bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Evolutionary recent dual obligatory symbiosis among adelgids indicates a transition between fungus and insect associated lifestyles

Gitta Szabó[1,2,*], Frederik Schulz[1,3], Alejandro Manzano-Marín[1], Elena Rebecca Toenshoff[1,4], Matthias Horn[1]

[1]Centre for Microbiology and Environmental Systems Science, University of Vienna, Vienna, Austria
[2]Current address: Department of Internal Medicine and Oncology, Semmelweis University, Budapest, Hungary
[3]Current address: US Department of Energy (DOE) Joint Genome Institute, Berkeley, California, USA
[4]Current address: Institute of Molecular Biology and Biophysics, ETH Zurich, Zurich, Switzerland

[*] Corresponding author

Running title: Evolutionary transition between fungus and insect endosymbionts

Competing Interests: The authors declare no competing interests

Corresponding author: Gitta Szabó; Email: gitta.szabo@yahoo.com

ABSTRACT

Adelgids (Insecta: Hemiptera: Adelgidae) form a small group of insects but harbor a surprisingly diverse set of bacteriocyte-associated endosymbionts, which suggest multiple replacement and acquisition of symbionts over evolutionary time. Specific pairs of symbionts have been associated with adelgid lineages specialized on

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

different secondary host conifers. Using a metagenomic approach, we investigated the symbiosis of the *Adelges laricis/tardus* species complex containing betaproteobacterial ('*Candidatus* Vallotia tarda') and gammaproteobacterial ('*Candidatus* Profftia tarda') symbionts. Genomic characteristics and metabolic pathway reconstructions revealed that *Vallotia* and *Profftia* are evolutionary young endosymbionts, which complement each other's role in essential amino acid production. Phylogenomic analyses and a high level of genomic synteny indicate an origin of the betaproteobacterial symbiont from endosymbionts of *Rhizopus* fungi. This evolutionary transition was accompanied with substantial loss of functions related to transcription regulation, secondary metabolite production, bacterial defense mechanisms, host infection and manipulation. The transition from fungus to insect endosymbionts extends our current framework about evolutionary trajectories of host-associated microbes.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Plant-sap feeding insects harbor bacterial endosymbionts, which are of great importance in their host ecology and serve as a model for studying microbe-host relationships and genome evolution of host restricted bacteria (1–3). Adelgids (Insecta: Hemiptera: Adelgidae) live on Pinaceae conifers and feed on phloem sap or parenchyma cells (4,5). The group has nearly seventy species and is sister to the families of phylloxerans (Phylloxeridae) and aphids (Aphididae) within the suborder Sternorrhyncha (6). Some adelgid species, such as the balsam woolly adelgid (*Adelges piceae*) and the hemlock woolly adelgid (*A. tsugae*) are well-known forest pests and represent severe threats to firs and hemlocks (7,8).

Adelgids have a complex multigenerational life cycle, which typically involves sexual generations and an alternation between spruce (*Picea*), which is the primary host, and another, secondary conifer host (*Abies, Pinus, Larix, Pseudotsuga,* or *Tsuga*). However, other adelgids reproduce asexually in all generations on either of the host genera (5).

Similarly to other plant-sap feeding insects, adelgids harbor maternally inherited bacterial symbionts within specialized cells, so-called bacteriocytes, which form a large bacteriome in the abdomen (9–14). Although the function of these bacterial partners remains largely unexplored, they are expected to provide essential amino acids and B vitamins scarce in the plant sap diet, similarly to obligate endosymbionts of other plant-sap feeding insects (2,15). Besides these obligate nutritional endosymbionts, non-essential facultative symbionts might also occur within the bacteriome (or in other tissues), which can provide selective fitness benefits to insects such as protection against parasites and fungal pathogens, increased

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

heat tolerance, or expansion of host plant range (16–18). Similarly to obligate mutualists, facultative symbionts are usually maternally inherited, but can also spread horizontally within and between insect species via mating (19), parasites (20), and food source such as plant tissues (21). Few examples of newly emerged bacteriocyte-associated symbionts of herbivorous insects pinpoint their source from plant-associated bacteria, such as *Erwinia* in *Cinara* aphids (22), gut bacteria, such as cultivable *Serratia symbiotica* strains colonizing the gut of *Aphis* aphids (23), and other free-living bacteria such as a *Sodalis* strain (HS) isolated from human wounds and being akin to primary endosymbionts of *Sitophilus* weevils (24).

Interestingly, two types of bacteriocyte-associated symbionts have been identified in all populations and life stages of most adelgid species (9–14). These symbionts belong to at least six different lineages within the Gammaproteobacteria or the Betaproteobacteria. *A. tsugae* populations also contain a universal *Pseudomonas* symbiont free in the hemocoel together with the bacteriocyte-associated symbionts (9,10,25). Remarkably, specific pairs of symbionts correspond to distinct lineages of adelgids specialized to one of the five secondary host tree genera (10,14). A gammaproteobacterial symbiont lineage involving '*Candidatus* Annandia adelgestsugas' and '*Candidatus* Annandia pinicola' (hereafter *Annandia*), is present in both *A. tsugae* and *Pineus* species, and was likely already associated with ancestral adelgids before diversification into the two major adelgid lineages, *Adelges* and *Pineus*, over 87 million years ago (5,9,10,14). Nevertheless, this putatively ancient symbiont lineage is missing from other adelgids, and the high diversity of symbionts within this small group of insects suggests an evolutionary history involving multiple acquisitions and replacement of bacterial partners (10,13,14). This is in sharp contrast

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

to the case in the aphid sister group, where most species have tightly co-evolved with a single obligate symbiont, *Buchnera aphidicola,* for over 180 million years (2). In the case of adelgids, it has been postulated that loss of the ancestral nutritional symbiont lineage and repeated replacements of bacterial partners might be due to fluctuating selective pressure on essential symbiotic functions during evolution as a consequence of repeated emergence of host-alternating lifestyles and feeding on nutrient-rich parenchyma versus nutrient-poor phloem on the primary and secondary host trees, respectively (10,25).

To date, whole genome sequences of adelgid endosymbionts are available for only one species: the hemlock woolly adelgid, *A. tsugae*. Metabolic potential and genomic characteristics of *Annandia* resemble those of long-term obligate intracellular symbionts. However, *Annandia* has lost many genes in essential amino acid synthesis. The accompanying, evolutionary more recent, *Pseudomonas* symbiont can complement these missing capabilities and thus has a co-obligatory status in maintaining the symbiosis (25). In addition to this obligate dual endosymbiotic system, analysis of a genome fragment of a gammaproteobacterial symbiont ('*Candidatus* Steffania adelgidicola') of the *Adelges nordmannianae/piceae* species complex revealed a metabolically versatile, putatively evolutionary young endosymbiont in this adelgid lineage (12). Further genomic data on the symbionts would help to infer the history of association of adelgids with distinct bacterial groups.

Here, we investigate the bacterial symbionts of the *A. laricis/tardus* species complex using a metagenomic approach and ask what is the function and putative origin of the dual symbiosis in this lineage of adelgids. *A. laricis* and *A. tardus* are

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

115 morphologically and genetically hardly distinguishable species of adelgids (4,26).
116 They contain betaproteobacterial and gammaproteobacterial symbionts, '*Candidatus*
117 Vallotia tarda' and '*Candidatus* Profftia tarda' (hereinafter *Vallotia* and *Profftia*),
118 respectively. Both symbionts are rod-shaped and are located intermingled inside the
119 same bacteriocytes. *Profftia*-related symbionts have only been found in larch-
120 associated, while *Vallotia* symbionts occur in both larch and Douglas-fir-associated
121 lineages of adelgids. Although host-symbiont co-speciation could not be fully
122 resolved with confidence yet, the dual obligatory status of *Profftia* and *Vallotia* in the
123 symbiosis seems to be possible given their common occurrence across different
124 populations and life stages of adelgids (10,13).

125 Our results demonstrate that both bacteriocyte-associated symbionts are
126 evolutionary recent partners of adelgids complementing each other's role in essential
127 amino acid biosynthesis. Notably, phylogenomic analyses revealed a close
128 relationship of *Vallotia* with endosymbionts of *Rhizopus* fungi. Detailed analysis of
129 genomic synteny and gene content indicated an evolutionary transition from fungus to
130 insect symbiosis accompanied by a substantial loss of functions in the insect symbiont
131 especially in transcription regulation, secondary metabolite production, host infection
132 and manipulation.

133 ## Materials and Methods

134 ### Sampling

135 Spruce (*Picea*) branches with galls of adelgids (4) were collected near
136 Klausen-Leopoldsdorf, Austria (Figure S1). Galls were stored at -80°C in the lab for
137 subsequent genomic DNA isolation.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

DNA isolation

Before DNA isolation, adelgids were collected from the galls using teasing needles. The insects were washed twice in buffer A + EDTA solution (35mM Tris-HCl, 250 mM sucrose, 25 mM KCl, 10 mM $MgCl_2$, 250 mM EDTA; pH 7.5) and were subsequently homogenized in fresh solution with a plastic pestle. The suspension was then sequentially filtered through 53 and 30 µm pore size meshes and 5 µm membrane syringe filters. Samples were centrifuged at 7,000 rpm for 5 min at 4°C and supernatants were discarded. Pellets were re-suspended in buffer A and centrifuged again at 7,000 rpm for 5 min at 4°C. This washing step was repeated once and pellets were next re-suspended in 1xTE buffer (10 mM Tris-HCl, 1 mM EDTA; pH 7.5). High-molecular-weight DNA was isolated by an SDS-based DNA extraction method using 1% cetyltrimethylammonium bromide and 1.5% polyvinylpyrrolidone in the extraction buffer (27). DNA samples were stored at -20°C.

Sequencing and genome assembly

A paired-end library was sequenced on a HighSeq 2000 Illumina sequencer. Sequencing reads were quality filtered and trimmed with PRINSEQ (28), and were assembled with SPAdes v3.1 (29). Using a subset of 30 million read pairs, a single contig representing the circular *Profftia* chromosome was obtained with 52-fold coverage, while the assembly of the *Vallotia* genome remained fragmented probably due to repetitive sequences. To improve this assembly, reads were mapped on the *Profftia* genome using the Burrows-Wheeler Alignment (BWA) tool and the BWA-MEM algorithm (30), and matching sequences were removed from further analysis. A novel assembly with the remaining reads resulted in 14 contigs longer than 1000 bp.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

These contigs were further analyzed against a custom protein database containing single-copy markers found in 99% of prokaryote genomes using blastx (31) and phylogenetic information of the best hits was assessed in Megan v4.70.4 (32). Ribosomal RNAs were inferred by RNAmmer (33). Based on these, eight contigs belonging to the *Vallotia* genome were identified. Seven contigs represent the *Vallotia* chromosome with ~212-fold coverage. In addition, a single contig obtained with 169-fold coverage corresponds to a putative circular plasmid of this endosymbiont. The remaining contigs, all shorter than 5,500 bp, were judged to belong to unrelated taxa, based on differences in GC content, coverage and taxonomic affiliation of best blastx hits in the NCBI non-redundant protein database (nr).

Genome analysis

The putative origin of replication was identified with GenSkew (http://genskew.csb.univie.ac.at/). We used the ConsPred genome annotation pipeline for gene prediction and annotation (34). Genome annotations were curated with the help of the UniPro UGENE software (35). We identified pseudogenes by using the intergenic and hypothetical protein regions as queries in blastx searches against nr and the UniProt Swiss-Prot database with an e-value < 1e-3 confidence threshold. Pseudogenes were identified as remains of genes, which were either truncated (having a length < 80% of the reference) or were interrupted by internal stop codons and/or frameshift mutations. Pseudogene coordinates were set according to the best blast hit in the UniProt Swiss-Prot database, if applicable.

The presence of mobile genetic elements was inferred with blastn and blastx searches against the ISfinder database (36). Metabolic pathways were explored with

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the help of the Ecocyc, Biocyc, and Metacyc databases (37) and the Pathway Tools software (38). Orthologous proteins shared by the relevant genomes or unique to either of the symbionts were identified by using OrthoMCL with a 1e-5 e-value threshold (39). Distribution of the predicted proteins among the main functional categories was explored by using eggNOG-mapper v2 (40,41) with the DIAMOND sequence comparison option and a 1e-3 e-value threshold (42). Genome alignments were performed by Mauve (43). The *Vallotia* contigs were reordered using the chromosome and plasmid sequences of the closely related fungus endosymbiont, *Mycetohabitans rhizoxinica,* as references (accession numbers: FR687359 and FR687360, respectively). Collinear genomic regions and genome rearrangements were visualized by Circos based on single-copy shared genes (44). A close-up of syntenic regions was created by using the Easyfig tool (45). The assembled genomes have been submitted to European Nucleotide Archive under accession number [submission in process, to be added].

Phylogenetic analyses

A phylogenomic approach was used to infer the phylogenetic positions of the endosymbionts. Protein sequences of closely related species within the Burkholderiales and the Enterobacteriales were collected from the Assembly database of NCBI. Single copy marker genes were identified by Phyla-AMPHORA (46) using the Brandon Seah (2014) Phylogenomics-tools (online: https://github.com/kbseah/phylogenomics-tools). Individual sets of genes were aligned with Muscle 3.8.31 (47). Poorly aligned positions were removed with Gblocks 0.91b (48) using default parameters apart from the following settings: allowed gap

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

207 positions with half, the minimum length of a block was 5. Alignments of 108 and 45
208 proteins were concatenated and used for the calculation of phylogenetic trees for
209 *Vallotia* and *Profftia*, respectively.

210 For both endosymbionts, we generated maximum likelihood trees with IQ-
211 TREE after selecting the best-fit substitution models with ModelFinder as available at
212 http://iqtree.cibiv.univie.ac.at (49–51). SH-like approximate likelihood ratio test (SH-
213 aLRT) and ultrafast bootstrap support values, both based on 1000 iterations, were
214 calculated (52). The best-fit models were LG+F+R5 and LG+R5 for the *Vallotia* and
215 *Profftia* tree, respectively. Besides, Bayesian phylogenetic analyses were performed
216 by MrBayes 3.2.7a (53) with the LG+I+G model and 4 gamma categories on the
217 CIPRES Science Gateway v.3.3. web interface (54). Two runs, each with 4 chains
218 were performed until convergence diagnostics fell below 0.01. A 50% majority
219 consensus tree was created with a relative burn-in of 25%.
220

221 ## Results and discussion

222 ### *Vallotia* and *Profftia* are evolutionary young symbionts of adelgids

223 #### *Intermediate states of genome reduction*

224 The complete sequence of the *Profftia* chromosome had a length of 1,225,795
225 bp and a G+C content of 31.9% (Table 1). It encoded for 645 proteins, one copy of
226 each rRNA, 35 tRNAs, and 10 ncRNAs. It had tRNAs and amino acid charging
227 potential for all 20 standard amino acids. However, protein-coding sequences made
228 up only 52.4% of the genome, and 21 pseudogenes indicated an ongoing gene
229 inactivation.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Currently, eight contigs with a total length of 1,183,315 bp represent the *Vallotia* genome. Seven contigs had an average G+C content and a coding density of 42.8% and 65.5%, respectively. However, a 72,431 bp long contig showed a characteristically lower G+C content (36.1%) and contained only 46.2% putative protein coding sequences (CDSs). This contig had identical repeats at its ends, and genome annotation revealed neighboring genes for a plasmid replication initiation protein, and ParA/ParB partitioning proteins, which function in plasmid and chromosome segregation between daughter cells before cell division (55). We thus assume that this contig corresponds to a circular plasmid of *Vallotia*. The 16S rRNA and the 5S plus 23S rRNA genes were encoded on two small contigs in the *Vallotia* assembly (1976 and 3571 bp, respectively) and were covered by nearly three times more sequence reads than the rest of the chromosomal contigs. This implies that *Vallotia* has three copies of each rRNA, similarly to its closest relative for which a complete genome sequence is available, *Mycetohabitans rhizoxinica* (56). In total, the *Vallotia* genome encoded for 780 proteins (29 on the putative plasmid), 37 tRNAs, and 52 predicted pseudogenes (5 on the putative plasmid).

The host-restricted lifestyle has a profound influence on bacterial genomes. Living in a stable, nutrient-rich niche relaxes purifying selection on many redundant functions, and small effective population size of the symbionts increases genetic drift. These can lead to the accumulation of slightly deleterious mutations, a proliferation of mobile genetic elements, and gene inactivation (57–60). Non-functional genomic regions and mobile genetic elements get subsequently lost, and ancient obligate endosymbionts typically have tiny (<<0.8 Mb), gene dense and stable genomes with AT-biased nucleotide composition (2,61,62). Facultative symbionts also possess

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

accelerated rates of sequence evolution but have larger genomes (>2Mb) with variable coding densities following the age of their host-restricted lifestyle (63). The only moderately reduced size and AT bias together with the low protein-coding density of the *Vallotia* and *Profftia* genomes was most similar to those of evolutionary young co-obligate partners of insects (63), for instance 'Ca. *Pseudomonas adelgestsugas'* in *A. tsugae* (25), *Serratia symbiotica* in *Cinara cedri* (64,65) and the *Sodalis*-like symbiont of *Philaenus spumarius*, the meadow spittlebug (66). However, compared to these systems involving a more ancient and a younger symbiont, similar genome characteristics of *Vallotia* and *Profftia* implied that both bacteria are evolutionarily recent symbionts in a phase of extensive gene inactivation typical for early stages of adaptation to an obligate host-restricted life-style (59,60).

Differential reduction of metabolic pathways

Although compared to their closest free-living relatives both *Vallotia* and *Profftia* have lost many genes in all functional categories, both retained a greater proportion of genes in translation-related functions and cell envelope biogenesis (Figures S2, S3). High retention of genes involved in central cellular functions such as translation, transcription, and replication is a typical feature of reduced genomes, even extremely tiny ones of long-term symbionts (62). However, ancient intracellular symbionts usually miss a substantial number of genes for the production of the cell envelope and might rely on host-derived membrane compounds (67–69).

Based on pathway reconstructions, both *Vallotia* and *Profftia* had a complete gene set for peptidoglycan, fatty acid, phospholipid biosynthesis, and retained most of the genes for the production of lipid A, LPS core, and the Lpt LPS transport machinery. Besides, we found a partial set of genes for O antigen biosynthesis in the

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

278 *Vallotia* genome. Regarding the membrane protein transport and assembly, both
279 adelgid endosymbionts had the necessary genes for Sec and signal recognition particle
280 (SRP) translocation, and the BAM outer membrane protein assembly complex.
281 *Profftia* also had a complete Lol lipoprotein trafficking machinery (lolABCDE),
282 which can deliver newly matured lipoproteins from the inner membrane to the outer
283 membrane (70). Besides, *Profftia* had a near-complete gene set for the Tol-Pal
284 system, however, *tolA* has been pseudogenized suggesting an ongoing reduction of
285 this complex. In addition, both adelgid endosymbionts have retained *mrdAB* and
286 *mreBCD* having a role in the maintenance of cell wall integrity and morphology
287 (71,72). The observed well-preserved cellular functions for cell envelope biogenesis
288 and integrity are consistent with the rod-shaped cell morphology of *Profftia* and
289 *Vallotia* (13), contrasting the spherical/pleomorphic cell shape of ancient
290 endosymbionts, such as *Annandia* in *A. tsugae* and *Pineus* species (9,10,14).

291 Regarding the central metabolism, *Vallotia* lacks 6-phosphofructokinase, but
292 has a complete gene set for gluconeogenesis and the tricarboxylic acid (TCA) cycle
293 cycle. TCA cycle genes are typically lost in long-term symbionts but are present in
294 facultative and evolutionary recent obligate endosymbionts (66,73,74). *Vallotia* does
295 not have any sugar transporter genes, similarly to its close relative, the fungus
296 symbiont *M. rhizoxinica* (56). A glycerol kinase gene next to a putative glycerol
297 uptake facilitator protein is present on its plasmid, however, it has a frameshift
298 mutation and a premature stop codon in the first 40% of the sequence and whether it
299 can still produce a functional protein remains unknown.

300 *Profftia* can convert acetyl-CoA to acetate for energy but lacks TCA cycle
301 genes, a feature characteristic to more reduced genomes, such, for instance, *Annandia*

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

302 in *A. tsugae* (25). *Profftia* has import systems for a variety of organic compounds,
303 such as murein tripeptides, phospholipids, thiamine, spermidine, and putrescine and 3-
304 phenylpropionate and two complete phosphotransferase systems for the uptake of
305 sugars. NADH dehydrogenase, ATP synthase, and cytochrome oxidases (*bo*/*bd*-1) are
306 encoded on both adelgid symbiont genomes.

307 *Profftia* retained more functions in inorganic ion transport and metabolism,
308 while *Vallotia* had a characteristically higher number of genes related to amino acid
309 biosynthesis (see its function below) and nucleotide transport and metabolism
310 (Figures S2, S3). For instance, *Profftia* can take up sulfate and use it for assimilatory
311 sulfate reduction and cysteine production, and it has also retained many genes for
312 heme biosynthesis. However it cannot produce inosine-5-phosphate (5'-IMP) and
313 uridine 5'-monophosphate (5'-UMP) precursors for the *de novo* synthesis of purine
314 and pyrimidine nucleotides thus would need to import these compounds.

315 Taken together their moderately reduced, gene-sparse genomes but still
316 versatile metabolic capabilities support that *Vallotia* and *Profftia* are both in an
317 intermediate stage of genome erosion and functional reduction similar to evolutionary
318 recently acquired endosymbionts.

319

320 **_Vallotia_ and _Profftia_ are both obligatory nutritional symbionts**

321 ***Complementary functions in essential amino acid provision***

322 *Vallotia* and *Profftia* complement each other's role in the essential amino acid
323 synthesis, thus have a co-obligatory status in the *A. laricis/tardus* symbiosis (Figure
324 1). Although *Vallotia* likely generates most essential amino acids, it can not provide

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

phenylalanine and tryptophan on its own. Solely, *Profftia* can produce chorismate, a key precursor for the synthesis of both amino acids. *Profftia* is likely responsible for the complete biosynthesis of phenylalanine as it has a full set of genes for this pathway. It can also convert chorismate to anthranilate, however further genes for tryptophan biosynthesis are only present in the *Vallotia* genome. Thus *Vallotia* likely takes up anthranilate for tryptophan biosynthesis. Anthranilate synthase (trpEG), is a subject to negative feedback regulation by tryptophan (75), thus partition of this rate-limiting step between the co-symbionts can enhance overproduction of the amino acid and might stabilize dual symbiotic partnerships at an early stage of coexistence. The production of tryptophan is partitioned between *Vallotia* and *Profftia* similarly as seen in other insect symbioses such as between *Buchnera* and *Serratia symbiotica* in co-obligatory partnerships in aphids (64,65) and between *Carsonella* and certain co-symbionts in psyllids (76). The tryptophan biosynthesis is also shared but is more redundant between the *Annandia* and *Pseudomonas* symbionts of *A. tsugae* (25). This association generally shows a higher level of functional overlap between the symbionts than the *Vallotia* - *Profftia* system, as redundant genes are present also in phenylalanine, threonine, lysine, and arginine synthesis in the *A. tsugae* symbiosis. Besides, the *Vallotia* - *Profftia* consortium is also more unbalanced than the *A. tsugae* system where *Annandia* can produce seven and the *Pseudomonas* partner five essential amino acids with the contribution of host genes (25).

The *Vallotia* genome encodes for all the enzymes for the synthesis of five essential amino acids (histidine, leucine, valine, lysine, threonine). Among the essential amino acid synthesis related genes, *argG* and *tyrB* of *Vallotia* are only present on the plasmid, which might have contributed to its maintenance in the

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

349 genome. However, neither of the endosymbionts can produce ornithine, 2-
350 oxobutanoate and homocysteine *de novo*, which are key for the biosynthesis of
351 arginine, isoleucine and methionine, respectively. *metC* and *argA* are still present as
352 pseudogenes in *Vallotia* suggesting a recent loss of these functions in methionine and
353 arginine biosynthesis, respectively. The corresponding functions are also missing
354 from the *Annandia* - *Pseudomonas* system (Weglarz et al., 2018). Ornithine, 2-
355 oxobutanoate, and homocysteine are thus likely supplied by the insect host, as seen for
356 instance in aphids, mealybugs, and psyllids, where the respective genes encoding for
357 cystathionine gamma and beta lyases and insect ornithine aminotransferase are
358 present in the insect genomes and are typically overexpressed within the bacteriome
359 (68,77,78). However, we can not confirm the presence of relevant genes in the *A.*
360 *laricis/tardus* genome, as our metagenome data were almost free from eukaryotic
361 sequences.

362 *Vallotia* and *Profftia* have more redundant functions in non-essential amino
363 acid production. Both symbionts can synthesize seven non-essential amino acids
364 mostly through a series of amino-acid conversions (Figure 1). Only *Profftia* can
365 produce cysteine and tyrosine, while none of the symbionts can build up glutamine
366 thus this latter amino acid is likely supplied by the insect bacteriocytes.

367 The presence of amino acid transporters can complement missing functions in
368 amino acid synthesis in the endosymbionts (Figure 1). For instance *Profftia* has a
369 high-affinity glutamine ABC transporter, and three symporters (BrnQ, Mtr, TdcC),
370 which can import isoleucine, leucine, valine, tryptophan, and threonine among the
371 essential amino acids that can be produced by *Vallotia*. *Vallotia* might excrete
372 isoleucine, valine, and leucine via AzICD, a putative branched-chain amino acid

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

efflux pump (79), and these amino acids could be taken up by *Profftia* via BrnQ and would be readily available also for the insect host.

B vitamin provision by Vallotia

Regarding the B vitamin synthesis, *Vallotia* should be able to produce thiamine (B_1), riboflavin (B_2), pantothenate (B_5), pyridoxine (B_6), biotin (B_7), and folic acid (B_9) (Figure S4). Although *Vallotia* misses some genes of the canonical pathways, alternative enzymes and host-derived compounds might bypass these reactions, as detailed in the supplementary material. *Profftia* has only a few genes related to B vitamin biosynthesis. Three pseudogenes (*ribAEC*) in the riboflavin synthesis pathway indicate that these functions might have been lost evolutionary recently in this symbiont.

In summary, *Profftia* and *Vallotia* are both obligate nutritional endosymbionts of adelgids, however, *Vallotia* has a pivotal role in essential amino acid and B vitamin provision.

Profftia **and** ***Vallotia*** **are related to free-living bacteria and fungus endosymbionts**

Previous 16S rRNA-based phylogenetic analyses suggested an affiliation of *Profftia* with free-living gammaproteobacteria and a close phylogenetic relationship between *Vallotia* and betaproteobacterial endosymbionts of *Rhizopus* fungi (13). Biased nucleotide composition and accelerated sequence evolution of endosymbiont genomes (2,3) often result in inconsistent phylogenies and may cause grouping of unrelated taxa (68,80,81). Thus to further investigate the phylogenetic relationships of

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

396 the *A. laricis/tardus* symbionts, we used conserved marker genes for maximum
397 likelihood and Bayesian phylogenetic analyses.

398 Phylogenetic analysis of 45 single-copy proteins demonstrated that *Profftia*
399 opens up a novel insect symbiont lineage most similar to *Hafnia* species and an isolate
400 from the human gastrointestinal tract within the Hafniaceae, which has been recently
401 designated as a distinct family within the Enterobacteriales (82) (Figure S5). *Hafnia*
402 strains are frequently found in the gastrointestinal tract of humans and animals
403 including insects, among others (83,84). The phylogenomic placement of *Profftia* in
404 our analysis is in agreement with previous 16S rRNA based analyses (13).

405 *Vallotia* formed a monophyletic group with *Mycetohabitans endofungorum*
406 and *M. rhizoxinica,* endosymbionts of *Rhizopus* fungi within the Burkholderiaceae
407 (85,86) with strong support in phylogenetic analyses based on a concatenated set of
408 108 proteins (Figures 2, S6; previous taxonomic assignments of the fungus
409 endosymbionts were as *Burkholderia/Paraburkholderia endofungorum* and
410 *rhizoxinica*, respectively). Interestingly, *Vallotia* and *M. endofungorum* appeared as
411 well-supported sister taxa within this clade. This implies a closer phylogenetic
412 relationship between *Vallotia* and *M. endofungorum*, and a common origin of adelgid
413 endosymbionts from within a clade of fungus endosymbionts. Lengths of branches
414 leading to the fungus endosymbionts were similar to those of free-living bacteria in
415 the data set, however *Vallotia* had a remarkably longer branch marking a rapid rate of
416 sequence evolution characteristic of obligate intracellular bacteria (2,3). *M.*
417 *endofungorum* and *M. rhizoxinica* have been identified in the cytosol of the
418 zygomycete *Rhizopus microsporus,* best known as the causative agent of rice seedling
419 blight (86,87). The necrotrophic fungus secrets potent toxins, rhizoxin and rhizonin,

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

which are produced by the endosymbionts (86,88). The bacterial partners are obligatory for their host as they tightly control its sporulation, while they benefit from host nutrients and spread with the fungal spores (89,90). Additionally, related bacterial strains have also been found in association with *Rhizopus* fungi worldwide in a diverse set of environments, including other plant species, soil, food, and even human tissues (91–93).

Taken together, phylogenomic analyses support that *Profftia* and *Vallotia* open up novel insect symbionts lineages most closely related to free-living bacteria within the Hafniaceae and a clade of fungus endosymbionts within the Burkholderiaceae, respectively. Given the well supported phylogenetic positioning of *'Candidatus* Vallotia tarda' nested within a clade formed by *Mycetohabitans* species, we propose the transfer of *'Candidatus* Vallotia tarda' to the *Mycetohabitans* genus, as *'Candidatus* Mycetohabitans vallotii' (a detailed proposal for the re-classification is given in the supplementary material).

The evolutionary link between *Vallotia* and fungus endosymbionts

High level of genomic synteny between Vallotia and M. rhizoxinica

The close phylogenetic relationship between *Vallotia* and *Rhizopus* symbionts offers a unique opportunity to gain insight into the early stages of genome reduction and to infer functional consequences of the partnerships of bacterial symbionts with insects and fungi, respectively. Among the *Rhizopus* endosymbionts, a closed genome is available for *M. rhizoxinica* (56). We therefore mostly focused on this fungus symbiont as a reference for comparison with *Vallotia*.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

443 A surprisingly high level of synteny between the genomes of *Vallotia* and *M.*
444 *rhizoxinica* provides further evidence for their shared ancestry. Seven contigs
445 representing the *Vallotia* genome showed a high level of collinearity with the
446 chromosome of *M. rhizoxinica* (Figure 3A). However, their cumulative size was only
447 ~40% of the fungus endosymbiont chromosome. The contig that corresponded to a
448 putative plasmid of *Vallotia* was perfectly syntenic with the larger of the two plasmids
449 of *M. rhizoxinica* (pBRH01), although the *Vallotia* plasmid was over 90% smaller in
450 size (72 431 bp vs. 822 304 bp) (56). Thus, the *Vallotia* plasmid showed a much
451 higher level of reduction than the chromosome, which together with its lower G+C
452 content and gene density suggest differential evolutionary constraints on these
453 replicons. The observed high level of genome synteny between *Vallotia* and *M.*
454 *rhizoxinica* genomes is consistent with the phylogenetic position of *Vallotia*
455 interleaved within the clade of *Rhizopus* endosymbionts and points towards a direct
456 evolutionary link between these symbioses and a symbiont transition between the
457 fungus and insect hosts.

458 The conservation of genome structure contrasts with the elevated number of
459 transposases and inactive derivatives making up ~6% of the fungus symbiont genome
460 (56). Transition to a host-restricted lifestyle is usually followed by a sharp
461 proliferation of mobile genetic elements coupled with many genomic rearrangement
462 and gene inactivation. As seen for instance in endosymbionts of grain weevils (94),
463 facultative and co-obligate *Serratia symbiotica* strains (95), and facultative
464 endosymbionts, such as *Hamiltonella defensa* and *Regiella insecticola* in aphids (74).
465 However, mobile genetic elements get subsequently purged out of the genomes of
466 strictly vertically transmitted symbionts via a mutational bias towards deletion and

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

because of lack of opportunity for horizontal acquisition of novel genetic elements (59,61). In contrast to the fungus symbiont, mobile elements are notably absent from the *Vallotia* genome, suggesting that they might have been lost early after the establishment of the adelgid symbiosis conserving high collinearity between the fungus and adelgid symbiont genomes.

Shrinkage of the insect symbiont genome

Deletion of large genomic fragments – spanning many functionally unrelated genes – represents an important driving force of genome erosion especially at early stages of symbioses when selection on many functions is weak (3,96). Besides, gene loss also occurs individually and is ongoing, albeit at a much lower rate, even in ancient symbionts with tiny genomes (62,97,98). Both small and large deletions could be seen when comparing the *Vallotia* and *M. rhizoxinica* genomes. Several small deletions as small as one gene were observed sparsely in the entire length of the *Vallotia* genome within otherwise syntenic regions. The largest genomic region missing from *Vallotia* encompassed 165 kbp on the *M. rhizoxinica* chromosome (Figure 3B). The corresponding intergenic spacer was only 3,843 bp long on the *Vallotia* genome between a phage shock protein and the Mfd transcription-repair-coupling factor, present both in *Vallotia* and *M. rhizoxinica*. Interestingly, this large genomic fragment included the large rhizoxin biosynthesis gene cluster (*rhiIGBCDHEF*), which is responsible for the production of rhizoxin, a potent antimitotic macrolide serving as a virulence factor for *R. microsporus*, the host of *M. rhizoxinica* (88). A homologous gene cluster is also present in *M. endofungorum* and *Pseudomonas fluorescens* and it has been suggested that the *rhi* cluster might have been horizontally acquired by *M. rhizoxinica* (56,88). Rhizoxin blocks microtubule

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

formation in various types of eukaryotic cells (88,99), thus lack of this gene cluster in ancestral *Vallotia* was likely a prerequisite for the establishment of the adelgid symbiosis. However, this large deleted genomic region also contained several transposases and many other genes, such as *argE* and *ilvA*, coding for the final enzymes for ornithine and 2-oxobutanoate productions, which were located adjacent to each other at the beginning of this fragment. The largest deletion between the plasmids encompassed nearly 137 kbp of the megaplasmid of *M. rhizoxinica* and involved several non-ribosomal peptide synthetases (NRPS), insecticidal toxin complex (Tc) proteins, and a high number of transposases among others. *M. rhizoxinica* harbors 15 NRPS genes clusters (56) in total, all of which are absent in *Vallotia*. NPRPs are large multienzyme machineries that assemble various peptides, which might function as antibiotics, signal molecules, or virulence factors (100). Insecticidal toxin complexes are bacterial protein toxins, which exhibit powerful insecticidal activity (101). Two of such proteins are also present in the large deleted chromosomal region in close proximity to the rhizoxin biosynthesis gene cluster (Figure 3B), however, their role in *M. rhizoxinica* remains elusive.

The Vallotia genome encodes for a subset of functions of the fungus endosymbionts

The number of protein coding genes of *Vallotia* is less than one-third of those of *M. rhizoxinica* and *M. endofungorum*, although metabolic functions are already reduced in the fungus endosymbionts compared to free-living *Burkholderia* (56). When compared to the two genomes of the fungus endosymbionts, only 53 proteins were specific to *Vallotia* (Figure S7). All of these were hypothetical proteins, and most of them showed no significant similarity to other proteins in public databases.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

However, several fall within regions syntenic to the *M. rhizoxinica* genome and even retained partial sequence similarity to intact genes present solely in the fungus endosymbiont. Thus we assume that at least some of these *Vallotia* specific hypothetical proteins might rather be remnants of degrading genes and over-annotated/non-functional open reading frames than orphan genes with a yet unknown function (102,103). Four genes were present in *Vallotia* and *M. rhizoxinica* but were missing in *M. endofungorum*. These encoded for BioA, BioD in biotin biosynthesis, NagZ in cell wall recycling, and an MFS transporter. Fifteen genes, including for instance the MreB rod-shape determining protein, glycosyltransferase and hit family proteins, genes in lipopolysaccharide, lipoate synthesis, and the oxidative pentose phosphate pathway were shared between *Vallotia* and *M. endofungorum* only. The rest of the *Vallotia* genes, coding for 91% of all of its proteins, were shared among the fungus endosymbionts and the insect endosymbiont.

Comparing the genes present in both the insect and the fungus endosymbionts to those shared only by the fungus endosymbionts (Figure S8), we can infer selective functions maintained or lost during transition to insect endosymbiosis. Translation related functions have been retained in the greatest measure in the group shared by all endosymbionts. Functions, where higher proportion of genes were specific to the fungus endosymbioses, were related to transcription, inorganic ion transport and metabolism, secondary metabolite biosynthesis, signal transduction, intracellular trafficking, secretion, vesicular transport and defense mechanisms. (Most of the proteins specific to either of the fungus endosymbionts were homologous to transposases and integrases, transcriptional regulators, or had an unknown function.)

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

538 Fungus endosymbionts encode for a high number of transcriptional regulators
539 (~5% of all genes in *M. rhizoxinica*) (56), but *Vallotia* has retained only a handful of
540 such genes, which is a feature similar to other insect symbionts and might contribute
541 to the overproduction of essential amino acids (62,104).

542 *M. rhizoxinica* is resistant against various β-lactams and has an arsenal of
543 efflux pumps which might provide defense against antibacterial fungal molecules: the
544 latter might also excrete virulence factors to the fungus cytosol (type I secretion) (56).
545 Besides, *M. rhizoxinica* has several genes for pilus formation, adhesion proteins, and
546 type II, type III, and type IV secretion systems which likely have a central role in host
547 infection and manipulation in the bacteria-fungus symbiosis (56,105,106). However,
548 all of the corresponding genes are missing in *Vallotia* thus neither of these
549 mechanisms likely plays a role in the operation of the adelgid symbiosis. We could
550 not even detect remnants of these genes in the *Vallotia* genome, except for a type II
551 secretion system protein as a pseudogene. Loss of these functions is consistent with a
552 strict vertical transmission of *Vallotia* between host generations, in contrast to *M.*
553 *rhizoxinca,* which can spread also horizontally among fungi and can re-infect cured
554 *Rhizopus* strains under laboratory conditions (86,87).

555 Additionally, the *M. rhizoxinica* genome encodes several predicted toxin-
556 antitoxin systems (56). Plasmid-associated toxin-antitoxin systems can act as
557 addiction molecules, which promote the maintenance of plasmids within bacterial
558 populations (107). However, most of these are missing from *Vallotia*, only PasT, the
559 toxic component of a chromosomal type II toxin-antitoxin system is present. Low
560 levels of PasT can enhance bacterial stress resistance and growth of free-living

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

561 bacteria, while high concentrations can induce persister cell formation (108).
562 However, the function of PasT in *Vallotia* remains unclear.
563

564 ## Conclusions

565 In most plant-sap feeding insects harboring a dual symbiotic system, typically
566 the more ancient symbiont provides most of the essential amino acids. However, due
567 to the ongoing genomic degradation characteristic for endosymbionts even genes
568 essential in the symbiosis can get inactivated (3,58). These events might lead to the
569 acquisition and fixation of an additional, younger symbiont, which can complement
570 these lost functions. For instance, among Auchenorrhyncha, the universal ancient
571 symbiont *Sulcia* provides seven or eight essential amino acids, while the rest is
572 supplied by different younger co-symbionts (109). As a consequence of a host-
573 restricted lifestyle, the genome of the newly arriving symbiont will also lose many
574 functions even among those key in the symbiosis but present in the other resident
575 symbiont (64,67,109). Co-obligate symbionts of *A. laricis/tardus* are both
576 evolutionary recent bacterial endosymbionts of adelgids with moderately reduced
577 genomes. This is following their occurrence in larch (*Profftia* and *Vallotia*) and
578 Douglas fir (*Vallotia*) associated lineages of adelgids, which likely diversified
579 relatively recently, ~47 and ~60 million years ago from the remaining clades of
580 adelgids, respectively (5). However, these recently gained adelgid symbionts show a
581 high level of metabolic complementarity and low functional redundancy in essential
582 amino acid synthesis. Given its presence in both larch and Douglas fir associated
583 adelgids, *Vallotia* might be the relatively older symbiont, and it can synthesize nine

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

584 essential amino acids with a putative contribution of insect delivered compounds.
585 Loss of functions in chorismate and anthranilate biosynthesis might have led to the
586 fixation of *Profftia* in the system. *Profftia* can produce phenyalanine, but has lost its
587 capabilities for synthetizing other essential amino acids. Host-derived compounds and
588 partition of tryptophan biosynthesis between the co-symbionts in *A. laricis/tardus* are
589 similar to other insect symbioses suggesting convergent evolution. However, the
590 *Vallotia* - *Profftia* system differs from the *Annandia* - *Pseudomonas* system in *A.*
591 *tsugae* where functions of the symbionts in essential amino acid synthesis are more
592 balanced and redundant. It has been suggested that repeated replacement of symbionts
593 among adelgids might be a consequence of periods with relaxed selection on
594 symbiont functions due to different feeding behavior of adelgids on primary and
595 secondary host trees – that is feeding on nutrient-rich parenchyma cells on spruce
596 versus nutrient-poor phloem sap on alternate hosts – and multiple origins of host-
597 alternating lifestyles (10). *Annandia*, the ancient symbiont of adelgids has lost many
598 functions in essential amino acid biosynthesis, which could support this hypothesis
599 (25), however the *Vallotia* - *Profftia* system does not follow this pattern.

600 One of the most remarkable findings of our study is the evolutionary link
601 between the betaproteobacterial insect symbiont, *Vallotia*, and endosymbionts of
602 *Rhizopus* fungi supported by their close phylogenetic relationships and a high-level of
603 genomic synteny. There are many possible scenarios that could explain the origin of
604 these symbioses. A common free-living ancestor could infect ancestral adelgids and
605 *Rhizopus* fungi independently or developed an intracellular lifestyle in either of these
606 hosts and got subsequently transmitted between them. We assume that a fungus-insect
607 symbiont transition is more likely than multiple origins of these associations as the

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

608 proliferation of mobile genetic elements typical in early stages of host restriction
609 would have resulted in extensive rearrangements and a substantially different genomic
610 structure (94,110), as seen, for instance, between very closely related *Serratia*
611 *symbiotica* strains in aphids (95). Alternative scenarios are also possible, but the
612 phylogenetic position of *Vallotia* interleaved within the clade of *Rhizopus*
613 endosymbionts and lack of functions specific to the adelgid symbiont point towards
614 the putative origin of *Vallotia* from the fungus endosymbionts. The origin of insect
615 symbionts from fungus endosymbionts is, according to our knowledge,
616 unprecedented. *Rhizopus* endosymbionts are equipped with many functions for
617 infection and overcoming host defense. Chitinase, chitosanase, and a putative chitin-
618 binding protein have also been found among the putatively *Sec* exported proteins of
619 *M. rhizoxinica* (56), which besides the infection of fungi could have had a role in the
620 transmission into an insect host. In addition, *Rhizopus* endosymbionts could be
621 maintained in pure cultures (86), thus, at least for a limited time, they might survive
622 also outside of their hosts in the environment. Their host, *Rhizopus microsporus*, is a
623 plant pathogen fungus with a broad environmental distribution. Thus a potential route
624 for acquisition of the symbionts by insects could have been via plant tissues, the food
625 source of adelgids, similar to plant-mediated symbiont transmission observed for
626 intracellular insect symbionts (21).

627 Taken together, our genomic analysis of co-obligate endosymbionts of
628 adelgids revealed a novel path for the evolution of bacteria-insect symbioses from a
629 clade of fungus-associated ancestors.

630

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

631 **Acknowledgements**

632 We would like to acknowledge Alexander Siegl and Thomas Penz for their
633 help in collecting adelgids and Irene Lichtscheidl-Schultz (Core Facility Cell Imaging
634 and Ultrastructure Research, Cell Imaging Lab, University of Vienna) for the close-up
635 photos of adelgids. This study was funded by the Austrian Science Fund (FWF)
636 project P22533-B17. Work in the lab of M. H. is supported by Austrian Science Fund
637 project DOC 69-B. A.M.M. was supported by a Marie Skłodowska-Curie Individual
638 Fellowship (840270, LEECHSYMBIO) of the European Union.

639

640 **Competing Interests:** The authors declare no competing interests

641

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

642 **References**

643 1. Douglas AE. Phloem-sap feeding by animals: problems and solutions. *J Exp*
644 *Bot* 2006; 57: 747–754.

645 2. Moran NA, McCutcheon JP, Nakabachi A. Genomics and Evolution of
646 Heritable Bacterial Symbionts. *Annu Rev Genet* 2008; 42: 165–190.

647 3. Wernegreen JJ. Endosymbiont evolution: Predictions from theory and surprises
648 from genomes. *Ann N Y Acad Sci* 2015; 1360: 16-35.

649 4. Havill NP, Foottit RG. Biology and evolution of adelgidae. *Annu Rev Entomol.*
650 2007; 52: 325–349.

651 5. Havill NP, Foottit RG, von Dohlen CD. Evolution of host specialization in the
652 Adelgidae (Insecta: Hemiptera) inferred from molecular phylogenetics. *Mol*
653 *Phylogenet Evol* 2007; 44: 357–370.

654 6. Favret C, Havill NP, Miller GL, Sano M, Victor B. Catalog of the adelgids of
655 the world (Hemiptera, Adelgidae). *ZooKeys* 2015; 534: 35–54.

656 7. Hrinkevich KH, Progar RA, Shaw DC. Climate Risk Modelling of Balsam
657 Woolly Adelgid Damage Severity in Subalpine Fir Stands of Western North
658 America. *PLOS ONE* 2016; 11: e0165094.

659 8. McClure MS, Cheah CAS-J. Reshaping the Ecology of Invading Populations of
660 Hemlock Woolly Adelgid, Adelges tsugae (Homoptera: Adelgidae), in Eastern
661 North America. *Biol Invasions* 1999; 1: 247–254.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

662 9. von Dohlen CD, Spaulding U, Shields K, Havill NP, Rosa C, Hoover K.
663 Diversity of proteobacterial endosymbionts in hemlock woolly adelgid
664 (Adelges tsugae) (Hemiptera: Adelgidae) from its native and introduced range.
665 *Environ Microbiol* 2013; 15: 2043-2062.

666 10. von Dohlen CD, Spaulding U, Patch KB, Weglarz KM, Foottit RG, Havill NP,
667 et al. Dynamic Acquisition and Loss of Dual-Obligate Symbionts in the Plant-
668 Sap-Feeding Adelgidae (Hemiptera: Sternorrhyncha: Aphidoidea). *Front*
669 *Microbiol* 2017; 8: 1037-1052.

670 11. Michalik A, Gołas A, Kot M, Wieczorek K, Szklarzewicz T. Endosymbiotic
671 microorganisms in Adelges (Sacchiphantes) viridis (Insecta, Hemiptera,
672 Adelgoidea: Adelgidae): Molecular characterization, ultrastructure and
673 transovarial transmission. *Arthropod Struct Dev* 2013; 42 : 531–538.

674 12. Toenshoff ER, Penz T, Narzt T, Collingro A, Schmitz-Esser S, Pfeiffer S, et al.
675 Bacteriocyte-associated gammaproteobacterial symbionts of the Adelges
676 nordmannianae/piceae complex (Hemiptera: Adelgidae). *ISME J* 2012; 6: 384–
677 396.

678 13. Toenshoff ER, Gruber D, Horn M. Co-evolution and symbiont replacement
679 shaped the symbiosis between adelgids (Hemiptera: Adelgidae) and their
680 bacterial symbionts. *Environ Microbiol* 2012; 14: 1284–1295.

681 14. Toenshoff ER, Szabó G, Gruber D, Horn M. The Pine Bark Adelgid, Pineus
682 strobi, Contains Two Novel Bacteriocyte-Associated Gammaproteobacterial
683 Symbionts. *Appl Env Microbiol* 2014; 80: 878–885.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

15. Gündüz EA, Douglas AE. Symbiotic bacteria enable insect to use a nutritionally inadequate diet. *Proc R Soc B Biol Sci* 2009; 276: 987–991.

16. Duron O, Hurst GD. Arthropods and inherited bacteria: from counting the symbionts to understanding how symbionts count. *BMC Biol* 2013; 11: 45–49.

17. Feldhaar H. Bacterial symbionts as mediators of ecologically important traits of insect hosts. *Ecol Entomol* 2011; 36: 533–543.

18. Oliver KM, Degnan PH, Burke GR, Moran NA. Facultative Symbionts in Aphids and the Horizontal Transfer of Ecologically Important Traits. *Annu Rev Entomol* 2010; 55: 247–266.

19. Moran NA, Dunbar HE. Sexual acquisition of beneficial symbionts in aphids. *Proc Natl Acad Sci* 2006; 103: 12803–12806.

20. Gehrer L, Vorburger C. Parasitoids as vectors of facultative bacterial endosymbionts in aphids. *Biol Lett* 2012; 8: 613–615.

21. Chrostek E, Pelz-Stelinski K, Hurst GDD, Hughes GL. Horizontal Transmission of Intracellular Insect Symbionts via Plants. *Front Microbiol* 2017; 8: 2237–2245.

22. Manzano-Marín A, Coeur D'acier A, Clamens A-L, Orvain C, Cruaud C, Barbe V, et al. Serial horizontal transfer of vitamin-biosynthetic genes enables the establishment of new nutritional symbionts in aphids' di-symbiotic systems. *ISME J* 2020; 14: 259–273.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

704 23. Pons I, Renoz F, Noël C, Hance T. Circulation of the Cultivable Symbiont
705 Serratia symbiotica in Aphids Is Mediated by Plants. *Front Microbiol* 2019; 10:
706 764–777.

707 24. Clayton AL, Oakeson KF, Gutin M, Pontes A, Dunn DM, Niederhausern AC
708 von, et al. A Novel Human-Infection-Derived Bacterium Provides Insights into
709 the Evolutionary Origins of Mutualistic Insect–Bacterial Symbioses. *PLOS*
710 *Genet* 2012; 8(11):e1002990.

711 25. Weglarz KM, Havill NP, Burke GR, von Dohlen CD. Partnering With a Pest:
712 Genomes of Hemlock Woolly Adelgid Symbionts Reveal Atypical Nutritional
713 Provisioning Patterns in Dual-Obligate Bacteria. *Genome Biol Evol* 2018; 10:
714 1607–1621.

715 26. Zurovcova M, Havelka J, Stary P, Vechtova P, Chundelova D, Jarosova A, et
716 al. “DNA barcoding” is of limited value for identifying adelgids (Hemiptera:
717 Adelgidae) but supports traditional morphological taxonomy. *Eur J Entomol*
718 2010; 107: 147–156.

719 27. Zhou J, Bruns MA, Tiedje JM. DNA recovery from soils of diverse
720 composition. *Appl Environ Microbiol* 1996; 62: 316–322.

721 28. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic
722 datasets. *Bioinformatics* 2011; 27: 863–864.

723 29. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al.
724 SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-
725 Cell Sequencing. *J Comput Biol* 2012; 19: 455–477.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

726 30. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler
727 transform. *Bioinformatics* 2009; 25: 1754–1760.

728 31. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al.
729 BLAST+: architecture and applications. *BMC Bioinformatics* 2009; 10: 421–
730 430.

731 32. Huson DH, Mitra S, Ruscheweyh H-J, Weber N, Schuster SC. Integrative
732 analysis of environmental sequences using MEGAN4. *Genome Res* 2011; 21:
733 1552–1560.

734 33. Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW.
735 RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic*
736 *Acids Res* 2007; 35: 3100–3108.

737 34. Weinmaier T, Platzer A, Frank J, Hellinger H-J, Tischler P, Rattei T. ConsPred:
738 a rule-based (re-)annotation framework for prokaryotic genomes. *Bioinforma*
739 *Oxf Engl* 2016; 32: 3327–3329.

740 35. Okonechnikov K, Golosova O, Fursov M, UGENE team. Unipro UGENE: a
741 unified bioinformatics toolkit. *Bioinforma Oxf Engl* 2012; 28: 1166–1167.

742 36. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the
743 reference centre for bacterial insertion sequences. *Nucleic Acids Res* 2006;
744 34(Database issue):D32-36.

745 37. Caspi R, Altman T, Dale JM, Dreher K, Fulcher CA, Gilham F, et al. The
746 MetaCyc database of metabolic pathways and enzymes and the BioCyc

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

747 collection of pathway/genome databases. *Nucleic Acids Res* 2010; 38(Database
748 issue):D473-479.

749 38. Karp PD, Paley SM, Krummenacker M, Latendresse M, Dale JM, Lee TJ, et al.
750 Pathway Tools version 13.0: integrated software for pathway/genome
751 informatics and systems biology. *Brief Bioinform* 2010; 11: 40–79.

752 39. Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for
753 eukaryotic genomes. *Genome Res* 2003; 13: 2178–2189.

754 40. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering
755 C, et al. Fast Genome-Wide Functional Annotation through Orthology
756 Assignment by eggNOG-Mapper. *Mol Biol Evol* 2017; 34: 2115–2122.

757 41. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK,
758 Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically
759 annotated orthology resource based on 5090 organisms and 2502 viruses.
760 *Nucleic Acids Res* 2019; 47(D1):D309–14.

761 42. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using
762 DIAMOND. *Nat Methods* 2015; 12: 59–60.

763 43. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment
764 with gene gain, loss and rearrangement. *PloS One*. 2010; 5(6):e11147.

765 44. Krzywinski MI, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D, et al.
766 Circos: An information aesthetic for comparative genomics. *Genome Res* 2009;
767 19:1639–1645.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

768 45. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer.
769 *Bioinformatics* 2011; 27: 1009–1010.

770 46. Wang Z, Wu M. A Phylum-Level Bacterial Phylogenetic Marker Database.
771 *Mol Biol Evol* 2013; 30: 1258–1262.

772 47. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and
773 high throughput. *Nucleic Acids Res* 2004; 32: 1792–1797.

774 48. Castresana J. Selection of Conserved Blocks from Multiple Alignments for
775 Their Use in Phylogenetic Analysis. *Mol Biol Evol* 2000; 17: 540–552.

776 49. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS.
777 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat*
778 *Methods* 2017; 14: 587–589.

779 50. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and
780 Effective Stochastic Algorithm for Estimating Maximum-Likelihood
781 Phylogenies. *Mol Biol Evol* 2015; 32: 268–274.

782 51. Trifinopoulos J, Nguyen L-T, von Haeseler A, Minh BQ. W-IQ-TREE: a fast
783 online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res*
784 2016; 44(W1):W232–235.

785 52. Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast Approximation for
786 Phylogenetic Bootstrap. *Mol Biol Evol* 2013; 30: 1188–1195.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

787 53. Ronquist F, Teslenko M, Mark P van der, Ayres DL, Darling A, Höhna S, et al.
788 MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice
789 Across a Large Model Space. *Syst Biol*. 2012; 61: 539–542.

790 54. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for
791 inference of large phylogenetic trees. In: 2010 Gateway Computing
792 Environments Workshop (GCE). 2010. p. 1–8.

793 55. Dubarry N, Pasta F, Lane D. ParABS Systems of the Four Replicons of
794 Burkholderia cenocepacia: New Chromosome Centromeres Confer Partition
795 Specificity. *J Bacteriol* 2006; 188: 1489–1496.

796 56. Lackner G, Moebius N, Partida-Martinez LP, Boland S, Hertweck C. Evolution
797 of an endofungal Lifestyle: Deductions from the Burkholderia rhizoxinica
798 Genome. *BMC Genomics* 2011;12: 210–223.

799 57. Bright M, Bulgheresi S. A complex journey: transmission of microbial
800 symbionts. *Nat Rev Microbiol* 2010 Mar; 8: 218–230.

801 58. Moran NA, Bennett GM. The tiniest tiny genomes. *Annu Rev Microbiol* 2014;
802 68: 195–215.

803 59. Moran NA, Plague GR. Genomic changes following host restriction in bacteria.
804 *Curr Opin Genet Dev* 2004;14: 627–633.

805 60. Wernegreen JJ. For better or worse: genomic consequences of intracellular
806 mutualism and parasitism. *Curr Opin Genet Dev* 2005; 15: 572–583.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

807 61. Kuo C-H, Ochman H. Deletional Bias across the Three Domains of Life.
808 *Genome Biol Evol*. 2009; 1: 145–152.

809 62. McCutcheon JP, Moran NA. Extreme genome reduction in symbiotic bacteria.
810 *Nat Rev Microbiol* 2012; 10: 13–26.

811 63. Lo W-S, Huang Y-Y, Kuo C-H. Winding paths to simplicity: genome evolution
812 in facultative insect symbionts. *FEMS Microbiol Rev* 2016; 40: 855–874.

813 64. Lamelas A, Gosalbes MJ, Manzano-Marín A, Peretó J, Moya A, Latorre A.
814 Serratia symbiotica from the Aphid Cinara cedri: A Missing Link from
815 Facultative to Obligate Insect Endosymbiont. *PLoS Genet*
816 2011;7(11):e1002357.

817 65. Manzano-Marín A, Latorre A. Snapshots of a shrinking partner: Genome
818 reduction in Serratia symbiotica. *Sci Rep* 2016; 6: 1–11.

819 66. Koga R, Moran NA. Swapping symbionts in spittlebugs: evolutionary
820 replacement of a reduced genome symbiont. *ISME J* 2014; 8: 1237–1246.

821 67. Bennett GM, McCutcheon JP, MacDonald BR, Romanovicz D, Moran NA.
822 Differential Genome Evolution Between Companion Symbionts in an Insect-
823 Bacterial Symbiosis. *mBio*. 2014; 5(5):e01697-14.

824 68. Husnik F, Nikoh N, Koga R, Ross L, Duncan RP, Fujie M, et al. Horizontal
825 Gene Transfer from Diverse Bacteria to an Insect Genome Enables a Tripartite
826 Nested Mealybug Symbiosis. *Cell* 2013; 153: 1567–1578.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

827 69. Shigenobu S, Wilson ACC. Genomic revelations of a mutualism: the pea aphid
828 and its obligate bacterial symbiont. *Cell Mol Life Sci* 2011; 68: 1297–309.

829 70. Kaplan E, Greene NP, Crow A, Koronakis V. Insights into bacterial lipoprotein
830 trafficking from a structure of LolA bound to the LolC periplasmic domain.
831 *Proc Natl Acad Sci U S A* 2018; 115: E7389–97.

832 71. Kruse T, Bork-Jensen J, Gerdes K. The morphogenetic MreBCD proteins of
833 Escherichia coli form an essential membrane-bound complex. *Mol Microbiol*
834 2005; 55: 78–89.

835 72. Tamaki S, Matsuzawa H, Matsuhashi M. Cluster of mrdA and mrdB genes
836 responsible for the rod shape and mecillinam sensitivity of Escherichia coli. *J*
837 *Bacteriol* 1980; 141: 52–57.

838 73. Burke GR, Moran NA. Massive Genomic Decay in Serratia symbiotica, a
839 Recently Evolved Symbiont of Aphids. *Genome Biol Evol* 2011; 3: 195–208.

840 74. Degnan PH, Leonardo TE, Cass BN, Hurwitz B, Stern D, Gibbs RA, et al.
841 Dynamics of genome evolution in facultative symbionts of aphids. *Environ*
842 *Microbiol* 2010; 12: 2060–2069.

843 75. Pabst MJ, Kuhn JC, Somerville RL. Feedback Regulation in the Anthranilate
844 Aggregate from Wild Type and Mutant Strains of Escherichia coli. *J Biol Chem*
845 1973; 248: 901–914.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

846 76. Sloan DB, Moran NA. Genome Reduction and Co-evolution between the
847 Primary and Secondary Bacterial Symbionts of Psyllids. *Mol Biol Evol* 2012;
848 29: 3781–3792.

849 77. Hansen AK, Moran NA. Aphid genome expression reveals host–symbiont
850 cooperation in the production of amino acids. *Proc Natl Acad Sci* 2011; 108:
851 2849–2854.

852 78. Sloan DB, Nakabachi A, Richards S, Qu J, Murali SC, Gibbs RA, et al. Parallel
853 histories of horizontal gene transfer facilitated extreme reduction of
854 endosymbiont genomes in sap-feeding insects. *Mol Biol Evol* 2014; 31: 857–
855 871.

856 79. Belitsky BR, Gustafsson MC, Sonenshein AL, Von Wachenfeldt C. An lrp-like
857 gene of Bacillus subtilis involved in branched-chain amino acid transport. *J*
858 *Bacteriol* 1997; 179:5448–5457.

859 80. Williams KP, Gillespie JJ, Sobral BWS, Nordberg EK, Snyder EE, Shallom
860 JM, et al. Phylogeny of Gammaproteobacteria. *J Bacteriol* 2010; 192: 2305–
861 2314.

862 81. Wu M, Eisen JA. A simple, fast, and accurate method of phylogenomic
863 inference. *Genome Biol* 2008; 9: R151.

864 82. Adeolu M, Alnajar S, Naushad S, S. Gupta R. Genome-based phylogeny and
865 taxonomy of the 'Enterobacteriales': proposal for Enterobacterales ord. nov.
866 divided into the families Enterobacteriaceae, Erwiniaceae fam. nov.,
867 Pectobacteriaceae fam. nov., Yersiniaceae fam. nov., Hafniaceae fam. nov.,

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Morganellaceae fam. nov., and Budviciaceae fam. nov. *Int J Syst Evol Microbiol* 2016; 66: 5575–5599.

83. Janda JM, Abbott SL. The Genus Hafnia: from Soup to Nuts. *Clin Microbiol Rev* 2006; 19: 12–28.

84. Lundgren JG, Lehman RM, Chee-sanford J. Bacterial Communities within Digestive Tracts of Ground Beetles (Coleoptera: Carabidae). *Ann Entomol Soc Am* 2007; 100: 275–282.

85. Estrada-de los Santos P, Palmer M, Chávez-Ramírez B, Beukes C, Steenkamp ET, Briscoe L, et al. Whole Genome Analyses Suggests that Burkholderia sensu lato Contains Two Additional Novel Genera (Mycetohabitans gen. nov., and Trinickia gen. nov.): Implications for the Evolution of Diazotrophy and Nodulation in the Burkholderiaceae. *Genes* 2018; 9: 389–412.

86. Partida-Martinez LP, Groth I, Schmitt I, Richter W, Roth M, Hertweck C. Burkholderia rhizoxinica sp. nov. and Burkholderia endofungorum sp. nov., bacterial endosymbionts of the plant-pathogenic fungus Rhizopus microsporus. *Int J Syst Evol Microbiol* 2007; 57: 2583–2590.

87. Partida-Martinez LP, Hertweck C. Pathogenic fungus harbours endosymbiotic bacteria for toxin production. *Nature* 2005; 437: 884–888.

88. Partida-Martinez LP, Hertweck C. A Gene Cluster Encoding Rhizoxin Biosynthesis in “Burkholderia rhizoxina”, the Bacterial Endosymbiont of the Fungus Rhizopus microsporus. *ChemBioChem* 2007; 8: 41–45.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

889 89. Moebius N, Üzüm Z, Dijksterhuis J, Lackner G, Hertweck C. Active invasion
890 of bacteria into living fungal cells. *eLife* 2014; 3:e03007

891 90. Partida-Martinez LP, Monajembashi S, Greulich K-O, Hertweck C.
892 Endosymbiont-Dependent Host Reproduction Maintains Bacterial-Fungal
893 Mutualism. *Curr Biol* 2007; 17: 773–777.

894 91. Dolatabadi S, Walther G, Gerrits van den Ende AHG, de Hoog GS. Diversity
895 and delimitation of Rhizopus microsporus. *Fungal Divers* 2014; 64: 145–163.

896 92. Lackner G, Hertweck C. Impact of Endofungal Bacteria on Infection Biology,
897 Food Safety, and Drug Development. *PLOS Pathog* 2011; 7(6):e1002096.

898 93. Lackner G, Möbius N, Scherlach K, Partida-Martinez LP, Winkler R, Schmitt I,
899 et al. Global Distribution and Evolution of a Toxinogenic Burkholderia-
900 Rhizopus Symbiosis. *Appl Environ Microbiol* 2009; 75: 2982–2986.

901 94. Plague GR, Dunbar HE, Tran PL, Moran NA. Extensive Proliferation of
902 Transposable Elements in Heritable Bacterial Symbionts. *J Bacteriol* 2008;
903 190: 777–779.

904 95. Manzano-Marín A, Latorre A. Settling Down: The Genome of Serratia
905 symbiotica from the Aphid Cinara tujafilina Zooms in on the Process of
906 Accommodation to a Cooperative Intracellular Life. *Genome Biol Evol* 2014; 6:
907 1683–1698.

908 96. Moran NA, Mira A. The process of genome shrinkage in the obligate symbiont
909 Buchnera aphidicola. *Genome Biol* 2001; 2(12):research0054.1.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

910 97. Moran NA. Tracing the evolution of gene loss in obligate bacterial symbionts.
911 *Curr Opin Microbiol* 2003; 6: 512–518.

912 98. Silva FJ, Latorre A, Moya A. Genome size reduction through multiple events of
913 gene disintegration in Buchnera APS. *Trends Genet* 2001; 17: 615–618.

914 99. Hamel E. Natural products which interact with tubulin in the vinca domain:
915 Maytansine, rhizoxin, phomopsin a, dolastatins 10 and 15 and halichondrin B.
916 *Pharmacol Ther* 1992; 55: 31–51.

917 100. Süssmuth RD, Mainz A. Nonribosomal Peptide Synthesis—Principles and
918 Prospects. *Angew Chem* Int Ed 2017; 56: 3770–3821.

919 101. Landsberg MJ, Jones SA, Rothnagel R, Busby JN, Marshall SDG, Simpson
920 RM, et al. 3D structure of the Yersinia entomophaga toxin complex and
921 implications for insecticidal activity. *Proc Natl Acad Sci* 2011; 108: 20544–
922 20549.

923 102. Ochman H. Distinguishing the ORFs from the ELFs: short bacterial genes and
924 the annotation of genomes. *Trends Genet* 2002; 18: 335–337.

925 103. Satoshi F, Nishikawa K. Estimation of the Number of Authentic Orphan Genes
926 in Bacterial Genomes. *DNA Res* 2004; 11: 219–231.

927 104. Moran NA, Dunbar HE, Wilcox JL. Regulation of Transcription in a Reduced
928 Bacterial Genome: Nutrient-Provisioning Genes of the Obligate Symbiont
929 Buchnera aphidicola. *J Bacteriol* 2005; 187: 4229–4237.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

930 105. Lackner G, Moebius N, Hertweck C. Endofungal bacterium controls its host by
931 an hrp type III secretion system. *ISME J* 2011; 5: 252–261.

932 106. Nivaskumar M, Francetic O. Type II secretion system: A magic beanstalk or a
933 protein escalator. *Biochim Biophys Acta BBA - Mol Cell Res* 2014; 1843: 1568–
934 1577.

935 107. Van Melderen L. Toxin–antitoxin systems: why so many, what for? *Curr Opin*
936 *Microbiol* 2010; 13: 781–785.

937 108. Norton JP, Mulvey MA. Toxin-Antitoxin Systems Are Important for Niche-
938 Specific Colonization and Stress Resistance of Uropathogenic Escherichia coli.
939 *PLOS Pathog* 2012; 8(10):e1002954.

940 109. McCutcheon JP, Moran NA. Functional Convergence in Reduced Genomes of
941 Bacterial Symbionts Spanning 200 My of Evolution. *Genome Biol Evol* 2010;
942 2: 708–718.

943 110. Bordenstein SR, Reznikoff WS. Mobile DNA in obligate intracellular bacteria.
944 *Nat Rev Microbiol* 2005; 3: 688–699.

945

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

946

947 **Figure Legends**

948 **Figure 1. Division of labor in amino acid biosynthesis and transport between**
949 ***Vallotia* and *Profftia* showing co-obligatory status of endosymbionts of *A.***
950 ***laricis/tardus*.** Amino acids produced by *Vallotia* and *Profftia* are shown in blue and
951 red, respectively. Bolded texts indicate essential amino acids. The insect host likely
952 supplies ornithine, homocysteine, 2-oxobutanoate and glutamine. Other compounds
953 that cannot be synthesized by the symbionts are shown in grey italics.

954

955 **Figure 2. Phylogenomic analysis showing the affiliation of the adelgid**
956 **endosymbiont '*Candidatus* Vallotia tarda' and its closest relatives, the fungus**
957 **endosymbionts *M. rhizoxinica* and *M. endofungorum* within the**
958 ***Burkholderiaceae*.** Free-living and pathogenic bacteria are colored in purple. Selected
959 members of *Oxalobacteraceae* (*Janthinobacterium agaricidamnosum* [HG322949],
960 *Collimonas pratensis* [CP013234] and *Herbaspirillum seropedicae* [CP011930]) were
961 used as outgroup. Maximum likelihood (IQ-TREE) and Bayesian analyses (MrBayes)
962 were performed based on a concatenated alignment of 108 proteins. Maximum
963 likelihood tree is shown. SH-aLRT support (%) and ultrafast bootstrap support (%)
964 values based on 1000 replicates, and Bayesian posterior probabilities are indicated on
965 the internal nodes. Asterisks stand for a maximal support in each analysis (100% / 1).

966

967 **Figure 3. (A) Synteny between the chromosome and plasmid of *Vallotia* and *M.***
968 ***rhizoxinica*, an endosymbiont of *Rhizopus* fungi**. The outermost and the middle

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

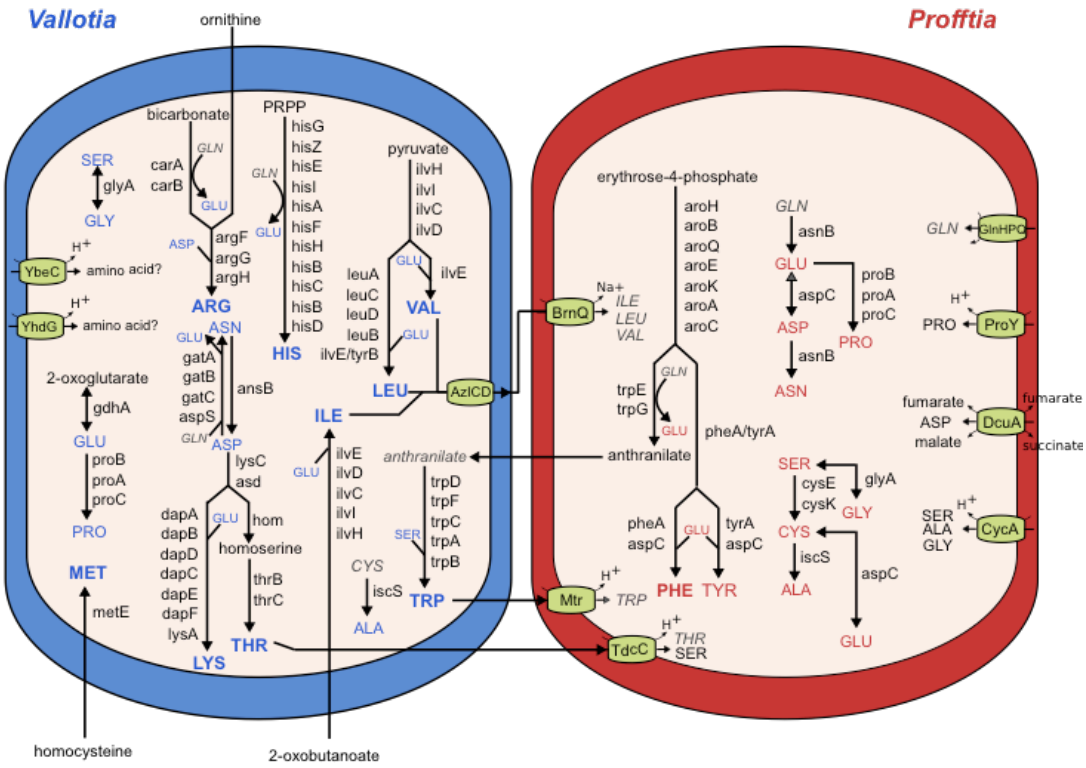
969 rings show genes in forward and reverse strand orientation, respectively. These
970 include rRNA genes in red and tRNA genes in dark orange. The innermost ring
971 indicates single-copy genes shared by *M. rhizoxinica* and *Vallotia* in black. Purple and
972 dark yellow lines connect forward and reverse matches between the genomes,
973 respectively. The two small contigs involving the rRNA genes of *Vallotia* are not
974 shown. **(B) Close-up of the largest deletion on the chromosome of *M. rhizoxinica***
975 **and the syntenic region on the *Vallotia* chromosome.** Genes are colored according
976 to COG categories. Yellow: secondary metabolite biosynthesis; red: transposase;
977 grey: unknown function; khaki: replication, recombination and repair; pink: lipid
978 transport and metabolism; brown: protein turnover and chaperones; dark green: amino
979 acid transport and metabolism; light green: cell envelope biogenesis; black:
980 transcription.
981

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

	Profftia	*Vallotia*	
Feature	**Chromosome**	**Chromosome**	**Plasmid**
Genome size (bp)	1,225,795	1,110,884	72,431
GC (%)	31.9	42.8	36.1
Coding density (%)	52.4	65.6	46.2
CDS	645	751	29
rRNA genes	3	3*	0
tRNAs	35	37	0
Pseudogenes	21	47	5

Table 1. Genomic features of *Profftia* and *Vallotia.* **Vallotia* likely has three copies of each rRNA gene based on the sequence coverage of the corresponding contigs involving these genes.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



988

989 **Figure 1. Division of labor in amino acid biosynthesis and transport between**
990 ***Vallotia* and *Profftia* showing co-obligatory status of endosymbionts of *A.***
991 ***laricis/tardus*.** Amino acids produced by *Vallotia* and *Profftia* are shown in blue and
992 red, respectively. Bolded texts indicate essential amino acids. The insect host likely
993 supplies ornithine, homocysteine, 2-oxobutanoate and glutamine. Other compounds
994 that cannot be synthesized by the symbionts are shown in grey italics.

995

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

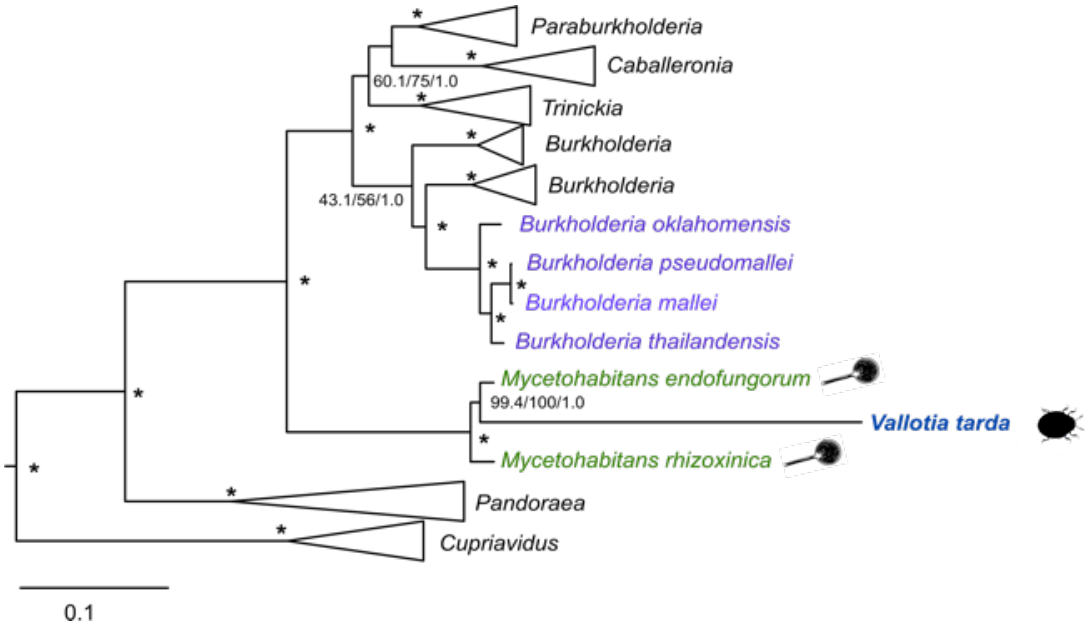


Figure 2. Phylogenomic analysis showing the affiliation of the adelgid endosymbiont '*Candidatus* Vallotia tarda' and its closest relatives, the fungus endosymbionts *M. rhizoxinica* and *M. endofungorum* within the *Burkholderiaceae*. Free-living and pathogenic bacteria are colored in purple. Selected members of *Oxalobacteraceae* (*Janthinobacterium agaricidamnosum* [HG322949], *Collimonas pratensis* [CP013234] and *Herbaspirillum seropedicae* [CP011930]) were used as outgroup. Maximum likelihood (IQ-TREE) and Bayesian analyses (MrBayes) were performed based on a concatenated alignment of 108 proteins. Maximum likelihood tree is shown. SH-aLRT support (%) and ultrafast bootstrap support (%) values based on 1000 replicates, and Bayesian posterior probabilities are indicated on the internal nodes. Asterisks stand for a maximal support in each analysis (100% / 1).

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.16.342642; this version posted October 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

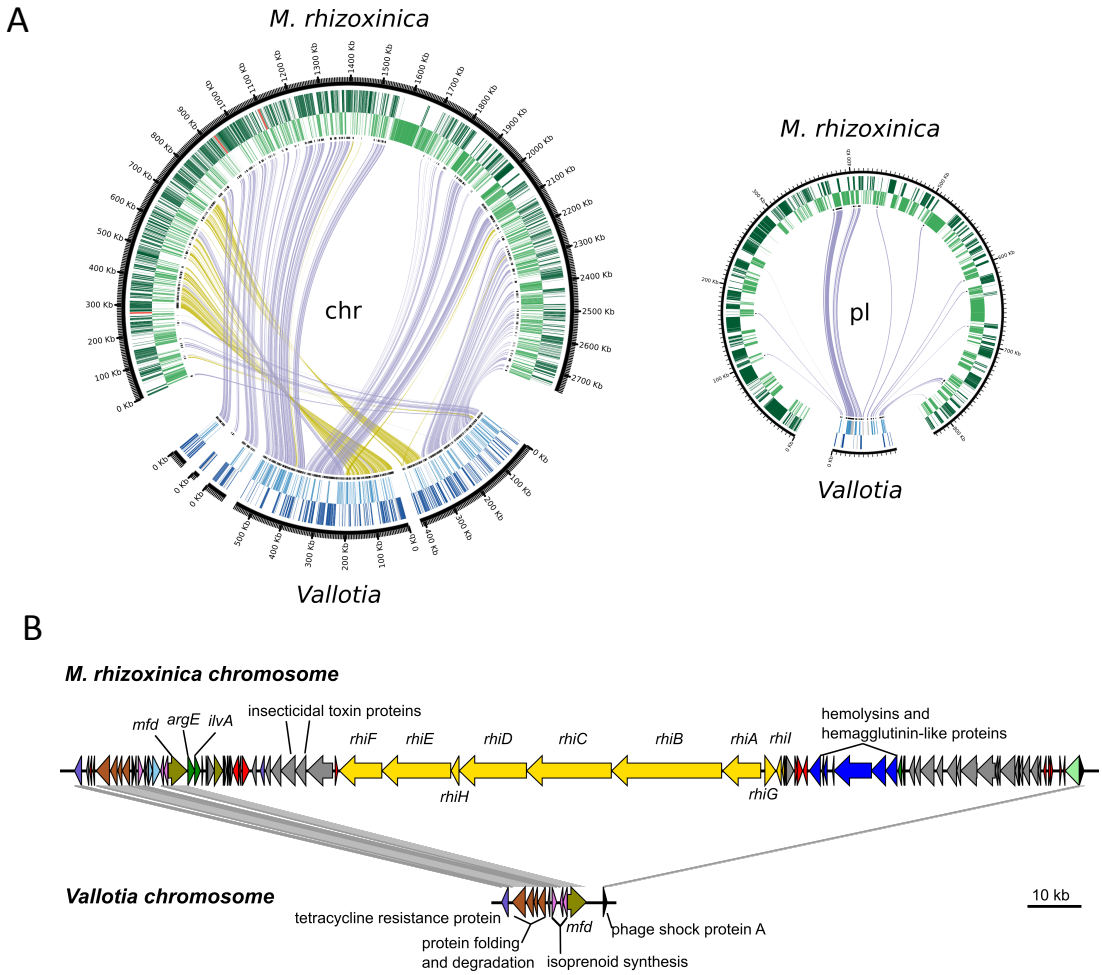


Figure 3. (A) Synteny between the chromosome and plasmid of *Vallotia* and *M. rhizoxinica*, an endosymbiont of *Rhizopus* fungi. The outermost and the middle rings show genes in forward and reverse strand orientation, respectively. These include rRNA genes in red and tRNA genes in dark orange. The innermost ring indicates single-copy genes shared by *M. rhizoxinica* and *Vallotia* in black. Purple and dark yellow lines connect forward and reverse matches between the genomes, respectively. The two small contigs involving the rRNA genes of *Vallotia* are not shown. **(B) Close-up of the largest deletion on the chromosome of *M. rhizoxinica* and the syntenic region on the *Vallotia* chromosome.** Genes are colored according to COG catergories. Yellow: secondary metabolite biosynthesis; red: transposase; grey: unknown function; khaki: replication, recombination and repair; pink: lipid transport and metabolism; brown: protein turnover and chaperones; dark green: amino acid transport and metabolism; light green: cell envelope biogenesis; black: transcription.