Honey bulk DNA metagenomic analysis to identify honey biological composition and monitor honeybee pathogens

Priit Paluoja[1,2,*], Kaarel Krjutškov[1,2], Andres Salumets[1,2,3], Kairi Raime[1,4]

[1] Celvia CC AS, Tartu, Estonia
[2] Department of Obstetrics and Gynaecology, Institute of Clinical Medicine, University of Tartu, Tartu, Estonia
[3] Division of Obstetrics and Gynecology, Department of Clinical Science, Intervention and Technology, Karolinska Institutet and Karolinska University Hospital, Stockholm, Sweden
[4] Institute of Molecular and Cell Biology, University of Tartu, Estonia

* Corresponding author

E-mail: priit.paluoja@ccht.ee

Abstract

Honeybees are effective environmental monitors due to their long-range foraging activities. Their hive products, particularly honey, reflect the environment of honeybees and honey production. Honey's DNA mixture originates from various organismal groups like plants, arthropods, fungi, bacteria, and viruses. Conventional methods like melissopalynological analysis and targeted honey DNA metabarcoding offer a limited view of honey's composition. We conducted a honey bulk DNA metagenomic analysis of 266 Estonian and 103 foreign centrifugally-extracted honey samples collected between 2020 and 2023. Honey bulk DNA was extracted, prepared, and massively parallel sequenced without the selection of preliminary target gene(s). Millions of honey-origin DNA sequences were analyzed by the taxonomic sequence classifier Kraken 2 to characterize the honey's taxonomic composition and by the Bracken statistical method to identify honeybee pathogens and parasites. In Estonian honey, 70.4% of the bulk DNA was derived from green plant families like *Brassicaceae*, *Rosaceae*, *Fabaceae*, *Pinaceae*, and *Salicaceae*. Geographical distribution analysis revealed distinct botanical compositions between Estonian mainland and island samples, although the most prevalent plant genera in honey were *Brassica, Picea, Trifolium, Rubus, and Salix*. The bacterial family *Lactobacillaceae* was prevalent overall, reflecting the leading proportion of DNA from honeybee microbiota in honey. Honey bulk DNA analysis reveals all DNA traces from other organisms that reflect the environment of honey production, e.g. honeybees, humans, bacteria, yeasts, domestic animals, and DNA viruses. We detected 12 honeybee pathogens and parasites, including *Paenibacillus larvae*, *Melissococcus plutonius, Nosema ceranae, Varroa destructor*, and *Aethina tumida*.

In conclusion, comprehensive honey bulk DNA metagenomic analysis highlights honey's diverse biological composition, including microbial, fungal, botanical, animal and pathogenic elements. The findings align with previous studies and reveal geographical variations in honey composition. The study underscores the potential of bulk DNA-based and non-targeted metagenomic approaches for monitoring honeybee health, environmental quality, and honey composition, origin, and authenticity.

Introduction

Honeybees are considered effective large-scale environmental monitors due to their large-scale foraging activities. Their hive products, especially honey, provide a snapshot of the honeybee and honey production environment, containing nectar and pollen DNA from various plant species and DNA sequences from arthropods, fungi, bacteria, and viruses (1,2). Previous studies focusing on North European honey biological composition have identified predominant floral sources such as *Brassica, Trifolium, Malus, Prunus, Fragaria, Medicago, Populus*, and *Solanum* (3,4), that are widely spread plant genera also in Estonian nature. *Apis mellifera*, as anticipated, is the most commonly detected arthropod species in honey DNA analyses (1,3). Additionally, DNA from several other arthropods from honeybee foraging environments, like plant-sucking and honeydew-producing insects, aphids from the order *Hemiptera* have been detected not only from honeydew honey but also from blossom honey (5). From viruses, mainly *Apis mellifera* filamentous virus (AmFV) has also been identified within honey DNA, which is known to be a ubiquitous dsDNA virus that affects many apiaries throughout Europe and can have mild pathogenetic effects on honeybees (6). Also, fungi, mostly yeasts, that are known to tolerate high sugar concentration and recognised for their roles in food and beverage production as fermentative agents, such as species from *Zygosaccharomyces*, and fungal pathogens affecting insects or plants, such as *Metarhizium* spp., *Aspergillus* spp., *Nosema ceranae*, *Bettsia alvei*, or *Alternaria alternata*, have also been observed in honey samples (1,3). Honey DNA has been found to contain common microorganisms from the honeybee gut microbiota, such as *Lactobacillus kunkeei*, as well as pathogens affecting honeybees or plants, and ubiquitous species like *Escherichia coli* (1). Honey DNA analysis has been used to detect several potential honeybee pathogens, such as *Paenibacillus larvae* – the causative agent of American Foulbrood, *Melissococcus plutonius* – the aetiological agent of European Foulbrood, and *Spiroplasma* species – the agent of the spiroplasmosis (1,3,7). Screening for pathogens is essential for several reasons. This detection aids in identifying and managing diseases affecting honeybee colonies that are already at an early stage. Colony losses have been linked to pathogens such as *Varroa destructor* or *Nosema ceranae* (8). Sensitive bulk DNA-based screening allows the detection of infections before visual symptoms appear. For hive health, early detection of pathogens can facilitate timely intervention,

potentially saving colonies from devastating diseases. Additionally, understanding the prevalence and spread of pathogens locally and on larger scales can help monitor and manage diseases and invasive honeybee parasites.

Considering the above, the honey composition reflects the surrounding ecological landscape. It helps detect pollinators and pathogens, map hive health, describe the honeybee foraging and honey production environment, and describe geographical peculiarities, creating a fingerprint of common regional honey and combating food fraud. Traditional methods, such as melissopalynological analysis or DNA metabarcoding, offer a limited view of honey composition. The melissopalynological analysis is restricted to detecting pollen plants, ignoring nectar and honeydew plants and other organisms, including pathogens, that leave DNA traces in honey (9). DNA metabarcoding expands this scope by targeting a broader range of organisms, but it remains a targeted approach, limited to detect only targeted taxa based on a few successfully preamplified genomic regions (10). To use an unbiased approach, we used shotgun metagenomics sequencing of all DNA extracted from honey sample, which describes the complexity of samples containing thousands of distinct species belonging to different kingdoms or phyla (10). We conducted a thorough all-DNA-sequencing-based metagenomic analysis on 266 Estonian and 103 foreign centrifugally-extracted honey samples to map the botanical composition of Estonian honey with geographical distribution. We conducted a comprehensive pathogen analysis of Estonian and foreign samples.

Materials and Methods

Honey samples

A total of 264 honey samples were collected from various regions across Estonia to describe the DNA composition of Estonian honey (**Table 1**). Additionally, two positive control samples from the hives with diagnosed American Foulbrood infection caused by *Paenibacillus larvae* were included, although their specific locations were not disclosed and are therefore included in honeybee pathogen analysis but not in the analysis of Estonian honey DNA botanical composition and geographical distribution (**Table 1,** labelled as undetermined). For honeybee pathogen analysis, in addition to the Estonian honey samples, 103 foreign samples were obtained directly from beekeepers, shops, or honey markets (**Table 1**). All samples were produced during the summers of 2020 to 2022 and collected for analysis between 2020 and 2023. It is important to note that all honey samples were collected from honey extracted and mixed from several honeycombs using a centrifugal extractor and not the honeycomb scraping method. Such samples contain DNA traces from several honeycombs and several hives in the apiary and provide a more comprehensive DNA taxonomical composition picture of the honeybees' foraging, hives, and honey production environment in an apiary.

107 **Table 1. Geographical distribution of 266 Estonian honey samples used in the study and origins of**
108 **103 non-Estonian honey samples utilised in the pathogen analysis.**

Estonian regions	N	Non-Estonian countries	N
Harju County	34	Austria	1
Hiiu County	13	Bulgaria	9
Ida-Viru County	6	China	10
Jõgeva County	10	Denmark	2
Järva County	5	England	2
Kihnu	2	Finland	3
Lääne County	3	Germany	1
Lääne-Viru County	25	Ghana	1
Muhu	3	Greece (including Rhodes Island)	4
Põlva County	4	India	4
Pärnu County	33	Italy	1
Rapla County	3	Kazakhstan	1
Ruhnu	1	Latvia	4
Saare County	13	Mix of EU and non-EU honey	18
Tartu County	26	Montenegro	1
Valga County	17	Non-EU honey (including Ukraine, Central and South American honey)	7
Viljandi County	28	Poland	3
Vilsandi	1	Portugal	1
Vormsi	4	Saudi Arabia	1
Võru County	33	Scotland	3
Undetermined	2	Spain	2
		Switzerland	2
		Ukraine	17
		United Arab Emirates	1
		Unspecified EU honey	2
		USA	1
		Yemen	1

109

110 # DNA extraction and sequencing

111 The honey sample was preheated at 40°C and homogenized by mixing with a clean spoon. 30 g honey
112 was weighed into a 50 ml centrifuge tube and diluted in 25 ml of preheated MilliQ water. After
113 centrifugation at 4000 rpm, the supernatant was removed, and bulk DNA from the pellet was extracted
114 by NucleoSpin Food Mini kit (MACHEREY-NAGEL). The DNA was fragmented down to 150-200 bp
115 fragments by Covaris M220 focused-ultrasonicator (Covaris) and concentrated by NucleoSpin Gel and
116 PCR Clean-up kit (MACHEREY-NAGEL). The quality and quantity of the DNA fragments were

117 assessed on Agilent 2200 TapeStation (Agilent Technologies). Illumina-compatible DNA libraries were
118 prepared using the Celvia CC AS in-house developed FOCUS protocol. Briefly, fragmented 25 µl honey
119 bulk DNA (1 ng/ µl) was end-repaired and A-tailed by a specific enzymatic mixture. Short double-
120 stranded and index-labelled DNA adapters were ligated to both ends of pre-treated DNA fragments. The
121 full adapter sequence and sufficient ready-made Illumina-compatible library were ensured by 12-cycle
122 PCR. 36 samples were pooled equimolarly, and the quality and quantity of the pool were assessed on
123 Agilent 2200 TapeStation (Agilent Technologies). The honey bulk DNA pooled library was sequenced
124 using the Illumina NextSeq 500 instrument (Illumina Inc.) and 85 bp single-read protocol. Past-filtered
125 sequencing read counts varied from 1 to 27 million, with a median of 13.7 million reads per sample.
126 Read counts were normalised by total read count to ensure comparability across samples.

127 Metagenomic analysis

128 To classify the taxonomic composition of the Estonian honey by assigning taxonomic labels to sequence
129 reads, we utilized the taxonomic sequence classifier Kraken 2 with a custom reference database (11).
130 The minimum hit groups required for classification were set to 3, and the confidence threshold for
131 taxonomic assignment was set to 0.5. The Kraken 2 custom reference database was built using the
132 reference sequences sourced from the three main collections: NCBI nt collection, The One Thousand
133 Plants Project, and NCBI's Sequence Read Archive (SRA) (12–14). Specifically, The One Thousand
134 Plants Project and NCBI's SRA were used to incorporate sequences of honey plants widely distributed
135 in Estonia but not represented in the NCBI nt collection.

136 To describe and analyse honey bee pathogens and parasites in honey DNA on the species level, we used
137 Bracken with the read length set to 80, taxonomic level to species, and threshold for the abundance
138 estimation set to 10 (15). We analysed the presence of following 20 honeybee related parasites and
139 pathogens: *Acarapis woodi, Acarus siro, Achroia grisella, Aethina tumida, Ascosphaera apis, Bettsia*
140 *alvei, Braula coeca, Forficula auricularia, Galleria mellonella, Melissococcus plutonius, Nosema apis,*
141 *Nosema ceranae, Oplostomus fuligineus, Paenibacillus larvae, Senotainia tricuspis, Spiroplasma apis,*
142 *Spiroplasma melliferum, Tropilaelaps clareae, Tropilaelaps mercedesae,* and *Varroa destructor*.

Results

Our study presents a metagenomic analysis of honey bulk DNA to identify its biological composition and monitor honeybee pathogens.

Estonian honey DNA biological composition

In our analysis of the DNA composition of Estonian honey, we characterised the proportions of bacteria, fungi, animals (Animalia, Metazoa), green plants (Viridiplantae), and viruses (**Fig 1**). As anticipated, most of the DNA was derived from green plants (70.4%), with bacteria constituting a secondary component (22.7%).

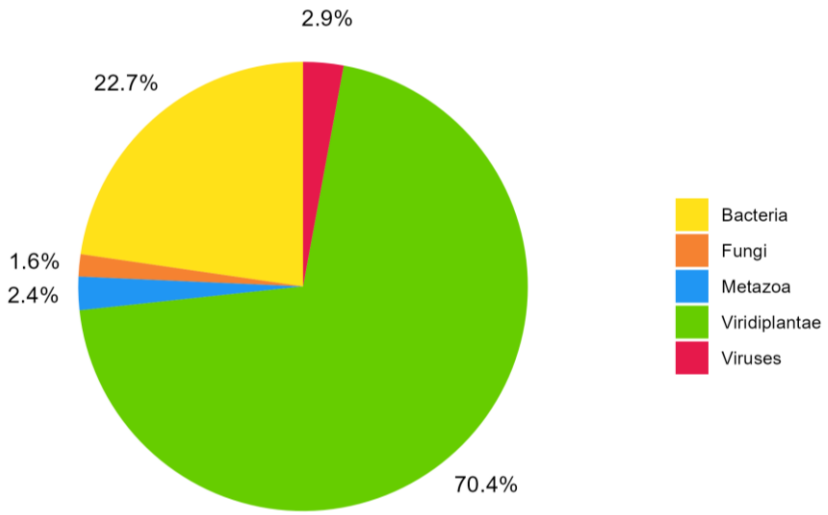


Fig 1. Bulk DNA taxonomic composition of Estonian honey.

Although Viridiplantae dominated the honey composition, the dominant family identified was the bacterial *Lactobacillaceae* (19.5%) (**Fig 2A**). Within the family *Lactobacillaceae*, the prevalent genus was *Apilactobacillus* (**S1 Fig**). The second next bacterial families were *Pseudomonadaceae* (1.7%) and *Erwiniaceae* (1.5%). The top five prevalent families of Viridiplantae in Estonian honey are *Brassicaceae* (19.1%), *Rosaceae* (13.1%), *Fabaceae* (12.0%), *Pinaceae* (9.1%), and *Salicaceae* (7.4%) (**Fig 2A)**. As expected, the most detected genera were *Brassica*, *Picea*, *Trifolium*, *Rubus*, and *Salix* (**S1 Fig**).

160 The prominent Animalia families detected in honey DNA were *Hominidae* and *Apidae*, containing
161 human (genus *Homo*), honeybee (genus *Apis*), and bumblebee (genus *Bombus*) DNA (**Fig 2A,B, S1**
162 **Fig**). Interestingly, the analysis revealed DNA traces belonging to the mammal families *Canidae* and
163 *Bovidae*, albeit in proportions under 0.2% (**Fig 2A,B, S1 Fig**). Also, DNA from arthropod families
164 containing honeybee or hive parasites or pests can be detected, e.g., *Varroidae*, *Pyralidae, and Vespidae*
165 (**Fig 2B**). The prominant fungal families detected in honey DNA were *Saccharomycetaceae* and
166 *Metschnikowiaceae*, mainly from yeasts' genera *Zygosaccharomyces*, *Saccharomyces,* and
167 *Metschnikowia* (**S1 Fig**). Viral DNA was predominantly from the *Apis mellifera filamentous virus (***S1**
168 **Fig)**.

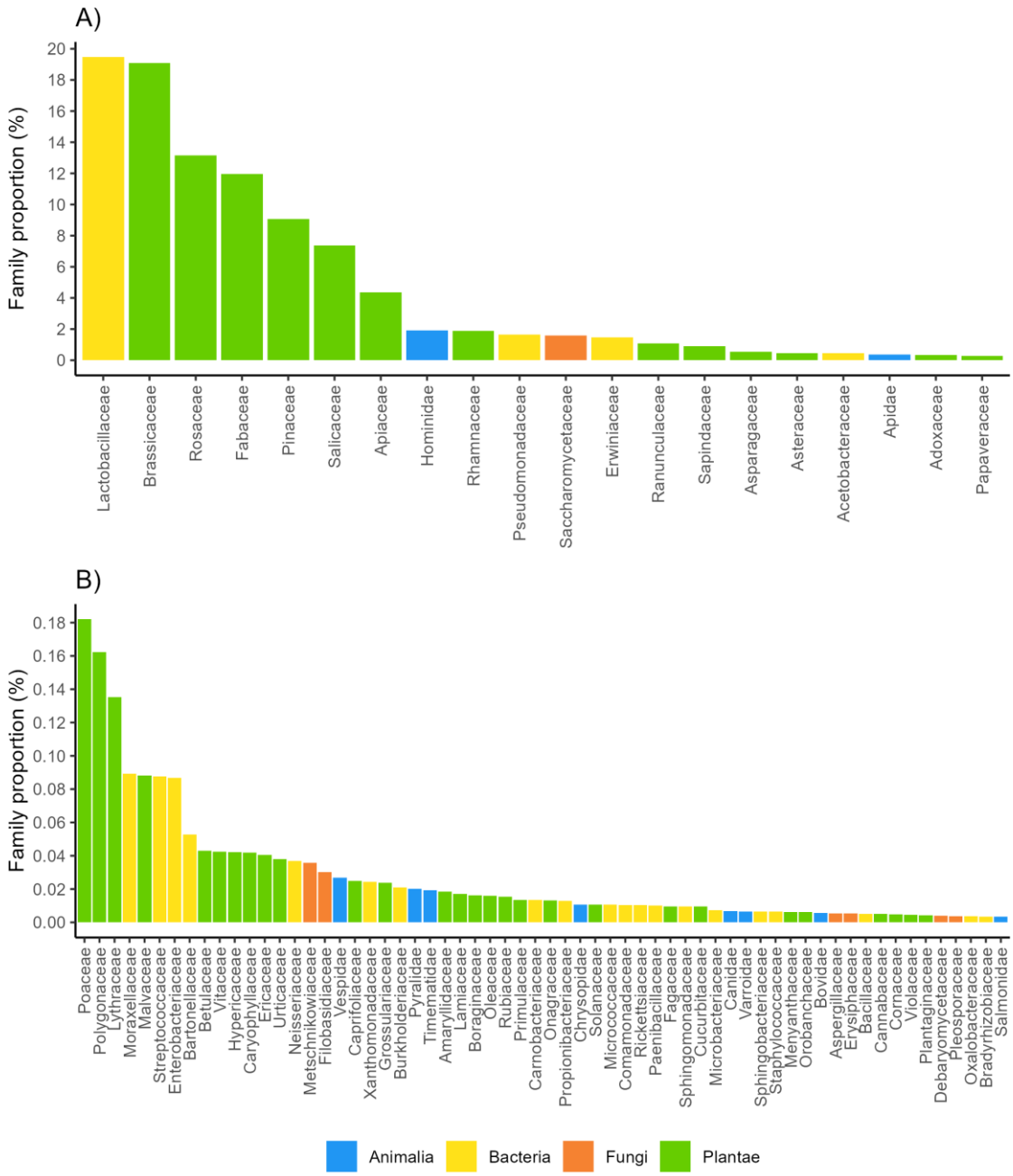


Fig 2. Proportions of bacteria, fungi, animals (Animalia), and green plants (Plantae) detected in Estonian honey. Panel A is over 0.2%, and Panel B is under 0.2%.

Estonian honey bulk DNA botanical composition and geographical distribution

We investigated the geographical distribution of different plant genera of the family Viridiplantae in Estonian honey samples (**Fig 3**). The widely distributed genus was *Brassica*, as confirmed by **Fig 2**, where the family *Brassicaceae* was the most common Viridiplantae. While *Brassica* was common and contributed in most areas, there were exceptions. For example, *Brassica* was not dominant on islands like Vilsandi, Ruhnu, Muhu, Kihnu, and Vormsi (**Fig 3**). Additionally, the islands had different prevalent genera compared to the mainland, such as *Frangula*, *Geum, Rhamnus*, and a considerable proportion of other plant genera (categorised as "Other") (**Fig 3**). From north to south, the mainland featured common genera such as *Brassica, Picea, Trifolium, Salix* and *Rubus*. From east to west, there was an increase in *Rhamnus* and *Frangula* prevalence. Other genera, such as *Aegopodium*, *Vicia*, and *Melilotus*, were also prevalent in Estonian honey.

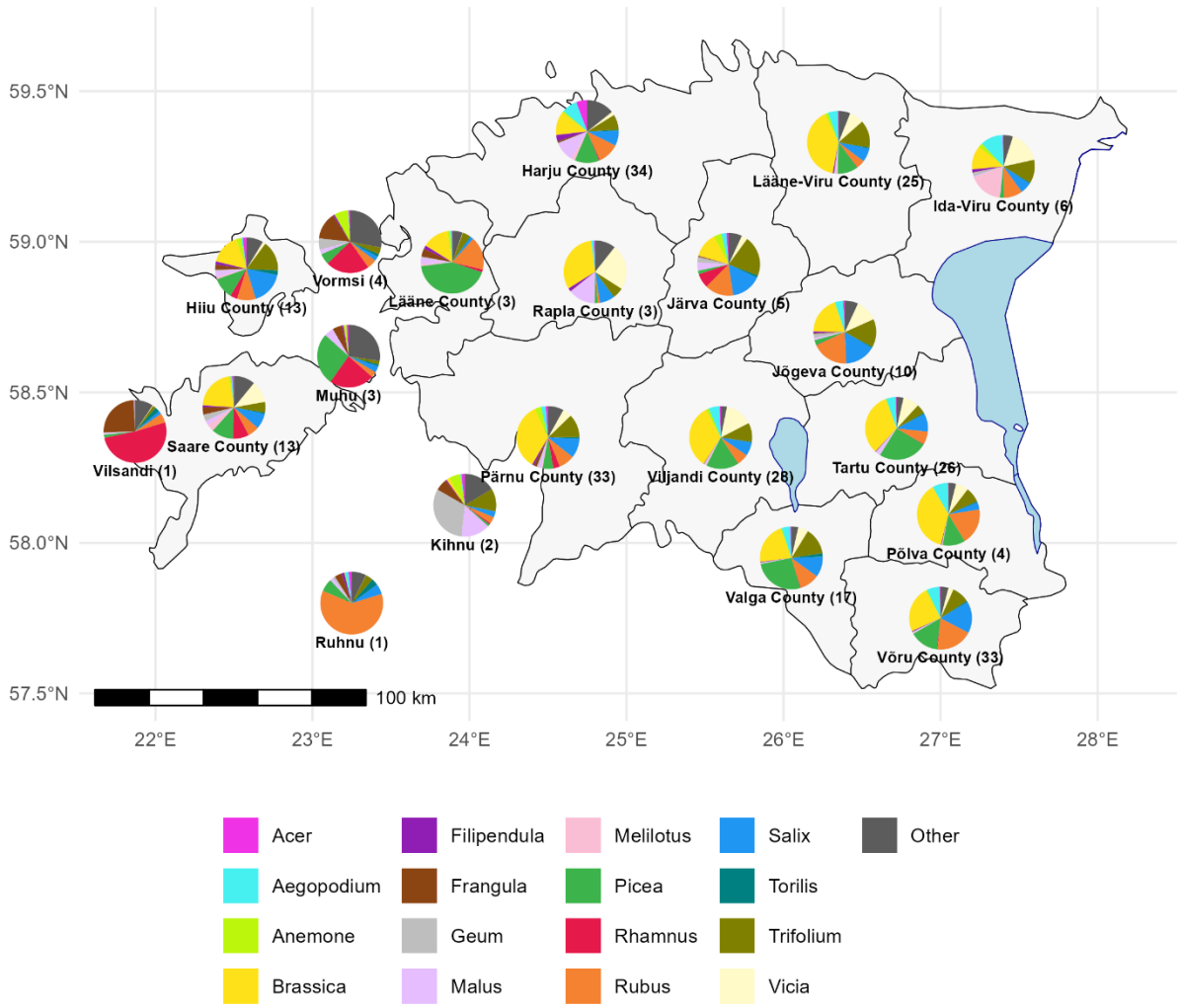


Fig 3. Honey bulk DNA botanical composition and geographical distribution.

Honey bee pathogens and parasites in honey bulk DNA

Our methodology detects DNA traces from honeybee-related pathogens and parasites. We pre-selected and monitored 20 honeybee pathogens and parasites in Estonian and foreign honey samples (see **Materials and Methods**). Specific DNA sequences from 12 pathogens or parasites were detected in numerous samples with either laboratory-confirmed pathogens, visually confirmed parasites, beekeeper-suspected issues, or samples without confirmation (**Fig 4**). For instance, DNA proof from the bacterium *Paenibacillus larvae*, which causes honeybee disease American Foulbrood, was detected in two laboratory-confirmed control honey samples, each with a fraction of sequencing reads exceeding 2%.

196 In all the samples where the microsporidian parasite *Nosema* sp. was detected, including two samples
197 from the hives suspected of nosematosis, only *Nosema ceranae* was detected but not *Nosema apis*. As
198 expected, DNA traces of *Aethina tumida* (small hive beetle) were only observed in some foreign
199 samples, as this beetle is not present in Estonia. DNA traces from flour mite *Acarus siro* were detected
200 in one Estonian honey sample. The widespread parasitic honeybee mite (*Varroa destructor*) and the
201 greater wax moth (*Galleria mellonella*) were found in Estonian and foreign apiaries. Also, DNA
202 sequences from honeybee pathogens or pests like *Ascosphaera apis* (fungus causing Chalkbrood),
203 *Melissococcus plutonius* (causing European Foulbrood), *Spiroplasma* species (related to spiroplasmosis,
204 May disease), *Bettsia alvei* (causing pollen mold), and even from *Forticula auricularia* (insect,
205 European earwig) were detected in numerous Estonian and foreign honey samples (**Fig 4**).

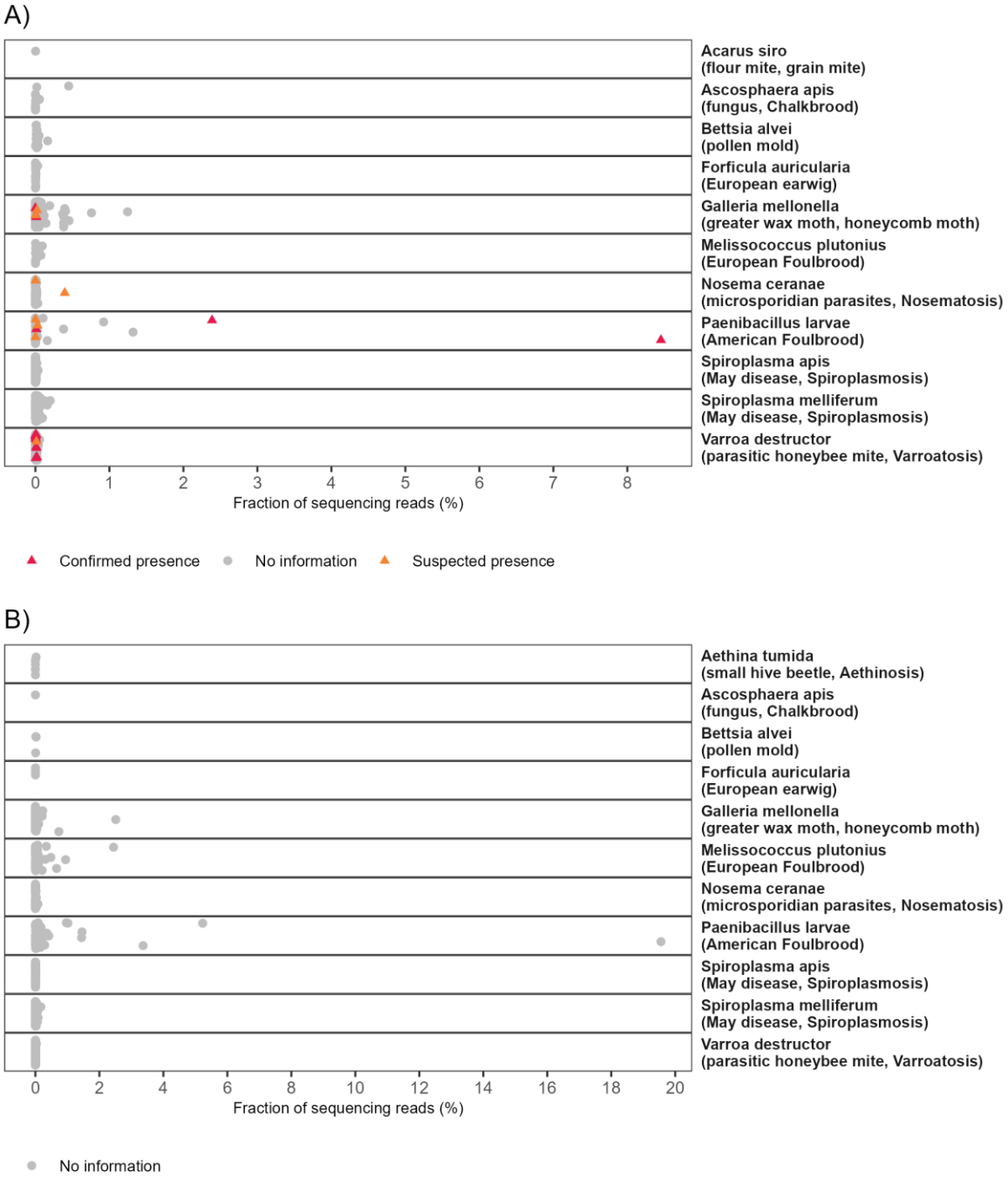


Fig 4. Detection of pathogens and parasites in Estonian (A) and foreign (B) honey samples. Red triangles indicate laboratory-confirmed pathogens or visually confirmed parasites, while orange triangles represent beekeeper-suspected issues. Grey points ("No information") depict samples without infection confirmation but containing sequencing reads belonging to known parasites or pathogens. Honey samples that did not yield any sequencing reads assigned to the pathogens listed in the Methods section are excluded from this figure. A fraction close to 0% signifies a very low proportion of sequencing reads assigned to a particular pathogen but indicates presence. For instance, the single sample containing *Acarus siro* in panel (A) had 11 reads assigned by Kraken, with an additional 102 reads assigned by Bracken, resulting in 0.002% of the total reads. Notably, certain pathogens were detected exclusively in either Estonian or foreign honey samples. For example, *Aethina tumida* presence was found only in foreign samples (B), whereas *Acarus siro* was detected in only one Estonian sample

(panel A). Sequencing reads originating from *Acarapis woodi* were not detected in any of the samples analysed.

Discussion

The honey bulk DNA metagenomic analysis provides an unbiased and non-restricted overview of honey's plant species and all other biological components that contain DNA. Unlike the DNA metabarcoding method, which targets limited selected gene(s) of the specific organism(s), the honey bulk DNA approach provides a comprehensive overview of honey botanical, microbial, fungal, entomological, and animal diversity, including honeybee pathogens and parasites (16). We conducted thorough analyses on 266 Estonian and 103 foreign honey samples. Unlike honeycomb-scrapped samples, these samples were collected from centrifugally extracted honey, which contains honey DNA from various honeycombs and hives of the apiary. The amount of at least one million DNA sequencing reads per honey sample enables us to describe the biological environment of honeybee foraging and honey production. In addition to the plant DNA from pollen, the method analyses all DNA traces in the sample, including cell-free DNA, which allows us to detect pollen and nectar and honeydew plants.

We demonstrate that green plants (Viridiplantae) constitute the majority of the DNA content in honey, accounting for 70.4% of the total honey DNA composition, with *Brassicaceae*, *Rosaceae*, *Fabaceae*, *Pinaceae*, and *Salicaceae* being the most common families identified in Estonian honey (**Fig 1, Fig 2**). The most common plant genera were expectedly *Brassica*, *Picea*, *Trifolium*, *Rubus*, and *Salix* (**S1 Fig**). These results concord with the observations made for the composition of honey pollen plants in Estonia (17), indicating that a significant part of plant DNA in honey may originate from plant pollen in honey and less from plant nectar or honeydew.

Interestingly, the most predominant genus detected in honey based on the amount of sequencing reads was not from the plant DNA but the bacterial genus *Apilactobacillus*, aligning with its known association with honeybee microbiota (**S1 Fig**), as also shown by the past study (3). Although in much lower proportions, also other notable bacterial families, like *Pseudomonadaceae* and *Erwiniaceae* (1.7% and 1.5%, respectively), were detected, both of which include species known for their roles in various

245 ecological functions and interactions with plants and insects (**Fig 2)** (18). These findings demonstrate
246 that the taxonomic diversity of plant genera in honey DNA surpasses that of bacterial genera.
247 Specifically, the DNA sequences from plants are distributed among a greater number of genera
248 compared to the bacterial DNA in the honey composition.

249 As expected, the most common Animalia families detected in honey DNA were the mammal's family
250 *Hominidae* and the arthropods family *Apidae*, containing mostly human (genus *Homo*), honeybee (genus
251 *Apis*), and bumblebee (genus *Bombus*) DNA from honeybee foraging and honey production
252 environment. Interestingly, DNA from arthropod families containing common honeybee or hive
253 parasites or pests from the honeybee or honey production environment can be detected in honey DNA,
254 e.g., *Varroidae*, *Pyralidae, and Vespidae* (**Fig 2B**). The family *Vespidae* includes species detrimental to
255 honey bees, such as hornets. Although hornet DNA detected in our samples was mainly from the
256 European hornet *Vespa crabro*, this finding could be valuable when searching methods for monitoring
257 and early detection of the Asian hornet (*Vespa velutina*), a species known to be devastating for honey
258 bee populations in warmer areas of Europe, but not yet detected in Estonia (19). The widespread parasitic
259 honeybee mite (*Varroa destructor*) from the arthropod family *Varroidae* and the greater wax moth
260 (*Galleria mellonella*) from the family *Pyralidae* were detected in many Estonian and foreign honey
261 samples (**Fig 4)** (20,21).

262 In contrast, the small hive beetle (*Aethina tumida*), known to cause colony collapses in weak colonies,
263 was only found in three samples, according to the label originating from the US, Spain, Ghana, and two
264 honey blends of undetermined geographical origins (**Fig 4)** (22). Importantly, *Aethina tumida*, known to
265 be absent in Estonia, was not detected in any Estonian honey samples (**Fig 4)**. This approach
266 demonstrates that the honey bulk DNA metagenomic analysis could be a valuable screening tool to
267 monitor agriculturally significant honeybee parasites' prevalence and geographical distribution.

268 Our analysis revealed the presence of several other honeybee-related pathogens and parasites (**Fig 4)**.
269 Notably, the bacteria species *Paenibacillus larvae*, which is known to cause American Foulbrood disease
270 in honeybees, was detected in several samples, including two control honey samples from the hives that
271 were confirmed to have American Foulbrood disease (23). In both control samples, a substantial

272 proportion of sequencing reads were attributed to *Paenibacillus larvae* (**Fig 4,** 8.5% and 2.4%). Also,
273 DNA traces from honeybee pathogens or parasites like *Ascosphaera apis* (fungus causing Chalkbrood),
274 *Melissococcus plutonius* (causing European Foulbrood), *Nosema ceranae* (microsporidian parasite,
275 causing Nosematosis), *Spiroplasma* species (related to spiroplasmosis, May disease), *Bettsia alvei*
276 (causing pollen mold), and even from *Forticula auricularia* (insect, European earwig) were detected in
277 several Estonian and foreign honey samples. We did not detect DNA of the following honeybee
278 pathogens or parasites in any analysed Estonian or foreign honey sample: *Acarapis woodi* (parasitic
279 honeybee mite, causes acarapiosis), *Achroia grisella* (lesser wax moth), *Braula coeca* (Braula fly, bee
280 louse), *Oplostomus fuligineus* (large African hive beetle), *Senotainia tricuspis* (fly, causes
281 senotainiosis), *Tropilaelaps clareae* (parasitic honeybee mite, causes tropilaelapsosis) and *Tropilaelaps*
282 *mercedesae* (parasitic honeybee mite, causes tropilaelapsosis). This might be because these important
283 honeybee pathogens and parasites species are not widespread worldwide, and none of these have been
284 seen in Estonian apiaries yet. We also did not detect the microsporidian parasite *Nosema apis* in our
285 samples, even though it has been identified as the primary *Nosema* species responsible for Nosematosis
286 in Estonia (24). Research has shown that *N. ceranae* has replaced *N. apis* in many countries including
287 Italy, Argentina or even northern countries such as Lithuania (24–28). Essentially, *N. ceranae* has spread
288 rapidly worldwide (24). Therefore, it is possible that *N. ceranae* has also replaced *N. apis* in Estonia by
289 now.

290 Interestingly, we even detected trace amounts of DNA sequences from mammals, probably originating
291 from domestic or pest animals, possibly due to the contamination DNA as honeybees often collect
292 brackish water enriched with mineral salts, which could be contaminated by mammal excreta (*Canidae*
293 and *Bovidae,* **Fig 2**) (29). This result shows the sensitivity of DNA analysis and indicates the possible
294 DNA transfer through honeybees' diet. This is in accordance with the study that has demonstrated the
295 presence of DNA from plant-sucking insects in honey DNA that produce the sticky excretion collected
296 by honeybees (5). DNA contamination from pest animals, such as mice representing <0.2% of
297 sequencing reads, may result from their contact with the honeycombs or the hive environment.

298 The fungal community was primarily represented by *Saccharomycetaceae* and *Metschnikowiaceae*,
299 families of yeasts, mainly genera *Zygosaccharomyces*, *Saccharomyces*, and *Metschnikowia*, commonly
300 involved in fermentation processes (**S1 Fig**). The presence of *Saccharomycetaceae* has also been
301 detected in previous honey related studies (1,3,30). We also detected viral DNA, predominantly from
302 the *Apis mellifera filamentous virus* (**S1 Fig**), which is known to infect honey bees but is little to no
303 pathogenic and has been detected in the past studies (6,31). The difference between our finding of 2.9%
304 sequencing reads assigned to DNA viruses, and the 40.2% (± 30.0%) as reported in (3) can be explained
305 by differences in the reference database and the number of samples analysed (**Fig 1**).

306 We also investigated Estonian honey DNA botanical composition with geographical distribution (**Fig**
307 **3**). Consistent with previous findings, we also observed frequent occurrences of *Brassica*, *Malus,* and
308 *Trifolium*, aligning with previous records from North European honey (**Fig 3)** (3,17,32). Interestingly,
309 we observe distinct differences in honey composition between the mainland and the islands, with the
310 islands showing a higher proportion of *Frangula* and species categorised as "Other" compared to the
311 mainland (**Fig 3)**. On small islands in Estonia, the proportion of *Brassica* was substantially lower
312 compared to the other regions. This could be explained by the lack of large agricultural fields on small
313 islands in Estonia. Furthermore, the diverse DNA taxonomical composition of honey creates a unique
314 fingerprint for every honey sample containing hundreds of different species of plants, bacteria, fungi,
315 insects and other organisms. Therefore, we hypothesise that metagenomic analysis of all extracted DNA
316 could be utilised to analyse the authenticity and geographical origin of honey (**Fig 2**, **Fig 3)**.

317 Metagenomic analysis of honey DNA presents inherent challenges, primarily because the accuracy of
318 the results heavily relies on the public reference database used for analysis, as also pointed out by other
319 researchers (33). If a genus is absent from the database, it can introduce biases and potentially reduce
320 the accuracy of the analysis (33). As comprehensive databases for plants are still under development and
321 there is a predominance of complete genome sequences for bacteria and viruses in existing databases,
322 we created a custom Kraken 2 reference database in our study (including partial genome sequences),
323 with the extended numbers of honey-related plants. Our Kraken 2 reference database was sourced from
324 three main collections: NCBI nt collection, The One Thousand Plants Project, and NCBI's Sequence

Read Archive (12–14). This approach enables the detection of an increased number of plants in honey DNA. In addition, the majority of foreign honey samples were acquired from shops, the contents of the honey jars were not validated, and we had to rely on the label information. However, as we were using foreign honey samples only for pathogen analysis, the accuracy of the label did not affect the proof-of-concept of detecting known pathogens in honey samples.

In conclusion, our metagenomic analysis of honey DNA provided a detailed and comprehensive overview of its biological composition, highlighting its significant microbial, botanical, and pathogenic diversity. This study mapped the botanical composition of Estonian honey with geographical distribution and conducted pathogen analysis, underscoring the potential of all DNA sequencing-based metagenomic approaches not only for describing the botanical composition of honey, monitoring honeybee health and apiary environment but also for identifying authenticity and origin of honey by using untargeted analysis of all DNA sequences extracted from honey.

Data availability

The data generated during this study is available in the Sequence Read Archive (SRA) repository under BioProject PRJNA1135913 (https://www.ncbi.nlm.nih.gov/sra/PRJNA1135913).

Funding

This work was supported by the Estonian Rural Development Plan (RDP) 2014-2020 (616219790085) and coordinated by the Agricultural Registers and Information Board in Estonia.

Acknowledgements

The authors would like to thank Dr. Mihkel Vaher and Dr. Hindrek Teder for their support and assistance.

References

1. Bovo S, Ribani A, Utzeri VJ, Schiavo G, Bertolini F, Fontanesi L. Shotgun metagenomics of honey DNA: Evaluation of a methodological approach to describe a multi-kingdom honey bee derived environmental DNA signature. PLoS One [Internet]. 2018 Oct 1 [cited 2024 Jul 8];13(10):e0205575. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0205575

2. Wirta H, Abrego N, Miller K, Roslin T, Vesterinen E. DNA traces the origin of honey by identifying plants, bacteria and fungi. Sci Reports 2021 111 [Internet]. 2021 Feb 26 [cited 2024 Jul 10];11(1):1–14. Available from: https://www.nature.com/articles/s41598-021-84174-0

3. Wirta HK, Bahram M, Miller K, Roslin T, Vesterinen E. Reconstructing the ecosystem context of a species: Honey-borne DNA reveals the roles of the honeybee. PLoS One [Internet]. 2022 Jul 1 [cited 2024 Jul 8];17(7):e0268250. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0268250

4. Varis A-L, Helenius J, Koivulehto K. Pollen spectrum of Finnish honey. Agric Food Sci [Internet]. 1982 Dec 1 [cited 2024 Jul 10];54(5):403–20. Available from: https://journal.fi/afs/article/view/72111

5. Utzeri VJ, Schiavo G, Ribani A, Tinarelli S, Bertolini F, Bovo S, et al. Entomological signatures in honey: an environmental DNA metabarcoding approach can disclose information on plant-sucking insects in agricultural and forest landscapes. Sci Rep [Internet]. 2018 Dec 1 [cited 2024 Jul 16];8(1):9996. Available from: /pmc/articles/PMC6030050/

6. Papp M, Tóth AG, Békési L, Farkas R, Makrai L, Maróti G, et al. Apis mellifera filamentous virus from a honey bee gut microbiome survey in Hungary. Sci Reports 2024 141 [Internet]. 2024 Mar 9 [cited 2024 Jul 4];14(1):1–8. Available from: https://www.nature.com/articles/s41598-024-56320-x

7. Ebeling J, Knispel H, Hertlein G, Fünfhaus A, Genersch E. Biology of Paenibacillus larvae, a

372 deadly pathogen of honey bee larvae. Appl Microbiol Biotechnol [Internet]. 2016 Sep 1 [cited
373 2024 Jul 10];100(17):7387–95. Available from:
374 https://link.springer.com/article/10.1007/s00253-016-7716-0

375 8. Ravoet J, Maharramov J, Meeus I, De Smet L, Wenseleers T, Smagghe G, et al. Comprehensive
376 Bee Pathogen Screening in Belgium Reveals Crithidia mellificae as a New Contributory Factor
377 to Winter Mortality. PLoS One [Internet]. 2013 Aug 26 [cited 2024 Jul 10];8(8):e72443.
378 Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0072443

379 9. Louveaux J, Maurizio A, Vorwohl G. Methods of Melissopalynology. Bee World [Internet].
380 1978 Jan [cited 2024 Jul 10];59(4):139–57. Available from:
381 https://www.tandfonline.com/doi/abs/10.1080/0005772X.1978.11097714

382 10. Porter TM, Hajibabaei M. Scaling up: A guide to high-throughput genomic approaches for
383 biodiversity analysis. Mol Ecol [Internet]. 2018 Jan 1 [cited 2024 Jul 10];27(2):313–38.
384 Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/mec.14478

385 11. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol
386 [Internet]. 2019 Nov 28 [cited 2024 Jul 10];20(1):1–13. Available from:
387 https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1891-0

388 12. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank. Nucleic Acids Res
389 [Internet]. 2007 Jan 1 [cited 2024 Jul 8];35(suppl_1):D21–5. Available from:
390 https://dx.doi.org/10.1093/nar/gkl986

391 13. Leebens-Mack JH, Barker MS, Carpenter EJ, Deyholos MK, Gitzendanner MA, Graham SW, et
392 al. One thousand plant transcriptomes and the phylogenomics of green plants. Nat 2019 5747780
393 [Internet]. 2019 Oct 23 [cited 2024 Jul 8];574(7780):679–85. Available from:
394 https://www.nature.com/articles/s41586-019-1693-2

395 14. Kodama Y, Collaboration on behalf of the INSD, Shumway M, Collaboration on behalf of the
396 INSD, Leinonen R, Collaboration on behalf of the INSD. The sequence read archive: explosive

growth of sequencing data. Nucleic Acids Res [Internet]. 2012 Jan 1 [cited 2024 Jul 8];40(D1):D54–6. Available from: https://dx.doi.org/10.1093/nar/gkr854

15. Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: Estimating species abundance in metagenomics data. PeerJ Comput Sci [Internet]. 2017 Jan 2 [cited 2024 Jul 10];2017(1):e104. Available from: https://peerj.com/articles/cs-104

16. Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. Towards next-generation biodiversity assessment using DNA metabarcoding. Mol Ecol [Internet]. 2012 Apr 1 [cited 2024 Jul 9];21(8):2045–50. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-294X.2012.05470.x

17. Puusepp L, Koff T. Pollen analysis of honey from the Baltic region, Estonia. Grana [Internet]. 2014 [cited 2024 Jul 9];53(1):54–61. Available from: https://www.tandfonline.com/doi/abs/10.1080/00173134.2013.850532

18. Cambronero-Heinrichs JC, Battisti A, Biedermann PHW, Cavaletto G, Castro-Gutierrez V, Favaro L, et al. Erwiniaceae bacteria play defensive and nutritional roles in two widespread ambrosia beetles. FEMS Microbiol Ecol [Internet]. 2023 Nov 13 [cited 2024 Jul 4];99(12):1–11. Available from: https://dx.doi.org/10.1093/femsec/fiad144

19. Requier F, Rome Q, Chiron G, Decante D, Marion S, Menard M, et al. Predation of the invasive Asian hornet affects foraging activity and survival probability of honey bees in Western Europe. J Pest Sci (2004) [Internet]. 2019 Mar 15 [cited 2024 Jul 11];92(2):567–78. Available from: https://link.springer.com/article/10.1007/s10340-018-1063-0

20. Rosenkranz P, Aumeier P, Ziegelmann B. Biology and control of Varroa destructor. J Invertebr Pathol. 2010 Jan 1;103(SUPPL. 1):S96–119.

21. Kwadha CA, Ong'amo GO, Ndegwa PN, Raina SK, Fombong AT, Stout MJ, et al. The Biology and Control of the Greater Wax Moth, Galleria mellonella. Insects 2017, Vol 8, Page 61 [Internet]. 2017 Jun 9 [cited 2024 Jul 9];8(2):61. Available from: https://www.mdpi.com/2075-

4450/8/2/61/htm

22. Claing G, Dubreuil P, Bernier M, Ferland J, L'homme Y, Rodriguez E, et al. Prevalence of pathogens in honey bee colonies and association with clinical signs in southwestern Quebec, Canada. Can J Vet Res [Internet]. 2024 Apr 1 [cited 2024 Jul 8];88(2):45. Available from: /pmc/articles/PMC11000428/

23. Hansen H, Brødsgaard CJ. American foulbrood: a review of its biology, diagnosis and control. Bee World [Internet]. 1999 [cited 2024 Jul 9];80(1):5–23. Available from: https://www.tandfonline.com/doi/abs/10.1080/0005772X.1999.11099415

24. Naudi S, Šteiselis J, Jürison M, Raimets R, Tummeleht L, Praakle K, et al. Variation in the Distribution of Nosema Species in Honeybees (Apis mellifera Linnaeus) between the Neighboring Countries Estonia and Latvia. Vet Sci 2021, Vol 8, Page 58 [Internet]. 2021 Apr 1 [cited 2024 Jul 31];8(4):58. Available from: https://www.mdpi.com/2306-7381/8/4/58/htm

25. Pacini A, Mira A, Molineri A, Giacobino A, Bulacio Cagnolo N, Aignasse A, et al. Distribution and prevalence of Nosema apis and N. ceranae in temperate and subtropical eco-regions of Argentina. J Invertebr Pathol. 2016 Nov 1;141:34–7.

26. Papini R, Mancianti F, Canovai R, Cosci F, Rocchigiani G, Benelli G, et al. Prevalence of the microsporidian Nosema ceranae in honeybee (Apis mellifera) apiaries in Central Italy. Saudi J Biol Sci [Internet]. 2017 Jul 1 [cited 2024 Jul 31];24(5):979–82. Available from: https://pubmed.ncbi.nlm.nih.gov/28663691/

27. Klee J, Besana AM, Genersch E, Gisder S, Nanetti A, Tam DQ, et al. Widespread dispersal of the microsporidian Nosema ceranae, an emergent pathogen of the western honey bee, Apis mellifera. J Invertebr Pathol. 2007 Sep 1;96(1):1–10.

28. Sinpoo C, Paxton RJ, Disayathanoowat T, Krongdang S, Chantawannakul P. Impact of Nosema ceranae and Nosema apis on individual worker bees of the two host species (Apis cerana and Apis mellifera) and regulation of host immune response. J Insect Physiol [Internet]. 2018 Feb 1

[cited 2024 Jul 31];105:1–8. Available from: https://pubmed.ncbi.nlm.nih.gov/29289505/

29. Mahefarisoa KL, Simon Delso N, Zaninotto V, Colin ME, Bonmatin JM. The threat of veterinary medicinal products and biocides on pollinators: A One Health perspective. One Heal. 2021 Jun 1;12:100237.

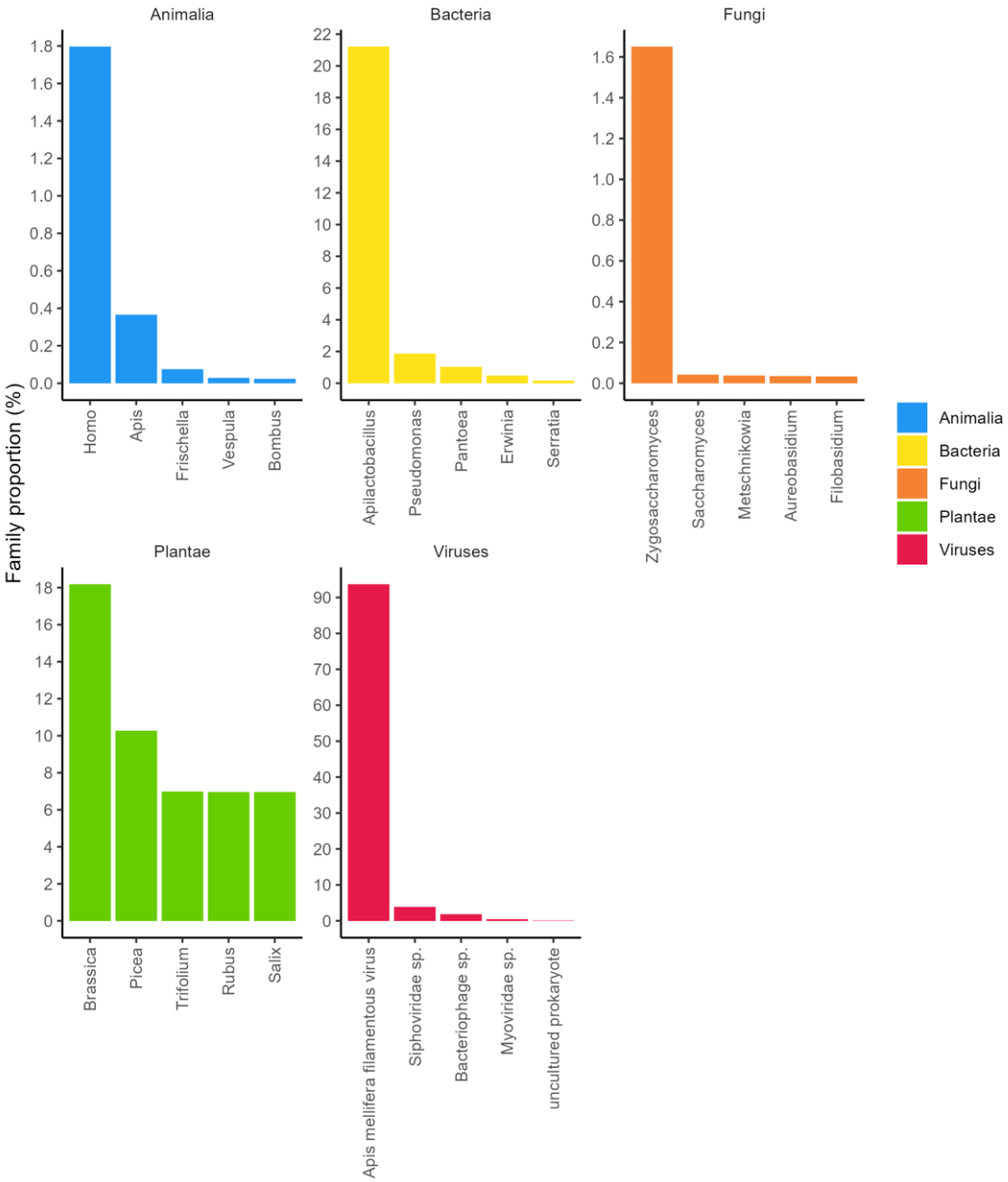
30. Marvin GE. The Occurrence and Characteristics of Certain Yeasts Found in Fermented Honey. J Econ Entomol [Internet]. 1928 Apr 1 [cited 2024 Jul 9];21(2):363–70. Available from: https://dx.doi.org/10.1093/jee/21.2.363

31. Hartmann U, Forsgren E, Charrière JD, Neumann P, Gauthier L. Dynamics of Apis mellifera Filamentous Virus (AmFV) Infections in Honey Bees and Relationships with Other Parasites. Viruses 2015, Vol 7, Pages 2654-2667 [Internet]. 2015 May 22 [cited 2024 Jul 8];7(5):2654–67. Available from: https://www.mdpi.com/1999-4915/7/5/2654/htm

32. Salonen A, Ollikka T, Grönlund E, Ruottinen L, Julkunen-Tiitto R. Pollen analyses of honey from Finland. Grana [Internet]. 2009 Dec [cited 2024 Jul 9];48(4):281–9. Available from: https://www.tandfonline.com/doi/abs/10.1080/00173130903363550

33. Lu J, Rincon N, Wood DE, Breitwieser FP, Pockrandt C, Langmead B, et al. Metagenome analysis using the Kraken software suite. Nat Protoc 2022 1712 [Internet]. 2022 Sep 28 [cited 2024 Jul 8];17(12):2815–39. Available from: https://www.nature.com/articles/s41596-022-00738-y

Supporting information



S1 Fig. Common genera of Bacteria, Fungi, Animalia, Plantae, and Viruses from the DNA of Estonian honey.