

1 Large-scale comparative genomics unravels great genomic  
2 diversity across the *Rickettsia* and *Ca. Megaira* genera and  
3 identifies Torix group as an evolutionarily distinct clade.  
4

5 Helen R Davison<sup>1</sup>, Jack Pilgrim<sup>1</sup>, Nicky Wybouw<sup>2</sup>, Joseph Parker<sup>3</sup>, Stacy Pirro<sup>4</sup>, Simon Hunter-Barnett<sup>1</sup>, Paul M  
6 Campbell<sup>5</sup>, Frances Blow<sup>6</sup>, Alistair C Darby<sup>1</sup>, Gregory D D Hurst<sup>1</sup> and Stefanos Siozios<sup>1\*</sup>

7 **Affiliations**

8 1. Department of Evolution, Ecology and Behaviour, Institute of Infection, Veterinary and Ecological sciences,  
9 University of Liverpool, Liverpool, L69 7ZB, United Kingdom

10 2. Terrestrial Ecology Unit, Department of Biology, Faculty of Sciences, Ghent University, Ghent, Belgium

11 3. Division of Biology and Biological Engineering, California Institute of Technology, 1200 E California  
12 Boulevard, Pasadena, CA 91125, USA

13 4. Iridian Genomes, Bethesda, MD, USA

14 5. School of Health and Life Sciences, Faculty of Biology Medicine and Health, the University of Manchester,  
15 Manchester, United Kingdom

16 6. Center for Genomics and Systems Biology, Department of Biology, New York University, New York, New  
17 York, USA

18 \* Corresponding author

19

## 20 Abstract

21 *Rickettsia* are intracellular bacteria originally described as arthropod borne pathogens that are  
22 emerging as a diverse group of often biologically important, non-pathogenic symbionts of  
23 invertebrates and microeukaryotes. However, sparse genomic resources for symbiotic strains and  
24 for the sister genus (*Candidatus* Megaira) inhibit our understanding of *Rickettsia* evolution and  
25 biology. Here, we present the first closed genomes of *Ca.* Megaira from an alga (*Mesostigma*  
26 *viride*), and Torix *Rickettsia* from midge (*Culicoides impunctatus*) and bed bug (*Cimex lectularius*)  
27 hosts. Additionally, we sequenced and constructed draft genomes for *Ca.* Megaira from another  
28 alga (*Carteria cerasiformis*), Transitional group *Rickettsia* from tsetse fly (*Glossina morsitans*  
29 *submorsitans*), and Torix *Rickettsia* from a spider mite (*Bryobia graminum*). We further extract 22  
30 draft genomes from arthropod genome sequencing projects, including 1 Adalia, 4 Transitional, 1  
31 Spotted Fever, 7 Torix, 7 Belli and the first Rhyzobius and Meloidae *Rickettsia* group genomes. We  
32 used new and existing *Rickettsia* genomes to estimate the phylogeny and metabolic potential  
33 across groups and reveal transitions in genomic properties. These data reveal Torix as unique  
34 amongst currently described *Rickettsia*, with highly distinct and diverse accessory genomes. We  
35 confirm the presence of a third subclade of Torix, previously only known from gene marker  
36 sequences. Further, Torix share an intact pentose phosphate pathway with *Ca.* Megaira, not  
37 observed in other *Rickettsia*. Considering the distinctness and diversity of Torix, we propose that  
38 the group be named *Candidatus* Tisiphia. The wide host range of *Ca.* Tisiphia symbionts  
39 necessitates onward research to understand the biological and physiological bases of *Ca.* Tisiphia-  
40 host interactions.

41

## 42 Importance statement

43

44 Members of the genus *Rickettsia* were originally identified as causative agents of mammalian vector-borne  
45 disease. In the last 25 years we have recognised that many *Rickettsia* are arthropod symbionts, and sit  
46 alongside a sister taxon, *Ca. Megaira*, which are symbiotic associates of microeukaryotes. The lack of  
47 genomic information for symbiotic strains affects our ability to determine the evolutionary relationships  
48 between strains and understand the biological underpinnings of the different symbioses. We clarify these  
49 relationships by assembling 26 genomes of *Rickettsia* from understudied groups, and the first two *Ca.*  
50 *Megaira*, from various insects and microeukaryotes. Of note, the accessory genome diversity and broad host  
51 range of Torix *Rickettsia* parallels all other *Rickettsia* combined. This diversity, alongside the breadth of host  
52 species, make the Torix clade an important hidden player in invertebrate biology and physiology. We argue  
53 this clade should be given its own genus status, for which we propose *Ca. Tisiphia*.

## 54 Introduction

55 Symbiotic bacteria are vital to the function of most living eukaryotes, including microeukaryotes,  
56 fungi, plants, and animals (Boettcher et al., 1996; Clay et al., 2005; Douglas, 2011; Fujishima &  
57 Kodama, 2012). The symbioses formed are often functionally important to the host with effects  
58 ranging from mutualistic to detrimental. Mutualistic symbionts may provide benefits through the  
59 biosynthesis of metabolites, or by protecting their hosts against pathogens and parasitoids (Hendry  
60 et al., 2014; Oliver et al., 2010). Meanwhile parasitic symbionts can be detrimental to the host due  
61 to resource exploitation or through reproductive manipulation that favours its own transmission  
62 over the host's (Engelstädter & Hurst, 2009; Leclair et al., 2017). Across these different symbiotic  
63 relationships, symbionts are often important determinants of host ecology and evolution.

64 The Rickettsiales (Alphaproteobacteria) represent an order of obligate intracellular bacteria that  
65 form symbioses with a variety of eukaryotes (Weinert et al., 2015). Within Rickettsiales, the family  
66 Rickettsiaceae represent a diverse collection of bacteria that infect a wide range of eukaryotic hosts  
67 and can act as symbionts, parasites, and pathogens. Perhaps the best-known clade of  
68 Rickettsiaceae is the genus *Rickettsia*, which was initially described as the cause of spotted fever  
69 and other rickettsioses in vertebrates that are transmitted by ticks, lice, fleas and mites (Angelakis  
70 & Raoult, 2017).

71 *Rickettsia* have been increasingly recognised as heritable arthropod symbionts. Since the first  
72 description of a maternally inherited male-killer in ladybirds (Werren et al., 1994), we now know  
73 that heritable *Rickettsia* are common in arthropods (Pilgrim et al., 2021; Weinert et al., 2009).  
74 Further, *Rickettsia*-host symbioses are diverse, with the symbiont capable of reproductive  
75 manipulation, nutritional and protective symbiosis, as well as influencing thermotolerance and  
76 pesticide susceptibility (Bodnar et al., 2018; Brumin et al., 2011; Chiel et al., 2009; Giorgini et al.,  
77 2010; Hurst et al., 1994; Kongsedalov et al., 2008; Łukasik et al., 2013).

78 Our understanding of the evolution and diversity of the genus *Rickettsia* and its allies has increased  
79 in recent years. Weinert et al. (2009) defined 13 different groups of *Rickettsia* with two early  
80 branching clades that appeared genetically distant from other members of the genus. The first of  
81 these was defined from a symbiont of *Hydra* and was named the Hydra group *Rickettsia*, which has  
82 since been assigned its own genus status, *Candidatus Megaira* (Schrallhammer et al., 2013). *Ca.*  
83 *Megaira* forms a sister clade to *Rickettsia* and is common in ciliate protists, amoebae, chlorophyte  
84 and streptophyte algae, and cnidarians (Lanzoni et al., 2019). Members of this clade are found in

85 hosts from aquatic, marine and soil habitats which include model organisms (e.g., *Paramecium*,  
86 *Volvox*) and economically important vertebrate parasites (e.g., *Ichthyophthirius multifiliis*, the ciliate  
87 that causes white spot disease in fish) (Lanzoni et al., 2019). Whilst symbioses between *Ca. Megaira*  
88 and microeukaryotes are pervasive, there is no complete publicly available genome and the impact  
89 of these symbioses on the host are poorly understood.

90 A second early branching clade was first described from *Torix tagoi* leeches and is commonly coined  
91 *Torix Rickettsia* (Kikuchi & Fukatsu, 2005). Symbionts in the *Torix* clade have since been found in a  
92 wide range of invertebrate hosts from midges to freshwater snails, and in a fish-parasitic amoeba  
93 (Pilgrim et al., 2021). The documented diversity of hosts is wider than other *Rickettsia* groups,  
94 which are to date only found in arthropods and their associated vertebrate or plant hosts (Weinert  
95 et al., 2009). *Torix* clade *Rickettsia* are known to be heritable symbionts, but their impact on host  
96 biology is poorly understood, despite the economic and medical importance of several hosts (inc.  
97 bed bugs, black flies, and biting midges). Rare studies have described the potential effects on the  
98 host, which include: larger body size in leeches (Kikuchi & Fukatsu, 2005); a small negative effect on  
99 growth rate and reproduction in bed bugs (Thongprem et al., 2020); and an association with  
100 parthenogenesis in *Empoasca* Leafhoppers (Aguin-Pombo et al., 2021).

101 Current data seems to suggest an emerging macroevolutionary scenario where the members of  
102 *Rickettsia*-*Megaira* clade originated as symbionts of microeukaryotes, before diversifying to infect  
103 invertebrate symbionts. The *Torix Rickettsia* retained a broad range of hosts from microeukaryotes  
104 to arthropods. The remaining members of the genus *Rickettsia* evolved to be arthropod heritable  
105 symbionts and vector-borne pathogens. However, a lack of genomic and functional information for  
106 symbiotic clades limits our understanding of evolutionary transitions within *Rickettsia* and its sister  
107 groups. No *Ca. Megaira* genome sequences are currently publicly available and of the 165 *Rickettsia*  
108 genome assemblies available on the NCBI (as of 29/04/21), only two derive from the *Torix* clade  
109 and these are both draft genomes. In addition, dedicated heritable symbiont clades of *Rickettsia*,  
110 such as the *Rhyzobius* group, have no available genomic data, and there is a single representative  
111 for the *Adalia* clade. Despite the likelihood that heritable symbiosis with microeukaryotes and  
112 invertebrates was the ancestral state for this group of intracellular bacteria, available genomic  
113 resources are heavily skewed towards pathogens of vertebrates.

114 In this study we establish a richer base of genomic information for heritable symbiont *Rickettsia*  
115 and *Ca. Megaira*, then use these resources to clarify the evolution of these groups. We broaden

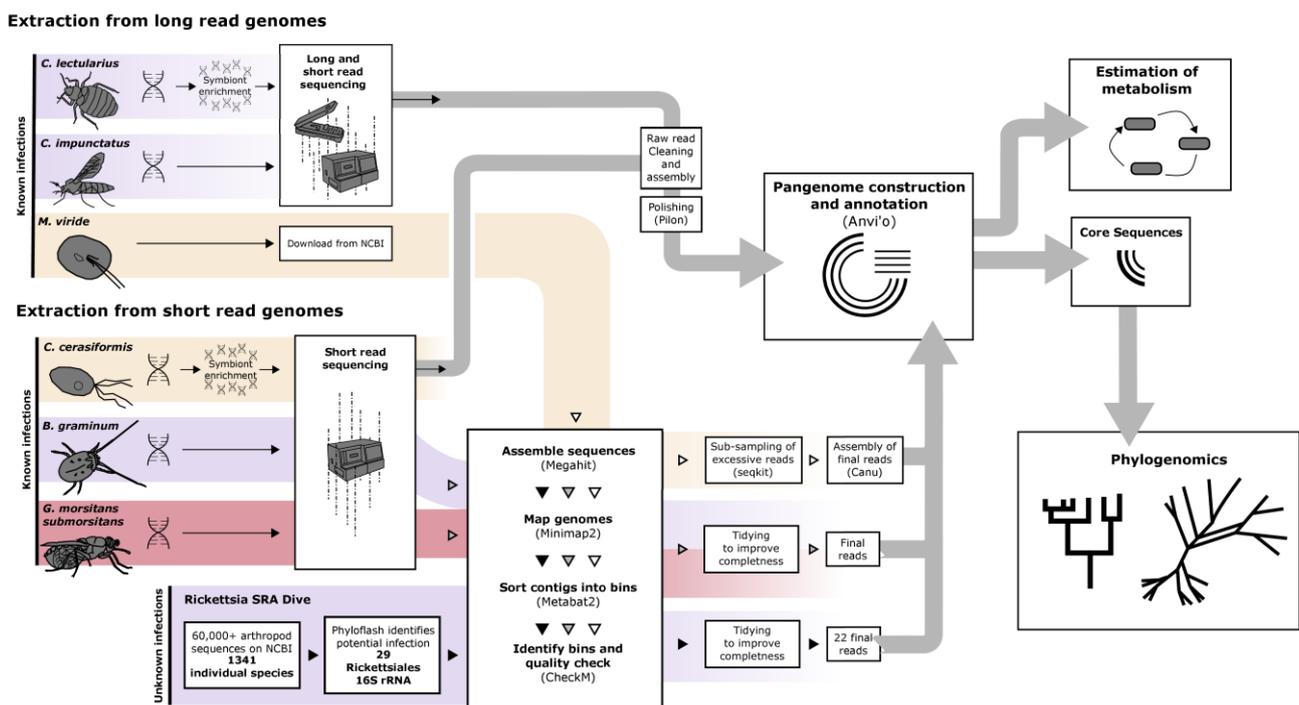
116 available genomic data through a combination of targeted sequencing of strains without complete  
117 genomes, and metagenomic assembly of *Rickettsia* strains from arthropod genome projects. We  
118 report the first closed circular genome of a *Ca. Megaira* symbiont from a streptophyte alga  
119 (*Mesostigma viride*) and provide a draft genome for a second *Ca. Megaira* from a chlorophyte  
120 (*Carteria cerasiformis*). In addition, we present the first complete genomes of two Torix *Rickettsia*  
121 from a midge (*Culicoides impunctatus*) and a bed bug (*Cimex lectularius*) as well as a draft genome  
122 for *Rickettsia* from a tsetse fly (*Glossina morsitans submorsitans*, an important vector species), and  
123 a new strain from a spider mite (*Bryobia graminum*). A metagenomic approach established a  
124 further 22 draft genomes for insect symbiotic strains, including the first Rhyzobius and Meloidae  
125 group draft genomes. We utilize these to carry out pangenomic, phylogenomic and metabolic  
126 analyses of our extracted genome assemblies, with comparisons to existing *Rickettsia*.

127

## 128 Methods

### 129 Genomic data collection and construction

130 We employed two different workflows to assemble genomes for *Ca. Megaira* and *Rickettsia*  
131 symbionts (Figure 1). A) Targeted sequencing and assembly of focal *Ca. Megaira* and Torix  
132 *Rickettsia*. B) Assembly from SRA deposits of *Ca. Megaira* from *Mesostigma viride* NIES296 and the  
133 29 arthropods identified in Pilgrim et al (2021) that potentially harbour *Rickettsia*. These were  
134 analysed alongside previously assembled genomes from the genus *Rickettsia*, and the outgroup  
135 taxon *Orientia tsutsugamushi*.



136 Figure 1. Workflow diagram for extraction, assembly and analyses performed in this study. Purple highlights Torix *Rickettsia* and  
137 orange highlights *Ca. Megaira* and red highlights Transitional *Rickettsia*. A full resolution version can be found here:  
138 <https://doi.org/10.6084/m9.figshare.15081975>.

139 DNA preparation, sequencing strategies and symbiont assembly methodologies varied between  
140 species. Methods are summarized in Figure 1 and detailed in supplementary material  
141 <https://doi.org/10.6084/m9.figshare.14865582>. The exact pipeline used to assemble genomes from  
142 Short Read Archive (SRA) data can be found here: <https://figshare.com/s/d1155765b523a6379443>.

### 143 Sample collection for targeted genome assembly

144 *Cimex lectularius* were acquired from the 'S1' isofemale colony maintained at the University of  
145 Bayreuth described in Thongprem et al (2020). *Culicoides impunctatus* females were collected from

146 a wild population in Kinlochleven, Scotland (56° 42' 50.7"N 4° 57' 34.9"W) on the evenings of the  
147 2nd and 3rd September 2020 by aspiration. *Carteria cerasiformis* strain NIES 425 was obtained from  
148 the Microbial Culture Collection at the National Institute for Environmental Studies, Japan. The  
149 *Glossinia morsitans submorsitans* specimen Gms8 was collected in Burkina Faso in 2010 and  
150 *Rickettsia* infection was present alongside other symbionts as described in Doudoumis et al. (2017).  
151 The assembly itself is a result of later thesis work (Blow, 2017).

152 A *Bryobia* mite community was sampled from herbaceous vegetation in Turku, Finland. The  
153 Moomin isofemale line was established by isolating a single adult female and was maintained on  
154 detached leaves of *Phaseolus vulgaris* L. cv Speedy at 25 °C, 60 % RH, and a 16:8 light:dark  
155 photoperiod. The Moomin spider mite line was morphologically identified as *Bryobia graminum* by  
156 Prof Eddie A. Ueckermann (North-West University).

#### 157 *Previously published Rickettsia genomes*

158 A total of 86 published *Rickettsia* genomes, and one genome from *Orientia tsutsugamushi* were  
159 retrieved from the European Nucleotide Archive and assessed with CheckM v1.0.13 (Parks et al.,  
160 2015). Inclusion criteria for genomes were high completeness (CheckM > 90%), low contamination  
161 (CheckM < 2%) and low strain heterogeneity (Check M < 50%) except in the case of Adalia for which  
162 there is only one genome (87.6% completeness). Filtering identified 76 high quality *Rickettsia*  
163 genomes that were used in all subsequent analyses (S1  
164 <https://figshare.com/s/198c88c6e3ea5553192e>).

#### 165 *Genome content comparison and pangenome construction*

166 Anvi'o 7 (Eren et al., 2021) was used to construct a pangenome for *Rickettsia*. Included in this were  
167 the 22 MAGs retrieved from SRA data, 2 *Ca. Megaira* genomes and 4 targeted Torix *Rickettsia*  
168 genomes, and one transitional group *Rickettsia* genome acquired in this study. To these were  
169 added the 76 published and 1 *Orientia* described above, giving a total of 104 genomes. Individual  
170 Anvi'o genome databases were additionally annotated with HMMER, KofamKOALA, and NCBI COG  
171 profiles (Aramaki et al., 2020; Eddy, 2011; Galperin et al., 2021). For the pangenome itself,  
172 orthologs were identified with NCBI blast, mcl inflation was set to 2, and minbit was 0.5. Genomes  
173 were arranged according to cluster presence absence and average nucleotide sequence identity  
174 was calculated using pyANI (Pritchard et al., 2016). See  
175 <https://figshare.com/s/d1155765b523a6379443> for the exact code used in this section.

176 KofamKOALA annotation (Aramaki et al., 2020) in Anvi-o 7 was used to estimate completeness of  
177 metabolic pathways. Then Pheatmap (Kolde, 2019) in R 3.4.4 (R Core Team, 2020) was used to  
178 produce heatmaps of metabolic potential (figure 7). Annotations for function and *Rickettsia* group  
179 were added *post hoc* in Inkscape.

180 The biotin operon found in the genome *Rhizobium Rickettsia* Oopac6 was identified from metabolic  
181 prediction (figure 7). To confirm Oopac6 carries a complete biotin pathway that shares ancestry  
182 with the existing *Rickettsia* biotin operon, Oopac6 biotin was compared to biotin pathways from  
183 five other related symbionts: *Cardinium*, *Lawsonia*, *Buchnera aphidicola*, *Rickettsia buchneri*, and  
184 *Wolbachia* (Seemann, 2014). Clinker (Gilchrist & Chooi, 2021) with default options was used to  
185 compare and visualise the similarity of genes within the biotin operon region of all 6 bacteria.

186 All generated draft and complete reference genomes were annotated using the NCBI's Prokaryotic  
187 Genome Annotation Pipeline (PGAP) (Tatusova et al., 2016). Secondary metabolite biosynthetic  
188 gene clusters were identified using AntiSMASH version 6.0 (Blin et al., 2021) along with Norine  
189 (Flissi et al., 2019) which searched for similarities to predicted non-ribosomal peptides.

190 Functional enrichment analyses between the main *Rickettsia* clade and the Torix – *Ca. Megaira*  
191 clades were performed using the Anvi'o program `anvi-get-enriched-functions-per-pan-group` and  
192 the "COG\_FUNCTION" as annotation source. A gene cluster presence - absence table was exported  
193 using the command "`anvi-export-tables`". This was used to create an UpSet plot using the R package  
194 `ComplexUpset` (Krassowski et al., 2020; Lex et al., 2014) to visualize unique and shared gene  
195 clusters between different *Rickettsia* groups. A gene cluster was considered unique to a specified  
196 *Rickettsia* group when it was present in at least one genome belonging to that group. Gene cluster  
197 accumulation curves were performed for the pan-, core- and unique-genomes based on the same  
198 presence-absence matrix using a custom-made R script (Siozios, 2021). In each case the cumulative  
199 number of gene clusters were computed based on randomly sampled genomes using 100  
200 permutations. The analysis was performed separately for each of the five major *Rickettsia* groups as  
201 well as the complete *Rickettsia* dataset. Curves were plotted using the `ggplot2` R package  
202 (Wickham, 2016).

203 All information on extra genomes can be found at <https://doi.org/10.6084/m9.figshare.14865582>  
204 and the code pipeline employed can be found at <https://figshare.com/s/d1155765b523a6379443>.

## 205 Phylogeny, Network, and recombination

206 The single-copy core of all 104 genomes was identified in Anvi'o 7 and is made up of 74 single-copy  
207 gene (SCG) clusters. Protein alignments from SCG were extracted and concatenated using the  
208 command "anvi-get-sequences-for-gene-clusters". Maximum likelihood phylogeny was constructed  
209 in IQ-TREE v2.1.2 (Nguyen et al., 2015). Additionally, 43 ribosomal proteins were identified through  
210 Anvi'o 7 to test phylogenomic relationships. These gene clusters were extracted from the  
211 pangenome and used for an independent phylogenetic analysis SUPPLEMENTARY FIG. The best  
212 model according to the Bayesian Information Criterion (BIC) was selected with Model Finder Plus  
213 (MFP) (Kalyaanamoorthy et al., 2017) as implemented in IQ-TREE; this was JTTDCMut+F+R6 for core  
214 gene clusters and JTTDCMut+F+R3 for ribosomal proteins. Both models were run with Ultrafast  
215 Bootstrapping (1000 UF bootstraps) (Hoang et al., 2018) with *Orientia tsutsugamushi* as the  
216 outgroup.

217 The taxonomic placement of Oopac6, Ppec13 and Dallo3 genomes within the Rhyzobius, Meloidae  
218 and Belli groups respectively were confirmed in a smaller phylogenetic analysis, performed as  
219 detailed in (Pilgrim et al. 2021) using reference MLST sequences (*gltA*, 16s rRNA, 17kDa, *COI*) from  
220 other previously identified *Rickettsia* profiles (S1 <https://figshare.com/s/198c88c6e3ea5553192e>). The  
221 selected models used in the concatenated partition scheme were as follows: 16S rRNA:  
222 TIM3e+I+G4; 17KDa: GTR+F+I+G4; COI: TPM3u+F+I+G4; *gltA*: K3Pu+F+I+G4a.

223 A nearest neighbour network was produced for core gene sets with default settings in Splitstree4 to  
224 further assess distances and relationships between *Rickettsia*, *Ca. Megaira* and Torix clades. All  
225 annotation was added post hoc in Inkscape. Furthermore, recombination signals were examined by  
226 applying the Pairwise Homoplasy Index (PHI) test to the DNA sequence of each core gene cluster  
227 extracted with Anvi'o-7. DNA sequences were aligned with MUSCLE (Edgar, 2004) and PHI scores  
228 calculated for each of the 74 core gene cluster with PhiPack (Bruen et al., 2006).

229 The taxonomic identity for new and newly expanded groups was established with GTDB-Tk  
230 (Chaumeil et al., 2020) to support the designation of new taxa through phylogenetic comparison of  
231 marker genes against an online reference database.

## 232 Results and Discussion

233 We have expanded the available genomic data for several *Rickettsia* groups through a combination  
234 of draft and complete genome assembly. This includes an eight-fold increase in available Torix-

235 group genomes, and the first available genomes for Meloidae and Rhyzobius groups. We further  
236 report the first reference genomes for *Ca. Megaira*.

### 237 Complete and closed reference genomes for *Torix Rickettsia* and *Ca. Megaira*

238 The use of long-read sequencing technologies produced the first complete genomes for two  
239 subclades of the *Torix* group (RiCimp-limoniae, RiClec-leech). Sequencing depth of the *Rickettsia*  
240 genomes from *C. impunctatus* (RiCimp) and *C. lectularius* (RiClec) were 18X and 52X respectively.  
241 The RiCimp genome provides the first evidence of plasmids in the *Torix* group (pRiCimp001 and  
242 pRiCimp002). In addition, we assembled the first complete closed reference genome of *Ca. Megaira*  
243 from *Mestostigma viride* (MegNEIS296) from previously published genome sequencing efforts.

244 General features of both genomes are consistent with previous genomic studies of the *Torix* group  
245 (Table 1). A single full set of rRNAs (16S, 5S and 23S) and a GC content of ~33% was observed.

246 Notably, the two complete *Torix* group genomes show a distinct lack of synteny (see S2

247 <https://doi.org/10.6084/m9.figshare.14866263>), a genomic feature that is compatible with our

248 phylogenetic analyses that placed these two lineages in different subclades (leech/limoniae)

249 (figures 2 and 3). Of note within the closed reference genomes MegNEIS296 and RiCimp, is the

250 presence of a putative non-ribosomal peptide synthetase (NRPS) and a hybrid non-ribosomal

251 peptide/polyketide synthetase (NRPS/PKS) respectively (see S3

252 <https://doi.org/10.6084/m9.figshare.14865570>). Although, the exact products of these putative

253 pathways are uncertain, *in silico* prediction by Norine suggests close similarity with both cytotoxic

254 and antimicrobial peptides hinting at a potential defensive role (see S3

255 <https://doi.org/10.6084/m9.figshare.14865570>). A hybrid NRPS/PKS cluster has previously been

256 reported in *Rickettsia buchneri* on a mobile genetic element, providing potential routes for

257 horizontal transmission (Hagen et al., 2018). In addition, putative toxin-antitoxin systems similar to

258 the one associated with cytoplasmic incompatibility in *Wolbachia* have recently been observed on

259 the plasmid of *Rickettsia felis* in a parthenogenetic booklouse (Gillespie et al., 2015, 2018). Toxin-

260 endotoxin systems are thought to be part of an extensive bacterial mobilome network associated

261 with reproductive parasitism (Gillespie et al., 2018). A BLAST search found a very similar protein in

262 Oopac6 to the putative large pLbAR toxin found in *R. felis* (88% aa identity), and a more distantly

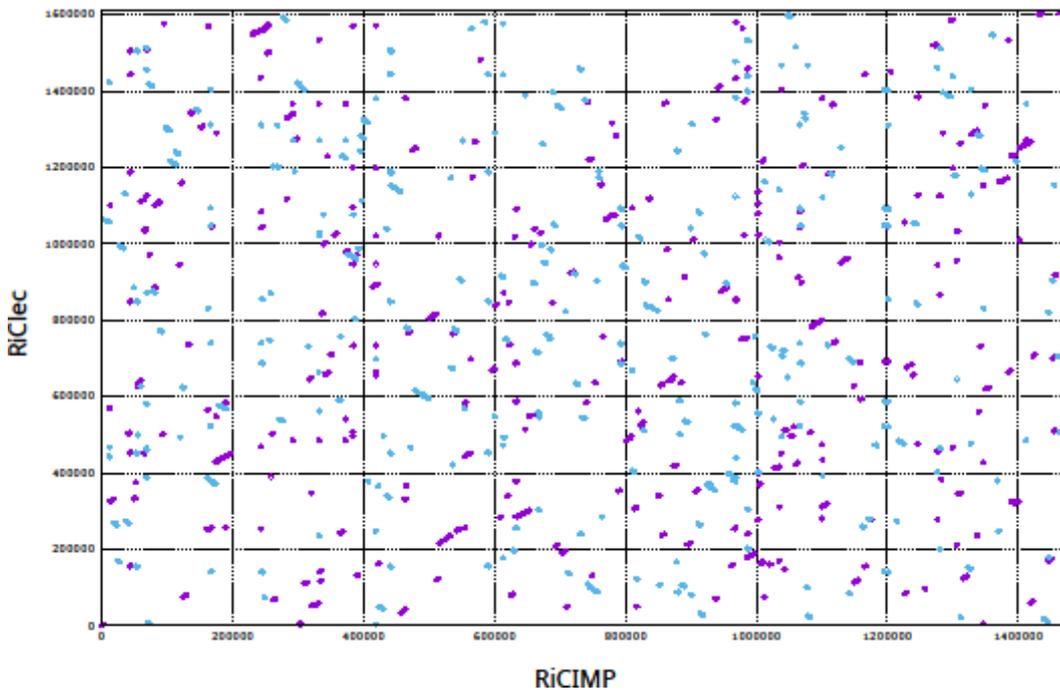
263 related protein in the *C. impunctatus* plasmid (25% aa identity).

264

265 Table 1. Summary of the complete *Ca. Megaira* and *Torix Rickettsia* genomes

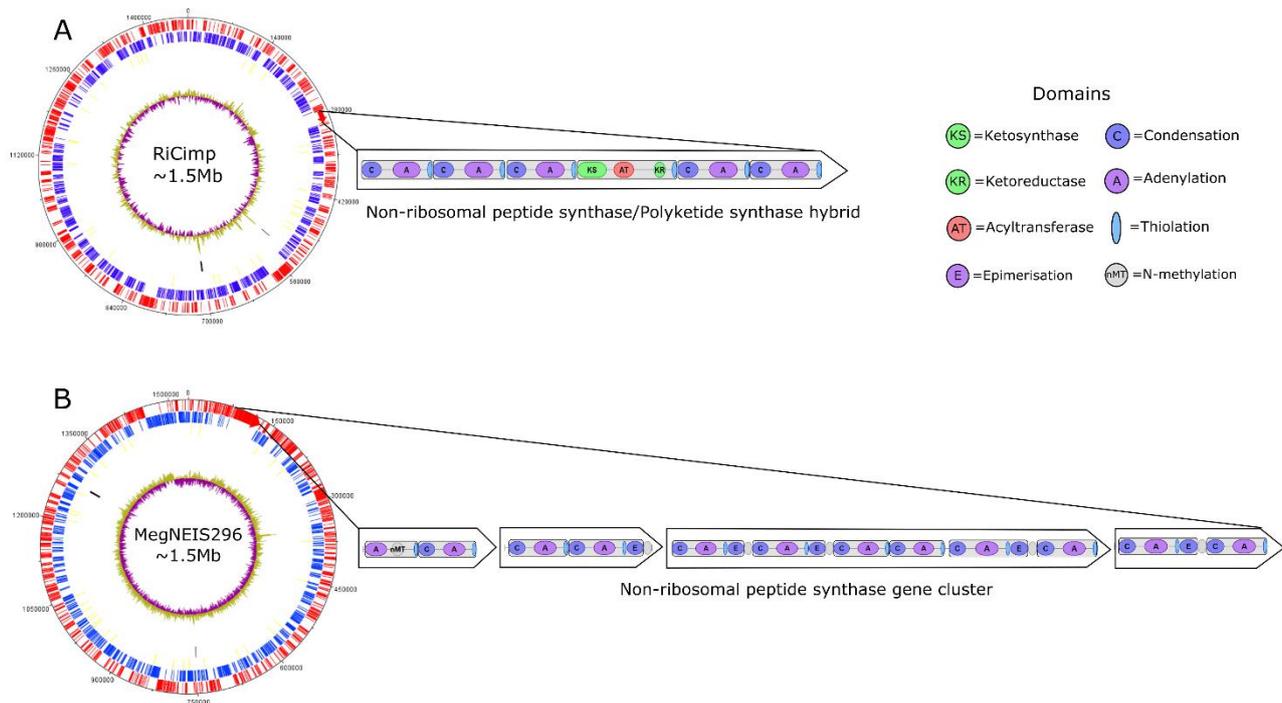
Group	<i>Ca. Megaira</i>	<i>Torix Rickettsia</i>	
Sub-group		Leech	Limoniae
Strain Name	MegNIES296	RiCimp	RiClec
Symbiont genome accession	CP084576-CP084577	CP084573-CP084575	CP084572
Host	Mesostigma viride NIES-296	Culicoides impunctatus	Cimex lectularius
Raw reads accession	SRR8439255, SRX5120346	SRR16018514, SRR16018513	SRR16018512, SRR16018511
Total nucleotides	1,532,409	1,566,468	1,611,726
Chromosome size (bp)	1,448,425	1,469,631	1,611,726
Plasmids	1 (83,984 bp)	2 (77550bp + 19287bp)	None
GC content (%)	33.9	32.9	32.8
Number of CDS	1,359	1,397	1,544
Avg. CDS length (bp)	998	900	874
Coding density (%)	88.5	86	84
rRNAs	3	3	3
tRNAs	34	34	35

266



267

268 S2 Whole genome alignment between the complete *Torix limoniae* (RiCIMP) and *Torix Leech* (RiClec) genomes reveals complete lack  
 269 of synteny. Magenta represents forward matches and blue reverse matches <https://doi.org/10.6084/m9.figshare.14866263>.



270

271 *S3. The circular chromosomes of A) a Torix group Rickettsia (RiCimp) and B) a Ca. Megaira sp. (MegNEIS296). From outside to in, the*  
 272 *circles represent: forward CDSs (Red), Reverse CDSs (blue), tRNAs (yellow) rRNAs (black), and GC content (green and magenta).*  
 273 *Highlighted are the predicted modules formed by non-ribosomal peptide synthase genes (domains) that define individual amino acids*  
 274 *in the synthesised peptide and show the catalytic domains within modules <https://doi.org/10.6084/m9.figshare.14865570>.*

275 **Sequencing and *de novo* assembly of other *Rickettsia* and *Ca. Megaira* genomes.**

276 Our direct sequencing efforts enabled assembly of draft genomes for a second *Ca. Megaira* strain  
 277 from the alga *Carteria cerasiformis*, and for *Rickettsia* associated with tsetse flies and *Bryobia* spider  
 278 mites. The Transitional *Rickettsia* from a wild caught Tsetse fly, RiTSETSE, is a potentially chimeric  
 279 assembly since we identified an excess of biallelic sites when the raw Illumina reads were mapped  
 280 back to the assembly. It is also likely that RiTSETSE is not a heritable symbiont but comes from  
 281 transient infection from a recent blood meal.

282 From the SRA accessions, the metagenomic pipeline extracted 29 full symbiont genomes for  
 283 Rickettsiales across 24 host species. Five of 29 were identified as *Wolbachia* and discarded from  
 284 further analysis, one was a *Rickettsia* discarded for low quality, and another was a previously  
 285 assembled Torix *Rickettsia*, RiCNE (Pilgrim et al., 2017). Thus, 22 high quality *Rickettsia*  
 286 metagenomes were obtained from 21 host species. One beetle (SRR6004191) carried coinfecting  
 287 *Rickettsia* Lappe3 and Lappe4 (Table 2). The high-quality *Rickettsia* covered the Belli, Torix,  
 288 Transitional, Rhyzobius, Meloidae and Spotted Fever Groups (Table 2 and S1  
 289 <https://figshare.com/s/198c88c6e3ea5553192e>).

290 Beetles, particularly rove beetle (Staphylinidae) species, appear in this study as a possible hotspot  
291 of *Rickettsia* infection. *Rickettsia* has historically been commonly associated with beetles, including  
292 ladybird beetles (*Adalia bipunctata*), diving beetles (*Deronectes* sp.) and bark beetles (*Scolytinae*)  
293 (Hurst et al., 1994; K uchler et al., 2009; Perlman et al., 2006; Weinert et al., 2009; Zchori-Fein et al.,  
294 2006). Though a plausible and likely hotspot, this observation needs be approached with caution as  
295 this could be an artefact of skewed sampling efforts.

296 All genome metadata and source information can be found here  
297 <https://figshare.com/s/198c88c6e3ea5553192e>.

298

299  
300

Table 2. Brief summary of draft genomes generated during the current project and their associated hosts. Full metadata can be found in S1 <https://figshare.com/s/198c88c6e3ea5553192e>.

Strain	Bacteria Biosample Accession	Group	Number of contigs	Total length (bp)	Host name	Order
<b>Blapp1</b>	SAMN21822536	Belli	171	1266633	<i>Bembidion lapponicum</i>	Coleoptera
<b>Btrans1</b>	SAMN21822537	Belli	241	1417452	<i>Bembidion nr. transversale</i> OSAC:DRMaddison DNA3205	Coleoptera
<b>Choog2</b>	SAMN21822538	Belli	16	1357829	<i>Columbicola hoogstraali</i>	Phthiraptera
<b>Cmasu2</b>	SAMN21822539	Transitional	196	1295004	<i>Ceroptres masudai</i>	Hymenoptera
<b>Dallo3</b>	SAMN21822540	Belli	196	990679	<i>Diachasma alloeum</i>	Hymenoptera
<b>Drufa1</b>	SAMN21822541	Belli	14	1364611	<i>Degeeriella rufa</i>	Phthiraptera
<b>Earac4</b>	SAMN21822542	Transitional	96	1350066	<i>Ecitomorpha arachnoides</i>	Coleoptera
<b>Econn1</b>	SAMN21822543	Transitional	238	1070326	<i>Eriopis connexa</i>	Coleoptera
<b>Gbili3</b>	SAMN21822544	Torix limoniae	171	1188102	<i>Gnoriste bilineata</i>	Diptera
<b>Gdoso1</b>	SAMN21822545	Belli	34	1420758	<i>Graphium doson</i>	Lepidoptera
<b>Lappe3</b>	SAMN21822558	Torix limoniae	122	1368980	<i>Labidopullus appendiculatus</i>	Coleoptera
<b>Lappe4</b>	SAMN21822559	Torix limoniae	154	1332357	<i>Labidopullus appendiculatus</i>	Coleoptera
<b>MegCarteria</b>	SAMN21822546	Ca. Megaira	72	1298707	<i>Carteria cerasiformis</i>	Chlamydomonadales
<b>Ofont3</b>	SAMN21822560	Adalia	91	1529137	<i>Omalisus fontisbellaquei</i>	Coleoptera
<b>Oopac6</b>	SAMN21822548	Rhyzobius	181	1497231	<i>Oxypoda opaca</i>	Coleoptera
<b>Pante1</b>	SAMN21822549	Torix limoniae	70	1472610	<i>Pseudomimeciton antennatum</i>	Coleoptera
<b>Pfluc4</b>	SAMN21822550	Spotted Fever Group	7	1251895	<i>Proechinophthirus fluctus</i>	Phthiraptera
<b>Ppec13</b>	SAMN21822551	Belli	90	1426047	<i>Pyrocoelia pectoralis</i>	Coleoptera
<b>Psono2</b>	SAMN21822552	Torix limoniae	163	1492063	<i>Platyusa sonomae</i>	Coleoptera
<b>RiTSETSE</b>	SAMN21822553	Transitional	172	1451997	<i>Glossina morsitans submorsitans</i>	Diptera
<b>S2</b>	SAMN21822554	Torix limoniae	103	1251484	<i>Sericostoma</i>	Trichoptera
<b>Sanch3</b>	SAMN21822555	Belli	181	1487154	<i>Stiretrus anchorago</i>	Hemiptera
<b>Slati1</b>	SAMN21822556	Transitional	109	1301763	<i>Sceptobius lativentris</i>	Coleoptera
<b>Ssp4</b>	SAMN21822557	Torix limoniae	87	1231013	<i>Sericostoma sp.</i> HW-2014	Trichoptera
<b>moomin</b>	SAMN21822560	Torix moomin	204	1137559	<i>Bryobia graminum</i>	Trombidiformes

301

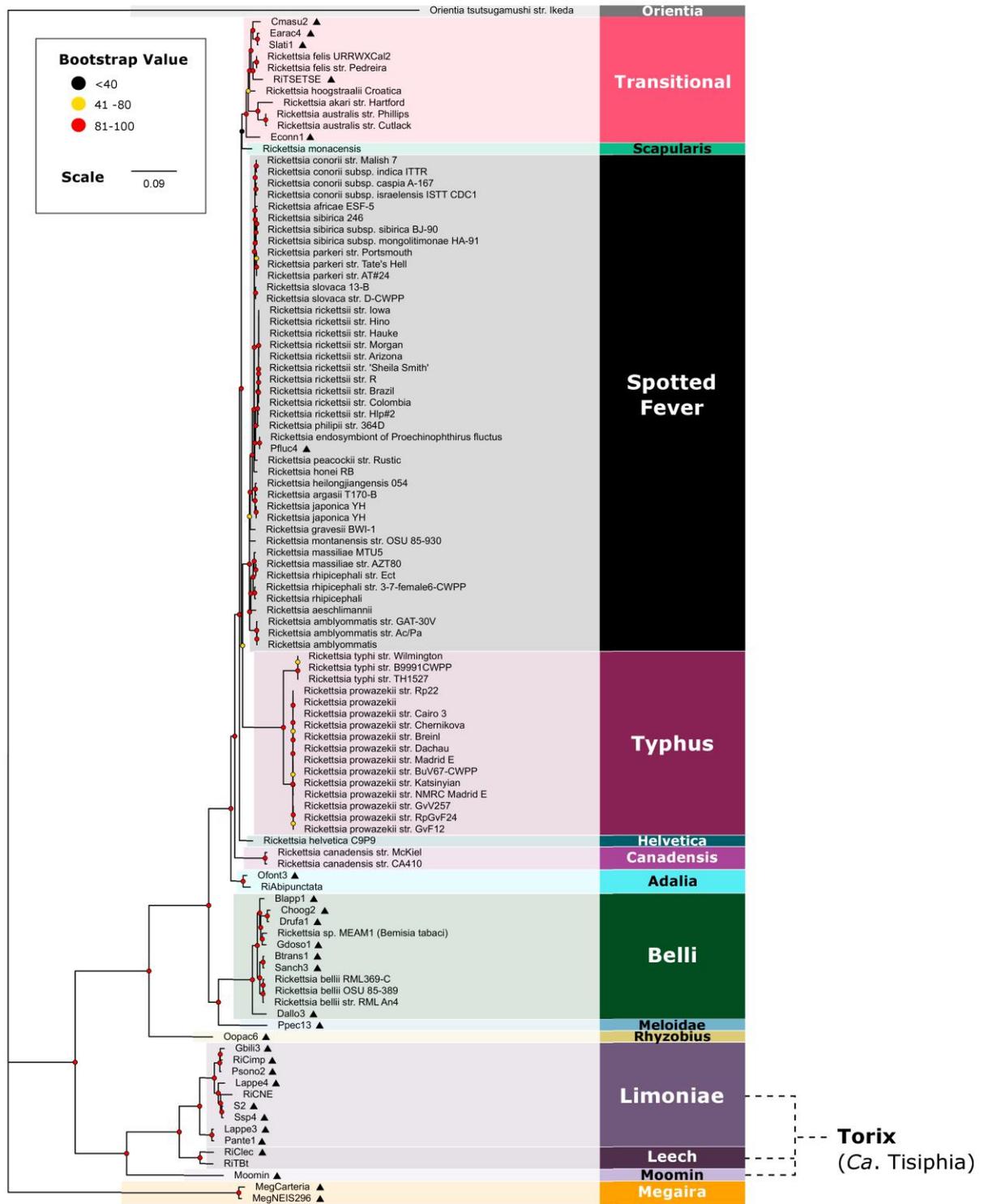
## 302 Phylogenomic analyses and taxonomic placement of newly assembled genomes

### 303 *Phylogeny, network, and recombination*

304 The network and phylogeny illustrate the distance of Torix from *Ca. Megaira* and other *Rickettsia*,  
305 along with an extremely high level of within-group diversity in Torix compared to any other group  
306 (Figures 2 and 3). The phylogenies generated using core genomes are consistent with previously  
307 identified *Rickettsia* and host associations using more limited genetic markers. For instance, Pfluc4  
308 from *Proechinophthirus fluctus* lice is grouped on the same branch as a previously sequenced  
309 *Rickettsia* from a different individual of *P. fluctus*. Four of 22 genomes from the SRA screen are  
310 identified as Transitional, 1 is in Spotted Fever Group, 1 is *Adalia*, 8 are *Belli* and 7 are *Torix*  
311 *limoniae*. Targeted sequences were confirmed as: *Torix limoniae* (RiCimp), *Torix leech* (RiClec),  
312 Transitional (RiTSETSE), *Ca. Megaira* (MegCarteria and MegNEIS296), and a new *Torix* clade,  
313 Moomin (Moomin). The new *Torix* include one double infection giving a total of 10 new genomes  
314 across 9 potential host species. The double infection is found within the rove beetle *Labidopullus*  
315 *appendiculatus*, forming two distinct lineages, Lappe3 and Lappe4 (Fig 2 and 3).

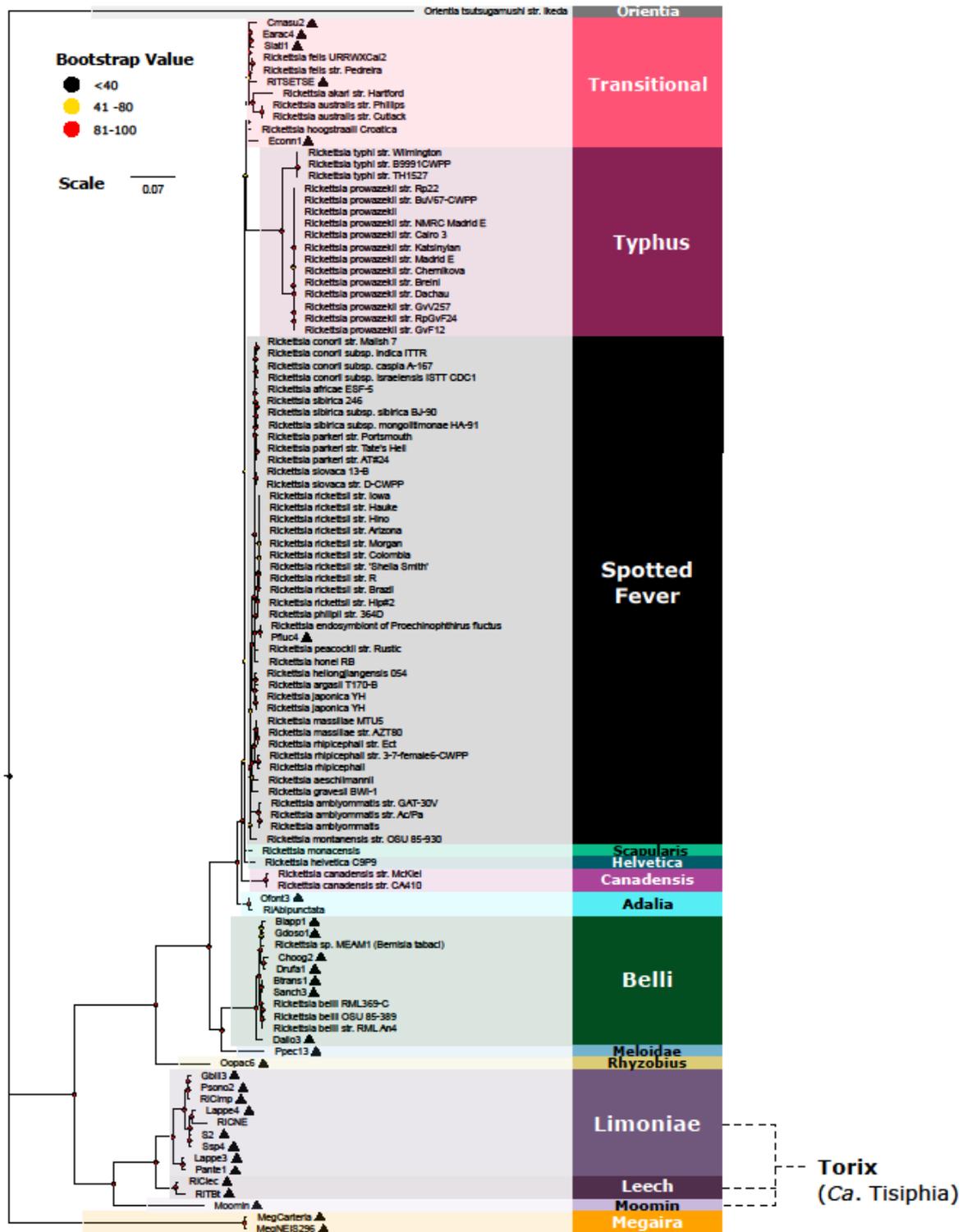
316 In addition, the pre-existing *Rickettsia helvetica*, which is typically cited as a member of the Spotted  
317 Fever group as a result of its first description in 1993 (Beati et al., 1993; Chisu et al., 2017), seems to  
318 form its own group in all trees and networks (figure 2, 3 and  
319 <https://doi.org/10.6084/m9.figshare.14865606> ). We conclude from this that *Rickettsia helvetica* is  
320 most similar to Scapularis group *Rickettsia*, but because it does not fall into the same clade in any  
321 tree or network, it is likely that the strain belongs to a distinct lineage of tick-borne *Rickettsia*.

## Core Phylogeny



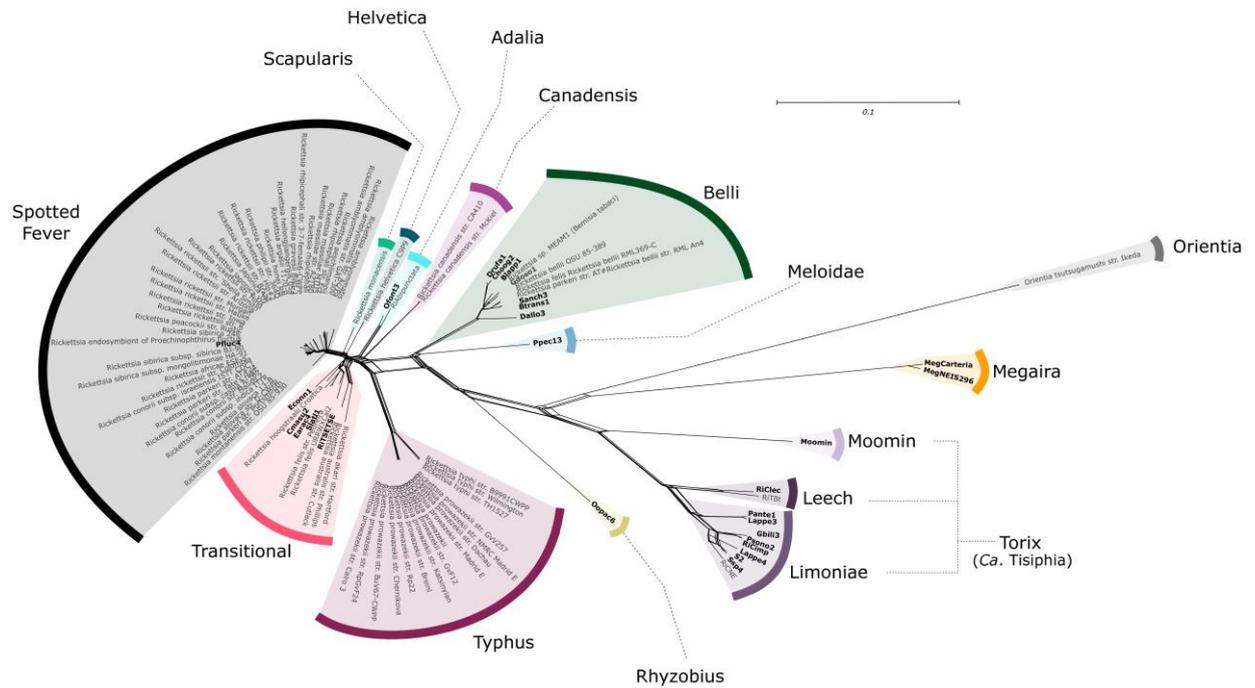
322 Figure 2. *Rickettsia* and *Ca. Megaira* maximum likelihood (ML) phylogeny constructed from 74 core gene clusters extracted from the  
 323 pangenome. New genomes are indicated by ▲ and bootstrap values based on 1000 replicates are indicated with coloured circles.  
 324 New complete genomes are: RiCimp, RiClec and MegNEIS296. A full resolution version can be found here:  
 325 <https://doi.org/10.6084/m9.figshare.15081975>.

## Ribosomal Phylogeny



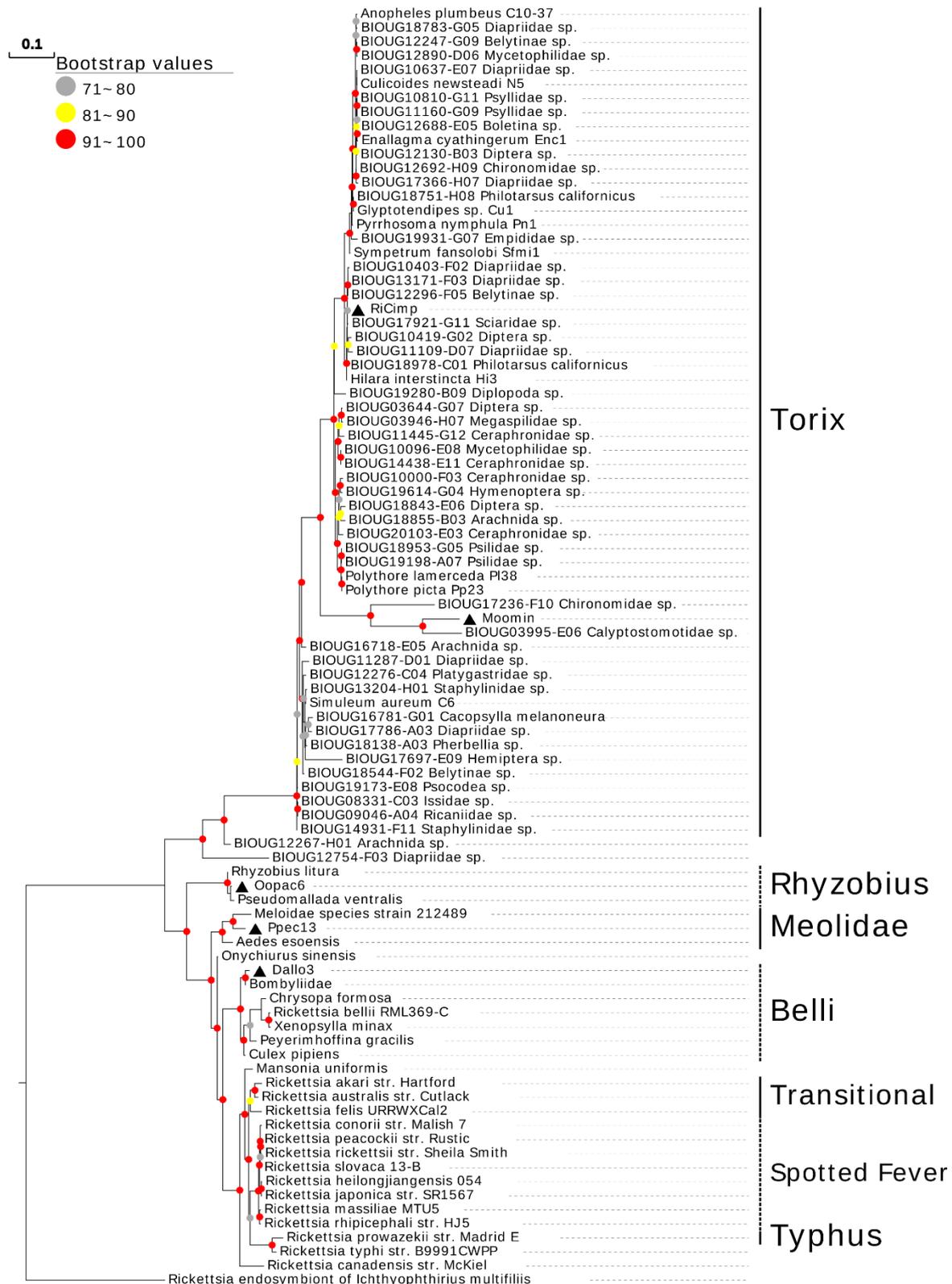
326

327 *S4. Rickettsia and Ca. Megaira maximum likelihood (ML) phylogeny constructed from 43 ribosomal protein gene clusters extracted*  
 328 *from the pangenome. New genomes are indicated by ▲ and bootstrap values based on 1000 replicates are indicated with coloured*  
 329 *circles. New complete genomes are: RiCimp, RiClec and MegNEIS296. <https://doi.org/10.6084/m9.figshare.14865606>*



330 Figure 3. Nearest Neighbour Network, displaying the distances between the 74 core gene sets across all 104 *Rickettsia*, *Ca. Megaira*  
 331 genomes, and the outgroup *Orientia*. New genomes are indicated with bold text. A full resolution version can be found here:  
 332 <https://doi.org/10.6084/m9.figshare.15081975>.

333 We also report the first putative *Rhyzobius Rickettsia* genomes extracted from the staphylinid  
 334 beetle *Oxypoda opaca* (Oopac6) and Meloidae *Rickettsia* from the firefly *Pyrocoelia pectoralis*  
 335 (Ppec13). They have high completeness (S1 <https://figshare.com/s/198c88c6e3ea5553192e>), low  
 336 pseudogenisation, and consistently group away from the other draft and completed genomes  
 337 (Figures 2 and 3). MLST analyses demonstrate that these bacteria are most like the *Rhyzobius* and  
 338 Meloidae groups described by Weinert *et al.* (2009) (see S5  
 339 <https://doi.org/10.6084/m9.figshare.14865600>). The pangenome and metabolic profile of this draft  
 340 genome suggests that Meloidae is a sister group to Belli and that *Rhyzobius Rickettsia* is  
 341 superficially similar to Belli and Transitional groups. The *Rhyzobius*-group symbiont is  
 342 phylogenetically distant from most *Rickettsia* and is potentially a sister clade linking Torix and the  
 343 main *Rickettsia* clades. Further genome construction will help clarify this taxon and its relationship  
 344 to the rest of *Rickettsia*.



345

346 *S5 Phylogram of a maximum likelihood (ML) tree of 90 Rickettsia mutilocus profiles. The tree is based on 4 loci, 16S rRNA, 17KDa,*  
347 *gltA, and COI, under a partition model (2,781 bp total). <https://doi.org/10.6084/m9.figshare.14865600>*

348 The sequencing data for the wasp, *Diachasma alloeum*, used here has previously been described to  
349 contain a pseudogenised nuclear insert of *Rickettsia* material, but not a complete *Rickettsia*  
350 genome (Tvedte et al., 2019). The construction of a full, non-pseudogenised genome with higher  
351 read depth than the insect contigs, low contamination (0.95%) and high completion (93.13%)  
352 suggests that these reads likely represent a viable *Rickettsia* infection in *D. alloeum*. However, these  
353 data do not exclude the presence of an additional nuclear insert. It is possible for a whole symbiont  
354 genome to be incorporated into the host's DNA (Hotopp et al., 2007), and there are recorded  
355 partial inserts of *Ca. Megaira* genomes in the *Volvox carteri* genome (Kawafune et al., 2015). The  
356 presence of both the insert and symbiont need confirmation through appropriate microscopy  
357 methods.

358 Recombination is low within the core genomes of *Rickettsia* and *Ca. Megaira*, but may occur  
359 between closely related clades that are not investigated here. Across all genomes, the PHI score is  
360 significant in 6 of the 74 core gene clusters, suggesting putative recombination events. However, it  
361 is reasonable to assume that most of these may be a result of systematic error due to the divergent  
362 evolutionary processes at work across *Rickettsia* genomes. Patterns of recombination can occur by  
363 chance rather than driven by evolution which cannot be differentiated by current phylogenetic  
364 methods (Murray et al., 2016). The function of each respective cluster can be found at  
365 <https://figshare.com/s/198c88c6e3ea5553192e>.

## 366 Gene content and pangenome analysis

### 367 Pangenome

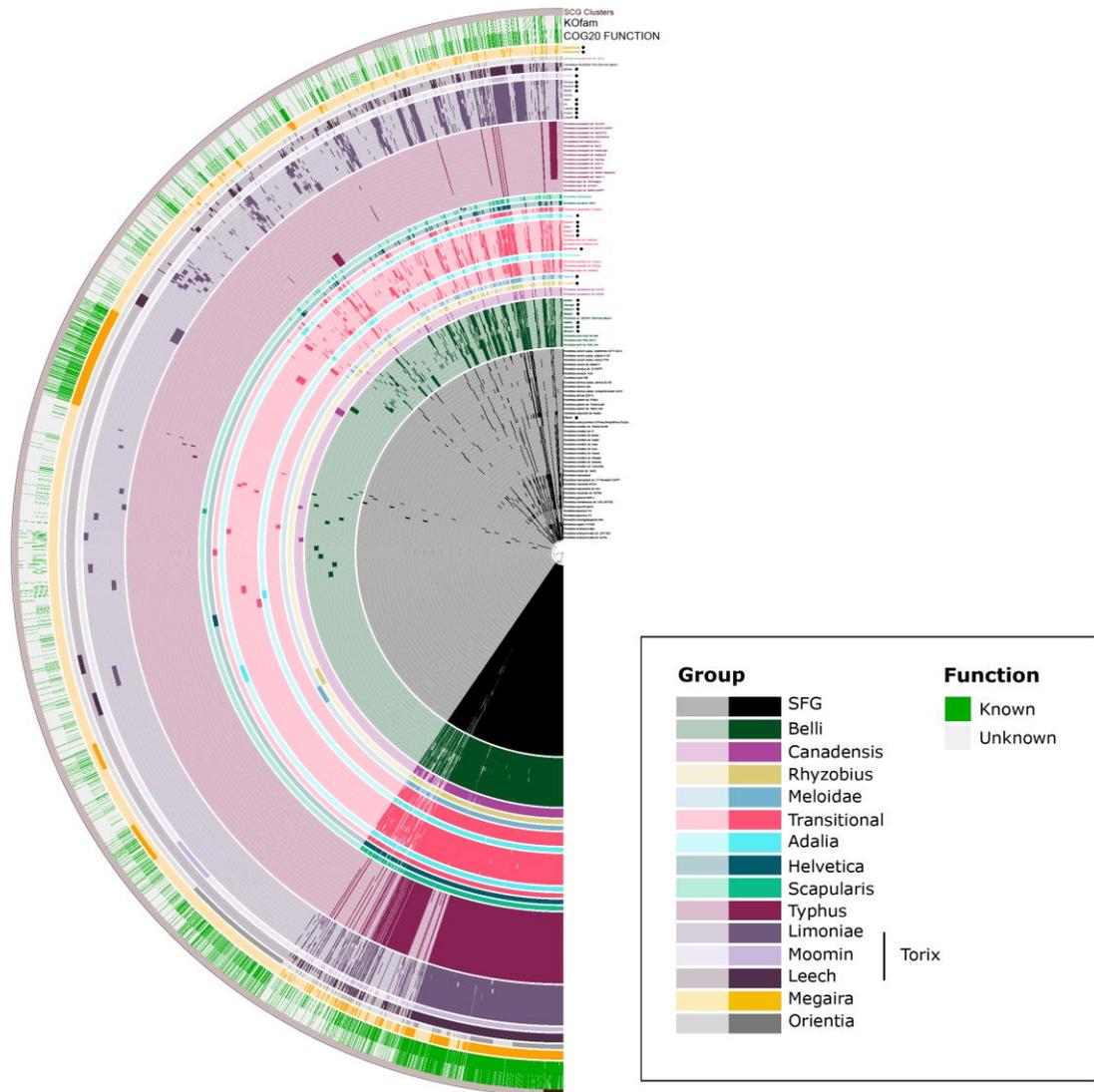
368 Across all 104 genomes used in the pangenome analysis (figure 2, full data in S6  
369 <https://doi.org/10.6084/m9.figshare.14865576>), Anvi'o identified 208 core gene clusters of which  
370 74 are represented by single-copy genes. Bacterial strains of the different *Rickettsia* groups,  
371 especially the neglected symbiotic Rickettsiaceae, seem to have large, open pangenomes an  
372 indication of rapid evolution. As expected, the more genomes that are included in analyses, the  
373 smaller the core genome extracted.

374 Torix is a distinctly separate clade sharing less than 70% ANI similarity to any *Rickettsia* or *Ca.*  
375 *Megaira* genomes. It contains at least three groups that reflect its highly diverse niche in the  
376 environment (figure 5) (Jain et al., 2018; Pilgrim et al., 2021; Rodriguez-R et al., 2021). Torix has the  
377 most unique genes out of all the clades in this study followed by *Ca. Megaira* and *Belli* clades (figure  
378 6). Rarefaction gene accumulation analysis suggest that Torix is the group where each additional

379 genome included increases the pangenome repertoire to the greatest extent (figure 7). Torix group  
380 is thus more diverse in terms of genome content and size of the pangenome than other *Rickettsia*  
381 groups.

382 *Rickettsia* lineages group together based on gene presence/absence and produce repeated patterns  
383 of accessory genes that reliably occur within each group (figure 2). ANI scores are also strongest  
384 within groups, while genomes tend to share lower similarity outside of their group (figure 4). This is  
385 particularly apparent in Torix and *Ca. Megaira* which are divergent from the main *Rickettsia* clade  
386 (figure 3 and 5).

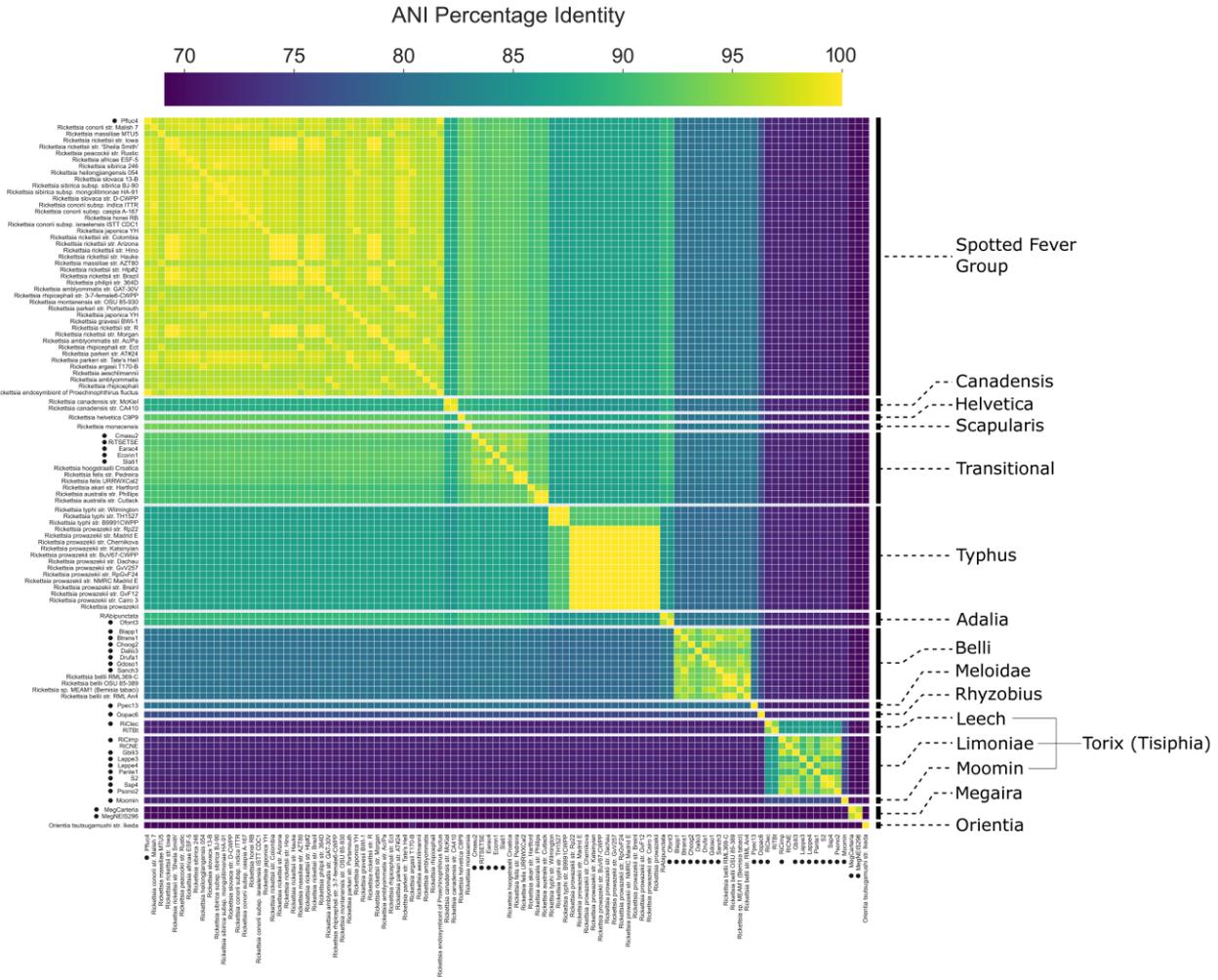
387



388

389 Figure 4. Pangenome of all 104 genomes including *Rickettsia*, *Torix*, *Ca. Megaira* and the outgroup *Orientia*. New genomes are  
390 indicated by ●. Each genome displays gene cluster presence/absence and is organised by gene cluster frequency. Group identity was  
391 assigned from phylogeny. SFG is Spotted Fever Group. A full resolution version can be found here:  
392 <https://doi.org/10.6084/m9.figshare.15081975>.

393



394

395 *Figure 5. Pairwise Average Nucleotide Identity percentage across all genomes. New genomes are indicated by a black circle. A full*  
 396 *resolution version can be found here: <https://doi.org/10.6084/m9.figshare.15081975>.*

397

398

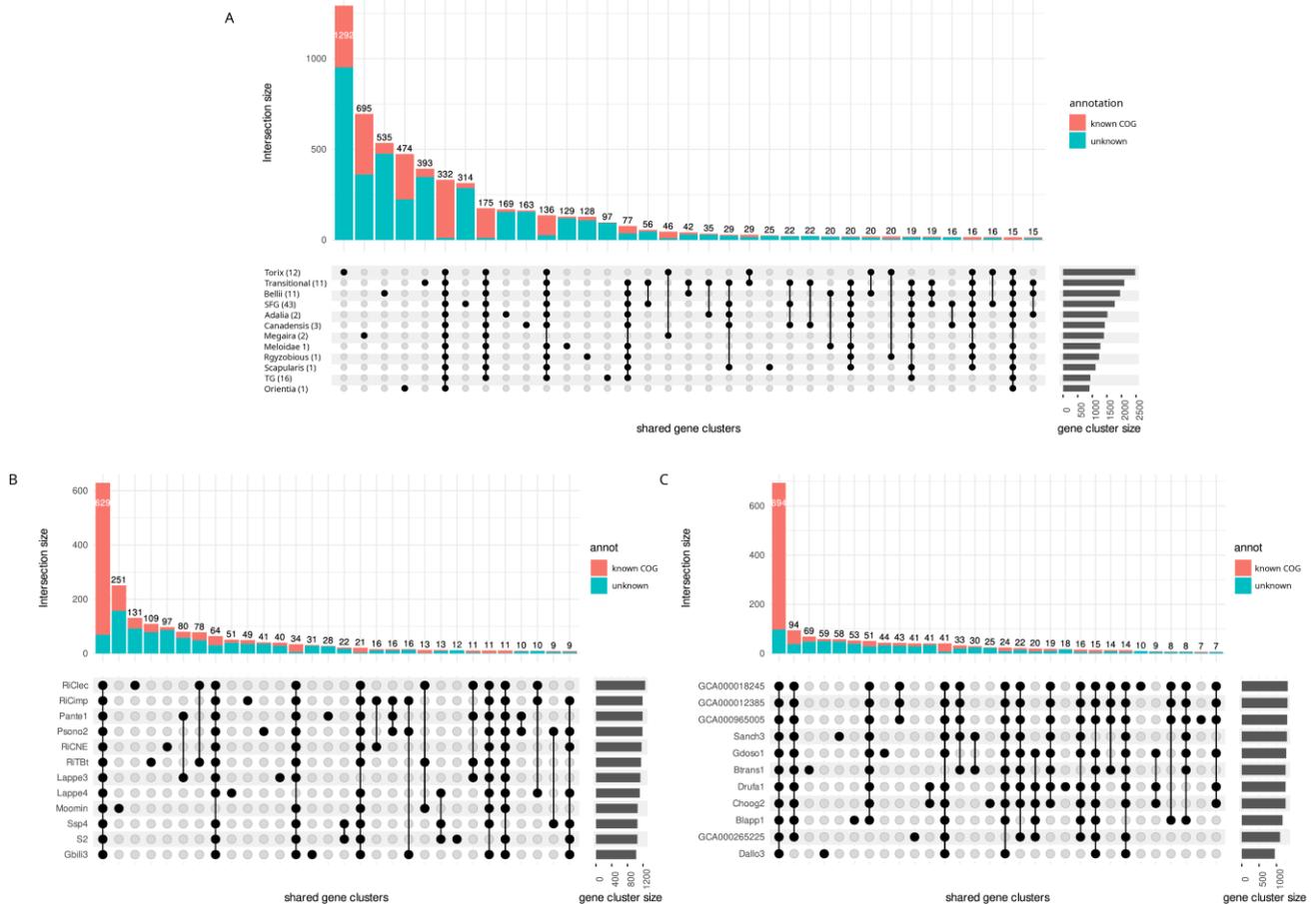
399

400

401

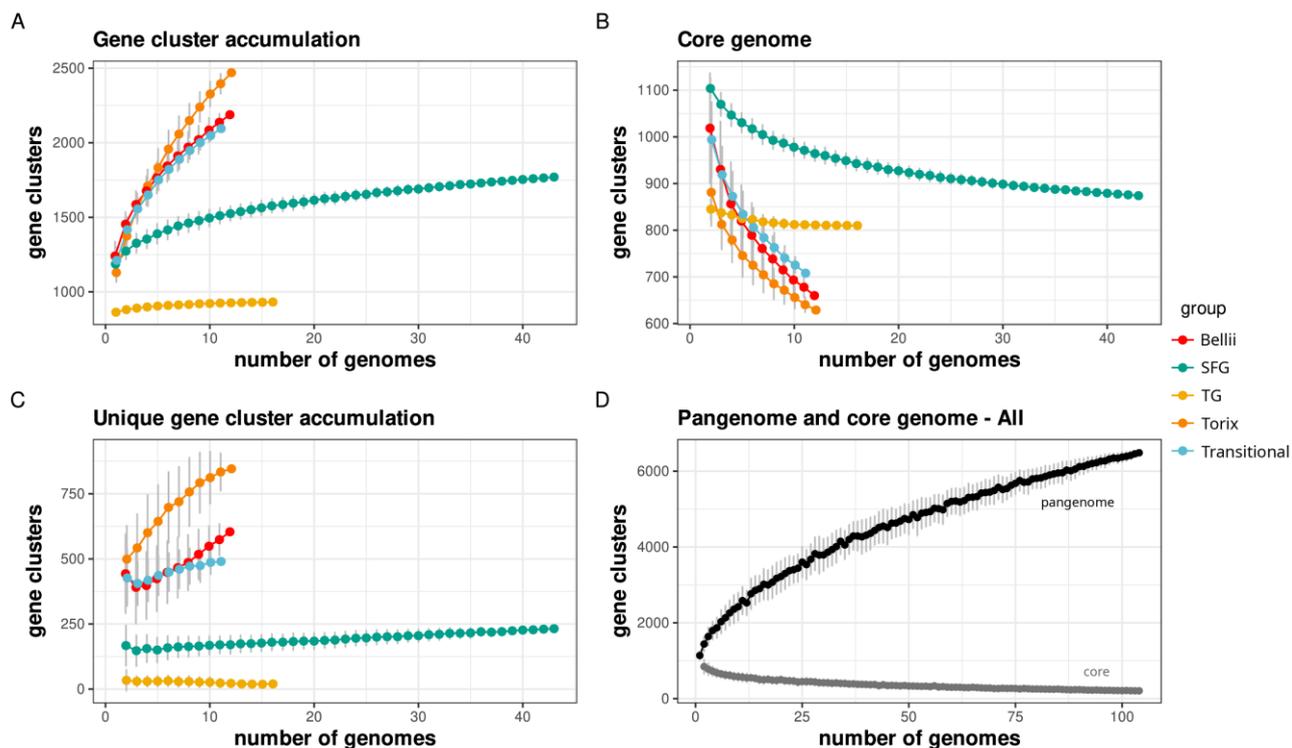
402

403



404

405 *Figure 6. Shared and unique gene clusters across A) All Rickettsia and Ca. Megaira genomes used in this study grouped by clade with*  
 406 *Orienta as an outgroup B) all individual Torix genomes, and C) all individual Belli genomes. Horizontal grey bars to the right of each*  
 407 *plot represent gene cluster size and vertical, coloured bars represent the size of intersections (the number of shared gene clusters)*  
 408 *between genomes in descending order with known COG functions displayed in red and unknown in blue. Black dots mean the cluster*  
 409 *is present and connected dots represent gene clusters that are present across groups. SFG is Spotted Fever Group and TG is Typhus*  
 410 *Group. A full resolution version can be found here: <https://doi.org/10.6084/m9.figshare.15081975>.*



411

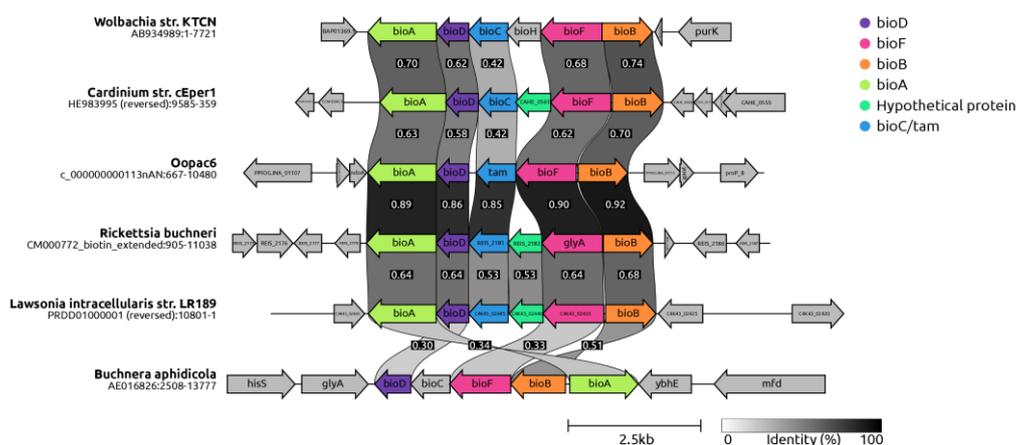
412 *Figure 7. Gene cluster accumulation curves for pangenome (A), core genome (B) and the unique genome (C) of the 5 largest Rickettsia*  
 413 *groups as a function of the number of genomes sequenced. The pangenome and the core genome accumulation curves for the*  
 414 *complete Rickettsia dataset is shown in panel D. Error bars represent  $\pm$  standard deviation based on 100 permutations. SFG is Spotted*  
 415 *Fever Group and TG is Typhus Group. A full resolution version can be found here: <https://doi.org/10.6084/m9.figshare.15081975>.*

#### 416 *Gene content and metabolic analyses*

417 Rickettsial genomes extracted from SRA samples are generally congruent with the metabolic  
 418 potential of their respective groups (Figure 8). Torix and *Ca. Megaira* have complete pentose  
 419 phosphate pathways (PPP); a unique marker for these groups which seems to have been lost in the  
 420 other *Rickettsia* clades. The PPP generates NADPH, precursors to amino acids, and is known to  
 421 protect against oxidative injury in some bacteria (Christodoulou et al., 2018), as well as conversion  
 422 of hexose monosaccharides into pentose used in nucleic acid and exopolysaccharide synthesis. The  
 423 PPP has also been associated with establishing symbiosis between the Alphaproteobacteria  
 424 *Sinorhizobium meliloti* and its plant host *Medicago sativa* (Hawkins et al., 2018). This pathway has  
 425 previously been highlighted in Torix (Pilgrim et al., 2017) and its presence in all newly assembled  
 426 Torix and *Ca. Megaira* draft genomes consolidates its importance as an identifying feature for these  
 427 groups (Figure 8, S1 <https://figshare.com/s/198c88c6e3ea5553192e>). The PPP is likely an ancestral  
 428 feature that was lost in the main *Rickettsia* clade.

429 Glycolysis, gluconeogenesis and cofactor and vitamin metabolism are absent or incomplete across  
 430 all *Rickettsia*, except the Rhyzobius group member, Oopac6 (Figure 8). Oopac6 has a complete

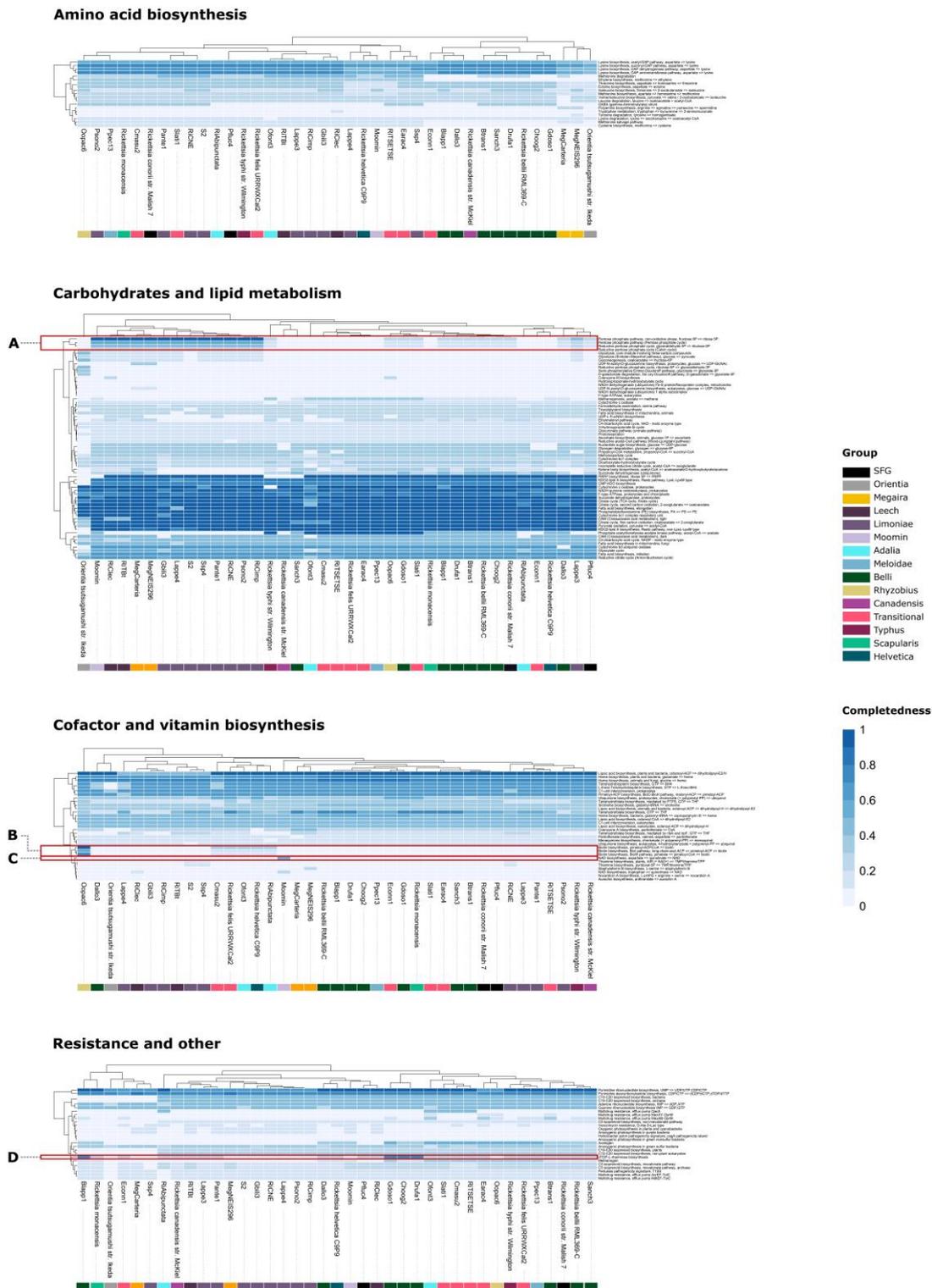
431 biotin synthesis pathway that is related to, but distinct from, the *Rickettsia* biotin pathway first  
 432 observed in *Rickettsia buchneri* (See S7 <https://doi.org/10.6084/m9.figshare.14865567>) (Gillespie  
 433 et al., 2012). Based on the gene cluster comparison plot and an independent blastx search, the *GlyA*  
 434 gene in *Rickettsia buchneri* appears to be a misidentified *bioF* gene (see S7  
 435 <https://doi.org/10.6084/m9.figshare.14865567>). Additionally, the insect SRA sample was not  
 436 infected with *Wolbachia*, making it unlikely that the presence of the biotin operon is a result of  
 437 misassembly. Animals can't synthesize B-vitamins, so they either acquire them from diet or from  
 438 microorganisms that can synthesize them. *Oopac6* has retained or acquired a complete biotin  
 439 operon where this operon is absent in other members of the genus. Biotin pathways in insect  
 440 symbionts can be an indicator of nutritional symbioses (Douglas, 2017), so Rhyzobius *Rickettsia*  
 441 could contribute to the feeding ecology of the beetle *O. opaca*. However, like other aleocharine  
 442 rove beetles, *O. opaca* is likely predaceous, omnivorous or fungivorous (analysis of gut contents  
 443 from a related species, *O. grandipennis*, revealed a high prevalence of yeasts: Klimaszewski et al.,  
 444 2013). We posit no obvious reason for how these beetles benefit from harbouring a biotin-  
 445 producing symbiont. One theory is that this operon could be a 'hangover' from a relatively recent  
 446 host shift event and may have been functionally important in the original host. Similarly, if the  
 447 symbiont is undergoing genome degradation, a once useful biotin pathway may be present but not  
 448 functional (Blow et al., 2020). As this is the only member of this group with a complete genome so  
 449 far, further research is required to firmly establish the presence and function of this pathway.



450 S7 Biotin operon of the Rhyzobius Rickettsia, Oopac6, and its surrounding genes compared with other known biotin pathways in other  
 451 related symbionts. Similarity scores in the black boxes refer to the percentage identity between the genes of the operon above or  
 452 below it, further illustrated by a greyscale bar. Operons are ordered by overall similarity, showing the closest relationships between  
 453 all 6. <https://doi.org/10.6084/m9.figshare.14865567>

454 A 75% complete dTDP-L-rhamnose biosynthesis pathway was observed in 4 of the draft belli  
 455 assemblies (Gdoso1, Choog2, Drufa1, Blapp1) (figure 8). Two host species are bird lice (*Columbicola*

456 *hoogstraali*, *Degeeriella rufa*), one is a butterfly (*Graphium doson*), and one is a ground beetle  
457 (*Bembidion lapponicum*). dTDP-L-rhamnose is an essential component of human pathogenic  
458 bacteria like *Pseudomonas*, *Streptococcus* and *Enterococcus*, where it is used in cell wall  
459 construction (van der Beek et al., 2019). This pathway has also been utilized in the synthesis of  
460 plant cell walls (Jiang et al., 2021), may be involved in the moulting process of *Caenorhabditis*  
461 *elegans* (Feng et al., 2016), and is a precursor to rhamnolipids that are used in quorum sensing  
462 (Daniels et al., 2004). In the root symbiont *Azospirillum*, disruption of this pathway alters root  
463 colonisation, lipopolysaccharide structure and exopolysaccharide production (Jofré et al., 2004). No  
464 *Rickettsia* from typically pathogenic groups assessed in figure 8 has this pathway, and the hosts of  
465 these four bacteria are not involved with human or mammalian disease. Presence in feather lice  
466 provides little opportunity for this *Rickettsia* to be pathogenic as feather lice are chewers rather  
467 than blood feeders, and Belli group *Rickettsia* more generally are rarely pathogenic. Further, this  
468 association does not explain its presence in a butterfly and ground beetle; it is most likely that this  
469 pathway, if functional, would be involved in establishing infection in the insect host or host-  
470 symbiont recognition.



471 Figure 8. Heatmaps of predicted KEGG pathway completion estimated in Anvi'o 7, separated by function and produced with  
 472 Pheatmap. Pathways of interest are highlighted: A) The pentose phosphate pathway only present in Torix and Ca. Megaira, B) the  
 473 biotin pathway present only in the Rhyzobius Rickettsia Oopac6, C) NAD biosynthesis only present in Moomin Rickettsia, D) dTDP-L-  
 474 rhamnose biosynthesis pathway in Gdoso1, Choog2, Drufa1, and Blapp1. SFG is Spotted Fever. A full resolution version can be found  
 475 here: <https://doi.org/10.6084/m9.figshare.15081975>.

## 476 Designation of *Ca. Tisiphia*

477 In all analyses, Torix consistently cluster away from the rest of *Rickettsia* as a sister taxon. Despite  
478 the relatively small number of Torix genomes, its within group diversity is greater than any  
479 divergence between previously described *Rickettsia* in any other group (figures 2, 3 and 5).  
480 Additionally, Torix shares characteristics with both *Ca. Megaira* and *Rickettsia*, but with many of its  
481 own unique features (figures 6 and 8). The distance of Torix from other *Rickettsia* and *Ca. Megaira*  
482 is confirmed in both the phylogenomic and metabolic function analyses to the extent that Torix  
483 should be separated from *Rickettsia* and assigned its own genus. This is supported by GTDB-Tk  
484 analysis which places all Torix genomes separate from *Rickettsia* (S1  
485 <https://figshare.com/s/198c88c6e3ea5553192e>) alongside ANI percentage similarity scores less  
486 than 70% in all cases. To this end, we propose the name *Candidatus Tisiphia* after the fury  
487 Tisiphone, reflecting the genus *Ca. Megaira* being named after her sister Megaera.

## 488 Conclusion

489 The bioinformatics approach has successfully extracted a substantial number of novel *Rickettsia*  
490 and *Ca. Megaira* genes from existing SRA data, including the first putative Rhyzobius *Rickettsia* and  
491 several *Ca. Tisiphia* (formerly Torix *Rickettsia*). Successful completion of two *Ca. Megaira* and two  
492 *Ca. Tisiphia* genomes provide solid reference points for the evolution of *Rickettsia* and its sister  
493 groups. From this, we can confirm the presence of a complete Pentose Phosphate Pathway in *Ca.*  
494 *Tisiphia* and *Ca. Megaira*, suggesting that this pathway was lost during *Rickettsia* evolution. We  
495 also describe the first Meloidae and Rhyzobius *Rickettsia* and show that Rhyzobius group *Rickettsia*  
496 has the potential to be a nutritional symbiont due to the presence of a complete biotin pathway.  
497 These new genomes provide a much-needed expansion of available data for symbiotic *Rickettsia*  
498 clades and clarification on the evolution of *Rickettsia* from *Ca. Megaira* and *Ca. Tisiphia*.

## 499 Supporting information

500 All original genomes and raw readsets produced in this study can be accessed at Bioproject accession  
501 PRJNA763820 and all assemblies produced from previously published third party data can be accessed at  
502 Bioproject PRJNA767332.

503 Supplementary data and full resolution figures can be accessed on figshare here:

504 <https://doi.org/10.6084/m9.figshare.c.5518182.v1>

## 505 Acknowledgements

506 HRD was supported by the NERC ACCE Doctoral Training Programme. Grant code: NE/L002450/1  
507 and NW was supported by a BOF post-doctoral fellowship (Ghent University, 01P03420) and by a

508 Research Foundation - Flanders (FWO) Research Grant (1513719N). Funding for tsetse fly genomics  
509 were to ACD IP BBSRC projects BB/J017698/1 and BB/K501773/1, the materials from which were  
510 provided by P. Solano (Institut de Recherche pour le Développement, Unité Mixte de Recherche  
511 Interactions Hôtes-Vecteurs-Parasites-Environnement dans les Maladies Tropicales Négligées Dues  
512 aux Trypanosomatides, 34398, Montpellier, France) and J.-B. Rayaisse (Centre International de  
513 Recherche-Développement sur l'Élevage en zone Subhumide (CIRDES), N°559, Rue 5-31 Avenue du  
514 Gouverneur Louveau, 01 BP 454, Bobo Dioulasso 01, Burkina Faso). Jean-Baptiste died a few years  
515 ago but he was a fantastic person to work with and a great field entomologist. We also wish to  
516 thank Dr David Montagnes for teaching skills associated with algal culture.

517 We wish to thank Dr Débora Pires Paula (Embrapa) for granting permission to use SRA data for  
518 sample number SRR5651504, Iridian Genomes for allowing use of their SRA data, and the Microbial  
519 Culture Collection at the National Institute for Environmental Studies, Japan for use of the sample  
520 *Carteria cerasiformis* NIES-425.

## 521 Contributions

522 Project concept: HRD, SS, JP and GH

523 Manuscript written by HRD, SS, JP and GDDH

524 SRA dive and metagenome assembly carried out by HRD with aid from SS.

525 Assembly of genome from SRA, pangenomics and phylogenomics carried out by HRD with advice from SS,  
526 GH

527 Metabolic analysis carried out by HRD, JP and SS

528 Sequencing and assembly of bacteria from *Cimex lectularius* and *Culicoides impunctatus* genomes by SS  
529 and JP.

530 Sequencing and assembly of symbionts from *Carteria* by SHB and SS, supervised by PC and GH.

531 Sequencing and construction of RiTSETSE carried out by FB as part of thesis work supervised by AD.

532 SP collected and sequenced staphylinid genomes that were released through NCBI by iridian genomics.

533 NW collected and sequenced the *Bryobia* Moomin strain and performed preliminary metagenomic analyses

## 534 References

535 Aguin-Pombo, D., Rodrigues, M. C. P. A., Voetdijk, B., & Breeuwer, J. A. J. (2021). Parthenogenesis  
536 and Sex-Ratio Distorting Bacteria in *Empoasca* (Hemiptera: Cicadellidae) Leafhoppers. *Annals*  
537 *of the Entomological Society of America*. <https://doi.org/10.1093/AESA/SAAB025>

538 Angelakis, E., & Raoult, D. (2017). *Rickettsia* and *Rickettsia*-Like Organisms. In *Infectious Diseases*

- 539 (pp. 1666-1675.e1). Elsevier. <https://doi.org/10.1016/b978-0-7020-6285-8.00187-8>
- 540 Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., & Ogata, H. (2020).  
541 KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold.  
542 *Bioinformatics*, 36(7), 2251–2252. <https://doi.org/10.1093/bioinformatics/btz859>
- 543 Beati, L., Peter, O., Burgdorfer, W., Aeschlimann, A., & Raoult, D. (1993). Confirmation that  
544 *Rickettsia helvetica* sp. nov. Is a Distinct Species of the Spotted Fever Group of Rickettsiae.  
545 *International Journal of Systematic Bacteriology*, 43(3), 521–526.  
546 <https://doi.org/10.1099/00207713-43-3-521>
- 547 Blin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., van Wezel, G. P., Medema, M. H., &  
548 Weber, T. (2021). antiSMASH 6.0: improving cluster detection and comparison capabilities.  
549 *Nucleic Acids Research*, 49(W1), W29–W35. <https://doi.org/10.1093/nar/gkab335>
- 550 Blow, F. (2017). *Variation in the structure and function of invertebrate-associated bacterial*  
551 *communities* [University of Liverpool]. <https://doi.org/10.17638/03009325>
- 552 Blow, F., Bueno, E., Clark, N., Zhu, D. T., Chung, S. H., Güllert, S., Schmitz, R. A., & Douglas, A. E.  
553 (2020). B-vitamin nutrition in the pea aphid-Buchnera symbiosis. *Journal of Insect Physiology*,  
554 126, 104092. <https://doi.org/10.1016/j.jinsphys.2020.104092>
- 555 Bodnar, J. L., Fitch, S., Rosati, A., & Zhong, J. (2018). The folA gene from the Rickettsia  
556 endosymbiont of Ixodes pacificus encodes a functional dihydrofolate reductase enzyme. *Ticks*  
557 *and Tick-Borne Diseases*, 9(3), 443–449. <https://doi.org/10.1016/j.ttbdis.2017.12.013>
- 558 Boettcher, K. J., Ruby, E. G., & McFall-Ngai, M. J. (1996). Bioluminescence in the symbiotic squid  
559 Euprymna scolopes is controlled by a daily biological rhythm. *Journal of Comparative*  
560 *Physiology A*, 179(1), 65–73. <https://doi.org/10.1007/BF00193435>
- 561 Bruen, T. C., Philippe, H., & Bryant, D. (2006). A Simple and Robust Statistical Test for Detecting the  
562 Presence of Recombination. *Genetics*, 172(4), 2665–2681.  
563 <https://doi.org/10.1534/genetics.105.048975>
- 564 Brumin, M., Kontsedalov, S., & Ghanim, M. (2011). Rickettsia influences thermotolerance in the  
565 whitefly Bemisia tabaci B biotype. *Insect Science*, 18(1), 57–66. [https://doi.org/10.1111/j.1744-](https://doi.org/10.1111/j.1744-7917.2010.01396.x)  
566 [7917.2010.01396.x](https://doi.org/10.1111/j.1744-7917.2010.01396.x)

- 567 Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2020). GTDB-Tk: a toolkit to classify  
568 genomes with the Genome Taxonomy Database. *Bioinformatics*, *36*(6), 1925–1927.  
569 <https://doi.org/10.1093/BIOINFORMATICS/BTZ848>
- 570 Chiel, E., Inbar, M., Mozes-Daube, N., White, J. A., Hunter, M. S., & Zchori-Fein, E. (2009).  
571 Assessments of Fitness Effects by the Facultative Symbiont *Rickettsia* in the  
572 Sweetpotato Whitefly (Hemiptera: Aleyrodidae). *Annals of the Entomological Society of*  
573 *America*, *102*(3), 413–418. <https://doi.org/10.1603/008.102.0309>
- 574 Chisu, V., Leulmi, H., Masala, G., Piredda, M., Foxi, C., & Parola, P. (2017). Detection of *Rickettsia*  
575 *hoogstraalii*, *Rickettsia helvetica*, *Rickettsia massiliae*, *Rickettsia slovacica* and *Rickettsia*  
576 *aeschlimannii* in ticks from Sardinia, Italy. *Ticks and Tick-Borne Diseases*, *8*(3), 347–352.  
577 <https://doi.org/10.1016/j.ttbdis.2016.12.007>
- 578 Christodoulou, D., Link, H., Fuhrer, T., Kochanowski, K., Gerosa, L., & Sauer, U. (2018). Reserve Flux  
579 Capacity in the Pentose Phosphate Pathway Enables *Escherichia coli*'s Rapid Response to  
580 Oxidative Stress. *Cell Systems*, *6*, 569–578. <https://doi.org/10.1016/j.cels.2018.04.009>
- 581 Clay, K., Holah, J., & Rudgers, J. A. (2005). Herbivores cause a rapid increase in hereditary symbiosis  
582 and alter plant community composition. *Proceedings of the National Academy of Sciences*,  
583 *102*(35), 12465–12470. <https://doi.org/10.1073/pnas.0503059102>
- 584 Daniels, R., Vanderleyden, J., & Michiels, J. (2004). Quorum sensing and swarming migration in  
585 bacteria. *FEMS Microbiology Reviews*, *28*(3), 261–289.  
586 <https://doi.org/10.1016/j.femsre.2003.09.004>
- 587 Doudoumis, V., Blow, F., Saridaki, A., Augustinos, A., Dyer, N. A., Goodhead, I., Solano, P., Rayaisse,  
588 J.-B., Takac, P., Mekonnen, S., Parker, A. G., Abd-Alla, A. M. M., Darby, A., Bourtzis, K., &  
589 Tsiamis, G. (2017). Challenging the *Wigglesworthia*, *Sodalis*, *Wolbachia* symbiosis dogma in  
590 tsetse flies: *Spiroplasma* is present in both laboratory and natural populations. *Scientific*  
591 *Reports* *2017* *7*:1, *7*(1), 1–13. <https://doi.org/10.1038/s41598-017-04740-3>
- 592 Douglas, A. E. (2011). Lessons from studying insect symbioses. In *Cell Host and Microbe* (Vol. 10,  
593 Issue 4, pp. 359–367). Cell Press. <https://doi.org/10.1016/j.chom.2011.09.001>
- 594 Douglas, A. E. (2017). The B vitamin nutrition of insects: the contributions of diet, microbiome and  
595 horizontally acquired genes. *Current Opinion in Insect Science*, *23*, 65–69.

- 596 <https://doi.org/10.1016/j.cois.2017.07.012>
- 597 Eddy, S. R. (2011). Accelerated Profile HMM Searches. *PLoS Computational Biology*, 7(10),  
598 e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>
- 599 Edgar, R. C. (2004). MUSCLE: A multiple sequence alignment method with reduced time and space  
600 complexity. *BMC Bioinformatics*, 5(1), 1–19. <https://doi.org/10.1186/1471-2105-5-113>
- 601 Engelstädter, J., & Hurst, G. D. D. (2009). The Ecology and Evolution of Microbes that Manipulate  
602 Host Reproduction. *Annual Review of Ecology, Evolution, and Systematics*, 40(1), 127–149.  
603 <https://doi.org/10.1146/annurev.ecolsys.110308.120206>
- 604 Eren, A. M., Kiefl, E., Shaiber, A., Veseli, I., Miller, S. E., Schechter, M. S., Fink, I., Pan, J. N., Yousef,  
605 M., Fogarty, E. C., Trigodet, F., Watson, A. R., Esen, Ö. C., Moore, R. M., Clayssen, Q., Lee, M.  
606 D., Kivenson, V., Graham, E. D., Merrill, B. D., ... Willis, A. D. (2021). Community-led, integrated,  
607 reproducible multi-omics with anvi'o. *Nature Microbiology*, 6(1), 3–6.  
608 <https://doi.org/10.1038/s41564-020-00834-3>
- 609 Feng, L., Shou, Q., & Butcher, R. A. (2016). Identification of a dTDP-rhamnose biosynthetic pathway  
610 that oscillates with the molting cycle in *Caenorhabditis elegans*. *Biochemical Journal*, 473(11),  
611 1507–1521. <https://doi.org/10.1042/BCJ20160142>
- 612 Flissi, A., Ricart, E., Campart, C., Chevalier, M., Dufresne, Y., Michalik, J., Jacques, P., Flahaut, C.,  
613 Lisacek, F., Leclère, V., & Pupin, M. (2019). OUP accepted manuscript. *Nucleic Acids Research*,  
614 48(D1), D465–D469. <https://doi.org/10.1093/nar/gkz1000>
- 615 Fujishima, M., & Kodama, Y. (2012). Endosymbionts in Paramecium. *European Journal of*  
616 *Protistology*, 48(2), 124–137. <https://doi.org/10.1016/j.ejop.2011.10.002>
- 617 Galperin, M. Y., Wolf, Y. I., Makarova, K. S., Vera Alvarez, R., Landsman, D., & Koonin, E. V. (2021).  
618 COG database update: focus on microbial diversity, model organisms, and widespread  
619 pathogens. *Nucleic Acids Research*, 49(D1), D274–D281.  
620 <https://doi.org/10.1093/nar/gkaa1018>
- 621 Gilchrist, C. L. M., & Chooi, Y.-H. (2021). clinker & clustermap.js: automatic generation of gene  
622 cluster comparison figures. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btab007>
- 623 Gillespie, J. J., Driscoll, T. P., Verhoeve, V. I., Rahman, M. S., Macaluso, K. R., & Azad, A. F. (2018). A

- 624 Tangled Web: Origins of Reproductive Parasitism. *Genome Biology and Evolution*, 10(9), 2292–  
625 2309. <https://doi.org/10.1093/gbe/evy159>
- 626 Gillespie, J. J., Driscoll, T. P., Verhoeve, V. I., Utsuki, T., Husseneder, C., Chouljenko, V. N., Azad, A.  
627 F., & Macaluso, K. R. (2015). Genomic Diversification in Strains of *Rickettsia felis* Isolated from  
628 Different Arthropods. *Genome Biology and Evolution*, 7(1), 35–56.  
629 <https://doi.org/10.1093/gbe/evu262>
- 630 Gillespie, J. J., Joardar, V., Williams, K. P., Driscoll, T., Hostetler, J. B., Nordberg, E., Shukla, M.,  
631 Walenz, B., Hill, C. A., Nene, V. M., Azad, A. F., Sobral, B. W., & Caler, E. (2012). A rickettsia  
632 genome overrun by mobile genetic elements provides insight into the acquisition of genes  
633 characteristic of an obligate intracellular lifestyle. *Journal of Bacteriology*, 194(2), 376–394.  
634 <https://doi.org/10.1128/JB.06244-11>
- 635 Giorgini, M., Bernardo, U., Monti, M. M., Nappo, A. G., & Gebiola, M. (2010). *Rickettsia* symbionts  
636 cause parthenogenetic reproduction in the parasitoid wasp *pnigalio soemius* (hymenoptera:  
637 Eulophidae). *Applied and Environmental Microbiology*, 76(8), 2589–2599.  
638 <https://doi.org/10.1128/AEM.03154-09>
- 639 Hagen, R., Verhoeve, V. I., Gillespie, J. J., & Driscoll, T. P. (2018). Conjugative Transposons and Their  
640 Cargo Genes Vary across Natural Populations of *Rickettsia buchneri* Infecting the Tick *Ixodes*  
641 *scapularis*. *Genome Biology and Evolution*, 10(12), 3218–3229.  
642 <https://doi.org/10.1093/gbe/evy247>
- 643 Hawkins, J. P., Ordonez, P. A., & Oresnik, I. J. (2018). Characterization of Mutations That Affect the  
644 Nonoxidative Pentose Phosphate Pathway in *Sinorhizobium meliloti*. *Journal of Bacteriology*,  
645 200(2). <https://doi.org/10.1128/JB.00436-17>
- 646 Hendry, T. A., Hunter, M. S., & Baltrus, D. A. (2014). The facultative symbiont *Rickettsia* protects an  
647 invasive whitefly against entomopathogenic *Pseudomonas syringae* strains. *Applied and*  
648 *Environmental Microbiology*, 80(23), 7161–7168. <https://doi.org/10.1128/AEM.02447-14>
- 649 Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving  
650 the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, 35(2), 518–522.  
651 <https://doi.org/10.5281/zenodo.854445>
- 652 Hotopp, J. C. D., Clark, M. E., Oliveira, D. C. S. G., Foster, J. M., Fischer, P., Torres, M. C. M., Giebel, J.

- 653 D., Kumar, N., Ishmael, N., Wang, S., Ingram, J., Nene, R. V., Shepard, J., Tomkins, J., Richards,  
654 S., Spiro, D. J., Ghedin, E., Slatko, B. E., Tettelin, H., & Werren, J. H. (2007). Widespread Lateral  
655 Gene Transfer from Intracellular Bacteria to Multicellular Eukaryotes. *Science*, *317*(5845),  
656 1753–1756. <https://doi.org/10.1126/science.1142490>
- 657 Hurst, G. D. D., Purvis, E. L., Sloggett, J. J., & Majerus, M. E. N. (1994). The effect of infection with  
658 male-killing *Rickettsia* on the demography of female *Adalia bipunctata* L. (two spot ladybird).  
659 *Heredity*, *73*(3), 309–316. <https://doi.org/10.1038/hdy.1994.138>
- 660 Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput  
661 ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature*  
662 *Communications*, *9*(1), 5114. <https://doi.org/10.1038/s41467-018-07641-9>
- 663 Jiang, N., Dillon, F. M., Silva, A., Gomez-Cano, L., & Grotewold, E. (2021). Rhamnose in plants - from  
664 biosynthesis to diverse functions. *Plant Science*, *302*, 110687.  
665 <https://doi.org/10.1016/j.plantsci.2020.110687>
- 666 Jofré, E., Lagares, A., & Mori, G. (2004). Disruption of dTDP-rhamnose biosynthesis modifies  
667 lipopolysaccharide core, exopolysaccharide production, and root colonization in *Azospirillum*  
668 *brasiliense*. *FEMS Microbiology Letters*, *231*(2), 267–275. [https://doi.org/10.1016/S0378-  
669 1097\(04\)00003-5](https://doi.org/10.1016/S0378-1097(04)00003-5)
- 670 Kalyanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., & Jermiin, L. S. (2017).  
671 ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*,  
672 *14*(6), 587–589. <https://doi.org/10.1038/nmeth.4285>
- 673 Kawafune, K., Hongoh, Y., Hamaji, T., Sakamoto, T., Kurata, T., Hirooka, S., Miyagishima, S., &  
674 Nozaki, H. (2015). Two Different *Rickettsial* Bacteria Invading *Volvox carteri*. *PLOS ONE*, *10*(2),  
675 e0116192. <https://doi.org/10.1371/JOURNAL.PONE.0116192>
- 676 Kikuchi, Y., & Fukatsu, T. (2005). *Rickettsia* Infection in Natural Leech Populations. *Microbial*  
677 *Ecology*, *49*(2), 265–271. <https://doi.org/10.1007/s00248-004-0140-5>
- 678 Klimaszewski, J., Morency, M.-J., Labrie, P., Seguin, A., Langor, D., Work, T., Bourdon, C., Thiffault,  
679 E., Pare, D., Newton, A., & Thayer, M. (2013). Molecular and microscopic analysis of the gut  
680 contents of abundant rove beetle species (Coleoptera, Staphylinidae) in the boreal balsam fir  
681 forest of Quebec, Canada. *ZooKeys*, *353*, 1–24. <https://doi.org/10.3897/zookeys.353.5991>

- 682 Kolde, R. (2019). *heatmap: Pretty Heatmaps. R package* (1.0.12).
- 683 Kontsedalov, S., Zchori-Fein, E., Chiel, E., Gottlieb, Y., Inbar, M., & Ghanim, M. (2008). The presence  
684 of Rickettsia is associated with increased susceptibility of Bemisia tabaci (Homoptera:  
685 Aleyrodidae) to insecticides. *Pest Management Science*, *64*(8), 789–792.  
686 <https://doi.org/10.1002/ps.1595>
- 687 Krassowski, M., Arts, M., & CyrilLagger. (2020). *ComplexUpset* (1.3.1).  
688 <https://doi.org/10.5281/zenodo.3700590>
- 689 KÜchler, S. M., Kehl, S., & Dettner, K. (2009). Characterization and localization of Rickettsia sp. in  
690 water beetles of genus Deronectes (Coleoptera: Dytiscidae). *FEMS Microbiology Ecology*,  
691 *68*(2), 201–211. <https://doi.org/10.1111/j.1574-6941.2009.00665.x>
- 692 Lanzoni, O., Sabaneyeva, E., Modeo, L., Castelli, M., Lebedeva, N., Verni, F., Schrallhammer, M.,  
693 Potekhin, A., & Petroni, G. (2019). Diversity and environmental distribution of the  
694 cosmopolitan endosymbiont “Candidatus Megaira”. *Scientific Reports*, *9*(1), 1179.  
695 <https://doi.org/10.1038/s41598-018-37629-w>
- 696 Leclair, M., Polin, S., Jousseume, T., Simon, J.-C., Sugio, A., Morlière, S., Fukatsu, T., Tsuchida, T., &  
697 Outreman, Y. (2017). Consequences of coinfection with protective symbionts on the host  
698 phenotype and symbiont titres in the pea aphid system. *Insect Science*, *24*(5), 798–808.  
699 <https://doi.org/10.1111/1744-7917.12380>
- 700 Lex, A., Gehlenborg, N., Strobel, H., Vuillemot, R., & Pfister, H. (2014). UpSet: Visualization of  
701 intersecting sets. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 1983–  
702 1992. <https://doi.org/10.1109/TVCG.2014.2346248>
- 703 Łukasik, P., Guo, H., van Asch, M., Ferrari, J., & Godfray, H. C. J. (2013). Protection against a fungal  
704 pathogen conferred by the aphid facultative endosymbionts Rickettsia and Spiroplasma is  
705 expressed in multiple host genotypes and species and is not influenced by co-infection with  
706 another symbiont. *Journal of Evolutionary Biology*, *26*(12), 2654–2661.  
707 <https://doi.org/10.1111/jeb.12260>
- 708 Murray, G. G. R., Weinert, L. A., Rhule, E. L., & Welch, J. J. (2016). The Phylogeny of Rickettsia Using  
709 Different Evolutionary Signatures: How Tree-Like is Bacterial Evolution? *Systematic Biology*,  
710 *65*(2), 265–279. <https://doi.org/10.1093/sysbio/syv084>

- 711 Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective  
712 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and*  
713 *Evolution*, 32(1), 268–274. <https://doi.org/10.1093/molbev/msu300>
- 714 Oliver, K. M., Degnan, P. H., Burke, G. R., & Moran, N. A. (2010). Facultative Symbionts in Aphids  
715 and the Horizontal Transfer of Ecologically Important Traits. *Annual Review of Entomology*,  
716 55(1), 247–266. <https://doi.org/10.1146/annurev-ento-112408-085305>
- 717 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM:  
718 assessing the quality of microbial genomes recovered from isolates, single cells, and  
719 metagenomes. *Genome Research*, 25(7), 1043–1055. <https://doi.org/10.1101/gr.186072.114>
- 720 Perlman, S. J., Hunter, M. S., & Zchori-Fein, E. (2006). The emerging diversity of Rickettsia.  
721 *Proceedings of the Royal Society B: Biological Sciences*, 273(1598), 2097–2106.  
722 <https://doi.org/10.1098/rspb.2006.3541>
- 723 Pilgrim, J., Ander, M., Garros, C., Baylis, M., Hurst, G. D. D., & Siozios, S. (2017). Torix group  
724 Rickettsia are widespread in Culicoides biting midges (Diptera: Ceratopogonidae), reach high  
725 frequency and carry unique genomic features. *Environmental Microbiology*, 19(10), 4238–  
726 4255. <https://doi.org/10.1111/1462-2920.13887>
- 727 Pilgrim, J., Thongprem, P., Davison, H. R., Siozios, S., Baylis, M., Zakharov, E. V., Ratnasingham, S.,  
728 deWaard, J. R., Macadam, C. R., Smith, M. A., & Hurst, G. D. D. (2021). Torix Rickettsia are  
729 widespread in arthropods and reflect a neglected symbiosis. *GigaScience*, 10(3), 1–19.  
730 <https://doi.org/10.1093/gigascience/giab021>
- 731 Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G., & Toth, I. K. (2016). Genomics and  
732 taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens.  
733 *Analytical Methods*, 8(1), 12–24. <https://doi.org/10.1039/C5AY02550H>
- 734 R Core Team. (2020). *R: A Language and Environment for Statistical Computing* (3.4.4). R  
735 Foundation for Statistical Computing. <https://www.r-project.org/>
- 736 Rodriguez-R, L. M., Jain, C., Conrad, R. E., Aluru, S., & Konstantinidis, K. T. (2021). Reply to: “Re-  
737 evaluating the evidence for a universal genetic boundary among microbial species.” *Nature*  
738 *Communications*, 12(1), 4060. <https://doi.org/10.1038/s41467-021-24129-1>

- 739 Schrallhammer, M., Ferrantini, F., Vannini, C., Galati, S., Schweikert, M., Görtz, H.-D., Verni, F., &  
740 Petroni, G. (2013). 'Candidatus Megaira polyxenophila' gen. nov., sp. nov.: Considerations on  
741 Evolutionary History, Host Range and Shift of Early Divergent Rickettsiae. *PLoS ONE*, *8*(8),  
742 e72581. <https://doi.org/10.1371/journal.pone.0072581>
- 743 Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068–  
744 2069. <https://doi.org/10.1093/bioinformatics/btu153>
- 745 Siozios, S. (2021). *SioStef/panplots: A small R script for generating pangenome accumulation curves*.  
746 <https://github.com/SioStef/panplots>
- 747 Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., Lomsadze, A.,  
748 Pruitt, K. D., Borodovsky, M., & Ostell, J. (2016). NCBI prokaryotic genome annotation pipeline.  
749 *Nucleic Acids Research*, *44*(14), 6614–6624. <https://doi.org/10.1093/nar/gkw569>
- 750 Thongprem, P., Evison, S. E. F., Hurst, G. D. D., & Otti, O. (2020). Transmission, Tropism, and  
751 Biological Impacts of *Torix Rickettsia* in the Common Bed Bug *Cimex lectularius* (Hemiptera:  
752 Cimicidae). *Frontiers in Microbiology*, *11*. <https://doi.org/10.3389/fmicb.2020.608763>
- 753 Tvedte, E. S., Walden, K. K. O., McElroy, K. E., Werren, J. H., Forbes, A. A., Hood, G. R., Logsdon, J.  
754 M., Feder, J. L., & Robertson, H. M. (2019). Genome of the Parasitoid Wasp *Diachasma*  
755 *alloeum*, an Emerging Model for Ecological Speciation and Transitions to Asexual  
756 Reproduction. *Genome Biology and Evolution*, *11*(10), 2767–2773.  
757 <https://doi.org/10.1093/gbe/evz205>
- 758 van der Beek, S. L., Zorzoli, A., Çanak, E., Chapman, R. N., Lucas, K., Meyer, B. H., Evangelopoulos,  
759 D., de Carvalho, L. P. S., Boons, G., Dorfmüller, H. C., & van Sorge, N. M. (2019). Streptococcal  
760 dTDP-L-rhamnose biosynthesis enzymes: functional characterization and lead compound  
761 identification. *Molecular Microbiology*, *111*(4), 951–964. <https://doi.org/10.1111/mmi.14197>
- 762 Weinert, L. A., Araujo-Jnr, E. V., Ahmed, M. Z., & Welch, J. J. (2015). The incidence of bacterial  
763 endosymbionts in terrestrial arthropods. *Proceedings of the Royal Society B: Biological*  
764 *Sciences*, *282*(1807), 20150249. <https://doi.org/10.1098/rspb.2015.0249>
- 765 Weinert, L. A., Werren, J. H., Aebi, A., Stone, G. N., & Jiggins, F. M. (2009). Evolution and diversity of  
766 Rickettsiabacteria. *BMC Biology*, *7*(1), 6. <https://doi.org/10.1186/1741-7007-7-6>

- 767 Werren, J. H., Hurst, G. D., Zhang, W., Breeuwer, J. A., Stouthamer, R., & Majerus, M. E. (1994).  
768 Rickettsial relative associated with male killing in the ladybird beetle (*Adalia bipunctata*).  
769 *Journal of Bacteriology*, 176(2), 388–394. <https://doi.org/10.1128/jb.176.2.388-394.1994>
- 770 Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.  
771 <https://ggplot2.tidyverse.org>
- 772 Zchori-Fein, E., Borad, C., & Harari, A. R. (2006). Oogenesis in the date stone beetle, *Coccotrypes*  
773 *dactyliperda*, depends on symbiotic bacteria. *Physiological Entomology*, 31(2), 164–169.  
774 <https://doi.org/10.1111/j.1365-3032.2006.00504.x>
- 775