Charles University

Faculty of Science

Study programme:

Molecular and Cellular Biology, Genetics and Virology

Branch of study:

D-MBBGV

Mgr. Dominika Kadlečková

Virome of Honey Bee
Virom včely medonosné

Type of thesis

Doctoral thesis

Supervisor:

doc. RNDr. Ruth Tachezy, Ph.D.

Prague, 2024

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracoval/a samostatně a že jsem uvedl/a všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 21.06.2024

Podpis

This study was funded by the Project National Institute of Virology and Bacteriology (Programme EXCELES, ID Project No. LX22NPO5103) - Funded by the European Union - Next Generation EU. And by the Ministry of Agriculture of the Czech Republic, Project No. QK1910018 (NAZV).

Acknowledgement

It always seemed to me that my existence consisted purely and exclusively of nothing but the most outrageous nonsense. That has been a problem I keep rediscovering and solving only to start again. It was never easy; I was not easy and I know that. But that is the reason I appreciate the help more. Because it is not easy to ask for help, and much harder to get it when you need it the most.

I want to thank everybody who helped me, with or without asking. It is the people who cared at least a little when they had no reason to. People who heard what I had to say and more importantly what I did not say. I want to thank them for hearing instead of listening, seeing instead of looking, and bearing me at my worst. Because it means a lot to me. And if you're not sure if you are on the list you probably are because even the smallest gesture of kindness means everything to me in this life nonsense.

Abstract

We explored the honey bee (*Apis mellifera*) virome through extensive metagenomic analysis across two major projects, focusing on virome stability and longitudinal changes. The first project, which included 39 samples from Czechia, aimed to understand virome stability within triplicates, revealing varying viral compositions of honey bee-infecting viruses and was published in 2022. The second project extended the analysis to 48 samples over three years, we explored recent tools in viral metagenomics bioinformatics which led to the discovery of novel DNA viruses: Apis mellifera filamentous-like virus (AmFLV) and Apis mellifera nudivirus (AmNV), along with nine new genomes from the *Parvoviridae* family, tentatively named Bee densoviruses 1 to 9. The longitudinal study highlighted significant viral abundance variability that may be influenced by factors such as pesticide exposure. An analysis comparing viral detection by proteomics and metagenomics was also conducted. These findings contribute to the understanding of honey bee virome dynamics and underscore the need for continued research into viral interactions and their ecological implications.

Keywords: metagenomics, viruses, honey bee, vMAGs

Abstrakt

Prozkoumali jsme virom včel medonosných (Apis mellifera) prostřednictvím rozsáhlé metagenomické analýzy v rámci dvou hlavních projektů, zaměřených na stabilitu viromu a jeho dlouhodobé změny. První projekt, který zahrnoval 39 vzorků z České republiky, si kladl za cíl pochopit stabilitu viromu v rámci triplikátů, přičemž byly odhaleny různé virové složení včelích virů a výsledky byly publikovány v roce 2022. Druhý projekt rozšířil analýzu na 48 vzorků během tří let, kde jsme prozkoumali nové nástroje v bioinformatice virové metagenomiky, což vedlo k objevu nových DNA virů: Apis mellifera filamentous-like virus (AmFLV) a Apis mellifera nudivirus (AmNV), spolu s devíti novými genomy z rodiny Parvoviridae, předběžně pojmenovanými Bee densoviruses 1 až 9. Dlouhodobá studie zdůraznila významnou variabilitu virové abundance, která může být ovlivněna faktory jako je expozice pesticidům. Rovněž byla provedena analýza porovnávající detekci virů pomocí proteomiky a metagenomiky. Tyto poznatky přispívají k pochopení dynamiky viromu včel medonosných a zdůrazňují potřebu pokračujícího výzkumu virových interakcí a jejich ekologických dopadů.

Klíčová slova: metagenomika, viry, včela, vMAGs

Abbreviations:

ABPV	Acute bee paralysis virus
ACE	Acetamiprid
AFV	Apis flavivirus
AmFLV	Apis mellifera filamentous-like virus
AmFV	Apis mellifera filamentous virus
AmNV	Apis mellifera nudivirus
BMLV	Bee macula-like virus
BQCV	Black queen cell virus
CBPV	Chronic bee paralysis virus
DWV	Deformed wing virus
IAPV	Israeli acute paralysis virus
IMI	Imidacloprid
ITR	Inverse terminal repeat
KBV	Kashmir bee virus
LSV	Lake Sinai viruses
NCBI	National centre for biotechnology information
NGS	Next generation sequencing
ORF	Open reading frame
PCR	Polymerase chain reaction
QC	Quality control
qPCR	Quantitative PCR
SBV	Sacbrood virus
vMAG	Viral metagenome-assembled genome

1. Introduction

Honey bees (*Apis mellifera* Linnaeus, 1758) are among the most vital pollinators in global ecosystems, playing a pivotal role in the pollination of a wide variety of plants, many of which are essential for human food production (1, 2). Beyond their ecological importance, honey bees have been domesticated and managed by humans for thousands of years, providing honey, beeswax, and other valuable products (3). However, in recent decades, honey bee populations have faced significant threats, with reports of colony collapses and large losses worldwide (1, 2, 4). Among the number of challenges, they face, viral infections stand out as one of the most insidious threats to honey bee health.

The term "virome" refers to the complete set of viruses present in and on an organism or at a certain place at a given moment (5). The honey bee virome, therefore, encompasses all the viruses that can infect honey bees, both those that cause overt disease and those that might exist in a more commensal or symbiotic relationship. Apart from eukaryotic viruses infecting bees the honey bee virome is composed of bacteriophages, transient viruses like those in pollen and viruses from other organisms living on/in bees (e.g., parasites). One of the most important characteristics of viromes is their fast mutability and diversity, making them perhaps the most variable component of the holobiont (6). It is very difficult to find a direct correlation between the very complex virome composition and the diseases that cause honey bee losses. However, several changes in the composition of the virome that play a role in disease development or progression were already identified (7).

Understanding the honey bee virome is crucial not only for the health and sustainability of honey bee populations but also for the broader ecosystems and agriculture that rely on them.

1.1. Factors responsible for losses of honey bees

The loss of honey bee populations is a multifaceted problem that defies simple attribution to a single cause (8, 9). Instead, it is a complex set of interacting biological and environmental factors that lead to losses (10). At the forefront of these contributing elements are several potential causes, ranging from pathogens like *Varroa destructor* with Deformed wing virus, the Acute bee paralysis virus to pesticides such as neonicotinoids, that have been banned in the EU to prevent bees and other pollinators losses (11–13). While each of these factors plays a role in the losses of honey bees, it is the intricate interplay and synergy among them that magnify the scope of the problem (overview in Fig. 1).

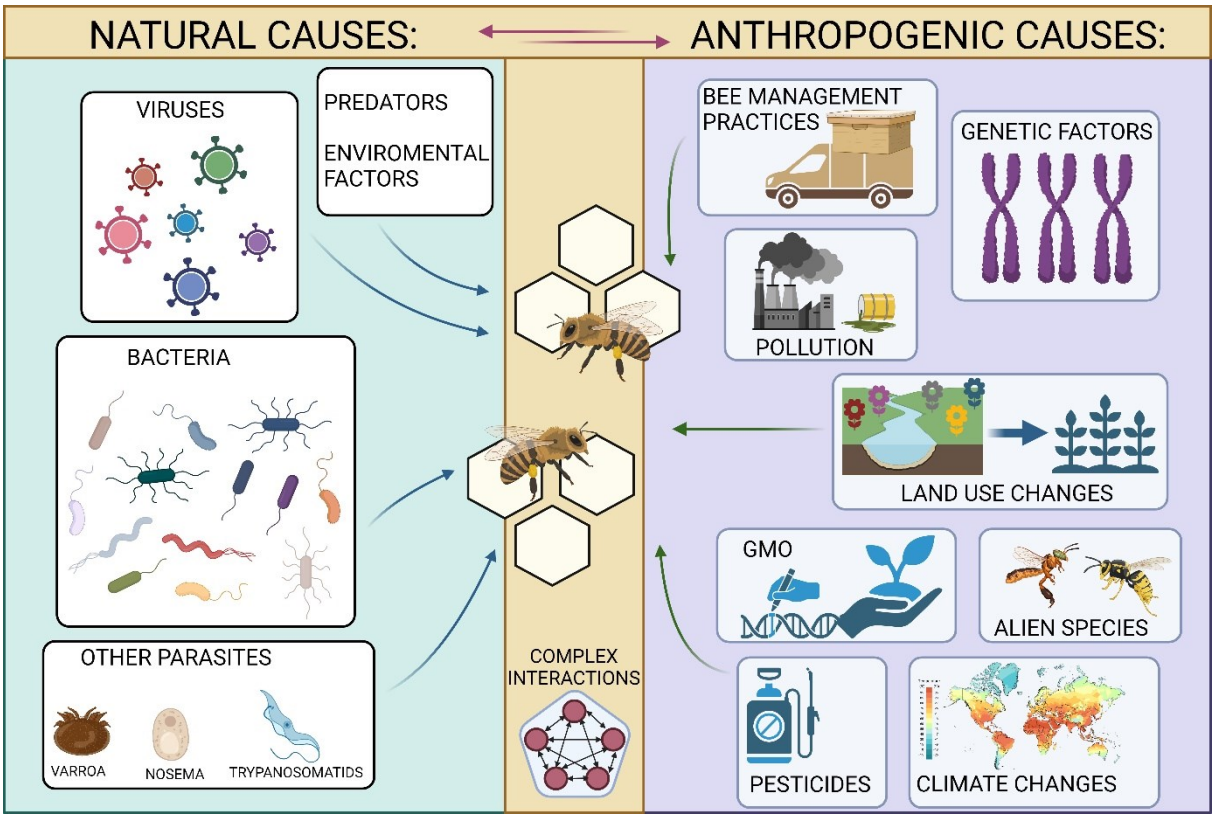


Figure 1- Overview of the factors responsible for honey bee losses. Created in BioRender.

Beyond natural stressors, anthropogenic factors exert significant pressure on honey bee colonies, exacerbating their losses. Pesticides, for example, have been identified as a high-risk driver, contributing to a high rate of mortality and altering various biological processes crucial for the bees' survival (see 1.1.1). Similarly, climate change poses a substantial threat, altering honey bee behaviour, physiology, and distribution, while inducing changes in the flora that are vital for their sustenance (14). The introduction of alien species further accelerates or complicates the issue by intensifying competition for food resources, leading to the decline of indigenous species and altering habitats (15–18).

Moreover, the proliferation of genetically modified crops (GMOs) has been linked to alterations in bee foraging behaviour (19), while land use and management practices result in habitat and forage loss, exacerbating competition among honey bees and wild bee populations (20). The management of bee colonies, including practices such as hybridization and migratory pollination, also contributes to the complex web of stressors faced by honey bees (21–23).

It is important to note that the interactions between these anthropogenic drivers and natural stressors are often poorly understood, representing a significant gap in current research (10, 24–27).

1.1.1. Pesticides

A significant portion of pesticides used in agriculture are also insecticides, posing a major threat to honey bee populations. These chemical agents, although tested before their widespread use, are harmful to honey bee colonies (28, 29), but even the adjuvants can be dangerous (30). However, it is important to acknowledge that the danger to honey bees is not solely confined to insecticides. Herbicides and fungicides, seemingly distant in the realm of chemical treatments, also harbour potential harm for honey bee colonies. Their adverse effects are often compounded when they interact with insecticides, creating a dangerous synergy (31) but the danger can be caused by fungicide alone (32, 33). For herbicides, the mechanism of harm can be attributed to their role in reshaping agricultural landscapes by nudging fields towards monocultures, but it was noted several times that they can be toxic by themselves (34, 35).

Additionally, honey bees grapple with a different category of chemicals, those administered to combat the *Varroa* mite. While designed to target the mite, these treatments can carry unintended consequences, as they may prove to be mildly toxic to honey bees themselves (36–38).

The losses of honey bees are further exacerbated by the intricate interplay among these chemicals. These cocktail effects of insecticides, herbicides, fungicides, and anti-parasitic treatments introduce a level of complexity that can result in unforeseen consequences for honey bee populations (39, 40). It is the synergy between these chemicals, their subtle interactions, and the resultant impact on honey bees that remains a subject of ongoing research and concern.

1.1.2. Varroa mite

Varroa destructor is the main factor contributing to the high annual losses of honey bee colonies (41–43). *Varroa* detrimental impact is particularly pronounced when considering its interactions with honey bee viruses (44, 45). *Varroa* can tip the balance from covert viral infection, characterized by asymptomatic and subclinical states, to overt infection, marked by severe symptoms, further described in Section 1.2. However, the precise mechanisms underpinning this shift remain a subject of ongoing inquiry, with several theories proposed but none definitively confirmed. From immunosuppression (46, 47), through protentional replication in vector (see 1.2.1 for more information about replication of DWV in mite) and ecological theories like Volterra's model (48) to toxins in *Varroa* saliva (49). The exact mechanisms driving this synergy between *Varroa* and viral infections remain a topic of active

research. It is clear that *Varroa* plays a pivotal role in the transmission of viruses, particularly DWV, and increasing the viral load within honey bee colonies.

The life cycle of *Varroa* is well-documented and, in part, contributes to its impact on honey bees. This ectoparasite reproduces within capped honey bee brood cells, where it feeds on developing honey bee pupae. Its life cycle is synchronized with that of the honey bee brood, allowing it to propagate effectively within colonies (50, 51).

One of the most concerning aspects of *Varroa* presence is its global spread. It has dispersed widely, from Apis ceranae originated in Asia (42, 52, 53). Notably, it has even appeared in recent years in Australia (54). What bears particular significance is the opportunity presented by the introduction of Varroa to Australia. With a baseline understanding of the health of Australian honey bee populations recorded before the arrival of *Varroa* (55), we are now positioned to monitor the evolving landscape of honey bee health and viral occurrence in the presence of this ectoparasite.

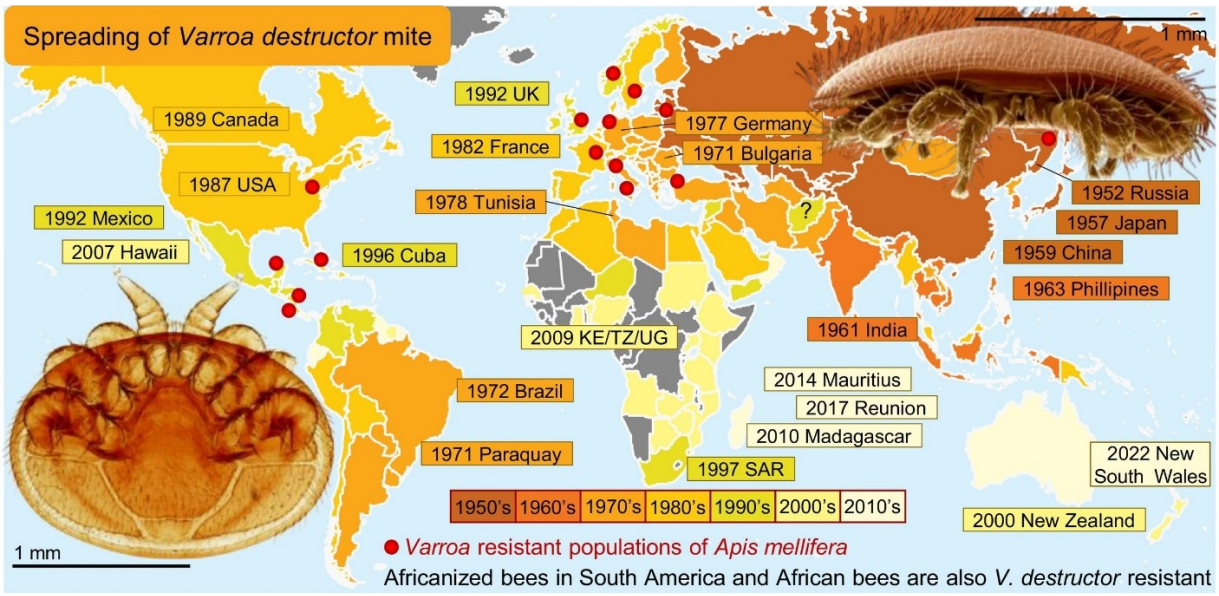


Figure 2- Spread of Varroa *from the 1950s to recent years. Taken from (54).*

1.1.3. Bacteria

Within the world of honey bee microbiota, a group of bacteria are the most abundant, collectively comprising over 95% of the entire bacterial community. These bacterial representatives include *Gilliamella*, *Snodgrassella*, *Lactobacillus* Firm 4, *Lactobacillus* Firm 5, and *Bifidobacterium* (56–59). These core members of the bacterial microbiome play pivotal roles in shaping the honey bee's microbial landscape, exerting a profound influence on their

health and well-being (60, 61). The second less abundant part is a few bacteria, mainly *Bartonella*, *Frischella*, and *Commensalibacter* (62).

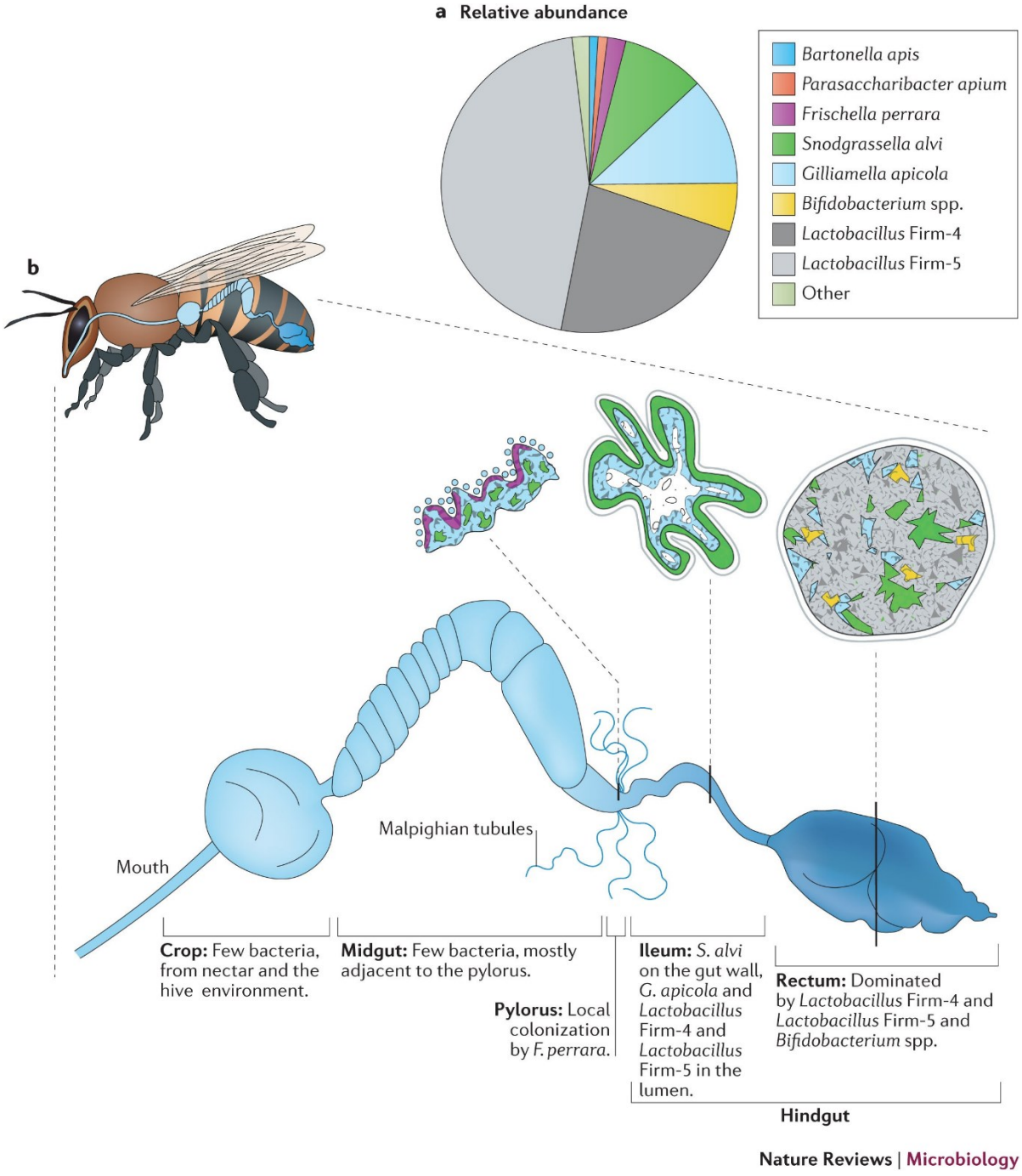


Figure 3- Overview of the bacterial composition *of honey bee intestines (63).*

Additionally, within the composition of the honey bee microbiome, there exist bacteria accounting for less than 5% of the overall bacterial community. These less-studied bacteria, while occupying a smaller slice of the bacteriome, present a unique and underexplored aspect of honey bee health (64–67).

However, not all bacteria in the honey bee microbiome contribute to the overall well-being of the hive. Two notably pathogenic bacteria are *Paenibacillus larvae* and *Melissococcus plutonius*. These agents are responsible for inflicting American Foulbrood and European Foulbrood, respectively, upon honey bee colonies (68, 69). These infections are severe, and once they take hold necessitate drastic measures like hive incineration to curtail further spread (70).

1.1.4. Other parasites

In addition to the parasites and pathogenic bacteria the honey bees face challenges from the fungus. Two distinct fungi, *Nosema apis* and *Nosema ceranae*, have been described. *Nosema ceranae* stands out as a more prevalent and concerning fungus for honey bees (71, 72). Its widespread presence has raised concerns among beekeepers and researchers alike, as it can significantly impact the vitality of honey bee colonies. This fungal pathogen can infiltrate the bee's gut, disrupting their digestive processes (73, 74) and compromising their overall well-being (74–76).

Beyond the fungi, honey bees are also faced with two noteworthy trypanosomatids: *Lotmaria passim* and the less prevalent *Crithidia mellificae* (77). *Lotmaria passim*, known commonly to infest honey bee colonies, can pose challenges to the health of these vital pollinators (78, 79). Its prevalence within honey bee populations underscores the complex interactions that shape the honey bee microbiome and their coexistence with various microorganisms. While less common, *Crithidia mellificae* is another trypanosomatid that has been identified within honey bee colonies. Its lower prevalence, compared to Lotmaria passim, suggests a less prominent role within honey bee ecosystems (80, 81).

1.2. Honey bee viruses

Apart from the interactions and dynamics of viruses, it is important to note that some of the honey bee viruses were identified in other pollinators. The threat from viral infections is increased by the massive beekeeping which leads to the spread and amplification of potentially harmful pathogens to wild pollinators (82–84).

The overview of the most studied and important viruses (in the scope of this work) is below. For fully referenced tables see supplement to my master’s thesis.

Table 1- Primary information *about the most important (in the scope of this work) bee viruses. LSV has been known for a relatively long time without known pathology and it is considered benign but further experiments preferably with isolated virus used on honey bee cell culture or controlled experiments where the honey bees are infected and observed, are needed to fully assign the virus as benign. Therefore the virus is noted as "benign?" till further experiments are possible and performed.*

Virus	Genome	Length [kb]	Taxonomy	Abbreviation	Pathogenicity
Acute bee paralysis virus	+ssRNA	9.5	*Dicistroviridae*	ABPV	Pathogenic
Israeli acute paralysis virus	+ssRNA	9.4	*Dicistroviridae*	IAPV	Pathogenic
Kashmir bee virus	+ssRNA	9.5	*Dicistroviridae*	KBV	Pathogenic
Apis mellifera filamentous virus	dsDNA	498.5	Unclassified	AmFV	Pathogenic
Black queen cell virus	+ssRNA	8.5	*Dicistroviridae*	BQCV	Pathogenic
Deformed wing virus (A/B/C/D)	+ssRNA	10	*Iflaviridae*	DWV	Pathogenic
Sacbrood virus	+ssRNA	8.7	*Iflaviridae*	SBV	Pathogenic
Lake Sinai viruses	+ssRNA	5.9	*Sinaiviridae*	LSV	Benign?
Chronic bee paralysis virus	+ssRNA (seg.)	5.9	Unclassified	CBPV	Pathogenic

Table 2- More information about the most important honey bee viruses. *P-positive, N- negative, O-overt infection, Tryp.-Trypanosomatidae, Under developmental stages is the presence and signs of infection. + means presence, - not found, ~ means inconclusive data (very little or inconsistent presence), ? is a sign of unknown information.*

Virus	Oral	Contact	*Varroa*	Vertical	*Varroa*	Nosema	Tryp.	Larvae	Pupae	Adult
Acute bee paralysis virus	+	-	+	+	P/O	?	?	+/~	+/+	+/+
Israeli acute paralysis virus	+	-	+	+	P/O	?	?	+/~	+/+	+/+
Kashmir bee virus	+	-	+	+	P/O	?	?	+/~	+/+	+/+
Apis mellifera filamentous virus	+	-	-	+	-	-	N	+/-	+/-	+/+
Black queen cell virus	+	-	-	+	-	P	?	+/-	+/+	+/-
Deformed wing virus (A/B/C/D)	+	-	+	+	P/O	N	?	+/-	+/+	+/+
Sacbrood virus	+	-	~	+	~/-	-	?	+/+	+/-	+/-
Lake Sinai viruses	+	?	~	+	-	?	?	?	?	+/?
Chronic bee paralysis virus	+	+	~	+	-	P	?	+/~	+/~	+/+

1.2.1. Deformed wing virus (DWV)

Deformed wing virus (DWV) is not a single entity; it comprises multiple variants, with four primary ones being described: DWV-A, DWV-B (also known as *Varroa destructor* virus 1), rare DWV-C described in the UK (85), and likely vanished DWV-D (also referred to as Egypt bee virus) (86). These variants, while sharing the DWV classification, may differ in their virulence, transmission dynamics, and interactions with both their bee hosts and the parasitic mites that often transmit them (87–94).

The virulence of these individual variants remains a topic of active research. While overt infections, characterized by symptoms like the deformed wings, are well-documented (95), the specific roles and impacts of each variant are less clear. For instance, the B variant, appears to be more dominant than DWV-A in several regions in recent years (96). This

dominance might be attributed to genetic recombination events between the two variants, leading to hybrid viruses with potentially altered virulence or transmission profiles (96–98). However, the exact nature and implications of these viruses and recombinant viruses remain unclear.

The relationship between DWV and the *Varroa* mite adds another layer of complexity to this.

The *Varroa* mite is a known vector for DWV transmission (45). However, the dynamic of DWV replication within these mites is not fully understood. Efforts to determine whether viruses replicate within *Varroa* have been undertaken through a variety of methods, yielding a range of results. Some of these approaches have offered indirect evidence suggestive of replication or passive transmission. However, the findings have been inconsistent, with a predominant focus on DWV. One factor contributing to this inconsistency is possibly the lack of differentiation between the common DWV genotypes, DWV-A and DWV-B, both of which might exhibit distinct behaviours within *Varroa* destructor mites, although the precise mechanisms underlying these differences remain unknown. For instance, fluorescence in situ hybridization experiments suggested that DWV-B may undergo replication within mites (99), whereas DWV-A has been considered non-propagative in these mites (99, 100). Additionally, some studies have indicated that viral replication within *Varroa* mites might be specific to certain individuals or populations of mites (101, 102). Recent research utilizing small RNA analysis has suggested that the transmission of both DWV genotypes within *Varroa* mites has been in a propagative manner (103). As for the other variants and their recombinants, the picture remains largely blank, underscoring the need for more focused research.

This virus has probably the strongest relationship to *Varroa*. Enough for the entire viral landscape to be affected by the introduction of *Varroa* mite. His spread changes the viral communities in favour of DWV (104, 105), but future studies of Australian bees (as *Varroa* spreads) will show how *Varroa* affects the viral communities in greater detail.

1.2.2. Acute bee paralysis virus (ABPV)

The name Acute bee paralysis virus (ABPV) is not merely descriptive; it encapsulates the primary clinical manifestation of the infection. Bees afflicted with ABPV often display sudden and severe paralysis, a symptom that can rapidly decimate bee populations within a hive (106).

Similar to DWV, ABPV transmission dynamics are linked to the *Varroa* mite (107). The mite, while feeding on bee, acts as a vector, introducing the virus directly into the bee's circulatory

system. This mode of transmission again not only facilitates the spread of the virus but also exacerbates the severity of the infection (108), leading to rapid and often lethal infections.

Adding to the complexity of understanding ABPV is its relationship with other closely related viruses, notably the Kashmir bee virus (KBV) and the Israeli acute bee paralysis virus (IAPV). These viruses share significant genetic and symptomatic similarities with ABPV, often leading to challenges in differential diagnosis (109). Interestingly, field observations and studies have indicated a phenomenon: typically, only one of these viruses predominates in a given location, generally ABPV in Europe, KBV in North America and New Zealand and IAPV in the Middle East and Australia (109). It is known that the presence of one bee virus might inhibit or outcompete the others, preventing their establishment or spread within a particular bee population (110).

1.2.3. Apis mellifera filamentous virus (AmFV)

The Apis mellifera filamentous virus (AmFV) stands unique within bee virology. As the only DNA virus described in bees till recent years, its size of over 450 kbp makes it an outlier not only in terms of its genetic material but also its length (111).

While its genetic makeup bears certain similarities to the Baculoviridae family, a group of viruses known to infect invertebrates, AmFV doesn't neatly fit into this or any other recognized viral category (111). This lack of clear classification underscores the virus's novelty and the gaps in our understanding of its origins, evolution, and behaviour.

The genome of AmFV reveals its potential modes of action and interactions with its bee hosts. Notably, the presence of *per os* infectivity factor proteins suggests a possible oral route of infection, reminiscent of some baculoviruses. Additionally, the virus appears to have acquired some host proteins present in other large DNA viruses, which might play roles in evading host defences or modulating host functions (111).

Clinically, AmFV presents another mystery. While the virus is relatively widespread among bee populations, overt symptoms seem to be the exception rather than the rule (111). The most characteristic manifestation of an AmFV infection is the appearance of bees with milky hemolymph, coupled with weakened, crawling bees near the hive entrance (112). Yet, these symptoms are rarely observed, suggesting that the virus might often establish covert infections or that other factors are required to trigger overt disease (113). However, there is no known cause, like the relation of *Varroa* mites and DWV which can turn the infection from covert to overt.

1.2.4. Other viruses

The relationship between honey bees and the parasitic mite, *Varroa*, is further complicated by the vast virosphere they share. To date, over 80 distinct viruses was identified in both of these organisms (GenBank search with keywords, 2023, see also (114)). This rich viral diversity, while offering insights into the health and interactions of its hosts, also presents challenges in understanding the specific dynamics and implications of each viral entity.

Determining the primary host for many of these viruses is not straightforward. The intimate association between honey bees and parasites, particularly *Varroa* mites, blurs the lines of viral host specificity. There seem to exist viruses that can replicate both in honey bees and their parasite. Some light can be shed on this issue with the use of small-RNA sequencing. This approach uncovers if the degradation process of small-RNA is random or active based on their profile, and additionally, the patterns are very different between honey bees and *Varroa* mites (103, 115, 116)

The limited research specifically focused on the *Varroa* virome further exacerbates these challenges. While honey bee virology has been a subject of extensive study, driven by the global importance of bees as pollinators and the observed declines in bee populations, the *Varroa* virome remains relatively underexplored (117, 118). This gap in knowledge underscores the need for more targeted investigations into the mite's virosphere, which could reveal novel insights into mite biology, behaviour, and its interactions with bees.

1.3. Metagenomics sequencing for viruses

In the field of microbiology, viruses represent one of the most diverse, abundant, and enigmatic entities. Despite their ubiquity and profound influence on global ecosystems, human health, and biogeochemical cycles, our understanding of the viral world has historically been limited by traditional culture-based methods (119).

Viral metagenomics refers to the direct analysis of viral genetic material from environmental samples without the need for prior cultivation. This approach is rooted in the broader field of metagenomics, which focuses on the analyses of a collective genome of microbial communities from a given environment. By specifically targeting viral communities, viral metagenomics offers a look into the vast and largely uncharted viral diversity present in various habitats (119).

1.3.1. Difficulties and advantages

One of the most salient advantages of viral metagenomics is its capacity for 'unbiased' detection. Traditional methods often hinge on prior knowledge of the target virus, but metagenomics allows us to discover previously unknown viruses, giving us a more holistic grasp of viral diversity. In contexts of health and disease, the ability of viral metagenomics to rapidly identify and characterize pathogens proves invaluable. But targeted sequencing is still a valuable method since it provides better coverage for the targeted viruses (120)

Yet, the viral metagenomics is not without its limits. Practical challenges start right at the sample preparation phase. Efficiently extracting viral DNA and/or RNA, while excluding host and bacterial genetic material, is a meticulous task that can introduce biases (121, 122). Then the amount of data it produces can be both an advantage and a challenge. Processing, analysing, and interpreting this vast amount of information, especially when faced with a significant proportion of unknown sequences, can be difficult. The field's heavy reliance on sophisticated computational tools necessitates a level of programming experience, which might be a learning curve for biologists. Additionally, the lack of standardization across various stages, from sample collection to sequencing and analysis, can make it difficult to compare results across different studies. And while metagenomics offers a snapshot of the viral community, it might not always capture the full dynamics of viral infections or provide clarity on the viability and infectivity of the detected viruses.

1.3.2. NetoVIR

The NetoVIR protocol represents a promising solution amidst the challenges of virome analysis, offering tailored enrichment techniques for viral metagenomics studies. Unlike other microbiome components, the virome's samples containing the majority of non-viral genetic material (123) and its lack of universally conserved genomic regions (124) necessitates precise methodologies. Key steps involve enrichment of capsid-protected viruses and random amplification, aiming to minimize bias and preserve viral diversity, though achieving this balance is challenging due to viral susceptibility during enrichment (for example filtering out bacteria can result in loss of large viruses, different groups of viruses are sensitive to different techniques). Studies assessing virome preparation protocols underscore the variability in viral recovery rates and the significant impact of different methods on outcomes (125–129).

One of NetoVIR's most salient strengths lies in its ability to provide a comprehensive view of the virome. By optimizing various sample preparation steps on mock metagenome, the protocol not only ensures the recovery of a wide range of viruses but also significantly

reduces bacterial contamination, allowing for a clearer and more focused analysis of viral communities. This is particularly crucial in the context of some samples, where bacterial populations are abundant and can easily overshadow the viral component in metagenomic studies. Furthermore, NetoVIR addresses a critical oversight in many existing protocols: the potential exclusion of large viruses (121, 122). One limitation is the design of the protocol for fecal samples, and the need to optimize it for other input materials.

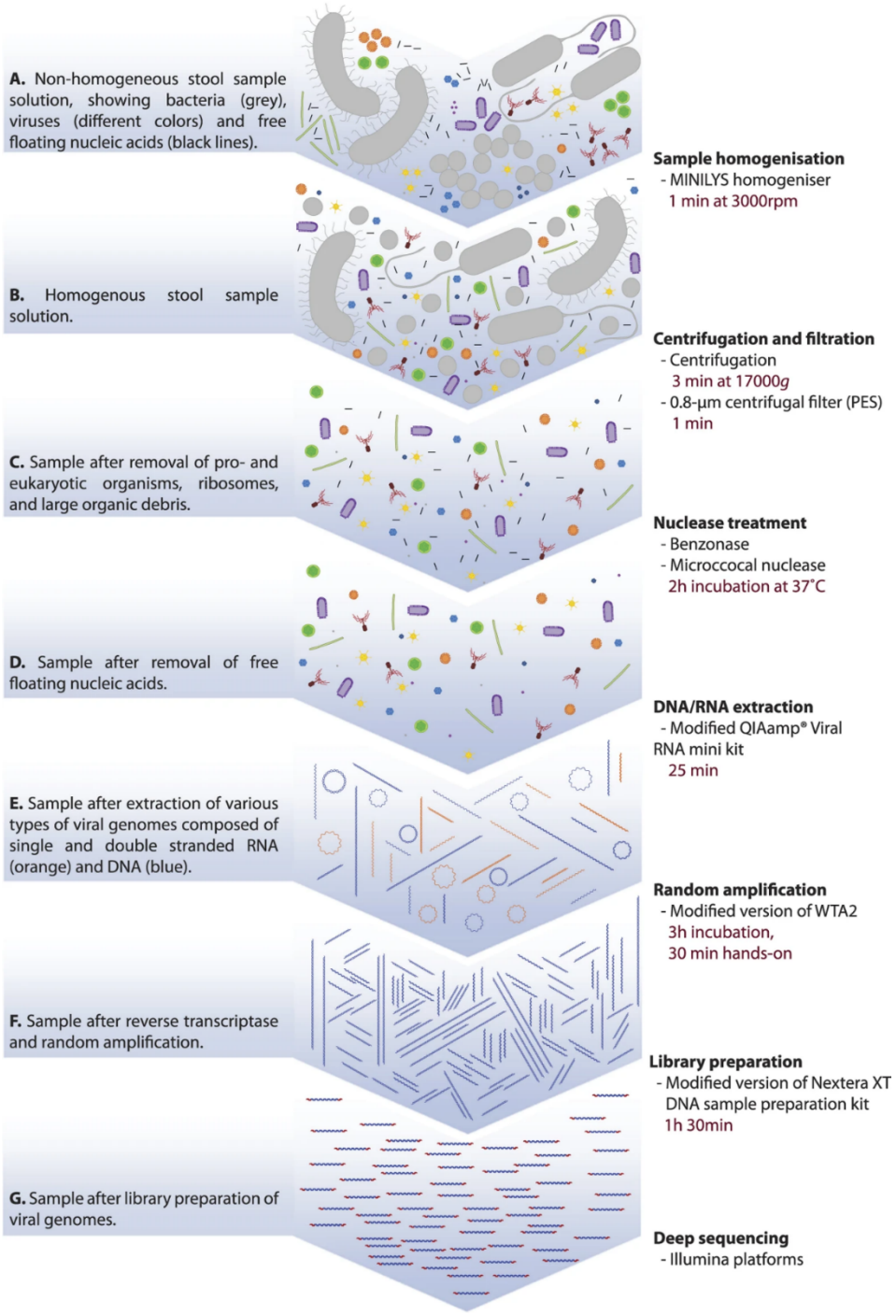


Figure 4- Overview of sample preparation with NetoVIR protocol. *Taken from (121)*

1.3.3. Bioinformatics

The pursuit of insights from such vast datasets is not without its share of difficulties. The very foundation of exploration lies in sequencing technology, a powerful yet imperfect tool. Since Illumina is still the main solution for many viral metagenomics studies, and was our solution to sequencing, the next part will be focused on work with Illumina-generated sequencing data. However, there are new promising approaches like a combination of Illumina and long-read technology, or only long-read sequencing (130).

Even with the advancing capabilities of platforms there exist regions of genomes that are inadequately sequenced or difficult to resolve which adds layer of challenge to the data from the very beginning.

Standardisation

A prominent challenge within the domain of viral metagenomics and metagenomics overall resides in the absence of standardized techniques (131–133). The field of viral metagenomics is marked by rapid evolution, with dedicated software tools emerging in recent years to address the unique demands of viral analysis. Notably, software like CheckV has been specifically designed for viral metagenomics (134), akin to how CheckM fulfils similar functions for bacteria and eukaryotes (135).

However, this accelerated pace of development and the absence of universally accepted methodologies poses a distinctive disadvantage. The diversity of software tools and methods employed within the field results in a veritable mosaic of approaches, with each research laboratory crafting its preferred methodologies for wet and dry labs. This heterogeneity is a more pronounced problem, especially in viral metagenomics.

Metagenomics assembly

One of the primary challenges in metagenomics assembly lies in the presence of numerous genomes within a single sample. These genomes exhibit varying levels of coverage, often falling outside of the optimal coverage obtained in isolated genome sequencing. Furthermore, the complexities extend beyond the diversity of genomes within a sample; they also encompass the disparities between different samples. Some samples may be composed of merely a handful of genomes, such as those stemming from acute infections with exceptionally high viral loads. In contrast, others may contain several hundred genomes (136, 137). Ideally, the aim is to achieve reliable and precise recovery of all genomes from both extreme types of samples. However, striking the right balance is delicate, as excessive

coverage can exacerbate errors, while insufficient coverage may result in fragmented genomes with large regions missing. The variations in coverage even manifest within a single genome sequencing, rendering metagenomics assembly an even more intricate task. Additionally, the presence of conservative regions or closely related strains, alongside repetitive segments, further compounds the challenges inherent in Illumina assembly (137).

One of the widely favoured and dependable choices for metagenomics assembly is SPAdes, specifically when executed with the "--meta" flag (138, 139). However, a more recent development in this assembler has introduced the "--metaviral" flag, tailored for viral metagenomics (140), although it is notable that lighter filtration for viral entities during this stage can yield more favourable outcomes based on the aim. Regardless, the "--metaviral" flag remains a valuable asset in subsequent assembly steps. Notably, SPAdes mandate meticulous quality control and trimming, as it exhibits sensitivity to technical sequences, as stated in SPAdes manual.

Another challenge arises from the reality that each sample may necessitate a somewhat distinct assembly approach (138). In scenarios with a limited number of rare samples, it proves beneficial to experiment with a number of assembly tools and approaches to extract the maximum value from the sequencing data. However, this expensive approach becomes unmanageable during large-scale sequencing projects. In such situations, a noteworthy strategy to maximize data utility involves a process known as co-assembly. This entails pooling several individual samples together and executing assembly on this combined dataset. Combining of results stemming from both individual assemblies and co-assemblies has been shown to yield the most robust outcomes (141).

Notably, the large genomes often need additional sequencing for resolving some regions/tandem repeats and others. As an extreme example of a virus can be named Pandoravirus which was assembled with a combination of Illumina, 454-Roche, and PacBio sequencing and then the terminal tandem repeats were resolved with another polymerase chain reaction (PCR) (142).

Mining viral sequences

Predicting viral sequences is a complex task primarily due to the absence of universally shared genes that typically exhibit high identity within genomes (124). The extensive diversity among viruses makes their analysis difficult. Another difficulty in the prediction

process is the reliance of software on known viral sequences, which can exclude entirely novel viruses that bear no resemblance to any previously observed entities.

Two main approaches are commonly employed for predicting viral entities. The first approach is based solely on sequence data, e.g. Virsorter/Virsorter2 (143, 144), while the second is an alignment-dependent process that relies on informative markers to deduce sequence characteristics e.g. VirFinder (145). Numerous software tools have been developed, each predominantly following one of these approaches, e.g. (146–149). However, a recent advance in this field comes in the form of software geNomad, which incorporates both strategies for predicting viruses and mobile elements. This comprehensive approach has demonstrated efficacy in viral sequence prediction (150). There is possibly a benefit in using more than one prediction tool to gain more robust results, one of the first was VIRify which uses several prediction tools, each performing exceptionally well on different groups of viruses (151). The whole integrated workflow of VIRify is shown in Fig. 8.

Quality and fragmented genomes

The challenge of assembly, previously discussed, is compounded by the fact that not all sequences maintain consistent representation across their entire length resulting in a single genome being fragmented into multiple contigs. This phenomenon is not uncommon, even in isolated genome sequencing, and becomes more pronounced in metagenomics sequencing (136). Because of that is essential process of the quality control (QC) of the obtained sequences, which helps distinguish between well-represented genomes and less significant fragments.

For QC of viral sequences, the above mentioned tool, CheckV, plays a crucial role (134). It assists in filtering out incomplete sequences. CheckV can predict the percentage of completeness, allowing for the selection of suitable cut-offs for subsequent analysis. But given the fact that one large genome can be fragmented into several pieces, methods have been devised to reconnect these contigs that originate from the same genome (152–154). This process is known as binning, wherein sequences are grouped based on their shared origin from a single genome. An example of the simplified inner workings of vRhyme is shown in Fig. 5. Notably it should also group segmented genomes of one virus.

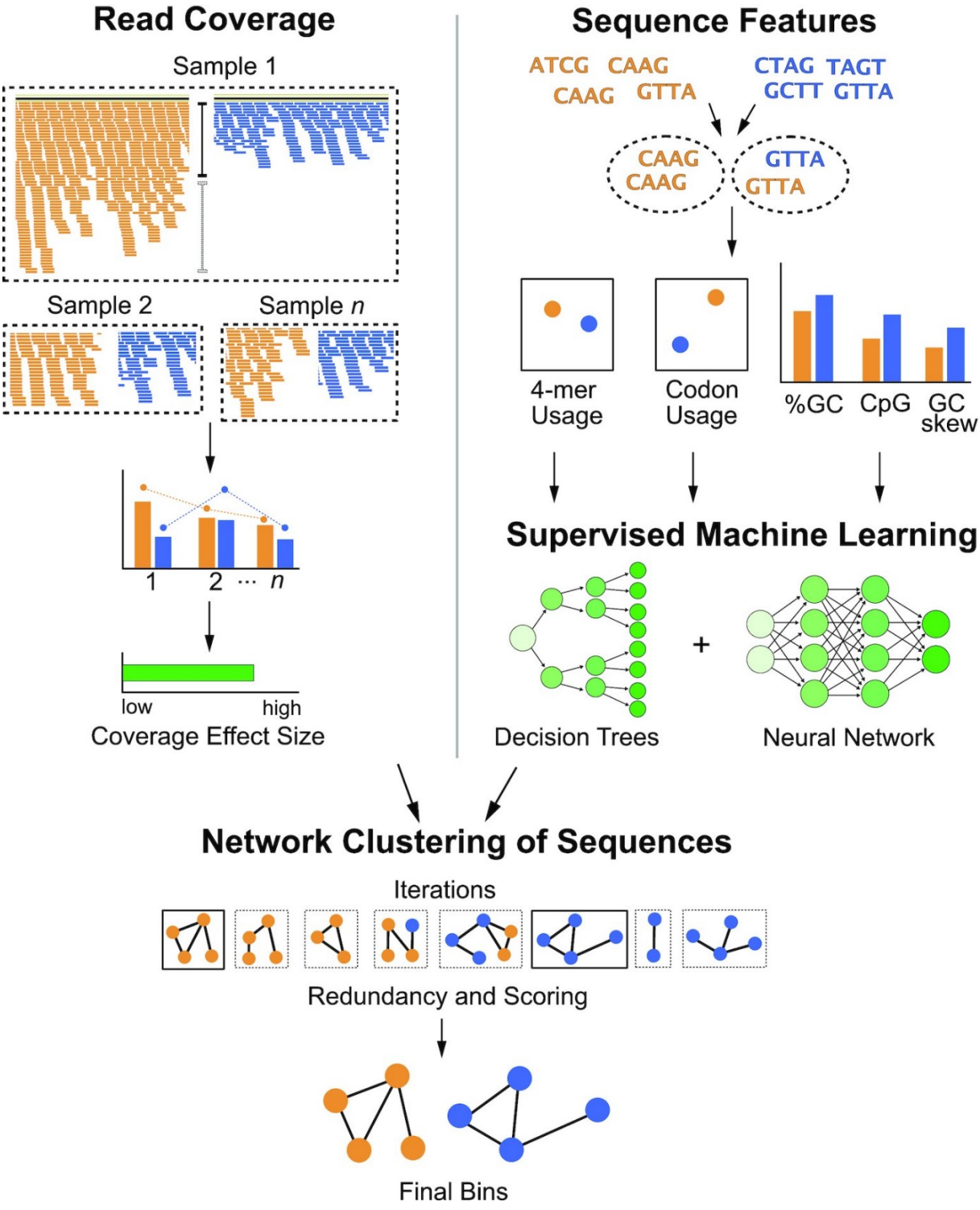


Figure 5- Overview of the principle of binning viral sequences as implemented by vRhyme. *Taken from (152).*

To facilitate this task, specialized software has been developed explicitly for binning viral contigs, outperforming traditional binning software primarily used for bacteria and eukaryotes (152, 153). However, there are still gaps in the available software, such as the absence of a bin-refining tool like MetaWRAP (154) that can leverage CheckV data or a QC tool capable of working with bins. Some limitations can be mitigated by connecting contigs using a

number of N's or by utilizing modified outputs from one software for analysis by another, albeit with adjustments to the initial settings. Although this approach is not without its imperfections, it enables the retention of maximum information while eliminating very incomplete data, thus optimizing the overall quality of the analysis.

Contamination

In the case of viral metagenomics where we are trying to enrich the sample as much as possible with various steps there are two problems, that are not unique to only viral metagenomics:

1) Contamination during preparation
2) Contamination of non-host sequences

The first problem is not only part of every low-abundant sample preparation but also the analysis afterwards (155). In some cases, even mildly contaminated samples can be filtered out against negative control that should be always included in the experiments. Even in the commercial kits for sample preparation contaminating sequences can be found. Because of that fact term "kitome" exists which is composed of all the sequences present in kits used during sample preparation. The less rich and abundant the viral population in the starting material the more contamination interferes. And another problem is cross-contamination between samples, called "splashome" (156, 157). During a sample preparation high standard of clean wet lab work should be employed, dividing work into several areas. A negative control should always be included to identify at minimum kitome and splashome.

The second problem is the presence of viruses that are ambiguous. In the case of bees, it concerns plant viruses (158), and in faeces, viruses originating from food can be present (159). But even more puzzling ones like sequences of phages in the human blood of healthy individuals, some of the sequences can be tracked to the intestine (e.g. to Escherichia coli), but their presence is still enigmatic (160). It's often difficult to determine what is still an important or significant sequence and who is the probable host of the virus.

Classification of viral sequences

The viruses are a very diverse group. Their genome can range from one polyprotein to a segmented genome with alternative proteins originating from one segment. One of the most used genes for the identification of viruses is their polymerase and other viral group-specific proteins like reverse transcriptase. But of course, the viral genome is composed of more than their polymerase and it's important to take the whole sequence into account when classifying

gained sequences. We are slowly moving away from a single 'best hit' classification to taking into account the whole sequence. Every software deals with this differently, one of the newest ones is above mentioned geNomad which evaluates the assignment of all the predicted genes in the sequence.

The working of the classification is quite simple. First, every gene gets its own classification:

Gene	Marker	Bitscore	Marker taxonomy
Gene 1	GENOMAD.007145.VV	175	*Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Autographiviridae*
Gene 2	GENOMAD.018431.VV	126	*Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Zobellviridae*
Gene 3	–	–	–
Gene 4	GENOMAD.006041.VV	268	*Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Zobellviridae*
Gene 5	GENOMAD.171320.VC	70	*Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Zobellviridae*
Gene 6	GENOMAD.153406.VP	116	*Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes*
Gene 7	GENOMAD.001922.VV	404	*Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes*

Figure 6- Initial assignment of taxonomy *to each gene identified.*

Weights are then determined for each taxon in the gene assignments by taking into account the bitscores obtained from the alignments.

Rank	Taxon	Gene 1	Gene 2	Gene 4	Gene 5	Gene 6	Gene 7	Taxon weight	Support (taxon weight / total weight)
Realm	*Duplodnaviria*	175	126	268	70	116	404	1159	**1159/1159 = 1.00**
Kingdom	*Heunggongvirae*	175	126	268	70	116	404	1159	**1159/1159 = 1.00**
Phylum	*Uroviricota*	175	126	268	70	116	404	1159	**1159/1159 = 1.00**
Class	*Caudoviricetes*	175	126	268	70	116	404	1159	**1159/1159 = 1.00**
Family	*Autographiviridae*	175	–	–	–	–	–	175	175/1159 = 0.15
	Zobellviridae	–	126	268	70	–	–	464	464/1159 = 0.40

Figure 7- Scoring *based on the previous assignment of taxonomy to each gene.*

In this case the final classification would be *Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes* (150). VIRify takes a similar approach, taking annotations for each gene in contig from several sources and then scores going through a voting system (151). The whole integrated workflow of VIRify including classification is in Fig.6.

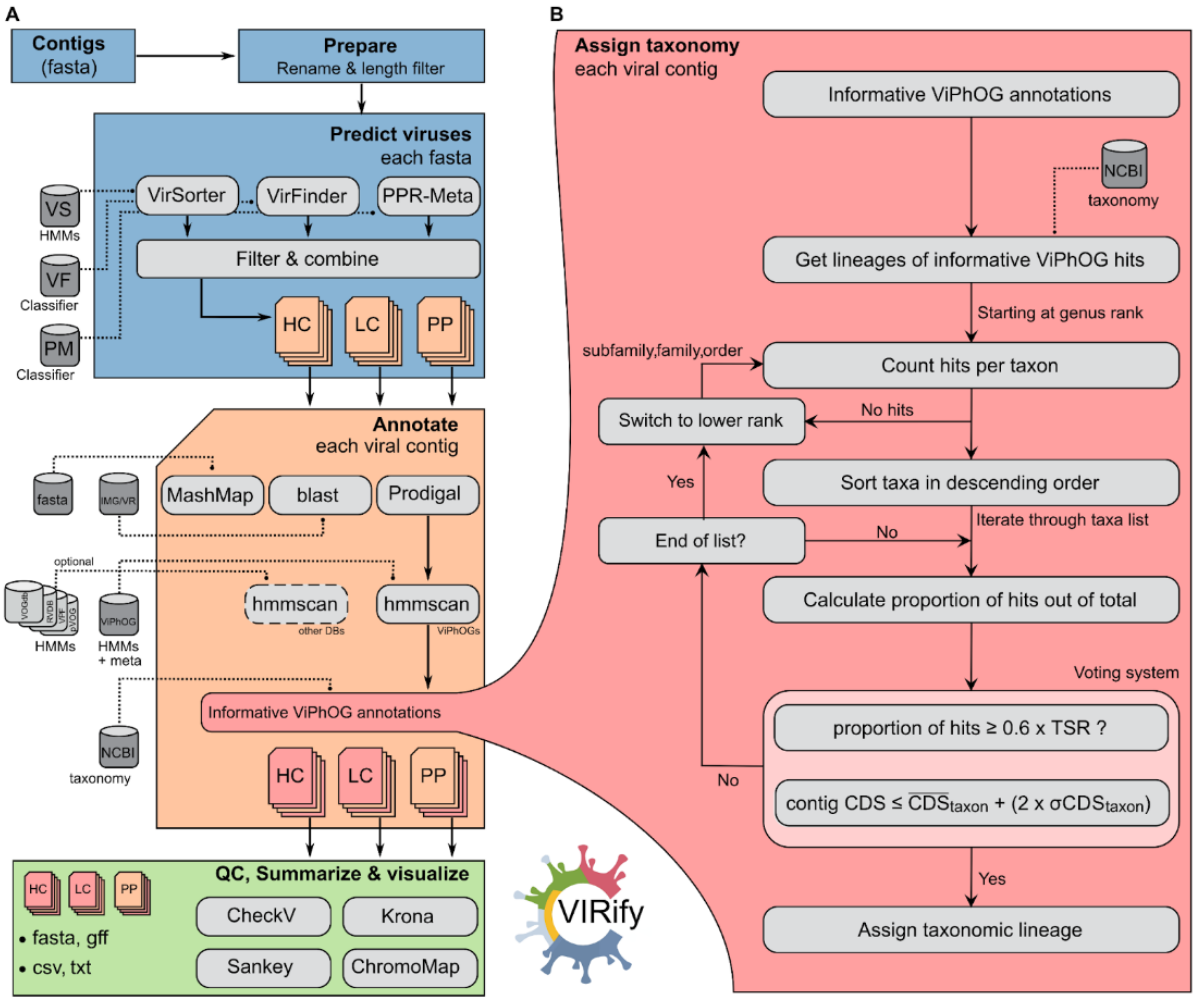


Figure 8- Approach of VIRify *(integrated tool for detection and classification of viral contigs from Metagenomics data)* ***to taxonomical classification****. Taken from (151).*

The resulting classification is often less specific or classifies fewer sequences but is more reliable in the assignment.

2. Aims

This study aims to comprehensively explore the honey bee virome, focusing on its diversity, dynamics, and potential implications for bee health. By tracking changes in the virome and assessing its interactions with other factors, such as bacterial components and environmental pollutants, we seek to gain a deeper understanding of honey bee well-being.

1. ***Honey Bee Virome:*** Our primary objective is to comprehensively characterize the viral communities inhabiting honey bee populations, focusing on their diversity, prevalence, and potential implications for bee health.
2. ***Virome Dynamics:*** We aim to track changes and fluctuations in the honey bee virome over three years to uncover patterns and trends that may provide insights into the resilience and adaptability of these essential pollinators.
3. ***Factors Impacting the Virome:*** Recognizing the multifaceted nature of honey bee health, we will examine factors such as bacterial and eukaryotic components within the honey bee microbiome. Additionally, we will assess the impact of environmental factors, such as pesticide concentrations, on the honey bee virome. Our goal is to identify potential correlations and dependencies between these elements and honey bee well-being.

3. *Methods:*

The main part of this thesis consists of two projects. In the first project, we explore the stability of the bee virome. We analyzed nine bees from each hive in biological replicates (Project 1). Based on findings from the first study and modification of the protocol we designed our second study with the aim of exploring the viromes of the same hives at different time points (Project 2).

The differences in sample preparation between the two studies are visualized in Fig 9.

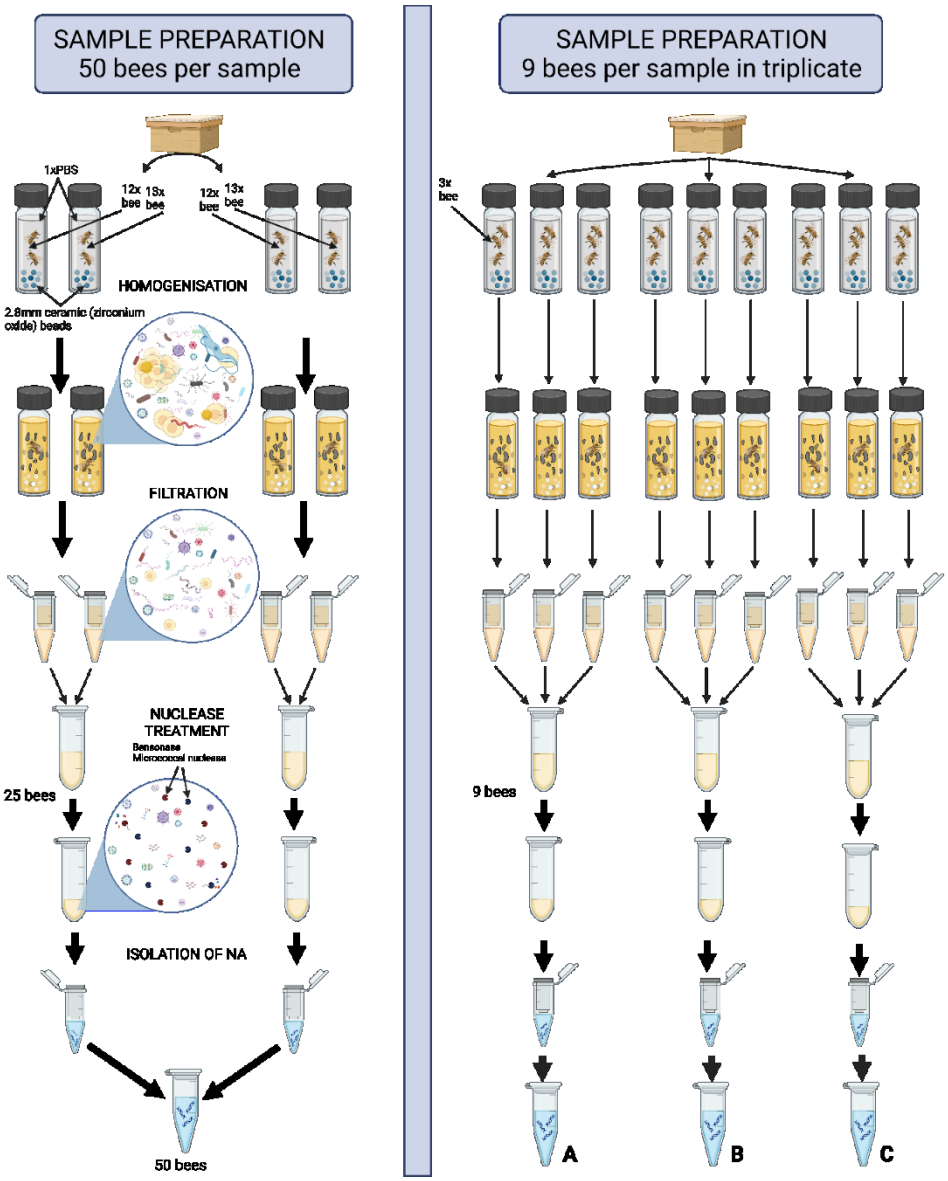


Figure 9- Overview of the sample preparation and the difference between the two studies. *Created in BioRender.*

3.1. Sample collection and preparation for sequencing

3.1.1. Project 1 - virome stability

We prepared 39 bee samples, three samples from one hive and altogether 13 hives from 9 locations (Brná, Prášily, Strakonice, Petrušov, Slezko, Brumov-Bylnice, Poysdorf, Ivančice, Lhotka u Telče) were analyzed. The bees were sampled from August to September 2018 and were immediately placed on dry ice. They were transported on dry ice and kept at -80°C.

Nine bees were randomly selected from each hive. Bees were divided into three 2ml tubes with 1.4mm ceramic (zirconium oxide) beads and homogenized in a MINILYS Personal Homogenizer (Bertin technologies, Montigny-le-bretonneux, Ile-de-France, France) in 1ml of 1xPBS, for 180s at the lowest speed (3 000 rpm). From each tube 400µl of homogenized bees was filtered through Vivaclear Centrifuge Filters with 0.8µm pores (Sartorius, Göttingen, Germany). For the consequent procedure, the filtered product of homogenate of 3x9 bees was pooled and 260µl of this filtered homogenate was incubated at 37°C with 14µl of 20x nuclease buffer (1M Tris, 100 mM CaCl2 and 30 mM MgCL2, pH=8), 4µl benzonase and 2µl of micrococcal nuclease (New England Biolabs, Ipswich, Massachusetts, United States). After two hours of incubation, the nucleic acid extraction was performed immediately. We used the QIAamp Viral RNA Mini Kit (QIAGEN, Hilden, Germany) without adding carrier RNA which allows the extraction of both RNA and DNA. Since we used a larger sample of bees, twice as much of the sample was used for the consequent nuclease treatment, and twice as much AVL buffer and ethanol was added.

After elution we continued according to the NetoVIR protocol; reverse transcription (RT) and first amplification with WTA2 (Sigma-Aldrich, St. Louis, Missouri, United States); library preparation with Nextera XT (Illumina, San Diego, California, United States), and QC/measurements steps in between steps.

The prepared libraries were shipped on dry ice to the KU Leuven Nucleomics Core (VIB) in Belgium for sequencing on the HiSeq2000 platform (Illumina, CA, USA) for 2x150-bp paired-sequencing.

3.1.2. Project 2- longitudinal virome changes

We continue with sampling 5 locations (Lisnice, Libechov, Brdy/Nerezin, Prasily and from Crop Research Institute abbreviated as VURV) for three years, with three hives per location. However, since not all the hives survived all sampling, we were forced to switch hives for some localities. E.g., hives from Brdy/Nerezin survived the first two years before collapsing

so we took other hives from the same location. Bees were removed from the hives and immediately placed on dry ice and stored at -80°C untill further processing.

The samples of bees collected in the year 2019 were divided into three parts. One part was used for the analysis of pesticides present in the localities, then for 16S/18S amplicon sequencing and lastly for virome analysis. I will focus solely on virome analysis in my methods section but the article with the methodology for of microbiome and pesticide analysis can be found in the manuscript in preparation. For the following years 2020 and 2021, the pesticide analysis wasn't done since only trace amounts of a very limited set of pesticides were found in our samples. The 16S/18S rRNA data for the years 2020 and 2021 sequenced by our collaborators did not pass QC and were not further included in the analyses.

We have randomly selected 50 bees from each hive. Bees were divided into four 7ml tubes with 1.4mm ceramic (zirconium oxide) beads (twice 12 bees and twice 13 bees) and homogenized in MINILYS Personal Homogenizer (Bertin technologies, Montigny-le-bretonneux, Ile-de-France, France) in 3ml of 1xPBS, for the duration of 180s on the lowest speed (3 000 rpm). From each tube 400µl of homogenized bees was filtered through Vivaclear Centrifuge Filters with 0.8µm pores (Sartorius, Göttingen, Germany). For the consequent procedure, the filtered product of homogenate of 12 bees and 13 bees were combined and 260µl of this filtered homogenate was incubated at 37°C with 14µl of 20x nuclease buffer (1M Tris, 100 mM CaCl2 and 30 mM MgCL2, pH=8), 4µl benzonase and 2µl of micrococcal nuclease (New England Biolabs, Ipswich, Massachusetts, United States). After two hours of incubation, the nucleic acid extraction was performed immediately. We used the QIAamp Viral RNA Mini Kit (QIAGEN, Hilden, Germany) without adding the carrier RNA which allows for the extraction of both RNA and DNA. Since we used a large sample of bees, twice the amount of the sample was treated with nucleases and twice the amount AVL buffer and ethanol was added. After elution, the two combined pools of bees were pooled together resulting in one sample of 50 bees.

WTA2 and library preparation was done according to the NetoVIR protocol.

Sequencing was done at the Institute of Molecular Genetics of the Czech Academy of Sciences, Czechia with different kits (High-Output/Mid Output) on NextSeq 500. We aimed for at least 10M reads per sample.

3.2. Bioinformatics analysis

With the improvements and availability of new bioinformatics tools, we improved our analysis and besides basic analysis, we also performed the steps in the advanced approach chapter but only on a later longitudinal study. We tested all steps of the analyses and selected the best available software based on the results of the bioinformatic analyses of samples from the year 2019 of Project 2. On data from Project 1 (stability study), 'basic' analysis was performed while data obtained in the consequent study were processed also with the 'advanced approach'. An overview of both approaches is shown in Fig 10.

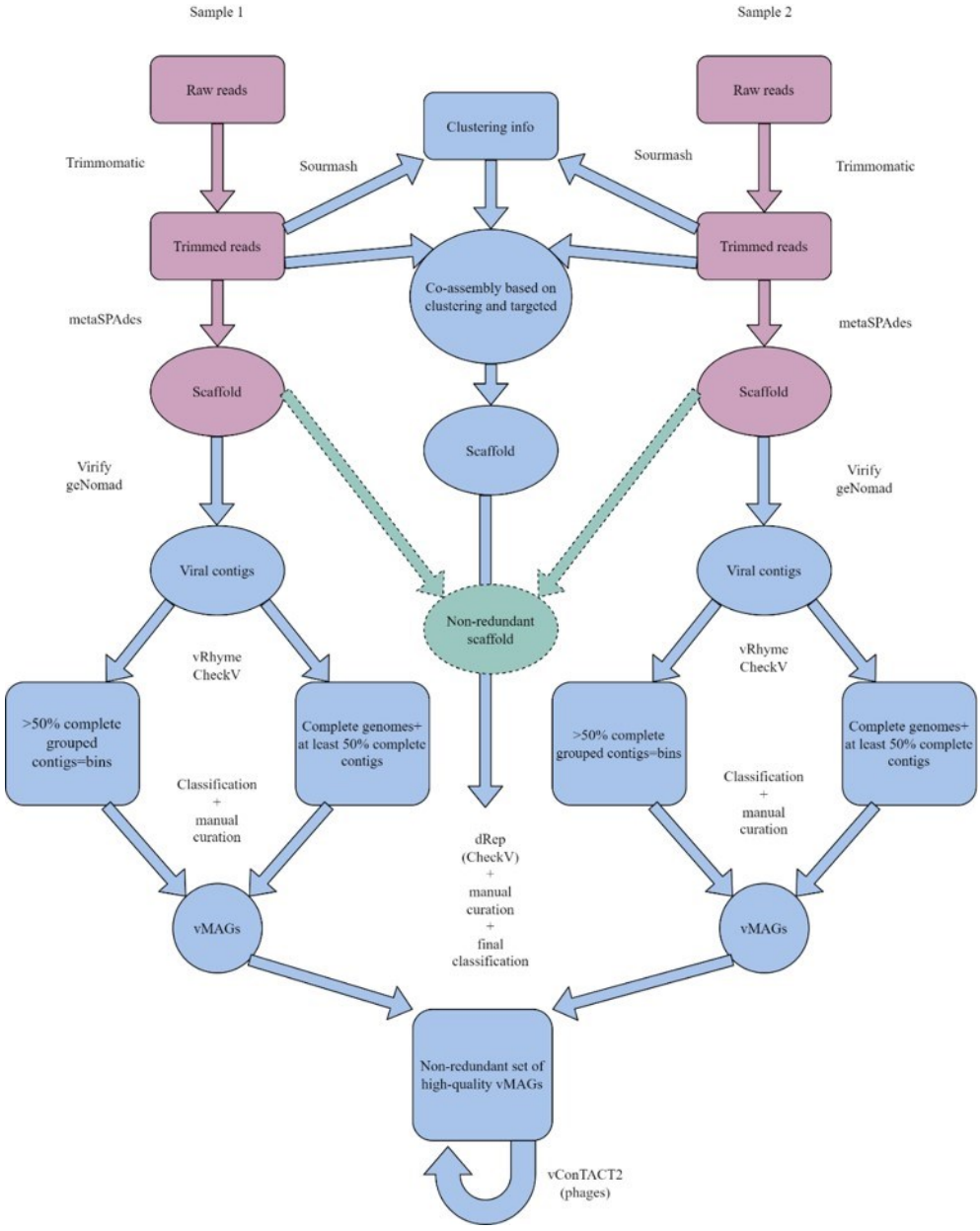


Figure 10- The sequencing processing pipeline for both projects. *Common steps are in pink, unique to Project 1 in green and unique to Project 2 in blue.*

3.2.1. Basics processing of sequencing data

In our study, the initial and crucial step was to ensure the quality of the sequencing reads, as this forms the foundation for all subsequent analyses. To achieve this, we employed FastQC v0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), a widely used tool that provides a comprehensive overview of raw sequence data. This allowed us to visually inspect various QC metrics and identify potential issues.

Following the initial quality assessment, we proceeded to trim and clean the reads using Trimmomatic v0.39.10 (161). This step was essential to remove any adapters or primers, specifically those associated with WTA2 and NEXTERA XT. The trimming process was set by several parameters: we utilized the ILLUMINACLIP option for adapter removal, HEADCROP was set to 19 to remove the first 19 bases from the start of the reads, and we set both LEADING and TRAILING to 15 to trim bases off the start or end of a read if below a threshold quality. Additionally, the SLIDINGWINDOW option was set to 4:20, which scans the read with a 4-base wide sliding window and trims when the average quality per base drops below 20. Lastly, we ensured that only reads with a minimum length of 50 bases post-trimming were retained using the MINLEN:50 setting.

To confirm the efficacy of our trimming process, we used FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), for a second round of QC. This step was pivotal to ensure that all low-quality bases and reads were effectively filtered out, leaving us with a high-quality dataset for assembly.

With cleaned and trimmed reads, we performed the assembly with SPAdes v3.15.3 (139). Given the metagenomic nature of our samples, we operated SPAdes with metagenomics flag. To optimize the assembly, we used multiple k-mer lengths, specifically 21, 35, 55, and 77.

Post-assembly, the generated contigs were subjected to alignment against the NR database from NCBI. We ensured that the database was regularly updated throughout the study, with the latest version being from 2023. For the alignment, we utilized diamond v2.0.11 (162) with the blastx option.

To further refine our analysis, we mapped the reads back to the assembled contigs using BWA-MEM2 v2.2.1 (163). This allowed us to gauge the coverage and depth of each contig. The specific count of reads mapping to each contig was extracted using CoverM v0.6.1 (https://github.com/wwood/CoverM). For a more intuitive and visual representation of our

data, we employed KronaTools v2.8.1 (164), which provided a multi-layered pie chart displaying the relative abundance of our reads.

3.2.2. Advanced approach

In our study spanning three years of sampling, it was essential to understand the similarities and differences in the sequencing reads across the samples from each year. To achieve this comparative analysis, we utilized Sourmash v3.3.0 (165), a tool adapt at rapidly comparing large datasets. We used three k-mer sizes: 21, 33, and 55 which allowed us to see the dataset in different resolutions, allowing us to select the resolution which gave us the best clusters.

Upon visual inspection of the comparison plot generated by Sourmash, we observed distinct groupings within the data (Fig 11). Specifically, four prominent clusters emerged, suggesting patterns or similarities among certain samples across the years. This clustering prompted us to further investigate the underlying genomic compositions of these grouped samples.

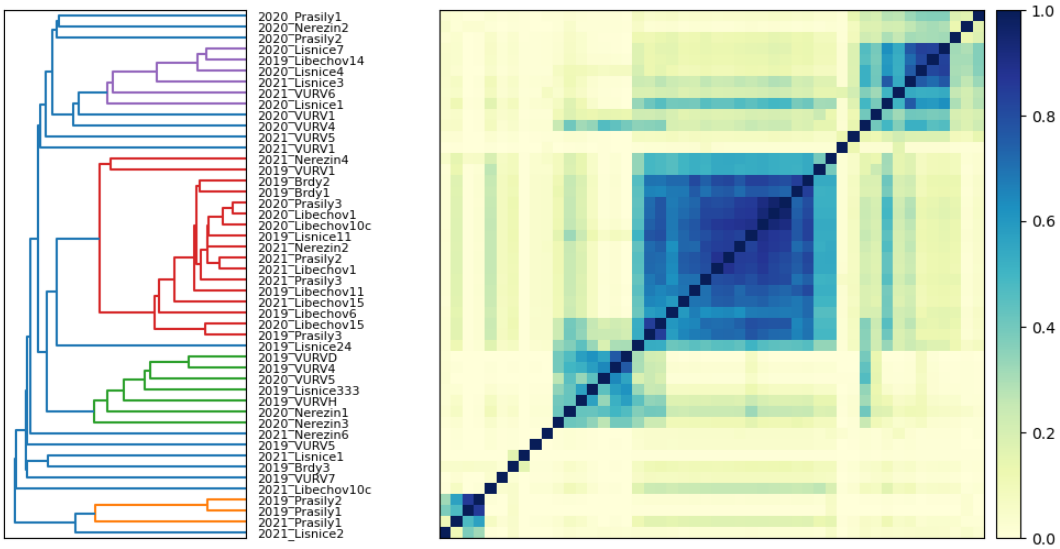


Figure 11- Sourmash comparison *for k-mer size 51, it shows distinct grouping in the data.*

To delve deeper into the composition of these clusters, we performed a co-assembly of the reads from samples within each cluster. By pooling the reads from the samples in each cluster, we aimed to generate a more comprehensive and representative genomic picture. For this co-assembly, we employed SPAdes, using the same parameters as previously described to ensure consistency in our assembly approach.

Beyond these primary co-assemblies, we also did on some targeted co-assemblies. These were designed to address specific questions or hypotheses that arose during our analysis.

We also did a relatively novel approach: viral binning and the creation of viral metagenome-assembled genomes (vMAGs).

Initially, we evaluated a range of binning software specifically designed for viruses, including Vamb v4.1.3 (166), CoCoNet v1.1.0 (153), and vRhyme v.1.1.0. (152) To ensure a comprehensive assessment, we juxtaposed the results from these tools with outputs from MetaWRAP (154), a well-established wrapper and bin refining tool. Historically, MetaWRAP, which employs metaBAT2 (167), CONCOCT (168), and MaxBin2 (169), has demonstrated efficacy with bacterial and eukaryotic datasets. However, its reliance on CheckM (135) for bin refinement renders it less advantageous for direct viral applications.

To further refine our approach, we assessed several viral sequence prediction tools, including geNomad (150), VirSorter (144), PPR-Meta (170), and VirFinder (145). Recognizing that each software exhibited strengths in predicting different types of viral sequences, we adopted a maximalist strategy. We employed Virify v0.4.0 (151), which integrates the capabilities of VirSorter, PPR-Meta, and VirFinder, and complemented this with geNomad v1.2.0. Our datasets indicated superior performance by geNomad, although other tools, excelled in specific viral categories, such as PPR-Meta for bacteriophages.

Post-prediction, we mapped the trimmed reads back to the contigs identified as viral. The binning process was primarily executed using vRhyme v1.1.0 (152). To ensure the quality and completeness of our scaffolds, we utilized CheckV v1.0.1 (134), which also aids in predicting complete contigs/genomes. The subsequent output underwent refinement with dRep v3.4.0 (171), ensuring the removal of suboptimal bins. We then subjected each bin to a rigorous validation process, employing blastx against the IMG/VR v4 (172) database and geNomad classification. For a temporal idea, we classified the bins and 'complete' genomes annually against the IMG/VR v4 and the current NR protein database from NCBI. Additionally, phage classifications were enhanced using vConTACT2 v0.11.3 (173) with the 'ProkaryoticViralRefSeq211-Merged' database.

Co-assemblies underwent a similar processing pipeline.

Upon sequencing completion over three years, we undertook a meticulous deduplication process for the cleaned, at least 50% complete, and classified vMAGs using CheckV and dRep. This yielded a non-redundant set of vMAGs, which underwent another round of classification validation. Specific sequences were further blasted to refine or expand

classifications, such as distinguishing between DWV-A and B variants. Finally, reads from all samples were mapped back to our curated set of vMAGs, and the coverage was extracted.

3.2.3. Data processing

By following the creator's manuals, data was processed mostly in R, using several packages like Phyloseq, MicrobiomeStats, SIAMCAT, MixOmics and other data manipulation packages. For Python others were used: Matplotlib, Pandas, Seaborn, NumPy and DNA feature viewer.

3.2.4. Data availability

The raw data of the current study were deposited in SRA under BioProject id PRJNA1008242, densoviruses OR553295-OR553303, AmFLV under OR553294 and AmNV under OR596894. Some more additional files are on GitHub (https://github.com/kadlck/NAZV19). The two other years of sequencing will be uploaded to SRA with upcoming articles and sequences of newly complete RNA viruses into GenBank as well.

4. Results

The findings presented in this section stem from two major sequencing projects, each described in the Methods section. These projects were conducted to explore the virome composition of honey bee populations and provided important insights into both RNA and DNA viruses that infect honey bees. We adapted and optimised the NetoVIR protocol for honey bee samples throughout the projects. Both projects focused on relatively healthy honey bees from Czechia, that showed no overt signs of diseases or parasitism.

The first project involved 39 samples from Czechia, consisting of three samples from nine bees per hive. This study focused on the stability of the virome within these triplicates, and the results were published in 2022. The second project expanded the scope by analysing 48 hives over three years. Initial sequencing of the first year's samples led to the discovery of sequences potentially belonging to large DNA viruses, which was unprecedented as only one DNA virus was previously known to infect honey bees. These findings and the completion of the vMAGs were published in 2024. The comprehensive longitudinal analysis of the three-year virome study is presented in this section.

In addition to these projects, we integrated several -omics datasets, that are currently in the process of manuscript preparation. Therefore, these data will be briefly discussed. We have also analysed the virome of hives used in an experimental study by our collaborators, the results of which will be prepared for publication in 2024. Beyond NGS, we performed qPCR on bees and *Varroa* mites from various experimental setups, one of which is described in detail here.

4.1. Project 1- virome stability

The Virome of Healthy Honey Bee Colonies: Ubiquitous Occurrence of Known and New Viruses in Bee Populations

In this study, we conducted an analysis of the honey bee virome in Czechia from different locations. An important finding was the fact that honey bee viruses didn’t cluster by location or hive they came from suggesting that each bee differs in the composition of honey bee-infecting viruses. The study emphasizes the stability and geographical dependence of the overall virome composition, which is primarly driven by bacteriophages and plant viruses, while honey bee-infecting viruses are much more variable among the hives and locations.

One noteworthy finding in the Czech samples is the prevalence of DWV-B, which is consistent with its recent global spread and dominance even in asymptomatic hives. This

underscores the persistence of DWV-B in colonies with low *Varroa* levels or those treated with miticides. Additionally, the study identifies the relationship of two related rhabdoviruses, BRV-1 and -2, with a unique distribution pattern suggesting a potential interdependence between these viruses.

LSVs are also discussed, with the detection of five distinct variants in Czech bees. This diversity highlights the complexity of LSV infections and suggests a long coexistence of LSV with honey bees.

We also compared the results with previously described viromes of Australian honey bees. The Czech viromes showed high differences from those in Australia. Importantly, DWV and ABPV were absent in Australian bees, while they were present in Czech honey bee populations. On the contrary, the virome of the Australian honey bees was rich in various Picornavirales, which were absent in the Czech honey bees. The difference might be caused by *Varroa* mite that was absent in Australia. Thus, interestingly, we proposed the possible replacement of the various Picornavirales by *Varroa* associated with increasing DWV and ABPV infections could occur in regions where the mite is widespread. This assumption is consistent with *Varroa* making certain viruses dominant.

4.2. Project 2

4.2.1. First year of Project 2 - new viruses

Discovery and characterization of novel DNA viruses in Apis mellifera: expanding the honey bee virome through metagenomic analysis

In recent years, the exploration of viruses in honey bees has yielded a number of new RNA viruses, but only one DNA virus has been identified: Apis mellifera filamentous-like virus (AmFV). In our study, we discovered and characterised two new large DNA viruses: Apis mellifera filamentous-like virus (AmFLV) and Apis mellifera nudivirus (AmNV) and nine other smaller DNA viruses (Bee densoviruses 1-9).

In our study, we employed binning and co-assembly techniques to reduce the number of contigs associated with large viruses, facilitating the design of primers for PCR amplification. The ‘bridges’ between the contigs from the PCR were then sequenced and the contigs were connected. After polishing we gained two complete genomes: ~152 kbp for AmFLV and ~129 kbp for AmNV.

AmFLV, characterised by linear DNA and inverted terminal repeats (ITR), shares similarities with AmFV, suggesting an evolutionary relationship. Our genomic analysis revealed a number of open reading frames (ORFs) within AmFLV, proteins potentially influencing host cell metabolism and apoptosis. This sheds light on the genetic complexity of AmFLV and its possible role in modulating bee physiology.

Similarly, our investigation into AmNV, classified within the Alphanudivirus genus, uncovered intriguing aspects of this novel nudivirus. With a circular genome containing multiple ORFs, AmNV shares genetic similarities with the core genes of the *Nudiviridae* family, suggesting its evolutionary relationship. Its widespread detection across multiple bee samples underscores its prevalence within honey bee populations, prompting questions about its ecological significance and potential implications for bee health.

4.2.2. Unpublished data of Project 2 - longitudinal analysis

We sequenced 48 samples, as follows:

1) 18 samples from 2019, three hives per site, in the case of the VURV site we added three more samples. Two were source bees for the experiment (abbreviated as VURVH and VURVD) and one unique hive that seemed resilient to viral infection (VURV7).
2) 15 samples from 2020, three hives per location.
3) 15 samples from 2021, three hives per location.
4) Three negative controls, one per year, and not included in the sample count.

Due to the loss of some observed hives over time, necessitating the substitution of hives at certain locations, subsequent analysis treated the location as repeated measurements of a singular entity. While this approach allowed longitudinal analysis, it also imposed limitations by obscuring the dynamics of individual hives, focusing instead on the viral dynamics within each location.

Statistics of sequencing results are shown in Table 3. In a large majority of samples, we had >10M reads even after trimming, but the percentage of viral reads was highly variable between samples, from 3.8% to 99.3%. This high variability in the number of viral reads is caused by a very high abundance of one virus (DWV or variant of LSV) which overshadows reads of other viruses and non-viral reads. This observation we made already in our first study in the case of DWV (158) and it was one of the reasons why we increased the number of bees to 50 individuals per sample.

Table 3- Basic statistics of sequenced samples.

Year	Sample	Reads (trimmed)	Viral reads	% of viral reads	Note	Year	Sample	Reads (trimmed)	Viral reads	% of viral reads	Note
2019	Lisnice11	31,076,599	10,550,554	34.0		2020	Prasily1	21,613,569	5,695,039	26.3	
	Lisnice24	19,091,873	5,188,533	27.2			Prasily2	11,995,591	2,621,285	21.9	
	Lisnice333	17,977,650	16,213,132	90.2			Prasily3	10,629,204	741,376	7.0	
	Brdy1	40,803,907	16,686,775	40.9	As Nerezin in 20/21		VURV1	11,432,528	2,862,727	25.0	
	Brdy2	23,117,498	9,194,386	39.8	As Nerezin in 20/21		VURV5	18,722,703	14,226,737	76.0	
	Brdy3	14,859,869	8,502,457	57.2	As Nerezin in 20/21		VURV4	6,559,249	478,904	7.3	
	Libechov11	27,607,013	14,476,774	52.4			Libechov1	18,820,999	1,697,991	9.0	
	Libechov14	12,922,552	891,042	6.9			Libechov10c	13,334,009	2,168,413	16.3	
	Libechov6	13,488,775	4,726,933	35.0			Libechov15	18,274,890	4,142,607	22.7	
	Prasily1	11,199,897	11,013,844	98.3		2021	Lisnice1	17,046,292	5,475,842	32.1	
	Prasily2	12,257,953	11,745,538	95.8			Lisnice2	20,032,318	8,858,139	44.2	
	Prasily3	35,351,758	17,478,510	49.4			Lisnice3	18,460,543	907,924	4.9	
	VURV1	20,311,029	13,020,637	64.1			Nerezin2	20,879,257	5,174,685	24.8	as Brdy in 2019
	VURV5	25,364,850	18,083,600	71.3			Nerezin4	20,373,718	9,349,718	45.9	as Brdy in 2019
	VURV4	37,424,653	35,597,081	95.1			Nerezin6	11,768,454	9,079,986	77.2	as Brdy in 2019
	VURV7	23,001,338	20,429,173	88.8			Libechov1	26,339,335	12,439,197	47.2	
	VURV_H	14,267,600	9,794,762	68.7	cage experiment		Libechov10c	21,579,342	17,643,814	81.8	
	VURV_D	9,841,001	9,775,180	99.3	cage experiment		Libechov15	17,411,657	8,411,945	48.3	
2020	Lisnice1	18,172,407	691,480	3.8			Prasily1	17,684,890	11,880,628	67.2	
	Lisnice4	6,517,007	363,215	5.6			Prasily2	21,556,595	9,543,288	44.3	
	Lisnice7	15,514,474	879,352	5.7			Prasily3	16,879,219	8,678,955	51.4	
	Nerezin1	20,954,359	12,146,800	58.0	as Brdy in 2019		VURV1	18,175,380	10,460,334	57.6	
	Nerezin2	17,574,456	3,035,360	17.3	as Brdy in 2019		VURV5	27,050,026	14,824,678	54.8	
	Nerezin3	15,321,032	13,268,303	86.6	as Brdy in 2019		VURV6	19,796,418	6,493,000	32.8	

Entire virome

Consistent with our earlier findings (158) the composition of the entire virome was significantly influenced by the location of sampling ($p < 0.001$***), with location accounting for approximately 16.7% of the observed variability. Additionally, there was a modest association between virome composition and the year of sampling ($p = 0.03$*). The results of dimensional reduction analysis are depicted in Figure 12, revealing clustering patterns possibly attributable to the presence of phages and plant viruses. This aligns with our previous observation (158) that the composition of these viral groups is strongly correlated with the sampling location. Notably, we observed low variability in virome composition exclusively at the Libechov location, while other sites exhibited higher levels of dispersion.

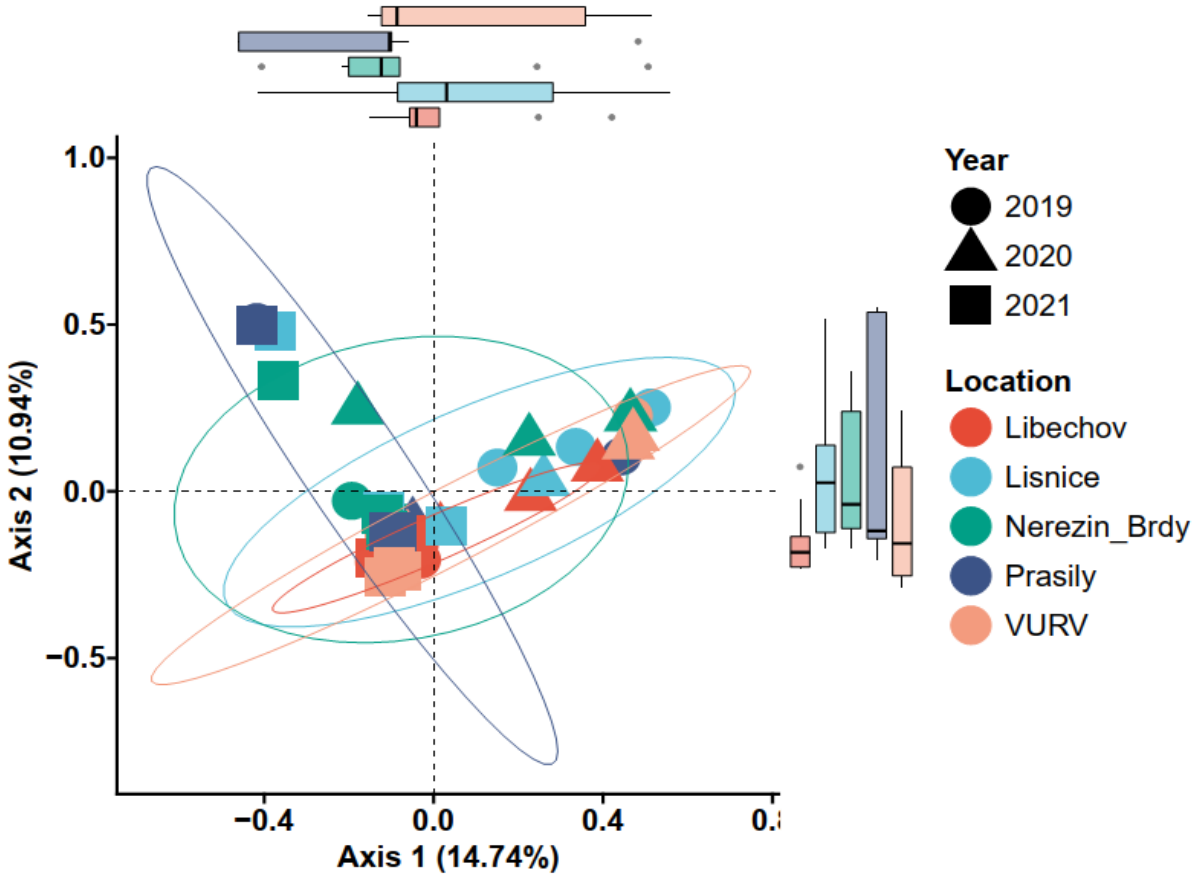


Figure 12- Dimensionality reduction *of gained high-quality vMAGs.. Boxplots on the side display the distribution of samples from the same hive on axes 1 and 2.*

Changes in the composition of the most abundant families within the whole virome were evident and are illustrated in Figure 13. In addition to honey bee-infecting viruses (e.g., *Dicistroviridae* like BQCV, *Iflaviridae* like DWV, *Sinhaliviridae*) phages (*Roundtreeviridae*), and plant viruses (*Partitiviridae*, *Secoviridae* and *Solemoviridae*), other viral families were present. It is noteworthy that the 'Unclassified' category constituted approximately 25% of the relative abundance. This high proportion can be attributed to incomplete viral taxonomy at certain hierarchical levels. For instance, at the time of publication of (158), the family *Sinhaliviridae*, to which Lake Sinai viruses (LSV) belong, had not yet been established. Furthermore, some phages lack extensive classification due to various reasons, such as limited availability of data or our focus on eukaryotic viruses that led to classification at taxonomical levels where certainty could be ensured, potentially resulting in restricted taxonomy due to the absence of thorough manual curation.

Another notable observation is the absence of the *Partitiviridae* family at the VURV location, despite its abundant presence at other sampling sites.

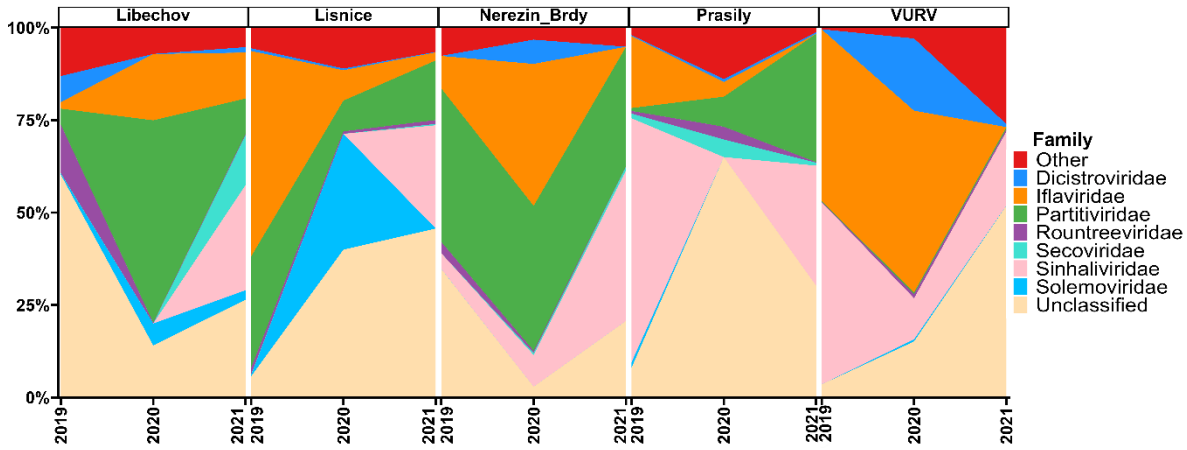


Figure 13- Changes in relative abundance of 8 most abundant families *of high-quality vMAGs; the 9th group "Other" is agglomerated and composed of every other family of high-quality vMAGs (low abundant families). Unclassified is a group composed of high-quality vMAGs that don't have classification on the 'Family' level.*

A subtle trend of increasing diversity over the years is discernible, albeit modest, with the lowest diversity observed in the year 2020 (Figure 14). This trend is consistent across all locations. However, it is essential to note that the pronounced peaks in the violin plot may be attributed to the presence of very high viral loads of specific viruses, such as LSV/DWV, which can skew the diversity metrics.

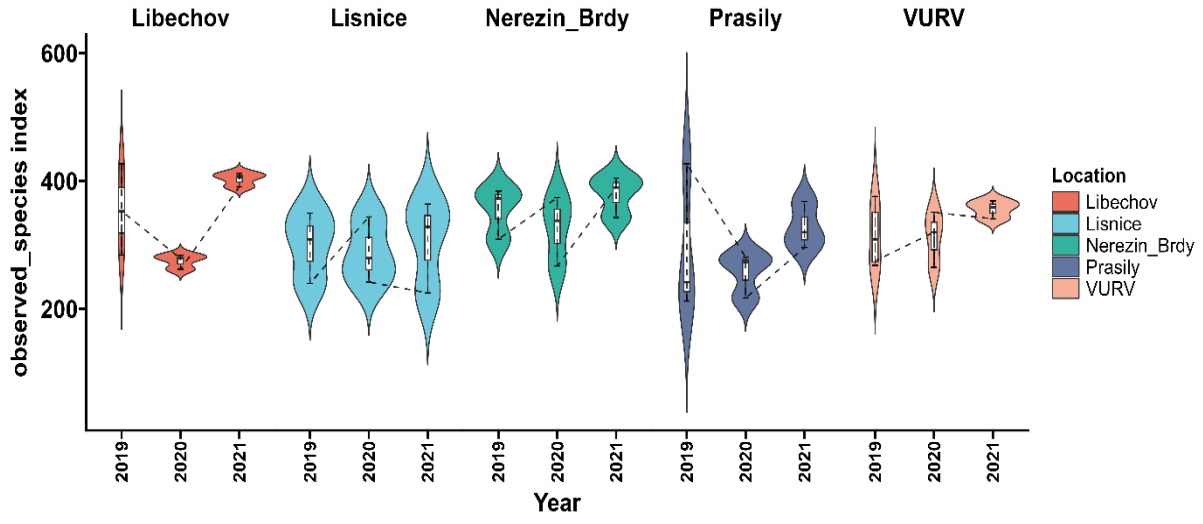


Figure 14- Violin plot of observed species diversity *over time, split by location. It comprises all high-quality vMAGs and observed species is used as a measure of diversity.*

We conducted an evaluation of the dynamics of the six most abundant viral families (one of them being the viruses without classification on the 'Family' level), revealing several noteworthy trends. The notable decline in relative viral abundance observed in the Iflaviridae family in 2021 is of particular interest. Remarkably, this decrease was consistent across all

locations, except for Libechov, where there was a discernible increase starting from a very low abundance at the commencement of sampling.

Similarly, the year 2020 exhibited a low abundance of *Sinhaliviridae*, contrasting with higher levels observed in other years. Additionally, there was an elevated abundance of *Dicistroviridae* in 2020 compared to other years. These trends are visually represented in Figure 15.

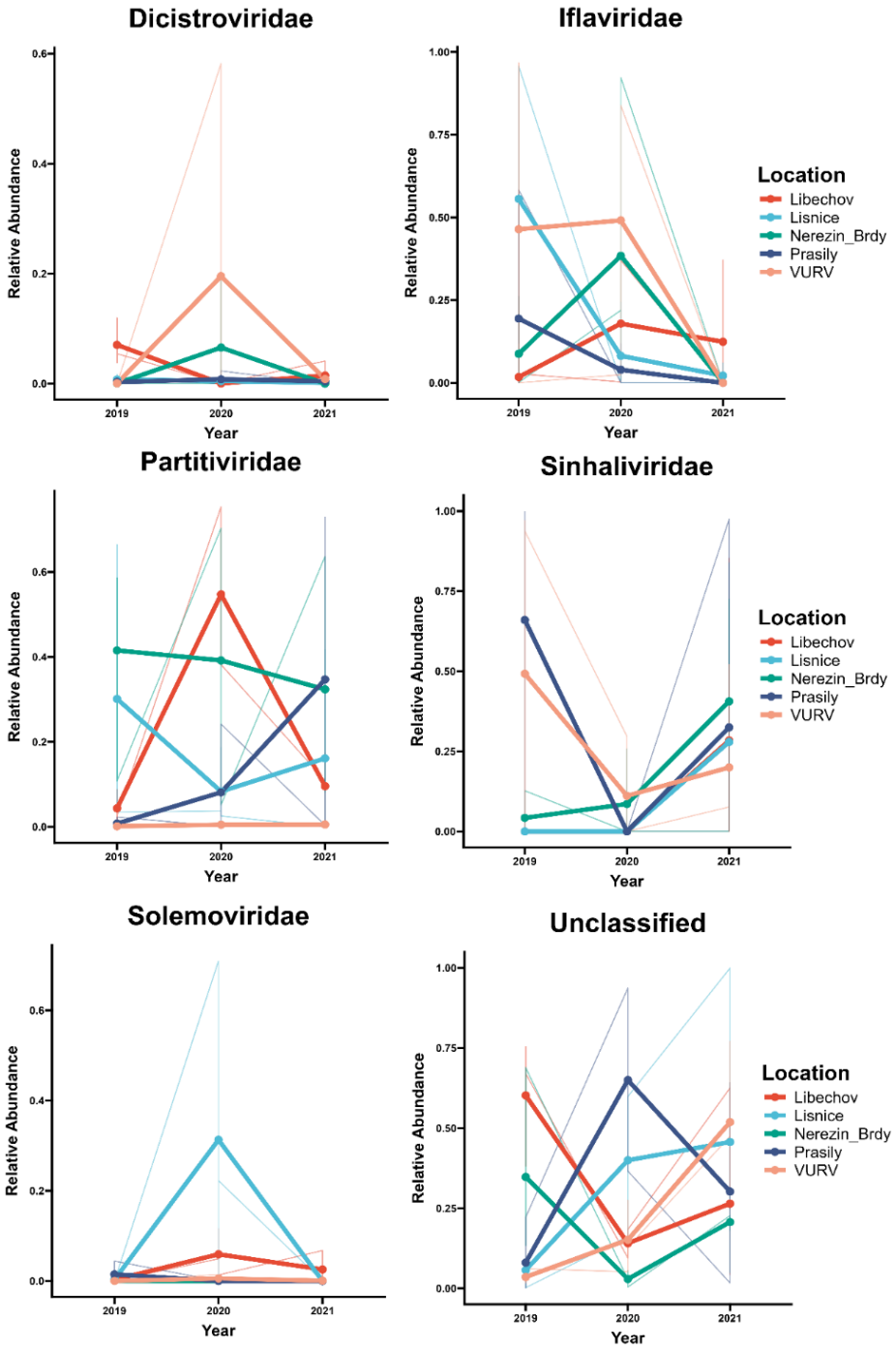


Figure 15- Line plots of trends in the six most abundant viral families *and their relative abundance change in abundance over time.*

Honey bee-infecting viruses

When focusing solely on bee-infecting viruses, no strong trends were observed. The diversity measures exhibited different patterns depending on the location, as depicted in Figure 16. While it might be anticipated that increases in the diversity of the whole virome over the years would correspond to increases in viral diversity specifically among bee viruses, this relationship is only mildly evident. Notably, the Nerezin_Brdy location exhibited a trend of decreasing diversity, contrary to expectations.

The overall trend of the diversity of only honey bee-infecting viruses followed the same trend as the overall diversity of the whole virome. With one notable exception, Nerezin_Brdy, where the diversity had the opposite trend and was reduced in two consecutive years. See Figure 16.

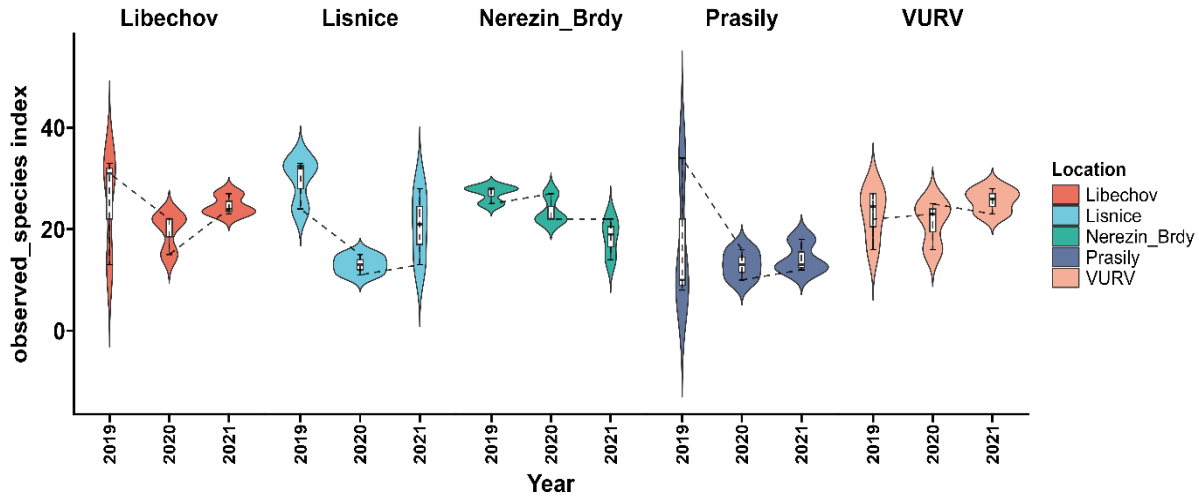


Figure 16- Violin plots of measure of bee-infecting virus diversity *between locations over the three years. It comprises honey-bee infecting high-quality vMAGs and observed species is used as a measure of diversity.*

In contrast to our previous study (158), we observed a dependency of honey bee virus prevalence and abundance on both the location (p = 0.002**, explaining 14.78% of the variance) and, to a lesser extent, the year of sampling (p = 0.018*). Dimensionality reduction analysis focusing solely on bee-infecting viruses is presented in Figure 17.

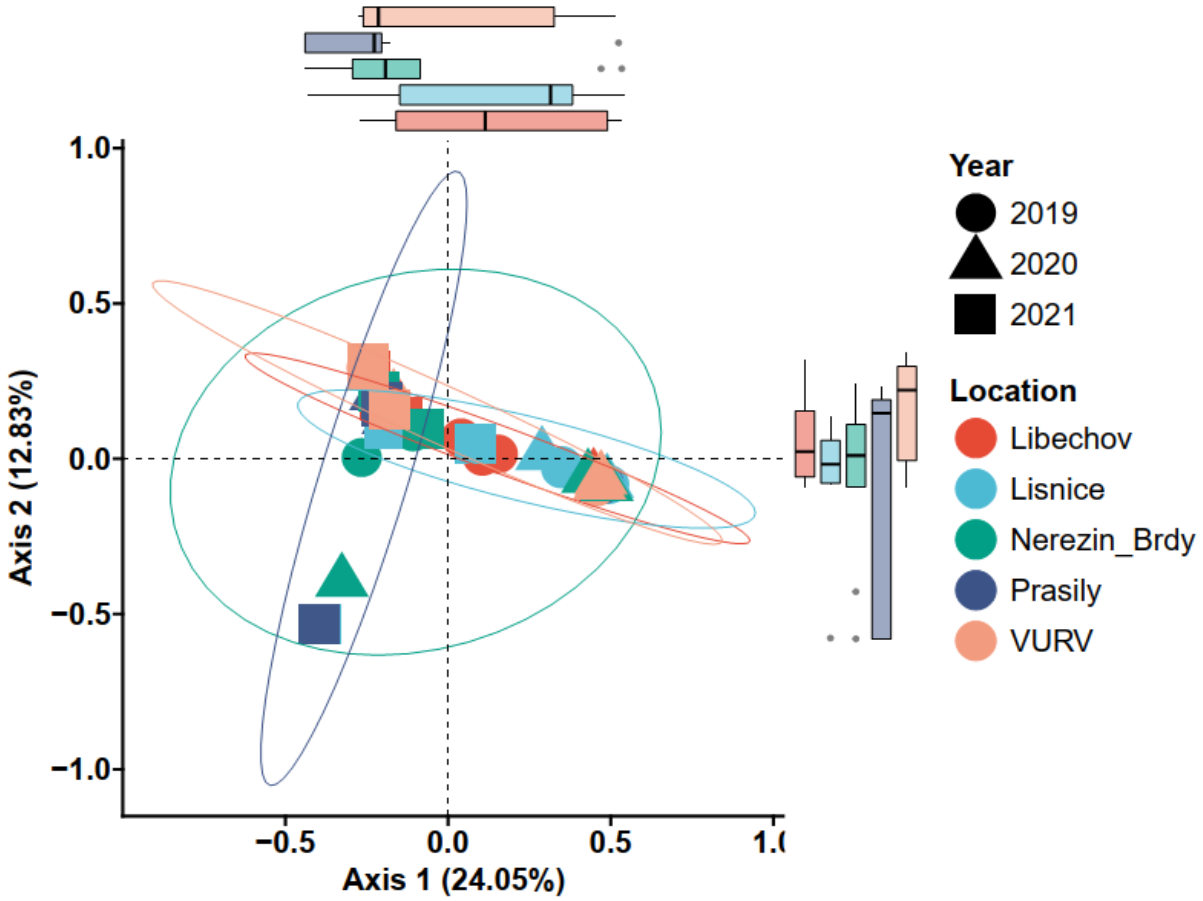


Figure 17- Dimensionality reduction for honey bee infecting high-quality vMAGs. *Boxplots on the side display the distribution of samples from the same hive on axes 1 and 2.*

One of the most intriguing observations pertains to the dynamics of the DWV over the study period. We noted a notable increase in DWV load followed by a subsequent decline (refer to Figures 19 and 20 for visualization; further details below). Although LSV also exhibited changes over the years, its prevalence was lower compared to DWV (refer to Figures 18 and 19).

Figure 20 highlights an interesting trend in viral abundance across most locations: the majority of viral families, except for *Iflaviridae*, displayed a trend resembling the shape of a ‘V,' indicating a drop in abundance in 2019. Conversely, *Iflaviridae* exhibited a trend resembling the shape of a 'Λ,' signifying an increase in abundance in 2020.

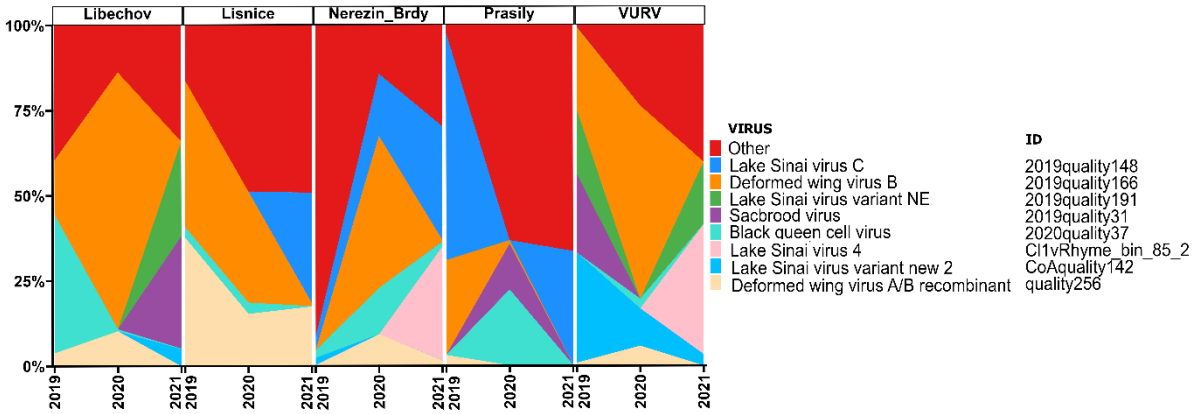


Figure 18- Changes in 8 most abundant families *of high-quality vMAGs that infect honey bees; the 9th group "Other" is agglomerated and composed of every other family of high-quality vMAGs (low abundant families). Unclassified is a group composed of high-quality vMAGs that don't have classification on the 'Family' level.*

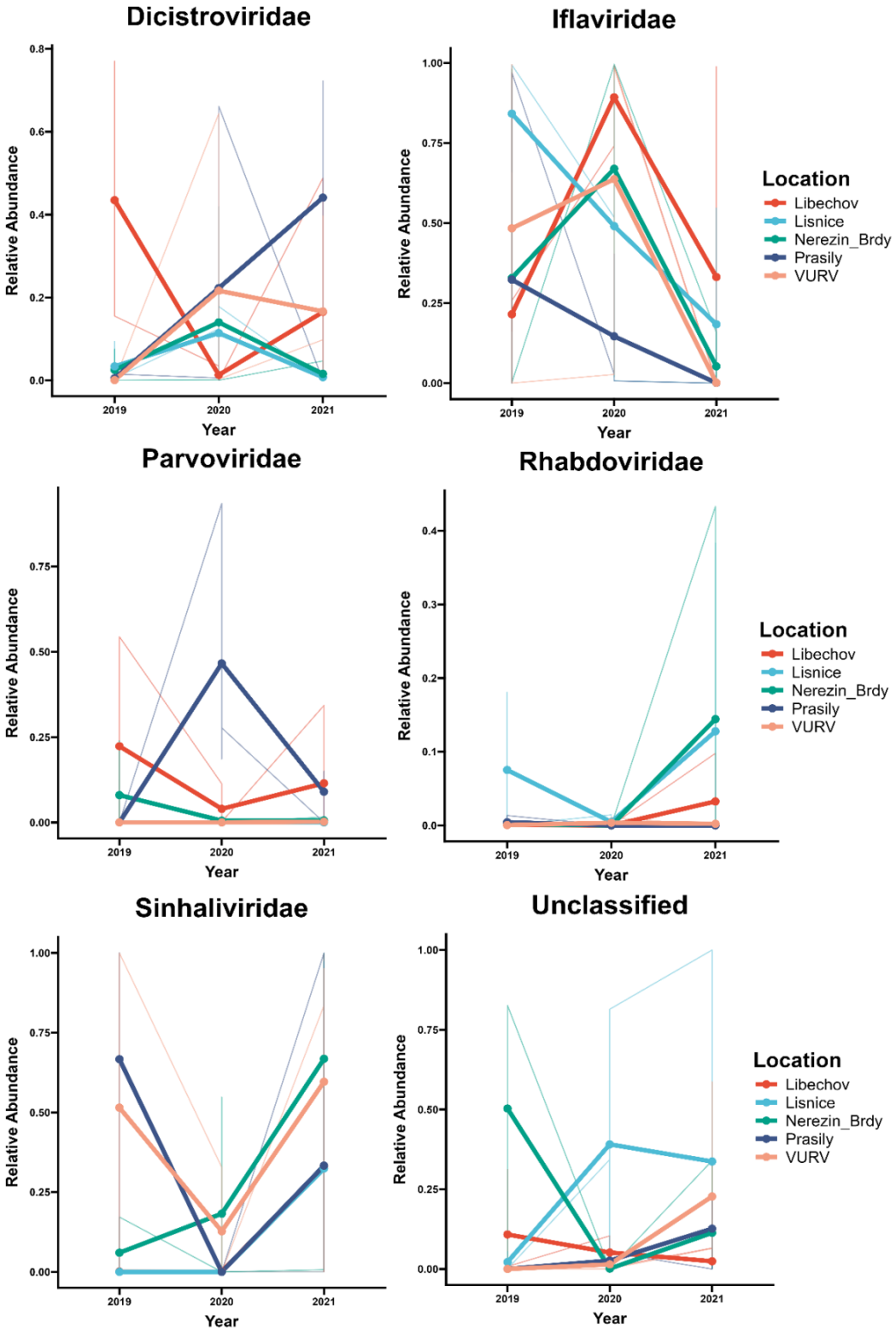


Figure 19- A closer look at trends over three years in 6 individual virus families that infect honey bees.

The heatmap analysis suggests the presence of three sample groups within the virome. Firstly, there is a sample group characterised by a high abundance of DWV. Secondly, another sample group exhibits a high abundance of LSV. Finally, there is a third sample group characterised by a low abundance of viruses but with high diversity. Locations are not clustering very well.

On another note, in comparison to our previous study (158), AmFV was almost absent. These observations are illustrated in Figure 20.

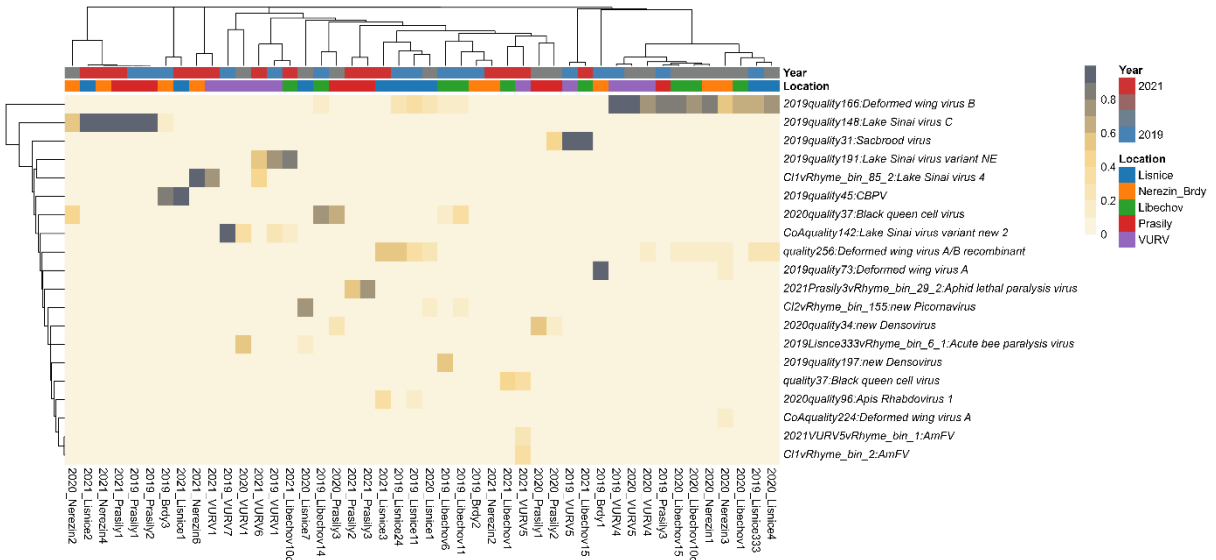


Figure 20- Viral heatmap of the top 20 *high-quality vMAGs that we classified as honey bee-infecting viruses. vMAGs are named with their unique ID and then the common name of the virus. On a scale of relative abundance.*

Notable associations were observed in the honey bee virus composition. In 2019, there was an increase in both the abundance and prevalence of Aphid Lethal Paralysis Virus (ALPV), Bee Rhabdovirus 1 (BRV-1), and LSV-C, while Black Queen Cell Virus (BQCV) exhibited decreased levels compared to other years. In 2020, there was a noticeable rise in DWV-B and ABPV, whereas BRV-1 and ALPV showed decreased abundance and prevalence. The year 2021 presented features including an increase in BQCV and LSV-4, and a decrease in one variant of DWV-A (CoAquality224); see Figure 21.

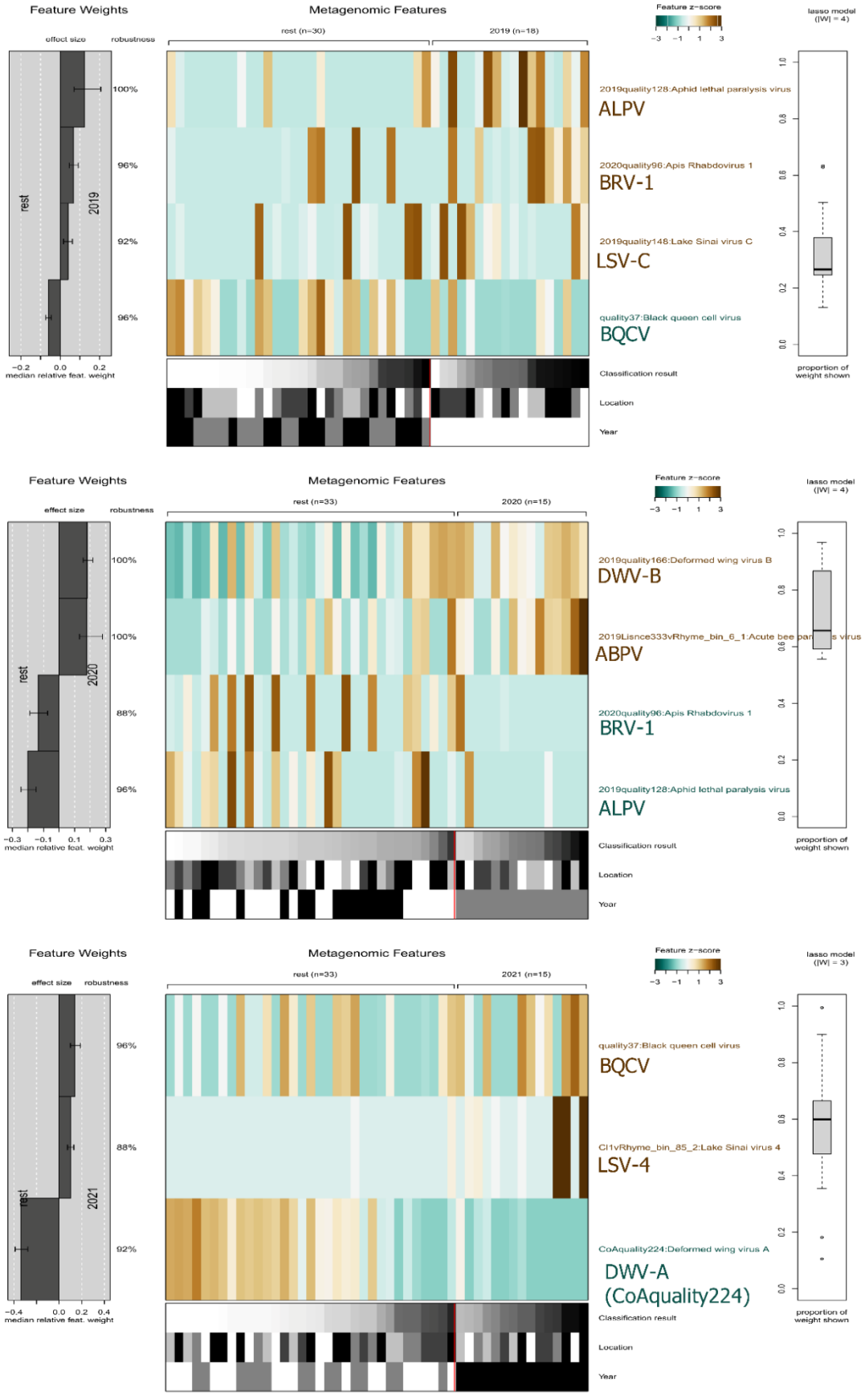


Figure 21- Interpretation plots *of SIAMCAT for all three years, from 2019 (the most upwards) to 2021 (lowest).*

Subsequently, we focused on examining differences in viral abundances of DWV and LSV over time and across different locations.

DWV

In our analysis of DWV, we identified two different variants of DWV-A, along with one variant of DWV-B and one recombinant of DWV-A and DWV-B. Notably, DWV-B was found to be the most abundant and prevalent virus in our samples. However, it was absent in certain samples where another virus, such as LSV, was exceptionally abundant (e.g., Prasily2 from 2019).

An interesting observation comes from one location: Nerezin_Brdy. Overall, the Nerezin_Brdy location displayed a distinct profile, characterised by a notable presence of DWV-A compared to other locations. One of the DWV-A variants (with unique ID of vMAG "2019quality73") appeared to be present independently of DWV-B, exhibiting high concentrations in two samples and a lower concentration in a third sample where DWV-A/B recombinant was also detected.

Three additional intriguing observations were made:

1) In cases where one variant of DWV was high in abundance, other DWV variants were found to be low in abundance, hinting at potential competition between DWV variants.
2) The abundance of DWV-B was significantly highest in 2020, whereas the abundance of other variants varied more over the study period.
3) The presence of DWV-A/B recombinant was noteworthy, as it was detected at most sites even in the absence of DWV-A and, in some cases, without DWV-B.

Refer to Figures 22 and 23 for visualisation of these observations.

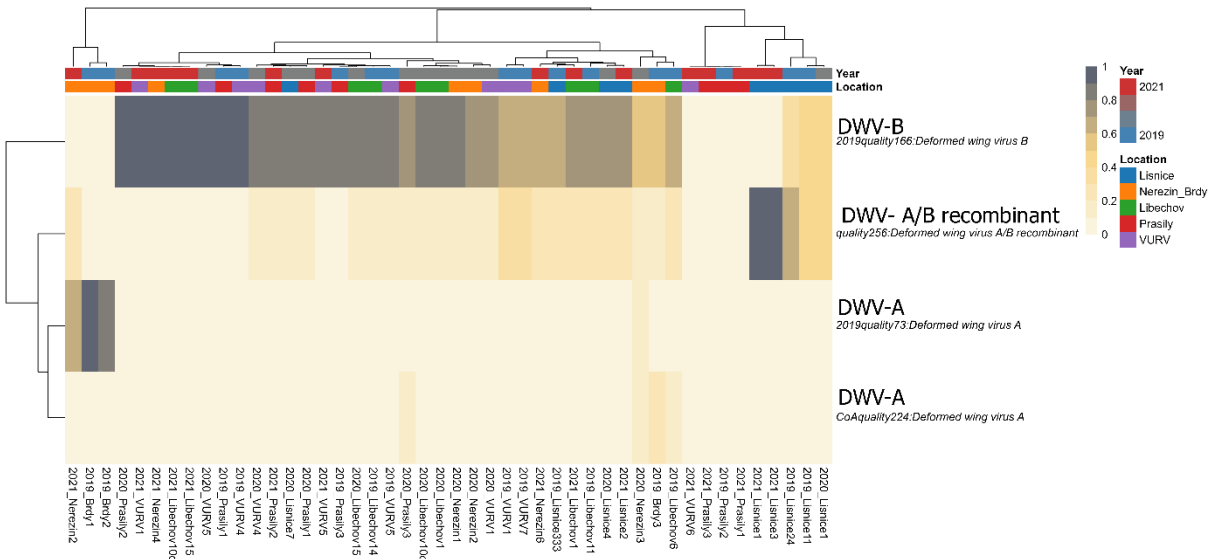


Figure 22- Heatmap of DWV viruses *present in our set of high-quality vMAGs, on a scale of relative abundance. Unique IDs are under the common name of the virus.*

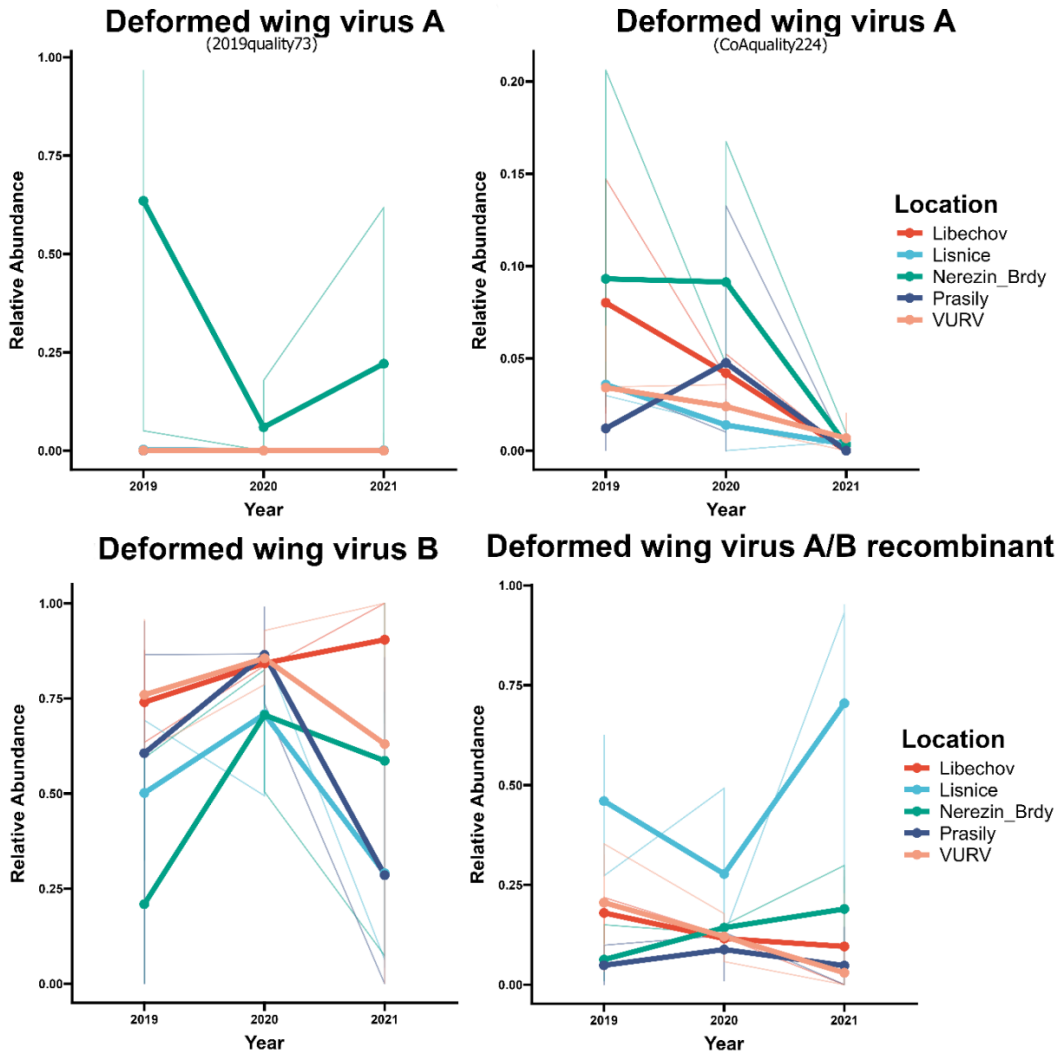


Figure 23- A closer look at trends over three years in changes of DWV *variants we found in our high-quality vMAGs. For DWV-A their specific IDs are under a common name.*

LSV

As shown in Tables 1 and 2, LSV is a benign virus without pathology but has many variants, 15 recognised by NCBI as of 16/05/2024 (over 30 differently named variants of genomes >5kbp). The naming is complex and not standardised. The names of variants vary from numbers to multiple letters. In (158) we used one-letter identifiers to distinguish between variants; in different studies, other identifications were used. Here we avoided naming the new variants for now and used the names which are already in use (closest variant). See Figure 24.

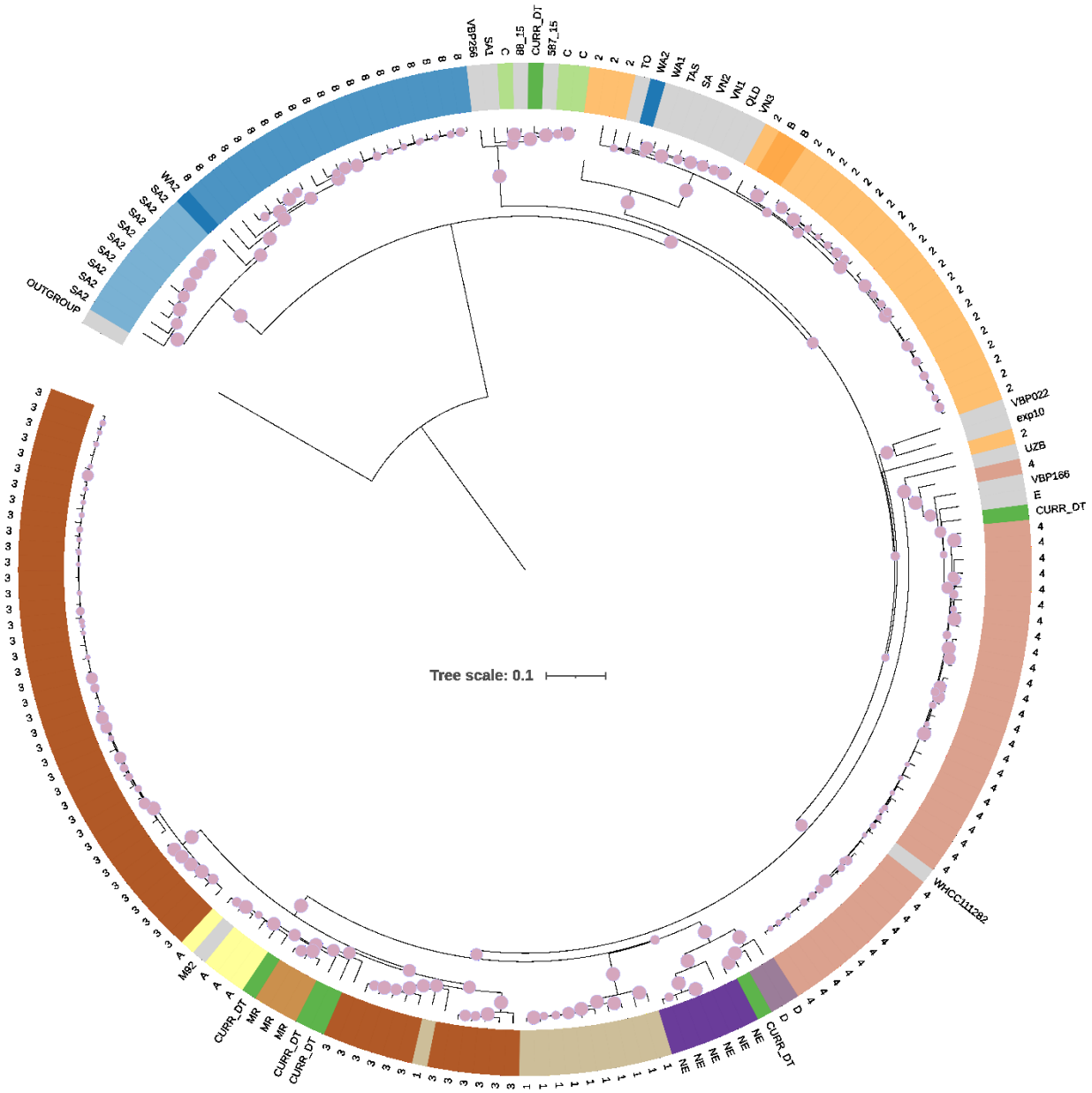


Figure 24- Phylogenetic analysis of LSV variants *>5kbp available to 16/05/2024. Existing variants selected from NCBI Virus, alignment with mafft version 7 (https://mafft.cbrc.jp/alignment/server/), from conservative sites, the tree was done with NJ and 100 bootstraps. Visualisation with iTOL (https://itol.embl.de/).*

Overall, we identified six variants of LSV, with two variants being previously not known (provisionally nicknamed LSV new and LSV new 2; ~91-92% nt identity to existing variants). Among these variants, LSV-C was the most abundant, consistent with our previous observations published in (158). Variants NE, 3, and 4 had been described previously in other countries. Notably, for LSV, it appears that the virus is present either in high concentrations or as a mixture of several genotypes, each with a relatively lower abundance.

Furthermore, the variability of LSV exhibited a potential drop in 2020, although high variability persisted in 2019 and 2021, as depicted in Figure 25.

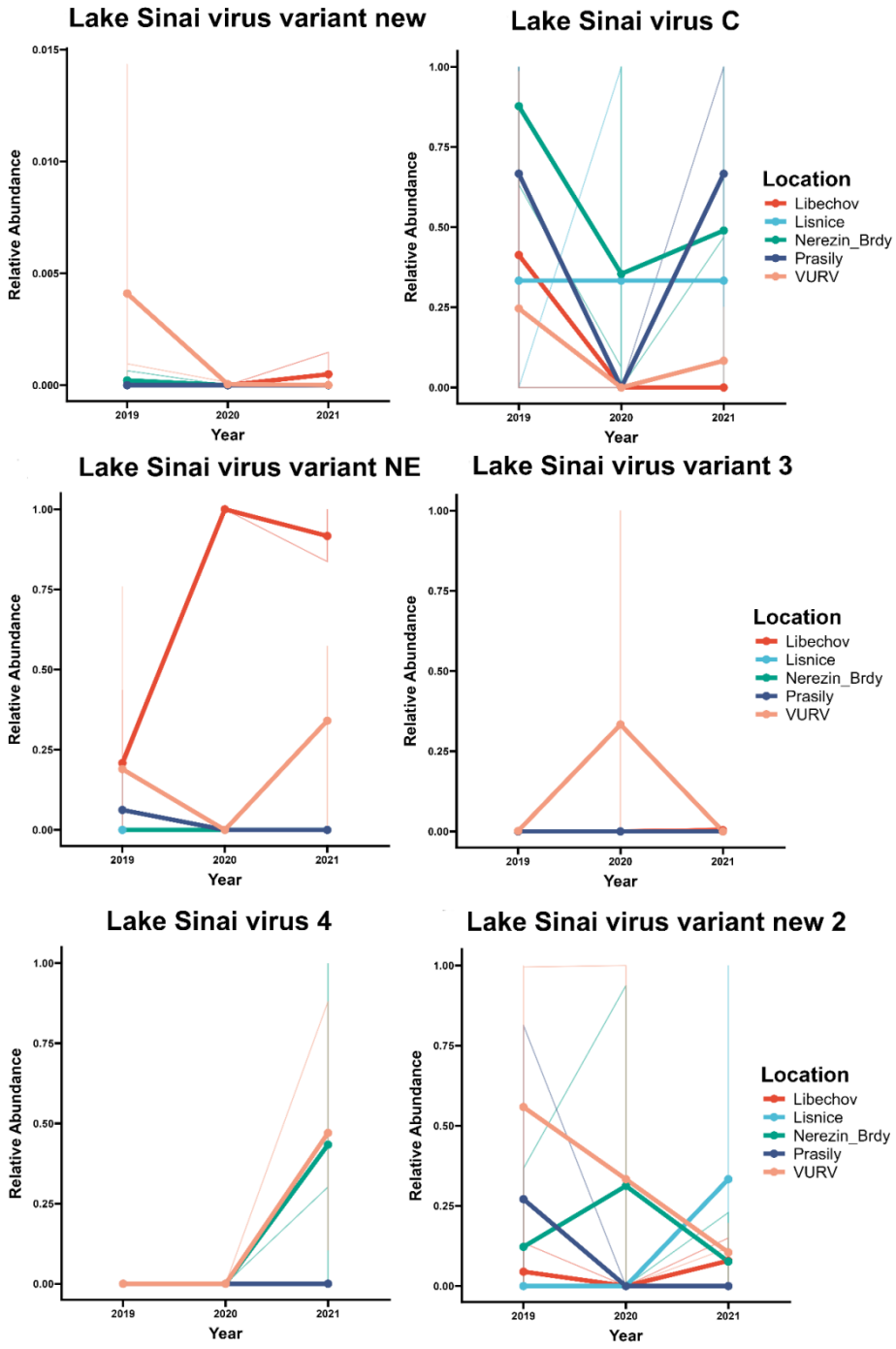


Figure 25- A closer look at trends over three years in changes in the abundance of LSV *variants we found in our high-quality vMAGs.*

New viruses

In addition to the viruses described in our previous article about novel DNA viruses, we identified several new RNA viruses. Among these, four new virus genomes were completed and characterised, as illustrated in Figure 26 and detailed in the blast results provided in Table 4.

Of particular interest are two picornaviruses that showed significant alignment with Australian bee viruses. Precisely it was Bee Picornavirus 1 with 31.7% identity over 73% of query sequence to Bundaberg bee virus 4. And then Bee Picornavirus 3 with 56% of sequence having 41.7% identity to Darwin bee virus 6.

However, it is important to note that while we detected other RNA viruses, their genomes were incomplete. For instance, we identified ~8 kb-long fragment of *Mononegavirales*, possibly representing the presence of a third Rhabdovirus.

Table 4- Results of BLAST *for the new viruses, all with the exception of Bee Picornavirus 2 are results of BLASTx (for Bee Picornavirus 2 it's BLASTn).*

Product [Virus]	Coverage	E value	identity	Accession
Bee Picornavirus 1				
polyprotein [Bundaberg bee virus 4]	73%	0	31.74%	AWK77860.1
putative polyprotein [Myrmica rubra picorna-like virus 8]	47%	0	28.71%	UXD80108.1
Bee Picornavirus 2				
hypothetical protein gene, complete cds [Insect picorna-like virus 1]	98%	0	86.67%	MN714669.1
Bee Picornavirus 3				
nonstructural polyprotein [Darwin bee virus 6]	56%	0	41.70%	AWK77846.1
nonstructural protein [Hovenia dulcis-associated virus 1]	57%	0	40.35%	QNT09302.1
Bee picornavirus 4				
polyprotein [Iflaviridae sp.]	76%	0	36.78%	ULF99891.1
RNA-dependent RNA polymerase [Riboviria sp.]	68%	0	35.95%	WKV33237.1

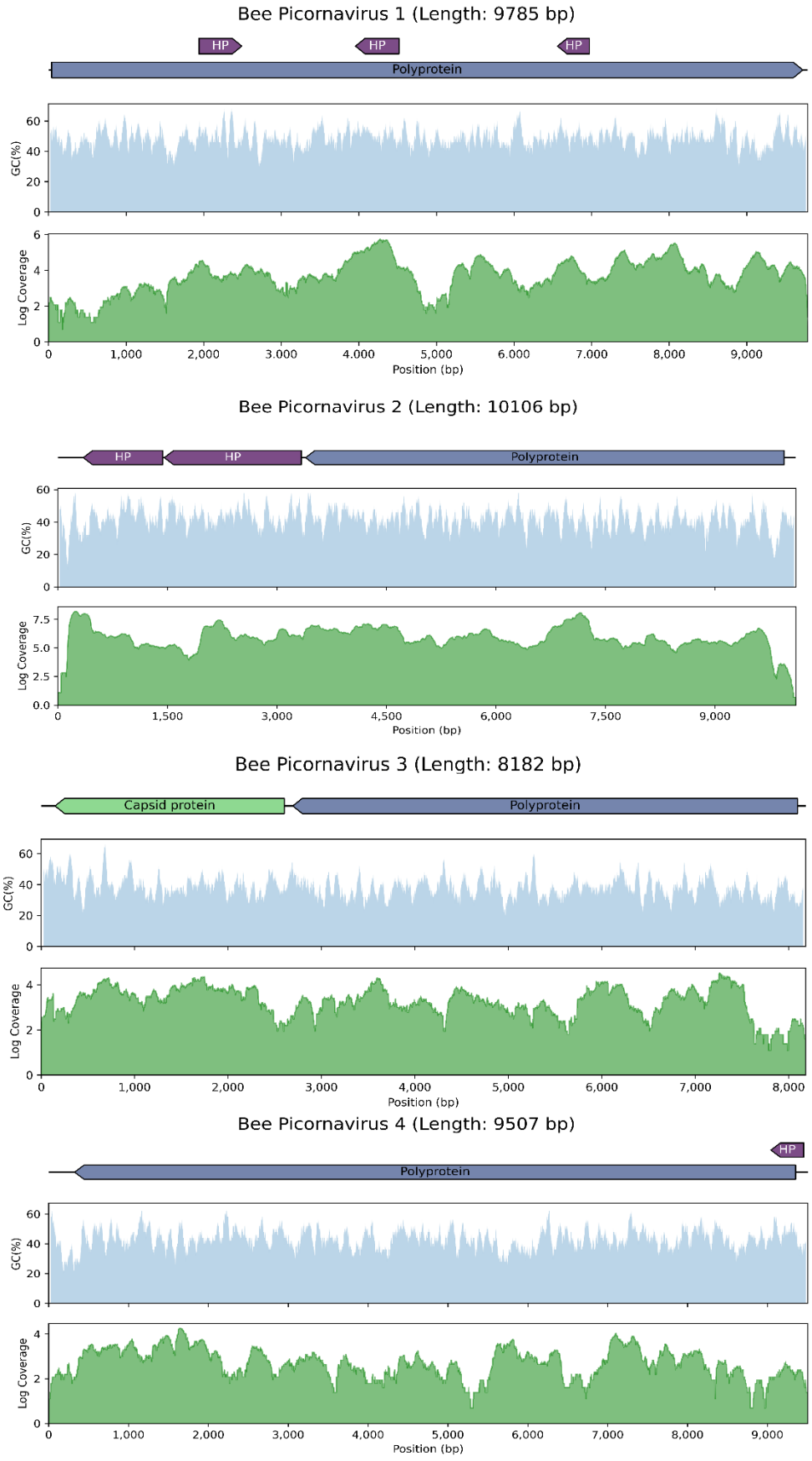


Figure 26- The basic information about the new viruses. *First, their genome is shown, under which are the GC content and the coverage in a log scale.*

4.2.3. Integration of -omics data

A manuscript containing a combined integration of the whole virome with bacteria and eukaryotes is in the preparation stages. Therefore, only a few results are discussed here.

We restricted the analysis to a honey bee-infecting viruses. The combined data clustered very well with the exceptions of Prasily and VURV locations. Prasily 3 was outliner in many features; while six VURV hives split into two groups, one closer to Lisnice and the other closer to Libechov.

Interestingly, apart from VURV, all sites had defining characteristics between viruses. For example, Prasily locations were defined by one variant of LSV (precisely LSV-C; the Czech variant we detected before (158)) and VURV was not defined by any specific viruses. In further analysis, we noted that each component (16S and 18S rRNA, viruses) had to a certain degree stable and a variable part (e.g., Figure 27 for viruses from BQCV to ABPV the expression of virus per location is stable while the rest of the viruses are variable on the locations).

Then we restricted the analysis to interactions between pesticides and selected bee-infecting viruses. The pesticides with the most interactions were: Mepiquat and Acetamiprid. They were negatively associated with the presence and abundance of several viruses (Figure 27). What is interesting is that they mostly did not belong to “high-abundant” honey bee viruses (DWV, ABPV, CBPV, BQCV, Sacbrood virus, some variants of LSV) but to several “low-abundant” viruses.

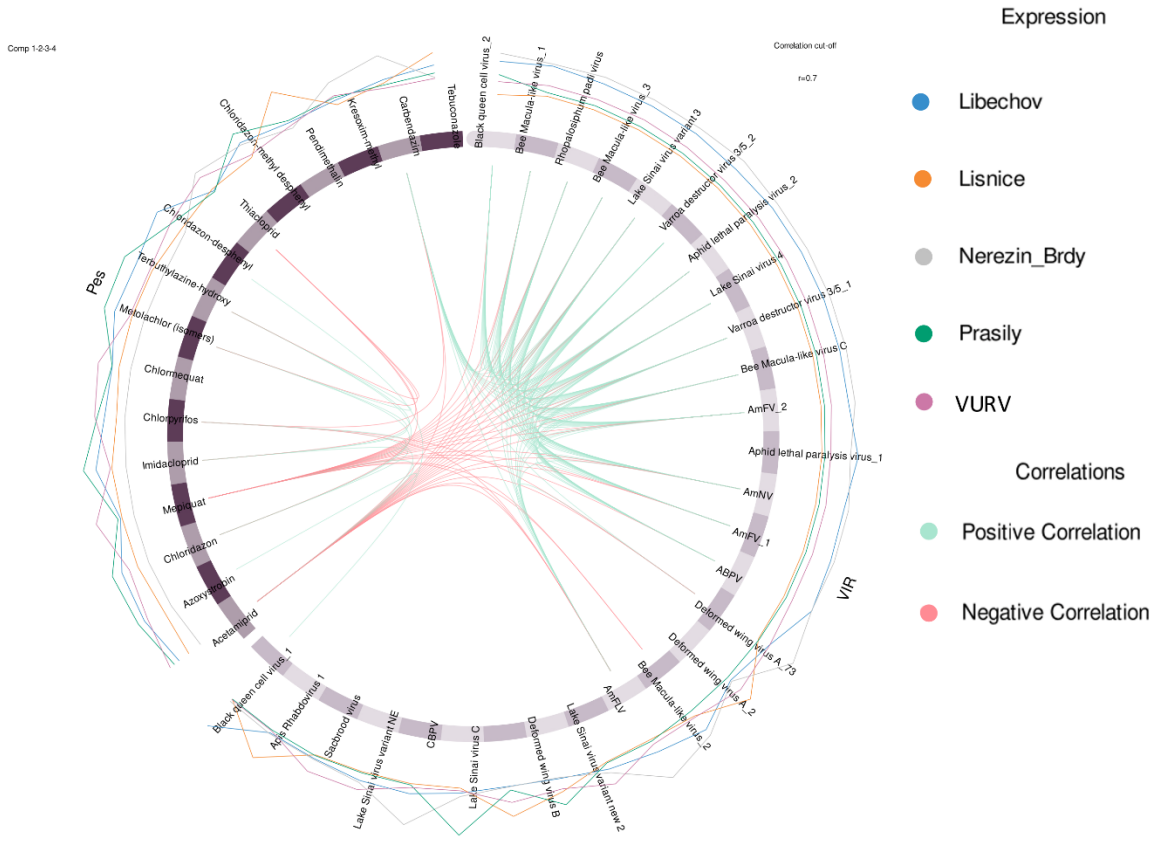


Figure 27- Circos diagram of interactions between individual components.

4.3. Other results

4.3.1. Proteomics

Varroa destructor parasitism and Deformed wing virus infection in honey bees are linked to peroxisome-induced pathways

In this study, our collaborators investigated the impact of *Varroa* mites, particularly their role in transmitting viruses like DWV, on honey bee colonies. By exposing newly emerged worker bees to *Varroa* for 72 hours, they observed significant changes in the bee proteome and identified various viral proteins. Notably, DWV was found at high levels in *Varroa*-exposed bees, indicating its transmission by *Varroa* and potential harm to bee health.

Proteomic analysis revealed DWV-B presence in all *Varroa*-exposed bees, while controls showed no DWV presence, confirming the association between *Varroa* infestation and DWV infection. Other viruses detected by proteomics included Apis Flavivirus (AFV), AmFV, Kashmir Bee Virus (KBV), and Bee Macula-like Virus (BMLV).

While DWV was the most prevalent, persistent infections of AFV and BMLV were also observed, possibly related to *Varroa* infestation. We performed an analysis of the virome in

the source bees for the experiment and found viruses that were described by proteomics (some of them in trace amounts) but it confirms the results. We found all the viruses except for AFV, which is a highly unusual sequence (see discussion).

In *Varroa*-exposed bees, several significant changes were observed by proteomic analyses:

1) Peroxisomal Metabolism: *Varroa*-DWV exposure upregulated peroxisomal proteins, suggesting alterations in fatty acid metabolism and ROS regulation.
2) Immune Response: Proteins associated with Toll-like receptor 4 (TLR4)-mediated viral infection were upregulated, indicating an activation of innate immune responses.
3) mTORC1 Pathway: Suppression of the mTORC1 pathway was observed, affecting protein synthesis and immune modulation, potentially facilitating viral replication.
4) Autophagy: Downregulation of the Ragulator complex protein LAMTOR1 suggested an increase in autophagy, particularly pexophagy, which may impact viral replication and cholesterol transport.
5) Cytoskeletal Remodelling: Changes in cytoskeletal and cuticular proteins were observed, possibly linked to *Varroa* feeding and DWV transmission dynamics.

My contribution was to the virome preparation and analysis of the source bees, writing this part of the manuscript, and reviewing the manuscript.

4.3.2. Determination of viral load changes in bees after exposure to pesticides

We used primers designed in (174) for DWV-A, DWV-B and Actin; and for ABPV we newly designed the primers. We designed standards and had them synthesised commercially, and we optimised the final concentration of primers in reactions and the cycling conditions. For DWV we used 375nM concentration of primers and two-step annealing-extension: annealing at 58.5 °C for 15 seconds and extension at 72 °C for 15 seconds, for ABPV 250nM and annealing and extension were done in one step at 60°C for 30 seconds. The rest of the cycling protocol followed the recommendation of the manufacturer. Reverse transcription was done with iScript™ cDNA Synthesis Kit and qPCR with SsoAdvanced Universal SYBR Green Supermix. Both from BioRad (Hercules, California, United States).

In this experiment, bees from the VURV location; not parasitized by *Varroa*, for our purposes referred to as Hive A were exposed to pesticide in 50% sugar solution or control without pesticide for 72 hours before being introduced into another hive from VURV which was highly infested with *Varroa* mites, referred to as Hive B. Bees from hive A were marked with colours on the thorax to identify controls and pesticide treatments. The two pesticides

Acetamiprid (ACE) and Imidacloprid (IMI) were tested at two different concentrations in the feed (IMI: 2.5 and 10 µg/L; ACE: 10 and 40 µg/L). The marked bees were collected from the hive after 15 days and in addition, unmarked bees from hive B were also collected, precisely two groups, old bees and younger ones. A total of 70 bees samples were tested, and we included negative controls for isolation, reverse transcription (RT), and qPCR to ensure contamination-free solutions and working space.

The results, depicted in Figure 28, yield few observations. Despite variations in pesticide dosages (ACE/IMI), there was no significant difference in viral loads compared to the no-pesticide control (Con). However, it's important to note that, in most instances, bees in the no-pesticide control group (from Hive A but not exposed to pesticides) exhibited slightly higher or the same mean viral loads as other manipulated bees.

Furthermore, a statistically slightly significant difference (p = 0.003**) was observed between bees from hive B (Bs: older bees, Bu: younger bees) and those manipulated from hive A (ACE/IMI/Con). This discrepancy suggests potential variations in viral loads influenced by hive-specific factors such as environmental conditions and age.

Of particular interest is the variability observed within hive B itself. Despite sharing a common hive environment, bees from different age groups (Bs: older bees, Bu: younger bees) displayed different viral loads. This highlights the nuanced dynamics of viral infections within bee colonies and underscores the potential influence of individual factors such as age and susceptibility.

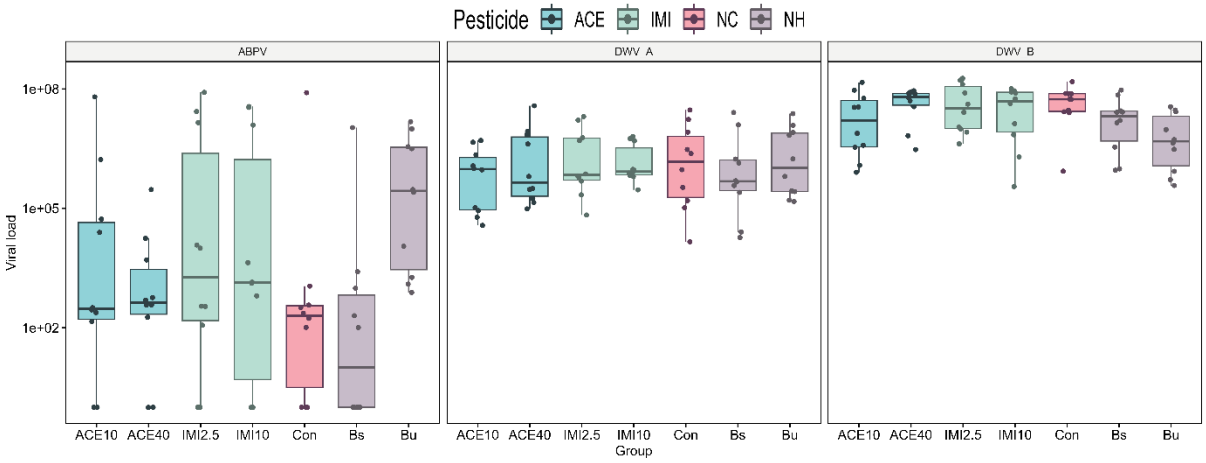


Figure 28- Barplot of viral loads per group of the experiment *(see text for explanation of abbreviation).*

In addition to the observations, individual bees exhibited varying proportions of present viruses, as depicted in Figure 29. The interconnectedness of individual bees suggests differences between

groups at this level, further emphasizing the intricate nature of viral dynamics within bee populations.

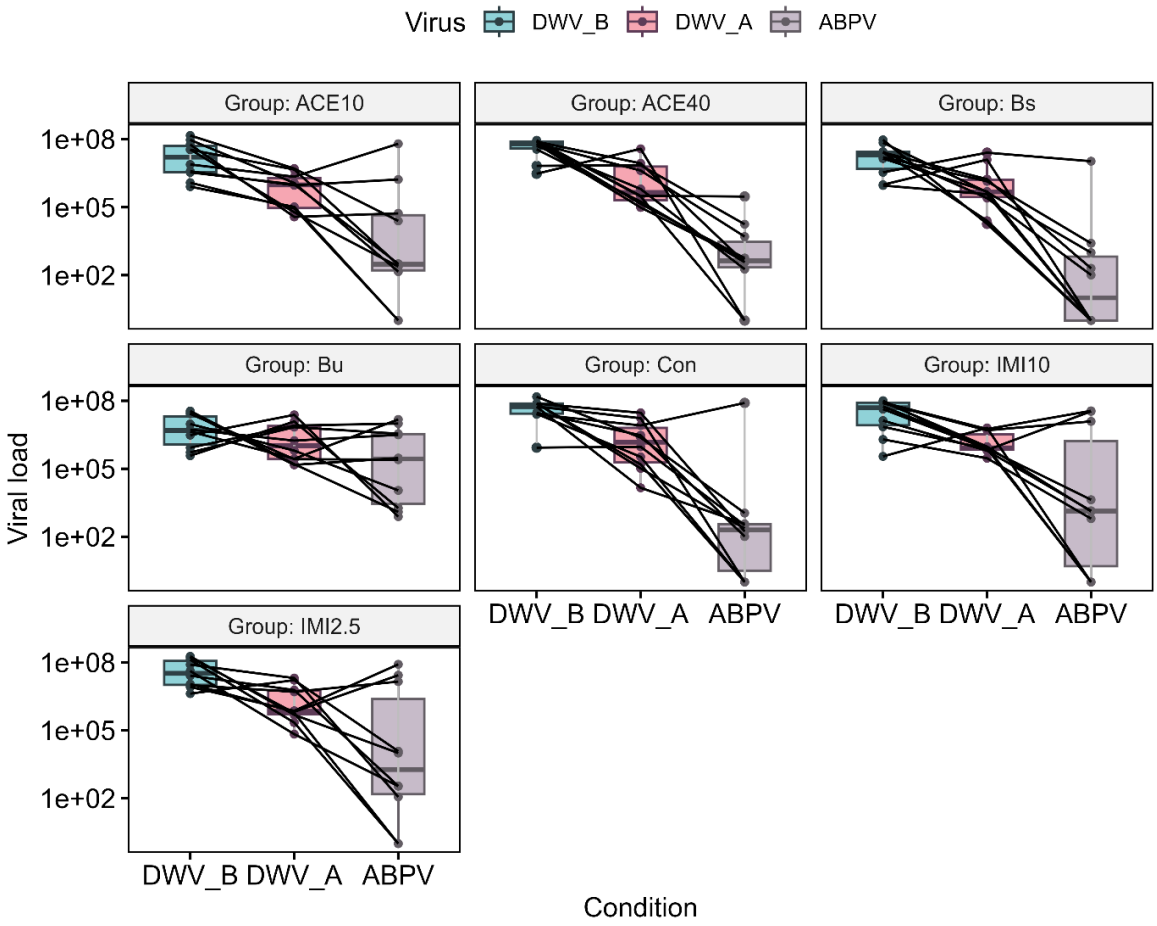


Figure 29- A paired graph showing relations of viral loads in different groups. *Each point represents bee values at different viruses and lines connect measurements of the same bee.*

5. Discussion

In the two projects, we focused on generally healthy bees with no overt signs of disease in the source colonies. In the first project, we sequenced 39 samples, each consisting of nine bees. We have taken triplicates from each hive (3x9 bees from one hive) from 13 locations. To our knowledge, this was the first study analysing virome replicates from one hive. It was interesting to find that when it came to honey bee-infecting viruses the replicates did not cluster. Each pool of nine bees had different bee-infecting viruses even though they came from the same hive. This could be due to by the unique set of bee-infecting viruses in each honey bee, even though they originate from the same environment. Apart from the unique bee virome in individuals, we saw in the first project skewed samples that consisted almost exclusively of one virus. In a few samples, DWV overshadowed others, including non-viral sequences and constitute ~90% of all reads. We hypothesized that very few individuals in the hive have very high viral loads and this may affect the resulting sample composition. These results led to a conclusion of the importance of analyzing a representative number of honey bees per hive in order to be able to see the whole virome, including the low-prevalent viruses. We expected that this approach would balance the high viral loads of a single virus in some individuals (158).

The increased size of the pools in the second project proved successful and we were able to detect several new viruses, including large DNA viruses (AmNV, AmFLV), small DNA viruses (Bee densovirus 1-9) (175), and four RNA viruses (Bee Picornavirus 1-4) two of which are discussed later. Additionally, many fragments of other viruses were detected but we were not able to complete their genomes.

However, we still saw some samples consisting of one virus (~90% of all reads). While our previous speculations about low-abundant honey bees with high viral loads (bees that are not present very frequently but have exceptionally high viral loads) still seems to be the only explanation, the presence of the skewed samples in the fifty bees pool might be explained by: either the individual bees having high viral loads are more prevalent or the viral loads are higher than we expected. In either case, the chances for skewed samples were lower when more bees were pooled together.

In the first project, we saw the skewed samples caused by outliner honey bees only in favour of DWV, we also saw this same phenomenon in the second project but there was one location that was prone to the same skewed samples but in this case in favour of LSV. LSV is a relatively recently described honey bee-infecting virus (176). Even though there is a chance that the virus

was described before as one of the similar viruses under the name: Bee virus X (described in 1974) or Bee virus Y (described in 1980) (177, 178). It was suggested that they are the same viruses based on several characteristics (from seasonality to capsid protein size) in two studies (179, 180). If that is the case then we know from the study where the was virus described, that it seems to have very little effect on the honey bees, except for slightly shortening the honey bees lives (177). A number of variants of the LSV were described, we know that several variants of this virus can infect honey bees at the same time, and viral loads of individual variants can be very variable and even very high, even so high to skew the samples in favour of one present variant.

Besides solving the methodological aspects, we aimed to explore the honey bee virome in Czechia, and its changes with hive, location and time. In the first project, we focused on the influence of hive and location i.e. stability across triplicates and variability across locations. In the second project, we mainly focused on changes in time and location. Furthermore, we analyzed the influence of pesticide presence. Later we explored the possibility of including other metadata concerning the climatic and environmental factors but gathering information retrospectively was not always possible.

In the results of the second project were several noteworthy observations:

Compared to the first project, where we observed no similarity in honey bee-infecting viruses in viromes based on the location, in the second project, the similarity of honey bee viral composition was identified based on the location, but the clustering was not very strong. This might be attributed to the methodical difference between the projects. In the second project, as already discussed above, 50 instead of 9 bees were pooled to one sample. Additionally, mapping of the samples was not done on the reference sequences of the known viruses instead they were mapped on all gained vMAGs. While mapping on known viruses allowed us a very sensitive detection of already known viruses, the *de-novo* approach was less sensitive but resulted in the detection of new viruses. Therefore, the number of viruses detected as honey-bee infecting viruses in our second project was greater than in the first one.

In our study, we did not identify the core virome defined as a set of viruses that is present across populations and time, a set of viruses that co-evolved with the host and might have a large impact on it. While being a well-defined concept, the core microbiome is hard to define precisely since the selection criteria differ. They can be based on abundance, prevalence, or both, and then studies differ in setting the cut-offs (181). The most likely candidates of the core virome of honey bees are DWV-B and DWV recombinants which are present in the majority of samples in and often in high

abundance. Another candidate might be AmFV or LSV with fluctuating abundances and variant composition. However, none of them complies with the condition of the definition of core microbiota we set for our project (>0.1% relative viral abundance over >80% of the samples; see Figure 30). Since honey bees have a very stable core bacterial microbiome, it is intriguing that they have such a variable virome. It is noteworthy that the core virome was detected (182) in mosquitoes inhabiting Guadeloupe island, but when the same research group examined viruses of mosquitoes in Europe the core virome was not present (183). The isolation might be a cause of the existence of core virome.

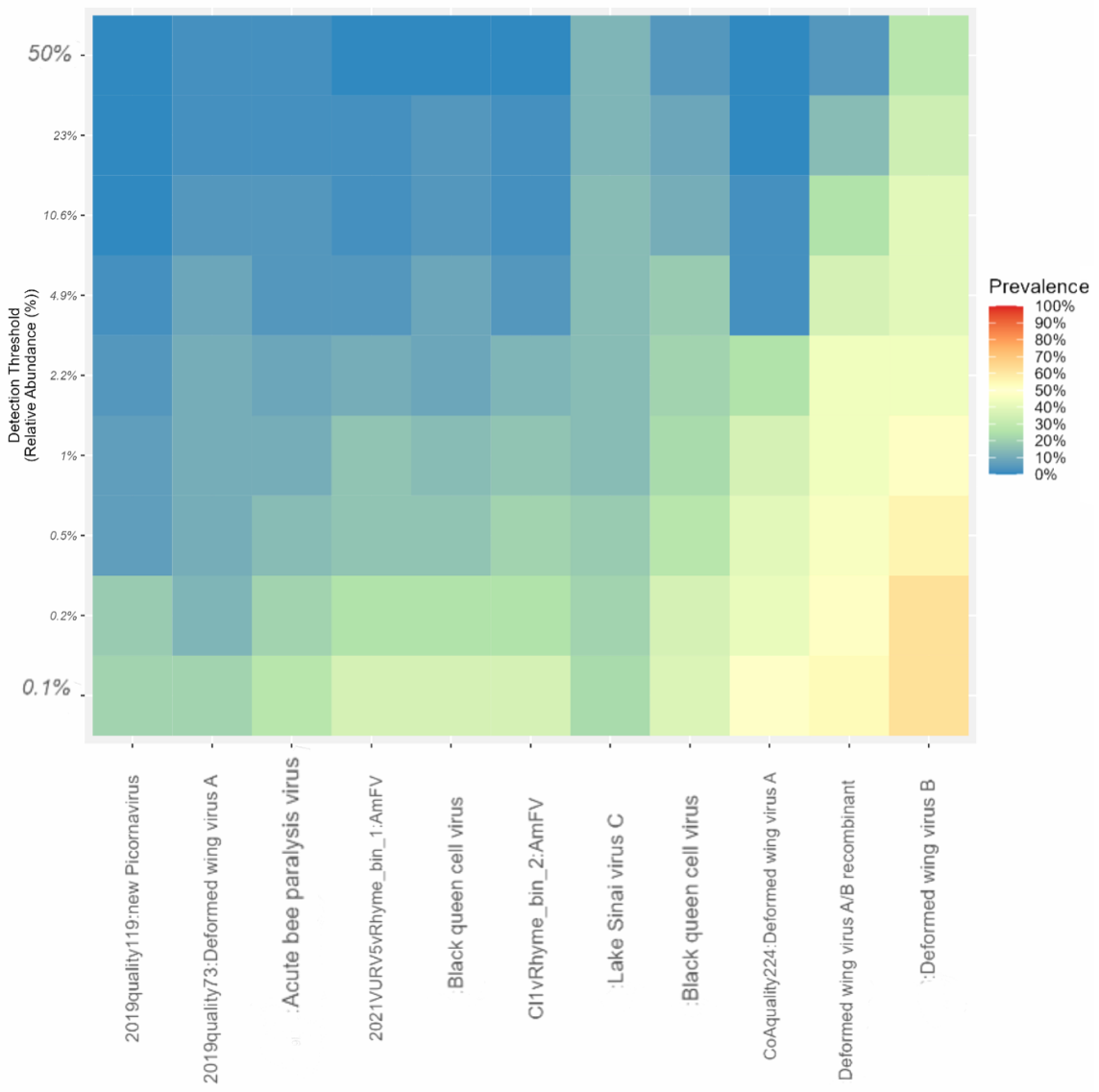


Figure 30- The heatmap with Relative Abundance and prevalence of 11 viruses. *On the x-axis are the viruses while on the y-axis is a threshold at which we count viruses as present, color shows the prevalence at each relative abundance for each virus.*

Regarding the changes of virome in time we have detected an increase in viral diversity of honey bee-infecting viruses over the three years. The relative abundance of bee-infecting viruses changed

from year to year, the changes were visible mainly for the 'high-abundant' honey bee-infecting viruses (DWV, ABPV, CBPV, BQCV, Sacbrood virus, and some variants of LSV). Furthermore, the virome composition changes not only with time but also with location. Therefore, we were not able to define the core virome of honey bees. We suggest that some viruses move through the honey bee populations, changing the virome composition in a time and location-dependent manner. Also, additional climatic and environmental factors (pathogens, pesticides, weather, etc.) will play a role. We hypothesize that it will not be possible to define it for honey bees in the future due to the abovementioned factors.

Apart from the 'high abundant' viruses infecting honey bees, there were more 'low-abundant' viruses (not often causing severe and well-defined overt symptoms). These viruses have been mostly recently discovered and are less well studied. Even though it seems that these viruses do not have a severe impact on honey bee health, we have shown that they are the ones more influenced by pesticides. Based on our observation we suggest that lowering viral diversity through the decrease of the prevalence of "low-abundant" viruses (or increase in "high-abundant" viruses and their competition with "low-abundant" viruses) can pesticides manifest their effect. It is also important to note that the division to 'high-abundant' and 'low-abundant' viruses we used is conditional: It is based on our current results and since the prevalences and abundances of honey bee viruses can be influenced by many factors, the cut-off for the distinction of these two categories of viruses can change accordingly.

Since the relationships between viruses in the virome of honey bees are so complex we want to explore the virome composition and viral interactions with the tool NetCoMi (184), which allows us to create and compare networks. This analysis is in progress.

5.1. Australian viruses Varroa naïve colonies vs under Varroa pressure

In the first study, when mapped samples on the set of known honey bee-infecting viruses, we noted was the absence of viruses identified previously in Australia. The Australian study reported a large number of RNA viruses such as Darwin bee virus 1-8, Bundaberg bee virus 1-8, Perth bee virus 1-9, Robinvale bee virus 1-9, Renmark bee virus 1-5, Hobart bee virus, Victoria bee virus 1-2, all of which belong to *Picornavirales,* mainly *Iflaviridae* and *Dicistroviridae* family (185). We then mapped data downloaded from publicly accessible databases to search for these viruses in studies from other parts of the world where *Varroa* is present. But again, we didn´t find any. Proving the absence is impossible and new data are being constantly generated, which means that some Australian or closely related viruses may be discovered in the *Varroa*-infected bees in the future.

In our second project, we found distant relatives of the Australian viruses, with the closest alignments to Bundaberg bee virus 4 and Darwin bee virus 6. That raises interesting questions about the evolution of honey bee virome under the other pathogens pressure.

It might be possible that the honey bee used to have these diverse RNA viruses. In Australia, the virome might stay diverse or even more diversified. In the rest of the world, it seems that the diversity of RNA viruses is decreased, which might be caused by the presence of *Varroa* that shifts the viral populations towards DWV and other *Varroa*-associated viruses. This could have resulted in the viral landscape we see now in countries outside of Australia.

However, any hypotheses about the evolution of honey bee virome are difficult since it is impossible to determine the original state of the virome (for example, in Europe before *Varroa* introduction). However, the introduction of *Varroa* to Australia now represents a unique opportunity to study changes of the virome in response to *Varroa* introduction. Another way that could elucidate these relationships is virome analysis of different insects and pollinators. As of now some of these data are available but not yet for large-scale meta-analysis (186)

5.2. Bioinformatics

We employed several bioinformatics tools to analyze the metagenomic data. We also tried tools for predicting viral sequences and viral binning for the first time on honey bee viruses. Through this analysis, we identified two large DNA viruses (AmNV, AmFLV). While viral binning allows us to detect and puzzle together large genomes it is a time-consuming approach, which proved to be beneficial, but given these facts:

Most honey bee-infecting viruses were complete from general assembly (individual sample assembly) since they are generally high-abundant. Those are the viruses that are generally of interest because of the linked pathology. While binning works well on some viruses, even the authors of vRhyme note that this approach for bacteriophages might be a bit more complicated. This means that ideally, bacteriophages should be predicted with a tool like VirSorter2 (143), and further analysed separately.

For identification of the new large viruses, AmNV and AmFLV, performing bining on only those samples where the viruses were present, will be a much more time-effective approach.

The basic analysis (QC, trimming, assembly and finding best alignment) can be done relatively fast, requires one database, and for most samples, it does not require high-end computers/access to the cluster. On the other hand, the advanced (predicting viral contigs, binning, co-assembly)

analyses can be more demanding on a high-end computing technology. The co-assembly alone with SPAdes is RAM greedy, depending on many factors but can grow over 200GB and higher. There exist alternative tools like megahit which is in contrast to SPAdes memory-efficient but sometimes can result in chimeric sequences, which can co-assembly exuberate. But in our experience co-assembly and read filtering proved to be advantageous since we completed two large genomes only with Illumina. So, while the co-assembly is very resource-demanding it reveals and allows to gain additional sequences.

There is a danger of 'bad' databases. Since most of the tool's databases come from community repositories (e.g., NCBI), where the number of sequences makes it impossible for perfect QC of each sequence. Even the databases that are carefully curated might contain chimeric regions, host regions and might not be precisely classified and even assembled. While blasting against the whole nr database (part of 'basic' analysis) these errors mostly disappear in the sheer amount and can be seen easily (it's possible to look at other best hits in the database and compare what is the sequence most similar to). Problems might appear with machine learning, where these programs learn on databases, often originating from community depositories and even through curated it might still contain mistakes. There is a certain degree of error that is acceptable, but it is not possible to avoid the fact that the more we do with the data the more errors we insert.

We suggest that the analysis should be very flexible and accommodate the type of samples, the research question and the resources available. The part we here call 'basic' to our experience works very well since it includes QC and thorough trimming. We would also suggest running geNomad beside diamond, the results of these two should be complementary and combining the data from these two sources could be beneficial. Other tools can be incorporated, one of them that proved very beneficial is CheckV which helps to predict contamination and completeness of the viral contigs. This is very handy for selecting relevant contigs from very incomplete fragments.

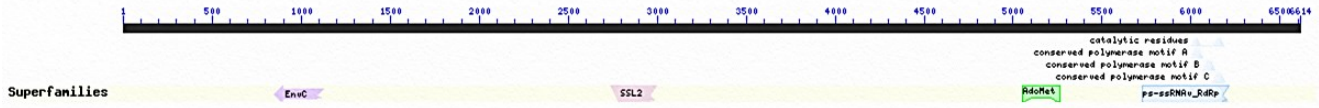
Therefore, our suggestion for binning use is targeted analysis rather than making it a part of the whole pipeline. When large viruses are detected and we want to pierce them together, binning is a beneficial approach. Additionally, it can be improved by combination with wet laboratory methods. But generally, RNA viruses and small/medium size DNA viruses can be completed without binning.

5.3. Other results- proteomics and qPCR

By proteomics, we identified several interesting pathway changes discussed in the article but during the data analysis, we came across a puzzling virus, AFV. We identified using NetoVir

protocol and sequencing almost all the viruses that were identified by proteomics even though some were present only in a very small amounts and low certainty (e.g., 50 reads mapping to the virus). The only large difference was AFV: Apis flavivirus. This virus is highly unusual. It belongs to the large *Flaviviridae* with a genome twice as long as members of the *Flaviviridae* family, but the virus has very few alignments and identities (best hit to Carrot flavi-like virus 1; 40.8% identity over 26% of the sequence, for preliminary screen of annotations see Figure 31) to existing sequences. We could not determine better sequence classification and identification than the current one. It would be interesting to further analyze this sequence. Mainly the functional annotations; and do this for all of the large Flaviviruses/ Flavi-like viruses. The larger genome means more coding potential. Therefore, finding the possible function of the additional sequence so much larger than in other Flaviviruses will be very interesting.

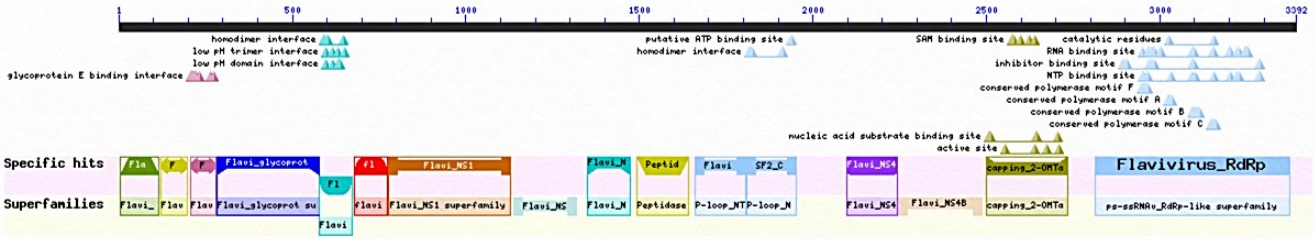
Apis flavivirus



List of domain hits

Name	Accession	Description	Interval	E-value
[+] ps-ssRNAv_RdRp-like super family	cl40470	conserved catalytic core domain of RNA-dependent RNA polymerase (RdRp) from the positive-sense ...	5728-6211	3.12e-10
[+] AdoMet_MTases super family	cl17173	S-adenosylmethionine-dependent methyltransferases (SAM or AdoMet-MTase), class I; ...	5052-5263	5.14e-08
[+] EnvC super family	cl34844	Septal ring factor EnvC, activator of murein hydrolases AmiA and AmiB [Cell cycle control, ...	847-1124	1.30e-06
[+] SSL2 super family	cl34083	Superfamily II DNA or RNA helicase [Transcription, Replication, recombination, and repair];	2734-2994	1.45e-05

Dengue virus 1



List of domain hits

Name	Accession	Description	Interval	E-value
[+] Flavivirus_RdRp	cd23204	catalytic core domain of RNA-dependent RNA polymerase (RdRp) in the genus Flavivirus, within ...	2812-3375	0e+00
[+] Flavi_NS1	pfam00948	Flavivirus non-structural Protein NS1; The NS1 protein is well conserved amongst the ...	777-1130	0e+00
[+] Flavi_glycoprot	pfam00869	Flavivirus glycoprotein, central and dimerization domains;	282-576	0e+00
[+] Flavi_NS4B super family	cl03175	Flavivirus non-structural protein NS4B; Flaviviruses encode a single polyprotein. This is ...	2245-2486	3.68e-135
[+] capping_2-OMTase_Flaviviridae	cd20761	Cap-0 specific (nucleoside-2'-O-)-methyltransferase of flaviviridae; Cap-0 specific ...	2501-2734	1.01e-129
[+] Flavi_DEAD	pfam07652	Flavivirus DEAD domain;	1661-1806	1.42e-95
[+] Flavi_NS2A super family	cl03066	Flavivirus non-structural protein NS2A; NS2A is a hydrophobic protein about 25 kDa is size. ...	1138-1320	7.58e-92
[+] Flavi_NS4A	pfam01350	Flavivirus non-structural protein NS4A; Flaviviruses encode a single polyprotein. This is ...	2099-2242	1.12e-77
[+] SF2_C_viral	cd18806	C-terminal helicase domain of viral helicase; Viral helicases in this family here are ...	1808-1952	1.01e-74
[+] Flavi_NS2B	pfam01002	Flavivirus non-structural protein NS2B; Flaviviruses encode a single polyprotein. This is ...	1349-1475	3.45e-70
[+] Flavi_glycop_C	pfam02832	Flavivirus glycoprotein, immunoglobulin-like domain;	578-673	4.40e-56
[+] Peptidase_S7	pfam00949	Peptidase S7, Flavivirus NS3 serine protease; The viral genome is a positive strand RNA that ...	1493-1642	8.94e-50
[+] flavi_E_stem	TIGR04240	flavivirus envelope glycoprotein E, stem/anchor domain; This model describes the C-terminal ...	679-775	2.56e-44
[+] Flavi_propep	pfam01570	Flavivirus polyprotein propeptide; The flaviviruses are small enveloped animal viruses ...	120-198	4.76e-43
[+] Flavi_capsid	pfam01003	Flavivirus capsid protein C; Flaviviruses are small enveloped viruses with virions comprised ...	5-114	1.15e-39
[+] Flavi_M	pfam01004	Flavivirus envelope glycoprotein M; Flaviviruses are small enveloped viruses with virions ...	207-280	1.11e-36

Figure 31- The annotations of two Flaviviridae, well-described Dengue virus 1 and Apis flavivirus. *Conserved domain search (CD-search at 17/6/24). The Apis Flavivirus has +ssRNA viral polymerase, helicase and different methyltransferase. We checked a few of the other large Flaviviruses and they look similar.*

In 2023 we tested bee workers with qPCR for the presence of three viruses to discern the effect of pesticides on honey bees' viral loads. It was conceived as a multifactorial experiment where *Varroa* and pesticides interact together on honey bees, and we explore if this interaction has an effect on the viral loads. We did not see much difference between the groups of pesticide-exposed bees and the control group (no pesticide but manipulated). Nevertheless, we have observed the difference between the manipulated honey bees (pesticide exposure and control) and honey bees from the hive in which they were kept.

The difference between this and many other pesticide experiments (done on caged bees) is the complexity of factors affecting the honey bees and the time of pesticide exposure. While they have been exposed to pesticides under controlled conditions in cages for three days since their emergence, when they are placed in a hive, the number of factors increases immensely under natural conditions.

In the experiment, we tested just three viruses associated with *Varroa* which are usually present in a high abundance. The high abundance might be a limitation because we might not be able to detect small changes in highly abundant viruses. Therefore, more complex analyses which would also include low-abundant viruses could shed light on this question.

It is possible that the relatively short exposure of 72 hours to the very low concentrations of pesticides (that are field realistic) and the return to natural conditions are not enough to impact honey bee viral loads long-term. It is possible that the bees were affected by the pesticides but recovered under natural conditions before being tested. This is interesting because it could mean that short-term exposure to sublethal realistic doses of pesticides may not be as lethal as thought and that the honey bees may be able to recover from exposure to neonicotinoid pesticides without severe lasting effects.

6. Conclusions, summary and future work

We investigated the virome of honey bees (Apis mellifera) through metagenomic analysis, revealing significant variability in viral loads across different colonies and seasons. We identified several novel DNA viruses, including Apis mellifera filamentous-like virus (AmFLV) and Apis mellifera nudivirus (AmNV), contributing to the understanding of the complex virome within honey bee colonies. Additionally, our analysis highlighted the need for representative sampling to accurately assess virome diversity. We discovered viruses related to those identified in Australia, which were previously undetected in the rest of the world, albeit with low identity over limited sequence lengths. This finding raises intriguing questions about the evolution of honey bee viruses. Furthermore, while the binning of viral sequences proved extremely useful for large viruses, it should be balanced with the specific needs of the analysis and the project. Combining necessary steps and using relevant tools as needed can optimize resources and time, maintaining the significant advantages these tools offer.

In the future, we would like to continue with:

1) The sequencing of the whole virome is no longer possible from the bees used for qPCR. But pooling the isolated RNA from the ten bees per group is. We would like to sequence RNA viromes in these pooled samples. That could reveal which viruses are present, what precisely we should test for and if there is a large difference in composition of virome between the groups which we could verify by PCR/qPCR.
2) Compare the hives that collapsed before the next sampling and hives that survived in their full viral complexity by looking at the present viruses as interconnected networks.
3) Compare the advantages of the software we here designated as ‘advanced’ and comparing the results with the ‘basic’ pipeline. We would like to see how/if the results change or the trends stay the same.
4) Explore the long *Flaviviridae* and attempt to find what is gained by the large elongation of genomes.

7. References

1. Hung K-LJ, Kingston JM, Albrecht M, Holway DA, Kohn JR. 2018. The worldwide importance of honey bees as pollinators in natural habitats. Proceedings of the Royal Society B: Biological Sciences 285:20172140.

2. Klein A-M, Vaissière BE, Cane JH, Steffan-Dewenter I, Cunningham SA, Kremen C, Tscharntke T. 2007. Importance of pollinators in changing landscapes for world crops. Proceedings of the Royal Society B: Biological Sciences 274:303–313.

3. Papa G, Maier R, Durazzo A, Lucarini M, Karabagias IK, Plutino M, Bianchetto E, Aromolo R, Pignatti G, Ambrogio A, Pellecchia M, Negri I. 2022. The Honey Bee Apis mellifera: An Insect at the Interface between Human and Ecosystem Health. 2. Biology 11:233.

4. Breeze TD, Vaissière BE, Bommarco R, Petanidou T, Seraphides N, Kozák L, Scheper J, Biesmeijer JC, Kleijn D, Gyldenkærne S, Moretti M, Holzschuh A, Steffan-Dewenter I, Stout JC, Pärtel M, Zobel M, Potts SG. 2014. Agricultural Policies Exacerbate Honeybee Pollination Service Supply-Demand Mismatches Across Europe. PLoS One 9:e82996.

5. Liang G, Bushman FD. 2021. The human virome: assembly, composition and host interactions. 8. Nat Rev Microbiol 19:514–527.

6. Virgin HW. 2014. The virome in mammalian physiology and disease. Cell 157:142–150.

7. Bai G-H, Lin S-C, Hsu Y-H, Chen S-Y. 2022. The Human Virome: Viral Metagenomics, Relations with Human Diseases, and Therapeutic Applications. Viruses 14:278.

8. Insolia L, Molinari R, Rogers SR, Williams GR, Chiaromonte F, Calovi M. 2022. Honey bee colony loss linked to parasites, pesticides and extreme weather across the United States. 1. Sci Rep 12:20787.

9. Hristov P, Shumkova R, Palova N, Neov B. 2020. Factors Associated with Honey Bee Colony Losses: A Mini-Review. Vet Sci 7:166.

10. Breda D, Frizzera D, Giordano G, Seffin E, Zanni V, Annoscia D, Topping CJ, Blanchini F, Nazzi F. 2022. A deeper understanding of system interactions can explain contradictory field results on pesticide impact on honey bees. Nat Commun 13:5720.

11. Blacquière T, van der Steen JJ. 2017. Three years of banning neonicotinoid insecticides based on sub-lethal effects: can we expect to see effects on bees? Pest Manag Sci 73:1299–1304.

12. Klingelhöfer D, Braun M, Brüggmann D, Groneberg DA. 2022. Neonicotinoids: A critical assessment of the global research landscape of the most extensively used insecticide. Environmental Research 213:113727.

13. Mamy L, Pesce S, Sanchez W, Aviron S, Bedos C, Berny P, Bertrand C, Betoulle S, Charles S, Chaumot A, Coeurdassier M, Coutellec M-A, Crouzet O, Faburé J, Fritsch C, Gonzalez P, Hedde M, Leboulanger C, Margoum C, Mougin C, Munaron D, Nélieu S, Pelosi C, Rault M, Sucré E, Thomas M, Tournebize J, Leenhardt S. 2023. Impacts of neonicotinoids on biodiversity: a critical review. Environ Sci Pollut Res Int https://doi.org/10.1007/s11356-023-31032-3.

14. Le Conte Y, Navajas M. 2008. Climate change: impact on honey bee populations and diseases. Rev Sci Tech 27:485–497, 499–510.

15. Laurino D, Lioy S, Carisio L, Manino A, Porporato M. 2020. Vespa velutina: An Alien Driver of Honey Bee Colony Losses. 1. Diversity 12:5.

16. Moritz RFA, Haddad N, Bataieneh A, Shalmon B, Hefetz A. 2010. Invasion of the dwarf honeybee Apis florea into the near East. Biol Invasions 12:1093–1099.

17. Brittain C, Williams N, Kremen C, Klein A-M. 2013. Synergistic effects of non-Apis bees and honey bees for pollination services. Proc Biol Sci 280:20122767.

18. Kenis M, Auger-Rozenberg M-A, Roques A, Timms L, Péré C, Cock MJW, Settele J, Augustin S, Lopez-Vaamonde C. 2009. Ecological effects of invasive alien insects. Biol Invasions 11:21–45.

19. Han P, Niu C-Y, Lei C-L, Cui J-J, Desneux N. 2010. Quantification of toxins in a Cry1Ac + CpTI cotton cultivar and its potential effects on the honey bee Apis mellifera L. Ecotoxicology 19:1452–1459.

20. Durant JL, Otto CRV. 2019. Feeling the sting? Addressing land-use changes can mitigate bee declines. Land Use Policy 87:104005.

21. Andrews E. 2019. To save the bees or not to save the bees: honey bee health in the Anthropocene. Agric Hum Values 36:891–902.

22. Dolezal AG, Toth AL. 2018. Feedbacks between nutrition and disease in honey bee health. Curr Opin Insect Sci 26:114–119.

23. Melicher D, Wilson ES, Bowsher JH, Peterson SS, Yocum GD, Rinehart JP. 2019. Long-Distance Transportation Causes Temperature Stress in the Honey Bee, Apis mellifera (Hymenoptera: Apidae). Environ Entomol 48:691–701.

24. Goulson D, Nicholls E, Botías C, Rotheray EL. 2015. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. Science 347:1255957.

25. Aufauvre J, Biron DG, Vidau C, Fontbonne R, Roudel M, Diogon M, Viguès B, Belzunces LP, Delbac F, Blot N. 2012. Parasite-insecticide interactions: a case study of Nosema ceranae and fipronil synergy on honeybee. Sci Rep 2:326.

26. Retschnig G, Neumann P, Williams GR. 2014. Thiacloprid-Nosema ceranae interactions in honey bees: host survivorship but not parasite reproduction is dependent on pesticide dose. J Invertebr Pathol 118:18–19.

27. Vanbergen AJ, Initiative the IP. 2013. Threats to an ecosystem service: pressures on pollinators. Frontiers in Ecology and the Environment 11:251–259.

28. Zhao H, Li G, Cui X, Wang H, Liu Z, Yang Y, Xu B. 2022. Review on effects of some insecticides on honey bee health. Pesticide Biochemistry and Physiology 188:105219.

29. Abay Z, Bezabeh A, Gela A, Tassew A. 2023. Evaluating the Impact of Commonly Used Pesticides on Honeybees (Apis mellifera) in North Gonder of Amhara Region, Ethiopia. J Toxicol 2023:2634158.

30. Shannon B, Jeon H, Johnson RM. 2023. Review: the risks of spray adjuvants to honey bees. J Insect Sci 23:20.

31. Raimets R, Karise R, Mänd M, Kaart T, Ponting S, Song J, Cresswell JE. 2018. Synergistic interactions between a variety of insecticides and an ergosterol biosynthesis inhibitor fungicide in dietary exposures of bumble bees (Bombus terrestris L.). Pest Management Science 74:541–546.

32. Bernauer OM, Gaines-Day HR, Steffan SA. 2015. Colonies of Bumble Bees (Bombus impatiens) Produce Fewer Workers, Less Bee Biomass, and Have Smaller Mother Queens Following Fungicide Exposure. 2. Insects 6:478–488.

33. DeGrandi-Hoffman G, Chen Y, Simonds R. 2013. The Effects of Pesticides on Queen Rearing and Virus Titers in Honey Bees (Apis mellifera L.). 1. Insects 4:71–89.

34. Faita MR, Oliveira E de M, Alves VV, Orth AI, Nodari RO. 2018. Changes in hypopharyngeal glands of nurse bees (Apis mellifera) induced by pollen-containing sublethal doses of the herbicide Roundup®. Chemosphere 211:566–572.

35. Abraham J, Benhotons GS, Krampah I, Tagba J, Amissah C, Abraham JD. 2018. Commercially formulated glyphosate can kill non-target pollinator bees under laboratory conditions. Entomologia Experimentalis et Applicata 166:695–702.

36. Glavan G, Novak S, Božič J, Jemec Kokalj A. 2020. Comparison of sublethal effects of natural acaricides carvacrol and thymol on honeybees. Pesticide Biochemistry and Physiology 166:104567.

37. Boncristiani H, Underwood R, Schwarz R, Evans JD, Pettis J, vanEngelsdorp D. 2012. Direct effect of acaricides on pathogen loads and gene expression levels in honey bees Apis mellifera. Journal of Insect Physiology 58:613–620.

38. Dulin F, Zatylny-Gaudin C, Ballandonne C, Guillet B, Bonafos R, Bureau R, Halm MP. 2014. Protecting honey bees: identification of a new varroacide by in silico, in vitro, and in vivo studies. Parasitol Res 113:4601–4610.

39. Johnson RM, Dahlgren L, Siegfried BD, Ellis MD. 2013. Acaricide, Fungicide and Drug Interactions in Honey Bees (Apis mellifera). PLOS ONE 8:e54092.

40. Zhu W, Schmehl DR, Mullin CA, Frazier JL. 2014. Four Common Pesticides, Their Mixtures and a Formulation Solvent in the Hive Environment Have High Oral Toxicity to Honey Bee Larvae. PLOS ONE 9:e77547.

41. Le Conte Y, Ellis M, Ritter W. 2010. *Varroa* mites and honey bee health: can *Varroa* explain part of the colony losses? Apidologie 41:353–363.

42. Traynor KS, Mondet F, de Miranda JR, Techer M, Kowallik V, Oddie MAY, Chantawannakul P, McAfee A. 2020. Varroa destructor: A Complex Parasite, Crippling Honey Bees Worldwide. Trends in Parasitology 36:592–606.

43. Warner S, Pokhrel LR, Akula SM, Ubah CS, Richards SL, Jensen H, Kearney GD. 2024. A scoping review on the effects of Varroa mite (Varroa destructor) on global honey bee decline. Sci Total Environ 906:167492.

44. Shen M, Cui L, Ostiguy N, Cox-Foster D. 2005. Intricate transmission routes and interactions between picorna-like viruses (Kashmir bee virus and sacbrood virus) with the honeybee host and the parasitic varroa mite. Journal of General Virology, 86:2281–2289.

45. Dainat B, Evans JD, Chen YP, Gauthier L, Neumann P. 2012. Dead or Alive: Deformed Wing Virus and Varroa destructor Reduce the Life Span of Winter Honeybees. Applied and Environmental Microbiology 78:981–987.

46. Nazzi F, Brown SP, Annoscia D, Piccolo FD, Prisco GD, Varricchio P, Vedova GD, Cattonaro F, Caprio E, Pennacchio F. 2012. Synergistic Parasite-Pathogen Interactions Mediated by Host Immunity Can Drive the Collapse of Honeybee Colonies. PLOS Pathogens 8:e1002735.

47. Yang X, Cox-Foster DL. 2005. Impact of an ectoparasite on the immunity and pathology of an invertebrate: Evidence for host immunosuppression and viral amplification. Proc Natl Acad Sci U S A 102:7470–7475.

48. Annoscia D, Brown SP, Di Prisco G, De Paoli E, Del Fabbro S, Frizzera D, Zanni V, Galbraith DA, Caprio E, Grozinger CM, Pennacchio F, Nazzi F. 2019. Haemolymph removal by Varroa mite destabilizes the dynamical interaction between immune effectors and virus in bees, as predicted by Volterra's model. Proceedings of the Royal Society B: Biological Sciences 286:20190331.

49. Zhang Y, Han R. 2018. A Saliva Protein of Varroa Mites Contributes to the Toxicity toward Apis cerana and the DWV Elevation in A. mellifera. Sci Rep 8:3387.

50. Rosenkranz P, Aumeier P, Ziegelmann B. 2010. Biology and control of Varroa destructor. J Invertebr Pathol 103 Suppl 1:S96-119.

51. Nazzi F, Le Conte Y. 2016. Ecology of Varroa destructor , the Major Ectoparasite of the Western Honey Bee, Apis mellifera. Annu Rev Entomol 61:417–432.

52. Oldroyd BP. 1999. Coevolution while you wait: Varroa jacobsoni, a new parasite of western honeybees. Trends in Ecology & Evolution 14:312–315.

53. Matheson A. 1995. First documented findings of Varroa jacobsoni outside its presumed natural range. Apiacta 30:1–8.

54. Krejčí AB, Votýpková K, Lukeš J, Votýpka J. 2023. Varroa destructor. Trends in Parasitology 39:487–488.

55. Roberts JMK, Anderson DL, Durr PA. 2017. Absence of deformed wing virus and Varroa destructor in Australia provides unique perspectives on honeybee viral landscapes and colony losses. Sci Rep 7:6925.

56. Moran NA, Hansen AK, Powell JE, Sabree ZL. 2012. Distinctive Gut Microbiota of Honey Bees Assessed Using Deep Sampling from Individual Worker Bees. PLOS ONE 7:e36393.

57. Martinson VG, Danforth BN, Minckley RL, Rueppell O, Tingek S, Moran NA. 2011. A simple and distinctive microbiota associated with honey bees and bumble bees. Mol Ecol 20:619–628.

58. Sabree ZL, Hansen AK, Moran NA. 2012. Independent Studies Using Deep Sequencing Resolve the Same Set of Core Bacterial Species Dominating Gut Communities of Honey Bees. PLOS ONE 7:e41250.

59. Corby-Harris V, Maes P, Anderson KE. 2014. The Bacterial Communities Associated with Honey Bee (Apis mellifera) Foragers. PLOS ONE 9:e95056.

60. Engel P, Martinson VG, Moran NA. 2012. Functional diversity within the simple gut microbiota of the honey bee. Proc Natl Acad Sci U S A 109:11002–11007.

61. Copeland DC, Maes PW, Mott BM, Anderson KE. 2022. Changes in gut microbiota and metabolism associated with phenotypic plasticity in the honey bee Apis mellifera. Frontiers in Microbiology 13.

62. Jeyaprakash A, Hoy MA, Allsopp MH. 2003. Bacterial diversity in worker adults of *Apis mellifera capensis* and *Apis mellifera scutellata* (Insecta: Hymenoptera) assessed using 16S rRNA sequences. Journal of Invertebrate Pathology 84:96–103.

63. Kwong WK, Moran NA. 2016. Gut microbial communities of social bees. Nat Rev Microbiol 14:374–384.

64. Gerth M, Saeed A, White JA, Bleidorn C. 2015. Extensive screen for bacterial endosymbionts reveals taxon-specific distribution patterns among bees (Hymenoptera, Anthophila). FEMS Microbiology Ecology 91:fiv047.

65. GERTH M, GEIßLER A, BLEIDORN C. 2011. Wolbachia infections in bees (Anthophila) and possible implications for DNA barcoding. Systematics and Biodiversity 9:319–327.

66. McFrederick QS, Mueller UG, James RR. 2014. Interactions between fungi and bacteria influence microbial community structure in the Megachile rotundata larval gut. Proceedings of the Royal Society B: Biological Sciences 281:20132653.

67. Anjum SI, Aldakheel F, Shah AH, Khan S, Ullah A, Hussain R, Khan H, Ansari MJ, Mahmoud AH, Mohammed OB. 2021. Honey bee gut an unexpected niche of human pathogen. Journal of King Saud University - Science 33:101247.

68. Matović K, Žarković A, Debeljak Z, Vidanović D, Vasković N, Tešović B, Ćirić J. 2023. American Foulbrood—Old and Always New Challenge. Vet Sci 10:180.

69. Forsgren E. 2010. European foulbrood in honey bees. Journal of Invertebrate Pathology 103:S5–S9.

70. Mejias E. 2019. American Foulbrood and the Risk in the Use of Antibiotics as a Treatment.

71. Paxton RJ, Klee J, Korpela S, Fries I. 2007. Nosema ceranae has infected Apis mellifera in Europe since at least 1998 and may be more virulent than Nosema apis. Apidologie 38:558–565.

72. Klee J, Besana AM, Genersch E, Gisder S, Nanetti A, Tam DQ, Chinh TX, Puerta F, Ruz JM, Kryger P, Message D, Hatjina F, Korpela S, Fries I, Paxton RJ. 2007. Widespread dispersal of the microsporidian Nosema ceranae, an emergent pathogen of the western honey bee, Apis mellifera. J Invertebr Pathol 96:1–10.

73. Gisder S, Möckel N, Linde A, Genersch E. 2011. A cell culture model for Nosema ceranae and Nosema apis allows new insights into the life cycle of these important honey bee-pathogenic microsporidia. Environ Microbiol 13:404–413.

74. Higes M, García-Palencia P, Martín-Hernández R, Meana A. 2007. Experimental infection of Apis mellifera honeybees with Nosema ceranae (Microsporidia). J Invertebr Pathol 94:211–217.

75. Martín-Hernández R, Meana A, Prieto L, Salvador AM, Garrido-Bailón E, Higes M. 2007. Outcome of colonization of Apis mellifera by Nosema ceranae. Appl Environ Microbiol 73:6331–6338.

76. Higes M, Martín-Hernández R, Botías C, Bailón EG, González-Porto AV, Barrios L, del Nozal MJ, Bernal JL, Jiménez JJ, Palencia PG, Meana A. 2008. How natural infection by Nosema ceranae causes honeybee colony collapse. Environmental Microbiology 10:2659–2669.

77. Schwarz RS, Bauchan GR, Murphy CA, Ravoet J, De Graaf DC, Evans JD. 2015. Characterization of two species of trypanosomatidae from the Honey Bee Apis mellifera: Crithidia mellificae Langridge and McGhee, and Lotmaria passim n. gen., n. sp. Journal of Eukaryotic Microbiology 62:567–583.

78. Ravoet J, Maharramov J, Meeus I, Smet LD, Wenseleers T, Smagghe G, Graaf DC de. 2013. Comprehensive Bee Pathogen Screening in Belgium Reveals Crithidia mellificae as a New Contributory Factor to Winter Mortality. PLOS ONE 8:e72443.

79. Schwarz RS, Moran NA, Evans JD. 2016. Early gut colonizers shape parasite susceptibility and microbiota composition in honey bee workers. Proceedings of the National Academy of Sciences 113:9345–9350.

80. Stevanovic J, Schwarz RS, Vejnovic B, Evans JD, Irwin RE, Glavinic U, Stanimirovic Z. 2016. Species-specific diagnostics of Apis mellifera trypanosomatids: A nine-year survey (2007–2015) for trypanosomatids and microsporidians in Serbian honey bees. Journal of Invertebrate Pathology 139:6–11.

81. Buendía M, Martín-Hernández R, Ornosa C, Barrios L, Bartolomé C, Higes M. 2018. Epidemiological study of honeybee pathogens in Europe: The results of Castilla-La Mancha (Spain). 2. Spanish Journal of Agricultural Research 16:e0502–e0502.

82. Fleites-Ayil FA, Medina-Medina LA, Quezada Euán JJG, Stolle E, Theodorou P, Tragust S, Paxton RJ. 2023. Trouble in the tropics: Pathogen spillover is a threat for native stingless bees. Biological Conservation 284:110150.

83. Nanetti A, Bortolotti L, Cilia G. 2021. Pathogens Spillover from Honey Bees to Other Arthropods. Pathogens 10:1044.

84. Streicher T, Tehel A, Tragust S, Paxton RJ. 2023. Experimental viral spillover can harm Bombus terrestris workers under field conditions. Ecological Entomology 48:81–89.

85. Mordecai GJ, Wilfert L, Martin SJ, Jones IM, Schroeder DC. 2016. Diversity in a honey bee pathogen: first report of a third master variant of the Deformed Wing Virus quasispecies. 5. The ISME Journal 10:1264–1273.

86. de Miranda JR, Brettell LE, Chejanovsky N, Childers AK, Dalmon A, Deboutte W, de Graaf DC, Doublet V, Gebremedhn H, Genersch E, Gisder S, Granberg F, Haddad NJ, Kaden R, Manley R, Matthijnssens J, Meeus I, Migdadi H, Milbrath MO, Mondet F, Remnant EJ, Roberts JMK, Ryabov EV, Sela N, Smagghe G, Somanathan H, Wilfert L, Wright ON, Martin SJ, Ball BV. 2022. Cold case: The disappearance of Egypt bee virus, a fourth distinct master strain of deformed wing virus linked to honeybee mortality in 1970's Egypt. Virology Journal 19:12.

87. Gisder S, Möckel N, Eisenhardt D, Genersch E. 2018. In vivo evolution of viral virulence: switching of deformed wing virus between hosts results in virulence changes and sequence shifts. Environmental Microbiology 20:4612–4628.

88. Penn HJ, Simone-Finstrom M, Lang S, Chen J, Healy K. 2021. Host Genotype and Tissue Type Determine DWV Infection Intensity. Frontiers in Insect Science 1.

89. McMahon DP, Natsopoulou ME, Doublet V, Fürst M, Weging S, Brown MJF, Gogol-Döring A, Paxton RJ. 2016. Elevated virulence of an emerging viral genotype as a driver of honeybee loss. Proc Biol Sci 283.

90. Barroso-Arévalo S, Vicente-Rubiano M, Molero F, Puerta F, Sánchez-Vizcaíno JM. 2019. Nucleotide sequence variations may be associated with virulence of deformed wing virus. Apidologie 50:482–496.

91. Ryabov EV, Wood GR, Fannon JM, Moore JD, Bull JC, Chandler D, Mead A, Burroughs N, Evans DJ. 2014. A Virulent Strain of Deformed Wing Virus (DWV) of Honeybees (Apis mellifera) Prevails after Varroa destructor-Mediated, or In Vitro, Transmission. PLOS Pathogens 10:e1004230.

92. Ryabov EV, Childers AK, Lopez D, Grubbs K, Posada-Florez F, Weaver D, Girten W, vanEngelsdorp D, Chen Y, Evans JD. 2019. Dynamic evolution in the key honey bee pathogen deformed wing virus: Novel insights into virulence and competition using reverse genetics. PLoS Biol 17.

93. Natsopoulou ME, McMahon DP, Doublet V, Frey E, Rosenkranz P, Paxton RJ. 2017. The virulent, emerging genotype B of Deformed wing virus is closely linked to overwinter honeybee worker loss. 1. Sci Rep 7:5242.

94. Ray AM, Davis SL, Rasgon JL, Grozinger CM. 2021. Simulated vector transmission differentially influences dynamics of two viral variants of deformed wing virus in honey bees (Apis mellifera). J Gen Virol 102:001687.

95. de Miranda JR, Genersch E. 2010. Deformed wing virus. Journal of Invertebrate Pathology 103:S48–S61.

96. Paxton RJ, Schäfer MO, Nazzi F, Zanni V, Annoscia D, Marroni F, Bigot D, Laws-Quinn ER, Panziera D, Jenkins C, Shafiey H. 2022. Epidemiology of a major honey bee pathogen, deformed wing virus: potential worldwide replacement of genotype A by genotype B. International Journal for Parasitology: Parasites and Wildlife 18:157–171.

97. Hasegawa N, Techer MA, Adjlane N, al-Hissnawi MS, Antúnez K, Beaurepaire A, Christmon K, Delatte H, Dukku UH, Eliash N, El-Niweiri MAA, Esnault O, Evans JD, Haddad NJ, Locke B, Muñoz I, Noël G, Panziera D, Roberts JMK, De la Rúa P, Shebl MA, Stanimirovic Z, Rasmussen DA, Mikheyev AS. 2023. Evolutionarily diverse origins of deformed wing viruses in western honey bees. Proceedings of the National Academy of Sciences 120:e2301258120.

98. Dalmon A, Desbiez C, Coulon M, Thomasson M, Le Conte Y, Alaux C, Vallon J, Moury B. 2017. Evidence for positive selection and recombination hotspots in Deformed wing virus (DWV). Sci Rep 7.

99. Gisder S, Genersch E. 2021. Direct Evidence for Infection of Varroa destructor Mites with the Bee-Pathogenic Deformed Wing Virus Variant B, but Not Variant A, via Fluorescence In Situ Hybridization Analysis. Journal of Virology 95:10.1128/jvi.01786-20.

100. Posada-Florez F, Childers AK, Heerman MC, Egekwu NI, Cook SC, Chen Y, Evans JD, Ryabov EV. 2019. Deformed wing virus type A, a major honey bee pathogen, is vectored by the mite Varroa destructor in a non-propagative manner. 1. Sci Rep 9:12445.

101. Yue C, Genersch E. 2005. RT-PCR analysis of Deformed wing virus in honeybees (Apis mellifera) and mites (Varroa destructor). J Gen Virol 86:3419–3424.

102. Gisder S, Aumeier P, Genersch E. 2009. Deformed wing virus: replication and viral load in mites (Varroa destructor). J Gen Virol 90:463–467.

103. Damayo JE, McKee RC, Buchmann G, Norton AM, Ashe A, Remnant EJ. 2023. Virus replication in the honey bee parasite, Varroa destructor. bioRxiv https://doi.org/10.1101/2023.07.16.549232.

104. Mondet F, de Miranda JR, Kretzschmar A, Le Conte Y, Mercer AR. 2014. On the Front Line: Quantitative Virus Dynamics in Honeybee (Apis mellifera L.) Colonies along a New Expansion Front of the Parasite Varroa destructor. PLoS Pathog 10.

105. Martin SJ, Highfield AC, Brettell L, Villalobos EM, Budge GE, Powell M, Nikaido S, Schroeder DC. 2012. Global honey bee viral landscape altered by a parasitic mite. Science 336:1304–1306.

106. Bailey L, Gibbs AJ, Woods RD. 1963. Two viruses from adult honey bees (Apis mellifera Linnaeus). Virology 21:390–395.

107. Brødsgaard CJ, Ritter W, Hansen H, Brødsgaard HF. 2000. Interactions among Varroa jacobsoni mites, acute paralysis virus, and Paenibacillus larvae larvae and their influence on mortality of larval honeybees in vitro. Apidologie 31:543–554.

108. Evans JD, Spivak M. 2010. Socialized medicine: individual and communal disease barriers in honey bees. J Invertebr Pathol 103 Suppl 1:S62-72.

109. de Miranda JR, Cordoni G, Budge G. 2010. The Acute bee paralysis virus–Kashmir bee virus–Israeli acute paralysis virus complex. Journal of Invertebrate Pathology 103:S30–S47.

110. Carrillo-Tripp J, Dolezal AG, Goblirsch MJ, Miller WA, Toth AL, Bonning BC. 2016. In vivo and in vitro infection dynamics of honey bee viruses. Sci Rep 6.

111. Gauthier L, Cornman S, Hartmann U, Cousserans F, Evans JD, de Miranda JR, Neumann P. 2015. The Apis mellifera Filamentous Virus Genome. Viruses 7:3798–3815.

112. Clark TB. 1978. A filamentous virus of the honey bee. Journal of Invertebrate Pathology 32:332–340.

113. Hartmann U, Forsgren E, Charrière J-D, Neumann P, Gauthier L. 2015. Dynamics of Apis mellifera Filamentous Virus (AmFV) Infections in Honey Bees and Relationships with Other Parasites. Viruses 7:2654–2667.

114. McMenamin AJ, Flenniken ML. 2018. Recently identified bee viruses and their impact on bee pollinators. Current Opinion in Insect Science 26:120–129.

115. Remnant EJ, Shi M, Buchmann G, Blacquière T, Holmes EC, Beekman M, Ashe A. 2017. A Diverse Range of Novel RNA Viruses in Geographically Distinct Honey Bee Populations. Journal of Virology 91:10.1128/jvi.00158-17.

116. Levin S, Galbraith D, Sela N, Erez T, Grozinger CM, Chejanovsky N. 2017. Presence of Apis rhabdovirus-1 in populations of pollinators and their parasites from two continents. Frontiers in Microbiology 8.

117. Lester PJ, Felden A, Baty JW, Bulgarella M, Haywood J, Mortensen AN, Remnant EJ, Smeele ZE. 2022. Viral communities in the parasite Varroa destructor and in colonies of their honey bee host (Apis mellifera) in New Zealand. Sci Rep 12:8809.

118. Levin S, Sela N, Erez T, Nestel D, Pettis J, Neumann P, Chejanovsky N. 2019. New Viruses from the Ectoparasite Mite Varroa destructor Infesting Apis mellifera and Apis cerana. Viruses 11:94.

119. Roux S, Matthijnssens J, Dutilh BE. 2021. Metagenomics in Virology. Encyclopedia of Virology 133–140.

120. Child HT, Airey G, Maloney DM, Parker A, Wild J, McGinley S, Evens N, Porter J, Templeton K, Paterson S, van Aerle R, Wade MJ, Jeffries AR, Bassano I. 2023. Comparison of metagenomic and targeted methods for sequencing human pathogenic viruses from wastewater. mBio 14:e01468-23.

121. Conceição-Neto N, Zeller M, Lefrère H, De Bruyn P, Beller L, Deboutte W, Yinda CK, Lavigne R, Maes P, Ranst MV, Heylen E, Matthijnssens J. 2015. Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. 1. Scientific Reports 5:16532.

122. Conceição-Neto N, Yinda KC, Van Ranst M, Matthijnssens J. 2018. NetoVIR: Modular Approach to Customize Sample Preparation Procedures for Viral Metagenomics. Methods Mol Biol 1838:85–95.

123. Reyes A, Semenkovich NP, Whiteson K, Rohwer F, Gordon JI. 2012. Going viral: next-generation sequencing applied to phage populations in the human gut. Nat Rev Microbiol 10:607–617.

124. Koonin EV, Dolja VV, Krupovic M. 2015. Origins and evolution of viruses of eukaryotes: The ultimate modularity. Virology 479–480:2–25.

125. Sachsenröder J, Twardziok S, Hammerl JA, Janczyk P, Wrede P, Hertwig S, Johne R. 2012. Simultaneous identification of DNA and RNA viruses present in pig faeces using process-controlled deep sequencing. PLoS One 7:e34631.

126. Hall RJ, Wang J, Todd AK, Bissielo AB, Yen S, Strydom H, Moore NE, Ren X, Huang QS, Carter PE, Peacey M. 2014. Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. J Virol Methods 195:194–204.

127. Li L, Deng X, Mee ET, Collot-Teixeira S, Anderson R, Schepelmann S, Minor PD, Delwart E. 2015. Comparing viral metagenomics methods using a highly multiplexed human viral pathogens reagent. J Virol Methods 213:139–146.

128. Kohl C, Brinkmann A, Dabrowski PW, Radonić A, Nitsche A, Kurth A. 2015. Protocol for metagenomic virus detection in clinical specimens. Emerg Infect Dis 21:48–57.

129. Kleiner M, Hooper LV, Duerkop BA. 2015. Evaluation of methods to purify virus-like particles for metagenomic sequencing of intestinal viromes. BMC Genomics 16:7.

130. Orellana L, Krüger K, Sidhu C, Amann R. 2023. Comparing genomes recovered from time-series metagenomes using long- and short-read sequencing technologies. Microbiome 11.

131. ten Hoopen P, Finn RD, Bongo LA, Corre E, Fosso B, Meyer F, Mitchell A, Pelletier E, Pesole G, Santamaria M, Willassen NP, Cochrane G. 2017. The metagenomic data life-cycle: standards and best practices. GigaScience 6:gix047.

132. Szóstak N, Szymanek A, Havránek J, Tomela K, Rakoczy M, Samelak-Czajka A, Schmidt M, Figlerowicz M, Majta J, Milanowska-Zabel K, Handschuh L, Philips A. 2022. The standardisation of the approach to metagenomic human gut analysis: from sample collection to microbiome profiling. Sci Rep 12:8470.

133. Tourlousse DM, Narita K, Miura T, Sakamoto M, Ohashi A, Shiina K, Matsuda M, Miura D, Shimamura M, Ohyama Y, Yamazoe A, Uchino Y, Kameyama K, Arioka S, Kataoka J, Hisada T, Fujii K, Takahashi S, Kuroiwa M, Rokushima M, Nishiyama M, Tanaka Y, Fuchikami T, Aoki H, Kira S, Koyanagi R, Naito T, Nishiwaki M, Kumagai H, Konda M, Kasahara K, Ohkuma M, Kawasaki H, Sekiguchi Y, Terauchi J. 2021. Validation and standardization of DNA extraction and library construction methods for metagenomics-based human fecal microbiome measurements. Microbiome 9:95.

134. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. 2021. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol 39:578–585.

135. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055.

136. García-López R, Vázquez-Castellanos JF, Moya A. 2015. Fragmentation and Coverage Variation in Viral Metagenome Assemblies, and Their Effect in Diversity Calculations. Front Bioeng Biotechnol 3.

137. Olson ND, Treangen TJ, Hill CM, Cepeda-Espinoza V, Ghurye J, Koren S, Pop M. 2017. Metagenomic assembly through the lens of validation: recent advances in assessing and improving the quality of genomes assembled from metagenomes. Brief Bioinform 20:1140–1150.

138. Sutton TDS, Clooney AG, Ryan FJ, Ross RP, Hill C. 2019. Choice of assembly software has a critical impact on virome characterisation. Microbiome 7:12.

139. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Res 27:824–834.

140. Antipov D, Raiko M, Lapidus A, Pevzner PA. 2020. Metaviral SPAdes: assembly of viruses from metagenomic data. Bioinformatics 36:4126–4129.

141. Delgado LF, Andersson AF. 2022. Evaluating metagenomic assembly approaches for biome-specific gene catalogues. Microbiome 10:72.

142. Philippe N, Legendre M, Doutre G, Couté Y, Poirot O, Lescot M, Arslan D, Seltzer V, Bertaux L, Bruley C, Garin J, Claverie J-M, Abergel C. 2013. Pandoraviruses: Amoeba Viruses with Genomes Up to 2.5 Mb Reaching That of Parasitic Eukaryotes. Science 341:281–286.

143. Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, Pratama AA, Gazitúa MC, Vik D, Sullivan MB, Roux S. 2021. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. Microbiome 9:37.

144. Roux S, Enault F, Hurwitz BL, Sullivan MB. 2015. VirSorter: mining viral signal from microbial genomic data. PeerJ 3:e985.

145. Ren J, Ahlgren NA, Lu YY, Fuhrman JA, Sun F. 2017. VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. Microbiome 5:69.

146. Miao Y, Liu F, Hou T, Liu Y. 2022. Virtifier: a deep learning-based identifier for viral sequences from metagenomes. Bioinformatics 38:1216–1222.

147. Kieft K, Zhou Z, Anantharaman K. 2020. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. Microbiome 8:90.

148. Bzhalava Z, Tampuu A, Bała P, Vicente R, Dillner J. 2018. Machine Learning for detection of viral sequences in human metagenomic datasets. BMC Bioinformatics 19:336.

149. Ren J, Song K, Deng C, Ahlgren NA, Fuhrman JA, Li Y, Xie X, Poplin R, Sun F. 2020. Identifying viruses from metagenomic data using deep learning. Quant Biol 8:64–77.

150. Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, Hu B, Chain PSG, Nayfach S, Kyrpides NC. 2023. Identification of mobile genetic elements with geNomad. Nat Biotechnol 1–10.

151. Rangel-Pineros G, Almeida A, Beracochea M, Sakharova E, Marz M, Muñoz AR, Hölzer M, Finn RD. 2022. VIRify: an integrated detection, annotation and taxonomic classification pipeline using virus-specific protein profile hidden Markov models. bioRxiv https://doi.org/10.1101/2022.08.22.504484.

152. Kieft K, Adams A, Salamzade R, Kalan L, Anantharaman K. 2022. vRhyme enables binning of viral genomes from metagenomes. Nucleic Acids Research 50:e83.

153. Liu J, Lin R, Wu G, Liu R, Luo Z, Fan X. 2022. CoCoNet: Coupled Contrastive Learning Network with Multi-level Feature Ensemble for Multi-modality Image Fusion. arXiv:2211.10960. arXiv https://doi.org/10.48550/arXiv.2211.10960.

154. Uritskiy GV, DiRuggiero J, Taylor J. 2018. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome 6:158.

155. Jurasz H, Pawłowski T, Perlejewski K. 2021. Contamination Issue in Viral Metagenomics: Problems, Solutions, and Clinical Perspectives. Front Microbiol 12.

156. Olomu IN, Pena-Cortes LC, Long RA, Vyas A, Krichevskiy O, Luellwitz R, Singh P, Mulks MH. 2020. Elimination of “kitome” and “splashome” contamination results in lack of detection of a unique placental microbiome. BMC Microbiology 20:157.

157. Paniagua Voirol LR, Valsamakis G, Yu M, Johnston PR, Hilker M. 2021. How the ‘kitome’ influences the characterization of bacterial communities in lepidopteran samples with low bacterial biomass. Journal of Applied Microbiology 130:1780–1793.

158. Kadlečková D, Tachezy R, Erban T, Deboutte W, Nunvář J, Saláková M, Matthijnssens J. 2022. The Virome of Healthy Honey Bee Colonies: Ubiquitous Occurrence of Known and New Viruses in Bee Populations. mSystems e0007222.

159. Li L, Victoria JG, Wang C, Jones M, Fellers GM, Kunz TH, Delwart E. 2010. Bat guano virome: predominance of dietary viruses from insects and plants plus novel mammalian viruses. J Virol 84:6955–6965.

160. Thijssen M, Khamisipour G, Maleki M, Devos T, Li G, Van Ranst M, Matthijnssens J, Pourkarim MR. 2023. Characterization of the Human Blood Virome in Iranian Multiple Transfused Patients. Viruses 15:1425.

161. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

162. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60.

163. Vasimuddin Md, Misra S, Li H, Aluru S. 2019. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems, p. 314–324. *In* 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).

164. Ondov BD, Bergman NH, Phillippy AM. 2011. Interactive metagenomic visualization in a Web browser. BMC Bioinformatics 12:385.

165. Brown CT, Irber L. 2016. sourmash: a library for MinHash sketching of DNA. Journal of Open Source Software 1:27.

166. Nissen JN, Johansen J, Allesøe RL, Sønderby CK, Armenteros JJA, Grønbech CH, Jensen LJ, Nielsen HB, Petersen TN, Winther O, Rasmussen S. 2021. Improved metagenome binning and assembly using deep variational autoencoders. 5. Nat Biotechnol 39:555–560.

167. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 7:e7359.

168. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. 2014. Binning metagenomic contigs by coverage and composition. Nat Methods 11:1144–1146.

169. Wu Y-W, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32:605–607.

170. Fang Z, Tan J, Wu S, Li M, Xu C, Xie Z, Zhu H. 2019. PPR-Meta: a tool for identifying phages and plasmids from metagenomic fragments using deep learning. GigaScience 8:giz066.

171. Olm MR, Brown CT, Brooks B, Banfield JF. 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J 11:2864–2868.

172. Camargo AP, Nayfach S, Chen I-MA, Palaniappan K, Ratner A, Chu K, Ritter SJ, Reddy TBK, Mukherjee S, Schulz F, Call L, Neches RY, Woyke T, Ivanova NN, Eloe-Fadrosh EA, Kyrpides NC, Roux S. 2023. IMG/VR v4: an expanded database of uncultivated virus genomes within a framework of extensive functional, taxonomic, and ecological metadata. Nucleic Acids Research 51:D733–D743.

173. Bolduc B, Jang HB, Doulcier G, You Z-Q, Roux S, Sullivan MB. 2017. vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. PeerJ 5:e3243.

174. Kevill JL, Highfield A, Mordecai GJ, Martin SJ, Schroeder DC. 2017. ABC Assay: Method Development and Application to Quantify the Role of Three DWV Master Variants in Overwinter Colony Losses of European Honey Bees. Viruses 9:E314.

175. Kadlečková D, Saláková M, Erban T, Tachezy R. 2024. Discovery and characterization of novel DNA viruses in Apis mellifera: expanding the honey bee virome through metagenomic analysis. mSystems 9:e00088-24.

176. Runckel C, Flenniken ML, Engel JC, Ruby JG, Ganem D, Andino R, DeRisi JL. 2011. Temporal Analysis of the Honey Bee Microbiome Reveals Four Novel Viruses and Seasonal Prevalence of Known Viruses, Nosema, and Crithidia. PLoS One 6.

177. Bailey L, Woods RD. 1974. Three previously undescribed viruses from the honey bee. J Gen Virol 25:175–186.

178. Bailey L, Carpenter JM, Govier DA, Woods RD. 1980. Bee Virus Y. Journal of General Virology, 51:405–407.

179. Thaduri S, Marupakula S, Terenius O, Onorati P, Tellgren-Roth C, Locke B, de Miranda JR. 2021. Global similarity, and some key differences, in the metagenomes of Swedish varroa-surviving and varroa-susceptible honeybees. Sci Rep 11:23214.

180. de Miranda JR, Bailey L, Ball BV, Blanchard P, Budge GE, Chejanovsky N, Chen Y-P, Gauthier L, Genersch E, de Graaf DC, Ribière M, Ryabov E, De Smet L, van der Steen JJM. 2013. Standard methods for virus research in Apis mellifera. Journal of Apicultural Research 52:1–56.

181. Custer GF, Gans M, van Diepen LTA, Dini-Andreote F, Buerkle CA. 2023. Comparative Analysis of Core Microbiome Assignments: Implications for Ecological Synthesis. mSystems 8:e01066-22.

182. Shi C, Beller L, Deboutte W, Yinda KC, Delang L, Vega-Rúa A, Failloux A-B, Matthijnssens J. 2019. Stable distinct core eukaryotic viromes in different mosquito species from Guadeloupe, using single mosquito viral metagenomics. Microbiome 7:121.

183. De Coninck L, Soto A, Wang L, De Wolf K, Smitz N, Deblauwe I, Mbigha Donfack KC, Müller R, Delang L, Matthijnssens J. 2024. Lack of abundant core virome in Culex mosquitoes from a temperate climate region despite a mosquito species-specific virome. mSystems 0:e00012-24.

184. Peschel S, Müller CL, von Mutius E, Boulesteix A-L, Depner M. 2021. NetCoMi: network construction and comparison for microbiome data in R. Briefings in Bioinformatics 22:bbaa290.

185. Roberts JMK, Anderson DL, Durr PA. 2018. Metagenomic analysis of Varroa-free Australian honey bees (Apis mellifera) shows a diverse Picornavirales virome. J Gen Virol 99:818–826.

186. Qi Y-H, Ye Z-X, Zhang C-X, Chen J-P, Li J-M. 2023. Diversity of RNA viruses in agricultural insects. Computational and Structural Biotechnology Journal 21:4312–4321.