Characterization of core bacterial species in the *Daphnia magna* microbiota using shotgun metagenomics

Reilly O. Cooper[1]* & Clayton E. Cressler[1]

[1]School of Biological Sciences, University of Nebraska, Lincoln, NE, USA

*Corresponding author: Reilly O. Cooper, Email: reilly.cooper@huskers.unl.edu

Clayton E. Cressler, Email: ccressler2@unl.edu

Abstract

Background: The keystone zooplankton *Daphnia magna* has recently been used as a model system for understanding host-microbiota interactions. However, the bacterial species present and functions associated with their genomes are not well understood. In order to understand potential functions of these species, we combined 16S rRNA sequencing and shotgun metagenomics to characterize the whole-organism microbiota of *Daphnia magna*.

Results: Five metagenome-assembled genomes (MAGs) were assembled from the *Daphnia magna* microbiota. Phylogenetic placement of these MAGs indicated that two belong to the *Limnohabitans* genus, one to *Polaromonas*, one to *Pedobacter*, and one unclassifiable below the Burkholderiaceae family. Average nucleotide identity of these MAGs to their closest sequenced relative was <95%, suggesting these may be new species in known genera. 16S rRNA community profiling shows that the *Daphnia magna* microbiota is distinct from its culture environment. Genes involved in host colonization and immune system evasion were detected across the MAGs. Some metabolic pathways were specific to some MAGs, including sulfur oxidation, nitrate reduction, and flagellar assembly. Threonine and arginine exporters were encoded by the

Limnohabitans and Burkholderiaceae MAGs, and pathways for key vitamin biosynthesis and export were identified across MAGs.

Conclusions: In this study, we characterize five metagenome-assembled bacterial genomes within the *Daphnia magna* microbiota. Our examination of functions associated with these genomes shows a diversity of nutrient acquisition and metabolism pathways present that may benefit the host, as well as genomic signatures of host association and immune system evasion.

Keywords: Metagenomics, shotgun sequencing, Daphnia magna, microbiota function

Introduction

Organisms are hosts to complex communities of microorganisms that live on any tissue in contact with the environment, collectively known as the microbiota. Species in the microbiota may provide beneficial functions to the host and to other species in the microbiota, including nutrient acquisition and uptake [1], production of host-accessible metabolites [2], host immune system priming [3], and direct pathogen protection [4]. However, characterizing these beneficial host-microbe and microbe-microbe interactions in biologically relevant systems can be difficult due to the number of core

45 species present in the microbiota and the variation in an individual organism's
46 microbiota over time from dietary changes or differential environmental exposure [5].
47
48 The zooplankton *Daphnia magna* provides a useful model for studying functional
49 relationships between microbes and their hosts. *Daphnia* species are used as a model
50 system in ecology, ecotoxicology, and host-parasite dynamics due to their well-
51 documented life cycle and rapid asexual reproduction [6]. The ability to raise *Daphnia*
52 *magna* clonally in the laboratory allows for genetically identical hosts to be used
53 experimentally, reducing the impact of genetic variation on the microbiota.
54 Furthermore, their indiscriminate filter feeding within the water column allows for
55 control over food input. *Daphnia* are colonized with bacteria throughout their entire
56 body cavity and gut [7,8]. Composition of the *Daphnia* microbiota appears to be similar
57 in spatially unique populations [8,9], suggesting mechanisms of acquisition and
58 cultivation of these microbes by the host. The *Daphnia* microbiota is relatively simple at
59 the class level, with β-proteobacteria, γ-proteobacteria, and Flavobacteriia consistently
60 identified at high relative abundances [9,10]. While some work on the *Daphnia*
61 microbiota has used shotgun sequences to examine potential bacterial symbionts [7],
62 the majority of studies have used coarse-level 16S rRNA analyses to profile the taxa
63 present. Species-level identification and subsequent functional profiling of the *Daphnia*

64 microbiota has not been achieved with these studies, as resolution of the microbiota
65 using 16S rRNA sequencing can only reliably identify bacterial genera. Much of the
66 genomic content of *Daphnia*-associated bacteria is unknown, and potentially novel taxa
67 may not be identified by standard 16S rRNA techniques.
68
69 Beyond identifying the composition of the microbiota, it is clear that this composition
70 affects host fitness. The microbiota as a whole has been implicated in nutrient
71 acquisition and breakdown of toxic compounds [11], and survival and growth are
72 affected by the presence of, and perturbations to, the microbiota [12-14]. Host
73 fecundity has been specifically tied to the most abundant genus in the *Daphnia*
74 microbiota, *Limnohabitans*. Entirely bacteria-free *Daphnia* experience significant
75 declines in fecundity, but monocolonization of bacteria-free *Daphnia* with
76 *Limnohabitans* restores fecundity to that of *Daphnia* with a complete microbiota [15].
77 However, it is unclear what metabolic functions *Limnohabitans* provides to achieve
78 these effects. How the other species in the *Daphnia* microbiota contribute to host life
79 history and what functions these species may be providing is entirely unclear.
80
81 Here, we use shotgun metagenomics to characterize the bacterial species present in
82 the *Daphnia magna* microbiota. The metagenome-assembled genomes generated from

this data were then used to examine potential metabolic functions of the microbiota in total and for each species assembled. This study is the first to report metagenome-assembled genomes of bacteria in the *Daphnia magna* microbiota and is the first to suggest functions based on gene content of these microbes. This increased resolution allows us to formulate testable hypotheses about the metabolic interactions happening between the host and microbes, and among microbes, that impact host fitness. This functional knowledge will provide a new lens for studying this important ecological model system.

Results

Shotgun sequencing, assembly, and binning

Shotgun sequencing of the adult and juvenile *Daphnia magna* microbiota resulted in 20.19Mbp of paired-end Illumina read data, which was then reduced to 9.64Mbp after quality trimming and host genome filtering. A co-assembly of all four samples generated 174,991 contigs (N50 of 1,923bp; longest assembled contig 226,522bp). Summaries of assembly statistics for the co-assemblies and the individual sample assemblies can be found in Table 1. Identification of high-quality reads and ≥1000bp assembled contigs using Kraken, Kaiju, and MetaPhlAn2 indicated 18 genera that were present as more than 1% of the sample (Supplementary Figure 1). Of the 18 genera,

only *Limnohabitans* was identified by all three tools in reads and contigs. Other abundant genera included *Pedobacter*, *Flavobacterium*, *Polaromonas* and other unclassified Burkholderiaceae. 16S rRNA sequencing of the library preparation and DNA extraction kits resulted in <50 reads (Figure 1A). 16S rRNA community profiles of the *Daphnia magna* food source, *Chlamydomonas reinhardtii*, the COMBO culture media *Daphnia magna* are raised in, and samples of 5 healthy adult *Daphnia magna* were found to have differences in composition (Figure 1A). The *Chlamydomonas* samples showed reduced relative abundance of Proteobacteria as compared to adult *Daphnia* and the culture media. The culture media showed higher relative abundance of Actinobacteria, and healthy *Daphnia magna* primarily were colonized by Proteobacteria and Bacteroidetes. These same community profiles were analyzed using an unweighted principal coordinate analysis (PCoA) on the unweighted UniFrac distances between samples, which represents the phylogenetic relatedness between samples based on ASV presence and absence (Supplementary Figure 2). The three sample groups were found to cluster separately, with no overlap.

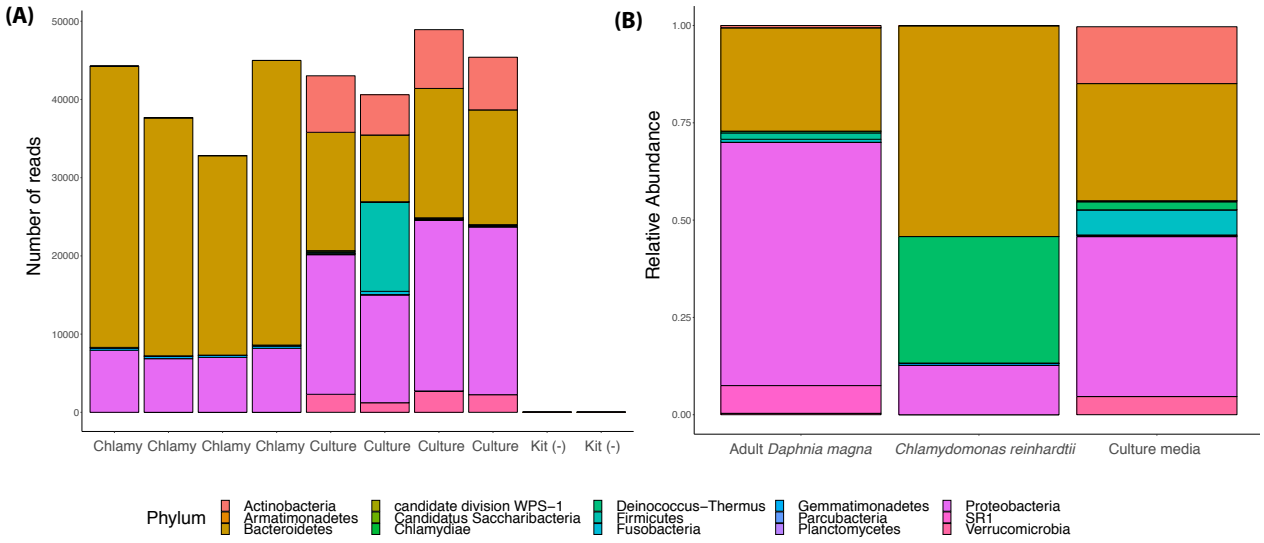


Figure 1. **(A)** Phylum-level 16S rRNA profiles of *Chlamydomonas reinhardtii* (Chlamy, n = 4), COMBO culture media (Culture, n= 4), and the DNA extraction kit and library preparation kit used for sequencing (Kit (-), n=2). **(B)** Relative abundance of phyla generated from 16S rRNA community profiles in *Chlamydomonas reinhardtii*, healthy adult *Daphnia magna*, and the culture media (all n=4).

Assembly	Total assembly length (bp)	Number of contigs	Number of contigs ≥1 kb	Largest contig (bp)	GC content (%)	N50
Master co-assembly	50,283,761	174,991	9,449	226,522	51.7	1,923
Adult co-assembly	27,616,512	93,427	4,021	103,086	53.61	1,535
Juvenile co-assembly	32,024,657	17,982	6,627	142,752	49.18	3,165
Adult sample 1	11,565,605	39,625	2,783	62,388	52.42	1,546

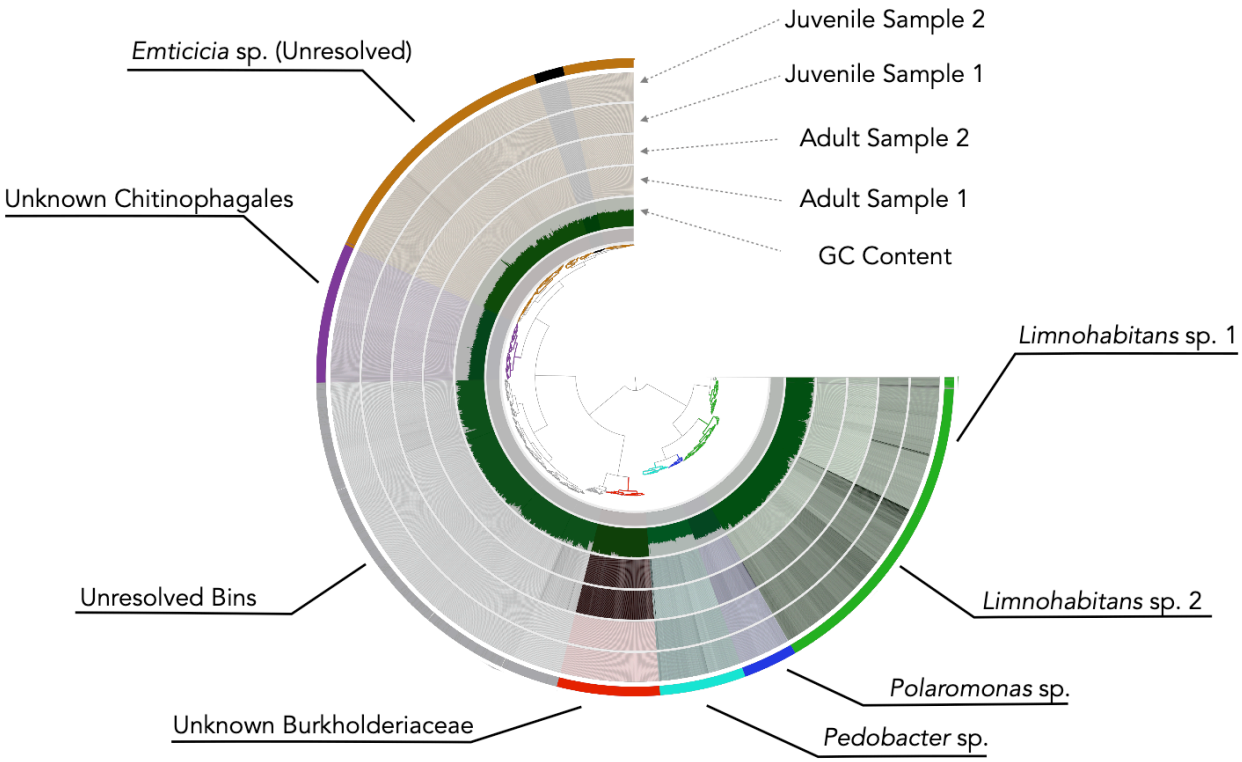
Adult sample 2	22,274,108	65,709	3,585	85,806	54.09	2,074
Juvenile sample 1	24,570,320	65,656	5,093	85,390	49.05	3,179
Juvenile sample 2	22,525,058	50,433	4695	65,631	49.35	3,256

Table 1. Summary of assembly statistics for individual sample assemblies and co-assemblies of all four samples (master co-assembly), the two adult samples (adult co-assembly), and the two juvenile samples (juvenile co-assembly).

CONCOCT clustering of contigs from the master co-assembly by tetranucleotide composition, GC content, and coverage resulted in 19 bins of draft and incomplete genomes. Manual curation, refinement, and merging using Anvi'o resulted in 15 bins, with 7 meeting the ≥50 quality metric. The adult-only co-assembly generated 12 bins, which further refined to 5 bins, with 4 meeting the ≥50 quality metric (Supplementary Figure 5). Three of the four high-quality bins were also present in the master co-assembly and one unique bin was identified. The juvenile-only co-assembly generated 15 bins, which further refined to 7 bins, with 4 meeting the ≥50 quality metric (Supplementary Figure 6). Because the master co-assembly generated bins with higher coverage or higher completeness, we used GTDB-Tk on bins refined in the master co-assembly only. Phylogenetic placement using GTDB-Tk for the seven high- and medium-quality bacterial MAGs identified from the master co-assembly placed two in the *Limnohabitans* genus (*Limnohabitans* MAG 1 and MAG 2), one in the *Polaromonas*

142 genus (*Polaromonas* MAG), one unable to be placed below the Burkholderiaceae family
143 (Burkholderiaceae MAG), one in the *Pedobacter* genus (*Pedobacter* MAG), one in the
144 *Emticicia* genus, and one unable to be placed below the Chitinophagales order
145 (Supplementary Table 1; Supplementary Figure 4). None of the MAGs mapped to
146 previously identified species in the GTDB-Tk database. Average nucleotide identity of
147 the two *Limnohabitans* MAGs was measured at 86.6% similarity. This is less than the
148 95% cutoff generally accepted for strain-levels similarity, suggesting the MAGs are
149 likely separate species [16].
150
151 To examine potential functions of the Daphnia magna microbiota, we focused on the
152 three most complete MAGs: unknown Burkholderiaceae, 99.28% complete, 11.5X
153 mean coverage; *Pedobacter* sp., 98.56% complete, 14.87X mean coverage; and
154 *Polaromonas* sp., 82.73% complete, 15.3X mean coverage. We also examined the two
155 *Limnohabitans* MAGs: sp1, 78.42% complete, 29.99X coverage; sp2, 60.43% complete,
156 14.16X coverage (Figure 2). Though the *Limnohabitans* MAGs were not over 90%
157 complete, according to read-based and contig-based identification tools they were the
158 most abundant, and coverage of both exceeded 30x in the juvenile *D. magna* samples.
159 Furthermore, prior work has indicated the presence and importance of this genus in the
160 *Daphnia magna* microbiota, suggesting unique functions may be present in these

161 genomes. We extracted the nearest matching reference genome for each of the five
162 MAGs above and calculated ANI for each to further investigate if these MAGs were
163 novel or just new strains of already sequenced species (Table 2).
164



165

166 **Figure 2. (A)** Anvi'o display of the two adult and two juvenile samples with the
167 metagenome-assembled genomes highlighted in color. The innermost dendogram
168 represents similarity among contigs based on sequence composition, GC content
169 (green inner ring), and differential coverage (black bars across the four grey rings). The
170 outermost layer shows the genome bins. Genome bin identification performed by
171 GTDB-Tk.

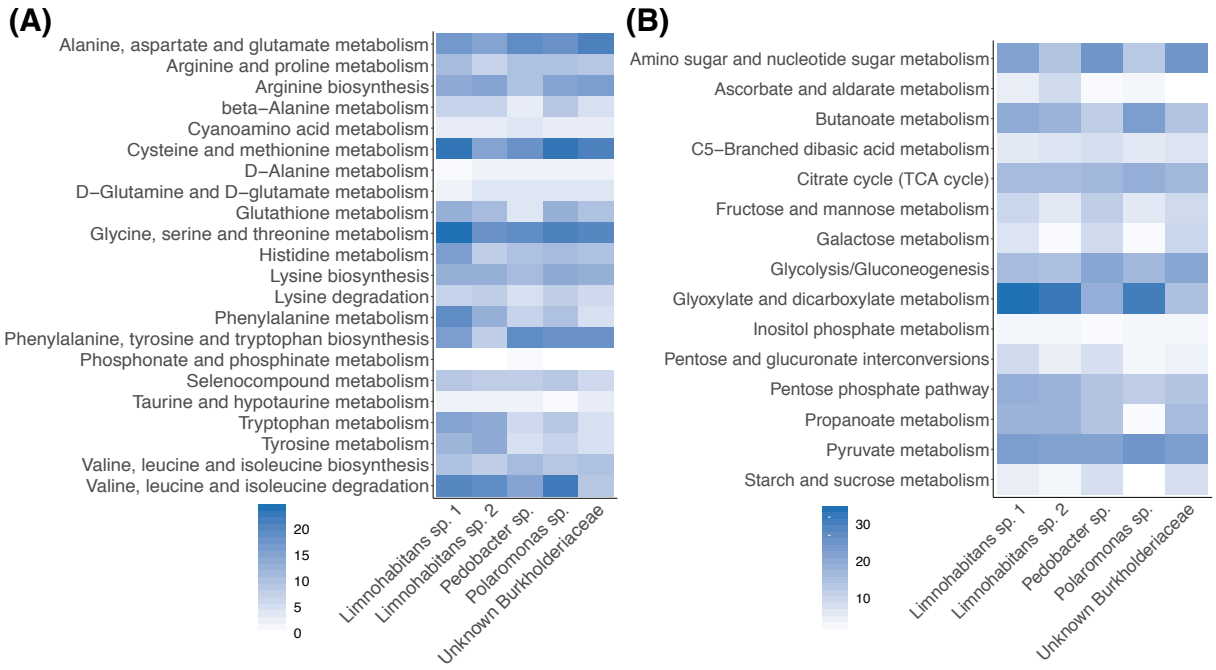
MAG	Closest Sequenced Relative	ANI
Limnohabitans sp. 1	*Limnohabitans* sp. 63ED37-2	0.879
Limnohabitans sp. 2	*Limnohabitans* sp. Rim28	0.820
Unknown Burkholderiaceae	*Pigmentiphaga* sp. NML080357	0.693
Pedobacter sp.	*Pedobacter ruber*	0.773
Polaromonas sp.	*Polaromonas* sp. A23	0.761

Table 2. Average nucleotide identity of the five MAGs of interest to their nearest sequenced relative. The closest sequenced relative was identified from placement within GTDB-Tk's reference phylogeny output.

Functional profiling of the five high- and medium-quality bacterial metagenome-assembled genomes

Contigs from the entire metagenome and from the 5 MAGs independently were analyzed for potential coding sequences (CDS) using Prokka. All identified CDS were then queried against the KEGG database using GhostKOALA and identified orthologs of functional genes were grouped and quantified into KEGG functional categories (Figure 3). We focused on pathways associated with respiration, carbohydrate metabolism, amino acid metabolism, other energy metabolism, and transport. These functional categories indicate not only general metabolism for these species, but also

could indicate what functions these species are providing to the host. Out of the 7,453 CDS from the five high-quality MAGs that mapped to KEGG pathways, 707 (9.48%) were associated with carbohydrate metabolism, 642 (8.71%) were associated with amino acid metabolism, and 358 (4.80%) were associated with other energy metabolism (Figure 3A-B; Supplementary Figure 3). Although *C. reinhardtii* has a relatively high concentration of lipids [17], only 188 (2.52%) of all CDS were associated with lipid metabolism. To understand how much genetic overlap there was among the five MAGs, we compared predicted genes identified with Prokka and clustered with OrthoVenn. The two *Limnohabitans* MAGs shared the highest number of genes (641), and across all MAGs 325 genes were shared (Figure 4). The *Pedobacter* MAG had the most unique genes in its gene set (117), while the Burkholderiaceae had so much overlap with other MAGs that it had relatively few unique genes (21).



200

201 **Figure 3. (A)** Number of genes associated with carbohydrate metabolism pathways in

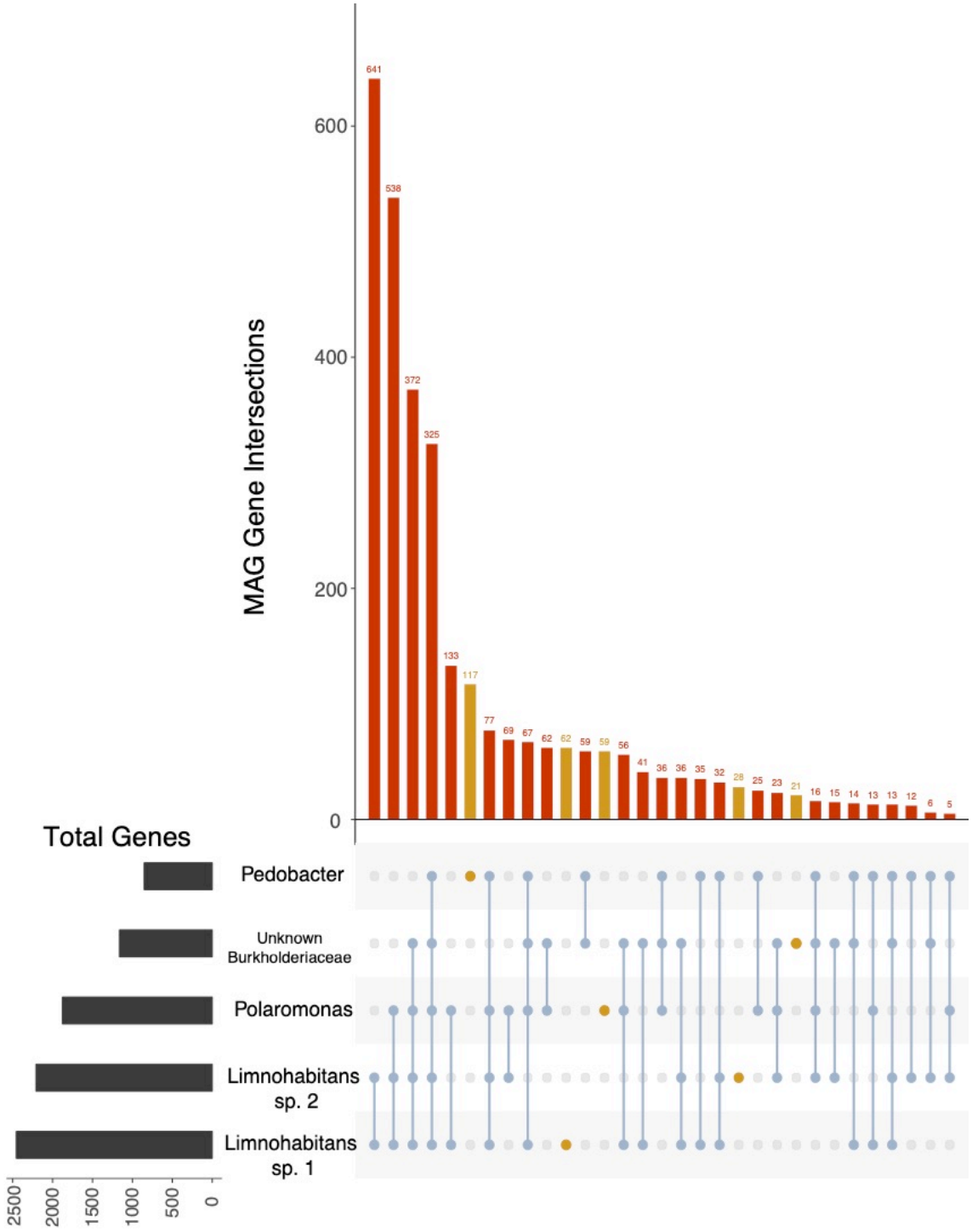
202 the five MAGs. Genes within each MAG bin were annotated with Prokka, assigned

203 KEGG orthology, and mapped to KEGG metabolic pathways. Genes assigned to

204 pathways were then counted. **(B)** Number of genes associated with KEGG amino acid

205 metabolism pathways in the five MAGs.

206



207

208 **Figure 4.** Shared and unique genes with KEGG orthologs in each MAG. Unique gene

209 sets are highlighted in gold; shared genes are in red. Intersections between MAGs, or

numbers of genes in each MAG's gene set shared with the other MAGs connected by lines, are noted in the bottom matrix. The total number of genes in each MAG is listed on the left side of the matrix.

(Table after references)

Table 3. Shared and unique complete KEGG pathways in the five high- and medium-quality MAGs. 'MAG' column indicates which MAG encodes that complete pathway; 'All' denotes that the pathway is shared among all five MAGs.

After examining genetic overlap and unique gene content of the five MAGs, we examined complete metabolic pathways annotated by KEGG present in the microbiota. Here, we describe some of the complete pathways annotated, specifically focusing on metabolic pathways involved in critical functions for bacteria such as nutrient uptake and biosynthesis as well as pathways that may be involved in host or environment interaction. A description and comparison of each MAG's set of complete pathways can be found in Supplementary Materials and in Supplemental Tables 2-5. We separate complete pathways into those encoded by all or multiple MAGs, indicating shared pathways that could indicate functional redundancy or common metabolites accessible to multiple species, and pathways uniquely encoded by single MAGs, indicating

229 potential niche differentiation within the microbiota [18]. Finally, we summarize some of
230 the described shared and unique functions and some functions specifically associated
231 with host association as genomic signals of host association, providing potential
232 explanations for why these genera are commonly found in the *Daphnia magna*
233 microbiota.
234
235 ## Shared functions of the *Daphnia magna* microbiota
236 *Nutrient uptake and major biosynthesis pathways.* All five MAGs shared genes involved
237 in some key metabolic pathways (Table 3, 'All' MAGs, Supplementary Tables 2 & 3). In
238 all MAGs, a complete TCA cycle was encoded. Genes encoding a cytochrome *c*
239 oxidase were identified in all MAGs, suggesting all have the capacity for aerobic
240 respiration. All encoded for lipopolysaccharide transport and lipoprotein release. All
241 MAGs except the *Pedobacter* encoded for a complete glyoxylate cycle, and the non-
242 oxidative phase of the pentose phosphate pathway was present in all MAGs except the
243 *Limnohabitans* species. Transporters for multiple TCA cycle intermediates are present
244 across the MAGs, including a C_4-dicarboxylate transport system in all MAGs that allows
245 for transport of multiple different molecules into bacterial cells. Other transporters for
246 TCA cycle intermediates encoded included those for alpha-glucosides, malate,
247 fumarate, 2-oxogultarate, succinate, and aspartate.

248

249 Transport and biosynthesis of other essential molecules were shared across MAGs as
250 well. All MAGs except *Pedobacter* shared transport systems for phospholipids,
251 phosphate, and branched-chain amino acids. All species except for *Limnohabitans*
252 MAG 2 encoded for the elongation step of fatty acid biosynthesis, but only the
253 Burkholderiaceae, *Polaromonas*, and *Pedobacter* MAGs encoded for fatty acid
254 biosynthesis initiation. The *Limnohabitans* MAGs shared multiple copies of sialic acid
255 TRAP transporter genes (*siaM*, *siaT*, *siaQ*).

256

257 All genomes encoded for the transport of multiple necessary vitamins, including
258 riboflavin (B_2, *ribX*, *ribY*, *ribZ*), pantothenate precursors (B_5, *panS*), and cobalamin (B_{12},
259 *btuB*). Genes involved in the salvage pathway for cobalamin were also present in both
260 *Limnohabitans* MAGs and in the *Polaromonas* MAGs (*cobO*, *cobP*). Some species were
261 also able to biosynthesize vitamins: the *Pedobacter* and *Polaromonas* MAGs encoded
262 for biotin biosynthesis, and all MAGs except *Pedobacter* encoded for tetrahydrofolate
263 biosynthesis. All MAGs encoded for a pyridoxine 5-phosphate synthase (PdxJ),
264 indicating likely biosynthesis of pyridoxine (B_6).

265

Biosynthesis and transport of amino acids varied across MAGs, with no shared amino acid transport systems present across all species. All MAGs except *Pedobacter* were able to biosynthesize arginine and ornithine. The Burkholderiacae, *Limnohabitans* 1, and *Polaromonas* MAGs encoded for threonine, cysteine, lysine, proline, phenylalanine, tyrosine, and glutathione biosynthesis. Valine, leucine, and tryptophan were also able to be biosynthesized by some of the MAGs.

Host and environment interaction. Superoxide dismutase (*sodA*, *sodB*) and catalase-peroxidase (*katG*) were found in all MAGs. These act as reactive oxygen species detoxifiers and scavengers. The universal minimal Tat system was encoded by all MAGs (*tatA*, *tatC*), and both *Limnohabitans* MAGs also encoded *tatB*, allowing all of these species to transport folded proteins across their cell membranes. Complete adhesin transport systems and a gene associated with type IV pilus biosynthesis were present in both *Limnohabitans* MAGs and the *Polaromonas* MAG (*pilQ*). These MAGs also encoded two membrane proteases related to aminoglycoside resistance (HtpX, FtsH) [19]. All MAGs except the *Pedobacter* encoded for the QseC-QseB quorum sensing regulatory system. Though the *Polaromonas* MAG was the only one to encode for complete flagellar assembly, the *Limnohabitans* MAGs encoded for multiple genes

284 involved in flagellar assembly (*flhA*, *flhB*, *flgB – L*). A gene for chemotaxis protein CheA
285 was present in these three MAGs as well.
286
287 As none of the MAGs encoded for all necessary amino acid biosynthesis pathways,
288 some amino acid importers were present across the MAGs. These include arginine
289 (*artM*), proline (*proP*), cystine (*yecS*) in *Limnohabitans* MAGs, glutamine (*glnQ*, *glnM*)
290 and glutathione (*gsiC*) in *Limnohabitans* MAGs and in the *Polaromonas* MAG, and
291 histidine (*hisP*, *hisM*, *hisQ*) in the Burkholderiaceae MAG and *Limnohabitans* MAG 1. It
292 is unclear whether food (e.g., *Chlamydomonas reinhardtii*) or export from other bacteria
293 is the source of these amino acids for these MAGs. As some MAGs are able to
294 biosynthesize amino acids that others must import, there may be cross-feeding among
295 species in the microbiota occurring.
296
297 Environmental stress tolerance mechanisms were shared among genomes as well. All
298 encoded for the phosphate starvation response regulatory system PhoR-PhoB. The
299 *Limnohabitans* MAGs and the *Polaromonas* MAG shared the EnvZ-OmpR osmotic
300 stress response system, allowing the bacterial cells to respond to changes in osmolality.
301 *Limnohabitans* MAG 2 and the *Polaromonas* MAG shared CusS-CusR, a copper
302 tolerance regulatory system.

303

304 ## MAG-specific functions of the *Daphnia magna* microbiota

305 *Nutrient uptake and major energy pathways.* There were potential differences in
306 respiration for some of the MAGs. The Burkholderiaceae encoded a nitrate/nitrite
307 transporter (*narK*) and a respiratory nitrate reductase (*narGHI*). Use of this system for
308 respiration was supported by the presence of genes encoding the nitrate respiration
309 two-component regulatory system NarX-NarL. The *Polaromonas* encoded for
310 thiosulfate transport (*cysT*, *cysW*, *cysP*) and potential use via a subunit of the Sox
311 complex (*soxA*). We found a supplementary energy pathway in the *Limnohabitans*
312 MAGs, which share a pathway for the biosynthesis of bacteriochlorophyll (*pufLM*, *bchY*).
313 This pathway has been documented in other *Limnohabitans* species [20].

314

315 Differences in carbohydrate metabolism among MAGs was apparent. The *Pedobacter*
316 MAG was the only high-quality genome to encode for glycolysis and glycogen
317 biosynthesis; the Burkholderiaceae MAG for the Leloir pathway; and *Limnohabitans*
318 MAG 2 for the Entner-Doudoroff pathway. The *Polaromonas* MAG encoded for a
319 complete beta-oxidation system, enabling fatty acid metabolism. Different
320 carbohydrate transporters were found in the MAGs as well. *Limnohabitans* MAG 1
321 uniquely encoded for the transport of glucose, mannose, and glycerol (*gtsABC*, *malK*).

322 It also encoded for transport of L-arabinose (*araPQ*) and for a semiSWEET general sugar
323 transporter. The Burkholderiaceae MAG encoded for methyl-galactoside transport
324 (*mglABC*) and a galactose processing pathway. The *Pedobacter* MAG can potentially
325 utilize chitin, as the MAG uniquely encoded for chitin degradation to fructose-6-
326 phosphate (*chiA*, *chb*, *nagB*).
327
328 Other pathways for vitamin and amino acid import, biosynthesis, and degradation were
329 uniquely present in some genomes. *Limnohabitans* MAG 1 was the only MAG to
330 encode taurine import (*tauA*, *tauB*). This MAG also uniquely encoded a pathway for the
331 degradation of histidine to glutamate. *Limnohabitans* MAG 2 could transport thiamine
332 into the cell via a putative thiamine transport system (KEGG Module M00192).
333 Uniquely, *Polaromonas* could synthesize cobalamin (*cobA*, *cobQ*, *cbiB*, *cobP, cobC*). It
334 may also synthesize riboflavin via the purine biosynthesis pathway.
335
336 *Host and environment interaction.* Multiple genes and pathways involved in antibiotic
337 resistance and detoxification were found in the bacterial MAGs. *Pedobacter* encoded
338 an MdlAB/SmdAB transporter as well as two genes involved in macrolide (*macA*, *macB*).
339 The *Polaromonas* MAG encoded an AcrAB-TolC/SmeDEF efflux pump, and the

Burkholderiaceae MAG encoded the BaeS-BaeR envelope stress response two-component regulatory system.

The MAGs encoded for different secretion systems and regulatory systems. The *Polaromonas* MAG was the only one to encode for a type I secretion system (RaxAB-RaxC). The Burkholderiaceae MAG encoded for an alpha-hemolysin and cyclolysin secretion system and a type IV secretion system (*virB1 – 5*, *10*, *11*). The Burkholderiaceae MAG encoded for an osmotically-inducible protein, OsmY, part of an osmoprotectant ABC transporter complex, and a putrescine transporter (*potFGHI*).

Genomic signs of host association in the *Daphnia magna* microbiota

Because the *Daphnia magna* microbiota has been shown to differ significantly from the surrounding aquatic environment at the genus level [9,21], we examined the MAGs for any potential indicators of host association. Multiple instances of host immune system evasion or tolerance were noted in the MAGs. Both *Limnohabitans* MAGs encoded for the microbial stealth protein CpsY, implicated in host immune system evasion [22]. Genes involved in dTDP-L-rhamnose biosynthesis were present in all MAGs, indicating these species are able to use rhamnose as an alternative cell wall polysaccharide. L-rhamnose has been linked to bacterial viability and virulence [23], and similarly

359 structured polysaccharides have been shown to modulate host immune systems [24]. A
360 gene involved in quorum quenching (signaling between other microbial species or
361 between microbes and the host) was identified in *Limnohabitans* MAG 1 and the
362 *Pedobacter* MAG (*ytnP*). Other virulence factors and regulators were identified in both
363 *Limnohabitans* and the *Pedobacter* MAG (*cvfB*, *bvgS/bvgA*).
364
365 These bacterial species may benefit the host via amino acid and vitamin biosynthesis
366 and export. Complex eukaryotes must acquire some essential vitamins and amino acids
367 from their diet or from heterotrophic microorganisms. Six amino acids have been
368 demonstrated as essential for *Daphnia magna's* close relative, *Daphnia pulex*: arginine,
369 histidine, leucine, phenylalanine, isoleucine, and tryptophan [25]. All of the dietary
370 essential amino acids are biosynthesized by at least one of the MAGs, and some may
371 be able to export these as well. The *Limnohabitans* MAGs and the Burkholderiaceae
372 MAG encode for an arginine exporter (*argO*), *Limnohabitans* sp. 1 encodes for a
373 threonine exporter (*rhtA*), and the Burkholderiaceae MAG a threonine-serine exchanger
374 (*steT)*. The *Chlamydomonas reinhardtii* food source used here does not contain
375 cobalamin [26], a beneficial vitamin for *Daphnia magna*, suggesting that *Daphnia*
376 *magna* raised in culture must acquire cobalamin from the culture media and from
377 microorganisms. In the *Daphnia* metagenome, the *Polaromonas* MAG encodes for

cobalamin biosynthesis. Other vitamins are biosynthesized by at least one of the MAGs, including tetrahydrofolate, biotin, and pyridoxine. Supplementation of *Daphnia magna* growth media with biotin and cobalamin has been shown to increase host fitness [27], though it is unclear if these vitamins are essential for *Daphnia* survival.

Other host-microbe interactions were also present in the MAGs. An essential nutrient for bacteria is iron, and catalase genes were found in all MAGs along with other ABC transporters for iron. The *Limnohabitans* MAG 2 encoded for heme binding proteins and a heme export system (*ccmB*, *ccmC*), systems implicated in bacterial use of host-synthesized heme [28]. The *Pedobacter* MAG encoded for an N-acetylneuraminate lyase (*nanA*) and all five MAGs contain genes involved in transport of sialic acids. As sialic acids are found in complex host tissue [29], this may indicate cleavage of sialic acids from host cells for import and use by the bacteria. Pathways involved in host invasion and colonization were also present in the MAGs. The *Polaromonas* MAG encoded a suite of type IV pilus and fimbriae associated genes, including type IV pilus biogenesis factors (*pilY1*, *pilQ*), fimbrial proteins (*pilE*), the type IV fimbriae synthesis two-component regulatory system PilS-PilR, and a complete set of flagellar assembly genes. This MAG, along with the two *Limnohabitans* Mags, also encoded for twitching motility (*pilT*). Both *Limnohabitans* MAGs encoded for swarming motility (*swrC*, *rssA*),

397 and all three for other genes involved in chemotaxis (*pctA*, *cheA*, *cheY*, *tar*, *cheB*,
398 *mcp4*, *tsr*, *cheW*). Though genes for adhesin production were not identified, genes
399 involved in adhesin transport were identified in both *Limnohabitans* MAGs and in the
400 Burkholderiaceae MAG (*bmaC*, *ehaG*, *ata*).
401

402 ## Discussion

403 To date, studies examining the *Daphnia magna* microbiota have only sequenced the
404 16S rRNA marker gene to understand broad-level interactions and functions of the
405 microbiota. Here, we were able to assemble the first metagenome-assembled genomes
406 from the *Daphnia magna* microbiota to elucidate genome-specific functions associated
407 with highly abundant members of the bacterial community. Our 16S rRNA gene level
408 data shows that the *Daphnia magna* bacterial community is structured differently than
409 the surrounding environment and from the microbiome of their food, which has been
410 also been corroborated by earlier studies [9,30]. We identified five MAGs that were
411 distinct from each other and distinct from their closest sequenced relatives based on
412 average nucleotide identity. We found that these MAGs were high- or medium-quality,
413 meaning they contained some or most of the single-copy genes found in all bacteria.
414 With metagenomic short-read shotgun sequencing it is unlikely that the MAGs
415 assembled here contained the full gene sets associated with these species; however,

416 bins of relatively high-quality suggest that we found many important genes within each
417 assembled genome. Most studies of the *Daphnia magna* microbiota using marker-
418 based sequencing have identified more OTUs or ASVs and higher diversity than the 12
419 bins we identified, likely due to the higher sequencing depth necessary for shotgun
420 sequencing to identify rare taxa [31]. However, the five MAGs assembled from our
421 shotgun sequencing are mostly consistent with genera found in 16S rRNA sequencing
422 from other studies and from our own sequencing.
423
424 The Burkholderiales order has been demonstrated as the most abundant in the
425 *Daphnia* gut and whole organism microbiota [9,13], and four of the five MAGs
426 assembled here were identified to families or genera within this order. Two
427 *Limnohabitans* MAGs were assembled and found at high abundances in adults and
428 juveniles. Within the Burkholderiales, the genus *Limnohabitans* has been reported to be
429 highly abundant in the *Daphnia* microbiota, and has been implicated in increasing host
430 fecundity [12,15]. Here, we show that there are two distinct *Limnohabitans* MAGs that
431 encode for different metabolic potential. Only one other study has definitively
432 identified more than one OTU in this genus [13]. We also identify two other MAGs in
433 the Burkholderiales order, including a *Polaromonas* species and one unclassified

434 Burkholderiaceae. Surprisingly, we also identify a *Pedobacter* MAG, which has
435 previously only been reported as a rare taxon in the *Daphnia* microbiota [32].
436
437 Analysis of annotated genes and pathways across the five MAGs showed overlap and
438 differences in metabolism. Key pathways such as the TCA cycle were shared across all
439 MAGs, and multiple ABC transporters of key vitamins and amino acids were identified.
440 Many genes encoding for the use of different carbohydrates were encoded across the
441 MAGs. Microalgae are generally rich in carbohydrates [33] and serve as a major food
442 source for *Daphnia magna* [34], potentially allowing these microbes easy access to
443 myriad nutrient sources. The two *Limnohabitans* MAGs shared a high proportion of
444 annotated genes, suggesting some functional similarity between them. Indeed,
445 multiple metabolic pathways were shared between these two MAGs, including the
446 glyoxylate cycle, bacteriochlorophyll biosynthesis, and biosynthesis of some vitamins
447 and amino acids. The *Limnohabitans* MAGs also shared many genomic features with
448 the *Polaromonas* MAG and the Burkholderiaceae MAG, including multiple TCA cycle
449 intermediate importers and several transport systems. There is also potential flexibility
450 among the taxa in encoded respiration, notably in the *Polaromonas* MAG's ability to
451 use thiosulfate and the Burkholderiaceae's ability to utilize nitrate under hypoxic
452 conditions. Along with the variety of two-component regulatory systems found within

453 each MAG, the wide range of potential respiratory pathways may allow functions to
454 sustain through different bacterial species even when stressful environments cause
455 fluctuations in the abundance of species within the microbiota.
456
457 The difference between taxa identified in the *Daphnia magna* microbiota and its
458 environment suggest that there are some key interactions between the host and its
459 associated microbes in order to establish and maintain these microbial populations.
460 Furthermore, there may be some interactions between bacterial species in the
461 microbiota that could impact the host. We found many genes in the *Limnohabitans*
462 MAGs and the *Polaromonas* MAG involved in flagellar assembly, type IV pilus
463 biogenesis and production, and biofilm formation, all of which have been implicated in
464 host colonization and successful adhesion to host-associated surfaces [35,36,37]. All
465 MAGs encode for l-rhamnose production, which has been implicated in adhesion to
466 other cells [23]. Genes in secretion systems implicated in host cell adhesion, particularly
467 Type I, II, and IV were also encoded [38]. We found many genes involved in host
468 immune system evasion or modification, which may allow these bacteria to persist
469 within the host species [39]. Notably, superoxide dismutase and catalase were encoded
470 by multiple MAGs, suggesting the bacteria could defend against radical oxygen species
471 produced by the host as a defense mechanism [40]. Also present across MAGs are

472 genes involved in the detoxification of antibiotics and toxins, including multidrug efflux
473 transporters and pumps (MdlAB/SmdAB, AcrAB-TolC/SmeDEF), macrolide export
474 (*macA*, *macB*), and stress tolerance to antibiotics (BaeS-BaeR).
475
476 Biosynthesis and provision of amino acids by bacteria to their host is a well-
477 documented set of interactions that is known to confer fitness benefits to the host [41-
478 43]. Here, we find that the *Limnohabitans* MAGs and the Burkholderiaceae MAG
479 encode for the export of arginine and threonine, essential amino acids for the host
480 *Daphnia* [25]. Similarly, biosynthesis or metabolism of vitamins and minerals by bacteria
481 and provisioning to the host has been well-documented in the microbiota of other
482 organisms [44-46]. Many genes involved in vitamin B biosynthesis were found across all
483 MAGs. Media supplemented with cobalamin (B_{12}) is used to successfully culture
484 *Daphnia magna* [47], and we find the *Polaromonas* MAG encoded for a complete
485 cobalamin biosynthesis pathway, suggesting potential vitamin provisioning to the host
486 from this species. We also find a potential microbe-microbe interaction, where the
487 *Pedobacter* MAG encodes for cleavage of sialic acids from host tissue, where it can
488 then be transported and utilized by other species as a carbohydrate source via a sialic
489 acid TRAP transporter. The breakdown of sialic acid to metabolites that are accessible
490 to the host by microbiome-associated species has been shown to increase host fitness

491 [48]. If *Limnohabitans* are able to use sialic acid as a nutrient source, this may be the
492 basis for a microbe-host-microbe interaction, where *Limnohabitans* provides essential
493 amino acids to the host using energy generated from metabolism of molecules
494 provisioned from the host.
495
496 In total, our data shows that there is much versatility in metabolism among the MAGs,
497 but some overlap in function. As *Daphnia magna* are indiscriminate filter feeders, they
498 may feed on a wide variety of particulate matter with variable nutrient profiles. The
499 versatility in metabolism encoded by these MAGs indicates that they are able to utilize
500 this unpredictable range of nutrients both in the digestive tract and on the carapace.
501 The specific functions of certain MAGs, particularly in amino acid and vitamin
502 biosynthesis and export, seem critical in providing nutritional benefits to the host
503 zooplankton.
504
505 ## Conclusions
506 *Daphnia magna* is an important model system for multiple facets of ecology, and has
507 recently become an organism of interest for understanding fundamental questions
508 about the microbiota. Our metagenomic sequencing and subsequent analysis
509 characterizes the *Daphnia magna* microbiota to the species level and finds some

genomic features that allow core bacterial species to acquire and biosynthesize nutrients, and to potentially interact with their host via amino acid and vitamin export. By examining this relatively simple microbiota via metagenome-assembled genomes, we can begin to investigate metabolic interactions between the host and its associated microbes. Future work to further elucidate functions of these MAGs will involve long-read metagenome sequencing to complete genome assemblies and pure, single-isolate sequencing to understand strain variation within the microbiota. Furthermore, transcriptomics and metabolomics could be used to understand which of these encoded genes are functioning under different environmental and host conditions, and will direct future hypotheses on host-microbiota interactions. For example, how much of the differences in *Daphnia* life history and population dynamics across food environments [49] can be attributed to differences in microbiota composition? These results will help to inform future work studying the effects of the microbiota on host health and population dynamics across ecological contexts. Moreover, as more populations of *Daphnia* and their microbiota are sequenced, it will become possible to examine the coevolutionary relationships between hosts and microbiota, and this functional information will be essential for making sense of those relationships.

Methods

529 *Sample collection and extraction*. Two samples of 100 21-day-old adult *Daphnia magna*
530 and two samples of 6-day-old juvenile Daphnia were collected from laboratory cultures
531 maintained in defined COMBO medium [47]. Laboratory cultures were fed a
532 standardized volume of green algae *Chlamydomonas reinhardtii* (CPCC 243) to provide
533 0.25 mg C/ml/day. Samples were immediately ground after collection using in sterile
534 1.5mL microcentrifuge tubes. To separate bacterial cells from host cells, a modified
535 protocol from Benson et al. 2014 was used [50]. Samples were suspended in 2mL PBE
536 buffer and layered on a cushion of 50% sucrose, then centrifuged at 4,000g for 10
537 minutes. DNA was extracted from the pellets at the base of the sucrose fractions
538 following the Qiagen DNEasy Blood & Tissue Kit spin-column protocol of total DNA
539 from animal tissues (Qiagen, Hilden, Germany).
540
541 *Library preparation and sequencing*. Shotgun sequencing libraries from the two adult
542 samples and two technical replicates of one juvenile sample were prepared and
543 multiplexed using the Illumina Nextera XT kit and protocol. Input DNA was quantified
544 using the Qubit dsDNA system. Libraries were checked using the Agilent Technology
545 2100 Bioanalyzer. Libraries were manually normalized due to a final library yield under
546 10nM. Paired-end sequencing was performed on an Illumina MiSeq using a MiSeq
547 Reagent Kit V2.

548

549 *Quality filtering and metagenome assembly.* Reads were demultiplexed using the built-
550 in Illumina MiSeq Reporter. Quality of demultiplexed reads was checked using FastQC
551 v0.11.5. Reads were trimmed using Trimmomatic v0.36 [51] with the commands:
552 ILLUMINACLIP:NexteraPE-PE.fa:2:30:10 LEADING:3 TRAILING:3
553 SLIDINGWINDOW:4:25 MINLEN:36 to remove adapter sequences and to remove
554 segments of reads where quality fell below 25. Trimmed reads were mapped to the
555 Daphnia magna draft genome [52] using BWA [53], Samtools [54], and BEDtools [55],
556 and reads with greater than 80% identity over 50% of the read were filtered. Remaining
557 paired unmapped reads were assembled de novo using metaSPAdes in SPAdes v3.11
558 [56]. A co-assembly of all four samples, a co-assembly of the adult samples, a co-
559 assembly of the juvenile samples, and individual assemblies for each sample were
560 created.

561

562 *Taxonomy-independent sequence binning and identification of individual*
563 *metagenome-assembled genomes.* To resolve genomes of identified organisms and to
564 discover genomes of organisms not present in read-based taxonomy identification
565 programs or entirely new organisms, contigs from the master co-assembly with reads
566 mapped from each sample were binned using CONCOCT within Anvi'o [57], accepting

567 contigs over 2500 bp in length. Contigs greater than 20kb in length were split into
568 20kb fragments prior to running CONCOCT. Bins from the master co-assembly were
569 assessed using Anvi'o, using Parks et al.'s quality score (genome completeness - 5x
570 estimated redundancy or contamination) of ≥50 as a cutoff for further refinement [58].
571 Bins meeting this quality cutoff were manually curated within Anvi'o, where contigs
572 within a bin that deviated dramatically from the mean GC content or mean coverage of
573 the bin were removed from the bin. Bins that increased completeness when merged,
574 did not increase redundancy above 10%, and were similar in GC content were merged.
575 Bins that did not have high (>90% completion) or medium quality (>50% completion)
576 after merging and refining were not analyzed further. After merging and refining, bins
577 that still met the quality score cutoff were assigned taxonomy using GTDB-Tk [59,60].
578 GTDB-Tk uses average nucleotide identity and genome topology to find the closest
579 genomic relative in its database. The same process was repeated for the juvenile co-
580 assembly and the adult co-assembly to confirm species presence and attempt to
581 resolve species identity at different host life stages. Similarity between MAGs was
582 calculated using the average nucleotide identity tool in Pyani [61].
583
584 *Read-based taxonomic classification.* Kaiju v1.5 [62], Kraken v1.0 [63], and MetaPhlAn2
585 v2.6 [64] were used to assign taxonomy to reads. We used all three identifiers due to

586 due to Kaiju's high rate of false identification [65] and MetaPhlAn2's use of specific
587 marker genes from reference organisms rather than entire genomes. All programs were
588 used with their built-in databases. Kraken results were confirmed via cross-comparison
589 of abundant species with both Kaiju and MetaPhlAn2. All three taxonomy profilers were
590 also used to assign taxonomy to contigs assembled from each sample. Visualization of
591 each sample's community composition was performed in R.
592
593 *Functional profiling.* Contigs from each metagenome-assembled genome and from the
594 adult, juvenile, and master co-assemblies were annotated using Prokka v1.12 [66].
595 Genes annotated in Prokka were assigned KOs (KEGG Orthologs) using the
596 GhostKOALA tool on the Kyoto Encyclopedia of Genes and Genomes [67]. KOs
597 identified from GhostKOALA were mapped to standard KEGG categories and
598 metabolic pathways using the KEGG Pathway Mapper & KEGG Module tools to
599 examine pathway completeness and identify pathways of interest [68]. Genes identified
600 using KEGG and GhostKOALA in each MAG were compared using OrthoVenn [69].
601 Overlapping and unique orthologs were compared using custom R scripts and with the
602 'UpSetR' package [70].
603

604 *16S rRNA gene sequencing and identification of contaminant taxa.* The V4
605 hypervariable region of the 16S rRNA gene was sequenced on the Illumina MiSeq using
606 a MiSeq Reagent Kit V2 and the same Qiagen DNEasy Blood & Tissue Kit and reagents
607 as in the shotgun sequencing sample processing. Four samples of five adult *Daphnia*
608 *magna* were sequenced to compare community composition found in 16S sequencing
609 to that found in shotgun sequencing, along with four samples of the COMBO media
610 *Daphnia magna* cultures are raised in, four samples of *Chlamydomonas reinhardtii*, and
611 two samples of the DNA sequencing kit and library preparation kit as negative controls.
612 Paired-end reads were analyzed in R using the 'dada2' package to trim primer
613 sequences, identify amplicon sequence variants, and assign taxonomy [71]. Taxonomy
614 was assigned using the RefSeq+RDP taxonomic training data set formatted for dada2
615 [72]. Further analysis of community composition and visualization were carried out using
616 the 'phyloseq' package in R [73].
617

618 **Declarations**

619 *Ethics approval and consent to participate*

620 Not applicable.

621 *Consent for publication*

622 Not applicable.

Availability of data and materials

The datasets supporting the conclusions of this article are available at the NCBI BioProject Portal under IDs PRJNA543317 (shotgun sequences) and PRJNA543842 (16S rRNA sequences. Raw metagenomic sequencing reads from each sample are deposited under accession numbers SAMN11660785 and SAMN11660786. 16S rRNA sequencing data can be found under accession numbers SAMN11784745 - SAMN11784837. All scripts for data analysis and visualization are available on GitHub (https://github.com/reillyowencooper/daphnia_magna_metagenome).

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by funds from the University of Nebraska and a Maude Hammond Fling Faculty Research Fellowship to CEC.

Authors' contributions

ROC and CEC designed the study. ROC collected and analyzed the metagenomic data. ROC wrote the first draft of the manuscript, and ROC and CEC revised the manuscript and approved its final form.

Acknowledgements

We are grateful for Dr. Andrew K. Benson, Mallory Van Haute, and Qinnan Yang's assistance with sequencing preparation and machine use, and to Brady Bathke for assistance with bioinformatic analyses. This work was completed utilizing the Holland Computing Center of the University of Nebraska, which receives support from the Nebraska Research Initiative.

References

1. Hacquard S, Garrido-Oter R, González A, Spaepen S, Ackermann G, Lebeis S, McHardy AC, Dangl JL, Knight R, Ley R, Schulze-Lefert P. Microbiota and Host Nutrition across Plant and Animal Kingdoms. Cell Host Microbe 2015; https://doi.org/10.1016/j.chom.2015.04.009
2. Wahlström A, Sayin SI, Marschall H, Bäckhed F. Intestinal Crosstalk between Bile Acids and Microbiota and Its Impact on Host Metabolism. Cell Metab. 2016; https://doi.org/10.1016/j.cmet.2016.05.005
3. Sansone CL, Cohen J, Yasunaga A, Xu J, Osborn G, Subramanian H, Gold B, Buchon N, Cherry S. Microbiota-Dependent Priming of Antiviral Intestinal Immunity in *Drosophila*. Cell Host Microbe 2015; https://doi.org/10.1016/j.chom.2015.10.010

4. Kaltenpoth M, Göttler W, Herzner G, Strohm E. Symbiotic Bacteria Protect Wasp Larvae from Fungal Infestation. Curr Biol. 2005;15:475-479.
5. Spor A, Koren O, Ley R. Unravelling the effects of the environment and host genotype on the gut microbiome. Nat. Rev. Microbiol. 2011;9:279-290.
6. Ebert, D. A Genome for the Environment. Science 2011;331:539-540.
7. Qi W, Nong G, Preston JF, Ben-Ami F, Ebert D. Comparative metagenomics of *Daphnia* symbionts. BMC Genom. 2009;10:21.
8. Eckert EM, Pernthaler J. Bacterial epibionts of *Daphnia*: a potential route for the transfer of dissolved organic carbon in freshwater food webs. ISME J. 2014;8(9):1808-1819.
9. Freese HM, Schink B. Composition and Stability of the Microbial Community inside the Digestive Tract of the Aquatic Crustacean *Daphnia magna*. Microb Ecol. 2011;62:882-894.
10. Sison-Mangus MP, Mushegian AA, Ebert D. Water fleas require microbiota for survival, growth and reproduction. ISME J. 2015;9:59-67.
11. Macke E, Callens M, Meester LD, Decaestecker E. Host-genotype dependent gut microbiota drives zooplankton tolerance to toxic cyanobacteria. Nat Commun. 2017;8:13.

12. Peerakietkhajorn S, Tsukado K, Kato Y, Matsuura T, Watanabe H. Symbiotic bacteria contribute to increasing the population size of a freshwater crustacean, *Daphnia magna*. Environ. Microbiol. Rep. 2015;7(2):364-372.

13. Motiei A, Brindefalk B, Ogonowski M, El-Shehawy R, Pastuszek P, Ek K, Liewenborg B, Udekwu K, Gorokhova E. Disparate effects of antibiotic-induced microbiome change and enhanced fitness in *Daphnia magna*. Preprint. https://doi.org/10.1101/586669

14. Callens M, Hajime W, Yasuhiko K, Miura J, Decaestecker E. Microbiota inoculum composition affects holobiont assembly and host growth in *Daphnia*. Microbiome 2018;6:56.

15. Peerakietkhajorn S, Kato Y, Kasalicky V, Matsuura T, Watanabe J. Betaproteobacteria *Limnohabitans* strains increase fecundity in the crustacean *Daphnia magna*: symbiotic relationship between major bacterioplankton and zooplankton in freshwater ecosystem. Environ Microbiol. 2016;18(8):2366-2374.

16. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U.S.A 2009;106(45):19126-19131.

17. Légeret B, Schulz-Raffelt M, Nguyen HM, Auroy P, Beisson F, Peltier G, Blanc G, Li-Beisson Y. Lipidomic and transcriptomic analyses of *Chlamydomonas*

reinhardtii under heat stress unveil a direct route for the conversion of membrane lipids into storage lipids. Plant Cell Environ. 2016;39:834-847.

18. Hernandez-Agreda A, Gates RD, Ainsworth TD. Defining the Core Microbiome in Corals' Microbial Soup. Trends Microbiol. 2017;25(2):125-140.

19. Hinz A, Lee S, Jacoby K, Manoil C. Membrane Proteases and Aminoglycoside Antibiotic Resistance. J Bacteriol. 2011;193(18):4790-4797.

20. Kasilicky V, Zeng Y, Piwosz K, Šimek K, Kratochvilová H, Koblížek. Aerobic Anoxygenic Photosynthesis Is Commonly Present within the Genus *Limnohabitans*. Environ Microbiol. 2018;84(1):13.

21. Mariadassou M, Pichon S, Ebert D. Microbial ecosystems are dominated by specialist taxa. Ecol Lett. 2015;18:974-982.

22. Sperisen P, Schmid CD, Bucher P, Zilian O. Stealth Proteins: In Silico Identification of a Novel Protein Family Rendering Bacterial Pathogens Invisible to Host Immune Defense. PLOS Comput Biol. 2005;1(6):0492-0499.

23. Mistou MY, Sutcliffe IC, van Sorge NM. Bacterial glycobiology: rhamnose-containing cell wall polysaccharides in Gram-positive bacteria. FEMS Microbiol Rev. 2016;40:464-479.

24. Fanning S, Hall LJ, Cronin M, Zomer A, MacSharry J, Goulding D, Motherway MO, Shanahan F, Nally K, Dougan G, van Sinderen D. Bifidobacterial surface-

715 exopolysaccharide facilitates commensal-host interaction through immune
716 modulation and pathogen protection. Proc Natl Acad Sci U.S.A
717 2012;109(6):2108-2113.
718 25. Fink P, Pfilitsch C, Marin K. Dietary Essential Amino Acids Affect the
719 Reproduction of the Keystone Herbivore *Daphnia pulex*. PLoS ONE
720 2011;6(12):8.
721 26. Helliwell KE, Wheeler GL, Leptos KC, Goldstein RE, Smith AG. Insights into the
722 Evolution of Vitamin B12 Auxotrophy from Sequenced Algal Genomes. Mol Biol
723 Evol. 2011;28(10):2921-2933.
724 27. Elendt BP, Bias WR. Trace nutrient deficiency in *Daphnia magna* cultured in
725 standard medium for toxicity testing. Effects of the optimization of culture
726 conditions on life history parameters of *D. magna*. Water Res. 1990;24(9):1157-
727 1167.
728 28. Tong Y, Guo M. Bacterial heme-transport proteins and their heme-coordination
729 modes. Arch Biochem Biophys. 2009;481(1):1-15.
730 29. Severi E, Hood DW, Thomas GH. Sialic acid utilization by bacterial pathogens.
731 Microbiology 2007;153:2817-2822.

30. Callens M, Macke E, Muylaert K, Bossier P, Lievens B, Waud M, Decaestecker E. Food availability affects the strength of mutualistic host-microbiota interactions in *Daphnia magna*. ISME J. 2016;10:911-920.

31. Ranjan R, Rani A, Metwally A, McGee HS, Perkins DL. Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. Biochem Biophysis. Res Commun. 2016;469(4):967-977.

32. Poehlein A, Daniel R, Simeonova DD. Genome sequence of *Pedobacter glucosidilyticus* DD6b, isolated from zooplankton *Daphnia magna*. Stand Genomic Sci. 2015;10:100.

33. Work VH, Radakovits R, Jinkerson RE, Meuser JE, Elliott LG, Vinyard DJ, Laurens LML, Dismukes GC, Posewitz MC. Increased Lipid Accumulation in the *Chlamydomonas reinhardtii sta7-10* Starchless Isoamylase Mutant and Increase Carbohydrate Synthesis in Complemented Strains. Eukaryot Cell 2010;9(8):1251-1261.

34. Taipale SJ, Galloway AWE, Aalto SL, Kahilainen KK, Strandberg U, Kankaala P. Terrestrial carbohydrates support freshwater zooplankton during phytoplankton deficiency. Sci Rep 2016;6:e30897.

35. Latasa C, Roux A, Toledo-Arana A, Ghigo JM, Gamazo C, Penadés JR, Lasa I. BapA, a large secreted protein required for biofilm formation and host

751 colonization of *Salmonella enterica* serovar Enteritidis. Mo Microbiol.
752 2005;58(5):1322-1339.
753 36. Rossez Y, Wolfson EB, Holmes A, Gally DL, Holden NJ. Bacterial Flagella: Twist
754 and Stick, or Dodge across the Kingdoms. PLoS Pathog. 2015;11(1):e1004483.
755 37. McLoughlin K, Schluter J, Rakoff-Nahoum S, Smith AL, Foster KR. Host Selection
756 of Microbiota via Differential Adhesion. Cell Host Microbe 2016;19:550-559.
757 38. Kline KA, Fälker S, Dahlberg S, Normark S, Henriques-Normark B. Bacterial
758 Adhesins in Host-Microbe Interactions. Cell Host Microbe 2009;5:580-592.
759 39. Foster TJ. Immune evasion by Staphylococci. Nat Rev Microbiol. 2005;3:948-
760 958.
761 40. Reveillaud J, Anderson R, Reves-Sohn S, Cavanaugh C, Huber JA. Metagenomic
762 investigation of vestimentiferan tubeworm endosymbionts from Mid-Cayman
763 Rise reveals new insights into metabolism and diversity. Microbiome 2018:6:19.
764 41. Hansen AK, Moran NA. Aphid genome expression reveals host-symbiont
765 cooperation in the production of amino acids. Proc Natl Acad Sci U.S.A
766 2011;108(7):2849-2854.
767 42. Pérez-Brocal V, Gil R, Ramos S, Lamelas A, Postigo M, Michelena JM, Silva FJ,
768 Moya A, Latorre A. A Small Microbial Genome: The End of a Long Symbiotic
769 Relationship? Science 2006;314:312-313.

770 43. Douglas AE. The microbial dimension in insect nutritional ecology. Funct Ecol.
771 2011;23:38-47.
772 44. Degnan PH, Taga ME, Goodman AL. Vitamin B12 as a Modulator of Gut
773 Microbial Ecology. Cell Metab. 2014;20:769-778.
774 45. Smith TA, Driscoll T, Gillespie JJ, Raghavan R. A *Coxiella*-Like Endosymbiont Is a
775 Potential Vitmain Source for the Lone Star Tick. Genome Biol Evol.
776 2015;7(3):831-838.
777 46. McKenney EA, Koelle K, Dunn RR, Yoder AD. The ecosystem services of animal
778 microbiomes. Mol Ecol. 2018;27:2164-2172.
779 47. Kilham SS, Kreeger DA, Lynn SG, Goulden CE, Herrera L. COMBO: a defined
780 freshwater culture medium for algae and zooplankton. Hydrobiologia
781 1998;377:147-159.
782 48. Castanys-Muñoz E, Martin MJ, Vazquez E. Building a Beneficial Microbiome
783 from Birth. Adv Nutr 2016;7:323-330.
784 49. Nelson WA, McCauley E, Wrona FJ. Multiple dynamics in a single predator-prey
785 system: experimental effects of food quality. Proc R Soc Lond B. 2001;268:1223-
786 1230.
787 50. Benson AK, David JRD, Gilbreth SE, Smith G, Nietfeldt J, Legge R, Kim J, Sinha
788 R, Duncan CE, Ma J, Singh I. Microbial successions are associated with changes

789 in chemical profiles of a model refrigerated fresh pork sausage during an 80-day
790 shelf-life study. Appl Environ Microbiol. 2014;80(17):5178-5194.
791 51. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina
792 sequence data. Bioinformatics 2014;30(15):2114-2120.
793 52. Daphnia Genomics. Genome Informatics Lab, Indiana University Biology
794 Department. http://wfleabase.org. Accessed 02 March 2018.
795 53. Li H. Aligning sequence reads, clone sequences and assembly contigs with
796 BWA-MEM. Preprint. https://arxiv.org/abs/1303.3997
797 54. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis
798 G, Durbin R, 1000 Genome Project Data Processing Subgroup. The Sequence
799 Alignment/Map format and SAMtools. Bioinformatics 2009;25(16):2078-2079.
800 55. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing
801 genomic features. Bioinformatics 2010;26(6):841-842.
802 56. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile
803 metagenomic assembler. Genome Res. 2017;27:824-834.
804 57. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO.
805 Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ
806 2015;3:e1319.

807 58. Parks DH, Rinke C, Chuvochina M, Chaumeil PA, Woodcroft BJ, Evans PN,
808 Hugenholtz P, Tyson GW. Recovery of nearly 8,000 metagenome-assembled
809 genomes substantially expands the tree of life. Nat Microbiol. 2017;2:1533-
810 1542.
811 59. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski, Chaumeil PA,
812 Hugenholtz P. A standardized bacterial taxonomy based on genome phylogeny
813 substantially revises the tree of life. Nat Biotechnol. 2018;36(10):996-1004.
814 60. Chaumeil PA, Hugenholtz P, Parks DH. GTDB-Tk: A toolkit to classify genomes
815 with the Genome Taxonomy Database.
816 https://github.com/Ecogenomics/GTDBTk
817 61. Pritchard L, Glover RH, Humphris S, Elphinstone GJ, Toth IK. Genomics and
818 taxonomy in diagnostics for food security: soft-rotting enterobacterial plant
819 pathogens. Anal Methods 2016;8:12-24.
820 62. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for
821 metagenomics with Kaiju. Nat Commun. 2016;7:11257.
822 63. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification
823 using exact alignments. Genome Biol. 2014;15:R46.

824 64. Segata N, Waldron L, Ballarini A, Narsimhan V, Jousson O, Huttenhower C.
825 Metagenomic microbial community profiling using unique clade-specific marker
826 genes. Nat Methods 2012;9(8):811-814.
827 65. Piro VC, Matschkowski M, Renard BY. MetaMeta: integrating metagenome
828 analysis tools to improve taxonomic profiling. Microbiome 2017;5:101.
829 66. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics
830 2014;30(14):2068-2069.
831 67. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA for Functional
832 Characterization of Genome and Metagenome Sequences. J Mol Biol.
833 2016;428:726-731.
834 68. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a
835 reference resource for gene and protein annotation. Nucleic Acids Res.
836 2016;44:D457-D462.
837 69. Wang Y, Coleman-Derr D, Chen G, Gu YQ. OrthoVenn: a web server for
838 genome wide comparison and annotation of orthologous clusters across
839 multiple species. Nucleic Acids Res. 2015;43:W78-W84.
840 70. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of
841 intersecting sets and their properties. Bioinformatics 2017;33(18):2938-2940.

842 71. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP.
843 DADA2: High-resolution sample inference from Illumina amplicon data. Nat
844 Methods 2016;13(7):581-583.
845 72. Alishum A. DADA2 formatted 16S rRNA gene sequences for both bacteria &
846 archaea. Version 1 Refseq+RDP. 2019; http://doi.org/10.5281/zenodo.2541239
847 73. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive
848 Analysis and Graphics of Microbiome Census Data. PLoS ONE
849 2013;8(4):e61217.

850

851 Table 3.

KEGG ID	Module	Pathway Name	MAG
M00307	Central carbohydrate metabolism	Pyruvate oxidation, pyruvate => acetyl-CoA	All
M00009	Central carbohydrate metabolism	Citrate cycle (TCA cycle, Krebs cycle)	All
M00010	Central carbohydrate metabolism	First carbon oxidation, oxaloacetate => 2-oxoglutarate	All
M00011	Central carbohydrate metabolism	Citrate cycle, second carbon oxidation, 2-oxoglutarate => oxaloacetate	All

M00005	Central carbohydrate metabolism	PRPP biosynthesis, ribose 5P => PRPP	All
M00120	Cofactor and vitamin metabolism	Coenzyme A biosynthesis, pantothenate => CoA	All
M00793	Polyketide sugar unit biosynthesis	dTDP-L-rhamnose biosynthesis	All
M00250	ABC-2 type and other transport systems	Lipopolysaccharide transport system	All
M00255	ABC-2 type and other transport systems	Lipoprotein-releasing system	All
M00336	Bacterial secretion system	Twin-arginine translocation (Tat) system	All
M00434	Two-component regulatory system	PhoR-PhoB (phosphate starvation response) two-component regulatory system	All
M00045	Histidine metabolism	Histidine degradation, histidine => N-formiminoglutamate => glutamate	Limnohabitans sp. 1
M00435	Mineral and organic ion transport system	Taurine transport system	Limnohabitans sp. 1
M00193	Mineral and organic ion transport system	Putative spermidine/putrescine transport system	Limnohabitans sp. 1
M00605	Saccharide, polyol, and lipid transport system	Glucose/mannose transport system	Limnohabitans sp. 1

M00607	Saccharide, polyol, and lipid transport system	Glycerol transport system	Limnohabitans sp. 1
M00008	Central carbohydrate metabolism	Entner-Doudoroff pathway, glucose-6P => glyceraldehyde-3P + pyruvate	Limnohabitans sp. 2
M00192	Mineral and organic ion transport system	Putative thiamine transport system	Limnohabitans sp. 2
M00259	ABC-2 type and other transport systems	Heme transport system	Limnohabitans sp. 2
M00001	Central carbohydrate metabolism	Glycolysis (Embden-Meyerhof pathway), glucose => pyruvate	Pedobacter sp.
M00854	Oher carbohydrate metabolism	Glycogen biosynthesis, glucose-1P => glycogen/starch	Pedobacter sp.
M00026	Histidine metabolism	Histidine biosynthesis, PRPP => histidine	Pedobacter sp.
M00096	Terpenoid backbone biosynthesis	C5 isoprenoid biosynthesis, non-mevalonate pathway	Pedobacter sp.
M00364	Terpenoid backbone biosynthesis	C10-C20 isoprenoid biosynthesis, bacteria	Pedobacter sp.
M00256	ABC-2 type and other transport systems	Cell division transport system	Pedobacter sp.
M00707	Drug efflux transporter/pump	Multidrug resistance, MdlAB/SmdAB transporter	Pedobacter sp.

M00595	Sulfur metabolism	Thiosulfate oxidation by SOX complex, thiosulfate => sulfate	Polaromonas sp.
M00087	Fatty acid biosynthesis and degradation	beta-Oxidation	Polaromonas sp.
M00020	Serine and threonine metabolism	Serine biosynthesis, glycerate-3P => serine	Polaromonas sp.
M00125	Cofactor and vitamin metabolism	Riboflavin biosynthesis, GTP => riboflavin/FMN/FAD	Polaromonas sp.
M00122	Cofactor and vitamin metabolism	Cobalamin biosynthesis, cobinamide => cobalamin	Polaromonas sp.
M00436	Mineral and organic ion transport system	Sulfonate transport system	Polaromonas sp.
M00669	Saccharide, polyol, and lipid transport system	gamma-Hexachlorocyclohexane transport system	Polaromonas sp.
M00320	ABC-2 type and other transport systems	Lipopolysaccharide export system	Polaromonas sp.
M00647	Drug efflux transporter/pump	Multidrug resistance, efflux pump AcrAB-TolC/SmeDEF	Polaromonas sp.
M00339	Bacterial secretion system	RaxAB-RaxC type I secretion system	Polaromonas sp.
M00454	Two-component regulatory system	KdpD-KdpE (potassium transport) two-component regulatory system	Polaromonas sp.

M00493	Two-component regulatory system	AlgZ-AlgR (alginate production) two-component regulatory system	Polaromonas sp.
M00501	Two-component regulatory system	PilS-PilR (type 4 fimbriae synthesis) two-component regulatory system	Polaromonas sp.
M00632	Oher carbohydrate metabolism	Galactose degradation, Leloir pathway, galactose => alpha-D-glucose-1P	Unknown Burkholderiaceae
M00549	Oher carbohydrate metabolism	Nucleotide sugar biosynthesis, glucose => UDP-glucose	Unknown Burkholderiaceae
M00554	Oher carbohydrate metabolism	Nucleotide sugar biosynthesis, galactose => UDP-galactose	Unknown Burkholderiaceae
M00579	Carbon fixation	Phosphate acetyltransferase-acetate kinase pathway, acetyl-CoA => acetate	Unknown Burkholderiaceae
M00183	RNA polymerase	RNA polymerase, bacteria	Unknown Burkholderiaceae
M00300	Mineral and organic ion transport system	Putrescine transport system	Unknown Burkholderiaceae
M00209	Mineral and organic ion transport system	Osmoprotectant transport system	Unknown Burkholderiaceae
M00214	Saccharide, polyol, and lipid transport system	Methyl-galactoside transport system	Unknown Burkholderiaceae
M00204	Saccharide, polyol, and lipid transport system	Trehalose/maltose transport system	Unknown Burkholderiaceae

M00325	Bacterial secretion system	alpha-Hemolysin/cyclolysin transport system	Unknown Burkholderiaceae
M00333	Bacterial secretion system	Type IV secretion system	Unknown Burkholderiaceae
M00449	Two-component regulatory system	CreC-CreB (phosphate regulation) two-component regulatory system	Unknown Burkholderiaceae
M00450	Two-component regulatory system	BaeS-BaeR (envelope stress response) two-component regulatory system	Unknown Burkholderiaceae
M00471	Two-component regulatory system	NarX-NarL (nitrate respiration) two-component regulatory system	Unknown Burkholderiaceae

852