1 Title:

2 Targeted Viromes and Total Metagenomes Capture Distinct Components of Bee Gut Phage

3 Communities

4 Authors:

5 Dino Lorenzo Sbardellati[1], Rachel Lee Vannette[2]

6 Author information:

7 [1]Microbiology Graduate Group, University of California Davis, Davis, California, USA.

8 [2]Department of Entomology and Nematology, University of California Davis, Davis, California,

9 USA

10 Corresponding author

11 Dino L. Sbardellati

12 Phone: 530-752-3379

13 Email: dlsbardellati@ucdavis.edu

14

Abstract:

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*. 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 abundance and low occupancy phage, while total metagenomes were biased towards sampling temperate phage and the most prominent 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. 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:

38 bacteriophage, bee microbiome, host-associated microbiome, microbial ecology, viral ecology

39

Introduction

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 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 is 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, sequenced using 'shotgun' or untargeted methods, assembled into metagenomes, and phage computationally mined[26,27]. This approach is cheap, easy to perform, and offers simultaneous characterization of all DNA in a sample (including bacteria). However, only a small minority of the total metagenomic data generated will originate from phage, resulting in 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[28,29]. With more sequencing 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[30,31]. For example, viromes remove bacterial cells prior to sequencing and, as a result, miss integrated temperate phage[32,33]. This is especially relevant in host-associated systems where temperate phage are abundant[32]. Total metagenomes may also be preferable to viromes logistically, given

86 that viromes demand a relatively large biomass and are labor-intensive to generate, which can
87 limit sample sizes in low-biomass systems.

88 Insects have become valuable models for exploring host-microbe interactions[34–36]. Social
89 bees in particular house a simple (5 - 9 taxa), highly conserved, socially transmitted, gut bacterial
90 community[37,38] which contributes to host nutrient acquisition and pathogen defense[39–41]. These
91 features position social bees as an excellent model for studying not only host-microbe
92 interactions[42,43], but tripartite host-microbe-phage interactions as well[44]. A series of recent studies
93 using targeted virome approaches have shown that honey bees host a diversity of novel phage
94 which target core bee gut bacteria[45–47]. While this work has advanced our understanding of bee
95 phage, no work has evaluated how sampling method influences recovered bee phage
96 communities. Given the large biomass required for viromes, total metagenomes may be
97 advantageous for surveying the phage communities associated with bees and other biomass
98 limited systems. Additionally, describing the phage associated with other social bees will develop
99 our understanding of how phage communities differ between related hosts. Lastly, broader work
100 comparing the role and diversity of phage among small invertebrate hosts will hinge on
101 appropriate sampling methods, so comparing the performance of viromes and total
102 metagenomics is particularly important for enabling future comparative study.

103 Here, we evaluate the phage communities inferred by applying total metagenomes and
104 targeted viromes to managed *Apis mellifera* and commercially raised *Bombus impatiens* gut
105 material. We hypothesize that targeted viromes will enrich for phage sequences and capture a
106 greater diversity of phage, relative to bioinformatic mining of total metagenomes, and that phage
107 will differ between host bee species. To interrogate these hypotheses, we first examine how host
108 bee species and sampling method impact sequencing read data and viral enrichment. We then

109 validate our sampling and computational methodology through rarefaction and by comparing our
110 phage sequences to those previously described in honey bees. Next, to test how phage
111 community differs across host bee and sampling methods, we compare phage community
112 diversity, structure, and composition. We then use gene content, occupancy-abundance plots,
113 qPCR, and bacterial community profiling to delve into what contributes to the apparent
114 differences in phage communities. Overall, we find that honey bees host more phage than do
115 bumble bees, but that the method which captured the most phage differed by bee species. This
116 suggests that viromes and total metagenomes are each biased in their own way and appear to
117 sample different populations of phage. As a result, we propose these techniques are
118 complementary in describing the full diversity of host-associated phage.
119

120 ## Methods

121 ### Bee rearing, sampling, and experimental design

122 A detailed description of our methods (including sample collection, and bench and
123 computational work) is provided in the supplemental materials. Briefly, bees were harvested
124 from three colonies of managed honey bees (*Apis mellifera*) and three colonies of commercially
125 reared bumble bees (*Bombus impatiens*, Koppert Biological Systems; Howell, MI, USA).
126 Bumble bee colonies were maintained in the laboratory and provided with honey bee collected
127 pollen and artificial nectar (Koppert Biological Systems) *ad libitum*, as per manufacturer's
128 recommendation. Honey bees were sampled from colonies maintained at the University of
129 California, Davis apiary and were collected from inside the colony using a handheld bee vacuum
130 between January 30th – February 13th 2023, 8:00am – 10:00am.

131 Bees were anesthetized via a 60 sec CO_2 exposure and then euthanized via decapitation.
132 Immediately following sacrifice, the mid-hindgut section was dissected from 100 bees (all
133 collected from the same colony) and pooled for same day virome extraction. Bees destined for
134 total metagenomic DNA extraction were collected the same day as virome samples and were
135 stored at -20°C pending mid-hindgut dissection on the following day. All dissections took place
136 in sterile PBS using ethanol and flame sterilized forceps. Each honey and bumble bee colony
137 generated one targeted virome and three total metagenomes (**Fig. 1**).
138
139 Sample preparation and DNA extraction
140 Phage enrichments for targeted virome sequencing were carried out following a protocol
141 adapted from previous publications[28,29,48] and is described fully in the supplement methods.
142 Briefly, bee gut pools were combined with 10 ml of Protein-supplemented phosphate-buffered
143 saline (PPBS: 2% BSA, 10% PBS, 1% K-citrate, 150 mM MgSO4), homogenized, and submitted
144 to a series of washes aimed at eluting phage particles. Phage were enriched via centrifugation,
145 0.22 um filtration, and ultracentrifugation. Lastly, phage pellets were resuspended in 200 uL
146 water and DNase treated.
147 For total metagenomes, mid-hind guts of individual bees were dissected and
148 homogenized and used for extraction. DNA was extracted from both virome and total
149 metagenome samples using a DNeasy PowerSoil Pro kit (Qiagen, Hilden, Germany) following
150 the manufacturer's instructions with the optional Qiagen Vortex Adapter 15 min bead beating
151 step.
152
153 Library construction and DNA sequencing

154 Extracted DNA from samples and kit blank negative controls was submitted to the
155 University of California Davis Genome Center for library prep and paired end 150 bp shotgun
156 metagenomic sequencing. Libraries were prepared using a Kapa Hyper prep kit (Kapa
157 Biosystems-Roche, Basel, Switzerland) with Illumina TruSeq adapters (Illumina, San Diego,
158 CA) and a target insert size of 350 bp. Sequencing took place on an Illumina NovaSeq. All
159 samples were sequenced on the same run.
160
161 Read processing, viral prediction, and vOTU table construction
162 Briefly, all demultiplexed *.fastq.gz* files from the Davis Genome Center were trimmed
163 and quality filtered using Trim-Galore[49] and Trimmomatic[50]. Reads aligning to negative samples
164 or the genomes of *A. mellifera* and *B. impatiens* were removed using Bowtie2[51]. Cleaned reads
165 were k-merized for complexity comparisons using Sourmash[52]. The program metaSPADES was
166 used to construct assemblies. Assemblies ≥5 kb were passed to Virsorter2[53] for phage prediction.
167 Phage with ≥95% nucleotide identity were collapsed into viral operational taxonomic units
168 (vOTUs) using CD-HIT[54] and were annotated with Prokka[55] using the PHROGs database[56].
169 vOTUs were classified as temperate when they were predicted to encode an integrase gene.
170 Finally, CoverM[57] was used to map cleaned reads against vOTUs. The resulting coverage table
171 was normalized to coverage per million reads.
172
173 Bacterial community and density measurements
174 Bacterial communities were predicted directly from cleaned total metagenomic reads
175 with Kraken2[58] and Bracken[59] using default parameters. Bacterial copy number was quantified
176 from the DNA extracts used for metagenome sequencing via a qPCR protocol adapted from

177 Christensen et al[60]. Each master mix solution contained: 5ul SSO Advanced Universal SYBR
178 Supermix 41 (Biorad, Hercules, CA, USA), 0.3ul of primers (10uM) targeting the 16S region of
179 the rRNA gene (799F= 5'-AACMGGATTAGATACCCKG-3';1115R= 37 5'-
180 AGGGTTGCGCTCGTTG-3'), 3.4ul Molecular grade water, and 1ul of template DNA (diluted
181 1:1000 in molecular grade water). Reactions were performed in triplicate for each sample.
182
183 Statistical and ecological analyses
184 Statistical analyses were conducted in R[61] v4.2.3. Briefly, alpha diversity, beta diversity,
185 and PERMANOVAs were compared between host and sampling method using vegan[62] and
186 phyloseq[63]. Linear mixed effect models were built using lme4[64], post-hoc tests were run using
187 emmeans[65], and T-tests were performed using base R. Gene-sharing networks were generated
188 outside of R using the program vConTACT2[66,67]. Genome alignment plots were constructed
189 using Clinker[68].
190 Phage-host predictions were performed using a custom CRISPR-spacer analysis. In short,
191 we built bacterial metagenome assembled genomes (MAGs) from our total metagenomic dataset
192 using metaSPADES and MetaBat2[69], combined these MAGs with NCBI assemblies of common
193 bee gut bacteria, mined all these sequences for CRISPRs using MINCED[70], and then aligned
194 those CRISPRs to our vOTUs using Blastn[71]. Lastly, to improve the number of vOTUs with host
195 assignments, we combined phage-host assignments generated by our CRISPR analysis with the
196 clusters predicted by our vConTACT2 network to infer the hosts of whole phage clusters,
197 following Rosso et al[45]. Bacterial copy number was compared between species using linear
198 models.
199

200 Results

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

202 Our sequencing effort generated a total of 419,884,845 raw read pairs, 2.47%
203 (10,385,967) of which came from extraction negatives. While the number of raw read pairs did
204 not differ significantly between samples (**Fig. S1;** Bee $t_{5.40}$=0.54 p=0.61, Type $t_{8.47e23}$=1.58
205 p=0.11, Bee:Type $t_{1.07e23}$=-1.38 p>0.17), there were differences in the number of high-quality
206 read pairs remaining after quality control (Bee $t_{5.81}$=2.87 p<0.05, Type $t_{2.75e25}$=-0.33 p=0.74,
207 Bee:Type $t_{2.75e25}$=4.69 p<0.001) (**Fig. 2A**). After filtering, honey bee viromes averaged 15.56
208 million high-quality read pairs per sample, while bumble bee viromes averaged only 5.49. Total
209 metagenome samples generated more similar numbers of high-quality reads per sample, with an
210 average of 8.59 and 5.85 million read pairs generated from honey and bumble bee samples,
211 respectively.

212

213 Bumble bee virome reads are less complex than other samples

214 Next, we measured the complexity of read libraries to further investigate how bee species
215 and sampling method influenced the data we generated. This was done by comparing the number
216 and occurrence of 31bp sized k-mers amongst each set of high-quality read libraries (**Fig. 2B**).
217 The results of this analysis suggest that, across the range of k-mer occurrences, bumble bee
218 viromes consistently contain fewer k-mers of a given occurrence than did other sample types.
219 Moreover, viewing only the y-intercept of this plot (**Fig. 2C**) shows that bumble bee viromes
220 contain a lower number of singleton k-mers, relative to all other sample types (Bee $t_{5.34}$=2.49
221 p=0.052, Type t_{16}=-6.47 p<0.001, t_{16}=5.14 p<0.001,). This decreased number of singleton k-

222 mers, and the tendency towards lower k-mer occurrences suggest that bumble bee viromes are
223 less complex compared to the other sample types in this dataset.
224
225 Targeted virome assemblies are enriched in phage
226 To evaluate the size and number of assemblies produced from each read library, and to
227 quantify the extent to which viromes enriched for phage sequences, we visualized the
228 distribution of assembly sizes in each sample, as well as the percentage of assemblies in a sample
229 predicted to encode phage sequences (**Fig. 2D**). A total of 4,278 and 37 contigs ≥5kb were
230 assembled from honey and bumble bee viromes, respectively. After passing to VirSorter2, these
231 contigs yielded 2,382 honey bee and 24 bumble bee putative phage sequences. Meanwhile,
232 honey and bumble bee total metagenomes produced 16,615 and 14,335 contigs ≥5kb,
233 respectively, which yielded 404 and 140 putative phage sequences. This means that 53.94% of
234 honey bee and 60.57% of bumble bee virome contigs ≥5kb were predicted to be viral, while only
235 2.42% and 2.13% of honey and bumble bee total metagenome contigs ≥5kb were annotated as
236 viral. The enrichment of phage sequences in viromes, relative to total metagenomes, provide
237 validation that viromes successfully enrich for phage when applied to bee guts.
238 Next, we used rarefaction to assess if the different number of putative phage sequences
239 recovered by honey and bumble bee viromes (**Fig. 2D**) was a result of differences in sampling
240 depth **(Fig. 2A**). This analysis shows that even at equal sampling depths, honey bee viromes
241 consistently recover >40x the number of phage recovered from bumble bee viromes (**Fig. S2A**
242 **and S2B**).
243
244 Phage target core bee bacteria and resemble previously described honey bee phage

245 After measuring the ability of viromes to enrich for phage sequences, we sought to
246 characterize the phage communities we detected in terms of similarity to previously described
247 bee phage, putative bacterial hosts, and predicted taxonomy. To do this, we first constructed
248 vOTUs by collapsing putative phage sequences at a 95% average nucleotide identity. This
249 reduced the number of putative phage from 2,950 to 2,560. We then built gene-sharing networks
250 comparing our vOTUs to those previously described by other bee phage studies (**Fig. 3A**). In this
251 network, many of the vOTUs from the current study clustered with those described by other bee
252 phage studies, suggesting they are related at roughly the genus level[66,67]. For example, viral
253 cluster 29 contains sequences recovered by the current study, Rosso et al[45], and Deboutte et al[46]
254 (**Fig. S3**). We interpret these results as further validating our computational methodologies and
255 further supporting the result that bumble bees host reduced diversity phage communities.
256 Next, we assigned bacterial hosts to our vOTUs. To focus this and all downstream
257 analyses on only the most abundant phage, we filtered vOTUs to retain only those with a
258 coverage ≥5, resulting in 2,481 vOTUs across 24 samples. Overall, we assigned hosts to 29.91%
259 (742) of the 2,481 vOTUs appearing in our dataset (**Fig. 3B**). In terms of relative abundance,
260 phage with assigned hosts made up an average of 89.34% and 44.75% of bumble bee viromes
261 and total metagenomes and 50.54% and 43.89% of honey bee viromes and total metagenomes,
262 respectively. Notably, the high proportion of phage with assigned hosts in bumble bee viromes is
263 likely confounded by the fact that these samples had an extraordinarily diminished phage
264 diversity.
265 Phage predicted to target the core social bee bacterial genera *Lactobacillus*,
266 *Bifidobacterium*, and *Gilliamella* were the most abundant of those with identified hosts in our
267 dataset (**Fig. 3B**). However, the relative abundance of phage targeting these genera varied by bee

268 species and sampling method (**Table S1)**. Phage targeting *Lactobacillus* were most abundant in
269 bumble bees, especially bumble bee viromes, while *Bifidobacterium* targeting phage were most
270 abundant in honey bee samples, particularly viromes. *Gilliamella* phage were not recovered in
271 bumble bee viromes, but were found in most bumble bee total metagenomes, and were
272 approximately 2x more abundant in honey bee total metagenomes than they were in honey bee
273 viromes. Phage targeting the core bumble bee symbiont *Schmidhempelia* were exclusively found
274 in bumble bee total metagenomes. We also found signatures of colony-level variation, for
275 example phage targeting *Bombilactobacillus* were detected in metagenomes from individuals
276 sampled from colony 1, but no individuals from colonies 2 or 3.
277 We assigned family level taxonomy to our phage if they clustered with reference phage in
278 gene-sharing networks. In total, we only assigned taxonomy to 9.83% (244) of 2,481 vOTUs
279 (**Fig. S4**). The large majority of classified vOTUs belonged to Siphoviridae (104), Myoviridae
280 (94), and Podoviridae (40), with a smaller number being classified as Autographiviridae (3) and
281 Zobellviridae (1) and multiple assignment (2).
282
283 <u>Bee species differ in phage diversity and composition</u>
284 Next we tested if vOTU diversity and community composition differed between bee
285 species and sampling method. While the diversity and richness of communities followed similar
286 trends (**Fig. 4A and 4B**), there were substantial quantitative differences between these two
287 metrics. When we compared Shannon's diversity, bee species, sampling method, and their
288 interaction all significantly affected vOTU diversity (**Fig. 4A**; Bee t_{20}=2.315, p=0.0313; Type
289 t_{20}=-5.992, p<0.001; Bee:Type t_{20}=8.004, p<0.001). Meanwhile, only the interaction between bee
290 species and sampling method significantly impacted community richness (**Fig. 4B**; Bee

291 t_{20}=0.690, p=0.498; Type t_{20}=-0.160, p=0.874; Bee:Type t_{20}=10.227, p<0.001). Overall, these
292 results suggest that sampling method can significantly influence the alpha diversity of inferred
293 phage communities, but that this can differ in magnitude and direction according to bee species.
294 In honey bees, 12.1% (290/2,392) of all vOTUs were shared between virome and total
295 metagenome samples. Similarly, in bumble bees, 12.9% (12/93) of vOTUs were detected by both
296 sampling methods. Meanwhile, only a few vOTUs were detected in both bee species (0.121%**;**
297 **Fig. 4C**). Additionally, vOTU community composition was predicted primarily by host bee
298 species, and secondarily based on sample method (**Figure 4D**; Bray-Curtis PERMANOVA Bee
299 R^2=0.143 $F_{1,20}$=4.17 p<0.001; Method R^2=0.0859 $F_{1,20}$=2.506 p<0.001; Bee:Method R^2=0.0859
300 $F_{1,20}$=2.505 p<0.001).
301
302 <u>Bacterial community variation predicts phage community variation</u>
303 We hypothesized that differences in phage richness and community composition between
304 bee species may be driven by differences in bacterial abundance and composition. First, we used
305 qPCR to compare the number of bacteria found in the guts of honey and bumble bees (**Fig. 5A**)
306 and found that the average mid-hindgut 16S gene copy number was significantly lower in
307 bumble bees than it was in honey bees ($t_{9.13}$=-3.54 p<0.01).
308 We then used Kraken2 and Bracken to estimate the diversity and taxonomic profile of
309 bacterial communities (**Fig. 5B, 5C, 5D**). At the species level, bumble bee bacterial communities
310 hosted lower bacterial richness and greater Shannon's diversity than honey bee bacterial
311 communities (Richness $t_{13.78}$=-6.27 p<0.001, Shannon $t_{11.70}$=-6.27 p<0.001), suggesting fewer
312 bacterial taxa were present in the bumble bees we sampled, but that bumble bee bacterial
313 communities were more evenly divided amongst their constituent bacteria.

314 Lastly, we used a Mantel test to compare the relationship between bacterial and phage
315 community structure (**Fig. S5**). We found a high degree of similarity between the bacterial and
316 phage communities ($R=0.74$ $p<0.001$), meaning that individual bees with similar bacterial
317 communities also have similar phage communities.
318
319 Total metagenomes capture prophage but miss rare phage
320 After identifying factors which explain some of the host-specific differences in phage
321 communities, we sought to evaluate why sampling method (virome vs total metagenome) also
322 led to significant differences in inferred phage communities. For this, we first compared the gene
323 content of vOTUs recovered from viromes and total metagenomes. Because total metagenomes
324 are primarily comprised of bacterial DNA, we hypothesized that this sampling method would
325 enrich for temperate phage. Supporting this hypothesis, we found that 189 of the 2075 (9.11%)
326 vOTUs identified in viromes were predicted to encode an integrase gene, while 89 of the 406
327 (21.92%) vOTUs recovered from total metagenomes encoded an integrase (**Fig. S6**). Moreover,
328 integrase encoding phage had a higher relative abundance in total metagenomes than in viromes
329 (**Fig. 6A**; Bee $t_{14.813}=-0.625$ $p=0.541$; Type $t_{16}=2.845$ $p<0.05$; Bee:Type Type $t_{16}=0.463$ $p=0.650$).
330 These results suggest that, relative to viromes, total metagenomes are biased towards sampling
331 prophage.
332 Lastly, to examine how sampling methods differ in recovery of rare vs abundant phage,
333 we visualized the occupancy-abundance relationships of virome-specific vOTUs vs vOTUs
334 recovered by both methods (**Fig. 6B and 6C**). This analysis revealed that vOTUs identified in
335 total metagenomes were among the most abundant and prevalent vOTUs found in viromes,

suggesting that total metagenomes recover more abundant and prevalent phage, while potentially missing rarer and lower occupancy phage.

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 of phage 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

359 honey bees sampled from North America and Europe[45–47]. For example, *Siphoviridae*,
360 *Myoviridae* and *Podoviridae* were among the most common viral families that we and these
361 previous studies identified. Similarly, *Lactobacillus*, *Gilliamella*, *Bifidobacterium*, and other core
362 bee gut bacteria were among the most predicted phage hosts in our and previous studies.
363 Notably, while previous bee phage studies found a number of phage predicted to target
364 *Bartonella*, none of our phage were predicted to target this bacterial genus. This is likely because
365 no available *Bartonella apis* assemblies met the strict criteria for inclusion in our CRISPR spacer
366 analysis.

367 Delving deeper, our gene-sharing network and clinker analyses show that many of our
368 phage share not only large genomic regions with previously described honey bee phage, but that
369 gene synteny in these geographically disparate viral genomes is also conserved. This
370 conservation between honey bee phage communities is similar to what was observed by Busby et
371 al[47]. Likewise, Rosso et al[45] showed that the genomes of honey bee-associated phage sampled
372 from Europe are able to recruit reads from Japanese honey bee total metagenomic datasets.

373 Altogether, these findings support the idea that, similar to the highly conserved nature of
374 some bee gut bacterial communities[37,38,43,72], some features of bee gut phage communities are
375 also conserved between bees sampled across geographically disparate populations. Future work
376 should focus on identifying what features of phage communities, whether individual genes,
377 clusters of genes, or perhaps whole genomes, are conserved between different populations of
378 bees.

379

380 <u>Host bacterial communities predict phage communities</u>

381 When compared to free-foraging managed honey bees, the commercially produced bumble
382 bees we sampled hosted significantly lower diversity phage communities. Although differences in
383 species biology and colony size may play a role, we attribute this low phage diversity primarily to
384 the low density and diversity bacterial communities these bumble bees hosted, suggesting that
385 phage diversity and abundance may track bacterial diversity and abundance across hosts. However,
386 whether such results are generalizable to wild bumble bees is unclear. In our study, bacterial 16S
387 rRNA copy number in bumble bee guts ranged from 10^6 to 10^7, whereas previous work in the same
388 species reported 10^8 - 10^9 copies of the 16S rRNA gene[37,73–75]. Given we used standard rearing
389 methods and diet, we hypothesize that lower bacterial titer may be due to the age of bees we
390 sampled. Previous work in bumble bees shows that bacterial density increases with worker age,
391 saturating approximately four days after worker emergence[73]. Future work could examine how
392 phage abundance and composition changes through worker development and if this reflects
393 previous observations of bacterial succession with worker age[17]. More generally, these findings
394 suggest that features of host bacterial communities (i.e. density, diversity, and structure) may be
395 used to predict phage community features. This is further supported by our Mantel analysis
396 showing a correlation between bacterial and phage community dissimilarity.
397
398 Viromes and total metagenomes reveal complementary phage communities.
399 There are several possible reasons as to why the phage communities inferred by viromes
400 and total metagenomes differ. One explanation is the way we generated viromes and how this
401 influenced the specific population of phage sampled. Research by Hoyles et al.[76] has shown that
402 passing human fecal samples through a 0.22 um filter reduces the number of phage recovered by
403 nearly half. As such, size filtration may have led some phage which are present in our bee guts to

404 have been excluded from viromes. This size exclusion may explain why we, and previous virome
405 vs total metagenome studies[16–18], consistently show that some phage are only found in total
406 metagenomes. The particular types of phage which are being excluded likely further explain the
407 total metagenome and virome differences we observed. We found that putative temperate phage
408 were more abundant and prevalent in total metagenomes than they were in viromes. This agrees
409 with existing literature which suggests that size filtration can specifically remove integrated
410 lysogenic phage, large jumbo phage, and phage adhered to bacterial cells and particles[31,33,77,78].
411 While it is tempting to expand these results to different environments by stating total
412 metagenomes always sample temperate phage missed by viromes, work in other systems implies
413 a more nuanced reality. In agricultural soils, Medellin et al.[16] identified only three phages that
414 were present in total metagenomes and absent from viromes. This suggests that the degree to
415 which sample processing skews viromes differs by environment. As a result, we suggest that the
416 relative utility of viromes and total metagenomes likely depends on the environment being
417 sampled. Samples with a high abundance of temperate phage, such as host-associated systems
418 and low-biomass environments, may benefit from total metagenomes. This is similar to the
419 conclusions made by Kosmopoulos et al.[17] which suggest that the choice of viromes vs total
420 metagenomes should be environment specific.

421 While our results highlight the ability of total metagenomes to recover phage missed by
422 viromes, they also showcase the capacity of viromes to sample a greater diversity of phage.
423 Similar to previous comparisons of viromes and total metagenomes in human gut, soil, and
424 freshwater environments[17,18,29], our honey bee viromes recovered a substantially larger number
425 of phage than did total metagenomes. Further, the occupancy-abundance relationships examined
426 here and by Medellin et al.[16] show that total metagenomes tend to be biased towards sampling

427 the most abundant and prevalent phage, while viromes are more successful at sampling rare
428 phage. Taken together, these results support previous work documenting that viromes produce a
429 deeper sampling of phage communities compared to total metagenomes[31,79,80].
430

431 ## Conclusions

432 By comparing bioinformatic mining of total metagenomes and targeted viromes across
433 two bee species, we found that sampling method significantly affected inferred phage
434 communities, but that the directionality of these differences can depend on the host being
435 sampled. In general, total metagenomes tended to be biased towards sampling both putatively
436 temperate phage and phage which were highly abundant. In samples with a relatively high
437 biomass (e.g., honey bees), viromes produced a greater diversity of phage and a better sampling
438 of rare phage. In contrast, when applied to relatively low biomass samples (e.g., bumble bees),
439 total metagenomes captured a greater diversity of phage than did viromes. Regardless of sample
440 biomass, there were always phage unique to both viromes and total metagenomes. In conclusion,
441 we suggest that these methods are complementary and recommend that both be used to capture
442 the full diversity of phage present in a gut environment. However, given that virome sampling is
443 not always feasible (i.e. in the case of field collected insects) total metagenomes may serve to
444 sample phage communities with the caveat that they will preferentially sample dominant and
445 temperate phage.
446

447 ## Acknowledgements

448 We thank Dr. Elina L. Nino, Joseph Tauzer, and the UC Davis Harry H. Laidlaw Jr.
449 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 and the UC Davis Genome center for sequencing.

Conflict of interest

The authors have no conflict of interest to declare.

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 DLS, as well as Hatch Multistate funding to RLV and NSF DEB #1929516 to RLV.

Data Availability

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 is publicly available at https://github.com/dsbard/Bee_phage_2023.

References

1. Bergh, Ø., BØrsheim, K. Y., Bratbak, G. & Heldal, M. High abundance of viruses found in aquatic environments. *Nature* **340**, 467–468 (1989).
2. Chibani-Chennoufi, S., Bruttin, A., Dillmann, M. L. & Brüssow, H. Phage-host interaction: An ecological perspective. *Journal of Bacteriology* vol. 186 3677–3686 Preprint at https://doi.org/10.1128/JB.186.12.3677-3686.2004 (2004).
3. Keen, E. C. A century of phage research: Bacteriophages and the shaping of modern biology. *BioEssays* **37**, 6–9 (2015).
4. Braga, L. P. P. *et al.* Impact of phages on soil bacterial communities and nitrogen availability under different assembly scenarios. *Microbiome* **8**, (2020).
5. Trubl, G. *et al.* Soil Viruses Are Underexplored Players in Ecosystem Carbon Processing. *mSystems* **3**, 1–21 (2018).
6. Emerson, J. B. *et al.* Host-linked soil viral ecology along a permafrost thaw gradient. *Nat Microbiol* **3**, 870–880 (2018).
7. Adriaenssens, E. M. *et al.* Environmental drivers of viral community composition in Antarctic soils identified by viromics. *Microbiome* **5**, (2017).
8. Fuhrman, J. A. *Marine Viruses and Their Biogeochemical and Ecological Effects*. *NATURE* vol. 399 www.nature.com (1999).
9. Suttle, C. A. Marine viruses - Major players in the global ecosystem. *Nat Rev Microbiol* **5**, 801–812 (2007).
10. Evans, C. *et al.* Shift from Carbon Flow through the Microbial Loop to the Viral Shunt in Coastal Antarctic Waters during Austral Summer. (2021) doi:10.3390/microorganisms.

489 11. Danovaro, R. *et al.* Major viral impact on the functioning of benthic deep-sea ecosystems.
490 *Nature* **454**, 1084–1087 (2008).
491 12. Tsao, Y. F. *et al.* Phage morons play an important role in pseudomonas aeruginosa
492 phenotypes. *J Bacteriol* **200**, (2018).
493 13. Heyerhoff, B., Engelen, B. & Bunse, C. Auxiliary Metabolic Gene Functions in Pelagic
494 and Benthic Viruses of the Baltic Sea. *Front Microbiol* **13**, (2022).
495 14. Dutilh, B. E. *et al.* A highly abundant bacteriophage discovered in the unknown sequences
496 of human faecal metagenomes. *Nat Commun* **5**, (2014).
497 15. Townsend, E. M. *et al.* The Human Gut Phageome: Origins and Roles in the Human Gut
498 Microbiome. *Frontiers in Cellular and Infection Microbiology* vol. 11 Preprint at
499 https://doi.org/10.3389/fcimb.2021.643214 (2021).
500 16. Santos-Medellin, C. *et al.* Viromes outperform total metagenomes in revealing the
501 spatiotemporal patterns of agricultural soil viral communities. *ISME Journal* (2021)
502 doi:10.1038/s41396-021-00897-y.
503 17. Kosmopoulos, J. C., Klier, K. M., Langwig, M. V, Tran, P. Q. & Anantharaman, K.
504 Viromes vs. mixed community metagenomes: choice of method dictates interpretation of
505 viral community ecology. *bioRxiv* (2023) doi:10.1101/2023.10.15.562385.
506 18. Zhang, F. *et al.* Critical Assessment of Whole Genome and Viral Enrichment Shotgun
507 Metagenome on the Characterization of Stool Total Virome in Hepatocellular Carcinoma
508 Patients. *Viruses* **15**, (2022).
509 19. Sender, R., Fuchs, S. & Milo, R. Revised Estimates for the Number of Human and
510 Bacteria Cells in the Body. *PLoS Biol* **14**, (2016).

511 20. Blum, W. E. H., Zechmeister-Boltenstern, S. & Keiblinger, K. M. Does soil contribute to
512 the human gut microbiome? *Microorganisms* **7**, (2019).
513 21. Garren, M. & Azam, F. New method for counting bacteria associated with coral mucus.
514 *Appl Environ Microbiol* **76**, 6128–6133 (2010).
515 22. Silveira, C. B. & Rohwer, F. L. Piggyback-The-Winner in host-Associated microbial
516 Communities. *npj Biofilms and Microbiomes* vol. 2 Preprint at
517 https://doi.org/10.1038/npjbiofilms.2016.10 (2016).
518 23. Knowles, B. *et al.* Lytic to temperate switching of viral communities. *Nature* **531**, 466–
519 470 (2016).
520 24. Mirzaei, M. K. & Maurice, C. F. Ménage à trois in the human gut: Interactions between
521 host, bacteria and phages. *Nature Reviews Microbiology* vol. 15 397–408 Preprint at
522 https://doi.org/10.1038/nrmicro.2017.30 (2017).
523 25. Chin, W. H. *et al.* Bacteriophages evolve enhanced persistence to a mucosal surface.
524 (2022) doi:10.1073/pnas.
525 26. Nelson, A. R. *et al.* Wildfire-dependent changes in soil microbiome diversity and function.
526 *Nat Microbiol* **7**, 1419–1430 (2022).
527 27. Waller, A. S. *et al.* Classification and quantification of bacteriophage taxa in human gut
528 metagenomes. *ISME Journal* **8**, 1391–1402 (2014).
529 28. Göller, P. C., Haro-Moreno, J. M., Rodriguez-Valera, F., Loessner, M. J. & Gómez-Sanz,
530 E. Uncovering a hidden diversity: Optimized protocols for the extraction of dsDNA
531 bacteriophages from soil. *Microbiome* **8**, (2020).

532 29. Santos-Medellín, C. *et al.* Spatial turnover of soil viral populations and genotypes overlain
533 by cohesive responses to moisture in grasslands. *Proceedings of the National Academy of*
534 *Sciences* **119**, e2209132119 (2022).
535 30. Parras-Moltó, M., Rodríguez-Galet, A., Suárez-Rodríguez, P. & López-Bueno, A.
536 Evaluation of bias induced by viral enrichment and random amplification protocols in
537 metagenomic surveys of saliva DNA viruses. *Microbiome* **6**, (2018).
538 31. Garmaeva, S. *et al.* Studying the gut virome in the metagenomic era: Challenges and
539 perspectives. *BMC Biology* vol. 17 Preprint at https://doi.org/10.1186/s12915-019-0704-y
540 (2019).
541 32. Gregory, A. C. *et al.* The Gut Virome Database Reveals Age-Dependent Patterns of
542 Virome Diversity in the Human Gut. *Cell Host Microbe* **28**, 724-740.e8 (2020).
543 33. Avellaneda-Franco, L., Dahlman, S. & Barr, J. J. The gut virome and the relevance of
544 temperate phages in human health. *Frontiers in Cellular and Infection Microbiology* vol.
545 13 Preprint at https://doi.org/10.3389/fcimb.2023.1241058 (2023).
546 34. Chandler, J. A., Lang, J., Bhatnagar, S., Eisen, J. A. & Kopp, A. Bacterial communities of
547 diverse Drosophila species: Ecological context of a host-microbe model system. *PLoS*
548 *Genet* **7**, (2011).
549 35. Douglas, A. E. The molecular basis of bacterial-insect symbiosis. *Journal of Molecular*
550 *Biology* vol. 426 3830–3837 Preprint at https://doi.org/10.1016/j.jmb.2014.04.005 (2014).
551 36. Pernice, M., Simpson, S. J. & Ponton, F. Towards an integrated understanding of gut
552 microbiota using insects as model systems. *J Insect Physiol* **69**, 12–18 (2014).
553 37. Hammer, T. J., Le, E., Martin, A. N. & Moran, N. A. The gut microbiota of bumblebees.
554 *Insectes Soc* **68**, 287–301 (2021).

555 38. Kwong, W. K. & Moran, N. A. Gut microbial communities of social bees. *Nature Reviews*
556 *Microbiology* vol. 14 374–384 Preprint at https://doi.org/10.1038/nrmicro.2016.43 (2016).
557 39. Raymann, K., Shaffer, Z. & Moran, N. A. Antibiotic exposure perturbs the gut microbiota
558 and elevates mortality in honeybees. *PLoS Biol* **15**, 1–22 (2017).
559 40. Marceau, T., Archer, C. R., Bulson, L. & Wilfert, L. Dose-dependent effects of antibiotic
560 intake on Bombus Terrestris (Linnaeus, 1758) dietary intake, survival and parasite
561 infection prevalence. *J Invertebr Pathol* **182**, (2021).
562 41. Koch, H. & Schmid-Hempel, P. Socially transmitted gut microbiota protect bumble bees
563 against an intestinal parasite. *Proc Natl Acad Sci U S A* **108**, 19288–19292 (2011).
564 42. Engel, P. *et al.* The bee microbiome: Impact on bee health and model for evolution and
565 ecology of host-microbe interactions. *mBio* **7**, 1–9 (2016).
566 43. Zheng, H., Steele, M. I., Leonard, S. P., Motta, E. V. S. & Moran, N. A. Honey bees as
567 models for gut microbiota research. *Lab Anim (NY)* **47**, 317–325 (2018).
568 44. Kwong, W. K. Bee microbiomes go viral. *Proceedings of the National Academy of*
569 *Sciences* **117**, 11197–11199 (2020).
570 45. Bonilla-Rosso, G., Steiner, T., Wichmann, F., Bexkens, E. & Engel, P. Honey bees harbor
571 a diverse gut virome engaging in nested strain-level interactions with the microbiota. *Proc*
572 *Natl Acad Sci U S A* **117**, 7355–7362 (2020).
573 46. Deboutte, W. *et al.* Honey-bee-associated prokaryotic viral communities reveal wide viral
574 diversity and a profound metabolic coding potential. doi:10.1073/pnas.1921859117/-
575 /DCSupplemental.
576 47. Busby, T. J., Miller, C. R., Moran, N. A. & Van Leuven, J. T. Global Composition of the
577 Bacteriophage Community in Honey Bees. *mSystems* **7**, (2022).

578 48. ter Horst, A. M., Fudyma, J. D., Sones, J. L. & Emerson, J. B. Dispersal, habitat filtering,
579 and eco-evolutionary dynamics as drivers of local and global wetland viral biogeography.
580 *ISME J* (2023) doi:10.1038/s41396-023-01516-8.
581 49. Krueger, F. TrimGalore. *Github* https://github.com/FelixKrueger/TrimGalore.
582 50. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina
583 sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
584 51. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods*
585 **9**, 357–359 (2012).
586 52. Pierce, N. T., Irber, L., Reiter, T., Brooks, P. & Brown, C. T. Large-scale sequence
587 comparisons with sourmash. *F1000Res* **8**, (2019).
588 53. Guo, J. *et al.* VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA
589 and RNA viruses. *Microbiome* **9**, 1–13 (2021).
590 54. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: Accelerated for clustering the next-
591 generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
592 55. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–
593 2069 (2014).
594 56. Terzian, P. *et al.* PHROG: Families of prokaryotic virus proteins clustered using remote
595 homology. *NAR Genom Bioinform* **3**, 1–12 (2021).
596 57. Woodcroft, B. J. CoverM. *Github* https://github.com/wwood/CoverM.
597 58. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2.
598 *Genome Biol* **20**, (2019).
599 59. Lu, J., Breitwieser, F. P., Thielen, P. & Salzberg, S. L. Bracken: Estimating species
600 abundance in metagenomics data. *PeerJ Comput Sci* **2017**, (2017).

601 60. Christensen, S. M. *et al.* Microbial Metamorphosis: Symbiotic bacteria and fungi
602 proliferate during diapause and may enhance overwintering survival in a solitary bee.
603 *bioRxiv* 2023.11.02.565352 (2023) doi:10.1101/2023.11.02.565352.
604 61. R Core Team. R: A Language and Environment for Statistical Computing. Preprint at
605 https://www.r-project.org/ (2019).
606 62. Oksanen, A. J. *et al.* Veganl: community ecology package. R package vegan, vers. 2.2-1.
607 (2015).
608 63. McMurdie, P. J. & Holmes, S. Phyloseq: An R Package for Reproducible Interactive
609 Analysis and Graphics of Microbiome Census Data. *PLoS One* **8**, (2013).
610 64. Bates, D., Mächler, M., Bolker, B. M. & Walker, S. C. Fitting linear mixed-effects models
611 using lme4. *J Stat Softw* **67**, (2015).
612 65. Russell V. Lenth. emmeans: Estimated Marginal Means, aka Least-Squares Means.
613 https://CRAN.R-project.org/package=emmeans.
614 66. Bolduc, B. *et al.* vConTACT: An iVirus tool to classify double-stranded DNA viruses that
615 infect Archaea and Bacteria. *PeerJ* **2017**, (2017).
616 67. Bin Jang, H. *et al.* Taxonomic assignment of uncultivated prokaryotic virus genomes is
617 enabled by gene-sharing networks. *Nat Biotechnol* **37**, 632–639 (2019).
618 68. Gilchrist, C. L. M. & Chooi, Y. H. Clinker & clustermap.js: Automatic generation of gene
619 cluster comparison figures. *Bioinformatics* **37**, 2473–2475 (2021).
620 69. Kang, D. D. *et al.* MetaBAT 2: An adaptive binning algorithm for robust and efficient
621 genome reconstruction from metagenome assemblies. *PeerJ* **2019**, (2019).
622 70. Skennerton, C. T. Minced. *Github* https://github.com/ctSkennerton/minced.

623 71. Camacho, C. *et al.* BLAST+: Architecture and applications. *BMC Bioinformatics* **10**,
624 (2009).
625 72. Handy, M. Y. *et al.* Incipiently social carpenter bees (Xylocopa) host distinctive gut
626 bacterial communities and display geographical structure as revealed by full-length
627 PacBio 16S rRNA sequencing. *Mol Ecol* **32**, 1530–1543 (2023).
628 73. Hammer, T. J., Easton-Calabria, A. & Moran, N. A. Microbiome assembly and
629 maintenance across the lifespan of bumble bee workers. *Mol Ecol* **32**, 724–740 (2023).
630 74. Wang, L. *et al.* Dynamic Changes of Gut Microbial Communities of Bumble Bee Queens
631 through Important Life Stages. (2019) doi:10.1128/mSystems.
632 75. Kwong, W. K. *et al.* Dynamic microbiome evolution in social bees. *Sci Adv* **3**, e1600513
633 (2017).
634 76. Hoyles, L. *et al.* Characterization of virus-like particles associated with the human faecal
635 and caecal microbiota. *Res Microbiol* **165**, 803–812 (2014).
636 77. López-Pérez, M., Haro-Moreno, J. M., Gonzalez-Serrano, R., Parras-Moltó, M. &
637 Rodriguez-Valera, F. Genome diversity of marine phages recovered from Mediterranean
638 metagenomes: Size matters. *PLoS Genet* **13**, (2017).
639 78. Gregory, A. C. *et al.* The Gut Virome Database Reveals Age-Dependent Patterns of
640 Virome Diversity in the Human Gut. *Cell Host Microbe* **28**, 724-740.e8 (2020).
641 79. Emerson, J. B. Soil Viruses: A New Hope. *mSystems* **4**, (2019).
642 80. Trubl, G., Hyman, P., Roux, S. & Abedon, S. T. Coming-of-age characterization of soil
643 viruses: A user's guide to virus isolation, detection within metagenomes, and viromics.
644 *Soil Systems* vol. 4 1–34 Preprint at https://doi.org/10.3390/soilsystems4020023 (2020).
645
646

Main Figures

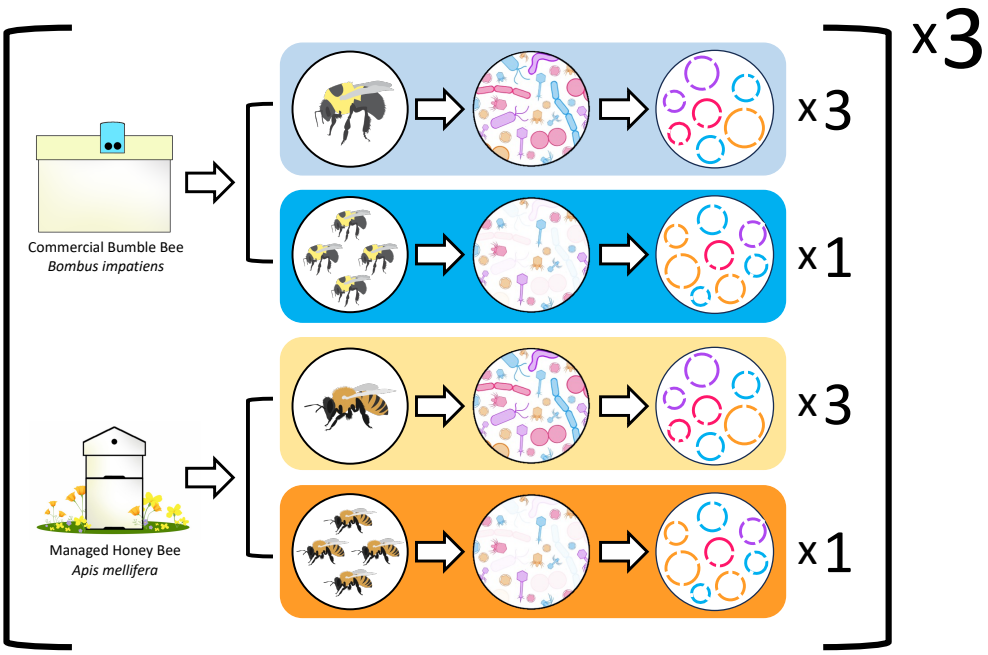

Figure 1: Graphical representation of the sampling scheme and methods used in this research. A total of three bumble bee and three honey bee colonies were sampled. From each colony of bee, we generated three total metagenomes and one targeted virome. Total metagenomes were sampled from individual bees, 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).

649

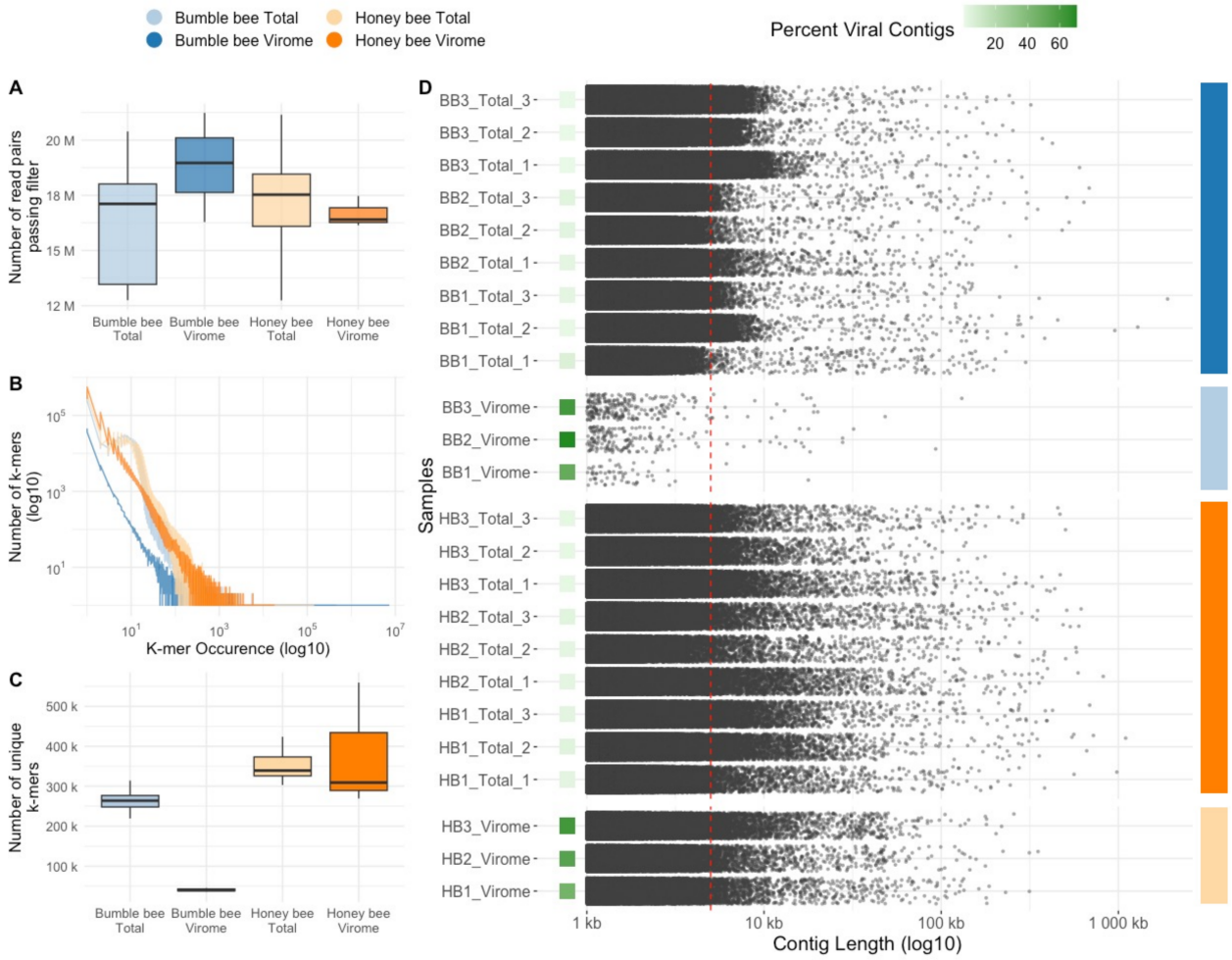

Figure 2: Figures describing sequencing and assembly quality. Color denotes sample groups. **A)** Boxplots describing sequencing depth. The x-axis shows sample groups. The y-axis displays the total number of high-quality reads produced. **B)** Line plots describing sequence complexity of the reads shown in panel **A**, as measured by the frequency distributions of 31bp 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 31bp 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 >= 1Kbp bp are shown. Individual samples are shown on the y-axis. The x-axis shows contig length. Each point represents a single contig. A dotted red line is drawn at 5Kbp. For each sample, a green square describes phage enrichment (the number of phage identified, divided by number of contigs assessed for phage, times 100).

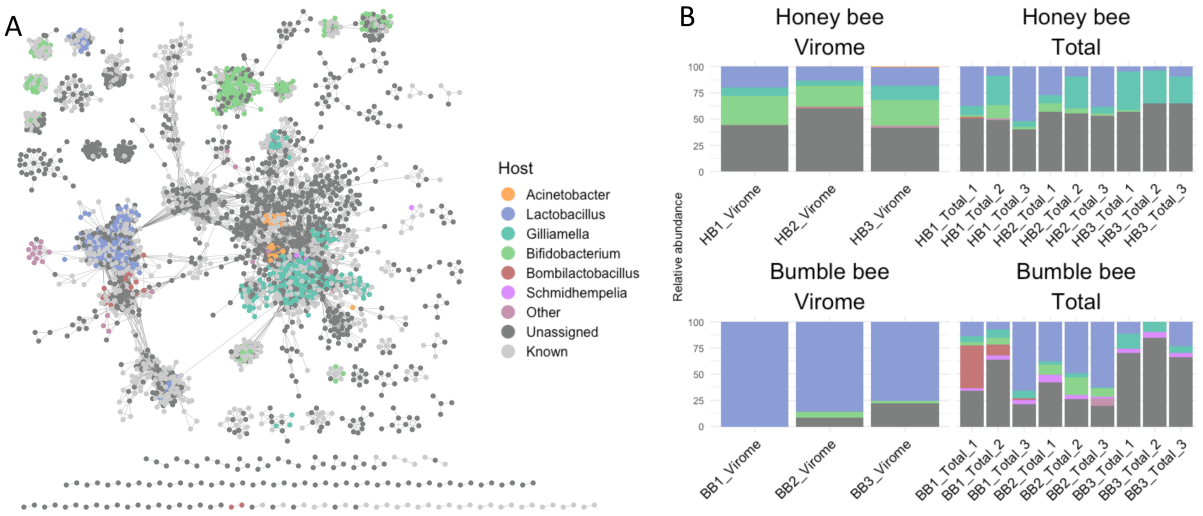

Figure 3: Figures describing phage identified in our dataset, relative to those identified previously, and the overall phage communities found in individual samples. **A)** Weighted gene-sharing network of all non-singleton vOTUs identified in our study. Individual nodes are phage. Nodes are connected by edges when phage share genes. Nodes are colored based on the predicted bacterial host of phage. Phage with "Known" as a predicted host are previously described bee phage. **B)** Stacked bar-plot describing the community of phage found in each sample based on predicted host. Both panel **A** and **B** use the same color palette.

650

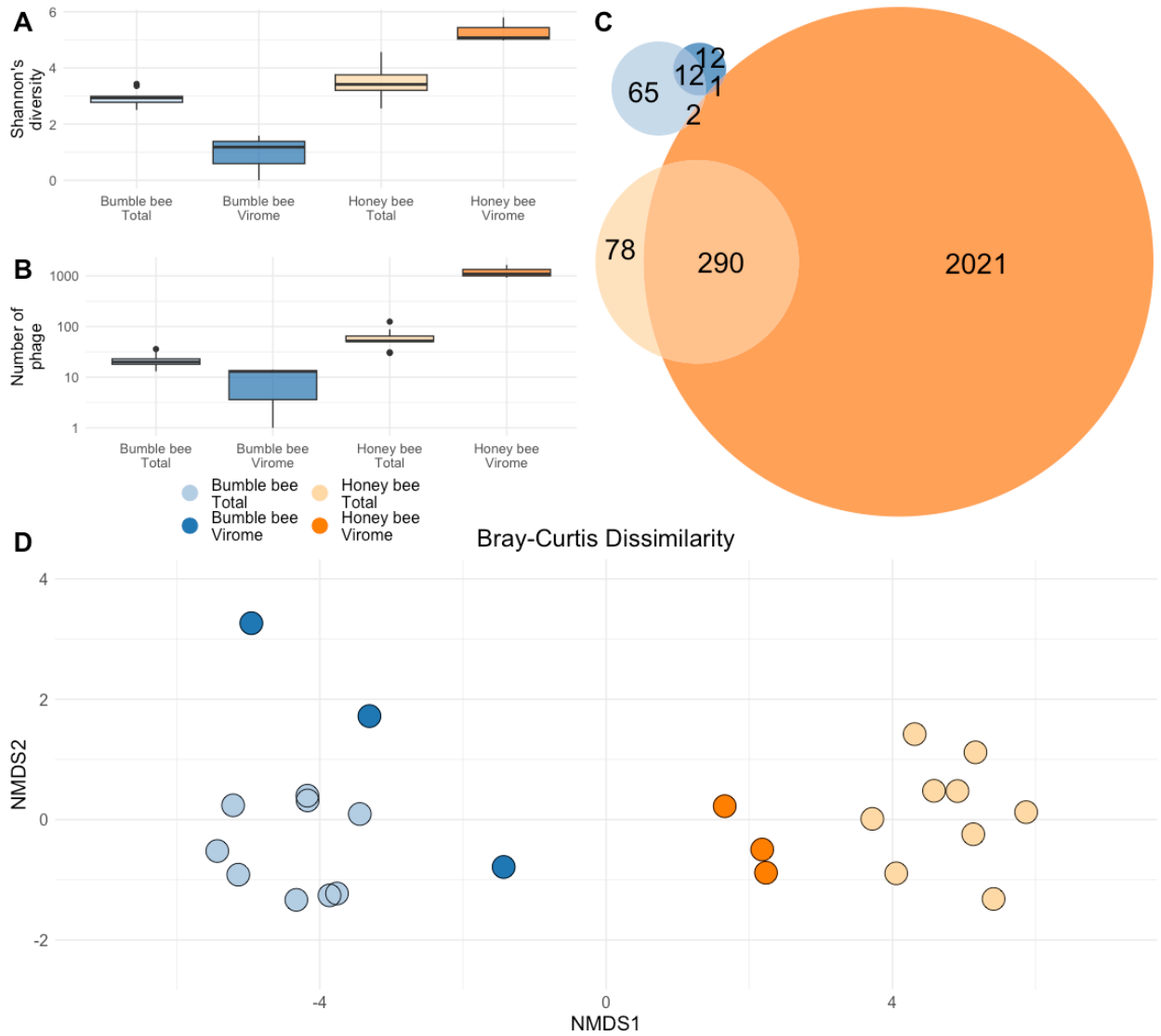

Figure 4: Figures describing the alpha and beta diversity of the phage communities identified in our samples. Blue represent samples taken from bumble bees. Orange represents samples taken from honey bees. Lighter colors are total metagenomes. Darker colors are viromes. **A)** Boxplots describing the **A)** Shannon's diversity and **B)** vOTU richness associated with each of our sample types. The x-axis groups 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. 290 phage were found in both honey bee viromes and total metagenomes). Circle size is proportional to number of phage. **C)** Non-metric multidimensional scaling (NMDS) ordination describing the Bray-Curtis dissimilarity of all samples on our dataset. Individual points are samples. Color represents sample type. The closer two points are to each other, the more similar they are in phage community.

651

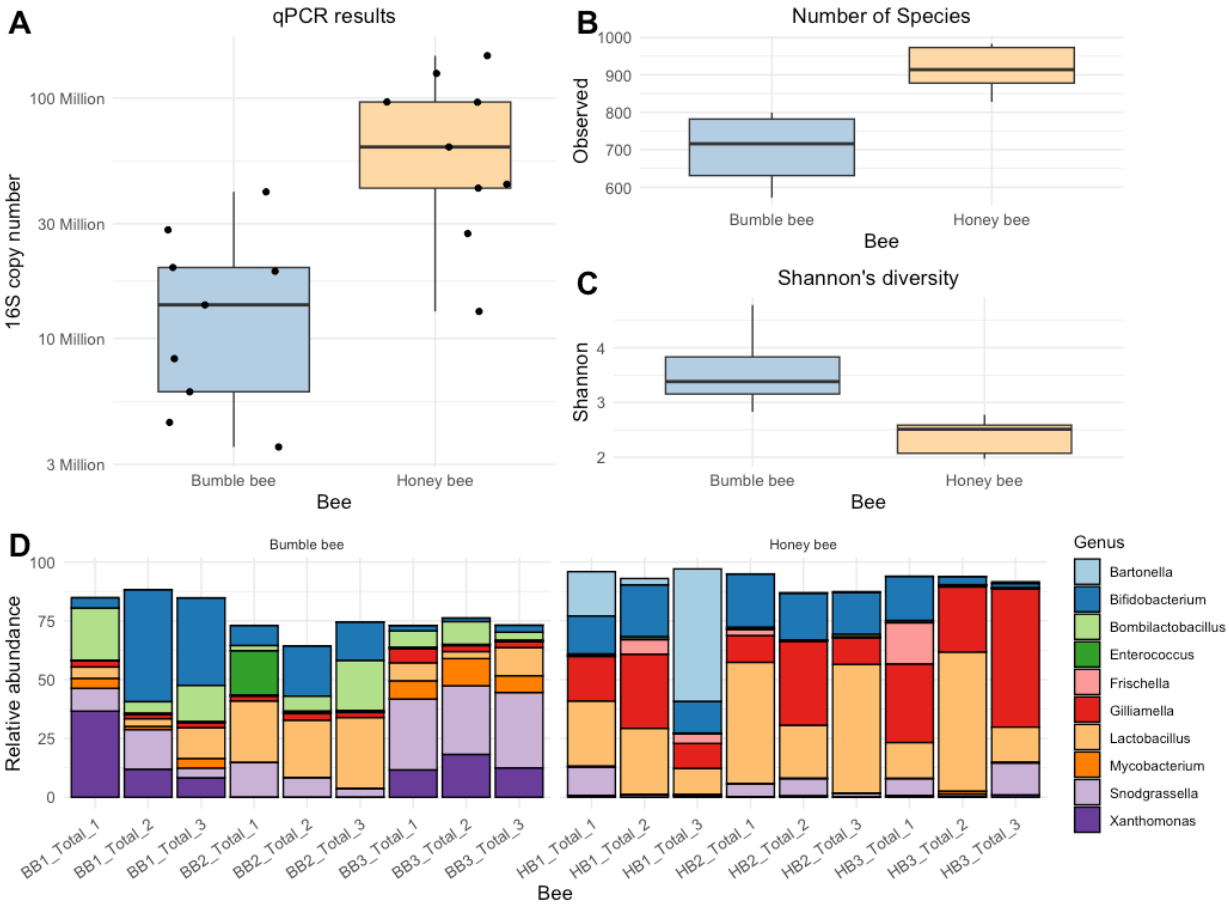

Figure 5: Honey and bumble bees differ in bacterial community density, diversity, and composition. **A)** Boxplots displaying the 16S rRNA gene copy number present in the mid-hindgut region of sampled honey and bumble bees. **B)** Boxplot presenting the number of observed bacterial species in honey and 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.

652

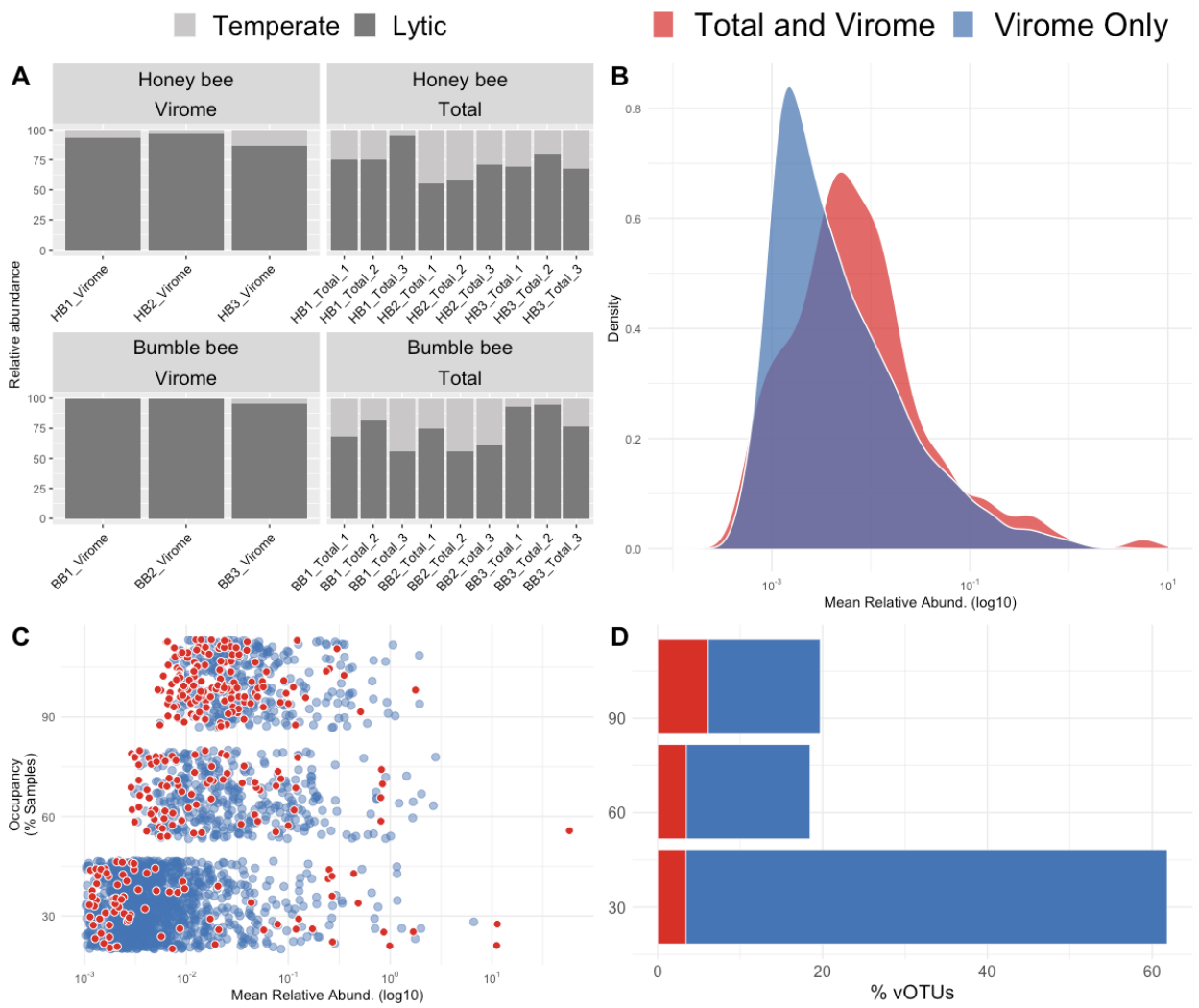

Figure 6: vOTUS sampled by viromes and total metagenomes differ in integrase content, occupancy, and abundance. **A)** Stacked bar plot showing the relative abundance of temperate and lytic vOTUs in virome and total metagenome samples. **B)** Density distributions showing the average relative abundance 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 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

653