Metagenomic profiling of ticks: identification of novel rickettsial genomes and detection of tick-borne canine parvovirus

Anuradha Ravi[1¶], Suheir Ereqat *[2,3¶], Amer Al-Jawabreh[4], Ziad Abdeen[3], Omar Abu Shamma[3], Holly Hall[5], Mark J. Pallen[1&], Abedelmajeed Nasereddin[3&]

[1]Quadram Institute, Norwich Research Park, Norwich, Norfolk, NR4 7UA, UK

[2]Biochemistry and Molecular Biology Department, Faculty of Medicine, Al-Quds University, Abu Deis, East Jerusalem, Palestine

[3]Al-Quds Nutrition and Health Research Institute, Faculty of Medicine, Al-Quds University Abu Deis, East Jerusalem, Palestine and Al-Quds Public Health Society, Abu Deis, East Jerusalem, Palestine.

[4]Arab American University-Palestine

[5]Warwick Medical School, University of Warwick, Coventry, UK.

* Corresponding author

E-mail: sereqat@staff.alquds.edu

[¶]These authors contributed equally to this work.

[&]These authors also contributed equally to this work.

Abstract

Background: Across the world, ticks act as vectors of human and animal pathogens. Ticks rely on bacterial endosymbionts, which often share close and complex evolutionary links with tick-borne pathogens. As the prevalence, diversity and virulence potential of tick-borne agents remain poorly understood, there is a pressing need for microbial surveillance of ticks as potential disease vectors.

Methodology/Principal Findings: We developed a two-stage protocol that includes 16S-amplicon screening of pooled samples of hard ticks collected from dogs, sheep and camels in Palestine, followed by shotgun metagenomics on individual ticks to detect and characterise tick-borne pathogens and endosymbionts. Two ticks isolated from sheep yielded an abundance of reads from the genus *Rickettsia*, which were assembled into draft genomes. One of the resulting genomes was highly similar to *Rickettsia massiliae* strain MTU5. Analysis of signature genes showed that the other represents the first genome sequence of the potential pathogen *Candidatus* Rickettsia barbariae. Ticks from a dog and a sheep yielded draft genome sequences of strains of the Coxiella-like endosymbiont *Candidatus* Coxeilla mudrowiae. A sheep tick yielded sequences from the sheep pathogen *Anaplasma ovis*, while *Hyalomma* ticks from camels yielded sequences belonging to *Francisella*-like endosymbionts. From the metagenome of a dog tick from Jericho, we generated a genome sequence of a canine parvovirus.

Significance: Here, we have shown how a cost-effective two-stage protocol can be used to detect and characterise tick-borne pathogens and endosymbionts. In recovering genome sequences from an unexpected pathogen (canine parvovirus) and a previously unsequenced pathogen (*Candidatus* Rickettsia barbariae), we demonstrate the open-ended nature of metagenomics. We also provide evidence

that ticks can carry canine parvovirus, raising the possibility that ticks might contribute to the spread of this troublesome virus.

Author Summary

We have shown how DNA sequencing can be used to detect and characterise potentially pathogenic microorganisms carried by ticks. We surveyed hard ticks collected from domesticated animals across the West Bank territory of Palestine. All the ticks came from species that are also capable of feeding on humans. We detected several important pathogens, including two species of *Rickettsia*, the sheep pathogen *Anaplasma ovis* and canine parvovirus. These findings highlight the importance of hard ticks and the hazards they present for human and animal health in Palestine and the opportunities presented by high-throughput sequencing and bioinformatics analyses of DNA sequences in this setting.

Keywords

Ticks, metagenomics, tick-borne pathogens, tick endosymbionts, Palestine.

Introduction

Ticks are ectoparasitic arthropods that feed exclusively on the blood of their vertebrate hosts. Across the world, ticks act as vectors of human and animal pathogens (including viruses, bacteria and protozoa), often mediating transfer of infection from one host species to another, including zoonotic infections of humans [1–3]. Aside from negative effects on human and animal morbidity and mortality, ticks and tick-borne diseases are responsible for huge global production losses, amounting to US$ 14-19 billion per annum [4].

Many tick-borne infections remain undiagnosed and the prevalence, diversity and virulence potential of tick-borne agents remain poorly understood. Furthermore, as most tick-borne pathogens are hard to grow in the laboratory and the ticks themselves may be hard to identify on morphological grounds, our understanding of pathogen and tick population structure and the fine-grained genomic epidemiology of tick-borne infections remains patchy. These gaps in our knowledge highlight the need for microbial surveillance in tick vectors to assess the risk of infection in humans and animals.

Ticks rely on bacterial endosymbionts to provide nutrients lacking from their highly restricted blood-centred diet, as evidenced by the observation that ticks treated with antibiotics show decreased fitness at all life stages [5–7]. Interestingly, some intracellular endosymbionts of ticks share close and complex evolutionary links with tick-borne pathogens of humans, such as *Coxiella burnetii* and *Francisella tularensis,* with *Coxiella*-like and *Francisella*-like endosymbionts (CLEs and FLEs, respectively) occurring in a wide range of ticks worldwide [8–20].

92 Transitions between pathogenic and symbiotic lifestyles occur in both directions—
93 thus, *Coxiella burnetii* is thought to have evolved from a vertically transmitted tick
94 endosymbiont, whereas most FLEs probably evolved from pathogenic strains of
95 *Francisella* [11,13,21]. Interactions with symbionts influence the abundance of
96 pathogens within ticks. For example, *Francisella*-like endosymbionts inhibit
97 colonisation with the spotted fever group rickettsias and with *Francisella novicida*,
98 while the endosymbiont *Rickettsia bellii* is inversely associated in abundance with
99 *Anaplasma marginale* [22]. Similarly, in North American ticks, the obligate
100 intracellular endosymbiont *Rickettsia peacockii*, inhibits its pathogenic relative *R.*
101 *rickettsia*, the causative agent of Rocky Mountain Spotted Fever [23].
102
103 Surveillance of ticks using taxon-specific PCR assays has shed light on the
104 distribution of tick-borne pathogens and endosymbionts in Palestine and in the
105 neighbouring state of Israel. We detected *Bartonella* in nearly 4% of hard ticks and
106 spotted-fever-group *Rickettsia* in 17% of hard ticks collected from domesticated
107 animals in Palestine [24,25]. Similarly, we have found apicomplexan parasites
108 including *Theileria, Babesia*, and *Hepatozoon* in ixodid ticks from Palestine [26].
109 Various pathogens have been detected in ticks collected in Israel, including *Ehrlichia*
110 *canis, Anaplasma spp., Babesia canis vogeli, Rickettsia massiliae, R. sibirica*
111 *mongolitimonae, R. africae, R. aeschlimannii,* [27–31]. *Rickettsia massiliae* is widely
112 disseminated throughout Israel in questing ticks, unattached to animals [32].
113 Endosymbionts have also been described in ticks from Israel, including the
114 mitochondrial endosymbiont *Midichloria mitochondrii* [30]. *Francisella*-like
115 endosymbionts have been reported in over 90% of *Hyalomma* ticks obtained from
116 domesticated animals and migratory birds in Israel [12], while *Coxiella*-like
117 endosymbionts have been identified and characterised in local *Rhipicephalus* ticks
118 [33,34].
119

120 Although amplification and sequencing of taxon-specific PCR products has proven
121 useful in microbial surveillance of ticks, this approach provides limited information on
122 the target microorganisms, plus it only finds what has been targeted in the assay.
123 Use of molecular barcodes such as 16S ribosomal gene sequences allows a more
124 open-ended approach, capable of detecting numerous species in the same sample
125 [24]. However, this approach provides limited insights into the biology, evolution and
126 spread of the species and lineages under examination. The advent of high-
127 throughput sequencing offers the potential for a powerful new open-ended
128 metagenomics approach to the identification and characterisation of pathogens and
129 endosymbionts [35]. This approach has already been used to identify novel viruses
130 and to detect bacterial pathogens and symbionts in ticks [34,36]. It also brings the
131 promise of sequence-based identification of ticks and hosts. However, widespread
132 application of this approach is still limited by cost.
133
134 Here, we present a two-stage tick surveillance strategy that combines an initial round
135 of 16S-amplicon screening of pooled tick samples, followed by focused shotgun
136 metagenomics investigations, which confirmed the identities of the ticks as well as
137 providing phylogenomic information on tick pathogens and endosymbionts, including
138 the first genome sequence of the potential pathogen *Candidatus* Rickettsia
139 barbariae.
140

141 Materials and Methods

142 Study design and sample collection

143 We employed a two-stage protocol that includes 16S-amplicon screening of pooled
144 tick samples followed by focused shotgun metagenomics investigations on individual
145 ticks (Fig. 1).

146

147 **Fig. 1 Study workflow**

148

149 Hard ticks were collected in 2015 from domesticated animals from seven
150 governorates of the West Bank, Palestine, including: two camels from Jericho; four
151 sheep from Nablus; three sheep from Tubas; and 23 dogs from all seven study sites
152 (Fig. 2). Ticks were gently removed from host animals by forceps or hand and
153 individually placed into small, labeled plastic tubes containing 70% ethanol for
154 morphological identification. Ticks were identified using standard taxonomic keys and
155 stored at −20°C until DNA extraction [24]. The animal owners were verbally informed
156 about the study objectives and sampling procedure. All owners gave their verbal
157 informed consent to collect ticks from their animals. The ethics committee at Al-Quds
158 University approved the study (EC number: ZA/196/013).

159

160 **Figure 2 Collection sites of pooled tick specimens from the West Bank,**
161 **Palestine**. Species identities according to morphological criteria. **Hyalomma*
162 *dromedarii* according to sequence identity; ∞unidentified *Haemaphysalis* sp.
163 according to sequence identity; # Nablus sheep tick 2 is *Rhipicephalus sanguineus*
164 according to sequence identity.

165

166 DNA Extraction

167 Individual ticks were washed with phosphate-buffered saline, air dried for 10 min on
168 tissue paper and separately sliced into small pieces by a sterile scalpel blade. They
169 were then manually homogenized with a sterile micro-pestle, re-suspended in 200 µl
170 of lysis buffer and 20 µl of proteinase K (20 µg/ml stock solution). After overnight
171 incubation at 56°C with continuous gentle shaking, DNA was extracted using the

172 QIAamp DNA tissue extraction kit (Qiagen, Hilden, Germany), following the
173 manufacturer's protocol. Eluted DNA (100 µl) was stored at -20°C until further use.
174

175 16S amplicon analyses

176 Forty tick-DNA pools were created, each containing equal concentrations of DNA
177 from five ticks isolated from the same region, host and tick species. Tick pools were
178 named after region and host and then given numerical identities (e.g. Nablus Dog
179 Tick Pool 1, Nablus Dog Tick Pool 2). Tick pools were subjected to 16S rRNA
180 sequence analyses on the Illumina platform, according to the manufacturer's
181 instructions [37]. In brief, V3–V4 hyper-variable regions were amplified using primers
182 tagged with Illumina-specific primers. Amplification was performed using X2 KAPA
183 HiFi HotStart ReadyMix (Kappa Biosystems). PCR conditions were as follows: 95°C
184 for 3 minutes, followed by 25 cycles of: 95°C for 30 seconds, 55°C for 30 seconds,
185 72°C for 30 seconds, then 72°C for 5 minutes and final hold at 4°C. Negative controls
186 containing nuclease-free water were used in each PCR run. PCR products were
187 purified using AMPure XP beads (X0.8) followed by the second round of amplification
188 that introduces the index sequences using the Nextera XT Index Kit (Illumina).
189 Amplicons were sequenced according to the manufacturer's recommendations. The
190 PhiX control (15%) was added to the denatured library. Paired-end sequencing was
191 carried out on the Illumina MiSeq platform.
192

193 16S rRNA amplicon sequences were analysed via the QIIME1 pipeline [38].
194 Amplicon sequences were demultiplexed into samples by barcode, then pre-
195 processed and quality-filtered, using QIIME's multiple_join_paired_ends.py and
196 split_libraries_fastq.py scripts, accepting only those with a quality score below 19
197 and a minimum sequence length of 350 bp. Sequences passing the quality filter were
198 clustered at 97% homology using QIIME's pick_open_reference_otus.py and

199 pick_otus.py scripts with the option “enable_rev_strand_match True”. Operational
200 taxonomic units (OTUs) were defined at 97% identity and then taxonomically
201 assigned using the Greengenes 13_8 database [39]. To remove spurious
202 sequences, OTUs that occurred only once in the data were removed. We then
203 calculated relative abundances for each OTU across the taxonomic levels from
204 phylum to family, focusing on taxonomic sequence identifications relevant to tick-
205 borne potential pathogens and endosymbionts (S1 Table).
206

207 Shotgun metagenomic sequencing

208 Fourteen individual tick DNA samples were selected for follow-up by shotgun
209 metagenomics, named after the pool they belonged to with a final numerical identifier
210 for each tick, e.g. Nablus sheep tick 4.1, Nablus sheep tick 4.2. DNA sequencing
211 libraries were prepared using the Nextera XT kit (Illumina), with DNA fragmented,
212 tagged, cleaned and normalized according to the manufacturer’s recommendations.
213 The quality of DNA in each library was evaluated using the Agilent 2200 TapeStation
214 (Agilent) and the quantity was measured using the Qubit (Invitrogen). Libraries were
215 sequenced on a NextSeq 500 with the mid-output setting and single-end reads. We
216 analysed shotgun metagenome sequences using the Cloud Infrastructure for
217 Microbial Bioinformatics [40]. Sequences were assessed for quality using FastQC
218 version 0.11.5 [41] and quality-filtered using Trimmomatic version 0.35 with default
219 parameters [42]. Trimmomatic’s Illuminaclip function was used to clip off Illumina
220 adapters. Metagenome sequences were deposited in the Sequence Read Archive
221 under Bioproject XYZ.
222

223 Sequence-based identification of tick and host

224 Reads from each metagenome were assembled into contigs and then subjected to
225 BLAST searches with mitochondrial signature genes from the potential tick and host

226 species (S2 Table). Coding sequences from the metagenomic assemblies that had
227 been identified from BLAST hits were then subjected to a second round of BLAST
228 searches to determine the percentage identity to signature genes from relevant taxa.
229

230 Taxonomic profiling of tick metagenomes

231 We performed a taxonomic assignment of reads in each sample using the Centrifuge
232 metagenome classifier Version 1.0.3 together with its associated p+h+v database
233 [43]. Centrifuge reports were converted to Kraken-style reports using Centrifuge's
234 kreport parameter and Pavian was used to visualize the reports [44]. Taxa with a
235 relative abundance of <5% were discarded as potential kit contaminants and we then
236 evaluated the reports for the presence of potential pathogens and endosymbionts.
237

238 Assembly of microbial genomes

239 We constructed a set of completed reference genomes for each genus of potential
240 pathogen or endosymbionts identified in at least one of our samples using Centrifuge
241 (S3 Table). Reference sequences were downloaded using the ncbi-genome-
242 download script [45]. We then mapped the metagenome from each sample against
243 all the reference genomes using BowTie2 version 2.3.4.1 [46]. We converted SAM
244 files to sorted BAM files using SAMtools and visualized stats using Qualimap 2
245 [47,48]. To allow us to bin species-specific reads for each microorganism of interest
246 from the metagenomic sequences, the set of metagenomic reads from each sample
247 that mapped to a taxon of interest were assembled into contigs using SPAdes
248 (version 3.11.1) [49] and annotated using Prokka (version 1.12). We then used
249 taxonomic profiles from Centrifuge for each contig to confirm the specificity of our
250 binning approach. The coverage of the resulting draft genome sequences was
251 calculated after mapping reads back to the assemblies using BowTie2. We
252 compared the gene content between our metagenome-derived draft bacterial

253 genomes and related reference genomes using Roary [50]. To confirm species
254 identity, average nucleotide identity was calculated from BLAST searches [51] or by
255 using the online ANI/AAI matrix tool [52].
256

257 Phylogenomic analyses

258 We aligned signature genes from selected microbial taxa (*Supplementary files*) with
259 homologues in metagenome-derived genome sequences using MAFFT (version
260 7.305b) [53] and performed maximum likelihood phylogenetic reconstructions using
261 RAxML (version 8.2.12) with 1000 bootstraps [54]. MEGA7 was used to visualize the
262 trees [55]. FastTree2.1 was used to construct phylogenetic trees using Generalized
263 time reversible (GTR) and the CAT approximation (estimated rate of evolution for
264 each site) model [56]. Bootstrapping with 1000 iterations was used to evaluate the
265 significance of branches within phylogenetic trees. SNP distance matrices between
266 assemblies and reference genomes were calculated using Snippy version 3.1
267 incorporating Freebayes version 1.1.0 for SNP detection [57,58].
268

269 Results

270 Taxonomic profiling of pooled tick samples

271 Two hundred hard ticks were collected from diverse animal hosts (23 dogs, 7 sheep,
272 and 2 camels) and were identified morphologically to the species level (Fig. 2). PCR-
273 amplification of 16S rRNA sequences was attempted on 40 pools of tick DNA. No
274 products were obtained from eight of the pools. Sequencing of amplification products
275 from the remaining 32 pools identified the major residents of the tick microbiome at
276 the level of bacterial family (S1 Table). Four families harbouring potential pathogens
277 or endosymbionts accounted for around 45% of all such sequences: the

278 *Coxiellaceae, the Francisellaceae, the Rickettsiaceae* and the *Anaplasmataceae*
279 (Fig. 3).

280 **Fig. 3 Relative abundances of potential pathogens and endosymbionts in**
281 **pooled tick samples.**

282

283 As expected from previous reports [33,34], the *Coxiellaceae* were detectable in all
284 dog and sheep ticks, although their relative abundance varied from <10% to >80%.
285 The *Rickettsiaceae* were detected as major residents (12–83% of reads) in all the
286 sheep tick pools and in one tick pool derived from dogs. One sheep-derived tick pool
287 also contained reads from the Anaplasmataceae. The two pools derived from ticks
288 isolated from camels were, in line with previous work [12], dominated by the
289 *Francisellaceae,* but lacked sequences from the *Coxiellaceae*.

290

291 Metagenomic profiling of individual ticks

292 Fourteen ticks were selected for follow-up studies on the basis of the 16S results:
293 two camel ticks, five sheep ticks, six dog ticks from the genus *Rhipicephalus* and two
294 dog ticks from the genus *Haemaphysalis*. Shotgun metagenome sequencing of DNA
295 from the fourteen ticks generated 15.8 million single-end reads, with >98% of reads
296 from each sample surviving trimming (S5 Table).

297

298 Sequence-based identification of the ticks based on mitochondrial signature gene
299 sequences generally confirmed the morphological identifications, with closest BLAST
300 hits typically showing sequence identities of >99% to sequences assigned to the
301 relevant tick species (Table 1). However, there were three discrepancies: Jericho
302 Camel Tick 2.1 identified on morphological grounds as *Hyalomma impeltatum* was
303 re-assigned on grounds of sequence identity to *Hyalomma dromedarii*; Nablus Sheep
304 Tick 2.1 was identified morphologically as *Rhipicephalus turanicus,* but showed

305 greatest sequence similarity to *Rhipicephalus sanguineus* dog ticks from Egypt and
306 Iran; and Ramallah Dog Ticks 1.1 and 1.2, assigned to the species *Haemaphysalis*
307 *parva* by morphology, both delivered a highest-scoring BLAST hit to a sequence from
308 the species *Haemaphysalis concinna*, but the low level of sequence identity (88%)
309 suggests that these ticks belong to a previously uncharacterised species within the
310 genus *Haemaphysalis*. Identification of the tick host via mitochondrial signature gene
311 sequences proved successful in six of the fourteen individual ticks (Table 1). This
312 varied success presumably reflects the variation in the state of engorgement of the
313 tick with the host's blood.
314

Tick	**Tick ID by morphology**	**Tick ID by sequence (% identity to closest hit)**	**Host ID by sequence**	**Key microbes in taxonomic profile**	**Microbe ID from metagenome**
Jericho Camel Tick 2.1	*Hyalomma impeltatum*	*Hyalomma dromedarii* (100)	undetermined	*Francisella*	*Francisella persica*
Nablus Sheep Tick 1.1	*R. turanicus*	*R. turanicus* (99.2)	undetermined	*Coxiella, Rickettsia*	*Rickettsia barbariae* strain Rb Nablus
Nablus Sheep Tick 2.1	*R. turanicus*	*R. sanguineus* (99.1)	*Ovis aries*	*Coxiella*	undetermined
Nablus Sheep Tick 3.1	*R. turanicus*	*R. turanicus* (99.2)	undetermined	*Coxiella, Rickettsia*	*Coxeilla mudrowiae* strain CRt Nablus *Rickettsia massiliae* strain Rm Nablus
Tubas Sheep Tick 3.1	*R. turanicus*	*R. turanicus* (99.3)	undetermined	*Coxiella, Anaplasma*	*Anaplasma ovis*
Tubas Sheep Tick 3.2	*R. turanicus*	*R. turanicus* (99.2)	undetermined	*Coxiella, Anaplasma*	*Anaplasma ovis*
Hebron Dog Tick 8.1	*R. sanguineus*	*R. sanguineus* (99.4)	*Canis lupus familiaris*	*Coxiella*	undetermined
Ramallah Dog Tick 1.1	*Haemaphysalis parva*	*Haemaphysalis concinna* (87.7)	undetermined	*Coxiella*	undetermined
Ramallah Dog Tick 1.2	*Haemaphysalis parva*	*Haemaphysalis concinna* (87.8)	*Canis lupus familiaris*	*Coxiella*	undetermined
Nablus Dog Tick 1.1	*R. sanguineus*	*R. sanguineus* (99.4)	undetermined	*Coxiella*	undetermined
Nablus Dog Tick 1.2	*R. sanguineus*	*R. sanguineus* (99.2)	*Canis lupus familiaris*	*Coxiella*	undetermined
Tubas Dog Tick 2.1	*R. sanguineus*	*R. sanguineus* (99.2)	*Canis lupus familiaris*	*Coxiella, canine parvovirus*	Canine parvovirus
Tubas Dog Tick 3.1	*R. sanguineus*	*R. sanguineus* (99.3)	undetermined	*Coxiella*	*Coxeilla mudrowiae* strain CRs Tubas
Tubas Dog Tick 3.2	*R. sanguineus*	*R. sanguineus* (99.9)	*Canis lupus familiaris*	*Coxiella*	undetermined

Table 1. Metagenomic analyses of individual tick samples from Palestine

319

320 When subjected to taxonomic profiling, 12.6 % of the metagenomic reads were
321 classified as bacterial and 0.015 % of reads as viral (S7 Table). The taxonomic
322 profiles confirmed the presence of the bacterial genera *Coxiella*, *Francisella*,
323 *Rickettsia* and *Anaplasma* in the relevant ticks (Fig. 4). In addition, reads from canine
324 parvovirus were identified in Tubas Dog Tick 2.1.

325

326 **Fig. 4. Relative abundance of potential pathogens and endosymbionts in**
327 **individual tick samples**

328

329 Phylogenomic characterisation of tick-borne potential
330 pathogens

331 Taxonomic profiling showed a high abundance of reads assigned to the genus
332 *Rickettsia* in Nablus Sheep Tick 1.1 and Nablus Sheep Tick 3.1, which are both *R.*
333 *turanicus*. In Nablus Sheep Tick 3.1, the *Rickettsia* reads (~8%) were outnumbered
334 by those from *Coxiella* (~30%). However in Nablus Sheep Tick 1.1, the relative
335 abundance of *Rickettsia* reads (~25%) was much higher than for *Coxiella* (~2%).
336 Metagenomic reads from these two ticks were mapped against a collection of all
337 complete rickettsial genome sequences (S3 Table). Reads mapping to any of the
338 rickettsial genomes were extracted from each sample and were assembled into a
339 draft genome sequence.

340

341 Rickettsial reads from Nablus Sheep Tick 3.1 assembled into 380 contigs, with an
342 average depth of coverage of 13X, a genome size of 1,342,424 base pairs and a GC
343 content of 32.73%. Comparisons to the *Rickettsia* reference genomes showed that
344 this metagenome-derived genome (which we term “Rm Nablus”) was highly similar to
345 the genome of *R. massiliae* strain MTU5 from Marseille, France [59], showing

346 extensive synteny and just 127 SNPs spanning the whole genome (Fig. 5, S8 Table).
347 The two strains also share a plasmid, *pRMA*. This high degree of genome
348 conservation between *R. massiliae* strains from opposite ends of the Mediterranean
349 region seems remarkable. However, both MTU5 and Rm Nablus strain are far less
350 closely related (SNP distance >16000) to the only other genome-sequenced strain
351 assigned to the species *R. massiliae*—the AZT80 strain from Arizona [60]. Instead,
352 our comparisons show that this US isolate is much more closely related (just 4 SNPs
353 different) to the strain ECT, which has been assigned to the species *R. rhipicephali*.
354

355 **Fig. 5. Genome comparisons for *Rickettsia massiliae* strain Rm Nablus**
356 *From centre outwards: Rickettsia massiliae* strain MTU5 (black circle:
357 GCF_000016625); GC content; GC skew; *Rickettsia rhipicephali* strain ECT
358 (GCF_000964905), *Rickettsia massiliae* strain AZT80; *Rickettsia massiliae* strain Rm
359 Nablus (this study).
360

361 Rickettsial reads from Nablus Sheep Tick 1.1 assembled into 47 contigs with an
362 average depth of coverage of 14X, a genome size of 1,246,042 base pairs and GC
363 content of 32.4%. Comparison with other rickettsial genome sequences showed that
364 the closest genome-sequenced relatives of the *Rickettsia* from Tick 9 are *R. africae*
365 str. ESF-5 and *R. parkeri* str. AT24. However, each are separated from this genome
366 by >10,000 SNPs (S2 Figure), suggesting that it represents a distinct species for
367 which no genome sequence yet exists. With that in mind, we performed comparisons
368 to sequences from the unidentified genome to a database of rickettsial signature
369 gene sequences that have been widely used to identify isolates to species level.
370 Here, we found 100% identity with the 16S, *gltA* and *recA* genes from *Candidatus* R.
371 barbariae and 582/584 residues identical in *ompB*. This leads us to conclude that our
372 metagenome-derived assembly (which we term “Rb Nablus”) represents the first
373 draft genome sequence from this species. Comparisons with the *R. africae* and *R.*

374 *parkeri* genomes suggest extensive synteny between all three genomes. The only
375 major exception appears to be the variable *tra* region, first identified in *R. massiliae*
376 [59], where *Candidatus* R. barbariae lacks the *tra* genes found in the other two
377 species (Fig. 6).
378
379 **Fig. 6: Genome comparisons for *Candidatus* Rickettsia barbariae strain Rb**
380 **Nablus**
381 *From centre outwards: Rickettsia parkeri* strain Tates Hell (black circle: NCBI ID
382 GCF_000965145); GC content; GC skew; *Rickettsia africae* strain ESF-3
383 (GCF_000023005); *Candidatus* Rickettsia barbariae strain Rb Nablus (this study).
384
385 Taxonomic profiling indicated a high abundance of reads assigned to the genus
386 *Anaplasma* in Tubas Sheep Ticks 1.1 and 1.2. Metagenomics reads from these two
387 ticks were mapped to a set of 44 completed *Anaplasma* genomes (S3 table). Reads
388 mapping to any *Anaplasma* genome were extracted and assembled into contigs. The
389 *Anaplasma* assembly from Tubas Sheep Tick 1.1 contained 155 contigs with a total
390 length of 312,716 bp, while that from Tubas Sheep Tick 1.2 contained 323 contigs
391 with a total length of 497,831 bp. BLAST comparisons to the *Anaplasma* reference
392 genomes showed that contigs from both ticks were highly similar (>99% sequence
393 identity) to the genome of *Anaplasma ovis*.
394
395 As some reads from Tubas Dog Tick 2.1 were taxonomically assigned to canine
396 parvovirus, we mapped this metagenome against the reference sequence for this
397 virus (NCBI Reference Sequence: NC_001539.1) and obtained 11X coverage of
398 canine parvovirus strain Tubas. BLAST comparisons with other canine parvovirus
399 gene sequences showed that the Tubas Dog Tick 2.1 virus was most closely related
400 to the lineage 2b strain CPV-LZ2 from Lanzhou in North-West China.
401

402

403 ## Phylogenomic characterisation of tick endosymbionts

404 Taxonomic profiling showed a high abundance (>30%) of reads assigned to the
405 genus *Coxiella* in Nablus Sheep Tick 3.1 and Tubas Dog Tick 3.1, so we mapped
406 metagenomic reads from these ticks against a set of all completed *Coxiella* genomes
407 (S3 Table). We retrieved all threads that mapped to any *Coxiella* genome from these
408 samples and, as before, performed genome assemblies.

409

410 *De novo* assembly of *Coxiella* reads from Nablus Sheep Tick 3.1—a specimen of *R.*
411 *turanicus*—resulted in 35 contigs with a combined size of 1,581,648 base pairs and a
412 GC content of 38.2% and a depth of coverage of 37X. Phylogenetic analyses
413 showed that the resulting draft genome (which we term "strain CRt Nablus") was
414 almost identical (with only 294 SNPs spanning the whole genome) to the Coxiella-
415 like endosymbiont *Candidatus* Coxiella mudrowiae, strain CRt (ASM107771v1),
416 derived from a *R. turanicus* sample collected at Kibbutz Hulda, Israel, 56 km from
417 Nablus.

418

419 **Fig. 8 Genome comparisons for** *Candidatus* Coxiella mudrowiae **strains CRt**
420 **Nablus and CRs Tubas.** Fig. 8 Legend.

421

422 *De novo* assembly of Coxiella reads from Tubas Dog Tick 3.1 resulted in 177 contigs
423 with a combined size of 1,529,439 base pairs and a GC content of 38.1%. The
424 resulting draft genome (which we term "strain CRs Tubas"; Fig. 8) is relatively
425 distantly related (>20,000 SNPs or an average nucleotide identity of only 98%) to the
426 only other known genome of a *Coxiella*-like endosymbiont from *Rhipicephalus*
427 *sanginueus*—strain CRs (NZ_CP024961.1)—which was derived from a sample
428 collected in Caesarea, Israel, 48 km from Tubas. However, signature genes from our

429 strain CRs Tubas all showed >99.5% nucleotide identity to those from other *Coxiella*-
430 like endosymbionts from *Rhipicephalus turanicus*, while they showed only ≤96%
431 identity to the same genes from similar endosymbionts derived from *Rhipicephalus*
432 *sanguineus* (S. 10 Table)—suggesting that endosymbionts from the these tick
433 species are vertically inherited and co-evolving with their hosts.
434
435 Taxonomic profiling of metagenomic reads from Jericho Camel Tick 2.1, a specimen
436 of the species *Hyalomma dromedarii,* confirmed the presence of *Francisella*-like
437 endosymbionts. However, when these reads were mapped against relevant
438 *Francisella* genomes the depth of coverage was insufficient for further phylogenetic
439 analysis.
440

441 Discussion

442 Here, we have shown how a cost-effective combination of 16S rRNA sequence
443 analysis and shotgun metagenomics can be used to detect and characterise tick-
444 borne pathogens and endosymbionts. According to morphological criteria, we
445 surveyed ticks from four different taxa [61]: the brown dog tick *Rhipicephalus*
446 *sanguineus,* which has a global distribution [62]; its close relative *Rhipicephalus*
447 *turanicus,* which we collected from sheep*;* a tick from the genus *Haemaphysalis,*
448 collected from a dog; and *Hyalomma* samples from camels*.* All these taxa have been
449 reported to feed on humans [61]. Interestingly, metagenomic sequencing allowed us
450 to confirm or refine the species identities assigned to the ticks on morphological
451 grounds, as well as in some cases confirming the identity of the host.
452
453 Mindful of the potential for reagent-based contamination of low-biomass samples
454 [63], here we have focused on taxa already known to contain tick-borne pathogens
455 and endosymbionts, rather than attempting to provide a definitive survey of tick

456 microbiomes. Our metagenomics strategy has delivered three important benefits.
457 First, in using sequence-based methods to confirm the identities of ticks (and their
458 hosts), we avoid the problems of using morphological criteria alone, which can be
459 problematic in tick identification and classification [61]. Second, in addition to simply
460 detecting or identifying tick-borne pathogens and endosymbionts, we, like others
461 [34,36], have used shotgun metagenomic sequencing to obtain in-depth genome-
462 resolved characterisation of key bacteria from this setting. Third, in recovering
463 genome sequences from an unexpected pathogens (canine parvovirus) and a
464 previously unsequenced pathogen (*R. barbariae*), we have demonstrated the open-
465 ended nature of diagnostic metagenomics [35].
466
467 *Rickettsia massiliae* has caused tick-borne spotted fever in humans in the Europe
468 and South America [64–66]. Although no cases of human infection have been
469 described in the Levant, this pathogen has been detected in ticks collected in
470 Palestine and in Israel [24,28,29,31,32]. *Candidatus* R. barbariae belongs to the
471 spotted-fever group of rickettsiae [67], but has yet to be associated with human
472 infection. However, it has been found in human-parasitising ticks [68] and may well
473 be pathogenic for humans. This candidate species has been detected previously in
474 ticks from Palestine, from Israel and from Lebanon [24,29,32,69].
475
476 Here, we provide a third draft genome from *Rickettsia massiliae* and the first draft
477 genome from *Candidatus* R. barbariae, which confirms its close similarity to the
478 established human pathogens, *R. africae* and *R. parkeri* and provides a starting point
479 for future population genomics studies of this potential pathogen. Comparison with
480 pre-existing examples reveal that our *Rickettsia massiliae* Rm Nablus strain shows
481 far greater sequence similarity to the MTU5 strain from Marseille than to the AZT50
482 strain from Arizona, hinting at an Old World/New World phylogeographic population

483 structure for this pathogen, which awaits clarification from additional genome
484 sequences.
485
486 Coxiella-like endosymbionts of *Rhipicephalus sanguineus* and *R. turanicus* have
487 been recently assigned to a new candidate species *Candidatus* C. mudrowiae, on
488 the basis of genome sequences derived from pools of ticks isolated in Israel [33,34].
489 We present genome sequences of *Candidatus* C. mudrowiae derived from individual
490 *Rhipicephalus sanguineus* and *R. turanicus* ticks, which were harvested in Palestine.
491 Unsurprisingly, *Candidatus* C. mudrowiae genomes from Palestine and Israel are
492 very similar. Interestingly, our two genome assemblies are each clearly more closely
493 related to one of the variants in each of the tick pools from Israel, pointing the way
494 towards a more sophisticated genome-informed documentation of phylogeographic
495 variation in this species, which will be relevant to the analysis of host and
496 endosymbiont co-evolution.
497
498 We detected two microorganisms by metagenomics that were somewhat
499 unexpected: *Anaplasma ovis* and canine parvovirus. *Anaplasma ovis* infects sheep
500 and goats in many regions of the world, but its precise prevalence and clinical
501 importance remains uncertain in most settings, including here in Palestine [70].
502 Enteritis caused by canine parvovirus has been a leading cause of morbidity and
503 mortality in dogs globally since it emerged as a new pathogen in the mid-1970s [71].
504 The usual route of transmission is thought to be faecal-oral. However, the infection is
505 characterised by a marked viraemia and canine parvovirus has been shown to
506 persist within hard ticks under laboratory conditions [72]. Our metagenomics survey
507 has provided the first evidence that ticks can carry the virus in the natural setting and
508 raises the possibility that ticks might contribute to the spread of canine parvovirus
509 among dogs in Palestine and more widely. However, it remains possible that the tick
510 is merely a dead-end host for this virus.

511

512 In closing, the results presented here hint at exciting opportunities for more ambitious
513 metagenomic surveys in the future, as sequencing technologies become more user-
514 friendly and cost-effective. Such ventures are likely to shed new light on the
515 transmission, evolution, and phylogeography of tick-borne pathogens and
516 endosymbionts.

517

518 Supporting information

519 S1 Table: 16S OTU table for tick pools

520 S2 Table: Tick and host mitochondrial sequences

521 S3 Table: Reference genomes for tick-borne pathogens and endosymbionts

522 S4 Table: Results of targeted PCRs on individual ticks

523 S5 Table: Shotgun Sequencing Statistics

524 S6 Table: Mapping to tick and host mitochondrial sequences

525 S7 Table: Taxonomic profiles of metagenomic sequences

526 S8 Table Mapping data for tick-borne pathogens and endosymbionts

527 S9 Table: SNP Tables for tick-borne pathogens and endosymbionts

528 S1 Figure. Multiple alignment of signature sequences confirming identity of
529 *Candidatus* R. barbariae Rb Nablus

530

Acknowledgments

We thank Yuval Gottlieb for reading and commenting on the manuscript.

Author Contributions

Conceived and designed the experiments: SE, AN, and MP. Identified the ticks and extracted DNA: AA-J and OA-S. Performed 16S PCR amplification: SE and AN. Performed 16S amplicon sequencing: OA-S and HH. Performed metagenomic sequencing: SE and AN. Bioinformatics analysis: AR and MP. Wrote the manuscript SE, MP, AR and AN. All authors approved the manuscript.

References

1. Pfäffle M, Littwin N, Muders SV, Petney TN (2013) The ecology of tick-borne diseases. Int J Parasitol 43: 1059-1077.
2. de la Fuente J, Antunes S, Bonnet S, Cabezas-Cruz A, Domingos AG, Estrada-Peña A et al. (2017) Tick-Pathogen Interactions and Vector Competence: Identification of Molecular Drivers for Tick-Borne Diseases. Front Cell Infect Microbiol 7: 114.
3. Jongejan F, Uilenberg G (2004) The global importance of ticks. Parasitology 129 Suppl: S3-14.
4. Ghosh S, Azhahianambi P, Yadav MP (2007) Upcoming and future strategies of tick control: a review. J Vector Borne Dis 44: 79-89.
5. Zhong J, Jasinskas A, Barbour AG (2007) Antibiotic treatment of the tick vector Amblyomma americanum reduced reproductive fitness. PLoS One 2: e405.
6. Guizzo MG, Parizi LF, Nunes RD, Schama R, Albano RM, Tirloni L et al. (2017) A Coxiella mutualist symbiont is essential to the development of Rhipicephalus

556 microplus. Sci Rep 7: 17554.
557 7. Li LH, Zhang Y, Zhu D (2018) Effects of antibiotic treatment on the fecundity of
558 Rhipicephalus haemaphysaloides ticks. Parasit Vectors 11: 242.
559 8. Niebylski ML, Peacock MG, Fischer ER, Porcella SF, Schwan TG (1997)
560 Characterization of an endosymbiont infecting wood ticks, Dermacentor
561 andersoni, as a member of the genus Francisella. Appl Environ Microbiol 63:
562 3933-3940.
563 9. Kreizinger Z, Hornok S, Dán A, Hresko S, Makrai L, Magyar T et al. (2013)
564 Prevalence of Francisella tularensis and Francisella-like endosymbionts in the
565 tick population of Hungary and the genetic variability of Francisella-like agents.
566 Vector Borne Zoonotic Dis 13: 160-163.
567 10. Szigeti A, Kreizinger Z, Hornok S, Abichu G, Gyuranecz M (2014) Detection of
568 Francisella-like endosymbiont in Hyalomma rufipes from Ethiopia. Ticks Tick
569 Borne Dis 5: 818-820.
570 11. Gerhart JG, Moses AS, Raghavan R (2016) A Francisella-like endosymbiont in
571 the Gulf Coast tick evolved from a mammalian pathogen. Sci Rep 6: 33670.
572 12. Azagi T, Klement E, Perlman G, Lustig Y, Mumcuoglu KY, Apanaskevich DA et
573 al. (2017) Francisella-Like Endosymbionts and Rickettsia Species in Local and
574 Imported Hyalomma Ticks. Appl Environ Microbiol 83:
575 13. Gerhart JG, Auguste Dutcher H, Brenner AE, Moses AS, Grubhoffer L,
576 Raghavan R (2018) Multiple Acquisitions of Pathogen-Derived Francisella
577 Endosymbionts in Soft Ticks. Genome Biol Evol 10: 607-615.
578 14. Smith TA, Driscoll T, Gillespie JJ, Raghavan R (2015) A Coxiella-like
579 endosymbiont is a potential vitamin source for the Lone Star tick. Genome Biol
580 Evol 7: 831-838.
581 15. Raele DA, Galante D, Pugliese N, De Simone E, Cafiero MA (2015) Coxiella-
582 like endosymbiont associated to the “Anatolian brown tick” Rhipicephalus bursa
583 in Southern Italy. Microbes Infect 17: 799-805.

584 16. Arthan W, Sumrandee C, Hirunkanokpun S, Kitthawee S, Baimai V,
585 Trinachartvanit W et al. (2015) Detection of Coxiella-like endosymbiont in
586 Haemaphysalis tick in Thailand. Ticks Tick Borne Dis 6: 63-68.
587 17. Al-Deeb MA, Frangoulidis D, Walter MC, Kömpf D, Fischer SF, Petney T et al.
588 (2016) Coxiella-like endosymbiont in argasid ticks (Ornithodoros muesebecki)
589 from a Socotra Cormorant colony in Umm Al Quwain, United Arab Emirates.
590 Ticks Tick Borne Dis 7: 166-171.
591 18. Michelet L, Joncour G, Devillers E, Torina A, Vayssier-Taussat M, Bonnet SI et
592 al. (2016) Tick species, tick-borne pathogens and symbionts in an insular
593 environment off the coast of Western France. Ticks Tick Borne Dis 7: 1109-
594 1115.
595 19. Papa A, Tsioka K, Kontana A, Papadopoulos C, Giadinis N (2017) Bacterial
596 pathogens and endosymbionts in ticks. Ticks Tick Borne Dis 8: 31-35.
597 20. Duron O, Binetruy F, Noël V, Cremaschi J, McCoy KD, Arnathau C et al. (2017)
598 Evolutionary changes in symbiont community structure in ticks. Mol Ecol 26:
599 2905-2921.
600 21. Duron O, Noël V, McCoy KD, Bonazzi M, Sidi-Boumedine K, Morel O et al.
601 (2015) The Recent Evolution of a Maternally-Inherited Endosymbiont of Ticks
602 Led to the Emergence of the Q Fever Pathogen, Coxiella burnetii. PLoS Pathog
603 11: e1004892.
604 22. Gall CA, Reif KE, Scoles GA, Mason KL, Mousel M, Noh SM et al. (2016) The
605 bacterial microbiome of Dermacentor andersoni ticks influences pathogen
606 susceptibility. ISME J 10: 1846-1855.
607 23. Felsheim RF, Kurtti TJ, Munderloh UG (2009) Genome sequence of the
608 endosymbiont Rickettsia peacockii and comparison with virulent Rickettsia
609 rickettsii: identification of virulence factors. PLoS One 4: e8361.
610 24. Ereqat S, Nasereddin A, Al-Jawabreh A, Azmi K, Harrus S, Mumcuoglu K et al.
611 (2016) Molecular Detection and Identification of Spotted Fever Group

612 Rickettsiae in Ticks Collected from the West Bank, Palestinian Territories.
613 PLoS Negl Trop Dis 10: e0004348.
614 25. Ereqat S, Nasereddin A, Vayssier-Taussat M, Abdelkader A, Al-Jawabreh A,
615 Zaid T et al. (2016) Molecular Evidence of Bartonella Species in Ixodid Ticks
616 and Domestic Animals in Palestine. Front Microbiol 7: 1217.
617 26. Azmi K, Ereqat S, Nasereddin A, Al-Jawabreh A, Baneth G, Abdeen Z (2016)
618 Molecular detection of Theileria, Babesia, and Hepatozoon spp. in ixodid ticks
619 from Palestine. Ticks Tick Borne Dis 7: 734-741.
620 27. Kleinerman G, Baneth G, Mumcuoglu KY, van Straten M, Berlin D,
621 Apanaskevich DA et al. (2013) Molecular detection of Rickettsia africae,
622 Rickettsia aeschlimannii, and Rickettsia sibirica mongolitimonae in camels and
623 Hyalomma spp. ticks from Israel. Vector Borne Zoonotic Dis 13: 851-856.
624 28. Keysary A, Eremeeva ME, Leitner M, Din AB, Wikswo ME, Mumcuoglu KY et al.
625 (2011) Spotted fever group rickettsiae in ticks collected from wild animals in
626 Israel. Am J Trop Med Hyg 85: 919-923.
627 29. Waner T, Keysary A, Eremeeva ME, Din AB, Mumcuoglu KY, King R et al.
628 (2014) Rickettsia africae and Candidatus Rickettsia barbariae in ticks in Israel.
629 Am J Trop Med Hyg 90: 920-922.
630 30. Harrus S, Perlman-Avrahami A, Mumcuoglu KY, Morick D, Eyal O, Baneth G
631 (2011) Molecular detection of Ehrlichia canis, Anaplasma bovis, Anaplasma
632 platys, Candidatus Midichloria mitochondrii and Babesia canis vogeli in ticks
633 from Israel. Clin Microbiol Infect 17: 459-463.
634 31. Harrus S, Perlman-Avrahami A, Mumcuoglu KY, Morick D, Baneth G (2011)
635 Molecular detection of Rickettsia massiliae, Rickettsia sibirica mongolitimonae
636 and Rickettsia conorii israelensis in ticks from Israel. Clin Microbiol Infect 17:
637 176-180.
638 32. Rose J, Nachum-Biala Y, Mumcuoglu KY, Alkhamis MA, Ben-Nun A, Lensky I
639 et al. (2017) Genetic characterization of spotted fever group rickettsiae in

640 questing ixodid ticks collected in Israel and environmental risk factors for their
641 infection. Parasitology 144: 1088-1101.
642 33. Gottlieb Y, Lalzar I, Klasson L (2015) Distinctive Genome Reduction Rates
643 Revealed by Genomic Analyses of Two Coxiella-Like Endosymbionts in Ticks.
644 Genome Biol Evol 7: 1779-1796.
645 34. Tsementzi D, Castro Gordillo J, Mahagna M, Gottlieb Y, Konstantinidis KT
646 (2018) Comparison of closely related, uncultivated Coxiella tick endosymbiont
647 population genomes reveals clues about the mechanisms of symbiosis. Environ
648 Microbiol
649 35. Pallen MJ (2014) Diagnostic metagenomics: potential applications to bacterial,
650 viral and parasitic infections. Parasitology 141: 1856-1862.
651 36. Bouquet J, Melgar M, Swei A, Delwart E, Lane RS, Chiu CY (2017)
652 Metagenomic-based Surveillance of Pacific Coast tick Dermacentor
653 occidentalis Identifies Two Novel Bunyaviruses and an Emerging Human
654 Ricksettsial Pathogen. Sci Rep 7: 12234.
655 37. Anonymous (2003) 16S Metagenomic Sequencing Library Preparation.
656 Available: https://www.illumina.com/content/dam/illumina-
657 support/documents/documentation/chemistry_documentation/16s/16s-
658 metagenomic-library-prep-guide-15044223-b.pdf via the Internet. Accessed x.
659 38. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello
660 EK et al. (2010) QIIME allows analysis of high-throughput community
661 sequencing data. Nat Methods 7: 335-336.
662 39. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K et al.
663 (2006) Greengenes, a chimera-checked 16S rRNA gene database and
664 workbench compatible with ARB. Appl Environ Microbiol 72: 5069-5072.
665 40. Connor TR, Loman NJ, Thompson S, Smith A, Southgate J, Poplawski R et al.
666 (2016) CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online
667 resource for the medical microbiology community. Microb Genom 2: e000086.

668 41. Andrews S FastQC: a quality control tool for high throughput sequence data.
669 Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc via the
670 Internet. Accessed x.
671 42. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for
672 Illumina sequence data. Bioinformatics 30: 2114-2120.
673 43. Kim D, Song L, Breitwieser FP, Salzberg SL (2016) Centrifuge: rapid and
674 sensitive classification of metagenomic sequences. Genome Res 26: 1721-
675 1729.
676 44. Breitwieser FP, Salzberg SL (2016) Pavian: Interactive analysis of
677 metagenomics data for microbiomics and pathogen identification. BioRxiv
678 https://doiorg/101101/084715
679 45. Blin K (2018) ncbi-genome-download. Available: https://github.com/kblin/ncbi-
680 genome-download via the Internet. Accessed x.
681 46. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2.
682 Nat Methods 9: 357-359.
683 47. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N et al. (2009) The
684 Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079.
685 48. Okonechnikov K, Conesa A, García-Alcalde F (2016) Qualimap 2: advanced
686 multi-sample quality control for high-throughput sequencing data.
687 Bioinformatics 32: 292-294.
688 49. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS et al.
689 (2012) SPAdes: a new genome assembly algorithm and its applications to
690 single-cell sequencing. J Comput Biol 19: 455-477.
691 50. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT et al. (2015)
692 Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31:
693 3691-3693.
694 51. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local
695 alignment search tool. J Mol Biol 215: 403-410.

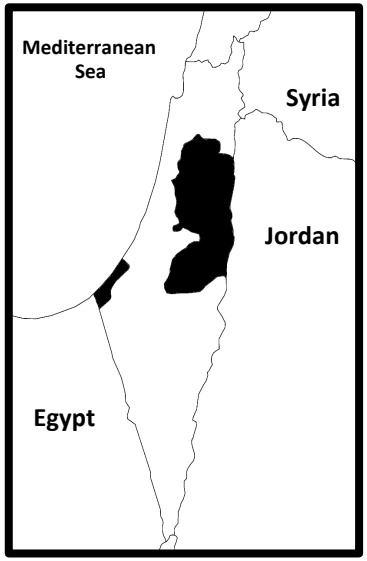
696 52. Rodriguez-R LM, Konstantinidis KT (2016) The enveomics collection: a toolbox
697 for specialized analyses of microbial genomes and metagenomes. PeerJ
698 Preprints 4: e1900v1.
699 53. Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid
700 multiple sequence alignment based on fast Fourier transform. Nucleic Acids
701 Res 30: 3059-3066.
702 54. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and
703 post-analysis of large phylogenies. Bioinformatics 30: 1312-1313.
704 55. Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary
705 Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33: 1870-1874.
706 56. Price MN, Dehal PS, Arkin AP (2010) FastTree 2--approximately maximum-
707 likelihood trees for large alignments. PLoS One 5: e9490.
708 57. Seemann T (2015) Snippy: rapid haploid variant calling and core SNP
709 phylogeny. Available: https://github.com/tseemann/snippy via the Internet.
710 Accessed x.
711 58. Garrison E, Marth G (2012) Haplotype-based variant detection from short-read
712 sequencing. arXiv preprint arXiv:12073907 [q-bioGN]
713 59. Blanc G, Ogata H, Robert C, Audic S, Claverie JM, Raoult D (2007) Lateral
714 gene transfer between obligate intracellular bacteria: evidence from the
715 Rickettsia massiliae genome. Genome Res 17: 1657-1664.
716 60. Eremeeva ME, Bosserman EA, Demma LJ, Zambrano ML, Blau DM, Dasch GA
717 (2006) Isolation and identification of Rickettsia massiliae from Rhipicephalus
718 sanguineus ticks collected in Arizona. Appl Environ Microbiol 72: 5569-5577.
719 61. (2018) Ticks of Europe and North Africa: A Guide to Species Identification.
720 Springer.
721 62. Gray J, Dantas-Torres F, Estrada-Peña A, Levin M (2013) Systematics and
722 ecology of the brown dog tick, Rhipicephalus sanguineus. Ticks Tick Borne Dis
723 4: 171-180.

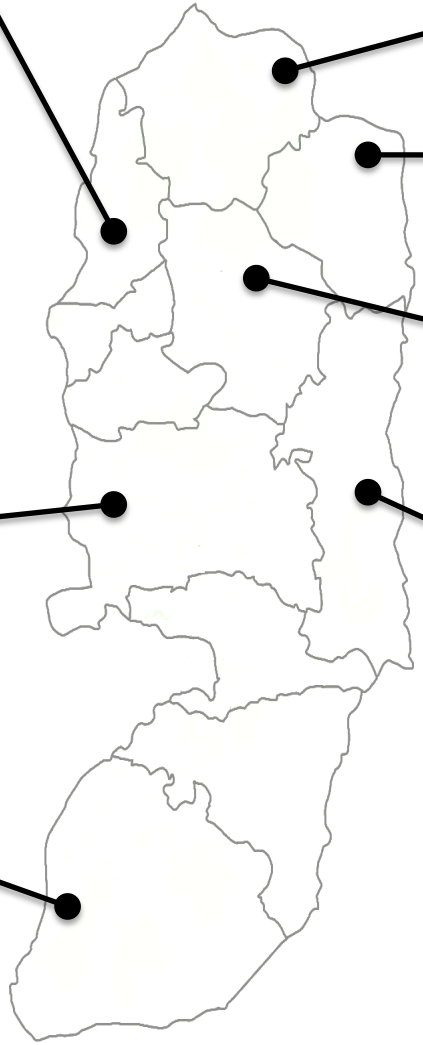
724 63. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF et al. (2014)
725 Reagent and laboratory contamination can critically impact sequence-based
726 microbiome analyses. BMC Biol 12: 87.
727 64. Dobler G, Wölfel R (2009) Typhus and other rickettsioses: emerging infections
728 in Germany. Dtsch Arztebl Int 106: 348-354.
729 65. García-García JC, Portillo A, Núñez MJ, Santibáñez S, Castro B, Oteo JA
730 (2010) A patient from Argentina infected with Rickettsia massiliae. Am J Trop
731 Med Hyg 82: 691-692.
732 66. Parola P, Socolovschi C, Jeanjean L, Bitam I, Fournier PE, Sotto A et al. (2008)
733 Warmer weather linked to tick attack and emergence of severe rickettsioses.
734 PLoS Negl Trop Dis 2: e338.
735 67. Mura A, Masala G, Tola S, Satta G, Fois F, Piras P et al. (2008) First direct
736 detection of rickettsial pathogens and a new rickettsia, 'Candidatus Rickettsia
737 barbariae', in ticks from Sardinia, Italy. Clin Microbiol Infect 14: 1028-1033.
738 68. Papa A, Xanthopoulou K, Kotriotsiou T, Papaioakim M, Sotiraki S, Chaligiannis
739 I et al. (2016) Rickettsia species in human-parasitizing ticks in Greece. Trans R
740 Soc Trop Med Hyg 110: 299-304.
741 69. Fernández de Mera IG, Blanda V, Torina A, Dabaja MF, El Romeh A, Cabezas-
742 Cruz A et al. (2018) Identification and molecular characterization of spotted
743 fever group rickettsiae in ticks collected from farm ruminants in Lebanon. Ticks
744 Tick Borne Dis 9: 104-108.
745 70. Renneker S, Abdo J, Salih DE, Karagenç T, Bilgiç H, Torina A et al. (2013) Can
746 Anaplasma ovis in small ruminants be neglected any longer. Transbound
747 Emerg Dis 60 Suppl 2: 105-112.
748 71. Mylonakis ME, Kalli I, Rallis TS (2016) Canine parvoviral enteritis: an update on
749 the clinical diagnosis, treatment, and prevention. Vet Med (Auckl) 7: 91-100.
750 72. Mori H, Tanaka T, Mochizuki M (2015) The widely distributed hard tick,
751 Haemaphysalis longicornis, can retain canine parvovirus, but not be infected in

752 laboratory condition. J Vet Med Sci 77: 405-411.

753

754





Tulkarem dog tick pool 1
Rhipicephalus sanguineus

Jenin dog tick pools 1-3
Rhipicephalus sanguineus

Tubas dog tick pools 1-5
Rhipicephalus sanguineus
Tubas sheep tick pools 1-3
Rhipicephalus turanicus

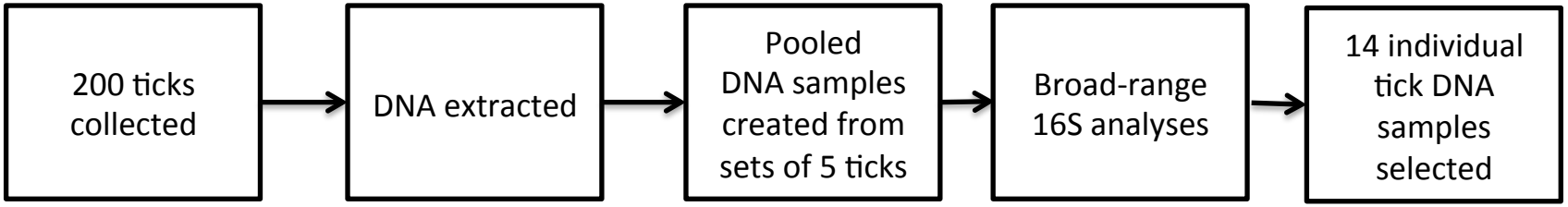
Nablus dog tick pools 1-4
Rhipicephalus sanguineus
Nablus sheep tick pools 1-4
Rhipicephalus turanicus#

Ramallah dog tick pools 1-4
Rhipicephalus sanguineus
Ramallah dog tick pool 5
Haemaphysalis parva∞

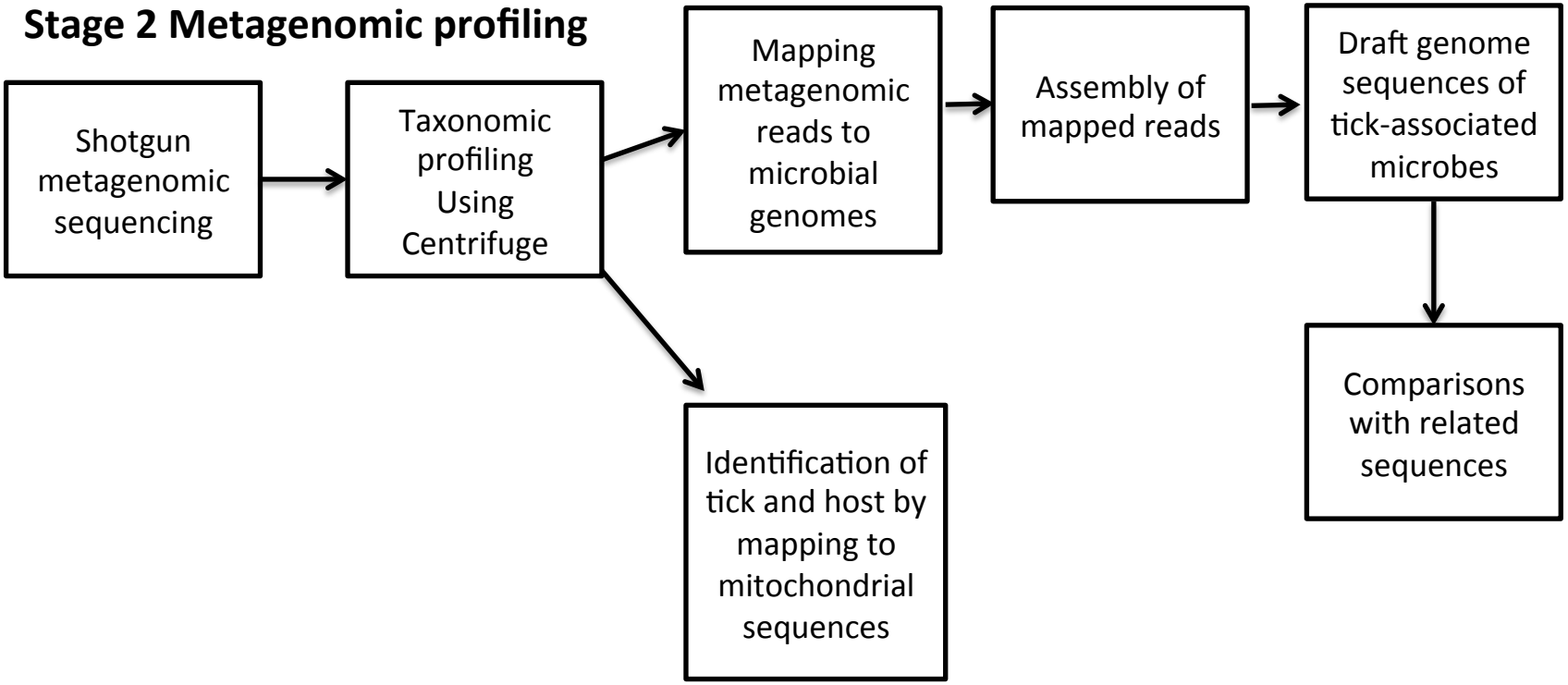
Jericho dog tick pools 1, 2
Rhipicephalus sanguineus
Jericho camel tick pool 1
Hyalomma dromedarii
Jericho camel tick pool 2
*Hyalomma impeltatum**

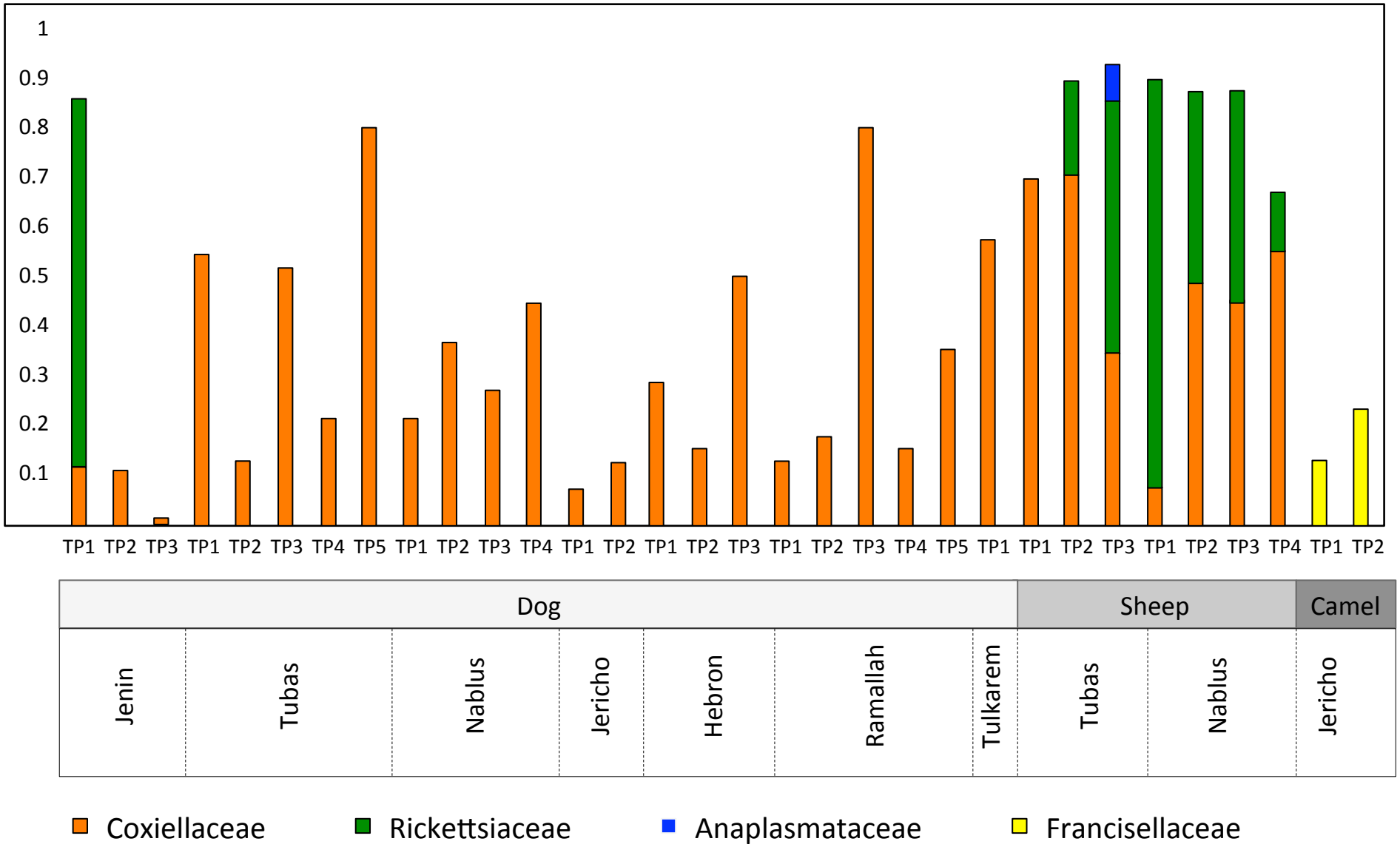
Hebron dog tick pools 1-3
Rhipicephalus sanguineus

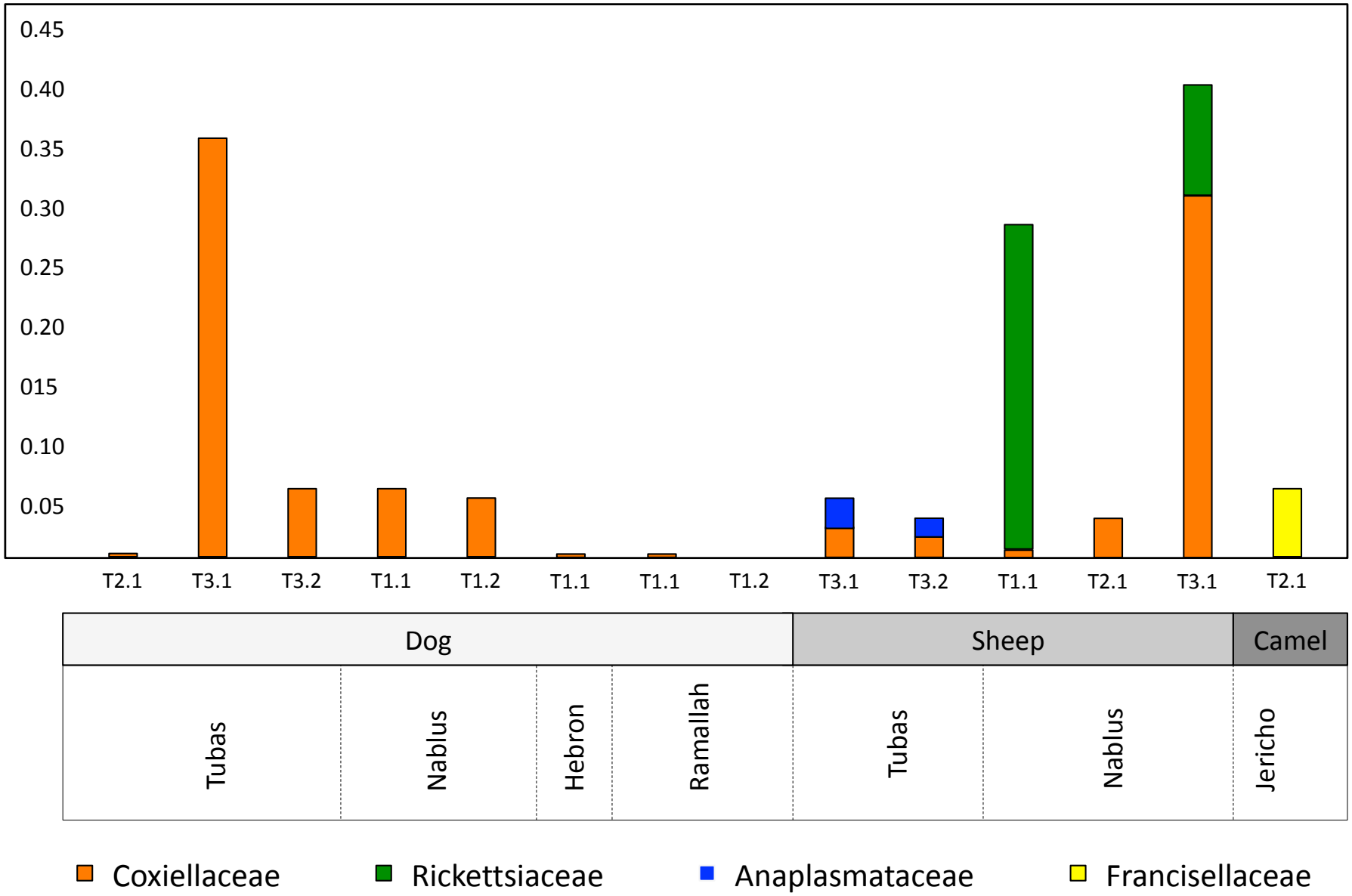
Stage 1 16S rRNA amplicon profiling

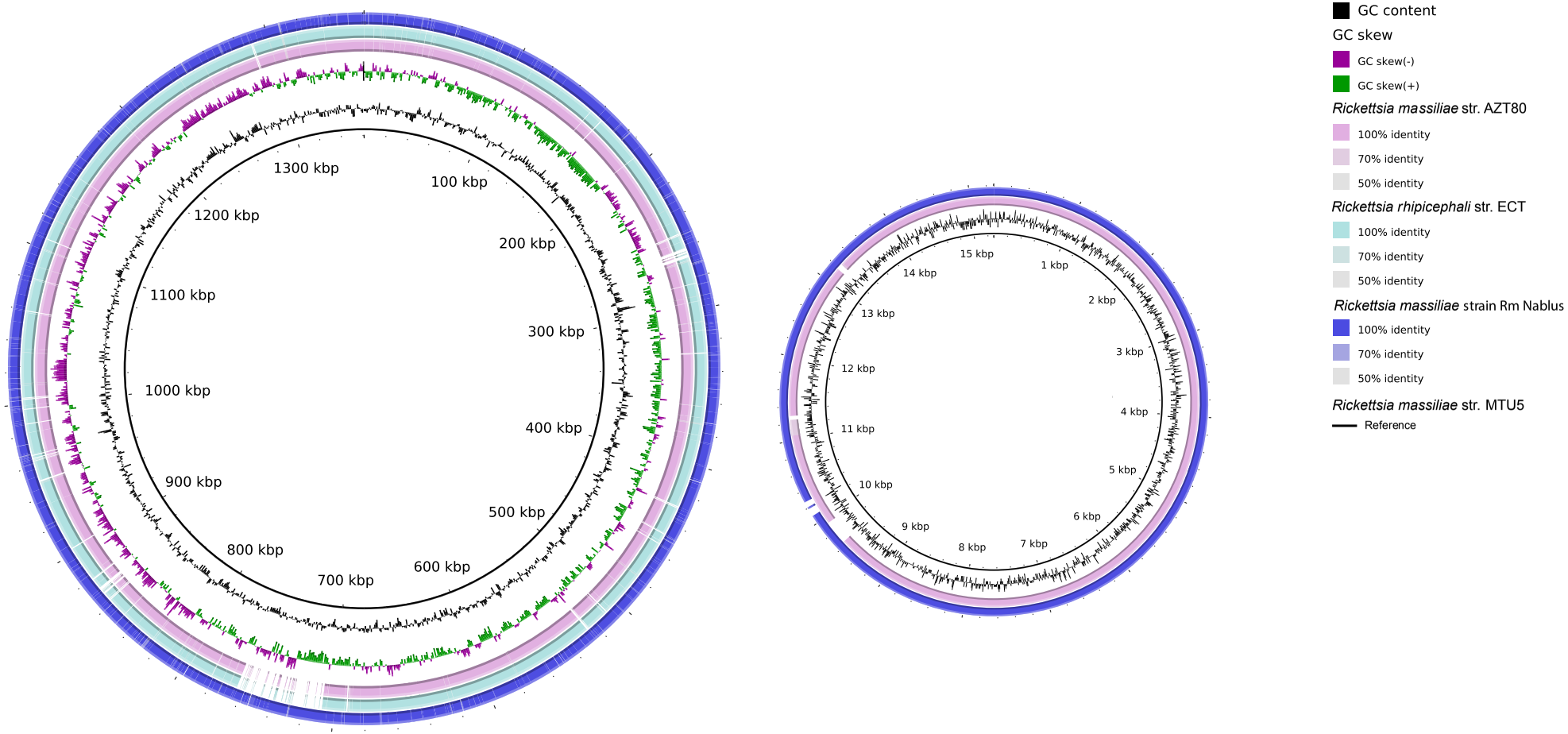


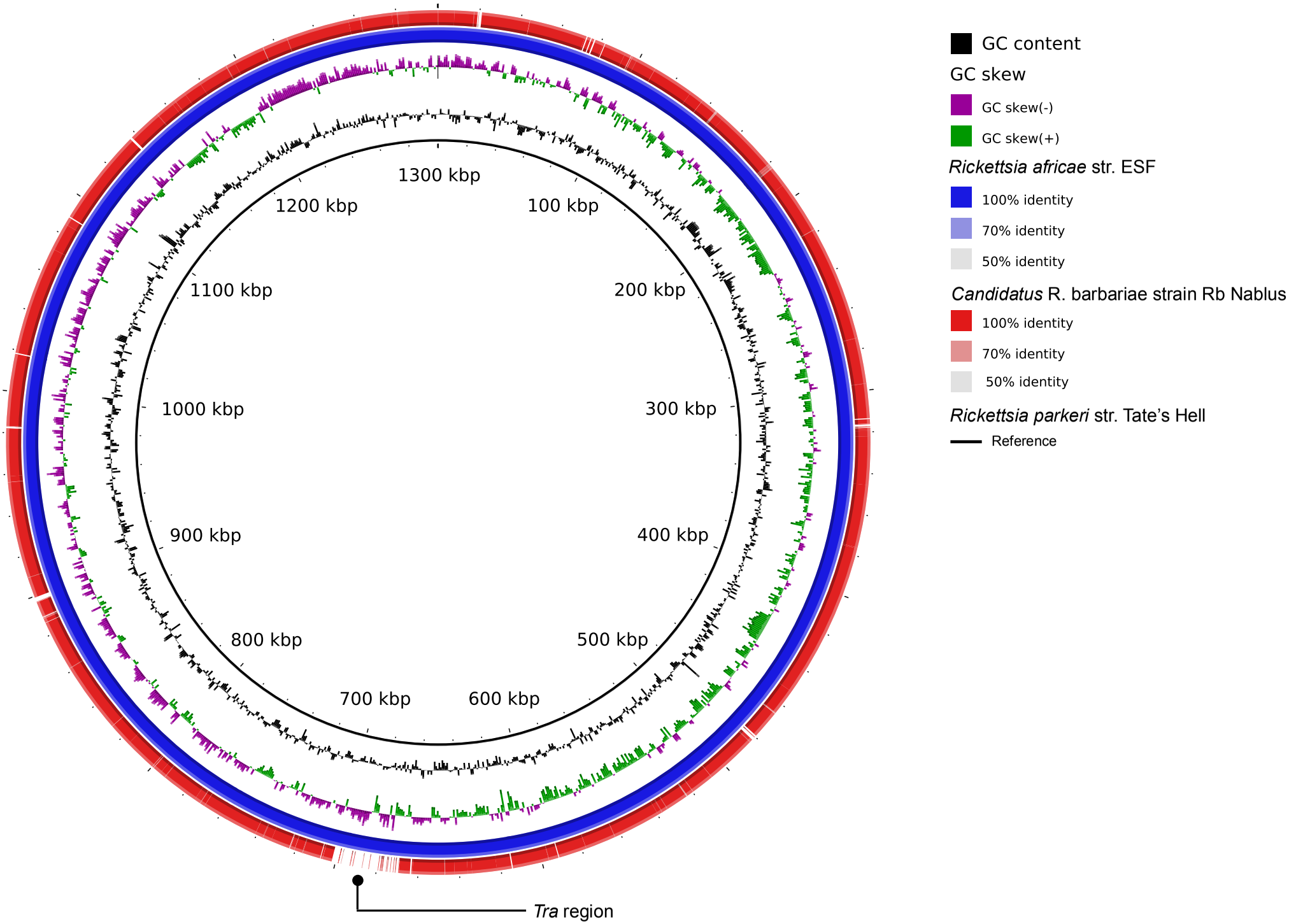
Stage 2 Metagenomic profiling

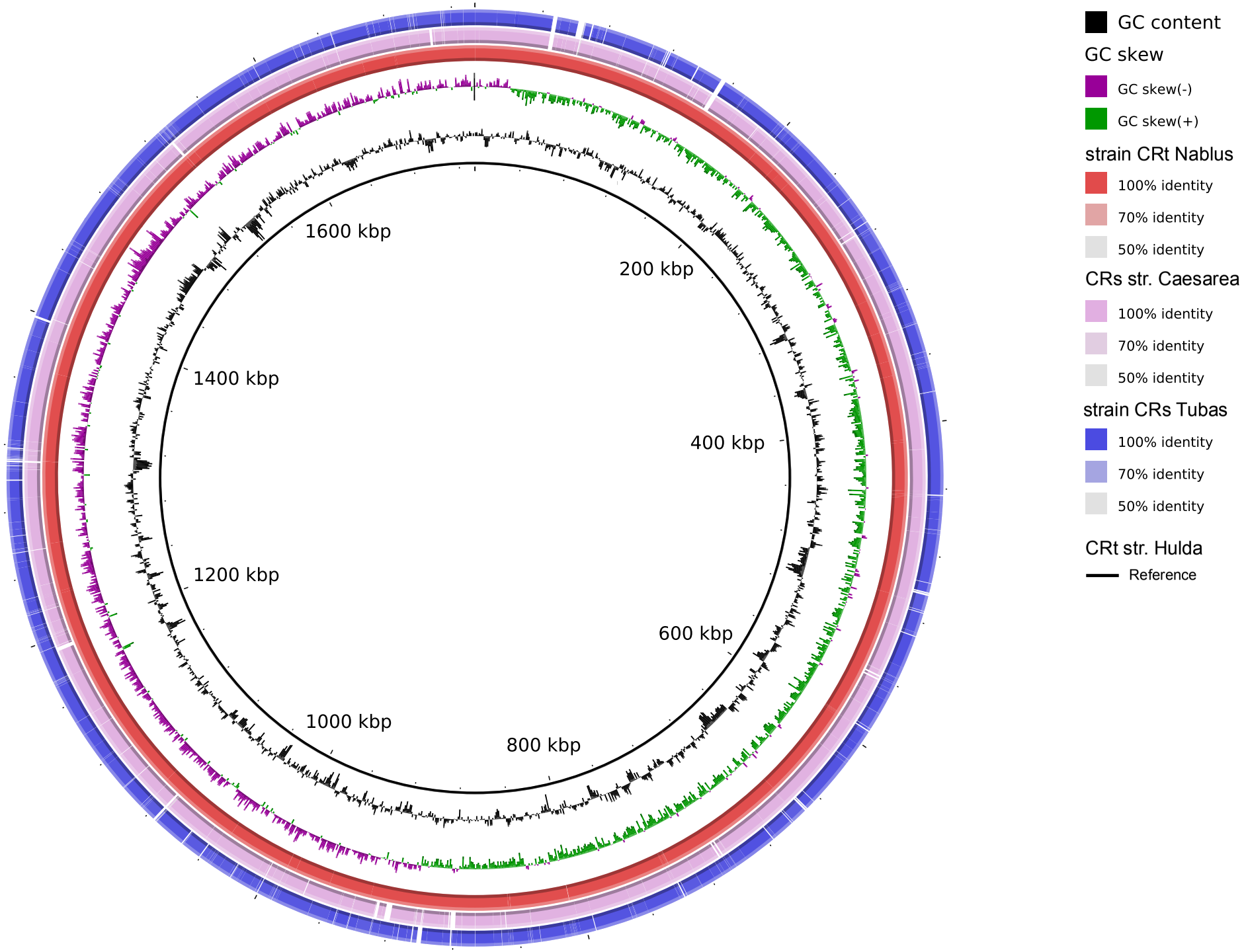



1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
TP1 TP2 TP3 TP1 TP2 TP3 TP4 TP5 TP1 TP2 TP3 TP4 TP1 TP2 TP1 TP2 TP3 TP1 TP2 TP3 TP4 TP5 TP1 TP1 TP2 TP3 TP1 TP2 TP3 TP4 TP1 TP2
Dog
Sheep
Camel
Jenin
Tubas
Nablus
Jericho
Hebron
Ramallah
Tulkarem
Tubas
Nablus
Jericho
Coxiellaceae
Rickettsiaceae
Anaplasmataceae
Francisellaceae


0.45
0.40
0.35
0.30
0.25
0.20
015
0.10
0.05
T2.1
T3.1
T3.2
T1.1
T1.2
T1.1
T1.1
T1.2
T3.1
T3.2
T1.1
T2.1
T3.1
T2.1
Dog
Sheep
Camel
Tubas
Nablus
Hebron
Ramallah
Tubas
Nablus
Jericho
Coxiellaceae
Rickettsiaceae
Anaplasmataceae
Francisellaceae


1300 kbp
100 kbp
1200 kbp
200 kbp
1100 kbp
300 kbp
1000 kbp
400 kbp
900 kbp
500 kbp
800 kbp
600 kbp
700 kbp
15 kbp
1 kbp
14 kbp
2 kbp
13 kbp
3 kbp
12 kbp
4 kbp
11 kbp
5 kbp
10 kbp
6 kbp
9 kbp
7 kbp
8 kbp
GC content
GC skew
GC skew(-)
GC skew(+)
Rickettsia massiliae str. AZT80
100% identity
70% identity
50% identity
Rickettsia rhipicephali str. ECT
100% identity
70% identity
50% identity
Rickettsia massiliae strain Rm Nablus
100% identity
70% identity
50% identity
Rickettsia massiliae str. MTU5
Reference


1300 kbp
100 kbp
200 kbp
300 kbp
400 kbp
500 kbp
600 kbp
700 kbp
800 kbp
900 kbp
1000 kbp
1100 kbp
1200 kbp
GC content
GC skew
GC skew(-)
GC skew(+)
Rickettsia africae str. ESF
100% identity
70% identity
50% identity
Candidatus R. barbariae strain Rb Nablus
100% identity
70% identity
50% identity
Rickettsia parkeri str. Tate's Hell
Reference
Tra region


1600 kbp
200 kbp
1400 kbp
400 kbp
1200 kbp
600 kbp
1000 kbp
800 kbp
GC content
GC skew
GC skew(-)
GC skew(+)
strain CRt Nablus
100% identity
70% identity
50% identity
CRs str. Caesarea
100% identity
70% identity
50% identity
strain CRs Tubas
100% identity
70% identity
50% identity
CRt str. Hulda
Reference