

## **Symbiotic Phylogenetics**

Article title: **Partner fidelity and environmental filtering preserve stage-specific turtle ant gut symbioses for over 40 million years**

Journal: **Ecological Monographs**

Authors: Yi Hu<sup>1,2\*</sup>, Catherine L. D'Amelio<sup>2</sup>, Benoît Béchade<sup>2</sup>, Christian S. Cabuslay<sup>2</sup>, Piotr Lukasik<sup>3</sup>, Jon G. Sanders<sup>4</sup>, Shauna Price<sup>4</sup>, Emily Fanwick<sup>2</sup>, Scott Powell<sup>5</sup>, Corrie S. Moreau<sup>4</sup>, Jacob A. Russell<sup>2</sup>

Affiliations: <sup>1</sup> State Key Laboratory of Earth Surface Processes and Resource Ecology and Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, College of Life Sciences, Beijing Normal University, Beijing 100875, China

<sup>2</sup> Department of Biology, Drexel University, Philadelphia, PA 19104, USA

<sup>3</sup> Institute of Environmental Sciences, Faculty of Biology, Jagiellonian University, Krakow 31-007, Poland

<sup>4</sup> Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA

<sup>5</sup> Department of Biological Sciences, George Washington University, Washington, DC 20052, USA

## Content

<b>1) Methods .....</b>	<b>3</b>
a. Symbiont phylogenetics .....	3
<b>2) Figures .....</b>	<b>6</b>
a. Figure S1: Maximum likelihood phylogenetic analysis of the Actinobacteria (Analysis #4) .....	6
b. Figure S2: Maximum likelihood phylogenetic analysis of the Lactobacillales (Analysis #5) .....	8
c. Figure S3: Maximum likelihood phylogenetic analysis of ambiguously classifying Pseudomonadales OTU046 (Analysis #6). ....	10
d. Figure S4: Maximum likelihood phylogenetic analysis of ambiguously classifying Sphingobacteriales OTU116 (Analysis #7). ....	12
e. Figure S5: Maximum likelihood phylogenetic analysis of other sequences with high abundance, and their relatives (Analysis #8). ....	14
<b>3) References.....</b>	<b>16</b>

## Supplementary methods:

### *Symbiotic phylogenetics*

In order to investigate whether these putatively non-specialized OTUs fall into previously unforeseen clades showing cephalotine exclusivity, we performed the following phylogenetic analysis. In our first approach, we prioritized the most abundant 97% OTUs from the larva-dominant orders Rhizobiales (Analysis #1), Enterobacterales (Analysis #2), and Lactobacillales (Analysis #3), identifying close relatives through BLASTn searches against the NCBI nr database. For query sequences we chose all unique 16S rRNA sequences from the focal OTUs with relative abundance  $> 0.01$  in at least one larval sequence library. We retained six to seven of the top BLASTn hits for each unique sequence for our phylogenetic analyses. In addition, we included 16S rRNA from related bacterial isolates cultured from *Cephalotes* larvae or adults [1]. We also included near-full-length 16S rRNA symbiont sequences obtained from *Cephalotes varians* by culture-independent Sanger sequencing of PCR products obtained with universal eubacterial primers (9Fa and 1513R). Larvae used in this direct sequencing were chosen upon inspection of amplicon sequencing data, and for each selected sample, the targeted OTU of interest made up  $\geq 0.90$  relative abundance, maximizing our chances of obtaining a ‘clean’ sequence.

The above-mentioned sequences were aligned using the Ribosomal Database Project Sequence aligner [2]. Alignments were uploaded to the CIPRES web portal [3], where maximum-likelihood phylogenies were inferred using RAXML-HPC BlackBox with default parameters [4]. Phylogenies were then visualized and annotated using the Interactive Tree of Life Website [5]. Using these figures, we searched for patterns suggestive of newly identified

cephalotine-specific lineages – i.e. containing representatives from multiple *Cephalotes* and/or *Procryptocerus* species in monophyletic groupings, and showing low relatedness to bacteria from non-cephalotine habitats.

While the three aforementioned taxa comprised substantial majorities of our larval sequence libraries, we detected a range of additional sequences in association with turtle ant immatures, with some occasionally reaching high abundance. We dissected the evolutionary origins of such common, though less ubiquitous larvae-associated bacteria through five additional analyses. The first (Analysis #4), focused on unique sequences (n=73) from the 25 OTUs unambiguously classifying to the phylum Actinobacteria – a group of interest due to their common production of antimicrobials. For the next analysis (Analysis #5), we broadened our focus on the Lactobacillales to encompass less abundant sequences not part of our initial phylogenetic inference for this order in Analysis #3. We, next, studied the ambiguously classified, but occasionally abundant proteobacterial sequences of OTU046 (Analysis #6). Additionally analyzed were sequences from OTU116 (Analysis #7), an ambiguously classifying group with a top BLASTn hit showing ~85% sequence identity to a specialized Sphingobacteriales symbiont from *Cephalotes*. For our final analysis (#8), we focused on a subset of the aforementioned 154 OTUs, ranking within the top 100 of all named OTUs that: 1) did not belong to an OTU assigning as a specialized cephalotine bacterium via BLAST; 2) had not been studied extensively in previously published phylogenetic analyses on ant-associated bacteria (e.g. *Wolbachia*, Entomoplasmatales – Russell et al. [6]; Funaro et al. [7]); and 3) were not included in phylogenetic Analyses #s 1-7. Top BLASTn hits were retrieved from NCBI for

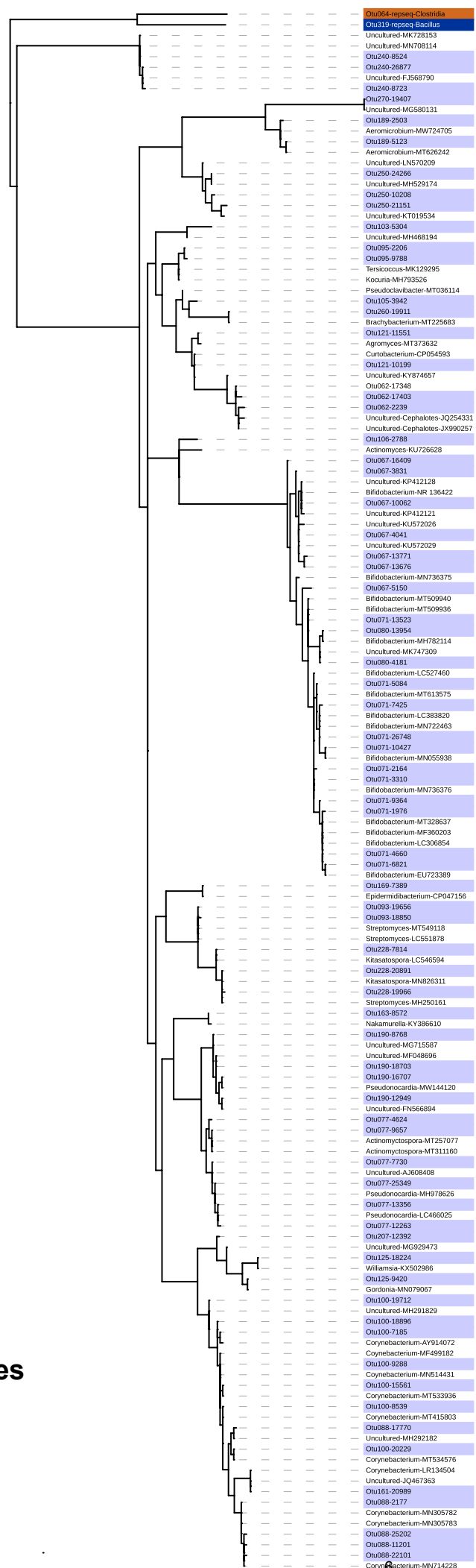


these analyses, with subsequent ClustalO alignment, and PhyML-implemented phylogenetic inference in SeaView. Additional details on methodology can be found in the supplementary figure legends for each inferred phylogeny.

Tree scale: 0.1

## Taxonomy of turtle ant-driven sequences

- Clostridiales
- Bacillales
- Actinobacteria

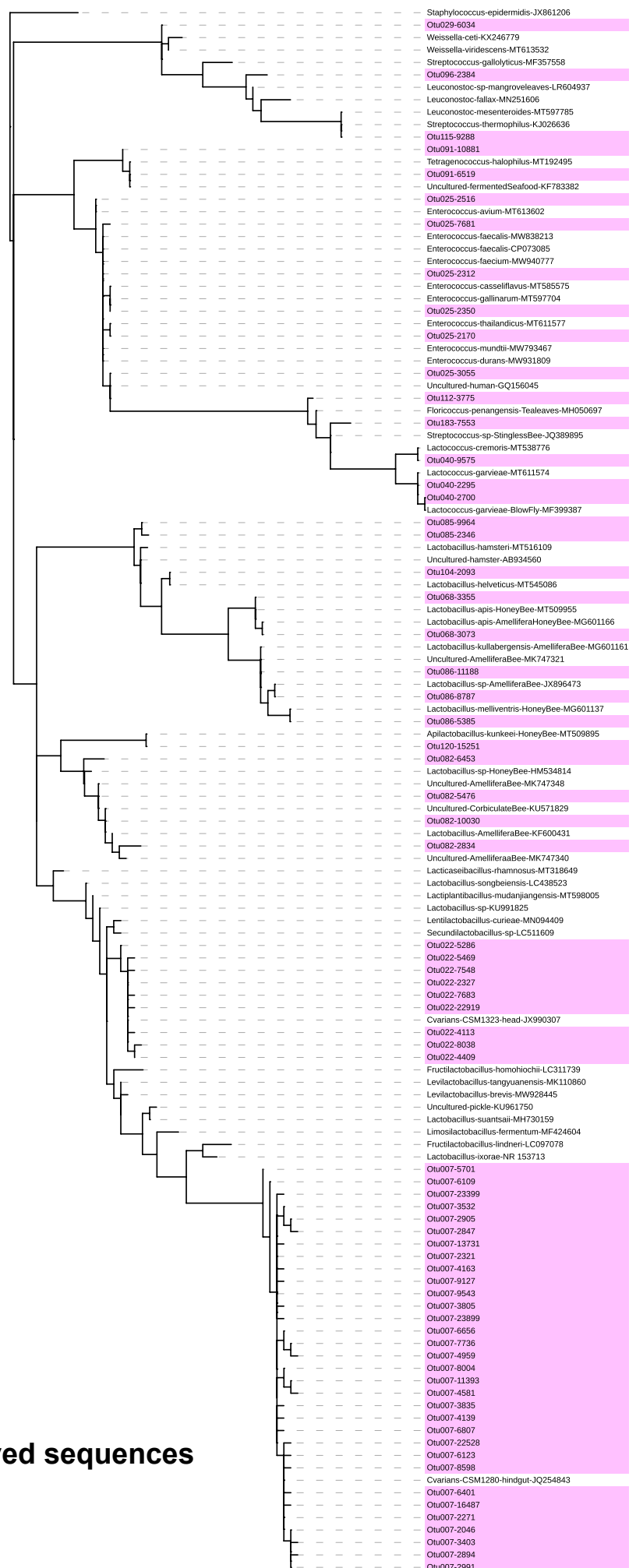


**Figure S1: Maximum likelihood phylogenetic analysis of the Actinobacteria (Analysis #4).** For this analysis (#4) we gathered all unique 16S rRNA sequences (n=73) coming from the 25 OTUs with unambiguous classifications to the phylum Actinobacteria. Top BLASTn hits were obtained for each unique sequence (vs. NCBI's nr database). If redundant with a prior top BLAST hit, we moved down the list to choose the most highly related, non-redundant match (as done for Analysis #s 5-9). After sequence alignment, we applied a Maximum Likelihood framework in Seaview, with the PhyML method and 100 bootstrap replicates. Phylogenies were visualized using the interactive Tree of Life website, and were inspected for monophyletic, cephalotine ant-specific clades showing distributions across multiple *Cephalotes* and/or *Procryptocerus* ant species.

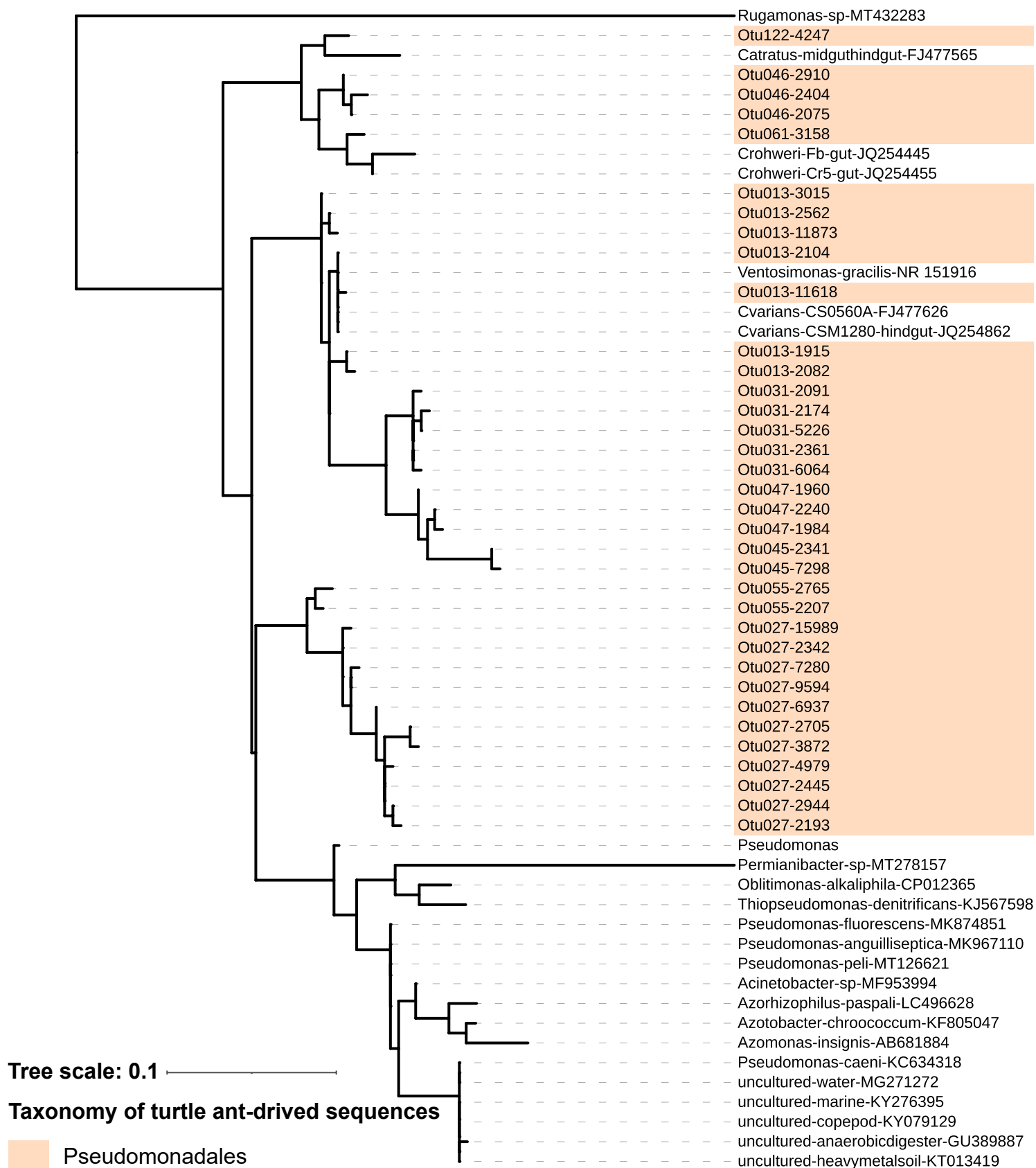
Tree scale: 0.1

## Taxonomy of turtle ant-driven sequences

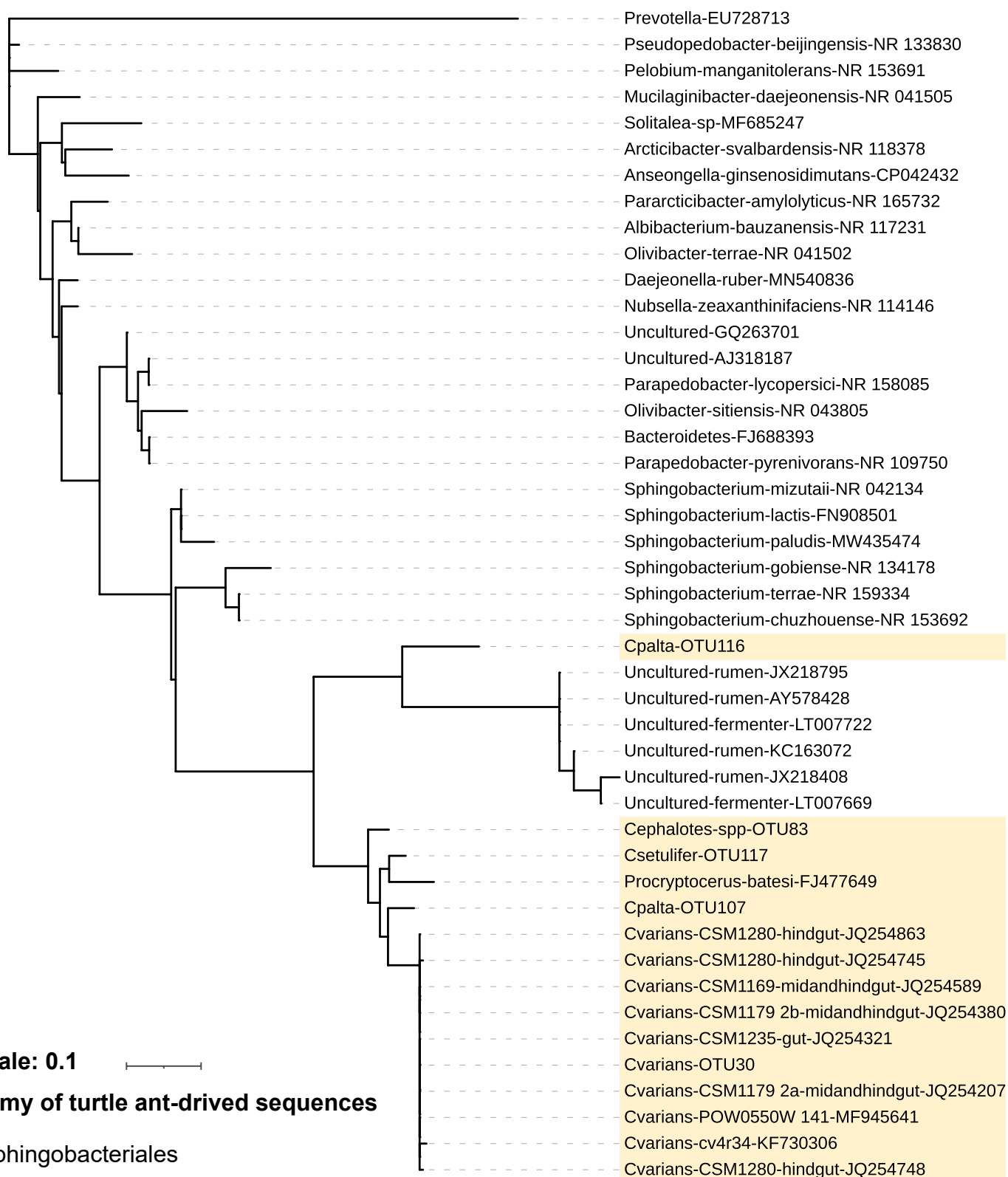
Lactobacillales



**Figure S2: Maximum likelihood phylogenetic analysis of the Lactobacillales (Analysis #5).** For this analysis we broadened our focus to encompass less abundant Lactobacillales sequences not part of initial phylogenetic inference for this order (Analysis #3). Included were all unique 16S rRNA sequences from the order with relative abundance >0.002994 in at least two sequence libraries or >0.05 in one library. Top BLASTn hits (vs. NCBI's nr database) were obtained for each unique sequence, and if redundant with a prior top BLAST hit, we moved down the list to choose the most highly related, non-redundant match. After sequence alignment, we applied a Maximum Likelihood framework in Seaview, with the PhyML method and 100 bootstrap replicates. Phylogenies were visualized using the interactive Tree of Life website, and were inspected for monophyletic, cephalotine ant-specific clades showing distributions across multiple *Cephalotes* and/or *Procryptocerus* ant species.

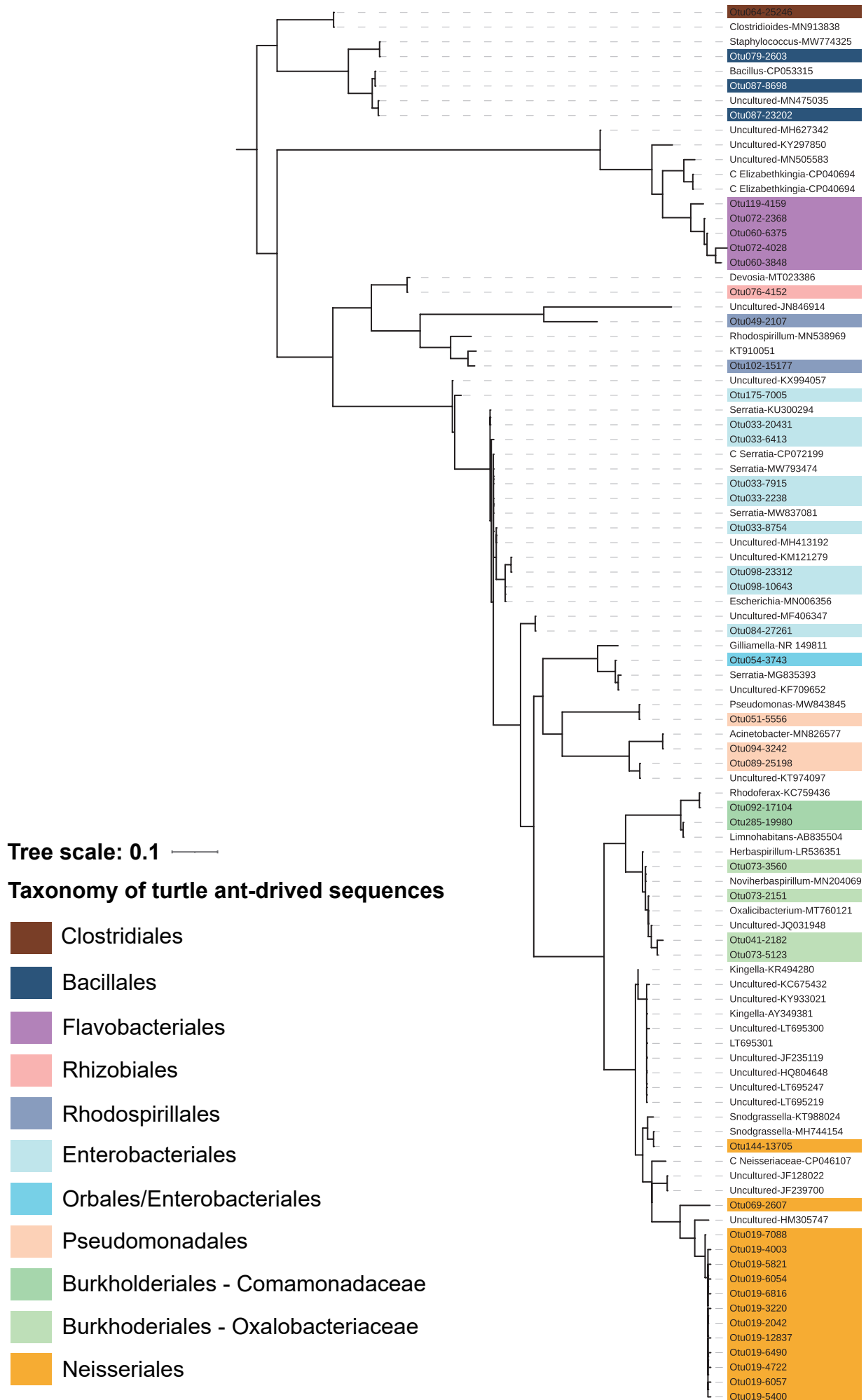


**Figure S3: Maximum likelihood phylogenetic analysis of ambiguously classifying Pseudomonadales OTU046 (Analysis #6).** To ascertain whether OTU046 clustered with cephalotine ant-specialized bacteria from the Pseudomonadales, we first selected all unique 16S rRNA sequences from this OTU with relative abundance  $>0.002994$  in at least two sequence libraries, or  $>0.05$  in one library. With these sequences as queries, we performed BLASTn searches against the NCBI nr database, and against the genera from the Pseudomonadaceae family found to be highly related to OTU046 based on initial BLAST results. In addition to our inclusion of top BLASTn hits, our phylogenetic analysis also included unique sequences from OTUs more confidently assigning as cephalotine-specialized Pseudomonadales through prior BLASTn searches. After sequence alignment, we applied a Maximum Likelihood framework in Seaview, with the PhyML method and 100 bootstrap replicates. Phylogenies were visualized using the interactive Tree of Life website, and were inspected for monophyletic, cephalotine ant-specific clades showing distributions across multiple *Cephalotes* and/or *Procryptocerus* ant species.





**Figure S4: Maximum likelihood phylogenetic analysis of ambiguously classifying Sphingobacteriales OTU116 (Analysis #7).** To characterize the evolution of an ambiguously classifying Sphingobacteriales OTU with a top BLASTn hit showing only ~85% sequence identity to a specialized Sphingobacteriales symbiont from *Cephalotes* ants, we performed additional phylogenetic analyses. First, we selected all unique 16S rRNA sequences from this OTU showing relative abundance >0.002994 in at least two sequence libraries, or >0.05 in one library – this equated to exactly n=1 sequence. We used this sequence for a BLASTn search against the NCBI nr database. We also included 1) related OTUs from the cephalotine-associated Sphingobacteriales identified in the present study (OTUs 038, 083, and 107); 2) top BLASTn hits to these 3 putatively related, cephalotine-specific Sphingobacteriales OTUs; and 3) relatives identified via BLASTn results targeting a range of genera and families in the Sphingobacteriales. After sequence alignment, we applied a Maximum Likelihood framework in Seaview, with the PhyML method and 100 bootstrap replicates. Phylogenies were visualized using the interactive Tree of Life website, and were inspected for monophyletic, cephalotine ant-specific clades showing distributions across multiple *Cephalotes* and/or *Procryptocerus* ant species.



**Figure S5: Maximum likelihood phylogenetic analysis of other sequences with high abundance, and their relatives (Analysis #8).** For this final phylogenetic analysis aimed at identifying specialized bacterial symbionts of cephalotine ants, we focused on subset of the 154 OTUs exceeding the required abundance threshold in ant sequence libraries – i.e. the most abundant left-overs, not analyzed in Analysis #1-7 above, with OTU IDs falling between #s 001 –100. These sequences did not belong to an OTU assigning as a specialized core cephalotine bacterium via BLAST, and had not been studied extensively in previously published phylogenetic analyses on ant-associated bacteria (e.g. *Wolbachia*, Entomoplasmatales – Russell et al. 2009; Funaro et al. 2011). This amounted to 19 selected OTUs, hailing from 8 orders. To maximize our chances of finding cephalotine-specific lineages we analyzed unique sequences from all OTUs in these 8 orders (i.e. not just the 19 selected OTUs) if they exhibited >0.002994 relative abundance in at least 2 *Cephalotes* sequence libraries, or if they were present in >0.05 relative abundance in at least one library. Top BLASTn hits were downloaded for each of these sequences. After sequence alignment, we applied a Maximum Likelihood framework in Seaview, with the PhyML method and 100 bootstrap replicates. Phylogenies were visualized using the interactive Tree of Life website, and were inspected for monophyletic, cephalotine ant-specific clades showing distributions across multiple *Cephalotes* and/or *Procryptocerus* ant species.

## Supplementary references

1. Hu Y, Sanders JG, Lukasik P, D'Amelio CL, Millar JS, Vann DR, Lan Y, Newton JA, Schotanus M, Kronauer DJC, et al: **Herbivorous turtle ants obtain essential nutrients from a conserved nitrogen-recycling gut microbiome.** *Nat Commun* 2018, **9**:964.
2. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, Tiedje JM: **The Ribosomal Database Project: improved alignments and new tools for rRNA analysis.** *Nucleic Acids Res* 2009, **37**:D141-D145.
3. Miller MA, Pfeiffer W, Schwartz T: **Creating the CIPRES Science Gateway for inference of large phylogenetic trees.** *2010 Gateway Computing Environments Workshop (GCE)* 2010:1-8.
4. Stamatakis A: **RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies.** *Bioinformatics* 2014, **30**:1312-1313.
5. Letunic I, Bork P: **Interactive Tree Of Life (iTOL) v4: recent updates and new developments.** *Nucleic Acids Res* 2019, **47**:W256-W259.
6. Russell JA, Goldman-Huertas B, Moreau CS, Baldo L, Stahlhut JK, Werren JH, Pierce NE: **Specialization and geographic isolation among *Wolbachia* symbionts from ants and lycaenid butterflies.** *Evolution* 2009, **63**:624-640.
7. Funaro CF, Kronauer DJ, Moreau CS, Goldman-Huertas B, Pierce NE, Russell JA: **Army ants harbor a host-specific clade of Entomoplasmatales bacteria.** *Appl Environ Microbiol* 2011, **77**:346-350.