Title: Sequence of a *Coxiella* endosymbiont of the tick *Amblyomma nuttalli* suggests a pattern of convergent genome reduction in the *Coxiella* genus

Tiago Nardi[1], Emanuela Olivieri[1], Edward Kariuki[2], Davide Sassera[1], Michele Castelli[1]*

1. Department of Biology and Biotechnology, University of Pavia, Pavia, Italy

2. Veterinary and Capture Service Department, Kenya Wildlife Service, Kenya

***Author for Correspondence:** Michele Castelli, Department of Biology and Biotechnology, University of Pavia, Pavia, Italy, miccast@tiscali.it

Abstract

Ticks require bacterial symbionts for the provision of necessary compounds that are absent in their hematophagous diet. Such symbionts are frequently vertically transmitted and, most commonly, belong to the *Coxiella* genus, which also includes the human pathogen *Coxiella burnetii*. This genus can be divided in four main clades, presenting partial but incomplete co-cladogenesis with the tick hosts. Here we report the genome sequence of a novel *Coxiella*, endosymbiont of the African tick *Amblyomma nuttalli,* and the ensuing comparative analyses. Its size (~1 Mb) is intermediate between symbionts of *Rhipicephalus* species and other *Amblyomma* species. Phylogenetic analyses show that the novel sequence is the first genome of the B clade, the only one for which no genomes were previously available. Accordingly, it allows to draw an enhanced scenario of the evolution of the genus, one of parallel genome reduction of different endosymbiont lineages, which are now at different stages of reduction from a more versatile ancestor. Gene content comparison allows to infer that the ancestor could be reminiscent of *Coxiella burnetii*. Interestingly, the convergent loss of mismatch repair could have been a major driver of such reductive evolution. Predicted metabolic profiles are rather homogenous among *Coxiella* endosymbionts, in particular vitamin biosynthesis, consistently with a host-supportive role. Concurrently, similarities among *Coxiella* endosymbionts according to host genus and despite phylogenetic unrelatedness hint at possible host-dependent effects.

Keywords: *Amblyomma nuttalli*, *Coxiella*-like endosymbiont, genome reduction, phylogeny, comparative genomics, symbiont evolution

Significance statement

The genus *Coxiella* includes the pathogen *Coxiella burnetii* and widespread nutritional mutualists in ticks. Current knowledge on their evolution is hampered by the limited genomic resources available.

Here we provide the first genome sequence of a *Coxiella* endosymbiont of clade B, the only clade for which none was available.

These data allow to infer an evolutionary scenario of parallel genome reduction among *Coxiella* endosymbionts, with similar constraints, leading to selective retention of biosynthetic pathways beneficial for the host. The combined predicted functional capabilities of the symbionts appear to be a subset of those of *C. burnetii*. Accordingly, this pathogen could be closer to an ancestral state of the endosymbionts, rather than being derived from an endosymbiotic ancestor, as previously hypothesized.

Introduction

Mutualistic associations with bacteria are widespread and can allow eukaryotes to colonise novel ecological niches (Bennett & Moran 2015). In arthropods, intracellular and maternally transmitted bacterial mutualists are particularly common (Douglas 1998). A typical role of these endosymbionts is providing essential nutrients that are absent in the host's diet (Sandström & Moran 1999). The classical example is that of sap-feeding insects, such as aphids, acquiring essential amino acids and vitamins from intracellular bacteria (Wernegreen 2012).

Similarly, blood-feeding arthropods also rely on an incomplete source of nutrients and harbour bacterial mutualists. For instance, "*Candidatus* Riesia" bacteria supply their host, the human lice, with several compounds, in particular B vitamins (Boyd et al. 2017; Sasaki-Fukatsu et al. 2006). A better understanding of these symbioses and their role can foster the development of control strategies for hematophagous arthropod vectors of diseases (Zindel et al. 2011). Ticks in particular are extensively studied due to their prominent role of disease vectors for humans and domestic animals (Dantas-Torres et al. 2012). The most important pathogens vectored by ticks include *Borrelia*, the tick-borne encephalitis virus, *Coxiella burnetii*, and multiple *Rickettsiales* bacteria (Kernif et al. 2016).

Most tick species present at least one bacterial endosymbiont, and many species more than one (Cafiso et al. 2016; Moutailler et al. 2016; Duron et al. 2017). Transmission of symbionts is mainly dependent on maternal inheritance through transovarial transmission, but horizontal transfer, possibly during co-feeding, also plays a role. Indeed, closely related tick species, and even different individuals of the same species, may harbour different sets of bacterial symbionts (Duron et al. 2017).

Experimental evidence indicates a major role of such symbionts in tick physiology, as their depletion resulted in impaired growth and reproduction in multiple species belonging to

different genera (Zhong et al. 2007; Guizzo et al. 2017; Ben-Yosef et al. 2020). Interestingly, in *Ornithodoros moubata* such developmental defects were rescued by supplementation with B vitamins (Duron et al. 2018). Collectively, these data suggest a common role as B vitamins providers of different, phylogenetically unrelated, symbionts. Other parallel roles have been hypothesized, including protection from pathogens, provision of energy, support for feeding, protection against oxidative and osmotic stress, and waste molecule recycling (Buysse et al. 2019; Olivieri et al. 2019).

Most of characterized tick symbionts are affiliated to *Coxiella* (*Gammaproteobacteria*) (Duron et al. 2017), and will now be abbreviated as CEs (*Coxiella* endosymbionts). Besides many CEs of ticks, this genus includes the tick-borne pathogen *Coxiella burnetii*, causative agent of the Q-fever (Angelakis & Raoult 2010). It has been hypothesized that the latter arose from a CE ancestor, but the origin of its virulence is unclear (Duron et al. 2015).

Currently, a total of seven genomes of CEs are available (Gottlieb et al. 2015; Smith et al. 2015; Guizzo et al. 2017; Ramaiah & Dasch 2018; Tsementzi et al. 2018). These present the hallmarks of genome reduction in bacterial symbionts (McCutcheon & Moran 2012), but at different stages. CEs of *Rhipicephalus* spp. have comparatively larger genomes (1.2-1.7 Mb) with a higher number of pseudogenes (Gottlieb et al. 2015; Guizzo et al. 2017; Ramaiah & Dasch 2018; Tsementzi et al. 2018), characteristics of relatively recent symbioses with ongoing genome reduction. Conversely, the CEs of *Amblyomma* ticks have more streamlined genomes (0.6-0.7 Mb), with high gene density and low number of mobile elements (Smith et al. 2015), typical of a later stage of symbiosis.

Nevertheless, all CEs possess the pathways for production of many B vitamins, consistently with the hypothesised role and the general trend observed in nutritional symbionts, which retain some host-supportive pathways even in case of severe genome reduction (Nakabachi et al. 2006; López-Madrigal et al. 2011).

Four phylogenetic clades (A-D) were identified in the genus *Coxiella*, exhibiting only partial congruence with their hosts' phylogeny (partial co-cladogenesis). Clade C displays a good

degree of co-cladogenesis with *Rhipicephalus* hosts, while CEs of two unrelated clades (B and D) were found in *Amblyomma* hosts (Duron et al. 2015). Interestingly, the known clade B *Amblyomma* hosts came from the African continent, while clade D hosts are American (Duron et al. 2015; Binetruy et al. 2020). Currently, genomes of representatives of three of the four *Coxiella* clades have been sequenced, i.e. CEs from clades C and D, and *Coxiella burnetii* from clade A.

Here, we sequenced the genome of a novel CE of *Amblyomma nuttalli* belonging to the fourth clade (B), and used it for comparative analyses, providing a basis for improving the understanding of the diversity and evolution of CEs.

Materials and methods

An adult female of *Amblyomma nuttalli* was collected from a white rhinoceros in the Masai Mara National Reserve, Kenya in February 2016. The tick was morphologically identified following standard taxonomic keys (Theiler & Salisbury 1959) and subjected to DNA extraction, using NucleoSpin® Tissue Kit (Macherey Nagel, Duren, Germany), according to the manufacturer's instructions. DNA was subjected to Illumina HiSeq X by Admera Health (South Plainfield, NJ, USA) using a Nextera XT library, obtaining 27,3511,224 150-nt paired-end reads.

The reads were assembled using SPAdes (3.6.0), and subjected to a modified version of the blobology pipeline (Kumar et al. 2013), in order to select only the symbiont sequences (for a complete description of the process see (Castelli et al. 2019)). Briefly, we selected contigs with a $\log_{10}$ coverage higher than 2.5, extracted and reassembled separately the reads mapping on those contigs (Langmead & Salzberg 2012), and extensively revised manually the results (Supplementary Figure 1-2).

The completeness level of the genome was confronted with all published CE genomes using BUSCO, using the *Gammaproteobacteria* lineage dataset (Seppey et al. 2019).

Genome annotation was performed using Prokka 1.11 (Seemann 2014) and manually curated by inspecting blastp hits of predicted ORFs on NCBI nr, Uniprot, and *Legionellales* sequences.

ISEScan (Xie & Tang 2017) and ISfinder (Siguier et al. 2006) were used to identify insertion sequences and PHASTER (Arndt et al. 2016) for prophages. Pseudogene prediction on the novel genome, all published CE genomes, and representative *C. burnetii* genomes (Supplementary table 1) was performed using Pseudo-finder (Syberg-Olsen & Husnik 2018).

COGs were predicted on the same dataset using the NCBI pipeline (Galperin et al. 2015) on validated genes (i.e. ORFs excluding predicted pseudogenes). COG repertoires were used for comparative analyses. Metabolic pathways were manually reconstructed employing the BioCyc database reference (Karp et al. 2019).

Two datasets were used for phylogeny. The first one involved a wide taxonomic sampling, analysed through MLST (multilocus sequence typing) as in (Duron et al. 2015), thus employing five genes and 96 organisms (published dataset plus all available CEs).

The second set was analysed by using a phylogenomic approach, and included the previous selection of *Coxiella* genomes, a representative selection of *Coxiellaceae*, including 1 MAG (metagenome assembled genome), and two other *Legionellales* as outgroup (Supplementary table 1). Using OrthoFinder (2.3.3) (Emms & Kelly 2019), 213 single copy conserved orthologs were identified.

Then, for the two sets, respectively the nucleotide and protein sequences of each single gene were aligned separately using Muscle (Edgar 2004), polished with Gblocks (Talavera & Castresana 2007), and finally concatenated (3,118 and 59,256 total positions, respectively).

For each set, we inferred the best model (GTR+I+G and LG+I+G, respectively) using modeltest-ng 0.1.3 (Darriba et al. 2019), built a maximum likelihood tree with RAxML 8.2.4 (Stamatakis 2014) with 1000 bootstrap pseudo-replicates, and a Bayesian inference tree with MrBayes (Ronquist et al. 2012) using three independent runs for 1 million and 250,000 generations, respectively, with a burn-in of 25%.

Results and discussion

The obtained genome assembly of the CE of *A. nuttalli* has a total length of 1,001,386 bp (9 contigs; N50: 229,733 bp; GC: 35,95%). BUSCO completeness score was 79.2%, similar to other CEs (Supplementary Figure 3). A total of 45 RNA genes (including 38 tRNAs and 3 rRNAs) and 730 ORFs were found. Among these, we identified 696 functional CDSs and 34 pseudogenes, accounting for a total of 658,600 bp (65.7%) coding (including structural RNA genes). Neither prophages nor ISs were found.

MLST phylogeny provided an overall consistent topology with most previous studies (Duron et al. 2015; Gottlieb et al. 2015), in particular for the major *Coxiella* clades and their relationships (clade A earliest divergent, clade B sister group of clades C+D), with moderate to high support (Figure 1A, Supplementary Figure 4-5). Consistently with what could be expected based on geographical origin, the CE of *A. nuttalli* lies in the clade B, composed by CEs of African ticks.

For the available *Coxiella* genomes, phylogenomic analysis showed the same relations of the MLST phylogeny (Figure 1B, Supplementary Figure 6). Interestingly, for all *Coxiella*, including the novel CE of *A. nuttalli*, branch lengths are proportional to the degree of genome reduction (Supplementary Table 2), consistently with previous analyses (Duron et al. 2015; Gottlieb et al. 2015). This would indicate a higher evolutionary rate in smaller genomes, as predicted for obligate symbionts by genome reduction models (McCutcheon & Moran 2012). Specifically, *C. burnetii*, a pathogen capable of living in different environments and hosts (Angelakis & Raoult 2010), presents the largest genome (2 Mb) and, with a high coding density (79.9%), the highest amount of coding DNA (1.8 Mb). All the host-restricted CEs have smaller genomes, with similar sizes within each clade: 1.2-1.7 Mb for CEs of *Rhipicephalus* (clade C), 1.0 Mb for the novel CE of *A. nuttalli* (clade B), and 0.6-0.7 Mb for CEs of other *Amblyomma* species (clade D). However, the degree of genome reduction does not correlate with the phylogenetic branching pattern between clades, in particular the CEs

with more reduced genomes (clades B, D) do not form a single monophyletic group (Figure 1B). Accordingly, considering also the novel CE of *A. nuttalli*, a scenario with parallel independent genome reduction in genus *Coxiella* (at least for monophyletic CEs of B-D clades together) appears plausible.

Interestingly, most size variation resides in the non-coding genome (from 88 Kb to 816 Kb), while the length of the functional coding genome is overall less variable in CEs, ranging from 540 Kb of CE of *A. sculptum* (Clade D), to 917 Kb of CE of *R. turanicus* (Clade C). These features as well are consistent with a recent and still ongoing parallel genome reduction of CEs under similar constraints, possibly due to an equivalent role for the host. Accordingly, the CE of *A. nuttalli* has traits of a relatively long-term obligate symbiont, at an intermediate stage among CEs for its genome size and coding density (Supplementary Table 2), and having no predicted mobile elements.

The functional capabilities, as represented by COG repertoires, are consistent with the observed pattern of genome reduction (Figure 2A). *C. burnetii* is the richest in COGs for all functional categories. CEs have lower COGs numbers, roughly proportional to the respective coding genomes. Many core functions are highly conserved, such as translation machinery (J), coenzyme (H) and nucleotide (F) metabolism, energy production (C), protein modification and chaperones (O), lipid synthesis (I), cell cycle regulation (D). This is consistent with their expected major role for bacterial survival and/or host-support (coenzymes). On the other side, all CEs are more pronouncedly depleted in accessory and regulative functions, including poorly characterized ones (R and S), signal transduction (T), secondary metabolite metabolism (Q), cell motility (N), secretion systems (U), extracellular and defense structures (W and V). Such functionalities are probably less important in strictly host-associated bacteria. Notable is the case of type IV secretion, probably ancestral in *Legionellales* (Hugoson et al. 2019) and an important virulence factor in *C. burnetii* (Luedtke et al. 2017), but absent in all CEs. Some functions display gradients of conservation along the genome size, e.g. membrane structure biogenesis (M), which correlates with decrease in

lipopolysaccharide complexity, while peptidoglycan synthesis is conserved.

Such scenario is reflected at the level of single COGs (Figure 2B), as *C. burnetii* (clade A) presents the highest number of unique COGs. CEs of all clades are a substantial subset of *C. burnetii*, which lacks only 60 of the 1207 total COGs in the dataset. Similar observations can be drawn among progressively more reduced CE clades, with highly streamlined clade D CEs almost as a subset of the other CEs.

Few lineage-specific peculiarities were found, and the functional significance of many of these is unclear, e.g. the *A. nuttalli*-specific COGs display redundant or poorly characterized functions (Supplementary table 3). However, some relevant variability was also observed. All CEs retain the predicted capability to perform glycolysis, Krebs cycle and oxidative phosphorylation (Supplementary table 4). However, contrarily to the other CEs, the symbiont of *A. nuttalli* does not possess a conventional citrate synthase, but instead presents a 2-methylcitrate synthase, which may also catalyse the same reaction (Patton et al. 1993). As *C. burnetii* has both genes, this partial redundancy could have been ancestral, and independently lost in different CE lineages.

In general, while biosynthetic abilities for amino acids are scarce in all *Coxiella* (E), those for vitamins and cofactors (H) are, as expected, abundant and highly conserved (Supplementary table 4). CEs are in particular rich in genes for the synthesis for riboflavin (B2), pantothenate (B5) and its derivative CoA, pyridoxine (B6), folic acid (B9), and biotin. For biotin (and lipid) synthesis, missing FabI functionality is possibly replaced by FabV (Massengo-Tiassé & Cronan 2008). Interestingly, despite overall smaller functional capabilities, the CEs of *Amblyomma* display complete biosynthetic pathways for thiamine (B1) and NAD (B3) (except the last thiamine step in clade D), while symbionts of *Rhipicephalus* retain only partial pathways (including final steps from thiamine phosphate and β-nicotinate D-ribonucleotide). Such differences can be explained by the presence of not yet identified transporters and/or non-canonical enzymes (Gottlieb et al. 2015). They might also indicate

different, not yet clarified, metabolic requirements of the tick hosts, with *Amblyomma* species requiring a full pathway while *Rhipicephalus* being permissive for the loss of some genes.

A similar scenario may hold for nucleotide metabolism (F), which is also more reduced in the CEs of *Rhipicephalus*, lacking the initial path for the synthesis of purines (up to 5-aminoimidazole ribonucleotide) (Supplementary table 4).

Consistently with their symbiotic condition (McCutcheon & Moran 2012), CEs are depleted in DNA repair abilities (L), with lineage-specific features (Supplementary table 4). For example both CEs of *R. microplus* are devoid of the RecFOR pathway and RecA, involved in homologous recombination (Kuzminov 1999). Interestingly, the MutSL pathway is fully absent in the smaller genomes of CEs of *Amblyomma* of clades B and D, but complete in *C. burnetii* (clade A) and in most members of clade C. Among those, the exception is the CE of *R. microplus* "2", found to have a full *mutL* gene, but a truncated *mutS* pseudogene, probably retaining only partial or no functionality. Parsimoniously, we can identify at least three multiple convergent losses of this pathway among CEs (complete in the clades B and D, and still ongoing within clade C). Considering the strong correlation with the degree of genome reduction, it is reasonable to hypothesise that this loss may have had a major evolutionary impact, possibly directly causing increased mutation rates (Schofield & Hsieh 2003), and eventually resulting in accelerated and more pronounced genome reduction. Thus, the seminal speculations by Gottlieb and co-authors (2015) on a smaller dataset are reinforced. This effect would be particularly evident from the differences in coding size and functional categories among the two closely related CEs of *R. microplus* (Figure 1B).

Conclusions

The novel sequence of CE of *Amblyomma nuttalli* expands the available diversity of

Coxiella genomes, being the first obtained from clade B. Despite reduced genome size, biosynthetic pathways for vitamins appear to be conserved, as in other CEs, supporting a role of CEs in dietary supplementation of these compounds to the hosts. At same time, some variations were found in vitamin and purine synthesis, possibly dependent on the host species.

Combining phylogenetic and genomic data, an evolutionary scenario of parallel genome reduction with analogous constraints among CEs (clades B-D) can be drawn. Consistently with previous observations (Gottlieb et al. 2015), the convergent loss of MutSL could have been a driver of such reduction, with the CE of *A. nuttalli* representing an intermediate level between clade C and D, and all CEs substantially being a subset of *C. burnetii*. Accordingly, and differently from previous views (Duron et al. 2015), CEs could have evolved from a more versatile *C. burnetii*-like ancestor, analogously to other unrelated symbionts (Taylor et al. 2005; Gerhart et al. 2016). Genomic and phylogenomic analyses of representatives of clade A other than *C. burnetii* may provide further insights.

Data Availability Statements

The data underlying this article are available in the NCBI GenBank Database at ncbi.nlm.nih.gov/, and can be accessed with JACBPR000000000 (CE genome sequence) and with SRR12168527 (total reads in SRA).

Acknowledgements

This work was supported by the Human Frontier Science Programme Grant RGY0075/2017 to DS, and by the Italian Ministry of Education, University and Research (MIUR): Dipartimenti di Eccellenza Programme (2018–2022)—Department of Biology and Biotechnology “L. Spallanzani”, University of Pavia to DS.

References

Angelakis E, Raoult D. 2010. Q fever. Vet. Microbiol. 140:297–309. doi: 10.1016/j.vetmic.2009.07.016.

Arndt D et al. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 44:W16-21. doi: 10.1093/nar/gkw387.

Bennett GM, Moran NA. 2015. Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. Proc. Natl. Acad. Sci. 112:10169–10176. doi: 10.1073/pnas.1421388112.

Ben-Yosef M et al. 2020. *Coxiella*-Like Endosymbiont of Rhipicephalus sanguineus Is Required for Physiological Processes During Ontogeny. Front. Microbiol. 11. doi: 10.3389/fmicb.2020.00493.

Binetruy F et al. 2020. Microbial community structure reveals instability of nutritional symbiosis during the evolutionary radiation of *Amblyomma* ticks. Mol. Ecol. 29:1016–1029. doi: 10.1111/mec.15373.

Boyd BM et al. 2017. Primates, Lice and Bacteria: Speciation and Genome Evolution in the Symbionts of Hominid Lice. Mol. Biol. Evol. 34:1743–1757. doi: 10.1093/molbev/msx117.

Buysse M, Plantard O, McCoy KD, Duron O, Menard C. 2019. Tissue localization of *Coxiella*-like endosymbionts in three European tick species through fluorescence in situ hybridization. Ticks Tick-Borne Dis. 10:798–804. doi: 10.1016/j.ttbdis.2019.03.014.

Cafiso A et al. 2016. Molecular screening for *Midichloria* in hard and soft ticks reveals variable prevalence levels and bacterial loads in different tick species. Ticks Tick-Borne Dis. 7:1186–1192. doi: 10.1016/j.ttbdis.2016.07.017.

Castelli M et al. 2019. *Deianiraea*, an extracellular bacterium associated with the ciliate *Paramecium*, suggests an alternative scenario for the evolution of *Rickettsiales*. ISME J. 13:2280–2294. doi: 10.1038/s41396-019-0433-9.

Dantas-Torres F, Chomel BB, Otranto D. 2012. Ticks and tick-borne diseases: a One Health

perspective. Trends Parasitol. 28:437–446. doi: 10.1016/j.pt.2012.07.003.

Darriba D et al. 2019. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. Mol. Biol. Evol. doi: 10.1093/molbev/msz189.

Douglas AE. 1998. Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria *Buchnera*. Annu. Rev. Entomol. 43:17–37. doi: 10.1146/annurev.ento.43.1.17.

Duron O et al. 2017. Evolutionary changes in symbiont community structure in ticks. Mol. Ecol. 26:2905–2921. doi: 10.1111/mec.14094.

Duron O et al. 2015. The Recent Evolution of a Maternally-Inherited Endosymbiont of Ticks Led to the Emergence of the Q Fever Pathogen, *Coxiella burnetii*. PLoS Pathog. 11. doi: 10.1371/journal.ppat.1004892.

Duron O et al. 2018. Tick-Bacteria Mutualism Depends on B Vitamin Synthesis Pathways. Curr. Biol. 28:1896-1902.e5. doi: 10.1016/j.cub.2018.04.038.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797. doi: 10.1093/nar/gkh340.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20:238. doi: 10.1186/s13059-019-1832-y.

Gerhart JG, Moses AS, Raghavan R. 2016. A *Francisella*-like endosymbiont in the Gulf Coast tick evolved from a mammalian pathogen. Sci. Rep. 6. doi: 10.1038/srep33670.

Gottlieb Y, Lalzar I, Klasson L. 2015. Distinctive Genome Reduction Rates Revealed by Genomic Analyses of Two *Coxiella*-Like Endosymbionts in Ticks. Genome Biol. Evol. 7:1779–1796. doi: 10.1093/gbe/evv108.

Guizzo MG et al. 2017. A *Coxiella* mutualist symbiont is essential to the development of *Rhipicephalus microplus*. Sci. Rep. 7. doi: 10.1038/s41598-017-17309-x.

Hugoson E, Ammunét T, Guy L. 2019. Host-adaptation in *Legionellales* is 2.4 Gya, coincident with eukaryogenesis. bioRxiv. 852004. doi: 10.1101/852004.

Kernif T, Leulmi H, Raoult D, Parola P. 2016. Emerging Tick-Borne Bacterial Pathogens.

Microbiol. Spectr. 4. doi: 10.1128/microbiolspec.EI10-0012-2016.

Kuzminov A. 1999. Recombinational Repair of DNA Damage in *Escherichia coli* and Bacteriophage λ. Microbiol. Mol. Biol. Rev. 63:751–813. doi: 10.1128/MMBR.63.4.751-813.1999.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods. 9:357–359. doi: 10.1038/nmeth.1923.

López-Madrigal S, Latorre A, Porcar M, Moya A, Gil R. 2011. Complete Genome Sequence of "*Candidatus* Tremblaya princeps" Strain PCVAL, an Intriguing Translational Machine below the Living-Cell Status. J. Bacteriol. 193:5587–5588. doi: 10.1128/JB.05749-11.

Luedtke BE, Mahapatra S, Lutter EI, Shaw EI. 2017. The *Coxiella burnetii* type IVB secretion system (T4BSS) component DotA is released/secreted during infection of host cells and during in vitro growth in a T4BSS-dependent manner. Pathog. Dis. 75. doi: 10.1093/femspd/ftx047.

Massengo-Tiassé RP, Cronan JE. 2008. *Vibrio cholerae* FabV Defines a New Class of Enoyl-Acyl Carrier Protein Reductase. J. Biol. Chem. 283:1308–1316. doi: 10.1074/jbc.M708171200.

McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. Nat. Rev. Microbiol. 10:13–26. doi: 10.1038/nrmicro2670.

Moutailler S et al. 2016. Co-infection of Ticks: The Rule Rather Than the Exception. PLoS Negl. Trop. Dis. 10. doi: 10.1371/journal.pntd.0004539.

Nakabachi A et al. 2006. The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. Science. 314:267. doi: 10.1126/science.1134196.

Olivieri E et al. 2019. Tissue tropism and metabolic pathways of *Midichloria mitochondrii* suggest tissue-specific functions in the symbiosis with *Ixodes ricinus*. Ticks Tick-Borne Dis. 10:1070–1077. doi: 10.1016/j.ttbdis.2019.05.019.

Patton AJ, Hough DW, Towner P, Danson MJ. 1993. Does *Escherichia coli* possess a second citrate synthase gene? Eur. J. Biochem. 214:75–81. doi: 10.1111/j.1432-

1033.1993.tb17898.x.

Ramaiah A, Dasch GA. 2018. Genome Sequence of *Coxiella*-Like Endosymbiont Strain CLE-RmD, a Bacterial Agent in the Cattle Tick (*Rhipicephalus microplus)* Deutsch Strain. Genome Announc. 6. doi: 10.1128/genomeA.00003-18.

Ronquist F et al. 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. Syst. Biol. 61:539–542. doi: 10.1093/sysbio/sys029.

Sandström J, Moran N. 1999. How nutritionally imbalanced is phloem sap for aphids? In: Proceedings of the 10th International Symposium on Insect-Plant Relationships. Simpson, SJ, Mordue, AJ, & Hardie, J, editors. Series Entomologica Springer Netherlands: Dordrecht pp. 203–210. doi: 10.1007/978-94-017-1890-5_26.

Sasaki-Fukatsu K et al. 2006. Symbiotic bacteria associated with stomach discs of human lice. Appl. Environ. Microbiol. 72:7349–7352. doi: 10.1128/AEM.01429-06.

Schofield MJ, Hsieh P. 2003. DNA mismatch repair: molecular mechanisms and biological function. Annu. Rev. Microbiol. 57:579–608. doi: 10.1146/annurev.micro.57.030502.090847.

Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 30:2068–2069. doi: 10.1093/bioinformatics/btu153.

Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: Assessing Genome Assembly and Annotation Completeness. Methods Mol. Biol. Clifton NJ. 1962:227–245. doi: 10.1007/978-1-4939-9173-0_14.

Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. 34:D32-36. doi: 10.1093/nar/gkj014.

Smith TA, Driscoll T, Gillespie JJ, Raghavan R. 2015. A *Coxiella*-Like Endosymbiont Is a Potential Vitamin Source for the Lone Star Tick. Genome Biol. Evol. 7:831–838. doi: 10.1093/gbe/evv016.

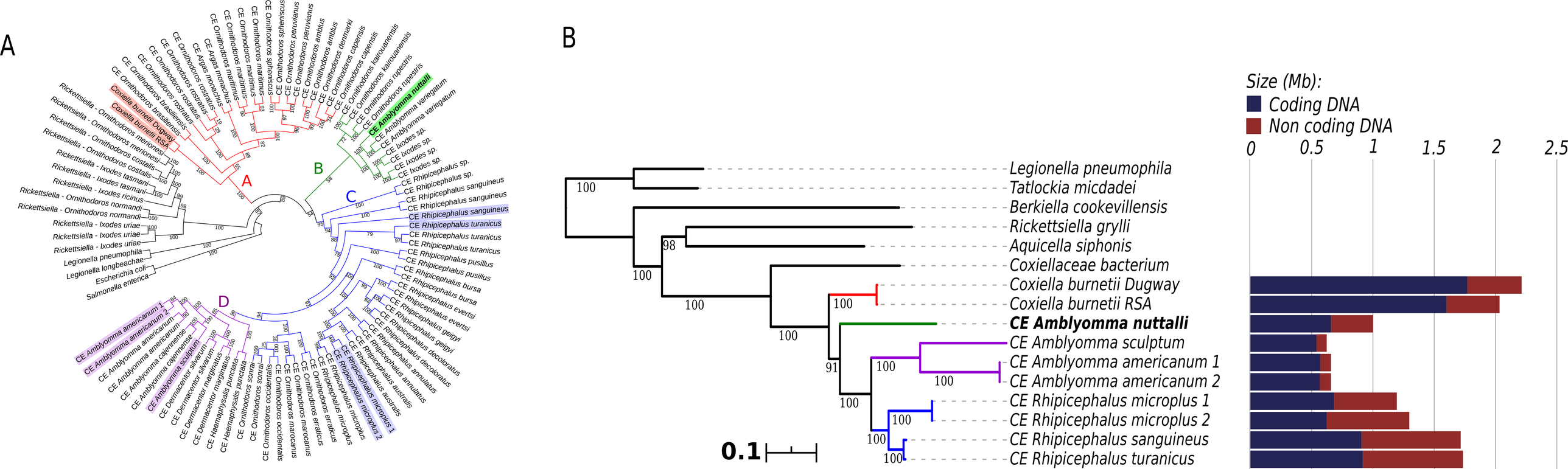
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30:1312–1313. doi: 10.1093/bioinformatics/btu033.

391 Syberg-Olsen M, Husnik F. 2018. Pseudofinder. GitHub repository: https://github.com/filip-
392 husnik/pseudo-finder/.
393 Talavera G, Castresana J. 2007. Improvement of Phylogenies after Removing Divergent and
394 Ambiguously Aligned Blocks from Protein Sequence Alignments. Syst. Biol. 56:564–577.
395 doi: 10.1080/10635150701472164.
396 Taylor MJ, Bandi C, Hoerauf A. 2005. *Wolbachia* bacterial endosymbionts of filarial
397 nematodes. Adv. Parasitol. 60:245–284. doi: 10.1016/S0065-308X(05)60004-8.
398 Theiler G, Salisbury L. 1959. Ticks in the South African Zoological Survey Collection. Part
399 IX. 'the *Amblyomma marmoreum* group'. Onderstepoort J. Vet. Res. 47–124.
400 Tsementzi D, Gordillo JC, Mahagna M, Gottlieb Y, Konstantinidis KT. 2018. Comparison of
401 closely related, uncultivated *Coxiella* tick endosymbiont population genomes reveals clues
402 about the mechanisms of symbiosis. Environ. Microbiol. 20:1751–1764. doi: 10.1111/1462-
403 2920.14104.
404 Wernegreen JJ. 2012. Endosymbiosis. Curr. Biol. 22:R555–R561. doi:
405 10.1016/j.cub.2012.06.010.
406 Xie Z, Tang H. 2017. ISEScan: automated identification of insertion sequence elements in
407 prokaryotic genomes. Bioinformatics. 33:3340–3347. doi: 10.1093/bioinformatics/btx433.
408 Zhong J, Jasinskas A, Barbour AG. 2007. Antibiotic Treatment of the Tick Vector
409 *Amblyomma americanum* Reduced Reproductive Fitness. PLoS ONE. 2. doi:
410 10.1371/journal.pone.0000405.
411 Zindel R, Gottlieb Y, Aebi A. 2011. Arthropod symbioses: a neglected parameter in pest- and
412 disease-control programmes. J. Appl. Ecol. 48:864–872. doi: 10.1111/j.1365-
413 2664.2011.01984.x.
414

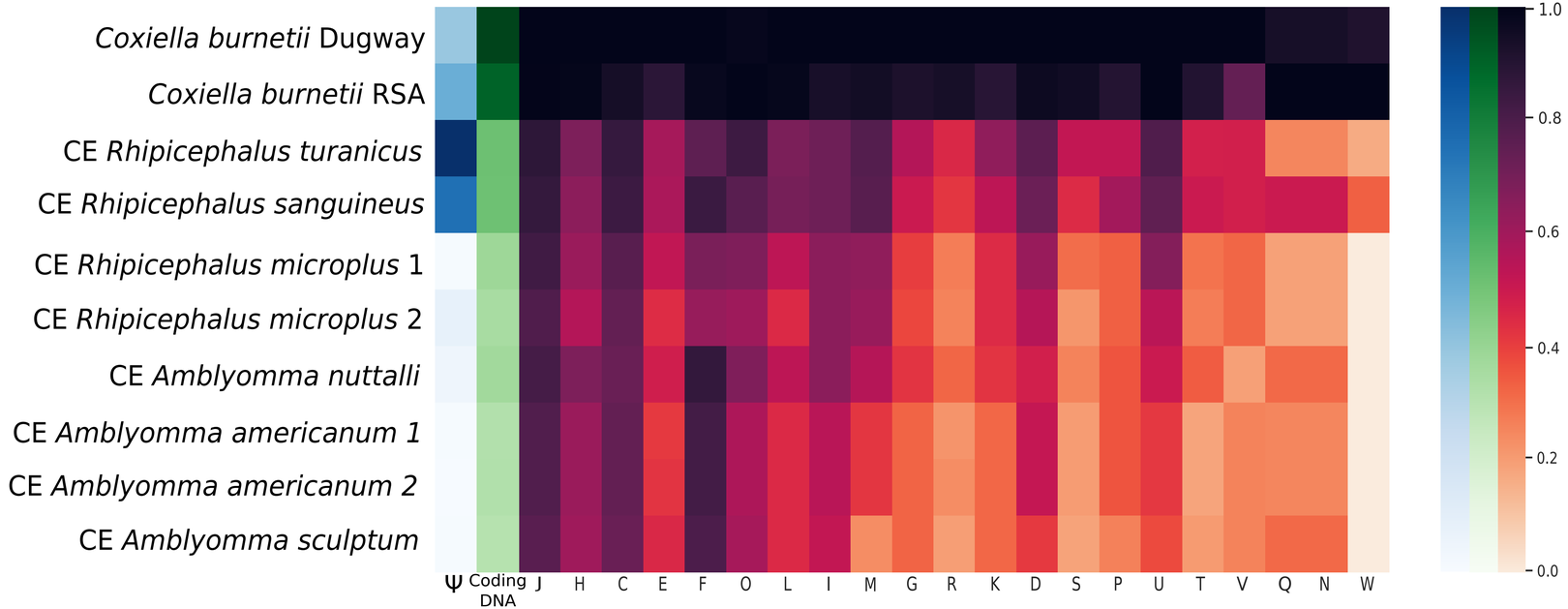
Figure legends

Fig. 1: (a) Maximum likelihood MLST phylogenetic tree of *Coxiella* and other *Legionellales* (full tree in Supplementary Figure 4), and (b) maximum likelihood phylogenomic tree of *Coxiellaceae*. In (a,b), the four major *Coxiella* clades are evidenced by different colors, numbers on branches stand for bootstrap support after 1000 pseudo-replicates, CE stands for *Coxiella* endosymbiont, and respective host organisms are indicated. In (a), symbionts with available genomes are highlighted. In (b), scale bar stands for estimated proportional sequence divergence, and the bar plot on the right shows the respective coding (blue) and non-coding (red) genome sizes.

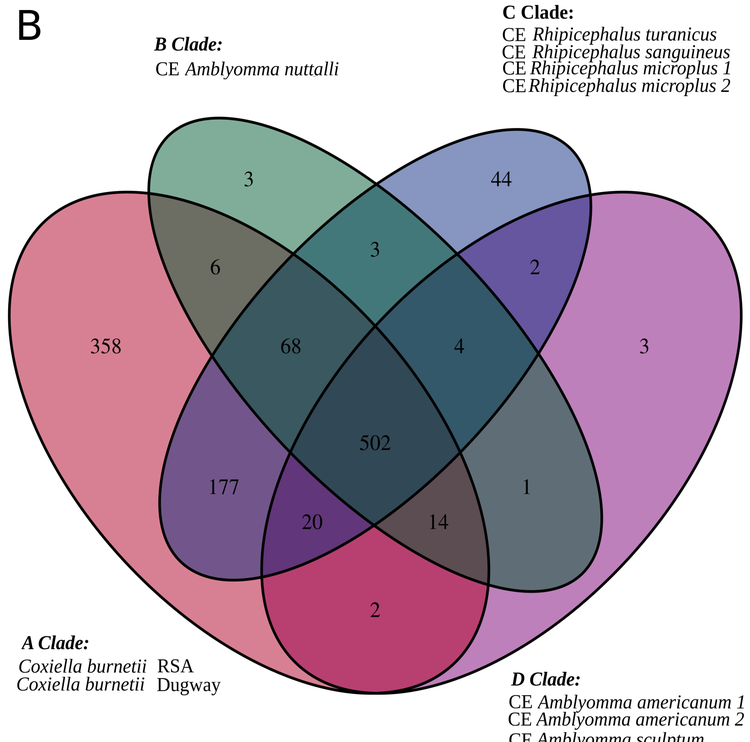
Figure 2. (a) Heatmap showing variations in number of pseudogenes (blue, indicated with the Ψ), genome size (green) and COG (Clusters of Orthologous Groups) repertoire for each functional category (orange-purple) in the *Coxiella* genus. CE stands for *Coxiella* endosymbiont. Organisms are grouped according to phylogenetic clades, in turn sorted according to coding DNA size. The color intensity is independently scaled for each column in proportion to its maximum value. (b) Venn diagram representing COG distribution on *Coxiella* clades. For each clade, COGs identified in at least one of the listed genomes are counted.


A
B
C
D
Coxiella burnetii Dugway
Coxiella burnetii RSA
CE Amblyomma nuttalli
Rickettsiella - Ornithodoros mesonesi
Rickettsiella - Ornithodoros costalis
Rickettsiella - Ixodes tasmani
Rickettsiella - Ixodes ricinus
Rickettsiella - Ornithodoros normandi
Rickettsiella - Ixodes uriae
Legionella pneumophila
Legionella longbeachae
Escherichia coli
Salmonella enterica
CE Amblyomma americanum 1
CE Amblyomma americanum 2
CE Amblyomma sculptum
CE Rhipicephalus microplus 1
CE Rhipicephalus microplus 2
CE Rhipicephalus sanguineus
CE Rhipicephalus turanicus
Legionella pneumophila
Tatlockia micdadei
Berkiella cookevillensis
Rickettsiella grylli
Aquicella siphonis
Coxiellaceae bacterium
Coxiella burnetii Dugway
Coxiella burnetii RSA
CE Amblyomma nuttalli
CE Amblyomma sculptum
CE Amblyomma americanum 1
CE Amblyomma americanum 2
CE Rhipicephalus microplus 1
CE Rhipicephalus microplus 2
CE Rhipicephalus sanguineus
CE Rhipicephalus turanicus
0.1
Size (Mb):
Coding DNA
Non coding DNA
0
0.5
1
1.5
2
2.5

A


Coxiella burnetii Dugway
Coxiella burnetii RSA
CE Rhipicephalus turanicus
CE Rhipicephalus sanguineus
CE Rhipicephalus microplus 1
CE Rhipicephalus microplus 2
CE Amblyomma nuttalli
CE Amblyomma americanum 1
CE Amblyomma americanum 2
CE Amblyomma sculptum
Ψ Coding DNA
J H C E F O L I M G R K D S P U T V Q N W
1.0
0.8
0.6
0.4
0.2
0.0

B


B Clade:
CE Amblyomma nuttalli
C Clade:
CE Rhipicephalus turanicus
CE Rhipicephalus sanguineus
CE Rhipicephalus microplus 1
CE Rhipicephalus microplus 2
3
44
3
6
2
358
68
4
3
502
177
1
20
14
2
A Clade:
Coxiella burnetii RSA
Coxiella burnetii Dugway
D Clade:
CE Amblyomma americanum 1
CE Amblyomma americanum 2
CE Amblyomma sculptum