Phylogenomic analysis and metabolic role reconstruction of mutualistic Rhizobiales hindgut symbionts of Acromyrmex leaf-cutting ants

Mariya Zhukova, Panagiotis Sapountzis, Morten Schiøtt, Jacobus Boomsma

FEMS Microbiology Ecology, 2022, 98, 1–13
DOI: 10.1093/femsec/fiac084
Advance access publication date: 29 July 2022
Research Article

Phylogenomic analysis and metabolic role reconstruction of mutualistic *Rhizobiales* hindgut symbionts of *Acromyrmex* leaf-cutting ants

Mariya Zhukova [1], Panagiotis Sapountzis [1,2], Morten Schiøtt[1,3], Jacobus J. Boomsma [1,*]

[1]Centre for Social Evolution, Department of Biology, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen, Denmark
[2]Present address: Université Clermont Auvergne, INRAE, UMR 0454 MEDIS, 63000 Clermont-Ferrand, France
[3]Present address: Unit for Fish and Shellfish Diseases, National Institute of Aquatic Resources, Technical University of Denmark, Kemitorvet Bldg. 202, 2800 Kgs. Lyngby, Denmark
*Corresponding author: Centre for Social Evolution, Department of Biology, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen, Denmark. E-mail: JJBoomsma@bio.ku.dk
One sentence summary: Analysis of genomes of Rhizobiales symbionts of Acromyrmex reveals a dependence of these bacteria on their host, while they supply ants with compounds that enhance the robustness of the mutualism.
Editor: Ville-Petri Friman

Abstract

Rhizobiales are well-known plant-root nitrogen-fixing symbionts, but the functions of insect-associated Rhizobiales are poorly understood. We obtained genomes of three strains associated with *Acromyrmex* leaf-cutting ants and show that, in spite of being extracellular gut symbionts, they lost all pathways for essential amino acid biosynthesis, making them fully dependent on their hosts. Comparison with 54 Rhizobiales genomes showed that all insect-associated Rhizobiales lost the ability to fix nitrogen and that the *Acromyrmex* symbionts had exceptionally also lost the urease genes. However, the *Acromyrmex* strains share biosynthesis pathways for riboflavin vitamin, queuosine and a wide range of antioxidant enzymes likely to be beneficial for the ant fungus-farming symbiosis. We infer that the Rhizobiales symbionts catabolize excess of fungus-garden-derived arginine to urea, supplementing complementary Mollicutes symbionts that turn arginine into ammonia and infer that these combined symbiont activities stabilize the fungus-farming mutualism. Similar to the Mollicutes symbionts, the Rhizobiales species have fully functional CRISPR/Cas and R-M phage defenses, suggesting that these symbionts are important enough for the ant hosts to have precluded the evolution of metabolically cheaper defenseless strains.

Keywords: antioxidants, complementary mutualistic functionality, CRISPR/Cas, domestication, gene loss, riboflavin

Introduction

The Rhizobiales are a highly diverse order of Gram-negative Alphaproteobacteria. They are best known as nitrogen-fixing root nodule symbionts of plants (Garrido-Oter et al. 2018) but also have several pathogenic lineages, most notably the genus *Bartonella* (Okaro et al. 2017). Since the turn of the millennium, new lineages have been identified with increasing efficiency. For example, lichens are now known to harbor strains of Rhizobiales that may provide crucial mutualistic services (Erlacher et al. 2015). Rhizobiales symbionts have also been increasingly identified as gut symbionts of insects, particularly of ants which are dominant arthropods in many terrestrial ecosystems (Hölldobler and Wilson 1990; Chapter 11). It was initially suggested that this related to many ants, particularly those living in the canopy, being functional herbivores in need of nitrogen supplementation (Davidson et al. 2003), a contention that appeared to be confirmed when several ant lineages could be shown to have independently acquired Rhizobiales gut symbionts (Russell et al. 2009). However, recent studies have failed to confirm the presence of nitrogen fixation genes in ant-associated Rhizobiales, both in predatory *Harpegnathos saltator* and in *Dolichoderus* and *Cephalotes* ants that are general scavengers as well as functional herbivores as they consume pollen, fungi, extrafloral nectar, or feed on exudates of Hemiptera (Neuvonen et al. 2016, Bisch et al. 2018, Hu et al. 2018). Instead, '*Candidatus* Tokpelaia hoelldoblerii', a symbiont of predatory *H. saltator* ants and '*Ca*. T. dolichoderi', a symbiont of herbivorous *Dolichoderus* ants were suggested to have roles in nitrogen recycling to ammonia via urease, which mediates hydrolysis of urea into CO_2 and NH_3 (Neuvonen et al. 2016, Bisch et al. 2018). Some of the Rhizobiales strains of *Cephalotes* ants had genes coding for both urease and arginase, suggesting a longer catalyzation reaction starting with converting arginine to urea and ornithine (Hu et al. 2018).

The attine fungus-farming ants are quintessential herbivorous ants unable to ingest any animal food as they feed on fungal tissues grown in their symbiotic fungus garden (Mueller, Rehner and Schultz 1998, Mueller et al. 2001). Across a set of attine ant genera that are exploited by attine-ant-associated agropredator and thief ant species of the genus *Megalomyrmex*, Rhizobiales gut symbionts were shown to be common and occasionally abundant, but identification did not proceed further than the 97% similarity of V4 16S rRNA fragment (Liberti et al. 2015). Other gut-microbiome studies focused on the terminal clade of the *Atta* and *Acromyrmex* leaf-cutting ants, showing that the diversity of the gut bacterial

Received: March 4, 2022. **Revised:** July 3, 2022. **Accepted:** July 27, 2022

community of adult *Acromyrmex* workers is remarkably low and almost completely dominated by *Wolbachia*, Mollicutes, and Rhizobiales bacteria (Sapountzis et al. 2015, 2019). To better understand the functional significance of these bacteria, which were all inferred to somehow provide mutualistic services to the farming symbiosis, we started to obtain genome sequences of these symbionts. This allowed us to sequence two of the dominant Mollicutes species, one typical for the leaf-cutting ants and another typical for the phylogenetically more basal attine lineages (Sapountzis et al. 2018).

The Rhizobiales associated with *Acromyrmex* are extracellular symbionts, forming biofilms supported by a polysaccharide matrix in the hindgut of large workers (Sapountzis et al. 2015). In an earlier study, we showed that these *Rhizobiales* bacteria usually represent a significant fraction of the gut microbiome in field samples of *A. octospinosus* workers (although data on presence/absence across individual ants are lacking), but that they are absent or only detectable in fractions < 1% in the guts of *A. echinatior* workers, larvae and pupae in the field (Sapountzis et al. 2015, Zhukova et al. 2017). We therefore suspect that the Rhizobiales species are secondary symbionts that have adaptive advantages but are not strictly required for host survival, at least in *A. echinatior* (Dale and Moran 2006). Our previous study showed colocalization of Rhizobiales bacteria and NifH protein production for dinitrogenase reductase in the hindgut of both *A. octospinosus* and *A. echinatior* workers (Sapountzis et al. 2015), which made us expect that these symbionts could have roles in the recycling or even fixation of nitrogen. We cultivated *in vitro* and sequenced the genome of a Rhizobiales associated with *A. octospinosus* and we reconstructed two Rhizobiales metagenome-assembled genomes from metagenomes obtained from ileums of two workers of *A. echinatior*. The alternative approaches of metagenomic and single-colony sequencing each have their advantages and disadvantages; cultivation allows obtaining pure DNA from one species, but many symbionts cannot grow *in vitro*. Metagenomic sequencing allows obtaining genomes for all gut bacteria, but may lead to chimeric contigs, especially when there are closely related bacterial strains. Applying both methods, our primary goal was to produce high quality draft genomes for both *Acromyrmex* species in order to: (i) compare the genomes of Rhizobiales associated with *Acromyrmex* guts with known genomes of other Rhizobiales symbionts in a phylogenetic context, (ii) predict the mutualistic functions of these symbionts, both in general and as far as variable across the two specific host ant species, and (iii) assess the potential strength of the antiphage defense systems maintained by the Rhizobiales gut symbionts.

Materials and methods

Collection, rearing, and sampling of ant colonies

Acromyrmex echinatior (colony Ae331) and *A. octospinosus* (colony Ao492) were collected in Panama in, respectively, 2007 (Gamboa) and 2010 (Tupper) and maintained at the Centre for Social Evolution, University of Copenhagen, at ca. 25åC and 70% relative humidity with a 12:12 h photoperiod. The two colonies had been kept under laboratory conditions for more than two years before sampling started.

Bacterial culturing and preparation for genome sequencing

Ten guts from workers of colony Ao492 were dissected in sterile PBS buffer and cautiously homogenized with a plastic pestle. The liquid obtained was spread on plates with YMA medium and incubated in a BBL GasPak 150 anaerobic chamber (Franklin Lukes, New Jersey, USA). Bacterial colonies grew after ca. 2 weeks and were identified by amplification and sequencing of the partial 16S rRNA gene. Single bacterial colonies were picked with sterile plastic tips to perform multiple displacement amplification (MDA).

Ten workers from colony Ae331 were anesthetized and surface sterilized by submergence in 70% ethanol for 1 min, after which they were rinsed in autoclaved MilliQ water, submerged in 50% bleach for 2 min, and rinsed again in autoclaved MilliQ water. The ileum, the gut section where we expected the highest abundance of Rhizobiales bacteria (Sapountzis et al. 2015), was dissected from each worker in autoclaved PBS buffer using sterile forceps and blades. DNA was extracted from each ileum individually using a DNeasy blood and tissue kit (Qiagen) according to the manufacturer's instructions, with an extra step at which glass beads of 0.5 mm were added to the lysis step and vortexed for 30 s. An additional centrifugation for 1 min at 20 000 g was performed to remove the AW2 washing buffer. All samples were eluted using 100 ml of AE elution buffer. DNA samples were subjected to PCR using the 16S generic primers 515F and 806R (Caporaso et al. 2011) as previously described (Sapountzis et al. 2015), purified using the Invitek kit (Westberg, Germany), and sent to MWG (Germany) for Sanger sequencing. Two samples with confirmed 16S amplicons of Rhizobiales origin and minimal signs of contamination with 16S rRNA gene sequences of other bacteria on our chromatographs were chosen to perform MDA reactions.

MDA reactions and DNA extraction

Live bacterial cells from culture (colony Ao492) and approximately 1 μl of extracted DNA (colony Ae331) were used for MDA reactions to obtain whole genome DNA using the Qiagen REPLI-g Midi Kit following the manufacturer's instructions. A blank reaction using sterile water as template instead of 1 μl of bacterial cell suspension or DNA was included in the same protocol to check for bacterial contaminations with eubacterial 515F/806R primers after the entire procedure was completed. This showed no detectable 16S rRNA gene amplicons that would indicate contamination. The MDA-amplified DNA was further purified using the Invitek kit (Westberg, Germany) following the manufacturer's instructions. The extracted DNA was then quantified for all samples using a spectrophotometer (Nanodrop) and sent to seqIT (Germany) where libraries were generated from 100 to 200 ng of DNA using the Nextera XT kit (Illumina, USA). MiSeq sequencing was performed at two times 250-bp read length, which generated 10 00000 paired reads for the colony Ao492 sample, and 30 00000 and 40 00000 for the two samples obtained from colony Ae331.

Assembly, annotation, quality control, and other genomic analyses

Samples from bacterial culture and metagenomic ileum dissections were processed with some differences necessary to obtain an optimal final assembly. The Nextera adaptors were removed from the fastq files using Trim Galore (Babraham Institute) for the sample from colony Ao492 and using Skewer (Jiang et al. 2014) for the two colony Ae331 samples after which the filtered reads were checked with FastQC (Andrews 2010). We used the SPAdes Genome Assembler (version 3.5.0) to generate *de novo* assemblies using the '–careful' option, which reduced the number of mismatches and short indels, before running MismatchCorrector with kmer sizes of 21, 33, 55, and 77 to obtain a consensus

assembly based on four individual assemblies for sequenced bacterial colony. We used SPAdes (version 3.11.1) with the '–meta' option recommended for metagenomic data sets with kmer sizes of 21, 33, 55, 77, 99, and 127 to obtain a consensus assembly for the metagenomic samples (Bankevich et al. 2012). Contigs with lengths of less than 250 bp and with coverage below 2.5x were removed from the subsequent analyses. To separate genome fragments of bacterial and ant origin within assembled metagenomes, we used MyCC with minimal contig length of 1500 bp for the first-step clustering, and minimum contig length of 250 bp afterwards (Lin and Liao 2016). Sequences of each cluster were then compared locally with the Uniref90 database using BLASTX v2.7.1+ (evalue <1e-15, percentage identities >30%) to assess phylogenetic binning performed by MyCC. *RAo1* contigs were excluded from this procedure because DNA came from a single Rhizobiales isolate and no binning was necessary. Genes for all contigs of the Rhizobiales cluster were predicted using the RAST annotation server (Aziz et al. 2008). Predicted amino acid sequences were then compared locally with the Uniref70 database, using BLASTP v2.7.1+ (evalue <1e-10, percentage identities >30%) to remove contigs of *Acromyrmex*, *Wolbachia*, and Mollicutes origin. The latter two represent the only other dominant elements of *Acromyrmex* gut microbiomes. In this procedure, we used a broad definition of microbiome that encompasses both intracellular and extracellular within-host habitat (Bordenstein and Theis 2015), because *Wolbachia* can both be intracellular and extracellular in *Acromyrmex* hosts (Andersen et al. 2012). Completeness of genome assembly of all samples was assessed using BUSCO v5 with a near-universal single-copy orthologs database for Rhizobiales bacteria (Manni et al. 2021) and CheckM v1, with a marker set for Alphaproteobacteria (Parks et al. 2015). BUSCO assesses the presence of gene orthologs, while CheckM evaluates the presence of co-located marker sets (operons) and was developed specifically for prokaryotes. The final sets of assembled contigs and raw reads were deposited in the NCBI database (Bioprojects PRJNA476860 and PRJNA476944).

As an extra precaution in evaluating the absence of the genes suggested as missing by BUSCO, we built a database of 16783 predicted proteins identified from Rhizobiales genomes using OrthoDB v10.1 (Kriventseva et al. 2019). The raw sequence reads that were used for assembling the *RAo1*, *RAe6*, and *RAe9* genomes (after removing the adapters) were compared locally to the Rhizobiales protein database using BLASTX v2.9.0. The minimal length of sequence similarity was set to 70 amino acids (equal to 210 nucleotides, ca. 85% length of an average sequence read), the minimal sequence identity was set to 70% (Halachev, Loman and Pallen 2011).

We functionally annotated the protein sequences obtained using a standalone version of InterproScan v-5.27–66.0 with default analyses (CDD-3.16, Coils-2.2.1, Gene3D-4.1.0, Hamap-2017_10, MobiDBLite-1.0, PANTHER-12.0, Pfam-31.0, PIRSF-3.02, PRINTS-42.0, ProDom-2006.1, ProSitePatterns-2017_09, ProSiteProfiles-2017_09, SFLD-3, SMART-7.1, SUPERFAMILY-1.75, TIGRFAM-15.0). To identify and compare metabolic pathways we used the BlastKOALA tool provided by the KEGG database (Kanehisa, Sato and Morishima 2016). CRISPRCasFinder was used to identify *cas* genes and adjacent spacers (Couvin et al. 2018). A search for genes of restriction-modification systems was performed in BLASTP (percentage identities >50%) using 'gold standard' database of REBASE (Roberts et al. 2015). Adhesin domains were determined using MOTIF Search (GenomeNet, Japan) after which topology was predicted with PRED-TMBB2 (Tsirigos, Elofsson and Bagos 2016). To compare the three sequenced Rhizobiales genomes associated with *Acromyrmex*, we used Mauve, a program for multiple genome alignment (Darling, Mau and Perna 2010).

Phylogenomic analyses

For the phylogenetic reconstructions, we first downloaded 54 Rhizobiales genomes from Ensembl (Zerbino et al. 2018) and NCBI databases (Table S1), which included symbionts of other social insects (bees and ants), pathogenic and free-living bacteria (accessed in January 2018). The *Bartonella* species except for *B. tamiae* and *B. apis*, will be referred to as the eubartonellae in accordance with previous studies (Zhu et al. 2014, Segers et al. 2017). We used the entire data set to determine the orthologous single-copy protein-coding genes using the OrthoFinder software (Emms and Kelly 2015), which resulted in 294 genes being available for our phylogenetic analyses. Both the nucleotide and amino acid sequences of these genes were extracted for all 57 genomes (the three Rhizobiales genomes obtained here and the 54 from the public databases), after which gene-specific alignments were constructed using MUSCLE v3.8.31 (Edgar 2004). Gaps in the alignments were removed with the trimAl software, using the automatic—gappyout function to determine optimal thresholds for column removals (Capella-Gutierrez, Silla-Martinez and Gabaldon 2009). The trimmed alignments were further tested for recombination using the Phipack software (Bruen, Philippe and Bryant 2006) and for nucleotide saturation using the Xie test implemented in DAMBE5 (Xia 2013). Individual alignments for the remaining genes were concatenated using Amas (Borowiec 2016).

We used the filtered alignments (both nucleotides and amino acids) for the three *Acromyrmex* symbionts and the comparable alignments for the other 54 Rhizobiales genomes to reconstruct phylogenies with maximum likelihood (ML) and Bayesian methods. Partitioning schemes and appropriate substitution models for the nucleotide and protein alignments were selected with PartitionFinder 2 (Lanfear et al. 2016). PartitionFinder produced 31 partitions for the concatenated nucleotide alignment using the GTR model with gamma distributed rate variation among sites and allowing for a proportion of invariable sites as a best-fit model of evolution for nucleotide alignment. For the amino acid alignment, 37 partitions with optimal models available in RaxML and MrBayes were chosen. For the ML analyses, we used the RaxML software v.8.2.10 (Stamatakis 2014), which applies a rapid bootstrapping algorithm performing 100 bootstrap sampling replications, after which bootstrapping is stopped automatically with the extended majority-rule consensus tree criterion. For the Bayesian inferences, we used the MrBayes v3.2.6 software (Ronquist and Huelsenbeck 2003). The concatenated alignments were specified for each partition and a variable rate of sequence evolution (ratepr) was allowed for each partition. Five chains were run for 10 00000 generations for nucleotide alignment and four chains for 50 0000 generations for the amino acid alignments. Appropriate average standard deviations of split frequencies (a measure which allowed us to assess whether two independent calculations of a tree converged onto a stationary distribution) were achieved for both nucleotide (0.0003) and protein alignments (0.0001), which were below the recommended 0.01 value and thus indicated sufficient convergence. All trees were further visualized using Mesquite 3.61 software (Maddison and Maddison 2019). To assess branch-specific gene gains and losses, orthogroups produced by OrthoFinder were mapped onto the phylogenetic tree using asymmetric Wagner parsimony implemented in COUNT(Csűrös 2010).

Results and discussion

Genome features, phylogenetic comparisons, and inferred metabolic functions of symbiotic Rhizobiales

The sequence reads of a single cultured bacterial colony (*RAo1*) isolated from the hindgut of an *A. octospinosus* worker were assembled *de novo* into 55 contigs with a total length of 1.96 Mb. The 16S rRNA sequence in this genome was 99% similar to the sequence of the *RhiAcro1* OTU identified earlier in *Acromyrmex* ants from the same field site (Sapountzis et al. 2015). The hindgut metagenomes of two *A. echinatior* workers from one lab colony were also individually isolated, sequenced and *de novo* assembled and binned based on GC content and coverage (Fig. S1), after which all bins were checked using BLAST for contigs with Rhizobiales, *Wolbachia*, and *Acromyrmex* genes to avoid missing relevant contigs or including contaminations. The two specific Rhizobiales bins (*RAe6* and *RAe9*) consisted of 52 and 40 contigs, respectively, with a total length of 1.86 and 1.87 Mb (for details, see Table S2). Genome coverage was 180x, 505x, and 372x for *RAo1, RAe6*, and *RAe9*, respectively. The 16S rRNA alignment showed 100% identity between *RAe6* and *RAe9* and 99% similarity with *RAo1*. Comparison of RAST annotations for these three genomes showed that the numbers of genes in each functional category were very similar (Table S2) except for the category 'Phages, prophages, transposable elements, plasmids' for which *RAo1* had 19 genes and *RAe6* and *RAe9* only five. We also found that the 16S rRNA sequence of the OTU *RhiAcro1* was 95% identical to the '*Ca.* T. hoelldoblerii' isolated from *H. saltator* ants and 94% identical to *Bartonella apis*, a gut symbiont of the honey bee *Apis mellifera*. We therefore taxonomically relate our Rhizobiales strains to the *Tokpelaia* genus, as a threshold of 94.5% 16S rRNA gene identity was suggested for bacterial genera (Yarza et al. 2014).

Fifty-four published Rhizobiales genomes were used for phylogenomic reconstruction, including two symbionts of social insects ('*Ca.* T. hoelldoblerii' and *B. apis*), mammalian *Bartonella* pathogens, root nodule symbionts of legumes and free-living species (Table S1). Analysis of 294 concatenated orthologous sequences produced nearly identical trees for maximum likelihood (Fig. 1, Fig. S2) and Bayesian analysis of both nucleotide and amino acid sequences (Fig. S2). The trees revealed that the three strains representing *RhiAcro1* are closely related to the Rhizobiales symbiont of a predatory ant '*Ca.* T. hoelldoblerii', consistent with our 16S rRNA assignments (Fig. 1, Fig. S2). '*Ca.* T. hoelldoblerii' and *RhiAcro1* came out as sister lineages when we used amino acid sequences, while '*Ca.* T. hoelldoblerii' diverged from the other *Bartonella* species prior to *RhiAcro1* in the analyses based on nucleotide sequences. Our phylogenetic trees match previous comparative studies (Neuvonen et al. 2016, Segers et al. 2017) showing that a branch of ant-associated symbionts (*RhiAcro1* and '*Ca.* T. hoelldoblerii') diverged prior to *B. tamiae* and *B. apis*, which form a basal lineage in the sister clade of eubartonellae mammalian pathogens (Zhu et al. 2014).

Wagner parsimony analysis implemented in Count (Csűrös 2010, Fig. 1) revealed a general trend of massive gene loss and much slower gene acquisition in the focal clade of our study, consistent with reductive genomic evolution of symbiotic bacteria (Dagan et al. 2006, McCutcheon and Moran 2012). The observed gene losses could be related to incomplete genome sequencing or improper *de novo* assembly. We therefore assessed genome completeness using BUSCO (Manni et al. 2021) and CheckM (Parks et al. 2015) and compared the results with values for the complete, closed genome of the '*Ca.* T. hoelldoblerii' strain *Hsal* (Table S2). *RAo1, RAe6* and *RAe9* had 89.67%, 89.2%, and 89.5% complete BUSCO orthologs, respectively, while the complete genome of '*Ca.* T. hoelldoblerii' had 96.09% of the BUSCO orthologs. Similar results were obtained using CheckM, where *RAo1, RAe6*, and *RAe9* had 90.54%, 90.76%, and 90.76% of the marker sets, while the complete genome of '*Ca.* T. hoelldoblerii' had 96.96% (Table S3). We subsequently compared orthologous genes and marker sets that were missing in each isolate (Tables S4 and S5). This showed that the missing genes were exactly the same for all three *RhiAcro1* isolates (65 orthologs in BUSCO), with addition of one missing marker set unique for *RAo1* found by CheckM (27 marker sets in *RAo1* vs. 26 marker sets in *RAe6* and *RAe9*). For CheckM, one missing marker set was common for *RhiAcro1* isolates and '*Ca.* T. hoelldoblerii' while six missing orthologs were the same in *RhiAcro1* isolates and '*Ca.* T. hoelldoblerii' in BUSCO analysis. BUSCO provided not only codes for the identified genes, but also nucleotide and amino acid sequences. This allowed us to further compare the raw sequence reads used to construct the three genomes against the predicted proteins of the gene orthologs of Rhizobiales that were suggested as missing in *RhiAcro1*. Sequence reads from all three *RhiAcro1* isolates showed homology to two known orthologs: a 'signal transduction response regulator, C-terminal effector' and a 'CLp, N terminal'. For *RAe6*, a few reads also showed similarity to pyridoxal phosphate-dependent transferase and for *RAe9* to leucyl/phenylalanyl-tRNA-protein transferase. These findings suggest that some genes may have been missing from the final assemblies, but taking into account that our sequenced genomes had high coverage (180x, 505x, and 372x for *RAo1, RAe6*, and *RAe9*) and 90.54%–90.76% completeness (based on CheckM analyses), we inferred that the majority of gene losses are likely to be true gene losses, also because the reads mapped both to missing genes and to other genomic regions with high confidence (98%–100% sequence identity, 98%–100% read length).

RhiAcro1 had more gene families (1748–1756 across its three isolates *RAo1*, *RAe6*, and *RAe9*) than '*Ca.* T. hoelldoblerii' (1617). Functional gene gains and losses included genes for flagellum construction, urease genes, and some genes for amino acid biosynthesis pathways that we will consider below. Among the genes lost by *RhiAcro1* in comparison with '*Ca.* T. hoelldoblerii' was a complete pathway for anaerobic fermentation of glucose to 2,3–butanediol, although the role of this compound for bacteria and hosts is unknown. We suggest that the *RhiAcro1* strains are microaerophilic, because these symbionts have genes for the complete citric acid cycle and for all protein complexes necessary for oxidative phosphorylation. Moreover, we obtained *RAo1* colonies in an anaerobic chamber with low oxygen and no growth was obtained under aerobic conditions. Similar to '*Ca.* T. hoelldoblerii', *RhiAcro1* uses *N*-acetylglucosamine (GlcNAc) and chitobiose (a dimer of glucosamine) for peptidoglycan biosynthesis when forming its cell walls. Both compounds are also part of fungal cell walls and thus readily available via the food that the fungus-growing ants ingest.

Nitrogen metabolism

It has been suggested that Rhizobiales bacteria can fix atmospheric nitrogen to supplement the nitrogen-poor diets of herbivorous ants (Russell et al. 2009). This conjecture relates to nitrogenase being a highly conserved enzyme with two components, a heterotetrameric core encoded by *nifD* and *nifK*, and a dinitrogenase reductase subunit encoded by *nifH*. Other nitrogenase genes have regulatory functions and can differ between bacterial species (Fischer 1994). In a previous study we detected and sequenced the *nifH* gene in the hindguts of *A. echinatior*, where *Rhi-*

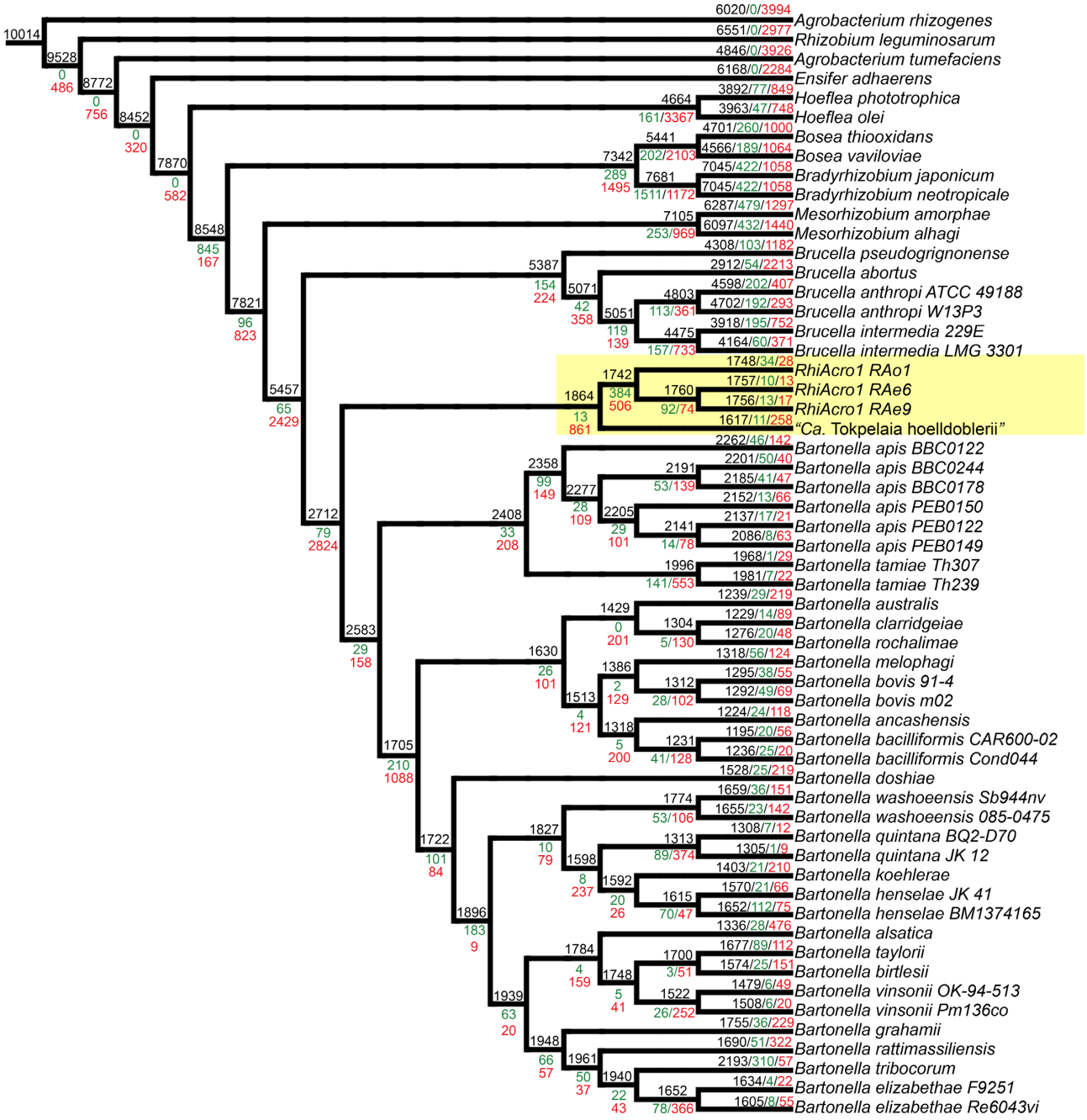

Figure 1. Rooted phylogenetic tree illustrating the evolutionary relationships among 57 fully sequenced Rhizobiales with the inferred number of genes gained and lost after divergence from the common ancestor. The maximum likelihood tree was constructed using a concatenated amino acid alignment of 294 single-copy orthologs with *Agrobacterium rhizogenes* as outgroup. Branch lengths are not drawn to scale. Black numbers correspond to the total number of genes having orthologs (94.2%), green numbers are genes gained, and red numbers are genes lost. The Rhizobiales gut symbionts obtained from *Acromyrmex octospinosis* (RAo1), *A. echinatior* (RAe6, RAe9) and *Harpegnathos saltator* ('*Ca.* T. hoeldoblerii') are highlighted in yellow.

Acro1, identified with 16S-MiSeq sequencing, was present in the exact same locations (Sapountzis et al. 2015). However, The *RhiAcro1* strains isolated in our present study from *A. echinatior* and *A. octospinosus* lacked the genes of the molybdenum-containing nitrogenase system (*nifH, nifD, nifK*), the iron-containing nitrogenase system (*anfH, anfD, anfG, anfK*), and the vanadium-containing nitrogenase system (*vnfD, vnfG, vnfK, vnfH*). We also checked for the presence of the urease system, which catalyzes hydrolysis of urea into CO_2 and NH_3 and was found previously in *B. apis* and '*Ca.* T. hoelldoblerii' (Neuvonen et al. 2016, Segers et al. 2017), but all urease-related genes were lacking in the symbiont samples that we examined (Fig. 2; Table S6). Urease genes were present in representatives of the Brucellaceae family investigated (*Brucella*), whereas all mammal-associated eubartanellae and *B. tamiae* have lost the urease genes. Bacteria that have a urease system can use ammonium obtained from urea hydrolysis, catalysed by urease, to synthesize glutamine, which *RhiAcro1* is also capable to do but with other sources of nitrogen.

Scanning for the presence of nitrogen-fixing and urease systems across all bacterial species included in our phylogenetic tree showed that both types of genes were present only in *Bradyrhizobium*, *Agrobacterium*, and in *Rhizobium* symbionts of legume root

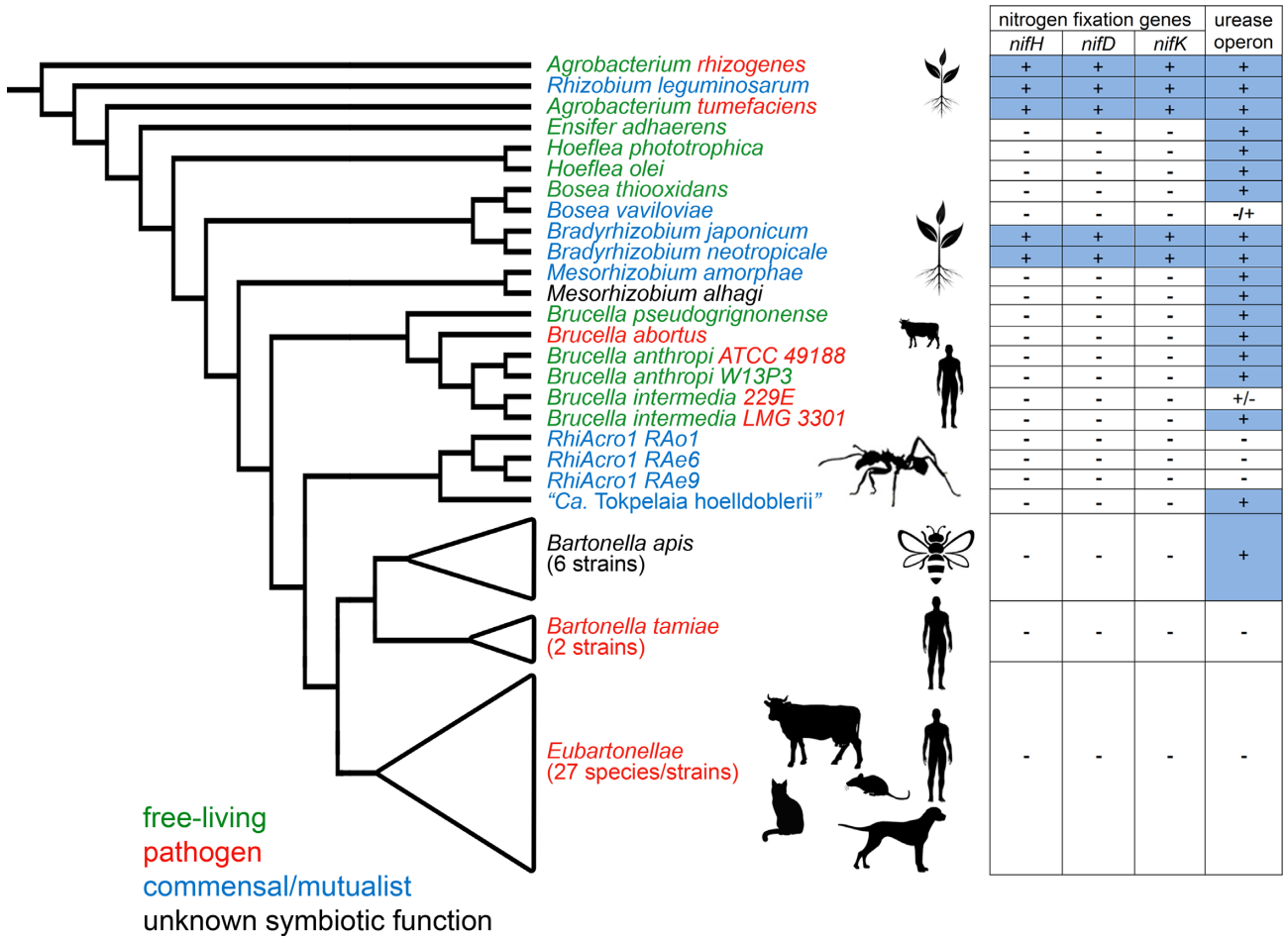

Figure 2. Losses (–) and gains (+) of genes for nitrogen fixation and urea conversion in Rhizobiales over evolutionary time. The tree was constructed based on maximum likelihood methods with *Agrobacterium rhizogenes* as outgroup. The complete list of species belonging to the unspecified triangle clades is given in Fig. 1 and Table S1. Colors highlight bacterial lifestyles; species names written in two colors means that bacteria can both be free-living and pathogenic. Branch lengths are not drawn to scale. For more detailed information about urease operon, see Table S6.

nodules (Fig. 2). Nitrogen-fixing systems were already lost in the Brucellaceae (*Brucella*), a branch that diverged from the legume root nodule symbionts before the *Tokpelaia* and *Bartonella* species, suggesting that none of the *Tokpelaia* and *Bartonella* symbionts from honey bees, ants or any mammalian hosts are likely to have nitrogen fixation genes unless they acquired them *de novo* by horizontal gene transfer. Considering that the V4 region of the 16S gene is fairly conserved so that rather many bacteria may be identical for this 16S sequence (Ellegaard and Engel 2016, Ellegaard et al. 2020) and that our previous study amplified and sequenced Rhizobiales-like *nifH* sequences (Sapountzis et al. 2015), it seems likely that there is at least one other *nifH* gene-carrying bacterium in the *Acromyrmex* guts. We will return to this issue in connection to the CRISPR results below, but infer already here that there may be an entire swarm of strains within the 97% identity OTU *RhiAcro1* with different complementary functions. Future studies should thus characterize *RhiAcro1* diversity more fully and examine the relative abundances of the different and potentially alternative strains across *Acromyrmex* species, while also taking variation in environmental conditions (field colonies sampled in different seasons; long term lab colonies, etc) into account. Our present results indicate that the *nifH+ RhiAcro1* bacteria may not be as abundant as previously thought (Sapountzis et al. 2015) or that they did not grow on the medium used to obtain the culture required for genome sequencing of the *A. octospinosus* gut symbiont.

Similar to Rhizobiales bacteria in *Cephalotes* ants and '*Ca.* T. hoelldoblerii' in *Harpegnathos*, *RhiAcro1* appears to be capable of converting arginine to urea and L-ornithine using arginase genes. The fungal cultivar of leaf-cutting ants is known to produce arginine, making it logical that both *Atta* and *Acromyrmex* farming ants have lost genes in the arginine biosynthesis pathway (Nygaard et al. 2011, Suen et al. 2011) and that these ants also domesticated gut symbionts to recycle nitrogen-rich excess arginine when their ingested food has reached their hindgut. We recently showed that another hindgut symbiont of *A. echinatior* belonging to the Mollicutes (Entomoplasmatales) also decomposes arginine while producing NH_3 (Sapountzis et al. 2018). Decomposition of arginine allows its nitrogen to be recovered and made available for fungus garden growth rather than being wasted when deposited directly on gardens in a form that the fungal symbiont can hardly handle (Sapountzis et al. 2018). It was shown that the fungus garden of *Acromyrmex* can grow using both ammonia or urea as exclusive source of nitrogen, but with addition of ammonia the fungal development was faster (Abril and Bucher 2004). If the primary function of Rhizobiales symbionts is to decompose excess arginine, it would seem reasonable to expect not only spatial but also functional niche segregation between the Mollicutes and Rhizobiales symbionts. While both Mollicutes and Rhizobiales are located in the hindgut, the Rhizobiales attach to the hindgut cuticle, while Mollicutes swarm in the lumen and become partially

deposited on the fungus garden with the ants' fecal droplets (Sapountzis et al. 2015). In field colonies of *Atta* leaf-cutting ants, Rhizobiales and Mollicutes symbionts are possibly mutual exclusive (Zhukova et al. 2017, Sapountzis et al. 2019)—workers with high Rhizobiales titers harbour very low amount of Mollicutes and the other way around—both at the individual level and at the level of the colony. In sympatric Panamanian *Acromyrmex*, the same mutual exclusiveness trend is visible, but less clear (Sapountzis et al. 2015). At the same time, Mollicutes symbionts can be 20–200 times more abundant than other bacterial symbiont in both *Acromyrmex* and *Atta* (Zhukova et al. 2017) similar to *Spiroplasma* in turtle ants (Kautz et al. 2013). This suggests that Mollicutes are the primary arginine-processing symbionts, so the question is why it is often but far from always beneficial to have a secondary Rhizobiales symbiont in the hindgut compartment to break down arginine into urea instead of NH_3. We hypothesize that a backup Rhizobiales symbiont was adaptive for the ants to domesticate because accumulation of NH_3 in the hindgut and fecal fluid can be toxic (O'Donnell and Donini 2017) and will increase pH to levels well in excess of the optimal pH of ca. 3–5 in rectum and around 5 in the fungus garden. These garden pH values are necessary for proper functioning of ferments such as xyloglucanase (Xeg1) from the fungal gondylidia ingested by the ants (Erthal et al. 2004, Semenova et al. 2011, Kooij et al. 2016). Urea is a water soluble compound that has nearly neutral pH in solutions, so that many organisms convert NH_3 to urea in spite of this being energetically costly (O'Donnell and Donini 2017). However, the main function of the ant hindgut is absorption, so the conversion of ammonia to urea in this compartment can only be performed by symbiotic bacteria (Holtof et al. 2019). The production of urea from arginine by the *RhiAcro1* symbiont is consistent with the *Acromyrmex*-associated Rhizobiales having lost the urease genes. Domestication of two complementary gut symbionts, whose titers can possibly be regulated by the host ants, would thus significantly increase the robustness of the entire fungus-farming symbiosis when facing variable leaf and flower resources over time. Further comparative studies are needed to test this hypothesis and to establish whether it is a unique symbiotic adaptation restricted to the evolutionarily derived leaf-cutting ants, though coexistence of the same Mollicutes and Rhizobiales OTUs were found in field colonies of *Trachymyrmex cornetzi* (Sapountzis et al. 2019).

Potential metabolic functions of *RhiAcro1* symbionts

Provisioning of essential amino acids and vitamins often explains the adaptive significance of insect bacterial symbionts and such specialized functions are often accompanied by significant losses of genes and pathways that were previously essential for a free-living existence (Dale and Moran 2006). It is interesting therefore that the *RhiAcro1* representatives that we sequenced have only retained four complete pathways for producing the non-essential amino acids Asp, Gln, Cys, and Gly (Table S7). For synthesizing the essential amino acid lysine, the *RhiAcro1* genomes lack only one step out of nine. These findings suggest that *RhiAcro1* is nutritionally completely dependent on the *Acromyrmex* host ants, in striking contrast with '*Ca*. T. hoelldoblerii' (associated with predatory ponerine ants) and *B. apis* (associated with corbiculate bees) that retained complete pathways for the biosynthesis of almost all amino acids (Table S7). We hypothesize that this complete dependence on host biosynthetic pathways could evolve because the exclusive fungal diet of leaf-cutting ants is more predictable than the diets of predatory ants and bees.

We also scanned the three *RhiAcro1* genomes that we obtained for the presence of vitamin biosynthesis pathways. This revealed a full set of genes (*ribBA*, *ribD*, *ribH*, *ribE*, *ribF*) to synthesise riboflavin (vitamin B_2) (Fig. 3; Table S8), which was the only vitamin biosynthetic pathway that was complete (BLASTKoala). Determining enzymes involved in the dephosphorylation of 5-amino-6-(5-phospho-D-ribitylamino)uracil to form 5-amino-6-(ribitylamino)uracil (enzyme 3 on Fig. S3) remains problematic in bacteria, although this conversion can be performed by multiple broad spectrum HAD-type hydrolases (Haase et al. 2013, London et al. 2015). Interestingly, we found a gene annotated as a HAD-type hydrolase next to *ribF*. The *ribD*, *ribE*, and *ribH* cluster was flanked by a ribonucleotide reductase regulator *nrdR* and an antitermination factor *nusB*, similar to what related pathogenic *Brucella abortus* and other bacteria have (García-Angulo 2017). *Acromyrmex* ants lack a biosynthetic pathway for riboflavin biosynthesis (Nygaard et al. 2011), so the *RhiAcro1* symbiont can supply the ant host with vitamin B_2. However, the fungus garden likely remains the primary source of riboflavin for the ants, because all fungi are capable of synthesizing riboflavin (Abbas and Sibirny 2011).

The three *RhiAcro1* genomes also had a complete pathway for queuosine (Q) biosynthesis, a modified nucleoside in the anticodon loop of tRNA that is ubiquitously found in many organisms (Fergus et al. 2015). Eukaryotes, including insects and fungi, are unable to synthesize Q *de novo* and rely on their diet and/or microbial symbionts to obtain the Q precursor (Zallot et al. 2014). It is interesting, therefore, that bacteria of the Brucellaceae family and '*Ca*. T. hoelldoblerii' share this full pathway for Q biosynthesis with the *RhiAcro1* species, while it was partially lost in all evolutionarily more derived *Bartonella* species, including *B. apis* (Table S8), a non-core gut symbiont of honeybees (Raymann and Moran 2018). Mollicutes, one of the dominant symbiont of *Acromyrmex*, lack genes for Q biosynthesis (Sapountzis et al. 2018).

When *Acromyrmex* workers ingest fungal cultivar material, they obtain a diversity of oxidoreductases that allows them to produce hydrogen peroxide that becomes operational when *Acromyrmex* fecal fluid is exposed to oxygen, providing the symbiosis with a powerful Fenton chemistry procedure for rapid decomposition of freshly chewed-up leaf-material (Schiøtt and Boomsma 2021). Spurious production of hydrogen peroxide in the ant gut may, however, pose a risk of inadvertent oxidation of gut tissue and beneficial microorganisms, and consequently the *Acromyrmex* genome appear to encode an increased number of glutathione S-transferase genes, known to be involved in detoxification of peroxidised lipids and other xenobiotics. Consistent with this, the *RhiAcro1* genomes encode extensive antioxidant machinery, including superoxide dismutase, catalase, glutathione reductase, glutathione S-transferase, and a complete glutathione biosynthesis pathway (Table S9), which likewise may serve to protect the bacteria themselves or their host against oxygen radicals produced prematurely in the gut by Fenton chemistry, or to protect against the oxidative stress induced by the ingestion of fresh plant food (Aucoin et al. 1995). In accordance, the '*Ca*. T. hoelldoblerii' symbiont harbored by predatory ants lost the catalase gene, an enzyme mediating the decomposition of hydrogen peroxide to water and oxygen, suggesting a correlation between antioxidants and a herbivorous lifestyle of the host, though it has other components of antioxidant machinery similar to *RhiAcro1* (Table S9). It is also noteworthy that in the Fenton reaction, hydrogen peroxide reacts with reduced iron to produce the very reactive hydroxyl radicals, which are able to degrade recalcitrant plant cell wall components. In this connection the riboflavin produced by *RhiAcro1* could play an important role, as riboflavin and iron availability seem to be strongly

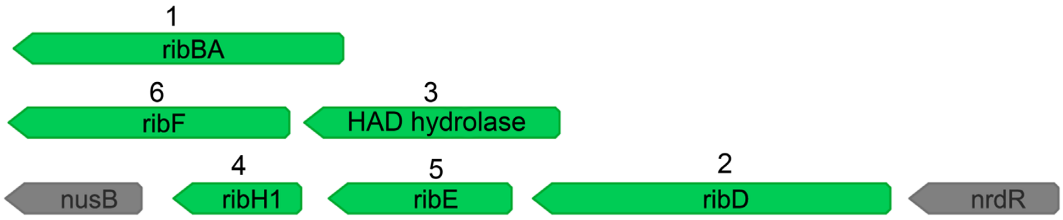

Figure 3. The full set of genes for riboflavin biosynthesis identified in *RhiAcro1*. Numbers above genes correspond to the step in the reaction pathway shown in Fig. S3 that is catalyzed by the enzyme.

interconnected (Cisternas, Salazar and García-Angulo 2018). Riboflavin appears to function both as an iron reduction catalyst and as a substitute for iron in many chemical reactions. Therefore, *RhiAcro1* might facilitate iron reduction with participation of flavins, ensuring a supply of reducing agents to allow the Fenton reaction to run, though all mechanisms and participants of iron reduction reaction are not known (Cisternas, Salazar and García-Angulo 2018). This hypothesis would fit with the finding that larvae and pupae of *A. echinatior* seem to have very few of these bacteria (Zhukova et al. 2017), as they do not partake in plant substrate preparation activities. Further investigations of this subject will be needed in order to verify this function of the Rhizobiales bacteria.

Lack of motility and biofilm formation

In the three isolates from two ant species that we studied, *RhiAcro1* had lost the genes for assembly of flagella, which is interesting because a full set of genes for flagellum assembly is present in the '*Ca.* T. hoelldoblerii' sister species associated with the guts of *H. saltator* ants, although more genomes of other ant symbionts would be necessary to generalize this result. Flagellar genes can be lost rapidly due to selection following phage attack, because flagellar proteins can serve as receptors for bacteriophage adsorption (Bertozzi Silva, Storms and Sauvageau 2016). Free-living Rhizobiales with at least short motile stages always have a set of genes for assembling functional flagella and the pathogenic eubartonellae use their flagellum to invade erythrocytes although flagellum loss has been quite frequent in this group. However, such losses are normally correlated with the acquisition of the *trw* gene, which produces hair-like (pilus) appendage subunits that facilitate direct interaction with erythrocyte receptors (Engel and Dehio 2009) so that pathogenicity is unlikely to be compromised. Loss of flagella genes in the *Acromyrmex* symbionts thus seems consistent with phage pressure and the presence of substantial defense mechanisms (see next section).

In our previous study, we showed that Rhizobiales bacteria associated with the large workers of *Acromyrmex* form a biofilm in the hindgut which was difficult to fully remove even after 35 days of antibiotics treatment (Sapountzis et al. 2015). Trimeric Autotransporter Adhesins (TAA) proteins are known to help Gram-negative bacteria adhere to extracellular matrix proteins or endothelial cells of hosts and to participate in the formation of biofilms and are found in all *Bartonella* species (Okaro et al. 2017). Members of this adhesin protein family have a head-stalk-anchor protein architecture, where the head domain is responsible for surface attachment. We found one copy of a gene coding for a TAA-like protein in *RhiAcro1* from *A. echinatior* (*RAe6*) and two copies in *RAo1* from *A. octospinosus* (Fig. 4; TAA-like proteins are referred to as YadA based on the name of an adhesin from *Yersinia*, the first member of this family to be characterized (El Tahir and Skurnik 2001)). Analysis of conserved domains showed that all proteins had anchor (YadA-like-C-terminal) and stalk parts, but that a typical TAA head domain was only present in one protein of *RAo1*. The second protein in *RAo1* had a domain from another adhesin family (filamentous haemagglutinin FhaB). The TAA-like protein sequence in *RAe6* was shorter than the sequences of *RAo1* (1067 aa vs. 1499 and 1322 aa) and did not have a head domain recognized by the NCBI Conserved Domain search tool (Marchler-Bauer et al. 2017). The presence of one TTA-like protein without a head domain in *RAe6* of *A. echinatior* and of two proteins in *RAo1* of *A. octospinosus* might explain the difference in the abundances of *RhiAcro1* in the two sympatric ant hosts. *RhiAcro1* is normally present in low abundance in large workers of *A. echinatior* in the field (< 1% of the total bacterial abundance per host body), while this OTU reaches considerably higher abundances (1%–37% abundance) in *A. octospinosus* (Sapountzis et al. 2015). The study of biosynthetic pathways for biofilm formation is a poorly explored field for *Rhizobiales*, so we cannot exclude the possibility that there might be other genes in species belonging to the *RhiAcro1* OTU that mediate biofilm formation in *A. echinatior.* Our results underline that in depth genome-level studies of co-occurring symbiont strains within the same OTU may offer surprising insights in complementary mutualistic services that cannot be inferred from 16S data, as was recently confirmed for bacterial gut symbionts of bees (Ellegaard and Engel 2016, Ellegaard et al. 2020).

Diversity of innate and adaptive immune defenses in Rhizobiales symbionts

Bacterial restriction-modification (R-M) and CRISPR (clustered regularly interspaced short palindromic repeats)-Cas systems are functionally analogous to the innate and adaptive immune systems of eukaryotes because they jointly defend bacteria against destruction by phages (Abedon 2012). CRISPR-Cas systems can also prevent horizontal gene transfer by targeting plasmids and integrative and conjugating elements (Bikard et al. 2012, Wheatley and MacLean 2020). Spacers, which are derived from the genome of the invading viruses, are incorporated into the CRISPR cassette, so that their diversity offers sequence-specific and heritable immunity against a range of phages with homologous sequences (Barrangou and Marraffini 2014). We scanned all 57 Rhizobiales genomes (Fig. 1) for the presence of CRISPR-Cas systems, which showed that most species (46/57) lacked these defence systems altogether. However, CRISPR-Cas (type I-C) was present in *Hoeflea olei*, a free-living diesel-oil-degrading Rhizobiales, and all *B. apis* strains had a partial set of *cas* genes. The closest relative of the three *RhiAcro1* strains '*Ca.* T. hoelldoblerii' has *cas* types I-C and II-C co-located with CRISPR cassettes, which suggests that mutualistic diet supplementation with riboflavin and urease recylcing offers sufficiently important fitness advantages to this predatory ant host to bear the immune defense costs on behalf of its symbiont (Neuvonen et al. 2016). The three *RhiAcro1* strains of *Acromyrmex* leaf-cutting ants had CRISPR cassettes with

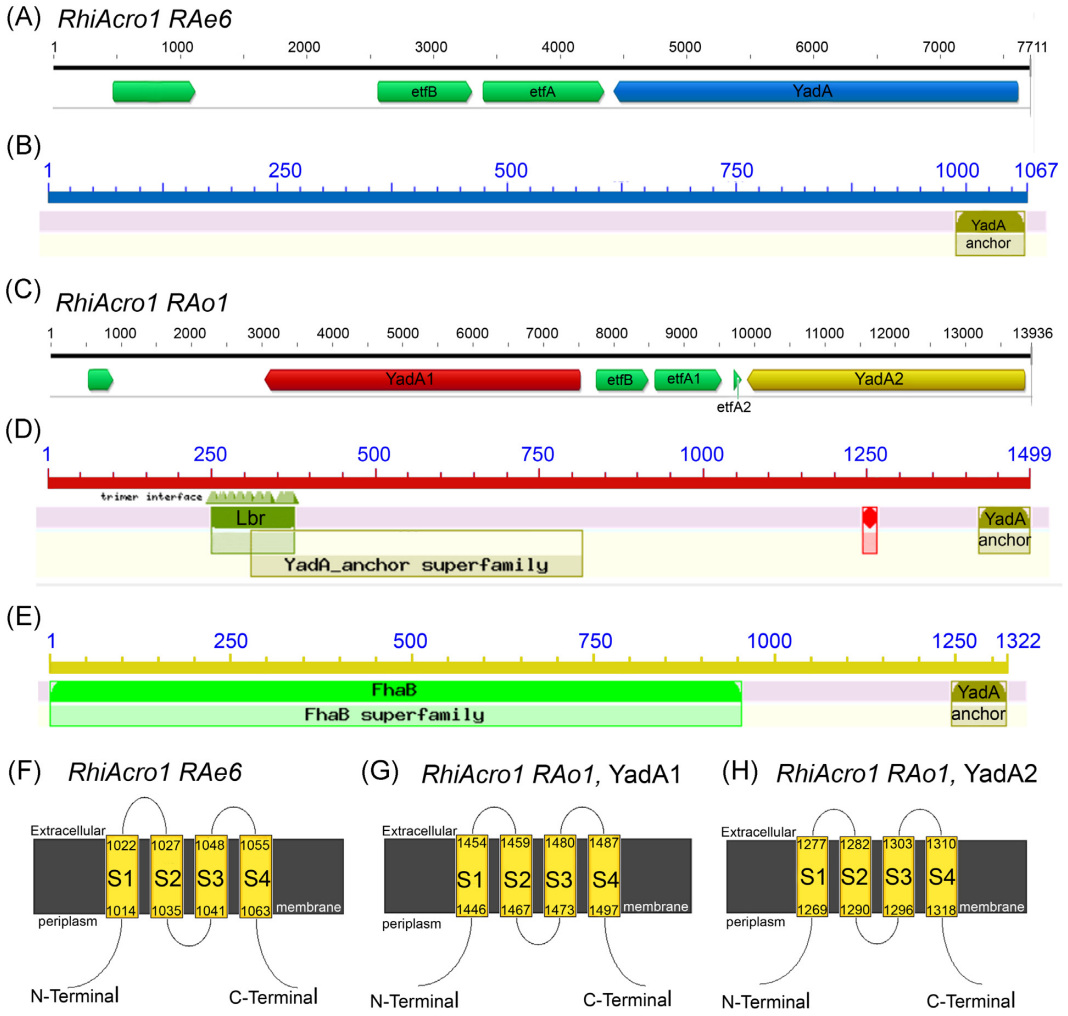

Figure 4. Trimeric Autotransporter Adhesin-like proteins encoded in the *RhiAcro1* genomes obtained from *A. echinatior* and *A. octospinosus*. **(A, C)** The positions of a single (*yadA*) and double (*yadA1* and *yadA2*) copy of the adhesin-like protein genes in the *A. echinatior* and *A. octospinosus* contigs obtained. The *etfA* and *etfB* sequences are the adjacent genes, with the former also occurring in two copies similar to *yadA*, consistent with a gene duplication having occurred in *RAo1*; **(B, D, E)** The conserved domains of TAA-like proteins in *RhiAcro1* based on the NCBI Conserved Domain search tool. Colors of rulers designate corresponding genes. The YadA-anchors represent the C-terminal of a family of surface-exposed bacterial proteins and FhaB is a large exoprotein involved in heme-utilization or adhesion. *Lbr*—YadA-like, a left-handed beta-roll, contains the collagen-binding domain and related cell surface proteins. The red square represents the coiled stalk of the trimeric autotransporter adhesin (YadA-stalk). **(F, G, H)** The C-terminal membrane anchor topology of the TAA proteins predicted with PRED-TMBB2.

an even higher number of spacers and possessed the most diverse *cas* genes of all Rhizobiales analysed, suggesting potential selection for immune protection against a range of bacteriophages and plasmids (Table 1). Our earlier study of Mollicutes symbionts of *Acromyrmex* found very similar extensive phage defences (Sapountzis et al. 2018), which suggest that phage defenses are generally necessary to maintain viable symbiosis between the few specialized bacterial gut symbionts and the *Acromyrmex* leaf-cutting ant hosts. Obligate bacterial symbionts generally lack CRISPR-Cas systems (Burstein et al. 2016) when they are intracellular and thus protected from phage attack by the eukaryote cell membranes surrounding them. This likely explains that intracellular mammalian *Bartonella* pathogens and the root-nodule Rhizobiales symbionts of legumes lack a CRISPR-Cas system (Table 1) while selection has maintained them in the specialized extracellular gut symbionts of *Acromyrmex* ants in spite of CRISPR-Cas systems potentially slowing down bacterial growth and imposing autoimmune fitness cost (Stern et al. 2010, Vale et al. 2015).

We also found a potentially interesting difference between the *RhiAcro1* strains from *A. octospinosus* (*RAo1*) and *A. echinatior* (*RAe6* and *RAe9*). The *RAo1* symbiont had *cas* types I-C, II-C, and IV-A co-located with CRISPR cassettes, while *RAe6* and *RAe9* only had the last two *cas* loci (type II-C and type IV-A) and positioned well away from CRISPR in their genomes. Remote *cas* loci and CRISPRs have been found in many bacterial species, and not all of them retained functional immune defences against phages (Zhang and Ye 2017). This difference in putative phage-defence efficiency could be related to the *RhiAcro1* OTU being always rare (<1% abundance) in field-sampled large workers of *A. echinatior* while it reaches abundances up to 37% of the total microbiome in *A. octospinosus* workers (Sapountzis et al. 2015, 2019).

Also the constitutive R-M system was different between the strains associated with the two sympatric *Acromyrmex* species in Panama (Table S10). Both *RAe6* and *RAe9* (isolated from two ants of the same colony) had proteins of a type II R-M system, but *RAe9* also had a complete set of genes for a type I R-M system, consisting of three polypeptides: R (restriction), M (modification), and

Table 1. The presence of CRISPR/Cas systems across the 57 Rhizobiales genomes analysed.

Bacterial species	CRISPR, number of spacers	*cas*	Cas type
Hoeflea olei	26	0	
	32	0	
	15	*cas3*, *cas3*, *cas5*, *cas8c*, *cas7*, *cas4*, *cas1*, *cas2*	Type I-C (complete)
Bartonella apis BBC0122	5	0	
	6	0	
	7	*cas1*, *cas9*	Type II-C (incomplete)
B. apis BBC0178	12	0	
	8	0	
	8	0	
	5	*cas1*, *cas9*	Type II-C (incomplete)
B. apis BBC0244	11	0	
	6	0	
	7	0	
	11	0	
	4	*cas1*, *cas9*	Type II-C (incomplete)
B. apis PEB0122	21	*cas1*, *cas9*	Type II-C (incomplete)
	7	0	
	4	0	
	20	0	
B. apis PEB0149	23	0	
	2	*cas1*, *cas9*	Type II-C (incomplete)
B. apis PEB0150	3	*cas9*	Type II-C (incomplete)
	6	0	
	6	0	
	9	0	
	10	0	
Tokpelaia hoelldoblerii	33	*cas3*, *cas5*, *cas8c*, *cas7*, *cas4*, *cas1*, *cas2*	Type I-C (complete)
	9	*cas1*, *cas2*, *cas9*	Type II-C (complete)
RhiAcro1 RAo1	29	0	
	63	*cas1*, *cas2*, *cas9*	Type II-C (complete)
	147	*cas3*, *cas5*, *cas8c*, *cas7*, *cas4*, *cas1*, *cas2*	Type I-C (complete)
	6	*cas13a*	Type IV-A (complete)
RhiAcro1 RAe6	37	0	
	43	0	
	17	0	
	0	*cas1*, *cas2*, *cas9*	Type II-C (complete)
	0	*cas13a*	Type IV-A (complete)
RhiAcro1 RAe9	37	*cas3*	Type I-C (incomplete)
	2	*cas13a*	Type IV-A (complete)
	81	0	
	19	0	
	0	*cas9*, *cas1*, *cas2*	Type II-C (complete)
	0	*cas13a*	Type IV-A (complete)

Most Rhizobiales species lacked CRISPR/Cas systems entirely so only the positive occurrences are given here.

S (specificity), whereas *RAe6* lacked the S gene. *RAo1* associated with *A. octospinosus* had a complete type II R-M system, consisting of several R and M proteins. Similar to *RAo1*, '*Ca.* T. hoelldoblerii' has genes for a functional type II R-M system, although their number is much lower than in the three *RhiAcro1* symbionts (2 in '*Ca.* T. hoelldoblerii' vs. 13, 16 and 15 in *RAe6*, *RAe9* and *RAo1*, respectively). The co-occurrence of different R-M systems in the same bacterial species is not uncommon because R-M systems behave as mobile elements and can be acquired through horizontal gene transfer (Kobayashi 2001). Overall, only a few of the 57 Rhizobiales genomes that we inspected lacked complete R-M systems (Table S10). Among these exceptions were the plant symbiont *Rhizobium leguminosarum* and several eubartonellae pathogens.

In general, the functional significance of most facultative and extracellular symbionts of insects are unknown (Nayfach et al. 2020), but this does not preclude that their roles are important and that functions can be inferred with proper omics approaches. Our present study infers several such putative functions for species of the *RhiAcro1* OTU that at first seemed to be a nitrogen-fixing symbiont, but now appears to have a more variable and broader set of functions. Many of these can be hypothesized as likely adaptive for the fungus-farming symbiosis as a whole, but only because a decade of research has clarified complementary functions of other bacterial OTUs and of the specialized fungal crop symbiont (Aylward et al. 2012, De Fine Licht, Boomsma and Tunlid 2014, Sapountzis et al. 2018, Schiøtt and Boomsma 2021). As it appears, communities of extracellular symbionts in insect guts are comparable to ecosystems on a leash (Foster et al. 2017), whose diversity and functional complementarity we are just beginning to understand. We suggest that the kinds of functionality will determine in large part how complex and condition-dependent metabolic functions of extracellular gut symbionts will be.

Supplementary data

Supplementary data are available at FEMSEC online.

Acknowledgements

We thank Sylvia Mathiasen for lab assistance. The Smithsonian Tropical Research Institute in Panama provided lab facilities during fieldwork, and the Autoridad Nacional del Ambiente of Panama issued collection and export permits.

Conflict of interest statement. None declared.

Funding

This work was supported by funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie Actions [no. 660255 to MZ, 300584 to PS], and by the European Research Council [Advanced Grant 323085 to JB] and the Danish National Research Foundation [DNRF57 to JB].

References

Abbas CA, Sibirny AA. Genetic control of biosynthesis and transport of riboflavin and flavin nucleotides and construction of robust biotechnological producers. *Microbiol Mol Biol Rev* 2011;**75**:321.

Abedon ST. Bacterial "immunity" against bacteriophages. *Bacteriophage* 2012;**2**:50–4.

Abril AB, Bucher EH. Nutritional sources of the fungus cultured by leaf-cutting ants. *Appl Soil Ecol* 2004;**26**:243–7.

Andersen SB, Boye M, Nash DR, *et al.* Dynamic wolbachia prevalence in Acromyrmex leaf-cutting ants: potential for a nutritional symbiosis. *J Evol Biol* 2012;**25**:1340–50.

Andrews S. FastQC: a quality control tool for high throughput sequence data [Online]. 2010. https://qubeshub.org/resources/fastqc.

Aucoin R, Guillet G, Murray C *et al.* How do insect herbivores cope with the extreme oxidative stress of phototoxic host plants? *Arch Insect Biochem Physiol* 1995;**29**:211–26.

Aylward FO, Burnum KE, Scott JJ *et al.* Metagenomic and metaproteomic insights into bacterial communities in leaf-cutter ant fungus gardens. *ISME J* 2012;**6**:1688–701.

Aziz RK, Bartels D, Best A *et al.* The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 2008;**9**:75, DOI: 10.1186/1471-2164-9-75.

Bankevich A, Nurk S, Antipov D *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;**19**:455–77.

Barrangou R, Marraffini LA. CRISPR-cas systems: prokaryotes upgrade to adaptive immunity. *Mol Cell* 2014;**54**:234–44.

Bertozzi Silva J, Storms Z, Sauvageau D. Host receptors for bacteriophage adsorption. *FEMS Microbiol Lett* 2016;**363**:2.

Bikard D, Hatoum-Aslan A, Mucida D *et al.* CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell Host Microbe* 2012;**12**:177–86.

Bisch G, Neuvonen MM, Pierce NE *et al.* Genome evolution of bartonellaceae symbionts of ants at the opposite ends of the trophic scale. *Genome Biol Evol* 2018;**10**:1687–704.

Bordenstein SR, Theis KR. Host biology in light of the microbiome: ten principles of holobionts and hologenomes. *PLOS Biol* 2015;**13**:e1002226.

Borowiec ML. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 2016;**2016**. DOI: 10.7717/peerj.1660.

Bruen TC, Philippe H, Bryant D. A simple and robust statistical test for detecting the presence of recombination. *Genetics* 2006;**172**:2665–81.

Burstein D, Sun CL, Brown CT *et al.* Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nat Commun* 2016;**7**:1–8.

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009;**25**:1972–3.

Caporaso JG, Lauber CL, Walters WA *et al.* Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci USA* 2011;**108**:4516–22.

Cisternas IS, Salazar JC, García-Angulo VA. Overview on the bacterial iron-riboflavin metabolic axis. *Front Microbiol* 2018;**9**:1478.

Couvin D, Bernheim A, Toffano-Nioche C *et al.* CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for cas proteins. *Nucleic Acids Res* 2018;**46**:W246–51.

Csűrös M. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* 2010;**26**::1910–2.

Dagan T, Blekhman R, Graur D. The "Domino theory" of gene death: gradual and mass gene extinction events in three lineages of obligate symbiotic bacterial pathogens. *Mol Biol Evol* 2006;**23**:310–6.

Dale C, Moran NA. Molecular interactions between bacterial symbionts and their hosts. *Cell* 2006;**126**:453–65.

Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 2010;**5**:e11147.

Davidson DW, Cook SC, Snelling RR *et al.* Explaining the abundance of ants in lowland tropical rainforest canopies. *Science (80-)* 2003;**300**:969–72.

De Fine Licht HH, Boomsma JJ, Tunlid A. Symbiotic adaptations in the fungal cultivar of leaf-cutting ants. *Nat Commun* 2014;**5**. DOI: 10.1038/ncomms6675.

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;**32**:1792–7.

El Tahir Y, Skurnik M. YadA, the multifaceted Yersinia adhesin. *Int J Med Microbiol* 2001;**291**:209–18.

Ellegaard KM, Engel P. Beyond 16S rRNA community profiling: intra-species diversity in the gut microbiota. *Front Microbiol* 2016;**7**:1475.

Ellegaard KM, Suenami S, Miyazaki R *et al.* Vast differences in strain-level diversity in the gut microbiota of two closely related honey bee species. *Curr Biol* 2020;**30**:2520–31.e7.

Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* 2015;**16**:157.

Engel P, Dehio C. Genomics of host-restricted pathogens of the genus bartonella. *Genome Dyn* 2009;**6**:158–69.

Erlacher A, Cernava T, Cardinale M *et al.* Rhizobiales as functional and endosymbiontic members in the lichen symbiosis of *Lobaria pulmonaria* L. *Front Microbiol* 2015;**6**:53.

Erthal M, Peres Silva C, Ian Samuels R. Digestive enzymes of leaf-cutting ants, *Acromyrmex subterraneus* (Hymenoptera: formicidae: attini): distribution in the gut of adult workers and partial characterization. *J Insect Physiol* 2004;**50**:881–91.

Fergus C, Barnes D, Alqasem MA *et al.* The queuine micronutrient: charting a course from microbe to man. *Nutrients* 2015;**7**:2897–929.

Fischer H-M. Genetic regulation of nitrogen fixation in rhizobia. *Microbiol Rev* 1994;**58**:352.

Foster KR, Schluter J, Coyte KZ *et al.* The evolution of the host microbiome as an ecosystem on a leash. *Nature* 2017;**548**:43–51.

García-Angulo VA. Overlapping riboflavin supply pathways in bacteria. *Crit Rev Microbiol* 2017;**43**:196–209.
Garrido-Oter R, Nakano RT, Dombrowski N *et al.* Modular traits of the rhizobiales root microbiota and their evolutionary relationship with symbiotic rhizobia. *Cell Host Microbe* 2018;**24**:155–67.e5.
Haase I, Sarge S, Illarionov B *et al.* Enzymes from the haloacid dehalogenase (HAD) superfamily catalyse the elusive dephosphorylation step of riboflavin biosynthesis. *ChemBioChem* 2013;**14**:2272–5.
Halachev MR, Loman NJ, Pallen MJ. Calculating orthologs in bacteria and archaea: a divide and conquer approach. *PLoS One* 2011;**6**, DOI: 10.1371/JOURNAL.PONE.0028388.
Hölldobler B, Wilson EO. Chapter 11. The organization of species communities. In: *The Ants*. Cambridge: Harvard University Press, 1990.
Holtof M, Lenaerts C, Cullen D *et al.* Extracellular nutrient digestion and absorption in the insect gut. *Cell Tissue Res* 2019;**377**:397–414.
Hu Y, Sanders JG, Łukasik P *et al.* Herbivorous turtle ants obtain essential nutrients from a conserved nitrogen-recycling gut microbiome. *Nat Commun* 2018;**9**, DOI: 10.1038/s41467-018-03357-y.
Jiang H, Lei R, Ding S-W *et al.* Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinf* 2014;**15**:182.
Kanehisa M, Sato Y, Morishima K. BlastKOALA and ghostkoala: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 2016;**428**:726–31.
Kautz S, Rubin BER, Russell JA *et al.* Surveying the microbiome of ants: comparing 454 pyrosequencing with traditional methods to uncover bacterial diversity. *Appl Environ Microbiol* 2013;**79**:525–34.
Kobayashi I. Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res* 2001;**29**:3742–56.
Kooij PW, Pullens JWM, Boomsma JJ *et al.* Ant mediated redistribution of a xyloglucanase enzyme in fungus gardens of *Acromyrmex echinatior*. *BMC Microbiol* 2016;**16**:1–9.
Kriventseva E V, Kuznetsov D, Tegenfeldt F *et al.* OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res* 2019;**47**:D807–11.
Lanfear R, Frandsen PB, Wright AM *et al.* PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol Biol Evol* 2016;**34**:msw260.
Liberti J, Sapountzis P, Hansen LH *et al.* Bacterial symbiont sharing in megalomyrmex social parasites and their fungus-growing ant hosts. *Mol Ecol* 2015;**24**:3151–69.
Lin H-H, Liao Y-C. Accurate binning of metagenomic contigs via automated clustering sequences using information of genomic signatures and marker genes. *Sci Rep* 2016;**6**:24175.
London N, Farelli JD, Brown SD *et al.* Covalent docking predicts substrates for haloalkanoate dehalogenase superfamily phosphatases. *Biochemistry* 2015;**54**:528–37.
Maddison WP, Maddison DR. Mesquite: a modular system for evolutionary analysis. Version 3.4. http://wwwMesquiteprojectOrg 2019.
Manni M, Berkeley MR, Seppey M *et al.* BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol* 2021;**38**:4647–54.
Marchler-Bauer A, Bo Y, Han L *et al.* CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* 2017;**45**:D200–3.
McCutcheon JP, Moran NA. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* 2012;**10**:13–26.
Mueller UG, Rehner SA, Schultz TR. The evolution of agriculture in ants. *Science (80-)* 1998;**281**:2034–8.
Mueller UG, Schultz TR, Currie CR *et al.* The origin of the attine ant-fungus mutualism. *Q Rev Biol* 2001;**76**:169–98.
Nayfach S, Roux S, Seshadri R *et al.* A genomic catalog of earth's microbiomes. *Nat Biotechnol* 2020. DOI: 10.1038/s41587-020-0718-6.
Neuvonen MM, Tamarit D, Näslund K *et al.* The genome of rhizobiales bacteria in predatory ants reveals urease gene functions but no genes for nitrogen fixation. *Sci Rep* 2016;**6**:1–11.
Nygaard S, Zhang G, Schiøtt M *et al.* The genome of the leaf-cutting ant Acromyrmex echinatior suggests key adaptations to advanced social life and fungus farming. *Genome Res* 2011;**21**:1339–48.
O'Donnell MJ, Donini A. Nitrogen excretion and metabolism in insects. *Acid-Base Balanc Nitrogen Excretion Invertebr* 2017:109–26.
Okaro U, Addisu A, Casanas B *et al.* Bartonella species, an emerging cause of blood-culture-negative endocarditis. *Clin Microbiol Rev* 2017;**30**:709–46.
Parks DH, Imelfort M, Skennerton CT *et al.* CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015;**25**:1043–55.
Raymann K, Moran NA. The role of the gut microbiome in health and disease of adult honey bee workers. *Curr Opin Insect Sci* 2018;**26**:97–104.
Roberts RJ, Vincze T, Posfai J *et al.* REBASE-a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res* 2015;**43**:D298–9.
Ronquist F, Huelsenbeck JP. MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics* 2003;**19**:1572–4.
Russell JA, Moreau CS, Goldman-Huertas B *et al.* Bacterial gut symbionts are tightly linked with the evolution of herbivory in ants. *Proc Natl Acad Sci USA* 2009;**106**:21236–41.
Sapountzis P, Nash DR, Schiøtt M *et al.* The evolution of abdominal microbiomes in fungus-growing ants. *Mol Ecol* 2019;**28**:879–99.
Sapountzis P, Zhukova M, Hansen LH *et al.* Acromyrmex leaf-cutting ants have simple gut microbiota with nitrogen-fixing potential. *Appl Environ Microbiol* 2015;**81**:AEM.00961–00915.
Sapountzis P, Zhukova M, Shik JZ *et al.* Reconstructing the functions of endosymbiotic mollicutes in fungus-growing ants. *Elife* 2018;**7**, DOI: 10.7554/eLife.39209.
Schiøtt M, Boomsma JJ. Proteomics reveals synergy between biomass degrading enzymes and inorganic fenton chemistry in leaf-cutting ant colonies. *Elife* 2021;**10**:1–39.
Segers FH, Kešnerová L, Kosoy M *et al.* Genomic changes associated with the evolutionary transition of an insect gut symbiont into a blood-borne pathogen. *ISME J* 2017;**11**:1232–44.
Semenova TA, Hughes DP, Boomsma JJ *et al.* Evolutionary patterns of proteinase activity in attine ant fungus gardens. *BMC Microbiol* 2011;**11**:15.
Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;**30**:1312–3.
Stern A, Keren L, Wurtzel O *et al.* Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet* 2010;**26**:335–40.
Suen G, Teiling C, Li L *et al.* The genome sequence of the leaf-cutter ant Atta cephalotes reveals insights into its obligate symbiotic lifestyle. *PLoS Genet* 2011;**7**, DOI: 10.1371/journal.pgen.1002007
Tsirigos KD, Elofsson A, Bagos PG. PRED-TMBB2: improved topology prediction and detection of beta-barrel outer membrane proteins. *Bioinformatics* 2016;**32**:i665–71.
Vale PF, Lafforgue G, Gatchitch F *et al.* Costs of CRISPR-Cas-mediated resistance in *Streptococcus thermophilus*. *Proc R Soc B Biol Sci* 2015;**282**, DOI: 10.1098/rspb.2015.1270.

Wheatley RM, MacLean RC. CRISPR-Cas systems restrict horizontal gene transfer in *Pseudomonas aeruginosa*. *ISME J* 2020 155 2020;**15**:1420–33.

Xia X. DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Mol Biol Evol* 2013;**30**:1720–8.

Yarza P, Yilmaz P, Pruesse E *et al.* Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol* 2014;**12**:635–45.

Zallot R, Brochier-Armanet C, Gaston KW *et al.* Plant, animal, and fungal micronutrient queuosine is salvaged by members of the DUF2419 protein family. *ACS Chem Biol* 2014;**9**: 1812–25.

Zerbino DR, Achuthan P, Akanni W *et al.* Ensembl 2018. *Nucleic Acids Res* 2018;**46**:D754–61.

Zhang Q, Ye Y. Not all predicted CRISPR–Cas systems are equal: isolated cas genes and classes of CRISPR like elements. *BMC Bioinf* 2017;**18**:92.

Zhu Q, Kosoy M, Olival KJ *et al.* Horizontal transfers and gene losses in the phospholipid pathway of Bartonella reveal clues about early ecological niches. *Genome Biol Evol* 2014;**6**:2156–69.

Zhukova M, Sapountzis P, Schiøtt M *et al.* Diversity and transmission of gut bacteria in Atta and Acromyrmex leafcutting ants during development. *Front Microbiol* 2017;**8**. DOI: 10.3389/fmicb.2017.01942.