Characterization of prophages in bacterial genomes from the honey bee (*Apis mellifera*) gut microbiome

Emma K Bueren [Corresp., 1], Alaina R Weinheimer [1], Frank O Aylward [1], Bryan B Hsu [1], David C Haak [2], Lisa K Belden [1]

[1] Department of Biological Sciences, Virginia Polytechnic Institute and State University (Virginia Tech), Blacksburg, VA, United States

[2] School of Plant and Environmental Sciences, Virginia Polytechnic Institute and State University (Virginia Tech), Blacksburg, VA, United States

Corresponding Author: Emma K Bueren
Email address: ebueren@vt.edu

The gut of the European honey bee (*Apis mellifera*) possesses a relatively simple bacterial community, but little is known about its community of prophages (temperate bacteriophages integrated into the bacterial genome). Although prophages may eventually begin replicating and kill their bacterial hosts, they can also sometimes be beneficial for their hosts by conferring protection from other phage infections or encoding genes in metabolic pathways and for toxins. In this study, we explored prophages in 17 species of core bacteria in the honey bee gut and two honey bee pathogens. Out of the 181 genomes examined, 431 putative prophage regions were predicted. Among core gut bacteria, the number of prophages per genome ranged from 0 – 7 and prophage composition (the compositional percentage of each bacterial genome attributable to prophages) ranged from 0 – 7%. *Snodgrassella alvi* and *Gilliamella apicola* had the highest median prophages per genome (3.0 ± 1.46; 3.0 ± 1.59), as well as the highest prophage composition (2.58% ± 1.4; 3.0% ± 1.59). The pathogen *Paenibacillus larvae* had a higher median number of prophages (8.0 ± 5.33) and prophage composition (6.40% ± 3.08) than the pathogen *Melissococcus plutonius* or any of the core bacteria. Prophage populations were highly specific to their bacterial host species, suggesting most prophages were acquired recently relative to the divergence of these bacterial groups. Furthermore, functional annotation of the predicted genes encoded within the prophage regions indicates that some prophages in the honey bee gut encode additional benefits to their bacterial hosts, such as genes in carbohydrate metabolism. Collectively, this survey suggests that prophages within the honey bee gut may contribute to the maintenance and stability of the honey bee gut microbiome and potentially modulate specific members of the bacterial community, particularly *S. alvi* and *G. apicola.*

1 Characterization of prophages in bacterial genomes from the honey bee (*Apis*
2 *mellifera*) gut microbiome

3 Emma K. Bueren[1], Alaina R. Weinheimer[1], Frank O. Aylward[1], Bryan B. Hsu[1], David C. Haak[2],
4 and Lisa K. Belden[1]

6 [1]Department of Biological Sciences, Virginia Tech, Blacksburg, VA, USA

7 [2]School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA, USA

9 Corresponding Author:

10 Emma K. Bueren[1]

11 Virginia Tech, Dept of Biological Sciences
12 2119 Derring Hall, 926 West Campus Drive
13 Blacksburg VA 24061, United States
14
15 Email: ebueren@vt.edu

Abstract

35 The gut of the European honey bee (*Apis mellifera)* possesses a relatively simple bacterial
36 community, but little is known about its community of prophages (temperate bacteriophages
37 integrated into the bacterial genome). Although prophages may eventually begin replicating and
38 kill their bacterial hosts, they can also sometimes be beneficial for their hosts by conferring
39 protection from other phage infections or encoding genes in metabolic pathways and for toxins. In
40 this study, we explored prophages in 17 species of core bacteria in the honey bee gut and two
41 honey bee pathogens. Out of the 181 genomes examined, 431 putative prophage regions were
42 predicted. Among core gut bacteria, the number of prophages per genome ranged from 0 – 7 and
43 prophage composition (the compositional percentage of each bacterial genome attributable to
44 prophages) ranged from 0 – 7%. *Snodgrassella alvi* and *Gilliamella apicola* had the highest median
45 prophages per genome (3.0 ± 1.46; 3.0 ± 1.59), as well as the highest prophage composition (2.58%
46 ± 1.4; $3.0\% \pm 1.59$). The pathogen *Paenibacillus larvae* had a higher median number of prophages
47 (8.0 ± 5.33) and prophage composition ($6.40\% \pm 3.08$) than the pathogen *Melissococcus plutonius*
48 or any of the core bacteria. Prophage populations were highly specific to their bacterial host
49 species, suggesting most prophages were acquired recently relative to the divergence of these
50 bacterial groups. Furthermore, functional annotation of the predicted genes encoded within the
51 prophage regions indicates that some prophages in the honey bee gut encode additional benefits to
52 their bacterial hosts, such as genes in carbohydrate metabolism. Collectively, this survey suggests
53 that prophages within the honey bee gut may contribute to the maintenance and stability of the
54 honey bee gut microbiome and potentially modulate specific members of the bacterial community,
55 particularly *S. alvi* and *G. apicola*.

Introduction

58 Bacteriophages, viruses that infect bacteria, are ubiquitous and the most numerous biological
59 entities on Earth (Hendrix, 2002). In animals, bacteriophages are known to shape microbial
60 communities, such as the gut or skin microbiome (Hannigan et al., 2015). Bacteriophages are
61 broadly classified in two forms, based on reproductive strategy. Lytic bacteriophages infect and
62 immediately kill their bacterial hosts via cell lysis. In contrast, temperate bacteriophages can
63 undergo either a lytic cycle or a lysogenic cycle, in which the phage integrates into the host genome
64 as a prophage and replicates along with the hosts genome until its triggered to revert to a lytic
65 lifecycle. Many phages found in animal microbial communities are lysogenic (Minot et al., 2011;
66 Kim & Bae, 2018). Some prophages encode genes in auxiliary metabolic pathways or encode non-
67 metabolic accessory genes ("morons"), like virulence factors, that are beneficial to their bacterial
68 hosts and can enhance bacterial fitness (Fortier & Sekulovic, 2013; Forcone et al., 2021; Huang et
69 al., 2021). For example, the human pathogens *Vibrio cholera* and *E. coli* O157:H7 can cause
70 disease via toxins encoded by prophages (Fortier & Sekulovic, 2013).

71 While prophages in pathogenic bacteria have been extensively studied, the roles of prophages in
72 other animal-associated bacteria are less clear. It is generally difficult to characterize phage-
73 bacteria interactions in the mammalian gut due to the complexity of these microbial communities,
74 although inroads have been made in murine models (Hsu et al., 2019). In contrast, the bacterial
75 community in the gut of honey bees (*Apis mellifera*) is significantly more constrained, consisting
76 primarily of a core group of nine bacterial phylotypes: *Bartonella apis, Bombella apis (*previously
77 known as *Parasaccharibacter apium), Frischella perrara, Snodgrassella alvi, Bifidobacterium*
78 spp*., Bombilactobacillus* spp*.* (also referred to as *Lactobacillus* Firmicutes-4), *Lactobacillus*

79 Firmicutes-5, *Gilliamella apicola,* and *Commensalibacter* spp. (also known as Alpha2.1) (Kwong
80 & Moran, 2016). Some of these core microbiota, like *Gilliamella apicola, Lactobacillus*
81 Firmicutes-5, *Bomblicatobacilus* spp., and *Bifidobacterium* spp., ferment a wide variety of
82 carbohydrates, while others, like *F. perrara,* stimulate bee immune function (Kwong & Moran,
83 2016; Emery, Schmidt & Engel, 2017).

84 Studies using both culture and sequence-based approaches reveal the numerous roles of phages in
85 honey bee-associated bacteria. For instance, some strains of the pathogen *Paenibacllus larvae,*
86 which causes American Foulbrood, acquire additional virulence through the prophage encoded
87 toxin *P1x1* (Ebeling, Fünfhaus & Genersch, 2021). Additionally, many phage targeting *P. larvae*
88 have been isolated and characterized for the purpose of experimental phage therapy (Beims et al.,
89 2015; Abraham et al., 2016; Stamereilers et al., 2018; Ribeiro et al., 2019). To capture the broader
90 diversity of phages present in honey bee guts and include those that cannot be easily isolated,
91 recent studies have used metagenomic surveys of viruses, in which all double-stranded DNA
92 viruses in a sample are sequenced to characterize the "virome". These virome studies have shown
93 that the phage community (consisting of both temperate and lytic phages) is far more diverse than
94 the bacterial community and appears to encode auxiliary metabolic genes (Bonilla-Rosso et al.,
95 2020; Deboutte et al., 2020; Busby et al., 2022). The majority of phages in these viromes were
96 novel, and thus were not taxonomically classified. Those that were classified were primarily from
97 the recently dissolved *Podoviridae, Myoviridae,* and *Siphoviridae* families (Bonilla-Rosso et al.,
98 2020; Deboutte et al., 2020), leaving their current taxonomic standing unknown. However,
99 viromes may not fully capture the phages present in the bee gut that exist primarily as prophages
100 because the preparation of viromes involves filtering out bacterial cells to enrich for free viral
101 particles (Huang et al., 2021).

102 As prophages can both protect the bacterial host from infection by other phages and improve host
103 fitness by expanded metabolic versatility, they have the potential to impact the physiology,
104 defense, and evolution of core honey bee gut bacteria (Deboutte et al., 2020). To understand the
105 roles of prophages in the regulation and function of honey bee gut bacteria, we examined publicly
106 available honey bee gut bacterial genomes for prophage regions and compared the occurrence of
107 prophages among bacterial species. We then assessed prophage diversity with pairwise average
108 nucleotide identity calculations and gene-sharing networks with reference phages. To explore
109 potential benefits the prophages confer their hosts, we functionally annotated their genes and
110 searched for those in relevant metabolic pathways.

111

Materials and Methods

113 **Bacterial Genomes.** Publicly available genomes (complete and fragmented) of bacteria originally
114 isolated from the gut of the honey bee, *A. mellifera*, were downloaded from GenBank between
115 March and July, 2020 (n = 181) (Table 1, Table S1). The nine core bacterial phylotypes are
116 represented in this dataset (n = 151 total strains across 17 species), as were two bacterial pathogens
117 (n=30). Each genome was from a uniquely named strain, and in the case of duplicate strain names
118 in NCBI, the most complete and recent genome was used.

119 **Prophage Identification.** To detect putative prophages, we used a combination of VirSorter2
120 (v2.2.2; (Guo et al., 2021, p. 2)), CheckV (v.0.7.0 (Nayfach et al., 2021)) and VIBRANT (v1.2.1;
121 (Kieft, Zhou & Anantharaman, 2020)). First, all genomes were analyzed with VirSorter2 (settings
122 --include-groups "dsDNAphage,ssDNA,NCLDV,laviviridae"). Resulting viral regions were

123 retained if they scored at least 0.5 for double stranded DNA phage (dsDNA phage, n = 539). Host
124 genome regions flanking these viral sequences were then trimmed with the CheckV
125 'contamination' command. To exclude highly degraded phages, trimmed sequences were retained
126 only if they were at least 5 kb in length (n = 462). Of these resulting sequences, a region was
127 considered a putative prophage if it scored at least 0.9 with VirSorter2 (n = 357). Additionally,
128 those with VirSorter scores of 0.5-0.9 were further analyzed with VIBRANT and were retained as
129 putative prophage if VIBRANT also classified these sequences as virus (n = 74). This resulted in
130 a final set of 431 putative prophages.

131 **Prophage Abundance and Prophage Composition in Honey Bee Bacterial Isolates.** Two
132 metrics for prophage presence were assessed: the absolute abundance of prophages (total number
133 of distinct regions) in a single bacterial genome and the portion of the bacterial genome that was
134 composed of prophage (referred to as prophage composition). Differences in prophage abundance
135 and composition among bacterial species were tested using Kruskal-Wallis tests. All graphs were
136 visualized using ggplot2 and PNWColors in R (Wickham, 2016; Lawlor, 2020).

137 Larger host genomes often contain more prophage regions (Touchon, Bernheim & Rocha, 2016).
138 We examined this pattern with host genome size vs. prophage abundance and host genome size
139 vs. prophage composition using Spearman's correlation with the R library package cocor
140 (Diedenhofen & Musch, 2015) and the Meng, Rosenthal and Rubin's z test (Meng, Rosenthal &
141 Rubin, 1992; Myers & Sirois).

142 **Functional Annotation of Prophage Region Genes and Detection of Intact Phages.** The 431
143 putative prophages were first annotated using prokka (v. 1.14.6) against the Prokaryotic Virus
144 Remote Homologous Groups database (PHROGs) database (downloaded from
145 http://millardlab.org/2021/11/21/phage-annotation-with-phrogs/ on July 20, 2022) (Seemann,
146 2014; Terzian et al., 2021). If an amino acid sequence had no hit to the PHROGs database (e value
147 $< 10^{-6}$), or was classified by PHROGs as "other" or "unknown function", additional functions were
148 then detected by aligning the amino acid sequences to hidden Markov profiles from the EggNOG
149 5.0 databases for bacteria, eukarya, archaea, and viruses (Huerta-Cepas et al., 2019) using
150 hmmsearch (-E 0.001) with HMMER (v. 3.1b2) (hmmer.org, (Eddy, 2011)). Hits with bitscores
151 above 30 were retained (Roux et al., 2016).

152 Functional composition of prophage genes were compared based on the COG category assignment
153 within the EggNOG annotation files for bacteria, eukarya, archaea, and viruses databases.
154 EggNOG hits were manually re-assigned to the category of "Phage-associated" if the descriptive
155 EggNOG name included the words: phage, baseplate, capsid, integrase, tail, tape, lysozyme, portal,
156 holin, N-Acetylmuramoyl-L-alanine amidase, transposase, virus, or viral. Sequences that did not
157 match to anything in the EggNOG database, or that were classified as belonging to multiple COG
158 categories, were re-classified as COG Category S, Function Unknown. Functional composition of
159 prophages using the combined results of PHROGs and EggNOG were then compared among
160 bacterial host species by summing the proteins of all prophages detected in genomes of that species
161 and visualized with the packages ggplot2 and microshades (Wickham, 2016; Dahl et al., 2022).

162 To estimate intactness of prophage regions, trimmed sequences were predicted as intact using the
163 following (i) those identified as intact by PHASTER (Arndt et al., 2016), (ii) those that encoded
164 at least 3 or more different phage hallmark genes (see: cornerstone genes in (Zhou et al., 2011))
165 and additionally encoded an integrase and/or transposase.

166 **Identification of Unique Putative Prophages.** To identify prophage regions likely belonging to
167 the same population (boundary defined in (Roux et al., 2019)), genomes of the 431 putative
168 prophages were dereplicated via pairwise average nucleotide identity (ANI) analysis using dRep
169 (v3.0.0 (Olm et al., 2017)) (settings: --ignoreGenomeQuality -l 5000 -pa 0.95 -sa 0.95 -nc 0.85)
170 with LASTn (v1080; (Kiełbasa et al., 2011)). Prophage pairs were considered to be from the same
171 population if the ANI was above 95% over 85% of the shorter sequence (Roux et al., 2019). The
172 representative prophage region for each population was selected with dRep. In one case, a
173 population cluster contained two prophages that were identified from different bacterial host
174 species. In that case, both prophage sequences were retained as representatives. As a result, a total
175 of 237 putative prophage regions were unique to themselves and did not share high similarity with
176 other sequences. The remaining 194 sequences had high similarity with at least one other prophage
177 sequence, resulting in 66 distinct populations of phage. After dereplication, a total of 303
178 representative phage were used for downstream functional annotation and taxonomic analysis.

179 **Viral Clustering and Classification.** Phages lack a universally conserved, high-resolution
180 phylogenetic marker gene to classify them. As such, prophages are often grouped into clusters via
181 gene-sharing network-based approaches or whole genome sequence alignments to reference phage
182 (Bolduc et al., 2017). Dereplicated prophages in this study (n = 303), along with 24,289 phage
183 reference genomes downloaded from the INfrastructure for a PHAge REference Database
184 (INPHARED) on January 25, 2023 (Cook et al., 2021) were first classified with gene-sharing
185 network approaches using vConTACT2 (v0.9.22; --rel-mode Diamond --db 'None' --pcs-mode
186 MCL --vcs-mode ClusterON). vContact2 (v0.9.22) groups phages into subclusters, indicating
187 phage are likely to be within the same viral genera, and broader viral clusters that indicate
188 subfamily relatedness (relatedness somewhere between family and genus level) (Bin Jang et al.,
189 2019). The resulting networks were visualized using edge-weighted spring-embedded models in
190 Cytoscape (Shannon et al., 2003; Bolduc et al., 2017; Bin Jang et al., 2019). Although
191 vConTACT2 can provide taxonomic classification if target phage cluster with reference phage, for
192 our dataset, few of the resulting clusters included references phages that enabled classification at
193 any taxonomic level. Therefore, while vConTACT2 gene-sharing networks were used to assess
194 the relationships among honey bee prophages, we ended up relying more on whole genome
195 alignments to reference phage to assign broader taxonomy.

196 To assign taxonomy based on similarities to reference viral groups, we compared the average
197 amino acid identity (AAI) of proteins between our dereplicated putative prophages and phage from
198 a dereplicated *Caudoviricetes* database (amino acids downloaded from the INPHARED on
199 November 5, 2022, taxonomic metadata downloaded from INPHARED on March 16, 2023) (Cook
200 et al., 2021). The 16,253 reference genomes were first dereplicated with dRep (--
201 ignoreGenomeQuality -p 32 -l 20000 -pa 0.90 –SkipSecondary inphared_derepout_90mash)
202 resulting in 2,600 representative reference genomes. Amino acid sequences of proteins encoded
203 by the reference were predicted using prodigal (default on each genome). A BLAST database was
204 made of the reference proteins (makeblastdb). The honey bee gut prophages protein predictions
205 (produced via prodigal, described above) were aligned to this reference database with BLASTp (E
206 value 0.001). Taxonomic class was assigned at the most precise rank possible using each
207 prophage's AAI similarity to its top hitting reference phage. If a putative prophage and a reference
208 phage shared greater than 70% AAI across 85% of proteins, it was classified as belonging to the
209 same genus as the reference (Turner, Kropinski & Adriaenssens, 2021). If a phage did not meet
210 this threshold but still shared at least 30% AAI over 50% of proteins, it was classified as the same
211 family as the reference phage (Turner, Kropinski & Adriaenssens, 2021). If a prophage could not

212 be classified using the above thresholds, but its top reference hit was to a *Caudoviricetes*
213 class phage, it was classified broadly as *Caudoviricetes*. Any putative prophage with a top
214 reference hit to a completely unclassified phage was labeled as Unclassified.

Results

216 **Distribution of Prophages Across Core Bacterial Hosts.** Bacterial host species significantly
217 varied in both the number and composition of prophages across all bacterial isolates, both core and
218 pathogens (Kruskal-Wallis, p < 0.0001; Fig. 1 & Table S2). A total of 269 high-confidence
219 prophage regions were predicted across the 151 core bacterial isolates before dereplication.
220 Bacterial genomes ranged in size from 1.3 Mb – 3.5 Mb (Table S2). Among the core bacterial
221 species, the number of prophages per isolate varied from 0 to 7 and prophage composition of the
222 bacterial genome varied from 0 – 7.01% (Table S2).

223 The core bacteria with the highest median number of prophages and prophage composition was *S.*
224 *alvi* (3.0 ± 1.46 prophages; 2.58% ± 1.40 prophage composition) and *G. apicola (*3.0 ± 1.58
225 prophages; 2.04% ± 1.48 prophage composition) (Fig. 1A, Fig. 1B, Table S2). Notably, while
226 *Bombella apis, Bartonella apis,* and *F. perrara* had the second highest median number of
227 prophages per genome, *L. melliventris* had the second highest prophage composition (Fig. 1A; Fig.
228 1B; Table S2). All other core species had a median number of 1 prophage or less and a prophage
229 composition of less than 1% (Fig. 1A; Fig. 1B, Table S2). These included *Bifidobacterium*
230 *asteroides, Lactobacillus mellifer, Lactobacillus apis,* and *Lactobacillus kimbladii,* which had a
231 median of 0 prophages per genome (Fig. 1A; Fig. 1B; Table S2).

232 Of these 269 putative prophage regions, 90 were estimated to be intact. Bacterial hosts with the
233 highest median number of estimated intact prophages were *G. apicola (*1.0 ± 1.1), *S. alvi* (1.0 ±
234 0.60) and *L. melliventris* (1 ± 0.58) (Fig. 1C; Table S2). All three isolates of *L. melliventris,* 71%
235 of *G. apicola* isolates (n = 38) and 56% of *S. alvi* isolates (n = 34) possessed at least one estimated
236 intact prophage (Table S3). The second highest median of intact prophage was found in *L.*
237 *kullabergensis* (0.5 ± 0.71), with one of the two *L. kullabergensis* isolates possessing an intact
238 prophage (Fig. 1C; Table S3). All other core species had a median of 0 intact prophages (Fig. 1C,
239 Table S2). However, intact prophages did still occur in these isolates at lower frequencies (Table
240 S3). Intact prophage composition varied slightly compared to prophage abundance, with *L.*
241 *melliventris* having the highest composition (1.81% ± 1.25), followed by *S. alvi* (1.16% ± 1.09),
242 *L. kullabergensis* (0.97% ± 1.37)*,* and *G. apicola* (0.89% ± 1.31) (Fig. 1D; Table S2).

243 **Distribution of Prophages Across Pathogenic Bacterial Hosts.** A total of 162 prophage regions
244 were predicted across the 30 bacterial isolates from the two pathogens, *P. larvae* and *M. plutonius,*
245 before dereplication. The bacterial isolate genomes ranged in size from 2.0 Mb – 4.8 Mb *(*Table
246 S2). The number of predicted prophage regions ranged from 1 to 22, and prophage composition
247 ranged from 0.96% - 13.85% of the bacterial genome (Table S2). *P. larvae* had the highest median
248 number of prophages per genome (8.0 ± 5.33) and prophage composition (6.40% ± 3.08) across
249 all bacterial species, including the core bacteria (Fig. 1A; Fig. 1B; Table S2). The median number
250 of prophages found per genome of *M. plutonius* was 2.0 ± 0.64, and its median prophage
251 composition was 2.77% ± 0.83 (Fig. 1A; Fig. 1B; Table S2).

252 Of the 13 *P. larvae* isolates analyzed, all had at least one intact prophage, with a median of 4 ±
253 1.98 intact prophage per isolate (TableS-Freq; Fig. 1C; Table S2). *M. plutonius* had a median of 1
254 ± 0.35 intact prophage, with 16 of 17 isolates possessing an intact prophage (Fig. 1C; Table S2;

255   TableS-Freq). The total intact prophage composition of *P. larvae* (4.57 % ± 3.10) was higher than
256   *M. plutonius* (1.71 % ± 0.97) (Fig. 1D; Table S2).

257   **Prophage Abundance Is More Correlated with Bacterial Genome Size than Prophage**
258   **Composition.** Bacterial genome size was positively correlated with both the number of prophages
259   (Spearman's ρ = 0.55, p <= 1.6e-15, Fig. 2A) and prophage composition (Spearman's ρ = 0.34, p
260   = 2e-6, Fig. 2B). Meng, Rosenthal and Rubin's correlation comparison determined that bacterial
261   genome size had a stronger correlation with the number of prophages than with prophage
262   composition (z= 5.9819, p-value < 0.001, 95% CI: 0.18-0.35) (Meng, Rosenthal & Rubin, 1992).

263   **Prophage-encoded genes varied by bacterial host.** Nearly half (48.4%) of all predicted protein-
264   encoding genes of the dereplicated prophages were classified broadly as phage-associated. A total
265   of 2.3% of all genes were classified by PHROGs specifically as "Moron, auxiliary metabolic
266   genes, and host takeover" (Table S5). *S. alvi* had the highest frequency of genes in this category
267   (66 genes, 3.3% frequency), due to its large collection of phage toxins and antitoxins (Fig. 3).
268   These toxins shared homology to Doc-like toxins (10 genes), BstA abortive infection systems (5
269   genes), HicB-like toxin-antitoxin systems (17 genes), and a RelE-like toxin (1 gene), among other
270   undescribed toxins (33 genes) (Table S7). *P. larvae* had the second highest number of genes (124
271   genes, 2.9% frequency), in the PHROGs category: moron, auxiliary metabolic genes, and host
272   takeover (Fig. 3). *P. larvae* prophages possessed a variety of phage toxins, antitoxins, and
273   additionally, genes associated with bacteriocin production or immunity (Table S7). Some *P. larvae*
274   prophages also possessed a gene associated with quaternary ammonium compound-resistance
275   proteins (2 genes), or to ABC transporters (7 genes), indicating the presence of antimicrobial
276   resistance genes (Table S7). One other antimicrobial-associated gene in a *P. larvae* prophage,
277   originally classified as "unknown" by PHROGS, was annotated as a bacitracin resistance protein,
278   BacA, by the eggNOG database (Table S7). Additionally, *P. larvae* prophages possessed genes
279   for flavodoxin, a NifU-like Fe-S cluster assembly protein, and a phosphoadenosine phosphosulfate
280   (PAPs) reductase (Table S7). The PAPs reductase also occurred frequently in prophages isolates
281   from *G. apicola (*13 genes), and once in a *B. asteroides* prophage (Table S7).

282   The majority of remaining predicted protein-encoding genes (49.0%) were classified broadly as
283   Function Unknown (COG Category S or no hits to the COG database) after failing to match to the
284   PHROGs database (Table S5). Coding regions that were able to be classified broadly by COG
285   were distributed into the categories of Information Storage and Processing (1.6%), Metabolism
286   (0.5%), or Cellular Processing and Signaling (0.5%) (Table S4). Within these broader categories,
287   COG category K (Transcription — Information Storage and Processing) occurred most frequently
288   at 0.8% (Table S5). Amino Acid Transport and Metabolism (COG category E) and Carbohydrate
289   Transport and Metabolism (COG category G) were the most commonly occurring metabolic
290   categories, each corresponding to 0.1% of genes (Table S5, Table S7). Genes associated with
291   amino acid metabolism were found in prophages of *G. apicola* (8 genes, 0.4% frequency) and *P.*
292   *larvae* (5 genes, 0.1% frequency) (Fig. 3, Table S6). Five different prophages of *P. larvae* appear
293   to possess a gene homologous for a thermolysin metallopeptidase, while two share a gene
294   associated with glycine amidinotransferase activity (Table S7). The amino acid genes associated
295   with *G. apicola* prophages are more varied but appear to be associated with arginosuccinate and
296   glutamate (Table S7). Prophage genes associated with carbohydrate metabolism were found in
297   several bacterial hosts: *M. plutonius (*3 genes, 0.8% frequency), *G. apis* (4 genes, 0.7%), *L. kunkeei*
298   *(*1 gene, 0.41% frequency), *Bartonella apis (*1 gene, 0.3% frequency), *G. apicola (*4 genes, 0.2%
299   frequency), and *P. larvae* (2 genes, <0.1% frequency) (Fig. 3, Table S6). Interestingly, a single

300 *M. plutonius* prophage possessed all 3 carbohydrate-associated proteins identified for that bacterial
301 host: a mannitol dehydrogenase, a MFS/sugar transport protein, and a protein associated with the
302 dehydration of D-mannotate (Table S7). Similarly, a single *G. apicola* prophage possessed all
303 four of the carbohydrate-associated proteins identified for that bacterial host species: three proteins
304 associated with the glycerate kinase family and one associated with alpha-ribazole phosphatase
305 activity (Table S7). In contrast, three separate prophages of *G. apis* possessed a gene for alpha-
306 galactosidase, with one also carrying a gene for an exopolysaccharide biosynthesis protein (Table
307 S7). Additionally, two separate prophage species of *P. larvae* both possessed a gene associated
308 with the D-gluconate metabolic process (Table S7). The remaining COG metabolic categories
309 appear to be more rare, although they did still occur (Fig. 3, Table S6). For example, the sole *B.*
310 *indicum* prophage that was identified possessed genes associated with nucleotide metabolism (a
311 purine-cytosine permease) and secondary metabolite metabolism (isochorismatase) (Table S7).

312 **Closely Related Prophages Typically Share Host Species.** Using vConTACT2, a total of 208
313 honey bee prophages from the dereplicated prophages (n = 303) were grouped into 56 subclusters,
314 with the remaining prophages detected as outliers (n=82) or unable to be incorporated into the
315 network (n = 6) (Fig. 4). The bacterial host strongly predicted clustering, with prophages of the
316 same host species typically forming exclusive subfamily clusters (Fig. 4A). Only 10 of the 62
317 predicted subclusters included a prophage from more than one host, via either multiple honey bee
318 bacterial species or the inclusion of non-honey bee reference phage (Fig. 4A). In several of these
319 mixed clusters, the bacterial host genus was shared among isolates even if the species was not
320 (Table S8).

321 In the resulting gene sharing network, honey bee prophages within the same clusters, and therefore
322 often from the same bacterial host genome, were most closely positioned to each other (Figure
323 4B). However, prophages from different host species sometimes appeared to share genetic
324 similarities. For example, prophages from *G. apicola* and *G. apis* were closely positioned, as were
325 some of the prophages from *S. alvi* and *Bombella apis,* and *Bartonella apis* (Fig. 4B). Interestingly,
326 most of the prophages from *Latobacillus* spp.*,* appear to share genes with each other, but prophages
327 of *L. kunkeei* appear very diverse potentially due to few genes shared among these prophages *(*Fig.
328 4B). Furthermore, *L. mellis* prophages appear to share some similarity to prophages isolated from
329 *M. plutonius* and *P. larvae* (Fig. 4B).

330 **Most Prophages from Honey Bee Symbionts are Unclassified *Caudoviricetes*.** Of the 303
331 distinct prophage populations, 59.4% (180 prophage regions) were only able to be classified as
332 *Caudoviricetes,* while an additional 31.7% prophages (96 prophage) were unable to be
333 taxonomically classified using top reference hits (Fig. 5). Only the remaining 27 prophage regions
334 (0.9%) were able to be classified at a family level or lower (Fig. 5). A total of 12 (16%) *G. apicola*
335 prophages (n = 75) belonged to the family *Peduoviridae (Class: Caudoviricetes),* while one of
336 three *F. perrara* prophages was predicted to belong to the *Mesyanzhinovviridae* family (Class:
337 *Caudoviricetes)* (Fig. 5). The most common taxonomic ranks assigned to prophage from *P. larvae*
338 isolates (n = 83*)* outside of *Caudoviricetes* (69 prophage, 66.3%) or unclassified (21 prophage,
339 20.2%) were the genera *Fernvirus (*7 prophage, 6.7%)*, Vegasvirus* (3 prophage, 2.9%) and
340 *Halyconevirus* (2 prophage, 1.9%) (Class: *Caudoviricetes)* (Fig 5)*.* A single *P. larvae* prophage
341 each (1%) was classified as either the genera *Lilyvirus* or *Dragolirvirus* (Class: *Caudovircetes)*.

342 **Discussion**

343   By analyzing publicly available genomes, 303 distinct putative prophage regions were predicted
344   from honey bee-associated bacteria. Although these predicted prophage regions have not been
345   empirically confirmed to be intact and active, this is the first targeted survey of dsDNA prophage
346   distribution among core honey bee bacterial symbionts. While some of these predicted prophage
347   regions may be genomic relics of prophages, the minimum threshold of 5 kb ensures that all
348   identified prophages, active or relic, could possess genes that may affect their bacterial hosts.
349   Additionally, it should be noted that this study targeted specifically dsDNA phage. The prevalence
350   of prophages with single stranded DNA genomes, like those found in the families *Microviridae* or
351   *Inoviridae,* have yet to be determined (Székely & Breitbart, 2016).

352   The wide presence of distinct prophage regions found in this study supports the broader trend of
353   high viral diversity in the honey bee gut. However, some of the most common hosts of prophage
354   identified in this study differ from the predicted hosts of viral particles revealed in metagenomic
355   studies of the gut virome (Bonilla-Rosso et al., 2020; Deboutte et al., 2020; Busby et al., 2022).
356   This is possibly because viral metagenomes filter out bacterial cells to enrich for free viral
357   particles. The differences in the host ranges identified with metagenomic approaches and our study
358   of individual isolate genomes may be partially due to the phenomenon of superinfection exclusion,
359   in which bacteria that contain lysogens are protected from additional infection of phages (Bondy-
360   Denomy et al., 2016). For instance, we found that *S. alvi* contains a median of three prophage
361   regions, and about half of all *S. alvi* isolates possessed at least one intact prophage. However, *S.*
362   *alvi* phages have been relatively rare in metagenomic virome analyses of the honey bee gut
363   (Bonilla-Rosso et al., 2020; Busby et al., 2022). It is possible that bacterial hosts like *S. alvi* may
364   be underrepresented in metagenomic studies due to the challenge of predicting hosts of viral
365   contigs using CRISPR spacers (Dion et al., 2021). However, *S. alvi* prophages may also limit the
366   population of lytic phages targeting *S. alvi,* or intact prophages in *S. alvi* may excise infrequently;
367   these factors could make detection of *S. alvi* prophage in metagenomic studies less likely.
368   Furthermore, in one virome study, 25% of the phage identified in the honey bee gut virome were
369   predicted to infect *Bifidobacterium* spp. (Bonilla-Rosso et al., 2020), but few *B. asteroides* and
370   other *Bifidobacterium* spp. genomes analyzed in our study had estimated intact prophage; only
371   three out of 12 *B. asterioides* isolates possessed an intact prophage, while the single *B. indicum*
372   isolate and the two *B. coryneforme* isolates did not possess any. This may suggest that some honey
373   bee symbionts, like *Bifidobacterium* spp.*,* are targeted predominantly by lytic phages, or
374   alternatively, highly active temperate phages that frequently excise from their host.

375   In contrast, G. apicola has a high number of prophage regions per genome, which is consistent
376   with other virome reports that assessed the abundance of phage (both lytic and temperate) that
377   existed outside of the bacterial cell at the time of sampling (Bonilla-Rosso et al., 2020; Busby et
378   al., 2022). In this study, *G. apicola* genomes had a higher frequency of intact prophage than many
379   other core bacteria. It is possible that unlike *S. alvi,* these prophages may frequently release
380   progeny phage into the host gut, contributing to the overall composition of the virome.
381   Alternatively, the prophage of *G. apicola* may fail to provide significant superinfection exclusion
382   and protection from lytic phage. A similar scenario may be true for *L. melliventris,* which had a
383   predicted intact prophage in every isolate and was also predominant in one of the same virome
384   studies (Busby et al., 2022). Investigating the lifestyle of phage identified in metagenomic studies
385   may further elucidate if certain bacterial hosts are primarily targeted by a specific lifestyle of
386   phage. In at least one study, the majority of *Bifodibacterium* spp. were targeted by phage predicted
387   to be lytic, while *Gilliamella* spp. phage were primarily predicted to be temperate (Bonilla-Rosso
388   et al., 2020). Exploration of additional bacterial metagenomic studies would also further clarify

389   the relationship between certain phylotypes of bacteria and prophage, as our study is limited by
390   having few sequenced isolates of several of the bacterial species. For example, conclusions about
391   *L. melliventris* should be interpreted cautiously, as only three genomes were publicly available at
392   the time of our analysis. Furthermore, additional genomes from species less represented in this
393   study, such as *F. perrara*, many of the other *Latcobacillus* spp., and *Bifidobacterium* spp., would
394   clarify which bacteria harbor prophages more frequently and to what degree these prophages
395   remain excisable. It should also be noted that the bacterial species with more sequenced isolates,
396   such as *Gillamellia sp.,* and *S. alvi,* typically had a higher median abundance of prophages,
397   possibly indicating a bias in this analysis.

398   The prevalence of prophages across bacterial species is also associated with bacterial traits, such
399   as pathogenicity (Knowles et al., 2016; Silveira & Rohwer, 2016; Touchon, Bernheim & Rocha,
400   2016). The two bacterial pathogens in our study, *P. larvae* (American Foulbrood) and *M. plutonius*
401   (European Foulbrood)*,* noticeably differ from each other in terms of both prophage number and
402   prophage composition. *P. larvae* has a higher number of prophages and prophage composition
403   compared to both *M. plutonius* and all core bacterial species. In contrast, the number of prophages
404   and prophage composition of *M. plutonius* is closer to the core species *S. alvi* and *L. melliventris*.
405   While the prophage-encoded toxin *P1x1* found in some strains of *P. larvae* is known to contribute
406   to virulence, it is not the sole cause of virulence; genotypes lacking *P1x1* can still cause severe
407   disease (Ebeling, Fünfhaus & Genersch, 2021). However, the high abundance and diversity of
408   putative prophage regions in nearly all *P. larvae* isolates may imply that there are more prophage-
409   encoded virulence factors to be discovered. For example, the presence of *P. larvae* prophage
410   regions containing possible antimicrobial resistance genes, as well as carbohydrate, iron and sulfur
411   metabolism genes, may indicate additional ways phage provide fitness to the pathogen (discussed
412   in Ribeiro et al., 2022). The presence of three separate carbohydrate-associated genes in a single
413   *M. plutonius* prophage indicates that prophage-encoded auxiliary metabolic pathways provide
414   additional benefit to *M. plutonius,* as well. This is particularly interesting because *M. plutonius*
415   strains are not typically able to metabolize the sugar mannitol, but can metabolize D-mannose
416   (Arai et al., 2012). The phage-encoded mannitol dehydrogenase may break down mannitol into D-
417   mannose, while a second phage-encoded protein may then dehydrate the D-mannose to release
418   energy. The third phage-encoded protein, a sugar transporter, could have a role in shuttling these
419   sugars in or out of the bacterial cell. Combined, these three prophage-encoded genes may serve to
420   enhance the pathogenicity of *M. plutonius* 82 by providing an additional sugar to exploit.
421   However, the lower prophage abundance and composition in *M. plutonius* may indicate that
422   prophages play a less significant role in the virulence of *M. plutonius* compared to *P. larvae*.

423   Functional annotation of the predicted prophages also suggests that some commensal bacterial,
424   such as *Gilliamellia* spp.*,* may receive auxiliary metabolic benefits from their prophages.
425   *Gilliamella* spp. contribute to the breakdown of pollen by primarily targeting pectin (Engel,
426   Martinson & Moran, 2012). It is possible that the glycerate kinases and the alpha-ribazole
427   phosphatase found in *G. apicola* prophage, or the alpha-galactosidase found in three separate *G.*
428   *apis* prophages assist with the breakdown of pectin or the production of other metabolites. Some
429   isoforms of alpha-galactosidase have been shown to be associated with pectin hydrolysis in fruits
430   (Soh, Ali & Lazan, 2006). The predicted glycerate kinase may be used to support bacterial
431   glycolysis pathways, while the alpha-ribazole phosphatase enzyme may contribute to cobalamin
432   biosynthesis, an important secondary metabolite (Doughty, Hayashi & Guenther, 1966; O'Toole,
433   Trzebiatowski & Escalante-Semerena, 1994). Additionally, an exopolysaccharide biosynthesis
434   coding region found in a *G. apis* prophage may contribute to biofilm formation. Biofilms produced

435   by *Gilliamellia* spp. may inhibit pathogen invasion in the midgut (Engel, Martinson & Moran,
436   2012). As a result, it is possible that prophage presence in *Gilliamella* spp. may directly benefit its
437   bacterial host, and consequently the honey bee. Additionally, several *G. apicola* prophages appear
438   to carry a PAPS reductase gene, which may enhance the host's ability to metabolize sulfur (Mara
439   et al., 2020).

440   Given that phage typically have a high host-specificity, along with the ability to recombine with
441   host DNA and other co-infecting phages, it is not surprising that the prophage regions found in the
442   same host species are most closely related to each other (Casjens, 2003; Brüssow, Canchaya &
443   Hardt, 2004). At the same time, phages from entirely different bacterial host species may
444   occasionally transfer genes horizontally to each other (Stecher, Maier & Hardt, 2013), and this
445   may be more likely to occur between phages with hosts that are physically near each other.
446   Prophages identified from bacterial hosts in the midgut and ileum (*Bombella apis, Bartonella apis,*
447   *Gilliamellia* spp*., F. perrara,* and *S. alvi)* were positioned more closely in the network, and
448   therefore more likely to share genes with each other, than prophages of bacteria in the rectum
449   (*Bifidobacterium* spp*., Lactobacillus* spp*.),* possibly because the phages are either more closely
450   related to one another or due to increased horizontal gene transfer between nearby phage
451   communities (Anderson & Ricigliano, 2017). The idea of spatially-separated phage communities
452   in the gut is supported by a recent study investigating the gut virome of the marine invertebrate
453   *Ciona intestinalis,* which found phage communities were localized to specific regions (Leigh et
454   al., 2018). Due to the inherent difficulty in classifying phages, and the limited number of
455   International Committee on Taxonomy of Viruses (ICTV) classified reference phage genomes
456   relative to the vast diversity of bacteriophages in nature, it is unsurprising that a large portion of
457   predicted prophage regions in our study remain unclassified (Bin Jang et al., 2019). Less than ten
458   percent of the predicted prophages were classifiable at a family level or lower. A few prophage
459   predicted from *P. larvae* shared taxonomic similarity to reference *P. larvae* phage isolated from
460   other studies, but the majority were only classifiable as belonging to *Caudoviricetes* order*,*
461   indicating that a greater diversity of *P. larvae* phage may exist (Stamereilers et al., 2018).

462   Although honey bee gut bacterial communities contain relatively few bacterial phylotypes,
463   prophages may contribute to the high bacterial strain diversity within the honey bee gut
464   microbiome (Ellegaard et al., 2015; Ellegaard & Engel, 2019), as 237 out of the 303 unique
465   prophage regions were detected in only one bacterial isolate. This may indicate that in some
466   bacterial species, susceptibility to specific prophage integration may vary depending on host strain
467   genotype, possibly driving both viral and bacterial evolution or diversification. By potentially
468   impacting their host's evolution, active honey bee gut prophages may affect general honey bee
469   health and dysbiosis, particularly within the context of antibiotic and pesticide exposure. In other
470   host-associated and environmental systems, antibiotics and other chemicals can induce prophage
471   excision (Danovaro et al., 2003; Allen et al., 2011). Frequent exposure to either antibiotics or
472   pesticides, something that honey bees must contend with (Kakumanu et al., 2016; Raymann,
473   Shaffer & Moran, 2017), may result in a gut state that is constantly perturbed, leading to prophage
474   induction and further gut perturbation. On the other hand, prophages in the honey bee gut may
475   stabilize bacterial communities against antimicrobials and pesticides, as prophages can also
476   enhance the fitness of their bacterial partners via beneficial accessory genes that provide resistance
477   to antibiotics or other environmental pollutants (Wang et al., 2010; Huang et al., 2021).

478   The diverse prophage community identified in our study supports the well-established idea that
479   prophages are common and present in a wide range of bacterial hosts (Touchon, Bernheim &

480 Rocha, 2016). However, some of the most common bacterial hosts of the prophage community
481 identified in this study do not entirely reflect that of the wider virome studies of honey bee guts,
482 indicating that prophages may remain cryptic despite more targeted viral sequencing approaches.
483 As a result, the role prophages play in bacterial evolution and community dynamics will require
484 further untangling both within the honey bee system and beyond.

485

Conclusion

487 As the first step in identifying how the prophage community interacts with and modulates bacterial
488 communities of the honey bee gut, a survey of prophage in publicly available honey bee gut-
489 associated bacterial genomes was conducted. This study found that prophage abundance and
490 composition vary across the core honey bee gut bacterial species, with *S. alvi* and *G. apicola*
491 possessing the highest prophage number and percent phage composition, and *L. melliventris*
492 having the highest frequency of likely intact prophage. Interestingly, there was some discrepancy
493 between the most commonly predicted bacterial hosts for prophages compared to phage isolated
494 from metagenomic studies. This may indicate that certain nuances of prophage-bacteria
495 interactions are not able to be captured by viromes, which primarily focus on free viral particles.
496 Additionally, the prophages of some commensal bacterial hosts, such as *Gilliamella* spp.*,* may
497 provide auxiliary metabolic genes for carbohydrate metabolism, which could possibly benefit the
498 honey bee host. These results set the foundation for targeted exploration of the prophage infecting
499 core bacterial species through culture-based methods.
500
501

Figure Legends

503 Table 1: Honey bee-associated bacterial isolates

504 The number of isolates (n) in each phylotype and species of the 181 publicly available genomes
505 of unique bacterial isolates from the NCBI GenBank that were downloaded March-July 2020.

506

507 Figure 1: Prophage frequency and composition across bacterial hosts

508 (A) The predicted number of total prophage regions, (B) total prophage composition (the percent
509 of the bacterial genome composed of prophages), (C) number of intact prophage regions, and (D)
510 intact prophage composition per bacterial isolate. Dots represent individual bacterial isolates,
511 while the center line indicates the median value for each bacterial species. The boxes span the
512 interquartile range and whiskers represent the 25th and 75th quartile ranges. Values below or above
513 the 25th and 75th quartiles are indicated by the larger colored circles behind individual dots. The
514 minimum, maximum, and median values are listed in Table S2.

515

516 Figure 2: Bacterial genome size correlates to prophage number and composition

517 (A) The number of prophages identified and the (B) percent prophage composition is positively
518 associated with bacterial genome size. Dots represent individual bacterial genomes.

519

520   Figure 3: Functional analysis of prophages found in honey bee-associated bacteria

521   The distribution of phage-associated and COG categories assigned to prophage genes, organized
522   by bacterial host. Similar COG categories are grouped together for visual clarity. X-axis indicates
523   percent. COG Category S (Unknown or No Hits) is not shown.

524

525   Figure 4: Viral clustering and network analysis based on shared genes

526   Nodes represent individual prophages. Bacterial hosts are indicated by the node's color and shape.
527   (A) Prophages are separated based on sub-family viral clusters. Numbers above clusters indicate
528   a specific subcluster. Distance between subclusters has no significance, but viral subclusters that
529   come from the same greater viral cluster (indicated by transparent oval overlay) are more related
530   to one another than phage originating from other clusters. Subclusters O6, 1, 7, 8, and 12 contain
531   reference *P. larvae* phage from external studies. Prophage within the outlier cluster are not related
532   to one another. (B) The layout of the network is based on shared genes, with strength indicated by
533   opacity of the network edges. The numbers on nodes indicate subcluster. Blank nodes indicate
534   outliers, while subclusters starting with "O" indicate unique overlapping subclusters.

535

536   Figure 5: Taxonomic classification of prophages

537   The lowest taxonomic rank (genus, family, or class) assigned to each prophage based on amino
538   acid similarity to its closest reference phage. Genus and family classifications were assigned if
539   prophages shared AAI across all proteins with a reference above respective thresholds. Prophage
540   which did not meet these thresholds but still matched to a *Caudoviricetes* reference were broadly
541   classified as *Caudoviricetes* class. Prophage which matched to unclassified reference phage were
542   grouped as Unclassified.

543   **Supplementary Figure Legends:**

544   Supplementary Figure 1: Prophage networks in the context of INPHARED

545   A vConTACT2 network created with both honey bee isolated prophages and reference phage from
546   the INPHARED database (January 2023). Visually, only reference phage that are third neighbors
547   or less to a honey bee prophage are represented. Nodes indicate individual phages and the opacity
548   of edges represent the strength of shared gene interactions. (A) Honey bee isolated prophages were
549   unable to be taxonomically classified by vConTACT2, but often clustered near other
550   taxonomically classified reference phage. Larger black circles are prophages identified from honey
551   bee symbionts, while smaller colored circles depict the taxonomic family of reference phages. (B)
552   Honey bee isolated prophages (host indicated by shape and color) tended to cluster near other
553   phage from the same bacterial host, but were not spatially separated from reference phages.

554

555   Supplementary Table 1 (.csv): NCBI Genbank accession numbers

556   A list of all NCBI bacterial isolates used and their accession numbers.

557

558   Supplementary Table 2 (.xlsx file): Summary statistics of prophages across bacterial hosts

559 The minimum, median ± SD, and maximum number of total and estimated intact prophage regions
560 per genome, as well as total and estimated intact prophage composition per genome. Additionally,
561 the minimum, median and maximum lengths of the bacterial genomes are included.

562

563 Supplementary Table 3 (.csv): Percentage of bacterial isolates with intact prophage

564 The percent of individual isolates within a bacterial species that possessed at least one estimated
565 intact prophage.

566

567 Supplementary Table 4 (.csv file): Frequency of broad COG categories across all prophages.

568

569 Supplementary Table 5 (.csv file): Frequency of specific COG categories across all prophages.

570

571 Supplementary Table 6 (.csv file): Frequency of specific COG categories across prophages,
572 grouped by bacterial species.

573

574 Supplementary Table 7 (.csv file): All predicted protein annotations, with a final column
575 "AMG_Moron", indicating whether a protein is predicted to be a potential AMG, a possible
576 moron, or neither.

577

578 Supplementary Table 8: vConTACT2 clusters

579 The vConTACT2 cluster assigned to each phage.

580

581 **Statements & Declarations**

582 Funding

583 This research was supported by the National Science Foundation (MCB-1817736).

584

585 Competing Interests

586 The authors declare that there is no conflict of interest or competing interests to disclose.

587

588 Data Availability

589 Sequences are available at Virginia Tech's Data Repository: https://doi.org/10.7294/22203001

590 Code and raw data are available at Github: https://github.com/ebueren/HBProphage_PeerJ

591

592 **References**

593

594   Abraham J, Bousquet A-C, Bruff E, Carson N, Clark A, Connell A, Davis Z, Dums J, Everington

595   C, Groth A, Hawes N, McArthur N, McKenney C, Oufkir A, Pearce B, Rampal S, Rozier

596   H, Schaff J, Slehria T, Carson S, Miller ES. 2016. *Paenibacillus larvae* Phage Tripp

597   Genome Has 378-Base-Pair Terminal Repeats. *Genome Announcements* 4:e01498-15,

598   /ga/4/1/e01498-15.atom. DOI: 10.1128/genomeA.01498-15.

599   Allen HK, Looft T, Bayles DO, Humphrey S, Levine UY, Alt D, Stanton TB. 2011. Antibiotics

600   in Feed Induce Prophages in Swine Fecal Microbiomes. *Mbio* 2:e00260-11. DOI:

601   10.1128/mBio.00260-11.

602   Anderson KE, Ricigliano VA. 2017. Honey bee gut dysbiosis: a novel context of disease

603   ecology. *Current Opinion in Insect Science* 22:125–132. DOI:

604   10.1016/j.cois.2017.05.020.

605   Arai R, Tominaga K, Wu M, Okura M, Ito K, Okamura N, Onishi H, Osaki M, Sugimura Y,

606   Yoshiyama M, Takamatsu D. 2012. Diversity of Melissococcus plutonius from Honeybee

607   Larvae in Japan and Experimental Reproduction of European Foulbrood with Cultured

608   Atypical Isolates. *PLoS ONE* 7:e33708. DOI: 10.1371/journal.pone.0033708.

609   Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better,

610   faster version of the PHAST phage search tool. *Nucleic Acids Research* 44:W16–W21.

611   DOI: 10.1093/nar/gkw387.

612   Beims H, Wittmann J, Bunk B, Spröer C, Rohde C, Günther G, Rohde M, von der Ohe W,

613   Steinert M. 2015. Paenibacillus larvae-Directed Bacteriophage HB10c2 and Its

614   Application in American Foulbrood-Affected Honey Bee Larvae. *Applied and*

615   *Environmental Microbiology* 81:5411–5419. DOI: 10.1128/AEM.00804-15.

616 Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, Brister JR, Kropinski

617   AM, Krupovic M, Lavigne R, Turner D, Sullivan MB. 2019. Taxonomic assignment of

618   uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nature*

619   *Biotechnology* 37:632–639. DOI: 10.1038/s41587-019-0100-8.

620 Bolduc B, Jang HB, Doulcier G, You Z-Q, Roux S, Sullivan MB. 2017. vConTACT: an iVirus

621   tool to classify double-stranded DNA viruses that infect *Archaea* and *Bacteria*. *PeerJ*

622   5:e3243. DOI: 10.7717/peerj.3243.

623 Bondy-Denomy J, Qian J, Westra ER, Buckling A, Guttman DS, Davidson AR, Maxwell KL.

624   2016. Prophages mediate defense against phage infection through diverse mechanisms.

625   *The ISME Journal* 10:2854–2866. DOI: 10.1038/ismej.2016.79.

626 Bonilla-Rosso G, Steiner T, Wichmann F, Bexkens E, Engel P. 2020. Honey bees harbor a

627   diverse gut virome engaging in nested strain-level interactions with the microbiota.

628   *Proceedings of the National Academy of Sciences*. DOI: 10.1073/pnas.2000228117.

629 Brüssow H, Canchaya C, Hardt W-D. 2004. Phages and the Evolution of Bacterial Pathogens:

630   from Genomic Rearrangements to Lysogenic Conversion. *Microbiology and Molecular*

631   *Biology Reviews* 68:560–602. DOI: 10.1128/MMBR.68.3.560-602.2004.

632 Busby TJ, Miller CR, Moran NA, Van Leuven JT. 2022. Global Composition of the

633   Bacteriophage Community in Honey Bees. *mSystems* 7:e01195-21. DOI:

634   10.1128/msystems.01195-21.

635 Casjens S. 2003. Prophages and bacterial genomics: what have we learned so far?: Prophage

636   genomics. *Molecular Microbiology* 49:277–300. DOI: 10.1046/j.1365-

637   2958.2003.03580.x.

638   Cook R, Brown N, Redgwell T, Rihtman B, Barnes M, Clokie M, Stekel DJ, Hobman J, Jones

639   MA, Millard A. 2021. INfrastructure for a PHAge REference Database: Identification of

640   Large-Scale Biases in the Current Collection of Cultured Phage Genomes. *PHAGE*

641   2:214–223. DOI: 10.1089/phage.2021.0007.

642   Dahl EM, Neer E, Bowie KR, Leung ET, Karstens L. 2022. microshades: An R Package for

643   Improving Color Accessibility and Organization of Microbiome Data. *Microbiology*

644   *Resource Announcements* 11:e00795-22. DOI: 10.1128/mra.00795-22.

645   Danovaro R, Armeni M, Corinaldesi C, Mei ML. 2003. Viruses and marine pollution. *Marine*

646   *Pollution Bulletin* 46:301–304. DOI: 10.1016/S0025-326X(02)00461-7.

647   Deboutte W, Beller L, Yinda CK, Maes P, Graaf DC de, Matthijnssens J. 2020. Honey-bee–

648   associated prokaryotic viral communities reveal wide viral diversity and a profound

649   metabolic coding potential. *Proceedings of the National Academy of Sciences*. DOI:

650   10.1073/pnas.1921859117.

651   Diedenhofen B, Musch J. 2015. cocor: A Comprehensive Solution for the Statistical Comparison

652   of Correlations. *PLOS ONE* 10:e0121945. DOI: 10.1371/journal.pone.0121945.

653   Dion MB, Plante P-L, Zufferey E, Shah SA, Corbeil J, Moineau S. 2021. Streamlining CRISPR

654   spacer-based bacterial host predictions to decipher the viral dark matter. *Nucleic Acids*

655   *Research* 49:3127–3138. DOI: 10.1093/nar/gkab133.

656   Doughty CC, Hayashi JA, Guenther HL. 1966. Purification and Properties of d-Glycerate 3-

657   Kinase from Escherichia coli. *Journal of Biological Chemistry* 241:568–572. DOI:

658   10.1016/S0021-9258(18)96874-2.

659   Ebeling J, Fünfhaus A, Gensersch E. 2021. The Buzz about ADP-Ribosylation Toxins from

660   Paenibacillus larvae, the Causative Agent of American Foulbrood in Honey Bees. *Toxins*

661   13:151. DOI: 10.3390/toxins13020151.

662   Eddy SR. 2011. Accelerated Profile HMM Searches. *PLOS Computational Biology* 7:e1002195.

663   DOI: 10.1371/journal.pcbi.1002195.

664   Ellegaard KM, Engel P. 2019. Genomic diversity landscape of the honey bee gut microbiota.

665   *Nature Communications* 10:1–13. DOI: 10.1038/s41467-019-08303-0.

666   Ellegaard KM, Tamarit D, Javelind E, Olofsson TC, Andersson SG, Vásquez A. 2015. Extensive

667   intra-phylotype diversity in lactobacilli and bifidobacteria from the honeybee gut. *BMC*

668   *Genomics* 16. DOI: 10.1186/s12864-015-1476-6.

669   Emery O, Schmidt K, Engel P. 2017. Immune system stimulation by the gut symbiont *Frischella*

670   *perrara* in the honey bee (*Apis mellifera*). *Molecular Ecology* 26:2576–2590. DOI:

671   10.1111/mec.14058.

672   Engel P, Martinson VG, Moran NA. 2012. Functional diversity within the simple gut microbiota

673   of the honey bee. *Proceedings of the National Academy of Sciences* 109:11002–11007.

674   DOI: 10.1073/pnas.1202970109.

675   Forcone K, Coutinho FH, Cavalcanti GS, Silveira CB. 2021. Prophage Genomics and Ecology in

676   the Family Rhodobacteraceae. *Microorganisms* 9:1115. DOI:

677   10.3390/microorganisms9061115.

678   Fortier L-C, Sekulovic O. 2013. Importance of prophages to evolution and virulence of bacterial

679   pathogens. *Virulence* 4:354–365. DOI: 10.4161/viru.24498.

680   Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, Pratama AA,

681   Gazitúa MC, Vik D, Sullivan MB, Roux S. 2021. VirSorter2: a multi-classifier, expert-

682   guided approach to detect diverse DNA and RNA viruses. *Microbiome* 9:37. DOI:

683   10.1186/s40168-020-00990-y.

684 Hannigan GD, Meisel JS, Tyldsley AS, Zheng Q, Hodkinson BP, SanMiguel AJ, Minot S,

685   Bushman FD, Grice EA. 2015. The Human Skin Double-Stranded DNA Virome:

686   Topographical and Temporal Diversity, Genetic Enrichment, and Dynamic Associations

687   with the Host Microbiome. *mBio* 6:e01578-15. DOI: 10.1128/mBio.01578-15.

688 Hendrix RW. 2002. Bacteriophages: Evolution of the Majority. *Theoretical Population Biology*

689   61:471–480. DOI: 10.1006/tpbi.2002.1590.

690 Hsu BB, Gibson TE, Yeliseyev V, Liu Q, Lyon L, Bry L, Silver PA, Gerber GK. 2019. Dynamic

691   Modulation of the Gut Microbiota and Metabolome by Bacteriophages in a Mouse

692   Model. *Cell Host & Microbe* 25:803-814.e5. DOI: 10.1016/j.chom.2019.05.001.

693 Huang D, Yu P, Ye M, Schwarz C, Jiang X, Alvarez PJJ. 2021. Enhanced mutualistic symbiosis

694   between soil phages and bacteria with elevated chromium-induced environmental stress.

695   *Microbiome* 9:150. DOI: 10.1186/s40168-021-01074-1.

696 Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR,

697   Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P. 2019. eggNOG 5.0: a hierarchical,

698   functionally and phylogenetically annotated orthology resource based on 5090 organisms

699   and 2502 viruses. *Nucleic Acids Research* 47:D309–D314. DOI: 10.1093/nar/gky1085.

700 Kakumanu ML, Reeves AM, Anderson TD, Rodrigues RR, Williams MA. 2016. Honey Bee Gut

701   Microbiome Is Altered by In-Hive Pesticide Exposures. *Frontiers in Microbiology* 7.

702   DOI: 10.3389/fmicb.2016.01255.

703   Kieft K, Zhou Z, Anantharaman K. 2020. VIBRANT: automated recovery, annotation and

704   curation of microbial viruses, and evaluation of viral community function from genomic

705   sequences. *Microbiome* 8:90. DOI: 10.1186/s40168-020-00867-0.

706   Kiełbasa SM, Wan R, Sato K, Horton P, Frith MC. 2011. Adaptive seeds tame genomic

707   sequence comparison. *Genome Research* 21:487–493. DOI: 10.1101/gr.113985.110.

708   Kim M-S, Bae J-W. 2018. Lysogeny is prevalent and widely distributed in the murine gut

709   microbiota. *The ISME Journal* 12:1127–1141. DOI: 10.1038/s41396-018-0061-9.

710   Knowles B, Silveira CB, Bailey BA, Barott K, Cantu VA, Cobián-Güemes AG, Coutinho FH,

711   Dinsdale EA, Felts B, Furby KA, George EE, Green KT, Gregoracci GB, Haas AF,

712   Haggerty JM, Hester ER, Hisakawa N, Kelly LW, Lim YW, Little M, Luque A, McDole-

713   Somera T, McNair K, de Oliveira LS, Quistad SD, Robinett NL, Sala E, Salamon P,

714   Sanchez SE, Sandin S, Silva GGZ, Smith J, Sullivan C, Thompson C, Vermeij MJA,

715   Youle M, Young C, Zgliczynski B, Brainard R, Edwards RA, Nulton J, Thompson F,

716   Rohwer F. 2016. Lytic to temperate switching of viral communities. *Nature* 531:466–

717   470. DOI: 10.1038/nature17193.

718   Kwong WK, Moran NA. 2016. Gut microbial communities of social bees. *Nature Reviews*

719   *Microbiology* 14:374–384. DOI: 10.1038/nrmicro.2016.43.

720   Lawlor J. 2020. jakelawlor/PNWColors: Initial Release. DOI: 10.5281/zenodo.3971033.

721   Leigh BA, Djurhuus A, Breitbart M, Dishaw LJ. 2018. The gut virome of the protochordate

722   model organism, Ciona intestinalis subtype A. *Virus Research* 244:137–146. DOI:

723   10.1016/j.virusres.2017.11.015.

724   Mara P, Vik D, Pachiadaki MG, Suter EA, Poulos B, Taylor GT, Sullivan MB, Edgcomb VP.

725   2020. Viral elements and their potential influence on microbial processes along the

726   permanently stratified Cariaco Basin redoxcline. *The ISME Journal* 14:3079–3092. DOI:

727   10.1038/s41396-020-00739-3.

728 Meng X, Rosenthal R, Rubin DB. 1992. Comparing correlated correlation coefficients.

729   *Psychological Bulletin* 111:172–175. DOI: 10.1037/0033-2909.111.1.172.

730 Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, Lewis JD, Bushman FD. 2011. The

731   human gut virome: Inter-individual variation and dynamic response to diet. *Genome*

732   *Research* 21:1616–1625. DOI: 10.1101/gr.122705.111.

733 Myers L, Sirois MJ. SPEARMAN RANK CORRELATION COEFFICIENT. :2.

734 Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. 2021. CheckV

735   assesses the quality and completeness of metagenome-assembled viral genomes. *Nature*

736   *Biotechnology* 39:578–585. DOI: 10.1038/s41587-020-00774-7.

737 Olm MR, Brown CT, Brooks B, Banfield JF. 2017. dRep: a tool for fast and accurate genomic

738   comparisons that enables improved genome recovery from metagenomes through de-

739   replication. *The ISME Journal* 11:2864–2868. DOI: 10.1038/ismej.2017.126.

740 O'Toole GA, Trzebiatowski JR, Escalante-Semerena JC. 1994. The cobC gene of Salmonella

741   typhimurium codes for a novel phosphatase involved in the assembly of the nucleotide

742   loop of cobalamin. *Journal of Biological Chemistry* 269:26503–26511. DOI:

743   10.1016/S0021-9258(18)47223-7.

744 Raymann K, Shaffer Z, Moran NA. 2017. Antibiotic exposure perturbs the gut microbiota and

745   elevates mortality in honeybees. *PLOS Biology* 15:e2001861. DOI:

746   10.1371/journal.pbio.2001861.

747 Ribeiro HG, Melo LDR, Oliveira H, Boon M, Lavigne R, Noben J-P, Azeredo J, Oliveira A.

748   2019. Characterization of a new podovirus infecting Paenibacillus larvae. *Scientific*

749   *Reports* 9:20355. DOI: 10.1038/s41598-019-56699-y.

750 Ribeiro HG, Nilsson A, Melo LDR, Oliveira A. 2022. Analysis of intact prophages in genomes

751   of Paenibacillus larvae: An important pathogen for bees. *Frontiers in Microbiology*

752   13:903861. DOI: 10.3389/fmicb.2022.903861.

753 Roux S, Adriaenssens EM, Dutilh BE, Koonin EV, Kropinski AM, Krupovic M, Kuhn JH,

754   Lavigne R, Brister JR, Varsani A, Amid C, Aziz RK, Bordenstein SR, Bork P, Breitbart

755   M, Cochrane GR, Daly RA, Desnues C, Duhaime MB, Emerson JB, Enault F, Fuhrman

756   JA, Hingamp P, Hugenholtz P, Hurwitz BL, Ivanova NN, Labonté JM, Lee K-B,

757   Malmstrom RR, Martinez-Garcia M, Mizrachi IK, Ogata H, Páez-Espino D, Petit M-A,

758   Putonti C, Rattei T, Reyes A, Rodriguez-Valera F, Rosario K, Schriml L, Schulz F,

759   Steward GF, Sullivan MB, Sunagawa S, Suttle CA, Temperton B, Tringe SG, Thurber

760   RV, Webster NS, Whiteson KL, Wilhelm SW, Wommack KE, Woyke T, Wrighton KC,

761   Yilmaz P, Yoshida T, Young MJ, Yutin N, Allen LZ, Kyrpides NC, Eloe-Fadrosh EA.

762   2019. Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nature*

763   *Biotechnology* 37:29–37. DOI: 10.1038/nbt.4306.

764 Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, Poulos BT, Solonenko N, Lara

765   E, Poulain J, Pesant S, Kandels-Lewis S, Dimier C, Picheral M, Searson S, Cruaud C,

766   Alberti A, Duarte CM, Gasol JM, Vaqué D, Bork P, Acinas SG, Wincker P, Sullivan

767   MB. 2016. Ecogenomics and potential biogeochemical impacts of globally abundant

768   ocean viruses. *Nature* 537:689–693. DOI: 10.1038/nature19366.

769   Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069.

770   DOI: 10.1093/bioinformatics/btu153.

771   Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B,

772   Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular

773   interaction networks. *Genome Research* 13:2498–2504. DOI: 10.1101/gr.1239303.

774   Silveira CB, Rohwer FL. 2016. Piggyback-the-Winner in host-associated microbial

775   communities. *npj Biofilms and Microbiomes* 2:1–5. DOI: 10.1038/npjbiofilms.2016.10.

776   Soh C-P, Ali ZM, Lazan H. 2006. Characterisation of an α-galactosidase with potential relevance

777   to ripening related texture changes. *Phytochemistry* 67:242–254. DOI:

778   10.1016/j.phytochem.2005.09.032.

779   Stamereilers C, Fajardo CP, Walker JK, Mendez KN, Castro-Nallar E, Grose JH, Hope S,

780   Tsourkas PK. 2018. Genomic Analysis of 48 Paenibacillus larvae Bacteriophages.

781   *Viruses* 10:377. DOI: 10.3390/v10070377.

782   Stecher B, Maier L, Hardt W-D. 2013. "Blooming" in the gut: how dysbiosis might contribute to

783   pathogen evolution. *Nature Reviews Microbiology* 11:277–284. DOI:

784   10.1038/nrmicro2989.

785   Székely AJ, Breitbart M. 2016. Single-stranded DNA phages: from early molecular biology tools

786   to recent revolutions in environmental microbiology. *FEMS Microbiology Letters*

787   363:fnw027. DOI: 10.1093/femsle/fnw027.

788   Terzian P, Olo Ndela E, Galiez C, Lossouarn J, Pérez Bucio RE, Mom R, Toussaint A, Petit M-

789   A, Enault F. 2021. PHROG: families of prokaryotic virus proteins clustered using remote

790   homology. *NAR Genomics and Bioinformatics* 3:lqab067. DOI: 10.1093/nargab/lqab067.

791   Touchon M, Bernheim A, Rocha EP. 2016. Genetic and life-history traits associated with the

792   distribution of prophages in bacteria. *The ISME Journal* 10:2744–2754. DOI:

793   10.1038/ismej.2016.47.

794   Turner D, Kropinski AM, Adriaenssens EM. 2021. A Roadmap for Genome-Based Phage

795   Taxonomy. *Viruses* 13:506. DOI: 10.3390/v13030506.

796   Wang X, Kim Y, Ma Q, Hong SH, Pokusaeva K, Sturino JM, Wood TK. 2010. Cryptic

797   prophages help bacteria cope with adverse environments. *Nature Communications* 1:147.

798   DOI: 10.1038/ncomms1146.

799   Wickham H. 2016. *ggplot2*. Cham: Springer International Publishing. DOI: 10.1007/978-3-319-

800   24277-4.

801   Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: A Fast Phage Search Tool.

802   *Nucleic Acids Research* 39:W347–W352. DOI: 10.1093/nar/gkr485.

803

804

Honey bee-associated bacterial isolates

The number of isolates (n) in each phylotype and species of the 181 publicly available genomes of unique bacterial isolates from NCBI GenBank that were downloaded March-July 2020.

	Phylotype	Species	Total Isolates (n)
Core, n=150	*Acetobacter*, n=6	*Bombella apis*	6
	Bartonella apis, n=6	*Bartonella apis*	6
	Bifidobacterium spp., n=15	*Bifidobacterium asteroides*	12
		Bifidobacterium coryneforme	2
		Bifidobacterium indicum	1
	Firmicutes-4, n=6	*Lactobacillus mellifer*	1
		Lactobacillus mellis	5
	Firmicutes-5, n=11	*Lactobacillus apis*	3
		Lactobacillus helsingborgensis	2
		Lactobacillus kimbladii	1
		Lactobacillus kullabergensis	2
		Lactobacillus melliventris	3
	Lactobacillus kunkeei, n=10	*Lactobacillus kunkeei*	10
	Frischella perrara, n=2	*Frischella perrara*	2
	Gilliamella spp., n=61	*Gilliamella apicola*	38
		Gilliamella api	23
	Snodgrassella alvi, n=34	*Snodgrassella alvi*	34
Pathogen, n=30	Pathogen, n=30	*Melissococcus plutonius*	17
		Paenibacillus larvae	13

Figure 1

Prophage frequency and composition across bacterial hosts

(A) The predicted number of total prophage regions, (B) total prophage composition (the percent of the bacterial genome composed of prophages), (C) number of intact prophage regions, and (D) intact prophage composition per bacterial isolate. Dots represent individual bacterial isolates, while the center line indicates the median value for each bacterial species. The boxes span the interquartile range and whiskers represent the 25th and 75th quartile ranges. Values below or above the 25th and 75th quartiles are indicated by the larger colored circles behind individual dots. The minimum, maximum, and median values are listed in Table S2.

A. Total Number of Predicted Prophage

B. Total Predicted Prophage Composition

C. Number of Estimated Intact Prophage

D. Estimated Intact Prophage Composition
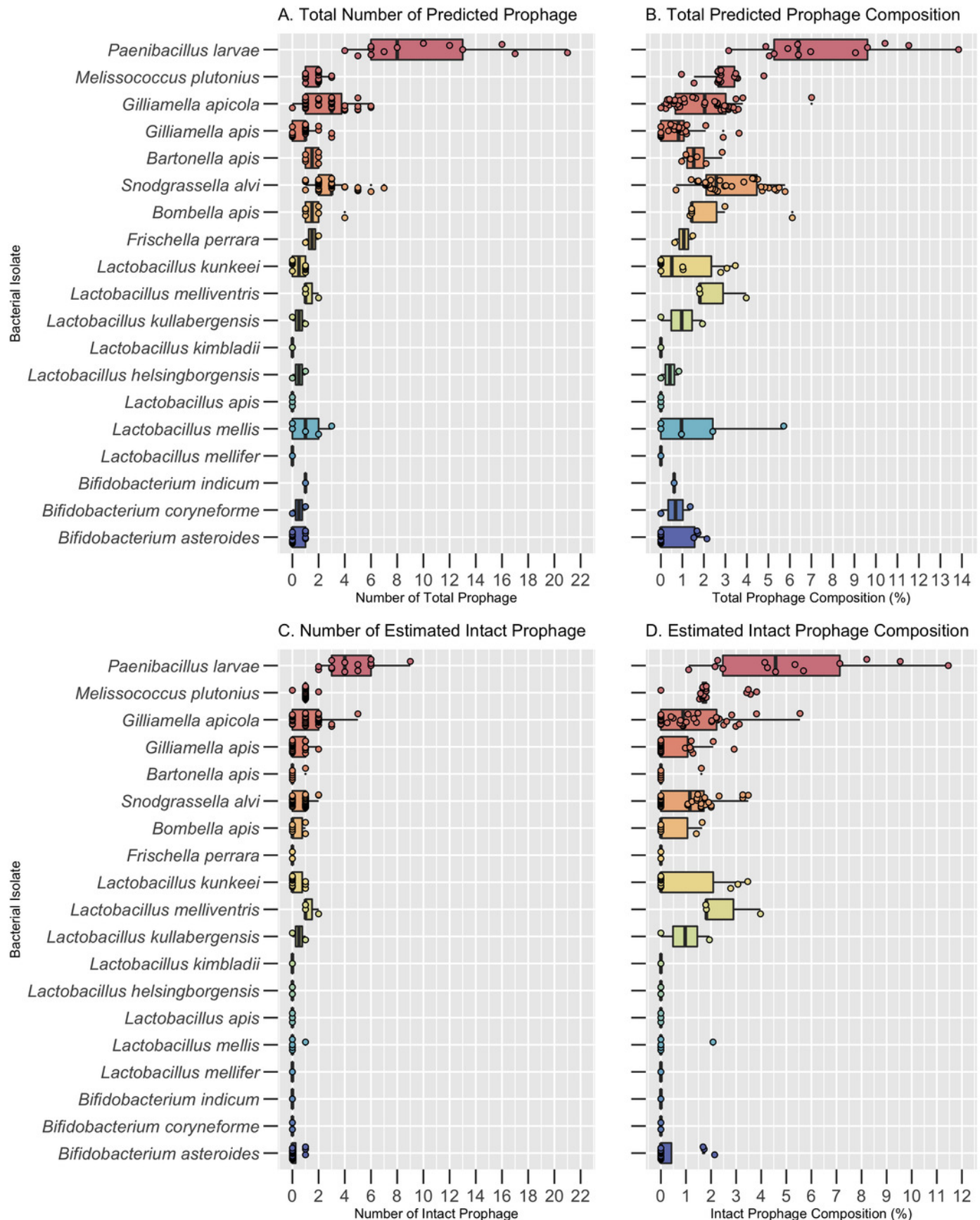
Figure 2

Bacterial genome size correlates to prophage number and composition

(A) The number of prophages identified and the (B) percent prophage composition is positively associated with bacterial genome size. Dots represent individual bacterial genomes.
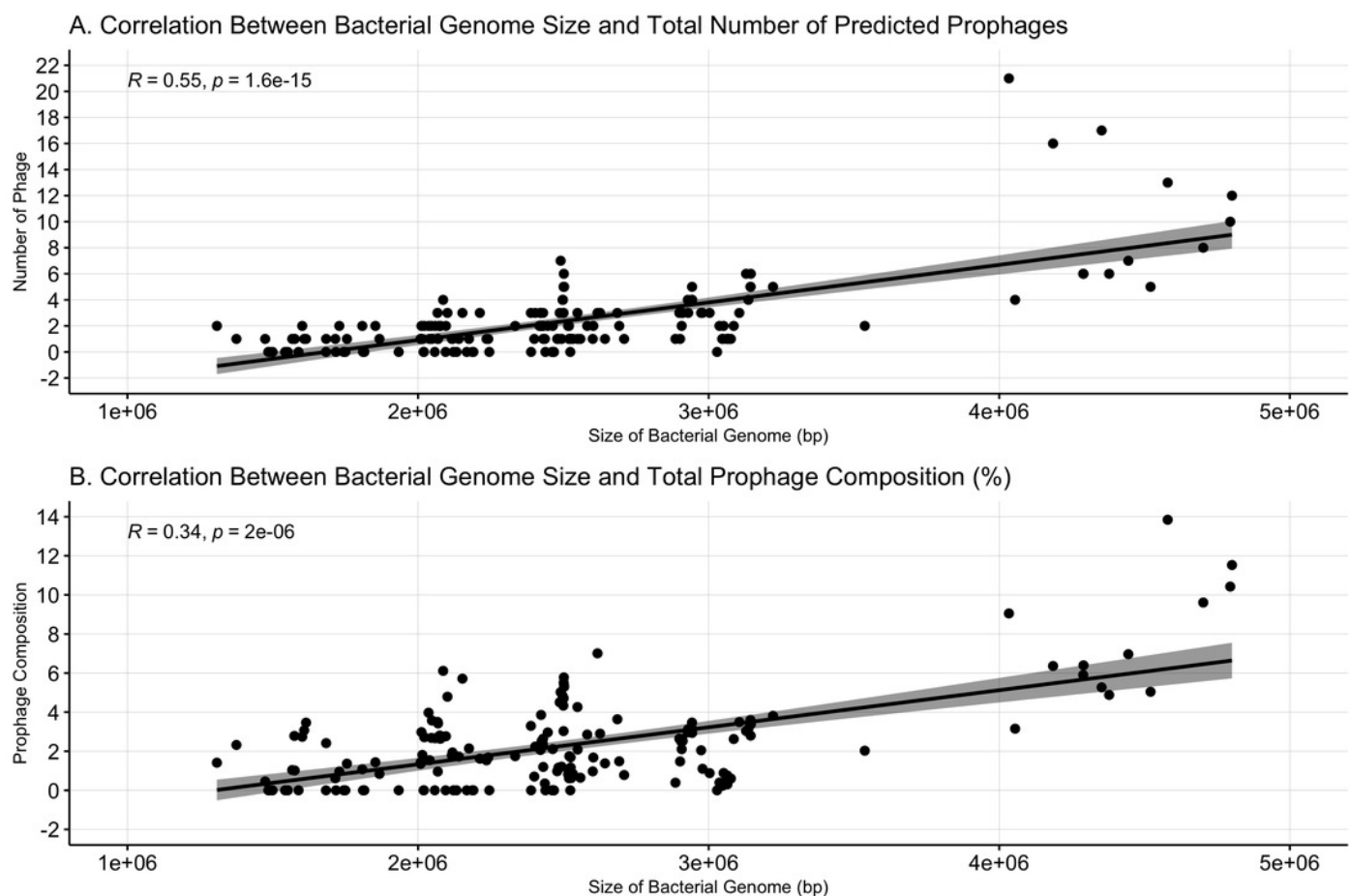
Figure 3

Functional analysis of prophages found in honey bee-associated bacteria

The distribution of phage-associated and COG categories assigned to prophage genes, organized by bacterial host. Similar COG categories are grouped together for visual clarity. X-axis indicates percent. COG Category S (Unknown or No Hits) is not shown.
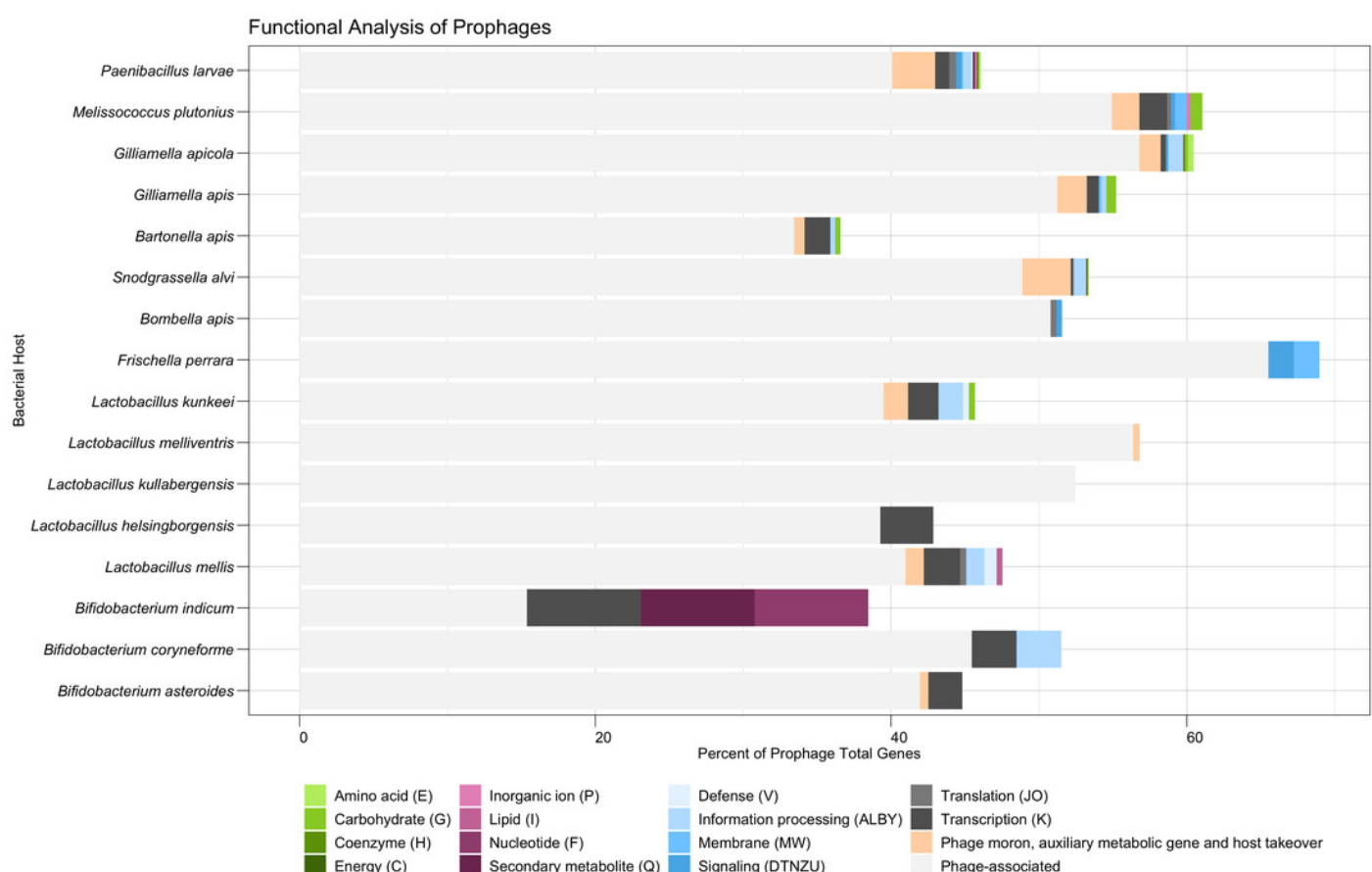
Figure 4

Viral clustering and network analysis based on shared genes

Nodes represent individual prophages. Bacterial hosts are indicated by the node's color and shape. (A) Prophages are separated based on sub-family viral clusters. Numbers above clusters indicate a specific subcluster. Distance between subclusters has no significance, but viral subclusters that come from the same greater viral cluster (indicated by transparent oval overlay) are more related to one another than phage originating from other clusters. Subclusters O6, 1, 7, 8, and 12 contain reference *P. larvae* phage from external studies. Prophage within the outlier cluster are not related to one another. (B) The layout of the network is based on shared genes, with strength indicated by opacity of the network edges. The numbers on nodes indicate subcluster. Blank nodes indicate outliers, while subclusters starting with "O" indicate unique overlapping subclusters.

Bacterial Host of the Prophage

▲ *Bartonella apis*   ■ *Frischella perrara*   ● *Lactobacillus helsingborgensis*   ○ *Lactobacillus melliventris*
▲ *Bifidobacterium asteroides*   ■ *Gilliamella apicola*   ● *Lactobacillus kullabergensis*   ■ *Paenibacillus larvae*
▲ *Bifidobacterium coryneforme* ■ *Gilliamella apis*   ● *Lactobacillus kunkeei*   ◰ *Melissococcus plutonius*
▲ *Bombella apis*   ■ *Snodgrasella alvi*   ● *Lactobacillus mellis*   ⋏ *Reference Phage*
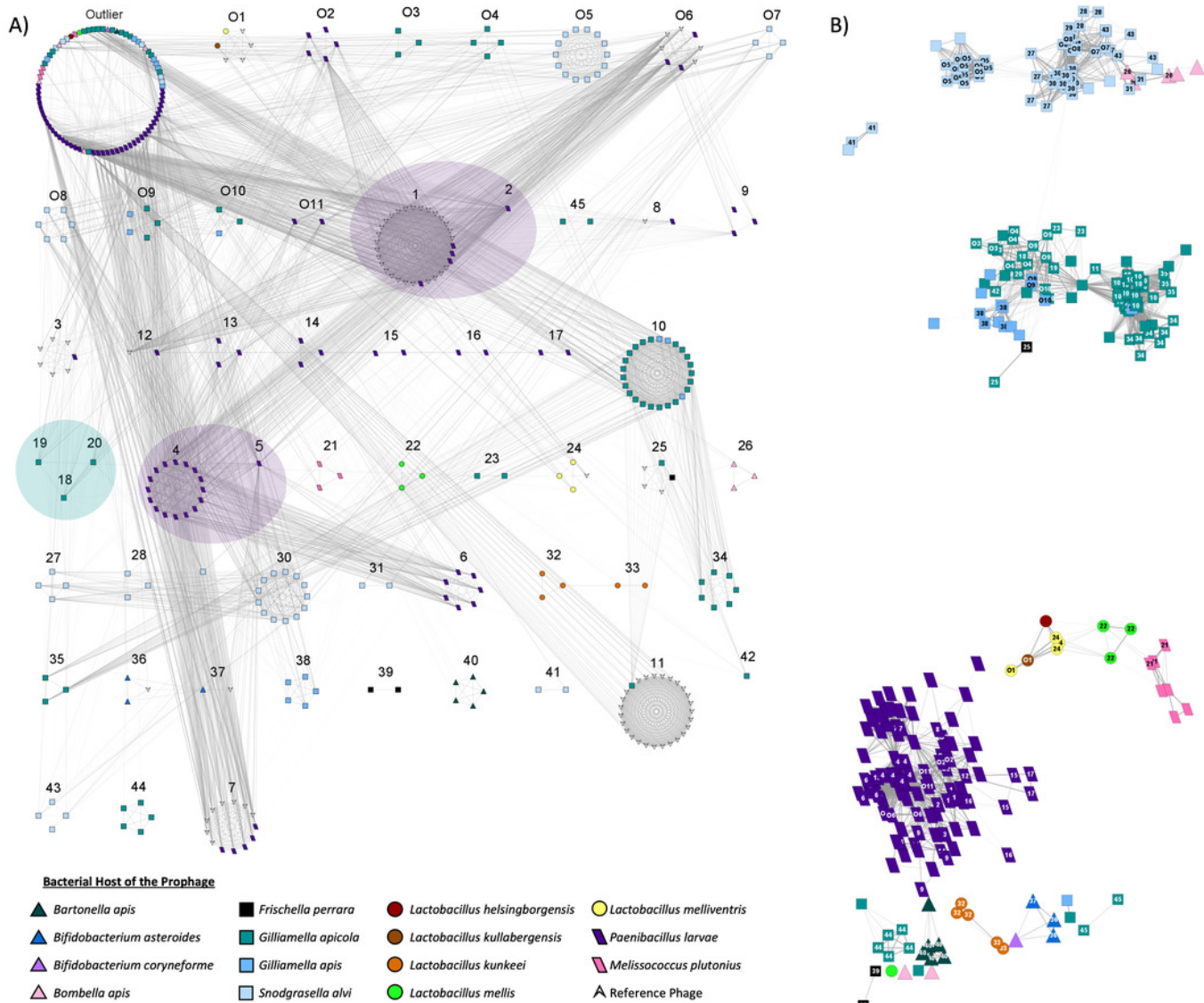
Figure 5

Taxonomic classification of prophages

The lowest taxonomic rank (genus, family, or class) assigned to each prophage based on amino acid similarity to its closest reference phage. Genus and family classifications were assigned if prophages shared AAI across all proteins with a reference above respective thresholds. Prophage which did not meet these thresholds but still matched to a *Caudoviricetes* reference were broadly classified as *Caudoviricetes* class. Prophage which matched to unclassified reference phage were grouped as Unclassified.