Supplementary Information for
Species divergence in gut-restricted bacteria of social bees

Yiyuan Li \textsuperscript{a}, Sean P. Leonard \textsuperscript{a}, J. Elijah Powell \textsuperscript{a}, Nancy A. Moran \textsuperscript{a}

\textsuperscript{a} Department of Integrative Biology, The University of Texas at Austin, Austin, TX 78712, USA

* Corresponding author: Yiyuan Li

Email: lyy0005@gmail.com

This PDF file includes:

- Supplementary Materials and Methods
- Figures S1 to S10
- Tables S1 to S3
- SI References
Supplementary Materials and Methods

Data collection. We downloaded all available Gilliamella and Snodgrassella genomes from NCBI. Low quality genome sequences were removed prior to analysis (Table S1). CheckM version 1.1.3 (1) was used to estimate completeness and contamination of the genomes based on 182 Proteobacteria specific single copy markers. Genomes were removed from downstream analyses if more than 10 markers were missing or if more than 10 markers had multiple copies.

Gene annotation. To annotate the gene structure and functions in a uniform way, we re-annotated all the genomes with Prodigal (2) implemented in anvi’o version 7 (3). We assigned genes as pseudogenes if the annotated coding sequences were short or with internal stop codons. We performed functional annotation using three databases, Clusters of Orthologous Genes (COGs) (4), Kyoto Encyclopedia of Genes and Genomes (KEGG) (5), and evolutionary genealogy of genes: non-supervised orthologous groups (eggNOG) (6). First, we assigned genes to COGs using anvi-run-ncbi-cogs (NCBI COG 2020 release). Then we performed KEGG annotations using both GhostKOALA version 2.2 (database updated on May 15, 2019) (7) and “anvi-run-kegg-kofams” in anvi’o. To retrieve any annotations that were missed by the first two methods, we ran eggnog-mapper version 2.1.2 (8) on the amino acid sequences of genes.

We also annotated certain gene groups of interest using specific annotation methods. We classified CAZyme genes using dbCan2 version 2.0.11 (9). To achieve accurate CAZyme annotations, we only used the annotations with supports from HMMER (10), DIAMOND (11), and Hotpep (12) with default e-values. For genes related to type VI secretion systems, we assign orthologous groups to functions based on the annotations from previous studies (13, 14). We annotated CRISPR spacers using CRISPRCasFinder version 2.0.3 (15). Identical CRISPR spacers between genomes were identified using BLASTN version 2.6.0+ (16) with e-value = 1e-5.

Orthologous group assignment, phylogenetic reconstruction and pangenome analysis. To investigate the core and accessory genomes, we performed ortholog assignment on Gilliamella and Snodgrassella, respectively, using anvi’o version 7. We used DIAMOND (11) and MCL (17), implemented in the anvi’o command “anvi-pan-genome”, on the amino acid sequences of the annotated genes. As the genomes belong to closely related species from the same genus, we assigned orthologous groups with increased sensitivity by setting the MCL inflation value to 10. We defined core gene families as single-copy gene families that existed in at least 80% of the species (i.e., 94 of the 117 Gilliamella genomes or 46 of the 57 Snodgrassella genomes).

To understand strain relationships, core genes were used to construct the phylogenies of Gilliamella and of Snodgrassella. All the selected single-copy orthologous groups were aligned using a codon-based approach with MAFFT version 7.407 (18) and in-house Perl scripts (https://github.com/lyy005/codon_alignment, last accessed Nov 16, 2020). Aligned nucleotide sequences of the orthologous genes were concatenated together. IQ-TREE version 2.0.6 (19) was used to find the best-fit model (20) and construct the phylogenetic trees using ultrafast bootstrap with 1,000 replicates (21).

To root the Gilliamella phylogenetic tree, we incorporated seven Orbaceae bacterial genomes available on NCBI RefSeq as outgroups, including four Frischella genomes, two Orbus genomes and one Schmidthempelia genome (Table S1). For the Snodgrassella phylogeny, we included one Neisseria meningitidis genome and one Kingella denitrificans genome as outgroups (Table S1). ggtree version 2.4.1 (22) was used to visualize the resulting trees.

To explore the gene content across Gilliamella and Snodgrassella strains, we performed principal components analysis on all the orthologous groups using pcomp in R version 4.0.3 (23). We also used “distance” function with “jaccard” distance in the R package phentropy version 0.5.0 (24) for calculating the distance based on gene presence and absence. We used R packages ape version 5.4.1 (25), vegan version 2.5.7 (26), and FigTree version 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/), last access July 19, 2021 for tree visualization.
Population delineation based on gene flow. To delineate populations in *Gilliamella* and *Snodgrassella*, we used PopCOGenT (27) to estimate the amount of recombination among strains. PopCOGenT measures homologous recombination by calculating the length distribution of identical sequences among all the pairwise whole genome alignments. Recombination is detected if the observed length distribution deviates from the null distribution where identical sequences in the genome were only affected by mutations. Genomes with less than 0.0355% divergence were collapsed into one group due to insufficient mutations to identify homologous recombination. Cytoscape version 3.8.2 (28) was used to visualize the gene flow networks.

We applied several approaches to validate the observed population boundaries. First, *Gilliamella* pop1 and pop6 showed low minimum gANI (gANI < 95% in Fig. 3). The lowest gANI within pop1 was between N-G2 and P83G (gANI = 92.6%). These two genomes were clustered together in the same main population cluster due to the gene flow between W8131 and M1-2G (gANI = 93.1%, length bias = 11.9 with 95% confidence interval from 10.6 to 13.3). The lowest gANI in pop6 was between Lep-s21 and Imp1-1 (gANI = 85.9%). These two genomes were clustered in the same main population cluster due to the gene flow between Choc5-1 and GillExp13 (gANI = 89.6%); Choc5-1 and Gris1-4 (gANI = 90.1%). To validate the finding of recombination among these relatively divergent genomes, we explored the long identical regions among these genomes. The lowest near-identical genomic regions between W8131 and M1-2G correspond to F-type H+-transporting ATPase subunits *atpAHFEB* (2,849 bp), translation elongation factor EF-Ts and 30S ribosomal protein S2 (1,193 bp). After removing the two scaffolds where the two identical regions reside, PopCOGenT did not detect sufficient recombination between W8131 and M1-2G, indicating that these identical regions were responsible for the detected recombination. Similarly, Choc5-1 and GillExp13 share a 7,200 bp near-identical region including genes related to histidine metabolism and amidohydrolase; Choc5-1 and Gris1-4 share a 5,736 bp identical region. These findings support that long identical regions lead to increased gene flow detected by PopCOGenT (27) and reflect real recombination even when gANI is as low as 86%.

Secondly, as genomes of pop2 branched among genomes of pop1 in the core gene phylogeny (Fig. 1), we investigated whether the recombination detected is due to artefacts. Based on length biases, the two genomes in pop2 (M6-3G and ESL0172) share high length bias (83.7 with 95% confidence interval from 71.2 to 102.0), whereas the second highest length bias between genomes in pop2 and pop1 (M6-3G and N-G2) is only 8.9 with 95% confidence interval from 8.1 to 10.0. In addition, the two genomes in pop2 were isolated, sequenced, and assembled independently from two different research groups at the University of Texas at Austin (29) and the University of Lausanne (30), so the observed recombination cannot reflect artefacts during sequencing or genome assembly.

We then confirmed the population delineation between pop1 and pop2 by investigating gene-specific sweeps in the core genome separating pop1 and pop2 using PopCOGenT (27). Gene-specific sweeps in the core genome were identified as low divergence alleles shared exclusively among members of each population, indicating population-specific adaptation. As pop2 only contains two genomes, we included two distant genomes ESL0177 and ESL0182 belonging to *Am. Gapis* clade as outgroups. We found extensive gene-specific sweeps separating pop1 and pop2 (Table S2). The same near-identical regions identified previously between W8131 and M1-2G, including F-type H+-transporting ATPase subunits *atpFEB*, translation elongation factor Ts and 30S ribosomal protein S2, were also identified as sweep regions in pop1. We built phylogenetic trees based on the near-identical regions containing atpFEB and its flanking regions. Resulting trees supported population specific-sweeps in the identical regions but not in the flanking regions (Fig. S11). Among the 131 genes undergoing sweeps, 15, 13, and 11 genes are related to coenzyme transport and metabolism, amino acid transport and metabolism, cell wall/membrane/envelope biogenesis, respectively. These sweeps likely reflect population-specific adaptation to ecological niche (Table S2).

Thirdly, we searched for gene-specific sweeps in flexible genomes of *Gilliamella* pop2 using PopCOGenT and found several important functional genes enriched in pop2. For example, genes...
related to antibacterial peptide secretion (cvaA and cvaB), fructose PTS system (fruA, fruAb, fruB), pectin degradation (kdgM, kdgR, and PL22), and peptide/nickel transport system substrate-binding proteins (ABC.PE.S, ABC.PE.P, and ddpF) are enriched in pop2 but not found in pop1 (Table S3).

To assess whether vertical descent substitutions could affect population delineation, we simulated genomes with the same divergence but without homologous recombination. We first built the phylogenetic tree based on the divergence estimated by PopCOGenT and BIONJ (31). Seq-Gen version 1.3.4 (32) with GTR model was used to simulate molecular sequences along the phylogenetic trees based on the genome length, nucleotide composition and average nucleotide diversity of Gilliamella wkB1 strain and Snodgrassella wkB2 strain.

Functional gene enrichment in different lineages. To investigate the variation of functional genes among different lineages, we used “anvi-get-enriched-functions-per-pan-group” command in anvi’o. A generalized linear model with the logit linkage function (3) was used to compute an enrichment score and p-value for each KEGG annotation and orthologous group. We summarized and visualized data using R version 4.0.3 (23), RStudio version 1.3.1093 (33) and Tidyverse version 1.3.1 (34). As similar genes were discovered by anvi’o KEGG enrichment analysis and PopCOGenT gene-specific sweeps in the flexible genome as described above (Table S2), we performed anvi’o KEGG enrichment analysis for each population of Gilliamella and Snodgrassella (Tables S4 and S5).

To explore essential genes and genes that facilitate host colonization identified from a previous transposon mutagenesis screen data of the Snodgrassella wkB2 strain (35), we associated the genes from the previous study to orthologous groups based on sequence similarity. Specifically, we downloaded the nucleotide sequences from all the essential and colonization-related genes. We then searched the essential and colonization-related gene sequences in the wkB2 genes annotated by anvi’o based on BLASTN version 2.6.0+ (16) with 100% identity and at least 88 bp alignment. Finally, we classified the orthologous groups that contained the wkB2 genes as essential or beneficial for colonization.

Urea transporters and urease gene phylogeny and synteny analysis. To understand the origins of the urea transporters and urease genes in Gilliamella, we searched the amino acid sequences of each gene in Gilliamella strain M6-3G to NCBI non-redundant protein database using NCBI BLASTP with default parameters (blast.ncbi.nlm.nih.gov, last accessed June 18, 2021). We built a phylogenetic tree for each gene based on the top 100 hits downloaded from NCBI.

Localization of Gilliamella populations. We confirmed the location of G. apis and G. apicola populations along with Snodgrassella using engineered strains expressing three different fluorescent proteins.

Bacterial culture. All cloning was performed using Escherichia coli strain DH5α, cultured in LB media and supplemented with Kanamycin (50 µg/mL) when appropriate. E. coli strain MFDpir was grown in LB media supplemented with 0.3 mM dianamopimelic acid (DAP) when appropriate. Bee gut bacteria used in this work (Table S6), were cultured as previously described (36). Briefly, frozen stocks were streaked onto Columbia blood agar and incubated at 35°C in an enriched 5% CO2 atmosphere. Plates were incubated for 48 hours or until visible colonies formed. To maintain plasmids in bee gut bacteria, agar plates were supplemented with Kanamycin (25 µg/mL).

Construction of plasmids. We designed three new plasmids for this work, and each expresses a different fluorescent protein from an identical promoter and on an identical plasmid backbone (Table S7). Plasmids were derived from plasmid pBTK402, a broad-host-range plasmid previously shown to replicate in multiple species of bee gut bacteria (37). This original plasmid is intended for use with a hierarchical Golden Gate assembly scheme as a “Type 8” origin part, and encodes Kanamycin resistance and a rfp reporter expressed at a low level. We replaced the
promoter and visible reporter via Hi-Fi Assembly (NEB #E2621S) to build the plasmids used in this work, resulting in three new plasmids: pSL1, pSL1-GFP, and pSL1-E2C. Each encodes a distinct reporter that can be differentiated by eye and by confocal fluorescence microscopy. These plasmids retain the original “Type 8” part overhangs of pBTK402, so they can still be used in the hierarchical assembly scheme for construction of broad-host-range plasmids.

**Construction of fluorescent bee gut bacteria.** We transformed plasmids pSL1, pSL1-GFP, and pSL1-E2C into *E. coli* strain MFDpir, which we then used to conjugate these plasmids into bee gut bacteria as previously described (37). Briefly, overnight cultures of donor and recipient bacteria were mixed in approximately equal ratios by optical density and then incubated overnight on a non-selective agar plate supplemented with DAP. The next day, we resuspended these mixtures in phosphate-buffered saline (PBS) and then plated dilutions of the conjugation mixture on selective media with Kanamycin (25 µg/mL) and without DAP. We passed visibly fluorescent colonies again on selective media to ensure stable transformation. We then made frozen stocks of these transconjugants and revived them for inoculation experiments.

**Honey bee husbandry and inoculation experiments.** Bees for inoculation experiments were sourced from three beehives owned by the Moran laboratory located on uncultivated pasture near Driftwood, TX and from beehives on a rooftop at the University of Texas, Austin. To obtain bees, we removed a brood frame from each hive and transported them to the laboratory. These frames were kept in an incubator at controlled temperature and humidity, and bees were allowed to emerge overnight. We collected these bees and randomly assigned to treatment groups that each received a different inoculum. Cohorts of twenty bees were briefly chilled and then exposed en masse to suspensions of bacterial inoculum mixed with a filter-sterilized 1:1 sucrose:water solution as described previously (38). Bees were then transferred to cup cages and given sterilized pollen and ad libitum access to sterilized sugar-water supplemented with Kanamycin (50 µg/mL), to ensure the plasmids were maintained in bacteria.

**Visualization of fluorescent bee gut bacteria.** After seven days, we immobilized bees with CO₂, placed them on ice, and then dissected out whole guts. We placed these guts on individual Ibidi µ-Dish 35 mm microscope dishes (CAT# 81156) and added 1 µL of PBS to prevent gut tissue from drying out. We imaged whole guts using a Zeiss 710 Laser Scanning Confocal Microscope with a 20× objective. All images were taken in “Lambda” mode in Zeiss software, which records a full spectral profile for each pixel. Using linear unmixing, each pixel was then assigned to a fluorescent marker (GFP, RCP, or E2-Crimson) or as background based on reference spectra captured using pure cultures of bacteria or uncolonized bees.

**Colonization experiment.** To measure the distribution of *G. apis* and *G. apicola* in bees with a conventional gut microbiome, we inoculated bees with gut homogenate from foragers. Bees were collected from brood frames from a single hive from beehives on a rooftop at University of Texas, Austin on August 9, 2021. Bees were allowed to emerge overnight and allocated into four cup cages (around 20 bees per cup cage) and fed sterile sucrose syrup and pollen. To inoculate bees with a conventional microbiome, bees were exposed to the gut homogenate from 12 foragers from outdoor hives. After seven days, 24 bees in total were collected and processed in the same way as the bees inoculated with engineered bacteria described above.

**Estimating Gilliamella abundance along the gut using qPCR.** To quantify the abundance of *G. apis* and *G. apicola* along the gut, we extracted DNA and performed qPCR on the pylorus and ileum by cutting in the middle of the ileum on 11 guts inoculated with *G. apis* and *G. apicola* and 24 guts with a conventional microbiome. We homogenized the pylorus and ileum parts with pestles and extracted DNA using the Qiagen DNEasy blood and tissue kit. We measured the abundance of *G. apis* by estimating the copy number of the *ureC* gene with primers designed in this study: *ureC* 282F (5'-AGCTATTGGTAAGCGGGAA-3') and *ureC* 345R (5'-TTCGCCGCAATAATTTAGTG-3'), yielding a product of 85bp. To control for the abundance of *Gilliamella* between the pylorus and ileum regions, we also measured the copy number of the 16S rRNA gene for *Gilliamella*: Gamma1-459-qF (5'-GTATCTAATAGGTGCATCAATT-3') and
These 16S primers amplify both *G. apis* and *G. apicola*. We performed qPCR using BioRad iTaq SYBR Green master mix on an Eppendorf Realplex Mastercycler and estimated the absolute copy number of *ureC* and 16S rRNA genes using standards ranging from $10^2$ to $10^8$ copies. To prepare standards, we amplified gene target regions, integrated them into a pGEM-T plasmid vector and extracted plasmids using the Monarch plasmid miniprep kit. We determined copy number based on the mass of extracted standards as determined by the Qubit HS dsDNA kit. The qPCR program we used included: 95 °C for 3 min, (95 °C for 5 s, 65 °C for 15 s, 68 °C for 20 s) $\times$ 5 cycles, and (95 °C for 5 s, 55 °C for 15 s, 68 °C for 20 s) $\times$ 35 cycles. We performed reactions in triplicate and used the average as the copy number.

To control for the abundance of *Gilliamella* between pylorus and ileum, we normalized the *ureC* copy number by dividing it by the *Gilliamella* 16S rRNA gene copy number. For the colonization experiment, we only included bees with successful *G. apis* or *G. apicola* colonization in the analysis. We identified the bee guts as successful colonization if the guts have a *ureC* gene copy number $\geq 1,000$ for ileum and have a 16S rRNA gene copy number $\geq 1,000$ for pylorus and ileum. We then performed Wilcoxon signed-rank test between pylorus and ileum samples using `wilcox.test` function in R.
Fig. S1. Gene-content trees of (A) *Gilliamella* and (B) *Snodgrassella* strains based on gene presence-absence. Red branches indicate strains from *Apis*. Blue branches represent strains from *Bombus*. Branches of outgroups are in black.
Fig. S2. Pairwise gANI among (A) Gillamella and (B) Snodgrassella strains. Blue indicates the gANI is >= 89% and red indicates the gANI is < 89%. Species are ordered based on the placement on the phylogenetic trees in Fig. 1. Populations and clades are colored as in Fig. 3.
**Fig. S3.** Number of sugar metabolism-related genes in *Gilliamella*. Species are ordered based on the placement on the phylogenetic trees in Fig. 1. Populations and clades are colored as in Fig. 3. Gene names in red indicate transporters. Gene names in bold indicate significant enrichment in Am_Gapis or Am_Gapicola.
**Fig. S4.** Distribution of urea transporters, urease genes, nitrite transporter and reductase genes in *Gilliameila*. Species are ordered based on the placement on the phylogenetic trees in Fig. 1. Populations and clades are colored as in Fig. 3. Gene names in bold indicate significant enrichment in Am_Gapis or Am_Gapiocola.
Fig. S5. Phylogenies of urea transporters and urease genes of *Gilliamella*, *Snodgrassella*, *Bartonella*, and other related proteins found on NCBI NR database. Branches in red indicate *Gilliamella* genes. Branches in blue indicate *Snodgrassella* and *Bartonella*-derived genes.
Fig. S6. Distribution of genes related to antagonism in (A) *Gilliamella* and (B) *Snodgrassella*. Species are ordered based on the placement on the phylogenetic trees in Fig. 1. Populations and clades are colored as in Fig. 3.
**Fig. S7.** CRISPR-Cas elements in *Gilliamella* populations. The CRISPR spacer network indicates the similarities of CRISPR spacers between genomes. The thickness of edges indicates the number of shared CRISPR spacers between genomes. Species are ordered based on the placement on the phylogenetic trees in Fig. 1. Populations and clades are colored as in Fig. 3.
Fig. S8. Number of (A) ureC gene copies, (B) 16S rRNA gene copies, and (C) normalized number of ureC copies based on 16S rRNA gene copies. Black lines connect pylorus and ileum from the same bee. (D) Ratio of normalized ureC copies between pylorus and ileum for 11 bees inoculated with engineered bacteria. Samples with ratio above 1 (red dotted line) have higher abundance of *G. apis* in pylorus than in ileum.
**Fig. S9.** Number of (A) _ureC_ gene copies, (B) 16S rRNA gene copies, and (C) normalized number of _ureC_ copies based on 16S rRNA gene copies between pylorus and ileum for 15 age-controlled bees with a conventional gut microbiome. Black lines connect pylorus and ileum from the same bee. Normalized ratios for pylorus to ileum are shown in Fig. 6F.
Fig. S10. Gene flow among simulated (A) Gilliamella and (B) Snodgrassella genomes. Genomes were simulated based on the genome-wide sequence divergence of Gilliamella and Snodgrassella genomes but without recombination. Simulated genomes are colored as in Fig. 3. No signal of gene flow was found among simulated genomes.
Fig. S11. Maximum likelihood trees of (A) gene-specific sweep region in pop1 (42,361-44,101 bp of W8131 genome scaffold 15), and its (B) left and (C) right flanking regions (30,209-42,360 bp and 44,102-66,064 of W8131 genome scaffold 15).
Table S1. Quality control of *Gilliamella* and *Snodgrassella* genomes using CheckM based on 182 Proteobacteria markers.

| Genome IDs          | NCBI accession number | Copy number of the 182 markers | Notes |
|---------------------|-----------------------|--------------------------------|-------|
| Gilliamella_A-1-24_Amellifera | ASM214151v1 | 0 181 1 0 0 0 |      |
| Gilliamella_A-12-12_Amellifera | ASM214167v1 | 0 181 1 0 0 0 |      |
| Gilliamella_A-2-24_Amellifera | ASM214156v1 | 0 181 1 0 0 0 |      |
| Gilliamella_A-4-12_Amellifera | ASM214186v1 | 0 181 1 0 0 0 |      |
| Gilliamella_A-7-12_Amellifera | ASM214165v1 | 0 181 1 0 0 0 |      |
| Gilliamella_A-7-24_Amellifera | ASM214166v1 | 0 181 1 0 0 0 |      |
| Gilliamella_A-8-12_Amellifera | ASM214171v1 | 0 181 1 0 0 0 |      |
| Gilliamella_A-9-12_Amellifera | ASM214174v1 | 0 181 1 0 0 0 |      |
| Gilliamella_ATSA1_Amellifera | ASM214203v1 | 0 181 1 0 0 0 |      |
| Gilliamella_ATSA2_Amellifera | ASM214205v1 | 0 181 1 0 0 0 |      |
| Gilliamella_ATSA3_Amellifera | ASM214206v1 | 0 181 1 0 0 0 |      |
| Gilliamella_ATSA4_Amellifera | ASM214213v1 | 0 181 1 0 0 0 |      |
| Gilliamella_A7_Amellifera | ASM214201v1 | 0 181 1 0 0 0 |      |
| Gilliamella_A8_Amellifera | ASM214155v1 | 0 181 1 0 0 0 |      |
| Gilliamella_A9_Amellifera | ASM214173v1 | 0 181 1 0 0 0 |      |
| Gilliamella_AM1_Amellifera | ASM214182v1 | 0 181 1 0 0 0 |      |
| Gilliamella_AM4_Amellifera | ASM214190v1 | 0 181 1 0 0 0 |      |
| Gilliamella_AM6_Amellifera | ASM214157v1 | 0 181 1 0 0 0 |      |
| Gilliamella_App2-1_Bappositus | ASM169025v1 | 0 182 0 0 0 0 |      |
| Gilliamella_App4-10_Bappositus | ASM169049v1 | 0 182 0 0 0 0 |      |
| Gilliamella_App6-5_Bappositus | ASM169052v1 | 1 180 1 0 0 0 |      |
| Gilliamella_Aw-17_Amellifera | ASM214159v1 | 0 181 1 0 0 0 |      |
| Gilliamella_AW11_Amellifera | ASM214181v1 | 1   | 180 | 1 | 0 | 0 | 0 |
|----------------------------|-------------|-----|-----|---|---|---|---|
| Gilliamella_AW13_Amellifera | ASM214175v1 | 0   | 181 | 1 | 0 | 0 | 0 |
| Gilliamella_B02_Amellifera  | sag_g1_b02_noCtrl | 69  | 108 | 3 | 1 | 1 | 0 |
| Gilliamella_Bif1-4_Bbifarius | ASM169053v1 | 0   | 181 | 1 | 0 | 0 | 0 |
| Gilliamella_Bim1-2_Bbimaculatus | ASM169083v1 | 0   | 182 | 0 | 0 | 0 | 0 |
| Gilliamella_Bim3-2_Bbimaculatus | ASM169057v1 | 0   | 151 | 3 | 0 | 0 | 0 |
| Gilliamella_Choc3-5_Bimpatiens | ASM169058v1 | 1   | 181 | 0 | 0 | 0 | 0 |
| Gilliamella_Choc4-2_Bimpatiens | ASM169035v1 | 1   | 181 | 0 | 0 | 0 | 0 |
| Gilliamella_Choc5-1_Bimpatiens | ASM169059v1 | 0   | 182 | 0 | 0 | 0 | 0 |
| Gilliamella_Choc6-1_Bimpatiens | ASM169038v1 | 1   | 181 | 0 | 0 | 0 | 0 |
| Gilliamella_DSM104097_Amellifera | ASM320145v1 | 0   | 181 | 1 | 0 | 0 | 0 |
| Gilliamella_ESL0169_Amellifera | ASM320287v1 | 1   | 180 | 1 | 0 | 0 | 0 |
| Gilliamella_ESL0172_Amellifera | ASM320277v1 | 0   | 181 | 1 | 0 | 0 | 0 |
| Gilliamella_ESL0177_Amellifera | ASM320291v1 | 1   | 180 | 1 | 0 | 0 | 0 |
| Gilliamella_ESL0178_Amellifera | ASM320281v1 | 0   | 181 | 1 | 0 | 0 | 0 |
| Gilliamella_ESL0182_Amellifera | ASM320265v1 | 1   | 180 | 1 | 0 | 0 | 0 |
| Gilliamella_ESL0232_Bterrestris | ASM1334688v1 | 0   | 182 | 0 | 0 | 0 | 0 |
| Gilliamella_ESL0250_Bpascuorum | ASM1334504v1 | 0   | 182 | 0 | 0 | 0 | 0 |
| Gilliamella_ESL0254_Bpascuorum | ASM1334498v1 | 0   | 182 | 0 | 0 | 0 | 0 |
| Gilliamella_ESL0405_Acerana | ASM1946920v1 | 0   | 182 | 0 | 0 | 0 | 0 |
| Gilliamella_ESL0441_Acerana | ASM1946918v1 | 0   | 182 | 0 | 0 | 0 | 0 |
| Gilliamella_ESL0443_Acerana | ASM1946916v1 | 1   | 181 | 0 | 0 | 0 | 0 |
| Gilliamella_Fer1-1_Bfervidus | ASM169082v1 | 1   | 181 | 0 | 0 | 0 | 0 |
| Gilliamella_Fer2-1_Bfervidus | ASM169030v1 | 30  | 136 | 1 | 6 | 0 | 0 |
| Gilliamella_Fer4-1_Bfervidus | ASM169051v1 | 1   | 181 | 0 | 0 | 0 | 0 |
| Gilliamella_GillExp13_Bimpatiens | ASM169081v1 | 0   | 182 | 0 | 0 | 0 | 0 |

*Removed as it was from single-cell sequencing*

*Contaminated, many CheckM genes have two copies*

*Low quality sequences identified by NCBI*
| Gilliamella_Gris1-4_Bgriseocollis | ASM169033v1 | 0 | 182 | 0 | 0 | 0 | 0 |
| Gilliamella_Gris3-2_Bgriseocollis | ASM169034v1 | 0 | 182 | 0 | 0 | 0 | 0 |
| Gilliamella_HK2_Bpensylvanicus | ASM169026v1 | 1 | 181 | 0 | 0 | 0 | 0 |
| Gilliamella_HK7_Bpensylvanicus | ASM169079v1 | 1 | 181 | 0 | 0 | 0 | 0 |
| Gilliamella_I20_Amellifera | sag_g1_i20_noCtrl | 46 | 132 | 4 | 0 | 0 | 0 |
| Gilliamella_Imp1-1_Bimpatiens | ASM169041v1 | 0 | 182 | 0 | 0 | 0 | 0 |
| Gilliamella_Imp1-6_Bimpatiens | ASM169042v1 | 30 | 152 | 0 | 0 | 0 | 0 |
| Gilliamella_Lep-s21_Blepidus | ASM979589v1 | 0 | 181 | 0 | 1 | 0 | 0 |
| Gilliamella_Lep-s35_Blepidus | ASM979584v1 | 0 | 181 | 0 | 1 | 0 | 0 |
| Gilliamella_Lep-s5_Blepidus | ASM979590v1 | 0 | 181 | 1 | 0 | 0 | 0 |
| Gilliamella_LMG28358_Blapidarius | IMG-taxon_2617270738_annotated_assembly | 0 | 182 | 0 | 0 | 0 | 0 |
| Gilliamella_LMG28359_Bpascuorum | IMG-taxon_2617270739_annotated_assembly | 0 | 182 | 0 | 0 | 0 | 0 |
| Gilliamella_LMG29879_Bterrestris | LMG29879 | 0 | 182 | 0 | 0 | 0 | 0 |
| Gilliamella_LMG29880_Bpascuorum | LMG29880 | 1 | 181 | 0 | 0 | 0 | 0 |
| Gilliamella_M1-2G_Amellifera | ASM169077v1 | 0 | 181 | 1 | 0 | 0 | 0 |
| Gilliamella_M6-3G_Amellifera | ASM172387v1 | 0 | 178 | 4 | 0 | 0 | 0 |
| Gilliamella_N-12-12_Amellifera | ASM208882v1 | 0 | 181 | 1 | 0 | 0 | 0 |
| Gilliamella_N-15-12_Amellifera | ASM208890v1 | 0 | 181 | 1 | 0 | 0 | 0 |
| Gilliamella_N-22_Amellifera | ASM214228v1 | 0 | 180 | 2 | 0 | 0 | 0 |
| Gilliamella_N-28_Amellifera | ASM214234v1 | 0 | 180 | 2 | 0 | 0 | 0 |
| Gilliamella_N-9-4_Amellifera | ASM208889v1 | 0 | 181 | 1 | 0 | 0 | 0 |
| Gilliamella_N-G1_Amellifera | ASM214218v1 | 0 | 180 | 2 | 0 | 0 | 0 |
| Gilliamella_N-G2_Amellifera | ASM214221v1 | 0 | 181 | 1 | 0 | 0 | 0 |
| Gilliamella_N-G3_Amellifera | ASM214222v1 | 0 | 180 | 2 | 0 | 0 | 0 |
| Gilliamella_N-G4_Amellifera | ASM214225v1 | 0 | 180 | 2 | 0 | 0 | 0 |
| Strain Name                      | Accession    | Ochre | Cyan | Blue | Red | Green |
|---------------------------------|--------------|-------|------|------|-----|-------|
| Gilliamella_N-G5_Amellifera     | ASM214226v1  | 0     | 181  | 1    | 0   | 0     |
| Gilliamella_N-W3_Amellifera     | ASM214227v1  | 0     | 181  | 1    | 0   | 0     |
| Gilliamella_N10_Amellifera      | ASM214178v1  | 0     | 181  | 1    | 0   | 0     |
| Gilliamella_N2_Amellifera       | ASM214185v1  | 0     | 181  | 1    | 0   | 0     |
| Gilliamella_N4_Amellifera       | ASM214163v1  | 0     | 181  | 1    | 0   | 0     |
| Gilliamella_N6_Amellifera       | ASM214188v1  | 0     | 181  | 1    | 0   | 0     |
| Gilliamella_Nev3-1_Bnevadensis  | ASM169043v1  | 1     | 181  | 0    | 0   | 0     |
| Gilliamella_Nev5-1_Bnevadensis  | ASM169089v1  | 1     | 181  | 0    | 0   | 0     |
| Gilliamella_Nev6-6_Bnevadensis  | ASM169060v1  | 1     | 180  | 1    | 0   | 0     |
| Gilliamella_NO1_Amellifera      | ASM214215v1  | 0     | 180  | 2    | 0   | 0     |
| Gilliamella_NO10_Amellifera     | ASM208884v1  | 0     | 180  | 2    | 0   | 0     |
| Gilliamella_NO12_Amellifera     | ASM214208v1  | 0     | 180  | 2    | 0   | 0     |
| Gilliamella_NO13_Amellifera     | ASM214197v1  | 0     | 179  | 3    | 0   | 0     |
| Gilliamella_NO14_Amellifera     | ASM214211v1  | 0     | 180  | 2    | 0   | 0     |
| Gilliamella_NO15_Amellifera     | ASM214193v1  | 0     | 180  | 2    | 0   | 0     |
| Gilliamella_NO16_Amellifera     | ASM214198v1  | 0     | 180  | 2    | 0   | 0     |
| Gilliamella_NO17_Amellifera     | ASM214215v1  | 0     | 180  | 2    | 0   | 0     |
| Gilliamella_NO3_Amellifera      | ASM214216v1  | 0     | 180  | 2    | 0   | 0     |
| Gilliamella_NO4_Amellifera      | ASM214216v1  | 0     | 180  | 2    | 0   | 0     |
| Gilliamella_NO5_Amellifera      | ASM208885v1  | 0     | 181  | 1    | 0   | 0     |
| Gilliamella_NO6_Amellifera      | ASM208879v1  | 0     | 181  | 1    | 0   | 0     |
| Gilliamella_NO8_Amellifera      | ASM208881v1  | 0     | 181  | 1    | 0   | 0     |
| Gilliamella_Occ3-1_Boccidentalis| ASM169065v1  | 0     | 182  | 0    | 0   | 0     |
| Gilliamella_Occ4-3_Boccidentalis| ASM169067v1  | 0     | 182  | 0    | 0   | 0     |
| Gilliamella_P17_Amellifera      | sag_g1_p17_noCtrl | 70  | 109  | 2    | 1   | 0     |
| Gilliamella_P46G_Amellifera     | ASM172600v1  | 0     | 24   | 9    | 6   | 3     |
| Gilliamella_P54G_Amellifera     | ASM169073v1  | 0     | 181  | 1    | 0   | 0     |

Removed as it was from single-cell sequencing
Removed due to large genome size
| Gene Name                  | Accession     | Start | End | Is ORF | ORF Length |
|---------------------------|---------------|-------|-----|--------|------------|
| Gilliamella_P62G_Amellifera | ASM169075v1   | 0     | 181 | 1      | 0          |
| Gilliamella_P83G_Amellifera | ASM169017v1   | 0     | 181 | 1      | 0          |
| Gilliamella_Pas-s25_Bpascuorum | ASM979588v1 | 0     | 182 | 0      | 0          |
| Gilliamella_Pas-s27_Bpascuorum | ASM979586v1 | 0     | 182 | 0      | 0          |
| Gilliamella_Pas-s95_Bpascuorum | ASM979579v1 | 0     | 182 | 0      | 0          |
| Gilliamella_Pra-s52_Bpratorum | ASM979580v1 | 0     | 182 | 0      | 0          |
| Gilliamella_Pra-s54_Bpratorum | ASM979578v1 | 0     | 182 | 0      | 0          |
| Gilliamella_Pra-s60_Bpratorum | ASM979575v1 | 0     | 182 | 0      | 0          |
| Gilliamella_Pra-s65_Bpratorum | ASM979574v1 | 0     | 182 | 0      | 0          |
| Gilliamella_W8127_Amellifera | ASM755914v1   | 0     | 181 | 1      | 0          |
| Gilliamella_W8131_Amellifera | ASM755916v1   | 0     | 181 | 1      | 0          |
| Gilliamella_WF3-4_Bpensylvanicus | ASM169027v1 | 1     | 181 | 0      | 0          |
| Gilliamella_wkB1_Amellifera | ASM599988v1   | 0     | 181 | 1      | 0          |
| Gilliamella_wkB108_Adorsata | ASM169019v1   | 0     | 182 | 0      | 0          |
| Gilliamella_wkB11_Bbimaculatus | GilApiwkB11v1.0 | 1     | 181 | 0      | 0          |
| Gilliamella_wkB112_Adorsata | ASM169068v1   | 0     | 182 | 0      | 0          |
| Gilliamella_wkB171_Aandreniformis | ASM169375v1 | 0     | 182 | 0      | 0          |
| Gilliamella_wkB178_Adorsata | ASM169070v1   | 0     | 182 | 0      | 0          |
| Gilliamella_wkB18_Bvagans | ASM169087v1   | 1     | 181 | 0      | 0          |
| Gilliamella_wkB195_Acerana | ASM169044v1   | 1     | 181 | 0      | 0          |
| Gilliamella_wkB292_Acerana | ASM169023v1   | 0     | 180 | 2      | 0          |
| Gilliamella_wkB30_Bvagans | GilApiwkB30v1.0 | 0     | 182 | 0      | 0          |
| Gilliamella_wkB308_Acerana | ASM169018v1   | 0     | 182 | 0      | 0          |
| Gilliamella_wkB7_Amellifera | ASM169343v1   | 0     | 181 | 1      | 0          |
| Gilliamella_wkB72_Acerana | ASM169374v1   | 1     | 181 | 0      | 0          |
| Snodgrassella_A-1-12_Amellifera | ASM208873v1 | 0     | 181 | 1      | 0          |
| Snodgrassella_A-10-12_Amellifera | ASM208840v1 | 0     | 181 | 1      | 0          |
| Strain Name                      | Assembly | Contigs | GC% | N50 | L50 | Removed due to large genome size |
|----------------------------------|----------|---------|-----|-----|-----|----------------------------------|
| Snodgrassella_A11-12_Amellifera | ASM208869v1 | 0 | 181 | 1 | 0 | 0 |
| Snodgrassella_A2-12_Amellifera  | ASM208858v1 | 0 | 181 | 1 | 0 | 0 |
| Snodgrassella_A5-24_Amellifera  | ASM208846v1 | 0 | 181 | 1 | 0 | 0 |
| Snodgrassella_A-9-24_Amellifera | ASM208851v1 | 0 | 181 | 1 | 0 | 0 |
| Snodgrassella_A11_Amellifera    | ASM208901v1 | 0 | 181 | 1 | 0 | 0 |
| Snodgrassella_A12_Amellifera    | ASM208867v1 | 0 | 181 | 1 | 0 | 0 |
| Snodgrassella_A2_Amellifera     | ASM208839v1 | 0 | 181 | 1 | 0 | 0 |
| Snodgrassella_A3_Amellifera     | ASM208845v1 | 0 | 181 | 1 | 0 | 0 |
| Snodgrassella_A5_Amellifera     | ASM208847v1 | 0 | 181 | 1 | 0 | 0 |
| Snodgrassella_App2-2_Bappositus | ASM277731v1 | 0 | 182 | 0 | 0 | 0 |
| Snodgrassella_App4-8_Bappositus | ASM277742v1 | 0 | 182 | 0 | 0 | 0 |
| Snodgrassella_App6-4_Bappositus | ASM277746v1 | 0 | 182 | 0 | 0 | 0 |
| Snodgrassella_Aw-18_Amellifera  | ASM208841v1 | 0 | 181 | 1 | 0 | 0 |
| Snodgrassella_Aw-20_Amellifera  | ASM201941v1 | 0 | 181 | 1 | 0 | 0 |
| Snodgrassella_E1_Amellifera     | ASM240664v1 | 3 | 178 | 1 | 0 | 0 |
| Snodgrassella_ESL0196_Amellifera| ASM320288v1 | 0 | 181 | 1 | 0 | 0 |
| Snodgrassella_ESL0253_Bpascuorum| ASM1334496v1 | 0 | 180 | 2 | 0 | 0 |
| Snodgrassella_ESL0304_Amellifera| ASM1334499v1 | 0 | 181 | 1 | 0 | 0 |
| Snodgrassella_ESL0323_Amellifera| ASM1334686v1 | 0 | 181 | 1 | 0 | 0 |
| Snodgrassella_ESL0324_Amellifera| ASM1334495v1 | 0 | 181 | 1 | 0 | 0 |
| Snodgrassella_Fer1-2_Bfervidus  | ASM277741v1 | 0 | 180 | 2 | 0 | 0 |
| Snodgrassella_Fer2-2_Bfervidus  | ASM277748v1 | 0 | 179 | 3 | 0 | 0 |
| Snodgrassella_Fer4-2_Bfervidus  | ASM277749v1 | 0 | 170 | 0 | 9 | 3 |
| Snodgrassella_Gris1-3_Bgriotlepis| ASM277752v1 | 0 | 177 | 5 | 0 | 0 |
| Snodgrassella_Gris1-6_Bgriotlepis| ASM277761v1 | 0 | 180 | 2 | 0 | 0 |
| Snodgrassella_Gris2-3-4_Bgriotlepis| ASM277733v1 | 0 | 180 | 2 | 0 | 0 |
| Snodgrassella_Gris3-4_Bgriotlepis| ASM277759v1 | 0 | 180 | 2 | 0 | 0 |
| Name                                           | Accession   | Total | BUSCOs | Contaminated | Notes                                                                 |
|------------------------------------------------|-------------|-------|--------|--------------|----------------------------------------------------------------------|
| Snodgrassella_HK3_Bpensylvanicus               | ASM277770v1 | 0     | 180    | 2            |                                                                      |
| Snodgrassella_HK9x_Bpensylvanicus              | ASM277779v1 | 0     | 179    | 3            |                                                                      |
| Snodgrassella_j21_Amellifera                   | sag_j21_noCtrl | 16   | 163    | 3            | Removed as it was from single-cell sequencing                         |
| Snodgrassella_MS1-3_Amellifera                 | ASM277792v1 | 0     | 181    | 1            |                                                                      |
| Snodgrassella_N-23_Amellifera                  | ASM208865v1 | 0     | 181    | 1            |                                                                      |
| Snodgrassella_N-S1_Amellifera                  | ASM208875v1 | 0     | 181    | 1            |                                                                      |
| Snodgrassella_N-S2_Amellifera                  | ASM208859v1 | 0     | 181    | 1            |                                                                      |
| Snodgrassella_N-S3_Amellifera                  | ASM208863v1 | 0     | 181    | 1            |                                                                      |
| Snodgrassella_N-S4_Amellifera                  | ASM208866v1 | 0     | 181    | 1            |                                                                      |
| Snodgrassella_N-S5_Amellifera                  | ASM208876v1 | 0     | 181    | 1            |                                                                      |
| Snodgrassella_N-W4_Amellifera                  | ASM208852v1 | 0     | 181    | 1            |                                                                      |
| Snodgrassella_N-W7_Amellifera                  | ASM208855v1 | 0     | 181    | 1            |                                                                      |
| Snodgrassella_N9_Amellifera                    | ASM208857v1 | 0     | 181    | 1            |                                                                      |
| Snodgrassella_Nev3CBA3_Bnevadensis             | ASM277757v1 | 0     | 175    | 7            |                                                                      |
| Snodgrassella_Nev4-2_Bnevadensis               | ASM277782v1 | 12    | 167    | 3            |                                                                      |
| Snodgrassella_O02_Amellifera                   | sag_o02_noCtrl | 50  | 120    | 2            | Removed as it was from single-cell sequencing                         |
| Snodgrassella_O11_Amellifera                   | sag_o11_noCtrl | 80  | 100    | 2            | Removed as it was from single-cell sequencing                         |
| Snodgrassella_Occ4-2_Boccidentalis             | ASM277763v1 | 0     | 180    | 2            |                                                                      |
| Snodgrassella_P14_Amellifera                   | sag_p14_noCtrl | 109 | 73     | 0            | Removed as it was from single-cell sequencing                         |
| Snodgrassella_PEB0171_Amellifera               | ASM277786v1 | 0     | 180    | 2            |                                                                      |
| Snodgrassella_PEB0178_Amellifera               | ASM277787v1 | 0     | 179    | 3            |                                                                      |
| Snodgrassella_Pens2-2-5_Bpensylvanicus         | ASM277732v1 | 0     | 180    | 2            |                                                                      |
| Snodgrassella_Ruf1-X_Brufocinctus              | ASM277767v1 | 0     | 168    | 2            | Contaminated, many BUSCOs/CheckM genes have two copies               |
| Snodgrassella_Snod2-1-5_Bimpatiens             | ASM277734v1 | 0     | 180    | 2            |                                                                      |
| Snodgrassella_WF3-3_Bpensylvanicus             | ASM277774v1 | 0     | 180    | 2            |                                                                      |
| Snodgrassella_wkB12_Bsp                        | SnoalwkB12v1.0 | 0   | 180    | 2            |                                                                      |
| Name                          | Accession   | Support | tColor | tAmino | sColor | sAmino |
|-------------------------------|-------------|---------|--------|--------|--------|--------|
| Snodgrassella_wkB2_Amellifera | ASM60000v1  | 181     | 1      | 0      | 0      | 0      |
| Snodgrassella_wkB237_Aandreniformis | ASM277777v1 | 182     | 0      | 0      | 0      | 0      |
| Snodgrassella_wkB273_Aflorea  | ASM277765v1 | 182     | 0      | 0      | 0      | 0      |
| Snodgrassella_wkB29_Bsp      | SnaolvkB29v1.0 | 180     | 2      | 0      | 0      | 0      |
| Snodgrassella_wkB298_Acerana | ASM277785v1 | 181     | 1      | 0      | 0      | 0      |
| Snodgrassella_wkB332_Amellifera | ASM277769v1 | 181     | 1      | 0      | 0      | 0      |
| Snodgrassella_wkB339_Amellifera | ASM277781v1 | 181     | 1      | 0      | 0      | 0      |
| Snodgrassella_wkB9_Amellifera | ASM277773v1 | 179     | 3      | 0      | 0      | 0      |
| Frischella_Ac13_Acerana      | GCF_014489845.1 | 181     | 0      | 0      | 0      | 0      |
| Frischella_DSM104328_Amellifera | GCF_003182045.1 | 180     | 1      | 0      | 0      | 0      |
| Frischella_ESL0167_Amellifera | GCF_003202705.1 | 180     | 1      | 0      | 0      | 0      |
| Frischella_PEB0191_Amellifera | GCF_000807275.1 | 180     | 1      | 0      | 0      | 0      |
| Orbus_hercynius_Sscrofa      | GCF_003634275.1 | 182     | 0      | 0      | 0      | 0      |
| Orbus_IPMB12_Zatratus        | GCF_011745665.1 | 182     | 0      | 0      | 0      | 0      |
| Schmidhempelia_bombi_Bimpatiens | GCF_000471645.3 | 180     | 1      | 0      | 0      | 0      |
| Kingella_denitrificans       | GCF_000190695.1 | 175     | 0      | 0      | 0      | 0      |
| Neisseria_meningitidis       | GCF_000008805.1 | 180     | 1      | 0      | 0      | 0      |
Table S2. Gene-specific sweeps in core genome of Gilliamella pop1. Location of sweep regions and gene IDs were based on Gilliamella pop1 W8131 genome.

| Scaffold ID | Start of the sweep region | End of the sweep region | W8131 gene ID | Orthologous group ID | KEGG ID | KEGG functions | COG | COG category |
|-------------|---------------------------|-------------------------|----------------|----------------------|---------|----------------|-----|--------------|
| 9           | 671,088                   | 669,070                 | 689            | GC_00000743          |         | pta; protease III [EC:3.4.24.55] | COG1396 | Transcription |
| 9           | 671,088                   | 669,070                 | 690            | GC_00000320          | K01407  | yggT; YggT family protein | COG1025 | Posttranslational modification, protein turnover, chaperones |
| 9           | 653,576                   | 652,846                 | 671            | GC_00000696          | K02221  | K09160; uncharacterized protein | COG0762 | Posttranslational modification, protein turnover, chaperones |
| 9           | 653,576                   | 652,846                 | 672            | GC_00000700          | K09160  | yggS, PROSC; PLP dependent protein | COG2983 | General function prediction only |
| 9           | 652,179                   | 651,005                 | 669            | GC_00000748          | K02669  | PilT; twitching motility protein PilT | COG2805 | Cell motility/Extracellular structures |
| 9           | 652,179                   | 651,005                 | 670            | GC_00000219          | K06997  | TC.MATE, SLC47A, norM, mdk, dinF; multidrug resistance protein, MATE family nikE, cntF; nickel transport system nikD, cntD; nickel transport system | COG0325 | Coenzyme transport and metabolism |
| 9           | 166,126                   | 166,963                 | 230            | GC_00001566          | K03327  | bglF, bglP; beta-glucoside PTS system EIICBA component | COG0534 | Defense mechanisms |
| 9           | 166,126                   | 166,963                 | 231            | GC_00001587          | K10824  | ATP-binding protein [EC:7.2.2.11] | COG1123 | Posttranslational modification, protein turnover, chaperones |
| 9           | 167,290                   | 167,841                 | 232            | GC_00001588          | K15587  | ATP-binding protein [EC:7.2.2.11] | COG0444 | Amino acid transport and metabolism/Inorganic ion transport and metabolism |
| 9           | 602,017                   | 601,351                 | 629            | GC_00000619          | K02757  | bglF, bglP; beta-glucoside PTS system EIICBA component | COG1263 | Carbohydrate transport and metabolism |
| Gene ID   | Start (bp) | End (bp) | Length (bp) | Description                                                                                           | COG ID   | Functional Category                               |
|-----------|------------|----------|-------------|-------------------------------------------------------------------------------------------------------|----------|--------------------------------------------------|
| 9 600,586 | 599,979    | 628      | GC_00001273 | K03489 yydK; GntR family transcriptional regulator, transcriptional regulator of bgfA dppB; 4-hydroxy-tetrahydrodipicolinate reductase [EC:1.17.1.8] tdaA; 4-hydroxytetrahydrodipicolinate reductase [EC:1.17.1.8] tdaA; 4-hydroxy-tetrahydrodipicolinate reductase [EC:1.17.1.8] tdaA; 4-hydroxy-tetrahydrodipicolinate reductase [EC:1.17.1.8] tdaA| COG2188 | Transcription                                     |
| 9 503,725 | 502,803    | 544      | GC_00000731 | K00215 tRNA(adenine34) deaminase [EC:3.5.4.33] hprT, hpt, HPRT1; hypoxanthine phosphoribosyltransferase [EC:2.4.2.8] arcD, lysl, lysP; arginine-ornithine antiporter / lysine permease yliC, trmX; tRNA1Val (adenosine37-N6)-methyltransferase [EC:2.1.1.223] K06915; uncharacterized protein hemA; glutamyl-tRNA reductase [EC:1.2.1.70]| COG0289 | Amino acid transport and metabolism                |
| 9 503,725 | 502,803    | 545      | GC_00000575 | K11991 tRNA(adenine34) deaminase [EC:3.5.4.33] hprT, hpt, HPRT1; hypoxanthine phosphoribosyltransferase [EC:2.4.2.8] arcD, lysl, lysP; arginine-ornithine antiporter / lysine permease yliC, trmX; tRNA1Val (adenosine37-N6)-methyltransferase [EC:2.1.1.223] K06915; uncharacterized protein hemA; glutamyl-tRNA reductase [EC:1.2.1.70]| COG0590 | Translation, ribosomal structure and biogenesis   |
| 9 503,725 | 502,803    | 546      | GC_00000499 | K00760 tRNA(adenine34) deaminase [EC:3.5.4.33] hprT, hpt, HPRT1; hypoxanthine phosphoribosyltransferase [EC:2.4.2.8] arcD, lysl, lysP; arginine-ornithine antiporter / lysine permease yliC, trmX; tRNA1Val (adenosine37-N6)-methyltransferase [EC:2.1.1.223] K06915; uncharacterized protein hemA; glutamyl-tRNA reductase [EC:1.2.1.70]| COG0634 | Nucleotide transport and metabolism               |
| 9 332,993 | 333,750    | 376      | GC_00001268 | K03758 tRNA(adenine34) deaminase [EC:3.5.4.33] hprT, hpt, HPRT1; hypoxanthine phosphoribosyltransferase [EC:2.4.2.8] arcD, lysl, lysP; arginine-ornithine antiporter / lysine permease yliC, trmX; tRNA1Val (adenosine37-N6)-methyltransferase [EC:2.1.1.223] K06915; uncharacterized protein hemA; glutamyl-tRNA reductase [EC:1.2.1.70]| COG0531 | Amino acid transport and metabolism               |
| 9 332,993 | 333,750    | 377      | GC_00000977 | K15460 tRNA(adenine34) deaminase [EC:3.5.4.33] hprT, hpt, HPRT1; hypoxanthine phosphoribosyltransferase [EC:2.4.2.8] arcD, lysl, lysP; arginine-ornithine antiporter / lysine permease yliC, trmX; tRNA1Val (adenosine37-N6)-methyltransferase [EC:2.1.1.223] K06915; uncharacterized protein hemA; glutamyl-tRNA reductase [EC:1.2.1.70]| COG4123 | Translation, ribosomal structure and biogenesis   |
| 9 445,401 | 444,617    | 481      | GC_00001121 | K06915 tRNA(adenine34) deaminase [EC:3.5.4.33] hprT, hpt, HPRT1; hypoxanthine phosphoribosyltransferase [EC:2.4.2.8] arcD, lysl, lysP; arginine-ornithine antiporter / lysine permease yliC, trmX; tRNA1Val (adenosine37-N6)-methyltransferase [EC:2.1.1.223] K06915; uncharacterized protein hemA; glutamyl-tRNA reductase [EC:1.2.1.70]| COG0433 | Replication, recombination and repair              |
| 9 323,933 | 321,816    | 366      | GC_00000995 | K02492 tRNA(adenine34) deaminase [EC:3.5.4.33] hprT, hpt, HPRT1; hypoxanthine phosphoribosyltransferase [EC:2.4.2.8] arcD, lysl, lysP; arginine-ornithine antiporter / lysine permease yliC, trmX; tRNA1Val (adenosine37-N6)-methyltransferase [EC:2.1.1.223] K06915; uncharacterized protein hemA; glutamyl-tRNA reductase [EC:1.2.1.70]| COG0373 | Coenzyme transport and metabolism                |
| 9 323,933 | 321,816    | 367      | GC_00001233 | K02494 tRNA(adenine34) deaminase [EC:3.5.4.33] hprT, hpt, HPRT1; hypoxanthine phosphoribosyltransferase [EC:2.4.2.8] arcD, lysl, lysP; arginine-ornithine antiporter / lysine permease yliC, trmX; tRNA1Val (adenosine37-N6)-methyltransferase [EC:2.1.1.223] K06915; uncharacterized protein hemA; glutamyl-tRNA reductase [EC:1.2.1.70]| COG3017 | Cell wall/membrane/envelope biogenesis            |
| 9 323,933 | 321,816    | 368      | GC_00000911 | K00919 tRNA(adenine34) deaminase [EC:3.5.4.33] hprT, hpt, HPRT1; hypoxanthine phosphoribosyltransferase [EC:2.4.2.8] arcD, lysl, lysP; arginine-ornithine antiporter / lysine permease yliC, trmX; tRNA1Val (adenosine37-N6)-methyltransferase [EC:2.1.1.223] K06915; uncharacterized protein hemA; glutamyl-tRNA reductase [EC:1.2.1.70]| COG1947 | Lipid transport and metabolism                   |
| 9 621,042 | 621,547    | 643      | GC_00000803 | K07053 tRNA(adenine34) deaminase [EC:3.5.4.33] hprT, hpt, HPRT1; hypoxanthine phosphoribosyltransferase [EC:2.4.2.8] arcD, lysl, lysP; arginine-ornithine antiporter / lysine permease yliC, trmX; tRNA1Val (adenosine37-N6)-methyltransferase [EC:2.1.1.223] K06915; uncharacterized protein hemA; glutamyl-tRNA reductase [EC:1.2.1.70]| COG0613 | Nucleotide transport and metabolism              |
| Gene | Start | End | Length | Description | COG | Function |
|------|-------|-----|--------|-------------|-----|----------|
| GC_00001107 | 696,242 | 697,671 | 711 | mscM, bspA; mechanosensitive channel xapB, MFS transporter, NHS family, xanthosine permease | K22051 | Cell wall/membrane/envelope biogenesis |
| GC_00000019 | 701,791 | 702,956 | 715 | xapB; MFS transporter, NHS family, xanthosine permease | K11537 | Carbohydrate transport and metabolism |
| GC_00001224 | 746,349 | 747,931 | 754 | tyR; transcriptional regulator of aroF, aroG, tyrA and aromatic amino acid transport nagZ, beta-N-acetylhexosaminidase [EC:3.2.1.52] | K03721 | Transcription|Amino acid transport and metabolism |
| GC_00000287 | 597,384 | 596,584 | 851 | lauroyltransferase/acyltransferase [EC:2.3.1.241 2.3.1.-] | K02517 | Lipid transport and metabolism |
| GC_00000935 | 48,148 | 49,658 | 852 | OmpR family, sensor histidine kinase BasS [EC:2.7.13.3] | K07643 | Signal transduction mechanisms |
| GC_00001426 | 48,148 | 49,658 | 851 | OmpR family, response regulator BasR | K07771 | Signal transduction mechanisms|Transcription |
| GC_00002406 | 55,396 | 57,063 | 859 | pdxA2; 4-phospho-D-threonate 3-dehydrogenase / 4-phospho-D-erythronate 3-dehydrogenase [EC:1.1.1.408 1.1.1.409] | K22024 | Coenzyme transport and metabolism |
| Accession | Start | End | Length | Description |
|-----------|-------|-----|--------|-------------|
| GC_00002412 | 55,396 | 57,063 | 860 | COG1349 | Transcription\|Carbohydrate transport and metabolism |
| GC_00001236 | 57,548 | 58,310 | 862 | COG4953 | Cell wall/membrane/envelope biogenesis |
| GC_00001178 | 61,438 | 62,035 | 862 | COG2373 | General function prediction only |
| GC_00001146 | 129,034 | 132,825 | 921 | COG0295 | Nucleotide transport and metabolism |
| GC_00001479 | 129,034 | 132,825 | 923 | COG0737 | Nucleotide transport and metabolism | Defense mechanisms |
| GC_00001277 | 128,767 | 128,965 | 920 | COG1882 | Energy production and conversion |
| GC_00000318 | 321,684 | 322,939 | 1088 | COG0661 | Coenzyme transport and metabolism | Signal transduction mechanisms |
| GC_00000513 | 321,684 | 322,939 | 1089 | COG2226 | Coenzyme transport and metabolism |
| ID  | Gene1       | Gene2       | Description                                                                 |
|-----|-------------|-------------|------------------------------------------------------------------------------|
| 12  | GC_00001014 | GC_00001021 | benzoquinol methylase [EC:2.1.1.163 2.1.1.201] ispA; farnesyl diphosphate    |
|     |             |             | synthase [EC:2.5.1.1 2.5.1.10] ddxs; 1-deoxy-D-xyulose-5-phosphate synthase [EC:2.2.1.7] rmuC; DNA recombination protein|
|     |             |             | rmuC: 23S rRNA (uracil1939-CS)-methyltransferase [EC:2.1.1.190] znuA; zinc |
|     |             |             | transport system substrate-binding protein mepM; murein DD-endopeptidase [EC:3.4.24.-] DARS2, aspS; asparyl-tRNA synthetase [EC:6.1.1.12] nubB, ntpA; dihydroneopterin triphosphate diphosphatase [EC:3.6.1.67] |
ATP-binding protein

clpX, CLPX; ATP-dependent Clp protease ATP-binding subunit ClpX

lon; ATP-dependent Lon protease [EC:3.4.21.53]

mukB; chromosome partition protein MukB

hemB, ALAD; porphobilinogen synthase [EC:4.2.1.24]

RP-S21, MRPS21, rpsU; small subunit ribosomal protein S21

mraY; phospho-N-acetylmuramoyl-pentapeptide-transferase [EC:2.7.8.13]

murD; UDP-N-acetylmuramoyl-alanine–D-glutamate ligase [EC:6.3.2.9]

ftsW, spoVE; cell division protein FtsW

murG; UDP-N-acetylgulcosamine-N-acetylmuramyl-pentapeptide pyrophosphoryl-undecaprenol N-acetylgulcosamine transferase [EC:2.4.1.227]

biosynthesis, transport and catabolism

Posttranslational modification, protein turnover, chaperones

Posttranslational modification, protein turnover, chaperones

Cell cycle control, cell division, chromosome partitioning

Coenzyme transport and metabolism

Translation, ribosomal structure and biogenesis

Cell wall/membrane/envelope biogenesis

Cell wall/membrane/envelope biogenesis

Cell cycle control, cell division, chromosome partitioning

Cell wall/membrane/envelope biogenesis
| Rank | Start | End | Genomic Position | Accession | Description | COG | Domain |
|------|-------|-----|------------------|-----------|-------------|-----|--------|
| 14   | 323,903 | 325,293 | 1646 | GC_00001176 | TerC; tellurite resistance protein | K05794 | COG0861 | Inorganic ion transport and metabolism |
| 14   | 323,903 | 325,293 | 1647 | GC_00000278 | dacB; serine-type D-Ala-D-Ala carboxypeptidase/endopeptidase (penicillin-binding protein 4) | K07259 | COG2027 | Cell wall/membrane/envelope biogenesis |
| 14   | 456,001 | 457,017 | 1769 | GC_00000988 | exodeoxyribonuclease V alpha subunit | K03581 | COG0507 | Replication, recombination and repair |
| 15   | 110    | 1,859  | 1,859 | GC_00001281 | recD; exodeoxyribonuclease V alpha subunit [EC:3.1.11.5] | K06181 | COG1187 | Translation, ribosomal structure and biogenesis |
| 15   | 110    | 1,859  | 1,859 | GC_00000242 | ftsY; fused signal recognition particle receptor | K03110 | COG0552 | Intracellular trafficking, secretion, and vesicular transport |
| 15   | 110    | 1,859  | 1,859 | GC_00000406 | ftsE; cell division transport system ATP-binding protein | K09812 | COG2884 | Cell cycle control, cell division, chromosome partitioning |
| 15   | 2,523  | 3,152  | 3,152 | GC_00001022 | ftsX; cell division transport system permease protein | K09811 | COG2177 | Cell cycle control, cell division, chromosome partitioning |
| 15   | 3,353  | 4,283  | 4,283 | GC_00000436 | rpoH; RNA polymerase sigma-32 factor | K03089 | COG0568 | Transcription |
| 15   | 7,012  | 7,664  | 7,664 | GC_00001125 | npdA; NAD-dependent deacetylase [EC:2.3.1.286] | K12410 | COG0846 | Posttranslational modification, protein turnover, chaperones |
| 15   | 7,012  | 7,664  | 7,664 | GC_00001388 | mtgA; monofunctional glycosyltransferase [EC:2.4.1.129] | K03814 | COG0744 | Cell wall/membrane/envelope biogenesis |
| 15   | 49,264 | 49,789 | 1836 | GC_00001090 | bioY; biotin transport system substrate-specific component | K03523 | COG1268 | Coenzyme transport and metabolism |
| Gene ID     | Transcript ID | Description                                                                 | COG ID | Categories                                      |
|------------|---------------|------------------------------------------------------------------------------|--------|-----------------------------------------------|
| K02109     | GC_00000555   | ATPF0B, atpF; F-type H+-transporting ATPase subunit b                        | COG0711| Energy production and conversion              |
| K02110     | GC_00000341   | ATPF0C, atpE; F-type H+-transporting ATPase subunit c                        | COG0636| Energy production and conversion              |
| K02108     | GC_00000427   | ATPF0A, atpB; F-type H+-transporting ATPase subunit a                        | COG0356| Energy production and conversion              |
| K03431     | GC_00000272   | phosphoglucoisomerase mutase [EC:5.4.2.1]                                   | COG1109| Carbohydrate transport and metabolism         |
| K00868     | GC_0001315    | pdxK, pdxY; pyridoxine kinase [EC:2.7.1.35]                                  | COG2240| Coenzyme transport and metabolism             |
| K15539     | GC_0002356    | rodZ; cytoskeleton protein RodZ gcpE, ispG; (E)-4-hydroxy-3-methylbut-2-enyl-
|            |               | diphosphate synthase [EC:1.17.7.1 1.17.7.3]                                  | COG1426| Cell cycle control, cell division, chromosome partitioning |
| K03526     | GC_0000465    | recC; recombination associated protein RecC rdpC                             | COG0821| Lipid transport and metabolism                |
| K03554     | GC_0000045    | recA, recB, recC; RecA family protein                                       | COG2974| Replication, recombination and repair         |
| K07274     | GC_00001154   | mipA, ompV; MipA family protein                                              | COG3713| Cell wall/membrane/envelope biogenesis        |
| K07250     | GC_00001106   | gabT; 4-aminobutyrate aminotransferase / (S)-3-amino-2-methylpropionate
|            |               | transaminase / S-5-aminovalerate transaminase [EC:2.6.1.19
|            |               | 2.6.1.22 2.6.1.48]                                                            | COG0160| Amino acid transport and metabolism           |
| ORG | START | STOP | YEAR | GenBank | K | Description |
|-----|-------|------|------|---------|---|-------------|
| 15  | 204,600 | 205,321 | 1991 | GC_00001157 | K03928 | yvaK; carboxylesterase [EC:3.1.1.1] |
| 15  | 204,600 | 205,321 | 1991 | GC_00001157 | K03928 | mtnX; 2-hydroxy-3-keto-5-methylthiopentyl-1-phosphate phosphatase [EC:3.1.3.87] |
| 15  | 214,698 | 215,477 | 1997 | GC_00000196 | K01626 | E2.5.1.54, aroF, aroG, aroH; 3-deoxy-7-phosphoheptulonate synthase [EC:2.5.1.54] |
| 15  | 215,737 | 216,993 | 1998 | GC_00000596 | K01735 | aroD; 3-dehydroquinate synthase [EC:4.2.3.4] |
| 15  | 215,737 | 216,993 | 1999 | GC_00001042 | K01950 | E6.3.5.1, NADSYN1, QNS1, nadE; NAD+ synthase (glutamine-hydrolysing) [EC:6.3.5.1] |
| 15  | 229,491 | 230,577 | 2011 | GC_00001275 | K08289 | purT; phosphoribosylglycinamid FORMYLTRANSFERASE 2 [EC:2.1.2.2] |
| 15  | 229,491 | 230,577 | 2012 | GC_00001572 | K04087 | hflC; modulator of FtsH protease HflC |
| 15  | 230,628 | 231,693 | 2013 | GC_00001510 | K04088 | hflK; modulator of FtsH protease HflK |
| 16  | 225 | 772 | 2020 | GC_00001404 | K04088 | flgE; flagellar hook protein FlgE |
| 16  | 3,155 | 4,276 | 2022 | GC_00000542 | K04087 | hflC; modulator of FtsH protease HflC |
| 16  | 3,155 | 4,276 | 2023 | GC_00000773 | K04088 | hflK; modulator of FtsH protease HflK |
| 16  | 55,470 | 56,254 | 2074 | GC_00000498 | K02390 | figE; flagellar hook protein FigE |

**Secondary metabolites biosynthesis, transport and catabolism**

**Amino acid transport and metabolism**

**Coenzyme transport and metabolism**

**Energy production and conversion**

**General function prediction only**

**Posttranslational modification, protein turnover, chaperones**

**Cell motility**
| Accession | Start | End | Description |
|-----------|-------|-----|-------------|
| GC_00000558 | 55,470 | 56,254 | flgD; flagellar basal-body rod modification protein FlgD |
| GC_00001387 | 168,195 | 169,316 | INV, sacA; beta-fructofuranosidase [EC:3.2.1.26] |
| GC_00000821 | 176,227 | 179,316 | tsf, TSFM; elongation factor Ts |
| GC_00000184 | 176,227 | 179,316 | RP-S2, MRPS2, rpsB; small subunit ribosomal protein S2 |
| GC_00000179 | 176,227 | 179,316 | phnA; protein PhnA |
| GC_00000177 | 179,547 | 180,246 | dnr; CRP/FNR family transcriptional regulator, dissimilatory nitrate respiration regulator pbuG, azgA, ghxP, ghxQ, adeQ; adenine/guanine/hypoxanthine permease |
| GC_00000067 | 190,786 | 191,207 | dnaX; DNA polymerase III subunit gamma/tau [EC:2.7.7.7] |
| GC_00000068 | 157,543 | 157,962 | licT, bglG; beta-glucoside operon |
| GC_00001103 | 32,752 | 33,490 | dnaX; DNA polymerase III subunit gamma/tau [EC:2.7.7.7] |

**Gene Ontology:**
- **Cell motility**
- **Transcription**
- **Carbohydrate transport and metabolism**
- **Translation, ribosomal structure and biogenesis**
- **General function prediction only**
- **Signal transduction mechanisms**
- **Replication, recombination and repair**
- **Nucleotide transport and metabolism**
- **Replication, recombination and repair**
- **Transcription**
transcriptional antiterminator

Amino acid transport and metabolism|Coenzyme transport and metabolism|Translation, ribosomal structure and biogenesis|Secondary metabolites biosynthesis, transport and catabolism

K01920

gshB; glutathione synthase [EC:6.3.2.3]

COG0189

rsmE; 16S rRNA (uracil1498-N3)-methyltransferase [EC:2.1.1.193]

por, nilU; pyruvateferredoxin/flavodoxin oxidoreductase [EC:1.2.7.1 1.2.7.-]

COG1385

Translation, ribosomal structure and biogenesis

K09761

Energy production and conversion

COG0674

Cell wall/membrane/envelope biogenesis

K03737

Carbohydrate transport and metabolism

COG0149

Cell cycle control, cell division, chromosome partitioning

K13695

Mobilome: prophages, transposons

COG0791

COG0149

COG03074

COG1396

COG0482

COG2915
| Gene ID | Start | End   | Description                                                                 | GO SL | Function                                    |
|---------|-------|-------|-----------------------------------------------------------------------------|-------|---------------------------------------------|
| GC_00000090 | 44,222 | 43,379 | Amino acid transport and metabolism                                           | COG1296 | |
| GC_00000759 | 72,391 | 73,126 | Amino acid transport and metabolism                                           | COG1448 | |
| GC_00000143 | 85,482 | 86,193 | Lipid transport and metabolism                                               | COG2937 | |
| GC_00000571 | 91,471 | 92,474 | Energy production and conversion                                              | COG0437 | |
| GC_00001116 | 93,051 | 93,409 | Energy production and conversion                                              | COG2864 | |
| GC_00000010 | 93,617 | 94,635 | General function prediction only (ABC transporter)                           | COG2984 | |
| GC_00000010 | 94,748 | 94,924 | General function prediction only (ABC transporter)                           | COG2984 | |
| GC_00000797 | 95,471 | 96,256 | General function prediction only (ABC transporter)                           | COG4120 | |
### Table S3. Flexible genes enriched in *Gilliamella* pop2. Most genes were detected by both PopCOGenT and anvi’o enrichment analysis. Genes not detected by anvi’o are due to missing KEGG annotation or the presence of other genes with the same KEGG function in the genome.

| Anvi’o orthologous group ID | M6-3G genome scaffold ID | Start position of the gene on the scaffold | End position of the gene on the scaffold | Detected by PopCoGenT gene-specific sweeps in flexible genome | Detected by anvi’o KEGG enrichment analysis | KEGG IDs | KEGG functions |
|-----------------------------|--------------------------|------------------------------------------|------------------------------------------|-----------------------------------------------------------|-------------------------------------------|---------|----------------|
| GC_00009082                 | 11                       | 74,795                                   | 75,826                                   | x                                                         |                                           | K00847  | E2.7.1.4, scrK; fructokinase [EC:2.7.1.4] |
| GC_00002301                 | 11                       | 77,099                                   | 77,965                                   | x                                                         |                                           |         |                                    |
| GC_00005323                 | 11                       | 79,090                                   | 79,854                                   | x                                                         |                                           |         |                                    |
| GC_00005287                 | 11                       | 81,036                                   | 81,491                                   | x                                                         | x                                         | K02768  | fruB; fructose PTS system EIIA component [EC:2.7.1.202] |
| GC_00005164                 | 11                       | 81,539                                   | 81,850                                   | x                                                         | x                                         | K02769  | fruAb; fructose PTS system EIIB component [EC:2.7.1.202] |
| GC_00005267                 | 11                       | 81,871                                   | 82,965                                   | x                                                         |                                           | K02770  | fruA; fructose PTS system EIIBC or EIIC component [EC:2.7.1.202] |
| GC_00005306                 | 11                       | 83,064                                   | 83,606                                   | x                                                         | x                                         | K09988  | lyxA; D-lyxose ketol-isomerase [EC:5.3.1.15] |
| GC_00002192                 | 11                       | 166,073                                  | 167,641                                  | x                                                         |                                           | K22110  | kdgM, kdgN, nanC, ompL; oligogalacturonate-specific porin family protein |
| GC_00002162                 | 11                       | 173,008                                  | 173,683                                  | x                                                         |                                           |         |                                    |
| GC_00002143                 | 11                       | 175,110                                  | 175,910                                  | x                                                         |                                           | K00046  | idnO; gluconate 5-dehydrogenase [EC:1.1.1.69] |
| GC_00002179                 | 11                       | 176,006                                  | 176,326                                  | x                                                         |                                           |         |                                    |
| GC_00002180                 | 11                       | 176,444                                  | 177,238                                  | x                                                         | x                                         | K19333  | kdgR; IclR family transcriptional regulator, KDG regulon repressor |
| GC_00004921                 | 13                       | 677                                      | 1,882                                    | x                                                         | x                                         | K12661  | LRA3, rhmD; L-rhamnate dehydratase [EC:4.2.1.90] |
| GC_00005249                 | 13                       | 1,911                                    | 2,669                                    | x                                                         |                                           |         |                                    |
| GC_00007220                 | 23                       | 43,742                                   | 44,170                                   | x                                                         |                                           |         |                                    |
| GC_00006791                 | 23                       | 44,559                                   | 45,759                                   | x                                                         |                                           |         |                                    |
| GC_00007038                 | 23                       | 45,090                                   | 45,209                                   | x                                                         |                                           |         |                                    |
| GC_00007208                 | 23                       | 45,421                                   | 46,080                                   | x                                                         |                                           |         |                                    |
| Accession     | Chromosome | Start Position | End Position | Type | Description                                                                 |
|--------------|------------|----------------|--------------|------|-----------------------------------------------------------------------------|
| GC_00002150  | 23         | 53,381         | 54,310       | x    | xuuR; GntR family transcription regulator, xu operon transcriptional repressor |
| GC_00005616  | 23         | 149,195        | 149,950      | x    | K13637                                                                     |
| GC_00002826  | 23         | 246,963        | 248,558      | x    |                                                                            |
| GC_00004448  | 23         | 300,515        | 301,261      | x    |                                                                            |
| GC_00001547  | 23         | 319,784        | 321,130      | x    |                                                                            |
| GC_00007987  | 24         | 16,247         | 16,498       | x    |                                                                            |
| GC_00005682  | 24         | 17,324         | 18,448       | x    |                                                                            |
| GC_00007496  | 24         | 50,303         | 51,961       | x    |                                                                            |
| GC_00003440  | 24         | 83,494         | 84,030       | x    |                                                                            |
| GC_00001365  | 24         | 268,223        | 269,200      | x    | K18446 yglF; triphosphatase [EC:3.6.1.25]                                   |
| GC_00007493  | 24         | 269,467        | 271,986      | x    |                                                                            |
| GC_00005682  | 24         | 354,872        | 355,438      | x    |                                                                            |
| GC_00002610  | 25         | 20,460         | 21,986       | x    |                                                                            |
| GC_00002658  | 25         | 22,466         | 22,750       | x    |                                                                            |
| GC_00003421  | 28         | 5,426          | 5,749        | x    | K09771 TC.SMR3; small multidrug resistance family-3 protein                 |
| GC_00006207  | 28         | 49,203         | 49,958       | x    | K12981 waaZ, rfaZ; KDO transferase III [EC:2.4.99.-]                        |
| GC_00003240  | 28         | 102,931        | 103,875      | x    |                                                                            |
| GC_00003302  | 28         | 106,638        | 106,904      | x    | K06940; uncharacterized protein                                           |
| GC_00002608  | 28         | 110,834        | 112,069      | x    | K06016 pydC; beta-ureidopropionase / N-carbamoyl-L-amino-acid hydrolase [EC:3.5.1.6 3.5.1.87] ptsG, glcA, glcB; glucose PTS system EIICBA or EIICB component [EC:2.7.1.199] rpe, RPE; ribulose-phosphate 3-epimerase [EC:5.1.3.1] |
| GC_00004981  | 31         | 12,909         | 14,498       | x    | K20118 rpiB; ribose 5-phosphate isomerase B [EC:5.3.1.6] rpiR, alsR; RpiR family transcriptional regulator, repressor of rpiB and als operon metY; O-acetylimoserine (thiol)-lyase [EC:2.5.1.49] |
| GC_00004912  | 31         | 14,731         | 15,384       | x    | K01783                                                                         |
| GC_00004400  | 31         | 15,377         | 15,829       | x    | K01808                                                                         |
| GC_00004198  | 31         | 16,001         | 16,894       | x    | K23238                                                                         |
| GC_00002090  | 32         | 68,306         | 69,583       | x    | K01740                                                                         |
| GC_00003023  | 32         | 114,730        | 117,270      | x    |                                                                            |
| Gene ID     | Strand Length | Start End | Strand | Proteins                                                                 |
|------------|---------------|-----------|--------|--------------------------------------------------------------------------|
| GC_00003085| 32            | 117,386   | 118,996| x K02471 bacA; vitamin B12/bleomycin/antimicrobial peptide transport system ATP-binding/permease protein |
| GC_00003300| 32            | 119,044   | 121,800| x x K07263 pqqL; zinc protease [EC:3.4.24.-]                                |
| GC_00006084| 35            | 15,052    | 15,366 | x                                                                                        |
| GC_00005614| 35            | 20,758    | 20,922 | x                                                                                        |
| GC_00004608| 35            | 104,516   | 105,439| x x K02842 FZD9_10, CD349_50; frizzled 9/10                                      |
| GC_000002800| 49         | 33,116    | 33,886 | x x K06897 7,8-dihydropterin-6-yl-methyl-4-(beta-D-ribofuranosyl)aminobenzene 5'-phosphate synthase [EC:2.5.1.105] |
| GC_00002407| 49            | 57,532    | 72,414 | x                                                                                        |
| GC_00002707| 52            | 154,231   | 155,175| x K09805 K09805; uncharacterized protein                                           |
| GC_00003609| 55            | 20,898    | 22,493 | x x K02035 ABC.PE.S; peptide/nickel transport system substrate-binding protein       |
| GC_00003591| 55            | 22,508    | 23,485 | x x K02033 ABC.PE.P; peptide/nickel transport system permease protein              |
| GC_00003608| 55            | 24,315    | 25,955 | x x K02032 ddpF; peptide/nickel transport system ATP-binding protein              |
| GC_00003606| 55            | 25,968    | 26,819 | x                                                                                        |
| GC_00000009| 7             | 50,171    | 50,422 | x                                                                                        |
| GC_00006509| 7             | 50,163    | 50,422 | x x K14731 mihB, chnC; epsilon-lactone hydrolase [EC:3.1.1.83]                    |
| GC_00001766| 7             | 132,055   | 132,450| x                                                                                        |
| GC_00002364| 7             | 147,357   | 147,605| x                                                                                        |
| GC_00002416| 7             | 148,002   | 149,294| x x K13408 raxA; membrane fusion protein                                            |
| GC_00002367| 7             | 149,287   | 151,398| x x K13409 raxB, cvaB; ATP-binding cassette, subfamily B, bacterial RaxB           |
| NA         | 7             | 253,092   | 253,277| x Not annotated                                                              |
| GC_00007053| 7             | 336,121   | 336,276| x                                                                                        |
| Accession  | Start | End   | Description                                                   | EC Number          |
|------------|-------|-------|---------------------------------------------------------------|--------------------|
| GC_00005362| 72    | 39,108| 2-dehydro-3-deoxy-L-rhamnose aldolase                        | EC:4.1.2.53        |
| GC_00004921| 72    | 41,380| L-rhamnose dehydratase                                       | EC:4.2.1.90        |
| GC_00005854| 85    | 6,741 | alpha-mannosidase                                            | EC:3.2.1.24        |
| GC_00006404| 85    | 10,995| L-asparaginase                                              | EC:3.5.1.1         |
Table S4. Genes related to metabolic pathways, inter-bacterial antagonistic interactions and CRISPR-Cas systems significantly enriched in Gilliamella populations.

| Clades                                                                 | Am_Gapis                  | Am_Gapicola               | Bm  |
|-----------------------------------------------------------------------|---------------------------|---------------------------|-----|
| Urease genes and urea transporters: ureABCDEFG, urtABCDE (K01428, K01429, K01430, K03187, K03188, K03189, K03190, K11959, K11960, K11961, K11962, K11963) | x                         |                           |     |
| Nickel transport system permease protein: nikBCDE (K10824, K15585, K15587, K15586) | x                         | x                         |     |
| Xylose metabolism: xylAB (K01805, K00854)                             | x                         | x                         |     |
| Nickel transport system permease protein: pseuGK (K16328, K16329)     |                           |                           |     |
| N-acetylgalactosamine PTS system and N-acetylglucosamine-6-phosphate deacetylase: agaAEFSVW (K02744, K02745, K02746 K02747, K02079, K02082) | x                         | x                         |     |
| fructose PTS system: fruAb, fruB (K02768, K02769)                     |                           |                           |     |
| Peptide/nickel transport system permease protein, ABC.PE.S, ABC.PE.P1, ABC.PE.P (K02033, K02034, K02035) |                           |                           |     |
| Extracellular pectin degradation: GH28, GH32, GH38 (GC_00002366, GC_00002401, GC_00002404) |                           |                           |     |
| Xylose and arabinose metabolism: xylE, araABCE (K08138, K01804, K00853, K02099, K02100) |                           |                           |     |
| Maltose-6'-phosphate glucosidase: glvAC (K01232, K027500)              |                           |                           |     |
| Dipeptide transport system: ddpABCEF (K12368, K12369, K12370, K12371, K12372) |                           |                           |     |
| Type I secretion system protein: cvaAB (K13408, K13409)                |                           |                           |     |
| Tight adherence export apparatus: cpaACEF, tadBCDFG, flp (K02278, K02280, K02282, K02283, K12510, K12511, K12512, K12514, K12515, K02651) |                           |                           |     |
| Type VI secretion systems - T6SS-1 (GC_00001745, GC_00001774, GC_00001777, GC_00001795, GC_00001797, GC_00001812, GC_00001824, GC_00001831, GC_00001836, GC_00001841, GC_00001847, GC_00001850, GC_00001851) | x | x | x | x | x | x |
| Capsular polysaccharide transporters: kpsCMET (K07266, K09688, K09689, K10107) | x | x | x |
| CRISPR-Cas system - Subtype I-F factors: cys1234 (K19127, K19128, K19129, K19130) | x | x | x | x | x | x |
| CRISPR-Cas system - Universal Cas proteins, Subtype I-E factors: cas2, casACDE (K09951, K19123, K19124, K19125, K19126) | x | x | x | x | x | x | x |
Table S5. Genes related to metabolic pathways and inter-bacterial antagonistic interactions significantly enriched in Snodgrassella populations.

| Clades                                                                 | Ap. mellifera derived | Bombus derived - Broad-host range | Bombus derived | Bombus derived - Host-specific | Other Apis species derived |
|------------------------------------------------------------------------|-----------------------|-----------------------------------|----------------|-------------------------------|---------------------------|
| Populations                                                            | 1                     | 2                                 | 3              | 4                            | 5                         | 6 | 7 | 8 |
| Urea transporter: utp (K08717)                                         | x                     | x                                 | x              |                               |                           | x | x |   |
| Urease genes: ureABCDEFGJ (K01430, K01429, K01428, K03190, K03187, K03188, K03189, K03192) | x                     | x                                 | x              |                               |                           | x | x |   |
| Arginine degradation genes: arcAC (K01478, K00926)                     |                       |                                   |                |                               |                           |   |   |   |
| Two component system - NtrC family, short-chain fatty acids transporter: atoCES (K07710, K07714, K02106) | x                     |                                   | x              |                               |                           | x | x |   |
| Biosynthesis of siderophore group nonribosomal peptides: entACE (K02361, K02363, K00216) | x                     | x                                 | x              |                               |                           | x | x |   |
| fhaB; filamentous hemagglutinin (K15125)                               | x                     |                                   |                |                               |                           |   |   |   |
| Type VI secretion systems - T6SS-1 (GC_00001925, GC_00001927, GC_00001924, GC_00001919, GC_00001918, GC_00001906, GC_00001916, GC_00001902, GC_00001903, GC_00001922, GC_00001904, GC_00001911, GC_00001880) | x                     | x                                 | x              |                               |                           | x | x |   |
| Type VI secretion systems - T6SS-2 (GC_00001996, GC_00002021, GC_00001998, GC_00002009) |                       |                                   |                |                               |                           | x | x |   |
| Type VI secretion systems - T6SS-3 (GC_00002008, GC_00002011, GC_00001989, GC_00002012, GC_00002028) |                       |                                   |                |                               |                           | x | x |   |
Table S6. Bacterial strains used in this study.

| Species and strain       | Source / Description | ID       | References |
|--------------------------|----------------------|----------|------------|
| *E. coli* MFDpir         | Ferrieres et al. 2010| N/A      | (40)       |
| *E. coli* DH5α           | Thermo-Fisher        | CAT# 11319-019 |           |
| *Snodgrassella alvi* wkB2| Kwong et al. 2013    | ATCC: BAA-2449 | (36)       |
| *Gilliamella apicola* wkB7| Kwong et al. 2014   | N/A      | (41)       |
| *Gilliamella apis* M6-3G  |                      | N/A      | (29)       |
| Name     | Use                                                                 | Source     |
|----------|----------------------------------------------------------------------|------------|
| pBTK402  | Type 8 origin of replication and origin of transfer plasmid, rsf1010 broad-host-range origin. Encodes Kanamycin resistance (aphA-1). | (37)       |
| pSL1     | Derived from pBTK402. Constitutively expresses red chromoprotein E1010 gene from CP25 promoter. | This work  |
| pSL1-GFP | Derived from pSL1. Replaced RCP to constitutively express GFP from CP25 promoter. | This work  |
| pSL1-E2C | Derived from pSL1. Replaced RCP to constitutively express E2-Crimson from CP25 promoter. | This work  |
SI References

1. D. H. Parks, M. Imelfort, C. T. Skennerton, P. Hugenholtz, G. W. Tyson, CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055 (2015).

2. D. Hyatt, *et al.*, Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11, 119 (2010).

3. A. M. Eren, *et al.*, Anvi’o: an advanced analysis and visualization platform for ‘omics data. *PeerJ* 3, e1319 (2015).

4. M. Y. Galperin, *et al.*, COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res.* 49, D274–D281 (2021).

5. M. Kanehisa, M. Furumichi, Y. Sato, M. Ishiguro-Watanabe, M. Tanabe, KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* 49, D309–D314 (2019).

6. J. Huerta-Cepas, *et al.*, eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314 (2019).

7. M. Kanehisa, Y. Sato, K. Morishima, BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726–731 (2016).

8. J. Huerta-Cepas, *et al.*, Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Mol. Biol. Evol.* 34, 2115–2122 (2017).

9. H. Zhang, *et al.*, dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 46, W95–W101 (2018).

10. R. D. Finn, J. Clements, S. R. Eddy, HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37 (2011).

11. B. Buchfink, C. Xie, D. H. Huson, Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60 (2015).

12. P. K. Busk, B. Pilgaard, M. J. Lezyk, A. S. Meyer, L. Lange, Homology to peptide pattern for annotation of carbohydrate-active enzymes and prediction of function. *BMC Bioinformatics* 18, 214 (2017).

13. M. I. Steele, W. K. Kwong, M. Whiteley, N. A. Moran, Diversification of type VI secretion system toxins reveals ancient antagonism among bee gut microbes. *mBio* 8 (2017).

14. M. I. Steele, N. A. Moran, Evolution of interbacterial antagonism in bee gut microbiota reflects host and symbiont diversification. *mSystems* 6, e00632-21, /mSystems/6/3/mSys.00632-21.atom (2021).

15. D. Couvin, *et al.*, CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* 46, W246–W251 (2018).
16. C. Camacho, et al., BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421 (2009).

17. S. van Dongen, C. Abreu-Goodger, “Using MCL to extract clusters from networks” in *Bacterial Molecular Networks: Methods and Protocols*, Methods in Molecular Biology., J. van Helden, A. Toussaint, D. Thieffry, Eds. (Springer, 2012), pp. 281–295.

18. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780 (2013).

19. B. Q. Minh, et al., IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534 (2020).

20. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin, ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589 (2017).

21. D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522 (2018).

22. G. Yu, D. Smith, H. Zhu, Y. Guan, T. T.-Y. Lam, ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods. Ecol. Evol.* 8, 28–36 (2017).

23. R Core Team, *R: a language and environment for statistical computing* (R Foundation for Statistical Computing, 2016).

24. H.-G. Drost, Philentropy: information theory and distance quantification with R. *Journal of Open Source Software* 3, 765 (2018).

25. E. Paradis, K. Schliep, ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528 (2019).

26. J. Oksanen, et al., vegan: community ecology package (2020).

27. P. Arevalo, D. VanInsberghe, J. Elsherbini, J. Gore, M. F. Polz, A reverse ecology approach based on a biological definition of microbial populations. *Cell* 178, 820-834.e14 (2019).

28. P. Shannon, et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504 (2003).

29. H. Zheng, et al., Metabolism of toxic sugars by strains of the bee gut symbiont Gilliamella apicola. *mBio* 7, e01326-16 (2016).

30. K. M. Ellegaard, P. Engel, New reference genome sequences for 17 bacterial strains of the honey bee gut microbiota. *Microbiol. Resour. Announc.* 7, e00834-18 (2018).

31. O. Gascuel, BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.* 14, 685–695 (1997).

32. A. Rambaut, N. C. Grass, Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Bioinformatics* 13, 235–238 (1997).

33. RStudio Team, *RStudio: integrated development environment for R* (RStudio, PBC, 2020).
34. H. Wickham, et al., Welcome to the Tidyverse. *Journal of Open Source Software* 4, 1686 (2019).

35. J. E. Powell, S. P. Leonard, W. K. Kwong, P. Engel, N. A. Moran, Genome-wide screen identifies host colonization determinants in a bacterial gut symbiont. *Proc. Natl. Acad. Sci. U.S.A.* 113, 13887–13892 (2016).

36. W. K. Kwong, N. A. Moran, Cultivation and characterization of the gut symbionts of honey bees and bumble bees: description of Snodgrassella alvi gen. nov., sp. nov., a member of the family Neisseriaceae of the Betaproteobacteria, and Gilliamella apicola gen. nov., sp. nov., a member of Orbaceae fam. nov., Orbales ord. nov., a sister taxon to the order ‘Enterobacteriales’ of the Gammaproteobacteria. *Int. J. Syst. Evol. Microbiol.* 63, 2008–2018 (2013).

37. S. P. Leonard, et al., Genetic engineering of bee gut microbiome bacteria with a toolkit for modular assembly of broad-host-range plasmids. *ACS Synth. Biol.* 7, 1279–1290 (2018).

38. S. P. Leonard, et al., Engineered symbionts activate honey bee immunity and limit pathogens. *Science* 367, 573–576 (2020).

39. V. G. Martinson, J. Moy, N. A. Moran, Establishment of characteristic gut bacteria during development of the honeybee worker. *Appl. Environ. Microbiol.* 78, 2830–2840 (2012).

40. L. Ferrières, et al., Silent mischief: bacteriophage Mu insertions contaminate products of *Escherichia coli* random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. *J. Bacteriol.* 192, 6418–6427 (2010).

41. W. K. Kwong, P. Engel, H. Koch, N. A. Moran, Genomics and host specialization of honey bee and bumble bee gut symbionts. *Proc. Natl. Acad. Sci. U.S.A.* 111, 11509–11514 (2014).