ISME Communications, 2026, 6(1), ycag079
https://doi.org/10.1093/ismeco/ycag079
Published: 27 March 2026
Original Article

Facultative mutualism between *Paramecium* and the intracellular *Rickettsiales* bacterium *Megaera* mediated by a horizontally acquired biotin operon

Michele Giovannini[1,2], Leandro Gammuto[3], Tania Alonso-Vásquez[2], Natalia Gogoleva[4], Greta Bellinzona[3], Alexey Potekhin[4], Giulio Petroni[1,*], Michele Castelli[3,5,*]

[1]Department of Biology, University of Pisa, Via Volta 4/6, 56100, Pisa, PI, Italy
[2]Department of Biology, University of Florence, Via Madonna del Piano, 6, 50019 Sesto Fiorentino, Florence, Italy
[3]Department of Biology and Biotechnology, University of Pavia, Via Ferrata 9, 27100 Pavia, PV, Italy
[4]Research Department for Limnology, University of Innsbruck, Mondseestrasse 9A-5310 Mondsee, Austria
[5]Present address: Department of Biology, University of Pisa, Via Volta 4/6, 56100, Pisa, PI, Italy
*Corresponding authors. Michele Castelli, Department of Biology, University of Pisa, Via Volta 4/6, Pisa, PI 56100, Italy. E-mail: michele.castelli@unipi.it; Giulio Petroni, Department of Biology, University of Pisa, Via Volta 4/6, Pisa, PI 56100, Italy. E-mail: giulio.petroni@unipi.it

Abstract

The *Rickettsiales* are an alphaproteobacterial lineage engaging in ancient associations with a variety of eukaryotic hosts and with a wide spectrum of effects. They include vector-borne pathogens, as well as *Wolbachia*, which ranges from a reproductive manipulator to a mutualist in arthropods and nematodes. The majority of *Rickettsiales* are associated with aquatic protists, but these interactions are poorly understood. Here, we explored by dual RNA-Seq the effect of the host-generalist *Rickettsiales* bacterium *Megaera polyxenophila* on the protist *Paramecium primaurelia*. *Megaera* induces substantial changes in host gene expression, in particular, increased expression levels of certain cell replication–related functions, consistent with the higher growth observed in previous experiments. Conversely, the co-occurring reduction of catabolism and energy metabolism can be explained by the capability of the bacterium to efficiently exert the same pathways also for the host's benefit. Therefore, *Megaera* likely behaves as a facultative mutualist, consistent with its predicted ability to provide the host with Adenosine triphosphate (ATP) and biotin, the latter synthesized, thanks to a recently horizontally acquired operon. At the same time, this bacterium expresses several genes involved in host cell invasion and possibly toxicity. Accordingly, it is envisioned that the overall effect of *Megaera* on its host is rather plastic, being the fine-tuned sum of supportive and parasitic actions, likely resulting in flexibility according to host and symbiont genotypes and environmental conditions, and subjected to evolutionary changes. Such flexibility may also explain the broad host range of *Megaera* and, from a more general perspective, hints for shared traits and analogies among other protist-associated *Rickettsiales*.

Keywords dual RNA-Seq, *Rickettsiales*, ciliate protists, biotin, nutritional mutualism, bacterial endosymbiont, phylogeny, comparative genomics transcriptomics, bacteria–host interactions, *Wolbachia*

Introduction

The *Rickettsiales* are a deep alphaproteobacterial lineage with an ancient association with eukaryotes [1, 2]. Their best studied representatives are arthropod-borne pathogens [3–6] and *Wolbachia*, hosted by arthropods and nematodes [7–10]. On the other hand, the still under-explored majority of known *Rickettsiales* are associated with protists, including amoebas, ciliates, algae, and other aquatic organisms, such as cnidarians [11–20]. Investigations on those aquatic members have been extremely useful for enlightening evolutionary trends among the *Rickettsiales*, including the emergence of traits typical of the pathogenic representatives [1, 2, 6, 21–23]. However, with only a few partial exceptions [13, 24, 25], these investigations typically consisted of a combination of microscopy observations and comparative genomics, without direct experimental data on how these bacteria interact with their hosts and on the reciprocal effects.

Here, we focused, as a model organism, on the *Rickettsiales* bacterium '*Candidatus* Megaera polyxenophila' (formerly '*Candidatus* Megaira polyxenophila' [26], from here on, *Megaera polyxenophila*), a widespread *Rickettsiaceae* bacterium, particularly in freshwater [27]. This species is closely related to the well-studied pathogen *Rickettsia*, and is found in diverse hosts such as ciliates, streptophytes, chlorophytes, ochrophytes, and cercozoans [15, 20, 27–32]. The host range of the *Megaera* genus is even broader, encompassing also hydras, corals, ascidians, and haplosporidians [33–36].

Received: 9 August 2025. **Revised:** 19 February 2026. **Accepted:** 25 March 2026
© The Author(s) 2026. Published by Oxford University Press on behalf of the International Society for Microbial Ecology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

While the interaction modes and mechanisms between *Megaera* and its hosts are poorly understood, the broad host range and the seemingly harmless long-term laboratory propagation of colonized hosts are suggestive of a mutualism/commensalism or a very specialized parasitism [37]. Exploratory experiments showed that, in laboratory conditions, symbiotic *Paramecium* ciliates reach higher cell densities than aposymbiotic lines, which can be indicative of some mutualistic action [25]. Genome data overall indicate a possible parasitic role, due to the massive scavenging of metabolic precursors, including intermediates of Krebs cycle [2], similarly to many other *Rickettsiales* [38]. On the other hand, some genome-encoded traits, particularly polyketide biosynthesis, could indicate a defensive symbiotic contribution [30].

Here, we performed a dual RNA-Seq (i.e. simultaneous bacterial and host transcriptome sequencing [39]) investigation on *Megaera* hosted in the cytoplasm of the freshwater ciliate *Paramecium primaurelia*, comparing symbiotic and aposymbiotic host lines.

Transcriptomics has been valuable in investigating bacterial–host interactions involving human pathogenic *Rickettsiales* [40, 41], as well as a few non-pathogenic others, including aquatic ones [13, 29, 42], and has also been employed on some occasions on symbionts of protists [43–47], including ciliates [48–50]. However, to our best knowledge, this is the first study to apply dual RNA-Seq to assess the impact of a *Rickettsiales* bacterium on a protist host.

Materials and methods

Strains employed and laboratory maintenance

In this work, we used previously established laboratory cultures of *P. primaurelia* Rio Lg_Jac 2III naturally harbouring *M. polyxenophila* (from here on, LgJac_Meg+), and a subclone of the same strain that had been treated with antibiotics to remove the bacterium (from here on, LgJac_Meg−) [25]. Strains were propagated at 19°C in the laboratory by regular feeding with Cerophyll medium inoculated with *Raoultella planticola* as previously described [25]. The presence/absence status of the symbiont was preliminarily confirmed by fluorescence *in situ* hybridization using the protocol by [51], as described previously [27].

Sequencing and assembly of the *Megaera* genome

A whole-genome amplification (WGA) was performed on ∼10 cells of *P. primaurelia* LgJac_Meg + as previously described [21]. The WGA product was processed by a TruSeq PCR-free library and sequenced with Illumina NovaSeq X by Macrogen Europe (Amsterdam, the Netherlands), generating 244 840 594 2 × 150 bp reads, as well as with Oxford Nanopore Technology (ONT) by Eurofins Genomics (Ebersberg, Germany) on a GridION machine, generating 755 646 921 bp output (219 911 reads, N50 = 4606 bp). Raw Illumina reads were then trimmed with Trimmomatic V0.39 [52], removing adapters and low-quality bases at both ends. The complete set of trimmed Illumina reads was assembled using SPAdes v3.6 [53]. The blobology pipeline [54] was employed to classify contigs based on GC content, sequencing coverage, and homology (Fig. S1), in order to help distinguish those belonging to *Megaera* from those of *Paramecium* and other bacterial species. Specifically, we selected the reads mapped with bowtie2 [55] on the contigs with coverage higher than 800, apart from those with the best megablast hit on *Paramecium* spp. These were separately reassembled with SPAdes, and the output was carefully manually curated to ensure an accurate assembly, as previously described (e.g. [21, 56]). Then, ONT reads were filtered with Filtlong v0.3.1 (https://github.com/rrwick/Filtlong) in two sequential steps, the first one without external reference and with –min_length 1000 –keep_percent 50 options, the second one using the Illumina assembly as reference, and the option –keep_percent 20. Then, a hybrid assembly was obtained from selected Illumina and ONT reads with Unicycler v0.5.0 [57] with the bold mode, followed by further manual refinement [2]. The resulting genome was annotated with Prokka v1.13.7 [58] with the –rfam option. The assembly quality and completeness were verified using BUSCO v6.0.0 [59] and the rickettsiaceae_odb12 reference dataset.

Pangenome analysis of *M. polyxenophila*

Twenty-one genome assemblies assigned to *M. polyxenophila* were selected for the analysis, including LgJac and others downloaded from NCBI. All genomes were annotated using Prokka, and their assembly quality was assessed using BUSCO as described above, applying cutoffs (>80% complete and single-copy and <2% duplicated reference orthologs) to obtain a final selection of 13 genomes (Table S1). Those were used as input for a pangenomic analysis with PPanGGOLiN [60], utilizing the 'all' subcommand. PPanGGOLiN identifies gene families and classifies them into three partitions according to their presence/absence in the dataset, namely, from the most to the least common, persistent, shell, and cloud. To refine this classification, the persistent partition was further subdivided into core (present in all the genomes), and soft core (not present in all the genomes) subpartitions. The resulting four classes (core, soft core, shell, cloud) were employed for the transcriptome enrichment analysis (see below).

Sample processing and dual transcriptome sequencing

The LgJac_Meg + and LgJac_Meg− *Paramecium* cultures were regularly fed (every second day) for three weeks with a proportional volume (0.25) of bacterized Cerophyll medium, then left unfed for 1 week. Then, they were cleaned and concentrated, adapting a previously established protocol [61]. For each culture, ∼300 ml were sieved through a nylon filter (pore size 100 μm), to remove the bigger clumps of debris and food bacteria (Fig. S2). Subsequently, to obtain highly concentrated and clean cultures, three centrifugation steps (at 120 g for 10′) were performed with a Rotofix 46 (Hettich) centrifuge. After each step, the supernatant, containing small debris, was discarded, and the pellet, containing the paramecia, was re-suspended in sterile medium. At the end of the last step, each pellet was resuspended in a small residual amount of medium (∼300 μl) and equally split into three parts, which were further processed and analysed separately, as replicates.

RNA extraction was performed using the RNeasy Mini kit (QIAGEN). Briefly, RTL buffer (600 μl) was added to each replicate, which was then homogenized using an RNAse-free syringe, and processed according to the manufacturer's protocol, including DNAse I treatment. Extracted RNA was quantified with the Qubit RNA High Sensitivity assay (Thermo Fisher Scientific). Subsequently, each sample was equally split into two parts. The first part was processed for the host RNA-seq, namely through a TruSeq stranded mRNA library including the Ribo-Zero Plus treatment for rRNA removal, and sequenced on

an Illumina NovaSeq 6000 Machine, with 2x100 bp reads ('poly(A)-enriched reads'). The second part of each sample was processed for the symbiont RNA-seq, using the NEBNext Magnetic Oligo d(T)25 beads (New England Biolabs) to deplete from eukaryotic mRNA. The manufacturer's instructions were followed up to the magnetic separation of the eukaryotic mRNA-bound beads from the supernatant, which contained other RNA species (including the targeted bacterial mRNA). Thus, RNA was recovered from the supernatant employing RNAClean XP beads (Beckman Coulter) following the manufacturer's instructions. Then, the samples were subjected to TruSeq stranded total RNA libraries with Ribo-Zero treatment and sequenced as above ('poly(A)-depleted reads').

All library preparation (including Ribo-Zero treatments) and sequencing were carried out at Macrogen Europe.

RNA-seq data analysis

The trimming procedure to eliminate adapter sequences present in our data was carried out by the sequencing company. The quality of the resulting reads was assessed using FastQC 0.11.7 [61]. For *P. primaurelia*, we downloaded from ParameciumDB (https://paramecium.i2bc.paris-saclay.fr/; accessed on 24 July 2024) the macronuclear genome, together with the respective annotation, to be used as reference. For *Megaera*, the *de novo* LgJac-assembled genome sequence annotated with Prokka was used as a reference. For each set of reads, and separately for the host and the symbiont, transcript abundance of each gene (in terms of TPM, Transcripts Per Million) was obtained with Salmon [62], using the following parameters in the quantification step: '–seqBias, –gcBias, –posBias, –incompatPrior 0, –recoverOrphans, –minScoreFraction 0.8, and –numGibbsSamples 110'.

Functional annotation of the reference genomes

These proteins encoded in the macronuclear genome of *P. primaurelia* AZ9-3 were functionally annotated with eggNOG-mapper 2.1.12 [63] using the default parameters of the online tool version. The assigned protein description and GO (Gene Ontology) term annotations were taken into account (Table S2). The GO terms annotation was further refined with GOATOOLS [64]. Specifically, it was parsed to retrieve the GO hierarchical structure based on the go-basic file (https://geneontology.org/docs/download-ontology/), and, for each GO term assigned, the corresponding hierarchical level 3 and 4 were individually selected, thus producing two distinct dataframes, each containing all GO terms annotated to the respective hierarchical level. Only genes with at least one annotated GO term were considered for the transcriptome enrichment analyses. In case a gene was associated with multiple GO terms at the same hierarchical level, we treated it as if there were as many separate genes, each assigned to a single GO term.

The proteins of *Megaera* LgJac were functionally annotated using eggNOG-mapper, with the following options: '-m diamond –target_orthologs one2one'. For the following cluster of orthologous gene (COG) enrichment analyses of the transcriptome, only genes with at least one annotated COG were considered. In case a gene was associated with multiple COG categories, we treated it as if there were as many separate genes, each assigned to a single category.

Gene expression analysis of *Paramecium*

In order to identify the differentially expressed genes (DEGs) between the LgJacMeg+ and LgJacMeg− strains, the TPM values of the respective three replicates were analysed using DESeq2 [65] with default parameters and the following thresholds: FDR-adjusted (False Discovery Rate) P-value $<.05$ and $|\log_2 FC| > 1$ (Fold Change). The reference condition was set as LgJacMeg-.

Gene clustering based on $\log_2 FC$ values across replicates was performed using pheatmap [66].

Functional enrichment analysis of GO terms for up- and down-regulated genes was performed using topGO [67]. For both up- and down-regulated genes, each of the three GO categories (Biological Process—BP; Molecular Function—MF; Cellular Component—CC) and each of the two selected hierarchical levels was analysed separately. The enrichment analysis was performed using the runTest() function, applying the classical method for managing the GO graph structure and the Fisher exact test for statistical evaluation, with the P-values adjusted for multiple testing using the Benjamini–Hochberg (BH) method. To identify significantly enriched GO terms, an adjusted P-value of $<.05$ was applied. Data visualization was attained with ggplot2 [68]. This analysis was integrated with a detailed inspection of individual DEGs.

Gene expression analysis of *Megaera*

The 'poly(A)-depleted' RNA samples sequenced for the *Megaera* RNA-Seq were expected to contain not only the targeted mRNA of this bacterium but also the non-coding RNAs (ncRNAs) of bacterium and host (besides those rRNAs effectively removed by the Ribo-Zero treatment). Therefore, to evaluate if sequencing depth was sufficient to comprehensively represent the *Megaera* transcriptome, at first, we inspected the amount of reads assigned to *Megaera*, based on the TPMs. Thus, following previous studies on bacterial symbionts [69], we performed a rarefaction analysis with vegan [70] on the reads of each replicate as well as of the three replicates merged. By this approach, it was possible to verify whether the number of expressed genes that were identified with random subsamples of each read set was comparable to those identified with the respective full set, indicative of sufficient sequencing depth. Accordingly, the combined 'poly(A)-depleted' reads were selected for the following analyses (see Results), and the TPMs were re-calculated for this merged read set as described above.

The *Megaera* gene expression was analysed by classifying the genes as follows. In a first classification ('active-inactive genes'), the dataset was separated into two subgroups, namely, active genes (TPM > 0) and inactive genes (TPM = 0) (Fig. S3). A second classification ('expression level') was focused on the transcriptionally active genes, which were further stratified into three subgroups, namely, highly expressed genes (TPM above the 75th percentile), moderately expressed genes (TPM between the 25th and 75th percentiles), and lowly expressed genes (TPM values below the 25th percentile) (Fig. S4).

A COG category enrichment analysis was performed separately for each classification using Fisher's exact test with BH-adjusted P-values, implemented in a custom R script. Similarly, for each classification, an enrichment analysis of the pangenome categories was also performed. For the 'expression-level' classification, the genes having TPM = 0 were excluded from the background of both enrichment analyses.

Horizontal gene transfer analyses

All the protein sequences annotated in the *Megaera* LgJac genome were scored for being potentially the result of horizontal acquisition events. Specifically, after a blastp search on nr with the following options: -outfmt "6 std staxids" -seg no -evalue 1e-5, the resulting hits were processed using AvP [71] to obtain the Alien Index (AI), setting

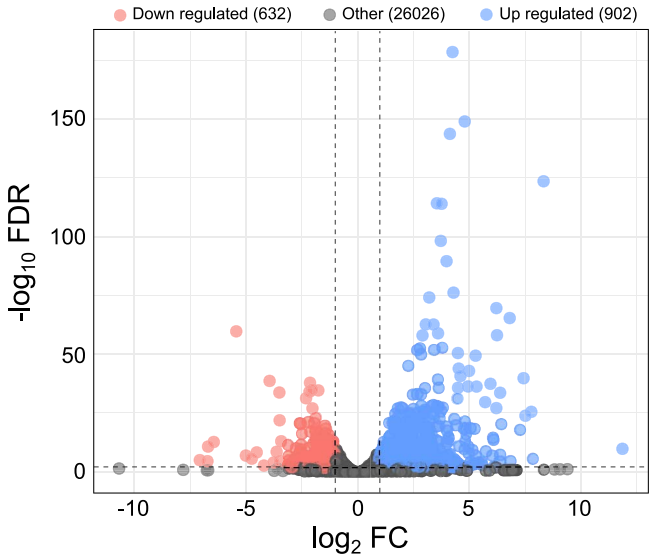


Figure 1 Volcano plot of the differential gene expression between *Paramecium* LgJac_Meg+ and LgJac_Meg−. The dashed horizontal and vertical lines indicate the thresholds applied for the FDR-adjusted *P*-value (.05) and for the $\log_2$FC (±1), respectively. Accordingly, DEGs that are up-regulated or down-regulated in LgJac_Meg+ with respect to LgJac_Meg− were selected from the other genes.

the ingroup as taxid = 988 779 ('*Candidatus* Megaera polyxenophila') and without any Excluded Genome Pool (EGP). These options allowed, respectively, to detect genes originating from non-*Megaera* taxa, and, aiming to focus on recent horizontal gene transfer (HGT) events, not to exclude any taxonomic group from the possible donors, including close relatives such as other *Rickettsiales*. Afterwards, an Alien Index threshold of >45 was applied to identify candidate HGT genes [72].

Biotin gene synthesis phylogenetic analyses

A representative set of 254 annotated bacterial genome assemblies was prepared, aiming for a phylogenetically balanced set of *Rickettsiales* together with a selection of the best online blastp hits of the six *Megaera* LgJac biotin synthesis proteins. The LgJac sequences were then blastp-queried on each proteome, and the corresponding sequences were selected after a manual inspection. Each gene was then aligned with MAFFT [73] and trimmed with BMGE [74]. After discarding organisms bearing less than three genes, the trimmed alignments were concatenated with AMAS [75], to get a more comprehensive and informative inference. Single-gene and concatenated phylogenies were performed with IQTREE v. 2.3.4 [76], using ModelFinder [77] for model selection, with 1000 ultrafast bootstraps and 1000 SH-aLRT replicates.

Results

Effects of *Megaera* on host gene expression

Our analyses revealed that *P. primaurelia* gene expression is appreciably influenced by the presence of *Megaera*, as numerous significant differentially expressed genes (DEGs) were identified in the presence of this symbiont, namely, 902 up-regulated and 632 down-regulated genes (Fig. 1; Table S3).

Functional enrichment analysis indicated 18 enriched GO terms among up-regulated genes and 32 among down-regulated genes (Fig. 2).

Many up-regulated genes were associated with regulatory functions (Fig. 2; Table S4), suggesting a complex and fine-tuned response of *Paramecium* to the presence of *Megaera*. Up-regulated processes include transcriptional and post-transcriptional (Argonaute proteins) regulation, protein turnover (a proteasome subunit, ubiquitin-protein ligases and ubiquitin transferases, Cdc48 proteins), and signal transduction (e.g. multiple protein kinases) (Tables S3 and S5). In terms of cell compartments, several up-regulated functions localize in the Golgi apparatus and pertain to the trans-Golgi network and vesicular dynamics, in particular vesicle-docking, and secretion (e.g. SNARE proteins, phosphoinositide 3-kinases, Yip1 domain proteins). Among vesicle-trafficking genes, Rab family proteins and other GTPases, as well as their respective prenyltransferases (e.g. geranyl-geranyltransferases) [78], are prominent and involved in multiple enriched GO terms (Tables S3 and S5). Several up-regulated GO terms and genes are related to nucleosome and chromatin organization and remodelling (multiple histones, histone lysine demethylases, Sir2 family histone deacetylase), being further consistent with a transcriptionally dynamic environment, also possibly indicating an overall shift towards a more condensed chromatin state. Interestingly, when inspecting in detail at the level of individual up-regulated DEGs, many of them were found to be connected to cell cycle regulation and/or division (e.g. cyclin, cyclin-kinase, anaphase-promoting complex components, Facilitates Chromatin Transcription (FACT) complex subunits, DNA replication licensing factor MCM4).

In addition, a number of up-regulated genes are linked to stress responses, including chaperones (Hsp70; HSP40/DnaJ, protein activator of Hsp90), glutathione (glutamate–cysteine ligase), and cytochrome P450 family proteins, as well as to DNA repair (Rad51).

On the other hand, most of the down-regulated GO terms are associated with metabolic processes, in particular catabolism of multiple compound classes and energy metabolism, and, consistently, with mitochondrial components (Table S4). Down-regulated pathways include general energy metabolism (Tables S3 and S6), namely Krebs cycle (e.g. 2-oxoglutarate/malate carrier, malate dehydrogenase, citrate synthases, isocitrate dehydrogenase; aconitate hydratase, fumarate hydratase and reductase, oxoglutarate dehydrogenase, succinate dehydrogenase) and oxidative phosphorylation (multiple subunits of the NADH (nicotinamide adenine dinucleotide + hydrogen)-quinone oxidoreductase and ATP - Adenosine triphosphate - synthase). Moreover, multiple compound-specific pathways are negatively affected by the presence of *Megaera*, namely, regarding carbohydrates, both glycolysis and gluconeogenesis (e.g. 6-phosphofructokinase, phosphoenolpyruvate carboxykinase, fructose-1,6-bisphosphatase, enolase, glucose-6-phosphate isomerase), as well as glycogen and glycerol degradation. Amino acid catabolism is also affected, regarding general enzymatic activities (multiple aminotransferases) and specific ones, such as degradation of glycine (cleavage system protein T), branched-chain (3-hydroxyisobutyrate dehydrogenase) or aromatic (homogentisate 1,2-dioxygenase, maleylacetoacetate isomerases, 4a-hydroxytetrahydrobiopterin dehydratase, 4-hydroxyphenylpyruvate dioxygenase) amino acids. Also, lipid catabolism is negatively affected (e.g. enoyl-CoA hydratase, acyl-CoA dehydrogenase/oxidase). Another negatively impacted function is NAD synthesis (glutamine-dependent NAD+ synthetase).

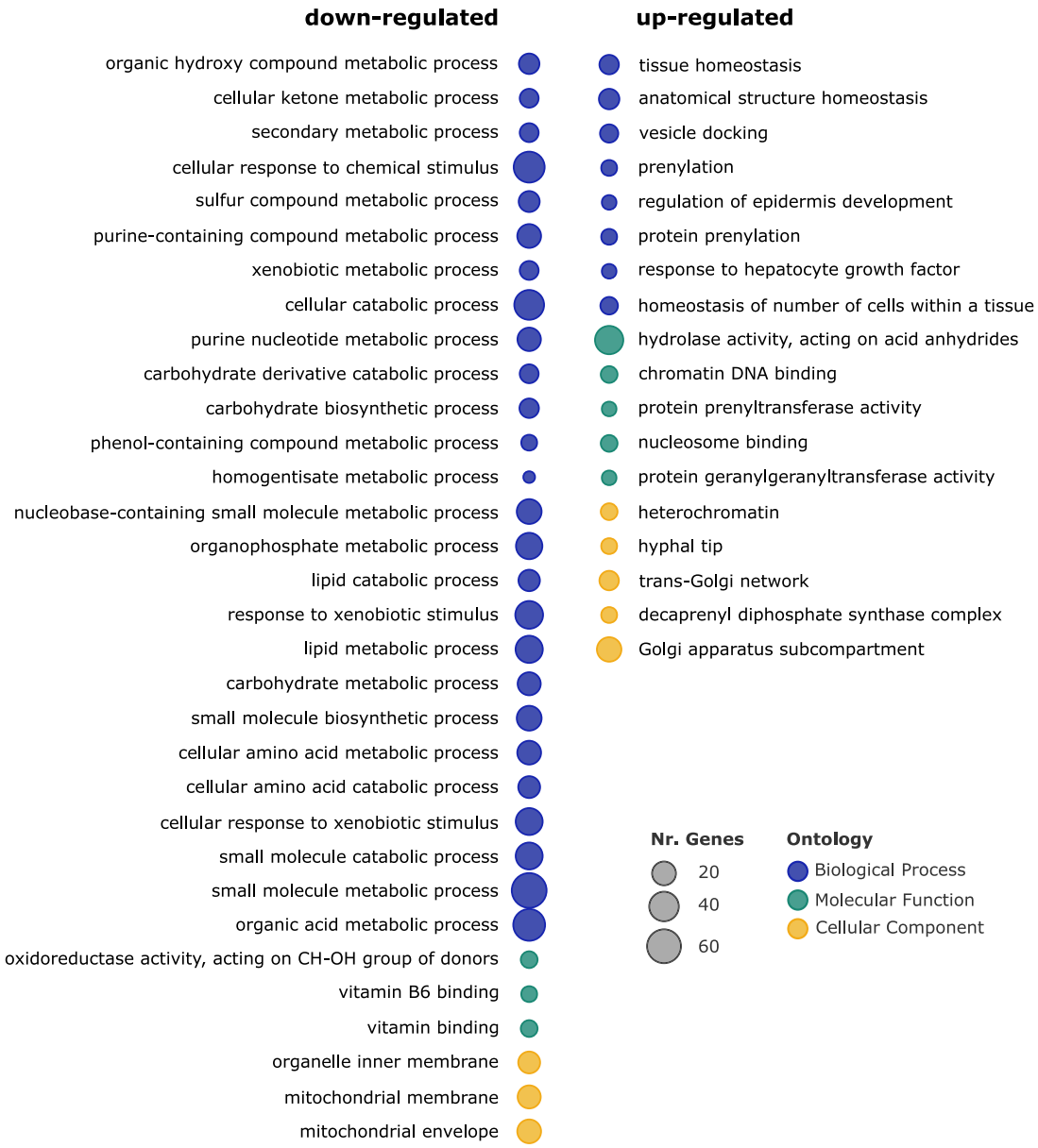


Figure 2 Enriched GO terms among down-regulated (left) and up-regulated (right) DEGs in LgJac_Meg+ with respect to LgJac_Meg−. GO terms are grouped according to their ontology category, namely, 'biological process', 'molecular function', and 'cellular component'. The circle sizes are proportional to the number of DEGs assigned to the respective GO term.

Gene expression analysis of Megaera

We analysed the transcriptome of *Megaera* to gather insights into the mechanisms and causes for the observed gene expression shifts in the host. To this purpose, the newly assembled *Megaera* LgJac genome was used as reference (four closed replicons: chromosome: 1 724 362 bp; three plasmids: 37693, 33 040, and 5131 bp), after preliminarily confirming that the assembly quality was very high and consistent with other genomes of the same species (Table S1).

The sequencing depth of two RNA-Seq replicates was preliminarily assessed to be insufficient; therefore, to reach an adequate depth, the three replicates were merged together (Fig. S5). In the absence of a reference *Megaera* condition for a DEG analysis, we performed a global assessment of its relative gene expression levels. First, we classified genes as either expressed or nonexpressed ('active/inactive genes'; Fig. S3). Out of 1755 annotated protein-coding genes, almost 80% ($n = 1377$) were found to be transcriptionally active. This set was strongly enriched in the COG category M (cell wall/membrane/envelope biogenesis), as well as both in the core and soft-core pangenome categories (Fig. S3; Table S7), thus indicating that highly conserved genes were preferentially expressed. On the other hand, the nonexpressed genes were complementarily enriched in the 'cloud' pangenome category, as well as in the COG categories L (replication, recombination, and repair) and S (function unknown) (Tables S3 and S7). Notably, L includes multiple recombinases, transposases, and DNA repair systems, which are indeed expected to be only conditionally expressed. On the other hand, several nonexpressed genes are unclassified hypothetical proteins, some of which potentially represent wrongly annotated CDSs (particularly the least conserved ones).

Next, we focused on the expressed genes, distinguishing three percentile groups ('expression level analyses'). The highly expressed genes (over 75th percentile) were enriched in multiple functions, namely, post-translational modification, protein turnover, and chaperones (O); secondary metabolite biosynthesis, transport, and catabolism (Q); transcription (K); signal transduction mechanisms (T); and function unknown (S) (Table S7), overall indicating a metabolically active and highly regulated state. This gene set was also significantly enriched in the shell and cloud pangenome categories (Fig. S4; Table S7), suggesting that strain-specific traits may be involved in those processes. The moderately expressed gene set (25th–75th percentile) was enriched in the category D (cell cycle control, cell division, and chromosome partitioning), consistent with a stable and active replication of the bacterium (Table S7). In the lowly expressed gene subset (<25th percentile), genes related to carbohydrate transport and metabolism (category G) were more represented, which may be indicative of a stable and relatively measured need for such functions within the host environment, also considering that many of those genes encode for membrane transporters. Conversely, no significant enrichment was detected in pangenome categories among the moderately and lowly expressed genes (Table S7).

Thus, we moved to an in-depth investigation at the single gene level. As expected for a *Rickettsiales* bacterium, *Megaera* LgJac expresses typical core (or soft core) components of secretion systems (Sec, Tat, and T4SS). Interestingly, an F-type conjugative pilus, moderately shared with other *Megaera* (mostly soft core/cloud), was also expressed. Among putative effectors, three ankyrin repeat proteins (two in the highly expressed genes) and five tetratricopeptide repeat proteins (one among the highly expressed genes), as well as one pentapeptide repeat protein are expressed, all of which are only poorly conserved (cloud). Besides, *Megaera* expresses several *Rickettsia*-like factors involved in host cell invasion, including Sca2, Sca4, patatin-like phospholipase and phospholipase D (core) [79, 80]. Interestingly, it also expresses multiple toxins and antitoxins, in particular four possible pairs among highly expressed genes. Most of those are not shared with other *Megaera* (cloud) and are likely involved in plasmid or other kinds of regulation but could also take part in host interactions [37]. *Megaera* also expresses five non-ribosomal peptide synthesis domain proteins (including thioesterase and gramicidin dehydrogenase-like), previously suggested to be possibly implied in defensive mutualism [30].

Megaera expresses several metabolic pathways that are down-regulated in *Paramecium* in its presence, in particular energy metabolism (pyruvate dehydrogenase; Krebs cycle; and oxidative phosphorylation complexes I, III, IV, and V, as well as the alternative bd oxidase). It also expresses five tlc nucleotide translocases, which may allow for the exchange of energy-carrying nucleotides with the host. Moreover, it bears transporters and enzymes (e.g. NAD-specific glutamate dehydrogenase, threonine dehydratase) enabling scavenging small compounds from the host and fuelling the Krebs cycle directly by its intermediates. When considering potentially host-supportive actions, *Megaera* is actively performing the synthesis of multiple vitamins and cofactors, namely, ubiquinone, lipoate, heme, folate, CoA, and biotin.

Megaera horizontally acquired a putatively host-supportive biotin synthesis operon

Subsequently, we focused on investigating the genome organization and evolutionary origin of the *Megaera* LgJac biotin synthesis genes. Indeed, while not within the highest expressed fraction (Table S7), these genes are almost absent in the other *Megaera* (shell). Conversely, the biosynthetic pathways of other vitamins and the biotin transporter *bioY* are all highly conserved (core or soft core), hinting at a possible HGT of biotin synthesis genes. Moreover, the eggNOG annotation of *P. primaurelia* indicates that this ciliate is devoid of such a pathway.

The six biotin synthesis genes are all in close proximity in the *Megaera* LgJac genome, namely, five of them consecutive on the same strand (*bioB* is consecutive but on the opposite strand), thus suggesting that they might constitute an operon. The same gene arrangement was repeatedly identified in bacterial symbionts of insects and other arthropods, with several inferred HGT events and a role in nutritional mutualism [81–83], and accordingly termed BOOM ('Biotin synthesis Operon of Obligate intracellular Microbes') [84]. Interestingly, the six biotin synthesis genes of *Megaera* LgJac had very high Alien Index scores (all >270, 3/6 with the top highest value; Table S8), which is further indicative of HGT. This was confirmed by phylogenies of every single gene and on the six concatenated genes, which were all consistent in placing the *Megaera* LgJac sequences within the same clade of BOOM sequences (hereafter, 'BOOM clade') with high support (ultrafast bootstrap ≥90% in all single proteins besides BioD, and 100% in the concatenated alignment; Figs 3, S6, and S7). This clade also encompasses numerous other *Rickettsiales* symbionts of ciliates and other protists (e.g. '*Candidatus* Fokinia spp.', *Lyticum sinuosum*, the symbiont of *Peranema trichophorum*, '*Candidatus* Trichorickettsia mobilis').

Within the BOOM clade, the relationships between *Rickettsiales* sequences and those from other bacteria, as well as among *Rickettsiales*, are quite interwoven and not consistent with organismal phylogeny (Figs 3 and S6). In particular, in the concatenated alignment, the *Megaera* LgJac sequences are not proximate with their closest relatives among the family *Rickettsiaceae* (*Rickettsia* and '*Candidatus* Trichorickettsia') but rather nested with very high support (bootstrap 97%) into a clade of '*Candidatus* Midichloriaceae' symbionts of *Paramecium*, namely, '*Candidatus* Fokinia spp.' and *L. sinuosum*. Single gene phylogenies are consistent and, remarkably, indicate that the only other biotin synthesis enzyme found in a *M. polyxenophila*, namely, a plasmid-encoded BioA from strain PCS, is very distant from LgJac and other *Rickettsiales* sequences in general (Fig. S7).

On the other hand, other *Rickettsiales* sequences, including those of symbionts of ciliates such as '*Candidatus* Grellia numerosa', form clades that are fully apart from the BOOM clade (Figs 3, S6, and S7).

Collectively, these results highly support that this operon was horizontally acquired by *Megaera* LgJac from other symbionts of *Paramecium*, probably quite recently, after the divergence from other *Megaera*. Consistent with this scenario, multiple genes linked to mobile elements (transposases, phages, plasmid toxins), which may have acted as carriers for the biotin synthesis genes, were also retrieved among the HGT candidates with high Alien Index scores (Table S8).

Discussion

Here, we investigated the effect of the *Rickettsiales* bacterium *Megaera* on its host *Paramecium* strain LgJac_Meg+, leveraging a combination of dual RNA-Seq and comparative genomics to understand the functional features of this interaction from an evolutionary perspective. While protists are prominent and ancestral hosts for *Rickettsiales* [11], to our best knowledge, this is the first study exploring a *Rickettsiales*–protist association through transcriptomics.

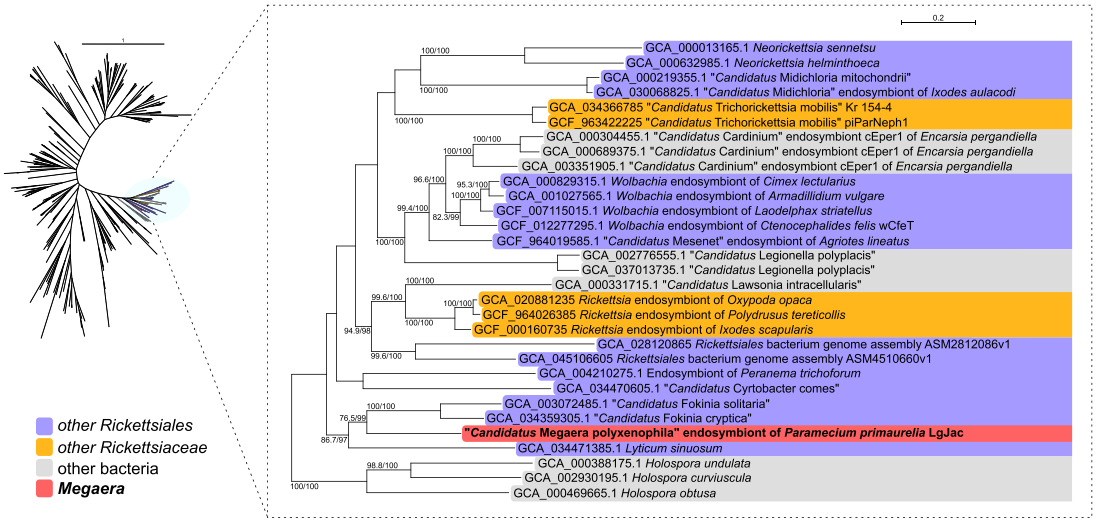


Figure 3 Maximum likelihood unlabelled tree of the concatenated biotin synthesis genes inferred with the Q.pfam + I + G4 model (left). The shaded branches, corresponding to the BOOM clade, are shown labelled in the dashed enlarged version on the right. On each branch, support values by SH-aLRT with 1000 replicates and by 1000 ultrafast bootstraps are reported (bootstrap values below 90 were omitted). The newly sequenced *Megaera* Lg_Jac is evidenced in bold. The branch colour (left) and background colour (right) indicate the organism lineages, namely, *Megaera*, other *Rickettsiaceae*, other *Rickettsiales*, and other bacteria. The scale bars stand for the estimated proportional sequence divergence. The full tree is shown in Fig. S6.

Overall, we identified 1534 *Paramecium* DEGs in response to the presence of *Megaera*. In *Paramecium*, a similar number of DEGs was detected in response to *Caedibacter taeniospiralis* (*Fastidiosibacteraceae*, *Gammaproteobacteria*) [48], while a markedly higher one was found in response to certain environmental changes [85], further suggesting that the effects of persistent bacterial symbionts on ciliates are relatively contained and fine-tuned. Nevertheless, the functions involved indicate quite substantial effects linked to the presence of *Megaera*, considering in particular the several up-regulated regulatory functions, including regulation over gene expression. The up-regulation of genes involved in chromatin condensation dynamics, DNA replication, and cell division, as well as in vesicular dynamics and Golgi-related processes, seems consistent with the previously observed higher growth rate of LgJac_Meg+ line as compared to LgJac_Meg− [25]. On the other hand, the prominent down-regulated functions are associated with catabolism of carbohydrates, amino acids, and lipids, as well as energy metabolism, namely, the Krebs cycle and oxidative phosphorylation. This seems unexpected at a first glance, considering that, as typical among *Rickettsiales* [2, 38], genomics would predict that *Megaera* relies heavily on host metabolism for energy, being supplied of intermediates for its energy metabolism as well as potentially importing ATP directly, thanks to its five tlc nucleotide translocases. However, this apparent paradox can be solved if one considers that *Megaera* is equipped with its own Krebs cycle and oxidative phosphorylation, with the produced ATP that could also become available for *Paramecium*, assuming the tlc transporters are acting in reverse [86]. This might effectively compensate the reduced expression of the corresponding host pathways, at least upon certain conditions. Accordingly, the higher *Paramecium* growth could be due to the fact that such surrogate of host metabolic functions by the symbiont might result in a more efficient exploitation of available resources, possibly due to compartmentalized functions [87], somehow analogous to cell organelles [88]. An alternative and nonmutually exclusive explanation is that such differences in growth observed by Pasqualetti and co-authors could be at least partly due to some long-term effect of the antibiotic treatment and/or subcloning [25]. In any case, the hypothesized involvement of *Megaera* would imply that this symbiont can behave as a well-integrated mutualist. This would occur provided that a sufficient food supply compensates for the overall energetic cost that *Megaera* is expected to burden the host with, given that this bacterium was not shown to broaden the range of molecules usable as food sources. Along the same line of thought, the possible implication of biotin seems noteworthy. While *Paramecium* is unable to produce this vitamin, *Megaera* expresses the biotin synthesis genes and a biotin transporter, so it could provide it to its host, further contributing to a mutualistic effect. This hypothesized biotin provision is reminiscent of the role of multiple nutritional symbionts of arthropods [81, 83, 84]. However, the bacterivorous diet of *Paramecium* and other ciliates is more varied [89], entailing that both the aposymbiotic LgJac_Meg− and at least partly the symbiotic LgJac_Meg+ may exploit biotin produced by food bacteria. Overall, this suggests a quantitative and facultative benefit provided by *Megaera*, with possible competitive advantages, which is fully consistent with experimental data, indicating a stable but nonobligate association [25].

Interestingly, *C. taeniospiralis* causes almost opposite effects on the gene expression of *Paramecium* (increased catabolism, reduced expression of proteins involved in the cell cycle) [48] and, consistently, has no tlc transporters or biotin synthesis or transporters [49]. On the other hand, '*Preeria caryophila*' (*Holosporaceae*, *Alphaproteobacteria*) was reported to have positive effects on the host growth rate, reminiscent of *Megaera* [90]. While no '*Preeria*' genome is currently publicly available, its close relatives, *Holospora* spp., have neither the Krebs cycle nor oxidative phosphorylation but do have biotin synthesis [91, 92], thus hinting at a possible similar host-supportive mechanism.

From an evolutionary perspective, all the other 12 analysed *M. polyxenophila* genomes do not encode for biotin synthesis, while the Alien Index and gene phylogenies robustly indicate that *Megaera* LgJac acquired quite recently the corresponding BOOM operon through HGT from other symbionts of *Paramecium*, such as '*Candidatus* Fokinia spp.' or *L. sinuosum* (Figs 3, S6, and S7). In general, multiple events of

transfer of BOOM have likely occurred between *Rickettsiales* and other bacterial symbionts, hosted by arthropods and protists. Interestingly, the presence of conjugal transfer machinery in many of those symbionts [2, 24, 93, 94], including LgJac and other *M. polyxenophila*, as well as of multiple mobile elements such as plasmids, transposases, and phages, suggests possible mechanisms that could support recurring exchanges of genetic material among them and/or with other bacteria. These results reinforce the notion that HGT events might be relevant for the evolution of host interactions of protist-associated *Rickettsiales* [18]. For what concerns biotin synthesis genes, these events are probably rather common and often followed by successive losses, due to changing evolutionary pressures according to host organisms and environmental conditions. This is also supported by the fact that just a single evolutionarily unrelated biotin synthesis gene (BioA) was found in only another *M. polyxenophila* strain and that another symbiont of *Paramecium* ('*Candidatus* Sarmatiella mevalonica') has a single distinct gene (BioB), which, in addition, is closely related to that of *Megaera* LgJac (Fig. S7).

Megaera also expresses nonribosomal protein synthases, some of which are expressed at very high levels, which have been previously suggested to be implied in defensive mutualisms [30]. In the context of laboratory cultivation, the role of these proteins is unclear.

Other than possible host-supportive functions, *Megaera* also bears and expresses genes relatable to exploitative actions on the host, such as homologs of *Rickettsia* host cell invasion proteins [79, 80] and a number of toxin–antitoxin modules, many of which are at a high level. Indeed, besides regulation of plasmid copy number and other bacterial cell physiological traits [95], for the latter, a role in the manipulation of host cells was suggested, in particular for tightening the association by preventing symbiont loss through addiction mechanisms [37]. The presence of secretion systems and effectors is also suggestive of an active modulation of the host cell by *Megaera*, possibly including a direct mediation of the observed gene expression variations. Among *Paramecium* up-regulated DEGs, we also found stress response chaperones such as Hsp70, which was previously found to be over-expressed by *Paramecium* in the presence of *C. taeniospiralis* and *Holospora elegans*, enabling the ciliate to better cope with heat shock [48, 96]. This finding is further consistent with a somewhat ambivalent role of *Megaera* for its host.

Overall, our results suggest multifaceted interactions between *Megaera* and its hosts, with possible physiological flexibility according to host–symbiont combinations and environmental factors, and potentially with adaptive evolutionary changes in gene content. Under this scenario, individual genes would exert multiple roles, e.g. tlc and bioY transporters may allow metabolite scavenging or alternatively be involved in their provision to the host. This suggests an intriguing parallelism with the well-studied case of *Wolbachia,* which is also a widely diffused *Rickettsiales* bacterium with a broad host range [7]. *Wolbachia* is able to addict its hosts through reproductive manipulation, but certain lineages have evolved into obligate mutualists, including biotin-providing ones, leading to its definition as a 'master manipulator of invertebrates' [97]. This evolutionary parallelism is also strengthened when considering other protist-associated *Rickettsiales.* Indeed, early studies showed that *Lyticum* (whose biotin operon is a close relative of the LgJac *Megaera* one) may behave both as a vitamin provider (folate) [98] and as a host killer [99, 100], which is an instance of an addictive manipulator [37].

Taking all the above into account, it is suggested that *Rickettsiales* in general (and potentially other lineages of the 'professional symbionts', such as *Holosporineae*, *Legionellales*, and *Chlamydiota* [101]) engage in complex and flexible interactions with protists. These bacteria, while having a primary capability to invade and exploit host cells, fulfil a fine-tuned sum of multiple actions, with the possibility to conditionally switch or evolve towards overall more positive interaction modes, including vitamin provision. This flexibility could be at the base of their wide diffusion and success in diverse hosts, in particular for clades such as *Megaera*, in which phylogenetically proximate bacteria are associated with a broad host range [27, 30].

Further studies are needed to functionally characterize the *Megaera–Paramecium* and other such interactions, in particular aiming to correlate the effect of physiological and environmental changes with potential molecular determinants, as well as accounting for a broader range of phenotypic effects, including behavioural ones. After seminal works initiated decades ago (e.g. [96–98]), thanks to the power of modern laboratory techniques and *in silico* approaches, a novel phase of experimental work on these symbiotic associations seems now possible, with a great potential for shedding light on their functional and ecological features, as well as on their evolution.

Acknowledgements

N.G. and A.P. acknowledge the University of Pisa visiting Fellow Program.

Supplementary material

Supplementary material is available at *ISME Communications* online.

Conflicts of interest

None declared.

Funding

This project was funded by the European Community's H2020 Programme H2020-MSCA-RISE 2019 under grant agreement n. 872767 to G.P. by the Italian Research Ministry PRIN project n. 2022283TH9 to M.C., and by the Italian Ministry of Universities and Research under the Department of Excellence 2023–2027 initiative.

Data availability

The genomic sequencing reads and assembled *Megaera* LgJac genome and the transcriptome reads have been deposited under the NCBI Bioproject PRJNA1297651.

References

1. Schön ME, Martijn J, Vosseberg J. *et al.* The evolutionary origin of host association in the *Rickettsiales*. *Nat Microbiol* 2022;**7**:1189–99. https://doi.org/10.1038/s41564-022-01169-x
2. Castelli M, Nardi T, Gammuto L. *et al.* Host association and intracellularity evolved multiple times independently in the *Rickettsiales*. *Nat Commun* 2024;**15**:1093. https://doi.org/10.1038/s41467-024-45351-7
3. Salje J. Cells within cells: *Rickettsiales* and the obligate intracellular bacterial lifestyle. *Nat Rev Microbiol* 2021;**19**:375–90. https://doi.org/10.1038/s41579-020-00507-2

4. Walker DH, Ismail N. Emerging and re-emerging rickettsioses: endothelial cell infection and early disease events. *Nat Rev Microbiol* 2008;**6**:375–86. https://doi.org/10.1038/nrmicro1866
5. Rikihisa Y. *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis*: subversive manipulators of host cells. *Nat Rev Microbiol* 2010;**8**:328–39. https://doi.org/10.1038/nrmicro2318
6. Lehman SS, Verhoeve VI, Driscoll TP. *et al.* Metagenome diversity illuminates the origins of pathogen effectors. *MBio* 2024;**15**:e00759–23. https://doi.org/10.1128/mbio.00759-23
7. Kaur R, Shropshire JD, Cross KL. *et al.* Living in the endosymbiotic world of *Wolbachia*: a centennial review. *Cell Host Microbe* 2021;**29**:879–93. https://doi.org/10.1016/j.chom.2021.03.006
8. Hoffmann AA, Montgomery BL, Popovici J. *et al.* Successful establishment of *Wolbachia* in *aedes* populations to suppress dengue transmission. *Nature* 2011;**476**:454–7. https://doi.org/10.1038/nature10356
9. Taylor MJ, Bandi C, Hoerauf A, Hoerauf A. *Wolbachia.* Bacterial endosymbionts of filarial nematodes. In: Baker JR, Muller R, Rollinson D (eds.), *Advances in Parasitology*. Academic Press, London-New York, 2005, 245–84, https://doi.org/10.1016/S0065-308X(05)60004-8.
10. Hoerauf A, Volkmann L, Hamelmann C. *et al.* Endosymbiotic bacteria in worms as targets for a novel chemotherapy in filariasis. *Lancet* 2000;**355**:1242–3. https://doi.org/10.1016/S0140-6736(00)02095-X
11. Castelli M, Sassera D, Petroni G. Biodiversity of "non-model" *Rickettsiales* and their association with aquatic organisms. *Rickettsiales: Biology, Molecular Biology, Epidemiology, and Vaccine Development* 2016;59–91, Cham, Springer International Publishing. https://doi.org/10.1007/978-3-319-46859-4_3
12. Carrier TJ, Leigh BA, Deaker DJ. *et al.* Microbiome reduction and endosymbiont gain from a switch in sea urchin life history. *Proc Natl Acad Sci USA* 2021;**118**:e2022023118. https://doi.org/10.1073/pnas.2022023118
13. Gruber-Vodicka HR, Leisch N, Kleiner M. *et al.* Two intracellular and cell type-specific bacterial symbionts in the placozoan *Trichoplax* H2. *Nat Microbiol* 2019;**4**:1465–74. https://doi.org/10.1038/s41564-019-0475-9
14. Yurchenko T, Ševčíková T, Přibyl P. *et al.* A gene transfer event suggests a long-term partnership between eustigmatophyte algae and a novel lineage of endosymbiotic bacteria. *ISME J* 2018;**12**:2163–75. https://doi.org/10.1038/s41396-018-0177-y
15. Hess S. Description of *Hyalodiscus flabellus* sp. nov. (Vampyrellida, Rhizaria) and identification of its bacterial endosymbiont, '*Candidatus* Megaira polyxenophila' (*Rickettsiales, Alphaproteobacteria*). *Protist* 2017;**168**:109–33. https://doi.org/10.1016/j.protis.2016.11.003
16. Galindo LJ, Torruella G, Moreira D. *et al.* Combined cultivation and single-cell approaches to the phylogenomics of nucleariid amoebae, close relatives of fungi. *Philos Trans R Soc B Biol Sci* 2019;**374**:20190094. https://doi.org/10.1098/rstb.2019.0094
17. Kwan JC, Schmidt EW. Bacterial endosymbiosis in a chordate host: long-term co-evolution and conservation of secondary metabolism. *PLoS One* 2013;**8**:e80822. https://doi.org/10.1371/journal.pone.0080822
18. Castelli M, Lanzoni O, Nardi T. *et al.* '*Candidatus* Sarmatiella mevalonica' endosymbiont of the ciliate *paramecium* provides insights on evolutionary plasticity among *Rickettsiales*. *Environ Microbiol* 2021;**23**:1684–701. https://doi.org/10.1111/1462-2920.15396
19. Klinges JG, Rosales SM, McMinds R. *et al.* Phylogenetic, genomic, and biogeographic characterization of a novel and ubiquitous marine invertebrate-associated *Rickettsiales* parasite, *Candidatus* Aquarickettsia rohweri, gen. nov., sp. nov. *ISME J* 2019;**13**:2938–53. https://doi.org/10.1038/s41396-019-0482-0
20. Kawafune K, Hongoh Y, Hamaji T. *et al.* Molecular identification of rickettsial endosymbionts in the non-phagotrophic volvocalean green algae. *PLoS One* 2012;**7**:e31749. https://doi.org/10.1371/journal.pone.0031749
21. Castelli M, Sabaneyeva E, Lanzoni O. *et al. Deianiraea*, an extracellular bacterium associated with the ciliate *Paramecium*, suggests an alternative scenario for the evolution of *Rickettsiales*. *ISME J* 2019;**13**:2280–94. https://doi.org/10.1038/s41396-019-0433-9
22. George EE, Husnik F, Tashyreva D. *et al.* Highly reduced genomes of protist endosymbionts show evolutionary convergence. *Curr Biol* 2020;**30**:925–933.e3. https://doi.org/10.1016/j.cub.2019.12.070
23. Kang Y-J, Diao X-N, Zhao G-Y. *et al.* Extensive diversity of *Rickettsiales* bacteria in two species of ticks from China and the evolution of the *Rickettsiales*. *BMC Evol Biol* 2014;**14**:167. https://doi.org/10.1186/s12862-014-0167-2
24. Schulz F, Martijn J, Wascher F. *et al.* A *Rickettsiales* symbiont of amoebae with ancient features. *Environ Microbiol* 2016;**18**:2326–42. https://doi.org/10.1111/1462-2920.12881
25. Pasqualetti C, Szokoli F, Rindi L. *et al.* The obligate symbiont "*Candidatus* Megaira polyxenophila" has variable effects on the growth of different host species. *Front Microbiol* 2020;**11**:11. https://doi.org/10.3389/fmicb.2020.01425
26. Candidatus OA, No L. 4: lists of names of prokaryotic *Candidatus* taxa. *Int J Syst Evol Microbiol* 2022;**72**:005545. https://doi.org/10.1099/ijsem.0.005545
27. Lanzoni O, Sabaneyeva E, Modeo L. *et al.* Diversity and environmental distribution of the cosmopolitan endosymbiont '*Candidatus* Megaira'. *Sci Rep* 2019;**9**:1179. https://doi.org/10.1038/s41598-018-37629-w
28. Schrallhammer M, Ferrantini F, Vannini C. *et al.* '*Candidatus* Megaira polyxenophila' gen. Nov., sp. nov.: considerations on evolutionary history, host range and shift of early divergent rickettsiae. *PLoS One* 2013;**8**:e72581. https://doi.org/10.1371/journal.pone.0072581
29. Yang A, Narechania A, Kim E. Rickettsial endosymbiont in the "early-diverging" streptophyte green alga *Mesostigma viride*. *J Phycol* 2016;**52**:219–29. https://doi.org/10.1111/jpy.12385
30. Davison HR, Hurst GDD, Siozios S. '*Candidatus* Megaira' are diverse symbionts of algae and ciliates with the potential for defensive symbiosis. *Microb Genom* 2023;**9**:000950. https://doi.org/10.1099/mgen.0.000950
31. Kawafune K, Hongoh Y, Hamaji T. *et al.* Two different rickettsial bacteria invading *Volvox carteri*. *PLoS One* 2015;**10**:e0116192. https://doi.org/10.1371/journal.pone.0116192
32. Zaila KE, Doak TG, Ellerbrock H. *et al.* Diversity and universality of endosymbiotic *rickettsia* in the fish parasite *Ichthyophthirius multifiliis*. *Front Microbiol* 2017;**8**:189. https://doi.org/10.3389/fmicb.2017.00189
33. Sunagawa S, DeSantis TZ, Piceno YM. *et al.* Bacterial diversity and white plague disease-associated community changes in the Caribbean coral *Montastraea faveolata*. *ISME J* 2009;**3**:512–21. https://doi.org/10.1038/ismej.2008.131
34. Fraune S, Bosch TCG. Long-term maintenance of species-specific bacterial microbiota in the basal metazoan *hydra*. *Proc Natl Acad Sci USA* 2007;**104**:13146–51. https://doi.org/10.1073/pnas.0703375104
35. Dishaw LJ, Flores-Torres J, Lax S. *et al.* The gut of geographically disparate *Ciona intestinalis* harbors a core microbiota. *PLoS One* 2014;**9**:e93386. https://doi.org/10.1371/journal.pone.0093386

36. Hine PM, Wakefield S, Diggles BK. *et al.* Ultrastructure of a haplosporidian containing Rickettsiae, associated with mortalities among cultured paua *Haliotis iris*. *Dis Aquat Org* 2002;**49**:207–19. https://doi.org/10.3354/dao049207
37. Castelli M, Nardi T, Giovannini M. *et al.* Addictive manipulation: a perspective on the role of reproductive parasitism in the evolution of bacteria–eukaryote symbioses. *Biol Lett* 2024;**20**:20240310. https://doi.org/10.1098/rsbl.2024.0310
38. Driscoll TP, Verhoeve VI, Guillotte ML. *et al.* Wholly *rickettsia*! Reconstructed metabolic profile of the quintessential bacterial parasite of eukaryotic cells. *MBio* 2017;**8**:e00859–17. https://doi.org/10.1128/mBio.00859-17
39. Westermann AJ, Barquist L, Vogel J. Resolving host–pathogen interactions by dual RNA-seq. *PLoS Pathog* 2017;**13**:e1006033. https://doi.org/10.1371/journal.ppat.1006033
40. Mika-Gospodorz B, Giengkam S, Westermann AJ. *et al.* Dual RNA-seq of *Orientia tsutsugamushi* informs on host-pathogen interactions for this neglected intracellular human pathogen. *Nat Commun* 2020;**11**:3363. https://doi.org/10.1038/s41467-020-17094-8
41. Pierlé SA, Hammac GK, Palmer GH. *et al.* Transcriptional pathways associated with the slow growth phenotype of transformed *Anaplasma marginale*. *BMC Genomics* 2013;**14**:272. https://doi.org/10.1186/1471-2164-14-272
42. Li Y-H, Peng J, Wu Q-J. *et al.* Transcriptome data reveal beneficial effects of *rickettsia* (*Rickettsiales: Rickettsiaceae*) on *Bemisia tabaci* (Hemiptera: Aleyrodidae) through nutritional factors and defense mechanisms. *J Econ Entomol* 2024;**117**:817–24. https://doi.org/10.1093/jee/toae066
43. Selberherr E, Penz T, König L. *et al.* The life cycle-dependent transcriptional profile of the obligate intracellular amoeba symbiont *Amoebophilus asiaticus*. *FEMS Microbiol Ecol* 2022;**98**:fiac001. https://doi.org/10.1093/femsec/fiac001
44. Hamann E, Gruber-Vodicka H, Kleiner M. *et al.* Environmental Breviatea harbour mutualistic *arcobacter* epibionts. *Nature* 2016;**534**:254–8. https://doi.org/10.1038/nature18297
45. Weissenmayer BA, Prendergast JGD, Lohan AJ. *et al.* Sequencing illustrates the transcriptional response of *legionella pneumophila* during infection and identifies seventy novel small non-coding RNAs. *PLoS One* 2011;**6**:e17570. https://doi.org/10.1371/journal.pone.0017570
46. Maita C, Matsushita M, Miyoshi M. *et al.* Amoebal endosymbiont *Neochlamydia* protects host amoebae against *legionella pneumophila* infection by preventing *legionella* entry. *Microbes Infect* 2018;**20**:236–44. https://doi.org/10.1016/j.micinf.2017.12.012
47. Stephens ME, Benjamino J, Graf J. *et al.* Simultaneous single-cell genome and transcriptome sequencing of termite hindgut protists reveals metabolic and evolutionary traits of their endosymbionts. *mSphere* 2022;**7**:e00021–2. https://doi.org/10.1128/msphere.00021-22
48. Grosser K, Ramasamy P, Amirabad AD. *et al.* More than the "killer trait": infection with the bacterial endosymbiont *Caedibacter taeniospiralis* causes transcriptomic modulation in *paramecium* host. *Genome Biol Evol* 2018;**10**:646–56. https://doi.org/10.1093/gbe/evy024
49. Pirritano M, Zaburannyi N, Grosser K. *et al.* Dual-Seq reveals genome and transcriptome of *Caedibacter taeniospiralis*, obligate endosymbiont of *paramecium*. *Sci Rep* 2020;**10**:9727. https://doi.org/10.1038/s41598-020-65894-1
50. Kodama Y, Suzuki H, Dohra H. *et al.* Comparison of gene expression of *Paramecium bursaria* with and without *Chlorella variabilis* symbionts. *BMC Genomics* 2014;**15**:183. https://doi.org/10.1186/1471-2164-15-183
51. Manz W, Amann R, Ludwig W. *et al.* Phylogenetic oligodeoxynucleotide probes for the major subclasses of *proteobacteria*: problems and solutions. *Syst Appl Microbiol* 1992;**15**:593–600. https://doi.org/10.1016/S0723-2020(11)80121-9
52. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;**30**:2114–20. https://doi.org/10.1093/bioinformatics/btu170
53. 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. https://doi.org/10.1089/cmb.2012.0021
54. Kumar S, Jones M, Koutsovoulos G. *et al.* Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. *Front Genet* 2013;**4**:237. https://doi.org/10.3389/fgene.2013.00237
55. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods* 2012;**9**:357–9. https://doi.org/10.1038/nmeth.1923
56. Nardi T, Olivieri E, Kariuki E. *et al.* Sequence of a *Coxiella* endosymbiont of the tick *Amblyomma nuttalli* suggests a pattern of convergent genome reduction in the *Coxiella* genus. *Genome Biol Evol* 2021;**13**:evaa253. https://doi.org/10.1093/gbe/evaa253
57. Wick RR, Judd LM, Gorrie CL. *et al.* Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017;**13**:e1005595. https://doi.org/10.1371/journal.pcbi.1005595
58. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;**30**:2068–9. https://doi.org/10.1093/bioinformatics/btu153
59. Simão FA, Waterhouse RM, Ioannidis P. *et al.* BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 2015;**31**:3210–2. https://doi.org/10.1093/bioinformatics/btv351
60. Gautreau G, Bazin A, Gachet M. *et al.* PPanGGOLiN: depicting microbial diversity via a partitioned pangenome graph. *PLoS Comput Biol* 2020;**16**:e1007732. https://doi.org/10.1371/journal.pcbi.1007732
61. Castelli M, Lanzoni O, Rossi L. *et al.* Evaluation of enrichment protocols for bacterial endosymbionts of ciliates by real-time PCR. *Curr Microbiol* 2016;**72**:723–32. https://doi.org/10.1007/s00284-016-1006-z
62. Patro R, Duggal G, Love MI. *et al.* Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 2017;**14**:417–9. https://doi.org/10.1038/nmeth.4197
63. Cantalapiedra CP, Hernández-Plaza A, Letunic I. *et al.* eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol* 2021;**38**:5825–9. https://doi.org/10.1093/molbev/msab293
64. Klopfenstein DV, Zhang L, Pedersen BS. *et al.* GOATOOLS: a python library for gene ontology analyses. *Sci Rep* 2018;**8**:10872. https://doi.org/10.1038/s41598-018-28948-z
65. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;**15**:550. https://doi.org/10.1186/s13059-014-0550-8
66. Kolde R. *Pheatmap: Pretty Heatmaps*, Vol. **2019**, 2019, 10.32614/CRAN.package.pheatmap.
67. Alexa A, Rahnenführer J. *Gene Set Enrichment Analysis with topGO*, 2024. https://doi.org/10.18129/B9.bioc.topGO. R package version 2.63.0, https://bioconductor.org/packages/topGO
68. H. Wickham *ggplot2*. Cham: Springer International Publishing, 2016, https://doi.org/10.1007/978-3-319-24277-4.
69. Chung M, Basting PJ, Patkus RS. *et al.* A meta-analysis of *Wolbachia* transcriptomics reveals a stage-specific *Wolbachia*

transcriptional response shared across different hosts. *G3 Genes|Genomes|Genetics* 2020;**10**:3243–60. https://doi.org/10.1534/g3.120.401534
70. Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci* 2003;**14**:927–30. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x
71. Koutsovoulos GD, Noriot SG, Bailly-Bechet M. *et al.* AvP: a software package for automatic phylogenetic detection of candidate horizontal gene transfers. *PLoS Comput Biol* 2022;**18**:e1010686. https://doi.org/10.1371/journal.pcbi.1010686
72. Gladyshev EA, Meselson M, Arkhipova IR. Massive horizontal gene transfer in bdelloid rotifers. *Science* 2008;**320**:1210–3. https://doi.org/10.1126/science.1156407
73. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;**30**:772–80. https://doi.org/10.1093/molbev/mst010
74. Criscuolo A, Gribaldo S. BMGE (block mapping and gathering with entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* 2010;**10**:210. https://doi.org/10.1186/1471-2148-10-210
75. Borowiec ML. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 2016;**4**:e1660. https://doi.org/10.7717/peerj.1660
76. Minh BQ, Schmidt HA, Chernomor O. *et al.* IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 2020;**37**:1530–4. https://doi.org/10.1093/molbev/msaa015
77. Kalyaanamoorthy S, Minh BQ, Wong TKF. *et al.* ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017;**14**:587–9. https://doi.org/10.1038/nmeth.4285
78. Bright LJ, Gout J-F, Lynch M. Early stages of functional diversification in the Rab GTPase gene family revealed by genomic and localization studies in *paramecium* species. *Mol Biol Cell* 2017;**28**:1101–10. https://doi.org/10.1091/mbc.E16-06-0361
79. Sears KT, Ceraul SM, Gillespie JJ. *et al.* Surface proteome analysis and characterization of surface cell antigen (Sca) or autotransporter family of *rickettsia typhi*. *PLoS Pathog* 2012;**8**:e1002856. https://doi.org/10.1371/journal.ppat.1002856
80. Gillespie JJ, Kaur SJ, Rahman MS. *et al.* Secretome of obligate intracellular *rickettsia*. *FEMS Microbiol Rev* 2015;**39**:47–80. https://doi.org/10.1111/1574-6976.12084
81. Nikoh N, Hosokawa T, Moriyama M. *et al.* Evolutionary origin of insect-*Wolbachia* nutritional mutualism. *Proc Natl Acad Sci USA* 2014;**111**:10257–62. https://doi.org/10.1073/pnas.1409284111
82. Gerth M, Bleidorn C. Comparative genomics provides a timeframe for *Wolbachia* evolution and exposes a recent biotin synthesis operon transfer. *Nat Microbiol* 2016;**2**:16241. https://doi.org/10.1038/nmicrobiol.2016.241
83. Duron O, Gottlieb Y. Convergence of nutritional symbioses in obligate blood feeders. *Trends Parasitol* 2020;**36**:816–25. https://doi.org/10.1016/j.pt.2020.07.007
84. Driscoll TP, Verhoeve VI, Brockway C. *et al.* Evolution of *Wolbachia* mutualism and reproductive parasitism: insight from two novel strains that co-infect cat fleas. *PeerJ* 2020;**8**:e10646. https://doi.org/10.7717/peerj.10646
85. Cheaib M, Dehghani Amirabad A, Nordström KJV. *et al.* Epigenetic regulation of serotype expression antagonizes transcriptome dynamics in *paramecium tetraurelia*. *DNA Res* 2015;**22**:293–305. https://doi.org/10.1093/dnares/dsv014
86. Mehta AP, Supekova L, Chen J-H. *et al.* Engineering yeast endosymbionts as a step toward the evolution of mitochondria. *Proc Natl Acad Sci* 2018;**115**:11796–801. https://doi.org/10.1073/pnas.1813143115
87. Zecchin A, Stapor PC, Goveia J. *et al.* Metabolic pathway compartmentalization: an underappreciated opportunity? *Curr Opin Biotechnol* 2015;**34**:73–81. https://doi.org/10.1016/j.copbio.2014.11.022
88. Gabaldón T, Pittis AA. Origin and evolution of metabolic sub-cellular compartmentalization in eukaryotes. *Biochimie* 2015;**119**:262–8. https://doi.org/10.1016/j.biochi.2015.03.021
89. Boscaro V, Felletti M, Vannini C. *et al. Polynucleobacter necessarius*, a model for genome reduction in both free-living and symbiotic bacteria. *Proc Natl Acad Sci USA* 2013;**110**:18590–5. https://doi.org/10.1073/pnas.1316687110
90. Bella C, Koehler L, Grosser K. *et al.* Fitness impact of obligate intranuclear bacterial symbionts depends on host growth phase. *Front Microbiol* 2016;**7**:7. https://doi.org/10.3389/fmicb.2016.02084
91. Dohra H, Tanaka K, Suzuki T. *et al.* Draft genome sequences of three *Holospora* species (*Holospora obtusa*, *Holospora undulata*, and *Holospora elegans*), endonuclear symbiotic bacteria of the ciliate *Paramecium caudatum*. *FEMS Microbiol Lett* 2014;**359**:16–8. https://doi.org/10.1111/1574-6968.12577
92. Beliavskaia AY, Predeus AV, Garushyants SK. *et al.* New intranuclear symbiotic bacteria from macronucleus of *Paramecium putrinum* - "*Candidatus* Gortzia yakutica". *Diversity* 2020;**12**:198. https://doi.org/10.3390/d12050198
93. Giannotti D, Boscaro V, Husnik F. *et al.* The 'other' *Rickettsiales*: an overview of the family '*Candidatus* Midichloriaceae'. *Appl Environ Microbiol* 2022;**88**:e0243221. https://doi.org/10.1128/aem.02432-21
94. Gillespie JJ, Joardar V, Williams KP. *et al.* A *rickettsia* genome overrun by mobile genetic elements provides insight into the acquisition of genes characteristic of an obligate intracellular lifestyle. *J Bacteriol* 2012;**194**:376–94. https://doi.org/10.1128/JB.06244-11
95. Harms A, Brodersen DE, Mitarai N. *et al.* Toxins, targets, and triggers: an overview of toxin-antitoxin biology. *Mol Cell* 2018;**70**:768–84. https://doi.org/10.1016/j.molcel.2018.01.003
96. Hori M, Fujishima M. The endosymbiotic bacterium *Holospora obtusa* enhances heat-shock gene expression of the host *Paramecium caudatum*. *J Eukaryot Microbiol* 2003;**50**:293–8. https://doi.org/10.1111/j.1550-7408.2003.tb00137.x
97. Werren JH, Baldo L, Clark ME. *Wolbachia*: master manipulators of invertebrate biology. *Nat Rev Microbiol* 2008;**6**:741–51. https://doi.org/10.1038/nrmicro1969
98. Soldo AT, Godoy GA. Observations on the production of folic acid by symbiont lambda particles of *Paramecium aurelia* stock 299. *J Protozool* 1973;**20**:502.
99. Jurand A, Rudman BM, Preer JRJ. Prelethal symptoms of the killing action of some non-kappa killers in the *Paramecium aurelia* complex. *J Exp Zool* 1982;**220**:135–45. https://doi.org/10.1002/jez.1402200202
100. Preer JR, Preer LB, Jurand A. Kappa and other endosymbionts in *paramecium aurelia*. *Bacteriol Rev* 1974;**38**:113–63. https://doi.org/10.1128/br.38.2.113-163.1974
101. Husnik F, Tashyreva D, Boscaro V. *et al.* Bacterial and archaeal symbioses with protists. *Curr Biol* 2021;**31**:R862–77. https://doi.org/10.1016/j.cub.2021.05.049

© The Author(s) 2026. Published by Oxford University Press on behalf of the International Society for Microbial Ecology. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
ISME Communications, 2026, **6(1)**, ycag079
https://doi.org/10.1093/ismeco/ycag079
Original Article