Genetics and Molecular Biology, 46, 3(suppl 1), e20230120 (2023)
Copyright © Sociedade Brasileira de Genética.
DOI: https://doi.org/10.1590/1678-4685-GMB-2023-0120

Short Communication
60 years of the PPGBM UFRGS – Special Issue

Tailed bacteriophages (Caudoviricetes) dominate the microbiome of a diseased stingless bee

Lilian Caesar[1] and Karen Luisa Haag[2]

[1]*Indiana University Bloomington, Department of Biology, Bloomington, IN, USA.*
[2]*Universidade Federal do Rio Grande do Sul, Departamento de Genética, Programa de Pós-Graduação em Genética e Biologia Molecular, Porto Alegre, RS, Brazil.*

Abstract

Bacteriophages, viruses that infect bacterial hosts, are known to rule the dynamics and diversity of bacterial populations in a number of ecosystems. Bacterial communities residing in the gut of animals, known as the gut microbiome, have revolutionized our understanding of many diseases. However, the gut phageome, while of apparent importance in this context, remains an underexplored area of research. Here we identify for the first time genomic sequences from tailed viruses (Caudoviricetes) that are associated with the microbiome of stingless bees (*Melipona quadrifasciata*). Both DNA and RNA were extracted from virus particles isolated from healthy and diseased forager bees, the latter showing symptoms from an annual syndrome that only affects *M. quadrifasciata*. Viral contigs from previously sequenced metagenomes of healthy and diseased forager bees were used for the analyses. Using conserved proteins deduced from their genomes, we found that Caudoviricetes were only present in the worker bee gut microbiome from diseased stingless bees. The most abundant phages are phylogenetically related to phages that infect Gram-positive bacteria from the order Lactobacillales and Gram-negative bacteria from the genus *Gilliamella* and *Bartonella*, that are common honey bee symbionts. The potential implication of these viruses in the *M. quadrifasciata* syndrome is discussed.

Keywords: Metagenome, stingless bee, bacteriophage, virus, disease.

Received: April 28, 2023; Accepted: December 06, 2023.

Metagenomics revealed that viruses are the most abundant biological entities on Earth (Edwards and Rohwer 2005). Most of these viruses are bacteriophages, predators of bacteria, and archaeal viruses. It is estimated that every 48 hours half of all bacteria on Earth are killed by them (Shkoporov *et al*., 2022). Bacteriophages have fundamental roles in niches as distinct as the oceans' sediments, where they participate in key biogeochemical processes, such as nutrient cycling (Fuhrman, 1999), and the gut microbiome, where they control bacterial densities (Townsend *et al*., 2021). In the human gut, phages interact directly with their prey, and indirectly with the human immune system, being major players in human health and disease (Seo and Kweon, 2019). Honey bees are models for the study of microbiomes (Zheng *et al*., 2018) and their gut 'phageome' has been recently characterized focusing on the diversity, host range and functional potential of bacteriophages (Bonilla-Rosso *et al*., 2020; Deboutte *et al*., 2020; Busby *et al*., 2022). It was found that most phages in the honey bee gut are virulent (Busby *et al*., 2022), viruses that do not insert themselves in bacterial genomes in the form of prophages, but instead hijack their host cells and use their resources to make new phages, causing the cell to lyse and die in the process. There is clear evidence that not all bacteria are infected by all phages, and that most phages can only infect a subset of bacterial species, *i.e.,* they show host specificity (Koskella and Meaden, 2013). Thus, phages are modulators of the gut microbiota, which is essential for bee development, pollen digestion and immunity (Deboutte *et al*., 2020).

In recent years we have been investigating the composition and dynamics of the bacterial and fungal gut communities of a stingless bee, *Melipona quadrifasciata* Lepeletier 1836. Some microbiome changes are associated with an annual syndrome that often leads the colony to collapse (Díaz *et al*., 2017; Haag *et al*., 2023). We also used metagenomics to identify differences in the virome composition of healthy and diseased stingless bees. Our data allowed us to characterize seven novel viruses with the potential of causing the neurological symptoms that we observed in some of the diseased colonies, but none of them was consistently associated with the disease outbreaks (Caesar *et al*., 2019). Gene expression studies on diseased *vs.* healthy *M. quadrifasciata*, rather suggest that the underlying causes of the annual syndrome are multifactorial, and involve a weakening process of the bee colony that culminates in March when outbreaks occur synchronously in different regions of southern Brazil (Caesar *et al*., 2022). Between January and March forager bees lose weight and bacterial counts increase in the gut, suggesting a relaxation in the bee immunologic mechanisms that regulate bacterial growth (Haag *et al*., 2023). Here we use the metagenomic data generated in our previous virome characterization (Caesar *et al*., 2019) to identify bacteriophages that may participate in the process of modulating the *M. quadrifasciata* gut microbiota.

Our analyses were run on four assemblies built from sequences generated by high throughput sequencing of virus particles separated by centrifugation (Caesar *et al*., 2019;

Send correspondence to Karen Luisa Haag. Universidade Federal do Rio Grande do Sul, Departamento de Genética, Programa de Pós-Graduação em Genética e Biologia Molecular, Av. Bento Gonçalves, 9500, 91501-970, Porto Alegre, RS, Brazil. E-mail: karen.haag@ufrgs.br

BioProject PRJNA960650). Two viral contig assemblies (DNA and RNA) came from a single pooled sample of worker bees from a diseased hive of *M. quadrifasciata*. The other two assemblies (DNA and RNA) are from a healthy hive of *M. quadrifasciata* sampled simultaneously at the same meliponary. For the identification of bacteriophage contigs we used the four sets as inputs on two softwares, geNomad (Camargo *et al*., 2023) and VirSorter2 (Guo *et al*., 2021). GeNomad was run end-to-end, using the flags *--cleanup --splits* 8 and the default virus score cutoff of >0.7. VirSorter2 was run with the flags *--min-score 0.5 --keep-original-seq --hallmark-required-on-short*. To confirm the identification of phages and check the quality of metagenome-assembled genomes (MAGs) we used CheckV (Nayfach *et al*., 2021) with default commands. All identified phage contigs were retained for the following analyses since they had 0 host genes, and the majority of them had at least 1 viral hallmark gene.

Using both geNomad and VirSorter a total of 193 viral contigs were identified as phages, and went through the downstream analyses. Within the four assemblies, bacteriophages were only found among the unhealthy DNA and RNA contigs (n=64 and 129, respectively; Table 1). GeNomad also performs gene prediction with a modified version of Prodigal (Hyatt *et al*., 2010) called prodigal-gv, and assigns the predicted proteins to geNomad's markers using MMseqs2 (Steinegger and Söding, 2017). The vast majority of the annotations of predicted genes from both datasets (DNA and RNA) were assigned to bacteriophage markers belonging to the Caudoviricetes (Tables 1 and S1), a class of virulent phages known as the tailed phages, which currently contains the majority of the total phage sequences in public databases (Zhu *et al*., 2022).

Next, contigs on each sample were de-replicated with dRep (Olm *et al*., 2017), using flags *--P_ani 0.95 --cov_thresh 0.85*, and coverage estimated by mapping the original reads against them with Bowtie2 (Langmead and Salzberg 2012), using the flag *--no-discordant* and *--very-sensitive.* The database for mapping reads included also non-phage contigs from the assembly to avoid non-specific or low quality reads mapping against the phage sequences. Coverage bam files were used as inputs along with the contigs for binning the phage genomes with vRhyme, using default commands (Kieft *et al*., 2022). Binned and non-binned sequences from both samples were then de-replicated with dRep using the same parameters that follow the standard thresholds for vOTU classification (Roux *et al*., 2019). Coverage of vOTUs in each sample was recovered again with Bowtie2 and command *depth -a* from Samtools (Li *et al*., 2009). Coverage and completeness estimates for the 182 vOTUs from both assemblies (DNA and RNA) are listed in Table S2. Coverage estimates ranged between 0.05 – 2704.15X (DNA) and 1.28 – 3134.75X (RNA). Completeness varied between 0.2 – 47.95% and 0.11 – 45.29%, for vOTUs from the DNA and RNA assemblies, respectively (Table 1). Thus, with the sequencing effort employed in our study, we assembled incomplete genomes, probably from the most abundant bacteriophages found in the gut of our diseased stingless bees (Figure 1).

We attempted to predict phage taxa and their putative hosts based on MAGs similarity to known phages, and on their match to CRISPR-spacers. First, proteins encoded by all phage vOTUs were predicted and primarily annotated with Prokka (Seemann, 2014), using flag -- *kingdom Viruses.* Proteins and contigs from the 5 largest vOTUs were used together with previously described bee phage sequences (Bonilla-Rosso *et al*., 2020; Deboutte *et al*., 2020; Busby *et al*., 2022) as inputs for vConTACT2 (Bin Jang *et al*., 2019). The program vConTACT2 was ran with default commands, using the 'ProkaryoticViralRefSeq211-Merged' database, and results were visualized with the R package Igraph. Only vOTU1 and vOTU3 could be connected to previously described phages, but with low confidence and none of the vOTUs clustered with any phage (Figure 2). Additionally, we used all phage contigs as queries in BLASTn, using flags *-evalue 1e-3 -ungapped -perc_identity 95,* against three CRISPR-spacer databases: CrisprOpenDB (Dion *et al*., 2021), spacers from honey bee microbiome-associated bacteria (Bonilla-Rosso *et al*., 2020), and spacers from stingless bees' microbiome-associated bacteria. The latter database was built based on spacers that we predicted in complete genomes of bacteria sequenced from six species of stingless bees (Sarton-Lohéac *et al*., 2023) using the 'CRISPRCasFinder' online tool, with default settings, and evidence level 3 or 4. Unfortunately, none of our vOTUs have matches with CRISPR spacers from any of the three searched databases. We think that traditional approaches to identify viral hosts that are based on genomic data from bacteria and their phages are of limited use in the context of our study, since there are so far only a few characterized genomes from stingless bee microbiomes (*e.g*. Sarton-Lohéac *et al*., 2023).

In a further attempt to infer the putative bacteria used as hosts by the dominant tailed phages inferred from our study (vOTUs 2, 5, 61, 62 and 65; see Figure 1 and Table S2) we used phylogenetic analyses. These MAGs showed the highest

Table 1 – Phage distribution, taxonomy and abundance among the four metagenomes analyzed in the present study.

	Healthy		Diseased	
	DNA	RNA	DNA	RNA
Total number of viral contigs[a]	3	15	343	1559
Number of phage contigs	0	0	64	129
Number of vOTUs	0	0	56	127
Phage taxonomy	–	–	Caudoviricetes	Caudoviricetes
Phage vOTU completeness	–	–	0.2 – 47.95 %	0.11 – 45.29 %
Phage vOTU coverage	–	–	0.05 – 2,704.15 X	1.28 – 3,134.75 X

[a]Caesar *et al*., 2019.

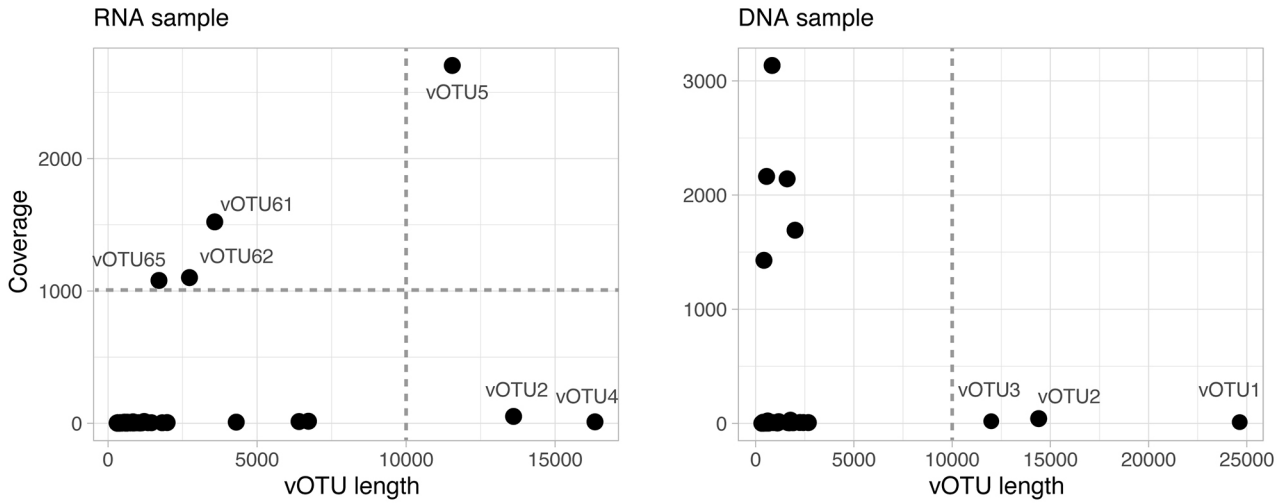

Figure 1 – Coverage estimates for vOTUs of variable lengths (bp) assembled from the unhealthy *M. quadrifasciata* DNA and RNA samples. Each dot on the chart corresponds to a phage MAG. Horizontal and vertical dashed lines mark the coverage and length thresholds, respectively, to keep vOTUs for the downstream analyses (see text for details).

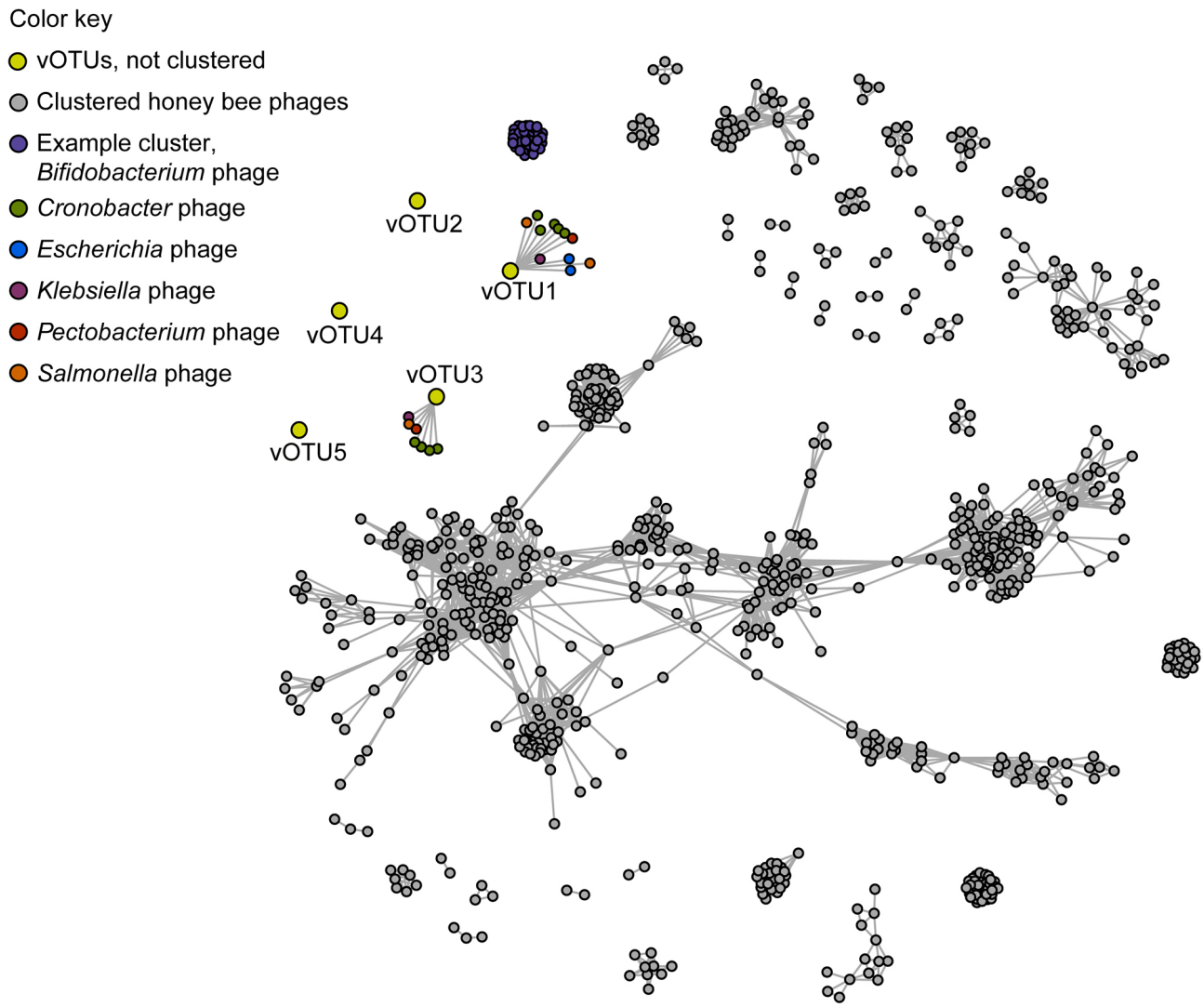

Figure 2 – Protein-sharing network displaying unclustered *M. quadrifasciata* phages and clustered phages of *Apis mellifera*. Nodes represent phage partial to complete genomes, and edges connecting them indicate a statistically significant similar protein profile between their genomes. No vOTU >10 kb clustered with known phages in the database, and only two had weak connections with known phages.

coverage in the RNA metagenome, which we interpreted as a proxy for their abundance in the unhealthy bee microbiome. Moreover, these MAGs contain hallmark genes that could be used for BLAST searches of related sequences. For vOTUs 2, 5 and 65 we used a gene encoding the DNA encapsidation protein (terminase) as phylogenetic marker. The protein deduced from vOTUs 5 and 65 was identical. For vOTUs 61 and 62 we used the predicted major capsid protein, which was also identical in both MAGs. Briefly, we selected reference protein sequences from GenBank (nr database) using BLASTp and aligned them to the corresponding sequences predicted from the vOTUs with MAFFT (Katoh and Standley, 2013). The alignment was used as input for phylogenetic analyses with PHYML (Guindon and Gascuel, 2003) implemented in Geneious Prime 2022.1.1 (Biomatters Ltd.) using the LG evolutionary model and 500 bootstrap replicates. Whereas vOTUs 2, 61 and 62 seem to be related to tailed viruses isolated from Gram-negative bacteria, vOTUs 5 and 65 cluster with virus sequences obtained from Gram-positive bacteria belonging to the order Lactobacillales (Figure 3). Viral OTU2, which appeared both in the DNA and RNA metagenomes, clusters with phages from *Bartonella*, a facultative symbiont of the honey bee, and vOTUs 61 and 62 are related to a *Gilliamella* phage isolated from bumble bees in China. Proteobacteria such as *Bartonella* and *Gilliamella* are not as abundant in the *M. quadrifasciata* microbiome as yet unidentified Acetobacteraceae, in contrast to the Lactobacillales, that are highly abundant and diverse (Díaz *et al*., 2017; Cerqueira *et al*., 2021; Haag *et al*., 2023; Sarton-Lohéac *et al*., 2023).

Surprisingly, none of the sequences from our phageomes could be classified as the same phages previously characterized in honey bees (Bonilla-Rosso *et al*., 2020; Deboutte *et al*., 2020; Busby *et al*., 2022). Nonetheless, Caudoviricetes were also the most abundant phages identified in the honey bee studies, which were not designed to assess matters related to bee health. Our study compares the phageomes of stingless bees from the same species sampled simultaneously from the same location, but differing in their health status. We only detected Caudoviricetes phages in diseased bees, leading us to reason that these phages might be implicated in their health. One hypothesis is that the virulent tailed phages opportunistically become more abundant in diseased *M. quadrifasciata* due to an overall increase in the amount of bacteria in the gut. Indeed, in our previous study we found that bacterial counts estimated by 16S qPCR increased between January and March, but we did not find significant differences between diseased and healthy bees (Haag *et al*., 2023). An alternative but not mutually exclusive hypothesis is that, due to host specificity, tailed viruses disrupt the equilibrium of the bacterial community, and that microbiome dysbiosis negatively affects host homeostasis. We did find in previous studies of the *M. quadrifasciata* microbiome based on metabarcoding that the Lactobacillales reached their lowest relative abundance during the outbreak period, but again there were no significant differences between diseased and healthy bees (Haag *et al*., 2023).

It is possible that the dynamics of the microbiome, including bacteriophages, have a longer-term effect on bee health that we did not bring to light by looking at a single outbreak. Indeed, by studying two successive outbreaks (2014 and 2015), we showed a significant interaction effect of sampling year and health status on microbiome composition (Díaz *et al*., 2017), implying that the microbiome profile of diseased bees is not the same for every outbreak. To obtain a more realistic view about the role of bacteriophages in the *M. quadrifasciata* microbiome, and its relationship with the annual syndrome, we need to work with deeper sequencing efforts. It would be possible to use a metagenomic strategy to characterize the phageomes and bacteriomes simultaneously in successive years. Nevertheless, to our knowledge, this is the first report of bacteriophages from a stingless bee microbiome. We expect that, in the near future, with proper sampling and deep sequencing, we will be able to clarify how phages control bacterial densities in the *M. quadrifasciata* gut and ultimately influence their health.

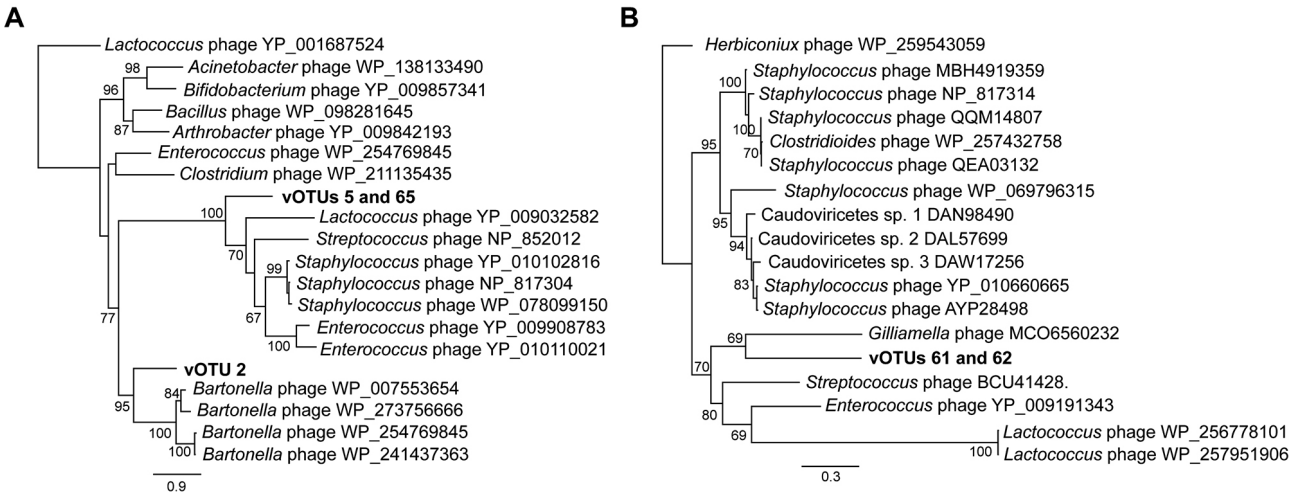

Figure 3 – Maximum Likelihood phylogenies obtained using DNA encapsidation (A) and major capsid (B) proteins (see text for details). Proteins predicted from vOTUs 5 and 65, as well as vOTUs 61 and 62 were identical, and therefore correspond to a single branch on their respective phylogenetic trees.

Acknowledgements

Our work is being supported by Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul FAPERGS PQG 19/2551-0001860-6.

Conflict of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicial to the impartiality of the reported research.

Authors Contributions

LC and KLH conceived and the study, analyzed the data and wrote the manuscript.

References

Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, Brister JR, Kropinski AM, Krupovic M, Lavigne R *et al*. (2019) Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. Nat Biotechnol 37:632–639.

Bonilla-Rosso G, Steiner T, Wichmann F, Bexkens E and Engel P (2020) Honey bees harbor a diverse gut virome engaging in nested strain-level interactions with the microbiota. Proc Natl Acad Sci U S A 117:7355–7362.

Busby TJ, Miller CR, Moran NA and Van Leuven JT (2022) Global composition of the bacteriophage community in honey bees. mSystems 7:e01195-21.

Caesar L, Cibulski SP, Canal CW, Blochtein B, Sattler A and Haag KL (2019) The virome of an endangered stingless bee suffering from annual mortality in southern Brazil. J Gen Virol 100:1153–1164.

Caesar L, Lopes AMC, Radaeski JN, Bauermann SG, Konzen ER, Pombert J-F, Sattler A, Blochtein B, Carvalho AT and Haag KL (2022) Longitudinal survey reveals delayed effects of low gene expression on stingless bee colony health. J Apic Res 61:654–663.

Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, Hu B, Chain PSG, Nayfach S and Kyrpides NC (2023) You can move, but you can't hide: Identification of mobile genetic elements with geNomad. bioRXiv:2023.03.05.531206.

Cerqueira AES, Hammer TJ, Moran NA, Santana WC, Kasuya MCM and da Silva CC (2021) Extinction of anciently associated gut bacterial symbionts in a clade of stingless bees. ISME J 15:2813–2816.

Deboutte W, Beller L, Yinda CK, Maes P, de Graaf DC and Matthijnssens J (2020) Honey-bee–associated prokaryotic viral communities reveal wide viral diversity and a profound metabolic coding potential. Proc Natl Acad Sci U S A 117:10511–10519.

Díaz S, de Souza Urbano S, Caesar L, Blochtein B, Sattler A, Zuge V and Haag KL (2017) Report on the microbiota of *Melipona quadrifasciata* affected by a recurrent disease. J Invertebr Pathol 143:35–39.

Dion MB, Plante P-L, Zufferey E, Shah SA, Corbeil J and Moineau S (2021) Streamlining CRISPR spacer-based bacterial host predictions to decipher the viral dark matter. Nucleic Acids Res 49:3127–3138.

Edwards RA and Rohwer F (2005) Viral metagenomics. Nat Rev Microbiol 3:504–510.

Fuhrman JA (1999) Marine viruses and their biogeochemical and ecological effects. Nature 399:541–548.

Guindon S and Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704.

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

Haag KL, Caesar L, da Silveira Regueira-Neto M, de Sousa DR, Montenegro Marcelino V, de Queiroz Balbino V and Torres Carvalho A (2023) Temporal changes in gut microbiota composition and pollen diet associated with colony weakness of a stingless bee. Microb Ecol 85:1514–1526.

Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW and Hauser LJ (2010) Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119.

Katoh K and Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780.

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

Koskella B and Meaden S (2013) Understanding bacteriophage specificity in natural microbial communities. Viruses 5:806–823.

Langmead B and Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R and 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

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

Olm MR, Brown CT, Brooks B and 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.

Roux S, Adriaenssens EM, Dutilh BE, Koonin EV, Kropinski AM, Krupovic M, Kuhn JH, Lavigne R, Brister JR, Varsani A *et al*. (2019) Minimum information about an uncultivated virus genome (MIUViG). Nat Biotechnol 37:29–37.

Sarton-Lohéac G, Nunes da Silva CG, Mazel F, Baud G, de Bakker V, Das S, El Chazli Y, Ellegaard K, Garcia-Garcera M, Glover N *et al*. (2023) Deep divergence and genomic diversification of gut symbionts of Neotropical stingless bees. mBio 14:e0353822.

Seemann T (2014) Prokka: Rapid prokaryotic genome annotation. Bioinforma Oxf Engl 30:2068–2069.

Seo S-U and Kweon M-N (2019) Virome–host interactions in intestinal health and disease. Curr Opin Virol 37:63–71.

Shkoporov AN, Turkington CJ and Hill C (2022) Mutualistic interplay between bacteriophages and bacteria in the human gut. Nat Rev Microbiol 20:737–749.

Steinegger M and Söding J (2017) MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol 35:1026–1028.

Townsend EM, Kelly L, Muscatt G, Box JD, Hargraves N, Lilley D and Jameson E (2021) The Human Gut Phageome: Origins and roles in the human gut microbiome. Front Cell Infect Microbiol 11:643214.

Zheng H, Steele MI, Leonard SP, Motta EVS and Moran NA (2018) Honey bees as models for gut microbiota research. Lab Anim 47:317–325.

Zhu Y, Shang J, Peng C and Sun Y (2022) Phage family classification under Caudoviricetes: A review of current tools using the latest ICTV classification framework. Front Microbiol 13:1032186.

Internet Resources

BioProject: Raw virome data, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA960650/ (accessed 3 January 2024)

Igraph for R, https://zenodo.org/record/824064. (accessed 10 September 2023)

Supplementary material

The following online material is available for this article:

Table S1 – Summary of DNA and RNA contig taxonomy.

Table S2 – Metagenome coverage and contig length.

Associate Editor: Loreta Brandão de Freitas

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License (type CC-BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original article is properly cited.