Acetobacteraceae in the honey bee gut comprise two distant clades with diverging metabolism and ecological niches

Bonilla-Rosso G[1], Paredes Juan C[2], Das S[1], Ellegaard KM[1], Emery O[1], Garcia-Garcera M[1], Glover N[3,4], Hadadi N[1], van der Meer JR[1], SAGE class 2017-18[5], Tagini F[6], Engel P[1]*

[1]Department of Fundamental Microbiology, University of Lausanne, 1015 Lausanne, Switzerland; [2]International Centre of Insect Physiology and Ecology (ICIPE), Kasarani, Nairobi, Kenya**;** [3]Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland; [4]Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland; [5] Master of Science in Molecular Life Sciences, Faculty of Biology and Medicine, University of Lausanne, Switzerland; [6]Institute of Microbiology, University Hospital Center and University of Lausanne, Lausanne, Switzerland.

***Author for Correspondence:**

Prof. Philipp Engel, Department of Fundamental Microbiology, University of Lausanne, CH-1015 Lausanne, Switzerland, Tel.: +41 (0)21 692 56 12, e-mail: philipp.engel@unil.ch

22 Key words: *Apis mellifera*, queen, honey bee, Alpha2.1, Alpha2.2, *Commensalibacter*,

23 *Bombella*, *Parasaccharibacter*, oxidative fermentation, respiration, genome

24 reduction

Abstract

Various bacteria of the family Acetobacteraceae are associated with the gut environment of insects. Honey bees harbor two distinct Acetobacteraceae in their gut, Alpha2.1 and Alpha2.2. While Alpha2.1 seems to be a gut specialist, Alpha2.2 is also found in the diet (e.g. royal jelly), the hypopharyngeal glands, and the larvae of honey bees. Here, we combined amplicon and genome sequencing to better understand functional differences associated with the ecology of Alpha2.1 and Alpha2.2. We find that the two phylotypes are differentially distributed along the worker and queen bee gut. Phylogenetic analysis shows that Alpha2.2 is nested within the acetic acid bacteria and consists of two separate sub-lineages, whereas Alpha2.1 belongs to a basal lineage with an unusual GC content for Acetobacteraceae. Gene content analysis revealed major differences in the central carbon and respiratory metabolism between the two phylotypes. While Alpha2.2 encodes two periplasmic dehydrogenases to carry out oxidative fermentation, Alpha2.1 lacks this capability, but instead harbors a diverse set of cytoplasmic dehydrogenases. These differences are accompanied by the loss of the TCA cycle in Alpha2.2, but not in Alpha2.1. We speculate that Alpha2.2 has specialized for fast-resource utilization through incomplete carbohydrate oxidation, giving it an advantage in sugar-rich environments such as royal jelly. On the contrary, the broader metabolic range of Alpha2.1 may provide an advantage in the worker bee hindgut, where competition with other bacteria and flexibility in resource utilization may be relevant for persistence. Our results show that bacteria belonging to the same family may utilize vastly different strategies to colonize niches associated with the animal gut.

Introduction

According to a recent review on the taxonomy of Alphaproteobacteria (Munoz-Gomez, et al. 2019) and the standardized genome phylogeny-based taxonomy of Parks et al. (Parks, et al. 2018), the family Acetobacteraceae (Rhodospirillales) is comprised of an externally branching acidophilic/neutrophilic group and an internal acetous group. The latter group includes acetic acid bacteria (AABs), which constitute the vast majority of the described taxa of the Acetobacteraceae (Komagata, et al. 2014).

AAB inhabit sugar-rich environments and use a rather exceptional strategy to gain energy. They oxidize sugars or sugar alcohols on the periplasmic side of the cell envelop with the help of membrane-bound dehydrogenases that are linked to the respiratory chain in a process known as oxidative fermentation (Matsushita and Matsutani 2016). This particular oxidative metabolism results in the accumulation of fermentation products (such as acetic acid) in the environment. AABs naturally occur in association with plants, flowers, and fruits (Bartowsky and Henschke 2008; Pedraza 2016; Yamada and Yukphan 2008). They also play key roles in food and beverage fermentations (De Roos and De Vuyst 2018). In addition, AABs are being increasingly described to be associated with different insect species that rely on sugar-based diets, such as fruit flies, mosquitoes, sugarcane leafhoppers, mealybugs, honey bees, and bumble bees (Chouaia, et al. 2014; Crotti, et al. 2010).

Honey bees feed on a highly sugar-rich diet composed of nectar, honey, and pollen (Brodschneider and Crailsheim 2010). While their gut microbiota is relatively simple,

most of the commonly found phylotypes are specialized to live in the bee gut environment (Bonilla-Rosso and Engel 2018; Kwong and Moran 2016). These phylotypes are consistently present in adult worker bees worldwide and belong to deep-branching phylogenetic lineages that have so far only been detected in different species of honey bees, bumble bees, and stingless bees (Kwong, et al. 2017). Two of these phylotypes, originally referred to as Alpha2.1 and Alpha2.2, belong to the Acetobacteraceae (Cox-Foster, et al. 2007; Martinson, et al. 2011). Both are frequently detected in adult worker bees. However, in contrast to the so-called core phylotypes of the honey bee gut microbiota (*Gilliamella*, *Snodgrassella*, *Lactobacillus* Firm4, *Lactobacillus* Firm5, and *Bifidobacterium*), Alpha2.1 and Alpha2.2 are not present in every individual worker bee and often their relative abundance is low (Kapheim, et al. 2015; Kešnerová, et al. 2019; Kwong, et al. 2017; Martinson, et al. 2011; Martinson, et al. 2012; Moran, et al. 2012; Powell, et al. 2018). Interestingly, they both belong to the few phylotypes that have repeatedly been found in the gut of the honey bee queen, which is the only reproductive female in a honey bee colony (Anderson, et al. 2018; Kapheim, et al. 2015; Powell, et al. 2018).

A recent study showed that Alpha2.2 was predominantly found in the mouth, midgut, and ileum of queens, while Alpha2.1 was more abundant in the rectum (Anderson, et al. 2018). Moreover, Alpha2.2 has also been detected in royal jelly, floral nectar, bee bread (i.e. pollen stores of honey bees), and in bee larvae, indicating that this phylotype can live in more diverse environments than Alpha2.1 (Anderson, et al. 2014; Corby-Harris, et al. 2014; Maes, et al. 2016; Vojvodic, et al. 2013) .

93 Several strains of Alpha2.2 have been isolated, one of which (strain MRM) was
94 described as a novel species, *Bombella apis* (Yun, et al. 2017). Other isolates of
95 Alpha2.2 have been referred to as *Candidatus* Parasaccharibacter apium (Corby-
96 Harris, et al. 2016; Corby-Harris, et al. 2014), but were never formally described as a
97 novel species. Some strains of Alpha2.2 have been suggested to increase larval
98 survival under experimental conditions indicating a possible beneficial role of this
99 phylotype (Corby-Harris, et al. 2014).
100 Genome analysis of a closely related species, *Bombella intestini*, isolated from the gut
101 of a bumble bee, revealed typical features of other Acetobacteraceae, including the
102 presence of genes for carrying out oxidative fermentation (Li, et al. 2016; Li, et al.
103 2015). Draft genomes of several strains of Alpha2.2 have been deposited in public
104 databases (Corby-Harris and Anderson 2018), and a comparison with a closely
105 related flower-associated strain, *Saccharibacter floricola,* revealed a number of
106 potentially adaptive changes (Smith and Newton 2018). However, an in-depth
107 functional analysis of the gene content and overall metabolic capabilities of Alpha2.2
108 relative to Alpha2.1 has not been carried out to date.
109 Compared to Alpha2.2, much less is known about the phylotype Alpha2.1.
110 Phylogenetic trees based on 16S rRNA gene sequences suggest that Alpha2.1 belongs
111 to a deep-branching lineage within the Acetobacteraceae (Martinson, et al. 2011). The
112 two most closely related species belong to the candidate genus *Commensalibacter*
113 (*Commensalibacter* sp. MX-Monarch01 and *Commensalibacter intestini* A911),
114 isolated from the gut of a butterfly and a fruit fly, respectively (Roh, et al. 2008;
115 Servin-Garciduenas, et al. 2014). Therefore, the phylotype Alpha2.1 is frequently

116 referred to as *Commensalibacter* sp. While genomes have been published for strains
117 of all three *Commensalibacter* species (Kim, et al. 2012; Servin-Garciduenas, et al.
118 2014; Siozios, et al. 2019), little is known about their gene content, metabolic
119 capabilities, and phylogenetic positioning in respect to other Acetobacteraceae.
120
121 The presence of two phylogenetically related phylotypes from a family known to be
122 optimized for fast growth in carbohydrate-rich environments prompted us to study
123 their metabolic niches in the honey bee gut environment. We carried out 16S rRNA
124 gene-based community analysis, sequenced 12 isolates of Alpha2.1 and Alpha2.2, and
125 carried out comparative genomic analysis including previously sequenced strains.
126 We find that the two phylotypes are differentially distributed along the worker and
127 queen bee gut, and confirm previous studies that show that they belong to distinct
128 phylogenetic lineages within the Acetobacteraceae. The comparative genome
129 analysis suggests that the two phylotypes have different strategies to metabolize
130 carbon sources and to harvest energy. Moreover, we identified several characteristics
131 of Alpha2.1 that are unique among Acetobacteraceae, which is in agreement with its
132 position on a deep-branching lineage within the Acetobacteraceae.

Materials and Methods

Bee sampling

Four queen bees and four worker bees of the European honey bee (*Apis mellifera*), were sampled in Summer 2017. Three of each came from different colonies at the University of Lausanne and one from a professional beekeeper in Western Switzerland (Imkerei Giger). The guts were dissected into gut compartments (honey crop, midgut, ileum, and rectum), and homogenized as described in (Ellegaard and Engel 2019). Each sample was split in two, one of which was used for amplicon sequencing and the other for bacterial isolation. For the samples used for bacterial isolation, the four different gut regions were pooled together prior to plating them on different media.

Amplicon sequencing

DNA from the different gut samples was isolated using an established CTAB/phenol extraction protocol (Kesnerova, et al. 2017). The 16S rRNA gene was amplified with primers 27F and 907R prior to sending the samples for amplicon sequencing analysis at Microsynth (Switzerland). At Microsynth, the V4 region of 16SrRNA gene was amplified using universal primers 515F/806R (Caporaso, et al. 2011), and the amplified fragments were purified and sequenced with the Illumina MiSeq platform (2x250bp). The number of 16S rRNA gene copies per host actin copy was quantified through qPCR with universal bacterial primers for honey bee gut (F 5'AGGATTAGATACCCTGGTAGTCC-3', R 5'-YCGTACTCCCCAGGCGG-3') following the method described by Kešnerová et al. (2017).

156

157 **Sequence processing and community analysis**

158 Raw reads were processed and reads with more than 75% of bases below a quality
159 score of 33 were filtered with FastX-Toolkit (Gordon & Hannon 2010, unpublished,
160 http://hannonlab. cshl. edu/fastx_toolkit). Remaining paired reads were merged
161 with PEAR (Zhang, et al. 2014). Paired reads were quality-filtered, dereplicated,
162 clustered into OTUs at 97% identity and chimera-filtered with VSEARCH (Rognes, et
163 al. 2016). OTU abundance was calculated by mapping the total quality-filtered paired
164 reads to the final clusters using VSEARCH --*usearch_global*. The resulting abundances
165 were normalized by the total 16S rRNA gene copy numbers as estimated through
166 qPCR.
167 Representative OTU sequences were assigned to taxonomic categories with SINA
168 (v.1.2.11) against SILVA_132_NR99 (Pruesse, et al. 2012), and composition was
169 visualized with PHINCH v.1 (Bik 2014). Differences in copy numbers across samples
170 were evaluated with an aligned Rank Transformation of a Factorial Model with the R
171 package ARTool (Kay and Wobbrock 2016). Comparative abundances of Alpha2.1 and
172 Alpha2.2 were expressed as log-ratios of the normalized copy numbers for all OTUs
173 assigned to Alpha2.1 (*Commensalibacter*) and Alpha2.2 (*Bombella*).

174

175 **Bacterial culturing, DNA isolation, and genome sequencing**

176 Serial dilutions of the gut homogenates were plated on Sabouraud Dextrose Agar
177 (SDA) or MRS + Mannitol agar and incubated at 35°C in 5% CO_2 incubator. After 3-5
178 days of incubation, single colonies were picked, restreaked on fresh agar, and

179 incubated for another 2-3 days. Isolates were genotyped with universal bacterial
180 primers and *rpoB*-specific primers (0937: GAAATTTATGCCGAGGCTGG; 0938:
181 GAAATTTATGCCGAGGCTGG) as described in Ellegaard et al (2019), and stocked in
182 MRS broth containing 25% glycerol at -80°C. A total of eight strains of Alpha2.1 and
183 Alpha2.2 were selected for genome sequencing. Four additional strains were selected
184 from a previous culturing effort using a similar culturing approach (see **Table S1**).
185
186 **Genome sequencing, assembly, and annotation.**
187 For Illumina genome sequencing, genomic DNA was isolated from fresh bacterial
188 cultures using the GenElute Bacterial Genomic DNA Kit (SIGMA) according to
189 manufacturer's instructions. Genome sequencing libraries were prepared with the
190 TruSeq DNA kit and sequenced on the MiSeq platform (Illumina) using the paired-
191 end 2x250-bp protocol at the Genomic Technology facility (GTF) of the University of
192 Lausanne. The genome sequence analysis was carried out as described in Ellegaard
193 et al. (2019). In short, the resulting sequence reads were quality-trimmed with
194 trimmomatic v0.33 (Bolger, et al. 2014) and assembled with SPAdes v.3.7.1
195 (Bankevich, et al. 2012). Small contigs (less than 500 bp) and contigs with low kmer
196 coverage (less than 5) were removed from the assemblies, resulting in 5-22 contigs
197 per assembly.
198 Two strains, ESL0284 and ESL0368, one from Alpha2.1 and Alpha2.2 each, were
199 selected for sequencing with PacBio 20K (Pacific Biosciences) single-molecule real-
200 time (SMRT) technology. High-molecular weight genomic DNA was extracted with the
201 previously established CTAB/phenol extraction protocol (Kesnerova, et al. 2017). De

202 novo genome assembly was done using the Hierarchical Genome Assembly Process
203 (HGAP) version 2.3 (Chin, et al. 2013). These two completely assembled genomes
204 served as references to order the contigs of the Illumina assemblies using MAUVE
205 v2.4 (Rissman, et al. 2009). The origin of replication was determined based on the GC-
206 skew and set to position 1 by cutting the contig at the corresponding location. The
207 same was done for the published complete genome of Alpah2.1 strain AMU001
208 (Siozios, et al. 2019), but not for any of the other published draft genomes that were
209 included in the analysis of the gene content (Corby-Harris and Anderson 2018; Smith
210 and Newton 2018). Assembly quality was checked by remapping reads to assemblies
211 with the Burrows-Wheeler Aligner (Durbin 2014). The genomes were annotated
212 using the 'Integrated Microbial Genomes and Microbiomes' (IMG/mer) system (Chen,
213 et al. 2017).
214
215 **Inference of gene families and genome-wide phylogeny.**
216 For the gene content analysis and the inference of the genome-wide phylogeny, we
217 determined gene families, i.e. sets of homologous genes, across 56 genomes from the
218 family Acetobacteraceae and other Alphaproteobactera using OrthoMCL (Li, et al.
219 2003). We included the 12 newly sequenced strains of Alpha2.1 and Alpha2.2, and 44
220 previously sequenced strains (including 9 previously sequenced Alpha2.1 and
221 Alpha2.2 strains and 35 genomes of other Acetobacteraceae and closely related
222 Alphaproteobacteria). Protein sequences of all CDS of the 56 genomes were searched
223 against each other using BLASTP. BLASTP hits with an e-value $<10^{-5}$ and a relative
224 alignment length of >50% of the length of the query and the hit CDS were retained.

225 The OrthoMCL analysis was carried out as recommended with the mcl program run
226 with the parameters '--abc -I 1.5'. A total of 11,451 gene families were identified. The
227 remaining CDS (14,453 CDS) were singletons, i.e. they had no detectable homolog in
228 any other genome in our dataset.
229 A total of 361 single copy orthologs were extracted from the OrthoMCL output (i.e.
230 gene families having exactly one representative in every genome in the analysis). The
231 protein sequences of each of these gene families were aligned with mafft (Katoh and
232 Standley 2013) and alignment columns represented by <50% of all sequences
233 removed. The single gene family alignments were concatenated and used as the basis
234 for inferring a core genome phylogeny using RAxML (Stamatakis 2014) with the
235 PROTCATWAG model and 100 bootstrap replicates.
236
237 **Comparison of genome structure and divergence.**
238 The R package genoplotR was used to compare and visualize whole genome
239 alignments (Guy, et al. 2010). Pairwise BlastN comparison files were generated using
240 command line blast v2.2.31+ using a bit score cutoff of 100. To estimate sequence
241 divergence between genomes, we calculated pairwise ANI with orthoani (Lee, et al.
242 2016) using the exectutable "OAT_cmd.jar" with the parameter "-method ani.".
243
244 **Comparison of the functional gene content.**
245 To analyze differences in gene content between Alpha2.1 and Alpha2.2, we extracted
246 all OrthoMCL gene families that contained a homolog in at least one of the 21 genomes
247 of the two phylotypes of interest and included also singletons, i.e. gene families that

had homologs in only one of the genomes. This resulted in a total of 3,275 gene families identified across 21 genomes of Alpha2.1 and Alpha2.2. These gene families were categorized into 'shared' and 'specific' core and pan genome subsets depending on their presence/absence in the genomes of Alpha2.1 and the two sub-lineages of Alpha2.2. To analyze functional differences between different subsets of core gene families, we assigned the IMG/mer annotation of one of the homologs (if possible the one of the complete reference genomes) to each gene family. These annotations were used to determine the distribution of gene families into COG categories and to identify the shared and phylotype-specific metabolic capabilities, biosynthetic pathways, and transport functions based on the analysis of KEGG pathways. For the analysis of the respiratory chain, we combined several approaches: The characterization of the electron acceptors was mainly based on KEGG annotations and the previous publication of the genome of *Bombella intestini* (Li, et al. 2016). To identify respiratory dehydrogenases we carried out keyword searches with 'dehydrogenase' and 'reductase', and identified all gene families that belonged to an enzyme class (EC number) that has been described in the literature as being a respiratory enzyme (Marreiros, et al. 2016). The TCA cycle analysis was based on KEGG annotations.

For inferring phylogenetic trees of the nitrate reductase subunit alpha (*narG*) and the nitric oxide reductase, the amino acid sequence of one of the homologs was searched against the nr database using BLASTP. From the top blast hits, a subset of homologs identified in divergent strains and species was selected to build a phylogeny using RAXML. The most similar homologs present among Acetobacteraceae were identified

by carrying out a second BLASTP search against nr by restricting hits to the family Acetobacteraceae (taxid:433).

Results

Community analysis of the gut microbiota of worker bees and queens suggests different niches of Alpha2.1 and Alpha2.2.

Bacteria of the Acetobacteraceae Alpha2.1 and Alpha2.2 have been reported to be dominant community members of the gut microbiota of adult honey bee queens (Anderson, et al. 2018; Corby-Harris, et al. 2014; Kapheim, et al. 2015; Tarpy, et al. 2015), and can also be abundant in the crop of worker bees (Corby-Harris, et al. 2014). To corroborate these previous results and obtain additional insights about the distribution and relative contribution of these phylotypes to the gut microbiota, we determined total bacterial loads and community composition in four different gut compartments (crop, midgut, ileum, and rectum) of honey bee queens and worker bees.

qPCR analyses revealed significant differences in total 16S rRNA gene copy number across castes (i.e. queens and workers) and gut regions (F_{caste}=15.7, F_{gut}=13.9, p<0.05, **Figure 1A**), with the largest difference being that queens had significantly smaller bacterial loads than workers. The largest bacterial loads were found in the rectum, and the smallest in the crops, as reported in previous studies (Martinson, et al. 2012; Powell, et al. 2018). The only gut region where the queens had significantly larger loads than the workers was the crop (**Figure 1A**).

The community composition differed markedly between castes and gut regions. Queen guts were characterized by the dominance of family Acetobacteraceae (i.e. Alpha2.1 and Alpha2.2) across all compartments and a low abundance of

295 *Snodgrassella* (*Neisseriaceae*) and *Gilliamella* (*Orbaceae*), corroborating previous
296 studies (**Figure 1C**) (Anderson, et al. 2018; Corby-Harris, et al. 2014; Kapheim, et al.
297 2015; Tarpy, et al. 2015). The family Acetobacteraceae was also present in all gut
298 compartments of worker bees, but was proportionally much more abundant in the
299 crop and midgut compared to midgut and rectum. Taking the bacterial loads per
300 compartment into consideration, the family Acetobacteraceae was most abundant,
301 and contributed the most to total community composition, in the rectum of queens
302 (**Figure 1A and 1C**).
303 None of the two Alpha2 phylotypes were exclusively associated with queen guts, but
304 instead were found in all gut compartments from both castes. The average 16S rRNA
305 gene copies per actin copy of both phylotypes were consistently higher in workers
306 than in queens, and overall Alpha2.1 was higher than Alpha2.2. Nevertheless, we
307 found differences in the ratios between Alpha2.2 to Alpha2.1 within gut
308 compartments: Alpha2.1 was more abundant in the rectum, and Alpha 2.2 was more
309 abundant in midguts of both queens and workers (**Figure 1B**). The ratios in the ileum
310 differed between castes, with Alpha2.2 being more abundant in the queens, and
311 Alpha2.1 being more abundant in the workers. As expected between individual bees,
312 we observed large variability across the four replicates, particularly for queen's
313 crops, but the pattern observed in queens is also in line with that reported in
314 Anderson et al. (2018).
315 In summary, both phylotypes are found across all gut compartments of queens and
316 worker bees, but while Alpha2.2 seems to be more abundant and contributes

317 proportionally more to the community in the queen crop and midgut, Alpha2.1 is
318 much more abundant in the rectum of both castes.
319
320 **Alpha2.1 and Alpha2.2 belong to two distinct phylogenetic clades within the**
321 **Acetobacteraceae.**
322 To facilitate functional and phylogenetic analysis, we sequenced the genomes of
323 seven Alpha2.1 and five Alpha2.2 strains (**Table S1**). Eight of the 12 strains were
324 isolated from the same worker and queen bees that were analyzed in the previous
325 section. The other four strains were isolated from other worker and queen bees. One
326 strain of each phylotype was assembled into a single circular chromosome (strain
327 ESL0284 for Alpha2.1, and strain ESL0368 for Alpha2.2) and served as reference for
328 our analysis. The assemblies of the other genomes consisted of 5-22 contigs. Genome
329 size varied little among strains and was comparable between the two phylotypes
330 ranging from 1.85-2.07 Mb. The overall genome structure was largely conserved in
331 both phylotypes as based on whole genome alignments with the complete genome of
332 the reference strain of each phylotype (**Figure S1**).
333
334 To better understand the phylogenetic relationship between Alpha2.1, Alpha2.2, and
335 other Acetobacteraceae, we inferred a genome-wide tree of the 12 newly sequenced
336 Alpha2.1 and Alpha2.2 strains, nine previously sequenced Alpha2.1 and Alpha2.2
337 strains (Corby-Harris and Anderson 2018; Siozios, et al. 2019), and 37 strains of other
338 Alphaproteobacteria (mainly Acetobacteraceae). This tree revealed that Alpha2.1 and

339 Alpha2.2 are polyphyletic, i.e. belong two distinct clades within the Acetobacteraceae
340 (**Figure 2**).
341
342 Alpha2.1 formed, together with the two previously identified strains of the genus
343 *Commensalibacter,* a deep-branching monophyletic lineage, which was basal to most
344 AABs within the Acetobacteraceae. However, a relatively long branch separated
345 Alpha2.1 from the two *Commensalibacter* strains (**Figure 2**) and the average
346 nucleotide identity (ANI) between their genomes was very low (~70%, see **Table S2**
347 **and Figure S2**) indicating deep divergence. In contrast, the analyzed strains of
348 Alpha2.1 were closely related to each other, as evident from the short branches
349 separating the different strains in the phylogenetic tree (**Figure 2**) and the high ANI
350 between their genomes (98-99%, see **Table S2 and Figure S2**). This agrees with a
351 recent metagenomic study, which found that Alpha2.1 belongs to the species with the
352 lowest extent of strain-level diversity within and between individual honey bees
353 (Ellegaard and Engel 2019). Our results suggest that Alpha2.1 presents a novel
354 species of the genus *Commensalibacter*. Notably, the genomes of Alpha2.1 and the two
355 strains of *Commensalibacter* had a very low GC content (~37%) relative to all other
356 sequenced Acetobacteraceae, which is usually between 50-60% (**Table S1**),
357 providing further evidence for their distinctive position within this family.
358
359 In contrast to the basal position of Alpha2.1 in the acetous group of the
360 Acetobacteraceae, our phylogeny revealed that Alpha2.2 is nested within the
361 *Gluconobacter* clade and is most closely related to *S. floricola* DSM 15669. The

362 sequenced Alpha2.2 strains clustered into two sub-lineages (**Figure 2**) that were
363 separated by a well-supported branch leading to *Bombella intestini* R-52487, a
364 species previously isolated from the gut of a bumble bee (**Figure 2**) } (Li, et al. 2015).
365 ANI values between strains of the same sub-lineage were relatively high (92-99%),
366 while ANI values between strains of the two different sub-lineages were low (74-
367 75%, see **Table S2 and Figure S2**). No genome sequence is currently available for the
368 designated type strain MRM of the described species *B. apis* (Yun, et al. 2017).
369 However, based on 16S rRNA sequence similarity, this strain could undoubtedly be
370 assigned to one of the two sub-lineages (**Figure S3**). We will refer to this sub-lineage
371 as '*Bombella apis*', while sub-lineage A2.2-2 may represent a new species of the genus
372 *Bombella*.
373 Taken together, we can conclude that while the sequenced *Bombella* strains clearly
374 belong to the acetous group, more precisely to the *Gluconobacter* clade, the deep-
375 branching basal position of Alpha2.1 suggests that it belongs either to an early
376 diverging clade within the acetous group, or to a yet unexplored third subgroup. This
377 calls for further taxonomic analyses, since the atypical GC content of the Apha2.1
378 genomes might confound the position in the phylogeny.
379
380 **Functional similarities between Alpha2.1 and Alpha2.2 based on the shared**
381 **core gene content.**
382 To compare the functional gene content of Alpha2.1 and Alpha2.2, we clustered
383 homologous genes of the two phylotypes into gene families and looked at their
384 distribution across the 21 analyzed genomes. A total of 3,275 gene families were

385 identified (including singletons), of which 1,006 were conserved across all strains,
386 hereafter referred to as the shared core genome of Alpha2.1 and Alpha2.2 (**Figure**
387 **3A, Table S3**).
388
389 Inspection of the shared core genome revealed that Alpha2.1 and Alpha2.2 have
390 similar biosynthetic capabilities as other proteobacterial bee gut symbionts (Kwong
391 and Moran 2016). Both phylotypes encode complete gene sets for the biosynthesis of
392 all amino acids (except for alanine), peptidoglycan, LPS, and five vitamins (riboflavin
393 (vitamin B_2), nicotinate (vitamin B_3), pantothenate (vitamin B_5), pyridoxine (vitamin
394 B_6), and tetrahydrofolate (vitamin B_9)) (**Figure 3C** and **3D**, **Figure S4**). In contrast to
395 most symbionts in the honey bee gut, Alpha2 lineages code for pathways for the *de*
396 *novo* synthesis of purine and pyrimidine nucleosides. The dihydroorotate
397 dehydrogenase (E.C. 1.2.5.2) in Alpha2.1 is quite divergent, with higher identity to its
398 homologues in *Gilliamella* (~64%) than those in Alpha2.1 (~24%). The gene is
399 adjacent to *uup*, involved in transposon excision in *E. coli* (Carlier, et al. 2012),
400 suggesting it has been acquired by HGT. Notwithstanding, all genomes from Alpha2.1
401 code for a symporter for the uptake of environmental orotate, suggesting it can also
402 satisfy its need for this nucleoside precursor extracellularly. The *de novo* nucleoside
403 biosynthesis is costly, and several gut symbionts lack these functional pathways,
404 preferring their uptake from the environment as evidenced by nucleoside depletion
405 in the bee gut in the presence of microbiota (Kesnerova, et al. 2017). Alpha2.2
406 genomes code for an adenosine importer, which in turn can be interconverted with
407 inosine, xanthine and guanosine. This suggests that Alpha2.2, like most other bee gut

symbionts, prefers to acquire these DNA and RNA building blocks from the environment rather than synthesizing them *de novo.*

All analyzed strains encoded a complete gene set of the Embden-Meyerhof-Parnas (EMP) pathway except for a homolog of the 6-phosphofructokinase gene (*pfkA*, EC:2.7.1.11), indicating that Alpha2.1 and Alpha2.2 are both capable of performing gluconeogenesis but incapable of carrying out glycolysis. However, all strains harbored the necessary genes to incorporate sugars and sugar acids via the Pentose phosphate pathway (PPP) into intermediate cellular metabolites (**Table S4, Figure 3C** and **3D**, **Figure S4**). In addition, all strains of Alpha2.2 carried a 6-phosphogluconate dehydratase (*edd*, EC 4.2.1.12) and a 2-keto-3-deoxy-6-phosphogluconate (*eda*, EC 4.1.2.14), which allows them to use the Entner-Doudoroff pathway (ED).

Analysis of the shared core genome content also revealed that both phylotypes - like most other Acetobacteraceae (Matsushita and Matsutani 2016)- carry out oxidative phosphorylation and rely on oxygen as the final electron acceptor (**Figure S4, Figure 4A, Table S4**). They encoded genes of the respiratory complex III ubiquinol:cytochrome *bc1* reductase, which can transfer electrons from the quinone pool to different cytochrome *c*-containing proteins. While they lacked genes for the complex IV COX-type cytochrome c terminal oxidase (*coxABC*), both lineages carried genes coding for two ubiquinol-dependent terminal oxidases: *ctaAB* and *cyaDCBA* for the cytochrome *bo3* oxidase, and *cydAB* for the low oxygen affinity cytochrome *bd* oxidase (Matsushita and Matsutani 2016). Overall, this suggest that both lineages heavily rely on utilizing the ubiquinone pool to transfer electrons.

One of the hallmarks of the Acetobacteraceae metabolism is the production of acetate from ethanol via acetaldehyde (De Roos and De Vuyst 2018). Although both lineages carried genes for NADH-producing alcohol dehydrogenases to convert ethanol into acetaldehyde (EC:1.1.1.1, EC:1.1.1.284,) none of them was a homolog of the ubiquinone-dependent PQQ alcohol dehydrogenases characteristic of acetate production and responsible in *Acetobacter pomorum* to induce the host insulin signaling pathway of *Drosophila* resulting in increased growth and development (Shin, et al. 2011). Moreover, none of the lineages carried genes for the subsequent production of acetaldehyde to acetate (EC:1.2.1.-), nor for acetate transport (COG1584). This means that none of the lineages produce ethanol through oxidative fermentation nor are they acetate producers, two characteristic reactions of Acetobacteraceae.

Functional differences between Alpha2.1 and Alpha2.2 based on the phylotype-specific gene content

Although sharing many metabolic capabilities, Alpha2.1 and Alpha 2.2 each harbored a considerable number of phylotype-specific core gene families, i.e. gene families present across all analyzed strains of one phylotype, but absent from all strains of the other phylotype (**Figure 3A, Table S5**): 483 and 407 gene families for Alpha2.1 and Alpha2.2, respectively. The number of core gene families specific to each of the two sub-lineages of Alpha2.2 was relatively small: 45 gene families for *B. apis* and 24 gene families for *Bombella sp.* All three groups (i.e. Alpha2.1, *B. apis,* and *B. sp.*) also harbored a relatively large flexible gene pool (i.e. gene families present in only a

454 subset of the strains), as indicated by the total number of gene families in each group
455 (i.e. pan genome) as compared to the core genome (**Figure 3A, Table S5**).
456
457 The 483 and 407 phylotype-specific core gene families belonged to a wide range of
458 COG categories suggesting differences in diverse metabolic functions. However, four
459 COG categories stood out as being particularly abundant among the Alpha2.1-specific
460 core gene content as compared to the shared core gene content (**Figure 3B**): 'Energy
461 production and conversion' (COG C) , 'Carbohydrate transport and metabolism' (COG
462 G), 'Coenzyme transport and metabolism' (COG H), and 'Inorganic ion transport and
463 metabolism' (COG P). Gene families of these particular categories encoded a relatively
464 large number of dehydrogenases/oxidoreductases, putative transporters for
465 nitrate/nitrite, sulfate, and iron, and nearly complete gene sets for the synthesis of
466 vitamin B_1 (thiamine), vitamin B12, and the co-factor molybdopterin, all of which
467 were absent from Alpha2.2. Interestingly, several genes linked to the TCA cycle were
468 also specific to Alpha2.1 (**Figure 3C and 3D**).
469
470 Among the Alpha2.2-specific core gene content two COG categories were particularly
471 abundant compared to the shared core genome (**Figure 3B**): 'Inorganic ion transport
472 and metabolism' (i.e. COG P) and 'Cell motility' (i.e. COG N). Almost all gene families
473 in the category 'Cell motility' coded for different flagella subunits, suggesting that
474 Alpha2.2, but not Alpha2.1, is motile (**Figure 3C and 3D**). Gene families in 'Inorganic
475 ion transport and metabolism' encoded for diverse transporters, in particular for iron
476 and phosphate. All *Bombella* strains also harbored genes encoding the redox cofactor

pyrroloquinoline quinone (PQQ) and a PQQ-dependent glucose dehydrogenase involved in respiration (see below). Another notable difference between the two phylotypes was that Alpha2.2 encoded an entire CRISPR/CAS9 system, while Alpha2.1 was lacking any homolog of these antiviral defense systems (**Figure 3C and 3D**).

Major differences in respiratory chain and TCA cycle between Alpha2.1 and Alpha2.2.

Both phylotypes seem to carry out aerobic respiration to gain energy. However, the presence of various dehydrogenases and TCA cycle genes among the phylotype-specific gene content suggested differences in the energy metabolism (**Table S5**). We identified 16 membrane-associated dehydrogenase/reductases likely to be involved in electron transport respiratory chain hence the production of energy (**Figure 4**, **Table S6**).

Six dehydrogenase/reductase gene families were present among the shared gene content of Alpha2.1 and Alpha2.2, three of which are electron donors (a D-lactate dehydrogenase, a complex I type-II NADH dehydrogenase, and a putative membrane-bound dehydrogenase), and three of which are electron acceptors, namely the ubiquinol:cytochrome *bc1* complex III and the two terminal electron acceptors cytochrome *bo3* and *bd* ubiquinol oxidases. and *bo3*). These consist of the common respiratory metabolism shared between the two lineages. The remaining 11 dehydrogenases/reductases belonged to the phylotype-specific core gene content (i.e. were present in all strains of one phylotype but absent from the other).

Eight of these phylotype-specific dehydrogenases/reductases were only present in Alpha2.1. Six dehydrogenases are electron donors to ubiquinone from oxidation of succinate, NADH (one *nuo*/type-I in addition to the shared *ndh*/type-II), glycerol-3-phosphate, malate and the aforementioned dihydroorotate dehydrogenase(**Figure 4A**, **Table S6**). The type I NADH dehydrogenase and the nitrate reductase are proton-pumping enzymes that directly contribute to the production of energy (Marreiros, et al. 2016). Both enzyme complexes are dependent on the cofactor molybdopterin, which explains the presence of the corresponding biosynthesis genes in the genomes of Alpha2.1, but not Alpha2.2. The genes encoding the co-factor and the nitrate reductase are located in the same genomic island and have best blast hits to Gamma- and Betaproteobacteria, suggesting acquisition by HGT (**Figure S5**). (**Figure 4A**). . The other two terminal oxidases are nitrate and nitric oxide reductases, which suggest that Alpha2.1 has the capability to carry out anaerobic respiration (Figure 4A). Homologs of the gene encoding the nitric oxide reductase were not present in any other closely related *Acetobacteriaeceae* strain (i.e. other *Commensalibacter* sp.) suggesting a specific role for Alpha2.1 and acquisition by HGT as is common for denitrification enzymes (Jones, et al. 2008) (**Figure S6**).

Only three dehydrogenases were specific to Alpha2.2: a different dihydroorotate dehydrogenase, the above mentioned PQQ-dependent glucose dehydrogenase, and a gluconate-2-dehydrogenase (**Figure 4B**). The last two catalyze the periplasmic conversion of glucose into gluconate and gluconate into 2-keto-gluconate, and are characteristic of the oxidative fermentation pathway common to most Acetobacteraceae. This is in contrast to Alpha2.1, which lacked genes encoding

523 periplasmic dehydrogenases and hence seems not able to carry out oxidative
524 fermentation.
525
526 We also identified major differences in the TCA cycle between the two phylotypes. All
527 strains of Alpha2.1 harbored the full gene set of the TCA cycle, including an
528 acetate:succinate CoA-transferase gene (*aarC*) for metabolizing acetate to acetyl-CoA
529 (Mullins, et al. 2008), a typical feature of the TCA cycle of Acetobacteraceae thought
530 to be an adaptation to the high amounts of acetate produced by some of them (**Figure**
531 **4C**). In contrast, the TCA cycle was incomplete across strains of Alpha2.2 to different
532 degrees. While in the sub-lineage of *B. apis* the enzymatic steps from succinyl-CoA to
533 fumarate (two steps) and from malate to oxaloacetate (one step) were missing, the
534 strains of the other sublineage were also missing genes for the conversion of acetyl-
535 CoA to 2-oxoglutarate (four steps). Several Acetobacteraceae have been reported to
536 harbor partial or modified TCA cycles, especially those that utilize oxidative
537 fermentation to gain energy (Brown and Wernegreen 2019; Mullins, et al. 2008).
538 However, in none of them the pathway seems to be as reduced as in this particular
539 sublineage of *Bombella* (**Figure 4D***)*.
540 Intriguingly, the fate of malate in the two phylotypes is markedly different. Alpha2.1
541 displays a malate:quinone oxidoreductase as part of the conversion of malate to
542 oxaloacetate in the TCA cycle. This enzyme is an alternative that contributes to both
543 the proton motive force through proton pumping and the ubiquinone pool with
544 electrons. Malate can also be converted into pyruvate and CO2 through malate
545 dehydrogenases. Alpha2.1 codes for the NADP-dependent *maeB* (EC 1.1.1.40), and

546 Alpha2.2 codes for a NAD-dependent *maeA* (EC 1.1.1.38). The latter is found adjacent
547 to the class-II fumarase *fumC* (EC 4.2.1.2) in the Alpha2.2 genomes, that also lack the
548 genes to generate fumarate from succinate, and hence must rely on fumarate
549 produced through other anabolic pathways such as the urea cycle or aspartate
550 metabolism.
551

Discussion

The two Acetobacteraceae, Alpha2.1 and Alpha2.2, are common members of the bee gut microbiota. They have been repeatedly identified by 16S rRNA gene sequence analyses in samples from the bee gut or the hive environment and several studies have provided important insights about their ecology (Cox-Foster, et al. 2007; Kapheim, et al. 2015; Kešnerová, et al. 2019; Kwong, et al. 2017; Martinson, et al. 2011; Martinson, et al. 2012; Moran, et al. 2012; Powell, et al. 2018; Vojvodic, et al. 2013). However, a comparative analysis between their genomes had not been carried out. Our study provides new insights about the phylogenetic relationship of Alpha2.1 and Alpha2.2, their functional gene content, and their metabolic and genomic differences.

We show that Alpha2.1 and Alpha2.2 belong to two distant phylogenetic clades within the Acetobacteraceae, confirming previous findings (Martinson, et al. 2011), and determining their taxonomic position more precisely. Alpha2.2 is part of the acetous group and falls within the genus *Bombella*, to which also *B. intestini* belongs (Li, et al. 2015). However, the Alpha2.2 strains isolated from honey bees belong to two divergent sub-lineages within the genus *Bombella.* The species formally described as *B. apis* (Yun, et al. 2017) and strains referred to as *Candidatus* 'Parasaccharibacter apium' and the undescribed strains *Saccharibacter* sp. 3.A.1 (Veress, et al. 2017) belong to the same sub-lineage, are identical or nearly identical in the 16S rRNA gene, and share high ANI with each other. While we acknowledge that the name 'Parasaccharibacter apium' has been proposed first (Corby-Harris, et al. 2014), it has

575 never been formally described as a novel genus, and thus we propose that these
576 closely related strains should be consistently referred to as *Bombella apis* (Yun et al.
577 2017).
578 The second sub-lineage of Alpha2.2 represents a putative novel species within the
579 genus *Bombella* based on its genomic divergence and paraphyletic position relative
580 to *B. apis*. The TCA cycle is almost completely lost in this sub-lineage suggesting
581 distinctive functional capabilities as compared to *B. apis*. How prevalent this sub-
582 lineage is across honey bee colonies, and whether it colonizes a different niche than
583 *Bombella apis* remains to be determined.
584
585 The second phylotype, Alpha2.1, belongs to a highly divergent lineage within the
586 Acetobacteraceae and is likely to represent a novel species within the genus
587 *Commensalibacter*. The only three taxa which are known from this lineage have been
588 isolated from insect guts, and all three display an atypical GC content compared to
589 other Acetobacteraceae. Future sampling of bacteria from this lineage will help to
590 solidify its position within the Acetobacteraceae phylogeny and to test if this lineage
591 harbors exclusively insect gut-associated symbionts.
592 Despite the fact that Alpha2.1 and Alpha2.2 are both Acetobacteraceae and can
593 colonize the same environment (i.e. the bee gut), we found major differences in their
594 metabolic and respiratory capabilities. While both rely on the ubiquinone pool and
595 oxygen- and ubiquinone-dependent terminal oxidases (and hence are aerobic),
596 Alpha2.1 is also able to respire nitrate and nitric oxide, making it a facultative
597 anaerobe.

598 The most important metabolic difference between the two phylotypes lies in their
599 energy and central carbon metabolism. Alpha2.2 exhibits a simple metabolism that
600 relies on the periplasmic oxidative fermentation of glucose and gluconate. Although
601 it carries genes to incorporate these intermediates through the ED or PP pathways,
602 this is unlikely, since most Acetobacteraceae accumulate 5-keto-D-gluconate
603 extracellularly (Bringer and Bott 2016), including *B. apis* (Yun, et al. 2017), and it has
604 been shown that a very small proportion of glucose and gluconate is incorporated
605 into the cell during oxidative fermentation in *Gluconobacter oxidans*.,. This, together
606 with a highly reduced TCA cycle that appears to retain only steps involved in other
607 biosynthetic pathways, suggest that Alpha2.2 has a streamlined and rigid metabolism
608 optimized for low-yield but rapid energy production from glucose without
609 incorporation into biomass. This has been shown to occur when growth rate is slower
610 than the reducing equivalent production rate in marine cyanobacteria (Braakman, et
611 al. 2017), and may provide Alpha2.2 strains an advantage in glucose-rich
612 environments, such as royal jelly (Simo and Christensen 1962). Worker bees secrete
613 glucose oxidase into royal jelly to partly convert glucose into gluconolactone and
614 hydrogen peroxide for antimicrobial purposes (Fratini, et al. 2016; Ohashi, et al.
615 1999). Alpha2.2 is able to further oxidize gluconolactone into gluconate (Smith and
616 Newton 2018) and enter the ED pathway, potentially facilitating its growth in royal
617 jelly.
618 In contrast to Alpha2.2, Alpha2.1 displays a wider range of primary electron donors
619 that includes several organic acids and NADH, but does not permit oxidative
620 fermentation. It carries the complete acetate-driven TCA cycle typically found in

621 Acetobacteraceae, which is a low-yield variant already optimized for rapid resource
622 consumption and replenishment of TCA intermediates while retaining steps needed
623 to provide precursors for multiple biosynthetic pathways. A larger range of carbon
624 substrate utilization fits a lifestyle that is more specialized to the honey bee hindgut.
625 In support of this, a recent study showed that while many transient bacteria
626 (including Alpha2.2) diminished or completely disappeared in long-lived winter bees,
627 Alpha2.1 persisted in the gut environment and even significantly increased relative
628 to foragers and nurses sampled in summer (Kešnerová, et al. 2019). The same pattern
629 was also observed for another community member of the bee gut microbiota,
630 *Bartonella apis*. Like Alpha2.1 this species relies on aerobic respiration (Segers, et al.
631 2017),in contrast to most other members, which are anaerobic fermenters. However,
632 the metabolic or physicochemical conditions in the hindgut of winter bees that favor
633 the growth of Alpha2.1 or *Bartonella apis* over other microbiota members is currently
634 unclear.
635 In summary, we have shown that the two honey bee-associated Acetobacteraceae
636 phylotypes both display adaptations towards fast growing metabolism, but in two
637 markedly different ways. Alpha2.1 harvests energy from a broad-range of substrates
638 and links substrate utilization with a flexible metabolism of oxidative and
639 biosynthetic pathways, whereas Alpha2.2 has streamlined its metabolism by using
640 oxidative fermentation for rapid energy harvesting from glucose, almost exclusively
641 because of the loss of alternative oxidative pathways. Our results exemplify how
642 physiological and metabolic differences may drive niche differentiation in terms of

643 resource utilization and spatial distribution, facilitating the coexistence of
644 phylogenetically related bacteria in the honey bee gut.
645
646

Acknowledgments

Participants of the Master in Molecular Life Sciences at the University of Lausanne, class 2017-2018 (Marie-Pierre Meurville, Maxime Brunner, Jonathan Klopfenstein, Baptiste Micheli, Tess Bonato, Alice Wallef, Alicia Zufferey, Andrea Dos Santos, Augustine Jaccard, Béatrice Tappy, Boris Schnider, Christèle Aubry, Clara Heiman, Dinis Barros, Emilie Ha, Giti Ghazi Soltani, Henri Paratte, Ilaria Bernabei, Imre Banlaki, Julien Dénéréaz, Karim Saied, Liza Darrous, Lucas Beurret, Lucas Serra Moncadas, Lucien Roesch, Ludivine Brandt, Lydia Horwath, Margot Delavy, Maria Chadiarakou, Maria Hadres, Mathieu Quinodoz, Maxence Trottet, Pascal Roth, Perrine Steffe, Robin Hofmeister, Béatrice Tappy, and Vishwachi Tripathi) conducted preliminary research as part of the 'Sequence-a-genome' (SAGE) graduate course. We thank the School of Biology of the University of Lausanne for financial support, and the Competence Centre in Bioinformatics and Computational Biology of the Swiss Institute of Bioinformatics (SIB) for providing access to high-performance computation through the Vital-IT cluster (http:vital-it.ch). This research was funded through the School of Biology of the University of Lausanne, and the European Research Council (ERC-StG 'MicroBeeOme'), the Swiss National Science Foundation (grant number 31003A_160345 and 31003A_179487) and HFSP (HFSP Young Investigator grant RGY0077/2016) received by PE.

Data deposition

Genome sequences and Illumina data underlying genome assemblies and 16S rRNA gene amplicon analysis have been deposited on NCBI under Bioproject accession PRJNA589199. Codes and data of the analysis can be found via the following SWITCHdrive link: https://drive.switch.ch/index.php/s/syuWZAQKKGmQn0h. These files will be moved to a public Zenodo repository upon publication.

References

Anderson KE, et al. 2014. Hive-stored pollen of honey bees: many lines of evidence are consistent with pollen preservation, not nutrient conversion. Mol Ecol 23: 5904-5917. doi: 10.1111/mec.12966

Anderson KE, et al. 2018. The queen's gut refines with age: longevity phenotypes in a social insect model. Microbiome 6: 108. doi: 10.1186/s40168-018-0489-1

Bankevich A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19: 455-477. doi: 10.1089/cmb.2012.0021

Bartowsky EJ, Henschke PA 2008. Acetic acid bacteria spoilage of bottled red wine -- a review. Int J Food Microbiol 125: 60-70. doi: 10.1016/j.ijfoodmicro.2007.10.016

Bik HM 2014. Phinch: An interactive, exploratory data visualization framework for –Omic datasets. bioRxiv: 009944. doi: 10.1101/009944

Bolger AM, Lohse M, Usadel B 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114-2120. doi: 10.1093/bioinformatics/btu170

Bonilla-Rosso G, Engel P 2018. Functional roles and metabolic niches in the honey bee gut microbiota. Curr Opin Microbiol 43: 69-76. doi: 10.1016/j.mib.2017.12.009

Braakman R, Follows MJ, Chisholm SW 2017. Metabolic evolution and the self-organization of ecosystems. Proc Natl Acad Sci U S A 114: E3091-E3100. doi: 10.1073/pnas.1619573114

695 Bringer S, Bott M. 2016. Central Carbon Metabolism and Respiration in
696 Gluconobacter oxydans. In: Matsushita K, Toyama H, Tonouchi N, Okamoto-
697 Kainuma A, editors. Acetic Acid Bacteria: Ecology and Physiology. Tokyo: Springer
698 Japan. p. 235-253.

699 Brodschneider R, Crailsheim K 2010. Nutrition and health in honey bees. Apidologie
700 41: 278-294. doi: 10.1051/apido/2010012

701 Brown BP, Wernegreen JJ 2019. Genomic erosion and extensive horizontal gene
702 transfer in gut-associated Acetobacteraceae. BMC Genomics 20: 472. doi:
703 10.1186/s12864-019-5844-5

704 Caporaso JG, et al. 2011. Global patterns of 16S rRNA diversity at a depth of millions
705 of sequences per sample. Proc Natl Acad Sci U S A 108 Suppl 1: 4516-4522. doi:
706 10.1073/pnas.1000080107

707 Carlier L, Haase AS, Burgos Zepeda MY, Dassa E, Lequin O 2012. The C-terminal
708 domain of the Uup protein is a DNA-binding coiled coil motif. J Struct Biol 180: 577-
709 584. doi: 10.1016/j.jsb.2012.09.005

710 Chen IA, et al. 2017. IMG/M: integrated genome and metagenome comparative data
711 analysis system. Nucleic Acids Res 45: D507-D516. doi: 10.1093/nar/gkw929

712 Chin CS, et al. 2013. Nonhybrid, finished microbial genome assemblies from long-
713 read SMRT sequencing data. Nat Methods 10: 563-569. doi: 10.1038/nmeth.2474

714 Chouaia B, et al. 2014. Acetic acid bacteria genomes reveal functional traits for
715 adaptation to life in insect guts. Genome Biol Evol 6: 912-920. doi:
716 10.1093/gbe/evu062

717 Corby-Harris V, Anderson KE 2018. Draft Genome Sequences of Four
718 Parasaccharibacter apium Strains Isolated from Honey Bees. Genome Announc 6.
719 doi: 10.1128/genomeA.00165-18

720 Corby-Harris V, et al. 2016. Parasaccharibacter apium, gen. nov., sp. nov., Improves
721 Honey Bee (Hymenoptera: Apidae) Resistance to Nosema. J Econ Entomol 109: 537-
722 543. doi: 10.1093/jee/tow012

723 Corby-Harris V, et al. 2014. Origin and effect of Alpha 2.2 Acetobacteraceae in honey
724 bee larvae and description of Parasaccharibacter apium gen. nov., sp. nov. Appl
725 Environ Microbiol 80: 7460-7472. doi: 10.1128/AEM.02043-14

726 Cox-Foster DL, et al. 2007. A metagenomic survey of microbes in honey bee colony
727 collapse disorder. Science 318: 283-287. doi: 10.1126/science.1146498

728 Crotti E, et al. 2010. Acetic acid bacteria, newly emerging symbionts of insects. Appl
729 Environ Microbiol 76: 6963-6970. doi: 10.1128/AEM.01336-10

730 De Roos J, De Vuyst L 2018. Acetic acid bacteria in fermented foods and beverages.
731 Curr Opin Biotechnol 49: 115-119. doi: 10.1016/j.copbio.2017.08.007

732 Durbin R 2014. Efficient haplotype matching and storage using the positional
733 Burrows-Wheeler transform (PBWT). Bioinformatics 30: 1266-1272. doi:
734 10.1093/bioinformatics/btu014

735 Ellegaard KM, Engel P 2019. Genomic diversity landscape of the honey bee gut
736 microbiota. Nat Commun 10: 446. doi: 10.1038/s41467-019-08303-0

737 Fratini F, Cilia G, Mancini S, Felicioli A 2016. Royal Jelly: An ancient remedy with
738 remarkable antibacterial properties. Microbiol Res 192: 130-141. doi:
739 10.1016/j.micres.2016.06.007

740 Guy L, Kultima JR, Andersson SG 2010. genoPlotR: comparative gene and genome
741 visualization in R. Bioinformatics 26: 2334-2335. doi:
742 10.1093/bioinformatics/btq413

743 Jones CM, Stres B, Rosenquist M, Hallin S 2008. Phylogenetic analysis of nitrite, nitric
744 oxide, and nitrous oxide respiratory enzymes reveal a complex evolutionary history
745 for denitrification. Mol Biol Evol 25: 1955-1966. doi: 10.1093/molbev/msn146

746 Kapheim KM, et al. 2015. Caste-specific differences in hindgut microbial
747 communities of honey bees (Apis mellifera). PLoS One 10: e0123911. doi:
748 10.1371/journal.pone.0123911

749 Katoh K, Standley DM 2013. MAFFT multiple sequence alignment software version
750 7: improvements in performance and usability. Mol Biol Evol 30: 772-780. doi:
751 10.1093/molbev/mst010

752 Kay M, Wobbrock J. 2016. ARTool: Aligned Rank Transform for Nonparametric
753 Factorial ANOVAs, R package version 0.10.0.

754 Kešnerová L, Emery O, Troilo M, Erkosar B, Engel P 2019. Gut microbiota structure
755 differs between honey bees in winter and summer. bioRxiv: 703512. doi:
756 10.1101/703512

757 Kesnerova L, et al. 2017. Disentangling metabolic functions of bacteria in the honey
758 bee gut. PLoS Biol 15: e2003467. doi: 10.1371/journal.pbio.2003467

759 Kim EK, et al. 2012. Draft genome sequence of Commensalibacter intestini A911T, a
760 symbiotic bacterium isolated from Drosophila melanogaster intestine. J Bacteriol
761 194: 1246. doi: 10.1128/JB.06669-11

762 Komagata K, Iino T, Yamada Y. 2014. The Family Acetobacteraceae. In: Rosenberg E,
763 DeLong EF, Lory S, Stackebrandt E, Thompson F, editors. The Prokaryotes:
764 Alphaproteobacteria and Betaproteobacteria. Berlin, Heidelberg: Springer Berlin
765 Heidelberg. p. 3-78.

766 Kwong WK, et al. 2017. Dynamic microbiome evolution in social bees. Sci Adv 3:
767 e1600513. doi: 10.1126/sciadv.1600513

768 Kwong WK, Moran NA 2016. Gut microbial communities of social bees. Nat Rev
769 Microbiol 14: 374-384. doi: 10.1038/nrmicro.2016.43

770 Lee I, Ouk Kim Y, Park SC, Chun J 2016. OrthoANI: An improved algorithm and
771 software for calculating average nucleotide identity. Int J Syst Evol Microbiol 66:
772 1100-1103. doi: 10.1099/ijsem.0.000760

773 Li L, et al. 2016. Whole-Genome Sequence Analysis of Bombella intestini LMG
774 28161T, a Novel Acetic Acid Bacterium Isolated from the Crop of a Red-Tailed
775 Bumble Bee, Bombus lapidarius. PLoS One 11: e0165611. doi:
776 10.1371/journal.pone.0165611

777 Li L, et al. 2015. Bombella intestini gen. nov., sp. nov., an acetic acid bacterium
778 isolated from bumble bee crop. Int J Syst Evol Microbiol 65: 267-273. doi:
779 10.1099/ijs.0.068049-0

780 Li L, Stoeckert CJ, Jr., Roos DS 2003. OrthoMCL: identification of ortholog groups for
781 eukaryotic genomes. Genome Res 13: 2178-2189. doi: 10.1101/gr.1224503

782 Maes PW, Rodrigues PA, Oliver R, Mott BM, Anderson KE 2016. Diet-related gut
783 bacterial dysbiosis correlates with impaired development, increased mortality and
784 Nosema disease in the honeybee (Apis mellifera). Mol Ecol 25: 5439-5450. doi:
785 10.1111/mec.13862

786 Marreiros BC, et al. 2016. Exploring membrane respiratory chains. Biochim Biophys
787 Acta 1857: 1039-1067. doi: 10.1016/j.bbabio.2016.03.028

788 Martinson VG, et al. 2011. A simple and distinctive microbiota associated with honey
789 bees and bumble bees. Mol Ecol 20: 619-628. doi: 10.1111/j.1365-
790 294X.2010.04959.x

791 Martinson VG, Moy J, Moran NA 2012. Establishment of characteristic gut bacteria
792 during development of the honeybee worker. Appl Environ Microbiol 78: 2830-
793 2840. doi: 10.1128/AEM.07810-11

794 Matsushita K, Matsutani M. 2016. Distribution, Evolution, and Physiology of
795 Oxidative Fermentation. In: Matsushita K, Toyama H, Tonouchi N, Okamoto-
796 Kainuma A, editors. Acetic Acid Bacteria: Ecology and Physiology. Tokyo: Springer
797 Japan. p. 159-178.

798 Moran NA, Hansen AK, Powell JE, Sabree ZL 2012. Distinctive gut microbiota of
799 honey bees assessed using deep sampling from individual worker bees. PLoS One 7:
800 e36393. doi: 10.1371/journal.pone.0036393

801 Mullins EA, Francois JA, Kappock TJ 2008. A specialized citric acid cycle requiring
802 succinyl-coenzyme A (CoA):acetate CoA-transferase (AarC) confers acetic acid
803 resistance on the acidophile Acetobacter aceti. J Bacteriol 190: 4933-4940. doi:
804 10.1128/JB.00405-08

805 Munoz-Gomez SA, et al. 2019. An updated phylogeny of the Alphaproteobacteria
806 reveals that the parasitic Rickettsiales and Holosporales have independent origins.
807 Elife 8. doi: 10.7554/eLife.42535

808 Ohashi K, Natori S, Kubo T 1999. Expression of amylase and glucose oxidase in the
809 hypopharyngeal gland with an age-dependent role change of the worker honeybee
810 (Apis mellifera L.). European Journal of Biochemistry 265: 127-133. doi:
811 10.1046/j.1432-1327.1999.00696.x

812 Parks DH, et al. 2018. A standardized bacterial taxonomy based on genome
813 phylogeny substantially revises the tree of life. Nat Biotechnol 36: 996-1004. doi:
814 10.1038/nbt.4229

815 Pedraza RO. 2016. Acetic Acid Bacteria as Plant Growth Promoters. In: Matsushita
816 K, Toyama H, Tonouchi N, Okamoto-Kainuma A, editors. Acetic Acid Bacteria:
817 Ecology and Physiology. Tokyo: Springer Japan. p. 101-120.

818 Powell JE, Eiri D, Moran NA, Rangel J 2018. Modulation of the honey bee queen
819 microbiota: Effects of early social contact. PLoS One 13: e0200527. doi:
820 10.1371/journal.pone.0200527

821 Pruesse E, Peplies J, Glockner FO 2012. SINA: accurate high-throughput multiple
822 sequence alignment of ribosomal RNA genes. Bioinformatics 28: 1823-1829. doi:
823 10.1093/bioinformatics/bts252

824 Rissman AI, et al. 2009. Reordering contigs of draft genomes using the Mauve
825 aligner. Bioinformatics 25: 2071-2073. doi: 10.1093/bioinformatics/btp356

826 Rognes T, Flouri T, Nichols B, Quince C, Mahe F 2016. VSEARCH: a versatile open
827 source tool for metagenomics. PeerJ 4: e2584. doi: 10.7717/peerj.2584

828 Roh SW, et al. 2008. Phylogenetic characterization of two novel commensal bacteria
829 involved with innate immune homeostasis in Drosophila melanogaster. Appl
830 Environ Microbiol 74: 6171-6177. doi: 10.1128/AEM.00301-08

831 Segers FH, Kesnerova L, Kosoy M, Engel P 2017. Genomic changes associated with
832 the evolutionary transition of an insect gut symbiont into a blood-borne pathogen.
833 ISME J 11: 1232-1244. doi: 10.1038/ismej.2016.201

834 Servin-Garciduenas LE, Sanchez-Quinto A, Martinez-Romero E 2014. Draft Genome
835 Sequence of Commensalibacter papalotli MX01, a Symbiont Identified from the Guts
836 of Overwintering Monarch Butterflies. Genome Announc 2. doi:
837 10.1128/genomeA.00128-14

838 Shin SC, et al. 2011. Drosophila microbiome modulates host developmental and
839 metabolic homeostasis via insulin signaling. Science 334: 670-674. doi:
840 10.1126/science.1212782

841 Simo K, Christensen GM 1962. Quantitative analysis of sugars in royal jelly. Nature
842 196: 1208-1209. doi: 10.1038/1961208a0

843 Siozios S, Moran J, Chege M, Hurst GDD, Paredes JC 2019. Complete Reference
844 Genome Assembly for Commensalibacter sp. Strain AMU001, an Acetic Acid
845 Bacterium Isolated from the Gut of Honey Bees. Microbiol Resour Announc 8. doi:
846 10.1128/MRA.01459-18

847 Smith EA, Newton ILG 2018. Genomic signatures of honey bee association in an
848 acetic acid symbiont. bioRxiv: 367490. doi: 10.1101/367490

849 Stamatakis A 2014. RAxML version 8: a tool for phylogenetic analysis and post-
850 analysis of large phylogenies. Bioinformatics 30: 1312-1313. doi:
851 10.1093/bioinformatics/btu033

852 Tarpy DR, Mattila HR, Newton IL 2015. Development of the honey bee gut
853 microbiome throughout the queen-rearing process. Appl Environ Microbiol 81:
854 3182-3191. doi: 10.1128/AEM.00307-15

855 Veress A, Wilk T, Kiss J, Olasz F, Papp PP 2017. Draft Genome Sequences of
856 Saccharibacter sp. Strains 3.A.1 and M18 Isolated from Honey and a Honey Bee
857 (Apis mellifera) Stomach. Genome Announc 5. doi: 10.1128/genomeA.00744-17

858 Vojvodic S, Rehan SM, Anderson KE 2013. Microbial gut diversity of Africanized and
859 European honey bee larval instars. PLoS One 8: e72106. doi:
860 10.1371/journal.pone.0072106

861 Yamada Y, Yukphan P 2008. Genera and species in acetic acid bacteria. Int J Food
862 Microbiol 125: 15-24. doi: 10.1016/j.ijfoodmicro.2007.11.077

863 Yun JH, Lee JY, Hyun DW, Jung MJ, Bae JW 2017. Bombella apis sp. nov., an acetic
864 acid bacterium isolated from the midgut of a honey bee. Int J Syst Evol Microbiol 67:
865 2184-2188. doi: 10.1099/ijsem.0.001921

866 Zhang J, Kobert K, Flouri T, Stamatakis A 2014. PEAR: a fast and accurate Illumina
867 Paired-End reAd mergeR. Bioinformatics 30: 614-620. doi:
868 10.1093/bioinformatics/btt593

869

870

Figure and Tables

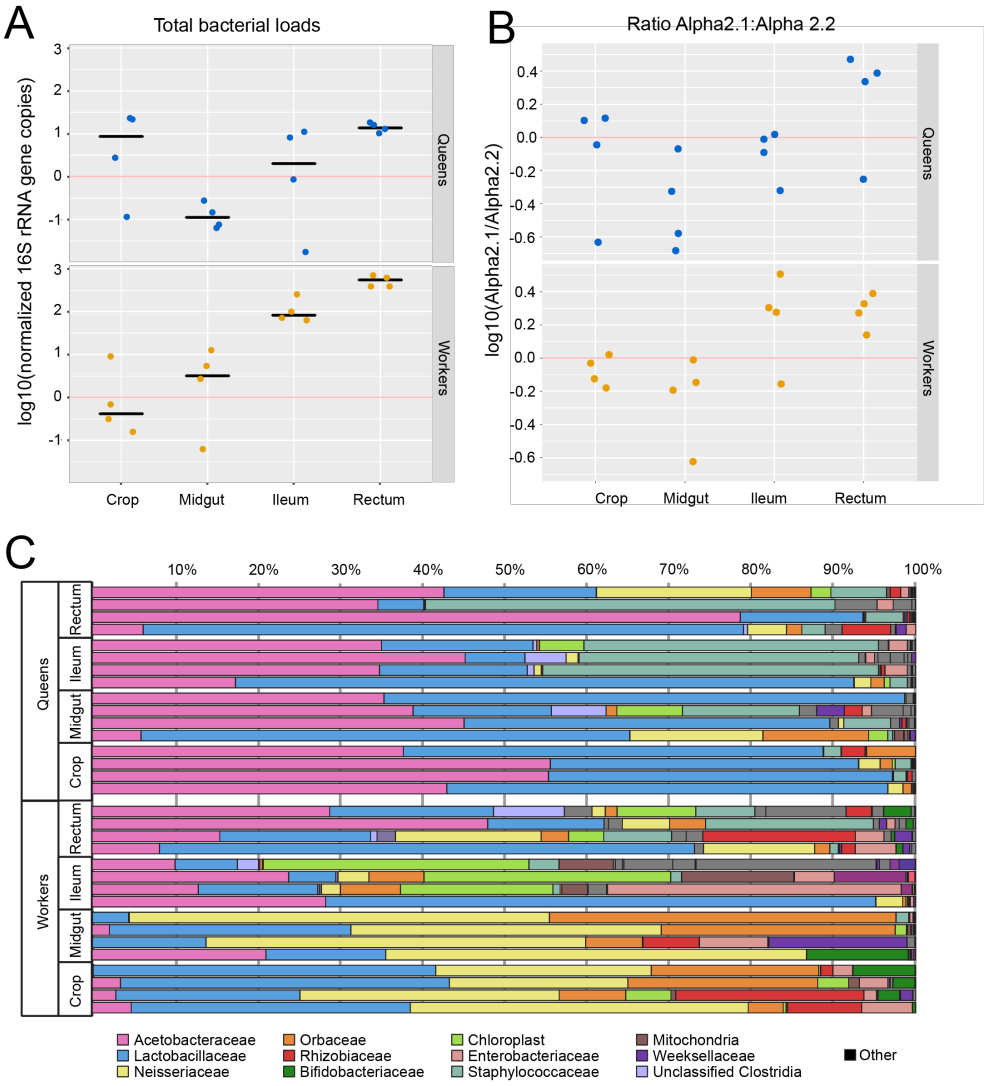

Figure 1. Community analysis of different gut compartments of adult worker and queen bees (n=4 for each category). (A) Total bacterial loads in different gut compartments assessed by qPCR using universal 16S rRNA gene primers and expressed as 16SrRNA copies per host actin copy. (B) Abundance of Alpha2.1 relative to Alpha2.2 in the four gut compartments of worker and queen bees. (C) Bar plot depicting the community composition per gut compartment and caste as relative abundance at the family level.

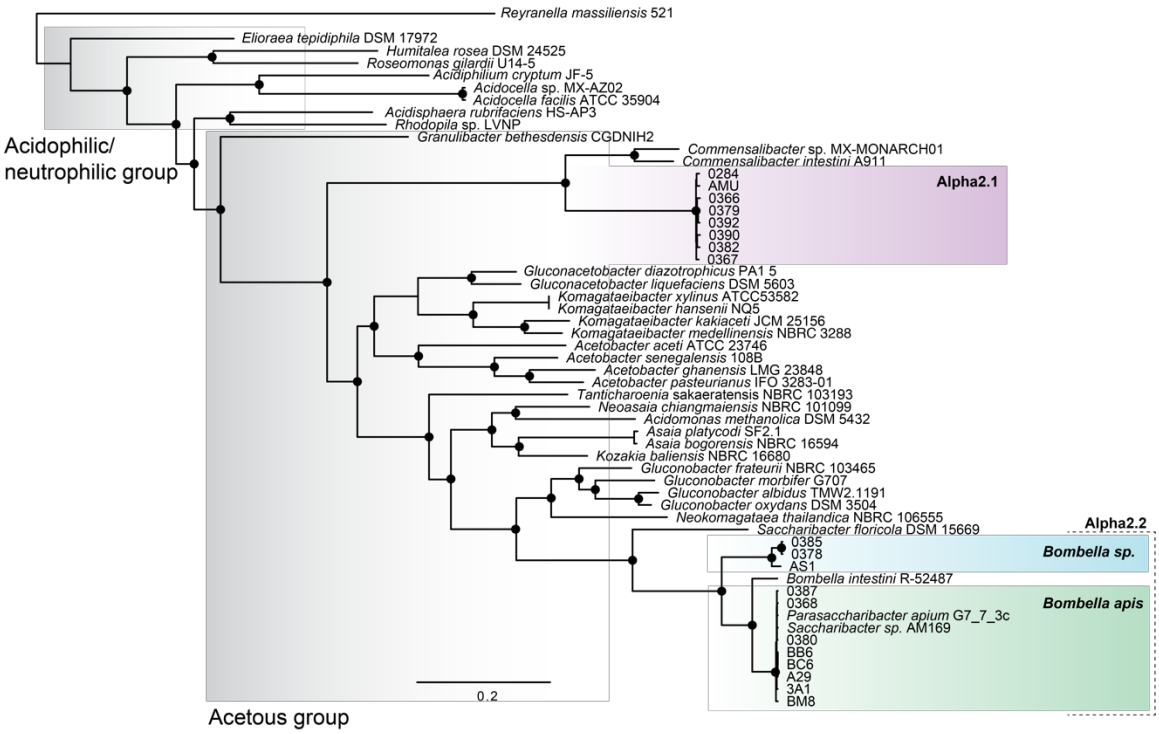

Figure 2. Genome-wide phylogeny of Acetobacteraceae based on 361 single-copy core genes. The tree was inferred using maximum likelihood. Filled circles indicate bootstrap value of 100 out of 100 replicates. Branches without circles indicate bootstrap values <80. The length of the bar indicates 0.2 amino acid substitutions per site. The three clades of interest are highlighted in magenta, blue, and green. The acetous and acidophilic/neutrophilic groups are indicated by shades of grey. Accession numbers of the Alpha2.1 and Alpha2.2 strains can be found in Table S1.

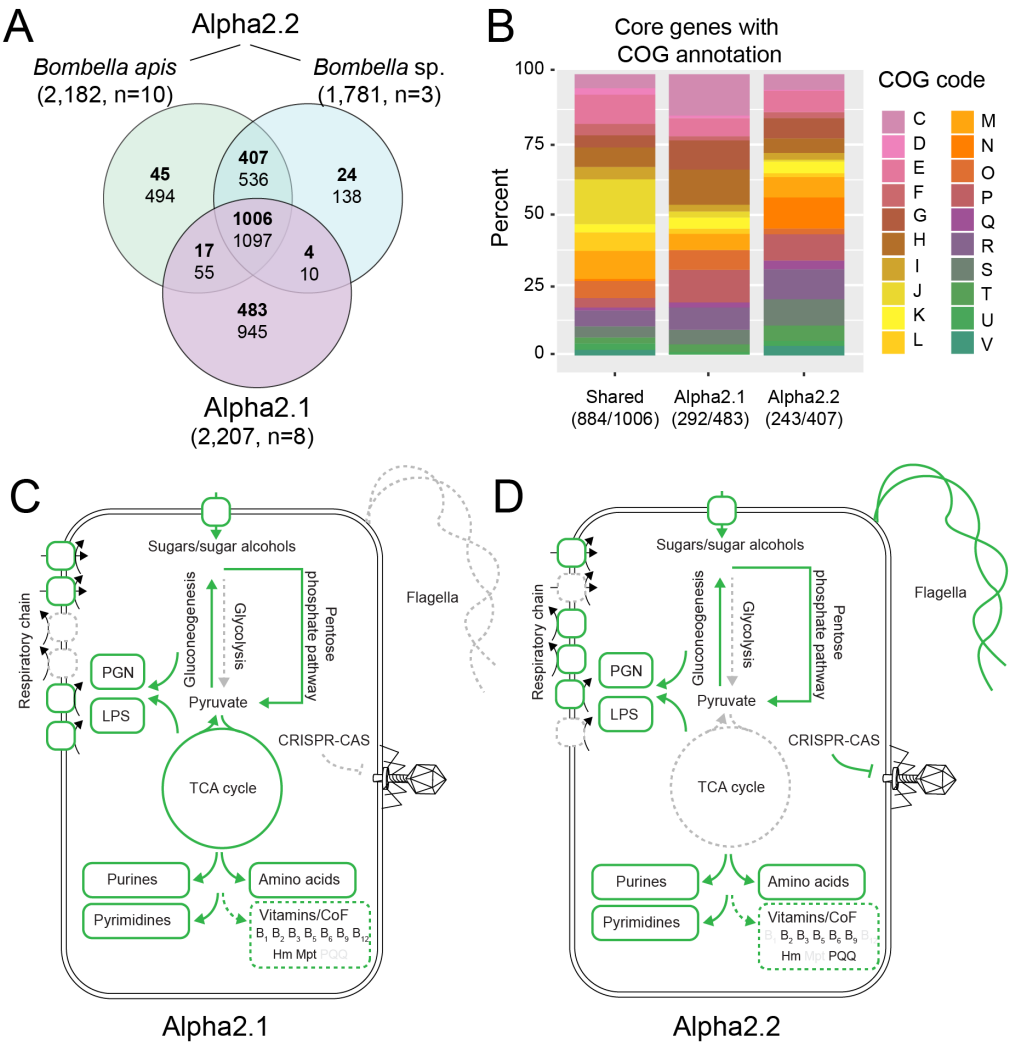

Figure 3. Distribution of gene families among the phylotypes Alpha2.1 and Alpha2.2 and overview of their major functional capabilities. (A) Venn diagram depicting shared and unique gene families in the three phylotypes (Alpha2.1 and the two Alpha2.2 sublineages *Bombella apis* and *Bombella sp.*). Gene families present in all genomes of a given group (i.e. core genes) are indicated in bold font, and gene families present in at least one genome (i.e. pangenome) are indicated in normal font. Numbers in brackets indicate total gene families found for each clade. Notably, the number of analyzed genomes (indicated by n) can influence the number of pan and

900 core gene families. (B) Distribution of shared and phylotype-specific core gene
901 families grouped by COG categories. Numbers in brackets indicate number of core
902 genes with annotation and total number of core genes. (C) and (D) Summary of the
903 major functional capabilities of Alpha2.1 and Alpha2.2 based on the shared and the
904 phylotype-specific gene content. Dashed lines indicated pathways and biosynthetic
905 capabilities missing in the respective phylotype. Solid lines indicate the presence of
906 pathway. LPS, lipopolysaccharide, PGN, peptidoglycan, CoF, cofactor; Hm, heme; MPT,
907 molybdopterin; PQQ, pyrroloquinoline quinone.

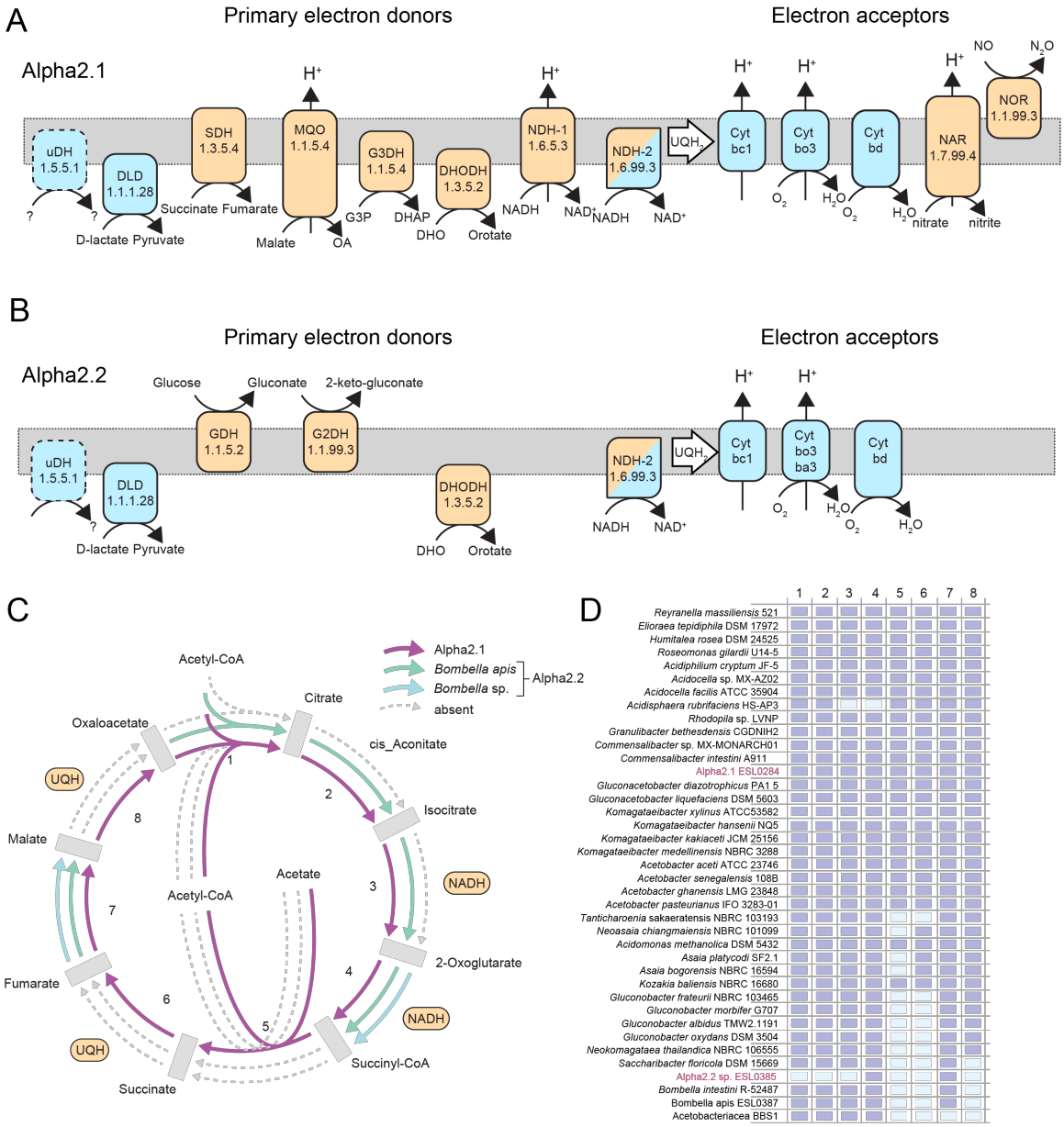

Figure 4. Respiratory chain of Alpha2.1 (A) and Alpha2.2 (B). Shared components are shown in blue, phylotype-specific functions are shown in tan. Dashed outlines indicate that this function is not conserved in all sequenced strains. Both colors indicate the presence of two copies of this function, one which is phylotype-specific, another one which is shared among the two. UQH_2 depicts the ubiquinone pool (C) TCA cycle completeness in Alpha2.1 and both Alpha2.2 sub-lineages, *Bombella apis*

916 and *Bombella* sp. Steps where electrons are transferred to ubiquinone (UQH) and
917 NADH are are indicated. (D) Completeness of the TCA cycle across Acetobacteraceae
918 strains shown in Figure 1, including a recently described strain isolated from the gut
919 of an ant, Acetobacteraceae BBS1, which also has an incomplete TCA cycle (Brown
920 and Wernegreen 2019). DLD, D-lactate dehydrogenase; SDH, succinate
921 dehydrogenase; MQH, maltate:quinone oxidoreductase; G3DH, glycerol-3-phosphate
922 dehydrogenase; DHOHD, dihydroorotate dehydrogenase; NDH-1, NADH
923 dehydrogenase type 1; NDH-2, NADH dehydrogenase type 2; NAR, nitrate reductase;
924 NOR, nitric oxide reductase; uDH, unknown flavoprotein dehydrogenase; GDH, PQQ-
925 dependent glucose dehydrogenase; G2DH, gluconate 2-dehydrogenase, UBQ_2.
926 ubiquinone-2; Cyt, cytochrome. 1, Citrate synthase; 2, Aconitate hydratase; 3,
927 Isocitrate dehydrogenase; 4, 2-oxoglutarate dehydrogenase; 5, Succinyl-CoA:acetate
928 CoA-transferase (*aarC*); 6, Succinate dehydrogenase; 7, Fumarase class II; 8, Malate
929 dehydrogease (*mqo*).
930