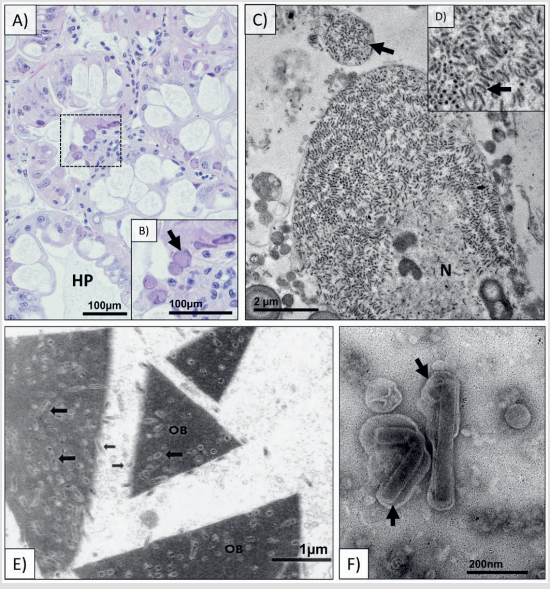


Figure I (in Box 1). Histological and electron microscopic images of nudiviral ultrastructure and development.
a) Multiple hepatopancreatic tubules (HP) display infected hepatopancreatocytes with hypertrophic nuceli, including a growing viroplasm (image obtained from *F. rusticus*; [179]). b) Inset image showing a higher magnified image of an infected nucleus (black arrow) (image obtained from *F. rusticus*; [179]). c) A transmission electron micrograph of Menippe mercenaria nudivirus (MmNV), in the nucleus of an infected hepatopancreatic epithelial cell, including a virions-containing vesicle (black arrow) (image reproduced from [180]). d) An enlarged view of the MmNV virions in the host nucleus (black arrow) (image reproduced from [180]). e) An example of a nudivirus from *Penaeus* sp., which produces occlusion bodies (OB) in pyramidal structures, containing virions (black arrows) (image reproduced from [173]). f) Negatively stained rod-shaped virions of HzNV-1 visualised via electron microscopy (purified from infected HZ-AM1 cells via sucrose cushion).

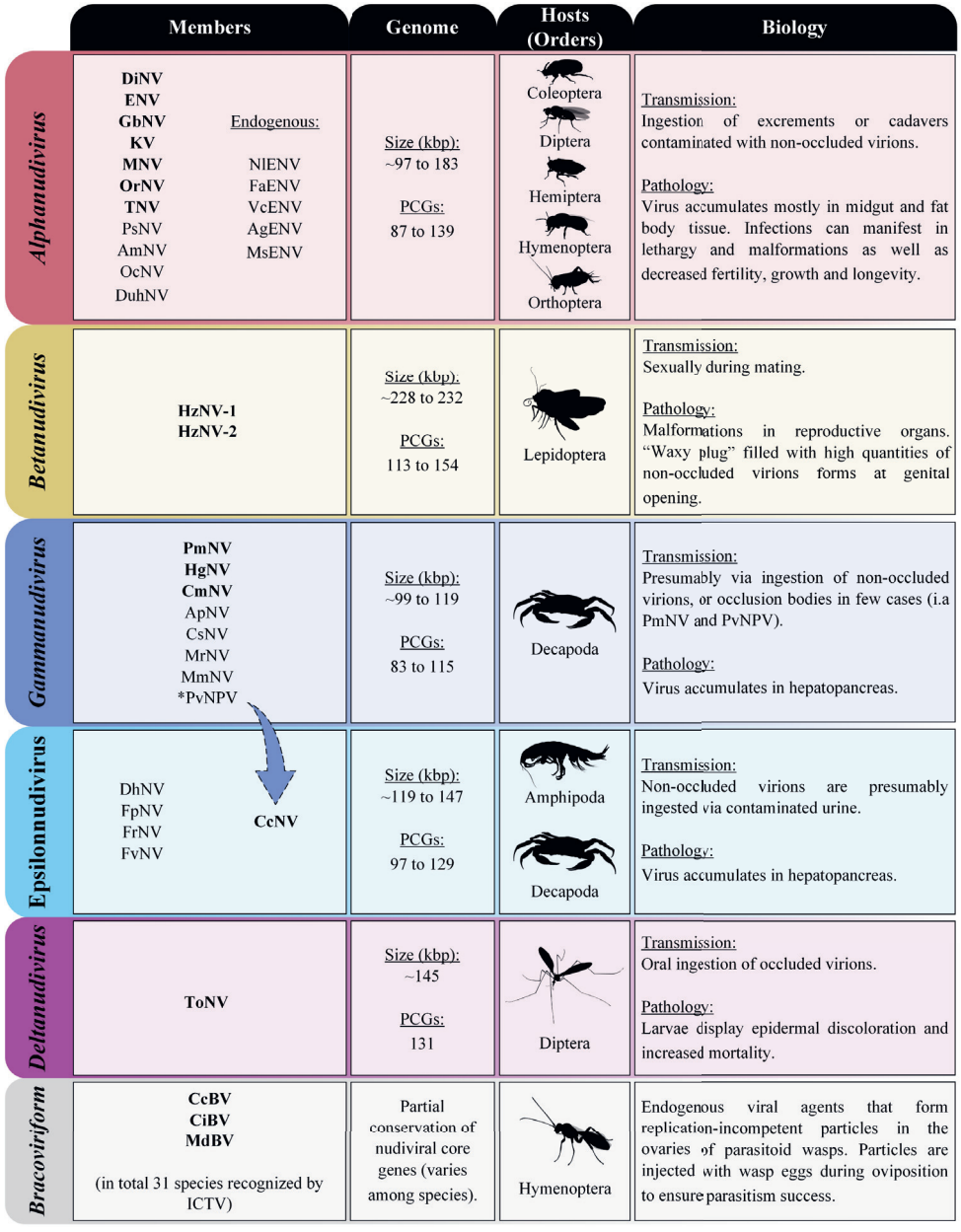


Figure 1. Overview of the family *Nudiviridae* covering official (*Alphanudivirus*, *Betanudivirus*, *Gammanudivirus*, *Deltanudivirus*) and unofficial (Epsilonnudivirus) genera as well as the related genus *Bracoviriform* (*Polydnaviriformidae*). Virus names in bold indicate official recognition by the ICTV with respective virus species names: DiNV, *Alphanudivirus droinnubilae*; ENV, *Alphanudivirus tertidromelanogasteris*; GbNV, *Alphanudivirus grybimaculati*; KV (*Alphanudivirus dromelanogasteris*); MNV (*Alphanudivirus quartudromelanogasteris*); OrNV (*Alphanudivirus oryrhinocerotis*); TNV (*Alphanudivirus alterdromelanogasteris*); HzNV-1/HzNV-2 (*Betanudivirus hezeae*); CmNV (*Gammanudivirus cameanadis*); CcNV (*Gammanudivirus cracrangonis*) is officially a gammanudivirus, but phylogenetic analyses support its assignment to Epsilonnudivirus (blue arrow); HgNV (*Gammanudivirus hogammari*); PmNV (*Gammanudivirus*

pemonodonis); ToNV (*Deltanudivirus tipoleraceae*). (*) PvNPV is currently assigned to the *Baculoviridae*, but shows clear phylogenetic relationship to the *Nudiviridae* (**Figure 2**).

Transcriptomic and metagenomic studies in invertebrate hosts have hinted at the existence of yet uncharacterised nudiviruses in other arthropods – for example, nudivirus-like transcripts were found in two additional species in the order Decapoda: *Eriocheir sinensis* and a *Charybdis* sp. [155, 181]. The bounty of next generation sequencing data from arthropods (593,792 public datasets – November 9th, 2023) in publicly available Sequence Read Archives (SRA) [182] represents a major resource for discovering more nudiviruses in order to address questions regarding taxonomy and evolutionary timeline.

Nudivirus genomes consist of a single circular dsDNA molecule, ranging in size from approximately 96 to 232 kilobase pairs (kbp), encoding between 87 to 154 protein coding genes (PCGs) [171]. All lefavirals encode homologs of genes involved in viral (late) gene transcription, including the four subunits of the DNA-directed RNA polymerase (*lef-4*, *lef-5*, *lef-8*, and *lef-9*). In nudiviruses, these four genes belong within a set of 28 'core' genes that can be found in all exogenous nudivirus genomes, whereas endogenous nudiviruses can have an incomplete repertoire of core genes [39, 69]. Nudiviral core genes fulfil crucial functions in virus replication, infectivity, virion assembly and morphology, but several have unresolved functions [171].

Underexplored nudiviral diversity

Numerous unclassified viral agents with nudivirus-like pathology and morphology exist. Two studies conducted in the 1970s identified nuclear viruses with nonoccluded, rod-shaped virions in *Solenopsis* sp. (Hymenoptera: Formicidae) and *Gyrinus natator* (Coleoptera: Gyrinidae). These viruses exhibited nudivirus-like symptoms [183, 184], but molecular data are missing. Another example involves the Tipula paludosa nucleopolyhedrovirus (TpNPV) [167], that is, based on single gene homology, more closely related to ToNV than to baculoviruses [40], emphasizing the possibility that unclassified viruses (some with occlusion bodies; **Box 1**) could be nudiviruses.

For the crustaceans, pathological studies have revealed nudivirus-like discoveries from (i) crabs: *Cancer pagurus* [160]; *Carcinus mediterraneus* (Tau-virus) [185]; *Paralithodes platypus* [159]; *Scylla serrata* [186]; and *Pinnotheres pisum* [187]; (ii) crayfish: *Astacus astacus* [188]; *Cherax quadricarinatus* [189]; *Pacifastacus leniusculus* [190]; *Cherax destructor* [156]; and *Astropotamobius pallipes* [157]; (iii) shrimp: *Farfantapenaeus* (=*Pandalus*) *montagui* [165]; *Penaeus japonicus* [191]; *Penaeus duorarum* [192]; *Penaeus plebejus* [193]; and (iv) various amphipods [194-197]. To date, all sequenced "baculovirus-like" isolates in crustaceans belong to the family *Nudiviridae* [180].

Evidence of nudiviruses hiding in cyberspace

In total, 2110 SRA datasets from arthropods were searched for nudivirus-like sequences (**Supplementary data, Table S1,** and online **Supplementary file S1**). Ten SRA submissions included large, circular dsDNA molecules derived from the nudivirus genomes (from insects: *Franciscoloa pallida* [198], *Franciscoloa roseicapillae* [198], *Myrsidea ptilorhynchi* [198], *Heterodoxus spiniger*, *Cuculoecus africanus*, *Echinophilopterus claytoni*, *Lagopoecus perplexus* [199], three datasets from *Ctenocephalides orientis* [200]); and eight included partial/complete core gene sets (from insects: *Culex pipiens* [201], *Goniodes lagopi* [199]; from crustaceans: *Gammarus pulex*, *Gammarus fossarum*, *Echinogammarus berilloni*, *Astacus leptodactylus* [202], *Eriocheir sinensis*, *Charybdis* sp. [155]). Additionally, putative nudiviral sequences for which fragmented/incomplete core gene sequences were obtained from five species (insects: *Aleuroclava psidii* [203], *Dasineura brassicae* [204], *Phlebotomus*

chinensis [205], *Trichocera saltator* [206, 207]; crustacean: *Niphargus hrabei* [208]) (**Supplementary data,** and online Supplementary file S1, S3 and S4).

Box 2. Endogenous nudiviruses among insect hosts

Whole-genome sequencing techniques have revealed the presence of nudivirus-derived endogenous viral elements (EVEs) in various arthropods [67]. Although the circumstantial conditions for nudiviral genome integration is scarcely studied, it can be assumed that such integration is driven by two possible scenarios: an active integration mechanism employed by nudiviruses to establish persistent infections, as it was observed for HzNV-1 [65]; or indirect integration as a consequence of induced DNA damage and repair during replication in the host cell nucleus [209]. Evolutionarily, the integration of nudiviral genetic material into host genomes has resulted in the emergence of several endogenous and "functional" nudiviruses – i.e. the bracoviruses. These endogenised viruses serve as invaluable "fossil records", offering insights into the evolutionary history of ancient viruses and their hosts [114]. Despite their ostensible frequency, nudiviral integration events produce only a few examples of functional endogenisations, and among those functional cases, there are still varying degrees of benefit to the host. For instance, FaENV and VcENV form virus-like particles (VLPs) in their parasitoid hosts (*Fopius arisanus* and *Venturia canescens*) that are injected together with the wasp eggs during parasitism-mediated oviposition. The particles of VcENV are considered to benefit the success of parasitism in its hymenopteran host (for instance via immune-protective coating of wasp eggs, carrying wasp-derived virulence proteins, and manipulation of parasitised host physiology), while FaENV particles carry DNA-empty nucleocapsids with unknown benefit to its parasitoid wasp. VLPs of VcENV and FaENV both possess viral envelopes with incorporated nudiviral proteins, but without a nucleic acid core. This stands in contrast with virions formed by bracoviruses, which are enveloped, and carry circular dsDNA molecules inside their nucleocapsids [171].

Next to these functional models of endogenous nudivirus-derived agents, there are scenarios where nudivirus integration events have led to non-functionality. The endogenous nudiviral agents of three hemipteran insects (*Nilaparvata lugens*: NlENV; *Melanaphis sacchari*: MsENV; *Aphis glycines*: AgENV), for instance, are unable to produce particles or at least have no evident function for the insect host [39, 69, 171]. Furthermore, a study by Cheng and co-authors [67] found a total of 359 nudivirus-like sequences in the genomes of various arthropods, the majority being present in Hemiptera and Hymenoptera, but with no confirmed functionalities.

Datasets from the oriental cat flea (*Ctenocephalides orientis*: Siphonaptera) led to the assembly of three distinct full nudivirus genomes that we refer to as Ctenocephalides orientis nudivirus isolates 1-3, CoNV1, CoNV2, and CoNV3, respectively (Genome range: 90,491 – 90,736 bp; PCGs: 86 – 87). Eight novel nudiviruses are from either the Amblycera or Ischnocera superfamily of ectoparasitic lice (Psocodea) (**Supplementary data, Table S1**). Genomic data disclosed seven complete louse nudiviral genomes, reflecting Franciscoloa pallida nudivirus (FrpNV); Franciscoloa roseicapillae nudivirus (FrrNV); Myrsidea ptilorhynchi nudivirus (MpNV); Heterodoxus spiniger nudivirus (HxsNV); Cuculoecus africanus nudivirus (CafNV); Echinophilopterus claytoni nudivirus (EcNV); and Lagopoecus perplexus nudivirus (LpNV). Goniodes lagopi nudivirus (GlNV) presented as two contiguous sequences, but encoded homologues of all expected virus genes. Genome size ranges: 83,884 – 97,269 bp; PCGs: 88 – 104 proteins.

The common house mosquito (*Culex pipiens*: Culicidae) in the study by Feng *et al*. [201] provided 102 predicted nudivirus genes through transcriptomic data, which is considered to represent the entire

collection of expressed genes for Culex pipiens nudivirus (CpNV). Transcriptomic data from three amphipods and three decapods also revealed new nudivirus-like sequences (**Supplementary data, Table S1**). We name these respectively, as: Gammarus pulex nudivirus (GpNV); Gammarus fossarum nudivirus (GfNV); Echinogammarus berilloni nudivirus (EbNV); Astacus leptodactylus nudivirus (AlNV); Eriocheir sinensis nudivirus (EsNV) [181] and Charybdis nudivirus (CharyNV). These encode 62 – 96 proteins.

Extensive nudivirus phylogeny leads to new evolutionary perspectives

Complete virus genomes are necessary for formal taxonomic inclusion by the International Committee on Taxonomy of Viruses (ICTV); however, most studies apply just the conserved genes in their phylogenetic analyses [171, 180]. By incorporating the core genes of incomplete viral genomes (or transcripts), we can achieve identical high-level taxonomic accuracy (**Figure 2**). We included 16 novel nudivirus isolates in our phylogenomic analysis, providing the first phylogeny of nearly 50 nudiviruses.

Our phylogeny supports the current genus demarcation (*Alphanudivrus* to *Deltanudivirus*) and emphasises the proposed Epsilonnudivirus genus as a monophyletic group through the addition of GpNV, GfNV and EbNV, as well as the most recently characterised crayfish nudiviruses FrNV, FpNV, and FvNV [179] (**Figure 2**). The new decapod nudiviruses (AlNV, EsNV and CharyNV) cluster with the other decapod nudiviruses in the genus *Gammanudivirus*. The fundamental monophyletic groups of the nudivirus phylogeny remain stable and well-supported, suggesting that the existing branches of the *Nudiviridae* family are resiliently defined using core proteins [53].

The CoNV isolates and CpNV group in the genus *Alphanudivirus*, which adds fleas and mosquitos to the already broad insect host range of this genus (**Figure 1-2**). The louse-infecting nudiviruses share a most recent common ancestor (MRCA) with the beta-, gamma-, delta- and epsilonudiviruses, but branch off as a distinct clade that could be considered to form a novel genus in the proposed subfamily Betanudivirinae. The nudiviruses from amblyceran lice (FrpNV, FrrNV, MpNV) form their own monophyletic group, except for HxsNV, which clusters as a related group with four nudiviruses isolated from ischnoceran lice (CafNV, EcNV, GlNV, LpNV).

The core gene synteny of alphanudiviruses is generally poorly conserved, with the exception of certain monophyletic groups, such as the *Drosophila*-infecting nudiviruses ENV, KV, DiNV, and MNV (**Figure 2**). Nudiviral taxonomy must become united to determine whether gene synteny should be considered an appropriate taxonomic criterion, where in this case, the large genus *Alphanudivirus* may be broken into multiple additional genera. Accordingly, the core gene organisation within other genera was more congruent, with less drastic rearrangements, except for the expanded group of Epsilonnudivirus members, which shows ~5 gene rearrangements (**Figure 2**).

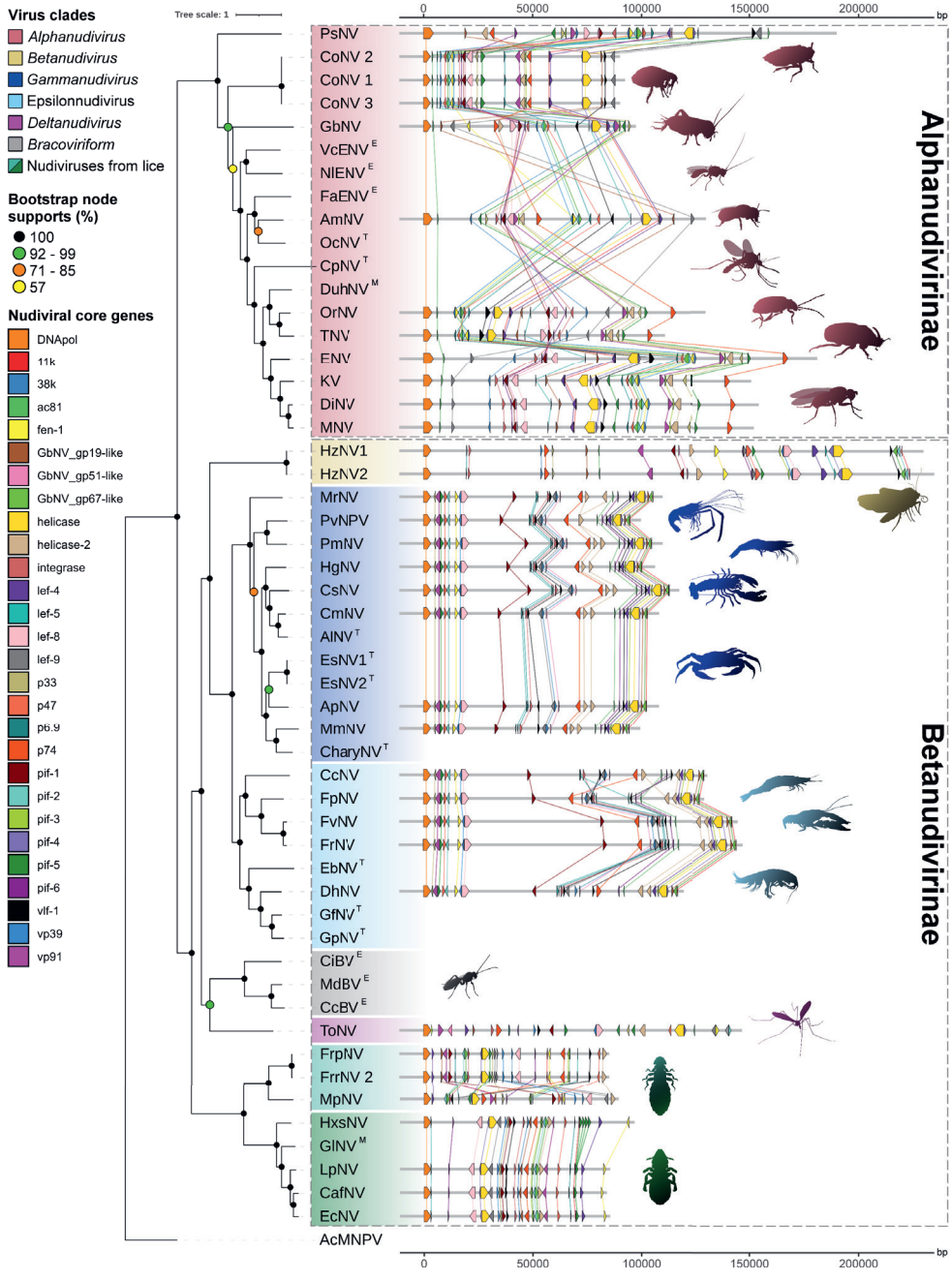
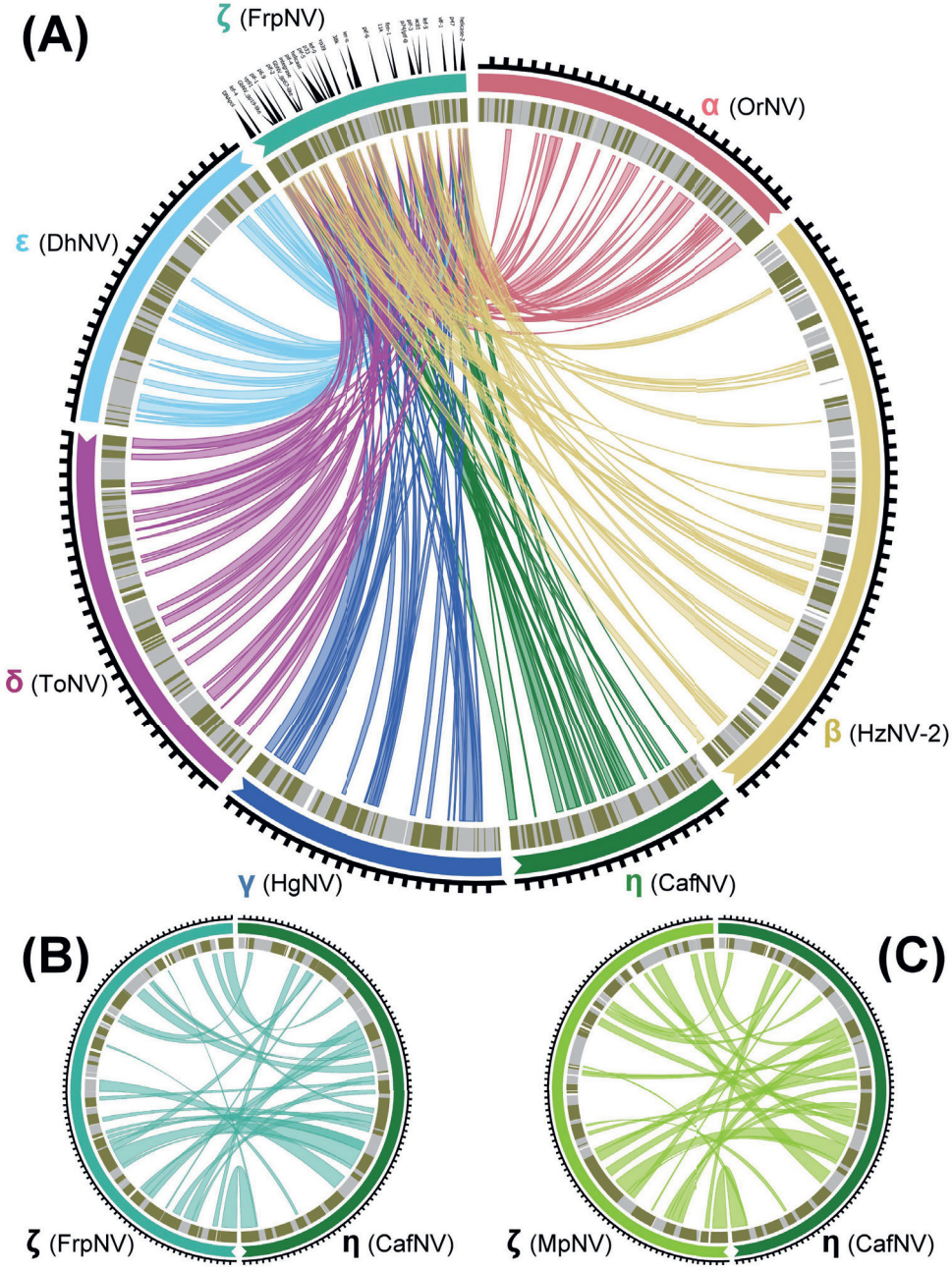


Figure 2. Phylogenetic tree with 49 nudiviruses, three bracoviruses and the baculovirus Autographa californica nucleopolyhedrovirus (AcMNPV) as outgroup, coupled with intragenus core gene synteny compared among nudiviruses with complete genomes. The amino acid sequences of 17 nudiviral core genes were aligned with the MAFFT alignment tool (v7.490) of Geneious (v2023.2.1). Best-fit substitution models were determined separately for each core gene based on the Bayesian information criterion (BIC) generated in IQ-Tree [210] (**Supplementary data**, **Table S2**). The multi-protein alignment file, containing 17,906 total sites, and the model

information were subjected to maximum-likelihood (ML) phylogenetic analysis with 1000 bootstrap replicates. The inferred treefile was visualised with the online tool iTOL [211]. Percentage values of bootstrap supports are indicated as coloured circles. On the right, linear synteny plots generated with the R packages ggplot2 (v3.4.4) and gggenes (v0.5.1) show core gene (coloured arrows) rearrangements compared among nudiviruses of the same genus in phylogenetic order. A scale bar (base-pairs) is presented at the top and bottom of the tree. Coloured lines between arrows connect homologous core genes of phylogenetic neighbours. Each dual synteny plot starts with the DNA polymerase of the respective virus. Virus species with superscript letters were not included in the gene synteny plot, either because they originate from transcriptomic data (T), are missing accurate genomes (M), or are endogenous (E). The original phylogenetic tree with node supports and additional references can be found in **Supplementary data**, and accession numbers of genes used from each virus are available online in the Supplementary file S1.

A similar observation could be made for the two monophyletic groups of louse-infecting nudiviruses (**Figure 2-3**). The group comprised of HxsNV, CafNV, EcNV, GlNV and LpNV shows nearly identical core gene synteny, though a slight gene shift can be observed in HxsNV due to multiple *pif-5* duplications. FrpNV and FrrNV share highly similar core gene synteny, while MpNV is most distinct in its gene organisation, relative to the other louse-infecting nudiviruses, perhaps indicating that MpNV may belong to a third clade of louse nudiviruses. Due to their monophyletic lineages and major differences in gene synteny (**Figure 3B-C**), the groups housing the louse nudiviruses might be considered as two novel genera for ICTV consideration, following the Greek alphabet - Zetanudivirus (FrpNV, FrrNV, MpNV) and Etanudivirus (HxsNV, CafNV, EcNV, GlNV, LpNV).

Figure 3. Inter-genus synteny of nudiviral core genes among representatives of nudivirus genera. **(A)** Circular plot showing the core gene synteny of nudivirus representatives from the genera *Alphanudivirus* (α), *Betanudivirus* (β), *Gammanudivirus* (γ), *Deltanudivirus* (δ), Epsilonnudivirus (ε) and Etanudivirus (η) compared to Zetanudivirus (ζ). The core genes of FrpNV are arranged in the following order (left to right): *dnapol*, *lef-4*, *GbNV_gp19-like*, *vp91*, *p6.9*, *pif-2*, *GbNV_gp67-like*, *integrase*, *pif-4*, *helicase*, *pif-5*, *p33*, *lef-9*, *vp39*, *38k*, *lef-8*, *pif-6*, *11k*, *fen-1*, *p74*/*pif-0*, *pif-3*, *ac81*, *lef-5*, *vlf-1*, *p47*, *helicase-2*. **(B)** Comparative core gene synteny of Zetanudivirus (FrpNV) and Etanudivirus (CafNV). **(C)** Comparative core gene synteny of Zetanudivirus (MpNV) and Etanudivirus (CafNV). The curved bars each represent the genome of the indicated nudivirus species, and are coloured according to the genus they belong to. Sites with indented triangles represent the start of the genome (start codon of DNA polymerase). Gene directions are visualised as golden (positive strand) and silver (negative strand) rectangles. Ribbons represent genome connections between core gene homologs of the respectively coloured nudiviruses. Distances between ticks represent 5 kb. The circular plots were generated with Circa (www.omgenomics.com/circa/). **See next page**.


(A)
ζ (FrpNV)
α (OrNV)
ε (DhNV)
δ (ToNV)
β (HzNV-2)
γ (HgNV)
η (CafNV)
(B)
ζ (FrpNV)
η (CafNV)
(C)
ζ (MpNV)
η (CafNV)

Molecular dating tree – nudiviral paleovirology

Using the available molecular information, we performed a molecular dating analysis with conserved nudiviral core genes from 49 nudiviruses, three bracoviruses, seven baculoviruses, and the hytrosavirus Musca domestica Salivary Gland Hypertrophy Virus (MdSGHV), to estimate ages of viral evolutionary events. A time-dated tree was generated via the RelTime Branch-Lengths method implemented in MEGA11 [212]. The RelTime method is a non-Bayesian approach that operates on a relative rate framework (RRF), computing branch-specific relative rates directly from the branch lengths of inferred non-clock trees, while relaxing the molecular clock throughout the tree [213]. In contrast to Bayesian methods, RelTime does not require a tree prior and a clock model as parameters, and calculates Confidence Intervals (CIs) of divergence times using the Tao *et al.* method [214]. Next to the non-clock tree, a single calibration constraint of 103.38 ± 4.41 Mya (red-circled Node 15; **Figure 4A**) was used as input under a log-normal distribution, representing the MRCA of Chelonus inanitus bracovirus (CiBV) and Cotesia congregata bracovirus (CcBV) [215]. We accounted for the possibility of strongly divergent evolutionary rates of the endogenised bracoviruses compared to the exogenous nudiviruses by performing an evolutionary rate analysis with HyPhy (v2.5.61) [216] under the Fixed Effects Likelihood (FEL) model [217]. The analysis of eight nudiviral core genes conserved among four bracoviruses and 32 exogenous nudiviruses resulted in ratios of non-synonymous to synonymous substitutions (dN/dS) < 1, which indicates a strong negative selection across all nudiviral core genes (**Supplementary data, Table S4**). The bracoviral genes encoding for virion components displayed higher evolutionary rates (dN/dS: 0.133 – 0.315) than those of exogenous nudiviruses (dN/dS: 0.048 – 0.118), which is congruent with findings from a previous study [218], while the transcription-associated genes evolve at more similar rates in both braco- and nudiviruses. Despite their endogenisation, the nudiviral core genes of bracoviruses appear to have a more relaxed mutation rates than free-living nudiviruses and a possible underestimation of the age estimates in the molecular dating analysis needs to be considered.

Our molecular dating analysis provided estimates of shared evolutionary events among members of the *Nudiviridae* and *Baculoviridae*. The MRCA of all nudiviruses including the bracoviruses (~280 Mya; the Palaeozoic Era, Permian) traces further back than the MRCA of baculoviruses (~222 Mya; the Mesozoic Era, Triassic), which is coherent with an earlier evolutionary study [114]. Node 1 indicates the time at which the MRCA of both nudiviruses and baculoviruses is estimated – 378 million years ago (Mya) (Devonian period of the Palaeozoic Era), while the divergence of holometabolous insects from polyneopteran insects is estimated at ~383 Mya according to TimeTree [219]. Hence, the divergence of these viral families likely happened early in insect evolution.

Another Palaeozoic hallmark involves the MRCA of all nudiviruses (~280 Mya), during the Permian, where the unofficial subfamilies Alphanudivirinae (including all alphanudiviruses) and Betanudivirinae (comprising all other nudiviruses, including the lice nudiviruses) split [113] (**Figure 4A**). In the evolutionary history of eumetabolous insects, hemipteran and psocodean insects diverged approximately 9 million years before the diversification of holometabolous insects (as inferred from **Figure 4B**). The emergence of the two nudiviral subfamilies could be linked to the diversification of these respective insect groups, with the theory that ancestral hemipteran and psocodean insects were hosts to the ancestors of the Alphanudivirinae and Betanudivirinae, respectively.

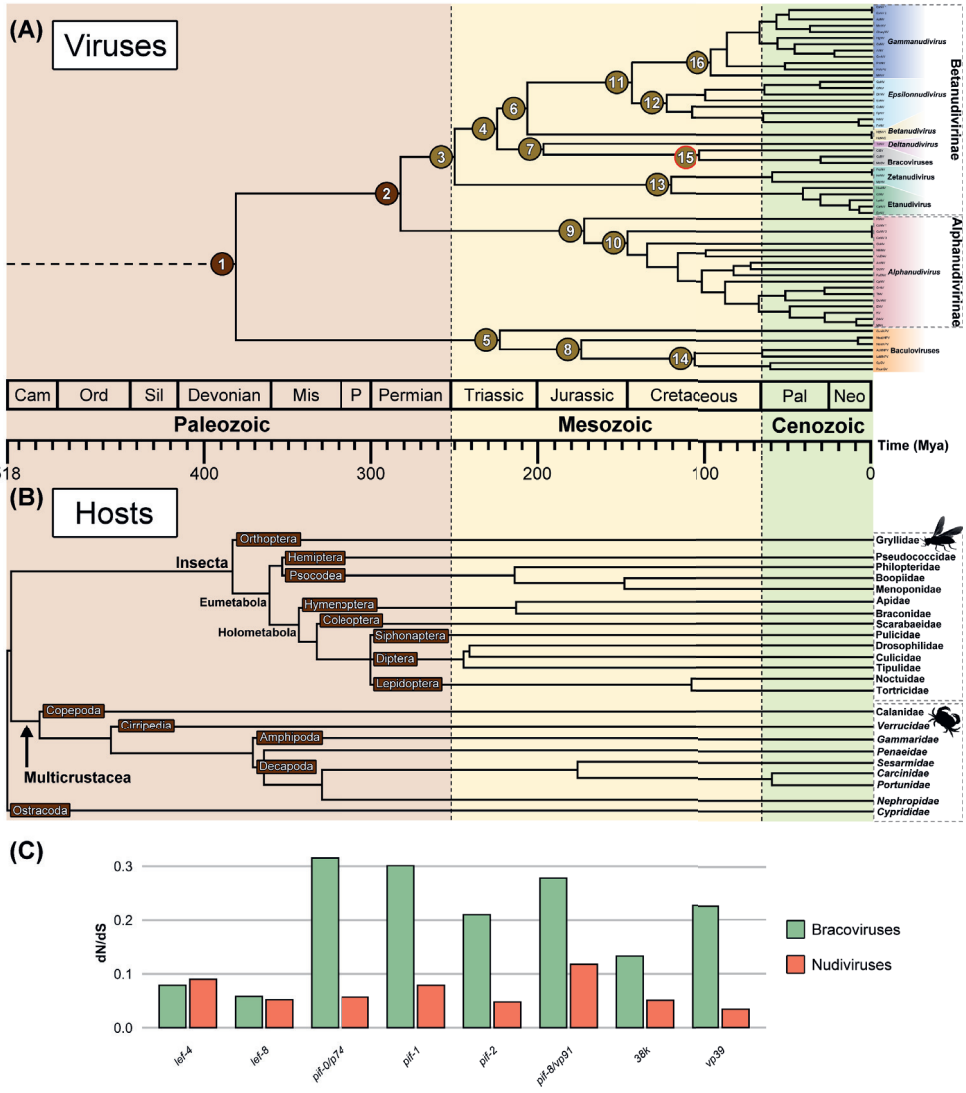


Figure 4. Molecular dating analysis of nudiviruses, baculoviruses and bracoviruses in relation to evolutionary timeline of crustacean and insect host clades. **(A)** Molecular dating tree from dsDNA viruses. Phylogenetic virus tree with optimised branch lengths served as input to compute the molecular dating tree with the MEGA11 (v11.0.10) software by using the RelTime-Branch Lengths option [212]. A single calibration constraint of 103.38 ± 4.41 Mya (red-circled Node 15) was specified with log-normal distribution, representing the most recent common ancestor (MRCA) of Chelonus inanitus bracovirus (CiBV) and Cotesia congregata bracovirus (CcBV) [215]. The age estimates and 95% CIs of the numbered nodes can be found in **Table 1**. The original molecular dating tree from MEGA11, including error bars and additional details on the methodology, can be found in **Supplementary data** (**Figure S13**). **(B)** Timeline of insect and crustacean hosts. The web-based TimeTree of Life (TToL5) resource [219] was used to build the host tree by providing a text file with relevant species names. The TToL5 resource is a publicly accessible database that infers divergence times of taxa based on the median value of existing evolutionary studies. **(C)** Bar chart showing the dN/dS values across eight nudiviral core genes from four bracoviruses and 32 exogenous nudiviruses. The dN/ds values were inferred from the codon-aligned nucleotide sequences with HyPhy (v2.5.61) under the Fixed Effects Likelihood (FEL) model (**Supplementary data, Table S4**).

This assumption is supported by the observation that the hemipteran *Alphanudivirus*, PsNV, and the psocodean nudiviruses (lice-infecting) branch off at the deepest nodes of their respective subfamilial clades (**Figure 4A**). This is further supported by the lack of hemipteran nudiviruses outside of the Alphanudivirinae, and no psocodean nudiviruses outside of the betanudivirinae, to date. Furthermore, the mosquito-infecting CpNV shares a MRCA with the other dipteran nudiviruses (e.g. ENV) and coleopteran nudiviruses (OrNV and DuhNV), suggesting that the earlier divergence of CpNV does not follow the evolutionary lineage of the dipteran and coleopteran hosts. We do not yet have enough data on the diversity of beta- and deltanudiviruses in holometabolous insects to explore the emergence of these lineages further – one plausible option may be that these hosts (*Tipula* spp. and *Heliothis* spp.) attained these viruses from other, yet unidentified, host groups.

The MRCA between the CoNV isolates (*Alphanudivirus*) and the rest of the alphanudiviruses (excluding PsNV) was estimated at 145 Mya (Node 10; 95% Confidence Interval: CI lower = 123 Mya, CI upper = 173 Mya), hinting that one of the oldest ancestors of alphanudiviruses might have been associated with an ancestral flea. Amniotes emerged ~325 Mya, during their colonisation of terrestrial environments [220]. This coincides with the emergence of the Siphonaptera, which emerged ~300 Mya (**Figure 4B**) and have been termed "dinosaur fleas", which are thought to have switched to mammals [221]. The late Jurassic and early Cretaceous sit within the 123 – 173 Mya period, a time when early mammals (e.g. marsupials) evolved alongside the diversification of boreid fleas [222]. The diversification of these flea nudiviruses, within the *Alphanudivirus* genus, could plausibly have occurred alongside the evolution of mammalian flea hosts.

The MRCA of all viruses in the Betanudivirinae clade can be traced back as the third deepest node among all observed events (Node 3; 248 Mya; **Figure 4A**), primarily including the early branching of the lice-infecting zeta- and etanudiviruses from the rest of the Betanudivirinae (Node 3). The diversification of the amblyceran zetanudiviruses and the ischnoceran etanudiviruses (HxsNV being the amblyceran exception) from one-another is estimated between 104 – 138 Mya (Node 13; 120 Mya; **Figure 4A**). The radiation of those two nudivirus lineages is coherent with the estimated origin of the clade Phthiraptera (115 Mya) [223] and the MRCA of its subclades Amblycera and Ischnocera between 105 – 110 Mya [224]. Suggestively, the evolutionary split of zeta- and etanudiviruses may have been driven by the diversification of the Phthiraptera clade into amblyceran and ischnoceran lice during the early-mid Cretaceous Epoch. Given the time period (early-mid Cretaceous), and the age of bird-like dinosaur evolution and diversification events (90 – 170 Mya) [225], this virological event could have taken place on the back of a feathered dinosaur, mimicking this same ecological system we now see in modern day birds, infested with lice, and hyperparasitised by nudiviruses.

Within the Betanudivirinae, the divergence of gammanudiviruses represents a relatively recent event in nudivirus evolutionary history. Since the crustacean diversification occurred in the Ordovician (~500 Mya) [219], it seems that this group may have acquired nudiviruses via a zoonotic event, possibly from an insect. Given that all crustacean nudiviruses identified to date (gamma- and epsilonnudiviruses) have a single ancestor, estimated at ~140 Mya (node 11; late Jurassic – early Cretaceous), it seems possible that a single host-switching event resulted in the crustacean infection capability that we see today. Perhaps the tissue tropism of crustacean-infecting compared to insect-infecting nudiviruses can explain this phenomenon. Crustacean nudiviruses target the gut, as do baculoviruses, but baculoviruses and other insect-infecting nudiviruses are also able to infect other tissue types – a trait missing in the *Gammanudivirus* and Epsilonnudivirus members. *Gammanudivirus* and Epsilonnudivirus appear to be a younger lineage of the *Nudiviridae*, beginning to parasitise the ancient arthropod group, Crustacea, during the late Jurassic.

Global distribution of nudiviruses and their hosts

When all nudiviruses are considered, their distribution extends across all continents (**Figure 5**). Marine, freshwater, and terrestrial hosts harbour nudiviruses – it seems that if insects or crustaceans occupy an area, nudiviruses likely occupy it as well. For insects, their environmental and ecological relationships are primarily associated with terrestrial systems, to a lesser extent freshwater systems; since we did not identify any nudiviruses from marine mammal-infesting insects. However, a handful of nudiviruses have been isolated from insects with a freshwater life stage (i.e. mosquitos), as well as a putative virus with nudivirus-like pathology from a water beetle (*G. natator*) [184]. Regarding terrestrial climatic zones, our findings complement the presence of nudiviruses across known biotopes (e.g. tropical: EcNV, FrpNV, FrrNV, MpNV; subequatorial: CpNV, HxsNV; equatorial: CafNV, CoNV), but also expand their habitat to the yet unrecognised sub-polar zones, with LpNV and GlNV from Alaska.

Insect hosts of nudiviruses from across Asia include: fleas (*C. orientis*) from infested dogs in Perak (Malaysia) [200]; the hemipteran, *A. psidii*, which infested guava trees (*Psidium guajava*) in Xinyang, China [203]; *H. spiniger* lice from wolves in India; and *C. pipiens* and *P. chinensis* from China [201, 205]. Siphonaptera, Hemiptera and Diptera samples had been collected across both Asia and Europe, while the lice hosts (Psocodea) cover a broader continental range.

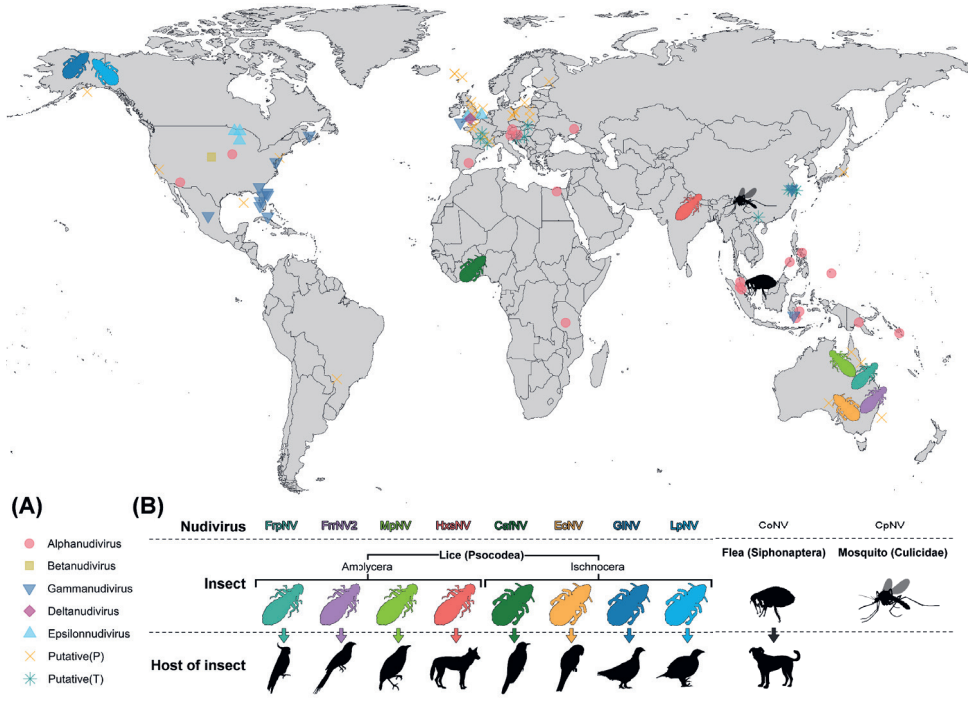


Figure 5. World map of confirmed and putative locations where nudivirus have been identified. Coordinates (latitude and longitude) of nudivirus-related find spots were used as input for the R packages maps (v3.4.2) and ggplot2 (v3.4.4) with support by the packages dplyr (v1.1.4) and ggpubr (v0.6.0). **(A)** Differently coloured shapes represent already identified nudiviruses with fully sequenced genomes, or putative yet uncharacterised nudiviruses based on pathological (P) or transcriptomic implications (T) from literature and personal data. **(B)** Newly characterised nudiviruses found in mosquito and ectoparasitic insects (flea and lice). Nudivirus names are shown at the top with their respectively coloured insect hosts below (separated by dashed lines). For the lice and flea, the warm-blooded animals that those ectoparasitic insects infest are displayed at the bottom (left to right): *Eolophus*

roseicapilla, *Centropus phasianinus*, *Ptilonorhynchus violaceus*, *Canis lupus*, *Chrysococcyx cupreus*, *Northiella haematogaster*, *Lagopus lagopus/muta*, *Tympanuchus phasianellus*, *Canis familiaris.* The list of nudiviruses and their coordinates can be found online in Supplementary file S1.

Most lice samples were gathered from bird hosts in New South Wales and Queensland (Australia; *F. pallida*, *F. roseicapillae*, *M. ptilorhynchi*, *E. claytoni*), while *L. perplexus* and *G. lagopi* were sampled in Alaska (North America), and another (*C. africanus*) in Ghana (Africa; **Figure 5B**). Given the ability of ectoparasites to traverse diverse environments by hitchhiking on the animals on which they feed (including birds), the widespread distribution of nudiviruses could be explained by migratory birds through evolutionary history.

In addition to the geographical dispersal, the observed nudiviral host diversity highlights further layers of ecological complexity (**Figure 5-6**). Next to their importance as pathogens of commercially important arthropods, and as biocontrol agents against certain insect pests (i.e. OrNV [226]; bracoviruses [227]), their presence in haematophagous (blood-feeding) insects emphasises a yet undescribed ecological niche this virus family fills. The eight psocodean species, the mosquitoes *C. orientis* and *C. pipiens,* as well as the sandfly *P. chinensis* all feed on the host tissue or blood of endothermic animals, including humans. However, *C. pipiens* and *P. chinensis* can complete their life cycles without the need of hematophagy [228, 229], while obligate ectoparasites such as lice and fleas spend most of their lives on the hosts from which they feed [230, 231]. The proximate interaction between ectoparasites and their hosts are usually accompanied by a long history of co-evolutionary adaptations [232, 233]. The discovery of these large dsDNA viruses in lice and fleas introduces a yet undescribed trophic level to these ectoparasite-host systems. Since viruses are intracellular parasites [234], one may consider this tripartite relationship a form of hyperparasitism [235-237] or, at least, showcase another parasitic strategy that members of the *Nudiviridae* have converged upon [238]. Another tripartite model is known from parasitoid wasps and their endogenised bracoviruses, whose nudiviral ancestor integrated into an ancestral wasp genome ~100 million years ago and co-evolved with the wasp to become a domesticated mutualist [239]. Our findings emphasise an expanding societal relevance of nudiviruses that might eventually reach beyond applications in biocontrol and arthropod farming, to the sectors of healthcare and veterinary medicine (**Figure 6**). Where phage therapy is used to control bacterial infection, perhaps "nuditherapy" might help animals to get rid of lice and flea pests in the future, providing that empirical data are obtained one day to show these newly discovered nudiviruses are pathogenic to their insect hosts.

Crustacean-infecting nudiviruses were found across marine and freshwater systems of China, Australia, Europe, inland USA, the Atlantic and Pacific coasts of Canada and the USA, as well as the Gulf of Mexico (**Figure 5**). The host diversity of these groups includes Decapoda and Amphipoda; however, all these hosts are omnivores (primarily detritivorous) in their freshwater, euryhaline, or marine habitat (**Figure 6**). Five crustacean hosts of nudiviruses are commonly present in fisheries, and three are involved in aquaculture efforts, highlighting the importance of understanding the interactions between crustacean and virus, including their economic impact (**Figure 6**). Examples include *C. sapidus* ("blue crab fishery") [240, 241]; *Homarus gammarus*, where HgNV has only been identified from hatcheries to date [92]; and the more intensive *P. monodon* aquaculture systems across Asia [242], where PmNV causes "wasting disease", lowering yield via a reduction in juvenile shrimp growth [41]. The remaining ten nudiviruses have been sampled from wild environments, where some are native components of co-evolved ecosystems [194, 243], while other nudiviruses infect invasive species [54, 83].

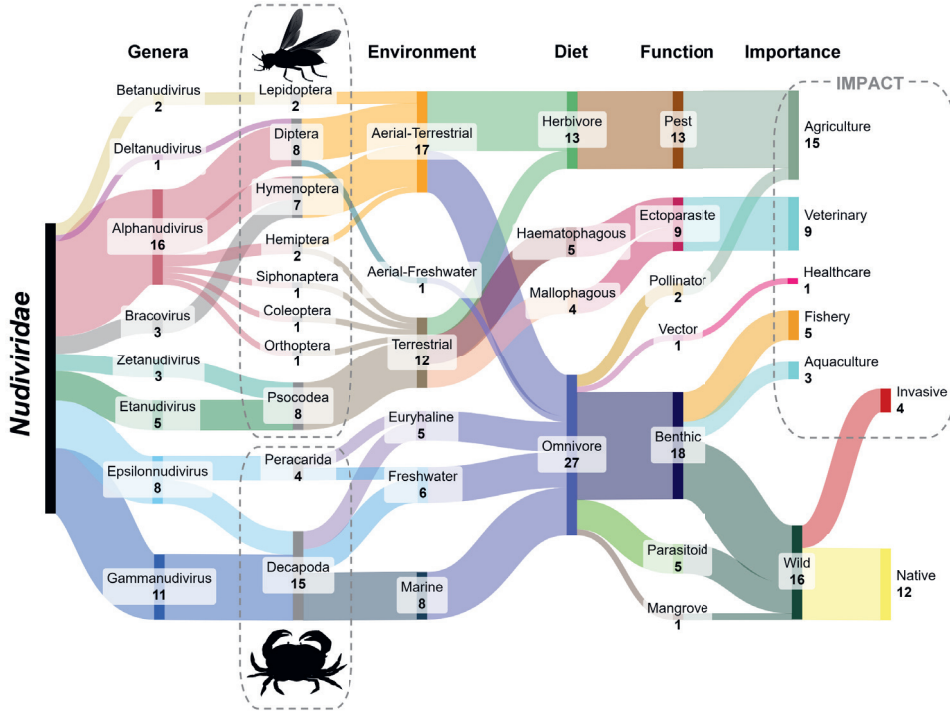


Figure 6. A Sankey diagram, elaborating on the host diversity, environmental diversity, host diet, ecological role, and potential impact associated with nudiviruses and their hosts (Crustacea and Hexapoda). The numbers indicate how many nudiviruses are associated with the respective trait. The noted haematophagous lice (Amblycera) also consume host skin and feathers (mallophagy), in contrast to the exclusively mallophagous ischnoceran lice. The Sankey diagram was generated with the web-based online tool (sankeymatic.com), supported by the online Supplementary file S1.

Ecologically, the 'wild' crustaceans that host nudiviruses all form part of the aquatic benthos. For example, amphipods are considered important nutrient recyclers [244] and the mangrove crab, *Aratus pisonii*, is an essential part of mangrove ecosystems as a detritivorous and semi-terrestrial species [245]. The four invasive non-native species that harbour nudiviruses include: the globally invasive *C. maenas* (aka. invasive green crab), *D. haemobaphes* (aka. the demon shrimp), and the invasive crayfish *F. propinquus* and *F. rusticus* [149, 179, 246, 247]. Some of the fishery and aquaculture species are also considered invasive species in certain locations, such as *C. sapidus* in Europe [248]; however, nudiviruses have not been screened-for in these invasive populations. The nudiviruses in invasive crustacean populations have the potential to increase in prevalence and they may to some extent modify host behaviour [149, 152]; however, there remains a need for broader ecological studies to untangle how these viruses might influence biological invasions – i.e. by potential transmission of nudiviruses to native crustacean populations, comparable to what happened when grey squirrels introduced squirrelpox to red squirrel populations [249]. Finally, there is one putative crustacean host with a parasitic lifestyle. The pea crab, *P. pisum*, parasitises mussels, and specimens with a histologically-demonstrated nudivirus infection were collected from the south coast of the UK [187]. However, no genomic data has been collected to confirm the relationship with the *Nudiviridae*.

Overall, a reasonable diversity of known nudiviruses are associated with facultative or obligate ectoparasitic/parasitoid hosts – fleas, lice, wasps, mosquito, pea crabs. Most terrestrial nudiviruses have been detected in insect hosts involved with agricultural systems, but their occurrence is likely much

broader, whereas several crustacean hosts are associated with fisheries and aquaculture. For the majority, these viruses sit within insect/crustacean taxa with an herbivorous or omnivorous lifestyle. The ecological role of the latter group are often prey species, or have a role in nutrient recycling.

Concluding remarks

Histology and electron microscopy identified the first nudiviruses, and now genomic data provides us with a vast resource of nudivirus diversity across invertebrate hosts. The complete genomes of eight nudiviruses (CafNV, CoNV, EcNV, FrpNV, FrrNV, HxsNV, LpNV, MpNV) were hidden in cyberspace from pre-published studies, while an additional eight nudiviruses were identified as partial gene sets or from transcriptomic data (AlNV, CharyNV, CpNV, EbNV, EsNV, GlNV, GpNV, GfNV). Five other resources hint at the presence of putative nudiviruses, but with insufficient assembly recovery (**Supplementary data, Table S1**). These new nudiviruses, coupled with the isolates sequenced between 2007 – 2023 [40, 41, 54, 59, 83, 92, 113, 175, 179, 180, 243], have allowed for a new taxonomic perspective, suggesting two new genera ('Etanudivirus' and 'Zetanudivirus') to house nudiviruses from ectoparasitic lice. Nudiviruses first emerged around 280 Mya – crustacean nudivirus infections may have once been considered the older lineage; however, they appear to have the youngest evolutionary history. Insect-infecting nudiviruses appear to be the first to emerge, including their potential evolution and diversification in lice on early birds in the Cretaceous period. Future studies are required to address questions regarding the morphology and pathology of these new nudiviruses, which may in turn reveal their ecological importance and new prospects for biocontrol of deleterious mammalian and avian ectoparasites, where chemical resistance is common [250]. Our findings merely scratch the surface, but we are hopeful that our study will act as a steppingstone for forthcoming exploration into nudiviral taxonomic reasoning to develop delineation criteria and explore further nudivirus diversity across the Arthropoda.

Acknowledgements

The writing of this paper has received funding from the European Union's Horizon 2020 research and innovation programme INSECT DOCTORS under the Marie Skłodowska-Curie grant agreement No. 859850. JB and ALB would like to thank Teesside University for PhD funding to support ALB. We would like to thank the Journal of Structural Biology and the journal Virology for the approval to reproduce pre-published images. JMP would like to thank the Wageningen Electron Microscopy Centre for the help in obtaining the negatively stained micrograph of HzNV-1.

Supplementary data

Table S1. Overview of SRA datasets that resulted in the assembly of full nudiviral genomes, or partial expected nudiviral genes from insect and crustacean datasets.

Organism (common name; Order)	SRA Accession Code	Platform	Source	Location of collection	Year of collection	Sample collected by
Nudiviruses found in datasets from insects						
Ctenocephalides orientis (oriental cat flea; Siphonaptera)	SRR23199466 SRR23199468 SRR23199472	Illumina	MG	Malaysia: Perak	2019	[200]
Ctenocephalides orientis (oriental cat flea; Siphonaptera)	SRR23199467 SRR23199469 SRR23199473	Minion	MG	Malaysia: Perak	2019	[200]
Cuculoecus africanus (chewing louse; Psocodea)	SRR5308372	Illumina	MG	Ghana: Goaso	2003	Kevin P. Johnson
Culex pipiens (mosquito; Diptera)	*SRR16905219	Illumina	MT	China: Qinghua cave	2018	[201]
Echinophilopterus claytoni (chewing louse; Psocodea)	SRR5308360	Illumina	MG	Australia: New South Wales	2001	Kevin P. Johnson
Franciscoloa roseicapillae (chewing louse; Psocodea)	SRR20711461 SRR20711469	Illumina	MG	Australia: Queensland	2019	[198]
Franciscoloa pallida (chewing louse; Psocodea)	SRR20711460	Illumina	MG	Australia: New South Wales	2019	[198]
Goniodes lagopi (chewing louse; Psocodea)	*SRR11871036[a] *SRR11871059[b]	Illumina	MG	USA: Alaska	2015[a] 2018[b]	[199]
Heterodoxus spiniger (chewing louse; Psocodea)	SRR5308125	Illumina	MG	India: Uttar	2001	Arun Kumar Saxena
Lagopoecus perplexus (chewing louse; Psocodea)	SRR11871040	Illumina	MG	USA: Alaska	2017	[199]
Myrsidea ptilorhynchi (chewing louse; Psocodea)	SRR20711472	Illumina	MG	Australia: Queensland	2019	[198]
Putative nudiviruses found in datasets from insects						
Aleuroclava psidii (Asian whitefly; Hemiptera)	SRR16114381	Illumina	MG	China: Xinyang	2019	[203]
Dasineura brassicae (brassica pod midge; Diptera)	SRR18053992	MGISEQ	MT	Göttingen: Germany	missing	[204]

Phlebotomus chinensis (sandfly; Diptera)	SRR19974249	Illumina	MT	China: Hejin	2019	[205]
Trichocera saltator (Diptera)	SRR921653	Illumina	MT	Germany: Jena	2011	[207, 251]
Nudiviruses found in datasets from crustaceans						
Gammarus pulex (freshwater shrimp; Amphipoda)	*SRR8089720	Illumina	MT	France: Cressoniere	2018	French National Research Agency
Gammarus fossarum (freshwater shrimp; Amphipoda)	*SRR8089724	Illumina	MT	-	2018	French National Research Agency
Echinogammarus berilloni (beach hopper; Amphipoda)	*SRR8089732	Illumina	MT	-	2018	French National Research Agency
Astacus leptodactylus (European crayfish; Decapoda)	*SRR650486	Illumina	MT	-	2012	[202]
Eriocheir sinensis (Chinese mitten crab; Decapoda)	*SRR9734062 *SRR9734063	Illumina	MG	China	2019	Soochow University
Eriocheir sinensis (Chinese mitten crab; Decapoda)	*SRR9822782 *SRR9822783 *SRR9822784	Illumina	MT	China	2019	Soochow University
Charybdis sp. (crab; Decapoda)	*SRR3401303	Illumina	MT	-	2016	[155]
Putative nudiviruses found in datasets from crustaceans						
Niphargus hrabei (freshwater shrimp; Amphipoda)	*SRR13297208 *SRR13297209	Illumina	MT	Hungary: Molnar Janos Cave	2020	[208]

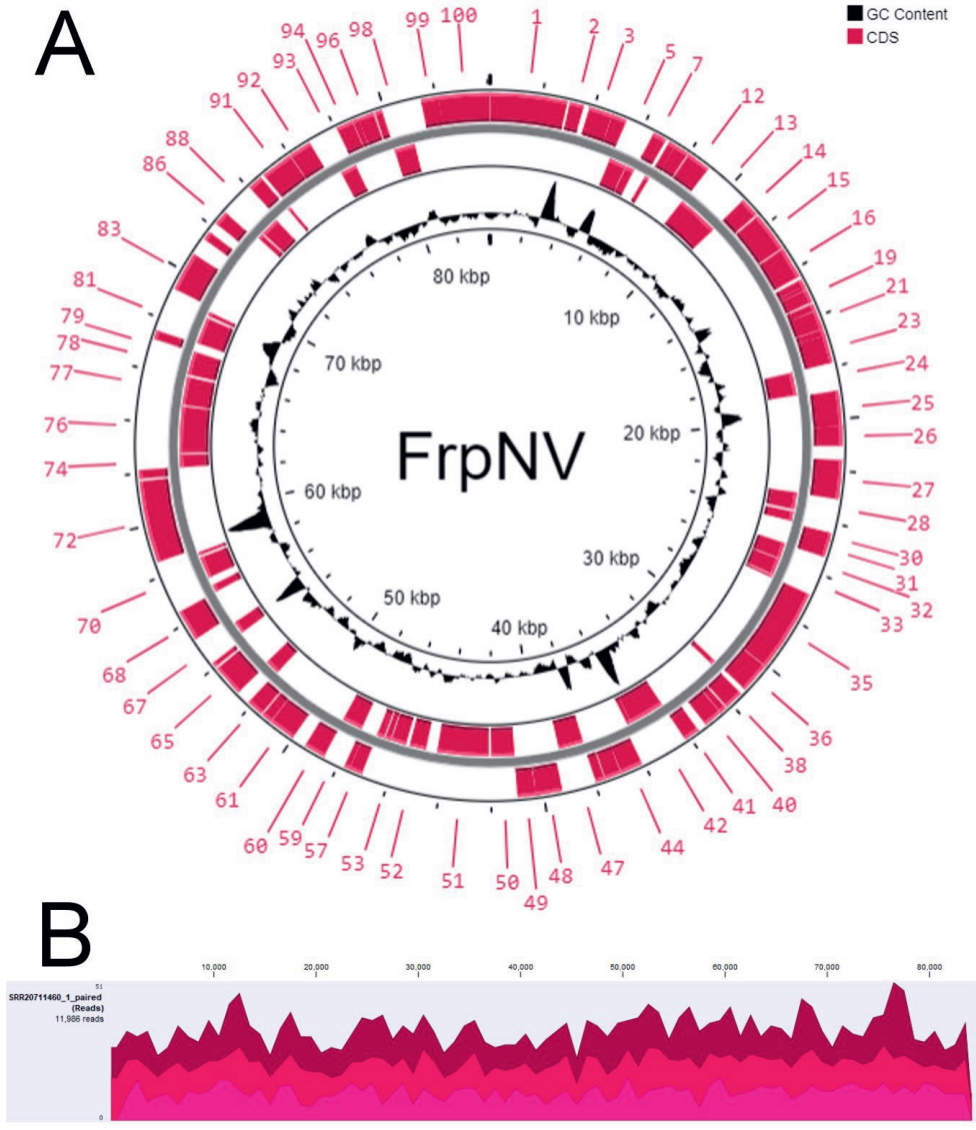


Figure S1. Circular genome (84,091 bp) of Franciscoloa pallida nudivirus (FrpNV). **(A)** First layer: locations of coding sequences (CDS) indicated with their respective gene numbers. Second layer: Genes coding on the positive strand. Third layer: Genes coding on the negative strand. Fourth layer: GC content. Fifth layer: Ticks indicating the location in the genome in kilobase pairs (kbp). **(B)** Coverage map from CLC genomics workbench v12 (Qiagen). Data acquired from SRR20711460 were re-assembled (genome accession: BK068076) using SPAdes v3.15.3 [252] and annotated using GeneMarkS (virus; [253]).

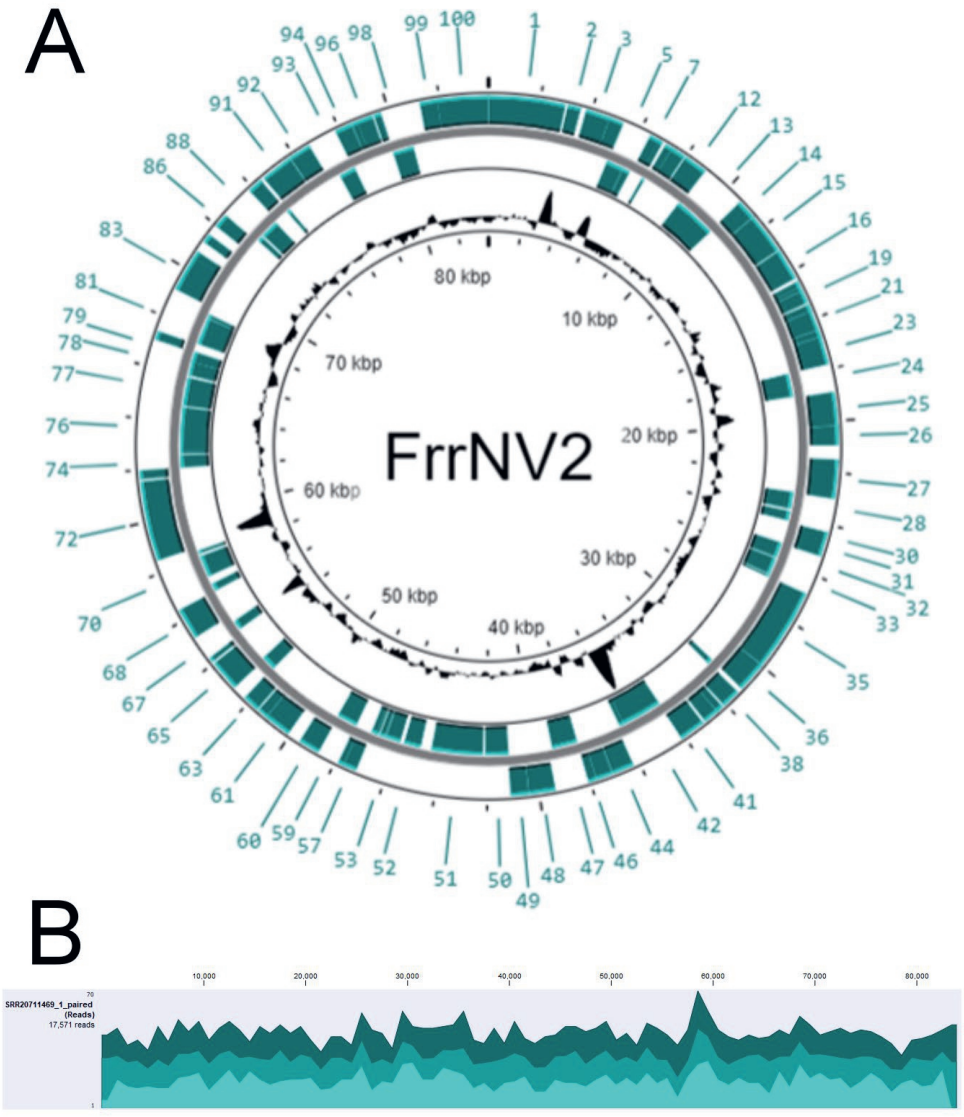


Figure S2. Circular genome (83,884 bp) of Franciscoloa roseicapillae nudivirus (FrrNV). **(A)** First layer: locations of coding sequences (CDS) indicated with their respective gene numbers. Second layer: Genes coding on the positive strand. Third layer: Genes coding on the negative strand. Fourth layer: GC content. Fifth layer: Ticks indicating the location in the genome in kilobase pairs (kbp). **(B)** Coverage map from CLC genomics workbench v12 (Qiagen). Data acquired from SRR20711469 were re-assembled (genome accession: BK068077) using SPAdes v3.15.3 [252] and annotated using GeneMarkS (virus; [253]).

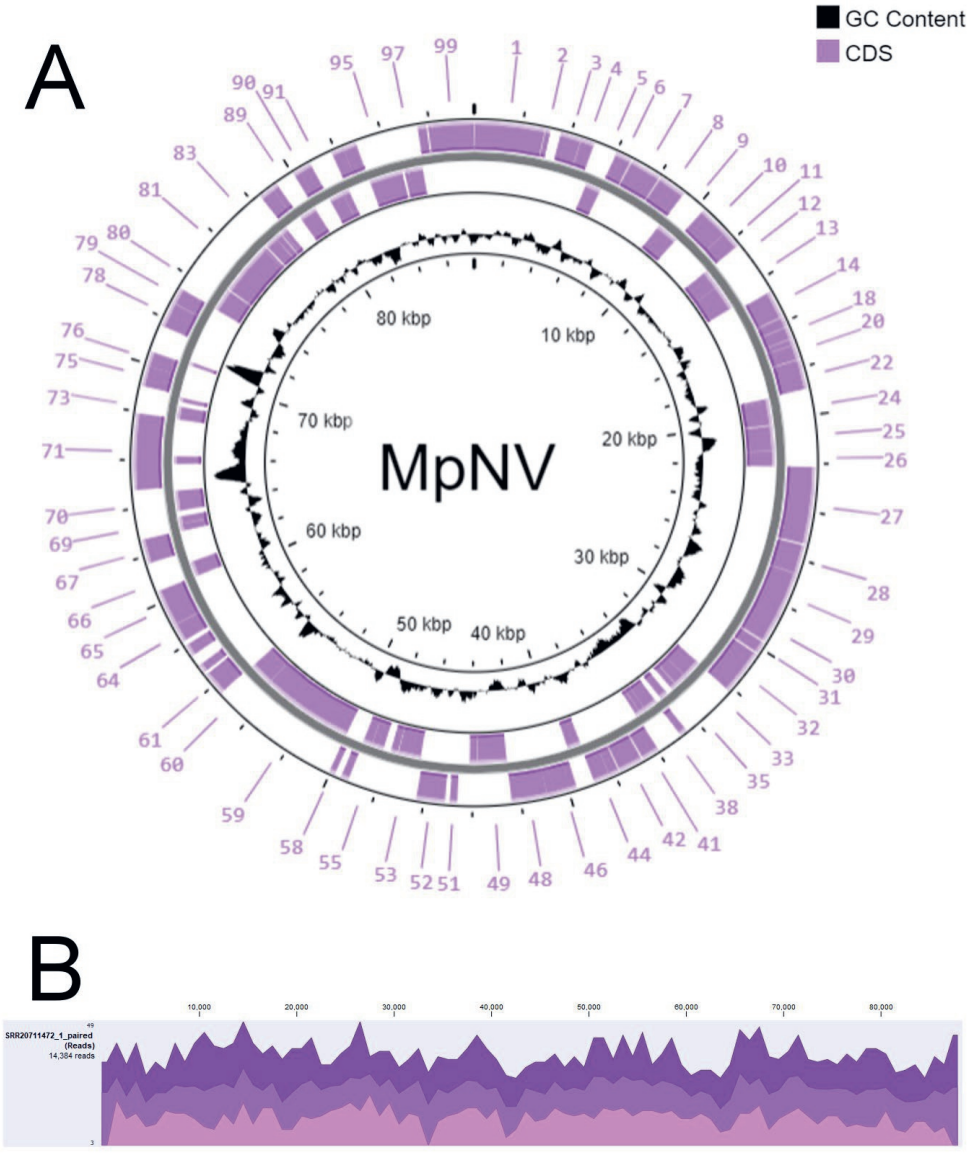


Figure S3. Circular genome (87,845 bp) of Myrsidea ptilorhynchi nudivirus (MpNV). **(A)** First layer: locations of coding sequences (CDS) indicated with their respective gene numbers. Second layer: Genes coding on the positive strand. Third layer: Genes coding on the negative strand. Fourth layer: GC content. Fifth layer: Ticks indicating the location in the genome in kilobase pairs (kbp). **(B)** Coverage map from CLC genomics workbench v12 (Qiagen). Data acquired from SRR20711472 were re-assembled (genome accession: BK068080) using SPAdes v3.15.3 [252] and annotated using GeneMarkS (virus; [253]).

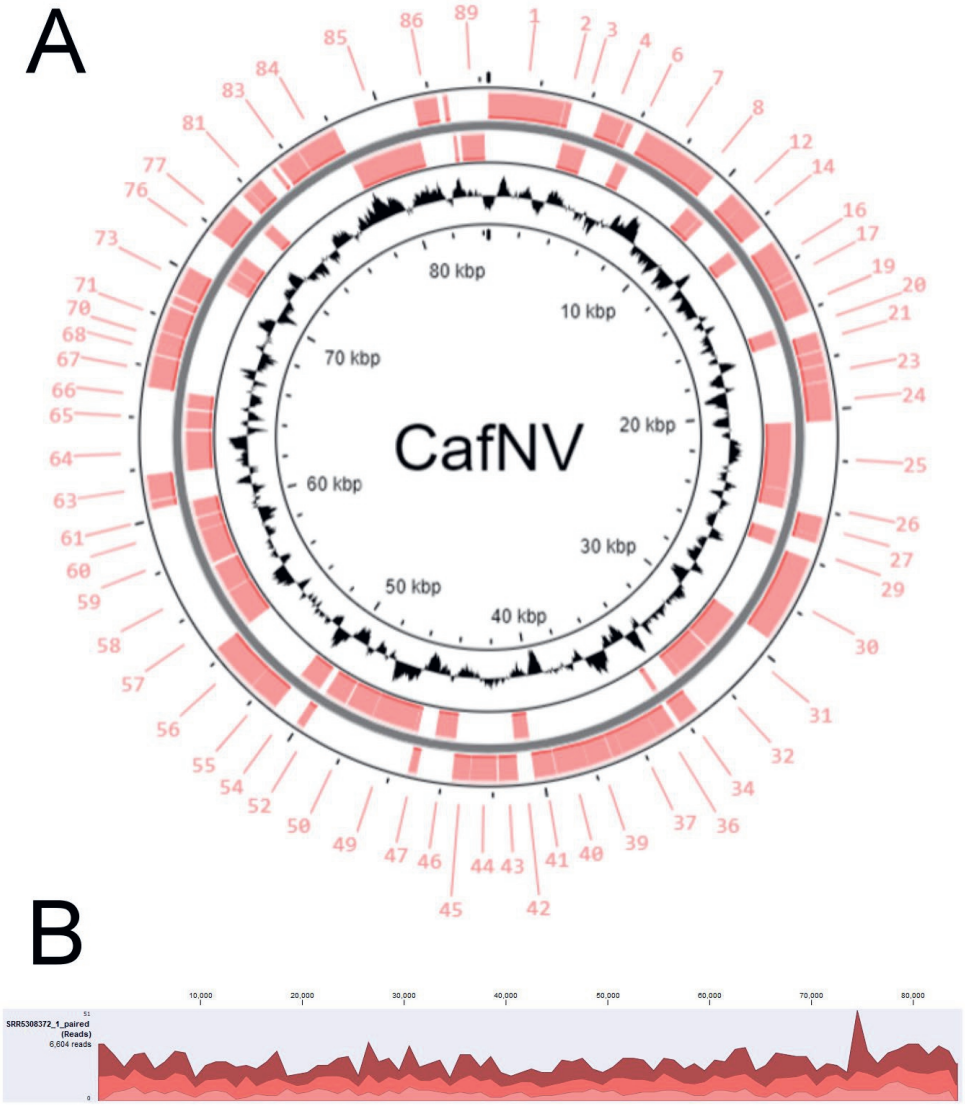


Figure S4. Circular genome (84,329 bp) of Cuculoecus africanus nudivirus (CafNV). **(A)** First layer: locations of coding sequences (CDS) indicated with their respective gene numbers. Second layer: Genes coding on the positive strand. Third layer: Genes coding on the negative strand. Fourth layer: GC content. Fifth layer: Ticks indicating the location in the genome in kilobase pairs (kbp). **(B)** Coverage map from CLC genomics workbench v12 (Qiagen). Data acquired from SRR5308372 were re-assembled (genome accession: BK068074) using SPAdes v3.15.3 [252] and annotated using GeneMarkS (virus; [253]).

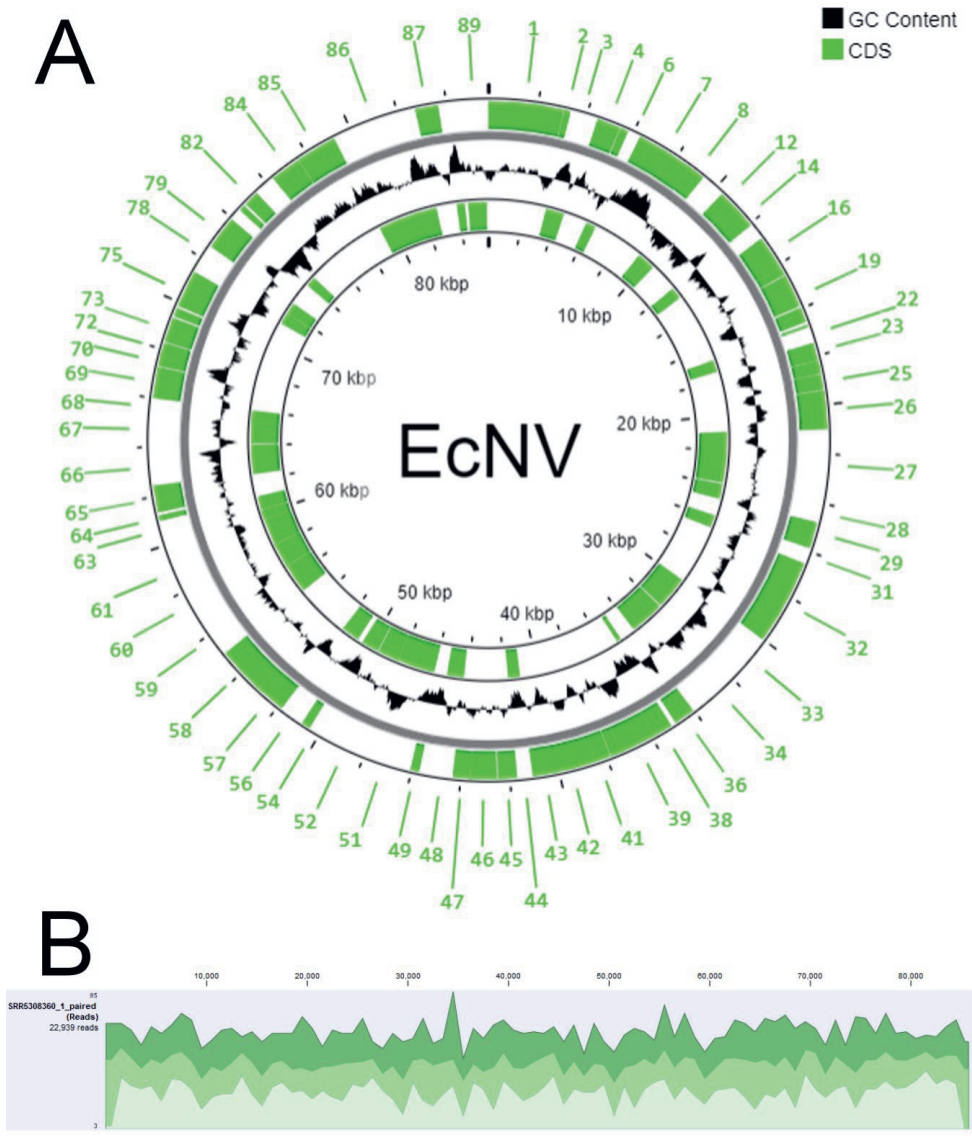


Figure S5. Circular genome (85,720 bp) of Echinophilopterus claytoni nudivirus (EcNV). **(A)** First layer: locations of coding sequences (CDS) indicated with their respective gene numbers. Second layer: Genes coding on the positive strand. Third layer: Genes coding on the negative strand. Fourth layer: GC content. Fifth layer: Ticks indicating the location in the genome in kilobase pairs (kbp). **(B)** Coverage map from CLC genomics workbench v12 (Qiagen). Data acquired from SRR5308360 were re-assembled (genome accession: BK068075) using SPAdes v3.15.3 [252] and annotated using GeneMarkS (virus; [253]).

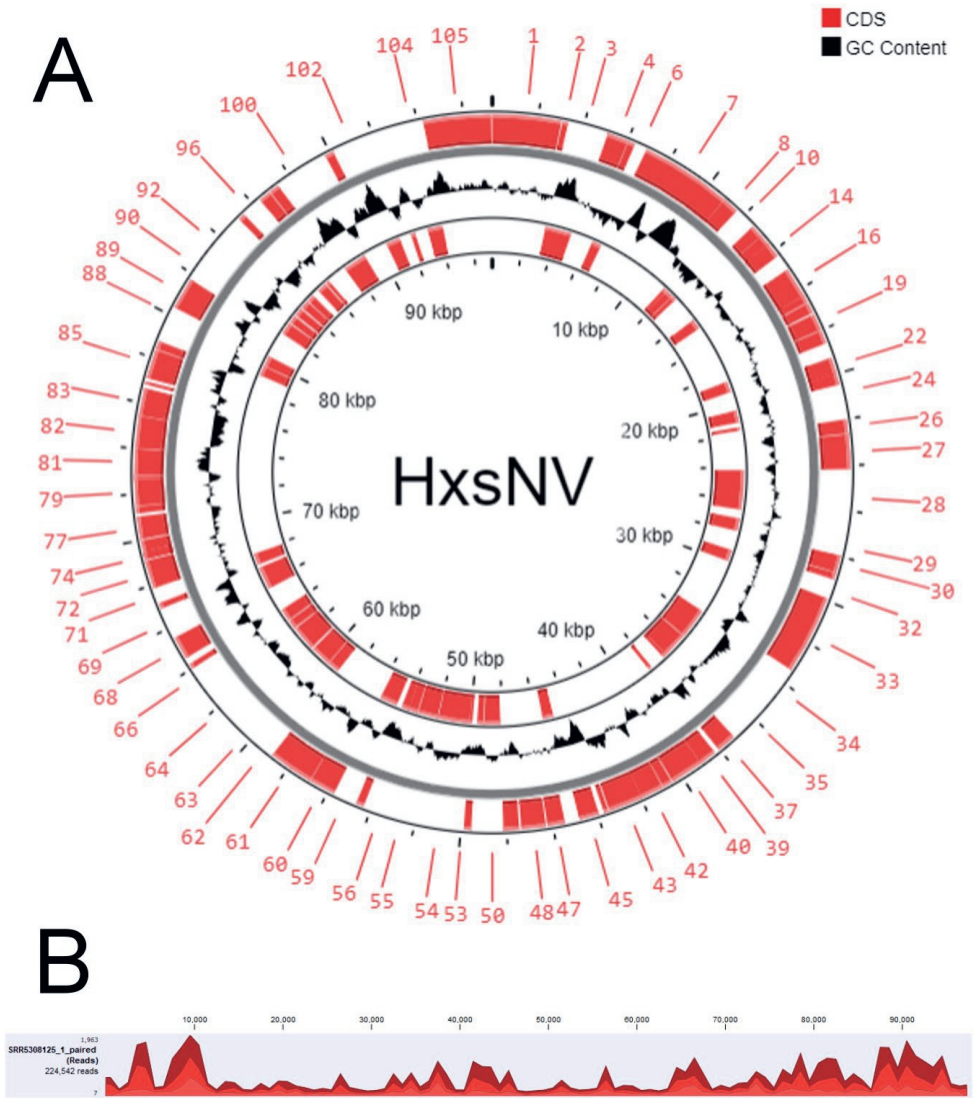


Figure S6. Circular genome (97,269 bp) of Heterodoxus spiniger nudivirus (HxsNV). **(A)** First layer: locations of coding sequences (CDS) indicated with their respective gene numbers. Second layer: Genes coding on the positive strand. Third layer: Genes coding on the negative strand. Fourth layer: GC content. Fifth layer: Ticks indicating the location in the genome in kilobase pairs (kbp). **(B)** Coverage map from CLC genomics workbench v12 (Qiagen). Data acquired from SRR5308125 were re-assembled (genome accession: BK068078) using SPAdes v3.15.3 [252] and annotated using GeneMarkS (virus; [253]).

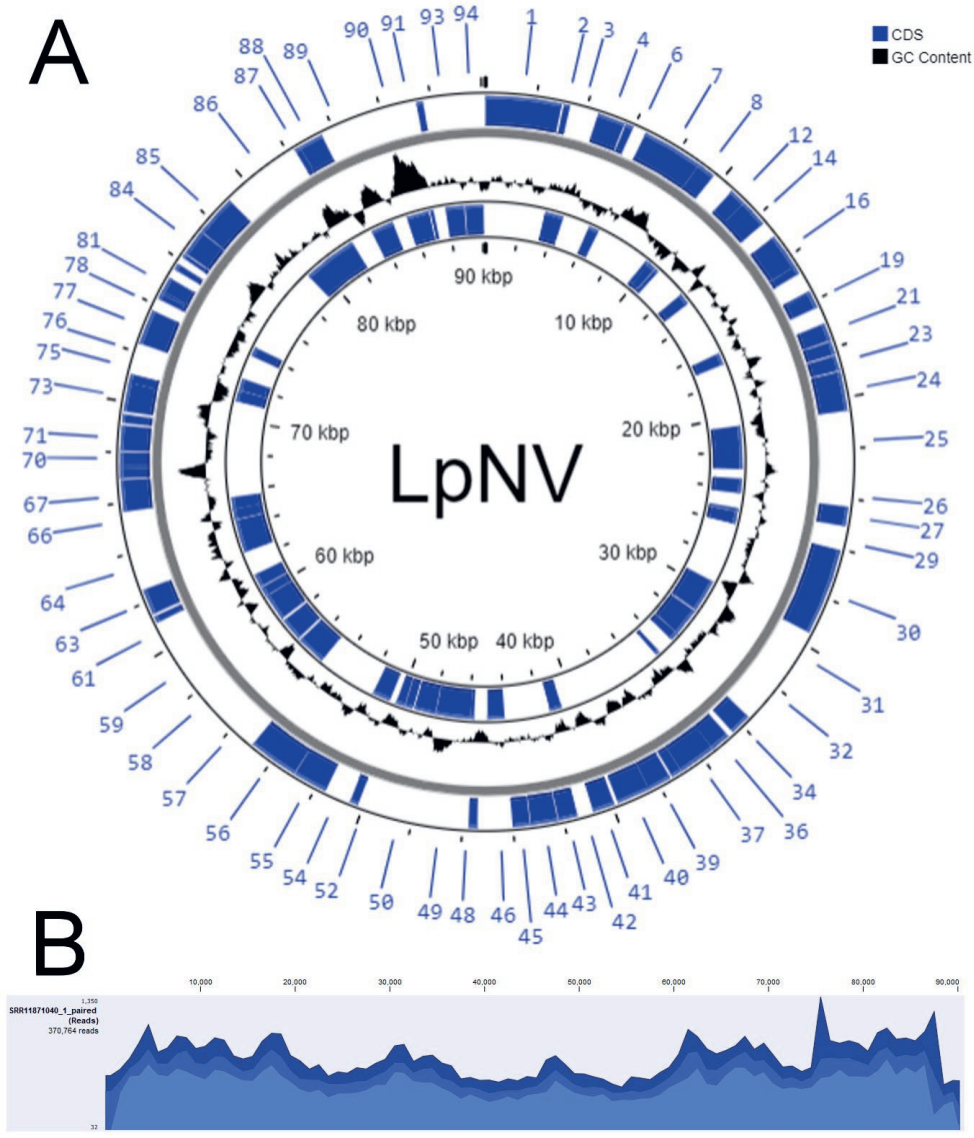


Figure S7. Circular genome (90,164 bp) of Lagopoecus perplexus nudivirus (LpNV). **(A)** First layer: locations of coding sequences (CDS) indicated with their respective gene numbers. Second layer: Genes coding on the positive strand. Third layer: Genes coding on the negative strand. Fourth layer: GC content. Fifth layer: Ticks indicating the location in the genome in kilobase pairs (kbp). **(B)** Coverage map from CLC genomics workbench v12 (Qiagen). Data acquired from SRR11871040 were re-assembled (genome accession: BK068079) using SPAdes v3.15.3 [252] and annotated using GeneMarkS (virus; [253]).

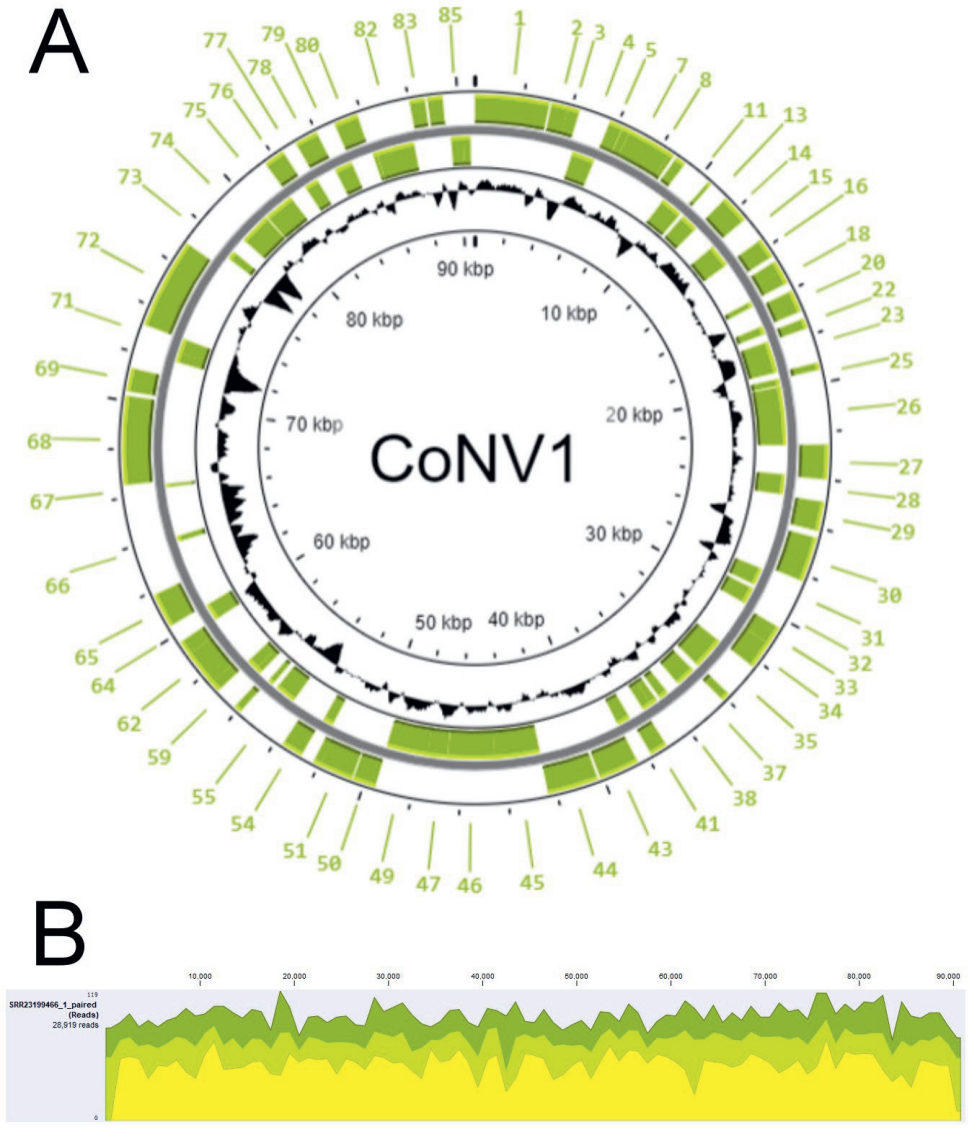


Figure S8. Circular genome (90,736 bp) of Ctenocephalides orientis nudivirus 1 (CoNV_1). **(A)** First layer: locations of coding sequences (CDS) indicated with their respective gene numbers. Second layer: Genes coding on the positive strand. Third layer: Genes coding on the negative strand. Fourth layer: GC content. Fifth layer: Ticks indicating the location in the genome in kilobase pairs (kbp). **(B)** Coverage map from CLC genomics workbench v12 (Qiagen). Data acquired from SRR23199466 were re-assembled (genome accession: BK068081) using SPAdes v3.15.3 [252] and annotated using GeneMarkS (virus; [253]).

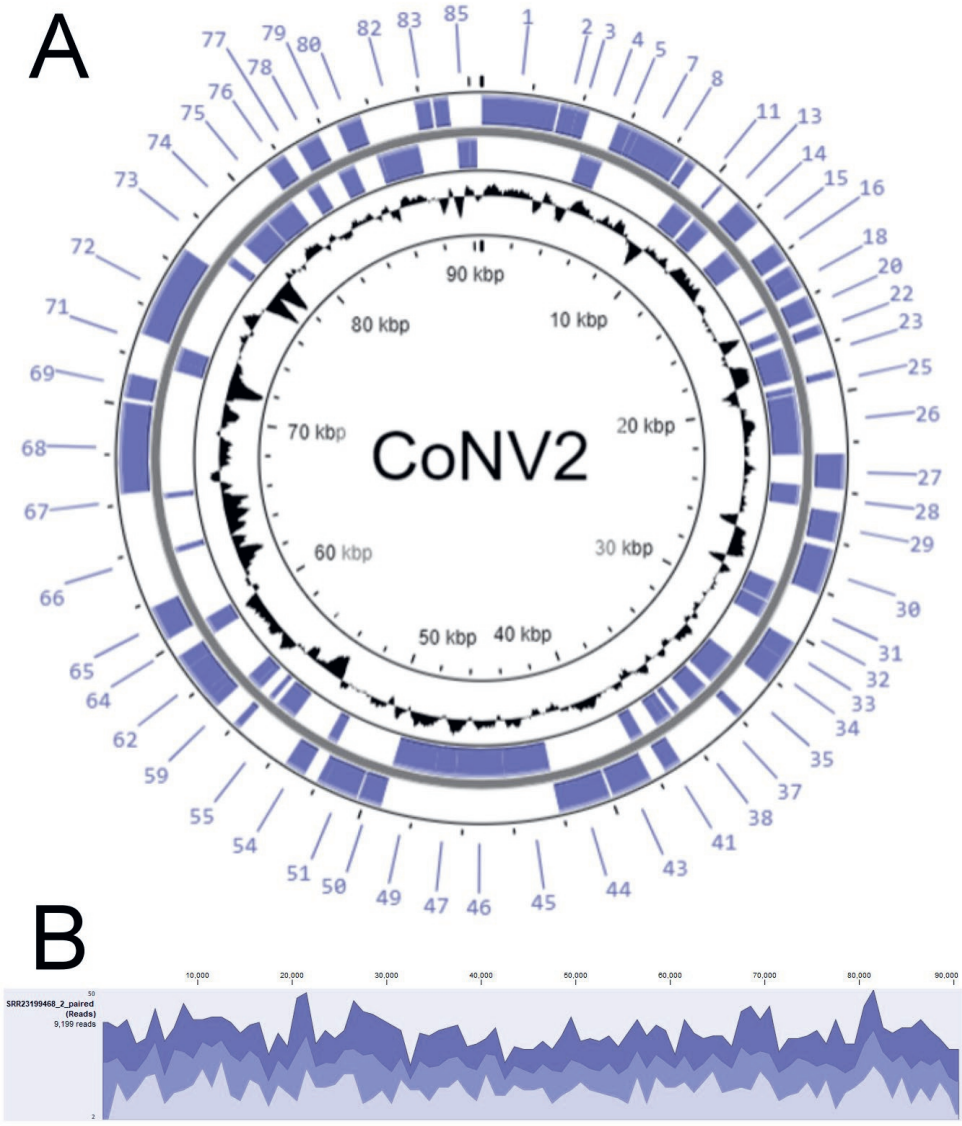


Figure S9. Circular genome (90,491 bp) of Ctenocephalides orientis nudivirus 2 (CoNV_2). **(A)** First layer: locations of coding sequences (CDS) indicated with their respective gene numbers. Second layer: Genes coding on the positive strand. Third layer: Genes coding on the negative strand. Fourth layer: GC content. Fifth layer: Ticks indicating the location in the genome in kilobase pairs (kbp). **(B)** Coverage map from CLC genomics workbench v12 (Qiagen). Data acquired from SRR23199468 were re-assembled (genome accession: BK068082) using SPAdes v3.15.3 [252] and annotated using GeneMarkS (virus; [253]).

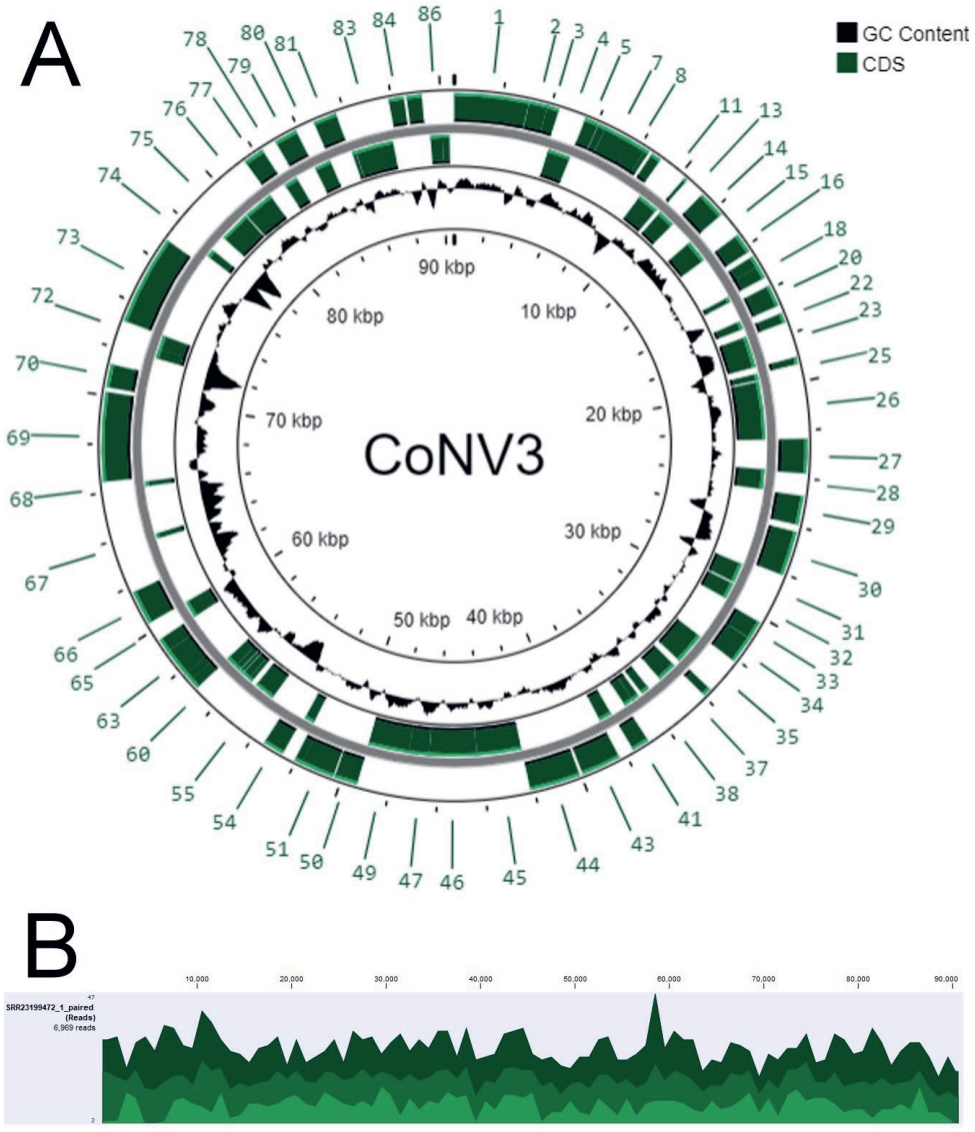


Figure S10. Circular genome (90,583 bp) of Ctenocephalides orientis nudivirus 3 (CoNV_3). **(A)** First layer: locations of coding sequences (CDS) indicated with their respective gene numbers. Second layer: Genes coding on the positive strand. Third layer: Genes coding on the negative strand. Fourth layer: GC content. Fifth layer: Ticks indicating the location in the genome in kilobase pairs (kbp). **(B)** Coverage map from CLC genomics workbench v12 (Qiagen). Data acquired from SRR23199472 were re-assembled (genome accession: BK068083) using SPAdes v3.15.3 [252] and annotated using GeneMarkS (virus; [253]).

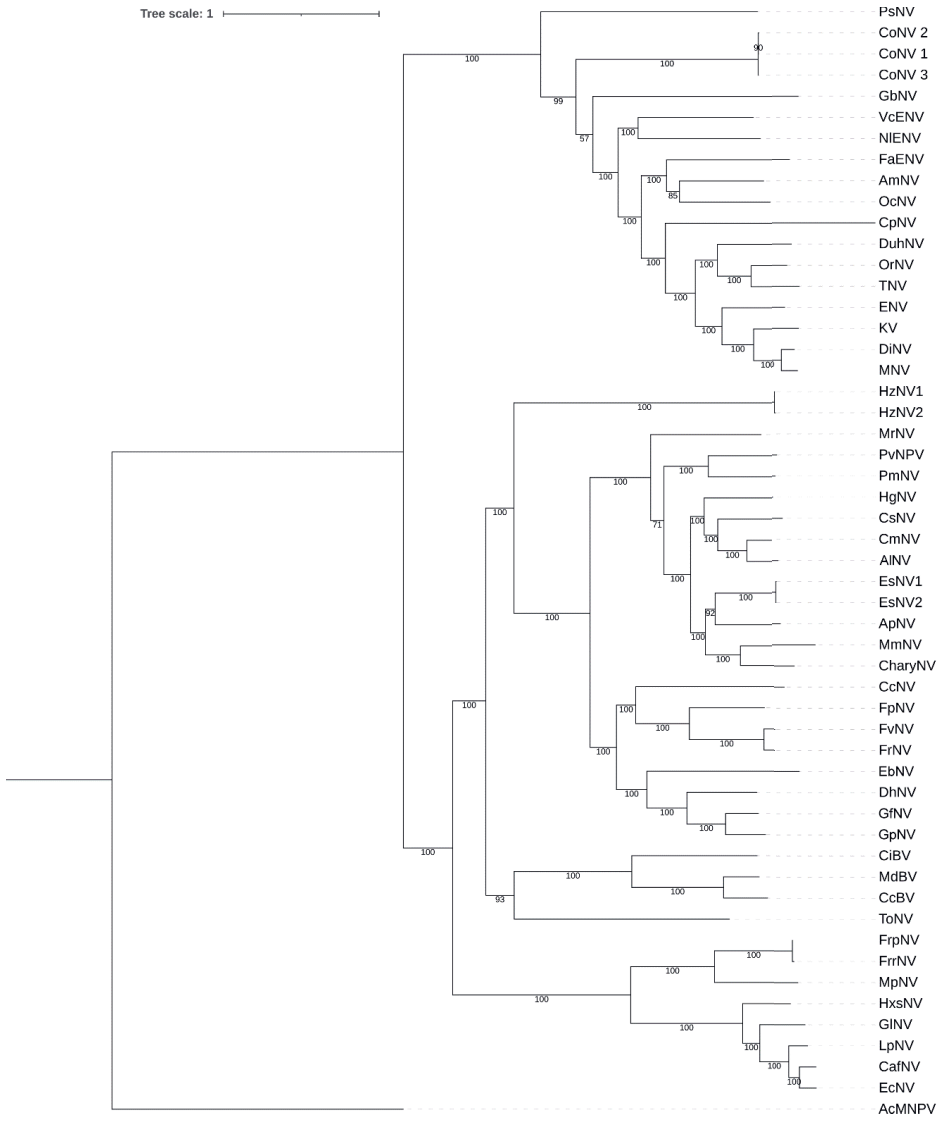


Figure S11. Phylogenetic tree with 49 nudiviruses, three bracoviruses and the baculovirus Autographa californica nucleopolyhedrovirus (AcMNPV) as outgroup, coupled with intragenus core gene synteny compared among nudiviruses with complete genomes. The amino acid sequences of 17 conserved nudiviral core genes were aligned with the MAFFT [254] alignment tool (v7.490) of Geneious (v2023.2.1) under default settings. Best-fit substitution models were determined separately for each core gene based on the Bayesian information criterion (BIC) generated by ModelFinder [255] in IQ-Tree [210] (Supplementary S2, Table S1). The multi-protein alignment file, containing 17,906 total sites, and the model information were subjected to maximum-likelihood (ML) phylogenetic analysis with 1000 bootstrap replicates. The inferred treefile was visualised with the online tool iTOL [211].

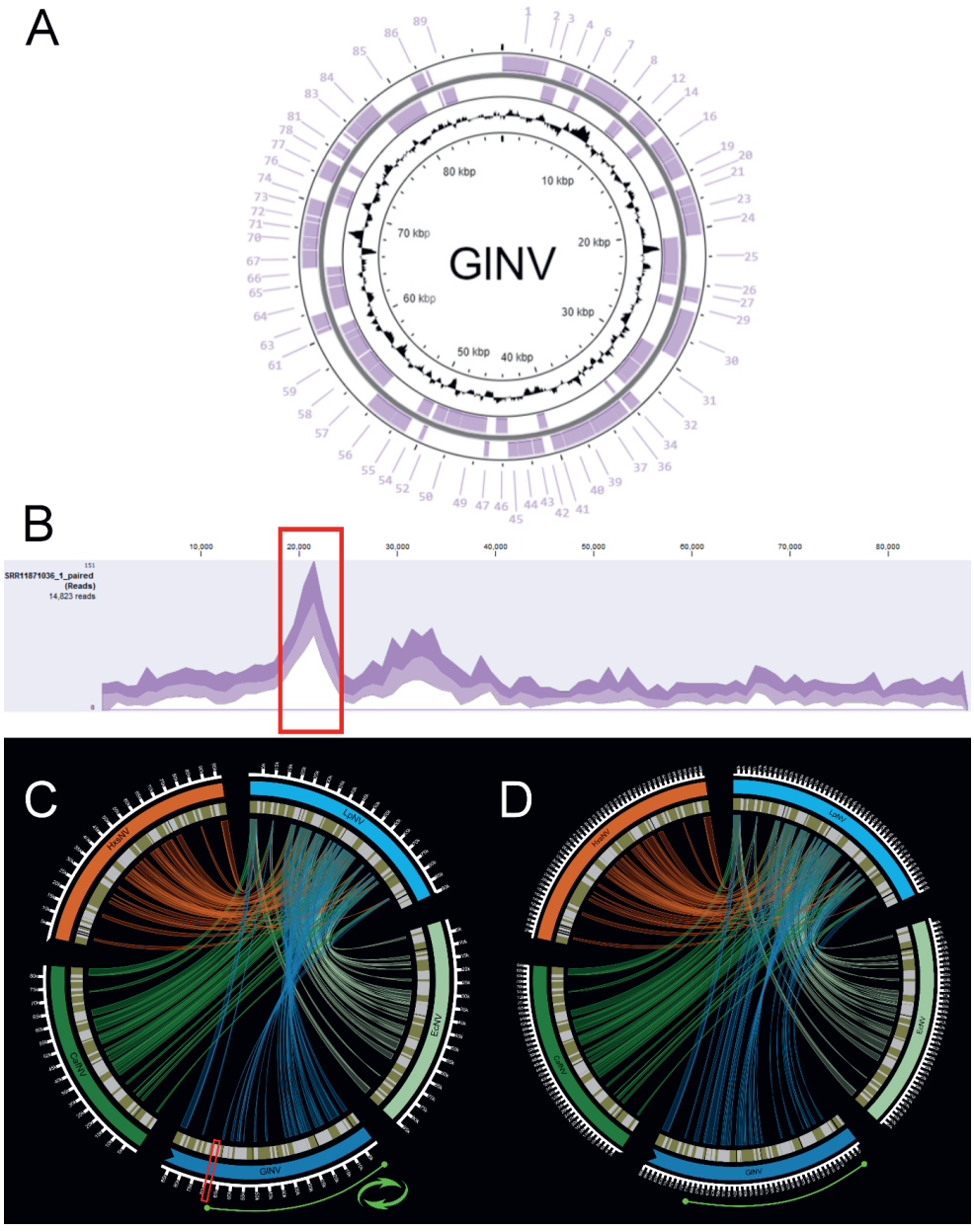


Figure S12. Schematic explanation of GlNV's potentially incomplete genome. **(A)** Putative circular genome (88,040 bp) of Goniodes lagopi nudivirus (GlNV). First layer: open reading frames (Genes) indicated with their respective gene numbers. Second layer: Genes coding on the positive strand. Third layer: Genes coding on the negative strand. Fourth layer: GC content. Fifth layer: Ticks indicating the location in the genome in kilobase pairs (kbp). **(B)** Coverage of mapped reads against the assembled GlNV genome resulted in an extraordinarily high peak (framed in red) that represents a TA-rich region in the genome. **(C)** Circular core gene synteny plot of etanudiviruses. In GlNV, after the TA-rich region onwards (indicated with green line) the arrangement of the nudiviral core genes is almost perfectly inverted compared to the other etanudiviruses. **(D)** Inversion of the putatively inverted genome part of GlNV results in a more congruent core gene synteny with the other

etanudiviruses. Other molecular analyses are needed (e.g. PCR) to elucidate the true nature of this region in GlNV's genome. Data acquired from SRR11871036 were re-assembled (contig accessions: BK068084, BK068085) using SPAdes v3.15.3 [252] and annotated using GeneMarkS (virus; [253]).

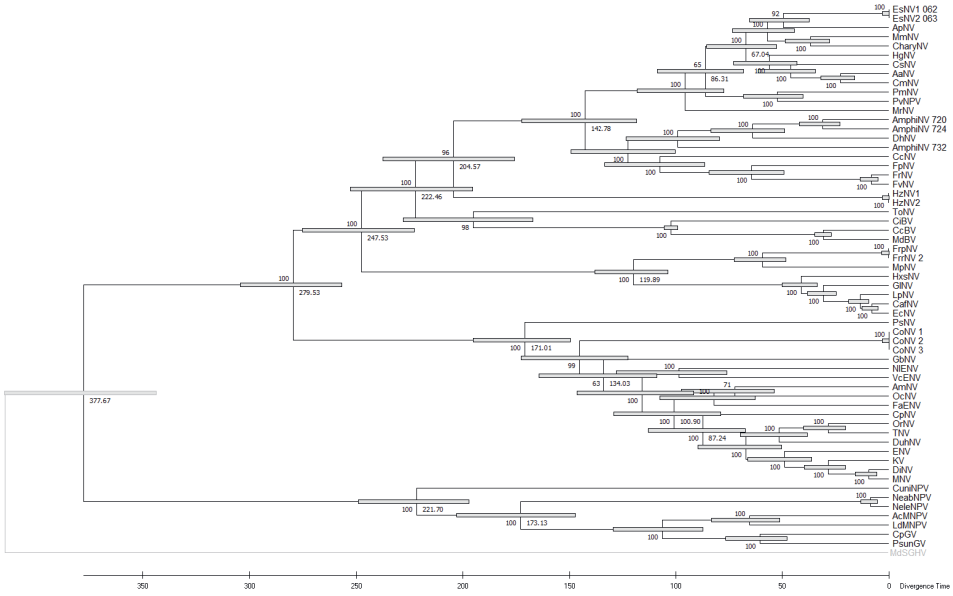


Figure S13. Original molecular dating tree of viruses computed in MEGA-X showing age estimates, and 95% CI values of numbered nodes as error bars. A phylogenetic tree with optimised branch lengths served as input to compute the molecular dating tree with the MEGA11 (v11.0.10) software [212] by using the Clocks feature, and RelTime-Branch Lengths option [256]. A single calibration constraint of 103.38 ± 4.41 Mya was specified with log-normal distribution which represents the most recent common ancestor (MRCA) of Chelonus inanitus bracovirus (CiBV) and Cotesia congregata bracovirus (CcBV) [215].

Methodology

Best-fit model for each nudiviral core gene for phylogenetic analysis. For both the non-clock and clock tree, the respectively best-fit substitution models for each of the 17 aligned protein datasets were determined based on the Bayesian information criterion (BIC) generated by ModelFinder (**Table S2 and S3**). If applicable, the ModelFinder-inferred numerical values of invariant sites (+I) and rate heterogeneity (+G) were included in the *.nexus file of the phylogenetic analysis. Using the multiprotein alignment file and nexus file with partitions as input for IQtree2, a phylogenetic tree was generated under 1000 bootstraps.

Table S2. Partitions per core gene used for the non-clock phylogenetic analysis of *Nudiviridae.*

Partition	Protein model and parameters
Pif-1, Pif-2, vp91, lef-9	WAG+F+I+G4
Pif-3, pif-6	Q.yeast+I+G4
Pif-4, 38k, p33	VT+F+I+G4
Vp39, ac81	LG+F+I+G4
DNApol, helicase, lef-4, lef-8, p47, p74	Q.pfam+F+I+G4

The *.treefile output generated by IQtree2 under the model information from Table S2 was used to create **Figure 2** in the main manuscript. The *.treefile contains the maximum-likelihood tree.

Table S3. Partitions per core gene used to infer the phylogenetic tree for the molecular dating analysis.

Partition	Protein model and parameters
pif-1, pif-2, vp91	WAG+F+I+G4
pif-3, pif-6	Q.yeast+I+G4
p33, lef-9	VT+F+I+G4
vp39, ac81	LG+F+I+G4
dnapol, helicase, lef-4, lef-8, p47, p74, Pif-4, 38k	Q.pfam+F+I+G4

The *.contree file generated by IQtree2 under the model information from Table S3 was used to create the virus time-tree in the main manuscript (**Figure 4A**). The *.contree file contains a consensus tree whose branch lengths were optimised based on the original alignment.

Molecular dating tree and molecular clock calibration. The *.contree file generated with IQtree2 was used as input for the RelTime-Branch Lengths tool in MEGA11. The hytrosavirus MdSGHV was specified as the outgroup, and the MRCA of CiBV and CcBV (103.38 ± 4.41 Mya) set with a log-normal distribution as the calibration node. The Maximum Rate Ratio was adjusted to the highest value of 100, and the molecular dating tree generated. The 95% confidence intervals (CIs) produced by RelTime were extracted from MEGA11 with "Export Timetree (Tabular)".

Testing for natural selection in nudiviral core genes of bracoviruses and exogenous nudiviruses using the dN/dS metric. Nudiviral core genes of bracoviruses and exogenous nudiviruses were individually codon-aligned utilizing the "Translation align" and "MAFFT alignment" tools in Geneious. In order to assess the individual evolutionary rates of the bracovirus and nudivirus lineage, the alignment files of each core gene were split into two versions, one containing only bracoviruses and one containing only nudiviruses. Those alignment files were then used as input for IQtree2 to generate individual trees, which could then be used together with the respective alignment file as input for the Fixed Effects Likelihood (FEL) [217] analysis in HyPhy [216].

Table S4 Nudiviral core genes and their dN/dS values calculated by HyPhy under the FEL model. In total, four bracoviruses and 32 exogenous nudiviruses were included in the analysis.

	dN/dS	
Nudiviral core gene	**Bracoviruses** (CcBV, MdBV, CiBV, Cotesia vestalis bracovirus)	**Nudiviruses** (AmNV, ApNV, CafNV, CcNV, CmNV, CoNV1-2, CsNV, DhNV, DiNV, EcNV, ENV, FrpNV, FrrNV, GbNV, HgNV, HxsNV, HzNV-1 and -2, KV, LpNV, MmNV, MNV, MpNV, MrNV, OrNV, PmNV, PsNV, PvNPV, TNV, ToNV)
lef-4	0.079	0.090
lef-8	0.058	0.052
pif-0/p74	0.315	0.057
pif-1	0.301	0.079
pif-2	0.210	0.048
pif-8/vp91	0.278	0.118
38k	0.133	0.051
vp39	0.227	0.034

“**I knew what she spoke of...**

To defeat the necromancer

I had to use love!

‘Stand back, Brian, I'm going to hug him’!”

– The Mystic Crystal by Ninja Sex Party (2020)

Chapter 4

Embracing the enemy: Macropinocytosis as the cell entry mechanism of Heliothis zea nudivirus 1

Jirka Manuel Petersen[1,2], Els Roode[1], Christine van Battum[1], Monique M. van Oers[1]

[1]Laboratory of Virology, Wageningen University and Research, 6708 PB Wageningen, the Netherlands.

[2]Institut de Recherche sur la Biologie de l'Insecte, UMR7261 CNRS - Université de Tours, 37200 Tours, France.

Abstract

The Heliothis zea nudivirus 1 (HzNV-1) and Helicoverpa zea nudivirus 2 (HzNV-2) are close relatives within the *Betanudivirus* genus (family *Nudiviridae*). In contrast to HzNV-2, HzNV-1 has lost its ability to infect living insects and can only replicate in cell lines. Still, HzNV-1 provides a valuable model for studying nudivirus infection mechanisms under controlled conditions, and studies on nudiviral pathology are scarce when compared to their well-studied relatives of the *Baculoviridae* family. Thus, by employing microscopic and molecular methods, we elucidated different aspects of HzNV-1's cytopathological and molecular dynamics in the ovarian HZ-AM1 cell line derived from *Helicoverpa zea*. Using electron microscopy, we observed that HzNV-1 virions enter ovarian cells in a macropinocytosis-like manner, similar to what was proposed for Oryctes rhinoceros nudivirus (OrNV) in 1984. To confirm this mechanism, we exposed cells to the macropinocytosis inhibitor imipramine (IMP) and inoculated the cell culture with HzNV-1, followed by quantitative PCR (qPCR) analysis to detect intracellular genome copies. This revealed a 72% reduction in viral DNA levels, supporting the involvement of macropinocytosis in viral entry. Additionally, we determined that the HzNV-1 DNA replication starts approximately 7 hours post-infection, which aligns with studies examining related virus clades, including baculoviruses and nimaviruses (*Nimaviridae*). These findings offer new insights into the cytopathological and molecular mechanisms of nudivirus infections, particularly regarding cell entry and the timing of viral replication, and lay the groundwork for transcriptomic experiments in which the choice for proper time points is crucial.

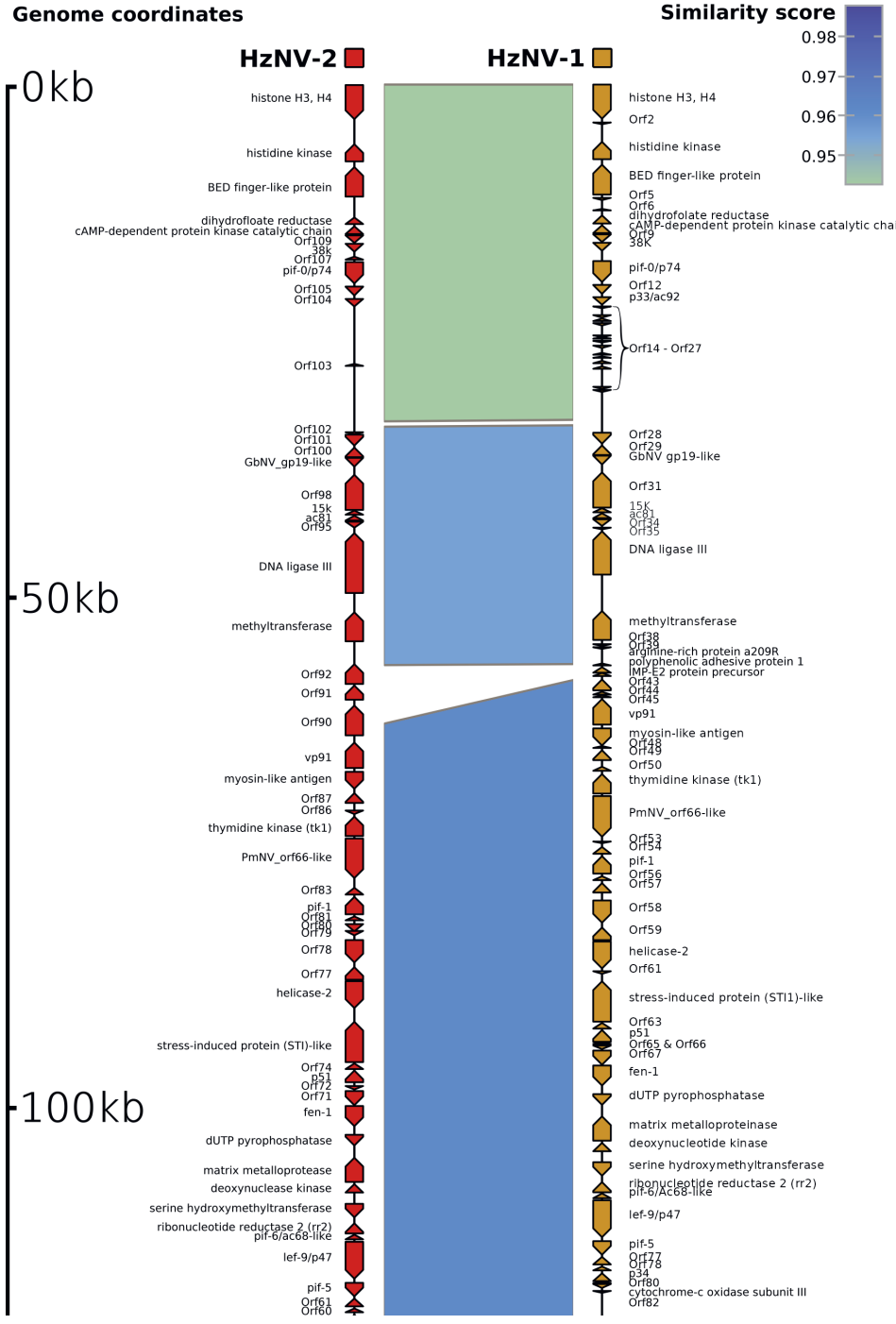
Introduction

The polyphagous larvae of the noctuid corn earworm moth, *Helicoverpa zea* (formerly *Heliothis zea*), feed on a variety of economically important crops [257]. This insect is a major agricultural pest across North and South America [258], causing annual damages exceeding 100 million dollars [259]. In the past, chemical pesticides have been commonly used to against *H. zea* infestations, but the overreliance on these chemicals has resulted in the development of resistant *H. zea* populations [260]. Exploring alternative and more sustainable pest management methods were therefore required. One such alternative involves biological control via natural enemies such as parasitoid wasps, fungi, and entomopathogenic viruses to suppress pest populations [261]. Among these biological control agents, baculoviruses (*Baculoviridae*) have been extensively studied for their pest control potential of many lepidopteran insects. Interestingly, a virus initially classified as a baculovirus was identified in *H. zea* [262] and stood out due to the visible symptoms it induces in adult moths. Studies later revealed that this virus was in fact a member of the *Nudiviridae* family and it is now known as Helicoverpa zea nudivirus 2 (HzNV-2).

Female moths infected with HzNV-2 were shown to suffer from a "waxy plug" at their genital openings, which was later found to be filled with high quantities of rod-shaped virions [64]. Despite being filled with HzNV-2 virions, the waxy plug is not essential for HzNV-2 transmission *per se*, given that infected males do not form a waxy plug and can still transmit the virus to uninfected females [263]. Once transmitted, HzNV-2 infections can either manifest overtly or covertly. Overt infections produce visible symptoms such as the formation of the waxy plug in females, gonadal atrophy in both sexes, and can even lead to sterile progeny [64]. Covertly infected individuals mainly serve as asymptomatic carriers of the virus [145]. The degree of disease is dose-dependent, with higher doses leading to sterility and agonadal phenotypes [263]. Additionally, HzNV-2 can influence the mating behaviour of infected individuals. Males infected with the virus show a reduction in mating efficiency and fail to induce the

cessation of female calling behaviour [264]. Moreover, infected females continue to emit mating calls even after copulation, a behaviour that favours continued mating and enhances transmission of HzNV-2 [265].

It is apparent that HzNV-2's ability to cause pathological effects makes it an attractive candidate for biocontrol, especially since it can persist over multiple generations in *H. zea* populations. This persistence of HzNV-2 is proposed to have led to the emergence of another version of this virus, identified in an ovarian cell line of *H. zea* well before HzNV-2 was described, and this virus, hence, has been given the name Heliothis zea nudivirus 1 (HzNV-1) [127]. Today, HzNV-1 and HzNV-2 are classified under the same species, *Betanudivirus hezeae*, within the genus *Betanudivirus*, a logical classification based on their highly similar genomic characteristics, and consequent phylogenetic relatedness. The genomes of HzNV-1 and HzNV-2 are among the largest of all known nudiviruses, with HzNV-2 being slightly larger than HzNV-1 (~232 kb and ~228 kb, respectively), and these genome sizes correlate with the length of their rod-shaped virions (~400 x 30 nm). Their overall genome similarity is 93.5%, but HzNV-1 lacks 14 open reading frames (ORFs) compared to HzNV-2 [144]. Aside from these differences, the two members of *Betanudivirus hezeae* display an almost identical genome organisation and gene synteny (**Figure 1**), supporting the hypothesis that HzNV-1 originated from HzNV-2, but underwent genomic adaptations after being cultured in cell lines over multiple generations.


Genome coordinates
Similarity score
0.98
0.97
0.96
0.95
HzNV-2
HzNV-1
0kb
50kb
100kb
histone H3, H4
histidine kinase
BED finger-like protein
dihydrofloate reductase
cAMP-dependent protein kinase catalytic chain
Orf109
38k
Orf107
pif-0/p74
Orf105
Orf104
Orf103
Orf102
Orf101
Orf100
GbNV_gp19-like
Orf98
15k
ac81
Orf95
DNA ligase III
methyltransferase
Orf92
Orf91
Orf90
vp91
myosin-like antigen
Orf87
Orf86
thymidine kinase (tk1)
PmNV_orf66-like
Orf83
pif-1
Orf81
Orf80
Orf79
Orf78
Orf77
helicase-2
stress-induced protein (STI)-like
Orf74
p51
Orf72
Orf71
fen-1
dUTP pyrophosphatase
matrix metalloprotease
deoxynuclease kinase
serine hydroxymethyltransferase
ribonucleotide reductase 2 (rr2)
pif-6/ac68-like
lef-9/p47
pif-5
Orf61
Orf60
histone H3, H4
Orf2
histidine kinase
BED finger-like protein
Orf5
Orf6
dihydrofolate reductase
cAMP-dependent protein kinase catalytic chai
Orf9
38K
pif-0/p74
Orf12
p33/ac92
Orf14 - Orf27
Orf28
Orf29
GbNV gp19-like
Orf31
15K
ac81
Orf34
Orf35
DNA ligase III
methyltransferase
Orf38
Orf39
arginine-rich protein a209R
polyphenolic adhesive protein 1
IMP-E2 protein precursor
Orf43
Orf44
Orf45
vp91
myosin-like antigen
Orf48
Orf49
Orf50
thymidine kinase (tk1)
PmNV_orf66-like
Orf53
Orf54
pif-1
Orf56
Orf57
Orf58
Orf59
helicase-2
Orf61
stress-induced protein (STI1)-like
Orf63
p51
Orf65 & Orf66
Orf67
fen-1
dUTP pyrophosphatase
matrix metalloproteinase
deoxynucleotide kinase
serine hydroxymethyltransferase
ribonucleotide reductase 2 (rr2)
pif-6/Ac68-like
lef-9/p47
pif-5
Orf77
Orf78
p34
Orf80
cytochrome-c oxidase subunit III
Orf82

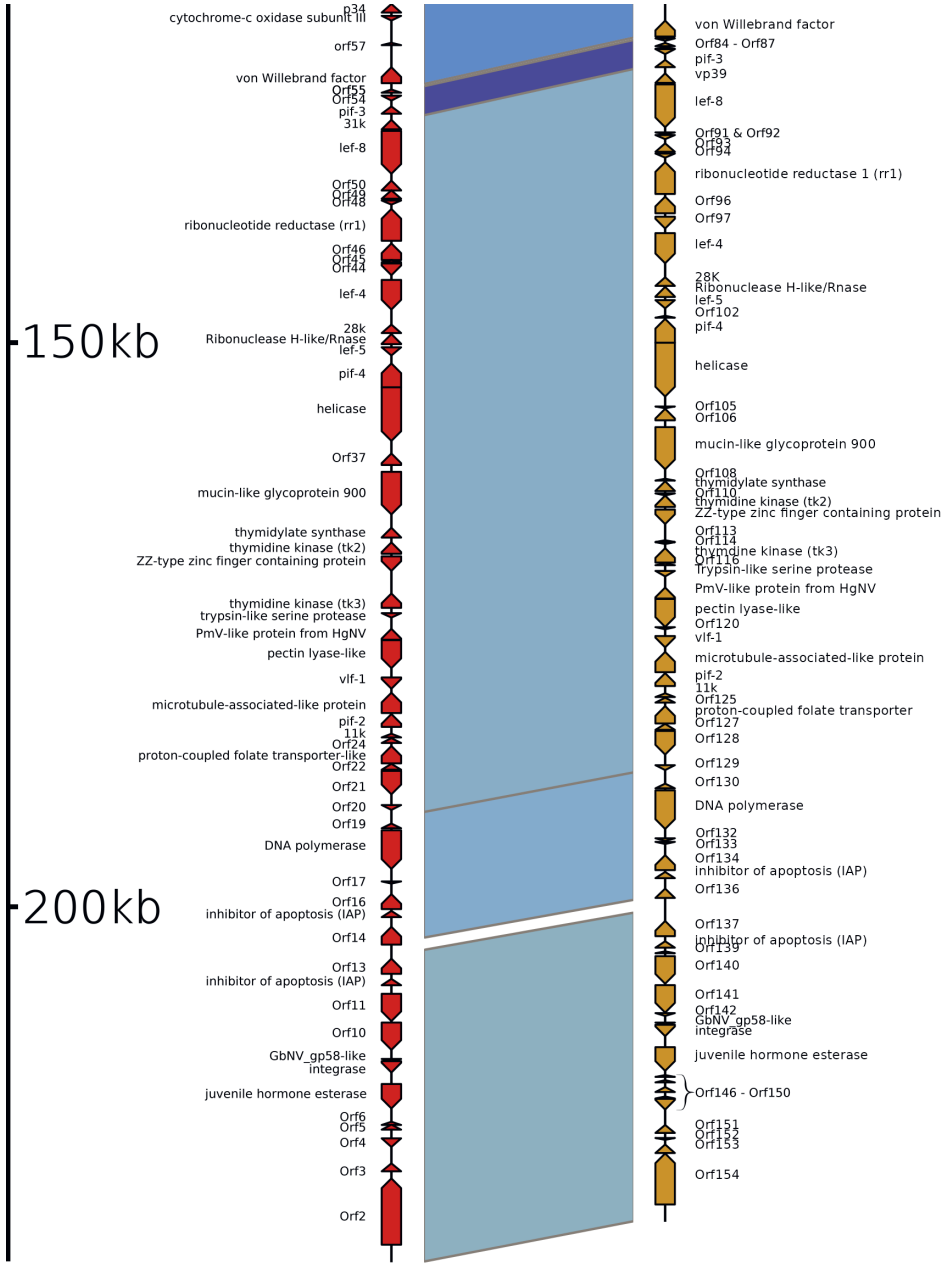


Figure 1. Linear genome alignment of Heliothis zea nudivirus 1 (HzNV-1) and Helicoverpa zea nudivirus 2 (HzNV-2) comparing gene synteny and sequence similarity. Alignment similarity blocks connect corresponding regions between the two genomes, with block colours representing sequence similarity scores based on the "map match/map length" ratio (the closer the score is to 1, the more similar the sequences). Gaps with missing similarity blocks represent regions of HzNV-1 protein-coding genes are displayed in yellow, while those of HzNV-2 are in red. Gene order and annotations are derived from GenBank (HzNV-1: AF451898.1, HzNV-2: JN418988.1). Homologous proteins from HzNV-1 and HzNV-2 were reassessed using BLAST [266] and InterPro [267], and

annotated consistently. For cases where the GenBank annotations were outdated or inconsistent with BLAST and InterPro, the annotations were updated. For example, HzNV-1's orf119 and HzNV-2's orf29 were both re-annotated as pectin lyase, while HzNV-1's orf4 and HzNV-2's orf112 are now identified as BED-finger-like proteins.

In spite of their highly identical genomic traits, these two viruses can be best distinguished by their pathological differences. While HzNV-2 is able to infect and spread among all developmental stages of its insect host, HzNV-1 lost this *in vivo* infectiousness and is now completely restricted to cell lines. However, HzNV-1 can still cause lytic and persistent infections in the HZ-AM1 cell line (initially designated as BCIRL-HZ-AM1), which originated from the ovarian tissue of *H. zea* [34], and it can infect and replicate in cell lines derived from other lepidopteran hosts [268]. Although studies in living insects are not possible with HzNV-1, this virus provides a convenient model for investigating the cytopathological dynamics of nudivirus infections in ovarian cells, which are also the primary target of HzNV-2 when initiating infection after being acquired through mating.

This chapter aims to elucidate the behaviour of HzNV-1 in the HZ-AM1 cell line, and to provide fundamental work for the experiments reported in **Chapter 5**. We assessed the cytopathological behaviour of HzNV-1 in the HZ-AM1 cell line, and visualised the nudiviral virions inside and outside of the cell via different microscopic methods. Furthermore, we gained insight into HzNV-1's cell entry mechanism and pinpointed the time point at which viral DNA replication is initiated. The groundwork of this chapter provides valuable information for downstream experiments by enhancing our understanding of HzNV-1's infection cycle under synchronised infection conditions.

Material and Methods

Comparative genome similarity plot of HzNV-1 and HzNV-2

The genome sequences of HzNV-1 and HzNV-2 in GenBank format (HzNV-1: AF451898.1, HzNV-2: JN418988.1) were converted to FASTA format using EMBOSS Seqret (seqret -sequence *.gb -outseq *.fna). Genome alignment was performed using Winnowmap [269] with the -cx asm10 option, which is optimised for sequence divergences of up to 10% (most fitting, since the overall sequence similarity between HzNV-1 and HzNV-2 equals 93.5%). The Winnowmap command (winnowmap -cx asm10 *.fna > *.paf) generated a PAF (Pairwise Mapping Format) alignment file. The genomic sequences, gene features, and alignment data were visualised using the R package gggenomes [270]. Genomic sequences and corresponding gene annotations were loaded from the GenBank files using the read_seqs() and read_feats() functions, respectively. Alignment links between the genomes were read from the previously generated PAF file using the read_links() function. To ensure consistent genome orientation, genomes were automatically synchronised based on the alignment links using the sync() function. The comparative genome plot was generated using the gggenomes command and saved as a high-resolution image with ggplot2's ggsave function [271].

Insect cell line and virus

The HzNV-1 infection system was established using a glycerol stock of HzNV-1 (P0), generously provided by Professor Yueh-Lung Wu (National Taiwan University, Department of Entomology) in 2020. In the very beginning, both HZ-AM1 and Sf9ET cells [272] were exposed to a 10x dilution of the P0 isolate. The Sf9ET cells received Sf900II medium, while the HZ-AM1 cells were cultured in CCM3 medium. The media of both cell lines were supplemented with 5% FBS and 50 µg/mL gentamicin, and the cells were incubated at 27°C. To subculture HZ-AM1 and Sf9ET cells, the attached cells at the bottom of the culture dish were detached either by gently scraping with a cell scraper or by

flushing with a bent glass Pasteur pipette, and the homogenous cell suspension was transferred to a fresh culture flask with an appropriate amount of medium. The seeded cells were then allowed to attach and grow overnight. The established Sf9ET cell line, derived from the ovarian tissue of the fall armyworm (*Spodoptera frugiperda*), is genetically modified to express enhanced green fluorescent protein (eGFP) upon baculovirus infection [272]. The infection in Sf9ET cells was performed to rule out any baculovirus contamination in the P0 stock. The potential expression of eGFP in the Sf9ET cells was examined using an inverted fluorescence microscope (Axio Observer 21, Zeiss). Electron microscopy (EM) imaging of sectioned Sf9ET cells provided additional confirmation regarding the presence of baculoviruses. Furthermore, a polymerase chain reaction (PCR) with the GoTaq® G2 DNA Polymerase (Promega) was performed at an annealing temperature of 60°C. The primers were specific for the *vp39* gene of Autographa californica multiple nucleopolyhedrovirus (AcMNPV) and would amplify a 108-bp PCR product. The forward primer qPCR-vp39-F1 had the sequence 5'-ACAACCCGATAAGAAGCAGTGACA-3', while the reverse primer qPCR-vp39-R1 had the sequence 5'-TAAGCGTTCTGTCCAGCTCACG-3'. Existing baculovirus contaminants were eliminated by passaging the P0 inoculum multiple times over HZ-AM1 cells. Initially, the supernatant from P0-infected HZ-AM1 cells was collected by centrifugation (15 min, 2.000 rpm) and filtered (syringe filter, 0.45 µm pore size) at two days post-infection (dpi), and used to reinfect a fresh culture of HZ-AM1 cells. The virus stock was passaged two more times in this way over HZ-AM1 cells resulting in the P3 isolate. The presence of baculovirus in the P3 isolate was then again tested via inoculation of Sf9ET. After confirming the absence of baculovirus, the viral titre of the P3 isolate was determined via end-point dilution assay (EPDA) as infectious units (IU) per mL [273] on HZ-AM1 cells. Continued passaging of the P3 isolate through HZ-AM1 subcultures allowed for the harvest of virus isolates (P4, P5, P6) with higher virus titres.

Infection of HZ-AM1 cells with HzNV-1

Prior to infecting fully attached HZ-AM1 cells, the culture medium was removed from the culture vessel. To prevent the cells from drying out, they were immediately covered with different volumes of virus inoculum, according to the surface area of the vessel (e.g., 500 µL for a 6-well plate, 1 mL for a T25 flask, and 3 mL for a T75 flask). Based on the respective cell density, the virus stock was diluted with CCM3 medium to the desired multiplicity of infection (MOI). After covering the HZ-AM1 cells with virus inoculum, the cells were incubated for 45-60 minutes at 27°C. Afterwards, the virus inoculum was removed, and fresh CCM3 medium was added. The cells were then incubated again at 27°C for the desired duration to allow virus replication.

HzNV-1 virion purification through ultracentrifugation with sucrose cushion

Four T25 flasks holding HzAM-1 cells at ~70% confluency were infected with HzNV-1 (MOI = 10), and incubated for 72 hours to ensure optimal virus yields. Afterwards, lysed cells that did not yet fully detach from the culture vessel were rinsed off with a micropipette to obtain homogenous suspensions. The supernatant of the cell suspension was then collected (15 min, 2000 rpm) and filtered (syringe filter, 0.45 µm pore size), resulting in a total volume of approximately 16 mL used for virus isolation through a sucrose cushion. A 60% (w/w) sucrose solution was prepared by dissolving 60 g of sucrose in 10 mM Tris-HCl (pH 8.0) to a final weight of 100 g. Beckman SW32.1 Ti ultracentrifuge tubes were pre-soaked overnight in Milli-Q water, and 5 mL of the 60% sucrose solution was carefully layered at the bottom of the tubes. Next, the filtered supernatant was carefully layered on top of the sucrose cushion, filling the tube to within 2 mm of the rim. The filled tubes were centrifuged in a Beckman L7-65 ultracentrifuge at 20,000 rpm for 1.5 hours at 4°C. After centrifugation, the cloudy fraction, containing the virus particles, was carefully extracted by piercing the polycarbonate centrifuge tube sideways with a needle and collecting the fraction into a syringe. The extracted fraction was then transferred to a new tube and

supplemented with Tris-HCl, followed by several inversions of the tube to gently homogenise the suspension. A second centrifugation step was then carried out at 25,000 rpm for 2 hours at 4°C to clean the sample from sucrose by pelleting the virus. Following this step, the supernatant was carefully decanted, and the tube was left upside-down to allow any remaining liquid to drain completely, lowering the risk of disturbing the pellet. Finally, the pellet was resuspended in 50-100 µL Tris-HCl and stored for further analysis.

Preparation of suspension and tissue samples for electron microscopy

For the visualisation of HzNV-1 virions in suspension, negative staining transmission electron microscopy (TEM) was employed. Purified virus particles, obtained after ultracentrifugation through a 60% sucrose cushion, were prepared for TEM by adding 5 µL of the well-mixed virus suspension onto a hydrophilic carbon-coated copper grid. The specimen was incubated at room temperature (RT) for 2 minutes, followed by a single wash with 5 µL of Milli-Q water. The specimen was negatively stained by applying 5 µL of 2% uranyl acetate (low pH) and then incubated for 30 seconds. The staining solution was carefully removed with filter paper, and the grid was allowed to air dry for 1 minute. The specimen could then be visualised under a JEM-1400Plus Electron Microscope (JEOL) for examination.

To visualise virus-infected cells, HZ-AM1 or SF9ET cells were exposed to the virus inoculum and incubated for as long as the specific experiment required. For instance, to study HzNV-1's cell entry mechanism, HZ-AM1 cells were infected with an MOI of 300, incubated with the virus inoculum for 30 minutes, and then immediately harvested for TEM preparation. For observing advanced virus infection, cells were infected with an MOI of 5 and incubated for at least 24 hours. After incubation, cells were detached and harvested by pelleting the cells via centrifugation (2000 rpm, 5 min) and resuspended in 1 mL of 2.5% glutaraldehyde in 0.1 M phosphate/citrate buffer. The sample was incubated for at least 1 hour at RT. After two washes in the same buffer, the cell pellet was embedded in warmed 3% gelatine and incubated at 4°C until solidified. Post-fixation was performed with 1% osmium tetroxide in 0.1M phosphate/citrate buffer for 1 hour at RT, followed by at least three washing steps with Milli-Q water. The sample was then dehydrated through a series of incubation steps with increasingly concentrated ethanol (30%-100%) and finally infiltrated with Spurr resin for embedding. The resin was polymerised in an oven at 70°C for a minimum of 8 hours. Thin sections of the resin-embedded cells were prepared with a diamond knife and visualised using a JEM-1400Plus Electron Microscope (JEOL) .

Inhibitory assay to study HzNV-1 cell entry mechanism

During the optimisation phase of this assay, HZ-AM1 cells were treated with the macropinocytosis-specific inhibitor imipramine (IMP) at varying concentrations (25–500 µM in water) and incubation durations (2.5–7.5 hours) to assess potential cytotoxicity. Cytotoxicity was evaluated both visually under an optical microscope (Labovert, Leitz, Elst NL) and through RNA quality analysis. For RNA extraction, the cells were washed twice with 1x PBS, after which Trizol reagent (Life Technologies) was added (500 µL per well of 6-well plates or 1 mL for T25 flasks). The cultures were thoroughly homogenised by scraping with bent Pasteur pipettes and homogenizing by pipetting. The samples were stored at 4°C overnight. The following day, RNA was extracted and treated with DNase using the Direct-zol RNA Miniprep Kit (Zymo Research), following the manufacturer's protocol. Finally, the extracted RNA was loaded on an agarose gel to check for degradation.

Finally, 200 µM of IMP was selected as an optimal concentration, with cells tolerating incubation for up to 7.5 hours without visible cytotoxicity. After incubation in CCM3 medium supplemented with 200 µM IMP for 7.5 hours, the cells were exposed to the HzNV-1 inoculum (MOI = 5), also supplemented with 200 µM IMP, for 45–60 minutes. Once the IMP-containing virus inoculum was removed, fresh

CCM3 medium with 200 µM IMP was added, and the cells were incubated for 24 hours at 27°C. Following this incubation, total DNA was extracted from triplicate samples of HZ-AM1 cells, both with and without IMP treatment (control), to assess the inhibitory effect of IMP on HzNV-1 cell entry.

Whole DNA extraction from HzNV-1-infected HZ-AM1 cells

After removing the medium from HZ-AM1 cells (approximately $1.5x10^6$ cells per well in a 6-well plate) 24 hours post-exposure to HzNV-1, 1 mL of 1x PBS was added in which the cells were detached with a bent Pasteur pipette. The cell suspension was then transferred to a 1.5 ml tube and the cells pelleted by centrifugation (2000 rpm, 5 min). The supernatant was discarded and the cell pellet resuspended in 200 µL ATL buffer of the DNeasy Blood & Tissue kit (Qiagen), followed by adding 20 µL of proteinase K and a short spin on a vortex mixer. The mixture was then incubated for 20 min at 56°C, before a premix comprising of 200 µL AL lysis buffer (from the DNeasy Blood & Tissue kit) and 200 µL ethanol was added, followed by another thorough mix. The mixture was incubated at RT for 10 min and then transferred to a spin column from the DNeasy Blood & Tissue kit. The last steps of the DNA extraction followed the manufacturers' protocol.

Relative viral DNA determination using quantitative PCR

Prior to conducting quantitative PCR (qPCR) analysis, two primer pairs were designed using SnapGene software (www.snapgene.com), with one set targeting the host and the other targeting the virus. Host-specific primers were designed to amplify a 199-bp fragment from the gene encoding Glyceraldehyde 3-phosphate dehydrogenase (GAPDH). The sequences for the host-specific primers were: qPCR_GAPDH_fw: 5'-GGTGGTGCTAAGAAGGTCATCATC-3' and qPCR_GAPDH_rev: 5'-CGTGTACAGTGGTCATCAGACC-3'. For the virus, primers were designed to bind in the coding region of the *vlf-1* gene, producing a 155-bp amplicon. The sequences for the virus-specific primers were: qPCR_vlf-1_fw: 5'-GTTCAAGGTGGAGGTCGACTC-3' and qPCR_vlf-1_rev: 5'-GCAAGCGATCCTGATTCTGTTC-3'. Both primer sets were designed with an annealing temperature of 59°C. Amplification efficiency of each primer pair was determined using a serial dilution of DNA extracted from HZNV-1-infected HZ-AM1 cells. The Cq values were measured in technical triplicates using the Bio-Rad CFX96 Touch Real-Time PCR system. The slope of the standard curve derived from these Cq values was used to calculate amplification efficiency with the formula: E = (10^(-1/slope) - 1) * 100. Host-specific primers targeting GAPDH showed an amplification efficiency of approximately 110%, while virus-specific primers achieved an efficiency of ~107%. Once primer efficiencies were established, two independent experiments were performed to quantify relative viral DNA levels in infected cells under various experimental conditions. The first experiment investigated the cell entry mechanism of HzNV-1, as outlined in a previous section (2.5. Inhibitory assay to impair HzNV-1 cell entry). The second experiment was a time-course study to monitor the onset of HzNV-1 DNA replication. In this time-course experiment, HZ-AM1 cells were infected with HzNV-1 at a multiplicity of infection (MOI) of 5, with triplicate samples collected at hourly intervals from 1 hour post-infection (hpi) to 10 hpi. Total DNA was extracted at each time point as described above (2.7). To determine the timing of viral DNA replication initiation, qPCR measurements from 2 hpi to 10 hpi were compared to the baseline value measured at 1 hpi. For both experiments, relative viral DNA levels were calculated using the Pfaffl method (Pfaffl, 2007), which is an enhanced version of the ΔΔCt method that accounts for varying primer efficiencies. The formula used was: Ratio = (E_target / E_reference) ^ (ΔCq_reference - ΔCq_target). For all qPCR measurements, a total of 60 ng DNA was used as the template for each individual well.

Results

Decontamination and enrichment of the HzNV-1 isolate through selective passaging

Initial infection experiments with the HzNV-1 P0 isolate resulted in the expression of eGFP in the exposed Sf9ET cells (**Figure 2A**), either implying a compatibility of the HzNV-1 transcriptional machinery with the baculoviral polyhedrin promoter incorporated in the cellular genome and driving eGFP expression, or a baculovirus contamination. Examining ultrathin TEM sections of the infected Sf9ET cells and a PCR with baculovirus-specific primers confirmed contamination with AcMNPV. The electron micrographs revealed the typical fibrillar structures formed by the baculoviral P10 protein (**Figure 2B**) in the cells treated with the P0 stock (but not the mock-infected cells). These cells also contained enveloped virions containing multiple nucleocapsids (ODVs) in the nucleus of the cell (**Figure 2C**), which are uncharacteristic of nudivirus infections, and typical for AcMNPV. Baculovirus occlusion bodies were not observed, hinting at a contamination with an AcMNPV-derived baculovirus expression vector in the received viral stock.

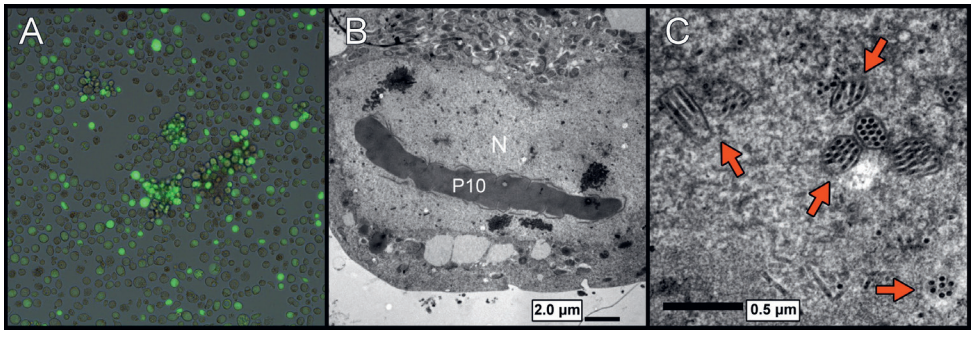


Figure 2. Sf9ET cells after exposure to the HzNV-1 P0 isolate (10x diluted), displaying contamination with a baculovirus. (A) Fluorescence imaging (magnification 40x) with an inverted microscope (Axio Observer 21, Zeiss) showing eGFP signal in Sf9ET cells exposed to the P0 isolate at 5 dpi, indicating baculovirus infection. (B) Ultrathin section of a P0-infected Sf9ET cell at 60 hpi visualised via EM using standard methods. Inside the nucleus (N) of the infected cell, a large fibrillar structure could be observed. This fibrillar structure is a common consequence of baculovirus infections, manifesting through the accumulation of the highly abundant P10 protein (Van Oers *et al*., 1994; reviewed by Carpentier and King, 2009), which nudiviruses lack. The bar corresponds to 2 µm. (C) Zoomed-in region inside the infected nucleus, showing virions strongly resembling occlusion derived virions (ODVs) with multiple nucleocapsids per virion, typical for certain baculoviruses (arrows). The bar indicates 500 nm. EM images were taken with a JEM-1400Plus Electron Microscope (JEOL) by Jan W. M. van Lent, Wageningen Electron Microscopy Centre (WEMC).

Knowing that the P0 isolate was contaminated with AcMNPV, a selective passaging approach was employed using multiple infection rounds of HZ-AM1 subcultures to eliminate this baculovirus. Since AcMNPV cannot propagate in HZ-AM1 cells (Mcintosh *et al*., 2005, Zhihong Hu & Manli Wang, pers. commun.), while HzNV-1 can, this approach was designed to selectively produce HzNV-1 while reducing AcMNPV titres with each passage. After passaging the P0 isolate three times through HZ-AM1 cells, the P3 isolate was obtained with a HzNV-1 titre of 3.56×10^7 IU/mL The infection of Sf9ET cells with the P3 isolate at the highest concentration resulted in no discernible cells expressing eGFP through baculovirus infection. Additionally, PCR with the AcMNPV-specific primer was repeated on the P3 isolate, which also resulted in no vp39 PCR amplicon, contrary to the P0 isolate (**Figure 3**).

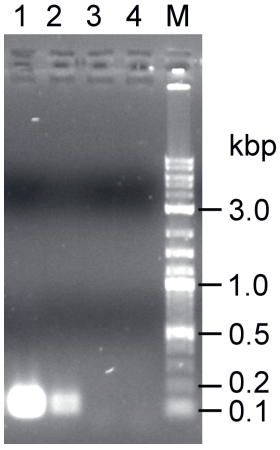


Figure 3. Agarose gel electrophoresis showing the results of a PCR with baculovirus-specific primers, confirming Autographa californica multiple nucleopolyhedrovirus (AcMNPV) contamination in the P0 isolate, and the absence of AcMNPV in the P3 isolate. The primers were specific for the *vp39* gene of AcMNPV and would amplify a 108-bp PCR product. The GoTaq® G2 DNA Polymerase (Promega) was used for the PCR with an annealing temperature of 60°C. **The differently numbered gel lanes represent PCR reactions with different templates. 1: Positive control, cDNA from caterpillars infected with AcMNPV (kindly provided by Simone Gasque); 2: DNA extracted from the P0 isolate; 3: DNA extracted from the P3 isolate; 4: Negative control with no DNA template; M: 5 µL of 1 kb Plus DNA ladder (NEB).**

Next to the cleansing of the virus isolate from AcMNPV, the selective passaging led to an increase of the HzNV-1 virus titre, reaching a maximum of $1.12x10^9$ IU/mL in the P6 isolate after six passages through the HZ-AM1 cells. In the end, four HzNV-1 stocks with different virus titres were acquired (**Table 1**) and stored at 4°C for follow-up experiments.

Table 1. Virus titres of different HzNV-1 isolates after passaging through HZ-AM1 cells. The virus isolates were harvested from infected HZ-AM1 cells at 48 hours post infection (hpi) and their titres (infectious units per mL) quantified via end-point dilution assay (EPDA).

HzNV-1 isolates	Passages from P0 to obtain isolate	Virus titre (IU/mL)
P3	3 passages	$3.56x10^7$
P4	4 passages	$3.56x10^7$
P5	5 passages	$1.12x10^8$
P6	6 passages	$1.12x10^9$

To investigate whether long-term storage affected HzNV-1 infectivity, the P4 isolate was used to infect HZ-AM1 cells, following refrigeration (4°C) of the virus isolate for over 12 months. Upon exposure to the P4 isolate, the HZ-AM1 cells were checked via light microscopy for cytopathogenic effects (**Figure 4**). No noticeable differences were observed between the mock-infected and P4-infected cells during the first 24 hours. However, at 48 hpi, cell lysis led to a reduction in cell density and the accumulation of cell debris in the culture. Although we did not reassess the virus titres of P4, this qualitative result confirmed that the P4 isolate was still infectious even after 12 months of storage at 4°C.

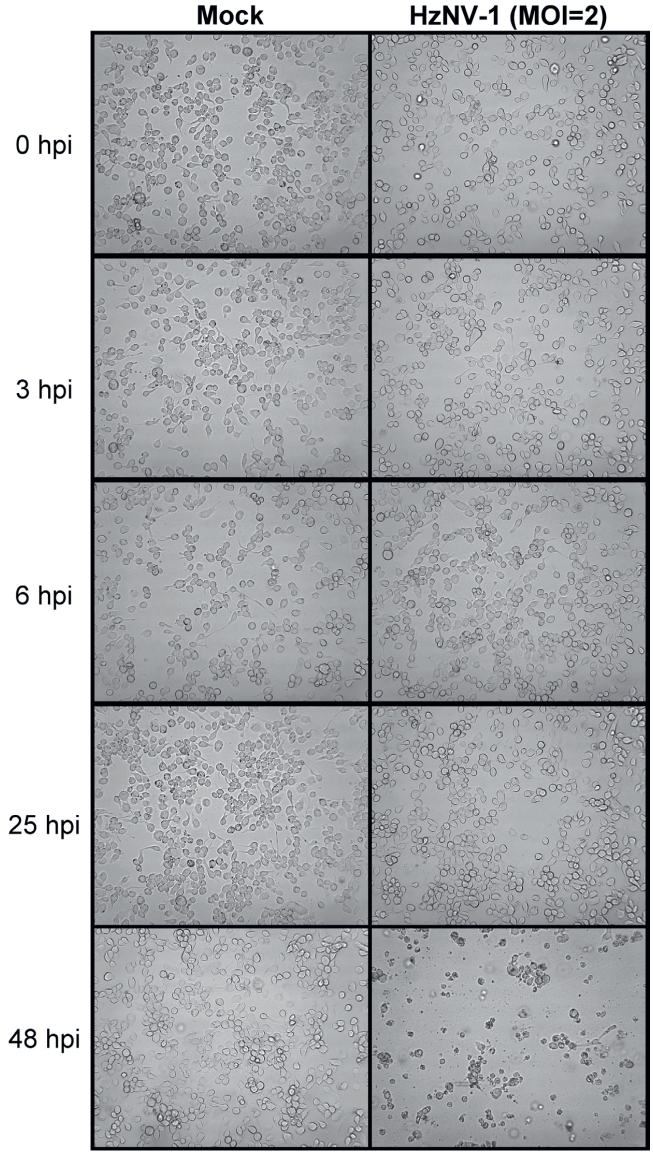


Figure 4. Light microscopy images illustrating HzNV-1 infection in HZ-AM1 cells over time. The left column shows mock-infected cells, while the right column represents cells infected with HzNV-1 at a multiplicity of infection (MOI) of 2. Brightfield images of the cells were taken at 0, 3, 6, 25 and 48 hpi at 40X magnification using a Labovert (Leitz) microscope.

Although light microscopy did not reveal obvious effects of HzNV-1 infection on the HZ-AM1 cells in the first 24 hours of infection, counting infected cells (MOI = 10) with a Neubauer chamber in a separate experiment provided more insight (**Figure 5**).

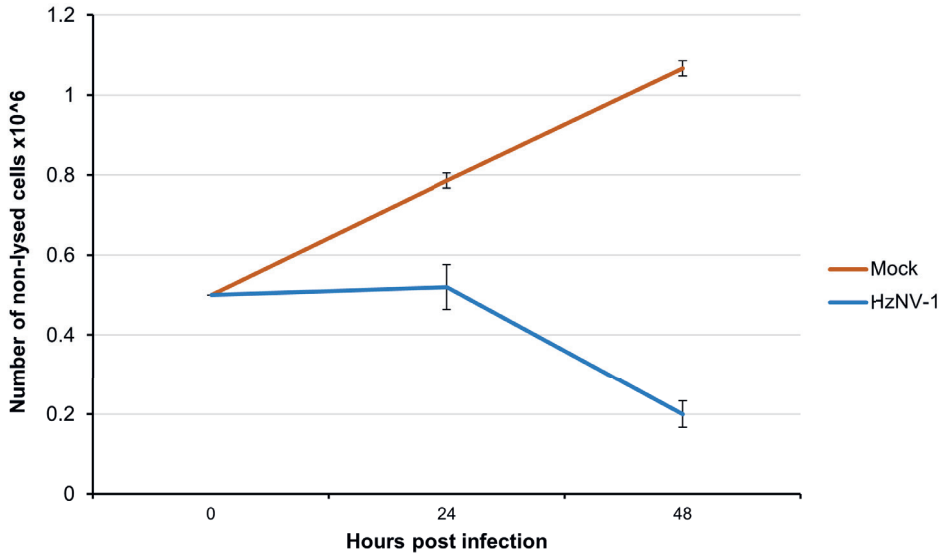


Figure 5. Linear plot comparing the differences in HZ-AM1 cell growth between mock-infected and HzNV-1 infected (MOI = 10) infected cells. The non-lysed cells were counted in triplicates with a counting chamber at three different time points: 0, 24 and 48 hpi. The error bars indicate the standard deviation among the triplicates.

While the cells in the mock-infected cultures continued to grow linearly, the HzNV-1-infected cells almost completely halted their multiplication within the first 24 hours of infection. It was only after this period that the virus-infected cells began to lyse and die as a result of the lytic infection caused by HzNV-1. This experiment showed that the HzNV-1 infection has an almost immediate pathogenic effect on host cell viability, though observable cell death occurs only after one day post-infection.

Purification and visualisation of extracellular HzNV-1 virions

The preliminary infection experiments revealed that it takes HzNV-1 at least 24 hours to initiate lysis in its host cell. It was therefore deduced that also the release of viral progeny into the extracellular space would peak sometime after that. Hence, the medium from highly lysed HZ-AM1 cell cultures was collected at 72 hpi in order to purify released HzNV-1 virions via ultracentrifugation over a sucrose cushion (**Figure 6**). After extracting the cloudy fraction from the tube (**Figure 6A**) and subsequent cleaning steps, the enriched HzNV-1 virions in suspension could be used as specimen for TEM analysis.

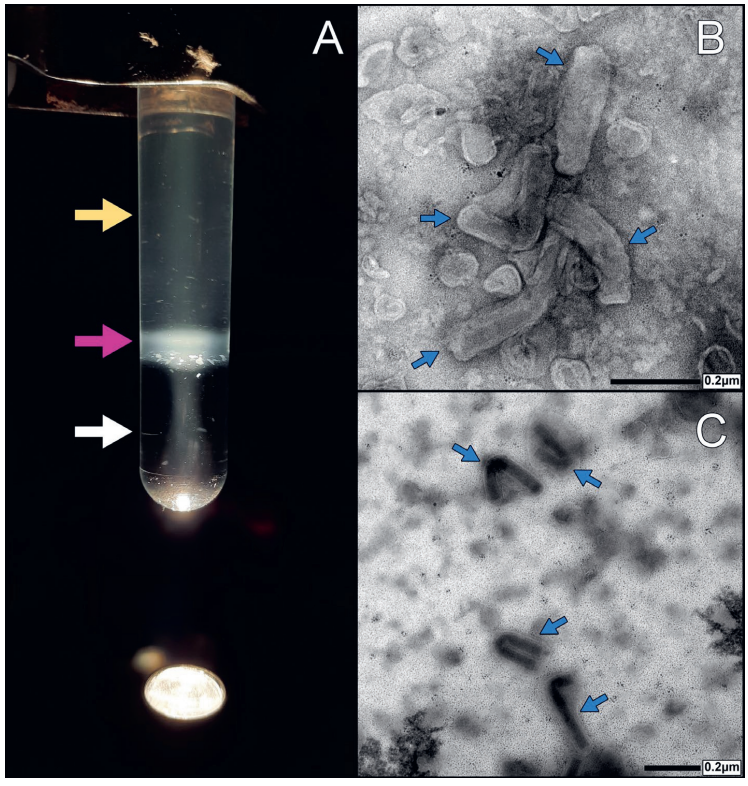


Figure 6. Electron microscopy-assisted visualisation of extracellular HzNV-1 virions, purified via sucrose cushion ultracentrifugation. **(A)** Post-ultracentrifugation sample showing the dense sucrose cushion layer at the bottom of the tube (white arrow) and the less dense medium layer at the top (yellow arrow). The cloudy, intermediately dense virion fraction (purple arrow) is visible between these layers and was collected by piercing the side of the tube with a syringe. The collected sample was washed with Tris-HCl and again ultracentrifuged, before being loaded on a copper grid and negatively stained with uranyl acetate for electron microscopy (EM). **(B)** Virions of HzNV-1 (blue arrows) outlined by the negative stain, showing their rod-shaped morphology with less pronounced, but still visible, nucleocapsids. Bar, 200 nm. **(C)** Fully stain-penetrated HzNV-1 virions with morphologically distinct nucleocapsids, surrounded by their viral envelopes. Bar, 200 nm. EM images were taken with the JEM-1400Plus Electron Microscope (JEOL) provided by Wageningen Electron Microscopy Centre (WEMC).

The negatively stained virus suspension provided two distinct views of the rod-shaped virions. In the first, the virions were only partially penetrated by the uranyl acetate stain (**Figure 6B**), resulting in outlined virions with less defined nucleocapsids, though their approximate length of ~400 nm could still be discerned. In contrast, when the HzNV-1 virions were fully stained by uranyl acetate, the protein-assembled nucleocapsids appeared clearly darkened (**Figure 6C**). Both views highlighted the flexibility of the rod-shaped virions, which could bend into forms resembling candy canes or horseshoes. Having confirmed the presence of HzNV-1 virions in the extracellular environment and examined their structural characteristics, the next step was to investigate the mechanism used by the nudivirus to enter the HZ-AM1 cell line of ovarian origin.

Macropinocytosis drives HzNV-1 cell entry into ovarian cells of *H. zea*

Cultured HZ-AM1 cells were infected with HzNV-1 and incubated for 30 minutes, providing enough time for the virus to attach to the plasma membrane and initiate cell entry. Given that each ultrathin section represents a two-dimensional cut through a three-dimensional cell, the probability of capturing a virion at the right location and moment of internalisation and at a proper angle for visualisation was low, even with multiple sections. Therefore, to increase the chances of visualizing a virion in the process of entering the cell, an enriched MOI of ~300 was used for infection. The approach using an enriched MOI successfully yielded ultrathin sections that captured notable structures in proximity to HzNV-1 virions at the cellular membrane (**Figure 7**).

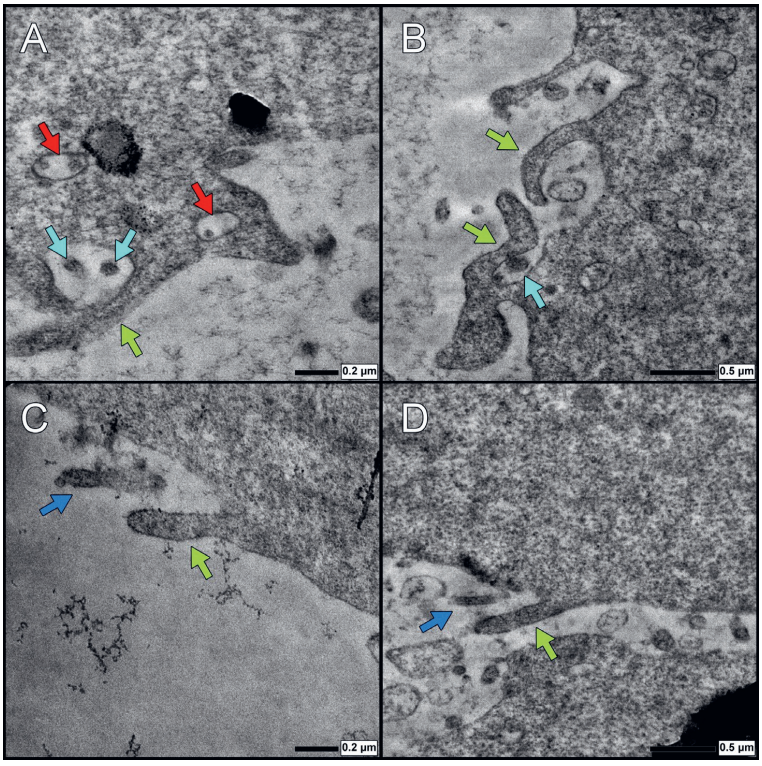


Figure 7. Electron microscopy images of ultrathin sections made from HZ-AM1 cells infected with HzNV-1 (MOI=300) and collected 30 minutes post exposure. At the plasma membrane of the HZ-AM1 cells, pseudopod-like structures (green arrows) were observed in proximity to HzNV-1 virions. **(A, B)** Cross-sectioned HzNV-1 virions (cyan arrows) engulfed by arm-like projections of an HZ-AM1 cell, resulting in virion-containing vesicles (red arrows) near the cell membrane. **(C, D)** Structures resembling longitudinally sectioned HzNV-1 virions (blue arrows) in between the plasma membrane and a manifesting pseudopod-like structure. EM images were taken with the JEM-1400Plus Electron Microscope (JEOL) during the practical course "Basic Transmission Electron Microscopy (TEM)" with Marcel Giesbers and Jelmer Vroom, Wageningen Electron Microscopy Centre (WEMC).

Cell projections resembling pseudopod-like structures appeared to engulf HzNV-1 virions (**Figure 7A**), which might eventually lead to their internalisation into the cell inside vesicles. Based on these observations, the assumption was made that HzNV-1 infects the ovarian cells of *H. zea* via macropinocytosis. This hypothesis was further supported by one of the pictures showing a bleb-like

protrusion of the plasma membrane (**Figure 7B**), which is another characteristic trait of macropinocytosis (Mercer and Helenius, 2009). Different angles of this internalisation process have been captured, also showing longitudinally sectioned virions between the cell membrane and forming pseudopod-like arms (**Figure 7C and 7D**).

To further investigate whether HzNV-1 utilises macropinocytosis to enter ovarian host cells, an additional experiment was conducted. This involved optimizing and implementing an inhibitory assay using imipramine (IMP), a chemical known to disrupt macropinocytosis. The preliminary optimisation focused on determining the tolerance of HZ-AM1 cells to varying concentrations of IMP, ensuring that cytotoxic effects were minimised. HZ-AM1 cells were exposed to IMP at concentrations ranging from 25 µM to 500 µM, with an incubation period of up to 7.5 hours. To evaluate cytotoxicity, cell viability was visually assessed under a light microscope (**Figure 8**). At the two highest concentrations, no significant detrimental effects or signs of severe apoptosis were observed. However, cells exposed to 500 µM of IMP for 7.5 hours displayed the accumulation of extracellular aggregates of unknown nature.

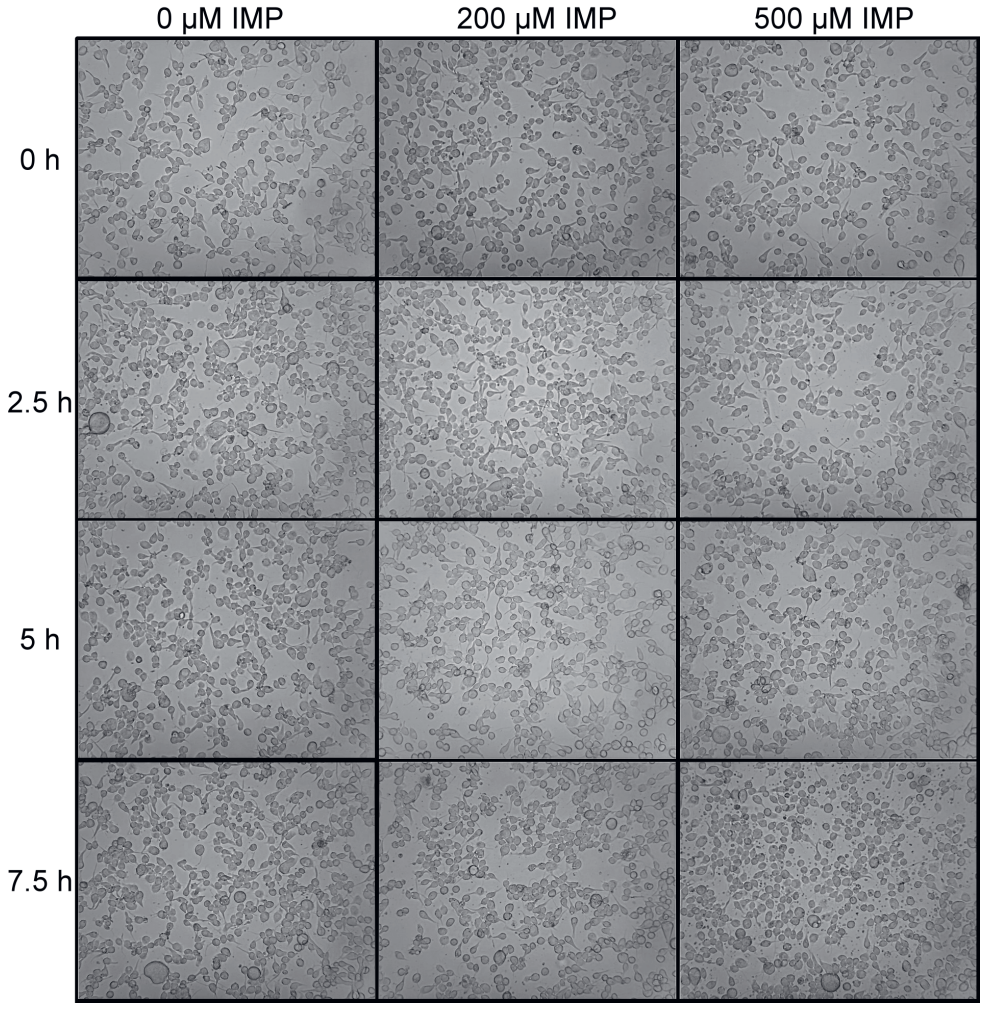


Figure 8. Effect of varying imipramine concentrations on HZ-AM1 cells observed at different time points using light microscopy. The cells were exposed to three concentrations of imipramine-containing medium (0 µM, 200 µM, 500 µM) and the cytopathological effects visually documented every 2.5 hours.

It is unclear what the small aggregates in the culture medium denoted, nonetheless, the cells themselves did not seem to suffer from any cytotoxic effects. Therefore, another experiment was conducted to test HZ-AM1 cell tolerance towards IMP on a molecular biological level. This time, HZ-AM1 cells were exposed to five concentrations of IMP (25 µM, 50 µM, 100 µM, 200 µM, 500 µM) and incubated for the maximum duration of 7.5 hours. After incubation, total RNA was extracted from the treated cells to examine whether RNA quality suffered from the IMP exposure (**Figure 9**).

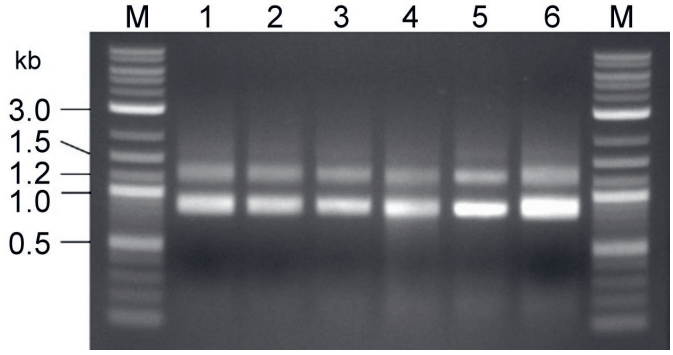


Figure 9. Image of a gel electrophoresis displaying the quality of RNA extracted from imipramine-treated HZ-AM1 cells. Per well, 300 ng of RNA were loaded on a 1% agarose gel and visualised using ethidium bromide. The marker lane (M) contained 5 µL of 1 kb Plus DNA ladder (NEB). The lane numbers correspond to different imipramine concentrations to which the HZ-AM1 cells were exposed for 7.5 hours. **1:** 0 µM; **2:** 25 µM, **3:** 50 µM, **4:** 100 µ.; **5:** 200 µM, **6:** 500 µM.

Inspection of the extracted RNA on agarose gel showed no signs of RNA degradation or impaired integrity, and the two highly abundant ribosomal subunits (top: 28S; bottom: 18S) were well pronounced with little to no signs of breakdown. Given that IMP-induced cell death would have caused severe RNA degradation [277], the results of the RNA extraction further supported that the established concentrations and incubation periods with IMP had no severe impact on cell viability.

Based on these findings, HZ-AM1 cells were infected with HzNV-1 under a constant exposure to 200 µM IMP. After 24 hours, total DNA was extracted from mock- and IMP-treated HZ-AM1 cells infected with HzNV-1. The extracted DNA was then submitted to qPCR analysis to examine differing viral DNA levels in the IMP-treated compared to the control without IMP treatment. The qPCR analysis revealed that the IMP-treated samples contained 0.28 times the amount of HzNV-1 DNA compared to the control group with non-overlapping error bars (**Figure 10**). The statistical significance of this reduction was confirmed using a one-tailed Welch's t-test, which indicated a significant decrease in viral DNA levels in the IMP-treated group compared to the control ($p = 0.0067$).

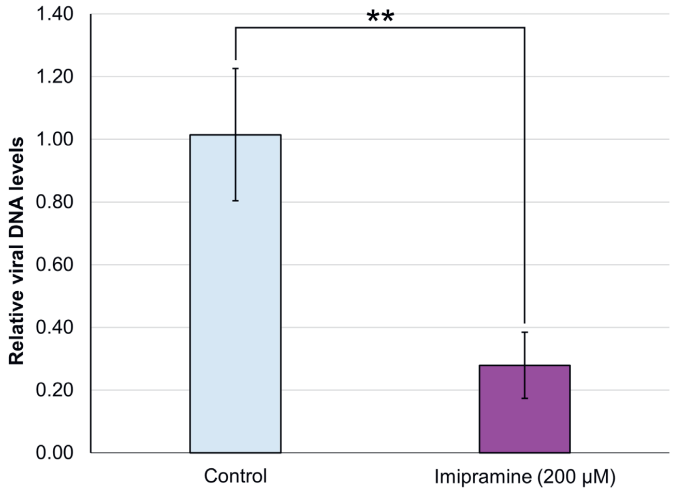


Figure 10. Bar graph displaying the relative DNA levels of HzNV-1 in HZ-AM1 cells exposed to the macropinocytosis inhibitor imipramine. Quantitative PCR (qPCR) data were normalised using primers that bind to the coding region of GAPDH as a reference, while the *vlf-1* gene was the targeted viral region. The relative viral DNA levels were calculated using the Pfaffl method [278], comparing imipramine-treated samples to control samples without imipramine in triplicates. The y-axis shows relative viral DNA levels, with error bars representing the standard deviation. Statistical significance was determined using a one-tailed Welch's t-test (** = $p < 0.01$)

Viral DNA Replication starts 7 hours after HzNV-1 infection

With an important mechanism of HzNV-1 entry into ovarian cells elucidated, the next focus was to determine when viral DNA replication begins following entry of the virus into the host cell. Whole DNA was extracted from triplicate samples of HzNV-1-infected HZ-AM1 cells at ten different time points (1 hpi to 10 hpi), and submitted to qPCR analysis. The qPCR values measured at 1 hpi served as the reference for comparison with the ensuing time points to identify when a notable increase in relative viral DNA levels occurred during infection (**Figure 11**).

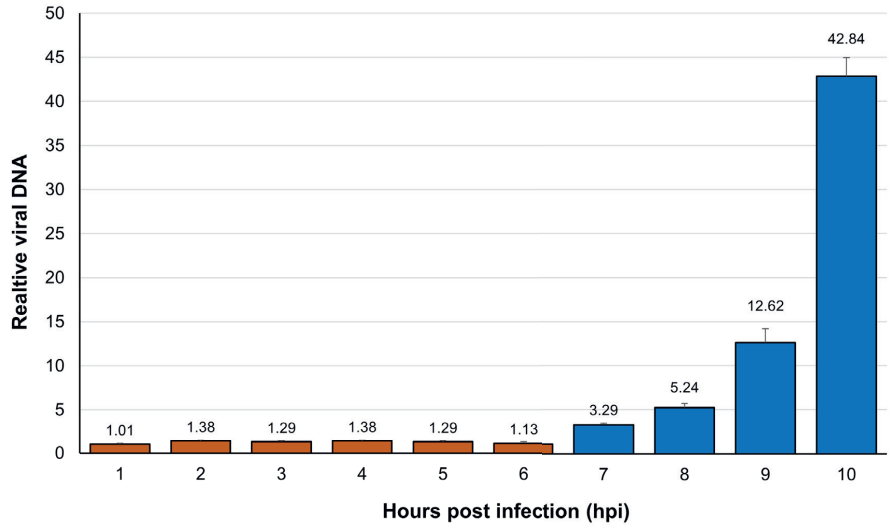


Figure 11. Bar graph illustrating the fold changes in relative viral DNA levels measured using quantitative PCR (qPCR) across different time points of HZ-AM1 cells infected with HzNV-1 (MOI = 5). The housekeeping gene of the host, GAPDH, was used as an internal reference to normalise the data, while the targeted viral gene region was *vlf-1*. The Pfaffl method [278] was employed to calculate the fold changes based on the difference to the reference time point 1 hpi (earliest infection time point). The y-axis represents the fold changes, with error bars indicating the standard deviation.

From 2 hpi to 6 hpi, no significant increases in viral DNA levels were observed. A discernible elevation in viral DNA was detected at 7 hpi, with levels increasing 3.29-fold compared to 1 hpi. From that point onward, the relative viral DNA levels grew exponentially and reached a level of 42.84 times the amount of HzNV-1 DNA at 10 hpi compared to 1 hpi. These results suggest that HzNV-1 requires a minimum of 6 hours to initiate DNA replication, at least under infection conditions with an MOI of 5 at 27°C.

Intracellular virus accumulation

In addition to visualizing extracellular HzNV-1 virions in suspension and during cell entry, another TEM-assisted experiment was conducted to investigate the pathological effects within HZ-AM1 cells at an advanced stage of HzNV-1 infection. For this purpose, HZ-AM1 cells were infected with HzNV-1 and harvested for sample preparation after 20, 40 and 60 hours. However, the cell pellets from the 20 and 40 hpi samples were lost during preparation, leaving only the 60 hpi sample for processing. This sample yielded ultrathin sections that were analysed under the TEM to visualise the interior of HzNV-1-infected cells (**Figure 12**). Different clusters of HzNV-1 virion bundles could be observed in the nucleus of an HZ-AM1 cell, either in longitudinal (**Figure 12B**) or cross section (**12C**). The dark-stained nucleocapsids are easily distinguishable from the nucleoplasm, and a faint outline around each nucleocapsid likely represents the viral envelope.

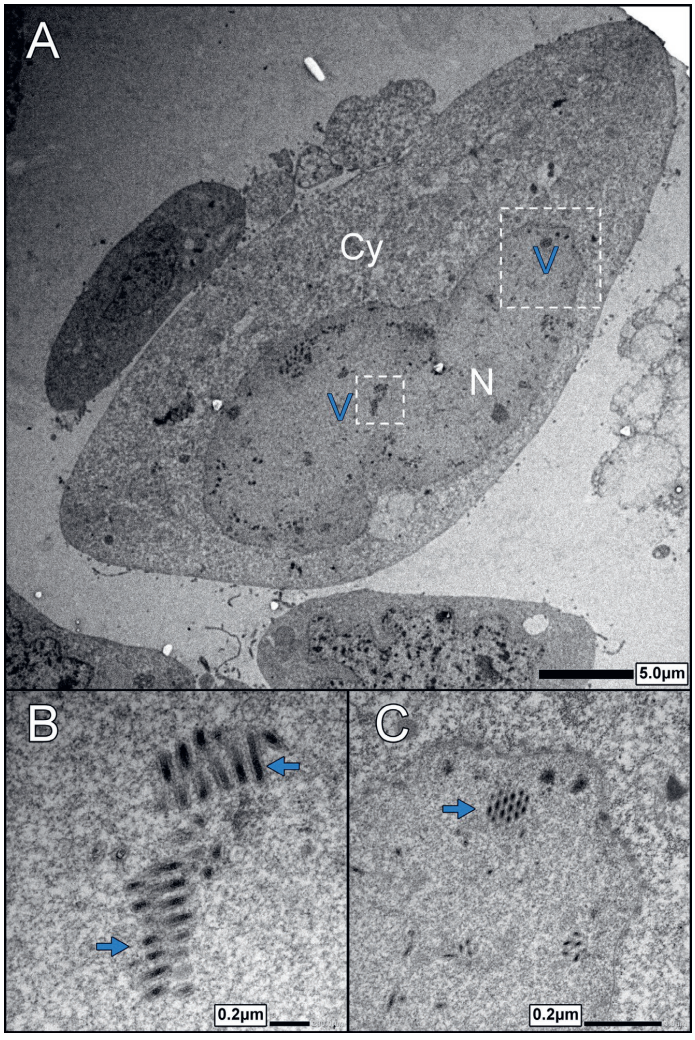


Figure 12. Electron microscopy (EM) image of sectioned HZ-AM1 cell infected with HzNV-1 after 60 hours of infection. **(A)** Whole view of an infected HZ-AM1 cell with indicated cytoplasm (Cy), nucleus (N) and regions with accumulated HzNV-1 virions (V). Bar, 5 µm. **(B)** Magnified region inside the smaller dashed box. Longitudinal section of intranuclear virion bundles. Bar, 200 nm. **(C)** Magnified region inside the larger dashed box. Cross-sectioned bundle of intranuclear virions (blue arrows) with a few individual HzNV-1 virions below. Bar, 1 µm. EM images were taken with the JEM-1400Plus Electron Microscope (JEOL) by Jan W. M. van Lent, Wageningen Electron Microscopy Centre (WEMC).

Both the cytoplasm and the nucleus of the cell appeared to be intact even after 60 hours of infection. Given the severe cell lysis observed at 48 hpi (**Figure 4**), it is possible that this cell was initially not-infected or was infected with an insufficient number of virions during the HzNV-1 exposure, but was later secondarily infected by viral progeny from the first replication cycle.

Discussion

The goal of this chapter was to better understand the behaviour of HzNV-1 in a controlled cell culture system. During the initial experimental setup, contamination with the baculovirus AcMNPV was detected in the P0 isolate, but baculovirus-free stock suspensions were eventually obtained by passaging the isolate through multiple subsequent cultures of HZ-AM1 cells. This method of selective passaging took advantage of the fact that AcMNPV cannot replicate in HZ-AM1 cells [276]. Consequently, unadsorbed baculovirus virions got successively removed with each washing and passaging of the HZ-AM1 cells. Next to baculovirus removal, continued passaging through HZ-AM1 subcultures facilitated the increase of HzNV-1 titres in the harvested isolates, up to $1.12x10^9$ IU/mL in the P6 isolate. This was an important finding as generation of defective interfering particles (DIP) could have been an intrusive factor for raising serial titres. DIPs, which carry defective viral genomes, can infect but not replicate in host cells and often arise from latent infections in baculoviral and nudiviral cell culture systems [65, 279]. Given that HzNV-1 arose from a persistent cell line infection, the deletions observed in the HzNV-1 genome, compared to HzNV-2 (**Figure 1**), likely resulted from a similar phenomenon that rendered the HzNV-1 virus defective in whole insects. However, in contrast to DIPs, HzNV-1 can still infect and replicate in noctuid cell lines. The largest deletion in the HzNV-1 genome emphasises that homologs to HzNV-2's orf91 and orf92 are missing, and the HzNV-1 orf90 homolog (orf43) has been truncated after an alanine at amino acid position 353 (Ala-353). The functions of the proteins encoded by HzNV-2's orf90, orf91 and orf92 are poorly understood, and a BLASTp search against all *Nudiviridae* members indicates that these genes are specific to the *Betanudivirus* genus, making them promising candidate genes for future studies to explore their roles for betanudivirus infection and importance for pathogenicity in whole insects.

When HZ-AM1 cells were infected with the purified HzNV-1 isolate at a high MOI, cell division almost fully halted within the first 24 hours (**Figure 5**). Stagnation of cell division upon virus infection is a common consequence observed also in baculoviruses, likely as a result of host transcriptional and translational processes being hijacked to promote gene expression and replication of the virus [280]. Alternatively, a study from 1998 showed that Sf9 cells were arrested in the G2/M phase during AcMNPV infection, also leading to virus-induced stagnation of cell proliferation [281]. Whether HzNV-1-induced proliferation arrest results from one of these mechanisms, or a different yet unidentified pathway requires, further experimental exploration.

Past this infection stage, severe cell death could be observed at 48 hpi (**Figure 4 and 5**), which eventually resulted in release of viral progeny into the cell culture medium. Electron micrographs of those extracellular HzNV-1 virions emphasised how the same type of virions can appear differently under TEM **(Figure 6B and 6C**), depending on variations in staining penetration, ultimately affecting resolution of structural details. However, it is important to mention that these differences do not indicate nucleocapsid-empty virions, but rather reflect differences in staining intensity, while still allowing for more or less discernible viral nucleocapsids and envelopes in both cases. The intracellular structures of HzNV-1-infected HZ-AM1 cells after 60 hpi revealed singly enveloped HzNV-1 nucleocapsids, either clustering together as bundles or laying individually in the nucleoplasm. A faint outline surrounding each nucleocapsid implies that HzNV-1 virions are both assembled and enveloped while still being in the nucleus. This observation aligns with findings in OrNV-infected DSIR-HA-1179 cells of coconut rhinoceros beetle embryos [30], where the host nuclear membrane became distorted, and membrane fragments are used to envelop nudiviral nucleocapsids [52], implying that this aspect of virion assembly may be conserved among coleopteran and lepidopteran nudiviruses.

Additionally, TEM revealed macropinocytosis-like projections engulfing HzNV-1 virions, likely leading to their internalisation into the cell via virion-containing vesicles. To support these observations,

an assay using the macropinocytosis inhibitor imipramine (IMP) was established. In line with this hypothesis, approximately 72% less viral DNA was detected by qPCR in cells treated with IMP where exposed to the virus compared to the control, showing that macropinocytosis is an important entry mechanism for HzNV-1 to enter HZ-AM1 cells. Interestingly, a similar pinocytosis-based entry mechanism was proposed about 40 years ago for OrNV in the DSIR-HA-1179 cell line, also suggesting a conserved entry mechanism between lepidopteran and coleopteran nudiviruses across different cell types. Despite these findings, it remains unclear which specific nudiviral protein(s) enable macropinocytosis as an entry mechanism. While *per os* infectivity factor (PIF) proteins are conserved across the order *Lefavirales*, which includes nudiviruses and baculoviruses, their exact role in nudivirus biology is still largely inferred from baculovirus studies. In baculoviruses, PIFs are known to facilitate primary infection in the midgut after oral ingestion [60, 282]. However, systemic baculoviral infections rely on budded viruses (BVs), which possess envelope proteins such as GP64 and/or F, enabling endocytosis-mediated entry into cells of numerous insect tissues and also into cultured insect cells [283-285]. Cell culture infection is severally hampered when baculovirus mutants lack budded virus envelope proteins, showing that the function of the budded virus envelope proteins cannot be complemented by the action ofPIF proteins, or only to a very limited extent [286]. Nudiviruses, on the other hand, lack F and GP64 homologues, making their entry mechanisms less understood. Although the involvement of PIFs in nudiviral entry is plausible based on sequence homology, further research is needed to confirm their role.

After entering the HZ-AM1 cells, HzNV-1 requires at least six hours to initiate DNA replication followed by an exponential increase in intracellular viral genome copies, as determined by measuring relative viral DNA levels at one-hour intervals using qPCR. Congruently, white spot syndrome virus (WSSV, family *Nimaviridae*) – a distinct relative of nudiviruses and baculoviruses within the shared *Naldaviricetes* class – was also shown to initiate DNA replication at 6 hpi in *Litopenaeus vannamei* cell lines at 27°C [287]. Similarly, a study by Rosinski *et al*. [288] reported findings with AcMNPV, where baculovirus replication began at 6 hpi in Sf9 cells. Interestingly, the same study described a two-fold increase in AcMNPV DNA levels every 1.7 hours, whereas growth in HzNV-1 DNA levels appeared to follow a more exponential trend, with viral DNA fold changes of 1.6x from 7 to 8 hpi, 2.4x from 8 to 9 hpi, and 3.4x from 9 to 10 hpi). Despite the lower MOI used for HzNV-1 infection, these findings suggest that naldaviricetes follow a similar timeline for DNA replication initiation, although the rate of nudiviral DNA replication appears to surpass that of baculoviruses.

The findings in this chapter suggest that nudiviruses use macropinocytosis as a primary cell entry route, and provide new insights into the timing of the infection process. As a future perspective on what kind of nudiviral factor(s) trigger macropinocytosis, it may be worthwhile to inspect other virus systems that employ macropinocytosis as their cell entry mechanism. For instance, vaccinia virus (VACV) and Ebola virus (EBOV) both utilise a process called apoptotic mimicry, where they expose phosphatidylserine on their surface to trick the host cell into internalizing them as if they were apoptotic debris, allowing these viruses to bypass an immunosuppressive mechanism and enter the cell in one go [289]. Furthermore, identifying the onset of HzNV-1 DNA replication provides important information for determining candidate time points for viral gene expression studies. Hence, the findings of this chapter are applied in **Chapter 5**, where the transcriptional dynamics during different stages of HzNV-1 infection are analysed.

Acknowledgements

I want to thank Els Roode for offering and sharing her excellent expertise on maintaining the insect cell cultures and HzNV-1 system. I also want to thank Christine van Battum for her outstanding assistance

in establishing and testing the methods described in this chapter during her Bachelor's thesis with me. Lastly, I would like to express my sincere gratitude to my promoters, Monique van Oers and Jean-Michel Drezen, for their constructive remarks and feedback on this chapter.

“**Nothing ever lasts forever, everybody wants to rule the world.**”

– *Everybody Wants To Rule The World* by Tears for Fears (1985)

Chapter 5

Transcriptional dynamics during Heliothis zea nudivirus 1 infection in an ovarian cell line from *Helicoverpa zea*

Jirka Manuel Petersen[1,2], Astrid Bryon[1], Annie Bézier[2], Jean-Michel Drezen[2], Monique M. van Oers[1]

[1]Laboratory of Virology, Wageningen University and Research, 6708 PB Wageningen, the Netherlands.

[2]Institut de Recherche sur la Biologie de l'Insecte, UMR7261 CNRS - Université de Tours, 37200 Tours, France.

Published as :

Petersen JM, Bryon A, Bézier A, Drezen J-M, van Oers MM. Transcriptional dynamics during Heliothis zea nudivirus 1 infection in an ovarian cell line from *Helicoverpa zea*. J Gen Virol. 2025. 106(1):002066. DOI: 10.1099/jgv.0.002066

Abstract

Nudiviruses (family *Nudiviridae*) are double-stranded DNA viruses that infect various insects and crustaceans. Among them, Helicoverpa zea nudivirus 1 (HzNV-1) represents the rare case of a lepidopteran nudivirus inducing a sexual pathology. However, studies about molecular pathological dynamics of HzNV-1 or other nudiviruses are scarce. Hence, this study aims to provide a transcriptomic profile of HzNV-1 in an ovary-derived cell line of *Helicoverpa zea* (HZ-AM1), during early (3, 6, 9 hours post-infection) and advanced (12, 24 hours post-infection) stages of infection. Total RNA was extracted from both virus- and mock-infected cells, and RNA-seq analysis was performed to examine both virus and host transcriptional dynamics. Hierarchical clustering was used to categorise viral genes while differential gene expression analysis was utilised to pinpoint host genes that are significantly affected by the infection. Hierarchical clustering classified the 154 HzNV-1 genes into four temporal phases, with early phases mainly involving transcription and replication genes and later phases including genes for virion assembly. In addition, a novel viral promoter motif was identified in the upstream region of early-expressed genes. Host gene analysis revealed significant upregulation of heat shock proteins (HSPs) and downregulation of histone genes. The identification of temporal patterns in viral gene expression enhances the molecular understanding of nudivirus pathology, while the identified host differentially expressed genes highlight the key pathways most hijacked by HzNV-1 infection.

Importance

Among the members of the order *Lefavirales*, nudiviruses have an exceptionally broad host range and diverse pathological dynamics, with new members being discovered frequently in insects and crustaceans. As nudiviruses are economically significant pathogens of invertebrates, these viruses warrant more in-depth studies to understand their molecular pathology and gene functions, but such studies are currently underrepresented. Our study offers insights into the gene expression profile of a member of the genus *Betanudivirus* with described potential as a biocontrol agent, and provides implications on how nudiviruses might modulate the cellular machinery and integrity of their host. These findings may pave the way for identifying targets to combat damaging nudivirus infections in crustaceans and insects rearing or improve nudivirus efficacy in biocontrol applications.

Introduction

Nudiviruses (family *Nudiviridae*) are insect- and crustacean-infecting viruses that share their taxonomic order (*Lefavirales*) with baculoviruses (*Baculoviridae*), hytrosaviruses (*Hytrosaviridae*), and the recently proposed family Filamentoviridae [35]. The large circular, double-stranded DNA (dsDNA) genomes that nudiviruses have in common with the other members of the *Lefavirales* can range from ~96 to ~231 kilobase pairs (kbp) among known nudivirus species. The rod-shaped virions of nudiviruses vary in length from ~120 nanometers (nm) to ~414 nm, while diameters range from ~30 nm to ~80 nm (reviewed by Petersen *et al*., 2022). In most cases, these nudiviral virions are perorally transmitted through ingestion of contaminated excrements or carcasses, e.g. through cannibalism [43, 44], but also sexual and transovarial transmissions have been observed for a member of the genus *Betanudivirus* [263, 290]. The genus *Betanudivirus* is one of the four officially recognised genera of *Nudiviridae* and is represented by two isolates belonging to the same virus species, *Betanudivirus hezeae* [53]. These two virus isolates, Heliothis zea nudivirus 1 (HzNV-1) and Helicoverpa zea nudivirus 2 (HzNV-2), are currently the only known lepidopteran-infecting nudiviruses and they can be mutually distinguished by particular characteristics. For instance, HzNV-2 can infect larvae of its natural host *Helicoverpa zea*

(Lepidoptera: Noctuidae), and cause malformations in the reproductive organs of adult moths [145], whereas HzNV-1 was originally found in a *Helicoverpa* (formerly *Heliothis*) *zea*-derived cell culture and has lost the ability to naturally infect insects [29, 127]. Nevertheless, HzNV-1 can infect a wide range of lepidopteran cell lines, including IPLB-1075 and HZ-AM1 (*H. zea*), IPLB-Sf-21 (*Spodoptera frugiperda*), IPLB-65Z (*Lymantria dispar*) and TN-368 (*Trichoplusia ni*), causing lytic infections with cytopathogenic effects (CPE) [29, 268, 291].

The most well-studied nudivirus, Oryctes rhinoceros nudivirus (OrNV), a member of the *Alphanudivirus* genus, was subject of several transcription profiling studies [138, 226, 292], which provided the first insights into nudivirus gene expression. However, these studies focused on single infection time points in *Oryctes rhinoceros* (Coleoptera: Scarabaeidae) populations, providing little information on the transcriptional changes over the course of a progressing nudivirus infection. On the other hand, progressive transcriptional changes have been investigated for bracoviruses, a clade of endogenous viruses found in braconid wasps, which originates from an ancestral nudivirus [74-76]. The available transcriptomic data on endogenous viruses of nudiviral origin involved in wasp parasitism provides valuable insights into the sequential expression of nudivirus core gene homologues integrated into the wasp genome during ovary development [82, 110, 293]. However, due to their unconventional life cycle, such bracoviral gene expression profiles in the wasp cannot simply be extrapolated to the situation in cells infected by exogenous nudiviruses. Nevertheless, it will be valuable to compare and discuss the expression pattern of pathogenic and endogenous nudivirus genes to determine to what extent these are similar.

In this study, we provide a transcriptional profile throughout early (3, 6 and 9 hpi) and advanced (12 and 24 hpi) HzNV-1 infection in HZ-AM1 cells to expand our knowledge on nudivirus cytopathology. This includes expression profiling of the 154 officially annotated HzNV-1 genes to determine their temporal expression patterns. Furthermore, we identify differentially expressed genes (DEGs) in the host at the mentioned time points to get insight into the interactions of HzNV-1 with the ovarian-derived host cells. Finally, the relevance of those transcripts for the success of the nudivirus infection versus the host immune response will be discussed.

Materials and methods

Insect cell line and virus

The main infectious agent of this study was the betanudivirus HzNV-1 [127]. The nudivirus was passaged in the cell line, HZ-AM1 (originally referred to as BCIRL-HZ-AM1), which is derived from ovarian tissue of *H. zea* [34]. The HZ-AM1 cells were grown at 28°C in HyClone™ CCM3 medium (cytiva) supplemented with 5% foetal bovine serum (FBS, Life Technologies) and 50 µg/mL gentamicin (gibco). Fresh HzNV-1 inoculum was produced from a frozen stock stored at -80 °C (P0 isolate) that was provided by Professor Yueh-Lung Wu (Department of Entomology, National Taiwan University, Taipei, Taiwan) in 2020, by collecting and filtrating (0.45 µm pore size) the supernatant of P0-infected HZ-AM1 cells two days post infection. This procedure was repeated over three passages of HZ-AM1 subcultures and the acquired HzNV-1 inoculum, hence designated as the P3 isolate. The virus titre of the P3 isolate was determined via end-point dilution assay [273].

Infection of HZ-AM1 cells with HzNV-1 for transcriptome analysis

HZ-AM1 cells were first grown in 75 cm^2 flasks (T75) as described above. Fully grown 12 mL cultures of HZ-AM1 cells were detached and resuspended in the medium to create a homogenous cell suspension. The cell suspension was then evenly distributed over up to twelve 25 cm^2 flasks (T25) in

1 mL aliquots and supplemented each with 3 mL of CCM3 medium as described above. The T25 flasks with HZ-AM1 cells were then left to grow for 24 h at 28°C. After removal of the medium from the attached cells, triplicates were incubated for 45 minutes at 28°C with the P3 isolate of HzNV-1 at a multiplicity of infection (MOI) of 5, while control groups were incubated in triplicates with virus-free medium. After virus or mock incubation, the supernatant (corresponding to virus inoculum or virus-free medium, respectively) was removed and 4 mL of fresh CCM3 medium added. This procedure was carried out for two separate infection experiments. In the first experiment, virus-infected cell cultures were harvested for RNA extraction at 3, 6, and 9 hours post-infection (hpi), each in triplicate. Mock-infected cells, serving as the control, were harvested at 0 hpi, also in triplicate. The second experiment involved both virus-infected and mock-infected cells, which were harvested at 12 and 24 hpi for RNA extraction, with all samples handled in triplicates.

Whole RNA extraction

The virus- and mock-infected cells were harvested at the defined time points for RNA extraction by removing the medium and washing them once in 1x phosphate-buffered saline (PBS). After PBS removal, 1 mL of TRIzol(R) (Thermo Fisher Scientific) was quickly added to the cells in the T25 flasks. The cells were detached with a bent glass pipette and at the same time resuspended in the TRIzol(R). A micropipette was then used to homogenise the cell-TRIzol-suspension by pipetting up and down. Next, the 1 mL cell-TRIzol-suspension was transferred to a 1.5 mL reaction tube, and incubated for 5 min at room temperature. After the addition of 200 uL chloroform to each sample, the suspension was mixed by vigorously shaking the tube. The sample was again incubated at room temperature for 5 min, and subsequently centrifuged for 15 min at 12,000 x g at 4°C. The upper colourless aqueous phase containing the RNA was then transferred to a new 1.5 mL tube and mixed with 500 µL isopropanol. The sample was incubated for 10 min at room temperature and subsequently centrifuged at 12,000 x g for 10 min at 4°C. The supernatant was removed and the RNA pellet was washed in 1 mL 75% ethanol, followed by a brief stir on the vortex mixer. Afterwards, the sample was centrifuged for 5 min at 7,500 x g at 4°C, and the ethanol was carefully removed from the pellet. To remove residual ethanol, the lid of the tube was left open for 5 to 10 minutes at room temperature. Finally, the air-dried pellet was resuspended in 30 to 50 uL RNAse-free water by flicking, and the RNA concentration was determined spectrophotometrically.

RNA-sequencing and differentially expressed genes analysis

All RNA samples were processed for sequencing according to the TruSeq stranded total RNA Ribo-Zero H/M/R Gold protocol (Illumina) and yielded RNA concentrations ranging from 3.33 to 20.90 ng/µL after library preparation. The resulting library was sequenced by Macrogen Inc. (Seoul, South Korea) using the NovaSeq6000 sequencing system to generate paired-end reads of 151 bp. The raw, paired-end reads were filtered and trimmed with fastp v0.23.2 [294] under default settings and the clean reads were used for further analyses. Host and virus reference genomes were indexed by using the "hisat2-build" command in HISAT2's v2.2.1 [295]. In the case of *H. zea*, two separate files containing lists of previously extracted exons and splice sites were provided to the hisat2-build command using the --exon and --ss parameters, along with the reference genome file. Afterward, the filtered paired-end reads were aligned to the indexed *H. zea* reference genome (ilHelZeax1.1, GenBank: GCA_022581195.1) using HISAT2. The alignment was performed with the -x option to specify the genome index, and the -1 and -2 options to provide the paired-end read files. The output was saved as a SAM file using the -S parameter, while unaligned read pairs were captured with the --un-conc option to save them into separate files. The unmapped reads were subsequently mapped with HISAT2 to the indexed HzNV-1 genome using the -x option to specify the genome index, and the -1 and -2 options to assign the unmapped paired-end reads, resulting in the generation of another SAM file [142]. The

obtained SAM files were sorted and converted to BAM files with SAMtools v1.15.1 [296]. Percentages of reads that aligned with the respective reference genome were extracted from the BAM files by using the "flagstat" command from SAMtools. The obtained BAM files were used as input for the transcript assembly tool StringTie [297] while specifying the -e option to estimate transcript abundances, and the -B option to create a Ballgown [298] table that stores the transcript-level abundance estimates, as well as including the --rf option for TruSeq stranded data. The resulting files were then submitted to the python script "prepDE.py" (https://github.com/gpertea/stringtie/blob/master/prepDE.py, accessed June 10th, 2022) to convert the host and virus read counts to their respective CSV files with read count matrices for genes and transcripts. Due to the small sample sizes in our experiments, we used the R package limma v3.50.3 [299] to test for differential host gene expression during early (3, 6 and 9 hpi) and advanced (12 and 24 hpi) HzNV-1 infection in HZ-AM1 cells. The early infection time points (3, 6 and 9 hpi) were all compared to a mock-infected sample at 0 hpi, while the two advanced infection time points (12 and 24 hpi) were compared to the respective mock-infected time points at 12 hpi and 24 hpi. Host genes with absolute read counts below six were considered lowly expressed and filtered out, while the remaining reads were normalised using the limma-voom method, which includes quantile normalisation to adjust for distributional differences [300]. For HzNV-1, the read counts for each sample were then normalised based on the total number of trimmed and filtered reads per sample (including both host and viral reads) and used for downstream analyses. Host genes were considered differently expressed if their false-discovery rate (FDR) values were lower than 0.05, and changes in their gene expression levels exceeded a log2 fold change (log2FC) = log2(1.5), accounting for an increase or reduction in gene expression by 50%. The proteome of *H. zea* was annotated with the "Annotate Your Proteome" function of the STRING: functional protein association networks webtool v12.0 [301]. The inferred proteome was created under the string taxid "STRG0A69IJE" and yielded a list of 23696 proteins, which is publicly accessible. This proteome was then supplied to the "Multiple proteins" function of STRING, together with the list of the 570 different host genes that were differently expressed during HzNV-1 infection. After inferring the protein network, protein clusters were determined by choosing the MCL clustering option in the "Clusters" tab and specifying a MCL inflation parameter of 1.5. The "Exports" tab was used to extract the functional enrichment terms with their assigned proteins. The protein networks were then imported from the STRING webtool into Cytoscape v3.10.2 [302], where network layouts were tailored and nodes (proteins) coloured based on the log2FC of their corresponding genes. Venn diagrams showing the conservation among DEGs of different infection time points were created with the vennDiagram() function of the aforementioned limma R package. Additionally, host DEGs were visualized in time point-specific volcano plots (**Figure S1 in Supplementary data**).

HzNV-1 gene transcription profiling

Hierarchical clustering, using the R package pheatmap (v1.0.12) [303] function with complete linkage and Euclidean distance, was applied to cluster the normalised viral read counts, enabling the assignment of transcriptional patterns to the HzNV-1 genes. Instead of performing hierarchical clustering on the read counts of all time points, we selected two time points (3 hpi and 6 hpi) to elucidate patterns in transcription initiation. For this approach, the viral read counts of the 3 hpi samples were first visualised with pheatmap, and the optimal number of clusters was determined using the NbClust v3.0.1 R package [304]. This led to the identification of 2 as the optimal number of clusters to distinguish between expressed (Phase 1) and non-expressed genes at this earliest time point of infection. The analysis was continued using the same R tools for the 6 hpi time point, first excluding the Phase 1 genes. Under these conditions, most of the NbClust indices returned 3 as the optimal number of clusters for the corresponding read counts, allowing the remaining three gene clusters (Phase 2 – 4) to be inferred. A figure showing the two different heatmaps of 3 hpi and 6 hpi, and their respective clusters can be found

in the **Supplementary data** (**Figure S2**). Finally, a heatmap with the data from all samples (3 hpi to 24 hpi) was inferred and genes were sorted based on the previously determined clusters (**Figure 2A**). The functional distribution of significant HzNV-1 genes from Phase 1 to Phase 4 was visualised as a stacked bar chart (**Figure 2B**). The locations of the Phase 1 to Phase 4 genes in the HzNV-1 genome were visualised using Snapgene software v7.1 (**Figure 2C**).

Promoter analysis

Baculovirus-associated promoter motifs have been previously identified in HzNV-1, including the late promoter [ATG] TAAG and the early TATA box promoters TATA [A/T] T [A/T] [142]. In this study, we aimed to search for novel HzNV-1 promoter sequences and test their congruence with our RNA-seq data. We therefore manually copied the 300-bp-upstream nucleotide sequences of each HzNV-1 gene from Snapgene into a text file, which was then used as input to perform a promoter analysis using the MEME Suite 5.5.4 tools [305], with the following settings: -dna (DNA alphabet), -zoops (zero or one occurrence per sequence), -minw 6, -maxw 20 (size of the motifs between 6 and 20 bp), -nmotifs 4 (search for 4 different motifs). The significant enrichment of motifs detected by MEME in HzNV-1 was further validated using the Analysis of Motif Enrichment (AME) tool [306]. The 300-bp-upstream nucleotide sequences from genes of nudiviruses belonging to different genera than HzNV-1 (OrNV: *Alphanudivirus*; Callinectes sapidus nudivirus, CsNV: *Gammanudivirus*; Tipula oleracea nudivirus, ToNV: *Deltanudivirus*; Franciscoloa pallida nudivirus, FrpNV: unclassified; Cuculoecus africanus nudivirus, CafNV: unclassified) and the baculovirus Autographa californica multiple nucleopolyhedrovirus (AcMNPV) were used as control sequences. The control sequences of each virus were supplied to individual AME analyses together with the 300-bp-upstream nucleotide sequences from HzNV-1, along with the identified motifs by MEME. Each AME analysis was run under the default parameters (Fisher's exact test). The Tomtom tool [307] was used with default parameters to identify binding site profiles that matched the enriched motifs, which were previously derived from the non-redundant JASPAR 2022 core DNA database. Finally, the Find Individual Motif Occurrences (FIMO) tool [308] was used to search for occurrences of the enriched motifs in the 500-bp-upstream nucleotide sequences of all protein-coding *H. zea* genes.

Prediction of transcription start sites in the HzNV-1 genome

Transcription start sites (TSS) of HzNV-1 were predicted using the R package TSSr [309] by following the standard workflow (https://github.com/Linlab-slu/TSSr, accessed April 19th, 2024) up to the annotation of TSS clusters (TCs). For this purpose, RNA-seq data from 15 paired-end BAM files representing short-reads mapped to the HzNV-1 genome at different time points post-infection (3, 6, 9, 12 and 24 hpi) were used. Specifically, the BAM files were loaded into R and used to create a TSSr object with the new() function. The TSSr object was then used as input for the gets() function to identify the genomic coordinates of transcription start sites. Biological replicates of each time point were merged with the mergeSamples() function to reduce variability. Identified TSS were filtered and normalised using a Poisson distribution under default settings, and then clustered into TCs with the clusterTSS() function using the "peakclu" method with a peak distance of 50 bp. Consensus clusters were determined with the consensusCluster() function under default settings and these clusters were annotated with the annotateCluster() function. Adjustments were made to the parameters of the annotateCluster() function to account for the compact nature of the viral genome: "upstream" was set between 130 and 200, "upstreamOverlap" was set to 20, and "downstream" was set between 0 and 100 to account for genes whose TSS localised downstream of their start codon. The visual assessment of the HzNV-1 BAM file coverage was performed in the Geneious software v2022.2.1 and visually compared to the TSS predicted by TSSr.

Results and discussion

Viral reads increase while host read abundance decreases with progressing HzNV-1 infection

Total RNA was extracted from HzNV-1 infected (MOI = 5) and mock-infected HZ-AM1 cells at different time points (3, 6, 9, 12, 24 hpi), and subjected to an RNA-seq analysis to investigate transcriptional dynamics of the host and virus over the course of infection. The reads were aligned to the host genome and reads that did not align to the host genome (unmapped reads) were extracted and aligned to the HzNV-1 genome. The percentages of reads aligned to the host or the virus genome are visualised in a line graph (**Figure 1**).The exact number of reads for each sample and the respective percentages of reads that aligned to the host or virus genome can be found in the **Supplementary data** (**Table S1**).

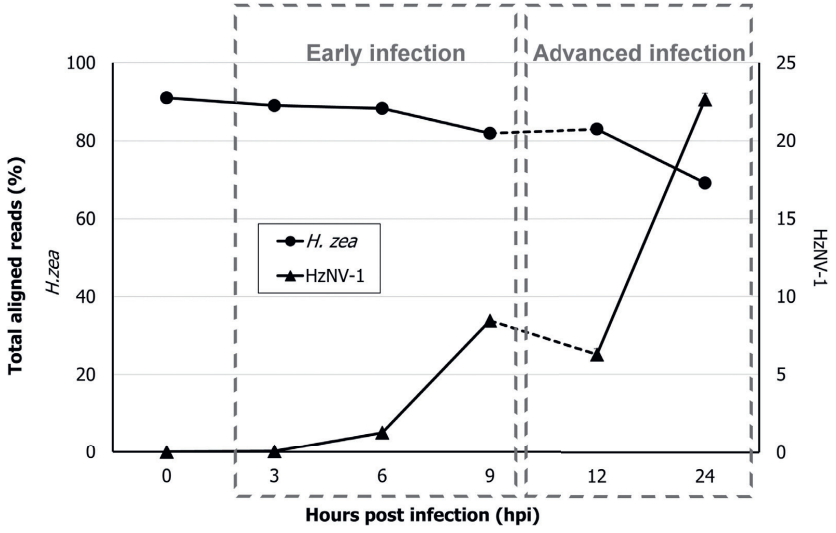


Figure 1. Time-dependent changes of mRNA reads from *H. zea* and HzNV-1 in percentages. The percentages of total reads mapped to the *H. zea* genome (triplicate average percentage of mRNA reads) are displayed as circles (left y-axis), and the average percentages of the total HzNV-1 reads are displayed as triangles (right y-axis). The dashed lines between 9 hpi and 12 hpi emphasise that those time point are from two distinct experiments (early infection: 0, 3, 6 and 9 hpi; advanced infection: 12 and 24 hpi).

Notably, although the RNA-seq analysis was conducted in two independent experiments (early versus advanced infection; **Figure 1**; see also Materials & Methods), both experiments showed a similar trend. However, the introduction of some uncertainty in the transition of aligned reads between the 9 to 12 hpi interval (early and advanced experiments respectively) needs to be considered, given the drop of aligned virus reads at 12 hpi compared to 9 hpi. Whether this decline is attributable to an antiviral response that the virus subsequently adapts to – resulting in a recovery of viral gene expression at 24 hpi – requires further experimental validation. The percentages of reads mapped to the host (*H. zea*) and the virus (HzNV-1) represent the proportion of reads that aligned to each genome, relative to the total number of filtered and trimmed reads. For instance, at the earliest measured stage of infection (3 hpi), 19,390 reads out of a total of 54,821,348 aligned to the viral genome, representing an average of 0.04% of the total reads (SEM ± 0.0095%). In contrast, the highest percentage of HzNV-1 aligned reads was observed at 24 hpi with 22.64% (SEM ± 0.38%) of the total reads mapping to the viral genome. During this interval,

the largest increase occurred from 3 hpi to 6 hpi with 35 times more viral transcripts. In contrast, host-specific reads gradually plummeted from a maximum of 91.04% (SEM ± 0.35%) at 0 hpi to a final low of 69.16% (SEM ± 0.56%) at 24 hpi. The decreasing percentages (and absolute values; **Table S1 in Supplementary data**) of host-specific reads over time suggest a reduction in overall host gene expression due to infection with HzNV-1. Similar observations have been made during bracovirus gene expression [310] and baculovirus infections [311], and also in viruses that are more distantly related to nudiviruses [312]. Impairing the global gene expression of the host can help the virus to obstruct immunity-associated responses, while the virus hijacks the host's cellular apparatus and resources to meet its requirements by up-regulating those host genes that are crucial for virus replication. However, further gene expression studies, incorporating normalisation with a stable host reference gene, are needed to validate this proposition.

Hierarchical clustering groups HzNV-1 genes into four temporal classes

We analysed transcriptional patterns of HzNV-1 genes by hierarchical clustering with Euclidean distance on normalised viral read counts at 3 hpi and 6 hpi to assign distinct temporal phases to the viral genes (see Materials & Methods). At 3 hpi, two gene clusters were identified, distinguishing between expressed (Phase 1) and non-expressed genes. At 6 hpi, excluding Phase 1 genes, the three remaining gene clusters were assigned (Phase 2, Phase 3, Phase 4). A heatmap with the read counts of all samples (3-24 hpi) was then created, and the HzNV-1 genes were grouped based on their assigned phases and sorted from largest to smallest number of reads (mean of triplicates; **Figure 2A**). The functional distribution of the phase-associated HzNV-1 genes is depicted in a stacked bar chart (**Figure 2B**), and their genomic locations in the HzNV-1 genome were displayed (**Figure 2C**). Based on our hierarchical clustering approach, we were able to group the 154 HzNV-1 protein coding genes into 37 Phase 1 genes (24.02%), 41 Phase 2 genes (24.02%), 43 Phase 3 genes (27.92%) and 33 Phase 4 genes (21.42%) (**Figure 2A**). In addition, the non-coding persistency associated gene (*pag1*) clusters in Phase 1, and is the transcript with the highest abundance of all transcripts in this phase, which may highlight an important role in early HzNV-1 infection, in addition to an earlier reported function in persistent infections [313]. The 28 nudiviral core genes are a set of genes that can be found in the genomes of all nudivirus isolates sequenced to date, of which 21 genes are homologous to baculovirus and 16 genes are homologous to hytrosavirus core genes. These core genes display an uneven dispersion across the four temporal classes, which likely reflects their different functions in the virus life cycle. Nudiviral core genes from Phase 1 include the transcription-associated late expression factors *lef-4*, *lef-5*, *lef-8* and *lef-9*/*p47*, and the DNA polymerase, *dnapol* (**Figure 2A**).

In other *Lefavirales* members and in bracoviruses, the *lef* genes are early expressed genes transcribed by the RNA polymerase of the infected cell essential for both late gene expression and virus replication [41, 239, 314, 315], whereas the DNA polymerases of baculoviruses and hytrosaviruses have also previously been described as early genes [315, 316], as they are required for the viral DNA replication. Unlike core genes, accessory genes are genes that are frequently but not universally present in the genomes of members within a clade. The accessory genes encoding a dUTPase (*orf69*) and thymidine kinase (TK1/*orf51*) also clustered in Phase 1. Those two proteins (and DNA polymerase B – the same type of DNA polymerase encoded by nudiviruses) were all proposed to be involved in a pathway that oversees misincorporation of uracil residues into viral DNA to ensure the fidelity of viral DNA genomes replication [317]. Congruently, two genes encoding for a deoxynucleoside kinase (DNK/*orf71*) and a thymidylate synthase (TS/*orf109*), both known to be involved in nucleotide metabolism [318-320], cluster in Phase 1. Furthermore, Phase 1 featured virus genes encoding for proteins with putative roles in host remodelling (*i.e.*, inhibition apoptosis or host gene expression),

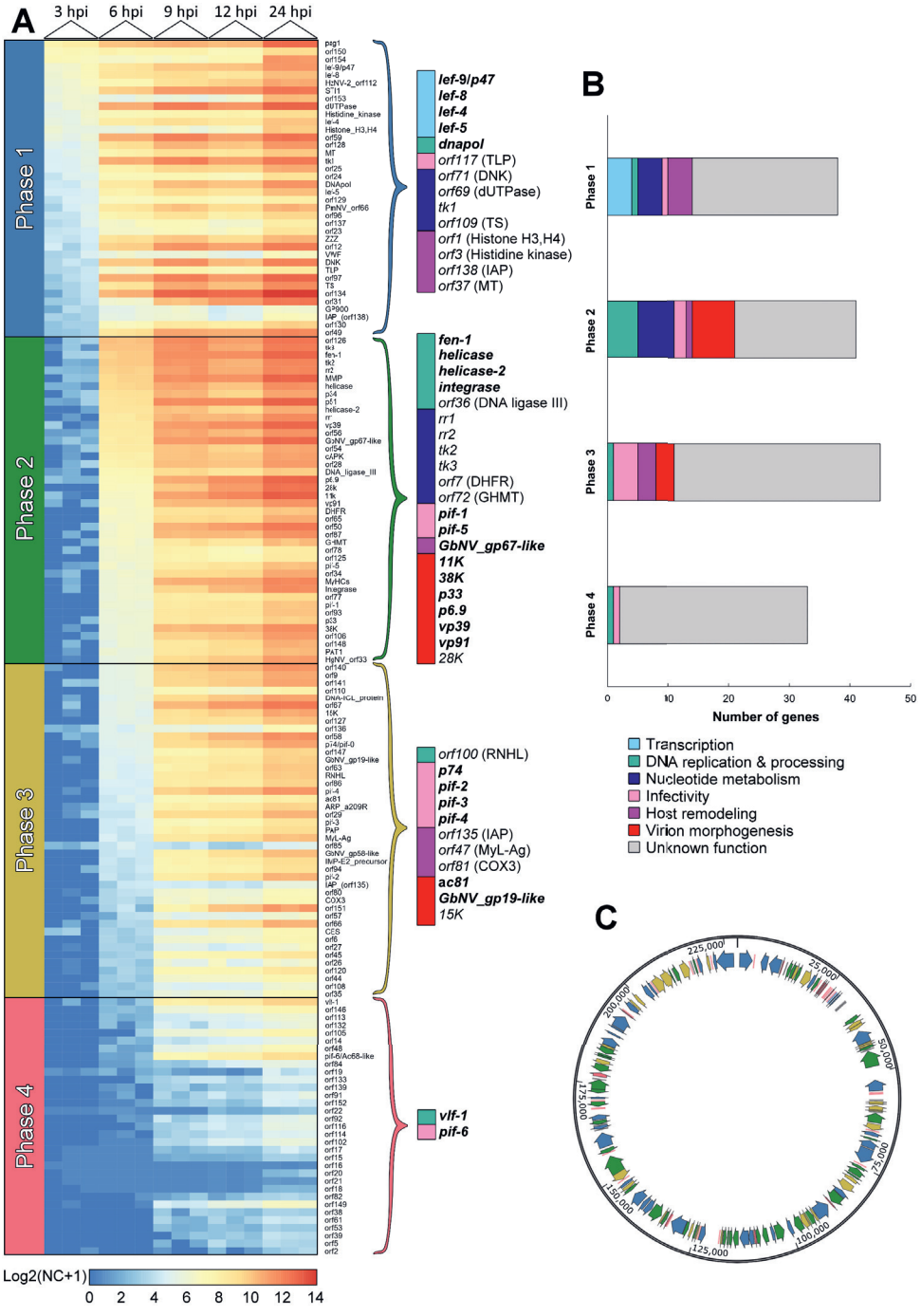

A
3 hpi
6 hpi
9 hpi
12 hpi
24 hpi
Phase 1
Phase 2
Phase 3
Phase 4
Log2(NC+1)
0 2 4 6 8 10 12 14
lef-9/p47
lef-8
lef-4
lef-5
dnapol
orf117 (TLP)
orf71 (DNK)
orf69 (dUTPase)
tk1
orf109 (TS)
orf1 (Histone H3,H4)
orf3 (Histidine kinase)
orf138 (IAP)
orf37 (MT)
fen-1
helicase
helicase-2
integrase
orf36 (DNA ligase III)
rr1
rr2
tk2
tk3
orf7 (DHFR)
orf72 (GHMT)
pif-1
pif-5
GbNV_gp67-like
11K
38K
p33
p6.9
vp39
vp91
28K
orf100 (RNHL)
p74
pif-2
pif-3
pif-4
orf135 (IAP)
orf47 (MyL-Ag)
orf81 (COX3)
ac81
GbNV_gp19-like
15K
vlf-1
pif-6
B
Phase 1
Phase 2
Phase 3
Phase 4
0 10 20 30 40 50
Number of genes
Transcription
DNA replication & processing
Nucleotide metabolism
Infectivity
Host remodeling
Virion morphogenesis
Unknown function
C
225,000
25,000
50,000
75,000
100,000
125,000
150,000
175,000
200,000

Fig 2. Gene expression profile of HzNV-1 genes over time. **(A)** Heatmap showing the log2 transformed, normalised counts (NC) from each triplicate of the respective time point shown in the indicated rows. Gene expression levels range from blue (low expression) to red (high expression). The 154 viral ORFs (excluding *pag1* and PAT1) were grouped into 37 Phase 1 genes, 41 Phase 2 genes, 43 Phase 3 genes and 33 Phase 4 genes. Tall curly brackets emphasise notable HzNV-1 genes from each phase (nudiviral core genes written in bold), with their predicted protein annotations in paratheses. **(B)** Stacked bar chart illustrates the proportional distribution of emphasised HzNV-1 genes across the four temporal classes (Phase 1 – 4) according to their respective functions. **(C)** HzNV-1 genes distributed along the genome, colour-coded according to the phase to which they are appended, based on our clustering approach. **See previous page**.

including a variant of the inhibitor of apoptosis type 3 protein (IAP/*orf138*), which is a protein with a histone-like domain (Histone H3,H4/*orf1*), a histidine kinase (*orf3*), and a methyltransferase (MT/*orf37).*

Almost half of the nudiviral core genes (12 in total) were assigned to Phase 2: *fen-1, helicase, helicase-2, integrase, pif-1, pif-5, GbNV_gp67-like, 11K, p33, p6.9, vp39* and *vp91*. Accessory genes assigned to Phase 2 encode for a DNA ligase III (*orf36*), two thymidine kinases (TK2/*orf111* and TK3/*orf115*) and two ribonucleotide reductases (RR1/*orf95* and RR2/*orf73*). Other notable genes within Phase 2 include a dihydrofolate reductase (DHFR/*orf7*), a glycine hydroxymethyltransferase (GHMT/*orf72*), and the virion structural protein 28K (*orf99*). Helicase, helicase-2, FEN-1, integrase and DNA ligase III have been labelled as proteins involved in DNA replication/processing [171], while orthologs of the proteins TK2 and TK3 [321], RR [322, 323], as well as DHFR and GHMT are all associated with nucleotide metabolism [324], which includes processes essential for deoxynucleotide synthesis and thus DNA replication. The nudivirus core gene *GbNV_gp67-like* is thought to encode a microtubule-associated protein with a putative function in viral replication by remodeling the microtubule network of infected cells [325, 326], but experimental data are needed to verify that this nudiviral protein acts in the same way as homologues in other lefavirals. The *pif-1* and *pif-5* genes encode proteins known as *per os* infectivity (PIF) factors (PIF-1 and PIF-5, respectively), which play an essential role in oral infectivity as structural components of the occlusion-derived virions (ODVs) in baculoviruses [60, 282]. PIF homologues from bracoviruses, hytrosaviruses, nudiviruses and the more distantly related nimaviruses (*Nimaviridae*) are thought to serve similar functions [171, 327]. Of the remaining genes mentioned above, some encode homologues of proteins known to be involved in nucleocapsid assembly (*vp39*, *vp91*), virion production, and envelopment (*p33*) [59, 90, 94, 328, 329], or in viral DNA processing (*p6.9*) [96, 171]. The core gene *11K* and the accessory gene *28K* both encode virion structural proteins [330]. In summary, 58.82% of genes (20 out of 34) expressed in the early stages of HzNV-1 infection (first 6 hours) are more or less directly associated with viral DNA replication and transcription, while 26.47% encode virion components. A significant increase in the relative level of viral DNA was measured from 7 hpi onwards (**Chapter 4, Figure 11**), which logically confirms the observed timeline and the genes found to be expressed in Phase 2: first the expression of all genes necessary for the virus replication process, followed by the replication of the virus.

Genes that are abundantly expressed in Phase 3, are mostly nudiviral core genes related to infectivity (*p74*, *pif-2*, *pif-3*, *pif-4*), virion morphogenesis (*38k*, *ac81*, *GbNV_gp19-like*), or genes with unknown functions (*GbNV_gp58-like*). Phase 3 accessory genes include *orf100,* which encodes a Ribonuclease H-like/RNase protein (RNHL), as well as *orf135*, which encodes a second inhibitor of apoptosis protein. The PIF-encoding genes (*pif-1* and *pif-5*), mentioned above in Phase 2, are still expressed in Phase 3. Homologues of all the mentioned-*pif* genes facilitate the fusion of baculoviral ODVs with host midgut cells [60, 171], while homologues of the proteins 38K, AC81 and 15K ensure proper nucleocapsid assembly and envelopment [40, 94, 328, 331]. While the function of the protein encoded by GbNV_gp19-like has not yet been experimentally characterised, it contains an α/β hydrolase domain,

as identified by InterPro [267]. Further analysis using the HMMER web tool [332] classified GbNV_gp19-like within the Carboxymethylbutenolide lactonase family, although this top match was supported by only a marginally significant E-value of 4.2E−03. Similar to *orf138* in Phase 2, the IAP protein encoded by *orf135* likely facilitates prolonged survival of host cells, providing the virus with additional time to replicate, a mechanism akin to that demonstrated for IAP homologues in baculoviruses [333].

In the latest stage, Phase 4, all the HzNV-1 genes are being expressed with the exception of the two core genes *pif-6* and *vlf-1*, that have no predicted function. All Phase 4 genes are expressed at low levels even at 24 hpi. In baculoviruses PIF-6 cooperates with PIF-4 to stabilise the ODV entry core complex for better proteolytic resistance [89]. Previous studies demonstrated that the baculovirus homologue of the very late expression factor 1 protein (VLF-1) is a structural component of the nucleocapsid. It is hypothesised to execute the crucial final step of processing large concatemeric viral DNA intermediates into lengths suitable for proper packaging within the nucleocapsid [97], which is consistent with VLF-1 expression occurring during the latest stage of viral infection.

Based on our clustering approach, we can deduce that HzNV-1 genes with transcription- and replication-associated functions initialise viral DNA replication in the earlier infection stages (Phase 1 and 2), while viral genes mostly involved in virion maturation and assembly showed expression at a more advanced stage of infection (Phase 3 and 4) (**Figure 2B**). Coherent with what is known from baculovirus studies, virus replication and transcription are needed to establish late gene expression through the viral RNA polymerase, while early genes are usually transcribed by a host RNA polymerase [334-336]. Early expressed accessory genes, encoding proteins such as viral thymidine kinases (*tk1*, *tk2*, *tk3*), viral ribonucleotide reductases (*rr1*, *rr2*), or deoxynucleotide kinase (*orf71*), allow the virus to partially circumvent the metabolic limitations of its host by using this arsenal to aid nucleic acid production [337]. The expression of these genes at the onset of HzNV-1 infection likely contributes to the creation of optimal conditions for viral DNA synthesis and, hence, replication.

Regarding the expression kinetics of the nudiviral core genes, similar results were previously obtained for the endogenous nudivirus harboured by the wasp *Cotesia congregata* (Cotesia congregata bracovirus, CcBV) [82]. Accordingly, transcription-associated genes (*p47*, *lef-9*, *lef-8*, *lef-4*, *lef-5*) were found to be more highly expressed than most other genes in the ovaries of the earliest developmental wasp stage, followed by genes involved in replication (*helicase*, *fen-1*), while structural nudiviral core genes reached their highest expression levels in the later developmental stages of the wasp. Thus, despite undergoing endogenisation, the nudiviral core genes maintained in bracoviruses, have largely retained the gene expression profiles of their free-living nudiviral ancestor.

A putative HzNV-1 promoter motif regulates early gene expression primarily associated with viral genes for transcription and replication

As part of the search for novel promoter sequences thought to be involved in early and late expression of the HzNV-1 genes, a comparison of the upstream sequences of all HzNV-1 genes was performed, with respect to the four identified phases. When the 300-bp upstream sequences of the genes assigned to Phase 1 to 3 were compared using MEME analysis, a motif corresponding to the consensus sequence ATA[G/C]G[G/C]TAT stood out significantly (E-value = 9.1E-012; Fig 3A). With no allowed mismatches, it was identified in the upstream sequences of 34 genes (**Supplementary data, Table S2**), representing 22.08% of all HzNV-1 genes. If a motif was located in the upstream sequences of two adjacent genes, it was counted separately for each gene, resulting in 15 genes in Phase 1, 13 genes in Phase 2, 6 genes in Phase 3, and none in Phase 4. Hence, this motif does not seem to be linked to a single temporal class of genes, however, it is preferentially found in early expressed genes (Phase 1 and

2, more rarely Phase 3). Additionally, using AME, the enriched ATA[G/C]G[G/C]TAT motif was tested for enrichment against sequences from related nudiviruses, including CafNV, CsNV, FrpNV, OrNV, ToNV, as well as the more distantly related baculovirus AcMNPV. The analyses indicated that this motif is significantly enriched in the upstream sequences of HzNV-1 when compared to the other nudiviruses and AcMNPV, as evidenced by significantly low p-values (1.21E-10 to 7.74E-09), whereas AME showed no significant difference in the occurrence of this motif between HzNV-1 and HzNV-2 (p-value = 0.496). This suggests that the motif is specific for the genus *Betanudivirus* and does not appear to be conserved in the other nudiviruses, which served as the comparative reference point in this analysis. The Tomtom-assisted search for binding site motifs with homology to the ATA[G/C]G[G/C]TAT motif resulted in multiple alignments with binding sites of the transcription factor family "sine oculis homeobox" (SIX; top score p-value = 8.97E-04). The family of SIX proteins is a group of transcription factors that fulfils developmental functions in a broad range of organisms, including both vertebrates and invertebrates, and the evolutionary lineage of this family extends back nearly 500 million years [338]. Although the p-value obtained is only weakly significant, members of the SIX family are known for their RNA polymerase II (RNAP II)-specific DNA-binding properties [339, 340], which would align with early gene expression.

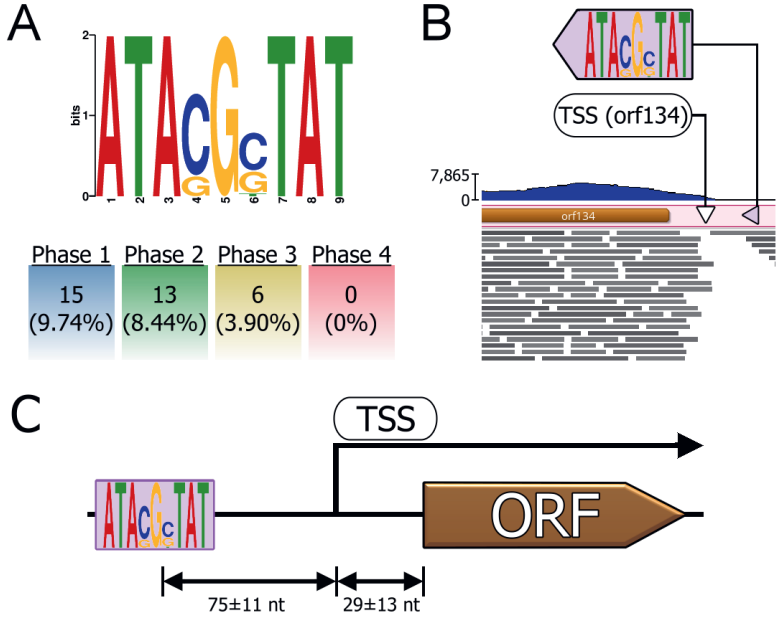


Fig 3. Analysis of the newly identified HzNV-1 motif, showing its prevalence and position relative to transcription start site (TSS) upstream of HzNV-1 genes. **(A)** Sequence logo of the significantly enriched motif "ATA[G/C]G[G/C]TAT" identified from the 300-bp upstream sequences of all HzNV-1 genes. The proportion of Phase 1, Phase 2, Phase 3 and Phase 4 genes among all 154 HzNV-1 genes with the motif are indicated as percentages in brackets. **(B)** Exemplary mapping of the paired-end RNA-seq reads (6 hpi) in the upstream region of HzNV-1's orf134 along with the TSS and "ATA[G/C]G[G/C]TAT" motif visualised with Geneious. The motif of orf134 is shown (purple triangles) to exemplify the distance of the motif and start codon to the TSS (white triangle) identified by TSSr. The read coverage graph is displayed as a blue curve with scale on the left and predicted ORF in brown. **(C)** Schematic showing the mean ± SEM nt distances between motif, start codon and TSS calculated from all ATA[G/C]G[G/C]TAT-containing HzNV-1 genes.

We further inspected the degree of proximity between TSS identified and start codons to the ATA[G/C]G[G/C]TAT motif for the respective HzNV-1 genes. The TSSr tool was able to predict the TSS for 28 of the 34 ATA[G/C]G[G/C]TAT motif-containing virus genes. We used this information to visualise (**Figure 3B**) and calculate the mean distances from those TSS to the motif of interest and the start codons of the respective genes (**Figure 3C**). The nucleotide (nt) distances between motif and TSS (Mean = 75, SEM = ±11) showed similar variability that the distance of the translational start codon to the TSS (Mean = 29, SEM = ±13). While these results offer initial insights on the spatial dynamics of the ATA[G/C]G[G/C]TAT motif to paired-end RNA-seq read coverages, experimental data using accurate sequencing techniques such as CAGE- and TSS-seq, will be necessary to confirm its role as a novel HzNV-1 motif involved in transcription.

Several genes with this motif appear to be involved in virus transcription and replication, including *lef-5*, *dnapol*, *helicase-2*, *tk1*, *tk2*, *rr1*, *rr2*, dUTPase (*orf69*) and deoxynucleotide kinase (DNK). Those genes are all of critical importance for early HzNV-1 infection, and essential to initiate late viral gene expression and establish an optimal environment with nucleic acid components for viral genome synthesis. It is particularly noteworthy that some of these viral genes containing that motif (*tk1*, *tk2*, *rr1*, *rr2*, *orf69*) have been described to be of eukaryotic host origin having been acquired by the virus through horizontal gene transfer [324]. In the context of the host, a FIMO-assisted search for the ATA[G/C]G[G/C]TAT motif (no mismatches allowed) in the upstream sequences of *H. zea* genes resulted in 251 hits, of which 170 corresponded to annotated *H. zea* genes. Among those, there were host genes with similar functions as those HzNV-1 genes with the ATA[G/C]G[G/C]TAT motif, including an RNA polymerase, two helicases, and seven zinc finger proteins. Although there is a possibility that HzNV-1 acquired these host genes – as it has been mentioned for *tk1*, *tk2*, *rr1*, *rr2* and *orf69* – these *H. zea* genes do not exhibit sequence homology with the corresponding HzNV-1 ORFs. This suggests that if these genes were indeed acquired from the host along with their regulatory motifs, they have greatly diverged within the virus genome beyond recognition . In addition to capturing host genes to expand the viral genomic arsenal, the acquisition of host genes together with their binding motifs may have helped to improve the binding efficiency of the host RNA polymerase II to HzNV-1 promoter regions to ensure proper expression of these critical genes.

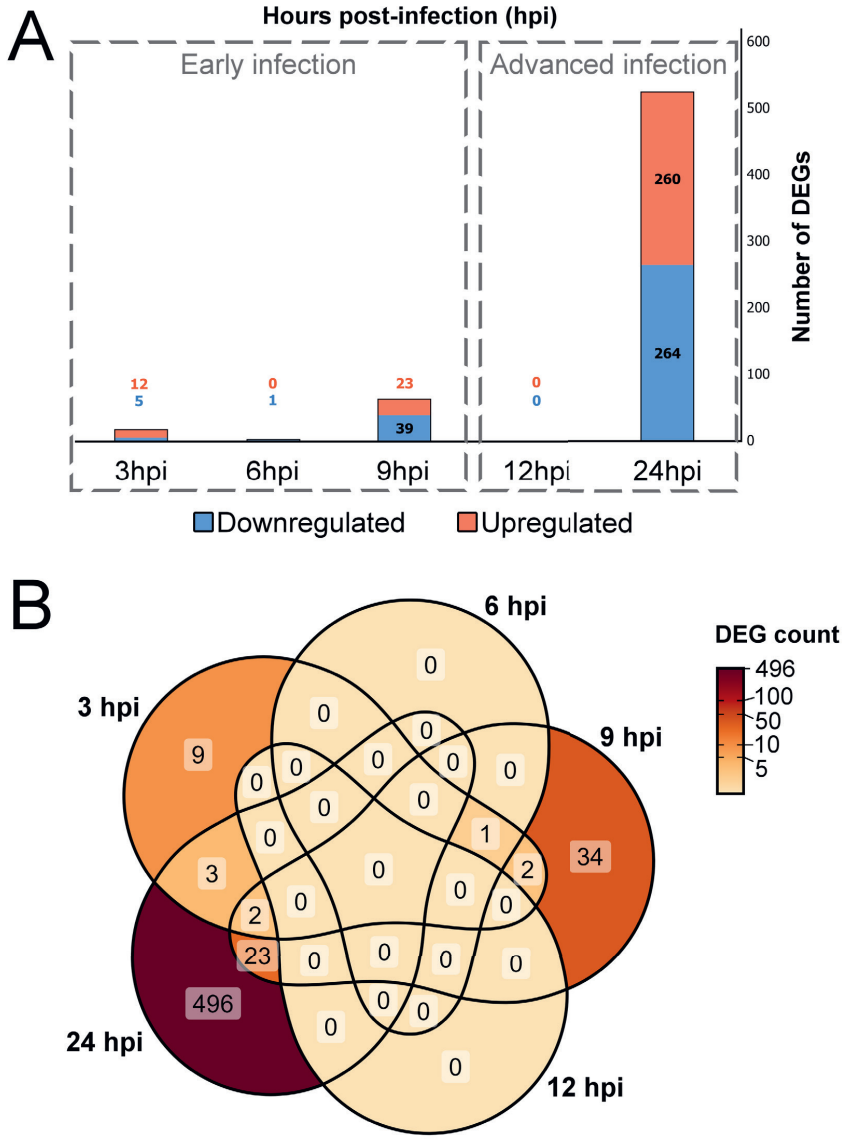
Numbers of host differentially expressed genes fluctuate over the course of HzNV-1 infection

Our analysis yielded a total of 570 distinct host DEGs, of which 74 were associated with the early stages of HzNV-1 infection (3, 6 and 9 hpi), and 524 DEGs were determined during advanced infection (12 and 24 hpi) when compared to mock-infected cells (**Figure 4A**). Genes with an FDR value < 0.05 and a log2(FC) > log2(1.5) were considered differentially expressed. The time point-specific numbers of DEGs varied with progressing virus infection, including 12 up- and 5 downregulated genes at 3 hpi, one downregulated gene at 6 hpi, 23 up- and 39 downregulated genes at 9 hpi , and no DEGs at 12 hpi. The number of DEGs reached its peak at 24 hpi with 260 up- and 264 downregulated genes, highlighting an overall greater amplitude of virus-induced transcriptional changes in the host cells during advanced infection. The “zigzag”-like fluctuations in the number of DEGs between 3 and 12 hpi resemble the well-known zigzag model in plant pathology, which depicts the interplay between host immune responses and viral counter-effectors over time [341]. Notably, like plants, insects also use RNA interference (RNAi) as an antiviral defense, and it is well-documented that the RNA silencing pathway plays a key role in the zigzag model [342]. However, while the original zigzag model is qualitative and lacks specific units for measurement [343], our findings would suggest that the number of DEGs could serve as a quantifiable indicator of these fluctuating defense amplitudes during host-virus interactions.

Naturally, additional experimental data would be needed to support this hypothesis. Moreover, any generalisation to other biological systems should be approached with great caution.

A total of 28 specific DEGs were identified in both early and advanced stages of infection (**Figure 4B**). Notably, none of the upregulated DEGs were shared across all three stages of early infection. However, four downregulated genes were common between the 3 hpi and 9 hpi time points. Interestingly, the only DEG identified at 6 hpi was also detected as DEG at 3 hpi and 9 hpi. This DEG encodes for an uncharacterised protein (LOC124637371), and based on InterPro analysis, it contains transmembrane helices and a cytoplasmic domain. This implies that this protein likely plays a role in interactions between the virus and host cellular membranes, potentially influencing viral replication or pathogenesis, especially during early infection. The majority of DEGs were unique to the advanced infection stage at 24 hpi, with only a few overlaps between early time points. The most significant overlap between early and advanced infection occurred between 9 hpi and 24 hpi, with five upregulated and 18 downregulated genes shared between these stages. This highlights that certain host genes may play crucial roles in both early and advanced stages of nudivirus infection. Particularly, the decreased expression of the Sprouty (Spry) protein gene (LOC124632226) may facilitate HzNV-1 replication, similar to findings for Bombyx mori nucleopolyhedrovirus (BmNPV), where its loss of function benefited the baculovirus and increased *Bombyx mori* mortality [344]. Three upregulated genes were found in common at both 3 hpi and 24 hpi, including two major heat shock 70 kDa proteins Ba-like (LOC124642626, LOC124642627) and a putative glutathione-specific gamma-glutamylcyclo-transferase 2 (LOC124640872). These genes may play important roles in both early stress responses and later stages of infection. Additionally, two genes were significantly downregulated across three early time points (3 hpi, 9 hpi, and 24 hpi), encoding phosphatidate phosphatase LPIN2 (LOC124644440) and interferon regulatory factor 2-binding protein 1 (LOC124636415), suggesting their consistent suppression is crucial for infection progression.

Fig 4. Overview of up- and downregulated host genes at different time points during HzNV-1 infection. **(A)** Changes in the number of up- (orange) and downregulated (blue) genes over the course of HzNV-1 infection summarised in a stacked column chart. **(B)** Venn diagram displaying the overlap of DEGs at 3, 6, 9, 12, and 24 hpi, showing distinct and shared gene expression profiles within and across the two individual experiments. The colour scale signifies the number of time point-overlapping DEGs. The colour scale represents the number of DEGs, with darker shades indicating a higher DEG count in the respective section. **See next page.**



The ratio between the number of different downregulated genes (283) and upregulated genes (287) only showed a slight difference over the course of global HzNV-1 infection. However, the log2FC of upregulated genes was of greater magnitude (highest measured log2FC = 6.84) compared to those of downregulated genes (lowest measured $\log_2$FC = -1.83) **(Table 1)**.

Table 1. Overview of the fifteen most upregulated and fifteen most downregulated DEGs with predicted protein functions considering all infection time points. These DEGs are listed from highest to lowest FC value. Of these DEGs, some genes may be differently expressed at multiple time points. In this case, the time point at which they are more strongly regulated is taken as the reference and the second time point is marked with an asterisk (*). **See next page.**

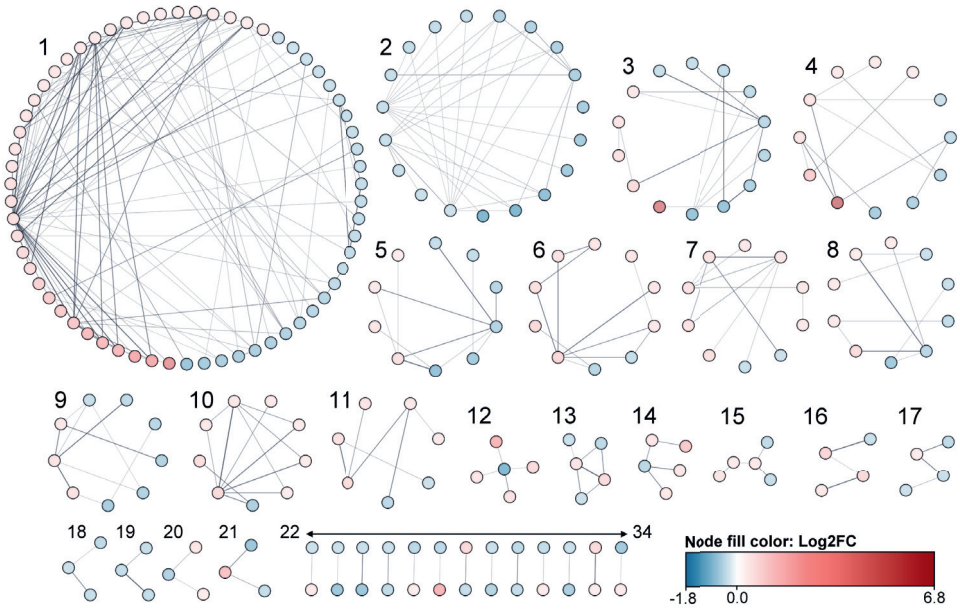
NCBI Gene Symbol	Predicted function	Time point	log2FC	Adjusted p-value
Upregulated genes				
LOC124642626	major heat shock 70 kDa protein Ba-like	24 hpi 3 hpi*	6.84 2.29*	2.10E-11 3.32E-02*
LOC124642627	major heat shock 70 kDa protein Ba-like	24 hpi 3 hpi*	6.65 2.43*	1.33E-11 3.33E-02*
LOC124630527	zinc finger protein 723-like	24 hpi	6.61	2.69E-04
LOC124643760	extracellular serine/threonine protein kinase four-jointed	9 hpi	6.25	1.61E-02
LOC124642339	F-box/LRR-repeat protein 14	9 hpi	5.47	1.50E-02
LOC124642779	lethal(2)essential for life-like (HSP)	24 hpi	4.65	3.95E-07
LOC124642484	lethal(2)essential for life-like (HSP)	24 hpi	4.57	1.53E-06
LOC124630722	heat shock protein 68-like	24 hpi	3.29	1.83E-12
LOC124633640	lethal(2)essential for life-like (HSP)	24 hpi	3.29	2.13E-13
LOC124642759	lethal(2)essential for life-like (HSP)	24 hpi	2.90	6.02E-12
LOC124645234	zinc finger protein 260-like	9 hpi	2.64	2.16E-02
LOC124642777	lethal(2)essential for life-like (HSP)	24 hpi	2.62	6.09E-10
LOC124634467	lachesin-like	24 hpi	2.52	4.10E-05

LOC124642780	lethal(2)essential for life-like (HSP)	24hpi	2.26	1.39E-10
LOC124640580	fibroblast growth factor receptor homologue 1-like	9 hpi	2.05	2.40E-02
	Downregulated genes			
LOC124629602	phosphoinositide 3-kinase adapter protein 1	24 hpi	-1.23	4.52E-05
LOC124645702	laminin subunit alpha-2-like	9 hpi	-1.24	2.86E-02
LOC124643685	E-26-specific DNA-binding protein *pokkuri*	24 hpi	-1.26	5.88E-08
LOC124644179	histone H1-like	9 hpi	-1.30	2.18E-02
LOC124644180	histone H1-like	9 hpi	-1.31	2.13E-02
LOC124645602	histone H1-like	9 hpi	-1.34	2.41E-02
LOC124645603	histone H1-like	9 hpi	-1.36	2.21E-02
LOC124645606	histone H2A	9 hpi	-1.38	2.21E-02
LOC124632937	mid1-interacting protein 1A	9 hpi	-1.44	3.27E-02
LOC124638833	probable ribonuclease ZC3H12B	9 hpi	-1.45	2.97E-02
LOC124637172	protein giant-lens	9 hpi	-1.58	1.50E-02
LOC124645608	histone H4	9 hpi	-1.60	4.90E-02
LOC124645605	histone H3	9 hpi	-1.63	2.21E-02
LOC124633107	ecdysone oxidase-like	24 hpi	-1.71	4.67E-02
LOC124632048	von Willebrand factor D and EGF domain-containing protein	9 hpi 3 hpi*	-1.72 -1.35	1.50E-02 3.32E-02

Notably, among the most upregulated DEGs with available protein annotations (**Table 1**), four genes were identified early in infection, *i.e*. 9 hpi. These early DEGs encode an extracellular serine/threonine protein kinase, an F-box/LRR-repeat protein, a zinc finger protein, and a fibroblast growth factor receptor. By 24 hpi, the remaining most upregulated DEGs predominantly encode several HSPs and an additional zinc finger protein. The most downregulated DEGs during viral infection were detected primarily at 9 hpi and included genes encoding laminin subunit alpha-2-like, most histone proteins (H1, H2A, H3, H4), mid1-interacting protein 1A, probable ribonuclease, protein giant-lens, and a protein containing the von Willebrand factor D and EGF domain. Three more downregulated DEGs encoding a phosphoinositide 3-kinase adapter protein, E-26-specific DNA-binding protein (*pokkuri*), and an ecdysone oxidase were identified by 24 hpi. To gain deeper insights into these host gene expression changes, the next section will explore the functional grouping and potential interactions of all identified DEGs, underscoring their importance in cellular pathways, and molecular processes, and suggesting biological functions impacted by the HzNV-1 infection.

Expression of host genes associated with cellular and metabolic pathways is modulated upon HzNV-1 infection

All the differentially expressed host genes upon HzNV-1 challenge were subjected to a functional enrichment analysis using the web-based STRING database. Database-available interactions among DEGs were determined with a significant protein-protein interaction (PPI) enrichment p-value of 0.0232 for the global PPI network (**Figure S3A in Supplementary data**). Based on the global STRING network, the most significant GO, KEGG and Local Network Cluster terms relate to protein processing and folding, and nuclear integrity. Other enriched terms relate to the transcriptional machinery and DNA damage response (DDR) (**Figure S3B in Supplementary data**). The MCL (Markov Cluster Algorithm)-based clustering assigned 239 of the 570 DEGs (41.93%) to clusters with direct protein interactions, however, enriched functionality terms were not available for all DEGs, mainly due to the incomplete annotated proteome of *H. zea*. The proteins with akin functionalities and direct interactions were grouped into 34 clusters, and an overview from the individually extracted cluster networks was created (**Figure 5**).



Cl #	Functional enrichment term	Cl #	Functional enrichment term
1	Protein processing and folding	18	Mixed, incl. fibroblast growth factor receptor binding, and SHC-related events triggered by IGF1R
2	Nuclear integrity and organization	19	-
3	DNA-binding and transcription factor activity, RNA polymerase II-specific	20	Guanylate cyclase complex, soluble
4	Mixed, incl. Myosin light chain binding, and stereocilium bundle	21	Peptide ligand-binding receptors
5	Ether lipid metabolism, and zymogen	22	Dystrophin-associated glycoprotein complex
6	Trehalose transport, and sugar transmembrane transporter activity	23	-
7	Lipid metabolism	24	Mitochondrial translation
8	Potassium channels	25	Negative regulation of epithelial cell differentiation, and positive regulation of photoreceptor cell differentiation
9	Nucleolus organization, and signaling by ROBO receptors	26	-
10	Glutathione metabolism	27	Apical cortex
11	Phototransduction	28	Mixed, incl. regulation of beta-cell development, and DNA binding, bending
12	Endoplasmic reticulum lumen	29	RNA polymerase III transcription initiation from type 2 promoter
13	Degradation of the extracellular matrix	30	Mixed, incl. actomyosin contractile ring, and positive regulation of peptidyl-threonine phosphorylation
14	DNA repair, and Fanconi anemia pathway	31	-
15	Developmental axis formation and signaling pathway regulation	32	Gap junction assembly, and heterophilic cell-cell adhesion via plasma membrane cell adhesion molecules
16	Nuclear lumen	33	Glutamate-cysteine ligase complex, and isoleucine catabolic process
17	Protein serine/threonine phosphatase complex	34	-

Figure 5. Clustered protein-protein interaction (PPI) networks with functional enrichment terms derived from host differentially expressed genes (DEGs) during HzNV-1 infection. MCL clustering produced 34 distinct clusters, representing 239 connected proteins, with clusters numbered by size. Node colours reflect the log2 fold change (log2FC) of each DEG, ranging from downregulated (blue) to upregulated (red), as indicated by the colour bar. The thickness of grey edges between nodes corresponds to the strength of the interaction evidence. Functional enrichment terms for each cluster are listed in the accompanying table.

The inferred 34 clusters ranged in size from 63 to 2 DEGs, with Cluster 1 and Cluster 2 containing the most DEGs of all clusters. Cluster 1 is the largest cluster and comprises DEGs involved in protein

processing and folding with 57.14% of DEGs being upregulated (red) and 42.86% being downregulated (blue). Cluster 2 has solely downregulated DEGs whose encoded proteins relate to nuclear and genome integrity, including structural constituents of chromatin, nucleosome assembly and organisation, as well as DDR. The two largest protein clusters will be thoroughly discussed in the two upcoming sections. The third largest cluster (Cluster 3) comprised proteins related to the term "DNA-binding and transcription factor activity, RNA polymerase II-specific", which implies that HzNV-1 infection modulates the transcriptional machinery of the host. Similar findings been reported for the African swine fever virus (ASFV), which promotes the expression of its genes by compartmentalizing viral mRNA, ribosomes, and cellular translation factors within the virus factory [345]. The proteins of the fourth largest cluster (Cluster 4) were enriched under the functional term "myosin binding and stereocilium bundle". The differential expression of genes encoding myosin- and stereocilium-related proteins during HzNV-1 is intriguing, since those proteins ensure proper stability of the cytoskeleton [346]. Congruently, multiple studies have shown that baculoviruses manipulate host cytoskeletal components to promote their intracellular trafficking and virion assembly [347-349], so it is possible that nudiviruses and baculoviruses share this pathological mechanism. Additionally, similar to what has been observed for baculoviruses [350, 351], metabolic pathways such as glucose, lipid and amino acid metabolism were affected by HzNV-1 infection (Cluster 5, 6, 7 and 10). Regulating and hijacking host metabolite synthesis is a common viral strategy, ensuring the virus has access to a sufficient pool of resources necessary for efficient propagation and replication [352-354].

Host protein folding and processing machinery is greatly affected during advanced HzNV-1 infection

The protein-protein association networks analysis revealed the largest cluster of DEGs (Cluster 1) encoding proteins involved in protein processing and folding, including various HSPs. Two HSPs (LOC124642627, LOC124642626) from the HSP70 family showed significant upregulation at 3 hpi and 24 hpi, with log2FCs of 6.65 and 6.84, respectively. Heat shock proteins, such as HSP70, function as molecular chaperones that fold various proteins or may play a role in DNA repair [355] and therefore, increased HSP-coding gene expression is a primary defense against stressors, including pathogen infections [356, 357]. For instance, *D. melanogaster* infected with Drosophila C Virus (*Dicistroviridae*) exhibited an increased heat shock response to limit infection, while loss of an essential heat shock transcription factor made flies hypersensitive to viral infection [358]. During infection of Sf9 cells with the baculovirus AcMNPV or of *Litopenaeus vannamei* shrimp with the white spot syndrome virus (WSSV; *Nimaviridae*), upregulation of HSP70 gene homologues was also observed [359, 360]. Other upregulated DEGs associated with the heat shock response at 24 hpi in our study included genes encoding HSP 68-like, 97 kDa HSP, HSP-12.2-like, HSP 83, and activator of 90 kDa HSP ATPase homologue 1. In addition, genes encoding several members of the DnaJ/HSP40 family, such as DnaJA1, DnaJC10, and Hdj1, also showed increased expression. Overexpression of genes encoding Hdj1 can inhibit Hepatitis B virus replication in humans, while DnaJA1 and DnaJC10/ERdj5 can benefit certain viruses, enhancing viral activities [361-364]. In general, it is a recurrent characteristic of viruses to hijack the host's chaperone machinery to promote their replication [365, 366].

Additionally, a total of six variants of the lethal(2)-essential-for-life [*l(2)efl*] gene encoding putative members of the HSP20 family were significantly upregulated at 24 hpi, with log2FCs ranging from 4.79 to 25.10. In *Drosophila*, these genes have been shown to be essential for viability, and their deficiency increases mortality [367, 368]. The expression of *l(2)efl* can be induced by various stressors, including virus infections. This gene was identified as a shared DEG in honey bee batches infected with a number of viruses [369] and was upregulated during Nora virus infection in *D. melanogaster* [370].

Overexpression of *l(2)efl* in *Aedes aegypti* inhibited Dengue virus replication, while silencing stimulated virus replication [371].

In summary, the largest cluster of HzNV-1 infection-induced DEGs related to the heat shock response and the protein folding machinery, highlighting the importance of this protein cluster in innate immune response against insect virus infections and its susceptibility to hijacking by viruses [366, 372, 373].

HzNV-1 alters expression of host genes involved in chromatin and nuclear dynamics during early and advanced infection

During HzNV-1 infection, the expression of genes encoding histone proteins was significantly downregulated (Cluster 2). Histone H1 variants and histone H3 showed reduced expression at both 9 and 24 hpi. Histone H2A and H4 encoding genes were downregulated at 9 hpi, while histone H2B encoding genes were downregulated at 24 hpi. Hence, the gene regulation for histones H2A, H3, and H4, which form the nucleosomal core, and linker-histone H1 [374], were all negatively affected by HzNV-1 infection.

In eukaryotes, histones are crucial for DNA condensation into chromatin for transcription and replication [375], and histones can fulfil roles in cell signalling, innate immunity, and antimicrobial activities [374, 376, 377]. Similarly, other members of the *Naldaviricetes (Lefavirales + Nimaviridae)* have been shown to downregulate histone gene expression, presumably to disrupt host transcription and facilitate virus replication. For instance, WSSV and Spodoptera litura nucleopolyhedrovirus (SpltNPV) infections were shown to reduce core histone levels (transcriptomic level, WSSV; proteomic level, SpltNPV) in shrimp and Sf21 cells, respectively [378, 379]. Moreover, AcMNPV infection decreased levels of transcripts encoding H2A, H3, and H4 in Sf cells [380]. Our findings further revealed a downregulation of two genes encoding histone-lysine methyltransferases (LOC124638756, LOC124641771) and the gene of the zinc finger protein Gfi-1 (LOC124646193). Gfi-1 is a conserved transcriptional repressor whose deletion in mice resulted in decreased histone H2B levels [381, 382]. Although little is known about the function of such enzymes in insects, the gene function might be similar. Thus, the downregulation of host histone-encoding genes during a HzNV-1 challenge is consistent with observations made for other naldaviruses.

Nevertheless, the differential expression of genes encoding proteins associated with nuclear integrity and organization indicates that nudivirus-induced processes may influence host chromatin dynamics and DNA packaging or even degrade the host cell nuclear lamina. The latter would align with observations made in granuloviruses (*Baculoviridae*) [383, 384]. Chromatin, a DNA-protein complex condensed by histones [385], is essential for maintaining the integrity of the nuclear membrane through its interaction with lamins [386]. Consequently, the virus-induced downregulation of histone-encoding genes may severely impair the structural organisation of the nuclear lamina, eventually leading to nuclear disintegration. The expected consequence of such a process is clearly visualised in electron microscopy (EM) images of an HzNV-1 infected HZ-AM1 cells at 60 hpi, showing that the nucleus and cytoplasm of the cell are no longer distinguishable from one another (**Figure 6**).

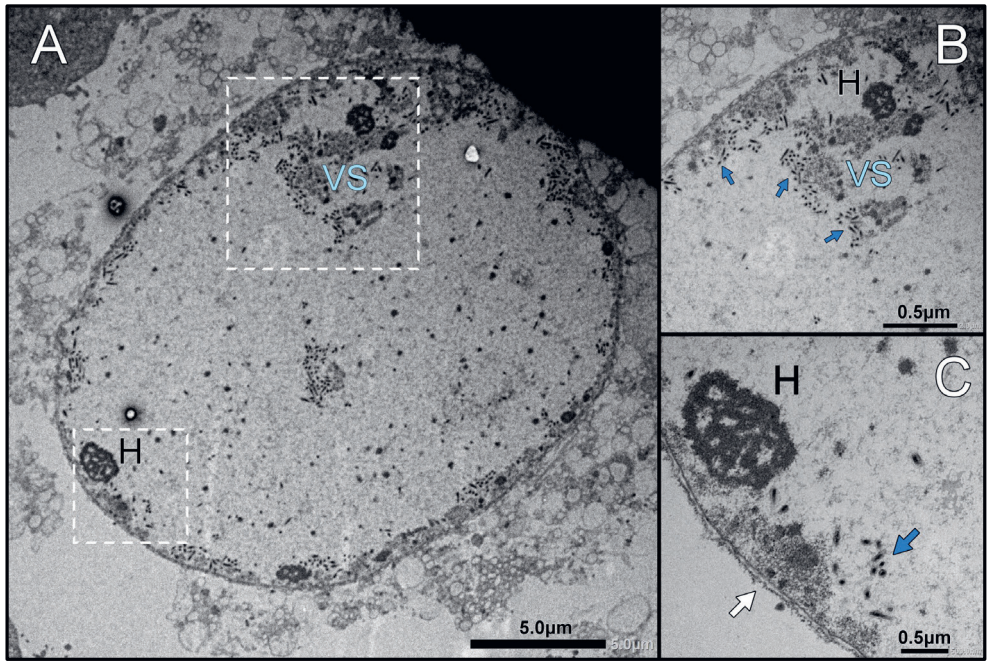


Figure 6. Electron microscopy (EM) image of sectioned HZ-AM1 cell infected with HzNV-1 after 60 hours of infection. **(A)** Whole view of an infected HZ-AM1 cell with indistinguishable nucleus and cytoplasm as well as electron dense host chromatin (H). Bar, 5 µm. **(B)** Magnified region inside the right dashed box. Clusters of intracellular HzNV-1 virions (blue arrows) in proximity to the less-electron-dense virogenic stroma (VS). Bar, 0.5 µm. **(C)** Magnified region inside the left dashed box. Extracellular host cell membrane (white arrow) with HzNV-1 virions (blue arrows). Bar, 0.5 µm. EM images courtesy Jan W. M. van Lent, Wageningen Electron Microscopy Centre (WEMC).

The disintegration of the host nuclear membrane by viruses with intranuclear DNA replication has been documented in multiple studies [387-391], but it has not yet been observed for a nudivirus, except in the case of bracovirus particle release [79, 392]. Our findings support the occurrence of this cytopathological process during HzNV-1 infection, similar to observations for related granuloviruses (*Baculoviridae*).

In this context, molecular mimicry is a common strategy used by viruses to interfere with host-specific cellular functions [393]. Notably, lepidopteran nudiviruses are currently the only known exogenous nudiviruses whose genomes encode a viral “histone mimic” protein (ORF1) consisting of a long N-terminal tail and two histone-fold domains (InterPro ID: IPR009072). Although the role of this viral histone is not fully understood, it might play a role in disrupting the integrity of the host nucleus. Such histone mimics have been identified in the evolutionary related bracoviruses with functions in suppressing host immunity [394, 395]. In more distantly related viruses, histone mimics have also been shown to fulfil functions in host immune suppression, as well as in viral genome condensation, and interaction with the host DDR [396, 397]. However, functional studies are required to determine whether the histone mimic encoded by *orf1* can play similar roles for HzNV-1.

Conclusions

Monitoring the transcriptional changes in HZ-AM1 cells infected with HzNV-1 across five time points (3, 6, 9, 12, and 24 hours post-infection) provided comprehensive gene expression profiles for both the virus and its host. Based on our study, we infer that the tipping point at which HzNV-1 finally takes over its host's cellular machinery and reprograms the host cell into a viral factory, occurs somewhere between 12 and 24 hpi. This observation is based on the drastic increase in detected DEGs at 24 hpi compared to the number of DEGs at the previously measured time points. Next to the total number of DEGs, we observed "zigzag"-like fluctuations in the quantities of DEGs between 3 and 12 hpi, suggesting that the number of DEGs might be a quantifiable indicator of changing defence amplitudes during insect-virus interactions.

We clustered the 154 HzNV-1 genes into four temporal classes, showing that genes linked to virus transcription and replication were mostly associated with the two earliest phases (Phase 1 and 2) of infection, while the phases of advanced infection (Phase 3 and 4) mostly harboured genes involved in virion assembly and maturation. Additionally, we identified a putative new promoter motif in the genome of HzNV-1 that was predominantly associated with early expressed genes, mainly involved in transcription and replication. During HzNV-1 infection, 570 DEGs of the host were identified, with notable association to protein processing, nuclear and cytoskeleton integrity, as well as metabolic pathways associated with glucose, lipid and amino acid metabolism. Significant upregulation of certain heat shock proteins and downregulation of histones were observed, indicating disruption of the host cellular machinery to facilitate viral replication and impair host defence mechanisms.

The lepidopteran betanudivirus, HzNV-1, is a particular example of a nudivirus, given its cell culture restricted pathogenicity, possession of a histone-like protein, tropism for reproductive organs, and extraordinary genome dimensions compared to other members of the *Nudiviridae*. Hence, comparisons of our results to nudiviruses of the other genera should be carefully contemplated. Our *in vitro* system does provide a detailed examination of HzNV-1 pathogenesis in an ovarian cell-derived cell line of *H. zea*, but is also limited compared to *in vivo* studies with HzNV-2. On the other hand, the cell line infection allowed for a much more synchronised infection, facilitating the discrimination of particular phases in the viral invasion (**Figure 2**).

Data availability

The raw reads of the RNA-seq experiment are reposited in the Sequence Read Archives (SRA) under the Bioproject number PRJNA1177232 (accessions from SRX26483606 to SRX26483629).

Acknowledgements

We are grateful for the excellent technical assistance from Els Roode, and would like to thank Professor Yueh-Lung Wu (National Taiwan University) for providing us with the HzNV-1 isolate that made this study possible. We also extend our gratitude to Gordon K. Smyth (University of Melbourne) for his invaluable expertise and support with the limma analysis, and to Zhenguo Lin (Saint Louis University) for insightful discussions on the appropriate use of the TSSr tool. We would like to thank Jan W. M. van Lent (Wageningen University & Research) and the Wageningen Electron Microscopy Centre (WEMC) for taking the electron microscopy images of the virus-infected HZ-AM1 cells. The work described in this paper has received funding from the European Union's Horizon 2020 research and innovation programme INSECT DOCTORS under the Marie Skłodowska-Curie grant agreement No 859850.

Supplementary data

Table S1. Short reads information of each sample with percentages of respective Hisat2 alignments to the host (*Helicoverpa zea*) or virus (Heliothis zea nudivirus 1) genome. The percentages of reads mapped to the host (*H. zea*) and the virus (HzNV-1) represent the proportion of reads that aligned to each genome, relative to the total number of reads after trimming. I = virus-infected, C = mock-infected

Sample	Cond.	No. of reads	No. of reads after trim	No. of mapped reads to *H. zea* genome	Reads aligning to *H. zea* genome (%)	No. of unmapped reads after alignment to *H. zea* genome	No. of unmapped reads aligning to HzNV-1 genome	Unmapped reads aligning to HzNV-1 genome (%)
0hpi_B1	C	59,524,704	58,269,522	53,438,512	91.709	4,831,010	495	0.001
0hpi_B2	C	59,347,242	58,260,440	52,934,136	90.858	5,326,304	889	0.002
0hpi_B3	C	56,381,136	55,223,040	50,000,277	90.542	5,222,763	723	0.001
3hpi_B1	I	58,093,300	54,919,506	48,372,612	88.079	6,546,894	16,453	0.030
3hpi_B2	I	56,188,186	54,821,348	49,576,424	90.433	5,244,924	19,973	0.036
3hpi_B3	I	54,225,532	51,967,510	46,065,199	88.642	5,902,311	21,743	0.042
6hpi_B1	I	51,540,092	50,249,346	44,686,300	88.929	5,563,046	592,350	1.179
6hpi_B2	I	56,715,806	54,677,330	47,644,315	87.137	7,033,015	700,530	1.281
6hpi_B3	I	59,405,502	58,072,560	51,573,070	88.808	6,499,490	772,297	1.330
9hpi_B1	I	54,594,030	53,009,374	43,128,955	81.361	9,880,419	4,576,232	8.633
9hpi_B2	I	57,145,186	55,945,306	46,294,618	82.750	9,650,688	4,279,422	7.649
9hpi_B3	I	58,199,552	56,793,144	46,388,701	81.680	10,404,443	5,163,680	9.092
12hpi_B1	C	39,392,600	38,172,802	34,416,059	90.159	3,756,743	415	0.001
12hpi_B2	C	41,923,364	40,627,852	36,358,437	89.491	4,269,415	387	0.001
12hpi_B3	C	40,808,366	39,583,558	35,808,525	90.463	3,775,033	370	0.001
12hpi_B1	I	43,416,158	41,498,866	34,446,330	83.005	7,052,536	2,579,886	6.217
12hpi_B2	I	52,134,144	50,433,336	41,937,868	83.155	8,495,468	3,185,963	6.317
12hpi_B3	I	40,473,120	38,940,882	32,211,087	82.718	6,729,795	2,439,157	6.264
24hpi_B1	C	73,008,924	71,611,450	64,547,053	90.135	7,064,397	917	0.001
24hpi_B2	C	48,128,340	47,468,798	42,919,072	90.415	4,549,726	567	0.001
24hpi_B3	C	39,487,824	38,502,658	34,974,237	90.836	3,528,421	445	0.001
24hpi_B1	I	40,951,296	39,525,334	27,118,719	68.611	12,406,615	9,003,317	22.779
24hpi_B2	I	52,434,608	51,439,572	36,356,744	70.679	15,082,828	11,469,085	22.296
24hpi_B3	I	43,149,140	41,752670	28,471,398	68.191	13,281,272	9,537,254	22.842

Table S2. Genes in the genome of HzNV-1 with the ATA[G/C]G[G/C]TAT motif. Motifs written in capital letters are on the negative strand. Framed motifs represent motifs in the upstream regions of two adjacent genes.

ORF	Annotation	Motif sequence	Location in HzNV-1 genome	P-value	Phase	Gene function
28		atacggtat	34,377 - 34,386	1.23E-05	2	
29		atacgctat	36,876 - 36,885	6.17E-06	3	
30	GbNV_gp19-like	ATACGCTAT	36,876 - 36,885	6.17E-06	3	Morphogenesis
33	ac81	atacgctat	43,089 - 43,098	6.17E-06	3	Morphogenesis
34		ATACGCTAT	43,089 - 43,098	6.17E-06	2	
49		ATACGCTAT	66,886 - 66,895	6.17E-06	1	
50		ATACGGTAT	67,776 - 67,784	1.23E-05	2	
51	tk1	atacgctat	70,062 - 70,071	6.17E-06	1	Nucleotide metabolism
52	PmNV_orf66	ATACGCTAT	70,062 - 70,071	6.17E-06	1	
59		atacggtat	84,369 - 84,378	1.23E-05	1	
60	helicase-2	ATACGGTAT	84,369 - 84,378	1.23E-05	2	DNA replication & processing
63		atacgctat	93,013 - 93,022	6.17E-06	3	
64	p51	ataggctat	94,476 - 94,485	1.85E-05	2	
65		ataggctat	94,476 - 94,485	1.85E-05	2	
66		ATAGGCTAT	94,476 - 94,485	1.85E-05	3	
69	dUTPase	atacggtat	99,252 - 99,261	1.23E-05	1	Nucleotide metabolism
71	DNK	ATACGCTAT	105,076 - 105,085	6.17E-06	1	Nucleotide metabolism
73	rr2	ataggctat	109,018 - 109,027	1.85E-05	2	Nucleotide metabolism
95	rr1	atacggtat	137,821 - 137,830	1.23E-05	2	Nucleotide metabolism
96		ATACGGTAT	139,681 - 139,690	1.23E-05	1	
97		atacggtat	139,681 - 139,690	1.23E-05	1	
100	RNHL	atagggtat	146,991 - 147,000	6.17E-06	3	Nucleotide metabolism
101	lef-5	ATAGGGTAT	146,991 - 147,000	6.17E-06	1	Transcription
106		ATACGCTAT	157,835 - 157,844	6.17E-06	2	
111	tk2	atagggtat	165,519 - 165,528	2.47E-05	2	Nucleotide metabolism
112	ZZZ	ATAGGGTAT	165,519 - 165,528	2.47E-05	1	
119	HgNV_orf33	atacggtat	173,507 - 173,516	1.23E-05	2	
126		ATACGCTAT	184,686 - 184,695	6.17E-06	2	
130		atacgctat	190,405 - 190,414	6.17E-06	1	
131	DNApol	ATACGCTAT	190,405 - 190,414	6.17E-06	1	DNA replication & processing
134		ATAGGCTAT	197,713 - 197,722	1.85E-05	1	
138	IAP (orf138)	ATAGGTTAT	204,513 - 204,522	2.14E-05	1	Host remodeling
142	p6.9	ATACGGTAT	210,227 - 210,236	1.23E-05	2	Morphogenesis
150		ATACGCTAT	217,773 - 217,782	6.17E-06	1	

Visualisation of differently expressed genes (DEGs) with volcano pots

Volcano plots showing the time point-specific distribution of up- and downregulated host genes during virus infection were inferred using the R package Glimma v2.4.0 [398].

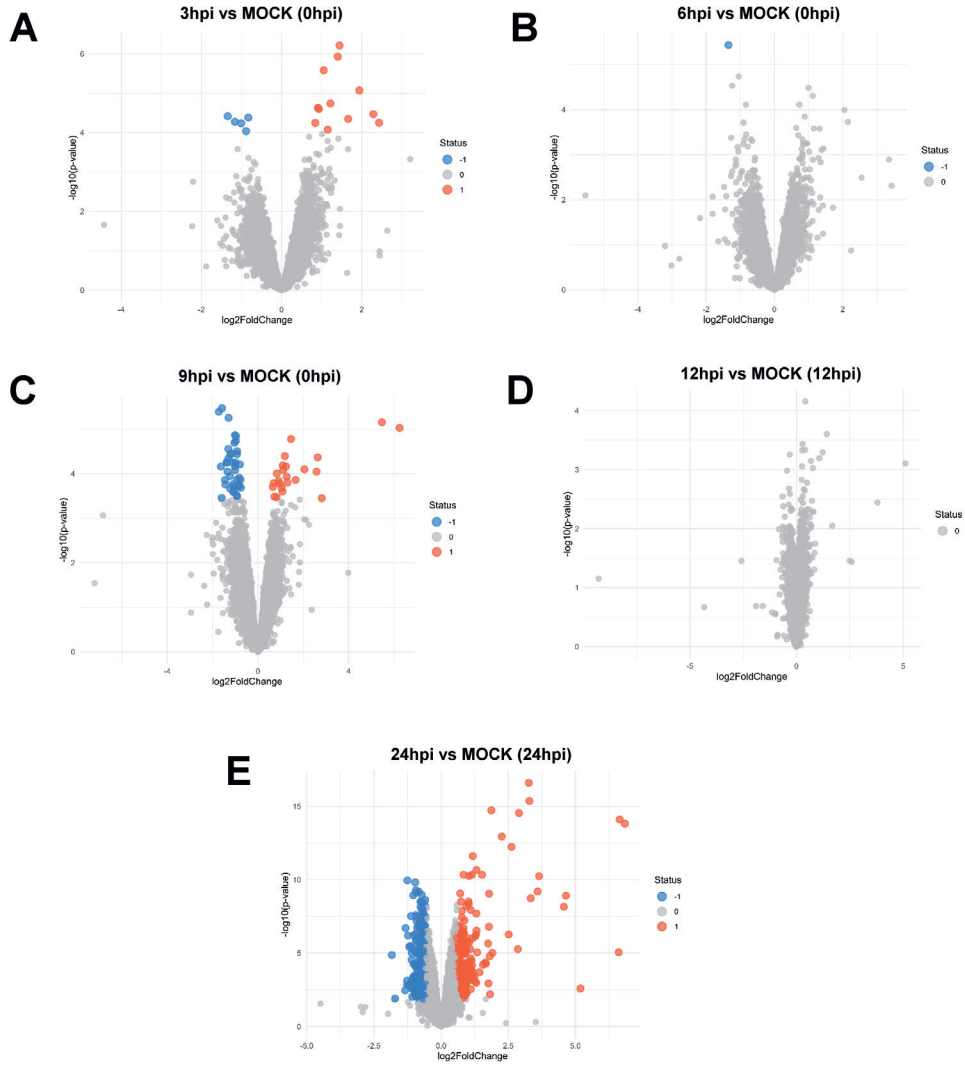


Figure S1. (**A – C**) Volcano plots showing the time point-specific distributions of up- (orange dots) and downregulated (blue dots) host genes, and their expression changes at 3, 6 and 9 hpi in comparison to mock-infected cells (0 hpi). (**D – E**) Volcano plots showing the distributions of up- and downregulated host genes in infected cells at 12 hpi and 24 hpi, and their expression changes in comparison to mock-infected cells (12 hpi or 24 hpi, respectively). In the volcano plots, the y-axis shows the negative log10(p-value), while the x-axis displays the log2-fold changes (log2FC). The differing significance cutoffs in the volcano plots result from limma's procedure to assign higher p-value cutoffs to contrasts with many DEGs, and lower p-value cutoffs to contrasts with fewer DEGs.

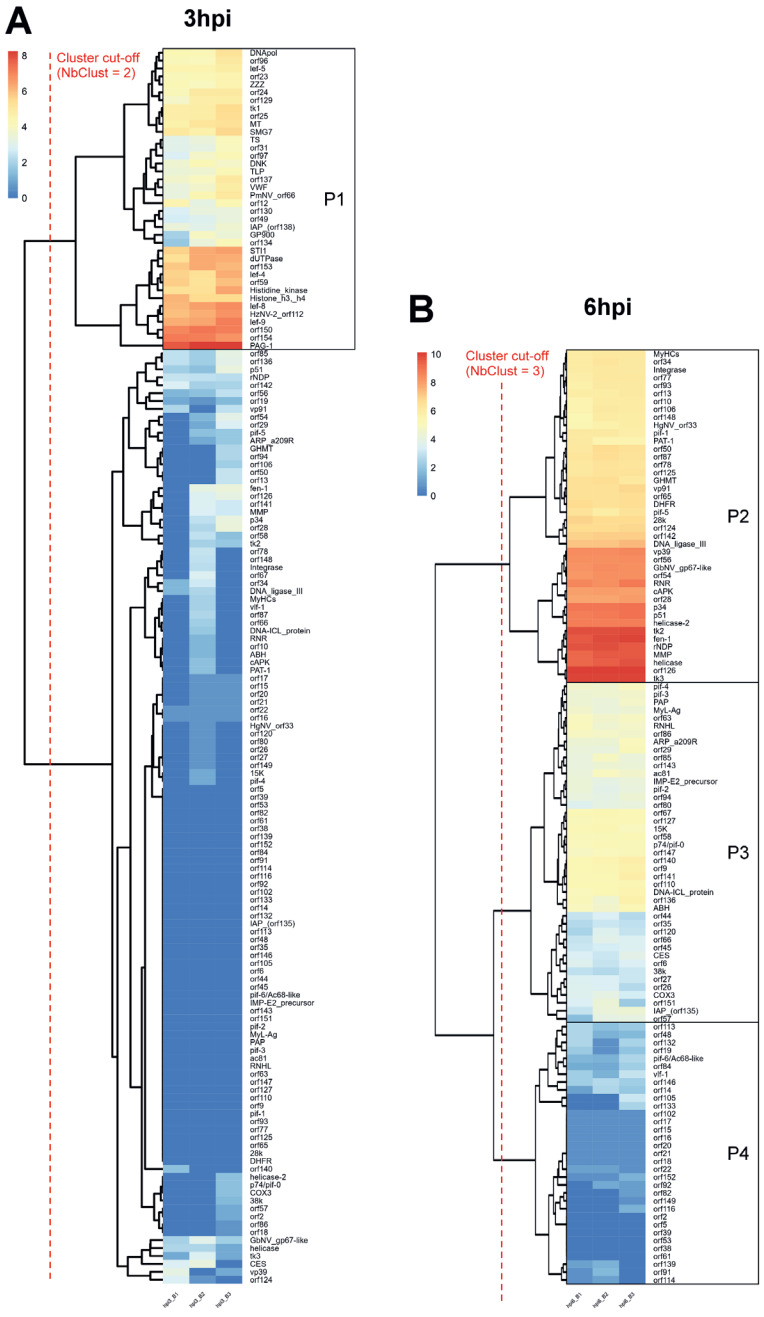


Figure S2. Hierarchical clustering of normalised HzNV-1 gene counts from infection time points **(A)** 3 hpi and **(B)** 6 hpi heatmaps showing the expression levels of the HzNV-1 genes in the two earliest stages of infection. The optimal numbers of clusters (red dotted line) for both individual heatmaps, and the temporal classes of genes were accordingly assigned.

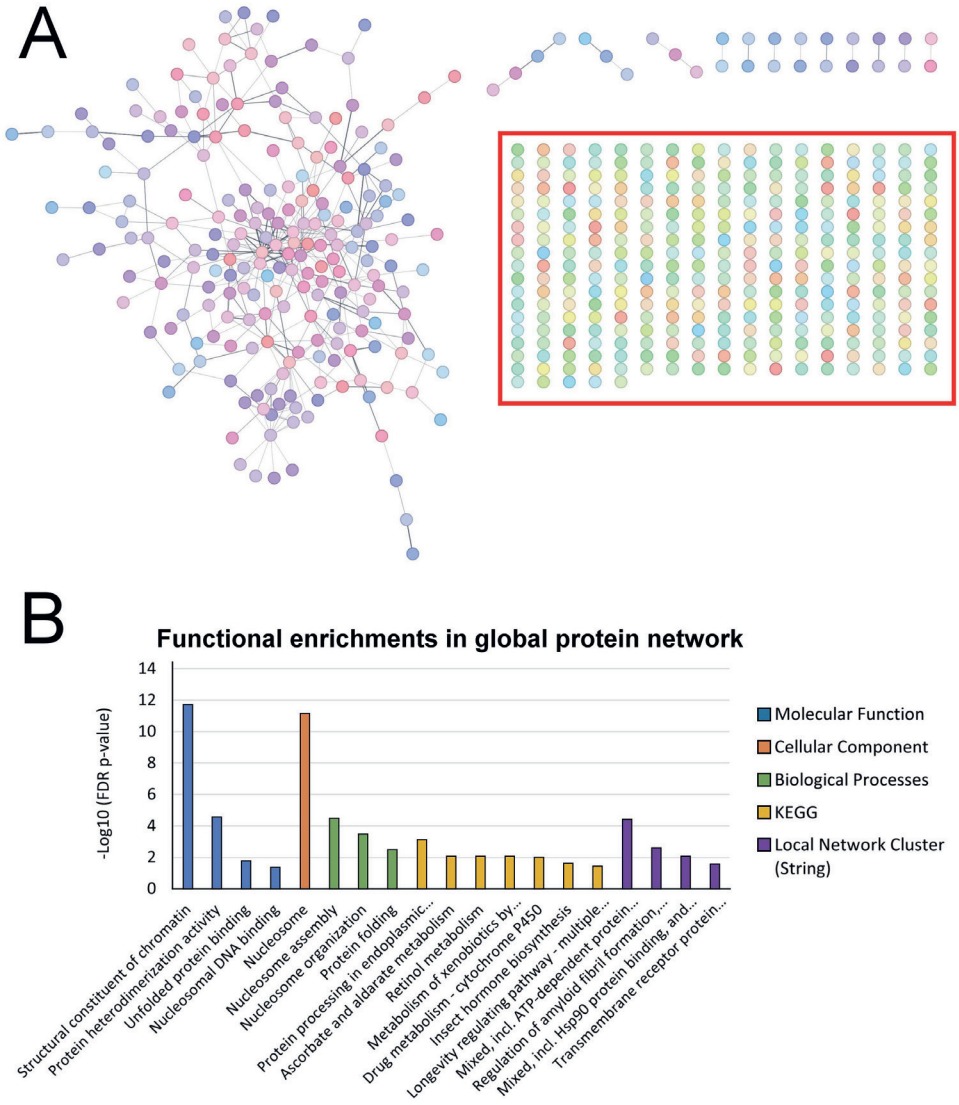


Figure S3. Protein-protein interaction (PPI) network analysis of all differently expressed genes (DEGs) during HzNV-1 infection. (A) Generated network before clustering, showing the global PPIs. Functional enrichments among nodes are colour-coded, and the thickness of the node-connecting lines represent the confidence of these interaction. The nodes in the red frame showed no supported interactions with any of the other proteins. The PPI was generated using the STRING database [301] and the supporting software Cytoscape [302]. (B) Functionally enriched gene annotation terms of the global protein network visualised in a bar chart against their respective negative log10 transformed false discovery rate (FDR) p-values.

“**La pire des bénédictions, la plus belle des malédictions.**”

– *Ma meilleure ennemie* by Luc Van Haver, Pomme, and Stromae (2024)

Chapter 6

La Route du RUM - PacBio sequencing-assisted localisation and identification of bracoviral replication unit motifs

Jirka Manuel Petersen[1,2], Nasir Mahmood Abbasi[2], Jean-Michel Drezen[2]

[1]Laboratory of Virology, Wageningen University and Research, 6708 PB Wageningen, the Netherlands.

[2]Institut de Recherche sur la Biologie de l'Insecte, UMR7261 CNRS - Université de Tours, 37200 Tours, France.

Abstract

Bracoviruses originate from an ancestral nudivirus (family *Nudiviridae*) integration into the genome of an ancestor wasp approximately 100 million years ago. These endogenous and mutualistic viruses have since then evolved alongside braconid wasps, playing essential roles in parasitism by suppressing host immune defences. However, the mechanisms behind their DNA replication, particularly the role of replication unit motifs (RUMs) that demarcate DNA amplification boundaries, remain to be fully understood. This study aimed to validate and identify RUMs previously predicted in *Cotesia congregata* (Microgastrinae) and identify those of *Toxoneuron nigriceps* (Cardiochilinae) using PacBio long-read sequencing, examining whether these regulatory sequences and, consequently, replication mechanisms are conserved across the bracoviruses of these distantly related wasp species. High molecular weight DNA was extracted from the bracovirus-producing tissues of both wasps, sequenced using PacBio long-read sequencing, and mapped onto the respective wasp genomes. Coverage plots were analysed to identify RUM locations and new motifs were identified using bioinformatic tools. By examining high read coverage regions in the wasp genomes, we confirmed the location of previously described RUMs in *C. congregata*, aligning with the amplification boundaries of the proviral loci. A new proviral locus, PL11, and its RUMs were elucidated, encoding a novel protein tyrosine phosphatase (PTP). In contrast, *T. nigriceps* proviral loci were flanked by DRJ-like motifs instead of RUMs, and DNA amplification did not extend beyond those DRJ-like motifs, indicating differences in DNA replication strategies between the two wasp species. No shared RUMs were found, except for conserved TA-rich regions. Our findings suggest that the evolutionary divergence between *C. congregata* and *T. nigriceps* promoted different replication strategies among their respective bracoviruses. Given that the common ancestor of Cardiochilinae dates back further than Microgastrinae, and that the mechanism of *T. nigriceps* appears simpler, we hypothesise that the bracoviruses of microgastrine wasps likely evolved by complexification of an ancestral replication mechanism conserved in Cardiochilinae.

Introduction

The clade of bracoviruses (BVs) originated from integration of a nudiviral ancestor into the genome of an ancestral wasp from the microgastroid complex roughly 100 million years ago (Mya) [114, 215]. This endogenisation events marked the start of a million years long evolutionary domestication of the nudiviral ancestor within its coevolving wasp hosts [399]. For more detailed information about bracoviruses and their biological and genomic traits, please refer to **Chapter 2**. The success of braconid wasps in parasitism, and consequently their entire life cycle, is highly dependent on the mutualistic relationship with their domesticated BVs. For instance, during the life cycle of the microgastrine wasp *C. congregata*, female wasps oviposit their eggs inside of lepidopteran hosts, while also injecting the virus particles of its domesticated BV (**Figure 1A**). These bracoviral particles contain different kinds of packaged DNA circles, encoding proteins from various virulence genes that serve to suppress the host immune response and ensure the proper development of the wasp progeny in the parasitised host. As a consequence of their domestication, bracoviruses lost their viral DNA polymerase. Thus, the amplification and processing of the bracoviral DNA circles rely strongly on the interplay between host DNA replication machinery and proteins encoded from the nudiviral cluster (**Figure 1B**). This complex interaction between wasp and viral components has been subject of numerous studies, aiming to experimentally unravel regulatory regions and mechanisms that guide bracovirus replication [82, 106, 109, 400]. While those studies provided important insights on bracovirus replication, additional evidence from complementary methods is needed for solidifying those findings. One such approach involves the identification of DNA amplification boundaries in the proviral loci (PL). The proviral regions that undergo DNA amplification are referred to as replication units (RU) and are each

demarcated by a pair of replication unit motifs (RUMs). Depending on the types of RU-flanking RUMs, bracoviral DNA can be amplified into various concatemeric intermediates, including head-to-head, tail-to-tail, or head-to-tail concatemers, as observed in two BV species [82, 109]. Although some RUM sequences and locations have been characterised, such as in *C. congregata* [106], further experimental validation is necessary to confirm exact amplification boundaries and whether they correspond precisely to the predicted RUMs (**Figure 1C**). Additionally, the conservation of these replication-associated regions across different braconid wasp species requires further investigations. Previous studies have demonstrated the conservation of RUM types between Cotesia congregata bracovirus (CcBV) and Microplitis demolitor bracovirus (MdBV) [82, 109], which belong to the same monophyletic group, the Microgastrinae subfamily of braconid wasps [401]. However, whether these similarities extend to all braconid wasps associated with bracoviruses is yet unresolved. For instance, the braconid wasp *T. nigriceps*, which belongs to the subfamily Cardiochilinae, represents a more distantly related species compared to *C. congregata* and *M. demolitor* [215], and although many studies have been performed to characterise the interactions of this parasitoid with its host, no scientific findings on the RUs and RUMs of its symbiotic virus, Toxoneuron nigriceps bracovirus (TnBV), are reported. Such data could provide valuable insights into whether DNA amplification mechanisms are conserved across more distantly related BVs.

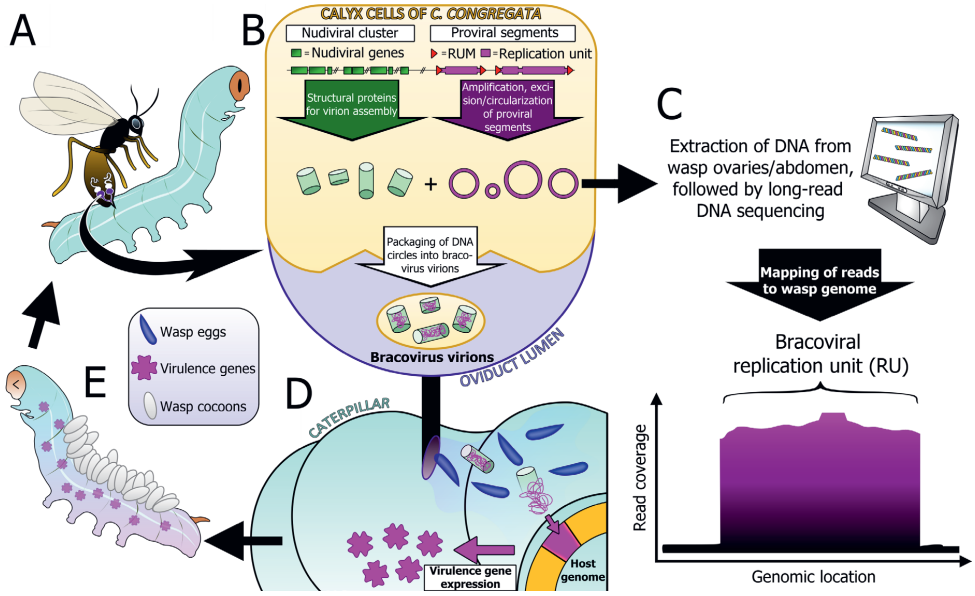


Figure 1. Schematic representation of the life cycle of Cotesia congregata and the role of its domesticated bracovirus (BV), Cotesia congregata bracovirus (CcBV). **(A)** A female braconid wasp parasitises a lepidopteran host by injecting eggs and BV particles. **(B)** Specialised calyx cells within the wasp produce infectious virions of the domesticated BV. The production of these virions involves two key genomic compartments: nudiviral genes, many of which are located in the nudiviral cluster, and an array of proviral loci (PL) [402]. The replication unit motifs (RUMs) define the boundaries of bracoviral replication units (RUs) within these PLs. RUs are double-stranded (ds)DNA molecules, but it is not known whether they are amplified linearly on the wasp chromosome or from a circular molecule that is excised from the genome prior amplification. **(C)** When DNA reads are mapped to the wasp genome, reads corresponding to the RUs appear at higher abundance relative to the wasp genome background, discernible as coverage peaks, highlighting the regions of amplified bracovirus DNA. After amplification, the RUs are resolved into dsDNA circles (“circularised”) and packaged into virions, these circles

carry various virulence genes. The expression of nudiviral genes begins in the wasp pupal stage and continues into adulthood *[82]*, producing the structural proteins essential for virions assembly. **(D)** Upon parasitizing a caterpillar, the wasp injects the replication-defective virions along with its eggs. The virions enter host cells, releasing viral DNA into the nuclei, which can integrate into the host genome [403]. The host's cellular machinery then transcribes the virulence genes, facilitating the development of wasp larvae by suppressing the host's immune response and growth [404]. **(E)** Ultimately, the caterpillar host is killed when the wasp larvae emerge to pupate, having consumed the host from within.

Hence, the aim of this study was to investigate the replication-associated regions of the BVs, CcBV and TnBV, within the genomes of *C. congregata* and *T. nigriceps* by utilizing PacBio long-read sequencing. Firstly, our main aim was to validate the locations of RUMs in the *C. congregata* genome – previously predicted by semi-quantitative PCR studies or homologies with MdBV RUMs [82, 106]. The RUMs of CcBV are currently divided into four different types based on nucleotide sequence similarity, namely Type 1 Head (T1_H_*), Type 1 Tail (T1_T_*), Type 2 Head (T2_H_*) and Type 2 Tail (T2_T_*). By extracting high molecular weight DNA from dissected ovaries, we obtained high-accuracy long reads and generated detailed coverage maps of the *C. congregata* genome. These coverage maps allowed us to correlate coverage patterns with the predicted locations of RUMs and to discover new motifs that align with these patterns. Secondly, we extended this analysis to long-read sequencing data retrieved from the abdomen of another parasitoid wasp, *T. nigriceps*, without any described RUM sequences, which were provided through a collaboration with Professor Patrizia Falabella (University of Basilicata, Italy). This comparative approach allowed us to explore whether the replication mechanisms of the respective BVs in those parasitoid wasps are similar.

Method & Materials

The insects: *Cotesia congregata* and *Manduca sexta*

For this study, a laboratory strain of *C. congregata* was reared on its lepidopteran host, the tobacco hornworm, *M. sexta* (Sphingidae; **Figure 2**). Cocoons of *C. congregata* were kindly provided and shipped to us by Karen Kester from the Virginia Commonwealth University (USA), which were collected from parasitised caterpillars. Upon arrival, the cocoons were incubated under standard conditions at room temperature in a 55 mm Petri dish until the wasps reached day 5 – a stage just before emergence when bracovirus replication activity is at its peak [400]. As soon as the first adult wasps started emerging from the cocoons, unhatched wasps were submitted to ovary dissection.

Ovary dissection from *C. congregata* pupae

Wasp cocoons were carefully opened with aid of a Leica MZ6 StereoZoom Microscope, using point tweezers and a pair of micro scissors (**Figure 3A**). The wasp pupae were removed from the opened cocoons (**Figure 3B**) and female wasps selected based on the typical morphological differences between male and female abdomens (**Figure 3C**). Several droplets of 1x PBS buffer were distributed over a 150 mm petri dish, and female pupae transferred over one by one. Each wasp abdomen was submerged with its tip into a droplet of PBS and the point tweezers used to pull out the ovaries directly into the droplet. Each dissected ovary was then immediately transferred to another 1x PBS droplet inside the cut-off cap of a 1.5 mL reaction tube chilled on ice. In total, 20 ovaries were pooled in the cooled PBS droplet and used for subsequent DNA extraction.

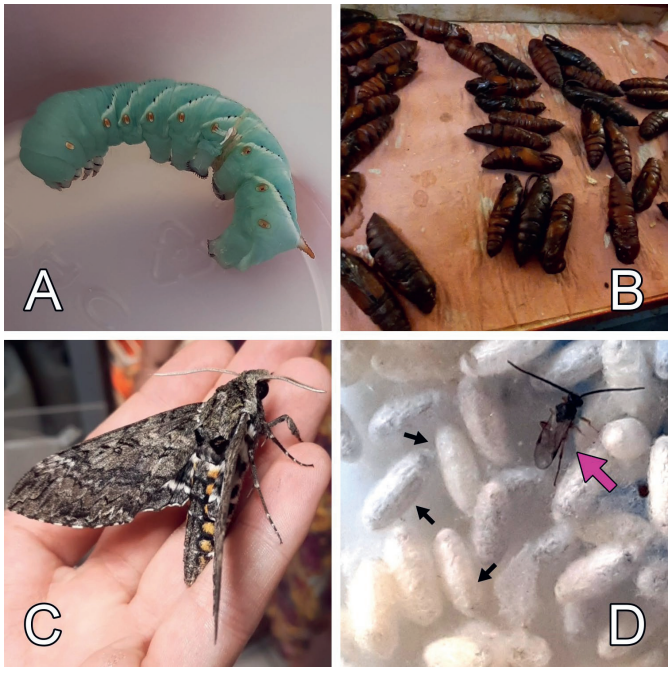


Figure 2. Overview of the host-parasitoid model involving *C. congregata*. A lepidopteran host, such as *M. sexta*, serves as the source organism for the rearing of parasitoid wasps in the laboratory. Under common circumstances, a healthy *M. sexta* larva **(A)** pupates **(B)** and emerges as an adult **(C)** from its matured pupa. When parasitised by a wasp, such as *C. congregata*, the oviposited eggs hatch inside the host and after 10 days of development the larvae will make holes through the caterpillar's cuticle and spin white cocoons on the back of the caterpillar **(D)** in which the parasitoid larvae pupate. The wasp cocoons (small black arrows) can then be picked from the outside of the caterpillar and collected, e.g. in a petri dish. Eventually, wasp adults emerge from the cocoons after 5 or 6 days (purple arrow) and the cycle can begin anew.

High molecular weight DNA extraction

The cut-off cap carrying the PBS droplet with the pooled 20 ovaries was used to close the 1.5 mL tube, and a brief spin in a tabletop mini centrifuge was performed to collect the ovaries in the PBS at the bottom of the reaction tube. High molecular weight (HMW) DNA was extracted from the ovaries using the MagAttract HMW DNA kit (Qiagen), following the manufacturer's instructions. The process began with the "Disruption/Lysis of Tissue" protocol, starting with the addition of 200 µL Buffer ATL and 20 µL proteinase K to the sample, followed by incubation at 56°C for 1 hour. After incubation, the extraction continued according to the "Manual Purification of High-Molecular-Weight Genomic DNA from Fresh or Frozen Tissue" protocol provided by the manufacturer. Extracting HMW DNA is crucial for successful long-read sequencing, as it minimises DNA fragmentation and ensures optimal DNA integrity for the PacBio sequencing method that we utilised in our study. Sufficient concentrations of HMW DNA (19.8 ng/µL) and quality were confirmed via Qubit 2.0 fluorometer measurement and agarose gel electrophoresis, respectively.

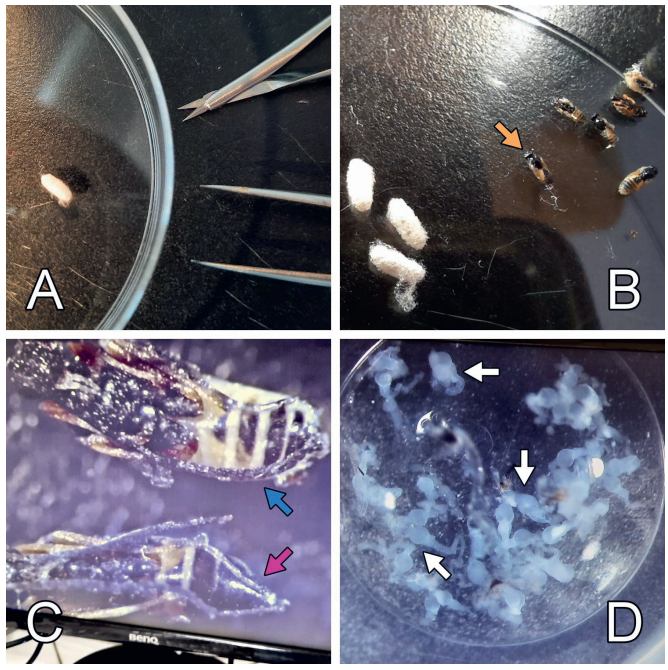


Figure 3. Photographs taken during dissection of *C. congregata* ovaries. The wasp cocoons were gently cut open using a pair of surgical scissors and point tweezers **(A)** and the wasp pupae removed **(B)**. Female wasps were distinguished from male wasps based on the morphological differences of their abdomens **(C**, blue arrow = male, purple arrow = female**)**. Dissected ovaries (white arrows) were collected and pooled together in a 1x PBS droplet on ice **(D)**.

Assembly of PacBio reads and mapping to *C. congregata* or *T. nigriceps* genome

The HMW DNA was sent for PacBio HiFi sequencing to Gentyane (Clermont-Ferrand, France). The received long-reads were *de novo* assembled using Canu [405], an assembler optimised for long reads, which bypasses correction and trimming steps due to the high accuracy of the HiFi data. BUSCO [406] was used to assess genome completeness, while QUAST [407] was applied to evaluate assembly metrics. Additional tools like Emboss-infoseq [408] and Inspector [409] were used to gather detailed sequence information. The genome was further analysed for structural variations and errors using Purge Haplotigs [410] and BlobToolKit [411] to ensure contiguous and proper assembly. Graphs generated by the abovementioned tools can be found in the Supplemental Data. After assembly, the PacBio reads were mapped back to the assembled genome using Minimap2 v2.24 (Li, 2018) for additional quality assessment. Additionally, the reads were aligned to the publicly available *C. congregata* genome (published on April 15th, 2020) from the BioInformatics Platform for Agroecosystem Arthropods (BIPAA, https://bipaa.genouest.org/). The provided long-read sequencing data from *T. nigriceps* were assembled under default settings with Hifiasm (v0.19.8), and the reads were mapped against the assembled scaffolds with Minimap2. Unmapped reads were removed from the resulting SAM file with SAMtools v1.15.1 [296]. The final coverage files for *C. congregata* and *T. nigriceps* were generated by converting the alignment map files into Browser Extensible Data (BED) format using BEDTools v2.27.1 [412]. Although the read-specific alignment information was removed through this conversion, the more compact BED files allowed for less computing draining visualisations in genome browsers such as JBrowse [413] or IGV [414].

Confirmation and identification of RUMs via coverage plot assessment and BLAST search in *C. congregata* and *T. nigriceps*

To create coverage plots of the PacBio reads mapped to the wasp genomes, we imported the relevant BED and General Feature Format (GFF) files containing read mappings and genome annotations, respectively. The BED file provided the genomic coordinates and coverage data, while the GFF file contained annotations for regions of interest in *C. congregata*, including the coordinates of experimentally determined RUMs from Louis *et al.* (2013) and Gauthier *et al.* (2021). The nucleotide sequences of the RUMs from the literature were submitted as queries for BLASTn searches against the *C. congregata* genome to identify potential new RUMs. With aid of IGV [414] and self-generated coverage plots, we examined whether the experimentally determined RUMs and putative RUMs from the BLAST search were located in proximity to sudden read coverage increases. Using the ggcoverage v1.4.0 [415] and ggplot2 v3.5.1 [416] packages in R, filled area plots were generated to visualise the read coverage across the genomic scaffolds of interest in the genomes of *C. congregata* and *T. nigriceps*, and annotations from the GFF file served to visualise the position of proposed RUMs in the *C. congregata* genome. Since there are no available RUM sequences for *T. nigriceps*, the 200-bp up- and downstream sequences surrounding regions of sudden coverage increase were extracted and submitted to the MEME webtool [305] to identify conserved motifs under default settings. The coverage plots were generated for each relevant scaffold, with the x-axis adjusted to reflect the scaffold's genomic coordinates. The plots were then saved with high resolution for publication and modified using the photo editing software Ulead Photoimpact x3. The transcriptomic data for the coverage plot of *C. congregata*'s newly identified proviral locus PL11 were retrieved as a GFF file from BIPAA via the JBrowse webtool [413] and used to generate a coverage as described above.

Results

Nucleotide alignment of newly identified RUMs from CcBV

The nucleotide sequences of the CcBV RUMs from the literature were aligned with newly identified RUMs from our BLAST search (**Figure 4**). For PL8, we found a new RUM with high similarity to the sequences of Type 1 Tail RUMs. A recently identified proviral locus, PL10 [82], is flanked by a motif of Type 1 Head and Type 1 Tail, implying that this segment might be amplified as a head-to-tail concatemer [106]. Another flanking site (FS) with nucleotide homology to the Type 1 Head RUMs was found on scaffold_7, but no Type 1 Tail RUM. The genomic region surrounding this newly identified RUM from scaffold_7 and its correlating read coverage will be further discussed in the following results section.

According to Gauthier *et al.* (2021) [82], the segment of PL3 is also flanked by Type 1 Head and Type 1 Tail RUMs. However, based on the new insights we gained from the long-read sequencing data, we propose two new RUMs, which we will further elaborate on in the following results section. No novel motifs were found for the Type 2 Head and Type 2 Tail RUMs.

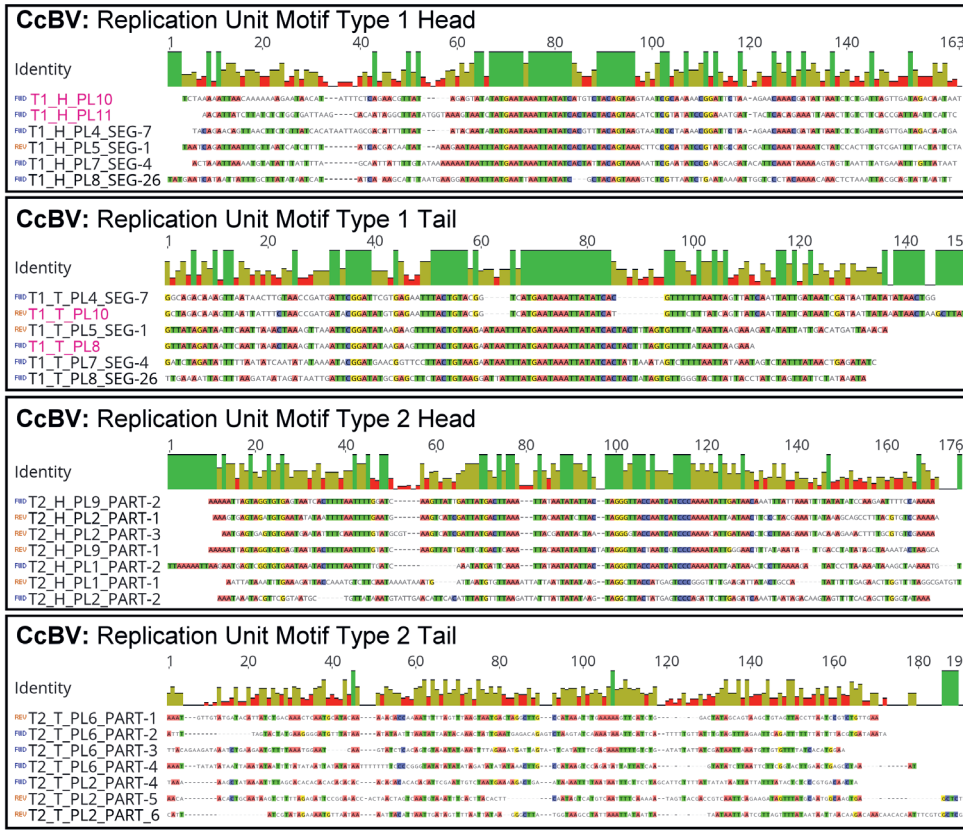


Figure 4. Nucleotide alignments of Type 1 and Type 2 replication unit motifs (RUMs) from Cotesia congregata bracovirus (CcBV). These motifs are classified into 'Head' and 'Tail' regions, reflecting their distinct roles in the amplification of bracoviral DNA as concatemers [106]. Sequences were analysed and aligned using the MAFFT alignment tool (v7.490) of Geneious (v2023.2.1), and the RUM sequences were obtained from Gauthier *et al.* (2021). Sequence names in purple indicate RUMs that have been newly identified in this study.

PacBio HiFi Long-Read Length Distribution of C. congregata and T. nigriceps

The reads generated by PacBio long-read sequencing were visualised in read length distribution plots for both parasitoid wasp species, *C. congregata* and *T. nigriceps*. Species-specific differences in the read length distributions likely arise because smaller fragments (8 kb instead of 15 kb) were selected to prepare the *T. nigriceps* sequencing library, as the DNA extracted from long-term -80°C stored insects was expected to have shorter average molecule length. Congruently, *C. congregata* reads peaked around 15,000 base pairs (bp), but also showed a secondary peak around 5,000-bp (**Figure 5A**). In contrast, *T. nigriceps* displayed a read length distribution towards shorter reads, peaking around 6,500 bp (**Figure 5B**).

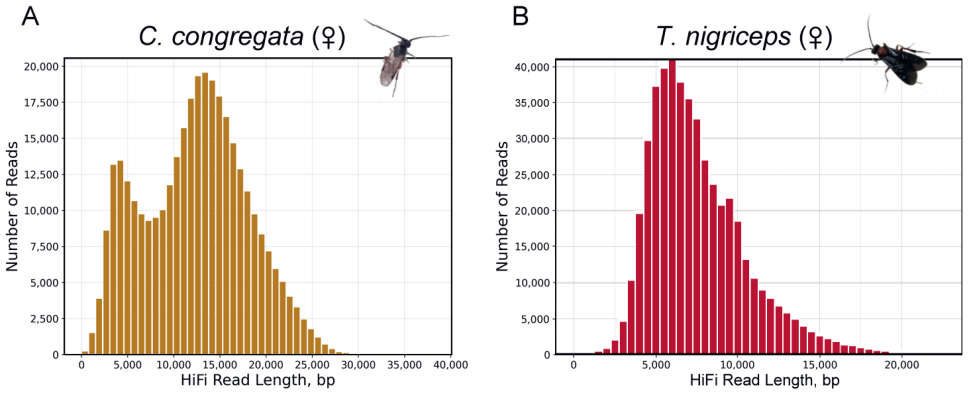


Figure 5. Read length distributions from PacBio HiFi long-read sequencing of *Cotesia congregata* and *Toxoneuron nigriceps*. **(A)** *C. congregata* sequencing reads, originating from dissected ovaries of female wasps. **(B)** *T. nigriceps* sequencing reads, derived from the whole abdomen of female wasps. The y-axis represents the number of reads, while the x-axis indicates the read length in base pairs. The photo of *T. nigriceps* was kindly provided by Professor Patrizia Falabella (Università degli Studi della Basilicata, Department of Sciences).

Long-read coverage patterns in C. congregata genome confirm proposed RUMs of PL1, PL2, PL4, PL5, PL6, PL7, PL8 and PL9

The long reads were aligned to the *C. congregata* genome (v2) retrieved from BIPAA, generating binary alignment map (BAM) files. To reduce file sizes and simplify visualisation in analysis tools such as JBrowse [413], IGV, and RStudio, these BAM files were converted to BED files, focusing on the coverage values of the mapped reads in the wasp genome. For further analysis, BED files containing only the coverage data from specific scaffolds – those with known proviral loci (scaffold_CcPL1_CcPL2_CcPL4, scaffold_CcPL3, scaffold_CcPL5_CcPL8, scaffold_CcPL6, scaffold_CcPL7, scaffold_CcPL9, scaffold_23) and scaffolds of interest based on the BLAST search for novel RUMs (scaffold_7, scaffold_333) – were used to generate individual coverage plots (**Figure 6 – 13**). The location of known RUMs from the literature as well as newly identified RUMs from a BLAST search (**Figure 4**) are indicated as arrows in the respective coverage plot to emphasise their locations to regions of coverage increase, aiming to support their involvement in BV DNA replication with long-read NGS data.

The first scaffold carries three proviral loci, namely PL1, PL2 and PL4 (**Figure 6A**) with PL2 having three separate RUs (**Figure 6B**). Based on Louis *et al*. (2015) and Gauthier *et al*. (2021), each RU of PL2 is flanked by a pair of RUMs: RU2.1 (T2_H_RU2_PART-1 and T2_H_RU2_PART-2), RU2.2 (T2_H_RU2_PART-3 and T2_T_RU2_PART-4) and RU2.3 (T2_T_RU2_PART-5 and T2_T_RU2_PART-6), respectively. The RUMs T2_H_RU2_PART-1 and T2_T_RU2_PART-6 precisely demarcated the entire high-coverage region of PL2. Similarly, the RUM pairs flanking PL1 (T2_H_RU1_PART-1 and T2_H_RU1_PART-2) and PL4 (T1_T_RU4_SEG-7 and T1_H_RU4_SEG-7) delineated their respective regions. However, the predicted regions of PL2's RUs (RU2.1, RU2.2 and RU2.3) do not seem to align with the observed coverage patterns, except for a minor drop in coverage adjacent to T2_H_RU2_PART-2 and T2_H_RU2_PART-3. Therefore, it appears that the regions flanked by T2_H_RU2_PART-2, T2_H_RU2_PART-3, T2_H_RU2_PART-4 and T2_H_RU2_PART-5 do not demarcate bracoviral RUs as previously stated by Louis *et al*. (2013) [106].

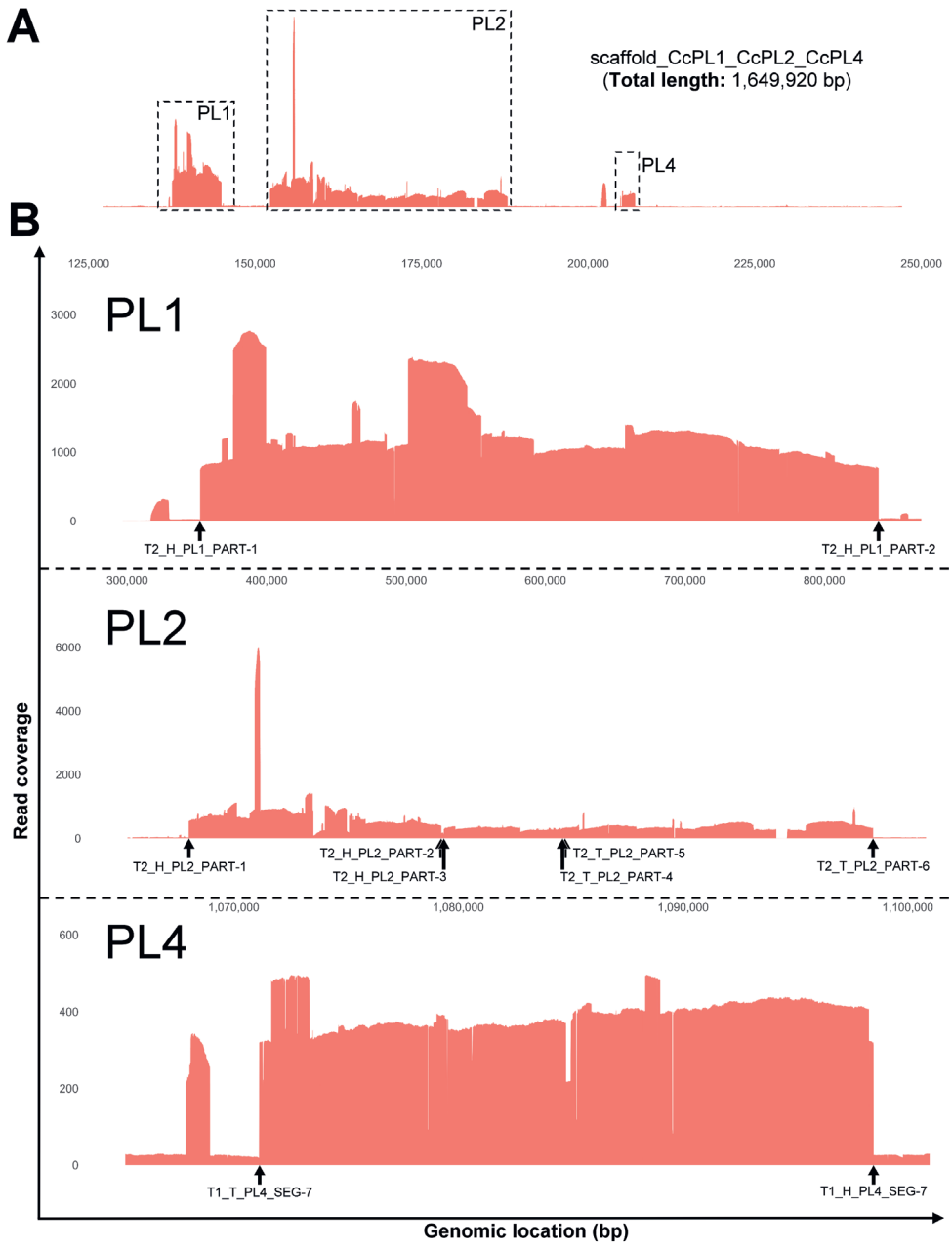


Figure 6. Filled area plot showing the PacBio read coverage across *C. congregata*'s scaffold_CcPL1_CcPL2_CcPL4. **(A)** Whole coverage plot of scaffold_CcPL1_CcPL2_CcPL4 with a total length of 1,649,920 bp, displaying the distribution of mapped PacBio reads. The boxed regions indicate the respective zoomed-in views of areas containing the proviral loci (PL1, PL2 and PL4). **(B)** Zoomed-in plots emphasizing the read coverage distribution across the respective proviral loci. The y-axis, with scales adjusted for each enlarged section, represents read coverages and the x-axis shows genomic locations in the scaffold. Black arrows mark where the RUMs from Gauthier *et al.* (2021) are located.

The coverage plot for scaffold_CcPL5_CcPL8 revealed distinct peaks corresponding to the proviral loci PL5 and PL8 (**Figure 7A**). The zoomed-in views (**Figure 7B**) demonstrated that the regions of highest coverage align precisely with the RUMs described by Gauthier *et al*. (2021) [82]. Similar to the pattern observed in scaffold_CcPL1_CcPL2_CcPL4, the RUMs flanking each RU within these loci precisely bordered the edges of their respective high-coverage areas, except for T1_H_PL8_SEG-26. Instead, it appears that PL8 is flanked by another Type 1 Tail motif (namely T1_T_PL8), in addition to T1_T_PL8_SEG-26.

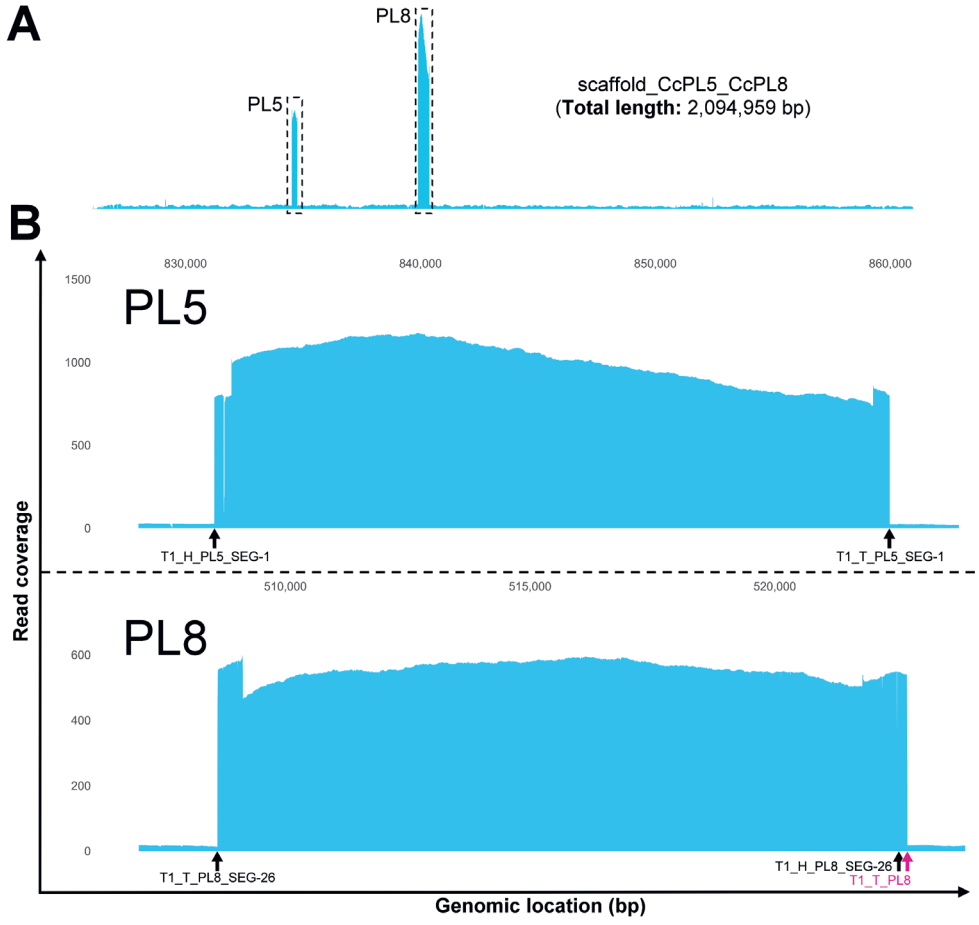


Figure 7. Filled area plot showing the PacBio read coverage across *C. congregata*'s scaffold_CcPL5_CcPL8. **(A)** Whole coverage plot of scaffold_CcPL5_CcPL8 with a total length of 2,094,959 bp, displaying the distribution of mapped PacBio reads. The boxed regions indicate the respective zoomed-in views of areas containing the proviral loci (PL5 and PL8). **(B)** Zoomed-in plots emphasizing the read coverage distribution across the respective proviral loci. The y-axis, with scales adjusted for each enlarged section, represents read coverages and the x-axis shows genomic locations in the scaffold. Purple arrows mark where newly identified RUMs are located. Black arrows mark where the RUMs from Gauthier *et al.* (2021) are located.

In scaffold_CcPL6 (**Figure 8**), the coverage peak corresponding to the proviral locus PL6 was notably lower compared to the higher peaks associated with nearby transposable elements (TE). Based on Louis *et al*. (2013), the PL6 region is split into two replication units, RU6.1 and RU6.2, flanked by specific

RUMs (T2_H_RU6_PART-1 and T2_T_RU6_PART-2 for RU6.1; T2_H_RU6_PART-3 and T2_T_RU6_PART-4 for RU6.2). Although T2_H_RU6_PART-1 and T2_H_RU6_PART-4 effectively delimit the outer rims of PL6, there were no distinguishable coverage regions correlating to RU6.1 and RU6.2, which therefore appear to constitute a single RU, RU6.

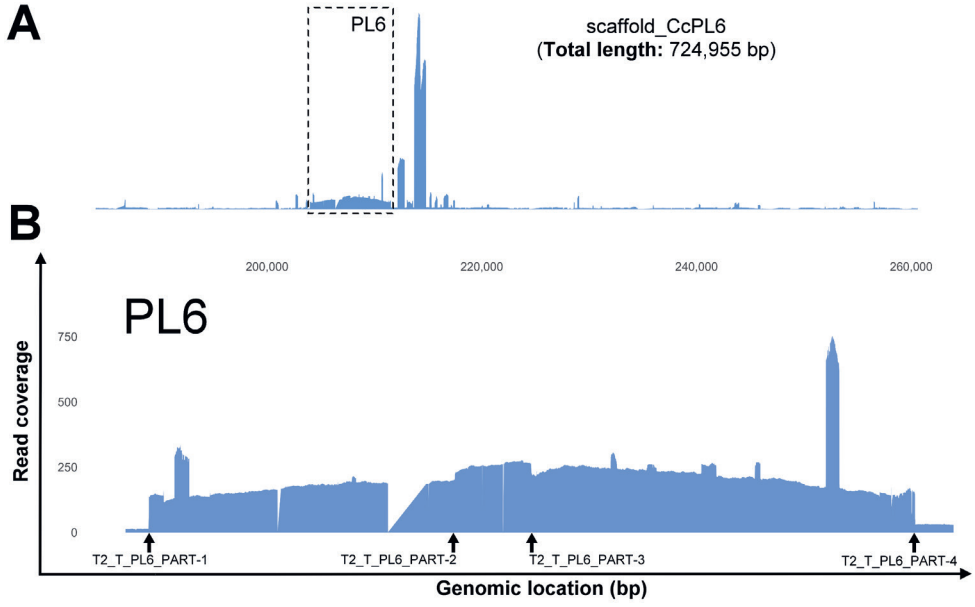


Figure 8. Filled area plot showing the PacBio read coverage across *C. congregata*'s scaffold_CcPL6. **(A)** Whole coverage plot of scaffold_CcPL6 with a total length of 724,955 bp, displaying the distribution of mapped PacBio reads. The boxed region indicates the respective zoomed-in view of the area containing the proviral locus (PL6). **(B)** Zoomed-in plots emphasizing the read coverage distribution across the respective proviral loci. The y-axis represents read coverages and the x-axis shows genomic locations in the scaffold. Black arrows mark where the RUMs from Gauthier *et al.* (2021) are located

The coverage analysis of scaffold_CcPL7 (**Figure 9A**) revealed a peak corresponding to PL7, which exhibited a similar coverage level to that of PL6. The literature-derived RUMs were located at the edges of this region with elevated read coverage (**Figure 9B**).

Also, for scaffold_CcPL9 (**Figure 10A**), an elevated read coverage was observed corresponding to the location of the PL9 locus. The zoomed-in view (**Figure 10B**) highlights how the RUMs effectively bracket the region of maximal coverage. Notably, from 1,268,987 to 1,270,904 bp a coverage decrease could be observed, however, the coverage levels are still way above the background of the wasp genome. This pattern supports the hypothesis that PL9 amplifies its RU as a single, large DNA molecule, which is subsequently processed into three distinct DNA circles.

The coverage plot for scaffold_CcPL10 (originally scaffold_23) (**Figure 11**) shows increased read coverage across the proviral locus PL10. Unlike the other proviral loci, the respective RUMs for PL10 were previously unknown and were identified through our BLASTn analysis. These newly identified RUMs, T1_H_PL10 and T1_T_PL10, are marked by purple arrows in **Figure 11B.**

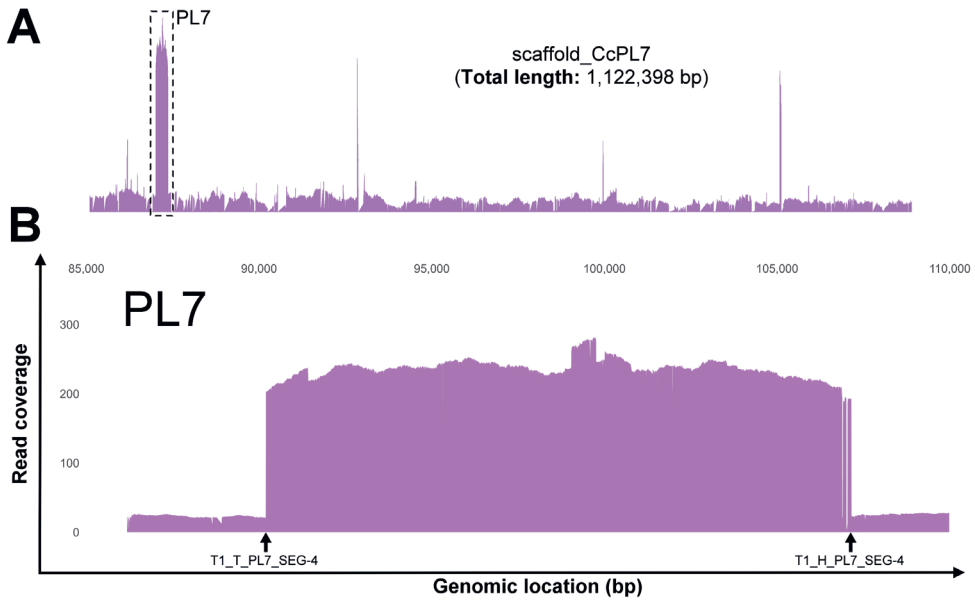


Figure 9. Filled area plot showing the PacBio read coverage across *C. congregata*'s scaffold_CcPL7. **(A)** Whole coverage plot of scaffold_CcPL7 with a total length of 1,122,398 bp, displaying the distribution of mapped PacBio reads. The boxed region indicates the respective zoomed-in view of the area containing the proviral locus (PL7). **(B)** Zoomed-in plots emphasizing the read coverage distribution across the respective proviral loci. The y-axis represents read coverages and the x-axis shows genomic locations in the scaffold. Black arrows mark where the RUMs from Gauthier *et al.* (2021) are located.

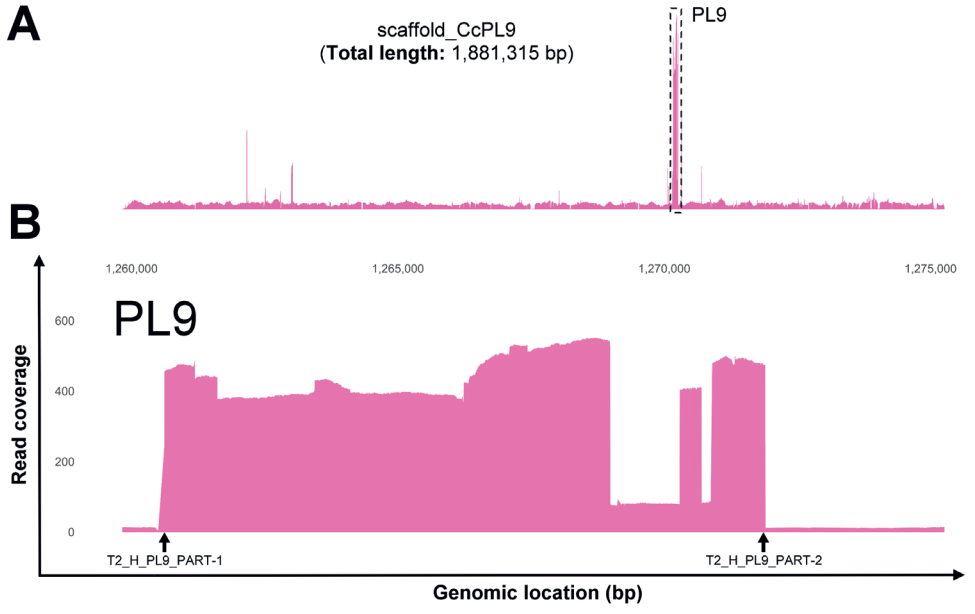


Figure 10. Filled area plot showing the PacBio read coverage across *C. congregata*'s scaffold_CcPL9. **(A)** Whole coverage plot of scaffold_CcPL9 with a total length of 1,881,315 bp, displaying the distribution of mapped PacBio

reads. The boxed region indicates the respective zoomed-in view of the area containing the proviral locus (PL9). **(B)** Zoomed-in plots emphasizing the read coverage distribution across the respective proviral loci. The y-axis represents read coverages and the x-axis shows genomic locations in the scaffold. Black arrows mark where the RUMs from Gauthier *et al.* (2021) are located.

The presence of a Head type and Tail type RUM may indicate that the RU of PL10 is amplified as a head-to-tail concatemer, similar to what has been proposed in the literature for PL4, PL5, PL7 and PL8 [82, 106]. It is noteworthy that the region from 701,968 to 704,421 bp contains a genomic locus with ambiguous nucleotides 'N', emphasising that T1_T_PL10 is adjacent to a yet unresolved region in the *C. congregata* genome.

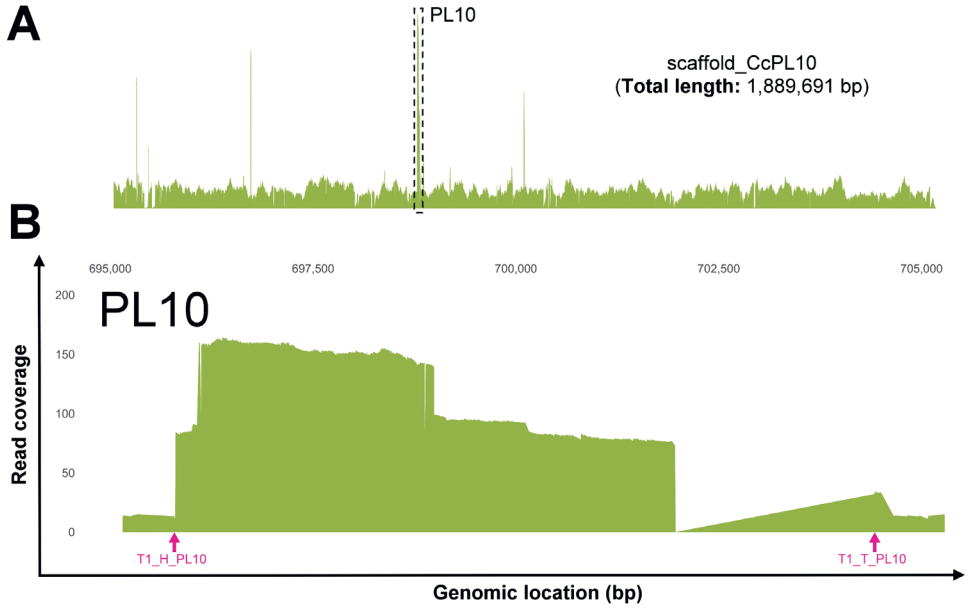


Figure 11. Filled area plot showing the PacBio read coverage across *C. congregata*'s scaffold_CcPL10 (originally scaffold_23). **(A)** Whole coverage plot of scaffold_CcPL10 with a total length of 1,889,691 bp, displaying the distribution of mapped PacBio reads. The boxed region indicates the respective zoomed-in view of the area containing the proviral locus (PL10). **(B)** Zoomed-in plots emphasizing the read coverage distribution across the respective proviral loci. The y-axis represents read coverages and the x-axis shows genomic locations in the scaffold. Purple arrows mark where the newly identified RUMs are located.

Overall, our PacBio sequencing-assisted approach supported and verified experimentally the locations and authenticity of most of the predicted RUMs from Louis *et al*. (2013) and Gauthier *et al*. (2021), except for those inside PLs (PL2 and PL6), while PL3 represents a particular case, which will be discussed in detail in the following section.

Long-read coverage patterns in Scaffold_CcPL3 of *C. congregata* indicate regulatory motifs different from the literature

The generated coverage plot for scaffold_CcPL3 indicated that the RUMs (T1_H_PL3_SEG-17 and T1_T_PL3_SEG17) specified by Gauthier *et al.* (2021) [82] do not align with the coverage patterns of this proviral segment (**Figure 12**, black arrows). Based on the regions of drastic coverage increase, we hypothesised that the FSs contain the actual regulatory motifs involved in PL3 DNA amplification. The nucleotide sequences flanking the abundant read coverage (FS_1_PL3 and FS_2_PL3) of

scaffold_CcPL3 (**Figure 12B**, purple arrows) were extracted and used as input for a BLASTn search in the *C. congregata* genome to screen for similar motifs. The BLASTn search revealed two FS that shared nearly identical sequence homology with FS_1_PL3 and FS_2_PL3. We will refer to these motifs as FS_1_PL3-like and FS_2_PL3-like, respectively (**Figure 12C**).

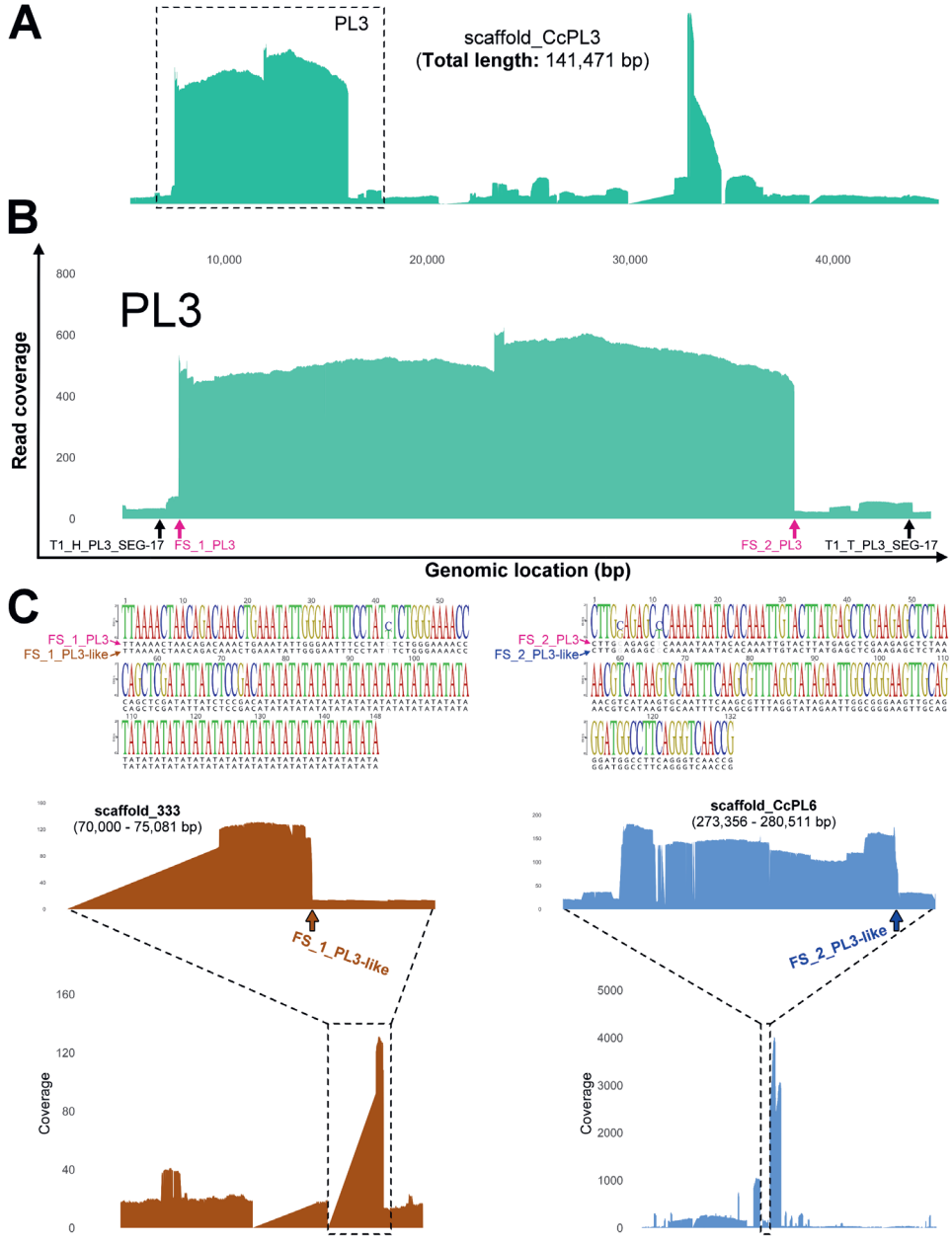


Figure 12. Filled area plot showing the PacBio read coverage across *C. congregata*'s scaffold_CcPL3. **(A)** Whole coverage plot of scaffold_CcPL3 with a total length of 141,471 bp, displaying the distribution of mapped PacBio reads. The boxed region indicates the respective zoomed-in view of the area containing the proviral locus (PL3).

(B) Zoomed-in plots emphasizing the read coverage distribution across the respective proviral loci. The y-axis represents read coverages and the x-axis shows genomic locations in the scaffold. Black arrows mark where the RUMs from Gauthier *et al.* (2021) are located, while purple arrows mark where the newly identified RUMs are located based on the sudden coverage. **(C)** Sequence alignment of PL3's newly proposed RUMs (FS_1_PL3 and FS_2_PL3) against two nearly identical motifs in proximity to sudden coverage increases (scaffold_333: FS_1_PL3-like, brown arrow; scaffold_CcPL6: FS_2_PL3-like, blue arrow), which can be seen in the coverage plots below.

On scaffold_333, we found a 148 bp region that was nearly identical to FS_1_PL3, which we designated FS_1_PL3-like. This region contained a TA-rich sequence and could not be found in any other parts of the wasp genome via BLAST search. While it has been suggested that TA-rich sequences may form hairpin structures and potentially function as origins of replication for BVs [106, 109], TA-rich regions have also been described to serve as insertion sites for some TEs [417]. Interestingly, we also found a 132 bp sequence with almost exact resemblance to FS_2_PL3 on the same scaffold as PL6, which contains multiple TEs (**Figure 8**) based on annotations of the BIPAA database. Therefore, the replication unit of PL3 is flanked by motifs with high similarity to sequences that flank regions in the wasp genome that are annotated as TEs.

A novel proviral locus (PL11) with Type 1 Head RUM in *C. congregata* encodes another bracoviral protein tyrosine phosphatase (PTP)

While searching for new RUMs in the *C. congregata* genome via BLAST, we detected a Type 1 Head-like RUM on scaffold_7, but no second RUM was identified on this scaffold. The inspection of the long-read coverage data revealed that the Type 1 Head-like RUM flanked a genomic region with increased DNA amplification, which led us to believe that scaffold_7 might harbour a potentially novel PL (**Figure 13A**). Further examination of this high-coverage region using the JBrowse webtool on BIPAA revealed an open reading frame (ORF) corresponding to an unannotated gene within that area. Submission of the amino acid sequence of the encoded protein to BLAST revealed significant similarity to bracoviral protein tyrosine phosphatases (PTPs). The expression of this PTP-encoding gene in wasp ovaries could be confirmed by inspecting transcriptomic data available on BIPPA (**Figure 13B**). The expression of the scaffold_7-asscoiated PTP genes supports that it may be packaged into bracovirus particles and benefit the success of wasp parasitism, similar to what has been observed for other proviral PTPs [403, 418]. Collectively, our findings support that scaffold_7 confines a novel PL, henceforth designated as proviral locus 11 (PL11), which encodes a newly identified bracoviral PTP (PTP_PL11) with putative importance for wasp parasitism success.

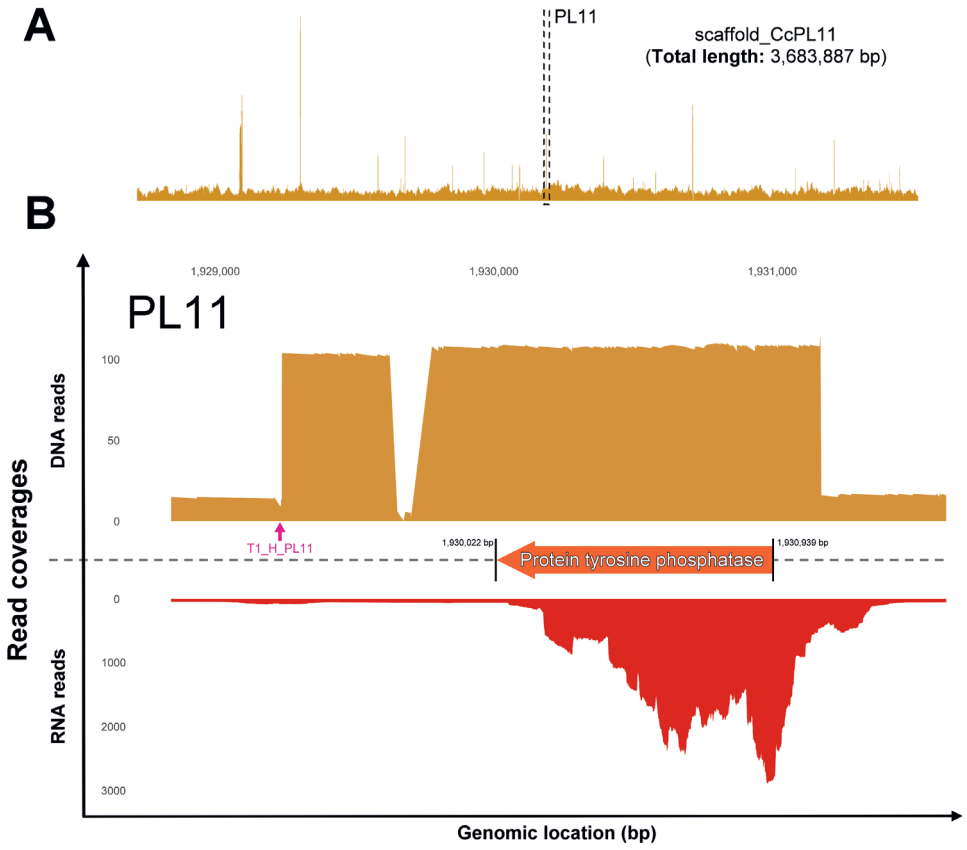


Figure 13. Filled area plot showing the PacBio read coverage across *C. congregata*'s scaffold_CcPL11 (originally scaffold_7). **(A)** Whole coverage plot of scaffold_CcPL11 with a total length of 3,683,887 bp, displaying the distribution of mapped PacBio DNA reads (gold filled area plot) and RNA reads (red filled area plot). The RNA reads were retrieved from BIPAA and incorporated in the plot to validate expression of the coding gene in this proviral locus. The boxed regions indicate the respective zoomed-in views of areas containing the newly found proviral loci (PL11). **(B)** Zoomed-in plots emphasizing two types of read coverage (DNA reads and RNA reads) distribution across the respective proviral locus. The y-axis represents read coverages and the x-axis shows genomic locations in the scaffold. Purple arrows mark where the newly identified RUM is located. The orange arrow indicates the open reading frame of a newly identified protein tyrosine phosphatase in this proviral segment.

It is commonly known that bracoviral DNA circles often encode PTPs that might function to impair the immune response of the parasitised host against oviposited wasp eggs [419, 420]. Homologs of bracoviral PTPs are found across multiple BV species and have been grouped into clades based on sequence similarities and conservation of PTP domains [421]. To determine the placement of PTP_PL11 among the bracoviral PTPs, we compared its sequence to the amino acid sequences of other bracoviral PTPs. An alignment of PTP_PL11 against 24 other PTPs from CcBV revealed that PTP_PL11 has greatly diverged, while still retaining most of the 10 PTP structural motifs characterised in human PTP domains using crystallography [422], except for Motif 1, Motif 8 and Motif 10 (**Figure 14**).

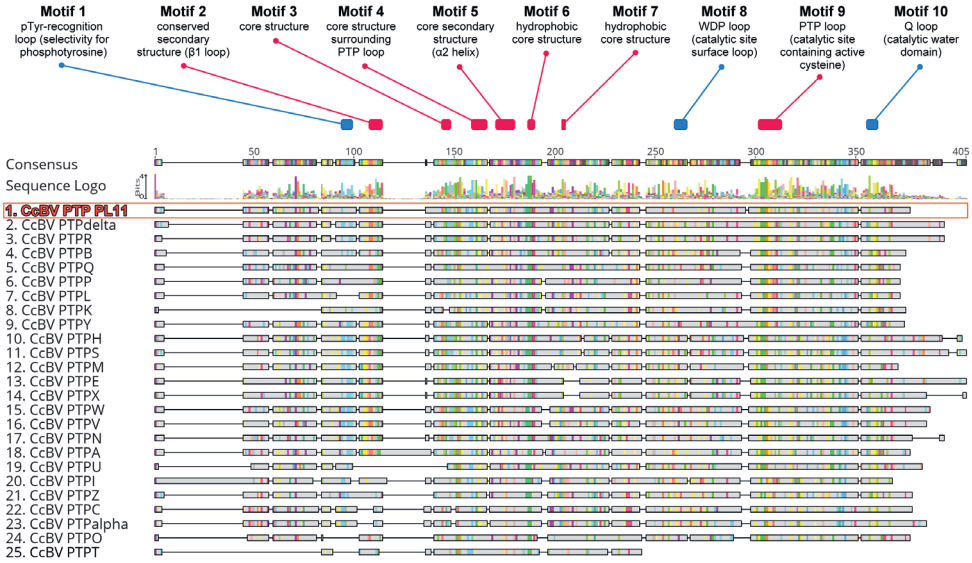


Figure 14. Protein sequence alignment of protein tyrosine phosphatases (PTPs) from Cotesia congregata bracovirus (CcBV). The amino acid sequences of 25 bracoviral PTPs, including the newly identified PTP from PL11 (red font and frame), were aligned with the MAFFT alignment tool (v7.490) of Geneious (v2023.2.1). The locations of conserved PTP motifs with structural and catalytic functions are displayed at the top, as identified in the study by Provost *et al.* (2004). Pink motifs are present in PTP_PL11, while blue motifs appear lost or only partially intact.

In addition to the alignment with PTPs from CcBV, we aligned the protein sequence of PTP_PL11 with a total of 146 bracoviral PTP sequences from multiple bracoviruses. The resulting alignment file was then used to generate a phylogenetic tree (**Figure 15**) to assess PTP_PL11's phylogenetic placement among these PTPs. This phylogenetic tree further illustrates the distinct diversification of PTP_PL11, showing that PTP_PL11 does not cluster within any of the monophyletic groups that have been characterised as clades based on Serbielle *et al*. (2012) [421]. This suggests that PTP_PL11 represents a unique lineage within the family of bracoviral PTPs.

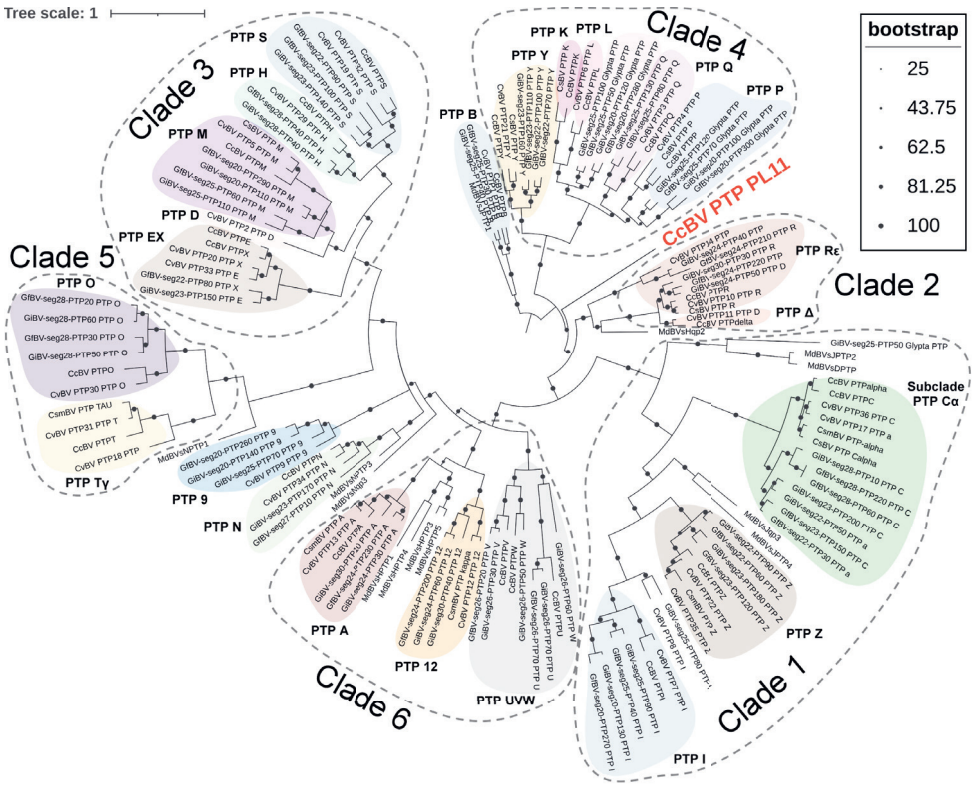


Figure 15. A molecular tree depicting the phylogenetic relationships among protein tyrosine phosphatases (PTPs) from various BVs. The tree was constructed from MAFFT-aligned amino acid sequences using IQ-TREE 2 [210] and includes 1,000 bootstrap replicates. The newly identified PTP_PL11 from Cotesia congregata bracovirus (CcBV) is highlighted in red and the obtained treefile was visualised in iTOL [211]. The PTP clades are highlighted in different colours. CcBV, Cotesia congregata bracovirus; CvBV, Cotesia vestalis bracovirus; CsBV, Cotesia sesamiae bracovirus; CsmBV, Cotesia sesamiae mombasa bracovirus; GfBV, Glyptapanteles indiensis bracovirus; GiBV, Glyptapanteles flavicoxis bracovirus; MdBV, Microplitis demolitor bracovirus.

Boundaries of amplified proviral segments in *T. nigriceps* are defined by direct repeat junction (DRJ)-like and TA-rich flanking sequences

The provided long reads were aligned to the assembled *T. nigriceps* genome and, similarly to *C. congregata*, and the resulting BAM files were converted to BED files. The inspection of the BED files using genomic visualisation tools such as IGV [414] allowed for the identification of high-coverage regions in the genome of *T. nigriceps*, which were expected to represent the amplified RU within the proviral segments. Strikingly, in contrast to CcBV and MdBV, there were no RUM sequences detected for *T. nigriceps*. Therefore, the 200-bp up- and downstream nucleotide sequences surrounding regions with elevated read coverage in the *T. nigriceps* genome were extracted from all proviral regions and screened for conserved motifs with MEME (**Figure 16**).

Figure 16. Sequence alignments from regions flanking high-read coverage areas within the *T. nigriceps* genome. Each sequence alignment is framed by a coloured border corresponding to conserved motifs identified across the sequences with MEME. Sequences within the red frame are characterised by a TA-rich motif, indicated by the sequence identifier "TA." The blue frame groups sequences are under the Type 1 (T1) motif, while the purple frame contains sequences with the Type 2 (T2) motif. Alignments within the yellow and green frames are designated as Ambiguous motifs, A1 and A2, respectively, due to their location-related improbability of functioning as replication unit motifs (RUMs).

The MEME analysis identified two predominant motif types among the FSs, including a motif we designated as type 1 (T1) and a motif with TA-rich region (TA). A smaller subset of sequences was

classified into three additional motif types (T2, A1 and A2). However, not all FSs submitted to the MEME analysis were found to have a conserved motif. In the coverage plots (**Figure 18–20**), the found motifs were marked with arrows, colour-coded according to the scheme shown in **Figure 16**. Except for TA-rich sequences, none of these motifs were found via BLASTn search in the genome of *C. congregata*. Additionally, all assembled *T. nigriceps* scaffold were submitted as query to a BLASTn search with the RUM sequences of CcBV as subject, but no RUMs similar to CcBV were found in the *T. nigriceps* scaffolds. Despite the absence of conserved RUM sequences between CcBV and TnBV, the T1 motifs of TnBV showed high sequence similarity to the direct repeat junctions (DRJs) of other BVs. Bracoviral DRJ regulatory sequences allow the excision or circularisation of the bracoviral dsDNA circles and their encapsidation into the virus particles [75, 424], and were shown to have in particular a characteristic AGCT motif shared among bracoviral DRJs [125]. The conservation of this motif is clearly distinguishable when aligning the type 1 motifs of TnBV with the DRJs of other BVs (**Figure 17**).

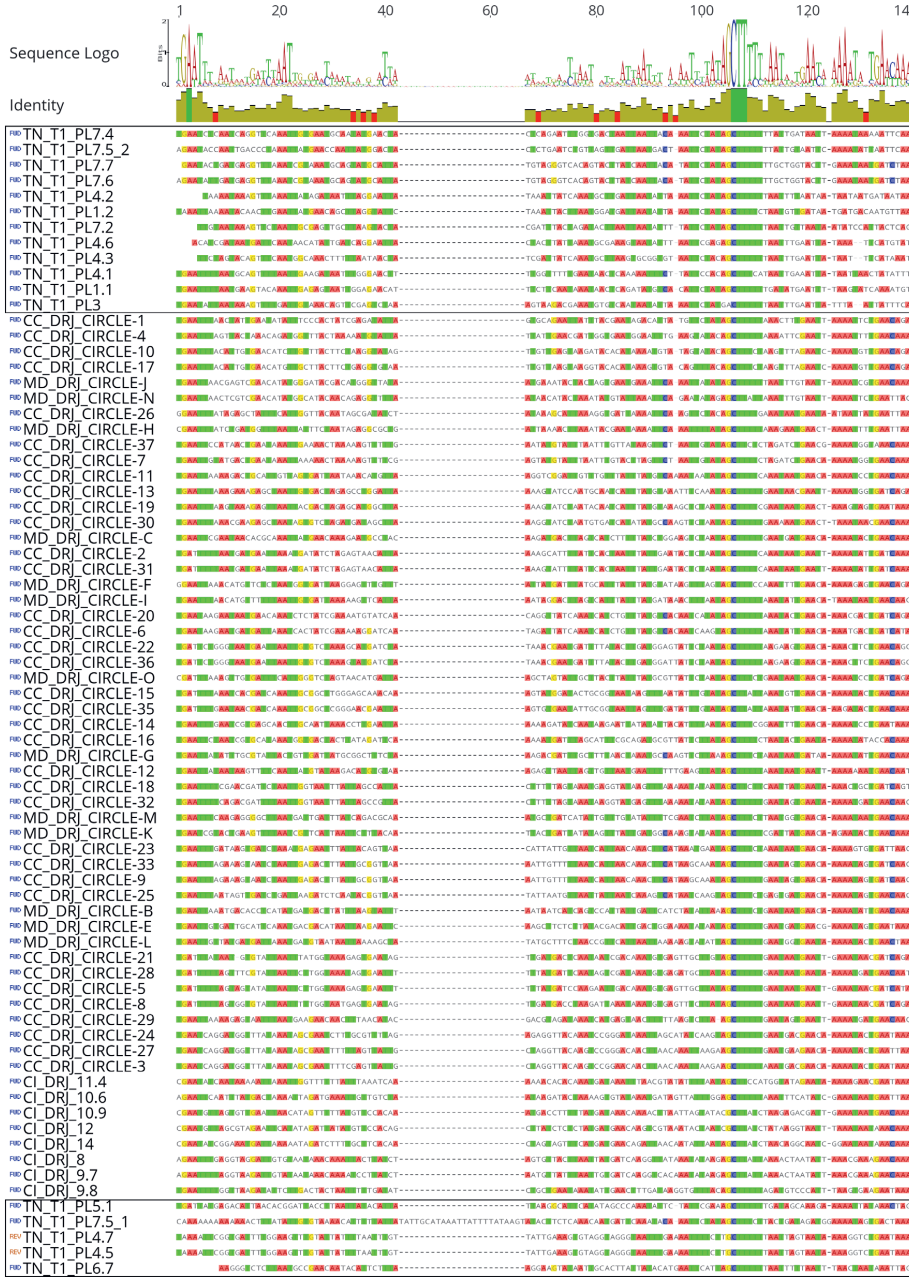


Figure 17. Nucleotide sequence alignment of bracoviral DRJs. The sequences of Cotesia congregata bracovirus and Microplitis demolitor bracovirus are indicated with CC and MD, respectively, and were obtained from a study by Gauthier *et al.* (2021) [82]. The DRJ names with CI indicate sequences from Chelonus inanitus bracovirus, which derive from a study by Cerquiera *et al*. (2022) [293]. The sequences of the type 1 motifs from TnBV are framed in black. The sequences were aligned using the MAFFT alignment tool (v7.490) of Geneious (v2023.2.1).

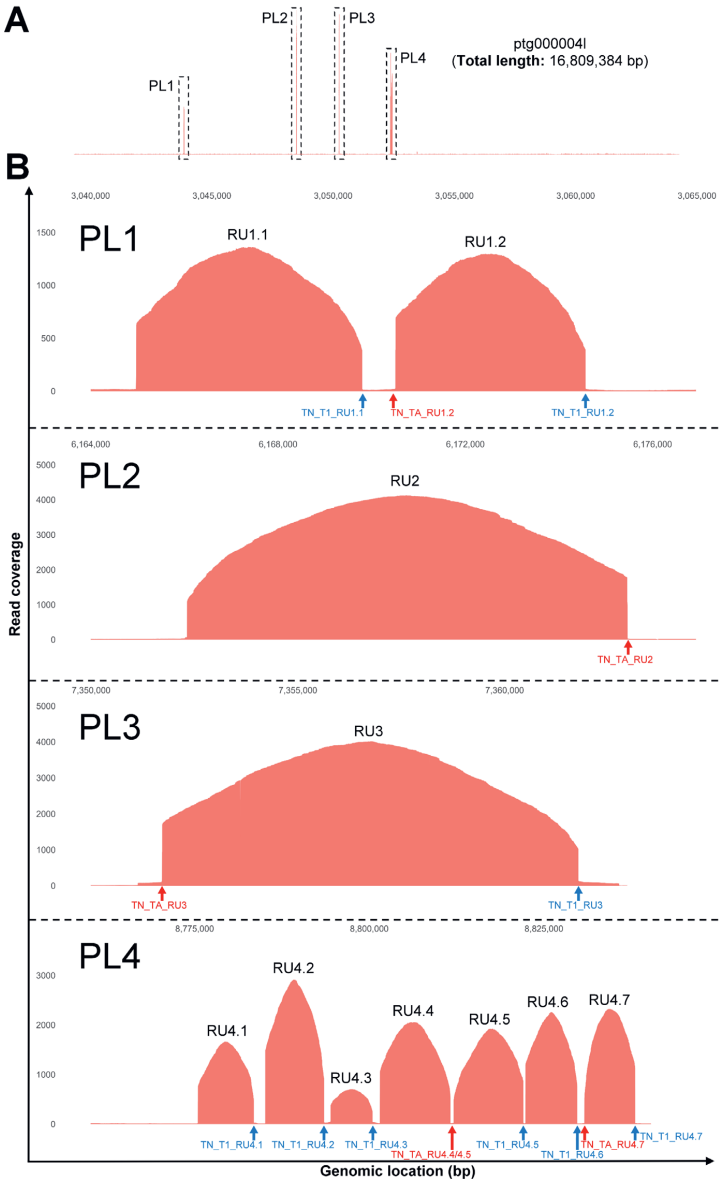


Figure 18. Filled area plot showing the PacBio read coverage across *T. nigriceps*' scaffold ptg000004l. **(A)** Whole coverage plot of ptg000004l with a total length of 16,809,384 bp, displaying the distribution of mapped PacBio reads. The boxed regions indicate the respective zoomed-in views of areas containing the proviral loci (PL1, PL2, Pl3 and PL4). **(B)** Zoomed-in plots emphasizing the read coverage distribution across the respective proviral loci. The y-axis, with scales adjusted for each enlarged section, represents read coverages and the x-axis shows genomic locations in the scaffold. Differently coloured arrows indicate the location of conserved motifs within the flanking sites of the respective regions with high coverage and can be found in the sequence alignment of **Figure 16**.

After identifying the different types of BV-like motifs in *T. nigriceps*, their locations were visualised in relation to the long-read coverage patterns across the *T. nigriceps* scaffolds. The *T. nigriceps* scaffold

ptg000004l features four proviral loci, namely PL1, PL2, PL3 and PL4 (**Figure 18A**). While the coverage plots of PL2 and PL3 both displayed single high-coverage peaks, the read coverages across PL1 and PL4 indicated multiple peaks that were divided by low coverage regions (**Figure 18B**), implying that PL1 is split into two RUs (RU1.1 and RU1.2) and PL4 is divided into seven RUs (RU4.1 to RU4.7) – each regulated by a distinct pair of regulatory sequences – while PL2 and PL3 appear to amplify only a single RU. Overall, the proviral loci of ptg000004l were shown to be flanked only by motifs of type TA and T1, some of which are possibly shared between RUs, for instance the TA-rich motif between PL4.4 and PL4.5.

The coverage plot for ptg000006l allowed for the clear distinction of PL5 from the rest of the wasp genome (**Figure 19A**). When inspecting the high coverage region corresponding to PL5 from a zoomed-in view, four distinct coverage peaks showed up. Those peaks were separated by low coverage regions of less than 1,400 bp length (**Figure 19B**). Across the coverage map of PL5, a single T1 FS was identified (TN_T1_RU5.1), while the remaining FSs all contained TA-rich sequences. Given the short distance between the four different coverage peaks, it can be assumed that PL5 features four distinct RUs.

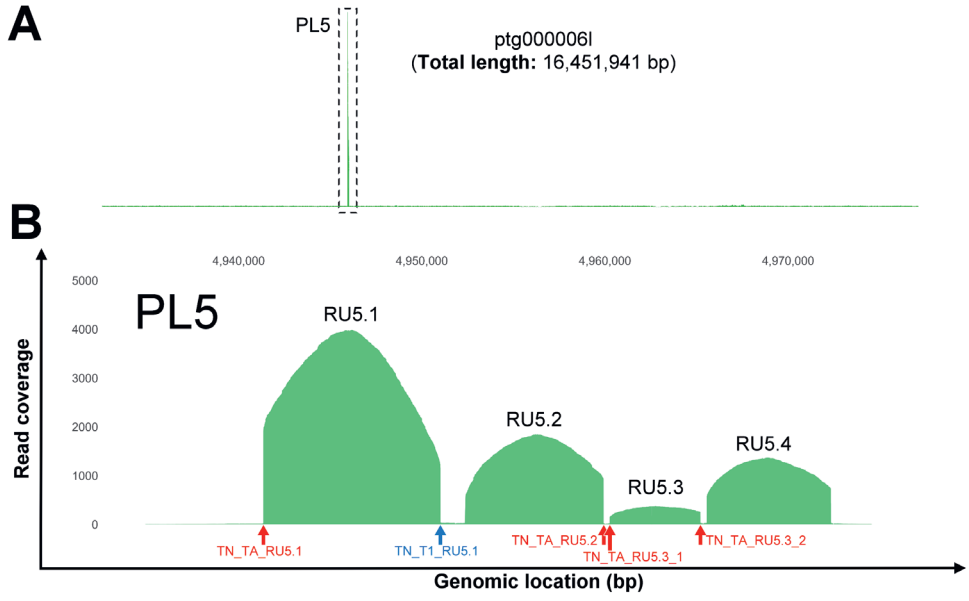


Figure 19. Filled area plot showing the PacBio read coverage across *T. nigriceps*' scaffold ptg000006l. **(A)** Whole coverage plot of ptg000006l with a total length of 16,451,941 bp, displaying the distribution of mapped PacBio reads. The boxed region indicates the zoomed-in view of the area containing the proviral locus (PL5). **(B)** Zoomed-in plot emphasizing the read coverage distribution across the proviral locus. The y-axis represents read coverages and the x-axis shows genomic locations in the scaffold. Differently coloured arrows indicate the location of conserved motifs within the flanking sites of the respective regions with high coverage and can be found in the sequence alignment of **Figure 16**.

Similar to the other proviral loci, PL6 and PL7 were easily discernible in the coverage plot from the surrounding *T. nigriceps* genome (**Figure 20A**). Identical to PL1, PL4 and PL5, they initially appeared as two single high-coverage peaks, but the magnified view revealed multiple coverage peaks spreading across those two proviral loci (**Figure 20B**). Both PL6 and PL7 appeared to consist of six RUs, commonly flanked by TA-rich regions and T1 motifs. Two notable decreases in coverage were observed

within RU6.5, flanking a 275-bp region where read coverage was elevated, following the expected coverage trend of RU6.5. When examining the adjacent sites around and within this region, we discovered two pairs of sequences that were almost identical two each other (yellow arrows: TN_A1_1 and TN_A1_2; green arrows: TN_A2_1 and TN_A2_2). Due to their unknown nature as putative regulatory sequences, they were designated as ambiguous motifs, hence, the identifiers "A1" and "A2". In PL7, a new type of motif was identified in three FSs of RU7.4, RU7.6 and RU7.7, which displayed nearly identical sequence similarity to one another (**Figure 17**). This motif was designated as type 2 (T2) motif and appeared to be specific for PL7.

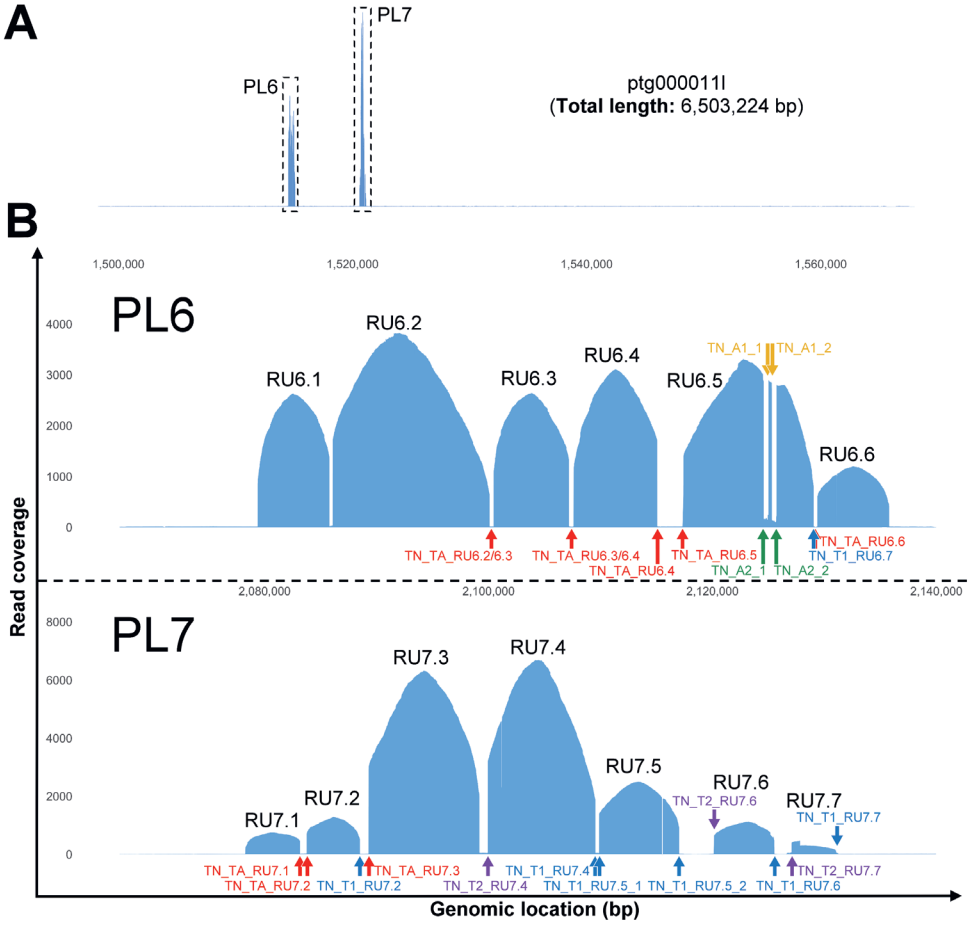


Figure 20. Filled area plot showing the PacBio read coverage across *T. nigriceps*' scaffold ptg000011l. **(A)** Whole coverage plot of ptg000011l with a total length of 6,503,224 bp, displaying the distribution of mapped PacBio reads. The boxed regions indicate the respective zoomed-in views of areas containing the proviral loci (PL6 and PL7). **(B)** Zoomed-in plots emphasizing the read coverage distribution across the respective proviral loci. The y-axis, with scales adjusted for each enlarged section, represents read coverages and the x-axis shows genomic locations in the scaffold. Differently coloured arrows indicate the location of conserved motifs within the flanking sites of the respective regions with high coverage and can be found in the sequence alignment of **Figure 16**.

In sum, we did not detect any RUM similarities between *T. nigriceps* and *C. congregata*, aside from the shared presence of TA-rich sequences. Instead, we found that the braconudviral DNA amplification

boundaries in *T. nigriceps* are delimited by sequences that share sequence similarity with DRJs of other BVs. A notable difference in the coverage data between *C. congregata* and *T. nigriceps* was the more distinct separation of TnBV RUs into multiple peaks in the coverage plots, whereas CcBV PLs with multiple assumed RUs (e.g. PL2 and PL6) were missing such clear amplification boundaries.

Discussion

Our PacBio sequencing approach revealed that most of the previously and newly identified RUs and RUMs from CcBV were well-defined and aligned with the long-read coverage patterns. The amplification boundaries of PL1, PL2, PL4, PL5, PL6, PL7, PL8, PL9, PL10, and the newly discovered PL11 were also supported by the long-read coverage in the genome of *C. congregata*. However, we found incongruency for PL8's and PL3's proposed RUMs, suggesting the involvement of previously undesignated RUMs. The suggested RUM for PL8, T1_H_PL8_SEG-26 did not align with the amplification boundaries of its RU. Instead, aided by our coverage data, we identified another motif with high similarity to other Type 1 Tail motifs. This finding suggests that the RU of PL8 is not amplified as a head-tail concatemer but rather as a tail-tail concatemer. When examining the FSs of PL3's coverage-inferred RU and searching for similar sequences in the genome of *C. congregata*, we identified nearly identical sequences on two other scaffolds (scaffold_333 and scaffold_CcPL6). The inspection of those sequences on the BIPAA database showed that they flank high-coverage regions that are annotated as TEs. This is an intriguing finding, given the suggestion by previous studies that the proviral segments of bracoviruses may be partially derived from TEs [82, 425]. The acquisition of TE-associated regulatory sequences for bracovirus replication may have been a possible facilitating factor for bracovirus domestication. TEs can relocate within a genome through copy-and-paste (retrotransposons) or cut-and-paste (DNA transposons) mechanisms [426]. Those TE-specific traits may have been co-opted by the bracoviral DNA, leading to their evolutionary divergence as components of the bracoviral domestication, which is a commonly known phenomenon observed for TEs [427].

Next to PL3's putatively TE-derived flanking motifs, we found that the proviral locus PL10 is flanked by a Type 1 Head and Type 1 Tail RUM, indicating that it might be amplified as a head-to-tail concatemer. Similarly, we discovered another Type 1 Head RUM at one of the amplification borders of a potentially novel proviral locus, which we designated PL11. While we could not identify a partnering RUM on the other site of PL11's amplification boundary, we found that PL11 contained the open reading frame of a yet uncharacterised protein tyrosine phosphatase (PTP), which we called PTP_PL11. Although PTPs are not a novelty within this virus clade, we showed that PTP_PL11 had some distinct characteristics compared to other bracoviral PTPs. An amino acid sequence alignment against other PTPs from CcBV showed that PTP_PL11 had retained most PTP motifs (as reported in Provost *et al*., 2004) [423], but not all, unlike some other BV PTPs. The lost or only partially intact motifs of PTP_PL11 are associated with selectivity for phosphotyrosine (Motif 1) and catalytic activities (Motif 8 and Motif 10). Despite the confirmed expression of PTP_PL11 based on RNA-seq data, the loss or partial preservation of those PTP motifs may argue against its enzymatic activity. The missing pTyr-recognition loop likely impairs its ability to specifically bind phosphotyrosine, while the compromised WPD loop could significantly reduce its catalytic efficiency. Furthermore, the partially intact Q loop may further hinder the enzyme's ability to stabilise the transition state during catalysis [428]. Even so, despite the partial loss of those catalytic motifs in PTP_PL11, the enzyme might still function similarly to viral inhibitors observed in baculoviruses, poxviruses, and vaccinia virus, where PTPs have been proposed to interfere with host signalling pathways by binding to phosphorylated proteins [429-431]. PTP_PL11 may have a similar role, potentially acting as a molecular trap that hinders host immune responses and other critical cellular functions during parasitism. This possibility suggests that, while

PTP_PL11 might not perform traditional phosphatase activities, it could still play an essential role in the virus' strategy to manipulate its host. However, PTP_PL11's phylogenetic position, remote from all other PTP clades, also raises the prospect that this protein has evolved distinct functions that differ from those of its counterparts within the BV clade. Further experimental studies are necessary to elucidate the function and importance of PTP_PL11 for CcBV as for the other bracoviral PTPs.

Furthermore, the distinct coverage patterns and the absence of conserved RUM sequences between TnBV and CcBV strongly suggest that these bracoviruses utilise distinct mechanisms for replicating their RUs. Conversely, the conservation of the DRJ motif among TnBV and other bracoviruses indicates that the mechanism for viral DNA excision and circularisation is more strongly conserved than the RUM-mediated replication mechanism described for CcBV and MdBV. Given this, the distinct demarcation of TnBV DNA amplification boundaries by FSs with DRJ motifs, and the corresponding long-read coverages, imply that DNA amplification does not extend beyond those DRJ-like motifs. This is further supported by the observed coverage between adjacent RUs in TnBV (intersegmental regions), which matched to the background coverage of the wasp genome. If higher coverage had been detected in these intersegmental regions, it might have indicated the amplification of a larger DNA molecule before excision and circularisation, similar to what has been described for CcBV and MdBV, where these intersegmental regions are referred to as amplified but not encapsidated sequences (ANES) [106]. Our findings, however, suggest that the RUs of TnBV are independently amplified, with their FSs possibly functioning as both DRJs and RUMs. This observation would also align with the initial replication model proposed for bracoviruses by Savary *et al.* (1997) [424], which suggested that bracoviral circles would first be excised – likely by a nudiviral recombinase [106, 125] – and then amplified by rolling circle amplification. In this context, it is unclear what function the RU-flanking TA-rich sequences fulfil. Previous studies suggested that they could serve BVs as origins of replication [106, 109]. Assuming that TnBV DNA amplification occurs post-excision of proviral regions, the TA-rich motifs would still need to be present in the excised DNA. However, this only applies to one of the identified motifs, TN_TA_RU3, which emphasises that our proposed replication model for TnBV requires further exploration – a task not covered in this thesis due to time constraints. Either way, the utilisation of such a mechanism for TnBV replication would provide a major difference to the BVs of *C. congregata* and *M. demolitor*, shedding new light on their evolutionary history. *T. nigriceps* belongs to the subfamily Cardiochilinae and is more distantly related to *C. congregata* and *M. demolitor* of the subfamily Microgastrinae. Of those two subfamilies, the evolutionary origin of the Cardiochilinae was estimated to be older (~67 million years ago, Mya) than the one of the Microgastrinae (~53 Mya). Therefore, it appears plausible that the amplification of multiple RUs in a single, large DNA molecule evolved specifically in the PLs of BVs from the Microgastrinae lineage, and possibly originated from a simpler mechanism, more similar to TnBV. Conclusively, the evolutionary distance between wasps of the Cardiochilinae and Microgastrinae lineage likely extended to their symbiotic viruses, leading to the development of different replication strategies within their respective bracoviruses.

Acknowledgements

The writing of this chapter has received funding from the European Union's Horizon 2020 research and innovation programme INSECT DOCTORS under the Marie Skłodowska-Curie grant agreement No. 859850. We are grateful to Rosanna Salvia (University of Basilicata, Italy) for extracting the DNA from *T. nigriceps*. The long-read sequencing of *T. nigriceps* was funded by Insect Biology Research Institute (IRBI) via the Cotebio ANR project. We would like to thank Vincent Pailler (Gentyane, France) for his help on the assembly of *T. nigriceps* and mapping of the reads to the assembled scaffolds, and Fabrice

Legeai (National Institute for Agricultural Research – INRAE, France) for making the *T. nigriceps* sequencing data available on BIPAA and performing the genome annotation.

“**Diseases come, diseases go. Welcome to the final show.**”

– *Every Generation* by Fury in the Slaughterhouse (2002)

Chapter 7

General discussion

Introduction

In my thesis, I elucidated various aspects of viruses in the *Nudiviridae*. The research field of large nuclear-invertebrate infecting DNA viruses has long been dominated by studies on baculoviruses (family *Baculoviridae)*. During the era of traditional virology, nudiviruses were first mistaken for baculoviruses due to similar pathological and morphological characteristics. It was not until the emergence of sequence-based virology that nudiviruses were officially distinguished from the family *Baculoviridae*. Today, nudiviruses are unified under the family *Nudiviridae* within the class *Naldaviricetes* and the order *Lefavirales* [36]. Along with the *Nudiviridae*, the order *Lefavirales* comprises other arthropod-infecting virus families with large double-stranded DNA (dsDNA) genomes, such as *Baculoviridae*, *Hytrosaviridae*, and the recently proposed Filamentoviridae [35]. Although the family *Nimaviridae* is also classified within the class *Naldaviricetes*, it is not part of the *Lefavirales* order (**Chapter 1**, **Figure 3**). However, all members of the class *Naldaviricetes* share seven core genes among each other, including multiple *per os* infectivity factors (*p74*/*pif-0*, *pif-1*, *pif-2*, *pif-3*, and *pif-5*), the sulfhydryloxidase (*p33*), and the DNA polymerase (*dnapol*) [35]. Among the lefavirals, the *Nudiviridae* and *Baculoviridae* exhibit the highest number of shared core genes, with two-thirds of nudiviral core genes conserved between them [171]. Despite this genetic similarity, phylogenetic analyses have confirmed that nudiviruses and baculoviruses from distinct monophyletic clades. Since its origins, virus taxonomy has been greatly reshaped through the introduction of sequence-based virology, but the pioneering achievements of traditional virology remain invaluable. In this thesis, I integrated both traditional and advanced virological methods to emphasise their co-dependence in unlocking the full potential of virological research.

My thesis covers a comprehensive review of existing literature and new experimental data to build an informational nexus on the *Nudiviridae* family. **Chapter 1** introduced the historical development of virology, highlighting the advances that led to the discovery and characterisation of nudiviruses. **Chapter 2** explored the relationship between nudiviruses, bracoviruses, and baculoviruses. Following the publication of this chapter in 2022, new studies, including our Data-Driven Virus Discovery (DDVD) approach, revealed additional nudivirus species in previously unrecognised arthropod hosts in NCBI's sequencing read archives (SRA). Based on these new findings, **Chapter 3** provided an updated nudiviral phylogeny, and analytical review on nudiviral diversity, genomics, evolution and ecology. Taken together, these chapters provided an informational foundation towards the currently available knowledge of the family *Nudiviridae* and its relatives. The subsequent chapters focused on studying the pathological and genomic dynamics of exogenous (or free-living) and endogenous virus systems. The Heliothis zea nudivirus 1 (HzNV-1) was the main subject of **Chapter 4 and 5**. As such, I examined its infection and transcriptional dynamics in an ovary-derived cell line of *Helicoverpa zea*, by employing a combination of microscopic, molecular biological, and NGS techniques. **Chapter 6** utilised long-read sequencing to investigate and compare the replication mechanisms of two bracoviruses from a microgastrine and a cardiochiline parasitoid wasp. In this general discussion of my thesis, I aim to synthesise the findings of the previous chapters into a global discussion, while also offering future directions for the field of nudivirus research.

Nudivirus taxonomy: What is the current status?

Since the official creation of the family *Nudiviridae* in 2014 [37], numerous new nudiviruses have been identified and their genetic information published. The data-mining of publicly available sequencing data certainly is one of the main drivers for new virus discoveries. In our DDVD study (**Chapter 3**), we assembled complete genomes of novel nudiviruses from sequencing data of arthropods that were

previously not recognised as nudiviral hosts, including flea- and louse-infecting nudiviruses. This allowed expansion of nudivirus diversity to 49 distinct viruses associated with insect hosts in the orders Coleoptera, Diptera, Hemiptera, Hymenoptera, Lepidoptera, Siphonaptera, Orthoptera, and Psocodea, as well as decapod and amphipod crustaceans. As a family, the nudiviruses currently comprise the widest host range within the *Naldaviricetes*, making them the most diverse group of viruses in terms of host interactions. While DDVD is without a doubt a powerful tool for the field of virology, it does not make traditional virology obsolete. Instead, traditional and sequence-based virology complement each other, one may say they are two sides of the same coin. In this context, the retrieval of full nudivirus genomes from the cyberspace of SRAs drives virus discovery and adds new molecular information, but *in vivo* studies are ultimately required to investigate morphological and pathological traits of these viruses, including their tissue tropism and transmission routes.

Nonetheless, the nudiviral genomes and gene sequences that we retrieved *in silico* allowed us to conduct new comprehensive phylogenetic analyses, prompting reclassifications within the *Nudiviridae*. Some monophyletic groups within the *Nudiviridae* grew in diversity through the addition of new members, such as the flea-infecting nudivirus within the genus *Alphanudivirus*, and newly discovered members within the genus *Gammanudivirus* and Epsilonnudivirus clade. Notably, we retrieved full genome sequences from eight members of a new monophyletic group of nudiviruses from the sequencing data of chewing lice. The phylogenetic position of these louse-associated nudiviruses made it clear that they belonged to a separate clade within the *Nudiviridae*. However, the examination of the core gene synteny from all nudiviruses with fully annotated genomes, led us to propose a subdivision of the louse-infecting nudiviruses into two new genera: Zetanudivirus and Etanudivirus. Although gene synteny is currently not an official criterion for virus classification, we argue that looking for conserved gene synteny profiles could very well serve as a supplementary feature in distinguishing related groups of nudiviruses within existing clades. Additionally, the taxonomical position within monophyletic groups of nudiviruses that display divergent gene synteny – such as the Myrsidea ptilorhynchi nudivirus (MpNV) from the proposed Zetanudivirus clade – may offer initial hints for future subdivisions within that group when more viruses are discovered.

The genus *Alphanudivirus* harbours the greatest diversity of insect-infecting nudiviruses, covering hosts across six insect orders. However, the classification of these viruses in one genus is primarily based on phylogeny, and gene synteny is poorly conserved within *Alphanudivirus*. An exception to this is the conserved gene synteny observed among certain dipteran-infecting viruses, including Esparto virus (ENV), Kallithea virus, Mauternbach virus (MNV), and Drosophila innubila nudivirus (DiNV). Apart from those, the genus *Alphanudivirus* comprises several free-living members with unique gene synteny patterns, including the orthopteran Gryllus bimaculatus nudivirus (GbNV), hemipteran Phenacoccus solenopsis nudivirus (PsNV), hymenopteran Apis mellifera nudivirus (AmNV), and the most recently added siphonapteran Ctenocephalides orientis nudivirus (CoNV). It is likely that the discovery of more nudiviruses from orthopteran, hemipteran, hymenopteran or siphonapteran hosts would reveal monophyletic subbranches within the genus *Alphanudivirus* with conserved gene syntenies. Thus, I support that the genus *Alphanudivirus* will eventually be subdivided into more genera, and that the current *Alphanudivirus* genus should become a subfamily with the name Alphanudivirinae as proposed by Liu *et al*. (2020) [69]. If that happens, then the monophyletic group comprising the genera *Betanudivirus*, *Gammanudivirus*, Epsilonnudivirus, Zetanudivirus, Etanudivirus, and the bracoviruses should eventually be unified under the subfamily Betanudivirinae.

The land of confusion: Bracovirus? Bracoviriform? Braconudivirus?

Among the above-mentioned clades, the taxonomic classification of bracoviruses remains one of the most debated and challenging issues in modern insect virology. During an undergraduate course that I supervised, students were tasked with writing a report to discuss whether bracoviruses and ichnoviruses are true viruses. This question is complex, even for experts, as it challenges traditional definitions. Originally grouped under the family *Polydnaviridae* due to their similar, unconventional life cycles [432], sequence-based virology later showed that bracoviruses and ichnoviruses do not share a common ancestor [433]. Instead, it was revealed that bracoviruses are closely related to nudiviruses [74], while ichnoviruses originate from a yet unknown viral ancestor [433-435]. On this basis, **Chapter 2** reviewed how the conservation of nudiviral core genes varies among members of the genus *Bracovirus*, but also showed that genes involved in transcription and virion assembly are well represented among the inspected bracoviral members. Despite the functional preservation of several nudiviral core genes and the infectivity of bracoviral particles, bracoviruses are unable to replicate outside their initial parasitoid hosts. This limitation is a key reason why their classification as true viruses remains a topic of debate. Some argue they should be termed "viriforms", or even advocate their exclusion from the virosphere [436]. Others suggest to classify them as Domesticated Endogenous Viruses (DEVs) to emphasise their evolutionary link to nudiviruses, and the conservation of the nudiviral core genes that have retained most of their viral functions even after millions of years of domestication [81]. I believe that whether bracoviruses are considered viruses or viriforms ultimately depends on the scientific community's definition of a virus. However, the challenge lies in maintaining unity in definitions and revising terminologies according to recent scientific advances.

Similarly, debates about the taxonomic placement of bracoviruses continue. Our phylogenetic analyses (**Chapter 3**) clearly support their grouping within the *Nudiviridae*, but referring to them as "braconudiviruses" still meets opposition. Phylogeny is a major criterion for distinguishing viral clades, but inconsistency arises when it is not uniformly applied across virus groups. The virological community must agree on the weight of phylogenetic relationships and evolutionary history in classifying viruses, ensuring consistent application across clades. On the basis of their evolutionary history, I assert that bracoviruses should be classified as members of the family *Nudiviridae*, despite their unconventional life cycle and inability to replicate outside their wasp host, once the bracovirus particles are formed and delivered to parasitised insects.

Evolutionary hallmarks of the *Nudiviridae*

In order to estimate the evolutionary timeline of the *Nudiviridae*, a molecular dating tree was inferred using the most recent common ancestor (MRCA) of Chelonus inanitus bracovirus (CiBV) and Cotesia congregata bracovirus (CcBV) as a calibration anchor at 103.38 ± 4.41 Mya [215], along with molecular information from their endogenised nudiviral core genes (**Chapter 3**). By combining this calibration point with the sequence data of nudiviral core genes from endogenous and exogenous viruses, we converted amino acid substitution rates into amino acid substitutions per million years. This approach allowed us to estimate the ages of divergence events across nudiviruses and baculoviruses (**Figure 1**). We showed that the MRCA of nudiviruses and baculoviruses dates back approximately 378 Mya, which aligns with the divergence of holometabolous insects from polyneopteran insects during the Devonian period of the Palaeozoic Era. Additionally, the newly incorporated molecular information from flea- and louse-infecting nudiviruses proved valuable in resolving divergence times between the two main

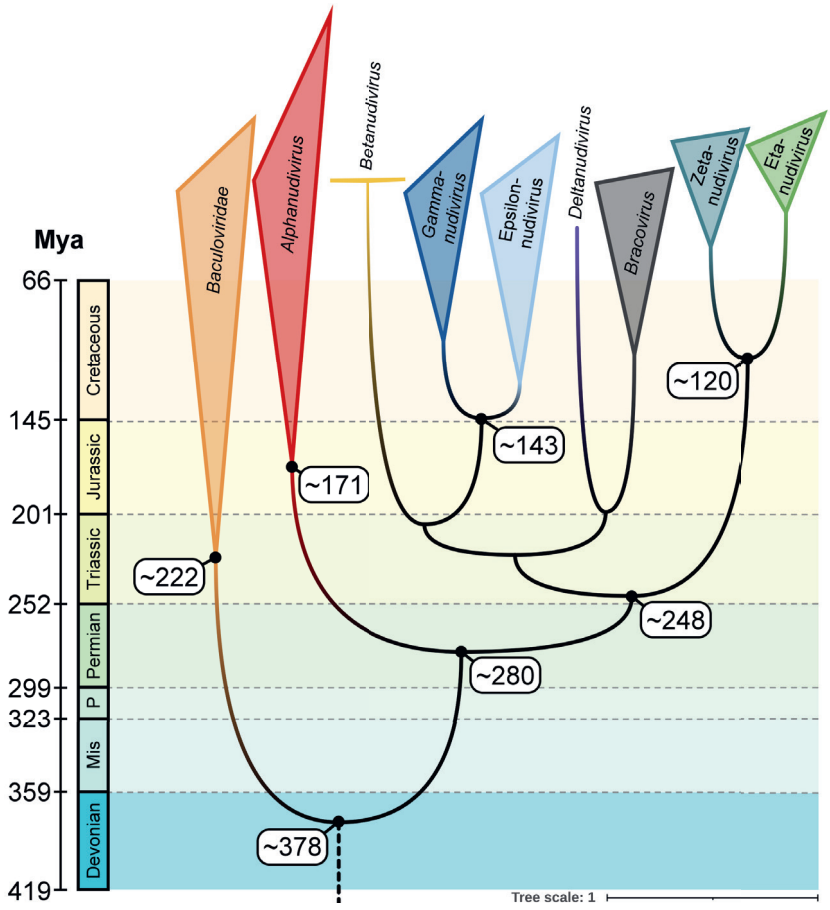


Figure 1. Phylogram showing evolutionary hallmarks in the timeline of the *Nudiviridae* and *Baculoviridae*. The scale on the left displays "million years ago" (Mya) with colour-coded geological periods corresponding to this timeframe. Numbered nodes indicate age estimates of the common ancestor from the respective clades, based on the molecular dating analysis from **Chapter 3**.

monophyletic groups within the *Nudiviridae*, proposed to be termed Alphanudivirinae and Betanudivirinae. While the MRCA of all nudiviruses dates back to approximately 280 Mya, the MRCA of all viruses in the Betanudivirinae (~248 Mya) precedes that of the alphanudivirins (~171 Mya). Although this suggests that psocodean nudiviruses emerged before hemipteran nudiviruses, it should be noted that our analysis included only one hemipteran nudivirus compared to eight psocodean nudiviruses. Therefore, adding more Alphanudivirinae members with older ancestors – presumably nudiviruses from hemipteran (true bugs), siphonapteran (fleas), or orthopteran (e.g., crickets) hosts – would likely revise those age estimates. Moreover, the molecular dating analysis estimated that the MRCA of the louse-associated nudiviruses emerged around 120 Mya, within the evolutionary timeframe of bird-like dinosaurs (approximately 90–170 Mya). It is intriguing to imagine that the diversification of these nudiviruses may have occurred within an ancestral louse, perhaps hitchhiking on the back of a feathered dinosaur. Additionally, the MRCA of crustacean-associated nudiviruses (*Gammanudivirus* and Epsilonnudivirus) dates back to approximately 143 Mya, supporting the hypothesis that the transmission of an insect-associated nudivirus to an ancestral crustacean host may

have led to the emergence of these aquatic nudiviruses. While we demonstrated that the nudiviral core genes in exogenous nudiviruses and endogenised bracoviruses are still subject to similar selective pressures, we were restricted to analysing only eight core genes due to the limited sequence availability of bracoviral genes.

Fantastic nudiviruses and where to find them

Following our discoveries in **Chapter 3**, I found myself wondering if the next new nudivirus might just be lurking around the corner. To investigate this thought, I collected cat fleas (*Ctenocephalides felis*) from a neighbourhood cat in Wageningen, Netherlands (**Figure 2**), hoping to gain both molecular and pathological insights into a potential new flea-infecting nudivirus *in vivo*. Given that the nudivirus associated with *Ctenocephalides orientis* originates from Perak, Malaysia, this side experiment offered an intriguing opportunity to investigate whether a nudivirus might also be present in a wild European population from a related species. Pooled fleas were processed for DNA extraction, and the extracted DNA sent for whole-genome sequencing, with the obtained reads assembled following the same approach as described in **Chapter 3**.

Figure 2. *Ctenocephalides felis* (cat fleas) magnified under a Leica Wild M3Z Stereozoom Microscope. Flea DNA was extracted and subjected to whole-genome sequencing to screen for potential nudivirus infection within the population. Specimens were collected in Wageningen, Netherlands.

Disappointingly, no nudiviral contigs were detected in the flea DNA, neither full nor partial sequences of nudiviral genes. This example emphasises that the discovery of new nudiviruses is not as straightforward as one might initially think. Frankly, even with our newly added findings to nudiviral host range, it was still a shot in the dark, given that I was not able to distinguish whether the collected fleas were showing symptoms that may imply a nudivirus infection. When looking for virus infections through traditional means, it is therefore all the more important to be familiar, or at least know someone who is familiar, with pathological diagnostics of the hosts one wants to screen for virus infections. Unlike RNA viruses, which exhibit higher mutation rates and thus greater adaptability across diverse host taxa [437], I assume that dsDNA viruses, such as nudiviruses, are generally not as prevalent as RNA viruses and less likely to shift hosts. Consequently, nudiviruses and other dsDNA viruses are more

challenging to find, especially if macroscopic symptoms are less obvious. Either way, searching natural arthropod populations for new nudiviruses is likely more fruitful than examining reared populations from industry, unless these populations exhibit pathological symptoms indicative of viral infections. This proposition is supported by our findings in **Chapter 3**, where new nudiviruses were detected only in sequencing data from wild-collected arthropods. Naturally, when establishing a new rearing or aquaculture facility based on wild-collected arthropods, it is advisable to screen these individuals for nudiviruses or other viruses at the beginning to ensure a pathogen-free stock and minimise the risk of viral outbreaks.

Nudiviruses and their ecological implications

The slower mutation rates of large dsDNA viruses [438], such as nudiviruses, may support a more stable co-evolution with specific hosts, making nudiviruses resilient and well-adapted within their established ecological niches [439-441]. I assume that this stability may strengthen species-specific interactions over time, contributing to a role of nudiviruses in maintaining long-term ecological dynamics within specific host populations. In particular, the tripartite relationship among nudiviruses, ectoparasitic insects, and their warm-blooded hosts offers an intriguing model for co-evolution, yet nothing is known about the virulence and fitness effects of flea- or louse-infecting nudiviruses. At this point, it remains speculative how far the influence of nudiviruses extends to ecological systems, since the virulence of nudiviruses can differ strongly from species to species, with some of them being more detrimental than others. Based on a few relatively well-studied models, including the Oryctes rhinoceros nudivirus (OrNV) or Helicoverpa zea nudivirus 2 (HzNV-2), it can be deduced that nudiviruses have the potential to affect their host populations on an ecological scale. The pathology of OrNV is apparently the most well-studied among all nudiviruses, due to its importance as a commonly applied biocontrol agent in Asia [442]. OrNV's host, the coconut rhinoceros beetle (*Oryctes rhinoceros*), is an agricultural pest known for damaging palm trees [443]. This pest has effectively been controlled by applying OrNV for several decades [444]. Multiple pathological studies have also been conducted on the betanudivirus HzNV-2, especially around the turn of the millennium [64, 145, 263, 265, 290, 445]. As described in **Chapter 4**, HzNV-2 is a rare example of a sexually transmitted virus that not only influences the mating behaviour of its lepidopteran host, but can also induces sterility in host progeny [64, 265]. HzNV-2's ability to impair *H. zea* reproduction could lead to declines in host populations, potentially altering ecological relationships and affecting predator and parasitoid species reliant on the host for food or reproduction. However, to better understand the ecological and evolutionary impact of nudiviruses, more in-depth studies on nudiviruses are needed, which are still very scarce compared to their baculoviral relatives.

Back to the future: New insights on nudivirus pathology through the HzNV-1 model

Since its discovery in 1978, HzNV-1 has served as a convenient model for nudivirus studies in cell lines [127], particularly due to its close relation to HzNV-2. Given the scarcity of HzNV-1 studies with modern methods, we conducted NGS and various cell biological analyses to elucidate different aspects of HzNV-1's pathology in an ovarian cell line derived from *H. zea*. By merging the experimental findings from **Chapters 4 and 5** into one schematic figure, I present a hypothetical model for the infection cycle of HzNV-1 (**Figure 2**). Based on what we found in **Chapter 4**, the model proposes that HzNV-1 virions enter ovarian cell lines of *H. zea* via macropinocytosis. This results in the formation of virion-filled macropinosomes, which are transported into the cell cytoplasm. Based on what has been

observed in baculoviruses [446], I assume that the nucleocapsid of HzNV-1 is then released into the cytoplasm through fusion of the macropinosome membrane and viral envelope. The nucleocapsid then moves, likely via actin filaments [447], to the nucleus and traverses towards the nuclear membrane to enter the nucleus through a nuclear pore, similar to what has been shown for the alphabaculovirus Autographa californica multiple nucleopolyhedrovirus (AcMNPV) [132]. From experimental observations with Plodia interpunctella granulovirus (a betabaculovirus), I presume that the dsDNA genome is then released into the nucleoplasm at the apical side of the nucleocapsid [448], occurring sometime before 3 hours post-infection (hpi). This is supported by our gene expression profiling from **Chapter 5**, which showed that the earliest HzNV-1 genes are expressed by 3 hpi, indicating that the viral genome must have been uncoated before this point. Among those early expressed HzNV-1 genes – which we referred to as Phase 1 genes – were the replication-associated *dnapol* gene, and the transcription-associated genes *p47*, *lef-9*, *lef-8*, *lef-4* and *lef-5*. The transcripts of *p47*, *lef-9*, *lef-8* and *lef-4* are eventually translated into the proteins that become part of the viral RNA polymerase, whereas the protein encoded by *lef-5* presumably functions as a transcription initiation factor [449]. These Phase 1 proteins are needed to transcribe the genes of Phase 2 and 3 between 6 to 9 hpi. In **Chapter 4**, we showed that it takes HzNV-1 at least 6 hours to initiate viral DNA replication. Congruently, HzNV-1 DNA replication commences upon expression of both Phase 1 and 2 genes, which are mostly associated with replication and nucleotide metabolism. Next to replication-associated genes, Phase 2 also comprised genes encoding for structural proteins required for virion assembly. Transcripts for most of the PIF proteins were found within 9 hpi, as well as transcripts for the major capsid protein VP39 and other virion components. The latest expressed core genes of Phase 4 included *pif-6* and *vlf*-1, which were expressed sometime after 9 hpi but within 12 hpi. Based on baculovirus studies, all PIF proteins are most likely integrated or closely connected to the viral envelope [60]. VLF-1 serves as a component of the nucleocapsid's basal region (opposite to the apical cap mentioned above) and presumably functions in viral DNA processing and packaging into the nucleocapsid [97]. Hence, I propose that nucleocapsid assembly must start somewhere between 9 to 12 hpi, but final packaging and envelopment of HzNV-1 virions likely occurs sometime after 12 hpi. I base this presumption on our RNA-seq study, which showed no detection of differently expressed genes (DEGs) associated with the host at 12 hpi when compared to the 0 hpi time point, but a drastic increase of host DEGs occurs at 24 hpi.

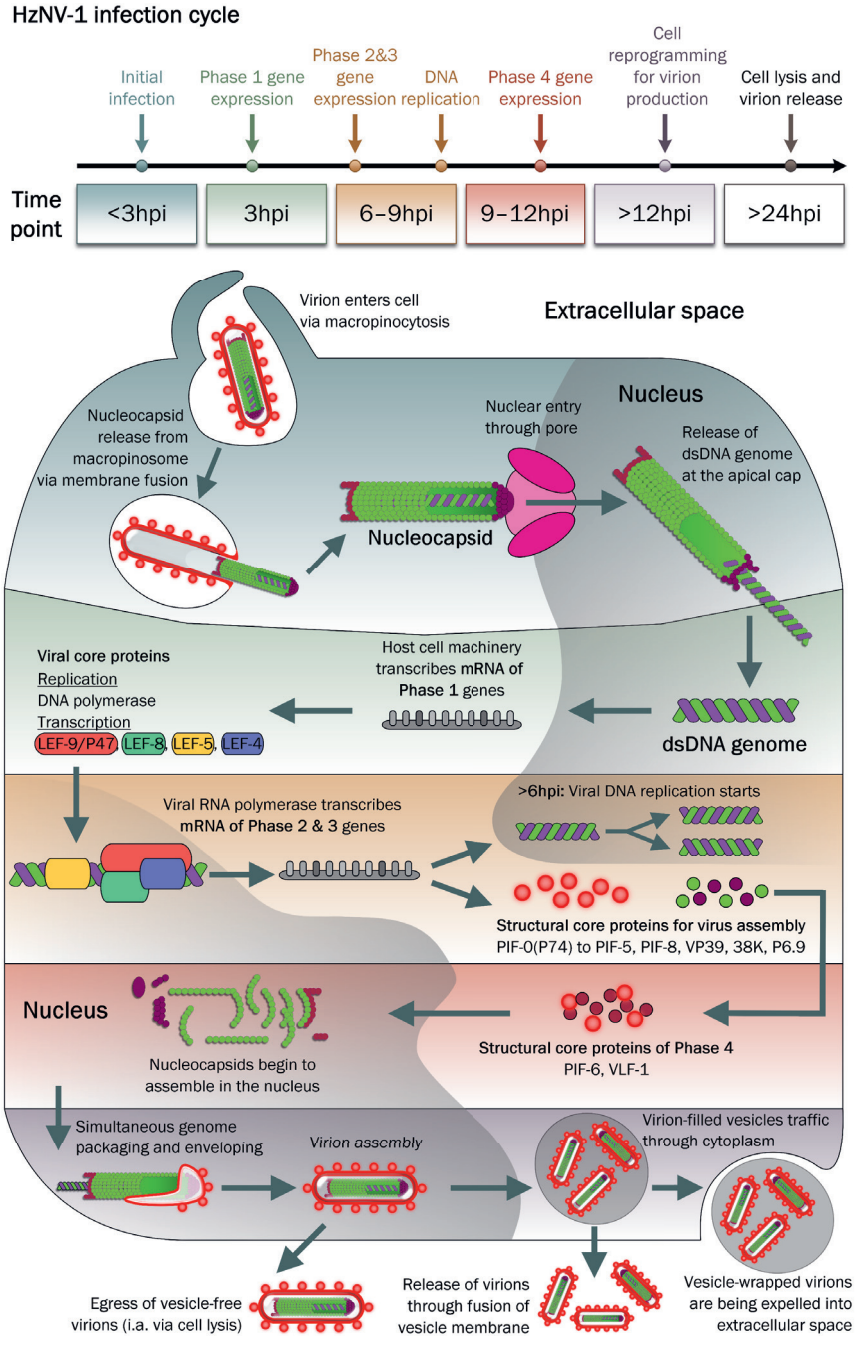


Figure 3. The top timeline highlights key milestones during the infection process at designated time points or intervals. Time points are colour-coded to correspond with specific regions in the cell model, indicating the timing and location of various infection-related processes. Dark-shaded areas within the cell represent nuclear processes, while non-darkened areas indicate cytoplasmic processes. The white area around the cell represents the extracellular space.

Therefore, I deduce that the tipping point in favour of HzNV-1 occurs sometime after 12 hpi, when the host cell is drastically reprogrammed to a viral factory, leading to the modification of the host protein folding machinery and other metabolic pathways, which are at this point likely hijacked for final virion maturation. Based on an electron microscopy-assisted study by Velamoor and co-authors with OrNV [52], it is likely that the encapsidation of the HzNV-1 genome and the envelopment of the nucleocapsids occur almost simultaneously in the nucleus. The same study on OrNV [52] and an older study on HzNV-2 [64] provide further implications of how HzNV-1 virions might egress from the host cell. Velamoor *et al*. showed that OrNV virions were released from the nucleus into the cytoplasm either via single or multiple membrane vesicles (MMVs) containing a number of virions. These virion-filled vesicles were then observed fusing with the plasma membrane, leading to the release of free virions when delivered as single membrane vesicles to the cell surface, or as virion-filled vesicles when delivered as MMVs into the extracellular space. High quantities of virion-filled vesicles have also been observed in the "waxy plug" of HzNV-2-infected *H. zea* adults. This waxy droplet, which forms at the genital opening of infected female adults, serves as the primary carrier for HzNV-2 transmission during mating [64]. Thus, I assume that HzNV-1 also releases a high proportion of its virions via vesicles, but the release of vesicle-free virions as a consequence of cell lysis is also probable. These virion release phenomena were not studied in this thesis.

The mystery of nudiviral endogenisation

In addition to affecting the protein-folding machinery and various metabolic pathways, the data presented in **Chapter 5** clearly indicated that HzNV-1 infection impairs host nuclear integrity and triggers a DNA damage response. Those findings are particularly noteworthy within the context of nudiviral endogenisation events, as the mechanisms that drive their integration into host genomes remain to be fully understood. As mentioned in **Chapter 3**, one potential mechanism for nudiviral integration could be an indirect consequence of DNA damage and repair induced during viral replication within the host cell [209], which aligns with our findings from **Chapter 5**. Interestingly, baculoviruses are also known to induce DNA damage during infection [450], but there are currently no known cases of endogenised baculoviruses. In **Chapter 2**, we reviewed a major distinction between nudiviruses and baculoviruses, which is the presence of the nudiviral core gene, *integrase*. While this integrase-encoding gene is present in all known free-living nudiviruses and in bracoviruses, it is absent in baculoviruses. For bracoviruses, the nudiviral integrase plays a crucial role in excising proviral segments from the host genome [76, 451] and/or in circularizing amplified bracoviral DNA [109]. The research presented in **Chapter 6** demonstrated that each Toxoneuron nigriceps bracovirus proviral segment functions as an independent replication unit (RU), typically delineated by direct repeat junction (DRJ)-like motifs, indicating that these segments are excised individually from the wasp genome and then replicated, presumably through rolling circle amplification. In contrast, multiple RUs of Cotesia congregata bracoviruses (CcBV) are initially amplified as large concatenated DNA molecules before being processed at their DRJs to form DNA circles. The role of the nudiviral integrase in the DNA replication of bracoviruses aligns with our **Chapter 2** findings that endogenous alphanudiviruses, such as Venturia canescens endogenous nudivirus (VcENV) and Fopius arisanus endogenous nudivirus (FaENV), lack the nudiviral integrase and produce nucleic acid-empty particles. Next to DNA damage repair associated events, these insights suggest a second possible mechanism for nudiviral endogenisation, involving an active integration process in which the nudiviral integrase facilitates stable integration of nudiviral genetic material into the host genome, potentially as a strategy for establishing persistent infections. I hypothesise that in nudivirus infections a combination of both DNA damage responses and the activity of nudiviral proteins with integrase or recombinase functions favour nudiviral endogenisation. Another aspect that may contribute to the relatively high frequency of nudiviral integration events, especially

when compared to baculoviruses, is the lower host mortality associated with nudiviral infections. The lower lethality of nudiviruses may allow a larger number of infected hosts to survive, creating selective pressure for nudiviruses to persist in the host through different measures – for instance, by utilizing the viral *integrase* gene for endogenisation. In contrast, the higher lethality of baculovirus infections might have removed the evolutionary need for *integrase*, resulting in its absence. This hypothesis would imply either that the common ancestor of nudiviruses and baculoviruses possessed the *integrase* gene, which was retained in nudiviruses as part of their adaptation strategy for persistence in the host genome, or that the common ancestor lacked the *integrase* gene, and nudiviruses acquired it later as part of their evolutionary path toward endogenisation.

Future directions and conclusion

Despite the many new discoveries described in this thesis, I believe we merely scratched the surface of the "nudiviral iceberg," and there are still many open questions and future directions left to explore. For instance, what is the true diversity of nudiviruses, and where and how should researchers continue searching for them? Fleas and lice offer intriguing candidates for studies, given the widespread habitation of ectoparasitic species that infest warm-blooded hosts across diverse environments. In particular, lice from aquatic hosts (Echinophthiriidae) may add invaluable information to the evolutionary history of *Nudiviridae*. Based on the findings in **Chapter 3**, discovering nudiviruses in marine lice could help to bridge the missing link between crustacean-infecting nudiviruses and how they emerged from an insect-infecting ancestor. Of course, also non-parasitic insects with aquatic life stages should be considered in this search for the missing link in nudivirus evolution. A fascinating thought that crossed my mind is the hypothetical presence of nudiviruses during the time of the "Black Death" from 1347 to 1351 [452]. The Black Death was one of the most devastating pandemics in human history, caused by *Yersinia pestis* [453]. The transmission of this bacterium to humans occurred mainly through bites from *Xenopsylla cheopis* (Oriental rat flea) [454], and various theories exist about how this pandemic ended so abruptly [455]. However, I could not find a single theory in literature that postulated the possibility of a naturally occurring virus in those past flea populations, with potential effect on the transmission of the bacterial pathogen. Considering that *Y. pestis* accumulates in the digestive tract of fleas [456], and nudiviruses commonly target the midgut of their insect hosts (as reviewed in **Chapter 2**), an hypothetical encounter of *Y. pestis* and a nudivirus in the flea's gut might have resulted in an antagonistic interaction that impaired the transmission of *Y. pestis*. With this in mind, we did not detect any nudivirus-like sequences in particular *X. cheopis* SRA datasets (**Chapter 3**), but virologists should still consider the potential presence of nudiviruses within this flea species as a subject for future endeavours.

Additionally, new nudiviruses might be hiding in hosts outside the insect and crustacean realm – for instance, in arthropods with close ecological relationships to insects – such as arachnids. Especially the discovery of new nudiviruses in arthropods with importance for society and economy – like ticks, mosquitoes, or parasitic insects that affect beneficial species – may expand our repertoire for biocontrol in agriculture or even the medical and veterinary sector. Revisiting older research could also be valuable, particularly studies that used electron microscopy and observed rod-shaped viruses in arthropods but lacked genetic sequencing data.

The discovery of further members within the monophyletic group *Nudiviridae*, in which I include bracoviruses, will naturally add resolution to phylogenetic and evolutionary analyses. However, the speed of new discoveries also makes it challenging for biological studies to keep up, which are essential to understand the possible pathogenic impact of these viruses, as well as their ecological meaning and biocontrol potential. A key gap in nudiviral research remains the functional analysis of nudiviral genes,

especially those unique to nudiviruses. While one may assume that the functions of genes shared between lefavirals are likely conserved, and can be inferred from baculovirus studies, there might still be differences as a consequence of deviating virus host interactions. Specific molecular biological tools would be very helpful to investigate the functions of other nudiviral genes. For instance, a bacmid derived from a nudivirus genome, similar to those developed for baculoviruses [457], would be invaluable for nudiviral gene function studies, enabling gene knockouts, mutational studies or the insertion of visual markers, such as the gene for the green fluorescent protein (GFP), for visualising virus localisation and infection progress. Until such tools are available, RNA interference (RNAi) remains an efficient, though more labour-intensive and technically demanding, approach for studying nudiviral gene functions,

Now that I am reaching the final paragraph of my thesis, I cannot help but reminisce about the beginnings of nudivirus research and the progress that has been made in this field, especially during my time as a PhD candidate. Although the initial project proposal for my thesis looked quite different from the final result, our findings surely serve as valuable jigsaw pieces in the complex puzzle of nudiviruses. My thesis focused primarily on fundamental research questions, yet it has been rewarding to delve into the complexities of nudiviruses and help to promote their visibility in the scientific community. As Edward Appleton once said [458], “The history of science has proved that fundamental research is the lifeblood of individual progress and that the ideas which lead to spectacular advances spring from it.” Looking back on my journey as an early-stage researcher in the INSECT DOCTORS program, this quote resonated strongly with me. I learned that unexpected outcomes can promote personal growth, and I experienced the joy of discovering new things just for the sake of discovery. As the collective nudivirus nexus expands, I am excited to see where future researchers will take the field. There is much left to discover, and I trust that my research will be important in advancing this field.

Acknowledgements

I would like to express my sincere gratitude to my promotors, Monique van Oers and Jean-Michel Drezen, for their constructive remarks and feedback on this chapter.

“**If children don’t grow up knowing about nature and appreciating it, they will not understand it. And if they don’t understand it, they won’t protect it.**”

David Attenborough

Appendices

Reference list

Summary

Acknowledgements

About the author

List of publications

Reference list

1. Boudin M, Green H. On the antagonism of miasmatic fever and pulmonary consumption, and the alleged incompatibility of ague or its causes with pulmonary consumption. Edinb Med Surg J. 1849; 71 (179): 344.
2. Kannadan A. History of the miasma theory of disease. Essai. 2018; 16 (1): 18. DOI:
3. Karamanou M, Panayiotakopoulos G, Tsoucalas G, Kousoulis AA, Androutsos G. From miasmas to germs: a historical approach to theories of infectious disease transmission. Infez Med. 2012; 20 (1): 58-62.
4. Smeele W. Grounding Miasma, or Anticipating the Germ Theory of Disease in Victorian Cholera Satire. J Midwest Mod Lang. 2016; 49 (2): 15-27. DOI: 10.1353/mml.2016.0046
5. Pasteur L, Ernst H. Germ theory and its applications to medicine and surgery: BiblioBytes; 1995.
6. Koch R. Die Ätiologie der Milzbrand-Krankheit, begründet auf die Entwicklungsgeschichte des Bacillus Anthracis (1876): Barth; 1910.
7. Koch R. The etiology of tuberculosis. Rev Infect Dis. 1982; 4 (6): 1270-4.
8. Ivanowski D. Concerning the mosaic disease of the tobacco plant. 1942.
9. Beijerinck MW. Concerning a contagium vivum fluidum as a cause of the spot-disease of Tobacco leaves. 1942.
10. Creager ANH. Tobacco Mosaic Virus and the History of Molecular Biology. Annu Rev Virol. 2022; 9 (1): 39-55. DOI: 10.1146/annurev-virology-100520-014520
11. Ruska E. The development of the electron microscope and of electron microscopy. Biosci Rep. 1987; 59 (3): 627. DOI: 10.1007/BF01127674
12. Murphy FA, Fauquet CM, Bishop DH, Ghabrial SA, Jarvis AW*, et al.* Introduction to the universal system of virus taxonomy. Virus Taxonomy: Classification and Nomenclature of Viruses Sixth Report of the International Committee on Taxonomy of Viruses. 1995: 1-13.
13. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977; 74 (12): 5463-7. DOI: 10.1073/pnas.74.12.5463
14. Lauber C, Seitz S. Opportunities and challenges of data-driven virus discovery. Biomolecules. 2022; 12 (8): 1073. DOI: 10.3390/biom12081073
15. Breitbart M. Marine viruses: truth or dare. Ann Rev Mar Sci. 2012; 4: 425-48. DOI: 10.1146/annurev-marine-120709-142805
16. Kuzyakov Y, Mason-Jones K. Viruses in soil: Nano-scale undead drivers of microbial life, biogeochemical turnover and ecosystem functions. Soil Biol Biochem. 2018; 127: 305-17. DOI: 10.1016/j.soilbio.2018.09.032
17. Dennehy JJ. What ecologists can tell virologists. Annu Rev Microbiol. 2014; 68: 117-35. DOI: 10.1146/annurev-micro-091313-103436
18. Raoult D, Forterre P. Redefining viruses: lessons from Mimivirus. Nat Rev Microbiol. 2008; 6 (4): 315-9. DOI: 10.1038/nrmicro1858
19. Hatfull GF, Dedrick RM, Schooley RT. Phage therapy for antibiotic-resistant bacterial infections. Annu Rev Med. 2022; 73: 197-211. DOI: 10.1146/annurev-med-080219-122208
20. Chroboczek J, Szurgot I, Szolajska E. Virus-like particles as vaccine. Acta Biochimica Polonica. 2014; 61 (3): 531-9.
21. Trombetta CM, Marchi S, Montomoli E. The baculovirus expression vector system: a modern technology for the future of influenza vaccine manufacturing. Expert Rev Vaccines. 2022; 21 (9): 1233-42. DOI: 10.1080/14760584.2022.2085565
22. van Oosten L. Innovations in virus-like particle vaccine production with the baculovirus expressionsystem: Wageningen University; 2024.
23. Korn J, Schackermann D, Kirmann T, Bertoglio F, Steinke S*, et al.* Baculovirus-free insect cell expression system for high yield antibody and antigen production. Sci Rep. 2020; 10 (1): 21393. DOI: 10.1038/s41598-020-78425-9
24. Hong Q, Liu J, Wei Y, Wei X. Application of Baculovirus expression vector system (BEVS) in vaccine development. Vaccines. 2023; 11 (7): 1218. DOI: 10.3390/vaccines11071218
25. El-Helaly A, Khattab M, El-Salamouny S, El-Sheikh M, Elnagar S. Promising additives to protect the activity of baculovirus biocontrol agent under field-sunlight conditions in Egypt. Journal of Life Sciences. 2013; 7 (5): 495.
26. Wagemans J, Holtappels D, Vainio E, Rabiey M, Marzachì C*, et al.* Going viral: virus-based biological control agents for plant protection. Annu Rev Phytopathol. 2022; 60: 21-42. DOI: 10.1146/annurev-phyto-021621-114208
27. Abd-Alla AM, Meki IK, Demirbas-Uzel G. Insect viruses as biocontrol agents: Challenges and opportunities. Cottage industry of biocontrol agents and their applications: Practical aspects to deal

biologically with pests and stresses facing strategic crops. 2020: 277-95. DOI: 10.1007/978-3-030-33161-0

28. Boucias D, Maruniak J, Pendland J. Characterization of a non-occluded baculovirus (subgroup C) from the field cricket, Gryllus rubens. Arch Virol. 1989; 106 (1-2): 93-102. DOI: 10.1007/BF01311041

29. Kelly DC, Lescott T, Ayres MD, Carey D, Coutts A*, et al.* Induction of a nonoccluded baculovirus persistently infecting *Heliothis zea* cells by Heliothis armigera and Trichoplusia ni nuclear polyhedrosis viruses. Virology. 1981; 112 (1): 174-89. DOI: 10.1016/0042-6822(81)90623-1

30. Crawford AM, Sheehan C. Replication of Oryctes Baculovirus in Cell Culture: Viral Morphogenesis, Infectivity and Protein Synthesis. J Gen Virol. 1985; 66 (3): 529-39. DOI: 10.1099/0022-1317-66-3-529

31. Johnson PT. A viral disease of the blue crab, Callinectes sapidus: histopathology and differential diagnosis. Journal of Invertebrate Pathology. 1977; 29 (2): 201-9.

32. ICTV Virus Taxonomy Profile: Nudiviridae [Internet]. 2021. Available from: https://talk.ictvonline.org/ictv-reports/ictv_online_report/dsdna-viruses/w/nudiviridae.

33. Vaughn J, Goodwin R, Tompkins G, McCawley P. The establishment of two cell lines from the insect Spodoptera frugiperda (Lepidoptera; Noctuidae). In vitro. 1977; 13 (4): 213-7. DOI: 10.1007/BF02615077

34. McIntosh AH, Ignoffo CM. Characterization of five cell lines established from species of *Heliothis*. Appl Entomol Zool. 1983; 18 (2): 262-9. DOI: 10.1303/aez.18.262

35. Guinet B, Leobold M, Herniou EA, Bloin P, Burlet N*, et al.* A novel and diverse family of filamentous DNA viruses associated with parasitic wasps. Virus Evol. 2024. DOI: 10.1093/ve/veae022

36. van Oers MM, Herniou EA, Jehle JA, Krell PJ, Abd-Alla AM*, et al.* Developments in the classification and nomenclature of arthropod-infecting large DNA viruses that contain pif genes. Arch Virol. 2023; 168 (7): 182. DOI: 10.1007/s00705-023-05793-8

37. Adams MJ, Lefkowitz EJ, King AM, Carstens EB. Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2014). Arch Virol. 2014; 159 (10): 2831-41. DOI: 10.1007/s00705-014-2114-3

38. Bedford GO. Advances in the control of rhinoceros beetle, Oryctes rhinoceros in oil palm. J Oil Palm Res. 2014; 26 (3): 183-94.

39. Cheng RL, Xi Y, Lou YH, Wang Z, Xu JY*, et al.* Brown planthopper nudivirus DNA integrated in its host genome. J Virol. 2014; 88 (10): 5310-8. DOI: 10.1128/JVI.03166-13

40. Bézier A, Theze J, Gavory F, Gaillard J, Poulain J*, et al.* The genome of the nucleopolyhedrosis-causing virus from *Tipula oleracea* sheds new light on the *Nudiviridae* family. J Virol. 2015; 89 (6): 3008-25. DOI: 10.1128/JVI.02884-14

41. Yang Y-T, Lee D-Y, Wang Y, Hu J-M, Li W-H*, et al.* The genome and occlusion bodies of marine *Penaeus monodon* nudivirus (PmNV, also known as MBV and PemoNPV) suggest that it should be assigned to a new nudivirus genus that is distinct from the terrestrial nudiviruses. BMC genomics. 2014; 15 (1): 628. DOI: 10.1186/1471-2164-15-628

42. Eilenberg J, Vlak JM, Nielsen-LeRoux C, Cappellozza S, Jensen AB. Diseases in insects produced for food and feed. J Insects Food Feed. 2015; 1 (2): 87-102. DOI: 10.3920/JIFF2014.0022

43. Huger AM. A new virus disease of crickets (Orthoptera: Gryllidae) causing macronucleosis of fatbody. J Invertebr Pathol. 1985; 45 (1): 108-11. DOI: 10.1016/0022-2011(85)90055-2

44. Huger AM. A virus disease of the Indian rhinoceros beetle, *Oryctes rhinoceros* (Linnaeus), caused by a new type of insect virus, Rhabdionvirus oryctes gen. n., sp. n. J Invertebr Pathol. 1966; 8 (1): 38-51. DOI: 10.1016/0022-2011(66)90101-7

45. Zelazny B. Studies on Rhabdionvirus oryctes: I. Effect on larvae of *Oryctes rhinoceros* and inactivation of the virus. J Invertebr Pathol. 1972; 20 (3): 235-41. DOI: 10.1016/0022-2011(72)90150-4

46. Deyrup ST, Stagnitti NC, Perpetua MJ, Wong-Deyrup SW. Drug Discovery Insights from Medicinal Beetles in Traditional Chinese Medicine. Biomol Ther. 2021; 29 (2): 105. DOI: 10.4062/biomolther.2020.229

47. Kwak KW, Choi HS, Nam SH, Han MS, Kim E*, et al.* The first report on the characterization of a virus isolated from Allomyrina dichotoma in the Republic of Korea. Entomol Res. 2019; 49 (12): 534-8. DOI: 10.1111/1748-5967.12401

48. Lee S, Kim H-G, Park K-h, Nam S-h, Kwak K-w*, et al.*, editors. Current status of viral disease in Korean horn beetle, Allomyrina dichotoma (Coleoptera: Scarabaeidae). 2015 한국응용곤충학회 임시총회 및 추계학술발표회; 2015.

49. Ros VID. Baculoviruses: General Features (*Baculoviridae*). In: Zuckerman M, Bamford DH, editors. Reference Module in Life Sciences. 1-5: Elsevier; 2020. p. 739-46.

50. van Oers MM. Baculoviruses: Molecular Biology and Replication (Baculoviridae). Encyclopedia of Virology2021. p. 747-58.

51. Create two new genera and eight new species (*Lefavirales*: *Nudiviridae*) [Internet]. 2020. Available from: https://ictv.global/ictv/proposals/2020.005D.R.Nudiviridae_2ngen_8nsp.zip. .

52. Velamoor S, Mitchell A, Humbel BM, Kim W, Pushparajan C*, et al.* Visualizing Nudivirus Assembly and Egress. mBio. 2020; 11 (4): e01333-20. DOI: 10.1128/mBio.01333-20

53. Harrison RL, Herniou EA, Bezier A, Jehle JA, Burand JP*, et al.* ICTV Virus Taxonomy Profile: *Nudiviridae*. J Gen Virol. 2020; 101 (1): 3-4. DOI: 10.1099/jgv.0.001381

54. Allain TW, Stentiford GD, Bass D, Behringer DC, Bojko J. A novel nudivirus infecting the invasive demon shrimp *Dikerogammarus haemobaphes* (Amphipoda). Sci Rep. 2020; 10 (1): 1-13. DOI: 10.1038/s41598-020-71776-3

55. Wang Y, Jehle JA. Nudiviruses and other large, double-stranded circular DNA viruses of invertebrates: new insights on an old topic. J Invertebr Pathol. 2009; 101 (3): 187-93. DOI: 10.1016/j.jip.2009.03.013

56. Zhao S, He G, Yang Y, Liang C. Nucleocapsid assembly of baculoviruses. Viruses. 2019; 11 (7). DOI: 10.3390/v11070595

57. Burand JP. Nudiviruses. The Insect Viruses: Springer; 1998. p. 69-90.

58. Long G, Pan X, Kormelink R, Vlak JM. Functional entry of baculovirus into insect and mammalian cells is dependent on clathrin-mediated endocytosis. J Virol. 2006; 80 (17): 8830-3. DOI: 10.1128/jvi.00880-06

59. Wang Y, Kleespies RG, Huger AM, Jehle JA. The genome of Gryllus bimaculatus nudivirus indicates an ancient diversification of baculovirus-related nonoccluded nudiviruses of insects. J Virol. 2007; 81 (10): 5395-406. DOI: 10.1128/JVI.02781-06

60. Boogaard B, van Oers MM, van Lent JWM. An Advanced View on Baculovirus *per os* Infectivity Factors. Insects. 2018; 9 (3): 84. DOI: 10.3390/insects9030084

61. Wang X, Shang Y, Chen C, Liu S, Chang M*, et al.* Baculovirus per os infectivity factor complex: components and assembly. J Virol. 2019; 93 (6): e02053-18. DOI: 10.1128/jvi.02053-18

62. Horton HM, Burand J. Saturable attachment sites for polyhedron-derived baculovirus on insect cells and evidence for entry via direct membrane fusion. J Virol. 1993; 67 (4): 1860-8. DOI: 10.1128/jvi.67.4.1860-1868.1993

63. Holter H. Pinocytosis. International review of cytology. 8: Elsevier; 1959. p. 481-504.

64. Hamm JJ, Carpenter JE, Styer EL. Oviposition day effect on incidence of agonadal progeny of Helicoverpa zea (Lepidoptera: Noctuidae) infected with a virus. Ann Entomol Soc Am. 1996; 89 (2): 266-75. DOI: 10.1093/aesa/89.2.266

65. Lin C-L, Lee J-C, Chen S-S, Wood HA, Li M-L*, et al.* Persistent Hz-1 virus infection in insect cells: evidence for insertion of viral DNA into host chromosomes and viral infection in a latent status. J Virol. 1999; 73 (1): 128-39. DOI: Doi 10.1128/Jvi.73.1.128-139.1999

66. Burke GR, Simmonds TJ, Sharanowski BJ, Geib SM. Rapid Viral Symbiogenesis via Changes in Parasitoid Wasp Genome Architecture. Mol Biol Evol. 2018; 35 (10): 2463-74. DOI: 10.1093/molbev/msy148

67. Cheng RL, Li XF, Zhang CX. Nudivirus Remnants in the Genomes of Arthropods. Genome Biol Evol. 2020; 12 (5): 578-88. DOI: 10.1093/gbe/evaa074

68. Leobold M, Bézier A, Pichon A, Herniou EA, Volkoff A-N*, et al.* The Domestication of a large DNA virus by the wasp *Venturia canescens* involves targeted genome reduction through pseudogenization. Genome Biol Evol. 2018; 10 (7): 1745-64. DOI: 10.1093/gbe/evy127

69. Liu S, Coates BS, Bonning BC. Endogenous viral elements integrated into the genome of the soybean aphid, Aphis glycines. Insect Biochem Mol Biol. 2020; 123: 103405. DOI: 10.1016/j.ibmb.2020.103405

70. Zhang Y, Wang J, Han G-Z. Chalcid wasp paleoviruses bridge the evolutionary gap between bracoviruses and nudiviruses. Virology. 2020; 542: 34-9. DOI: 10.1016/j.virol.2020.01.007

71. Drezen JM, Josse T, Bézier A, Gauthier J, Huguet E*, et al.* Impact of Lateral Transfers on the Genomes of Lepidoptera. Genes (Basel). 2017; 8 (11). DOI: 10.3390/genes8110315

72. Pichon A, Bézier A, Urbach S, Aury J-M, Jouan V*, et al.* Recurrent DNA virus domestication leading to different parasite virulence strategies. Sci Adv. 2015; 1 (10): e1501150. DOI: 10.1126/sciadv.1501150

73. Volkoff AN, Cusson M. The Unconventional Viruses of Ichneumonid Parasitoid Wasps. Viruses. 2020; 12 (10). DOI: 10.3390/v12101170

74. Bézier A, Annaheim M, Herbiniere J, Wetterwald C, Gyapay G*, et al.* Polydnaviruses of braconid wasps derive from an ancestral nudivirus. Science. 2009; 323 (5916): 926-30. DOI: 10.1126/science.1166788

75. Bézier A, Louis F, Jancek S, Periquet G, Thézé J*, et al.* Functional endogenous viral elements in the genome of the parasitoid wasp *Cotesia congregata*: insights into the evolutionary dynamics of bracoviruses. Philosophical Transactions of the Royal Society B: Biological Sciences. 2013; 368 (1626): 20130047. DOI: 10.1098/rstb.2013.0047

76. Burke GR, Thomas SA, Eum JH, Strand MR. Mutualistic polydnaviruses share essential replication gene functions with pathogenic ancestors. PLoS Pathog. 2013; 9 (5): e1003348. DOI: 10.1371/journal.ppat.1003348

77. Dupuy C, Gundersen-rindal D, Cusson M. Genomics and Replication of Polydnaviruses. Parasitoid Viruses: Symbionts and Pathogens. 2012: 47-61. DOI:

78. Belle E, Beckage NE, Rousselet J, Poirié M, Lemeunier F*, et al.* Visualization of polydnavirus sequences in a parasitoid wasp chromosome. J Virol. 2002; 76 (11): 5793-6. DOI: 10.1128/jvi.76.11.5793-5796.2002

79. Bezier A, Herbiniere J, Lanzrein B, Drezen JM. Polydnavirus hidden face: the genes producing virus particles of parasitic wasps. J Invertebr Pathol. 2009; 101 (3): 194-203. DOI: 10.1016/j.jip.2009.04.006

80. Kuhn J, Dolja V, Krupovic M, Adriaenssens E, Di Serio F*, et al.* Expand, amend, and emend the International Code of Virus Classification and Nomenclature (ICVCN;"the Code") and the statutes to clearly define the remit of the ICTV. International Committee for Taxonomy of Viruses proposal (Taxoprop) number. 2020. DOI:

81. Drezen J-M, Bézier A, Burke GR, Strand MR. Bracoviruses, ichnoviruses, and virus-like particles from parasitoid wasps retain many features of their virus ancestors. Curr Opin Insect Sci. 2022; 49: 93-100. DOI: 10.1016/j.cois.2021.12.003

82. Gauthier J, Boulain H, van Vugt JJ, Baudry L, Persyn E*, et al.* Chromosomal scale assembly of parasitic wasp genome reveals symbiotic virus colonization. Commun Biol. 2021; 4 (1): 1-15. DOI: 10.1038/s42003-021-02480-9

83. Bateman KS, Kerr R, Stentiford GD, Bean TP, Hooper C*, et al.* Identification and Full Characterisation of Two Novel Crustacean Infecting Members of the Family Nudiviridae Provides Support for Two Subfamilies. Viruses. 2021; 13 (9): 1694. DOI: 10.3390/v13091694

84. Burke GR. Common themes in three independently derived endogenous nudivirus elements in parasitoid wasps. Curr Opin Insect Sci. 2019; 32: 28-35. DOI: 10.1016/j.cois.2018.10.005

85. Wang Y, Bininda-Emonds OR, van Oers MM, Vlak JM, Jehle JA. The genome of Oryctes rhinoceros nudivirus provides novel insight into the evolution of nuclear arthropod-specific large circular double-stranded DNA viruses. Virus genes. 2011; 42 (3): 444-56. DOI: 10.1007/s11262-011-0589-5

86. Boogaard B, van Lent JW, van Oers MM. Functional analysis of the baculovirus per os infectivity factors 3 and 9 by imaging the interaction between fluorescently labelled virions and isolated midgut cells. J Gen Virol. 2020; 101 (7): 778-84. DOI: 10.1099/jgv.0.001430

87. Haas-Stapleton EJ, Washburn JO, Volkman LE. P74 mediates specific binding of *Autographa californica* M nucleopolyhedrovirus occlusion-derived virus to primary cellular targets in the midgut epithelia of *Heliothis virescens* larvae. J Virol. 2004; 78 (13): 6786-91. DOI: 10.1128/JVI.78.13.6786-6791.2004

88. Ohkawa T, Washburn JO, Sitapara R, Sid E, Volkman LE. Specific binding of *Autographa californica* M nucleopolyhedrovirus occlusion-derived virus to midgut cells of *Heliothis virescens l*arvae is mediated by products of pif genes Ac119 and Ac022 but not by Ac115. Journal of Virology. 2005; 79 (24): 15258-64.

89. Boogaard B, van Lent JWM, Theilmann DA, Erlandson MA, van Oers MM. Baculoviruses require an intact ODV entry-complex to resist proteolytic degradation of per os infectivity factors by co-occluded proteases from the larval host. J Gen Virol. 2017; 98 (12): 3101-10. DOI: 10.1099/jgv.0.000974

90. Jiantao L, Zhu L, Zhang S, Deng Z, Huang Z*, et al.* The Autographa californica multiple nucleopolyhedrovirus ac110 gene encodes a new *per os* infectivity factor. Virus Res. 2016; 221: 30-7. DOI: 10.1016/j.virusres.2016.05.017

91. Abd-Alla AM, Cousserans F, Parker AG, Jehle JA, Parker NJ*, et al.* Genome analysis of a Glossina pallidipes salivary gland hypertrophy virus reveals a novel, large, double-stranded circular DNA virus. J Virol. 2008; 82 (9): 4595-611. DOI: 10.1128/JVI.02588-07

92. Holt CC, Stone M, Bass D, Bateman KS, van Aerle R*, et al.* The first clawed lobster virus *Homarus gammarus* nudivirus (HgNV n. sp.) expands the diversity of the *Nudiviridae*. Sci Rep. 2019; 9 (1): 1-15. DOI: 10.1038/s41598-019-46008-y

93. Wetterwald C, Roth T, Kaeslin M, Annaheim M, Wespi G*, et al.* Identification of bracovirus particle proteins and analysis of their transcript levels at the stage of virion formation. J Gen Virol. 2010; 91 (10): 2610-9. DOI: 10.1099/vir.0.022699-0

94. Wu W, Lin T, Pan L, Yu M, Li Z*, et al.* Autographa californica multiple nucleopolyhedrovirus nucleocapsid assembly is interrupted upon deletion of the 38K gene. J Virol. 2006; 80 (23): 11475-85. DOI: 10.1128/JVI.01155-06

95. Wu W, Passarelli AL. *Autographa californica* multiple nucleopolyhedrovirus Ac92 (ORF92, P33) is required for budded virus production and multiply enveloped occlusion-derived virus formation. J Virol. 2010; 84 (23): 12351-61. DOI: 10.1128/Jvi.01598-10

96. Wang M, Tuladhar E, Shen S, Wang H, van Oers MM*, et al.* Specificity of baculovirus P6.9 basic DNA-binding proteins and critical role of the C terminus in virion formation. J Virol. 2010; 84 (17): 8821-8. DOI: 10.1128/JVI.00072-10

97. Vanarsdall AL, Okano K, Rohrmann GF. Characterization of the role of very late expression factor 1 in baculovirus capsid structure and DNA processing. J Virol. 2006; 80 (4): 1724-33. DOI: 10.1128/jvi.80.4.1724-1733.2006

98. Hill T, Unckless RL. The dynamic evolution of Drosophila innubila Nudivirus. Infect Genet Evol. 2018; 57: 151-7. DOI: 10.1016/j.meegid.2017.11.013

99. Argaud O, Croizier L, L√≥pez-Ferber M, Croizier G. Two key mutations in the host-range specificity domain of the *p143* gene of *Autographa californica* nucleopolyhedrovirus are required to kill *Bombyx mori* larvae. J Gen Virol. 1998; 79 (4): 931-5. DOI: 10.1099/0022-1317-79-4-931

100. Croizier G, Croizier L, Argaud O, Poudevigne D. Extension of *Autographa californica* nuclear polyhedrosis virus host range by interspecific replacement of a short DNA sequence in the *p143* helicase gene. Proc Natl Acad Sci U S A. 1994; 91 (1): 48-52. DOI: 10.1073/pnas.91.1.48

101. Rohrmann GF. Baculovirus molecular biology. 2019.

102. Jehle JA, Abd-Alla AM, Wang Y. Phylogeny and evolution of Hytrosaviridae. J Invertebr Pathol. 2013; 112: S62-S7. DOI: 10.1016/j.jip.2012.07.015

103. Pearson MN, Rohrmann GF. Lymantria dispar nuclear polyhedrosis virus homologous regions: characterization of their ability to function as replication origins. J Virol. 1995; 69 (1): 213-21. DOI: 10.1128/Jvi.69.1.213-221.1995

104. Guarino LA, Summers MD. Interspersed homologous DNA of Autographa californica nuclear polyhedrosis virus enhances delayed-early gene expression. J Virol. 1986; 60 (1): 215-23. DOI: 10.1128/JVI.60.1.215-223.1986

105. Garcia-Maruniak A, Maruniak JE, Farmerie W, Boucias DG. Sequence analysis of a non-classified, non-occluded DNA virus that causes salivary gland hypertrophy of Musca domestica, MdSGHV. Virology. 2008; 377 (1): 184-96. DOI: 10.1016/j.virol.2008.04.010

106. Louis F, Bézier A, Periquet G, Ferras C, Drezen J-M*, et al.* The bracovirus genome of the parasitoid wasp *Cotesia congregata* is amplified within 13 replication units, including sequences not packaged in the particles. J Virol. 2013; 87 (17): 9649-60. DOI: 10.1128/jvi.00886-13

107. Gundersen-Rindal D, Dupuy C, Huguet E, Drezen J-M. Parasitoid polydnaviruses: evolution, pathology and applications: Dedicated to the memory of Nancy E. Beckage. Biocontrol Science and Technology. 2013; 23 (1): 1-61.

108. Drezen JM, Bézier A, Lesobre J, Huguet E, Cattolico L*, et al.* The few virus-like genes of Cotesia congregata bracovirus. Arch Insect Biochem Physiol. 2006; 61 (3): 110-22. DOI: 10.1002/arch.20108

109. Burke GR, Simmonds TJ, Thomas SA, Strand MR. *Microplitis demolitor* bracovirus proviral loci and clustered replication genes exhibit distinct DNA amplification patterns during replication. J Virol. 2015; 89 (18): 9511-23. DOI: 10.1128/JVI.01388-15

110. Burke GR, Strand MR. Deep sequencing identifies viral and wasp genes with potential roles in replication of Microplitis demolitor bracovirus. J Virol. 2012; 86 (6): 3293-306. DOI: 10.1128/JVI.06434-11

111. Francino MP. An adaptive radiation model for the origin of new gene functions. Nat Genet. 2005; 37 (6): 573-7. DOI: 10.1038/ng1579

112. Van Eynde B, Christiaens O, Delbare D, Shi C, Vanhulle E*, et al.* Exploration of the virome of the European brown shrimp (*Crangon crangon*). J Gen Virol. 2020: jgv001412. DOI: 10.1099/jgv.0.001412

113. Liu S, Sappington TW, Coates BS, Bonning BC. Nudivirus Sequences Identified from the Southern and Western Corn Rootworms (Coleoptera: Chrysomelidae). Viruses. 2021; 13 (2): 269. DOI: 10.3390/v13020269

114. Thézé J, Bézier A, Periquet G, Drezen J-M, Herniou EA. Paleozoic origin of insect large dsDNA viruses. Proc Natl Acad Sci U S A. 2011; 108 (38): 15931-5. DOI: 10.1073/pnas.1105580108

115. Ikeda M, Hamajima R, Kobayashi M. Baculoviruses: diversity, evolution and manipulation of insects. Entomol Sci. 2015; 18 (1): 1-20.

116. Adler PH, Courtney GW. Ecological and Societal Services of Aquatic Diptera. Insects. 2019; 10 (3): 70. DOI: 10.3390/insects10030070

117. Evenhuis NL. Catalog of the Diptera of the Australasian and Oceanian Regions: Bishop Museum Press; 1989.

118. de Jong H, Oosterbroek P, Gelhaus J, Reusch H, Young C. Global diversity of craneflies (Insecta, Diptera: Tipulidea or Tipulidae sensu lato) in freshwater. Freshwater Animal Diversity Assessment. 2007: 457-67.

119. Pritchard G. Biology of tipulidae. Annu Rev Entomol. 1983; 28 (1): 1-22. DOI: 10.1146/annurev.en.28.010183.000245

120. Foster GN, Bilton DT. The conservation of predaceous diving beetles: knowns, unknowns and anecdotes. Ecology, systematics, and the natural history of predaceous diving beetles (Coleoptera: Dytiscidae): Springer; 2014. p. 437-62.

121. Yee DA. An Introduction to the Dytiscidae: their diversity, historical importance, cultural significance, and other musings. Ecology, Systematics, and the Natural History of Predaceous Diving Beetles (Coleoptera: Dytiscidae): Springer; 2014. p. 1-16.

122. Glenner H, Thomsen PF, Hebsgaard MB, Sørensen MV, Willerslev E. The origin of insects. Science. 2006; 314 (5807): 1883-4. DOI: 10.1126/science.1129844

123. Hill T, Unckless RL. Baculovirus molecular evolution via gene turnover and recurrent positive selection of key genes. J Virol. 2017; 91 (22): e01319-17. DOI: 10.1128/jvi.01319-17

124. Cameron CE, Gotte M, Raney K. Viral Genome Replication: Springer US; 2009.

125. Wang Z, Ye X, Zhou Y, Wu X, Hu R*, et al.* Bracoviruses recruit host integrases for their integration into caterpillar's genome. Plos Genet. 2021; 17 (9): e1009751. DOI: 10.1371/journal.pgen.1009751

126. Gilbert C, Chateigner A, Ernenwein L, Barbe V, Bézier A*, et al.* Population genomics supports baculoviruses as vectors of horizontal transfer of insect transposons. Nat Commun. 2014; 5 (1): 1-9. DOI: 10.1038/ncomms4348

127. Granados RR, Nguyen T, Cato B. An insect cell line persistently infected with a baculovirus-like particle. Intervirology. 1978; 10 (5): 309-17. DOI: 10.1159/000148993

128. Ohkawa T, Volkman LE, Welch MD. Actin-based motility drives baculovirus transit to the nucleus and cell surface. J Cell Biol. 2010; 190 (2): 187-95. DOI: 10.1083/jcb.201001162

129. Au S, Wu W, Pante N. Baculovirus nuclear import: open, nuclear pore complex (NPC) sesame. Viruses. 2013; 5 (7): 1885-900. DOI: 10.3390/v5071885

130. Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J*, et al.* Determining the architectures of macromolecular assemblies. Nature. 2007; 450 (7170): 683-94. DOI: 10.1038/nature06404

131. Jehle JA, Blissard G, Bonning B, Cory J, Herniou E*, et al.* On the classification and nomenclature of baculoviruses: a proposal for revision. Archives of virology. 2006; 151 (7): 1257-66. DOI: 10.1007/s00705-006-0763-6

132. Au S, Panté N. Nuclear transport of baculovirus: revealing the nuclear pore complex passage. J Struct Biol. 2012; 177 (1): 90-8. DOI: 10.1016/j.jsb.2011.11.006

133. Unckless RL. A DNA virus of Drosophila. PloS one. 2011; 6 (10): e26564. DOI: 10.1371/journal.pone.0026564

134. Palmer WH, Medd NC, Beard PM, Obbard DJ. Isolation of a natural DNA virus of Drosophila melanogaster, and characterisation of host resistance and immune responses. PLoS Pathog. 2018; 14 (6): e1007050. DOI: 10.1371/journal.ppat.1007050

135. Webster CL, Waldron FM, Robertson S, Crowson D, Ferrari G*, et al.* The discovery, distribution, and evolution of viruses associated with Drosophila melanogaster. PLoS Biol. 2015; 13 (7): e1002210. DOI: 10.1371/journal.pbio.1002210

136. Kapun M, Barrón MG, Staubach F, Obbard DJ, Wiberg RAW*, et al.* Genomic analysis of European Drosophila melanogaster populations reveals longitudinal structure, continent-wide selection, and previously unknown DNA viruses. Mol Biol Evol. 2020; 37 (9): 2661-78. DOI: 10.1093/molbev/msaa120

137. Wallace MA, Coffman KA, Gilbert C, Ravindran S, Albery GF*, et al.* The discovery, distribution and diversity of DNA viruses associated with Drosophila melanogaster in Europe. bioRxiv. 2020.

138. Etebari K, Parry R, Beltran MJB, Furlong MJ. Genomic structural and transcriptional variation of Oryctes rhinoceros nudivirus (OrNV) in Coconut Rhinoceros Beetle. bioRxiv. 2020: 2020.05. 27.119867. DOI: 10.1128/jvi.01097-20

139. Huger AM. The Oryctes virus: its detection, identification, and implementation in biological control of the coconut palm rhinoceros beetle, Oryctes rhinoceros (Coleoptera: Scarabaeidae). J Invertebr Pathol. 2005; 89 (1): 78-84. DOI: 10.1016/j.jip.2005.02.010

140. Kim K, Kitajima E. Nonoccluded baculovirus-and filamentous virus-like particles in the spotted cucumber beetle, Diabrotica undecimpunctata (coleoptera: chrysomelid). Journal of invertebrate pathology. 1984; 43 (2): 234-41.

141. Burand P. Persistent baculovirus infection. Curr Top Microbiol. 1986; 1: 159-75.

142. Cheng C-H, Liu S-M, Chow T-Y, Hsiao Y-Y, Wang D-P*, et al.* Analysis of the complete genome sequence of the Hz-1 virus suggests that it is related to members of the *Baculoviridae*. J Virol. 2002; 76 (18): 9024-34. DOI: 10.1128/jvi.76.18.9024-9034.2002

143. Langridge W. Biochemical properties of a persistent nonoccluded baculovirus isolated from Heliothis zea cells. virology. 1981; 112 (2): 770-4. DOI: 10.1016/0042-6822(81)90324-x

144. Burand JP, Kim W, Afonso CL, Tulman ER, Kutish GF*, et al.* Analysis of the genome of the sexually transmitted insect virus Helicoverpa zea nudivirus 2. Viruses. 2012; 4 (1): 28-61. DOI: 10.3390/v4010028

145. Raina AK, Adams JR, Lupiani B, Lynn DE, Kim W*, et al.* Further characterization of the gonad-specific virus of corn earworm, *Helicoverpa zea*. J Invertebr Pathol. 2000; 76 (1): 6-12. DOI: 10.1006/jipa.2000.4942

146. Mari J, Bonami J-R, Poulos B, Lightner D. Penaeus monodon (PmSNPV= MBV). Dis Aquat Org. 1993; 16: 207-15.

147. Rajendran KV, Makesh M, Karunasagar I. Monodon baculovirus of shrimp. Indian J Virol. 2012; 23 (2): 149-60. DOI: 10.1007/s13337-012-0086-z

148. Vickers J, Webb R, Young P. Monodon baculovirus from Australia: ultrastructural observations. Dis Aquat Organ. 2000; 39 (3): 169-76. DOI: 10.3354/dao039169

149. Bojko J, Stebbing PD, Dunn AM, Bateman KS, Clark F*, et al.* Green crab Carcinus maenas symbiont profiles along a North Atlantic invasion route. Dis Aquat Organ. 2018; 128 (2): 147-68. DOI: 10.3354/dao03216

150. Mari J, Bonami J. Les infections virales du crabe Carcinus mediterraneus Czerniavski, 1884. Pathology in Marine Aquaculture. 1986: 283-93.

151. Stentiford G, Bateman K, Feist S. Pathology and ultrastructure of an intranuclear bacilliform virus (IBV) infecting brown shrimp Crangon crangon (Decapoda: Crangonidae). Dis Aquat Organ. 2004; 58 (2-3): 89-97. DOI: 10.3354/dao058089

152. Bojko J, Stentiford GD, Stebbing PD, Hassall C, Deacon A*, et al.* Pathogens of Dikerogammarus haemobaphes regulate host activity and survival, but also threaten native amphipod populations in the UK. Dis Aquat Organ. 2019; 136 (1): 63-78. DOI: 10.3354/dao03195

153. Bézier A, Harichaux G, Musset K, Labas V, Herniou EA. Qualitative proteomic analysis of Tipula oleracea nudivirus occlusion bodies. J Gen Virol. 2017; 98 (2): 284-95. DOI: 10.1099/jgv.0.000661

154. Schoonvaere K, Smagghe G, Francis F, de Graaf DC. Study of the metatranscriptome of eight social and solitary wild bee species reveals novel viruses and bee parasites. Front Microbiol. 2018; 9: 177. DOI: 10.3389/fmicb.2018.00177

155. Porter AF, Shi M, Eden JS, Zhang YZ, Holmes EC. Diversity and Evolution of Novel Invertebrate DNA Viruses Revealed by Meta-Transcriptomics. Viruses. 2019; 11 (12). DOI: 10.3390/v11121092

156. Edgerton B, Paasonen P, Henttonen P, Owens L. Description of a bacilliform virus from the freshwater crayfish, Astacus astacus. J Invertebr Pathol. 1996; 68 (2): 187-90. DOI: 10.1006/jipa.1996.0081

157. Edgerton B, Watt H, Becheras JM, Bonami JR. An intranuclear bacilliform virus associated with near extirpation of Austropotamobius pallipes Lereboullet from the Nant watershed in Ardèche, France. J Fish Dis 2002; 25 (9): 523-31. DOI: 10.1046/j.1365-2761.2002.00395.x

158. Sano T, NISHIMURA T, OGUMA K, MOMOYAMA K, TAKENO N. Baculovirus infection of cultured Kuruma shrimp, Penaeus japonicus in Japan. Fish Pathology. 1981; 15 (3-4): 185-91.

159. Johnson PT. The rod-shaped nuclear viruses of crustaceans: gut-infecting species. Dis Aquat Org. 1988; 4: 123-41.

160. Bateman KS, Stentiford GD. Cancer pagurus bacilliform virus (CpBV) infecting juvenile European edible crabs C. pagurus from UK waters. Dis Aquat Organ. 2008; 79 (2): 147-51. DOI: 10.3354/dao01900

161. Edgerton B. A new bacilliform virus in Australian Cherax destructor (Decapoda: Parastacidae) with notes on Cherax quadricarinatus bacilliform virus (= Cherax baculovirus). Diseases of aquatic organisms. 1996; 27 (1): 43-52.

162. Anderson I, Prior H. Baculovirus infections in the mud crab, Scylla serrata, and a freshwater crayfish, Cherax quadricarinatus, from Australia. J Invertebr Pathol. 1992; 60 (3): 265-73. DOI: 10.1016/0022-2011(92)90008-R

163. Bojko J, Bącela-Spychalska K, Stebbing PD, Dunn AM, Grabowski M*, et al.* Parasites, pathogens and commensals in the "low-impact" non-native amphipod host Gammarus roeselii. Parasites & vectors. 2017; 10 (1): 1-15. DOI: 10.1186/s13071-017-2108-6

164. Hedrick R, McDowell T, Friedman C. Baculoviruses found in two species of crayfish from California. Aquaculture. 1995; 95: 135.

165. Bateman KS, Stentiford GD. A taxonomic review of viruses infecting crustaceans with an emphasis on wild hosts. J Invertebr Pathol. 2017; 147: 86-110. DOI: 10.1016/j.jip.2017.01.010

166. Meynardier G, Ricou G, Bergoin M. Virose à corps d'inclusion chez Tipula paludosa (Diptera) en France. Re Pathol Vég Entomol Agric Fr. 1964; 43: 113-8.

167. Smith KM, Xeros N. An unusual virus disease of a dipterous larva. Nature. 1954; 173 (4410): 866-7. DOI: 10.1038/173866a0

168. Koonin EV, Dolja VV, Krupovic M, Varsani A, Wolf YI*, et al.* Global organization and proposed megataxonomy of the virus world. Microbiology and molecular biology reviews. 2020; 84 (2): e00061-19. DOI: 10.1128/MMBR.00061-19

169. van Oers MM, Herniou EA, Jehle JA, Krell PJ, Abd-Alla AM*, et al.* Correction to: Developments in the classification and nomenclature of arthropod-infecting large DNA viruses that contain pif genes. Archives of Virology. 2023; 168 (10): 255.

170. Wang H-C, Hirono I, Maningas MBB, Somboonwiwat K, Stentiford G*, et al.* ICTV virus taxonomy profile: Nimaviridae. J Gen Virol. 2019; 100 (7): 1053-4. DOI: 10.1099/jgv.0.001248

171. Petersen JM, Bézier A, Drezen J-M, van Oers MM. The naked truth: An updated review on nudiviruses and their relationship to bracoviruses and baculoviruses. J Invertebr Pathol. 2022; 189: 107718. DOI: 10.1016/j.jip.2022.107718

172. Hegna TA, Luque J, Wolfe JM. The fossil record of the Pancrustacea. The natural history of Crustacea. 2020; 8: 21-52.

173. Bonami J-R, Aubert H, Mari J, Poulos BT, Lightner DV. The polyhedra of the occluded baculoviruses of marine decapod crustacea: a unique structure, crystal organization, and proposed model. J Struct Biol. 1997; 120 (2): 134-45. DOI: 10.1006/jsbi.1997.3905

174. Groff J, McDowell T, Friedman C, Hedrick R. Detection of a nonoccluded baculovirus in the freshwater crayfish Cherax quadricarinatus in North America. Journal of Aquatic Animal Health. 1993; 5 (4): 275-9.

175. Bojko J, Walters E, Burgess A, Behringer DC. Rediscovering "Baculovirus-A"(Johnson, 1976): The complete genome of 'Callinectes sapidus nudivirus'. J Invertebr Pathol. 2022; 194: 107822. DOI: 10.1016/j.jip.2022.107822

176. Chiu E, Coulibaly F, Metcalf P. Insect virus polyhedra, infectious protein crystals that contain virus particles. Curr Opin Struct Biol. 2012; 22 (2): 234-40. DOI: 10.1016/j.sbi.2012.02.003

177. Francki R. Fifth report of the International Committee on Taxonomy of Viruses. Arch Virol Suppl. 1991; 2: 1-450.

178. Keown JR, Crawshaw AD, Trincao J, Carrique L, Gildea RJ*, et al.* Atomic structure of a nudivirus occlusion body protein determined from a 70-year-old crystal sample. Nat Commun. 2023; 14 (1): 4160. DOI: 10.1038/s41467-023-39819-1

179. Stratton CE, Reisinger LS, Behringer DC, Gray SN, Larson ER*, et al.* North American crayfish harbour diverse members of the *Nudiviridae*. Virology. 2024; 598: 110183. DOI: 10.1016/j.virol.2024.110183

180. Bojko J, Duermit-Moreau E, Gandy R, Behringer DC. A new member of the Nudiviridae from the Florida stone crab (Menippe mercenaria). Virology. 2023; 588: 109910. DOI: 10.1016/j.virol.2023.109910

181. Shen Z, Kumar D, Liu X, Yan B, Fang P*, et al.* Metatranscriptomic Analysis Reveals an Imbalance of Hepatopancreatic Flora of Chinese Mitten Crab Eriocheir sinensis with Hepatopancreatic Necrosis Disease. Biology. 2021; 10 (6): 462. DOI: 10.3390/biology10060462

182. Leinonen R, Sugawara H, Shumway M, Collaboration INSD. The sequence read archive. Nucleic Acids Res. 2010; 39 (suppl_1): D19-D21. DOI: 10.1093/nar/gkq1019

183. Avery S, Jouvenaz D, Banks W, Anthony D. Virus-like particles in a fire ant, Solenopsis sp.,(Hymenoptera: Formicidae) from Brazil. Florida Entomologist. 1977: 17-20. DOI: 10.2307/3494478

184. Gouranton J. Development of an intranuclear nonoccluded rod-shaped virus in some midgut cells of an adult insect, Gyrinus natator L.(Coleoptera). J Ultra Mol Struct R. 1972; 39 (3-4): 281-94. DOI: 10.1016/S0022-5320(72)90023-8

185. Pappalardo R, Mari J, Bonami J-R. τ (tau) virus infection of Carcinus mediterraneus: histology, cytopathology, and experimental transmission of the disease. J Invertebr Pathol. 1986; 47 (3): 361-8. DOI: 10.1016/0022-2011(86)90107-2

186. Owens L, Liessmann L, La Fauce K, Nyguyen T, Zeng C. Intranuclear bacilliform virus and hepatopancreatic parvovirus (PmergDNV) in the mud crab Scylla serrata (Forskal) of Australia. Aquaculture. 2010; 310 (1-2): 47-51. DOI: 10.1016/j.aquaculture.2010.10.028

187. Longshaw M, Feist SW, Bateman KS. Parasites and pathogens of the endosymbiotic pea crab (Pinnotheres pisum) from blue mussels (Mytilus edulis) in England. J Invertebr Pathol. 2012; 109 (2): 235-42. DOI: 10.1016/j.jip.2011.11.011

188. Edgerton BF, Henttonen P, Jussila J, Mannonen A, Paasonen P*, et al.* Understanding the causes of disease in European freshwater crayfish. Conserv Biol. 2004; 18 (6): 1466-74. DOI: 10.1111/j.1523-1739.2004.00436.x

189. Edgerton B, Owens L. Age at first infection of Cherax quadricarinatus by Cherax quadricarinatus bacilliform virus and Cherax giardiavirus-like virus, and production of putative virus-free crayfish. Aquaculture. 1997; 152 (1-4): 1-12. DOI: Doi 10.1016/S0044-8486(97)00006-9

190. Anderson LG, Bojko J, Bateman KS, Stebbing PD, Stentiford GD*, et al.* Patterns of infection in a native and an invasive crayfish across the UK. J Invertebr Pathol. 2021; 184: 107595. DOI: 10.1016/j.jip.2021.107595

191. Takahashi Y, Itami T, Kondo M, Maeda M, Fujii R*, et al.* Electron microscopic evidence of bacilliform virus infection in kuruma shrimp (Penaeus japonicus). Fish Pathology. 1994; 29 (2): 121-5. DOI: 10.3147/jsfp.29.121

192. Couch JA. An enzootic nuclear polyhedrosis virus of pink shrimp: ultrastructure, prevalence, and enhancement. J Invertebr Pathol. 1974; 24 (3): 311-31. DOI: 10.1016/0022-2011(74)90139-6

193. Lester R, Doubrovsky A, Paynter J, Sambhi S, Atherton J. Light and electron microscope evidence of baculovirus infection in the prawn Penaeus plebejus. Dis Aquat Organ. 1987; 3 (3): 217-9. DOI: 10.3354/dao003217

194. Bojko J, Ovcharenko M. Pathogens and other symbionts of the Amphipoda: taxonomic diversity and pathological significance. Dis Aquat Organ. 2019; 136 (1): 3-36. DOI: 10.3354/dao03321

195. Bojko J, Stebbing P, Bateman K, Meatyard J, Bacela-Spychalska K*, et al.* Baseline histopathological survey of a recently invading island population of 'killer shrimp', Dikerogammarus villosus. Dis Aquat Organ. 2013; 106 (3): 241-53. DOI: 10.3354/dao02658

196. Warren DA, Burgess AL, Karemera F, Bacela-Spychalska K, Stentiford GD*, et al.* Histopathological survey for parasite groups in Gammarus varsoviensis (Amphipoda). Dis Aquat Organ. 2022; 149: 47-51. DOI: 10.3354/dao03658

197. Warren DA, Burgess AL, Prati S, Bacela-Spychalska K, Rogers MS*, et al.* Histopathological screening of Pontogammarus robustoides (Amphipoda), an invader on route to the United Kingdom. J Invertebr Pathol. 2023; 200: 107970. DOI: 10.1016/j.jip.2023.107970

198. Dong Y. Mitochondrial genome evolution and molecular phylogeny of parasitic lice (infraorder: Phthiraptera): University of the Sunshine Coast, Queensland; 2022.

199. Sweet AD, Wilson RE, Sonsthagen SA, Johnson KP. Lousy grouse: Comparing evolutionary patterns in Alaska galliform lice to understand host evolution and host–parasite interactions. Ecol Evol. 2020; 10 (15): 8379-93. DOI: 10.1002/ece3.6545

200. Beliavskaia A, Tan KK, Sinha A, Husin NA, Lim FS*, et al.* Metagenomics of culture isolates and insect tissue illuminate the evolution of Wolbachia, Rickettsia and Bartonella symbionts in Ctenocephalides spp. fleas. Microb Genom. 2023; 9 (7): 2023.02. 06.527313. DOI: 10.1099/mgen.0.001045

201. Feng Y, Gou QY, Yang WH, Wu WC, Wang J*, et al.* A time-series meta-transcriptomic analysis reveals the seasonal, host, and gender structure of mosquito viromes. Virus Evol. 2022; 8 (1): veac006. DOI: 10.1093/ve/veac006

202. Tom M, Manfrin C, Giulianini PG, Pallavicini A. Crustacean oxi-reductases protein sequences derived from a functional genomic project potentially involved in ecdysteroid hormones metabolism - a starting point for function examination. Gen Comp Endocrinol. 2013; 194: 71-80. DOI: 10.1016/j.ygcen.2013.09.003

203. Song N, Zhang H, Bai R-e, Meng H-g. The Mitogenome of Aleuroclava Psidii (Singh, 1931)(Hemiptera: Aleyrodidae) and Increased Number of Mitochondrial Gene Rearrangements in Whiteflies. Frontiers in Bioscience-Landmark. 2022; 27 (5): 154. DOI: 10.31083/j.fbl2705154

204. Sana S, Vollhardt I, Kubon K, Rostás M, Scholten S. De novo transcriptome assemblies of five major European oilseed rape insect pests. BMC Genomic Data. 2023; 24 (1): 15. DOI: 10.1186/s12863-023-01115-8

205. Wang J, Gou Q-y, Luo G-y, Hou X, Liang G*, et al.* Total RNA sequencing of Phlebotomus chinensis sandflies in China revealed viral, bacterial, and eukaryotic microbes potentially pathogenic to humans. Emerg Microbes Infect. 2022; 11 (1): 2080-92. DOI: 10.1080/22221751.2022.2109516

206. Käfer S, Paraskevopoulou S, Zirkel F, Wieseke N, Donath A*, et al.* Re-assessing the diversity of negative strand RNA viruses in insects. PLoS Pathog. 2019; 15 (12): e1008224. DOI: 10.1371/journal.ppat.1008224

207. Misof B, Liu S, Meusemann K, Peters RS, Donath A*, et al.* Phylogenomics resolves the timing and pattern of insect evolution. Science. 2014; 346 (6210): 763-7. DOI: 10.1126/science.1257570

208. Pérez-Moreno JL, Balázs G, Bracken-Grissom HD. Transcriptomic insights into the loss of vision in Molnár János Cave's crustaceans. Integrative and Comparative Biology. 2018; 58 (3): 452-64.

209. Weitzman MD, Fradet-Turcotte A. Virus DNA replication and the host DNA damage response. Annu Rev Virol. 2018; 5: 141-64. DOI: 10.1146/annurev-virology-092917-043534

210. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD*, et al.* IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Molecular biology and evolution. 2020; 37 (5): 1530-4.

211. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021; 49 (W1): W293-W6. DOI: 10.1093/nar/gkab301

212. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol. 2021; 38 (7): 3022-7. DOI: 10.1093/molbev/msab120

213. Tamura K, Tao Q, Kumar S. Theoretical foundation of the RelTime method for estimating divergence times from variable evolutionary rates. Mol Biol Evol. 2018; 35 (7): 1770-82. DOI: 10.1093/molbev/msy044

214. Tao Q, Tamura K, Mello B, Kumar S. Reliable confidence intervals for RelTime estimates of evolutionary divergence times. Mol Biol Evol. 2020; 37 (1): 280-90. DOI: 10.1093/molbev/msz236

215. Murphy N, Banks JC, Whitfield JB, Austin AD. Phylogeny of the parasitic microgastroid subfamilies (Hymenoptera: Braconidae) based on sequence data from seven genes, with an improved time estimate of the origin of the lineage. Mol Phylogenet Evol. 2008; 47 (1): 378-95. DOI: 10.1016/j.ympev.2008.01.022

216. Kosakovsky Pond SL, Poon AFY, Velazquez R, Weaver S, Hepler NL*, et al.* HyPhy 2.5-A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. Mol Biol Evol. 2020; 37 (1): 295-9. DOI: 10.1093/molbev/msz197

217. Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Molecular biology and evolution. 2005; 22 (5): 1208-22.

218. Herniou EA, Huguet E, Thézé J, Bézier A, Periquet G*, et al.* When parasitic wasps hijacked viruses: genomic and functional evolution of polydnaviruses. Philos T R Soc B. 2013; 368 (1626): 20130051. DOI: 10.1098/rstb.2013.0051

219. Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M*, et al.* TimeTree 5: An Expanded Resource for Species Divergence Times. Mol Biol Evol. 2022; 39 (8): msac174. DOI: 10.1093/molbev/msac174

220. Brocklehurst N, Benson RJ. Multiple paths to morphological diversification during the origin of amniotes. Nat Ecol Evol. 2021; 5 (9): 1243-9. DOI: 10.1038/s41559-021-01516-x

221. Watson V, Rothschild B. Deep origin of parasitic disease in vertebrates. The Evolution and Fossil Record of Parasitism: Coevolution and Paleoparasitological Techniques. 2021: 317-58.

222. Zhu Q, Hastriter MW, Whiting MF, Dittmar K. Fleas (Siphonaptera) are Cretaceous, and evolved with Theria. Mol Phylogenet Evol. 2015; 90: 129-39. DOI: 10.1016/j.ympev.2015.04.027

223. de Moya RS, Yoshizawa K, Walden KKO, Sweet AD, Dietrich CH*, et al.* Phylogenomics of Parasitic and Nonparasitic Lice (Insecta: Psocodea): Combining Sequence Data and Exploring Compositional Bias Solutions in Next Generation Data Sets. Syst Biol. 2021; 70 (4): 719-38. DOI: 10.1093/sysbio/syaa075

224. Zhang Y, Rasnitsyn AP, Zhang W, Song F, Shih C*, et al.* Stem chewing lice on Cretaceous feathers preserved in amber. Curr Biol. 2024. DOI: 10.1016/j.cub.2024.01.027

225. Yu Y, Zhang C, Xu X. Deep time diversity and the early radiations of birds. Proc Natl Acad Sci U S A. 2021; 118 (10): e2019865118. DOI: 10.1073/pnas.2019865118

226. Etebari K, Parry R, Beltran MJB, Furlong MJ. Transcription Profile and Genomic Variations of Oryctes Rhinoceros Nudivirus in Coconut Rhinoceros Beetles. J Virol. 2020; 94 (22): e01097-20. DOI: 10.1128/JVI.01097-20

227. Cusumano A, Volkoff AN. Influence of parasitoid-associated viral symbionts on plant-insect interactions and biological control. Curr Opin Insect Sci. 2021; 44: 64-71. DOI: 10.1016/j.cois.2021.03.009

228. Haba Y, McBride L. Origin and status of Culex pipiens mosquito ecotypes. Curr Biol. 2022; 32 (5): R237-R46. DOI: 10.1016/j.cub.2022.01.062

229. Safyanova V. Laboratory cultivation of sandflies (Diptera; Phlebotominae). Bull World Health Organ. 1964; 31 (4): 573.

230. Ahmad A, Banerjee K, Gupta D, Chandra K. Insecta: Phthiraptera. Faunal Diversity of Biogeographic Zones of India: North-East. 2021: 271-80.

231. Zhang Y, Fu Y-T, Yao C, Deng Y-P, Nie Y*, et al.* Mitochondrial phylogenomics provides insights into the taxonomy and phylogeny of fleas. Parasites & Vectors. 2022; 15 (1): 223. DOI: 10.1186/s13071-022-05334-3

232. Appelgren ASC, Saladin V, Richner H, Doligez B, McCoy KD. Gene flow and adaptive potential in a generalist ectoparasite. BMC Evol Biol. 2018; 18 (1): 99. DOI: 10.1186/s12862-018-1205-2

233. Harnos A, Lang Z, Petras D, Bush SE, Szabo K*, et al.* Size matters for lice on birds: Coevolutionary allometry of host and parasite body size. Evolution. 2017; 71 (2): 421-31. DOI: 10.1111/evo.13147

234. Gelderblom HR. Structure and classification of viruses. Medical Microbiology 4th edition. 1996.

235. Sandhu SK, Morozov AY, Holt RD, Barfield M. Revisiting the role of hyperparasitism in the evolution of virulence. Am Nat. 2021; 197 (2): 216-35. DOI: 10.1086/712351

236. Toguebaye BS, Quilichini Y, Diagne PM, Marchand B. Ultrastructure and development of Nosema podocotyloidis n. sp.(Microsporidia), a hyperparasite of Podocotyloides magnatestis (Trematoda), a parasite of Parapristipoma octolineatum (Teleostei). Parasite. 2014; 21. DOI: 10.1051/parasite/2014044

237. Wood J, Ashby B. Hyperparasitism and the evolution of parasite virulence. Evolution. 2023; 77 (12): 2631-41. DOI: 10.1093/evolut/qpad178

238. Poulin R, Randhawa HS. Evolution of parasitism along convergent lines: from ecology to genomics. Parasitology. 2015; 142 (S1): S6-S15. DOI: 10.1017/S0031182013001674

239. Burke GR, Walden KK, Whitfield JB, Robertson HM, Strand MR. Widespread genome reorganization of an obligate virus mutualist. PLoS Genet. 2014; 10 (9): e1004660. DOI: 10.1371/journal.pgen.1004660

240. Perry HM, Anderson J, Collins L. Status and Management of the Blue Crab Fishery in the Gulf of Mexico. N Am J Fish Manage. 2022; 42 (1): 164-79. DOI: 10.1002/nafm.10727

241. Walters EA, Bojko J, Crowley CE, Gandy RL, Martin CW*, et al.* Salinity and temperature affect the symbiont profile and host condition of Florida USA blue crabs Callinectes sapidus. J Invertebr Pathol. 2023; 198: 107930. DOI: 10.1016/j.jip.2023.107930

242. Arulmoorthy MP, Anandajothi E, Vasudevan S, Suresh E. Major viral diseases in culturable penaeid shrimps: a review. Aquacult Int. 2020; 28 (5): 1939-67. DOI: 10.1007/s10499-020-00568-3

243. Bojko J, Burgess AL, Allain TW, Ross EP, Pharo D*, et al.* Pathology and genetic connectedness of the mangrove crab (Aratus pisonii)–a foundation for understanding mangrove disease ecology. Anim Dis. 2022; 2 (1): 1-16. DOI: 10.1186/s44149-022-00039-7

244. Ritter CJ, Bourne DG. Marine amphipods as integral members of global ocean ecosystems. J Exp Mar Biol Ecol. 2024; 572: 151985. DOI: 10.1016/j.jembe.2023.151985

245. Sandoval LA, Mancera-Pineda JE, Delgado-Huertas A, Blanco-Libreros JF, Leal-Flórez J. Leaf consumption and experimental discrimination of stable isotopes between mangrove leaves and the tree-climbing crab (Brachyura: Sesarmidae: Aratus pisonii). Estuar Coast Shelf S. 2022; 274: 107906. DOI: 10.1016/j.ecss.2022.107906

246. Bojko J, Burgess AL, Baker AG, Orr CH. Invasive non-native crustacean symbionts: diversity and impact. J Invertebr Pathol. 2021; 186: 107482. DOI: 10.1016/j.jip.2020.107482

247. Wood L, Smith E, Bojko J, Stebbing P. Options for the control of Dikerogammarus villosus (killer shrimp) and other invasive amphipods: Invasive Amphipod Control. Manag Biol Invasion. 2021; 12. DOI: 10.3391/mbi.2021.12.3.10

248. Mancinelli G, Bardelli R, Zenetos A. A global occurrence database of the Atlantic blue crab Callinectes sapidus. Sci Data. 2021; 8 (1): 111. DOI: 10.1038/s41597-021-00888-w

249. Shuttleworth CM, Everest D, Holmes P, Bell S, Cripps R. An opportunistic assessment of the impact of squirrelpox disease outbreaks upon a red squirrel population sympatric with grey squirrels in Wales. Animals. 2022; 12 (1): 99. DOI: 10.3390/ani12010099

250. Rust MK. Recent advancements in the control of cat fleas. Insects. 2020; 11 (10): 668. DOI: 10.3390/insects11100668

251. Käfer S, Paraskevopoulou S, Zirkel F, Wieseke N, Donath A*, et al.* Re-assessing the diversity of negative strand RNA viruses in insects. PLoS Pathog. 2019; 15 (12): e1008224. DOI: 10.1371/journal.ppat.1008224

252. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M*, et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012; 19 (5): 455-77. DOI: 10.1089/cmb.2012.0021

253. Besemer J, Lomsadze A, Borodovsky M. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. 2001; 29 (12): 2607-18. DOI: 10.1093/nar/29.12.2607

254. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013; 30 (4): 772-80. DOI: 10.1093/molbev/mst010

255. Kalyaanamoorthy S, Minh BQ, Wong TK, Von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017; 14 (6): 587-9. DOI: 10.1038/nmeth.4285

256. Mello B. Estimating timetrees with MEGA and the TimeTree resource. Mol Biol Evol. 2018; 35 (9): 2334-42. DOI: 10.1093/molbev/msy133

257. Light DM, Flath RA, Buttery RG, Zalom FG, Rice RE*, et al.* Host-plant green-leaf volatiles synergize the synthetic sex pheromones of the corn earworm and codling moth (Lepidoptera). Chemoecology. 1993; 4: 145-52.

258. Health EPoP, Bragard C, Dehnen-Schmutz K, Di Serio F, Gonthier P*, et al.* Pest categorisation of Helicoverpa zea. EFSA J. 2020; 18 (7): e06177. DOI: 10.2903/j.efsa.2020.6177

259. Capinera J. Handbook of vegetable pests: Academic press; 2020.

260. North HL, Fu Z, Metz R, Stull MA, Johnson CD*, et al.* Rapid Adaptation and Interspecific Introgression in the North American Crop Pest Helicoverpa zea. Mol Biol Evol. 2024; 41 (7): msae129. DOI: 10.1093/molbev/msae129

261. Koller J, Sutter L, Gonthier J, Collatz J, Norgrove L. Entomopathogens and Parasitoids Allied in Biocontrol: A Systematic Review. Pathogens. 2023; 12 (7): 957. DOI: 10.3390/pathogens12070957

262. Raina AK, Adams JR. Gonad-specific virus of corn earworm. Nature. 1995; 374 (6525): 770-. DOI: 10.1038/374770a0

263. Rallis CP, Burand JP. Pathology and ultrastructure of Hz-2V infection in the agonadal female corn earworm, *Helicoverpa zea*. J Invertebr Pathol. 2002; 81 (1): 33-44. DOI: 10.1016/s0022-2011(02)00113-1

264. Kingan TG, Bodnar WM, Raina AK, Shabanowitz J, Hunt DF. The loss of female sex pheromone after mating in the corn earworm moth Helicoverpa zea: identification of a male pheromonostatic peptide. Proc Natl Acad Sci U S A. 1995; 92 (11): 5082-6. DOI: 10.1073/pnas.92.11.5082

265. Burand JP, Tan W, Kim W, Nojima S, Roelofs W. Infection with the insect virus Hz-2v alters mating behavior and pheromone production in female Helicoverpa zea moths. J Insect Sci. 2005; 5 (1): 6. DOI: 10.1093/jis/5.1.6

266. Madden T. The BLAST sequence analysis tool. The NCBI handbook. 2013; 2: 425-36.

267. Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Bateman A*, et al.* New developments in the InterPro database. Nucleic Acids Res. 2007; 35 (Database issue): D224-8. DOI: 10.1093/nar/gkl841

268. Ralston AL, Huang YS, Kawanishi CY. Cell culture studies with the IMC-Hz-1 nonoccluded virus. Virology. 1981; 115 (1): 33-44. DOI: 10.1016/0042-6822(81)90086-6

269. Jain C, Rhie A, Zhang H, Chu C, Walenz BP*, et al.* Weighted minimizer sampling improves long read mapping. Bioinformatics. 2020; 36 (Suppl_1): i111-i8. DOI: 10.1093/bioinformatics/btaa435

270. Hackl T, Ankenbrand MJ, van Adrichem B. gggenomes: A Grammar of Graphics for Comparative Genomics. R package version 1.0.1 ed2024.

271. Wilkins D, Kurtz Z. gggenes: draw gene arrow maps in 'ggplot2'. R package version 04 0. 2019; 342.

272. Hopkins R, Esposito D. A rapid method for titrating baculovirus stocks using the Sf-9 Easy Titer cell line. Biotechniques. 2009; 47 (3): 785-8. DOI: 10.2144/000113238

273. Reed LJ, Muench H. A simple method of estimating fifty per cent endpoints. American journal of epidemiology. 1938; 27 (3): 493-7.

274. Carpentier DC, King LA. The long road to understanding the baculovirus P10 protein. Virologica Sinica. 2009; 24: 227-42.

275. Van Oers MM, Flipsen JT, Reusken CB, Vlak JM. Specificity of baculovirus p10 functions. Virology. 1994; 200 (2): 513-23. DOI: 10.1006/viro.1994.1214

276. McIntosh AH, Grasela JJ, Popham HJ. AcMNPV in permissive, semipermissive, and nonpermissive cell lines from Arthropoda. In Vitro Cell Dev Biol Anim. 2005; 41 (8-9): 298-304. DOI: 10.1290/0412083R.1

277. Del Prete MJ, Robles MS, Guío A, Martínez-A C, Izquierdo M*, et al.* Degradation of cellular mRNA is a general early apoptosis-induced event. The FASEB journal. 2002; 16 (14): 2003-5.

278. Pfaffl MW. Relative quantification. Real-time PCR: Taylor & Francis; 2007. p. 89-108.

279. Wickham T, Davis T, Granados R, Hammer D, Shuler M*, et al.* Baculovirus defective interfering particles are responsible for variations in recombinant protein production as a function of multiplicity of infection. Biotechnol Lett. 1991; 13: 483-8. DOI: 10.1007/Bf01049204

280. Zhang YH, Enden G, Merchuk JC. Insect cells–Baculovirus system: factors affecting growth and low MOI infection. Biochem Eng J. 2005; 27 (1): 8-16. DOI: 10.1016/j.bej.2005.05.013

281. Braunagel SC, Parr R, Belyavskyi M, Summers MD. Autographa californicanucleopolyhedrovirus infection results in sf9 cell cycle arrest at g2/m phase. Virology. 1998; 244 (1): 195-211. DOI: 10.1006/viro.1998.9097

282. Wang X, Liu X, Makalliwa GA, Li J, Wang H*, et al.* Per os infectivity factors: a complicated and evolutionarily conserved entry machinery of baculovirus. Sci China Life Sci. 2017; 60 (8): 806-15. DOI: 10.1007/s11427-017-9127-1

283. WF IJ, Westenberg M, Goldbach RW, Blissard GW, Vlak JM*, et al.* A novel baculovirus envelope fusion protein with a proprotein convertase cleavage site. Virology. 2000; 275 (1): 30-41. DOI: 10.1006/viro.2000.0483

284. Pearson MN, Groten C, Rohrmann GF. Identification of the lymantria dispar nucleopolyhedrovirus envelope fusion protein provides evidence for a phylogenetic division of the Baculoviridae. J Virol. 2000; 74 (13): 6126-31. DOI: 10.1128/jvi.74.13.6126-6131.2000

285. Hefferon KL, Oomens AG, Monsma SA, Finnerty CM, Blissard GW. Host cell receptor binding by baculovirus GP64 and kinetics of virion entry. Virology. 1999; 258 (2): 455-68. DOI: 10.1006/viro.1999.9758

286. Monsma SA, Oomens A, Blissard GW. The GP64 envelope fusion protein is an essential baculovirus protein required for cell-to-cell transmission of infection. J Virol. 1996; 70 (7): 4607-16. DOI: 10.1128/Jvi.70.7.4607-4616.1996

287. Li W, Desmarets LMB, De Gryse GMA, Theuns S, Van Tuan V*, et al.* Virus replication cycle of white spot syndrome virus in secondary cell cultures from the lymphoid organ of Litopenaeus vannamei. J Gen Virol. 2015; 96 (9): 2844-54. DOI: 10.1099/vir.0.000217

288. Rosinski M, Reid S, Nielsen LK. Kinetics of baculovirus replication and release using real-time quantitative polymerase chain reaction. Biotechnol Bioeng. 2002; 77 (4): 476-80.

289. Amara A, Mercer J. Viral apoptotic mimicry. Nat Rev Microbiol. 2015; 13 (8): 461-9. DOI: 10.1038/nrmicro3469

290. Burand JP, Rallis CP, Tan W. Horizontal transmission of Hz-2V by virus infected *Helicoverpa zea* moths. J Invertebr Pathol. 2004; 85 (2): 128-31. DOI: 10.1016/j.jip.2004.01.004

291. Granados RR, Lawler KA. *In vivo* pathway of Autographa californica baculovirus invasion and infection. Virology. 1981; 108 (2): 297-308. DOI: 10.1016/0042-6822(81)90438-4

292. Etebari K, Gharuka M, Asgari S, Furlong MJ. Diverse Host Immune Responses of Different Geographical Populations of the Coconut Rhinoceros Beetle to Oryctes Rhinoceros Nudivirus (OrNV) Infection. Microbiol Spectr. 2021; 9 (2): e0068621. DOI: 10.1128/Spectrum.00686-21

293. Cerqueira de Araujo A, Leobold M, Bezier A, Musset K, Uzbekov R*, et al.* Conserved Viral Transcription Plays a Key Role in Virus-Like Particle Production of the Parasitoid Wasp *Venturia canescens*. J Virol. 2022; 96 (13): e0052422. DOI: 10.1128/jvi.00524-22

294. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018; 34 (17): i884-i90. DOI: 10.1093/bioinformatics/bty560

295. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015; 12 (4): 357-60. DOI: 10.1038/nmeth.3317

296. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V*, et al.* Twelve years of SAMtools and BCFtools. Gigascience. 2021; 10 (2): giab008. DOI: 10.1093/gigascience/giab008

297. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT*, et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015; 33 (3): 290-5. DOI: 10.1038/nbt.3122

298. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc. 2016; 11 (9): 1650-67. DOI: 10.1038/nprot.2016.095

299. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW*, et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015; 43 (7): e47. DOI: 10.1093/nar/gkv007

300. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 2014; 15 (2): R29. DOI: 10.1186/gb-2014-15-2-r29

301. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F*, et al.* The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. 2023; 51 (D1): D638-D46. DOI: 10.1093/nar/gkac1000

302. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT*, et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13 (11): 2498-504. DOI: 10.1101/gr.1239303

303. Kolde R, Kolde MR. Package 'pheatmap'. R package. 2015; 1 (7): 790.

304. Charrad M, Ghazzali N, Boiteau V, Niknafs A. NbClust: an R package for determining the relevant number of clusters in a data set. J Stat Softw. 2014; 61 (6): 1-36. DOI: 10.18637/jss.v061.i06

305. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. Nucleic Acids Res. 2015; 43 (W1): W39-49. DOI: 10.1093/nar/gkv416

306. McLeay RC, Bailey TL. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. BMC Bioinformatics. 2010; 11 (1): 165. DOI: 10.1186/1471-2105-11-165

307. Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. Quantifying similarity between motifs. Genome Biol. 2007; 8 (2): 1-9. DOI: 10.1186/gb-2007-8-2-r24

308. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. Bioinformatics. 2011; 27 (7): 1017-8. DOI: 10.1093/bioinformatics/btr064

309. Lu Z, Berry K, Hu Z, Zhan Y, Ahn TH*, et al.* TSSr: an R package for comprehensive analyses of TSS sequencing data. NAR Genom Bioinform. 2021; 3 (4): lqab108. DOI: 10.1093/nargab/lqab108

310. Chevignon G, Theze J, Cambier S, Poulain J, Da Silva C*, et al.* Functional annotation of Cotesia congregata bracovirus: identification of viral genes expressed in parasitized host immune tissues. J Virol. 2014; 88 (16): 8795-812. DOI: 10.1128/JVI.00209-14

311. Shrestha A, Bao K, Chen W, Wang P, Fei Z*, et al.* Transcriptional Responses of the *Trichoplusia ni* Midgut to Oral Infection by the Baculovirus Autographa californica Multiple Nucleopolyhedrovirus. J Virol. 2019; 93 (14): 10.1128/jvi. 00353-19. DOI: 10.1128/JVI.00353-19

312. Rivas HG, Schmaling SK, Gaglia MM. Shutoff of Host Gene Expression in Influenza A Virus and Herpesviruses: Similar Mechanisms and Common Themes. Viruses. 2016; 8 (4): 102. DOI: 10.3390/v8040102

313. Wu YL, Wu CP, Liu CY, Lee ST, Lee HP*, et al.* Heliothis zea nudivirus 1 gene hhi1 induces apoptosis which is blocked by the Hz-iap2 gene and a noncoding gene, pag1. J Virol. 2011; 85 (14): 6856-66. DOI: 10.1128/JVI.01843-10

314. Hefferon KL. Baculovirus late expression factors. J Mol Microbiol Biotechnol. 2004; 7 (3): 89-101. DOI: 10.1159/000078652

315. Kariithi HM, Meki IK, Boucias DG, Abd-Alla AM. Hytrosaviruses: current status and perspective. Curr Opin Insect Sci. 2017; 22: 71-8. DOI: 10.1016/j.cois.2017.05.009

316. Tomalski MD, Wu JG, Miller LK. The location, sequence, transcription, and regulation of a baculovirus DNA polymerase gene. Virology. 1988; 167 (2): 591-600. DOI: 10.1016/0042-6822(88)90122-5

317. Chen R, Wang H, Mansky LM. Roles of uracil-DNA glycosylase and dUTPase in virus replication. J Gen Virol. 2002; 83 (Pt 10): 2339-45. DOI: 10.1099/0022-1317-83-10-2339

318. Chu E, Allegra CJ. The role of thymidylate synthase as an RNA binding protein. Bioessays. 1996; 18 (3): 191-8. DOI: 10.1002/bies.950180306

319. Chen HH, Tso DJ, Yeh WB, Cheng HJ, Wu TF. The thymidylate synthase gene of Hz-1 virus: a gene captured from its lepidopteran host. Insect Mol Biol. 2001; 10 (5): 495-503. DOI: 10.1046/j.0962-1075.2001.00289.x

320. Eriksson S, Munch-Petersen B, Johansson K, Eklund H. Structure and function of cellular deoxyribonucleoside kinases. Cell Mol Life Sci. 2002; 59 (8): 1327-46. DOI: 10.1007/s00018-002-8511-x

321. Wintersberger E. Regulation and biological function of thymidine kinase. Biochem Soc Trans. 1997; 25 (1): 303-8. DOI: 10.1042/bst0250303

322. Reichard P. Interactions between deoxyribonucleotide and DNA synthesis. Annu Rev Biochem. 1988; 57 (1): 349-74. DOI: 10.1146/annurev.bi.57.070188.002025

323. Reichard P. Ribonucleotide reductase and deoxyribonucleotide pools. Basic Life Sci. 1985; 31: 33-45. DOI: 10.1007/978-1-4613-2449-2_3

324. Wu DD, Zhang YP. Eukaryotic origin of a metabolic pathway in virus by horizontal gene transfer. Genomics. 2011; 98 (5): 367-9. DOI: 10.1016/j.ygeno.2011.08.006

325. Slack J, Arif BM. The baculoviruses occlusion-derived virus: virion structure and function. Adv Virus Res. 2006; 69: 99-165. DOI: 10.1016/S0065-3527(06)69003-9

326. Seo D, Gammon DB. Manipulation of Host Microtubule Networks by Viral Microtubule-Associated Proteins. Viruses. 2022; 14 (5): 979. DOI: 10.3390/v14050979

327. Kariithi HM, van Oers MM, Vlak JM, Vreysen MJ, Parker AG*, et al.* Virology, Epidemiology and Pathology of Glossina Hytrosavirus, and Its Control Prospects in Laboratory Colonies of the Tsetse Fly, *Glossina pallidipes* (Diptera; Glossinidae). Insects. 2013; 4 (3): 287-319. DOI: 10.3390/insects4030287

328. Zhang H, Kuang W, Fu C, Li J, Wang M*, et al.* AC81 Is a Putative Disulfide Isomerase Involved in Baculoviral Disulfide Bond Formation. J Virol. 2022; 96 (24): e01167-22. DOI: 10.1128/jvi.01167-22

329. Russell R, Rohrmann G. Characterization of P91, a Protein Associated with Virions of an *Orgyia pseudotsugata* Baculovirus. Virology. 1997; 233 (1): 210-23. DOI: 10.1006/viro.1997.8599

330. Kim W. Characterization of genome and structural proteins of the sexually transmitted insect virus, Hz-2V: University of Massachusetts Amherst; 2009.

331. Dong F, Wang J, Deng R, Wang X. Autographa californica multiple nucleopolyhedrovirus gene ac81 is required for nucleocapsid envelopment. Virus Res. 2016; 221: 47-57. DOI: 10.1016/j.virusres.2016.05.005

332. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011; 39 (Web Server issue): W29-37. DOI: 10.1093/nar/gkr367

333. Byers NM, Vandergaast RL, Friesen PD. Baculovirus Inhibitor-of-Apoptosis Op-IAP3 Blocks Apoptosis by Interaction with and Stabilization of a Host Insect Cellular IAP. J Virol. 2016; 90 (1): 533-44. DOI: 10.1128/JVI.02320-15

334. Guarino LA, Xu B, Jin J, Dong W. A virus-encoded RNA polymerase purified from baculovirus-infected cells. J Virol. 1998; 72 (10): 7985-91. DOI: 10.1128/JVI.72.10.7985-7991.1998

335. Berretta MF, Ferrelli ML, Salvador R, Sciocco A, Romanowski V. Baculovirus gene expression. Current issues in molecular virology-viral genetics and biotechnological applications: IntechOpen; 2013.

336. Kelly BJ, King LA, Possee RD. Introduction to Baculovirus Molecular Biology. Methods Mol Biol. 2016; 1350: 25-50. DOI: 10.1007/978-1-4939-3043-2_2

337. Chaitanya KV. Structure and Organization of Virus Genomes. Genome and Genomics. Singapur: Springer; 2019. p. 1-30.

338. Kumar J. The sine oculis homeobox (SIX) family of transcription factors as regulators of development and disease. Cell Mol Life Sci. 2009; 66: 565-83. DOI: 10.1007/s00018-008-8335-4

339. Jusiak B, Wang F, Karandikar UC, Kwak SJ, Wang H*, et al.* Genome-wide DNA binding pattern of the homeodomain transcription factor Sine oculis (So) in the developing eye of *Drosophila melanogaster*. Genom Data. 2014; 2: 153-5. DOI: 10.1016/j.gdata.2014.06.016

340. Oliver G, Mailhos A, Wehr R, Copeland NG, Jenkins NA*, et al.* Six3, a murine homologue of the sine oculis gene, demarcates the most anterior border of the developing neural plate and is expressed during eye development. Development. 1995; 121 (12): 4045-55. DOI: 10.1242/dev.121.12.4045

341. Jones JD, Dangl JL. The plant immune system. Nature. 2006; 444 (7117): 323-9. DOI: 10.1038/nature05286

342. Fei Q, Zhang Y, Xia R, Meyers BC. Small RNAs Add Zing to the Zig-Zag-Zig Model of Plant Defenses. Mol Plant Microbe Interact. 2016; 29 (3): 165-9. DOI: 10.1094/MPMI-09-15-0212-FI

343. Pritchard L, Birch PR. The zigzag model of plant–microbe interactions: is it time to move on? Mol Plant Pathol. 2014; 15 (9): 865. DOI: 10.1111/mpp.12210

344. Jin S, Cheng T, Jiang L, Lin P, Yang Q*, et al.* Identification of a new Sprouty protein responsible for the inhibition of the Bombyx mori nucleopolyhedrovirus reproduction. PLoS One. 2014; 9 (6): e99200. DOI: 10.1371/journal.pone.0099200

345. Sánchez EG, Quintas A, Nogal M, Castelló A, Revilla Y. African swine fever virus controls the host transcription and cellular machinery of protein synthesis. Virus Res. 2013; 173 (1): 58-75. DOI: 10.1016/j.virusres.2012.10.025

346. McGrath J, Roy P, Perrin BJ, editors. Stereocilia morphogenesis and maintenance through regulation of actin stability. Seminars in Cell & Developmental Biology; 2017: Elsevier. DOI: 10.1016/j.semcdb.2016.08.017

347. Volkman LE. Baculovirus infectivity and the actin cytoskeleton. Curr Drug Targets. 2007; 8 (10): 1075-83. DOI: 10.2174/138945007782151379

348. Hornikova L, Brustikova K, Huerfano S, Forstova J. Nuclear Cytoskeleton in Virus Infection. Int J Mol Sci. 2022; 23 (1): 578. DOI: 10.3390/ijms23010578

349. Khorramnejad A, Perdomo HD, Palatini U, Bonizzoni M, Gasmi L. Cross talk between viruses and insect cells cytoskeleton. Viruses. 2021; 13 (8): 1658. DOI: 10.3390/v13081658

350. Xing L, Yuan C, Wang M, Lin Z, Shen B*, et al.* Dynamics of the interaction between cotton bollworm *Helicoverpa armigera* and nucleopolyhedrovirus as revealed by integrated transcriptomic and proteomic analyses. Mol Cell Proteomics. 2017; 16 (6): 1009-28. DOI: 10.1074/mcp.M116.062547

351. Feng M, Fei S, Xia J, Zhang M, Wu H*, et al.* Global metabolic profiling of baculovirus infection in silkworm hemolymph shows the importance of amino-acid metabolism. Viruses. 2021; 13 (5): 841. DOI: 10.3390/v13050841

352. Mayer KA, Stöckl J, Zlabinger GJ, Gualdoni GA. Hijacking the supplies: metabolism as a novel facet of virus-host interaction. Front Immunol. 2019; 10: 1533. DOI: 10.3389/fimmu.2019.01533

353. Thaker SK, Ch'ng J, Christofk HR. Viral hijacking of cellular metabolism. BMC Biol. 2019; 17 (1): 59. DOI: 10.1186/s12915-019-0678-9

354. Sanchez EL, Lagunoff M. Viral activation of cellular metabolism. Virology. 2015; 479-480: 609-18. DOI: 10.1016/j.virol.2015.02.038

355. Dubrez L, Causse S, Borges Bonan N, Dumetier B, Garrido C. Heat-shock proteins: chaperoning DNA repair. Oncogene. 2020; 39 (3): 516-29. DOI: 10.1038/s41388-019-1016-y

356. Haslbeck M, Vierling E. A first line of stress defense: small heat shock proteins and their function in protein homeostasis. J Mol Biol. 2015; 427 (7): 1537-48. DOI: 10.1016/j.jmb.2015.02.002

357. Zhao L, Jones W. Expression of heat shock protein genes in insect stress responses. Invertebr Surviv J. 2012; 9 (1): 93-101.

358. Merkling SH, Overheul GJ, van Mierlo JT, Arends D, Gilissen C*, et al.* The heat shock response restricts virus infection in *Drosophila*. Sci Rep. 2015; 5 (1): 12758. DOI: 10.1038/srep12758

359. Janewanthanakul S, Supungul P, Tang S, Tassanakajon A. Heat shock protein 70 from *Litopenaeus vannamei* (LvHSP70) is involved in the innate immune response against white spot syndrome virus (WSSV) infection. Dev Comp Immunol. 2020; 102: 103476. DOI: 10.1016/j.dci.2019.103476

360. Lyupina YV, Zatsepina OG, Timokhova AV, Orlova OV, Kostyuchenko MV*, et al.* New insights into the induction of the heat shock proteins in baculovirus infected insect cells. Virology. 2011; 421 (1): 34-41. DOI: 10.1016/j.virol.2011.09.010

361. Cao M, Wei C, Zhao L, Wang J, Jia Q*, et al.* DnaJA1/Hsp40 is co-opted by influenza A virus to enhance its viral RNA polymerase activity. J Virol. 2014; 88 (24): 14078-89. DOI: 10.1128/jvi.02475-14

362. Khachatoorian R, French SW. Chaperones in hepatitis C virus infection. World J Hepatol. 2016; 8 (1): 9-35. DOI: 10.4254/wjh.v8.i1.9

363. Oka OBV, Pringle MA, Schopp IM, Braakman I, Bulleid NJ. ERdj5 is the ER reductase that catalyzes the removal of non-native disulfides and correct folding of the LDL receptor. Mol Cell. 2013; 50 (6): 793-804. DOI: 10.1016/j.molcel.2013.05.014

364. Sohn S-Y, Kim J-H, Baek K-W, Ryu W-S, Ahn B-Y. Turnover of hepatitis B virus X protein is facilitated by Hdj1, a human Hsp40/DnaJ protein. Biochem Biophys Res Commun. 2006; 347 (3): 764-8. DOI: 10.1016/j.bbrc.2006.06.158

365. Lyupina YV, Dmitrieva SB, Timokhova AV, Beljelarskaya SN, Zatsepina OG*, et al.* An important role of the heat shock response in infected cells for replication of baculoviruses. Virology. 2010; 406 (2): 336-41. DOI: 10.1016/j.virol.2010.07.039

366. Xiao A, Wong J, Luo H. Viral interaction with molecular chaperones: role in regulating viral infection. Arch Virol. 2010; 155 (7): 1021-31. DOI: 10.1007/s00705-010-0691-3

367. Kurzik-Dumke U, Lohmann E. Sequence of the new *Drosophila melanogaster* small heat-shock-related gene, lethal (2) essential for life [l (2) efl], at locus 59F4, 5. Gene. 1995; 154 (2): 171-5. DOI: 10.1016/0378-1119(94)00827-F

368. Raut S, Mallik B, Parichha A, Amrutha V, Sahi C*, et al.* RNAi-Mediated Reverse Genetic Screen Identified *Drosophila* Chaperones Regulating Eye and Neuromuscular Junction Morphology. G3: Genes Genomes Genet. 2017; 7 (7): 2023-38. DOI: 10.1534/g3.117.041632

369. Brutscher LM, Daughenbaugh KF, Flenniken ML. Virus and dsRNA-triggered transcriptional responses reveal key components of honey bee antiviral defense. Sci Rep. 2017; 7 (1): 6448. DOI: 10.1038/s41598-017-06623-z

370. Lopez W, Page AM, Carlson DJ, Ericson BL, Cserhati MF*, et al.* Analysis of immune-related genes during Nora virus infection of Drosophila melanogaster using next generation sequencing. AIMS Microbiol. 2018; 4 (1): 123-39. DOI: 10.3934/microbiol.2018.1.123

371. Runtuwene LR, Kawashima S, Pijoh VD, Tuda JSB, Hayashida K*, et al.* The Lethal(2)-Essential-for-Life [L(2)EFL] Gene Family Modulates Dengue Virus Infection in Aedes aegypti. Int J Mol Sci. 2020; 21 (20): 7520. DOI: 10.3390/ijms21207520

372. Mayer MP. Recruitment of Hsp70 chaperones: a crucial part of viral survival strategies. Rev Physiol Biochem Pharmacol. 2005; 153: 1-46. DOI: 10.1007/s10254-004-0025-5

373. Yingsunthonwattana W, Junprung W, Supungul P, Tassanakajon A. Heat shock protein 90 of Pacific white shrimp (*Litopenaeus vannamei*) is possibly involved in promoting white spot syndrome virus infection. Fish Shellfish Immunol. 2022; 128: 405-18. DOI: 10.1016/j.fsi.2022.08.016

374. Nikapitiya C, Dorrington T, Gómez-Chiarri M. The role of histones in the immune responses of aquatic invertebrates. Invertebr Surviv J. 2013; 10 (1): 94-101.

375. Wolffe AP, Hayes JJ. Chromatin disruption and modification. Nucleic Acids Res. 1999; 27 (3): 711-20. DOI: 10.1093/nar/27.3.711

376. Brinkmann V, Zychlinsky A. Neutrophil extracellular traps: is immunity the second function of chromatin? J Cell Biol. 2012; 198 (5): 773-83. DOI: 10.1083/jcb.201203170

377. Parseghian MH, Luhrs KA. Beyond the walls of the nucleus: the role of histones in cellular signaling and innate immunity. Biochem Cell Biol. 2006; 84 (4): 589-604. DOI: 10.1139/o06-082

378. Encinas-Garcia T, Loreto-Quiroz DL, Mendoza-Cano F, Pena-Rodriguez A, Fimbres-Olivarria D*, et al.* White spot syndrome virus down-regulates expression of histones H2A and H4 of *Penaeus vannamei* to promote viral replication. Dis Aquat Organ. 2019; 137 (1): 73-9. DOI: 10.3354/dao03428

379. Nayyar N, Kaur I, Malhotra P, Bhatnagar RK. Quantitative proteomics of Sf 21 cells during Baculovirus infection reveals progressive host proteome changes and its regulation by viral miRNA. Sci Rep. 2017; 7 (1): 10902. DOI: 10.1038/s41598-017-10787-z

380. Ooi BG, Miller LK. Regulation of host RNA levels during baculovirus infection. Virology. 1988; 166 (2): 515-23. DOI: 10.1016/0042-6822(88)90522-3

381. Foudi A, Hochedlinger K, Van Buren D, Schindler JW, Jaenisch R*, et al.* Analysis of histone 2B-GFP retention reveals slowly cycling hematopoietic stem cells. Nat Biotechnol. 2009; 27 (1): 84-90. DOI: 10.1038/nbt.1517

382. Kasai S, Scott JG. A house fly gene homologous to the zinc finger proto-oncogene Gfi-1. Biochem Biophys Res Commun. 2001; 283 (3): 644-7. DOI: 10.1006/bbrc.2001.4826

383. Tanada Y, Hess RT. *Baculoviridae*. Granulosis Viruses. Atlas of Invertebrate Viruses: CRC Press; 2017. p. 227-57.

384. Winstanley D, Crook NE. Replication of Cydia pomonella granulosis virus in cell cultures. J Gen Virol. 1993; 74 (Pt 8) (8): 1599-609. DOI: 10.1099/0022-1317-74-8-1599

385. Henikoff S, Ahmad K. Assembly of variant histones into chromatin. Annu Rev Cell Dev Biol. 2005; 21 (1): 133-53. DOI: 10.1146/annurev.cellbio.21.012704.133518

386. Lee GE, Byun J, Lee CJ, Cho YY. Molecular Mechanisms for the Regulation of Nuclear Membrane Integrity. Int J Mol Sci. 2023; 24 (20): 15497. DOI: 10.3390/ijms242015497

387. Simoes M, Rino J, Pinheiro I, Martins C, Ferreira F. Alterations of Nuclear Architecture and Epigenetic Signatures during African Swine Fever Virus Infection. Viruses. 2015; 7 (9): 4978-96. DOI: 10.3390/v7092858

388. Porwal M, Cohen S, Snoussi K, Popa-Wagner R, Anderson F*, et al.* Parvoviruses cause nuclear envelope breakdown by activating key enzymes of mitosis. PLoS Pathog. 2013; 9 (10): e1003671. DOI: 10.1371/journal.ppat.1003671

389. Silva L, Cliffe A, Chang L, Knipe DM. Role for A-type lamins in herpesviral DNA targeting and heterochromatin modulation. PLoS Pathog. 2008; 4 (5): e1000071. DOI: 10.1371/journal.ppat.1000071

390. Cibulka J, Fraiberk M, Forstova J. Nuclear actin and lamins in viral infections. Viruses. 2012; 4 (3): 325-47. DOI: 10.3390/v4030325

391. Snoussi K, Kann M. Interaction of parvoviruses with the nuclear envelope. Adv Biol Regul. 2014; 54: 39-49. DOI: 10.1016/j.jbior.2013.09.008

392. Wyler T, Lanzrein B. Ovary development and polydnavirus morphogenesis in the parasitic wasp Chelonus inanitus. II. Ultrastructural analysis of calyx cell development, virion formation and release. J Gen Virol. 2003; 84 (Pt 5): 1151-63. DOI: 10.1099/vir.0.18830-0

393. Christen U. Molecular mimicry. Autoantibodies: Elsevier; 2014. p. 35-42.

394. Gad W, Kim Y. A viral histone H4 encoded by Cotesia plutellae bracovirus inhibits haemocyte-spreading behaviour of the diamondback moth, *Plutella xylostella*. J Gen Virol. 2008; 89 (Pt 4): 931-8. DOI: 10.1099/vir.0.83585-0

395. Kim J, Kim Y. Transient expression of a viral histone H4, CpBV-H4, suppresses immune-associated genes of *Plutella xylostella* and *Spodoptera exigua*. J Asia Pac Entomol. 2010; 13 (4): 313-8. DOI: 10.1016/j.aspen.2010.05.003

396. Avgousti DC, Della Fera AN, Otter CJ, Herrmann C, Pancholi NJ*, et al.* Adenovirus Core Protein VII Downregulates the DNA Damage Response on the Host Genome. J Virol. 2017; 91 (20): 10.1128/jvi. 01089-17. DOI: 10.1128/JVI.01089-17

397. Kulej K, Avgousti DC, Sidoli S, Herrmann C, Della Fera AN*, et al.* Time-resolved Global and Chromatin Proteomics during Herpes Simplex Virus Type 1 (HSV-1) Infection. Mol Cell Proteomics. 2017; 16 (4 suppl 1): S92-S107. DOI: 10.1074/mcp.M116.065987

398. Kariyawasam H, Su S, Voogd O, Ritchie ME, Law CW. Dashboard-style interactive plots for RNA-seq analysis are R Markdown ready with Glimma 2.0. NAR Genom Bioinform. 2021; 3 (4): lqab116. DOI: 10.1093/nargab/lqab116

399. Gauthier J, Drezen J-M, Herniou EA. The recurrent domestication of viruses: major evolutionary transitions in parasitic wasps. Parasitology. 2018; 145 (6): 713-23. DOI: 10.1017/S0031182017000725

400. Pasquier-Barre F, Dupuy C, Huguet E, Monteiro F, Moreau A*, et al.* Polydnavirus replication: the EP1 segment of the parasitoid wasp Cotesia congregata is amplified within a larger precursor molecule. J Gen Virol. 2002; 83 (8): 2035-45. DOI: 10.1099/0022-1317-83-8-2035

401. Whitfield JB, Cameron SA, Huson DH, Steel MA. Filtered Z-closure supernetworks for extracting and visualizing recurrent signal from incongruent gene trees. Syst Biol. 2008; 57 (6): 939-47. DOI: 10.1080/10635150802552849

402. Dupuy C, Huguet E, J.M. D. Unfolding the evolutionary story of polydnaviruses. Virus Res. 2006; 117: 81-9. DOI: 10.1016/j.virusres.2006.01.001

403. Chevignon G, Periquet G, Gyapay G, Vega-Czarny N, Musset K*, et al.* Cotesia congregata bracovirus circles encoding PTP and Ankyrin genes integrate into the DNA of parasitized Manduca sexta hemocytes. J Virol. 2018; 92 (15): 10.1128/jvi. 00438-18. DOI: 10.1128/JVI.00438-18

404. Chevignon G, Cambier S, Da Silva C, Poulain J, Drezen JM*, et al.* Transcriptomic response of Manduca sexta immune tissues to parasitization by the bracovirus associated wasp Cotesia congregata. Insect Biochem Mol Biol. 2015; 62: 86-99. DOI: 10.1016/j.ibmb.2014.12.008

405. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH*, et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017; 27 (5): 722-36. DOI: 10.1101/gr.215087.116

406. Seppey M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. Methods Mol Biol. 2019: 227-45. DOI: 10.1007/978-1-4939-9173-0_14

407. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013; 29 (8): 1072-5. DOI: 10.1093/bioinformatics/btt086

408. Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. Trends Genet. 2000; 16 (6): 276-7. DOI: 10.1016/s0168-9525(00)02024-2

409. Chen Y, Zhang Y, Wang AY, Gao M, Chong Z. Accurate long-read de novo assembly evaluation with Inspector. Genome Biol. 2021; 22 (1): 312. DOI: 10.1186/s13059-021-02527-4

410. Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC bioinformatics. 2018; 19: 1-10. DOI: 10.1186/s12859-018-2485-7

411. Laetsch DR, Blaxter ML. BlobTools: Interrogation of genome assemblies. F1000Research. 2017; 6: 1287.

412. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26 (6): 841-2. DOI: 10.1093/bioinformatics/btq033

413. Buels R, Yao E, Diesh CM, Hayes RD, Munoz-Torres M*, et al.* JBrowse: a dynamic web platform for genome visualization and analysis. Genome Biol. 2016; 17: 1-12. DOI: 10.1186/s13059-016-0924-1

414. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013; 14 (2): 178-92. DOI: 10.1093/bib/bbs017

415. Song Y, Wang J. ggcoverage: an R package to visualize and annotate genome coverage for various NGS data. BMC bioinformatics. 2023; 24 (1): 309. DOI: 10.1186/s12859-023-05438-2

416. Wickham H, Wickham H. Data analysis. ggplot2. Use R!: Springer International Publishing; 2016.

417. Zeng C-W, Sheu J-C, Tsai H-J. Genomic Structure, Protein Character, Phylogenic Implication, and Embryonic Expression Pattern of a Zebrafish New Member of Zinc Finger BED-Type Gene Family. Genes. 2023; 14 (1): 179. DOI: 10.3390/genes14010179

418. Gornard S, Venon P, Lasfont F, Balliau T, Kaiser L*, et al.* Characterizing virulence differences in a parasitoid wasp through comparative transcriptomic and proteomic. BMC genomics. 2024; 25 (1): 940. DOI: 10.1186/s12864-024-10694-4

419. Suderman RJ, Pruijssers AJ, Strand MR. Protein tyrosine phosphatase-H2 from a polydnavirus induces apoptosis of insect cells. J Gen Virol. 2008; 89 (6): 1411-20. DOI: 10.1099/vir.0.2008/000307-0

420. Pruijssers AJ, Strand MR. PTP-H2 and PTP-H3 from Microplitis demolitor bracovirus localize to focal adhesions and are antiphagocytic in insect immune cells. J Virol. 2007; 81 (3): 1209-19. DOI: 10.1128/JVI.02189-06

421. Serbielle C, Dupas S, Perdereau E, Héricourt F, Dupuy C*, et al.* Evolutionary mechanisms driving the evolution of a large polydnavirus gene family coding for protein tyrosine phosphatases. BMC Evol Biol. 2012; 12: 1-19. DOI: 10.1186/1471-2148-12-253

422. Andersen JN, Mortensen OH, Peters GnH, Drake PG, Iversen LF*, et al.* Structural and evolutionary relationships among protein tyrosine phosphatase domains. Molecular and cellular biology. 2001; 21 (21): 7117-36. DOI: 10.1128/MCB.21.21.7117-7136.2001

423. Provost B, Varricchio P, Arana E, Espagne E, Falabella P*, et al.* Bracoviruses contain a large multigene family coding for protein tyrosine phosphatases. Journal of virology. 2004; 78 (23): 13090-103. DOI: 10.1128/JVI.78.23.13090-13103.2004

424. Savary S, Beckage N, Tan F, Periquet G, Drezen J-M. Excision of the polydnavirus chromosomal integrated EP1 sequence of the parasitoid wasp Cotesia congregata (Braconidae, Microgastinae) at potential recombinase binding sites. J Gen Virol. 1997; 78 (12): 3125-34. DOI: 10.1099/0022-1317-78-12-3125

425. Zhang F, Zhang J, Yang Y, Wu Y. A chromosome-level genome assembly for the beet armyworm (Spodoptera exigua) using PacBio and Hi-C sequencing. BioRxiv. 2019: 2019.12. 26.889121. DOI:

426. Schrader L, Schmitz J. The impact of transposable elements in adaptive evolution. Mol Ecol. 2019; 28 (6): 1537-49. DOI: 10.1111/mec.14794

427. Hayward A, Gilbert C. Transposable elements. Curr Biol. 2022; 32 (17): R904-R9. DOI: 10.1016/j.cub.2022.07.044

428. Welsh CL, Pandey P, Ahuja LG. Protein Tyrosine Phosphatases: A new paradigm in an old signaling system? Adv Cancer Res. 2021; 152: 263-303. DOI: 10.1016/bs.acr.2021.06.001

429. Harvey R, McCaughan C, Wise LM, Mercer AA, Fleming SB. Orf virus inhibits interferon stimulated gene expression and modulates the JAK/STAT signalling pathway. Virus Res. 2015; 208: 180-8. DOI: 10.1016/j.virusres.2015.06.014

430. Wang F, Xue R, Li X, Hu C, Xia Q. Characterization of a protein tyrosine phosphatase as a host factor promoting baculovirus replication in silkworm, Bombyx mori. Developmental & Comparative Immunology. 2016; 57: 31-7. DOI: 10.1016/j.dci.2015.12.002

431. Najarro P, Traktman P, Lewis JA. Vaccinia virus blocks gamma interferon signal transduction: viral VH1 phosphatase reverses Stat1 activation. J Virol. 2001; 75 (7): 3185-96. DOI: 10.1128/JVI.75.7.3185-3196.2001

432. Stoltz DB, Krell P, Summers MD, Vinson B. Polydnaviridae–a proposed family of insect viruses with segmented, double-stranded, circular DNA genomes. Intervirology. 1984; 21 (1): 1-4. DOI: 10.1159/000149497

433. Volkoff AN, Jouan V, Urbach S, Samain S, Bergoin M*, et al.* Analysis of virion structural components reveals vestiges of the ancestral ichnovirus genome. PLoS Pathog. 2010; 6 (5): e1000923. DOI: 10.1371/journal.ppat.1000923

434. Béliveau C, Cohen A, Stewart D, Periquet G, Djoumad A*, et al.* Genomic and proteomic analyses indicate that banchine and campoplegine polydnaviruses have similar, if not identical, viral ancestors. J Virol. 2015; 89 (17): 8909-21. DOI: 10.1128/jvi.01001-15

435. Burke GR, Hines HM, Sharanowski BJ. The Presence of Ancient Core Genes Reveals Endogenization from Diverse Viral Ancestors in Parasitoid Wasps. Genome Biol Evol. 2021; 13 (7): evab105. DOI: 10.1093/gbe/evab105

436. Koonin EV, Dolja VV, Krupovic M, Kuhn JH. Viruses defined by the position of the virosphere within the replicator space. Microbiol Mol Biol Rev. 2021; 85 (4): e00193-20. DOI: 10.1128/MMBR.00193-20

437. Combe M, Sanjuán R. Variability in the mutation rates of RNA viruses. Future Virology. 2014; 9 (6): 605-15. DOI: 10.2217/fvl.14.41

438. Duffy S, Shackelton LA, Holmes EC. Rates of evolutionary change in viruses: patterns and determinants. Nat Rev Genet. 2008; 9 (4): 267-76. DOI: 10.1038/nrg2323

439. Diaz-Munoz SL. Viral coinfection is shaped by host ecology and virus–virus interactions across diverse microbial taxa and environments. Virus evol. 2017; 3 (1): vex011. DOI: 10.1093/ve/vex011

440. Kariithi HM, Boucias DG, Murungi EK, Meki IK, Demirbaş-Uzel G*, et al.* Coevolution of hytrosaviruses and host immune responses. BMC microbiology. 2018; 18: 209-30. DOI: 10.1186/s12866-018-1296-3

441. Kaján GL, Doszpoly A, Tarján ZL, Vidovszky MZ, Papp T. Virus–host coevolution with a focus on animal and human DNA viruses. J Mol Evol. 2020; 88: 41-56. DOI: 10.1007/s00239-019-09913-4

442. Jackson TA, Crawford AM, Glare TR. *Oryctes* virus—time for a new look at a useful biocontrol agent. J Invertebr Pathol. 2005; 89 (1): 91-4. DOI: 10.1016/j.jip.2005.03.009

443. Manjeri G, Muhamad R, Tan SG. Oryctes rhinoceros beetles, an oil palm pest in Malaysia. Annu Res Rev Biol. 2014; 4 (22): 3429-39. DOI: 10.9734/ARRB/2014/11023

444. Bedford GO. Biology and management of palm dynastid beetles: recent advances. Annu Rev Entomol. 2013; 58 (1): 353-72. DOI: 10.1146/annurev-ento-120710-100547

445. Burand JP, Tan W. Mate preference and mating behavior of male Helicoverpa zea (Lepidoptera: Noctuidae) infected with the sexually transmitted insect virus Hz-2V. Annals of the Entomological Society of America. 2006; 99 (5): 969-73. DOI: 10.1603/0013-8746(2006)99[969:MPAMBO]2.0.CO;2

446. Kataoka C, Kaname Y, Taguwa S, Abe T, Fukuhara T*, et al.* Baculovirus GP64-mediated entry into mammalian cells. J Virol. 2012; 86 (5): 2610-20. DOI: 10.1128/jvi.06704-11

447. Ohkawa T, Volkman LE, Welch MD. Actin-based motility drives baculovirus transit to the nucleus and cell surface. J Cell Biol. 2010; 190 (2): 187-95. DOI: 10.1083/jcb.201001162

448. Funk CJ, Consigli RA. Evidence for zinc binding by two structural proteins of Plodia interpunctella granulosis virus. J Virol. 1992; 66 (5): 3168-71. DOI: 10.1128/JVI.66.5.3168-3171.1992

449. Guarino LA, Dong W, Jin J. In vitro activity of the baculovirus late expression factor LEF-5. J Virol. 2002; 76 (24): 12663-75. DOI: 10.1128/jvi.76.24.12663-12675.2002

450. Huang N, Wu W, Yang K, Passarelli AL, Rohrmann GF*, et al.* Baculovirus infection induces a DNA damage response that is required for efficient viral replication. J Virol. 2011; 85 (23): 12547-56. DOI: 10.1128/jvi.05766-11

451. Strand MR, Burke GR. Polydnavirus-wasp associations: evolution, genome organization, and function. Curr Opin Virol. 2013; 3 (5): 587-94. DOI: 10.1016/j.coviro.2013.06.004

452. Gottfried RS. Black death: Simon and Schuster; 2010.

453. Christakos G, Olea RA, Serre ML, Wang L-L, Yu H-L. Interdisciplinary public health reasoning and epidemic modelling: the case of black death: Springer; 2005.

454. Zhang Y, Dai X, Wang Q, Chen H, Meng W*, et al.* Transmission efficiency of the plague pathogen (Y. pestis) by the flea, Xenopsylla skrjabini, to mice and great gerbils. Parasit Vectors. 2015; 8: 1-11. DOI: 10.1186/s13071-015-0852-z

455. Kolata G. How pandemics end. The Covid-19 Reader: Routledge; 2020. p. 252-6.

456. Hinnebusch BJ. The evolution of flea-borne transmission in Yersinia pestis. Curr Issues Mol Biol. 2005; 7 (2): 197-212. DOI: 10.21775/cimb.007.197

457. Luckow VA, Lee S, Barry G, Olins P. Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in Escherichia coli. J Virol. 1993; 67 (8): 4566-79. DOI: 10.1128/jvi.67.8.4566-4579.1993

458. Holmstrom JE. Records and research in engineering and industrial science: a guide to the sources, processing and storekeeping of technical knowledge, with a chapter on translating. (No Title).

Summary

Nudiviruses (family *Nudiviridae*) are double-stranded (ds)DNA viruses that infect a variety of insect and crustacean hosts. In the past, many nudiviruses have been mischaracterised as baculoviruses (family *Baculoviridae*), mainly due to their similar virion morphologies, genetic material and pathogenicity to insects. With the advances of sequence-based virology, nudiviruses were eventually revealed to be monophyletically distinct from baculoviruses, however, they still share a number of core genes with each other and form phylogenetic sister clades. In 2014, nudiviruses were officially recognised as the family *Nudiviridae*, but nudivirus research is still outshined by the vast number of baculovirus-related studies. Although history shows that early science favoured baculoviruses, it was mainly their efficient use as biocontrol agents and the development of baculoviral molecular tools for protein production and gene function analyses that promoted baculovirus research. Some nudiviruses, such as Oryctes rhinoceros nudivirus (OrNV) are also used in biocontrol, but extensive studies on nudiviruses beyond OrNV are scarce. Potential biocontrol applications exist for other nudiviruses, like Helicoverpa zea nudivirus (HzNV-2), but, unlike OrNV, these viruses are still underexplored on an applied scale.

More recently, nudiviruses were grouped within the class *Naldaviricetes*, alongside four other virus families: *Baculoviridae*, *Hytrosaviridae*, *Nimaviridae*, and the very recently proposed Filamentoviridae. Within this class, the *Nudiviridae*, *Baculoviridae*, *Hytrosaviridae* and Filamentoviridae belong to the order *Lefavirales*, while *Nimaviridae* does not. Phylogenetic and evolutionary analyses have further revealed that bracoviruses – endogenised viruses in braconid wasps – originate from an ancestral nudivirus. This ancient nudivirus integrated its genetic material into the genome of an ancestral wasp approximately 100 million years ago (Mya), eventually evolving into bracoviruses, which function as a virulence delivery system within the wasp's ovaries to facilitate parasitism. However, debates about their taxonomic classification and placement are still ongoing, despite their confirmed evolutionary origin from nudiviruses. This thesis addressed the existing knowledge gaps surrounding nudiviruses by conducting a review of current literature, analysing available data, and performing experimental studies on nudivirus biology and pathology. Moreover, due to their evolutionary connection, one experimental chapter focused on the replication mechanisms of bracoviruses.

The general introduction (**Chapter 1**) served to provide a more global view on the history of virology and how viruses have originally been discovered, and how this original era of virology paved the way for the discovery of nudiviruses. However, this chapter also elucidated the advances that occurred in the field of virology, which eventually allowed for the distinction between nudiviruses and baculoviruses.

Following the general introduction, **Chapter 2** explored the phylogenetic, pathological, and genomic features of nudiviruses, baculoviruses, and bracoviruses. Although bracoviruses were initially classified under *Polydnaviridae*, phylogenetic analysis based on nudiviral core genes indicated that bracoviruses group within *Nudiviridae*. This chapter also reviewed the process of nudiviral host genome integration – a feature absent in baculoviruses – and evaluated core gene conservation across nudiviruses, baculoviruses, and hytrosaviruses (family *Hytrosaviridae*).

In **Chapter 3**, a data-driven virus discovery (DDVD) study expanded the known diversity of nudiviruses by examining publicly available sequencing data of insects and crustaceans for hidden nudiviral sequences. This approach led to the identification of eight novel nudivirus genomes from ectoparasitic insect hosts (lice and a flea) and the retrieval of several partial nudivirus gene sets from other insect and crustacean datasets. The newly assembled nudivirus sequences prompted an updated phylogenetic analysis comprising 49 nudiviruses, which was further complemented by examining

nudiviral gene synteny, molecular dating estimates, and evolutionary rates in both free-living nudiviruses and endogenised bracoviruses. The molecular dating analysis estimated that the most recent common ancestor (MRCA) of *Nudiviridae* and *Baculoviridae* diverged approximately 378 Mya, while the MRCA of all nudiviruses dates to around 280 Mya. The age estimates also supported the hypothesis that crustacean nudiviruses originated from an ancestral insect nudivirus approximately 143 Mya. Moreover, the louse-associated nudiviruses clearly branched as a distinct monophyletic group within the *Nudiviridae*, but divergent gene syntenies prompted division of these into two new genera: Zetanudivirus and Etanudivirus. In the end of this chapter, the geographic distribution and ecological implications of nudiviruses were evaluated, while also highlighting their potential significance for biocontrol in the veterinary sector.

Chapter 4 serves as an experimental chapter to introduce Heliothis zea nudivirus 1 (HzNV-1), a close relative to HzNV-2. Although the genomes of these two viruses are nearly identical, they differ greatly in their pathological abilities. While HzNV-2 is sexually transmitted in lepidopteran host populations, and can cause sterility in host progeny, HzNV-1 lost its ability to infect actual insects and is now restricted to cell lines. Nonetheless, due to its convenience as a controlled virus system, the experimental work of this chapter focused on HzNV-1 and its behaviour in a cell line derived from *Helicoverpa zea* ovaries (HZ-AM1). This chapter also laid the foundation for the experimental work presented in the subsequent chapter. The initial experimental work included the purification of the HzNV-1 stock from potential contaminants and the analysis of HzNV-1's cytopathology in the HZ-AM1 cells. With the HzNV-1 system established, electron microscopy (EM) was employed to visualise HzNV-1 at different stages of infection. This EM approach suggested that HzNV-1 employs a macropinocytosis-like mechanism to enter the host cell, and an inhibitory assay with impramine, followed by quantitative PCR (qPCR) showed that viral DNA levels were significantly reduced at 24 hpi when macropinocytosis was inhibited.

Chapter 5 explored the transcriptional dynamics of HzNV-1 infection in Hz-AM1 cells over time (3, 6, 9, 12, and 24 hpi). RNA extracted from mock-infected and virus-infected cells was subjected to RNA sequencing (RNA-seq), and the data were analysed to characterise expression profiles of the 154 HzNV-1 genes and host differentially expressed genes (DEGs) across different infection stages. Hierarchical clustering revealed four temporal phases of HzNV-1 gene expression, showing that replication- and transcription-associated genes were expressed prior to structural proteins needed for virion assembly. In the host, 570 distinct DEGs were identified and clustered into functional interaction networks associated with protein processing and other metabolic pathways that the virus likely hijacks to complete its infection cycle. Interestingly, HzNV-1 infection also led to major impairment of genes encoding for proteins involved in nuclear integrity, which was supported by EM-observed nuclear disintegration in HzNV-1-infected cells towards the end of infection.

Chapter 6 examined the replication mechanisms of two bracoviruses – Cotesia congregata bracovirus (CcBV) and Toxoneuron nigriceps bracovirus (TnBV). During the process of endogenization in the wasp genome, the nudiviral ancestors of bracoviruses lost several core genes, including the viral DNA polymerase, which made them dependent on their wasp hosts' replication machinery. The experimental studies in this chapter elucidated the mechanisms of bracoviral DNA amplification from the proviral loci (PL) of the wasp genome, while focusing on replication unit motifs (RUMs) that determine where and how DNA amplification of CcBV is initiated. By using PacBio long-read sequencing, this study confirmed most previously proposed RUMs in CcBV. In TnBV, where RUMs have not been described yet, amplification boundaries were instead commonly flanked by direct repeat junctions (DRJs), which are known for their role in DNA circularisation and excision. These findings suggested that TnBV first excises its DNA from the wasp genome and then amplifies it, potentially using rolling circle

amplification. In this context, the divergent replication strategies between microgastrine bracoviruses (like CcBV) and cardiochiline bracoviruses (like TnBV) may be explained by distinct evolutionary adaptations to their respective wasp hosts.

In conclusion, this thesis offers a comprehensive overview on the recent advances of nudivirus research, by reviewing available literature, screening genomic databases, and providing newly obtained experimental insights into their interaction with host cells and their pathology. Furthermore, this thesis emphasises the potential that nudiviruses hold and the future directions that could be taken to expand our understanding of the nudivirus nexus.

Acknowledgements

As I reach the end of this incredible journey, I find it difficult to put into words just how grateful I am to all the people who have accompanied and supported me along the way. A PhD is not just a personal endeavour; it is shaped by the countless interactions, encouragements, and guidance received from mentors, colleagues, friends, and family. Every conversation, every moment of support, and every shared experience has left a mark on this work, and for that, I am deeply thankful.
I want to begin with a quote from the German movie *Der Junge muss an die frische Luft*:

„Es hat eine Weile gedauert bis ich es verstehen konnte. Das alles ist es was ich bin. Ich bin meine Mutter. Mein Vater. Meine Geschwister und meine Großeltern. Ich bin ihr Lachen, und ihr Schmerz. … Ich bin die Richtung in die mich meine Mutter im Kinderwagen geschoben hat. Ich bin die gescheckte Kuh auf der Weide. Das gelbe Korn auf dem Feld. Und der rote Mohn am Wegesrand. Ich bin der wolkenlose Himmel. Ich bin wach".

First of all, **Monique**, I am incredibly thankful to have had you as my *Doktormutter*. Throughout my PhD, I never felt alone in my project – even though I often enjoyed immersing myself in solitude – because I always knew I could count on you. Your insightfulness and empathy are qualities that every PhD student would be lucky to have in a supervisor. Not only did you guide me through my time as an INSECT DOCTOR, but you also gave me the freedom to explore my scientific ideas and creativity, something I deeply appreciate. I hope to stay in touch with you after my graduation, not only as my former promotor but also as a good friend.
Jean-Michel, *mon père de doctorat*, I am immensely grateful for your mentorship and guidance, especially during my time in France. You always had an open ear and supported my decisions for the project, even when they were not easy ones. It was encouraging to have you as a promotor, knowing that you trusted my scientific instincts. Your passion for science and research is truly inspiring, and I am grateful to have had the chance to learn from you. Your enthusiasm for research is inspiring, and I hope to carry that same energy into my own career.
Annie, thank you for being my co-promotor. I truly appreciate your dedication in reviewing our work and providing invaluable feedback – always with impressive speed and efficiency. Your sharp eye for detail and constructive input have been immensely helpful. I wish you and your family all the best and look forward to future collaborations on nudiviruses.
Els, **Dorothy**, **Marleen**, **Corinne**, **Gwen**, **Dennis**, **Mark**, and **Joppe**, thank you for all your help, patience, and support in keeping the lab running. Without you, many of my experiments would not have been possible. Your dedication and expertise have made a huge difference in my work, and I am incredibly grateful for the many times you helped me troubleshoot experiments and ensured that everything ran smoothly.
Jan, **Jelmer**, **Marcel**, and **Esther**, thank you for always being there to help with my EM-related experiments and for teaching me how to use the TEM. Your guidance made all the difference, and I am truly grateful for the time and effort you invested in helping me acquire the skills I needed.
Melanie and **Marry**, thank you for handling all the organizational aspects and planning with such efficiency. You pulled everything together behind the scenes, and I truly appreciate it. Your ability to make everything look effortless is something I greatly admire.
Anne, **Max**, and **Christine**, thank you for your great work and persistence in assisting with my PhD project. Your contributions and dedication meant a lot to me, and I truly hope you continue your journeys in science with the same curiosity and determination.

To everyone from the Arbovirus Group, past or current members – **Gorben**, thank you for the stimulating conversations on blood-feeding insects. I will keep looking for nudiviruses in mosquitoes! **Jelke**, **Joyce**, **Tessy**, **Jerome**, and **Linda** van O., thanks for the great time we had at the ECV in Gdańsk – true brothers and sisters in arms, *pewpew*. Also, a big thank you to **Sophie**, **Linda** de J., and **Dennis** for the laughs and good times.

To everyone from the Plant Virology Group – **Emilyn** and **Cristina**, thank you for your guidance during the practical courses we supervised together and for the nice talks. **Richard**, **René**, **Mitsuru**, **Mandy**, **Rizko**, **Janna**, **Mathjis**, **Sjors**, and every past or current member, thank you for your support and for making the lab a great place to work.

In the Insect Virology Group – **Astrid** and **Vera**, thank you for always having an open ear and door. Your experience and support helped me improve as a scientist, and I much enjoyed the times when we got together at your homes with everyone. **Ahmed**, my lab bro and fellow anime enthusiast – thank you for all the laughs, deep scientific discussions, and passionate padel matches. I wish you all the best for your future career path, but I am certain you will have great success wherever you go. Let us stay in touch! **Esmée**, **Laura** and **Sofia**, you've just started your PhD journeys, so we have not shared many moments together, but I always enjoyed our conversations and I am sure you will do fantastically in the Insect Virology group. I wish you all the best! **Simone**, thank you for the unforgettable birthday parties at your place – especially when I had just moved to Wageningen and didn't know anyone. You made me feel welcome and helped me connect with so many great people.

I also want to thank everyone at IRBI for their inclusiveness and kindness while I stayed in France. **Elisabeth**, thank you for always checking on us in our office in Tours and staying for those stimulating chitchats, and the BBQ at your place. Also, thank you to **Carlos** for catching insects at night with us during our INSECT DOCTORS meeting in France. **Camille**, **Fernando**, **Matthieu**, **Karine**, **Christophe**, and everyone at IRBI, thank you for the good times!

Tom, **Ben**, **Caroline**, and everyone else at Koppert Biological Systems – thank you for your openness and support during my few months conducting experiments. I felt very comfortable working with you and always well-supported!

Jamie, my fellow nudivirus enthusiast from across the North Sea – thank you for the collaborations and for always having an open ear, both personally and scientifically. I look forward to future projects with you on discovering new nudiviruses and beyond!

Sara, thank you for introducing me to Wageningen and to the apartment I called home for more than four years. You made my arrival so much easier, and even during the pandemic, you ensured I never felt alone. From Netflix nights to empanada-making, long bike rides, or racing down the dike on a sled with a beer in hand, I feel lucky to have had you as a flatmate and friend. *Bueno*!

Of course, a huge thank you goes out to my two paranymphs who took the time and effort to help me organise so many things around my PhD graduation.
Annamaria, my friend from the opposite side of the table. My PhD journey and time in Wageningen would have been so much less bright without you. Whether it was road trips to the beach or the Nederrijn during Dutch summer, backyard drinks, or simply hanging out in Lombardi, you made everything better with your joyful presence. I wish you all the best for the last stretch of your PhD and I'm excited to see where your path will lead you!
Sam, my friend and former INSECT DOCTORS companion. Whether it was watching *Castlevania*, baking lye bread, or going for a run, I've truly enjoyed our scientific mind trips and deep discussions about where to search for new viruses. Talking with you about science on such a passionate level has always been a pleasure. I'm looking forward doing future side projects with you, and I wish you all the best for your future endeavours. One day, we will catch those mole fleas!
Again, thank you both, for standing by my side at the final stage of my PhD.

To all my other fellow ESRs from the INSECT DOCTORS program – thank you for the unforgettable memories we've created across different countries and occasions. It feels like just yesterday we met for the first time in Paris, and now here we are, years later.

Jozsef, thank you for being a worthy *Risk* opponent and always reminding me to have my daily dose of Pálinka. Your wedding in Hungary was superb, and I wish you and Nora all the best for your future together.
Alessandro, thank you for the honour of being your *partynymph*. Binge-watching *Demon Slayer*, cooking carbonara, discussing the newest LEGO models, and battling it out on the Nintendo Switch over pizza are cherished memories of mine. I wish you and Linda all the best for your new jobs and future.
Carlotta, thank you for bringing calmness and laughter into my life, and for our coffee chats in the Lombardi kitchen. We need to go to another water park together sometime, like the one in Debrecen!
Luis, thank you for the challenging padel and squash games, and for taking such good care of my apartment while I was in France. I wish you and Marina all the best for the future!
Melissa, my fellow *Game Grumps* enthusiast – thank you for all our chats and laughter over coffee and in the lab. All the best for the last stretch of your PhD project in Wageningen!
Jennifer, ich hoffe du arbeitest nach wie vor an deinem Deutsch, so dass wir bei unserem nächsten Treffen ausgelassen schnacken können! I wish you success for your new position in Copenhagen and I'm curious to see where your work will take you!
Lim, thank you for teaching us all the ways of the Nanopore in Paris. Even though we only worked together in the lab that one time at the course in Paris, I always enjoyed our conversations. I hope we find the time to collaborate on some future projects together.
Robert and **Loretta**, thank you for making my time in Tours even more enjoyable. I wish you and your families all the best, and I hope to see you soon!
Hannah, thank you for always being up for a *Klönschnack*. I hope we can soon meet again in Wageningen, whether for drinks or walks along the dyke – surely the next BBQ weather is right around the corner. Good luck for the last stretch of your PhD project!
Pascal, it was an honour to have walked the path of an INSECT DOCTOR with you, and I wish you all the best for your career in Switzerland!
Edouard, thank you for always making me laugh. I still remember having to leave the lecture in Paris because I couldn't hold it together – STEGOSAURUS!!!
Anna, thank you for making every INSECT DOCTORS get-together unforgettable with your legendary speeches. I wish you and your partner all the best for your life in the Netherlands!
Even after this project, I hope we all stay connected – maybe even organise an INSECT DOCTORS reunion in Italy, France, or maybe the high north of Germany.
Of course, I also want to thank the entire INSECT DOCTORS consortium for mentoring and guiding us on our journey. **Jørgen**, **Helen**, **Margreet** and **Nicolai**, thank you for always keeping us on track with organisation-related matters and ensuring the whole project ran smoothly.

Heartfelt thanks go out to my family and friends in Germany. You are the feeling that engulfs me when I think of home, and what I always look forward to when I return to Nordfriesland.
Marec and **Maria**, I am grateful to have you as my siblings and that we have such a harmonic relationship with each other. Looking at the paths each of us have taken in our lives and where we each stand now, I am very proud of us. I know I can always count on you, and likewise, you will always have my back if you need me.
Aaron, my best friend. I deeply treasure our friendship. From the first Pokémon generation to now, so much has changed, yet our connection has remained. Even though life has taken us in different directions, I'm deeply grateful that we're still in touch. I'm incredibly happy for you, Anita, and Lina-Louisa, and of course, for your families. I'm wishing you endless happiness as you create memories in your new family home!
Mama, words cannot fully express my gratitude for you. You have always supported me and given me nothing but unconditional love. Thank you for pointing me in the right direction from a very young age and supporting the decisions and paths I took since then. I wouldn't be the person I am today if it wasn't for you, and it was your meaningful influence that enabled me to discover and pursue my passion for science, which finally led to this doctoral thesis. In diesem Sinne, danke, Mama, ich hab dich lieb.
Michael, danke, dass du ein maßgeblicher Teil meiner Kindheit warst und bis heute bist. In vielen Gewohnheiten, die mich täglich begleiten, erkenne ich deinen väterlichen Einfluss wieder. Ich freue mich schon auf Südseecamp und Rammstein dieses Jahr zusammen!

Oma, auch wenn uns oft viele Kilometer getrennt haben, hoffe ich, dass ich dir mit all den Souvenirs aus meinen Reisen ein kleines Stück von meinen Abenteuern mitbringen konnte. Danke, dass du immer für mich da bist. Ich habe dich lieb!
Dirk, **Silke**, **Christopher** und **Jule**, danke auch an Euch, dass ihr stets Anteilnahme und Interesse an meiner Zeit als Doktorand im Ausland genommen habt, ob es nun über Telefonate oder Textnachrichten war.

And last but certainly not least, I want to express my heartfelt gratitude to my girlfriend. **Sophia**, having you in my life has been one of my greatest joys, and I am so grateful that you walked this final stretch of my PhD journey with me. You always know when to remind me to take a step back, to breathe, and to find balance – something I needed more than ever in the last phase of this project. You are my dearest Player-2, whether we're navigating stormy seas in *Ship of Fools*, strategizing over *Dice Throne*, or climbing mountainous heights. No matter the adventure, I am endlessly thankful to have you by my side. With all my love, Jirka.

To everyone who has been part of this journey – thank you.

About the author

Born on the North Sea coast of Germany, **Jirka Manuel Petersen** grew up in a village of fewer than 1,000 people, where his days were filled with soccer training, Nintendo 64 marathons, and camping trips. His academic journey began in Hamburg, where he earned both his Bachelor's degree in Biology and his Master's degree in Molecular Plant Science. However, his path took a brief detour when he explored illustration and design for a year, considering a career in video game design. Eventually, his passion for biology won out, but to this day, he integrates his artistic skills into both his research and personal projects.

Jirka was convinced from an early age that he would become a marine biologist – one reason why, as a teenager, he interned at an aquarium on the island of Sylt, close to his home town. Yet, as he progressed through his studies, his focus shifted toward molecular biology, particularly virology, which sparked a fascination with infection biology and pathology. After completing his Master's thesis on a hypovirulent mycovirus, he set his sights beyond Germany, eager to expand his academic horizons by pursuing a PhD abroad.

His journey led him to Wageningen in the Netherlands, where he joined the European joint doctoral program INSECT DOCTORS, investigating the replication mechanisms of nudiviruses (*Nudiviridae*). His research took him across Europe, spending months at research institutes in Tours (France) and Rotterdam (Netherlands), conducting experiments and collaborating with fellow scientists. Yet, no matter where he travels, something always draws him back to the North Sea – perhaps explaining his interest in nudiviruses, which infect not only insects but also aquatic crustaceans, including the brown shrimp (*Crangon crangon*), a familiar resident of his home waters.

During his PhD, Jirka became deeply engaged with his research on nudiviruses and their diversity. He believes that a vast, hidden world of undiscovered nudiviruses exists in ectoparasitic insects, arachnids and other arthropods – viruses that could reveal unknown pathological, ecological, and evolutionary relationships, with roots tracing back to the age of the dinosaurs. If given unlimited resources, he would embark on a global arthropod collection expedition to uncover this viral diversity. Beyond research, Jirka enjoys teaching both theoretical and practical courses and plays an active role in organising a newly launched virology seminar group under the Netherlands Association of Virus Ecology (NAVE).

In his free time, he is learning to play the kalimba and piano and enjoys both indoor and outdoor sports, including running, swimming, and padel. He stays connected to family and friends through online video games and is a devoted board game enthusiast – most recently obsessed with *Dice Throne* and *Pandemic*. His comfort media consists of 70s and 80s movies and music, and when he first visited the U.S., he made it a priority to see the *Ghostbusters* firehouse in New York (left photo). At home, he cares for a small indoor jungle of houseplants and dreams of one day having a terrarium full of carnivorous plants. Despite the often grey and windy weather of the North Sea coast, he can spend hours by the steel-blue water, letting his thoughts drift and finding inspiration in nature.

List of publications

Petersen JM, Bryon A, Bézier A, Drezen J-M, van Oers MM. (2025) Transcriptional dynamics during Heliothis zea nudivirus 1 infection in an ovarian cell line from Helicoverpa zea. Journal of General Virology 106: 002066, DOI: 10.1099/jgv.0.002066

Petersen JM, Burgess AL, van Oers MM, Herniou EA, Bojko J. (2024) Nudiviruses in free-living and parasitic arthropods: evolutionary taxonomy. Trends in Parasitology 40: 744-762, DOI: 10.1016/j.pt.2024.06.009

Petersen JM, Bézier A, Drezen J-M, van Oers MM. (2022) The naked truth: An updated review on nudiviruses and their relationship to bracoviruses and baculoviruses. Journal of Invertebrate Pathology 189: 107718, DOI: 10.1016/j.jip.2022.107718

Ros VI, Panziera D, Nalcacioglu R, **Petersen JM**, Ryabov E, van Oers MM. (2022) Viral diseases of insects. In: Rowley AF, Coates CJ, Whitten MW (Eds.), Invertebrate Pathology Feb: 249-285.

Lutz T, **Petersen JM**, Yanık C, de Oliveira C, Heinze C. (2021) Processing of the capsid proteins of the Betachrysovirus Fusarium graminearum virus-China 9 (FgV-ch9). Virology 563: 50-57. DOI: 10.1016/j.virol.2021.08.007

Under review

Dearlove E, van Gestel CAM, Loureiro S, Svendsen C, Lloyd M, Mugo-Kamiri L, **Petersen JM**, *et al*. Determining multiple stressor interactions in mass reared insects based on principles of ecotoxicology (Under review in *Journal of Insects as Food and Feed*)

Santos ER, **Petersen JM**, Santana TDD, Harrison RL, Ardisson-Araújo DMP, Alphanudiviral segments found in transcriptomes of the two-spotted spider mite, Tetranychus urticae (Acari: Tetranychidae) (Under review in *Virus genes*)

PE&RC Training and Education Statement

With the training and education activities listed below the PhD candidate has complied with the requirements set by the Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 30 ECTS (= 20 weeks of activities)

Review/project proposal (6 ECTS)

- The naked truth: An updated review on nudiviruses and their relationship to bracoviruses and baculoviruses (Review)
- Nudivirus replication mechanisms and gene functions (Project proposal)

Post-graduate courses (8.5 ECTS)

- Concepts and terms in insect pathology and their placement in the general context of diseases and health University of Copenhagen, Online (2020)
- The value and pitfalls of metagenomics in pathogen detection and discovery, Centre National de la Recherche Scientifique, Paris, France (2021)
- Using mixture models to predict expected joint effects of multiple pathogens, UK Centre for Ecology & Hydrology, Aveiro, Portugal (2022)
- Introduction to R and R Studio, PE&RC, Online (2020)
- Introduction to the analysis of NGS data, Vlaams Instituut voor Biotechnologie, Online (2021)
- Phylogenetic comparative methods, Physalia, Online (2020)

Laboratory training and working visits (4.5 ECTS)

- Laboratory methods (diagnostics, bio-assays) in insect pathology across organisms, Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement, Paris, France (2021)
- Basic Transmission Electron Microscopy, WUR (2021)

Invited review of journal manuscripts (4 ECTS)

- MDPI – Viruses, Envelope-Fusion-Syncytium Formation in Microplitis bicoloratus bracovirus Maturation (2022)
- Journal of General Virology, Complete genome sequence of Penaeus vannamei singly enveloped nuclear polyhedrosis virus (PvSNPV) reveals more similarity to nudiviruses than baculoviruses (2022)
- Journal of Invertebrate Pathology, Rediscovering “Baculovirus-A” (Johnson, 1976): the complete genome of ‘Callinectes sapidus nudivirus’ (Nudiviridae: Gammanudivirus) (2022)
- Journal of Invertebrate Pathology, Exploring the spectrum of entomopathogenic viruses (2024)

Competence, skills and career-oriented activities (6.9 ECTS)

- Scientific Publishing, WGS (2020)
- Workshop on Individual research projects, WUR (2021)
- Scientific writing, WGS (2021)
- Writing Grant Proposals, WGS (2023)
- French 1, course level A1 on the CEFR, WUR, In’to Language, Online (2020)

Scientific Integrity/Ethics in science activities (1.5 ECTS)

- Ethics and Animal Sciences, WUR (2020)
- Scientific Integrity, WUR (2020)
- EMSTU-MIPTIS-SSBCV - Intégrité scientifique et recherche, École Doctorale Santé, Sciences Biologiques et Chimie du Vivant, Tours, France (2023)

PE&RC Annual meetings, seminars and PE&RC weekend/retreat (1.2 ECTS)

- PE&RC Afternoon event (2020)
- PE&RC First years online event (2020)

National scientific meetings, local seminars, and discussion groups (4.1 ECTS)

- NAVE Virus Interest Group (2022/2023)
- Insects for Food and Feed and the practical problems encountered (2023)
- 1st Annual Workshop of the Netherlands Association for Virus Ecology, Wageningen (2022)
- 3rd Annual Workshop of the Netherlands Association for Virus Ecology, Groningen (2024)
- "Insect Doctors" and stakeholder symposium, INRAE, France (2023)

International symposia, workshops and conferences (10.1 ECTS)

- International Congress on Invertebrate Pathology and Microbial Control, online, South Africa (2022)
- NEV Entomology Day, Ede-Wageningen, Netherlands (2022)
- 8th European Congress of Virology, Gdańsk, Poland (2023)
- SIP Annual Conference, Maryland, USA (2023)
- SIP Annual conference, Vienna, Austria (2024)

Lecturing/supervision of practicals/tutorials (8.1 ECTS)

- Molecular Virology – Case study, Polydnaviruses (2020, 2021, 2022, 2023)
- Molecular Virology – Case study, Herpesviruses (2022, 2023)
- Fundamental and Applied Virology – Practical Course (2023)

BSc/MSc thesis supervision (6 ECTS)

- Research topic 1: BSc thesis project: *Heliothis zea* nudivirus 1 cell entry and ORF1 modification by host proteases
- Research topic 2: MSc thesis project: Generation of St-2-1 aptamer-fused viral DNA constructs: a first step towards unravelling the DNA replication initiation of *Heliothis zea* nudivirus 1
- Research topic 3: MSc thesis project: Exploring the potential of a bacmid system for the HzNV-1 genome

The research described in this thesis was financially supported by the European Union's Horizon 2020 Research and Innovation Programme INSECT DOCTORS under the Marie Skłodowska-Curie grant agreement No 859850.

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover design by Jirka Manuel Petersen – the cover was designed and generated using Ulead PhotoImpact X3

Printed by ProefschriftMaken