Division of labour within psyllids: Metagenomics reveals an ancient dual endosymbiosis with metabolic complementarity in the genus *Cacopsylla*

Jessica Dittmer[a,b]#, Erika Corretto[a], Liliya Štarhová Serbina[a], Anna Michalik[c], Eva Nováková[d], Hannes Schuler[a,e]#

[a] Faculty of Agricultural, Environmental and Food Sciences, Free University of Bozen-Bolzano, Bolzano, Italy

[b] UMR 1345, Université d'Angers, Institut Agro, INRAE, IRHS, SFR Quasav, Beaucouzé, France

[c] Department of Developmental Biology and Morphology of Invertebrates, Institute of Zoology and Biomedical Research, Faculty of Biology, Jagiellonian University, Krakow, Poland

[d] Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic

[e] Competence Centre for Plant Health, Free University of Bozen-Bolzano, Bolzano, Italy

#Corresponding Authors:

Jessica Dittmer, UMR 1345, Université d'Angers, Institut Agro, INRAE, IRHS, SFR Quasav, Beaucouzé, France. Email: jessica.dittmer@inrae.fr

Hannes Schuler, Faculty of Agricultural, Environmental and Food Sciences, Free University of Bozen-Bolzano, Bolzano, Italy. Email: hannes.schuler@unibz.it

Running title: Ancient dual endosymbiosis in psyllids

Word count Abstract: 229

Word count text: 4667

ABSTRACT

Hemipteran insects are well-known for their ancient associations with beneficial bacterial endosymbionts, particularly nutritional symbionts providing the host with essential nutrients such as amino acids or vitamins lacking from the host's diet. Thereby, these primary endosymbionts enable the exploitation of nutrient-poor food sources such as plant sap or vertebrate blood. In turn, the strictly host-associated lifestyle strongly impacts the genome evolution of the endosymbionts, resulting in small and degraded genomes. Over time, even the essential nutritional functions can be compromised, leading to the complementation or replacement of an ancient endosymbiont by another, more functionally versatile, bacterium. Herein, we provide evidence for a dual primary endosymbiosis in several psyllid species. Using metagenome sequencing, we produced the complete genome sequences of both the primary endosymbiont '*Candidatus* Carsonella ruddii' and an as yet uncharacterized *Enterobacteriaceae* bacterium from four species of the genus *Cacopsylla*. The latter represents a new psyllid-associated endosymbiont clade for which we propose the name '*Candidatus* Psyllophila symbiotica'. Fluorescent *in situ* hybridisation confirmed the co-localization of both endosymbionts in the bacteriome. The metabolic repertoire of *Psyllophila* is highly conserved across host species and complements the tryptophan biosynthesis pathway that is incomplete in the co-occurring *Carsonella.* Unlike co-primary endosymbionts in other insects, the genome of *Psyllophila* is almost as small as the one of *Carsonella*, indicating an ancient co-obligate endosymbiosis rather than a recent association to rescue a degrading primary endosymbiont.

IMPORTANCE

Heritable beneficial bacterial endosymbionts have been crucial for the evolutionary success of numerous insects, enabling the exploitation of nutritionally limited food sources such as vertebrate blood and plant sap. Herein, we describe a previously unknown dual endosymbiosis in the psyllid genus *Cacospylla*, consisting in the primary endosymbiont '*Candidatus* Carsonella ruddii' and a co-occurring *Enterobacteriaceae* bacterium for which we propose the name '*Candidatus* Psyllophila symbiotica'. Its localization within the bacteriome and its small genome size confirm that *Psyllophila* is a co-primary endosymbiont widespread within the genus *Cacopsylla.* Despite its highly eroded genome, *Psyllophila* complements the tryptophan biosynthesis pathway that is incomplete in the co-occurring *Carsonella.* Moreover, the genome of *Psyllophila* is almost as small as the one of *Carsonella*, indicating an ancient dual endosymbiosis rather than a recent acquisition of a new symbiont. Hence, our results shed light on the dynamic interactions of psyllids and their endosymbionts over evolutionary time.

INTRODUCTION

Numerous insects maintain long-lasting associations with heritable bacterial endosymbionts that provide the host with essential nutrients lacking from its diet (1). Plant sap-feeding and blood-feeding insects in particular are well-known to harbour nutrient-providing endosymbionts in specialized cells called bacteriocytes, which may form a tissular structure called a bacteriome (2, 3). These so-called primary endosymbionts are obligatory for host survival and reproduction, as they provide essential amino acids and/or vitamins that the host cannot produce or obtain from its food source (4-9). Hence, these bacteria have been crucial for the evolutionary success of numerous insects, enabling the exploitation of nutritionally unbalanced food sources such as vertebrate blood and plant sap.

In turn, the host-associated lifestyle has a strong impact on the genome evolution of the endosymbionts: Their strictly intracellular environment, small effective population size and frequent bottlenecks due to vertical transmission result in genomic decay through the accumulation of deleterious mutations (Muller's ratchet) and the loss of genes that are no longer needed (10-12). Over evolutionary time, this has produced some of the smallest bacterial genomes known to date (6), streamlined for the production of nutrients required by the host. However, eventually even these pathways can be degraded, leading either to the complementation or the replacement of the ancient endosymbiont by another, more functionally versatile, bacterium (13-18).

This dynamic can be observed in several plant sap-feeding hemipterans which rely on more than one primary endosymbiont to produce all necessary nutrients. Notably, the Auchenorrhyncha (cicadas, planthoppers, spittlebugs) are well-known for their ancient dual endosymbiotic consortia where two co-primary endosymbionts jointly produce the complete set of essential nutrients required by the host, resulting in an intricate metabolic

interdependence between the different partners (19). Nonetheless, multiple endosymbiont replacements occurred over time to compensate for the extreme genome erosion of the ancient symbionts (6, 16, 20-25). A similar pattern occurs in aphids (Sternorrhyncha): Whereas most species harbour a single primary endosymbiont, *Buchnera aphidicola*, which provides the host with the ten essential amino acids and the vitamins biotin and riboflavin (26), dual-endosymbiotic systems have evolved repeatedly in multiple aphid lineages to compensate for lost pathways in *B. aphidicola* (13, 14, 18, 27-29).

Similar dual primary endosymbioses may be widespread in psyllids (Hemiptera: Psylloidea), a species-rich group of phloem-feeding jumping plant lice. Like other plant sap-feeding insects, psyllids harbour a bacteriocyte-associated primary endosymbiont ('*Candidatus* Carsonella ruddii', hereafter *Carsonella*), which provides the host with essential amino acids (30-33). *Carsonella* is present in all investigated psyllid species and exhibits strict host-symbiont co-divergence, suggesting a single infection of a common ancestor of all extant psyllids (30, 34, 35). Its genome is extremely streamlined and figures among the smallest bacterial genomes known to date (157-175 Kbp) (31). Due to this extreme genome reduction, some *Carsonella* strains are no longer able to produce the full complement of essential amino acids, questioning their ability to fulfil their symbiotic function without compensation from host genes or co-occurring symbiotic bacteria (33, 36).

Additional endosymbionts have indeed been observed to co-inhabit the bacteriome with *Carsonella* in several species (3, 37-39). In these cases, *Carsonella* is located in bacteriocytes surrounding the bacteriome, while a second bacterium occurs in the syncytium at the center of the bacteriome. Importantly, the taxonomy of the co-primary endosymbiont varies depending on the psyllid species: Whereas the syncytium-symbiont ('Y-symbiont') of the mulberry psyllid *Anomoneura mori* is an uncharacterized *Enterobacteriaceae* bacterium (*Gammaproteobacteria*) (37) whose symbiotic role is unknown, the citrus psyllid *Diaphorina*

citri harbours '*Candidatus* Profftella armatura' (*Betaproteobacteria*). The latter is a defensive and nutritional endosymbiont which produces vitamins, carotenoids and a polyketide toxin, i.e. metabolites that are not provided by *Carsonella* (39, 40). In contrast, the psyllid species *Ctenarytaina eucalypti* and *Heteropsylla cubana* harbour symbionts closely related to the insect endosymbionts '*Ca*. Moranella endobia' and *Sodalis*, whose genomes precisely complement several amino acid biosynthesis pathways missing from the co-occurring *Carsonella* strains (33). Despite typical hallmarks of vertically transmitted intracellular bacteria, the genomes of both endosymbionts are less reduced (>1 Mbp), suggesting a more recent acquisition relative to *Carsonella,* presumably to compensate for lost functions in the latter. In addition, numerous psyllid microbiome studies revealed highly abundant but yet uncharacterized *Enterobacteriaceae* bacteria in diverse species from several psyllid families (41-46), suggesting that dual primary endosymbioses may be more widespread in psyllids than previously thought.

Herein, we aim to elucidate the evolutionary and metabolic relationships between psyllids of the genus *Cacopsylla* (Psyllidae) and their *Enterobacteriaceae* endosymbionts. Many *Cacopsylla* species have indeed been shown to harbour highly abundant *Enterobacteriaceae* endosymbionts that are closely-related to the Y-symbiont co-inhabiting the syncytium of the bacteriome in *A. mori* (42, 45, 46). Furthermore, these symbionts were present in all tested individuals of a given species, suggesting that they may represent co-primary endosymbionts widespread in this genus. In this study, we produced the complete genome sequences of both *Carsonella* and the *Enterobacteriaceae* endosymbionts of four *Cacopsylla* species (*C. melanoneura, C. picta, C. pyri* and *C. pyricola*) known to harbour closely-related endosymbionts from our previous metabarcoding studies (45, 46). Fluorescent *in situ* hybridisation confirmed the co-localization of both endosymbionts in the bacteriome. Comparative genomic analyses revealed that the *Enterobacteriaceae* endosymbionts

represent a psyllid-associated clade among other insect endosymbionts. Its genome is almost as small as that of *Carsonella* and complements the tryptophan biosynthesis pathway that is compromised in the co-occurring *Carsonella*.

RESULTS

All four *Cacopsylla* species harbour two endosymbionts with tiny genomes

To investigate endosymbiont genetic diversity across different *Cacopsylla* species and genotypes, 12 insect metagenomes were sequenced. These metagenomes encompassed four different host species: *C. melanoneura* and *C. picta,* which complete their development on apple (or hawthorn in the case of *C. melanoneura*) and the pear psyllids *C. pyri* and *C. pyricola.* Multiple metagenomes were sequenced for *C. melanoneura* (N=8) and *C. picta* (N=2), covering different Cytochrome Oxidase I (COI) haplotypes and, in the case of *C. melanoneura,* different regions of origin (Aosta Valley vs. South Tyrol, Italy) and different host plants (apple vs. hawthorn) (Table 1). As expected, the majority of the Nanopore reads belonged to the insect genome and only about 5% of the reads (range: 2.94-9.80%) corresponded to non-host reads. Nonetheless, circular genomes of the primary endosymbiont *Carsonella* could be assembled from all metagenomes (Table 1) with coverages of 85-408x. Genome size ranged from 169,917 to 171,920 bp with 14.98-15.53% GC content, similar to previously sequenced *Carsonella* genomes from other psyllid genera (31, 33, 39, 40). The genomes encoded 182-190 protein-coding genes, one ribosomal rRNA operon and 26-27 tRNAs (Table 1). Synteny and gene content were highly conserved across all genomes, with 161 out of 184 orthogroups (87.5%) shared across all twelve genomes (Fig. 1a, b).

In addition to *Carsonella*, a second circular genome could be assembled from 10 out of the 12 metagenomes (Table 1) with coverages of 19-431x. These genomes belonged to the uncharacterized *Enterobacteriaceae* endosymbiont previously identified through 16S rRNA

gene metabarcoding (45, 46). Contigs of this symbiont were also present in the two remaining metagenomes (both from *C. melanoneura*), but the coverage was insufficient to assemble complete genomes. The ten complete chromosomes of the *Enterobacteriaceae* endosymbiont ranged from 221,413 bp in *C. pyri* to 237,114 bp in a strain from *C. melanoneura* from Aosta Valley (strain PSmelAO1, Table 1). GC content varied from 17.30-18.60%. Despite the variations in genome size, synteny and gene content were highly conserved across all *Enterobacteriaceae* genomes (Fig. 1c, d). They contained 205-208 protein-coding genes, 1-3 pseudogenes, one ribosomal rRNA operon, 27 tRNAs and 2 ncRNAs (Table 1). Moreover, 196 out of 209 orthogroups (93.78%) were shared across all ten genomes (Fig. 1d), indicating that the functional repertoire is highly similar across all four host species. Taken together, all four *Cacopsylla* species harbour two endosymbionts with typical hallmarks of a long intracellular symbiotic lifestyle, such as extremely small genomes and low GC content.

The *Enterobacteriaceae* symbionts represent a new clade of insect endosymbionts

To determine the phylogenetic position of the newly-sequenced psyllid endosymbionts, we performed a Maximum Likelihood phylogenomic analysis based on 67 single-copy genes present in 46 genomes, namely the 10 *Enterobacteriaceae* endosymbionts of *Cacopsylla* spp., 33 insect endosymbionts from the *Gammaproteobacteria* and three *Pseudomonas entomophila* strains as outgroup (Fig. 2). The insect endosymbionts included the two previously sequenced endosymbionts of the psyllid species *C. eucalypti* and *H. cubana* as well as both obligate and facultative endosymbionts of diverse hemipterans (aphids, adelgids, leafhoppers, mealybugs, stinkbugs), beetles (reef beetles and weevils) and the tsetse fly (Table S1). Interestingly, the *Enterobacteriaceae* endosymbionts of *Cacopsylla* spp. were not closely-related to the previously sequenced endosymbionts of the psyllids *C. eucalypti* and *H. cubana* (33) (Fig. 2). Instead, they formed a clade with full bootstrap support that was most

196 closely-related to '*Ca.* Annandia adelgestsuga' and '*Ca.* Annandia pinicola', nutritional
197 endosymbionts of adelgids (47) as well as '*Ca.* Nardonella sp.', ancient endosymbionts of
198 weevils (48) (Fig. 2). Hence, the *Enterobacteriaceae* endosymbionts of *Cacopsylla* spp.
199 represent a new psyllid-associated clade of insect endosymbionts for which we propose the
200 name '*Ca.* Psyllophila symbiotica' (hereafter *Psyllophila*).
201
202 **Both *Cacopsylla* endosymbionts are localized in the bacteriome**
203 Fluorescence *in situ* hybridization with *Carsonella* and *Psyllophila*-specific probes revealed
204 that all *Cacopsylla* species exhibit the same pattern of endosymbiont co-localization in the
205 same bacteriome (Fig. 3). The bacteriomes are large, paired organs localized in the insect's
206 abdomen. A single bacteriome contains two distinct parts: central and peripheral.
207 Uninucleated bacteriocytes filled with *Carsonella* are located in the peripheral zone of the
208 bacteriome (Fig. 3), whereas the central part is occupied by a multinucleated syncytium filled
209 with *Psyllophila* cells as well as some bacteriocytes containing *Carsonella* (Fig. 3).
210
211 **Metabolic complementarity between *Carsonella* and *Psyllophila***
212 The COG category "Amino acid transport and metabolism" was enriched in all sequenced
213 *Carsonella* genomes (Fig. 4a), in line with its role as a nutritional symbiont. Indeed, based on
214 the KEGG pathway annotation, the biosynthesis pathways for eight of the ten essential amino
215 acids are complete or almost complete in all 12 *Carsonella* strains from the four *Cacopsylla*
216 species (Fig. 4b). Most of the missing functions (*hisN* in the histidine pathway, *dapC* in the
217 lysine pathway, *thrB* in the threonine pathway and *aroE* in the Shikimate pathway, Fig. 4b)
218 are also missing in all previously sequenced *Carsonella* genomes (Table S2, S3). The same
219 applies to the methionine biosynthesis pathway, for which only the last reaction (*metE*) is
220 present in all sequenced *Carsonella* genomes from this and previous studies (Fig. 4b, Table

S3). The only difference between the *Cacopsylla*-associated *Carsonella* strains was the absence of *aroB* in the stains from *C. picta* and *C. pyri*, whereas this gene is present in all *Carsonella* strains from *C. melanoneura* and *C. pyricola* (Fig. 4b, Table S3). Interestingly, the tryptophan biosynthesis pathway was incomplete in all 12 *Carsonella* strains from *Cacopsylla* spp., in that only *trpE* and *trpG* were present, whereas the rest of the pathway was missing (Fig. 4b).

In contrast to *Carsonella,* the genomes of *Psyllophila* have lost almost all genes involved in amino acid synthesis ((Fig. 4a). Only four genes were retained and these precisely complement the incomplete tryptophan biosynthesis pathway in *Carsonella*, namely *trpD, trpCF, trpB* and *trpA* (Fig. 4b). The four genes were arranged consecutively in the genomes. In addition, all *Psyllophila* genomes encoded partial biosynthesis pathways for the vitamins biotin (*bioA, bioB, bioD*) and riboflavin (*ribA, ribB, ribD, ribH*) as well as all necessary genes for the biosynthesis of carotenoids (*crtB, crtI, crtY*) (Fig. 4b).

Repeated gene losses throughout *Carsonella* evolution

Apart from the *Carsonella* genomes presented herein, complete genome sequences are available for 11 *Carsonella* strains from nine psyllid species representing five genera and three families (Aphalaridae, Psyllidae and Triozidae) (Table S2). The functional repertoire of these genomes is quite conserved, since 135 out of 197 orthogroups (68.5%) were shared across all 33 genomes and specific orthogroups occurring only in strains from particular host species or genera were rare (14/197) (Fig. 5a). In contrast, host lineage-specific losses of orthogroups were more common. For instance, 12 orthogroups were specifically absent from the three *Carsonella* strains from *H. texana, P. celtidis* and *P. venusta* (Fig. 5a). Similarly, six orthogroups were specifically absent from the *Carsonella* strains from *Ctenarytaina* spp., four orthogroups were absent from strains from *Cacopsylla* spp. and three orthogroups were

absent from strains from *Pachypsylla* spp. (Fig. 5a). These differences are also reflected in repeated losses of genes or entirely pathways involved in essential amino acid biosynthesis across the *Carsonella* phylogeny (Fig. 5b). Notably, the tryptophan pathway has been lost at least three times independently, as it is incomplete or missing in all *Carsonella* strains associated with the genera *Cacopsylla* and *Heteropsylla* (Psyllidae) as well as *Ctenarytaina* and *Pachypsylla* (Aphalaridae) (Fig. 5b, Table S3). In contrast, this pathway is complete in the *Carsonella* strains from *Bactericera* spp. (Triozidae) and *Diaphorina citri* (Liviidae) (Fig. 5b, Table S3). Other repeatedly lost functions include the histidine biosynthesis pathway as well as the genes *aroB* and *dapE,* implicated in the Shikimate and lysine pathways, respectively (Fig. 5b, Table S3). Concomitantly, co-primary endosymbionts complementing the missing amino acid biosynthesis pathways have been identified in several species, i.e. complementing tryptophan in *Cacopsylla* spp. (this study) and *H. cubana* (33) and both tryptophan and arginine in *Ct. eucalypti* (33) (Fig. 5b).

DISCUSSION

Herein, we present the complete genome sequences of both *Carsonella* and the uncharacterized *Enterobacteriaceae* endosymbionts of four *Cacopsylla* species from different host plants. The *Enterobacteriaceae* endosymbionts represent a psyllid-associated clade among other insect endosymbionts, for which we propose the name '*Ca.* Psyllophila symbiotica'. Both endosymbionts co-occur within the bacteriome, exhibiting the same co-localization pattern (*Carsonella* in peripheral bacteriocytes, *Psyllophila* in the central syncytium) as for other dual endosymbioses in *D. citri* and *A. mori* (37-39). In combination with a small and AT-rich genome, the bacteriome localization confirms that *Psyllophila* is a co-primary endosymbiont widespread within the genus *Cacopsylla.* Interestingly, unlike co-occurring endosymbionts in other psyllid species (33, 39, 40), the *Psyllophila* genome is

almost as small as the genome of *Carsonella*, indicating an ancient dual endosymbiosis rather than a recent acquisition of a more versatile symbiont to rescue a degrading primary endosymbiont.

Despite having a tiny and functionally limited genome, *Psyllophila* has retained the necessary genes to complement the tryptophan biosynthesis pathway that is compromised in the co-occurring *Carsonella.* This appears to be a recurring theme across *Carsonella* evolution, since the tryptophan pathway is the most frequently lost amino acid biosynthesis pathway based on the genomes available to date. Specifically, this pathway has been lost multiple times independently, namely in the *Carsonella* strains associated with species from the genera *Pachypsylla* and *Ctenarytaina* (Aphalaridae), and the psyllid lineage leading to both *Heteropsylla* and *Cacopsylla* (Psyllidae). Apart from tryptophan, the arginine and histidine pathways have also been lost in specific *Carsonella* strains, albeit less frequently (33). Concomitantly, co-primary endosymbionts complementing the missing amino acid biosynthesis pathways have been identified in several species, namely *Psyllophila* and a *Sodalis*-like symbiont complementing tryptophan in *Cacopsylla* spp. (this study) and *H. cubana* (33), respectively, and another *Sodalis*-like symbiont complementing both tryptophan and arginine in *Ct. eucalypti* (33). Intriguing cases in this context are the psyllid species *H. texana, P. celtidis* and *P. venusta*, whose *Carsonella* strains have lost both the histidine and tryptophan pathways, but no co-primary endosymbionts have been observed to date (33). Possible alternative scenarios are that the missing genes are encoded by the host, e.g. after horizontal transfers of bacterial genes to the host genome, or that the amino acids in question are present in sufficient quantities in the phloem sap of the psyllid's host plant. Although numerous genes of bacterial origin have indeed been identified in the genome of *P. venusta*, they do not restore the missing amino acid pathways (36).

Based on its genome sequence, *Psyllophila* not only rescues tryptophan biosynthesis, it also encodes partial biosynthesis pathways for the vitamins biotin and riboflavin, as well as all necessary genes for the synthesis of carotenoids, pigments that may protect against oxidative damage of DNA (49). This represents a striking convergence with '*Ca.* Profftella armatura', the co-primary endosymbiont in several *Diaphorina* species, which has an almost identical gene set for these pathways as *Psyllophila* (40). As in *Psyllophila,* the last step in the riboflavin pathway is missing in '*Ca.* Profftella armatura', but the relevant gene has been detected in the genomes of the psyllids *D. citri* and *P. venusta*, likely due to a horizontal transfer from an unknown bacterium (36). Likewise, we identified a similar riboflavin synthase gene encoded in the genomes of all four *Cacopsylla* species via blast searches of the *D. citri* gene against preliminary assemblies of the insect genomes from our metagenomic datasets. Hence, it is likely that riboflavin can be jointly synthesized by *Psyllophila* and its psyllid hosts, just like in the symbiosis of *D. citri* and *Profftella*. In any case, the functional similarity between two distantly-related psyllid endosymbionts highlights the importance of these metabolites for the psyllid hosts and/or the endosymbionts.

Taken together, our data shed light on the dynamic interactions of psyllids and their endosymbionts over evolutionary time. Notably, the tiny and highly eroded genome of *Psyllophila* suggests a long-lasting dual endosymbiosis of *Carsonella* and *Psyllophila* within the genus *Cacopsylla*. However, this dual endosymbiosis has likely reached a highly labile state, since no functional redundancy exists between the two endosymbionts and any additional gene loss would destabilise the symbiotic system. Considering the diversity of predominant psyllid-associated bacteria revealed by previous studies (34, 41-46, 50), it is likely that the ancient endosymbiont *Psyllophila* has already been replaced by younger symbionts in some psyllid lineages. For instance, this may have been the case in *H. cubana* and *C. eucalypti*, which harbour more recently acquired co-primary endosymbionts with

larger genomes (33). It is tempting to speculate that species which do not harbour co-primary endosymbionts today (e.g. *P. venusta*) may have harboured a similar dual endosymbiosis in the past, but the co-symbiont was lost without replacement, maybe because its functions were no longer required after a change in ecological conditions (e.g. change of host plant, evolution of gall-forming behaviour). This could also explain why the *Carsonella* strains in these species have lost similar genes and pathways as the strains existing in dual endosymbiotic systems today.

This raises the question whether these pathways were lost before or after the establishment of the dual primary endosymbiosis. According to the Black Queen Theory on the evolution of dependencies within bacterial communities (51), it is advantageous for bacteria to lose costly metabolic functions (i.e. to streamline their genomes), as long as another species within the community still produces these metabolites as "common goods". Applying this concept to a community with two partners would imply that any essential pathway can be lost in only one of them but has to be retained in the other, to maintain all essential functions in the system. Hence, in the dual endosymbioses in psyllids, *Carsonella* may have lost the tryptophan pathway, since it was encoded by its symbiotic partner. In turn, the co-primary endosymbionts lost all other genes involved in amino acid biosynthesis, since these were maintained in *Carsonella*, thus establishing the existing metabolic complementarities in different psyllids. However, only a few psyllid endosymbionts have been characterized at the genomic level and more studies across the psyllid tree of life will be necessary to obtain a more complete picture of the evolutionary dynamics of psyllids and their primary endosymbionts.

MATERIALS AND METHODS

Psyllid samples

Genomic data was obtained from four psyllid species: the apple psyllids *Cacopsylla melanoneura* and *C. picta* as well as the pear psyllids *C. pyri* and *C. pyricola*. All four species are vectors of plant pathogens, namely '*Ca.* Phytoplasma mali' and '*Ca.* Phytoplasma pyri', respectively causing Apple proliferation and Pear decline (52, 53). Remigrants (i.e. adults that return to their host plants for reproduction after overwintering on shelter plants) of *C. melanoneura* were captured in various apple orchards in two Italian regions (Aosta Valley and South Tyrol) in March 2020 and March 2021. Additional *C. melanoneura* specimens were sampled on hawthorn (*Crataegus* sp.) in a single location (Aosta Valley) in March 2021. Remigrants of *C. picta* were captured from apple orchards in Trentino (Italy) in April 2021. Adults of *C. pyri* and *C. pyricola* were collected in pear orchards in Litenčice (Czech Republic) in December 2019 and in Starý Lískovec (Czech Republic) in July 2020, respectively. Sampling was done using the beating tray method.

Since psyllids are too small to obtain sufficient DNA for long-read sequencing from a single individual, several specimens need to be pooled, which introduces genetic variation that can hinder genome assembly. We used two different strategies to reduce the genetic variation among the pooled individuals, depending on the host plant of the different species. For *C. melanoneura* and *C. picta* collected on apple trees, we applied the same experimental design as in (52): In the green house, the field-caught adults were sorted into mating couples and each couple was caged on a branch of an apple tree (cultivar Golden Delicious) using nylon nets. Once the offspring of the mating couples had reached adulthood, all newly-emerged siblings were collected and stored at -20°C. Since the primary endosymbionts are vertically transmitted from mother to offspring, all siblings harbour genetically identical endosymbionts and can therefore be pooled without introducing genetic variation for the endosymbionts. In addition, the Cytochrome Oxidase I (COI) haplotype was determined for two individuals per sibling group according to (54), to determine the genetic diversity among the different

populations and mating couples. For all psyllids that do not develop on apple (*C. melanoneura* from hawthorn and the pear psyllids *C. pyri* and *C. pyricola*), adults collected in the field were immediately stored at -20°C. Subsequently, the COI haplotype was determined for numerous individuals of each species, in order to select individuals with identical COI haplotype for pooling.

DNA extraction

For each sibling group of the apple psyllids *C. melanoneura* and *C. picta* selected for long-read metagenome sequencing, DNA was extracted from two pools, each containing 4-6 whole females. For the field-caught *C. melanoneura* from hawthorn and *C. pyricola* from pears, DNA was first extracted from individual females and subjected to COI haplotype determination as outlined above. Subsequently, two pools, each combining the DNA extracts of five females with identical haplotypes, were established for each species. Only females were used since they are larger and hence provide more DNA and because we reasoned that their endosymbiont titers may be higher since the endosymbionts are harboured in two tissues, the bacteriome and the ovaries. DNA extraction was performed using a modified protocol of the PureGene Tissue kit (Qiagen). Whole insects were ground in 100 µl of Cell Lysis Solution and 5 µl Proteinase K solution and incubated at 56°C for three hours, followed by an incubation with 1.5 µl RNase A at 37°C for 30 minutes. Subsequently, proteins were precipitated by adding 35 µl of Protein Precipitation Solution. DNA was then extracted with one volume chloroform/isoamyl alcohol (24:1 v/v) and precipitated in one volume of isopropanol after overnight incubation at -20°C. The DNA pellet was resuspended in 40 µl of sterile water and incubated at 65°C for one hour to increase DNA rehydration. For *C. pyri,* DNA was extracted from a single female using the QIAamp DNA Micro Kit (Qiagen) according to the manufacturer's instructions.

395

396 **Metagenome sequencing and assembly**

397 Long-read metagenome sequencing using an Oxford Nanopore-Illumina hybrid approach was
398 performed for *C. melanoneura, C. picta* and *C. pyricola*. For each sample, one pool was used
399 for long-read sequencing on the MinION (Oxford Nanopore Technologies, UK) and the
400 second pool was used for 2 x 150 bp paired-end sequencing on an Illumina NovaSeq
401 (Macrogen). About 1.5 µg of DNA was used for library preparation using the Oxford
402 Nanopore Ligation Sequencing kit SQK-LSK 109 (Oxford Nanopore Technologies, UK).
403 Each library was sequenced on an entire R9.4 flowcell for 43-72 hours, depending on pore
404 activity. Basecalling was done using Guppy v5.0.11 (Oxford Nanopore Technologies, UK) in
405 high-accuracy mode. Low quality (< Q7) and short (< 500 bp) reads were discarded and host
406 reads were removed via mapping against a genome scaffold of *C. melanoneura* (J. M. Howie
407 & O. Rota-Stabelli, unpublished data) using Minimap2 v2.15 (55). The remaining non-host
408 reads ≥ 500 bp were assembled using Flye v2.9 (56) with the --metagenome option. Contigs
409 belonging to the endosymbionts were identified using blast (57). Reads were mapped back
410 onto the endosymbiont contigs using Minimap2 v2.15 and all mapped reads were assembled
411 again with Flye v2.9 using the same parameters. This produced two circular genomes for
412 most datasets. These genomes were first polished with Nanopore reads using Medaka v1.5.0
413 (https://github.com/nanoporetech/medaka) and subsequently with Illumina reads using
414 several iterations of Polca, a genome polisher integrated in the MaSuRCa toolkit v4.0.7 (58),
415 until no more errors were found. It is important to note that the two endosymbiont genomes
416 need to be polished together to avoid the introduction of errors in highly conserved regions
417 (e.g. ribosomal RNA operon) of the endosymbiont genome with lower coverage. In rare
418 cases, two rounds of Flye assemblies did not produce circular endosymbiont genomes. Two
419 of these genomes (CRmelAO2 and PSmelET) could be finished using alternative assembly

420 approaches: (i) Assembly with Canu v2.1.1 (59), polishing with Medaka and Polca as
421 outlined above, followed by scaffolding and gap-closing with Redundans v0.14 (60) and (ii)
422 Nanopore and Illumina reads mapping onto the complete endosymbiont genomes were
423 assembled together using SPAdes v3.15.1 (61). Genome coverage was estimated by mapping
424 the Nanopore reads onto the finished genomes during the polishing step with Medaka.
425 The metagenome of *C. pyri* was assembled from 42 million 2 x 250 paired-end reads from a
426 single female sequenced on an Illumina NovaSeq (University of Illinois, Urbana-Champaign,
427 USA). The metagenome was assembled using SPAdes v3.15.1 (61) with the --meta option
428 and default kmers. Endosymbiont contigs were identified based on coverage, which initially
429 produced two contigs for *Carsonella* and three contigs for the *Enterobacteriaceae* symbiont.
430 These contigs were ordered based on the complete genomes obtained using long-read
431 sequencing and closed after scaffolding and gap-closing with Redundans v0.14 (60). The
432 completeness of all genomes was assessed using BUSCO (gammaproteobacteria_odb10
433 dataset) (62).
434
435 **Functional genome analyses**
436 All complete endosymbiont genomes were annotated using the NCBI Prokaryotic Genome
437 Annotation Pipeline (PGAP) version 2021-07-01.build5508 (63). Circular plots of conserved
438 protein-coding genes were produced using MGCplotter
439 (https://github.com/moshi4/MGCplotter). For each endosymbiont (i.e. *Carsonella ruddii* and
440 *Psyllophila symbiotica*), all protein-coding genes were assigned to orthogroups using
441 Orthofinder v2.5.2 (64) and shared orthogroups were plotted using the package UpsetR (65).
442 Clusters of Orthologous Genes (COG) categories were determined using eggNOG-mapper
443 v2.1.7 (66) and KEGG pathway annotations were obtained using BlastKOALA v2.2 (67).
444

Phylogenomics

Orthofinder v2.5.2 (64) was used to identify single-copy orthologous genes shared (i) between all available *Carsonella* genomes and (ii) between the *Enterobacteriaceae* endosymbionts of *Cacopsylla* spp., 33 other nutritional endosymbionts from the *Gammaproteobacteria* and three *Pseudomonas entomophila* strains as outgroup. The amino acid sequences of each conserved gene were aligned using Muscle v3.1.31 (68) and the alignments were concatenated into a partitioned supermatrix using the script geneStitcher.py (https://github.com/ballesterus/Utensils/blob/master/geneStitcher.py). IQ-TREE v1.6.1 (69) was used to predict the optimal amino acid substitution model for each gene partition (70, 71) and to produce a Maximum Likelihood phylogenetic tree with 1000 bootstrap iterations. The tree was visualized in FigTree v1.4.4 (https://github.com/rambaut/figtree).

Fluorescence *in situ* hybridisation

Fluorescence *in situ* hybridization was conducted with symbiont-specific probes complementary to their 16S rRNA gene sequences (*Carsonella*: Probe Carso107 'Cy3-ATACTAAAAGGCAGATTCTTG', *Psyllophila*: Probe Psyllo118 'Cy5-TCCATTGAGTAGTTTCCCAG'). For *Carsonella*, two helper probes were used to increase the signal (helperFCarso107: 'AGCGAACGGGTGAGTAATATG', helperRCarso107: 'ACATTTCTATATACTTTCCA').

Insects preserved in ethanol were rehydrated and then postfixed in 4% paraformaldehyde for two hours at room temperature. Next, the specimens were dehydrated again by incubation in increased concentrations of ethanol and acetone, embedded in Technovit 8100 resin (Kulzer, Wehrheim, Germany), and cut into semithin sections (1 µm). The sections were then incubated overnight at room temperature in hybridization buffer containing the specific probes at a final concentration of 100 nM. After hybridization, the slides were washed three

470 times in PBS, dried, covered with ProLong Gold Antifade Reagent (Life Technologies) and
471 observed using a Zeiss LSM 900 Airyscan 2 confocal laser scanning microscope.
472
473 **DATA AVAILABILITY**
474 The genomes produced in this work are accessible in the NCBI database under BioProject
475 accessions PRJNA803426 (endosymbionts of *Cacopsylla melanoneura*), PRJNA853274
476 (endosymbionts of *Cacopsylla picta*), PRJNA853282 (endosymbionts of *Cacopsylla*
477 *pyricola*) and PRJNA853726 (endosymbionts of *Cacopsylla pyri*). New COI haplotypes for
478 *C. melanoneura, C. picta* and *C. pyricola* were deposited under accessions OQ106377 –
479 OQ106379.
480
481 **ACKNOWLEDGEMENTS**
482 We thank Domagoj Gajski, Michal Štarha and Valentina Candian for help collecting psyllids
483 in the field as well as Massimiliano Trenti for help with the psyllid rearing in the greenhouse.
484 This study was funded by a joint project of the province of Bolzano and the Austrian Science
485 Fund FWF (FIGHToplasma, I 4639-B). Anna Michalik was supported by the Polish National
486 Science Centre grant 2017/26/D/NZ8/00799.
487
488 **FIGURES AND TABLES**
489
490 **Table 1.** Properties of the endosymbiont genomes obtained in this study.
491
492 **Fig. 1. Endosymbiont genomes are highly conserved across *Cacopsylla* host species.** (a, c)
493 Circular genome plots of the twelve *Carsonella* genomes (a) and the ten *Psyllophila* genomes
494 (c) produced in this study. The three outer-most circles represent forward CDS, reverse CDS

and the ribosomal RNA operon of a reference genome (CRmelAO1 and PSmelAO1, respectively). The inner circles represent the conserved genes in all other genomes of the same taxon (the order is identical to b and d), the shading indicating the degree of sequence similarity compared to the reference genome. (b, d) Intersection plots showing the number of shared orthogroups across all *Carsonella* (b) and *Psyllophila* (d) genomes. The matrix lines are coloured according to host species.

Fig. 2. The *Enterobacteriaceae* symbiont represents a new psyllid-associated genus. Maximum-likelihood tree based on the concatenated amino acid sequence alignment of 67 single-copy orthologous genes from 46 genomes, namely the 10 *Enterobacteriaceae* endosymbionts of *Cacopsylla* spp., 33 insect endosymbionts from the *Gammaproteobacteria* and three *Pseudomonas entomophila* strains as outgroup. The *Enterobacteriaceae* endosymbionts of *Cacopsylla* spp. are colour-coded based on host species. Branch support is based on 1000 bootstrap iterations. Blue dots on branches indicate full bootstrap support.

Fig. 3. Both *Cacopsylla* endosymbionts are localized in the bacteriome. Fluorescent *in situ* hybridisation of *Carsonella* (Cy3, red) and *Psyllophila* (Cy5, green) symbionts in the bacteriomes of *Cacopsylla pyri* (A, B) and *C. melanoneura* (C). Blue represents DAPI. Scale bar = 10 µm.

Fig. 4. Metabolic complementarity between *Carsonella* and *Psyllophila.* (a) COG functional categories for the *Carsonella* and *Psyllophila* genomes show different proportions of genes involved in amino acid transport and metabolism (red). (b) Schematic representation of the metabolic complementarity between the two symbionts for the biosynthesis of essential

519 amino acids, vitamins and carotenoids. Genes present in *Carsonella* genomes are shown in
520 blue, genes present in *Psyllophila* in red.
521
522 **Fig. 5. Repeated gene losses throughout *Carsonella* evolution.** (a) Intersection plot
523 showing the distribution of orthogroups across 33 *Carsonella* genomes depending on host
524 species. (b) Maximum-likelihood tree based on the concatenated amino acid sequence
525 alignment of 119 single-copy orthologous genes present in all 33 *Carsonella* genomes. The
526 *Carsonella* strains of *Cacopsylla* spp. are colour-coded based on host species. Branch support
527 is based on 1000 bootstrap iterations. Blue dots on branches indicate full bootstrap support.
528 Losses of genes or pathways involved in the biosynthesis of essential amino acids are
529 indicated on the branches. (See Table S3 for a detailed list of genes identified based on
530 KEGG pathway annotation). The presence of known co-primary endosymbionts is indicated
531 using blue dots for amino acid-providing nutritional co-primary endosymbionts and using red
532 dots for the defensive and nutritional endosymbiont '*Ca.* Profftella armatura'.
533
534 **SUPPLEMENTARY MATERIAL**
535
536 **Supplementary Table S1.** List of insect endosymbiont genomes included in the
537 phylogenomics analysis of '*Ca*. Psyllophila symbiotica'.
538
539 **Supplementary Table S2.** Table of previously published *Carsonella* genomes used for
540 comparative and phylogenomic analyses.
541

Supplementary Table S3. Presence of genes involved in the biosynthesis of essential amino acids in all available *Carsonella* genomes based on KEGG pathway annotations using BlastKoala.

REFERENCES

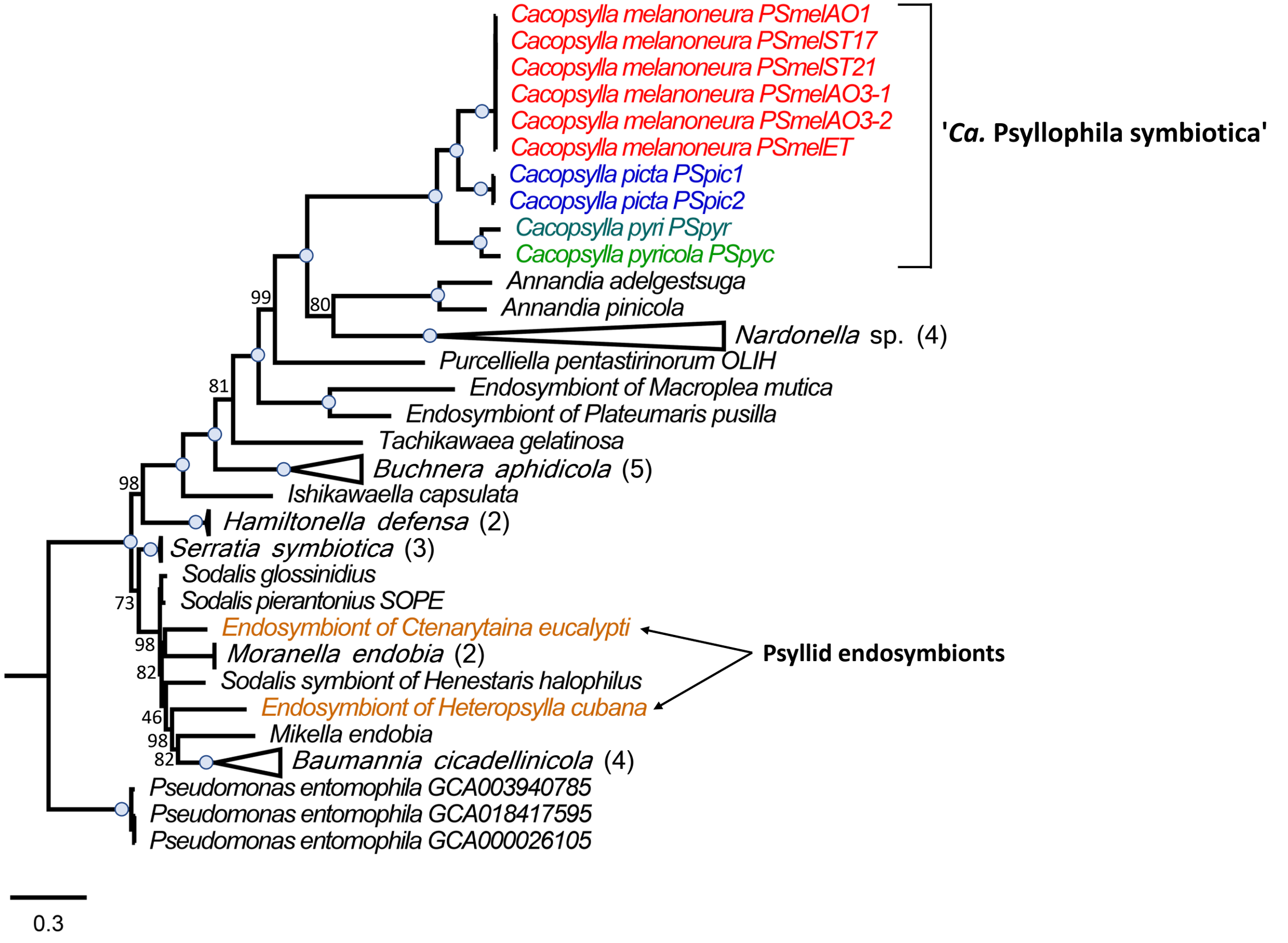
1. Moran NA, McCutcheon JP, Nakabachi A. 2008. Genomics and evolution of heritable bacterial symbionts. Annu Rev Genet 42:165-90.
2. Buchner P. 1965. Endosymbiosis of animals with plant microorganisms. Interscience Publishers, New York.
3. Baumann P. 2005. Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. Annu Rev Microbiol 59:155-89.
4. Hansen AK, Moran NA. 2011. Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. Proc Natl Acad Sci U S A 108:2849-54.
5. McCutcheon JP, von Dohlen CD. 2011. An interdependent metabolic patchwork in the nested symbiosis of mealybugs. Curr Biol 21:1366-72.
6. Bennett GM, Moran NA. 2013. Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a Phloem-feeding insect. Genome Biol Evol 5:1675-88.
7. Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M, Aksoy S. 2002. Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. Nat Genet 32:402-7.
8. Novakova E, Husnik F, Sochova E, Hypsa V. 2015. *Arsenophonus* and *Sodalis* Symbionts in Louse Flies: an Analogy to the *Wigglesworthia* and *Sodalis* System in Tsetse Flies. Appl Environ Microbiol 81:6189-99.
9. Reis F, Kirsch R, Pauchet Y, Bauer E, Bilz LC, Fukumori K, Fukatsu T, Kolsch G, Kaltenpoth M. 2020. Bacterial symbionts support larval sap feeding and adult folivory in (semi-)aquatic reed beetles. Nat Commun 11:2964.
10. McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol 10:13-26.
11. Toft C, Andersson SG. 2010. Evolutionary microbial genomics: insights into bacterial host adaptation. Nat Rev Genet 11:465-75.
12. Manzano-Marin A, Latorre A. 2016. Snapshots of a shrinking partner: Genome reduction in *Serratia symbiotica*. Sci Rep 6:32590.
13. Perez-Brocal V, Gil R, Ramos S, Lamelas A, Postigo M, Michelena JM, Silva FJ, Moya A, Latorre A. 2006. A small microbial genome: the end of a long symbiotic relationship? Science 314:312-3.
14. Monnin D, Jackson R, Kiers ET, Bunker M, Ellers J, Henry LM. 2020. Parallel Evolution in the Integration of a Co-obligate Aphid Symbiosis. Curr Biol 30:1949-1957 e6.
15. Koga R, Bennett GM, Cryan JR, Moran NA. 2013. Evolutionary replacement of obligate symbionts in an ancient and diverse insect lineage. Environ Microbiol 15:2073-81.
16. Koga R, Moran NA. 2014. Swapping symbionts in spittlebugs: evolutionary replacement of a reduced genome symbiont. ISME J 8:1237-46.
17. McCutcheon JP, Boyd BM, Dale C. 2019. The Life of an Insect Endosymbiont from the Cradle to the Grave. Curr Biol 29:R485-R495.
18. Manzano-Marin A, Szabo G, Simon JC, Horn M, Latorre A. 2017. Happens in the best of subfamilies: establishment and repeated replacements of co-obligate secondary endosymbionts within Lachninae aphids. Environ Microbiol 19:393-408.

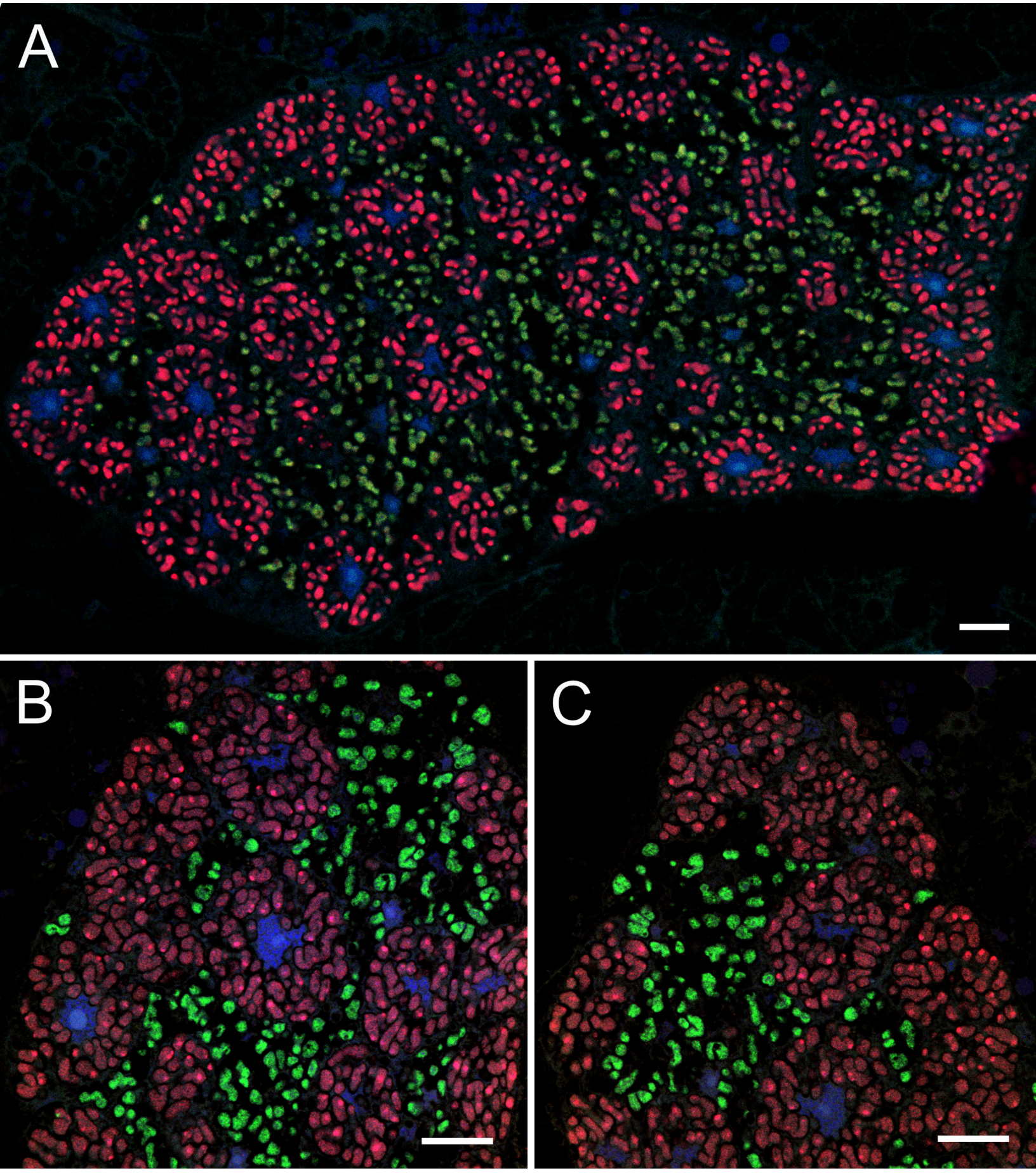
19. Ankrah NYD, Chouaia B, Douglas AE. 2018. The cost of metabolic interactions in symbioses between insects and bacteria with reduced genomes. mBio 9.
20. McCutcheon JP, Moran NA. 2007. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. Proc Natl Acad Sci U S A 104:19392-7.
21. McCutcheon JP, McDonald BR, Moran NA. 2009. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. Proc Natl Acad Sci U S A 106:15394-9.
22. McCutcheon JP, Moran NA. 2010. Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. Genome Biol Evol 2:708-18.
23. Mao M, Yang X, Poff K, Bennett G. 2017. Comparative Genomics of the Dual-Obligate Symbionts from the Treehopper, *Entylia carinata* (Hemiptera: Membracidae), Provide Insight into the Origins and Evolution of an Ancient Symbiosis. Genome Biol Evol 9:1803-1815.
24. Bennett GM, Mao M. 2018. Comparative genomics of a quadripartite symbiosis in a planthopper host reveals the origins and rearranged nutritional responsibilities of anciently diverged bacterial lineages. Environ Microbiol 20:4461-4472.
25. Michalik A, Castillo Franco D, Kobialka M, Szklarzewicz T, Stroinski A, Lukasik P. 2021. Alternative transmission patterns in independently acquired nutritional cosymbionts of *Dictyopharidae* planthoppers. mBio 12:e0122821.
26. Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. Nature 407:81-6.
27. Lamelas A, Gosalbes MJ, Manzano-Marin A, Pereto J, Moya A, Latorre A. 2011. *Serratia symbiotica* from the aphid *Cinara cedri:* a missing link from facultative to obligate insect endosymbiont. PLoS Genet 7:e1002357.
28. Manzano-Marin A, Simon JC, Latorre A. 2016. Reinventing the wheel and making it round again: Evolutionary convergence in *Buchnera-Serratia* symbiotic consortia between the distantly related *Lachninae* aphids *Tuberolachnus salignus* and *Cinara cedri*. Genome Biol Evol 8:1440-58.
29. Yorimoto S, Hattori M, Kondo M, Shigenobu S. 2022. Complex host/symbiont integration of a multi-partner symbiotic system in the eusocial aphid *Ceratovacuna japonica*. iScience 25:105478.
30. Thao ML, Moran NA, Abbot P, Brennan EB, Burckhardt DH, Baumann P. 2000. Cospeciation of psyllids and their primary prokaryotic endosymbionts. Appl Environ Microbiol 66:2898-905.
31. Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, Moran NA, Hattori M. 2006. The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. Science 314:267.
32. Tamames J, Gil R, Latorre A, Pereto J, Silva FJ, Moya A. 2007. The frontier between cell and organelle: genome analysis of *Candidatus* Carsonella ruddii. BMC Evol Biol 7:181.
33. Sloan DB, Moran NA. 2012. Genome reduction and co-evolution between the primary and secondary bacterial symbionts of psyllids. Mol Biol Evol 29:3781-92.
34. Hall AA, Morrow JL, Fromont C, Steinbauer MJ, Taylor GS, Johnson SN, Cook JM, Riegler M. 2016. Codivergence of the primary bacterial endosymbiont of psyllids versus host switches and replacement of their secondary bacterial endosymbionts. Environ Microbiol 18:2591-603.
35. Fromont C, Riegler M, Cook JM. 2016. Phylogeographic analyses of bacterial endosymbionts in fig homotomids (Hemiptera: Psylloidea) reveal codiversification of both primary and secondary endosymbionts. FEMS Microbiol Ecol 92.
36. Sloan DB, Nakabachi A, Richards S, Qu J, Murali SC, Gibbs RA, Moran NA. 2014. Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. Mol Biol Evol 31:857-71.
37. Fukatsu T, Nikoh N. 1998. Two intracellular symbiotic bacteria from the mulberry psyllid *Anomoneura mori* (Insecta, Homoptera). Appl Environ Microbiol 64:3599-606.
38. Subandiyah S, Nikoh N, Tsuyumu S, Somowiyarjo S, Fukatsu T. 2000. Complex endosymbiotic microbiota of the citrus psyllid *Diaphorina citri* (Homoptera: Psylloidea). Zoolog Sci 17:983-989.
39. Nakabachi A, Ueoka R, Oshima K, Teta R, Mangoni A, Gurgui M, Oldham NJ, van Echten-Deckert G, Okamura K, Yamamoto K, Inoue H, Ohkuma M, Hongoh Y, Miyagishima SY,

644 Hattori M, Piel J, Fukatsu T. 2013. Defensive bacteriome symbiont with a drastically reduced
645 genome. Curr Biol 23:1478-84.
646 40. Nakabachi A, Piel J, Malenovsky I, Hirose Y. 2020. Comparative genomics underlines
647 multiple roles of *Profftella*, an obligate symbiont of psyllids: Providing toxins, vitamins, and
648 carotenoids. Genome Biol Evol 12:1975-1987.
649 41. Morrow JL, Hall AAG, Riegler M. 2017. Symbionts in waiting: the dynamics of incipient
650 endosymbiont complementation and replacement in minimal bacterial communities of
651 psyllids. Microbiome 5:58.
652 42. Morrow JL, Om N, Beattie GAC, Chambers GA, Donovan NJ, Liefting LW, Riegler M,
653 Holford P. 2020. Characterization of the bacterial communities of psyllids associated with
654 Rutaceae in Bhutan by high throughput sequencing. BMC Microbiol 20:215.
655 43. Kwak Y, Sun P, Meduri VR, Percy DM, Mauck KE, Hansen AK. 2021. Uncovering
656 symbionts across the psyllid tree of life and the discovery of a new *Liberibacter* species,
657 "*Candidatus* Liberibacter capsica". Front Microbiol 12:739763.
658 44. Nakabachi A, Inoue H, Hirose Y. 2022. Microbiome analyses of 12 psyllid species of the
659 family Psyllidae identified various bacteria including *Fukatsuia* and *Serratia symbiotica*,
660 known as secondary symbionts of aphids. BMC Microbiol 22:15.
661 45. Štarhová Serbina L, Gajski D, Pafco B, Zurek L, Malenovsky I, Novakova E, Schuler H,
662 Dittmer J. 2022. Microbiome of pear psyllids: A tale about closely related species sharing
663 their endosymbionts. Environ Microbiol doi:10.1111/1462-2920.16180.
664 46. Schuler H, Dittmer J, Borruso L, Galli J, Fischnaller S, Anfora G, Rota-Stabelli O, Weil T,
665 Janik K. 2022. Investigating the microbial community of *Cacopsylla* spp. as potential factor
666 in vector competence of phytoplasma. Environ Microbiol 24:4771-4786.
667 47. Dial DT, Weglarz KM, Aremu AO, Havill NP, Pearson TA, Burke GR, von Dohlen CD.
668 2022. Transitional genomes and nutritional role reversals identified for dual symbionts of
669 adelgids (Aphidoidea: Adelgidae). ISME J 16:642-654.
670 48. Anbutsu H, Moriyama M, Nikoh N, Hosokawa T, Futahashi R, Tanahashi M, Meng XY,
671 Kuriwada T, Mori N, Oshima K, Hattori M, Fujie M, Satoh N, Maeda T, Shigenobu S, Koga
672 R, Fukatsu T. 2017. Small genome symbiont underlies cuticle hardness in beetles. Proc Natl
673 Acad Sci U S A 114:E8382-E8391.
674 49. Sloan DB, Moran NA. 2012. Endosymbiotic bacteria as a source of carotenoids in whiteflies.
675 Biol Lett 8:986-9.
676 50. Thao ML, Clark MA, Baumann L, Brennan EB, Moran NA, Baumann P. 2000. Secondary
677 endosymbionts of psyllids have been acquired multiple times. Curr Microbiol 41:300-4.
678 51. Morris JJ, Lenski RE, Zinser ER. 2012. The Black Queen Hypothesis: evolution of
679 dependencies through adaptive gene loss. MBio 3.
680 52. Corretto E, Trenti M, Štarhová Serbina L, Howie JM, Dittmer J, Kerschbamer C, Candian V,
681 Tedeschi R, Janik K, Schuler H. 2023. Multiple factors driving the acquisition efficiency of
682 apple proliferation phytoplasma in *Cacopsylla melanoneura*. Research Square preprint
683 doi:https://doi.org/10.21203/rs.3.rs-2646791/v1.
684 53. Jarausch B, Tedeschi R, Sauvion N, Gross J, Jarausch W. 2019. Psyllid vectors, p 53–78. *In*
685 Bertaccini A, Weintraub P, Rao R, N. M (ed), Phytoplasmas: Plant Pathogenic Bacteria II.
686 Springer, Singapore.
687 54. Oettl S, Schlink K. 2015. Molecular identification of two vector species, *Cacopsylla*
688 *melanoneura* and *Cacopsylla picta* (Hemiptera: Psyllidae), of Apple Proliferation disease and
689 further common psyllids of Northern Italy. J Econ Entomol 108:2174-83.
690 55. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094-
691 3100.
692 56. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, Kuhn K, Yuan J,
693 Polevikov E, Smith TPL, Pevzner PA. 2020. metaFlye: scalable long-read metagenome
694 assembly using repeat graphs. Nat Methods 17:1103-1110.
695 57. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search
696 tool. J Mol Biol 215:403-10.
697 58. Zimin AV, Salzberg SL. 2020. The genome polishing tool POLCA makes fast and accurate
698 corrections in genome assemblies. PLoS Comput Biol 16:e1007981.

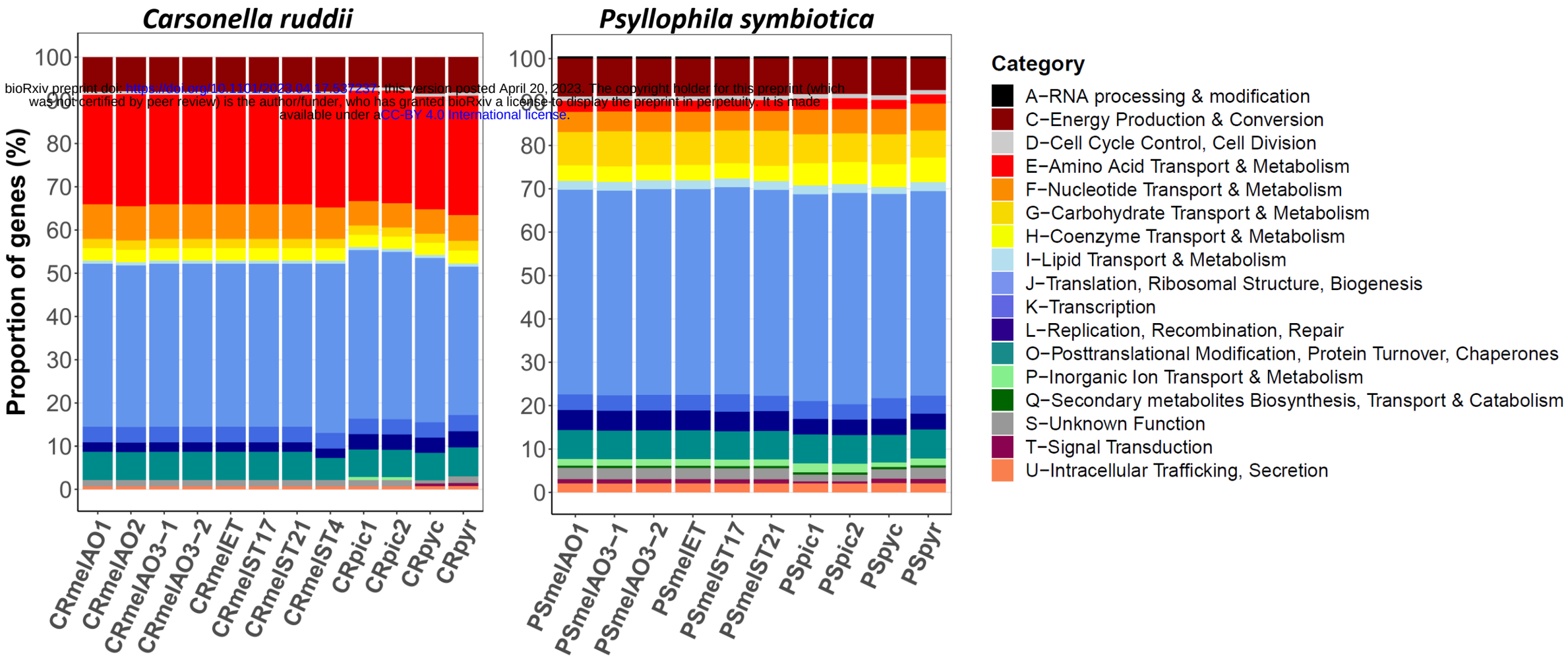
699 59. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable
700 and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome
701 Res 27:722-736.
702 60. Pryszcz LP, Gabaldon T. 2016. Redundans: an assembly pipeline for highly heterozygous
703 genomes. Nucleic Acids Res 44:e113.
704 61. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes De
705 Novo Assembler. Curr Protoc Bioinformatics 70:e102.
706 62. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO:
707 assessing genome assembly and annotation completeness with single-copy orthologs.
708 Bioinformatics 31:3210-2.
709 63. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze
710 A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline.
711 Nucleic Acids Res 44:6614-24.
712 64. Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative
713 genomics. Genome Biol 20:238.
714 65. R Core Team. 2021. R: A language and environment for statistical computing, R Foundation
715 for Statistical Computing, Vienna, Austria, https://www.R-project.org/.
716 66. Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-
717 mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the
718 Metagenomic Scale. Mol Biol Evol 38:5825-5829.
719 67. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG Tools for
720 Functional Characterization of Genome and Metagenome Sequences. J Mol Biol 428:726-
721 731.
722 68. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
723 throughput. Nucleic Acids Res 32:1792-7.
724 69. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A,
725 Lanfear R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference
726 in the Genomic Era. Mol Biol Evol 37:1530-1534.
727 70. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder:
728 fast model selection for accurate phylogenetic estimates. Nat Methods 14:587-589.
729 71. Chernomor O, von Haeseler A, Minh BQ. 2016. Terrace Aware Data Structure for
730 Phylogenomic Inference from Supermatrices. Syst Biol 65:997-1008.

731

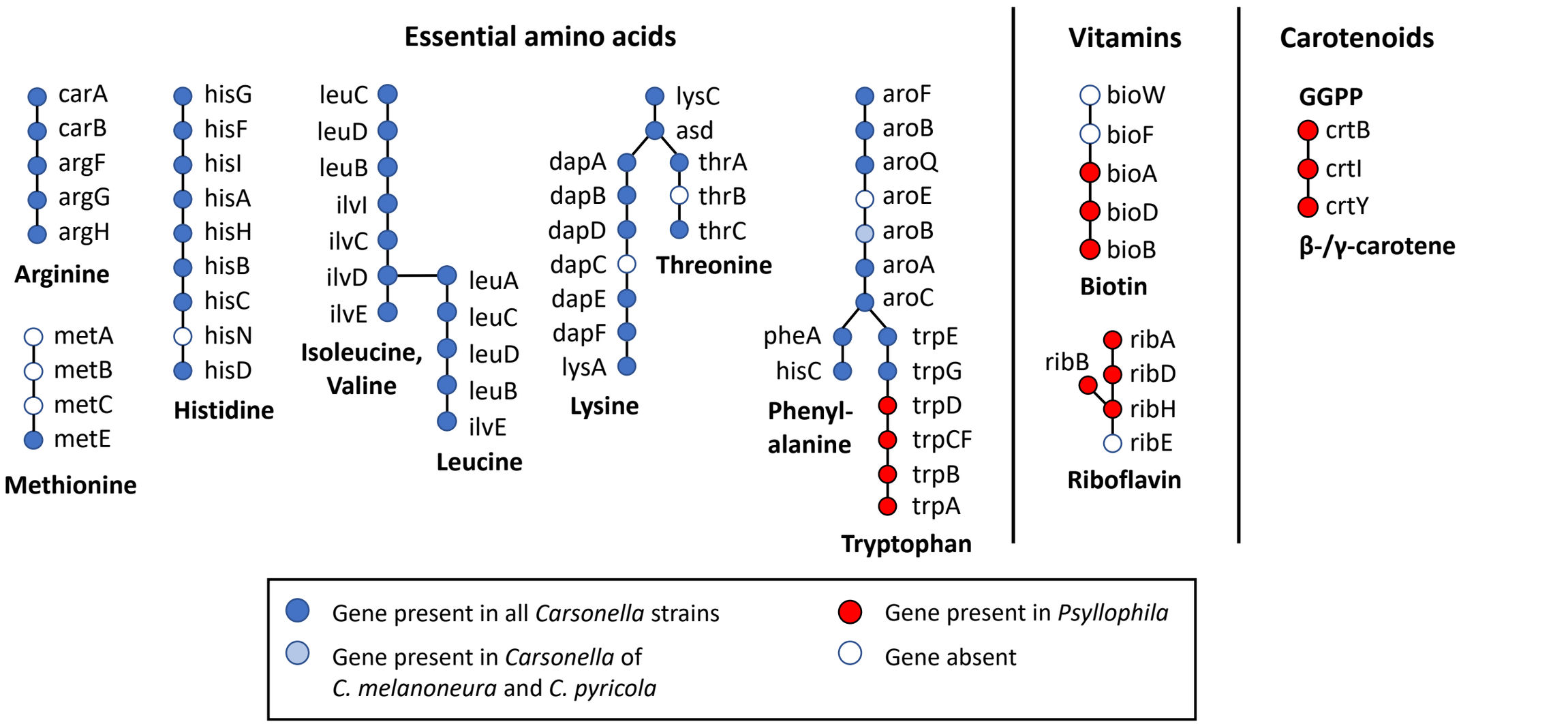

Cacopsylla melanoneura PSmelAO1
Cacopsylla melanoneura PSmelST17
Cacopsylla melanoneura PSmelST21
Cacopsylla melanoneura PSmelAO3-1
Cacopsylla melanoneura PSmelAO3-2
Cacopsylla melanoneura PSmelET
Cacopsylla picta PSpic1
Cacopsylla picta PSpic2
Cacopsylla pyri PSpyr
Cacopsylla pyricola PSpyc
'Ca. Psyllophila symbiotica'
Annandia adelgestsuga
Annandia pinicola
Nardonella sp. (4)
Purcelliella pentastirinorum OLIH
Endosymbiont of Macroplea mutica
Endosymbiont of Plateumaris pusilla
Tachikawaea gelatinosa
Buchnera aphidicola (5)
Ishikawaella capsulata
Hamiltonella defensa (2)
Serratia symbiotica (3)
Sodalis glossinidius
Sodalis pierantonius SOPE
Endosymbiont of Ctenarytaina eucalypti
Moranella endobia (2)
Sodalis symbiont of Henestaris halophilus
Endosymbiont of Heteropsylla cubana
Mikella endobia
Baumannia cicadellinicola (4)
Psyllid endosymbionts
Pseudomonas entomophila GCA003940785
Pseudomonas entomophila GCA018417595
Pseudomonas entomophila GCA000026105
99
80
81
98
73
98
82
46
98
82
0.3



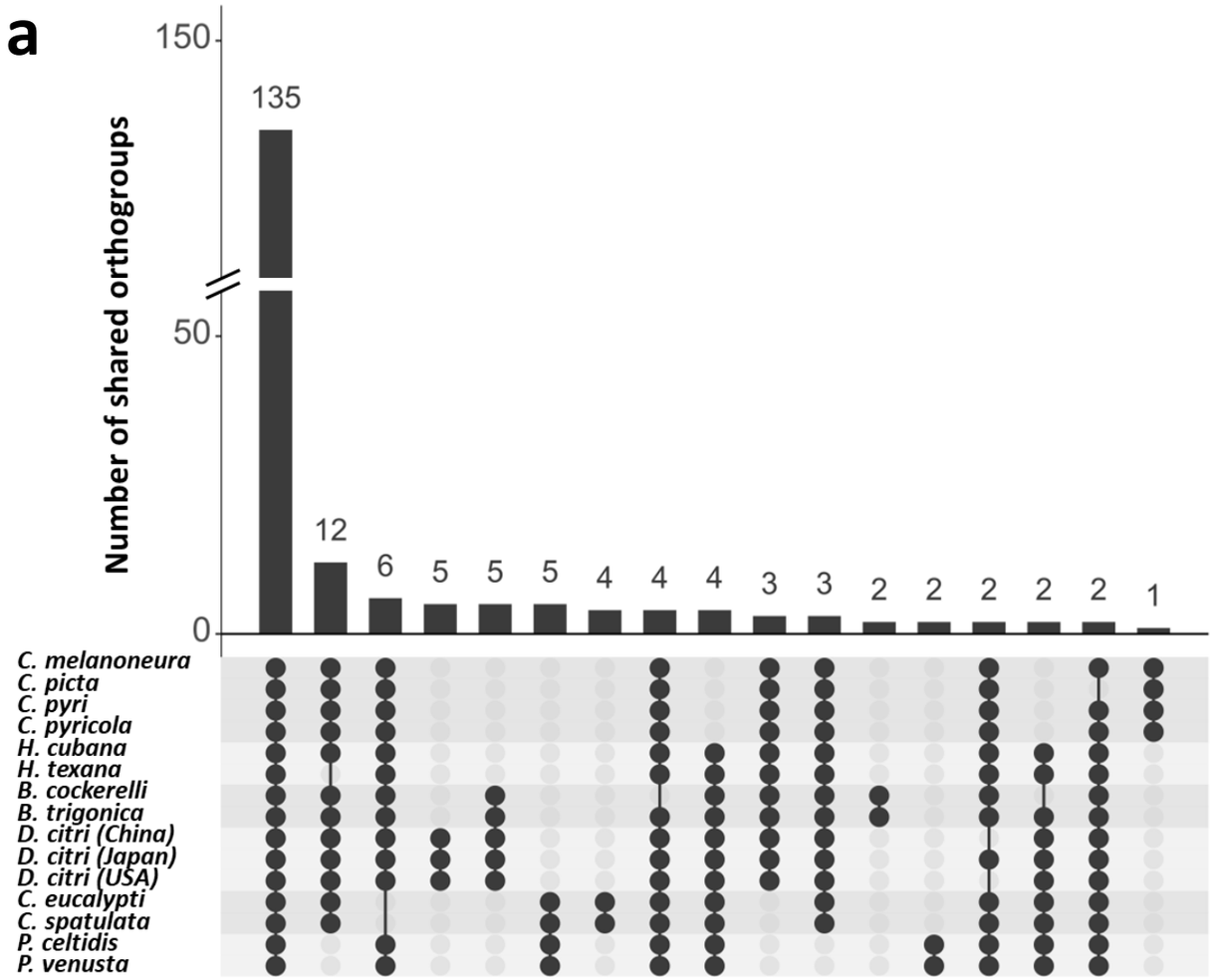
a



b



a


Number of shared orthogroups
150
50
0
135
12
6
5
5
5
4
4
4
3
3
2
2
2
2
2
1
C. melanoneura
C. picta
C. pyri
C. pyricola
H. cubana
H. texana
B. cockerelli
B. trigonica
D. citri (China)
D. citri (Japan)
D. citri (USA)
C. eucalypti
C. spatulata
P. celtidis
P. venusta

b

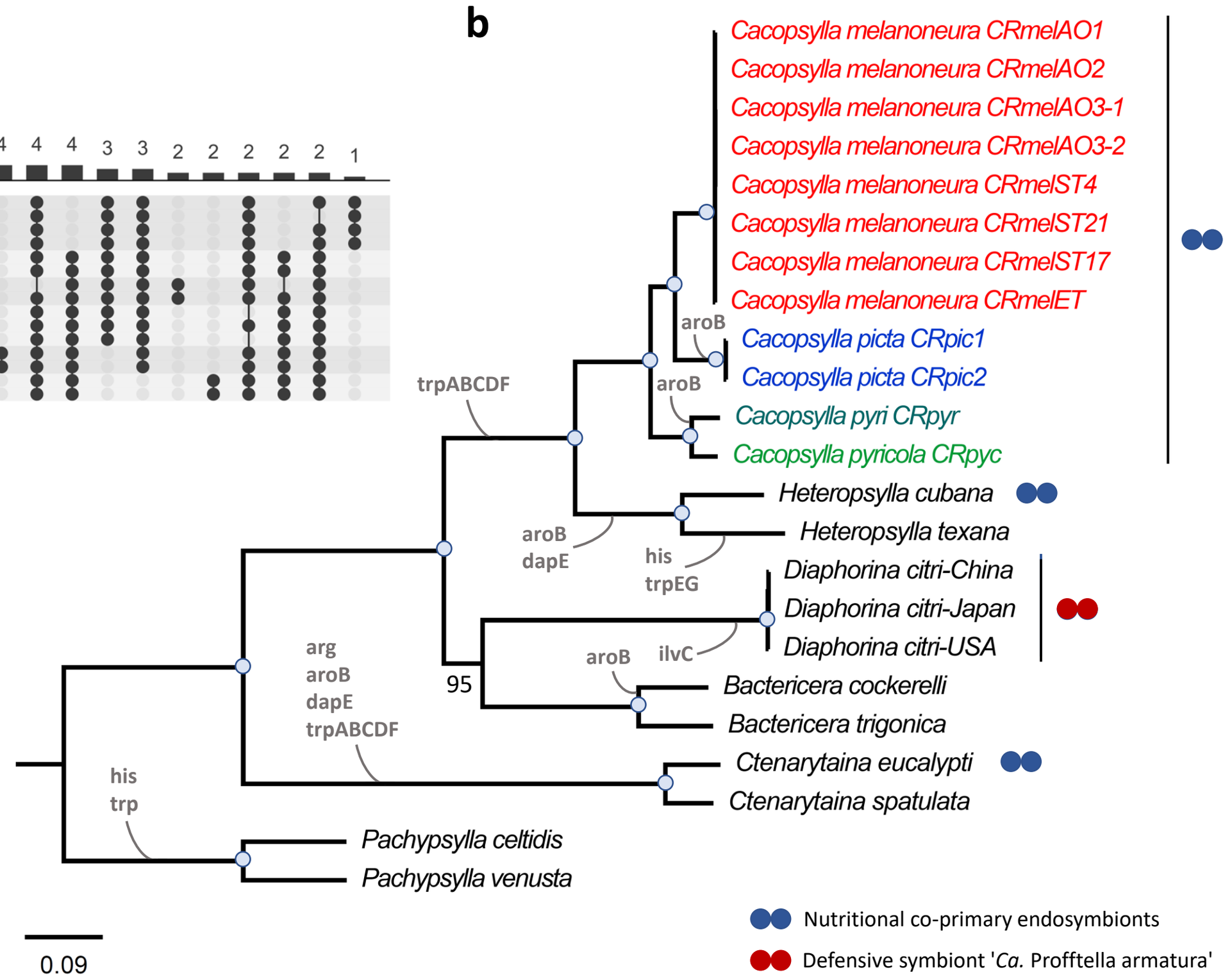

Cacopsylla melanoneura CRmelAO1
Cacopsylla melanoneura CRmelAO2
Cacopsylla melanoneura CRmelAO3-1
Cacopsylla melanoneura CRmelAO3-2
Cacopsylla melanoneura CRmelST4
Cacopsylla melanoneura CRmelST21
Cacopsylla melanoneura CRmelST17
Cacopsylla melanoneura CRmelET
Cacopsylla picta CRpic1
Cacopsylla picta CRpic2
Cacopsylla pyri CRpyr
Cacopsylla pyricola CRpyc
Heteropsylla cubana
Heteropsylla texana
Diaphorina citri-China
Diaphorina citri-Japan
Diaphorina citri-USA
Bactericera cockerelli
Bactericera trigonica
Ctenarytaina eucalypti
Ctenarytaina spatulata
Pachypsylla celtidis
Pachypsylla venusta
aroB
aroB
trpABCDF
aroB
dapE
his
trpEG
aroB
ilvC
95
arg
aroB
dapE
trpABCDF
his
trp
0.09
Nutritional co-primary endosymbionts
Defensive symbiont 'Ca. Profftella armatura'

Table 1. Properties of the complete endosymbiont genomes obtained in this study. IT=Italy, CZ= Czech Republic. ND=Not determined. Empty columns for *Psyllophila* indicate that the genome could not be assembled due to insufficient coverage.

Host species	***Cacopsylla melanoneura***								***Cacopsylla picta***		***Cacopsylla pyricola***	***Cacopsylla pyri***
COI haplotype	mel01[a]	mel01[a]	mel01[a]	mel01[a]	mel01[a]	mel01[a]	mel21[b]	mel09[c]	pic18[d]	pic01[e]	pyc01[f]	ND
Origin	Aosta Valley (IT)	Aosta Valley (IT)	Aosta Valley (IT)	Aosta Valley (IT)	South Tyrol (IT)	South Tyrol (IT)	South Tyrol (IT)	South Tyrol (IT)	Trentino (IT)	Trentino (IT)	Starý Lískovec (CZ)	Litenčice (CZ)
Host plant	Apple	Apple	Hawthorn	Hawthorn	Apple	Apple	Apple	Apple	Apple	Apple	Pear	Pear
Sequencing method	Nanopore + Illumina	Nanopore + Illumina	Nanopore + Illumina	Nanopore + Illumina	Nanopore + Illumina	Nanopore + Illumina	Nanopore + Illumina	Nanopore + Illumina	Nanopore + Illumina	Nanopore + Illumina	Nanopore + Illumina	Illumina only
***Carsonella* Strain**	**CRmelAO1**	**CRmelAO2**	**CRmelAO3-1**	**CRmelAO3-2**	**CRmelET**	**CRmelST4**	**CRmelST17**	**CRmelST21**	**CRpic1**	**CRpic2**	**CRpyc**	**CRpyr**
Length (bp)	169,165	169,273	169,079	169,081	169,051	169,046	169,064	169,056	171,569	171,920	168,917	169,391
%GC	15.52	15.53	15.52	15.52	15.52	15.52	15.52	15.52	15.42	15.40	14.98	15.19
Genes	216	216	215	215	215	216	215	214	212	211	220	214
CDS	186	186	185	185	185	183	185	184	183	182	190	184
Pseudogenes	0	0	0	0	0	3	0	0	0	0	0	0
rRNA	3	3	3	3	3	3	3	3	3	3	3	3
tRNA	27	27	27	27	27	27	27	27	26	26	27	27
***Psyllophila* strain**	**PSmelAO1**		**PSmelAO3-1**	**PSmelAO3-2**	**PSmelET**		**PSmelST17**	**PSmelST21**	**PSpic1**	**PSpic2**	**PSpyc**	**PSpyr**
Length (bp)	237,114		236,266	235,918	231,767		236,072	235,744	225,730	225,865	222,755	221,414
%GC	18.26		18.29	18.33	18.60		18.32	18.34	17.80	17.79	17.36	17.30
Genes	239		240	239	239		242	243	240	240	241	240
CDS	205		206	205	205		207	208	207	207	207	205
Pseudogenes	2		2	2	2		3	3	1	1	2	3
rRNA	3		3	3	3		3	3	3	3	3	3
tRNA	27		27	27	27		27	27	27	27	27	27
ncRNA	2		2	2	2		2	2	2	2	2	2

[a] Identical to Accession no. KM206163

[b] Accession no. OQ631077

[c] Identical to Accession no. KM206167

[d] Accession no. OQ631078

[e] Identical to Accession no. KM206174

[f] Accession no. OQ631079