

RESEARCH

Open Access



Targeted viromes and total metagenomes capture distinct components of bee gut phage communities

Dino Lorenzo Sbardellati^{1*} and Rachel Lee Vannette¹

Abstract

Background Despite being among the most abundant biological entities on earth, bacteriophage (phage) remain an understudied component of host-associated systems. One limitation to studying host-associated phage is the lack of consensus on methods for sampling phage communities. Here, we compare paired total metagenomes and viral size fraction metagenomes (viromes) as methods for investigating the dsDNA viral communities associated with the GI tract of two bee species: the European honey bee *Apis mellifera* and the eastern bumble bee *Bombus impatiens*.

Results We find that viromes successfully enriched for phage, thereby increasing phage recovery, but only in honey bees. In contrast, for bumble bees, total metagenomes recovered greater phage diversity. Across both bee species, viromes better sampled low occupancy phage, while total metagenomes were biased towards sampling temperate phage. Additionally, many of the phage captured by total metagenomes were absent altogether from viromes. Comparing between bees, we show that phage communities in commercially reared bumble bees are significantly reduced in diversity compared to honey bees, likely reflecting differences in bacterial titer and diversity. In a broader context, these results highlight the complementary nature of total metagenomes and targeted viromes, especially when applied to host-associated environments.

Conclusions Overall, we suggest that studies interested in assessing total communities of host-associated phage should consider using both approaches. However, given the constraints of virome sampling, total metagenomes may serve to sample phage communities with the understanding that they will preferentially sample dominant and temperate phage.

Keywords Bacteriophage, Bee microbiome, Host-associated microbiome, Microbial ecology, Viral ecology

Background

Bacteriophage (phage) are the viruses which target bacteria. Phage are hypothesized to be among the most abundant biological entities on earth [1–3] and to be important modulators of microbial communities [2, 4]. In

marine and soil environments, where these prokaryotic viruses are relatively well studied, phage shape the microbiomes they infect by lysing dominant bacteria [5–7], influencing nutrient turnover and cycling [8, 9], and vectoring functional genes between bacterial hosts [10–13]. While the basis for much of our understanding of phage ecology stems from the study of marine and soil environments, there is a growing interest in understanding the role of phage in host-associated systems [14, 15]. However, host-associated environments and free-living soil and marine environments represent different microbial

*Correspondence:

Dino Lorenzo Sbardellati
dlsbardellati@ucdavis.edu

¹ Department of Entomology and Nematology, University of California Davis, Davis, CA, USA



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

ecosystems with different selective pressures. As such, insights gained from studying free-living phage communities may not always apply to communities of host-associated phage. Specifically, while the best methods for sampling free-living phage communities have been empirically compared [16, 17], little work has evaluated the best methods for sampling host-associated phage [18]. Moreover, given the importance of microbiomes to host health, understanding phage community dynamics and how phage differ between related animal hosts is a key priority.

Like their environmental counterparts, phage infecting host-associated bacteria (hereafter “host-associated” phage) can affect the abundance, composition, and function of host-derived microbial communities. However, unlike free-living phage, host-associated phage must also adapt to their close association with a host and its microbiome. For example, the comparatively high-density bacterial communities associated with the human gut [19, 20], mucosal layer of corals [21], and other host environments are suggested to favor integrase encoding temperate phage, as opposed to obligately lytic phage [22–24]. Phage targeting bacteria adhered to gut mucosa have also been shown to evolve the ability to adhere to and persist in the animal mucosa where their target bacteria reside [25]. Given the fundamental differences in the ecology of free-living and host-associated phage communities [22], it remains unclear whether sampling methods developed for free-living communities are also appropriate for other habitat types.

Two methods are frequently used for describing dsDNA viral communities: bioinformatic mining of total metagenomes and targeted viral enrichments (viromes). In bioinformatic mining of total metagenomes, total genomic DNA is extracted from a sample, amplified when necessary, sequenced using “shotgun” or untargeted methods, assembled into metagenomes, and phage computationally mined [26, 27]. While this approach is easy to perform and offers simultaneous characterization of all DNA in a sample (including bacteria), only a small minority of the total metagenomic data generated will originate from phage, producing a relatively shallow sampling of phage communities. An alternative method, targeted virome sequencing, leverages the physical characteristics of viruses to select for the phage and virus-like particle fraction of a sample prior to nucleic acid extraction and sequencing [28–30]. With more sequencing space devoted to viruses, targeted viromes can recover a greater diversity of phage relative to total metagenomes. This is illustrated by previous comparisons of viromes and total metagenomes in soil and freshwater samples [16, 17]. However, whether viromes always outperform total metagenomes is not clear. Work applying these two

methods to host-associated (human gut) and low-biomass (deep-sea marine) samples has produced conflicting results [17, 18]. One explanation for these discrepancies is the biases associated with total metagenomes and targeted viromes [31, 32]. For example, viromes remove bacterial cells prior to sequencing and, as a result, may select against integrated, non-replicating, temperate phage [33, 34]. This is especially relevant in host-associated systems where temperate phage are abundant [33]. Total metagenomes can also be preferable to viromes logistically. Viromes typically demand a relatively large amount of biomass to generate enough nucleic acids for sequencing. While protocols, such as the NetoVIR protocol [30], have successfully been used to circumvent this limitation, it is important to note that the random amplification used in these approaches can introduce another form of bias. Moreover, the pooling of sample material from multiple individuals used in this and other approaches can obscure inter-individual variation in viral communities.

Insects have become valuable models for exploring host-microbe interactions [35–37]. Social bees in particular house a simple (5–9 taxa), highly conserved, socially transmitted, gut bacterial community [38–41] which contributes to host nutrient acquisition and pathogen defense [42–44]. These features position social bees as an excellent model for studying not only host-microbe interactions [45, 46] but tripartite host-microbe-phage interactions as well [47]. A series of recent studies using targeted virome approaches have shown that honey bees host a diversity of novel phage which target core bee gut bacteria [48–50]. While this work has advanced our understanding of bee phage, no work has evaluated how sampling method influences recovered bee phage communities. Given the large biomass required for viromes, total metagenomes may be advantageous for surveying the phage communities associated with bees and other biomass limited systems. Additionally, describing the phage associated with other social bees will develop our understanding of how phage communities differ between related hosts. Lastly, broader work comparing the role and diversity of phage among small invertebrate hosts will hinge on appropriate sampling methods, so comparing the performance of viromes and total metagenomes is particularly important for enabling future comparative study.

Here, we evaluate the phage communities inferred by applying total metagenomes and targeted viromes to managed *Apis mellifera* and commercially raised *Bombus impatiens* gut material. We hypothesize that targeted viromes will enrich for phage sequences and capture a greater diversity of phage, relative to bioinformatic mining of total metagenomes, and that phage will differ between host bee species. To interrogate these

hypotheses, we first examine how host bee species and sampling method impact sequencing read data and viral enrichment. We then validate our sampling and computational methodology through rarefaction and by comparing our phage sequences to those previously described in honey bees. Next, to test how phage community differs across host bee and sampling methods, we compare phage community diversity, structure, and composition. We then use gene content, occupancy-abundance plots, qPCR, and bacterial community profiling to delve into what contributes to the apparent differences in phage communities. Overall, we find that honey bees host more phage than do bumble bees, but that the method which captured the most phage differed by bee species. This suggests that viromes and total metagenomes are each biased in their own way and appear to sample different populations of phage. As a result, we propose these techniques are complementary in describing the full diversity of host-associated phage.

Methods

Bee rearing, sampling, and experimental design

A detailed description of our methods (including sample collection and bench and computational work) is provided in the supplemental materials. Briefly, bees were harvested from three colonies of managed honey bees (*Apis mellifera*) and three colonies of commercially

reared bumble bees (*Bombus impatiens*, Koppert Biological Systems; Howell, MI, USA). Bumble bee colonies were maintained in the laboratory and provided with honey bee collected pollen and artificial nectar (Koppert Biological Systems) *ad libitum*, as per manufacturer’s recommendation. Honey bees were sampled from colonies maintained at the University of California, Davis, apiary and were collected from inside the colony using a handheld bee vacuum between January 30 and February 13, 2023, 8:00 am–10:00 am.

Bees were anesthetized via a 60-s CO₂ exposure and then euthanized via decapitation. Immediately following sacrifice, the mid-hindgut section was dissected from 100 bees (all collected from the same colony) and pooled for same-day virome extraction. Bees destined for total metagenomic DNA extraction were collected the same day as virome samples and were stored at –20 °C pending mid-hindgut dissection on the following day. All dissections took place in sterile PBS using ethanol and flame sterilized forceps. Each honey and bumble bee colony generated one targeted virome and three total metagenomes (Fig. 1).

Sample preparation and DNA extraction

Phage enrichments for targeted virome sequencing were carried out following a protocol adapted from previous publications [28, 29, 51] and is described fully

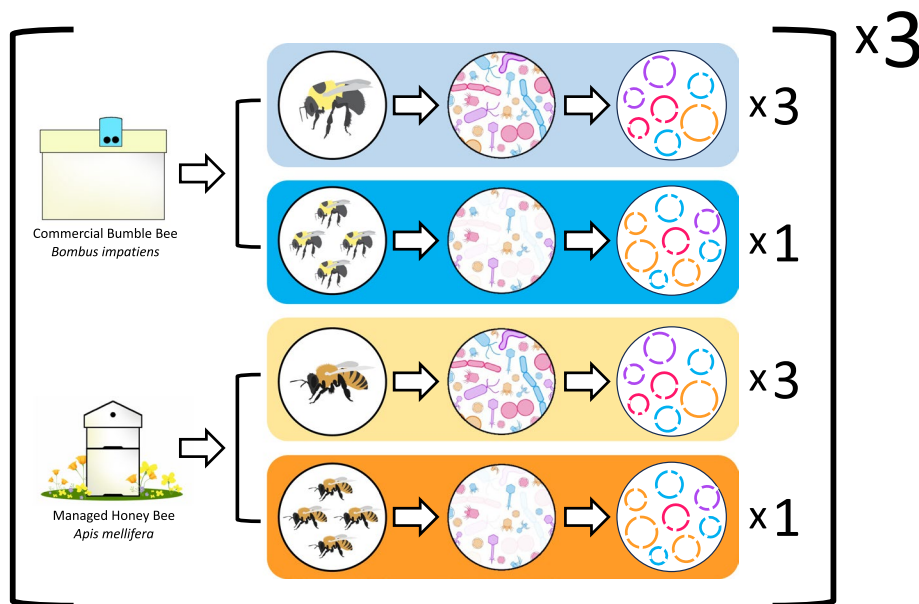


Fig. 1 Graphical representation of the sampling scheme and methods used in this research. A total of three bumble bee and three honey bee colonies were sampled. From each colony of bee, we generated three total metagenomes and one targeted virome. Total metagenomes were sampled from individual bees, while targeted viromes were produced from the pooled guts of 100 bees. This sampling resulted in nine bumble bee total metagenomes (light blue), three bumble bee viromes (dark blue), nine honey bee total metagenomes (light orange), and three honey bee viromes (dark orange)

in the supplemental methods. Briefly, bee gut pools were combined with 10 ml of protein-supplemented phosphate-buffered saline (PPBS: 2% BSA, 10% PBS, 1% K-citrate, 150-mM MgSO₄), homogenized, and submitted to a series of washes aimed at eluting phage particles. Phage were enriched via centrifugation, 0.22- μ m filtration, and ultracentrifugation. Lastly, phage pellets were resuspended in 200- μ L water and DNase treated.

For total metagenomes, mid-hind guts of individual bees were dissected and homogenized and used for extraction. DNA was extracted from both virome and total metagenome samples using a DNeasy PowerSoil Pro kit (Qiagen, Hilden, Germany) following the manufacturer's instructions with the optional Qiagen Vortex Adapter 15-min bead beating step.

Library construction and DNA sequencing

Extracted DNA from samples and kit blank-negative controls was submitted to the University of California Davis Genome Center for library prep and paired-end 150-bp shotgun metagenomic sequencing. Libraries were prepared using a KAPA HyperPrep Kit (Kapa Biosystems, Roche, Basel, Switzerland) with Illumina TruSeq adapters (Illumina, San Diego, CA, USA) and a target insert size of 350 bp. Sequencing took place on an Illumina NovaSeq. All samples were sequenced on the same run.

Read processing, viral prediction, and vOTU table construction

Briefly, all demultiplexed *.fastq.gz* files from the Davis Genome Center were trimmed and quality filtered using Trim-Galore [52] and Trimmomatic [53]. Reads aligning to negative samples, or the genomes of *A. mellifera* (GCA_000002195.1) and *B. impatiens* (GCF_000188095.3), were removed using Bowtie2 [54]. Cleaned reads were k-merized for complexity comparisons using Sourmash [55]. The program metaSPADES was used to construct assemblies. Assemblies \geq 5 kb were passed to geNomad [56] for phage prediction. Putative phage sequences predicted to be < 50% complete by CheckV [57] were dropped. Remaining sequences \geq 95% average nucleotide identity were clustered into viral operational taxonomic units (vOTUs) using dRep [58] and annotated with Prokka [59] using the PHROGs database [60]. The programs BACPHLIP [61] and geNomad were then used to predict the lifestyle and taxonomy of identified vOTUs, respectively. Finally, CoverM [62] was used to map cleaned reads against vOTUs. The resulting coverage table was normalized to coverage per million reads.

Bacterial community and density measurements

Bacterial communities were predicted directly from cleaned total metagenomic reads with Kraken2 [63] and Bracken [64] using default parameters. Bacterial copy number was quantified from the DNA extracts used for metagenome sequencing via a qPCR protocol adapted from Christensen et al. [65]. Each master mix solution contained the following: 5- μ L SsoAdvanced Universal SYBR Supermix 41 (Bio-Rad, Hercules, CA, USA), 0.3 μ L of primers (10 μ M) targeting the 16S region of the rRNA gene (799F = 5'-AACMGGATTAGATACCCCKG-3'; 1115R = 37 5'-AGGGTTGCGCTCGTTG-3'), 3.4- μ L molecular grade water, and 1 μ L of template DNA (diluted 1:1000 in molecular grade water). Reactions were performed in triplicate for each sample.

Statistical and ecological analyses

Statistical analyses were conducted in R [66] v4.2.3. Briefly, alpha diversity, beta diversity, and PERMANOVAs were compared between host and sampling method using vegan [67] and phyloseq [68]. Linear mixed effect models were built using lme4 [69], and post hoc tests were run using emmeans [70]. *T*-tests were performed using base R. Gene-sharing networks were generated outside of R using the program vConTACT2 [71, 72]. Genome alignment plots were constructed using clinker [73].

Phage-host predictions were performed using a custom CRISPR-spacer analysis. In short, we built bacterial metagenome-assembled genomes (MAGs) from our total metagenomic dataset using metaSPADES, MetaBat2 [74], and dRep. We then combined these MAGs with NCBI assemblies of common bee gut bacteria, mined all these sequences for CRISPRs using MINCED [75], and then aligned those CRISPRs to our vOTUs using Blastn [76] following the parameters suggested by Edwards et al. [77]. Lastly, to improve the number of vOTUs with bacterial host and viral taxonomic assignments, we combined our CRISPR-based phage-host assignments and geNomad taxonomic predictions with the clusters produced by vConTACT2 to infer the bacterial hosts and viral taxonomies of whole phage clusters, similar to Bonilla-Rosso et al. [48]. Bacterial 16S copy number was compared between bee species using linear mixed effect models.

Results

Bumble bee viromes produce fewer high-quality reads than honey bee viromes

Our sequencing effort generated a total of 419,884,845 raw read pairs, 2.47% (10,385,967) of which came from extraction negatives. While the number of raw read pairs did not differ significantly between samples (Fig. S1;

Bee $t_{5.40} = 0.54, p = 0.61$, Type $t_{8.47e23} = 1.58, p = 0.11$, Bee:Type $t_{1.07e23} = -1.38, p > 0.17$, there were differences in the number of high-quality read pairs remaining after quality control (Bee $t_{5.81} = 2.87, p < 0.05$, Type $t_{2.75e25} = -0.33, p = 0.74$, Bee:Type $t_{2.75e25} = 4.69, p < 0.001$) (Fig. 2A). After filtering, honey bee viromes averaged 15.56 million high-quality read pairs per sample, while bumble bee viromes averaged only 5.49. Total metagenome samples generated more similar numbers

of high-quality reads per sample, with an average of 8.57 and 5.85 million read pairs generated from honey and bumble bee samples, respectively.

Bumble bee virome reads are less complex than other samples

Next, we measured the complexity of read libraries to further investigate how bee species and sampling method influenced the data we generated. This was done by

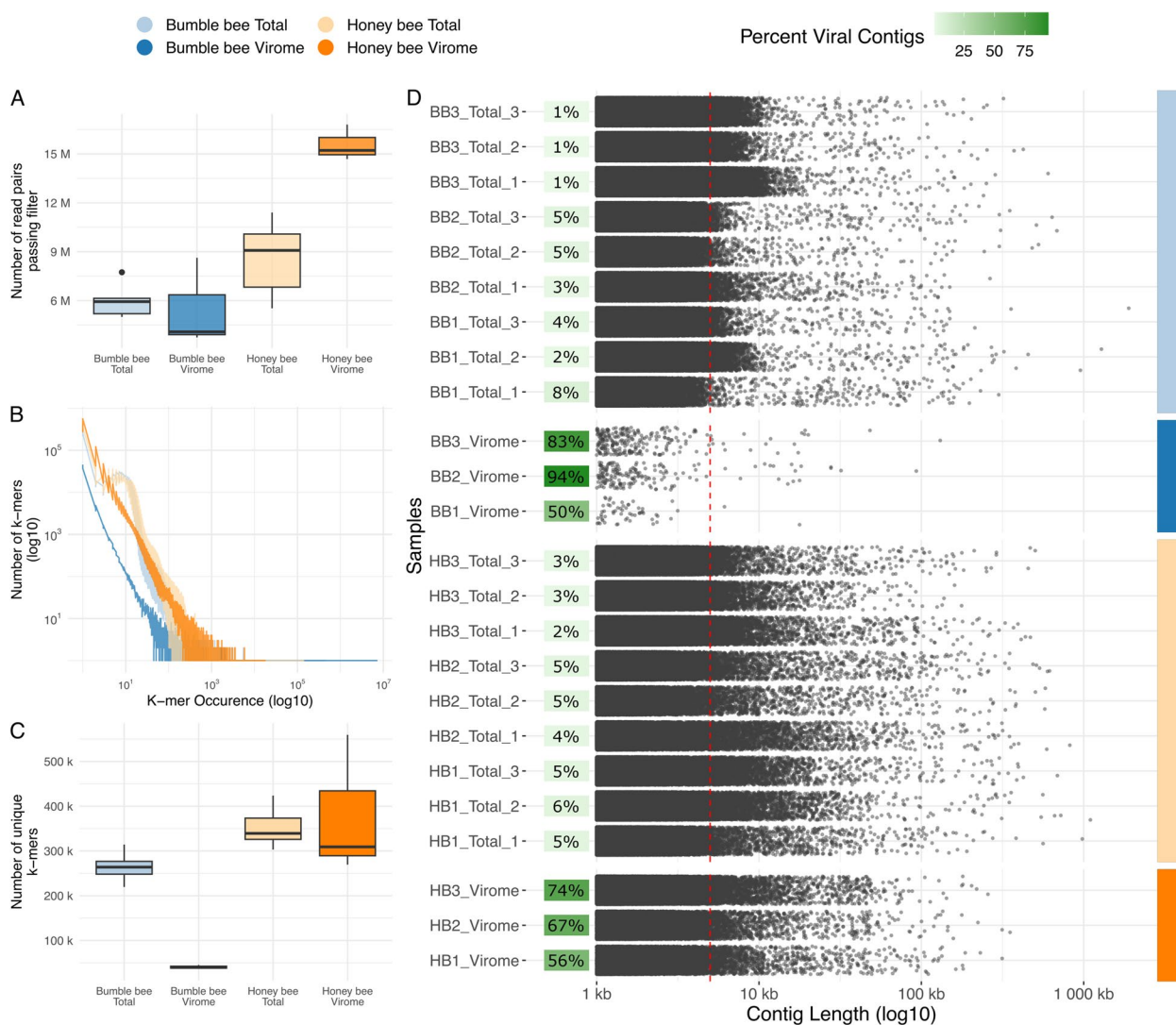


Fig. 2 Figures describing sequencing and assembly quality. Color denotes sample groups. **A** Boxplots describing sequencing depth. The x-axis shows sample groups. The y-axis displays the total number of high-quality reads produced. **B** Line plots describing sequence complexity of the reads shown in **A**, as measured by the frequency distributions of 31-bp-sized k-mers. The x-axis presents k-mer occurrence (i.e., how abundant a particular k-mer was in a given sample's read), while the y-axis shows the number of k-mers with a certain occurrence. **C** Boxplots describing the number of unique 31-bp-sized k-mers present in the read libraries of each group of samples. The x-axis shows sample groups. The y-axis shows the number of unique k-mers. **D** Jittered dot plots describing the length distributions of contigs assembled from each sample. Only contigs > 1 kbp are shown. Individual samples are shown on the y-axis. The x-axis shows contig length (\log_{10} scale). Each point represents a single contig. A dotted red line is drawn at 5 Kbp. For each sample, a green square is plotted phage enrichment (the number of phage identified, divided by number of contigs assessed for phage, times 100)

comparing the number and occurrence of 31-bp-sized k-mers among each set of high-quality read libraries (Fig. 2B). The results of this analysis suggest that across the range of k-mer occurrences, bumble bee viromes consistently contain fewer k-mers of a given occurrence than did other sample types. Moreover, viewing only the y -intercept of this plot (Fig. 2C) shows that bumble bee viromes contain a lower number of singleton k-mers, relative to all other sample types (Bee $t_{5.34} = 2.49$, $p = 0.052$, Type $t_{16} = -6.47$, $p < 0.001$, Bee:Type $t_{16} = 5.14$, $p < 0.001$). This decreased number of singleton k-mers and the tendency towards lower k-mer occurrences suggest that bumble bee viromes are less complex compared to the other sample types in this dataset.

Targeted virome assemblies are enriched in phage

To evaluate the size and number of assemblies produced from each read library, and to quantify the extent to which viromes enriched for phage sequences, we visualized the distribution of assembly sizes in each sample, as well as the percentage of assemblies in a sample predicted to encode phage sequences (Fig. 2D). A total of 4278 and 37 contigs ≥ 5 kb were assembled from honey and bumble bee viromes, respectively. After passing to geNomad, these contigs yielded 2883 honey bee and 32 bumble bee putative phage sequences. Meanwhile, honey and bumble bee total metagenomes produced 16,615 and 14,335 contigs ≥ 5 kb, respectively, which yielded 662 and 276 putative phage sequences. This means that 65.74% of honey bee and 75.82% of bumble bee virome contigs ≥ 5 kb were predicted to be viral, while only 4.06% and 3.43% of honey and bumble bee total metagenome contigs ≥ 5 kb were annotated as viral. The enrichment of phage sequences in viromes, relative to total metagenomes, provides validation that viromes successfully enrich for phage when applied to bee guts.

Next, we used rarefaction to assess if the different number of putative phage sequences recovered by honey and bumble bee viromes (Fig. 2D) was a result of differences in sampling depth (Fig. 2A). This analysis shows that even at equal sampling depths, honey bee viromes consistently recover approximately 20 \times the number of phage recovered from bumble bee viromes (Fig. S2A and B).

Phage target core bee bacteria and resemble previously described honey bee phage

After measuring the ability of viromes to enrich for phage sequences, we next sought to characterize the phage communities we detected in terms of similarity to previously described bee phage communities, putative bacterial hosts, and predicted viral taxonomy. To do this, we first removed phage predicted to be low quality ($< 50\%$ complete). This reduced the total number of sequences

recovered from 3853 to 655. Next, we constructed vOTUs by collapsing the remaining high-quality phage sequences at a 95% average nucleotide identity, further reducing the number of putative phage from 655 to 609.

We then built gene-sharing networks comparing our vOTUs to those previously described by other bee phage studies (Fig. 3A), as well as reference phage taken from vConTACT2's NCBI's RefSeq database (Fig. S3). In both network, many of the vOTUs from the current study clustered with those described by other bee phage studies, suggesting they are related at roughly the genus level [71, 72]. For example, viral cluster 149 contains sequences recovered by the current study, Bonilla-Rosso et al. [48], Deboutte et al. [49], and Busby et al. [50] (Fig. S4). We interpret these results as further validating our sample preparation and computational methodologies.

Next, we assigned bacterial hosts to our vOTUs. To focus this and all downstream analyses on only the most abundant phage, we filtered vOTUs to retain only those with a normalized coverage ≥ 1 , resulting in 477 vOTUs across 24 samples. Overall, we assigned hosts to 51.99% (248) of the 477 vOTUs appearing in our dataset (Fig. 3B). In terms of relative abundance, phage with assigned hosts made up an average of 4.03% and 18.81% of bumble bee viromes and total metagenomes and 60.92% and 65.97% of honey bee viromes and total metagenomes, respectively. Notably, the high proportion of bumble bee phage with unassigned hosts, especially in viromes, is likely confounded by the fact that these samples had an extraordinarily diminished phage diversity.

Phage predicted to target the core social bee bacterial genera *Lactobacillus*, *Bifidobacterium*, and *Gilliamella* were among the most abundant of those with identified hosts in our dataset (Fig. 3B and Table S1). Phage targeting *Lactobacillus* tended to be most abundant in honey bees, especially honey bee total metagenomes, though did not differ significantly (Bee $t_{14.16} = -1.38$, $p = 0.19$, Type $t_{16} = 1.00$, $p = 0.33$, Bee:Type $t_{16} = -0.01$, $p = 0.99$). Meanwhile, *Gilliamella* phage, which only appeared in honey bee, were most abundant in total metagenomes (Bee $t_{13.31} = -0.57$, $p = 0.58$, Type $t_{16} = 2.89$, $p < 0.05$, Bee:Type $t_{16} = -2.04$, $p = 0.058$). *Bifidobacterium* targeting phage were more abundant in honey bee viromes than they were in any other sample type (Bee $t_{20} = -6.42$, $p < 0.001$, Type $t_{20} = -5.83$, $p < 0.001$, Bee:Type $t_{20} = 4.38$, $p < 0.001$). Phage targeting the core bumble bee symbiont *Schmidhempelia* were exclusively found in bumble bee total metagenomes, though these differences were not statistically significant (Bee $t_{20} = 0$, $p = 1$, Type $t_{20} = 0$, $p = 1$, Bee:Type $t_{20} = 1.96$, $p = 0.064$). We also found signatures of colony-level variation. For example, *Bombilactobacillus* phage did not differ significantly across samples (Bee $t_{18.29} = -0.036$, $p = 0.97$, Type $t_{16} = -0.046$, $p = 0.96$,

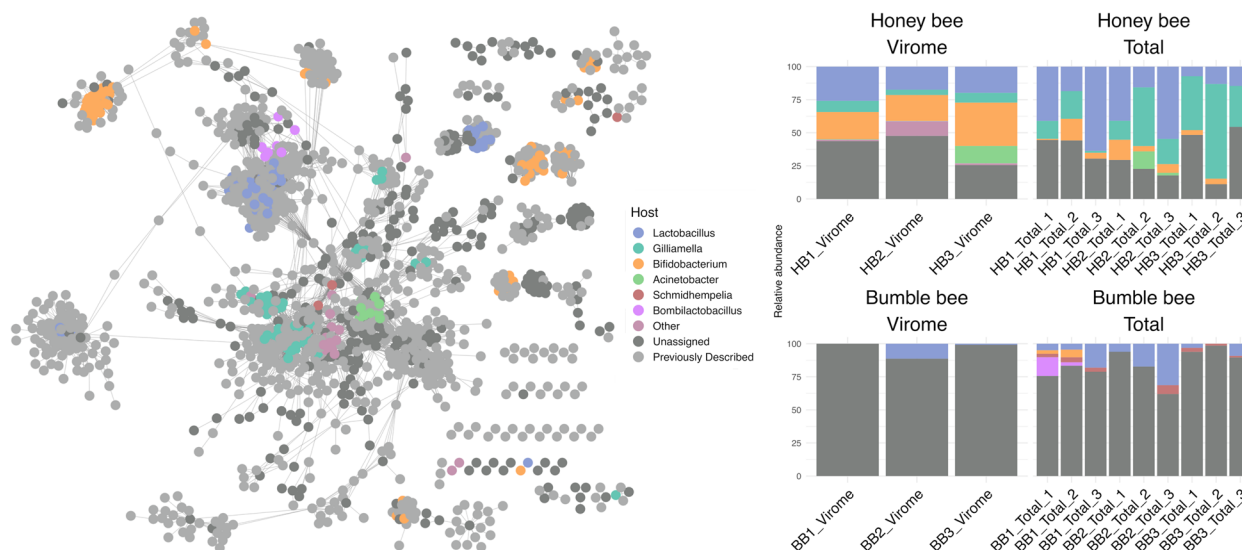


Fig. 3 Figures describing gene sharing among the phage identified in our dataset and the overall phage communities found in individual samples. **A** Weighted gene-sharing network of all non-singleton vOTUs identified in our study and by previous bee phage studies. Individual nodes are vOTUs. Nodes are connected by edges when vOTUs share genes. Nodes are colored based on the predicted bacterial host of phage. Nodes with “previously described” as a predicted host are the vOTUs previously described by other bee phage papers [48–50]. **B** Stacked bar plot describing the community of phage found in each sample based on predicted host. Both **A** and **B** use the same color palette

Bee:Type $t_{16} = 0.72$, $p = 0.48$) but were only detected in total metagenomes sampled from bumble bee colony 1.

In total, we assigned family level taxonomy to 9.64% (46) of our 490 vOTUs (Fig. S4). The majority of classified vOTUs belonged to either Rountreeviridae (24) or Herelleviridae (10), with a smaller number being classified as Inoviridae (4), Autographiviridae (3), and Microviridae (1). Notably, four vOTUs predicted to belong to the double-stranded RNA virus family Totiviridae were detected only in bumble bee total metagenomes.

Bee species differ in phage diversity and composition

Next, we tested if vOTU diversity and community composition differed between bee species and sampling method. While the diversity and richness of communities followed similar trends (Fig. 4A and B), there were substantial quantitative differences between these two metrics. When we compared Shannon’s diversity, sampling method and the interaction of bee species and sampling method significantly affected vOTU diversity, while bee species alone did not (Fig. 4A; Bee $t_{5,93}=2.16$ $p=0.074$; Type $t_{16}=-2.39$ $p<0.05$; Bee:Type $t_{16}=5.09$ $p<0.001$). Meanwhile, only the interaction between bee species and sampling method significantly impacted community richness (Fig. 4B; Bee $t_{20} = 0.75$, $p = 0.46$; Type $t_{20} = -0.85$, $p = 0.41$; Bee:Type $t_{20} = 19.15$, $p < 0.001$). Overall, these results suggest that sampling method can significantly influence the alpha diversity of inferred phage

communities, but that this can differ in magnitude and direction according to bee species.

In honey bees, 21.56% (91/422) of all vOTUs were shared between virome and total metagenome samples. Similarly, in bumble bees, 23.21% (13/56) of vOTUs were detected by both sampling methods. Only one vOTUs was detected in both bee species (Fig. 4C). vOTU community composition was predicted primarily by host bee species and secondarily based on sample method (Figure 4D; Bray-Curtis PERMANOVA Bee $R^2 = 0.21$, $F_{1,20} = 7.390$, $p < 0.001$; Method $R^2 = 0.11$, $F_{1,20} = 3.93$, $p < 0.001$; Bee:Method $R^2 = 0.11$, $F_{1,20} = 3.93$, $p < 0.001$).

Bacterial community variation predicts phage community variation

We hypothesized that differences in phage richness and community composition between bee species may be driven by differences in bacterial abundance and composition. First, we used qPCR to compare the number of bacteria found in the guts of honey and bumble bees (Fig. 5A) and found that the average mid-hindgut 16S gene copy number was significantly lower in bumble bees than it was in honey bees ($t_{9,13} = -3.54$, $p < 0.01$).

We then used Kraken2 and Bracken to estimate the diversity and taxonomic profile of bacterial communities (Fig. 5B, C, D). At the species level, bumble bee bacterial communities hosted lower bacterial richness and greater Shannon’s diversity than honey bee bacterial communities (Richness $t_{13,78} = -6.27$, $p < 0.001$,

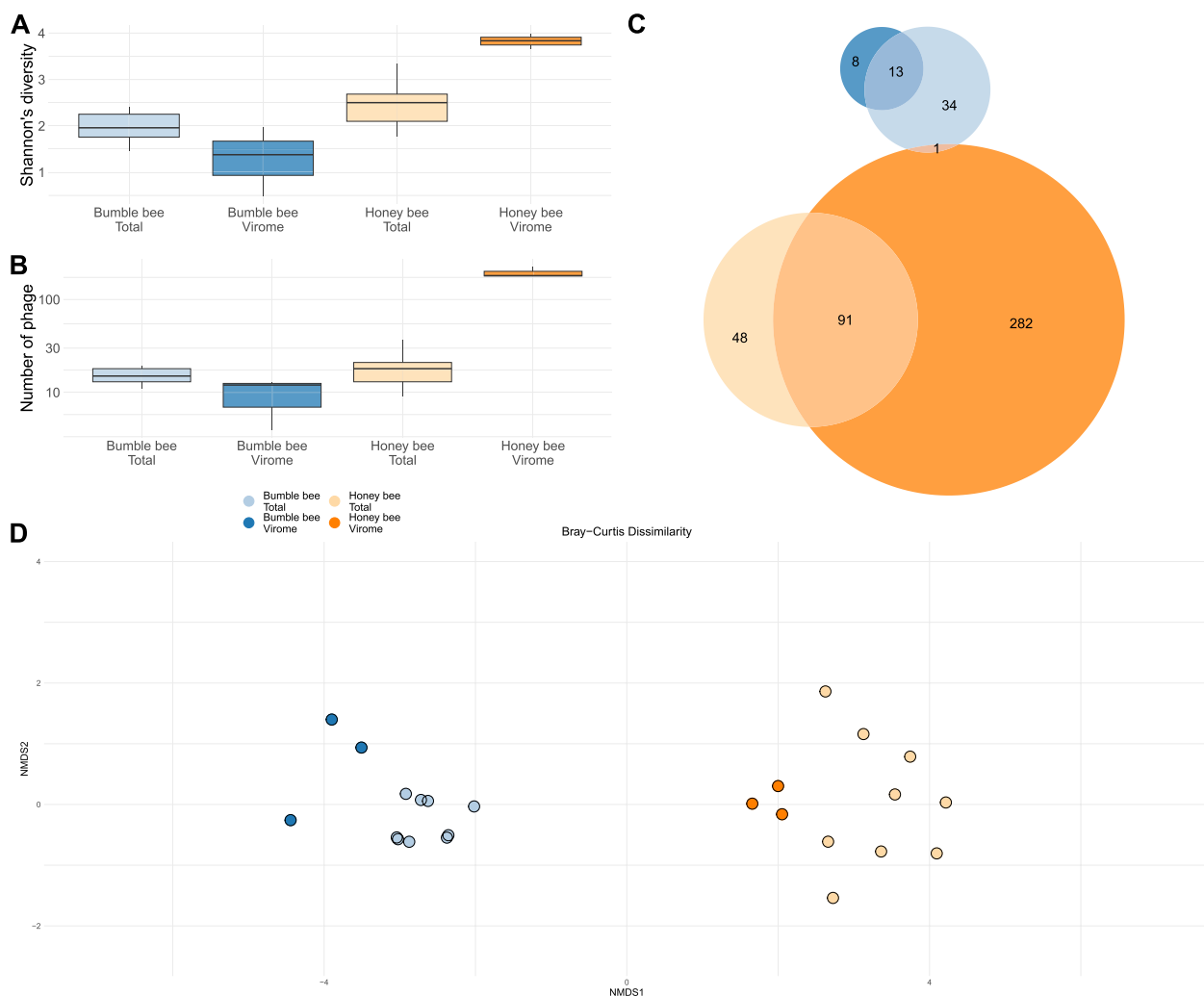


Fig. 4 Figures describing the alpha and beta diversity of the phage communities identified in our samples. Blue represents samples taken from bumble bees. Orange represents samples taken from honey bees. Lighter colors are total metagenomes. Darker colors are viromes. **A** and **B** Boxplots describing Shannon's diversity and vOTU richness associated with each of our sample types. The x-axis group samples. The y-axis shows diversity and richness scores. **C** Euler diagram describing phage community overlaps between each of our sample types. Numbers correspond to the number of phage present in each section of the graph (i.e., 91 phage were found in both honey bee viromes and total metagenomes). Circle size is proportional to number of phage. **D** Nonmetric multidimensional scaling (NMS) ordination describing the Bray-Curtis dissimilarity of all samples on our dataset. Individual points are samples. Color represents sample type. The closer 2 points are to each other, the more similar they are in phage community

Shannon $t_{11.70} = -6.27$, $p < 0.001$), suggesting fewer bacterial taxa were present in the bumble bees we sampled, but that bumble bee bacterial communities were more evenly divided among their constituent bacteria.

Lastly, we used a Mantel test to compare the relationship between bacterial and phage community structure (Fig. S5). We found a high degree of similarity between the bacterial and phage communities ($R = 0.59$, $p < 0.001$), meaning that individual bees with similar bacterial communities also have similar phage communities.

Total metagenomes capture prophage but miss rare phage

After identifying factors which explain some of the host-specific differences in phage communities, we sought to evaluate why sampling method (virome vs total metagenome) also led to significant differences in inferred phage communities. For this, we first compared the gene content of vOTUs recovered from viromes and total metagenomes. Because total metagenomes are primarily comprised of bacterial DNA, we hypothesized that this sampling method would enrich for temperate phage. In both honey bees and total metagenomes, temperate

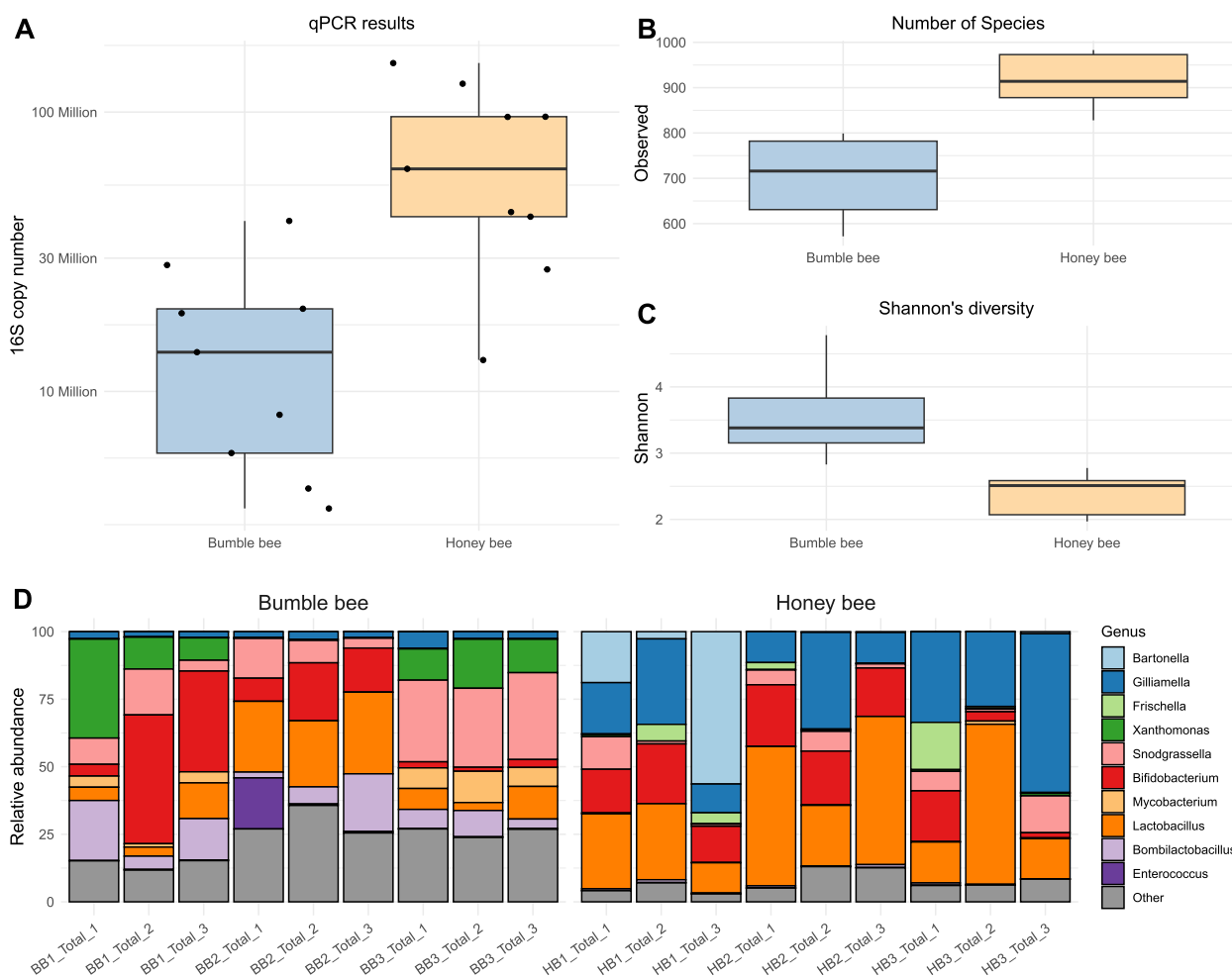


Fig. 5 Honey and bumble bees differ in bacterial community density, diversity, and composition. **A** Boxplots displaying the 16S rRNA gene copy number present in the mid-hindgut region of sampled honey and bumble bees. **B** Boxplot presenting the number of observed bacterial species in honey and bumble bee. **C** Boxplot showing the species level Shannon's diversity of honey and bumble bee gut bacterial communities. **D** Stacked bar plot describing honey and bumble bee bacterial communities at the genus level

phage comprised a greater proportion of the entire vOTU community, relative to bumble bees and viromes (Fig. S6; Bee $t_{20} = 3.88$, $p < 0.001$, Type $t_{20} = -2.32$, $p < 0.05$, Bee:Type $t_{20} = 0.31$, $p = 0.76$). In terms of relative abundance, integrase encoding phage were more abundant in total metagenomes than they were in viromes (Fig. 6A; Bee $t_{17.36} = -1.56$, $p = 0.14$; Type $t_{16} = 3.80$, $p < 0.01$; Bee:Type $t_{16} = -1.53$, $p = 0.12$). Overall, these results suggest that, in this system, total metagenomes are biased towards sampling temperate phage.

Lastly, to examine how sampling methods differ in recovery of rare vs abundant phage, we visualized the occupancy-abundance relationships of virome-specific vOTUs vs vOTUs recovered by both methods. This analysis revealed that while there is little to no difference between the relative abundance of vOTUs identified in viromes vs total metagenomes (Fig. 6B), viromes

are better able to capture vOTUs with low occupancy (Fig. 6C and D).

Discussion

Although phage are integral members of complex microbiomes, methods for describing phage communities are still being developed. Here, we compare how two common forms of phage community sampling, bioinformatic mining of total metagenomes and targeted sequencing of viromes, influence the phage communities recovered from the guts of two commercially important bee species: *Apis mellifera* and *Bombus impatiens*. Similar to the results of previous studies [16–18], we show that sampling method significantly affected alpha diversity, beta diversity, composition, and structure of sampled phage communities. In particular, we show that viromes outperform total metagenomes in terms of the number

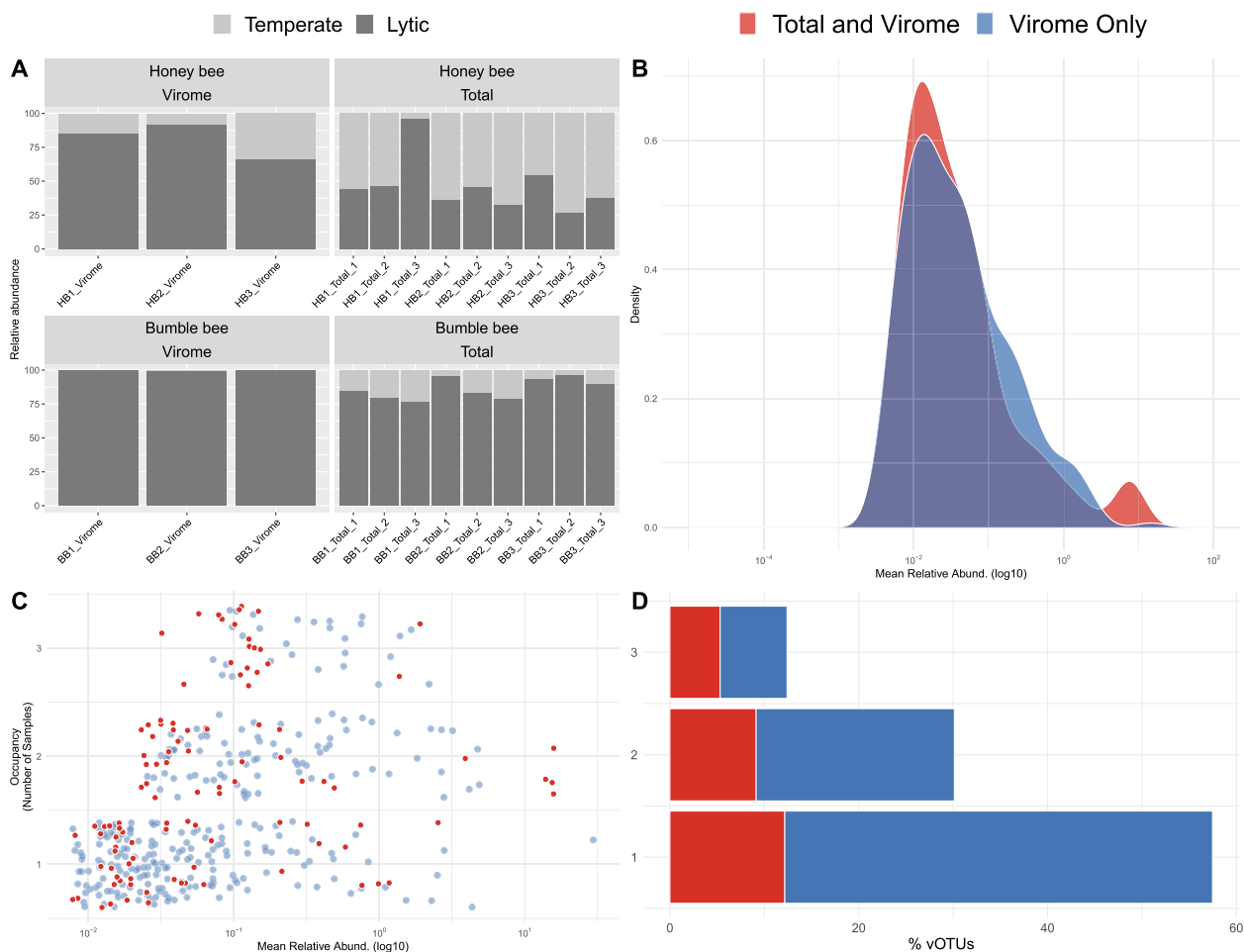


Fig. 6 vOTUs sampled by viromes and total metagenomes differ in integrase content, occupancy, and abundance. **A** Stacked bar plot showing the relative abundance of temperate and lytic vOTUs in virome and total metagenome samples. **B** Density distributions showing the average relative abundance of all vOTUs found in viromes. vOTUs unique to viromes are shown in blue. vOTUs found in both viromes and total metagenomes are shown in red. **C** Dot plot showing the occupancy-abundance relationship of all virome vOTUs. The x-axis describes average relative abundance, while the y-axis shows the percent occupancy of each vOTU. As before, vOTUs unique to viromes are shown in blue, and those found in both viromes and total metagenomes are shown in red. **D** Stacked bar plot describing the percent of vOTUs found in one, two, or three samples. The x-axis represents the percent of vOTUs. The y-axis shows the occupancy. Similar to **B** and **C**, vOTUs unique to viromes are shown in blue, and those found in both viromes and total metagenomes are shown in red

of phages recovered, but that this can depend on the specific environment being sampled. Overall, viromes were better at sampling rare and less prevalent phage when bacterial and phage diversity was high, while total metagenomes captured a greater diversity of phage in low biomass samples. We also show that regardless of sample biomass, total metagenomes were enriched for temperate phage, compared to viromes, and were consistently able to recover phage not found in viromes. Given these results, we suggest that viromes and total metagenomes each have limitations, are complementary, and that choice of one method over the other likely depends on the environment being sampled.

Honey bees host a core phage community

Viewing our results within the broader context of bee phage, our analyses show that many features of the phage communities identified here resemble those previously described in honey bees sampled from North America and Europe [48–50]. For example, Herelleviridae (formerly Myoviridae [78]) and Rountreeviridae (formerly Podoviridae [79]) were among the most abundant viral families that we and these previous studies identified. Similarly, *Lactobacillus*, *Gilliamella*, *Bifidobacterium*, and other core bee gut bacteria, were among the most predicted phage hosts in our and previous studies. Interestingly, four of the viruses identified in our study were

predicted to belong to Totiviridae, a family of viruses known to target yeasts and other eukaryotes [80]. Given the importance of yeasts in maintaining bee health [81], especially in bumble bees [82], this suggests that yeast targeting viruses may be an important member of bee gut microbial communities. Notably, while previous bee phage studies found a number of phage predicted to target *Bartonella*, none of our phage was predicted to target this bacterial genus. This is likely because no available *Bartonella apis* assemblies met the strict criteria for inclusion in our CRISPR spacer analysis.

Delving deeper, our gene-sharing network and clinker analyses show not only that many of our phage share large genomic regions with previously described honey bee phage but also that gene order in these geographically disparate viral genomes is also conserved (Fig. S4). This conservation between honey bee phage communities is similar to what was observed by Busby et al. [50]. Likewise, Rosso et al. [48] showed that the genomes of honey bee-associated phage sampled from Europe are able to recruit reads from Japanese honey bee total metagenomic datasets.

Altogether, these findings support the idea that similar to the highly conserved nature of some bee gut bacterial communities [38, 39, 46, 83], some features of bee gut phage communities are also conserved between bees sampled across geographically disparate populations. Future work should focus on identifying what features of phage communities, whether individual genes, clusters of genes, or perhaps whole genomes, are conserved between different populations of bees.

Host bacterial communities predict phage communities

When compared to free-foraging managed honey bees, the commercially produced bumble bees we sampled hosted significantly lower diversity phage communities. Although differences in species biology and colony size may play a role, we attribute this low phage diversity primarily to the low density and diversity bacterial communities these bumble bees hosted, suggesting that phage diversity and abundance may track bacterial diversity and abundance across hosts. However, whether such results are generalizable to wild bumble bees is unclear. In our study, bacterial 16S rRNA copy number in bumble bee guts ranged from 10^6 to 10^7 , whereas previous work in the same species reported 10^8 – 10^9 copies of the 16S rRNA gene [38, 84–86]. Given we used standard rearing methods and diet, we hypothesize that lower bacterial titer may be due to the age of bees we sampled. Previous work in bumble bees shows that bacterial density increases with worker age, saturating approximately 4 days after worker emergence [84]. Future work could examine how phage abundance and composition changes

through worker development and if this reflects previous observations of bacterial succession with worker age [17]. More generally, these findings suggest that features of host bacterial communities (i.e., density, diversity, and structure) may be used to predict phage community features. This is further supported by our Mantel analysis showing a correlation between bacterial and phage community dissimilarity.

Viromes and total metagenomes reveal complementary phage communities

There are several possible reasons as to why the phage communities inferred by viromes and total metagenomes differ. One explanation is the way we generated viromes and how this influenced the specific population of phage sampled. Research by Hoyles et al. [87] has shown that passing human fecal samples through a 0.22- μ m filter reduces the number of phage recovered by nearly half. As such, size filtration may have led some phage which are present in our bee guts to have been excluded from viromes. This size exclusion may explain why we, and previous virome vs total metagenome studies [16–18], consistently show that some phage are only found in total metagenomes. Other potential explanations include sheer random chance (two independent sampling events each recovering unique elements) or sample processing (total metagenome samples being frozen prior to DNA extraction). While random chance cannot be fully ruled out, given that total metagenomes did not receive a DNase treatment prior to sequencing, disruption of phage capsid proteins by freezing likely did not have a substantial impact on inferred phage communities. Instead, the particular types of phage which are excluded by size filtration likely further explain the total metagenome and virome differences we observed. We found that putative temperate phage were more abundant and prevalent in total metagenomes than they were in viromes. This agrees with existing literature which suggests that size filtration can specifically remove integrated lysogenic phage, large jumbo phage, and phage adhered to bacterial cells and particles [32, 34, 88, 89]. While it is tempting to expand these results to different environments by stating total metagenomes always sample temperate phage missed by viromes, work in other systems implies a more nuanced reality. In agricultural soils, Medellin et al. [16] identified only three phages that were present in total metagenomes and absent from viromes. This suggests that the degree to which sample processing skews viromes differs by environment. As a result, we suggest that the relative utility of viromes and total metagenomes likely depends on the environment being sampled. Samples with a high abundance of temperate phage, such as host-associated systems and low-biomass environments,

may benefit from total metagenomes. This is similar to the conclusions made by Kosmopoulos et al. [17] which suggest that the choice of viromes vs total metagenomes should be environment specific.

While our results highlight the ability of total metagenomes to recover phage missed by viromes, they also showcase the capacity of viromes to sample a greater diversity of phage. Similar to previous comparisons of viromes and total metagenomes in human gut, soil, and freshwater environments [17, 18, 29], our honey bee viromes recovered a substantially larger number of phage than did total metagenomes. Further, the occupancy-abundance relationships examined here and by Medellin et al. [16] show that total metagenomes tend to be biased towards sampling the most abundant and prevalent phage, while viromes are more successful at sampling rare phage. Taken together, these results support previous work documenting that viromes produce a deeper sampling of phage communities compared to total metagenomes [32, 90, 91].

Conclusions

By comparing bioinformatic mining of total metagenomes and targeted viromes across two bee species, we found that sampling method significantly affected inferred phage communities, but that the directionality of these differences can depend on the host being sampled. In general, total metagenomes tended to be biased towards sampling putatively temperate phage. In samples with a relatively high biomass (e.g., honey bees), viromes produced a greater diversity of phage and a better sampling of rare phage. In contrast, when applied to relatively low biomass samples (e.g., bumble bees), total metagenomes captured a greater diversity of phage than did viromes. Regardless of sample biomass, there were always phage unique to both viromes and total metagenomes. In conclusion, we suggest that these methods are complementary and recommend that both be used to capture the full diversity of phage present in a gut environment. However, given that virome sampling is not always feasible (i.e., in the case of field collected insects), total metagenomes may serve to sample phage communities with the caveat that they will preferentially sample dominant and temperate phage.

Abbreviations

qPCR	Quantitative polymerase chain reaction
PBS	Phosphate buffered saline
PPBS	Protein-supplemented phosphate buffered saline
vOTU	Viral operational taxonomic unit

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40168-024-01875-0>.

Additional file 1: Figure S1. Boxplots describing the number of raw read pairs generated per sample type. Sample type is shown along the x-axis. Total number of reads pairs, in millions, generated is represented on the y-axis. All samples produce a roughly equal number of raw reads. Figure S2. When sampled at equal depths, honey bees recover a greater number of phage sequences. A) Species accumulation curve showing the number of phage recovered from honey and bumble bees as the number of reads considered increases. B) The number of phage recovered from honey and bumble bees when sampled at an equal depth (3.5 million read pairs). Figure S3. Weighted vOTU gene-sharing network. Nodes represent vOTUs or reference viral sequences. Nodes are connected by edges whose viral sequences share genes. Nodes corresponding to vOTUs recovered from the current study are red, vOTUs from previous bee-phage studies are blue, and nodes representing reference phage sequences present in vConTACT2's "ProkaryoticViralRefSeq211-Merged" database are colored gray. Figure S4. Clinker plot displaying gene sharing amongst all the phage in a single gene-sharing network cluster. The cluster shown (VC_149) is predicted to target *Gilliamella*. Each row is an individual phage in this cluster. Arrows represent the genes present in those phage. Connections between phage represent homologous genes. Connections are colored based on average amino acid identity. Phage are labeled with the study they were identified in. Virome and total metagenome phage were identified in our virome and total metagenome samples, respectively. Figure S5. Bar plots showing the number of vOTUs successfully assigned to different viral families. Figure S6. Scatterplot visualizing the mantel test performed. Each point represents the pairwise distance between two samples in beta diversity space. The x-axis measures bacterial community distance. The y-axis measures viral community distance. Points are colored based on whether the sample pairs being compared are both bumble bees (red), honey bees (green) or two different types of bees (blue). There is a positive relationship between phage and bacterial community structure. Figure S7. Boxplots describing the percent of vOTUs in each sample type predicted to be temperate. The x-axis describes sample group, the y-axis describes percent temperate vOTUs. Table S1. Table describing the distribution of phage predicted to target common bee associated bacteria. Letters represent results of emmeans posthoc test.

Additional file 2. Supplemental methods: Bee rearing, sampling, and experimental design. Sample preparation and DNA extraction. Read generation, pre-processing, and decontamination. Computational phage identification. vOTU clustering, annotation, and table generation. Statistical and ecological analyses. Bacterial community and density measurements. Network generation. Phage host prediction.

Acknowledgements

We thank Dr. Elina L. Nino, Joseph Tauzer, and the UC Davis Harry H. Laidlaw Jr. Honey Bee Research Facility staff for help sampling managed honey bee colonies, Samantha R. Eck and members of the Vannette, Emerson, and Brown labs for comments on the manuscript, Anoushka Basu for help dissecting bees, reviewers for their insightful feedback, and the UC Davis Genome center for sequencing.

Authors' contributions

DLs devised and carried out the research, including experimental design, sampling effort, sample processing, and computational analyses. RLV assisted with experimental design, supervised progress, and provided manuscript feedback.

Funding

This research was supported in part by the Henry A. Jastro Scholarship and the United States Department of Agriculture: National Institute of Food and Agriculture, Agriculture, and Food Research Initiative (award number 2023-67011-40501) awarded to D. L. S., as well as Hatch Multistate funding to R. L. V. and N. S. F. and DEB no. 1929516 to R. L. V.

Availability of data and materials

The raw sequencing data generated in this study is available on NCBI under the Bioproject accession number PRJNA1072153. All scripts used to process this data are publicly available at https://github.com/dsbard/Bee_page_2024.

Declarations**Ethics approval and consent to participate**

NA.

Consent for publication

NA.

Competing interests

The authors declare no competing interests.

Received: 19 February 2024 Accepted: 16 July 2024

Published online: 23 August 2024

References

- Bergh Ø, Børsheim KY, Bratbak G, Heldal M. High abundance of viruses found in aquatic environments. *Nature*. 1989;340:467–8.
- Chibani-Chennoufi S, Bruttin A, Dillmann ML, Brüssow H. Phage-host interaction: an ecological perspective. *J Bacteriol*. 2004;186:3677–3686. Preprint at <https://doi.org/10.1128/JB.186.12.3677-3686.2004>.
- Keen EC. A century of phage research: bacteriophages and the shaping of modern biology. *BioEssays*. 2015;37:6–9.
- Braga LPP, et al. Impact of phages on soil bacterial communities and nitrogen availability under different assembly scenarios. *Microbiome*. 2020;8:52.
- Trubl G, et al. Soil viruses are underexplored players in ecosystem carbon processing. *mSystems*. 2018;3:1–21.
- Emerson JB, et al. Host-linked soil viral ecology along a permafrost thaw gradient. *Nat Microbiol*. 2018;3:870–80.
- Adriaenssens EM, et al. Environmental drivers of viral community composition in Antarctic soils identified by viromics. *Microbiome*. 2017;5:83.
- Fuhrman JA. Marine viruses and their biogeochemical and ecological effects. *Nature*. 1999;399. www.nature.com.
- Suttle CA. Marine viruses - major players in the global ecosystem. *Nat Rev Microbiol*. 2007;5:801–12.
- Evans C et al. Shift from carbon flow through the microbial loop to the viral shunt in coastal Antarctic waters during austral summer. 2021. <https://doi.org/10.3390/microorganisms>.
- Danovaro R, et al. Major viral impact on the functioning of benthic deep-sea ecosystems. *Nature*. 2008;454:1084–7.
- Tsao YF, et al. Phage morons play an important role in *Pseudomonas aeruginosa* phenotypes. *J Bacteriol*. 2018;200:e00189.
- Heyerhoff B, Engelen B, Bunse C. Auxiliary metabolic gene functions in pelagic and benthic viruses of the Baltic Sea. *Front Microbiol*. 2022;13:863620.
- Dutilh BE, et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat Commun*. 2014;5:4498.
- Townsend EM et al. The human gut phageome: origins and roles in the human gut microbiome. *Front Cell Infect Microbiol*. 2021;11. Preprint at <https://doi.org/10.3389/fcimb.2021.643214>.
- Santos-Medellin C. et al. Viromes outperform total metagenomes in revealing the spatiotemporal patterns of agricultural soil viral communities. *ISME Journal*. 2021. <https://doi.org/10.1038/s41396-021-00897-y>.
- Kosmopoulos JC, Klier KM, Langwig MV, Tran PQ, Anantharaman, K. Viromes vs. mixed community metagenomes: choice of method dictates interpretation of viral community ecology. *bioRxiv*. 2023. <https://doi.org/10.1101/2023.10.15.562385>.
- Zhang F, et al. Critical assessment of whole genome and viral enrichment shotgun metagenome on the characterization of stool total virome in hepatocellular carcinoma patients. *Viruses*. 2022;15:53.
- Sender R, Fuchs S, Milo R. Revised estimates for the number of human and bacteria cells in the body. *PLoS Biol*. 2016;14:e1002533.
- Blum WEH, Zechmeister-Boltenstern S, Keiblinger KM. Does soil contribute to the human gut microbiome? *Microorganisms*. 2019;7:287.
- Garren M, Azam F. New method for counting bacteria associated with coral mucus. *Appl Environ Microbiol*. 2010;76:6128–33.
- Silveira CB, Rohwer FL. Piggyback-the-Winner in host-associated microbial communities. *NPJ Biofilms Microbiomes*. 2016;2. Preprint at <https://doi.org/10.1038/npjbiofilms.2016.10>.
- Knowles B, et al. Lytic to temperate switching of viral communities. *Nature*. 2016;531:466–70.
- Mirzaei MK, Maurice CF. Ménage à trois in the human gut: interactions between host, bacteria and phages. *Nat Rev Microbiol*. 2017;15:397–408. Preprint at <https://doi.org/10.1038/nrmicro.2017.30>.
- Chin WH et al. Bacteriophages evolve enhanced persistence to a mucosal surface. 2022. <https://doi.org/10.1073/pnas>.
- Nelson AR, et al. Wildfire-dependent changes in soil microbiome diversity and function. *Nat Microbiol*. 2022;7:1419–30.
- Waller AS, et al. Classification and quantification of bacteriophage taxa in human gut metagenomes. *ISME J*. 2014;8:1391–402.
- Göller PC, Haro-Moreno JM, Rodriguez-Valera F, Loessner MJ, Gómez-Sanz E. Uncovering a hidden diversity: optimized protocols for the extraction of dsDNA bacteriophages from soil. *Microbiome*. 2020;8:17.
- Santos-Medellin C, et al. Spatial turnover of soil viral populations and genotypes overlain by cohesive responses to moisture in grasslands. *Proceedings of the National Academy of Sciences*. 2022;119:e2209132119.
- Conceição-Neto N, et al. Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. *Sci Rep*. 2015;5:16532.
- Parras-Moltó M, Rodríguez-Galet A, Suárez-Rodríguez P, López-Bueno A. Evaluation of bias induced by viral enrichment and random amplification protocols in metagenomic surveys of saliva DNA viruses. *Microbiome*. 2018;6:119.
- Garmaeva S et al. Studying the gut virome in the metagenomic era: challenges and perspectives. *BMC Biology*. 2019;17. Preprint at <https://doi.org/10.1186/s12915-019-0704-y>.
- Gregory AC, et al. The gut virome database reveals age-dependent patterns of virome diversity in the human gut. *Cell Host Microbe*. 2020;28:724–740.e8.
- Avellaneda-Franco L, Dahlman S, Barr J J. The gut virome and the relevance of temperate phages in human health. *Front Cell Infect Microbiol*. 2023;13. Preprint at <https://doi.org/10.3389/fcimb.2023.1241058>.
- Chandler JA, Lang J, Bhatnagar S, Eisen JA, Kopp A. Bacterial communities of diverse *Drosophila* species: ecological context of a host-microbe model system. *PLoS Genet*. 2011;7:e1002272.
- Douglas AE. The molecular basis of bacterial-insect symbiosis. *J Mol Biol*. 2014;426:3830–3837. Preprint at <https://doi.org/10.1016/j.jmb.2014.04.005>.
- Pernice M, Simpson SJ, Ponton F. Towards an integrated understanding of gut microbiota using insects as model systems. *J Insect Physiol*. 2014;69:12–8.
- Hammer TJ, Le E, Martin AN, Moran NA. The gut microbiota of bumblebees. *Insectes Soc*. 2021;68:287–301.
- Kwong WK, Moran NA. Gut microbial communities of social bees. *Nat Rev Microbiol*. 2016;14:374–384. Preprint at <https://doi.org/10.1038/nrmicro.2016.43>.
- Motta EVS, Moran NA. The honeybee microbiota and its impact on health and disease. *Nat Rev Microbiol*. 2024;22:122–137. Preprint at <https://doi.org/10.1038/s41579-023-00990-3>.
- Cabirol A, Moriano-Gutierrez S, Engel P. Neuroactive metabolites modulated by the gut microbiota in honey bees. *Mol Microbiol*. 2023. Preprint at <https://doi.org/10.1111/mmi.15167>.
- Raymann K, Shaffer Z, Moran NA. Antibiotic exposure perturbs the gut microbiota and elevates mortality in honeybees. *PLoS Biol*. 2017;15:1–22.
- Marceau T, Archer CR, Bulson L, Wilfert L. Dose-dependent effects of antibiotic intake on *Bombus terrestris* (Linnaeus, 1758) dietary intake, survival and parasite infection prevalence. *J Invertebr Pathol*. 2021;182:107580.
- Koch H, Schmid-Hempel P. Socially transmitted gut microbiota protect bumble bees against an intestinal parasite. *Proc Natl Acad Sci U S A*. 2011;108:19288–92.

45. Engel P, et al. The bee microbiome: impact on bee health and model for evolution and ecology of host-microbe interactions. *mBio*. 2016;7:1–9.
46. Zheng H, Steele MI, Leonard SP, Motta EVS, Moran NA. Honey bees as models for gut microbiota research. *Lab Anim (NY)*. 2018;47:317–25.
47. Kwong WK. Bee microbiomes go viral. *Proc Natl Acad Sci U S A*. 2020;117:11197–9.
48. Bonilla-Rosso G, Steiner T, Wichmann F, Bexkens E, Engel P. Honey bees harbor a diverse gut virome engaging in nested strain-level interactions with the microbiota. *Proc Natl Acad Sci U S A*. 2020;117:7355–62.
49. Deboutte W et al. Honey-bee-associated prokaryotic viral communities reveal wide viral diversity and a profound metabolic coding potential. <https://doi.org/10.1073/pnas.1921859117/-/DCSupplemental>.
50. Busby TJ, Miller CR, Moran NA, Van Leuven JT. Global composition of the bacteriophage community in honey bees. *mSystems*. 2022;7:e0119521.
51. ter Horst AM, Fudyma JD, Sones JL, Emerson JB. Dispersal, habitat filtering, and eco-evolutionary dynamics as drivers of local and global wetland viral biogeography. *ISME J*. 2023. <https://doi.org/10.1038/s41396-023-01516-8>.
52. Krueger F. TrimGalore. Github <https://github.com/FelixKrueger/TrimGalore>.
53. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
54. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
55. Pierce NT, Irber L, Reiter T, Brooks P, Brown CT. Large-scale sequence comparisons with sourmash. *F1000Res*. 2019;8:1006.
56. Camargo AP et al. Identification of mobile genetic elements with geNomad. *Nat Biotechnol*. 2023. <https://doi.org/10.1038/s41587-023-01953-y>.
57. Nayfach S, et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol*. 2021;39:578–85.
58. Olm MR, Brown CT, Brooks B, Banfield JF. DRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME Journal*. 2017;11:2864–8.
59. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30:2068–9.
60. Terzian P, et al. PHROG: families of prokaryotic virus proteins clustered using remote homology. *NAR Genom Bioinform*. 2021;3:1–12.
61. Hockenberry AJ, Wilke CO. BACPHLIP: predicting bacteriophage lifestyle from conserved protein domains. *PeerJ*. 2021;9:e11396.
62. Woodcroft BJ, CoverM. Github <https://github.com/wwood/CoverM>.
63. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol*. 2019;20:257.
64. Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: estimating species abundance in metagenomics data. *PeerJ Comput Sci*. 2017;3:104.
65. Christensen SM et al. Microbial metamorphosis: symbiotic bacteria and fungi proliferate during diapause and may enhance overwintering survival in a solitary bee. *bioRxiv*. 2023. <https://doi.org/10.1101/2023.11.02.565352>.
66. R Core Team. R: a language and environment for statistical computing. 2019. Preprint at <https://www.r-project.org/>.
67. Oksanen AJ et al. Vegan: community ecology package. R package version 2.6-4; 2022. <https://CRAN.R-project.org/package=vegan>.
68. McMurdie PJ, Holmes S. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*. 2013;8:e61217.
69. Bates D, Mächler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. *J Stat Softw*. 2015;67(1):1–48.
70. Russell V. Lenth. emmeans: estimated marginal means, aka least-squares means. <https://CRAN.R-project.org/package=emmeans>.
71. Bolduc B, et al. vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. *PeerJ*. 2017;5:e3243.
72. Bin Jang H, et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat Biotechnol*. 2019;37:632–9.
73. Gilchrist CLM, Chooi YH. Clinker & clustermapjs: automatic generation of gene cluster comparison figures. *Bioinformatics*. 2021;37:2473–5.
74. Kang DD, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*. 2019;7:e7359.
75. Skennerton CT. Minced. Github <https://github.com/ctSkennerton/minced>.
76. Camacho C, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10:421.
77. Edwards RA, McNair K, Faust K, Raes J, Dutilh BE. Computational approaches to predict bacteriophage-host relationships. *FEMS Microbiol Rev*. 2016;40:258–72.
78. Adriaenssens EM, et al. Taxonomy of prokaryotic viruses: 2018–2019 update from the ICTV Bacterial and Archaeal Viruses Subcommittee. *Arch Virol*. 2020;165:1253–60.
79. Zerbini FM, et al. Changes to virus taxonomy and the ICTV statutes ratified by the International Committee on Taxonomy of Viruses (2023). *Arch Virol*. 2023;168:175.
80. Rowley PA. The frenemies within: viruses, retrotransposons and plasmids that naturally infect *Saccharomyces* yeasts. *Yeast*. 2017;34:279–92.
81. Rutkowski D, Weston M, Vannette RL. Bees just wanna have fungi: a review of bee associations with nonpathogenic fungi. *FEMS Microbiol Ecol*. 2023;99. Preprint at <https://doi.org/10.1093/femsec/fiad077>.
82. Rutkowski D, Litsey E, Maalouf I, Vannette RL. Bee-associated fungi mediate effects of fungicides on bumble bees. *Ecol Entomol*. 2022;47:411–22.
83. Handy MY, et al. Incipiently social carpenter bees (Xylocopa) host distinctive gut bacterial communities and display geographical structure as revealed by full-length PacBio 16S rRNA sequencing. *Mol Ecol*. 2023;32:1530–43.
84. Hammer TJ, Easton-Calabria A, Moran NA. Microbiome assembly and maintenance across the lifespan of bumble bee workers. *Mol Ecol*. 2023;32:724–40.
85. Wang L et al. Dynamic changes of gut microbial communities of bumble bee queens through important life stages. 2019. <https://doi.org/10.1128/mSystems>.
86. Kwong WK, et al. Dynamic microbiome evolution in social bees. *Sci Adv*. 2017;3:e1600513.
87. Hoyles L, et al. Characterization of virus-like particles associated with the human faecal and caecal microbiota. *Res Microbiol*. 2014;165:803–12.
88. López-Pérez M, Haro-Moreno JM, Gonzalez-Serrano R, Parras-Moltó M, Rodríguez-Valera F. Genome diversity of marine phages recovered from Mediterranean metagenomes: size matters. *PLoS Genet*. 2017;13:e1007018.
89. Gregory AC, et al. The gut virome database reveals age-dependent patterns of virome diversity in the human gut. *Cell Host Microbe*. 2020;28:724–740.e8.
90. Emerson JB. Soil viruses: a new hope. *mSystems*. 2019;4:e00120.
91. Trubl G, Hyman P, Roux S, Abedon ST. Coming-of-age characterization of soil viruses: a user's guide to virus isolation, detection within metagenomes, and viromics. *Soil Syst*. 2020;4:1–34. Preprint at <https://doi.org/10.3390/soilsystems4020023>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.