Torix Rickettsia are widespread in arthropods and reflect a neglected symbiosis

Jack Pilgrim[1], Panupong Thongprem[1], Helen Davison[1], Stefanos Siozios[1], Matthew Baylis[1], Evgeny Zakharov[2], Sujeevan Ratnasingham[2], Jeremy deWaard[2], Alex Smith[2], and Gregory Hurst[1]

[1]University of Liverpool Faculty of Health and Life Sciences
[2]University of Guelph

July 21, 2020

Abstract

Rickettsia are intracellular bacteria best known as the causative agents of human and animal diseases. Although these medically important Rickettsia are often transmitted via haematophagous arthropods, other Rickettsia, such as those in the Torix group, appear to reside exclusively in invertebrates and protists with no secondary vertebrate host. Importantly, little is known about the diversity or host range of Torix group Rickettsia. This study describes the serendipitous discovery of Rickettsia amplicons in the Barcode of Life Data System (BOLD), a sequence database specifically designed for the curation of mtDNA barcodes. Out of 184,585 barcode sequences analysed, Rickettsia is observed in approximately 0.41% of barcode submissions and is more likely to be found than Wolbachia (0.17%). The Torix group of Rickettsia are shown to account for 95% of all unintended amplifications from the genus, with a multilocus analysis of these strains revealing this symbiont commonly shifts between distantly related host taxa. A further targeted PCR screen of 1,612 individuals from 169 terrestrial and aquatic arthropod species identified mostly Torix strains (14/16) and supports the "aquatic hotspot" hypothesis for Torix infection. Furthermore, the analysis of Sequence Read Archive (SRA) deposits indicates Torix infections represent a significant proportion of all Rickettsia symbioses. This combination of methods reveals a broad host diversity associated with Torix Rickettsia including phloem-feeding bugs, parasitoid wasps, forest detritivores and vectors of disease. The unknown host effects and transmission strategies of these endosymbionts makes these newly discovered associations important to inform future directions of investigation involving the understudied Torix Rickettsia.

Introduction

It is now widely recognized that animals live in a microbial world, and that many aspects of animal biology, ecology and evolution are a product of their symbioses with microorganisms (McFall-Ngai et al., 2013). In invertebrates, these symbioses may be particularly intimate, and involve transmission of the microbe from parent to offspring (Hurst, 2017). The alignment of host reproduction with symbiont transmission produces a correlation between the fitness interests of the parties, reflected in symbionts evolving to play a number of physiological roles within the host, from defence (Jaenike, Unckless, Cockburn, Boelio, & Perlman, 2010; Łukasik, Guo, van Asch, Ferrari, & Godfray, 2013; Teixeira, Ferreira, & Ashburner, 2008) through to core anabolic and digestive functions (Douglas, 2009; Rio, Attardo, & Weiss, 2016). However, the maternal inheritance of these microbes has led to the retention of parasitic phenotypes associated with distortion of reproduction, with symbiont phenotypes including biases towards daughter production and cytoplasmic incompatibility (Hurst & Frost, 2015). These diverse individual impacts alter the ecology and evolution of the host, in terms of diet, dynamics of interaction with natural enemies, sexual selection and speciation.

Heritable symbioses have evolved on multiple occasions amongst microbial taxa. In some cases, the microbial

lineage is limited to a single clade of related animal hosts, such as *Buchnera* in aphids (Munson, Baumann, & Kinsey, 1991). In other cases, particular heritable microbes are found across a wide range of arthropod species. *Wolbachia*represents the most common associate, considered to infect nearly half of all species (Zug & Hammerstein, 2012), and this commonness is a function in part of the ability of *Wolbachia* to transfer to a broad range of new host species and spread within them (host shift events) (Siozios, Gerth, Griffin, & Hurst, 2018; Turelli et al., 2018). Aside *Wolbachia* , other microbes are found commonly as heritable symbionts of arthropod hosts (Duron et al., 2008). *Cardinium* and*Rickettsia* , for instance, have been estimated at being present in 13-55% and 20-42% of species respectively (Weinert, Araujo-Jnr, Ahmed, & Welch, 2015).

In this paper, we address the diversity and commonness of symbioses between *Rickettsia* and arthropods. The *Rickettsia* have increasingly been recognized as a genus of bacteria with diverse interactions with arthropods (Perlman, Hunter, & Zchori-Fein, 2006; Weinert, Werren, Aebi, Stone, & Jiggins, 2009). First discovered as the agents underlying several diseases of humans vectored by haematophagous arthropods (da Rocha-Lima, 1916; Ricketts, 1909), our understanding of the group changed in the 1990s with the recognition that *Rickettsia* were commonly arthropod symbionts (Chen, Campbell, & Purcell, 1996; Fukatsu & Shimada, 1999; Werren et al., 1994).*Rickettsia* were recognized first as male-killing reproductive parasites (Hurst, Walker, & Majerus, 1996; Werren et al., 1994) and then later as beneficial partners (Hendry, Hunter, & Baltrus, 2014; Himler et al., 2011; Łukasik et al., 2013).

Following this extension of our understanding of*Rickettsia* -arthropod interactions, a new clade of*Rickettsia* was discovered from work in *Torix* leeches (Kikuchi & Fukatsu, 2005; Kikuchi, Sameshima, Kitade, Kojima, & Fukatsu, 2002). This clade was sister to all other *Rickettsia* genera, with no evidence to date of any strain having a vertebrate pathogen phase. The host range for Torix *Rickettsia* is broader than that for other members of the genus, going beyond arthropods to include leech and amoeba hosts (Dyková, Veverková, Fiala, Macháčková, & Pecková, 2003; Galindo et al., 2019; Kikuchi & Fukatsu, 2005). Targeted PCR based screening have revealed Torix group *Rickettsia* as particularly common in three groups with aquatic association: *Culicoides* biting midges, deronectid beetles and odonates (Küchler, Kehl and Dettner, 2009; Pilgrim et al., 2017; Thongprem *et al.*2020). However, some previous hypothesis-free PCR screens that aimed to detect *Rickettsia* in arthropods have likely missed these symbioses, due to divergence of the marker sequence and mismatch with the primers (Weinert, 2015).

During our previous work on Torix *Rickettsia* in biting midges (Pilgrim et al., 2017), we became aware of the presence of*Rickettsia CoxA* sequences deposited in Genbank that derived from studies where the intended target of amplification/sequencing was cytochrome *c* oxidase I (*COI*), the mitochondrial equivalent of *CoxA* . These deposits derived from studies using mtDNA for phylogeographic inference (Lagrue, Joannes, Poulin, & Blasco-Costa, 2016), or in barcoding based species identification approaches (Ceccarelli, Haddad, & Ramírez, 2016; Řezáč, Gasparo, Král, & Heneberg, 2014). Non-target amplification of *Rickettsia COI* using mitochondrial *COI* barcoding primers has been reported in spiders (Ceccarelli et al., 2016; Řezáč et al., 2014) and freshwater amphipods (Lagrue et al., 2016; Park & Poulin, 2020). Furthermore, we have noted two cases in our lab where amplicons obtained for mtDNA barcoding of an insect have, on sequence analysis, revealed *Rickettsia COI*amplification (Belli group *Rickettsia* from Collembola, and Torix group *Rickettsia* from *Cimex lectularius* bedbugs). Previous work had established barcoding approaches may amplify *COI* from*Wolbachia* symbionts (Smith et al., 2012), and the data above indicate that non-target *Rickettsia COI* may be likewise amplified during this PCR amplification for mitochondrial*COI* .

In this paper, we use three approaches to reveal the diversity and commonness of Torix *Rickettsia* in arthropods. First, we probed the contaminant bin of the Barcode of Life Data System (BOLD) for*Rickettsia* sequences and used the template from these projects to define the diversity of *Rickettsia* observed using a multilocus approach. Second, we screened DNA templates from multiple individuals from 169 arthropod species for *Rickettsia* presence using PCR assays that function more broadly than previously utilized in screens. Within this, we included a wider range of aquatic taxa, to investigate if the previous work highlighting particular aquatic taxa as hotspots for *Rickettsia* symbiosis (water beetles, biting midges, damselflies) reflects a wider higher incidence in species from this habitat. Finally, we used bioinformatic approaches to

examine the Sequence Read Archive (SRA) depositions for one individual from 1,342 arthropod species for the presence of *Rickettsia* and used this as a means of estimating the relative balance of Torix group to other*Rickettsia* within symbioses.

Materials and Methods

Interrogation of the Barcode of Life Data System (BOLD)

BOLD datasets acquisition

Access was permitted to analyse *COI* barcoding data deriving from a BOLD screening project totaling 184,585 arthropod specimens from 21 countries and collected between 2010 and 2014. *COI* sequences provided by BOLD were generally derived from templates created from somatic tissues (legs are often used in order to retain most of the specimen for further analyses if necessary), but also rarely included abdominal tissues. The first dataset made available included 3,817 sequences deemed as contaminant sequences, defined as not matching initial morphotaxa assignment. The second dataset included 55,366 specimens judged to not contain non-target amplicons ([dataset] Zakharov, Ratnasingham, deWaard & Smith, 2020). A remaining 125,402 specimens were not made available, and the 55,366 subsample was used as a representative sample from which the contaminants had originated (Figure 1).

Assessment of non-target microbe amplicons

Contaminant pool sequences were then compared against a variety of potential contaminants including *Wolbachia* and human homologues. Bacterial identity was then refined by phylogenetic placement. To this end, barcodes confirmed as microbial sequences were aligned using the "L-INS-I" algorithm in MAFFT v7.4 (Katoh & Standley, 2013) before using Gblocks (Castresana, 2000) to exclude areas of the alignment with excessive gaps or poor alignment. ModelFinder (Kalyaanamoorthy, Minh, Wong, von Haeseler, & Jermiin, 2017) then determined the TIM3+F+I+G4 model to be used after selection based on default "auto" parameters using the Bayesian information criteria. A maximum likelihood (ML) phylogeny was then estimated with IQTree (Nguyen, Schmidt, Von Haeseler, & Minh, 2015) using an alignment of 561 nucleotides and 1000 ultrafast bootstraps (Hoang, Chernomor, Haeseler, Minh, & Vinh, 2017). The Rickettsiales genera*Anaplasma* , *Neorickettsia* , *Rickettsia* and*Wolbachia* (Supergroups A, B, E, F, H) were included in the analysis as references. Finally, phylogenetic trees were drawn and annotated based on host taxa (order) using the EvolView (He et al., 2016) online tree annotation and visualisation tools.

A determining factor for non-target amplification of bacteria is primer site matching to microbial associates. Subsequently, pairwise homology of the primer set predominantly used for BOLD barcode screening was compared to *Rickettsia* and *Wolbachia COI* genes.

Further phylogenetic analysis

COI sequence alone provides an impression of the frequency with which *Rickettsia* associates are found in barcoding studies. However, they have limited value in describing the diversity of the*Rickettsia* found. To provide further insight into the diversity of *Rickettsia* using a multilocus approach, we obtained 186 DNA extracts from the archive at the Centre for Biodiversity Genomics (University of Guelph, Canada) that had provided *Rickettsia*amplicons in the previous screen. Templates were chosen based on varied collection site, host order and phylogenetic placement. Multilocus PCR screening and phylogenetic analysis of *Rickettsia* was then completed, using the methodology in Pilgrim et al. 2017. However, slight variations include the exclusion of the *atpA* gene due to observed recombination at this locus. Furthermore, the amplification conditions for the *17KDa* locus was changed because a Torix *Rickettsia*reference DNA extract (Host: *Simulium aureum*) failed to amplify with the primer set Ri_17KD_F/ Ri_17KD_R from Pilgrim et al. 2017. Subsequently, a *17KDa* alignment from genomes spanning the Spotted fever, Typhus, Transitional, Belli, Limoniae groups, and the genus*Megaira* was generated to design a new set of primers using the online tool PriFi (Fredslund, Schauser, Madsen, Sandal, & Stougaard, 2005).

Once multilocus profiles of the *Rickettsia* had been established, we tested for recombination within and

between loci using RDP v4 (Martin, Murrell, Golden, Khoosal, & Muhire, 2015) using the MaxChi, RDP, Chimaera, Bootscan and GENECONV algorithms with the following criteria to assess a true recombination positive: a p-value of <0.001; sequences were considered linear with 1000 permutations being performed. Samples amplifying at least 3 out of 4 genes (*16S rRNA* , *17KDa* , *COI* and *gltA*) were then concatenated and their relatedness estimated using maximum likelihood as previously described. The selected models used in the concatenated partition scheme (Chernomor, von Haeseler, & Minh, 2016) were as follows: *16S rRNA* : TIM3+F+R2; *17KDa* : GTR+F+I+G4;*COI:* TVM+F+I+G4; *gltA:* TVM+F+I+G4. Accession numbers for all sequences used in phylogenetic analyses can be found in Table S1.

Re-barcoding Rickettsia-containing BOLD DNA extracts

Aside from phylogenetic placement of these *Rickettsia* -containing samples, attempts were made to extract an mtDNA barcode from these taxa in order to identify the hosts of infected specimens. Previous non-target amplification of *Rickettsia* through DNA barcoding of arthropod DNA extracts had occurred in the bed bug *Cimex lectularius* , with a recovery of the true barcode after using the primer set C1-J-1718/HCO1490, which amplifies a shortened 455 bp sequence within the *COI* locus. Subsequently, all samples were screened using these primers or a further set of secondary *COI* primers (LCOt_1490/ MLepR1 and LepF1/C_ANTMR1D) if the first failed to give an adequate host barcode. All *COI* and *Rickettsia* multilocus screening primer details, including references, are available in Table S2.

Cycling conditions for *COI* PCRs were as follows: initial denaturation at 95°C for 5 min, followed by 35 cycles of denaturation (94°C, 30 sec), annealing (50°C, 60 sec), extension (72°C, 90 sec), and a final extension at 72°C for 7 min. *Rickettsia* and host amplicons identified by gel electrophoresis were subsequently purified enzymatically (ExoSAP) and Sanger sequenced through both strands using a BigDye® Terminator v3.1 kit (Thermo Scientific, Waltham, USA), and capillary sequenced on a 3500 xL Genetic Analyser (Applied Biosystems, Austin, USA). Forward and reverse reads were assessed in UGENE (Okonechnikov, Golosova, & Fursov, 2012) to create a consensus sequence by eye with a cut-off phred (Q) score (Ewing, Hillier, Wendl, & Green, 1998) of 20. Primer regions were trimmed from barcodes before being matched to Genbank and BOLD databases by BLAST based on default parameters and an e-value threshold of <1e-85. Host taxonomy was determined by a barcode-based assignment of the closest BLAST hit, under the following criteria modified from Ramage et al. (2017):

1) Species level designation for at least 98% sequence identity.

2) Genus level designation for at least 95% sequence identity.

3) Family level designation for at least 85% sequence identity.

Additionally, all sequences were required to be at least >200 bp in length.

Assessment of barcoding success

One of the factors determining a successful *COI* bacterial amplification is the initial failure of an extract to amplify mtDNA. Subsequently, to determine the likelihood of this event within taxa, we used the 55,366 specimen representative data subset to evaluate failure rates. To this end, all orders of host which gave at least one non-target *Rickettsia COI* hit were assessed. The barcoding success rate was determined as the proportion of specimens which matched initial morphotaxa assignment and were not removed after BOLD quality control (Ratnasingham & Hebert, 2007). As the total *Rickettsia* count was from a larger dataset than the one made available, an adjusted prevalence for each taxon was calculated based on the representative data subset.

Targeted and bioinformatic *Rickettsia* screens

Targeted screen of aquatic and terrestrial arthropods

A targeted *Rickettsia* screen was undertaken as an adjunct to the BOLD screen to increase our understanding of *Rickettsia* ecology. Overall, 1,612 individuals from 169 species, including both terrestrial (DNA templates

derived from European material, mostly from Duron et al., 2008) and aquatic invertebrates (largely acquired from the UK between 2016-2018), were screened. mtDNA *COI* amplification was conducted as a control for DNA quality. To investigate symbiont infection status, rickettsial-specific primers based on *gltA* and *16S rRNA*genes were used for conventional PCR screening (Pilgrim et al., 2017), with Sanger sequences obtained from at least one specimen per*Rickettsia* positive species to identify any misamplification false positives. Subsequently, the number of species containing at least one Torix *Rickettsia* positive individual were compared between terrestrial and aquatic environments using the Fisher's Exact test. Additionally, newly identified hosts of interest from BOLD and targeted screens were also placed phylogenetically (see sections above) before being mapped by lifestyle and diet.

Search of the Sequence Read Archive (SRA) and NCBI

Further insights into the balance of *Rickettsia* groups within arthropod symbioses was obtained through searching for Rickettsia presence in Illumina datasets associated with arthropod whole genome sequence (WGS) projects in the SRA (60,409 records as of the 20th May 2019). To reduce the bias from over-represented laboratory model species (e.g. *Drosophila* spp., *Anopheles* spp.) a single dataset per species was examined, and where multiple data sets existed for a species, that with the largest read counts was retained. The resultant dataset, representing 1,342 arthropod species (Table S3), was then screened with phyloFlash (Gruber-Vodicka, Seah, & Pruesse, 2019) which finds, extracts and identifies *16S rRNA* sequences.

Reconstructed full *16S rRNA* sequences affiliated to*Rickettsia* were extracted and compared to sequences derived from the targeted screen phylogenetically (see sections above) to assess group representation within the genus. The microbial composition of all SRA datasets that did not result in a reconstructed *Rickettsia 16S rRNA* with phyloFlash were re-evaluated using Kraken2 (Wood, Lu, & Langmead, 2019), a k-mer based taxonomic classifier for short DNA sequences. A cut-off of at least 40k reads assigned to *Rickettsia*taxa was applied for reporting potential infections (theoretical genome coverage of ~ 1 – 4X assuming an average genome size of ~1.5Mb).

We also examined NCBI for *Rickettsia* sequences deposited as invertebrate *COI* barcodes. To this end, a BLAST search of Torix*Rickettsia COI* sequences from previous studies (Ceccarelli et al., 2016; Pilgrim et al., 2017) was conducted on the 29th June 2020. Sequences were initially considered belonging to the Torix group if their similarity was >90% and subsequently confirmed phylogenetically as described above.

Results

Torix Rickettsia is the most common bacterial contaminant produced during barcoding projects

Out of 3,817 sequences considered contaminants, 1,126 of these were deemed by BOLD to be bacterial in origin (Figure 1, Table S4). Phylogenetic placement supported the correct designation of these sequences as of microbial origin (Figure 2). The dominant genus was*Rickettsia* with 753 (66.9%) amplifications, compared to*Wolbachia* with 306 (27.2%). Of the remaining 67 non-target sequences, 16 formed a monophyletic group with other Anaplasmataceae and 51 were undesignated proteobacteria. When considering the 184,585 specimens in the total project, this analysis gave an overall*Rickettsia* and *Wolbachia* prevalence of 0.41% and 0.17% respectively within the dataset. Through later access to the 55,366 representative data subset from where the contaminants originated, further unique bacteria contaminants were also detected (possibly missed by BOLD's automated contaminant filtering system). This suggests these prevalences are conservative estimates.

BOLD *Rickettsia* contaminants were dominated by amplicons from the Torix group of *Rickettsia* (716/753; 95.1%) (Figure 3). The remaining 37 *Rickettsia* clustered with Transitional/Spotted Fever (n=15), Belli (n=9), Rhyzobius (n=1) groups, while 12 sequences formed two unique clades (Table S4). Across arthropod hosts: 292 (38.8%) were derived from Hymenoptera; 189 (25.1%) from Diptera; 177 from Hemiptera (23.5%); 41 from Psocoptera (5.4%); 40 from Coleoptera (5.3%); 7 from Arachnida (0.9%); 4 from Trichoptera (0.5%); and single cases of Thysanoptera, Diplopoda and Dermaptera (0.1% each). Mapping the 753*Rickettsia* to collection site (Figure S1) revealed arthropod infections predominantly from Canada with other locations in

South/Central America, Europe, Africa and Asia.

We observed that two sets of *COI* primers were responsible for 99% of *Rickettsia* amplifications (Table S5) with a majority (89%) amplifying with the primer combination C_LepFolF/C_LepFolR (Hernández-Triana et al., 2014). Torix *Rickettsia COI* showed a stronger match to these primers at the 3' end (the site responsible for efficient primer annealing) compared to *Wolbachia* and other*Rickettsia* groups. Whilst all contained a SNP at the 3' priming end of C_LepFolR, Torix *Rickettsia* (*Rickettsia*endosymbiont of *Culicoides newsteadi* ; MWZE00000000) was the only sequence to not contain a similar SNP at the 3' priming site of C_LepFolF (Tables S6.1 and S6.2).

Rickettsia multilocus phylogenetic analysis

To better resolve the phylogenetic relationships between BOLD*Rickettsia* contaminants, a multilocus approach was employed on a subsample of 186 *Rickettsia* -containing samples. To this end, 2 further housekeeping genes (*16S rRNA* , *gltA*) and the antigenic *17KDa* protein gene were amplified from the respective templates.

Overall, 135 extracts successfully amplified and gave a high-quality sequence for at least one gene. No intragenic or intergenic recombination was detected for any of the gene profiles. A phylogram, including 99 multilocus profiles containing at least 3 of the 4*Rickettsia* genes of interest (including *COI*), allocated strains to both Limoniae and Leech subclades of the Torix group (Figure 4) and these subclades were derived from similar hosts. For example, specific families (Hemiptera: Psyliidae and Hymenoptera: Diapriidae) were present in both Leech and Limoniae groups. A full list of multilocus profiles and *Rickettsia* group designation can be found in Table S7.

The multilocus study also provided evidence of co-infection with*Rickettsia* . During Sanger chromatogram analysis, double peaks were occasionally found at third codon sites from protein coding genes. This pattern was observed in 6/10 *Philotarsus californicus*individuals and in one member of each of the Psilidae, Sciaridae , Chironomidae and Diapriidae (Table S7). Where double peaks were observed, this was found consistently across markers within an individual specimen. This pattern indicates co-infecting*Rickettsia* strains in hosts is a widespread phenomenon of the Torix group.

Barcoding success of host taxa

An available subset of attempted barcodes associated with the contaminants contained 55,366 out of 184,585 arthropods originally used in the overall study. The three classes of Insecta (n=49,688), Arachnida (n=3,626) and Collembola (n=1,957), accounted for >99.8% of total specimens (Figure 1). Successful amplification and sequencing of *COI* was achieved in 43,246 specimens (78.1%) of the genomic extracts, but when assessed at the order level success rates varied (Table S8). The likely explanation for this variation is taxa-specific divergence of sequences at priming sites.

The number of each taxonomic order giving at least one *Rickettsia*amplification was then calculated and adjusted based on the total number of specimens in the project to allow for a prevalence estimate. Overall, Hymenoptera, Diptera and Hemiptera were the three taxa most likely to be associated with *Rickettsia COI* amplification (87.4%). Similarly, on assessment of a subsample from the project where the contaminants originated, a majority (77.7%) of the dataset were also accounted for by these three orders. After adjusting the prevalence to take into account the number of inaccessible specimens, Trichoptera (2.45%), Dermaptera (1.89%) and Psocodea (1.67%) were the most likely taxa to give an inadvertent *Rickettsia* amplification. Despite Hemiptera and Diptera having a similar estimated prevalence (0.58% and 0.56%), Hemiptera were much more likely to fail to barcode (67.2% vs 93.3%) indicating the true dipteran prevalence is likely to be higher, as a barcoding failure is necessary to amplify non-target bacteria*COI* . Attempts to re-barcode 186 *Rickettsia* -containing DNA templates of interest from BOLD resulted in 90 successful arthropod host barcodes (Table S7).

Rickettsia bacterial diversity detected by Targeted and SRA screens

The targeted *Rickettsia* screen of 1,612 individuals from 169 invertebrate species detected infections in 16 species (9.47%) including both aquatic and terrestrial taxa (Table 1). Of these, 14 profiles clustered within the Torix group with the remaining two placed in the Belli and Rhyzobius groups (Figure 5). Comparisons of Torix*Rickettsia* frequency between aquatic/semiaquatic vs terrestrial arthropods revealed evidence for a higher representation of Torix*Rickettsia* infected species in the aquatic biome (Fisher's Exact, $P = 0.019$).

During the SRA search, phyloFlash flagged 29 *Rickettsia* sequences in the groups: Bellii (n=12), Torix (n=8), Transitional (n=6), Rhyzobius (n=2), and Spotted Fever (n=1) (Figure 5). In addition, Kraken identified nine *Rickettsia* -containing arthropod SRA datasets missed by phyloFlash. Two of these were from the Torix group, in phantom midge hosts (Diptera: Chaobaridae: *Mochlonyx cinctipes* and*Chaoborus trivitattus*), with the remaining seven placed in Belli and Spotted Fever groups (Table S3). The search of NCBI revealed 11 deposits ascribed to host mtDNA that were in fact Torix*Rickettsia* sequences (Table S9 and Figure S2).

The hidden host diversity of Torix Rickettsia

Overall, novel Torix hosts detected from all three screening methods included taxa from the orders Gastropoda, Trichoptera and Trombidiformes. Additionally, new Torix-associated families, genera and species were identified. These included haematophagous flies (*Simulim aureum; Anopheles plumbeus; Protocalliphora azurea* ; Tabanidae), several parasitoid wasp families (e.g. Ceraphronidae; Diapriidae; Mymaridae), forest detritivores (e.g. Sciaridae; Mycetophilidae; Staphylinidae) and phloem-feeding bugs (Psyllidae; Ricaniidae). Feeding habits such as phloem-feeding, predation, detritivory or haematophagy were not correlated with any particular Torix *Rickettsia* subclade (Figure 6). Furthermore, parasitoid and aquatic lifestyles were seen across the phylogeny. All newly discovered Torix *Rickettsia* host taxa are described in Table 2, alongside previously discovered hosts in order to give an up to date overview of Torix-associated taxa.

Discussion

Symbiotic interactions between hosts and microbes are important drivers of host phenotype, with symbionts both contributing to, and degrading, host performance. Heritable microbes are particularly important contributors to arthropod biology, with marked attention focused on*Wolbachia* , the most common associate (Hilgenboecker, Hammerstein, Schlattmann, Telschow, & Werren, 2008; Zug & Hammerstein, 2012). Members of the Rickettsiales, like *Wolbachia* , share an evolutionary history with mitochondria (Ferla, Thrash, Giovannoni, & Patrick, 2013; Wang & Wu, 2015), such that a previous screen of BOLD submissions of mtDNA submissions observed *Wolbachia* as the main bacterial contaminant associated with DNA barcoding (Smith et al., 2012). However, our BOLD screen found that *Rickettsia* were more likely to be amplified than *Wolbachia* (0.41% vs 0.17% of deposits). Furthermore, Torix group *Rickettsia* were overrepresented in barcode misamplifications (95%) when compared to other groups within the genus. A comparison of the most commonly used barcoding primers to *Wolbachia* and *Rickettsia* genomes suggest homology of the forward primer 3' end was likely responsible for this bias towards Torix *Rickettsia* amplification. To gain a clearer understanding of the relative balance of Torix group to other*Rickettsia* within symbioses and habitats, a targeted screen and bioinformatic approach was also undertaken. Through these three screens, a broad range of host diversity associated with Torix *Rickettsia*was uncovered.

As the *in silico* and empirical evidence suggests *Rickettsia COI* amplification is not uncommon (Řezáč et al., 2014; Ceccarelli et al., 2016; Park & Poulin, 2020), why has this phenomenon not been described more widely before? The conduction of a previous large-scale non-target *COI* study using BOLD submissions (Smith et al., 2012), revealed only *Wolbachia* hits. This screen involved comparison to a *Wolbachia* -specific reference library and was thus likely to miss *Rickettsia* . Additionally, there has been a lack of Torix *Rickettsia COI* homologues to compare barcodes to until recently, where a multilocus identification system, including *COI*was devised (Pilgrim et al., 2017). Indeed, out of the contaminant dataset received in this study, some of the *Rickettsia* contaminants were tentatively described by BOLD as *Wolbachia* due to the previous absence of publicly available *Rickettsia COI* to compare.

Although *Rickettsia* will only interfere with barcoding in a minority of cases (˜0.4%), it is likely that alternate screening primers for some studies will need to be considered. In a demonstration of how unintended

Rickettsia amplifications can affect phylogeographic studies relying on DNA barcoding, a*Rickettsia COI* was conflated with the mtDNA *COI* of a species of freshwater amphipod, *Paracalliope fluvitalis*(Lagrue et al., 2016). Subsequently, supposed unique mtDNA haplotypes were allocated to a particular collection site, whereas this merely demonstrated the presence of Torix *Rickettsia* in host individuals in this lake. Contrastingly, non-target *Rickettsia* amplification can also allow for the elucidation of a novel host range of the symbiont (Ceccarelli et al., 2016; Park & Poulin, 2020; Řezáč et al., 2014) and this has been exemplified with our probing of BOLD.

Previously, several host orders have been associated with Torix*Rickettsia* , including Araneae, Coleoptera, Diptera, Hemiptera and Odonata (Goodacre et al., 2006; Küchler, Kehl and Dettner, 2009; Machtelinckx et al., 2012; Martin et al., 2013; Thongprem et al. 2020). However, newly uncovered host orders from this study include Gastropoda (snails), Trichoptera (caddis flies) and Trombidiformes (mites) (Table 2). Caution needs to be taken when interpreting what these newly found associations mean, as mere presence of *Rickettsia* DNA does not definitively indicate an endosymbiotic association. Indeed, parasitism or ingestion of symbiont-infected biota can also result in PCR detection (Le Clec'h et al., 2013; Plantard et al., 2012; Ramage et al., 2017). Additionally, by calculating barcode success rate at an order level, Hemiptera were deemed to fail barcoding (either lack of amplification and/or quality sequence) more commonly than Diptera despite having a similar estimated *Rickettsia* prevalence (Table S8). As an increased barcoding failure rate is correlated with non-target*COI* amplification, it is probable there is a higher overall proportion of Torix *Rickettsia* -associated Diptera than Hemiptera in BOLD and likely in nature.

Model-based estimation techniques suggest *Rickettsia* are present in between 20-42% of arthropod species (Weinert et al., 2015). However, targeted screens often underestimate the incidence of*Rickettsia* hosts due to various methodological biases including small within-species sample sizes (missing low-prevalence infections) and the use of non-conserved primers (Weinert, 2015). Importantly, the inclusion and exclusion of specific ecological niches can also lead to a skewed view of *Rickettsia* symbioses. A previous review of *Rickettsia* bacterial and host diversity by Weinert et al. (2009) suggested a possible (true) bias towards aquatic taxa in the Torix group. In accordance with this, our targeted screen demonstrated Torix*Rickettsia* infections were more prevalent in aquatic arthropod species compared to terrestrial. However, our observed over-representation of Torix group *Rickettsia* (14/16 strains) contrasts with Weinert's findings which show a predominance of Belli infections and is likely due to the latter study's absence of screened aquatic taxa. Furthermore, through the additional use of a bioinformatics approach, our SRA search appears to confirm that Belli and Torix are two of the most common *Rickettsia* groups among arthropods. Overall, these multiple screening methods suggest Torix*Rickettsia* are more widespread than previously thought and their biological significance underestimated.

Previous studies have used either one or two markers to identify the relatedness of strains found in distinct hosts. In this study, we use the multilocus approach developed in Pilgrim et al. (2017) to understand the affiliation of Torix *Rickettsia* from diverse invertebrate hosts. Our analysis of Torix strains indicates that closely related strains are found in distantly related taxa. Closely related*Rickettsia* are also found in hosts from different niches and habitats – for instance, the *Rickettsia* strains found in terrestrial blood feeders do not lie in a single clade, but rather are allied to strains found in non-blood feeding host species. Likewise, strains in phloem feeding insects are diverse rather than commonly shared.

The distribution of Torix *Rickettsia* across a broad host range suggests host shifts are occurring between distantly related taxa. It is notable that parasitoid wasps are commonly infected with*Rickettsia* and have been associated with enabling symbiont host shifts (Ahmed et al., 2015; Vavre, Fleury, Lepetit, Fouillet, & Bouletreau, 1999). Aside from endoparasitoids, it is also possible that plant-feeding can allow for endosymbiont horizontal transmission (Caspi-Fluger et al., 2012; Gonella et al., 2015; S.-J. Li et al., 2017). For example, *Rickettsia* horizontal transmission has been demonstrated in *Bemisia* whiteflies infected by phloem-feeding (Caspi-Fluger et al., 2012; Y.-H. Li et al., 2017). Finally, ectoparasites like the Torix-infected water mites of the Calyptostomatidae family, could also play a role in establishing novel*Rickettsia* -host associations, as feeding by mites has been observed to lead to host shifts for other endosymbiont taxa (Jaenike, Polak, Fiskin,

Helou, & Minhas, 2007). Indeed, if multiple horizontal transmission paths do exist, this could account for the diverse plethora of infected taxa, as well as arthropods identified in this study which harbour more than one strain of symbiont (Morrow, Frommer, Shearman, & Riegler, 2014; Vavre et al., 1999).

The finding that Torix *Rickettsia* are associated with a broad range of invertebrates leads to an obvious question: what is the impact and importance of these symbiotic associations? Previous work has established Torix *Rickettsia* represent heritable symbionts and it is likely that this is true generally. There have, however, been few studies on their impact on the host. In the earliest studies (Kikuchi & Fukatsu, 2005; Kikuchi et al., 2002), *Torix* spp. leeches infected with *Rickettsia* were observed to be substantially larger than their uninfected counterparts. Since then, the only observation of note, pertaining to the Torix group, is the reduced ballooning (dispersal) behaviour observed in infected *Erigone atra* money spiders (Goodacre et al., 2009). Overall, the incongruencies in host and Torix *Rickettsia*phylogenies (suggesting a lack of co-speciation and obligate mutualism), along with the lack of observed sex bias in carrying the symbiont, indicate facultative benefits are the most likely symbiotic relationship (Jaenike, 2012; Weinert, 2015). However, *Rickettsia* induction of thelytokous parthenogenesis should not be discounted in Torix infected parasitoid wasps identified in this study (Giorgini, Bernardo, Monti, Nappo, & Gebiola, 2010; Hagimori, Abe, Date, & Miura, 2006). To add to the challenge of understanding Torix *Rickettsia*symbioses, the challenges of laboratory rearing of many Torix*Rickettsia* hosts has led to difficulties in identifying model systems to work with. However, the large expansion of our Torix group host knowledge can now allow for a focus on cultivatable hosts (e.g phloem-feeding bugs).

A particularly important group for study are haematophagous host species. Our discovery of *Rickettsia* -associated tabanid and simulid flies, alongside *Anopheles plumbeus* mosquitoes, add to existing blood-feeders previously identified as Torix group hosts which include sand flies (Kaili Li et al., 2016; Reeves, Kato, & Gilchriest, 2008), fleas (Song et al., 2018), ticks (Floris et al., 2008) bed bugs (Potts, Molina, Sheele, & Pietri, 2020) and biting midges (Pilgrim et al., 2017). Some *Rickettsia* strains are known to be transmitted to vertebrates via haematophagy (Parola, Paddock, & Raoult, 2005). However, there is no evidence to date for vertebrate pathogenic potential for the Torix group. Despite this, Torix*Rickettsia* could still play a significant role in the ecology of vectors of disease. A key avenue of research is whether these endosymbionts alter vectorial capacity, as found for other associations (Bourtzis et al., 2014; Hoffmann, Ross, & Rašić, 2015). In contrast to the widely reported virus blocking phenotype observed in *Wolbachia* -infected vectors (Moreira et al., 2009; van den Hurk et al., 2012; Walker et al., 2011),*Rickettsia* have been associated with a virus potentiating effect in *Bemisia* white flies vectoring Tomato yellow leaf curl virus (Kliot, Cilia, Czosnek, & Ghanim, 2014). Additionally, we uncovered a*Rickettsia* -infected psyllid (*Cacopsylla melanoneura*) which is a vector of *Phytoplasma mali* (apple proliferation) (Tedeschi, Visentin, Alam, & Bosco, 2003). Thus, the question of Torix*Rickettsia* vector-competence effects is clearly of widespread relevance and deserves further attention.

To conclude, we have shown that large-scale DNA barcoding initiatives of arthropods can include non-target amplification of Torix*Rickettsia* . By examining these non-target sequences, alongside a targeted screen and SRA search, we have uncovered numerous previously undetected host associations. Our findings lay bare multiple new avenues of inquiry for Torix *Rickettsia* symbioses.

Acknowledgments

We would like to thank Dr. Michael Gerth for kindly providing comments on the manuscript. This work was supported by: a BBSRC Doctoral Training Partnership studentship (BB/M011186/1) awarded to JP; a Development and Promotion of Science and Technology Talents Project (DPST) of the Institute for the Promotion of Teaching Science and Technology, Thailand to PT and a Harry Smith Vacation studentship (Microbiology society) to HD.

References

Ahmed, M. Z., Li, S.-J., Xue, X., Yin, X.-J., Ren, S.-X., Jiggins, F. M., ... Qiu, B.-L. (2015). The intracellular bacterium *Wolbachia* uses parasitoid wasps as phoretic vectors for efficient horizontal transmission.

PLOS Pathogens , *11* (2), e1004672. doi: 10.1371/journal.ppat.1004672

Bourtzis, K., Dobson, S. L., Xi, Z., Rasgon, J. L., Calvitti, M., Moreira, L. A., ... Gilles, J. R. L. (2014). Harnessing mosquito-*Wolbachia* symbiosis for vector and disease control.*Acta Tropica* , *132* (1). doi: 10.1016/j.actatropica.2013.11.004

Caspi-Fluger, A., Inbar, M., Mozes-Daube, N., Katzir, N., Portnoy, V., Belausov, E., ... Zchori-Fein, E. (2012). Horizontal transmission of the insect symbiont *Rickettsia* is plant-mediated.*Proceedings of the Royal Society B: Biological Sciences* ,*279* (1734), 1791–1796. doi: 10.1098/rspb.2011.2095

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* , *17* (4), 540–552. doi: 10.1093/oxfordjournals.molbev.a026334

Ceccarelli, F. S., Haddad, C. R., & Ramírez, M. J. (2016). Endosymbiotic Rickettsiales (Alphaproteobacteria) from the spider genus*Amaurobioides* (Araneae: Anyphaenidae). *Journal of Arachnology* , *44* (2), 251–253. doi: 10.1636/JoA-S-15-009

Chen, D.-Q., Campbell, B. C., & Purcell, A. H. (1996). A new*Rickettsia* from a herbivorous insect, the pea aphid*Acyrthosiphon pisum* (Harris). *Current Microbiology* ,*33* (2), 123–128. doi: 10.1007/s002849900086

Chernomor, O., von Haeseler, A., & Minh, B. Q. (2016). Terrace aware data structure for phylogenomic inference from supermatrices.*Systematic Biology* , *65* (6), 997–1008. doi: 10.1093/sysbio/syw037

da Rocha-Lima, H. (1916). Zur aetiologie des Fleckfiebers.*Deutsche Medizinische Wochenschrift* , *53* , 567–569. doi: 10.1055/s-0028-1144228

Douglas, A. E. (2009). The microbial dimension in insect nutritional ecology. *Functional Ecology* , *23* (1), 38–47. doi: 10.1111/j.1365-2435.2008.01442.x

Drummond, A. J., Newcomb, R. D., Buckley, T. R., Xie, D., Dopheide, A., Potter, B. C. M., ... Nelson, N. (2015). Evaluating a multigene environmental DNA approach for biodiversity assessment. *GigaScience* , *4* (1), 46. doi: 10.1186/s13742-015-0086-1

Drury, J. P., Okamoto, K. W., Anderson, C. N., & Grether, G. F. (2015). Reproductive interference explains persistence of aggression between species. *Proceedings of the Royal Society B: Biological Sciences* ,*282* (1804), 20142256. doi: 10.1098/rspb.2014.2256

Duron, O., Bouchon, D., Boutin, S., Bellamy, L., Zhou, L., Engelstädter, J., & Hurst, G. D. D. (2008). The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. *BMC Biology* , *6* (1), 27. doi: 10.1186/1741-7007-6-27

Dyková, I., Veverková, M., Fiala, I., Macháčková, B., & Pecková, H. (2003). *Nuclearia pattersoni* sp. n. (Filosea), a new species of amphizoic amoeba isolated from gills of roach (*Rutilus rutilus*), and its rickettsial endosymbiont. *Folia Parasitologica* ,*50* (3), 161–170.

Ewing, B., Hillier, L., Wendl, M. C., & Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment.*Genome Research* , *8* (3), 175–185. doi: 10.1101/gr.8.3.175

Ferla, M. P., Thrash, J. C., Giovannoni, S. J., & Patrick, W. M. (2013). New rRNA gene-based phylogenies of the Alphaproteobacteria provide perspective on major groups, mitochondrial ancestry and phylogenetic instability. *PLoS ONE* , *8* (12), e83383. doi: 10.1371/journal.pone.0083383

Floris, R., Yurtman, A. N., Margoni, E. F., Mignozzi, K., Boemo, B., Altobelli, A., & Cinco, M. (2008). Detection and identification of*Rickettsia* species in the Northeast of Italy. *Vector-Borne and Zoonotic Diseases* , *8* (6), 777–782. doi: 10.1089/vbz.2008.0006

Fredslund, J., Schauser, L., Madsen, L. H., Sandal, N., & Stougaard, J. (2005). PriFi: using a multiple alignment of related sequences to find primers for amplification of homologs. *Nucleic Acids Research* ,*33* , W516–W520. doi: 10.1093/nar/gki425

Fukatsu, T., & Shimada, M. (1999). Molecular characterization of*Rickettsia* sp. in a bruchid beetle, *Kytorhinus sharpianus*(Coleoptera : Bruchidae). *Applied Entomology and Zoology* ,*34* (3), 391–397. doi: 10.1303/aez.34.391

Galindo, L. J., Torruella, G., Moreira, D., Eglit, Y., Simpson, A. G. B., Volcker, E., ... Lopez-Garcia, P. (2019). Combined cultivation and single-cell approaches to the phylogenomics of nucleariid amoebae, close relatives of fungi. *Philosophical Transactions of the Royal Society B: Biological Sciences* , *374* (1786), 20190094. doi: 10.1098/rstb.2019.0094

Gerth, M., Wolf, R., Bleidorn, C., Richter, J., Sontowski, R., Unrein, J., ... Gruppe, A. (2017). Green lacewings (Neuroptera: Chrysopidae) are commonly associated with a diversity of rickettsial endosymbionts. *Zoological Letters* , *3* (1), 12. doi: 10.1186/s40851-017-0072-9

Giorgini, M., Bernardo, U., Monti, M. M., Nappo, A. G., & Gebiola, M. (2010). *Rickettsia* symbionts cause parthenogenetic reproduction in the parasitoid wasp *Pnigalio soemius* (Hymenoptera: Eulophidae). *Applied and Environmental Microbiology* ,*76* (8), 2589–2599. doi: 10.1128/AEM.03154-09

Gonella, E., Pajoro, M., Marzorati, M., Crotti, E., Mandrioli, M., Pontini, M., ... Alma, A. (2015). Plant-mediated interspecific horizontal transmission of an intracellular symbiont in insects.*Scientific Reports* , *5* (1), 15811. doi: 10.1038/srep15811

Goodacre, S. L., Martin, O. Y., Bonte, D., Hutchings, L., Woolley, C., Ibrahim, K., ... Hewitt, G. M. (2009). Microbial modification of host long-distance dispersal capacity. *BMC Biology* , *7* , 32. doi: 10.1186/1741-7007-7-32

Goodacre, S. L., Martin, O. Y., Thomas, C. F. G., & Hewitt, G. M. (2006). *Wolbachia* and other endosymbiont infections in spiders.*Molecular Ecology* , *15* (2), 517–527. doi: 10.1111/j.1365-294X.2005.02802.x

Gruber-Vodicka, H. R., Seah, B. K., & Pruesse, E. (2019). phyloFlash — Rapid SSU rRNA profiling and targeted assembly from metagenomes.*BioRxiv* . doi: 10.1101/521922

Gualtieri, L., Nugnes, F., Nappo, A. G., Gebiola, M., & Bernardo, U. (2017). Life inside a gall: closeness does not favour horizontal transmission of *Rickettsia* between a gall wasp and its parasitoid. *FEMS Microbiology Ecology* , *93* (7). doi: 10.1093/femsec/fix087

Hagimori, T., Abe, Y., Date, S., & Miura, K. (2006). The first finding of a *Rickettsia* bacterium associated with parthenogenesis induction among insects. *Current Microbiology* , *52* (2), 97–101. doi: 10.1007/s00284-005-0092-0

He, Z., Zhang, H., Gao, S., Lercher, M. J., Chen, W.-H., & Hu, S. (2016). Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Research* , *44* , W236–W241. doi: 10.1093/nar/gkw370

Hendry, T. A., Hunter, M. S., & Baltrus, D. A. (2014). The facultative symbiont *Rickettsia* protects an invasive whitefly against entomopathogenic *Pseudomonas syringae* strains. *Applied and Environmental Microbiology* , *80* (23), 7161–7168. doi: 10.1128/AEM.02447-14

Hernandez-Triana, L. M., Prosser, S. W., Rodriguez-Perez, M. A., Chaverri, L. G., Hebert, P. D. N., & Ryan Gregory, T. (2014). Recovery of DNA barcodes from blackfly museum specimens (Diptera: Simuliidae) using primer sets that target a variety of sequence lengths.*Molecular Ecology Resources* , *14* (3), 508–518. doi: 10.1111/1755-0998.12208

Hilgenboecker, K., Hammerstein, P., Schlattmann, P., Telschow, A., & Werren, J. H. (2008). How many species are infected with*Wolbachia* ? A statistical analysis of current data. *FEMS Microbiology Letters* , *281* (2), 215–220. doi: 10.1111/j.1574-6968.2008.01110.x

Himler, A. G., Adachi-Hagimori, T., Bergen, J. E., Kozuch, A., Kelly, S. E., Tabashnik, B. E., ... Hunter, M. S. (2011). Rapid spread of a bacterial symbiont in an invasive whitefly is driven by fitness benefits and

female bias. *Science* , *332* (6026), 254–256. doi: 10.1126/science.1199410

Hoang, D. T., Chernomor, O., Haeseler, A. von, Minh, B. Q., & Vinh, L. S. (2017). UFBoot2: Improving the ultrafast bootstrap approximation.*Molecular Biology and Evolution* , *35* (2), 518–522. doi: 10.1101/153916

Hoffmann, A. A., Ross, P. A., & Rašić, G. (2015). *Wolbachia*strains for disease control: Ecological and evolutionary considerations.*Evolutionary Applications* , *8* (8), 751–768. doi: 10.1111/eva.12286

Hurst, G. D. D. (2017). Extended genomes: Symbiosis and evolution.*Interface Focus* , *7* (5), 20170001. doi: 10.1098/rsfs.2017.0001

Hurst, G. D. D., & Frost, C. L. (2015). Reproductive parasitism: Maternally inherited symbionts in a biparental world. *Cold Spring Harbor Perspectives in Biology* , *7* (5), a017699. doi: 10.1101/cshperspect.a017699

Hurst, G. D. D., Walker, L. E., & Majerus, M. E. N. (1996). Bacterial infections of hemocytes associated with the maternally inherited male-killing trait in British populations of the two spot ladybird,*Adalia bipunctata* . *Journal of Invertebrate Pathology* ,*68* (3), 286–292. doi: 10.1006/jipa.1996.0098

Jaenike, J. (2012). Population genetics of beneficial heritable symbionts. *Trends in Ecology & Evolution* , *27* (4), 226–232. doi: 10.1016/j.tree.2011.10.005

Jaenike, J., Polak, M., Fiskin, A., Helou, M., & Minhas, M. (2007). Interspecific transmission of endosymbiotic Spiroplasma by mites.*Biology Letters* , *3* (1), 23–25. doi: 10.1098/rsbl.2006.0577

Jaenike, J., Unckless, R., Cockburn, S. N., Boelio, L. M., & Perlman, S. J. (2010). Adaptation via symbiosis: Recent spread of a*Drosophila* defensive symbiont. *Science* , *329* (5988), 212–215. doi: 10.1126/science.1188235

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods* , *14* (6), 587–589. doi: 10.1038/nmeth.4285

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability.*Molecular Biology and Evolution* , *30* (4), 772–780. doi: 10.1093/molbev/mst010

Kikuchi, Y., & Fukatsu, T. (2005). *Rickettsia* infection in natural leech populations. *Microbial Ecology* , *49* (2), 265–271. doi: 10.1007/s00248-004-0140-5

Kikuchi, Yoshitomo, Sameshima, S., Kitade, O., Kojima, J., & Fukatsu, T. (2002). Novel clade of *Rickettsia* spp. from leeches.*Applied and Environmental Microbiology* , *68* (2), 999–1004. doi: 10.1128/AEM.68.2.999-1004.2002

Kliot, A., Cilia, M., Czosnek, H., & Ghanim, M. (2014). Implication of the bacterial endosymbiont *Rickettsia* spp. in interactions of the whitefly *Bemisia tabaci* with tomato yellow leaf curl virus.*Journal of Virology* , *88* (10), 5652–5660. doi: 10.1128/JVI.00071-14

Kobiałka, M., Michalik, A., Świerczewski, D., & Szklarzewicz, T. (2020). Complex symbiotic systems of two treehopper species:*Centrotus cornutus* (Linnaeus, 1758) and *Gargara genistae*(Fabricius, 1775) (Hemiptera: Cicadomorpha: Membracoidea: Membracidae).*Protoplasma* , *257* (3), 819–831. doi: 10.1007/s00709-019-01466-z

Kölsch, G., & Synefiaridou, D. (2012). Shared ancestry of symbionts? Sagrinae and Donaciinae (Coleoptera, Chrysomelidae) harbor similar bacteria. *Insects* , *3* (2), 473–491. doi: 10.3390/insects3020473

Küchler, S. M., Kehl, S., & Dettner, K. (2009). Characterization and localization of *Rickettsia* sp. in water beetles of genus*Deronectes* (Coleoptera: Dytiscidae). *FEMS Microbiology Ecology* , *68* (2), 201–211. doi: 10.1111/j.1574-6941.2009.00665.x

Lagrue, C., Joannes, A., Poulin, R., & Blasco-Costa, I. (2016). Genetic structure and host-parasite co-divergence: evidence for trait-specific local adaptation. *Biological Journal of the Linnean Society* ,*118* (2), 344–358. doi: 10.1111/bij.12722

Le Clec'h, W., Chevalier, F. D., Genty, L., Bertaux, J., Bouchon, D., & Sicard, M. (2013). Cannibalism and predation as paths for horizontal passage of *Wolbachia* between terrestrial isopods. *PLoS ONE* , *8* (4), e60232. doi: 10.1371/journal.pone.0060232

Li, Kaili, Chen, H., Jiang, J., Li, X., Xu, J., & Ma, Y. (2016). Diversity of bacteriome associated with *Phlebotomus chinensis*(Diptera: Psychodidae) sand flies in two wild populations from China.*Scientific Reports* , *6* (1), 36406. doi: 10.1038/srep36406

Li, Kun, Stanojević, M., Stamenković, G., Ilić, B., Paunović, M., Lu, M., ... Zhang, Y. (2019). Insight into diversity of bacteria belonging to the order Rickettsiales in 9 arthropods species collected in Serbia. *Scientific Reports* , *9* (*1*), 18680. doi: 10.1038/s41598-019-55077-y

Li, S.-J., Ahmed, M. Z., Lv, N., Shi, P.-Q., Wang, X.-M., Huang, J.-L., & Qiu, B.-L. (2017). Plantmediated horizontal transmission of*Wolbachia* between whiteflies. *The ISME Journal* ,*11* (4), 1019–1028. doi: 10.1038/ismej.2016.164

Li, Y.-H., Ahmed, M. Z., Li, S.-J., Lv, N., Shi, P.-Q., Chen, X.-S., & Qiu, B.-L. (2017). Plant-mediated horizontal transmission of*Rickettsia* endosymbiont between different whitefly species.*FEMS Microbiology Ecology* , *93* (12). doi: 10.1093/femsec/fix138

Łukasik, P., Guo, H., van Asch, M., Ferrari, J., & Godfray, H. C. J. (2013). Protection against a fungal pathogen conferred by the aphid facultative endosymbionts *Rickettsia* and *Spiroplasma* is expressed in multiple host genotypes and species and is not influenced by co-infection with another symbiont. *Journal of Evolutionary Biology* , *26* (12), 2654–2661. doi: 10.1111/jeb.12260

Machtelinckx, T., Van Leeuwen, T., Van De Wiele, T., Boon, N., De Vos, W. H., Sanchez, J.-A., ... De Clercq, P. (2012). Microbial community of predatory bugs of the genus *Macrolophus* (Hemiptera: Miridae). *BMC Microbiology* , *12* (Suppl 1), S9. doi: 10.1186/1471-2180-12-S1-S9

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution* , *1* (1), 1–5. doi: 10.1093/ve/vev003

Martin, O. Y., Puniamoorthy, N., Gubler, A., Wimmer, C., & Bernasconi, M. V. (2013). Infections with *Wolbachia* , *Spiroplasma* , and*Rickettsia* in the Dolichopodidae and other Empidoidea.*Infection, Genetics and Evolution* , *13* , 317–330. doi: 10.1016/j.meegid.2012.11.005

McFall-Ngai, M., Hadfield, M. G., Bosch, T. C. G., Carey, H. V., Domazet-Lošo, T., Douglas, A. E., ... Wernegreen, J. J. (2013). Animals in a bacterial world, a new imperative for the life sciences.*Proceedings of the National Academy of Sciences* , *110* (9), 3229–3236. doi: 10.1073/pnas.1218525110

Moreira, L. A., Iturbe-Ormaetxe, I., Jeffery, J. A., Lu, G., Pyke, A. T., Hedges, L. M., ... O'Neill, S. L. (2009). A *Wolbachia*symbiont in *Aedes aegypti* limits infection with dengue, chikungunya, and plasmodium. *Cell* , *139* (7), 1268–1278. doi: 10.1016/j.cell.2009.11.042

Morrow, J. L., Frommer, M., Shearman, D. C. A., & Riegler, M. (2014). Tropical tephritid fruit fly community with high incidence of shared *Wolbachia* strains as platform for horizontal transmission of endosymbionts. *Environmental Microbiology* , *16* (12), 3622–3637. doi: 10.1111/1462-2920.12382

Munson, M. A., Baumann, P., & Kinsey, M. G. (1991). *Buchnera*gen. nov. and *Buchnera aphidicola* sp. nov., a taxon consisting of the mycetocyte-associated, primary endosymbionts of aphids.*International Journal of Systematic Bacteriology* , *41* (4), 566–568. doi: 10.1099/00207713-41-4-566

Murakami, T., Segawa, T., Bodington, D., Dial, R., Takeuchi, N., Kohshima, S., & Hongoh, Y. (2015).

Census of bacterial microbiota associated with the glacier ice worm *Mesenchytraeus solifugus* .*FEMS Microbiology Ecology* , *91* (3). doi: 10.1093/femsec/fiv003

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* ,*32* (1), 268–274. doi: 10.1093/molbev/msu300

Noda, H., Watanabe, K., Kawai, S., Yukuhiro, F., Miyoshi, T., Tomizawa, M., ... Fukatsu, T. (2012). Bacteriome-associated endosymbionts of the green rice leafhopper *Nephotettix cincticeps* (Hemiptera: Cicadellidae). *Applied Entomology and Zoology* , *47* (3), 217–225. doi: 10.1007/s13355-012-0110-1

Okonechnikov, K., Golosova, O., & Fursov, M. (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* , *28* (8), 1166–1167. doi: 10.1093/bioinformatics/bts091

Park, E., & Poulin, R. (2020). Torix group *Rickettsia* are widespread in New Zealand freshwater amphipods: using blocking primers to rescue host *COI* sequences. *BioRxiv* . doi: https://doi.org/10.1101/2020.05.28.120196

Parola, P., Paddock, C. D., & Raoult, D. (2005). Tick-borne rickettsioses around the world: Emerging diseases challenging old concepts. *Clinical Microbiology Reviews* , *18* (4), 719–756. doi: 10.1128/CMR.18.4.719-756.2005

Perlman, S. J., Hunter, M. S., & Zchori-Fein, E. (2006). The emerging diversity of *Rickettsia* . *Proceedings of the Royal Society B* , *273* (1598), 2097–2106. doi: 10.1098/rspb.2006.3541

Perotti, M. A., Clarke, H. K., Turner, B. D., & Braig, H. R. (2006).*Rickettsia* as obligate and mycetomic bacteria. *The FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology* , *20* (13), 2372–2374. doi: 10.1096/fj.06-5870fje

Pilgrim, J., Ander, M., Garros, C., Baylis, M., Hurst, G. D. D., & Siozios, S. (2017). Torix group *Rickettsia* are widespread in*Culicoides* biting midges (Diptera: Ceratopogonidae), reach high frequency and carry unique genomic features. *Environmental Microbiology* , *19* (10), 4238–4255. doi: 10.1111/1462-2920.13887

Plantard, O., Bouju-Albert, A., Malard, M. A., Hermouet, A., Capron, G., & Verheyden, H. (2012). Detection of *Wolbachia* in the tick*Ixodes ricinus* is due to the presence of the hymenoptera endoparasitoid *Ixodiphagus hookeri* . *PLoS ONE* . doi: 10.1371/journal.pone.0030692

Põldmaa, K., Jürgenstein, S., Bahram, M., Teder, T., & Kurina, O. (2015). Host diversity and trophic status as determinants of species richness and community composition of fungus gnats. *Basic and Applied Ecology* , *16* (1), 46–53. doi: 10.1016/j.baae.2014.10.004

Potts, R., Molina, I., Sheele, J. M., & Pietri, J. E. (2020). Molecular detection of *Rickettsia* infection in field-collected bed bugs.*New Microbes and New Infections* , *34* , 100646. doi: 10.1016/j.nmni.2019.100646

Ramage, T., Martins-Simoes, P., Mialdea, G., Allemand, R., Duplouy, A., Rousse, P., ... Charlat, S. (2017). A DNA barcode-based survey of terrestrial arthropods in the Society Islands of French Polynesia: host diversity within the SymbioCode Project. *European Journal of Taxonomy* , (272). doi: 10.5852/ejt.2017.272

Rangel López, J. Á., Husemann, M., Schmitt, T., Kramp, K., & Habel, J. C. (2018). Mountain barriers and trans-Saharan connections shape the genetic structure of *Pimelia* darkling beetles (Coleoptera: Tenebrionidae). *Biological Journal of the Linnean Society* ,*124* (3), 547–556. doi: 10.1093/biolinnean/bly053

Ratnasingham, S., & Hebert, P. D. N. (2007). BOLD: The Barcode of Life Data System. *Molecular Ecology Notes* , *7* (3), 355–364. doi: 10.1111/j.1471-8286.2007.01678.x

Reeves, W. K., Kato, C. Y., & Gilchriest, T. (2008). Pathogen screening and bionomics of *Lutzomyia apache* (Diptera: Psychodidae) in Wyoming, USA. *Journal of the American Mosquito Control Association* , *24* (3), 444–447. doi: 10.2987/5745.1

Řezáč, M., Gasparo, F., Král, J., & Heneberg, P. (2014). Integrative taxonomy and evolutionary history of a newly revealed spider*Dysdera ninnii* complex (Araneae: Dysderidae). *Zoological Journal of the Linnean Society* , *172* (2), 451–474. doi: 10.1111/zoj.12177

Ricketts, H. T. (1909). A micro-organism which apparently has a specific relationship to Rocky Mountain spotted fever. *Journal of the American Medical Association* , *52* , 379–380. doi: 10.1001/jama.1909.02540320063015

Rio, R. V. M., Attardo, G. M., & Weiss, B. L. (2016). Grandeur alliances: Symbiont metabolic integration and obligate arthropod hematophagy. *Trends in Parasitology* , *32* (9), 739–749. doi: 10.1016/j.pt.2016.05.002

Siozios, S., Gerth, M., Griffin, J. S., & Hurst, G. D. D. (2018). Symbiosis: *Wolbachia* host shifts in the fast lane. *Current Biology* , *28* (6), R269–R271. doi: 10.1016/j.cub.2018.02.008

Smith, M. A., Bertrand, C., Crosby, K., Eveleigh, E. S., Fernandez-Triana, J., Fisher, B. L., ... Zhou, X. (2012). *Wolbachia* and DNA barcoding insects: Patterns, potential, and problems. *PLoS ONE* , *7* (5), e36514. doi: 10.1371/journal.pone.0036514

Song, S., Chen, C., Yang, M., Zhao, S., Wang, B., Hornok, S., ... Wang, Y. (2018). Diversity of *Rickettsia* species in border regions of northwestern China. *Parasites & Vectors* ,*11* (1), 634. doi: 10.1186/s13071-018-3233-6

Tedeschi, R., Visentin, C., Alam, A., & Bosco, D. (2003). Epidemiology of apple proliferation (AP) in northwestern Italy: evaluation of the frequency of AP-positive psyllids in naturally infected populations of*Cacopsylla melanoneura* (Homoptera: Psyllidae). *Annals of Applied Biology* , *142* (3), 285–290. doi: 10.1111/j.1744-7348.2003.tb00252.x

Teixeira, L., Ferreira, A., & Ashburner, M. (2008). The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in*Drosophila melanogaster* . *PLoS Biology* , *6* (12), 2753–2763. doi: 10.1371/journal.pbio.1000002

Thongprem, P., Davison, H., Ferreira, S., Cabello, O. L., & Hurst, G. D. D. (2020). Incidence and diversity of Torix *Rickettsia* – Odonate symbioses. *Manuscript Submitted for Publication* .

Turelli, M., Cooper, B. S., Richardson, K. M., Ginsberg, P. S., Peckenpaugh, B., Antelope, C. X., ... Hoffmann, A. A. (2018). Rapid global spread of wRi-like *Wolbachia* across multiple*Drosophila* . *Current Biology* , *28* (6), 963-971.e8. doi: 10.1016/j.cub.2018.02.015

van den Hurk, A. F., Hall-Mendelin, S., Pyke, A. T., Frentiu, F. D., McElroy, K., Day, A., ... O'Neill, S. L. (2012). Impact of*Wolbachia* on infection with chikungunya and yellow fever viruses in the mosquito vector *Aedes aegypti*. *PLoS Neglected Tropical Diseases* , *6* (11). doi: 10.1371/journal.pntd.0001892

Vavre, F., Fleury, F., Lepetit, D., Fouillet, P., & Bouletreau, M. (1999). Phylogenetic evidence for horizontal transmission of*Wolbachia* in host-parasitoid associations. *Molecular Biology and Evolution* , *16* (12), 1711–1723. doi: 10.1093/oxfordjournals.molbev.a026084

Walker, T., Johnson, P. H., Moreira, L. a, Iturbe-Ormaetxe, I., Frentiu, F. D., McMeniman, C. J., ... Hoffmann, A. A. (2011). The wMel *Wolbachia* strain blocks dengue and invades caged *Aedes aegypti* populations. *Nature* , *476* (7361), 450–453. doi: 10.1038/nature10355

Wang, H., Lei, T., Wang, X., Maruthi, M. N., Zhu, D., Cameron, S. L., ... Liu, S. (2020). A newly recorded *Rickettsia* of the Torix group is a recent intruder and an endosymbiont in the whitefly*Bemisia tabaci* . *Environmental Microbiology* , *22* (4), 1207–1221. doi: 10.1111/1462-2920.14927

Wang, Z., & Wu, M. (2015). An integrated phylogenomic approach toward pinpointing the origin of mitochondria. *Scientific Reports* ,*5* (1), 7949. doi: 10.1038/srep07949

Watanabe, K., Yukuhiro, F., Matsuura, Y., Fukatsu, T., & Noda, H. (2014). Intrasperm vertical symbiont transmission. *Proceedings of the National Academy of Sciences* , *111* (20), 7433–7437. doi:

10.1073/pnas.1402476111

Weinert, L. A. (2015). The diversity and phylogeny of *Rickettsia* . In S. Morand, B. R. Krasnov, & D. T. J. Littlewood (Eds.),*Parasite Diversity and Diversification* (pp. 150–181). doi: 10.1017/CBO9781139794749.012

Weinert, L. A., Araujo-Jnr, E. V, Ahmed, M. Z., & Welch, J. J. (2015). The incidence of bacterial endosymbionts in terrestrial arthropods.*Proceedings of the Royal Society B: Biological Sciences* ,*282* (1807), 20150249. doi: 10.1098/rspb.2015.0249

Weinert, L. A., Werren, J. H., Aebi, A., Stone, G. N., & Jiggins, F. M. (2009). Evolution and diversity of *Rickettsia* bacteria. *BMC Biology* , *7* , 6. doi: 10.1186/1741-7007-7-6

Werren, J. H., Hurst, G. D. D., Zhang, W., Breeuwer, J. A., Stouthamer, R., & Majerus, M. E. (1994). Rickettsial relative associated with male killing in the ladybird beetle (*Adalia bipunctata*). *Journal of Bacteriology* , *176* (2), 388–394. doi: 10.1128/jb.176.2.388-394.1994

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology* , *20* (1), 257. doi: 10.1186/s13059-019-1891-0

Yan, P., Qiu, Z., Zhang, T., Li, Y., Wang, W., Li, M., ... Liu, J. (2019). Microbial diversity in the tick *Argas japonicus* (Acari: Argasidae) with a focus on *Rickettsia* pathogens. *Medical and Veterinary Entomology* , *33* (3), 327–335. doi: 10.1111/mve.12373

Zheng, Z., Wang, D., He, H., & Wei, C. (2017). Bacterial diversity of bacteriomes and organs of reproductive, digestive and excretory systems in two cicada species (Hemiptera: Cicadidae). *PLOS ONE* ,*12* (4), e0175903. doi: 10.1371/journal.pone.0175903

Zakharov, E. V., Ratnasingham, S., deWaard, J. R., Smith, M. A. (2020). BOLD non-contaminant dataset; Barcoding Of Life Data System; http://dx.doi.org/10.5883/DS-RICKET [dataset]

Zouache, K., Voronin, D., Tran-Van, V., & Mavingui, P. (2009). Composition of bacterial communities associated with natural and laboratory populations of *Asobara tabida* Infected with *Wolbachia* . *Applied and Environmental Microbiology* ,*75(11)* , 3755–3764. doi: 10.1128/AEM.02964-08

Zug, R., & Hammerstein, P. (2012). Still a host of hosts for *Wolbachia* : Analysis of recent data suggests that 40% of terrestrial arthropod species are infected. *PLoS ONE* ,*7* (6), e38544. doi: 10.1371/journal.pone.0038544

Data Accessibility Statement

The data that support the findings of this study are openly available in BOLD at http://dx.doi.org/10.5883/DS-RICKET. For DNA sequences, EMBL accession codes are: LR798809-LR800243; LR812141-LR812260; LR812269-LR812283; LR812678; LR813674-LR813676; LR813730

Author contributions

JP, GH, MB and MAS assisted in the conception and design of the study. MAS, EZ, SR and JD assisted in assembling BOLD datasets and providing DNA extracts for laboratory experiments. Field and Laboratory work was undertaken by JP and PT. SRA work was undertaken by HD and SS. Analysis and interpretation of the data were undertaken by JP, PT, HD, GH, MB and SS, as well as drafting of the manuscript. All authors assisted in critical revision of the manuscript.

Figure and Table Legends

Figure 1. Workflow of the BOLD project demonstrating the acquisition and fates of contaminant and non-contaminant *COI*barcoding sequences.

Figure 2. Cladogram of the maximum likelihood (ML) tree of 1,126 proteobacteria *COI* contaminants retrieved from a BOLD project incorporating 184,585 arthropod specimens. The tree is based on 561 bp and is rooted with the free-living alphaproteobacteria*Pelagibacter ubique* . Parantheses indicate the number

of BOLD contaminants present in each group. Tips are labelled by BOLD processing ID and host arthropod taxonomy. No colour=Non-BOLD Reference. The Rickettsiales sequences of *Anaplasma* , *Neorickettsia* ,*Rickettsia* and *Wolbachia* supergroups (A, B, E, F and H) are included as references (Accession numbers: Table S1).

Figure 3. Cladogram of a maximum likelihood (ML) tree of 753 *COI Rickettsia* contaminants retrieved from a BOLD project incorporating 184,585 arthropod specimens. The tree is based on 561 bp and is rooted by the *Rickettsia* endosymbiont of *Ichthyophthirius multifiliis* (Candidatus Megaira) using the TVM+F+I+G4 model. Parantheses indicate the number of BOLD contaminants present in Torix and non-Torix *Rickettsia* groups. Tips are labelled by BOLD processing ID and host arthropod taxonomy. No colour=Non-BOLD reference sequence unless designated by a circle (Dermaptera), star (Diplopoda), triangle (Thysanoptera). The *Rickettsia* groups: Spotted fever, Transitional, Belli, Typhus, Rhyzobius and Torix are included as references (Accession numbers: Table S1).

Figure 4. Phylogram of the maximum likelihood (ML) tree of 99 *COI Rickettsia* contaminants (prefix "BIOUG") used for further phylogenetic analysis and 53 Non-BOLD reference profiles (Accession numbers: Table S1). The tree is based on the concatenation of 4 loci; *16S rRNA* , *17KDa* , *gltA* and *COI* under a partition model, with profiles containing at least 3 out of 4 sites included in the tree (2,834 bp total) and is rooted by *Rickettsia* endosymbiont of *Ichthyophthirius multifiliis* (Candidatus Megaira). Tips are labelled by host arthropod taxonomy.

Figure 5. *16S rRNA* and *gltA* concatenated maximum likelihood (ML) phylogram (1,834 bp total) including *Rickettsia* hosts from SRA (Triangles) and targeted screens (Stars). The TIM3+F+R2 (16S) and K3Pu+F+G4 (gltA) models were chosen as best fitting models. Rooting is with *Orientia tsutsugamushi* . Accession numbers found in Table S1.

Figure 6. Phylogram of a maximum likelihood (ML) tree of *COI Rickettsia* contaminants (prefix "BIOUG") giving a host barcode and 43 Non-BOLD reference profiles. The tree is based on 4 loci; *16S rRNA* , *17KDa* , *gltA* and *COI* under a partition model with profiles containing at least 2 out of 4 sites included in the tree (2,781 bp total) and is rooted by the *Rickettsia* endosymbiont of *Ichthyophthirius multifiliis* (Candidatus Megaira). The habitats and lifestyles of the host are given to the right of the phylogeny. Accession numbers found in Table S1.

Figure S1. Collection sites of the 753 *COIRickettsia* contaminants retrieved from BOLD projects.

Figure S2. Phylogram of a maximum likelihood (ML) tree of *COI Rickettsia* found in the NCBI database erroneously identified as mtDNA barcodes based on 577 bp. The HKY+F+G4 model was chosen as the best fitting model using Modelfinder with the Bayesian information criterion (BIC) (Kalyaanamoorthy et al., 2017).

Table 1.1 Targeted *Rickettsia* screen of aquatic invertebrates. A species was deemed positive through PCR and designated to *Rickettsia* group after Sanger sequencing and phylogenetic placement (Figure 5). All strains belong to the Torix group.

Table 1.2 . Targeted *Rickettsia* screen of terrestrial invertebrates. A species was deemed positive through PCR and designated to *Rickettsia* group after Sanger sequencing and phylogenetic placement (Figure 5). All strains belong to the Torix group except +=Rhyzobius and ++=Belli.

Table 2. Torix *Rickettsia* hosts known to date alongside screening method. Bold entries indicate hosts identified in this study. FISH=fluoresence *in-situ* hybridisation; TEM=transmission electron microscopy; SRA=sequence read archive.

Table S1. Accession numbers used for phylogenetic analyses (Figures 2, 3, 4 ,5 and 6). Accession numbers generated in this study are marked in BOLD.

Table S2. Mitochondrial *COI* and bacterial gene primers used for re-barcoding and multilocus phylogenetic analysis.

Table S3. List of SRA datasets analysed with phyloFlash and Kraken2.

Table S4. BOLD contaminant datasets

Table S5. Primer pairs involved in the unintended amplification of 753 *Rickettsia COI* from BOLD project.

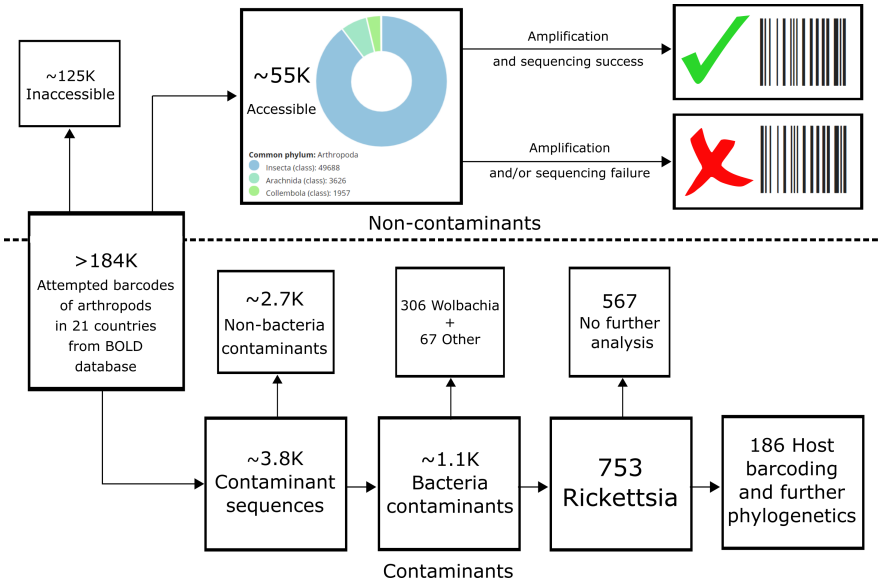
Table S6.1. Homology of *Rickettsia* groups and *Wolbachia* to the most common forward primer (C_-LepFolF) attributed to bacterial *COI* amplification from arthropod DNA extracts.

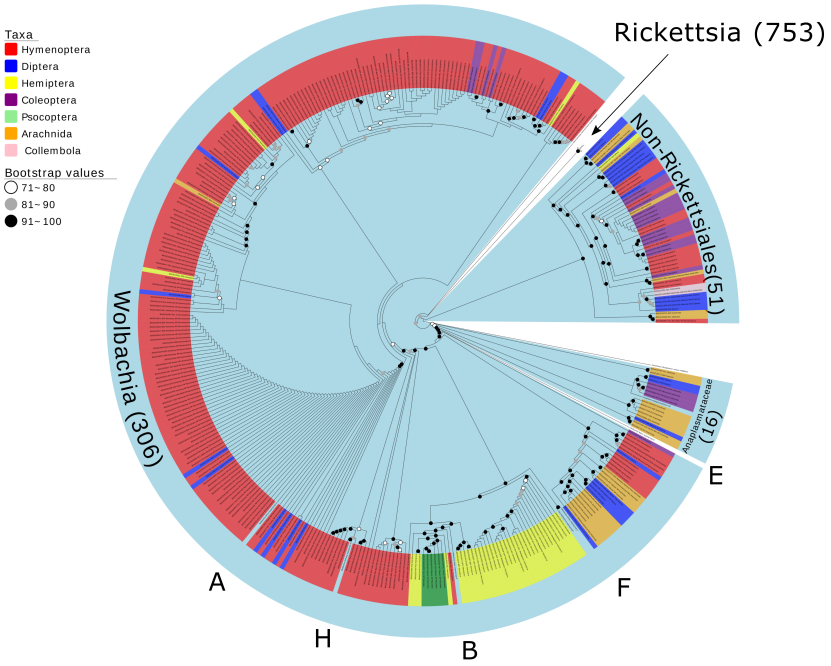
Table S6.2. Homology of *Rickettsia* groups and *Wolbachia* to the most common forward reverse (C_-LepFolR) attributed to bacterial *COI* amplification from arthropod DNA extracts.

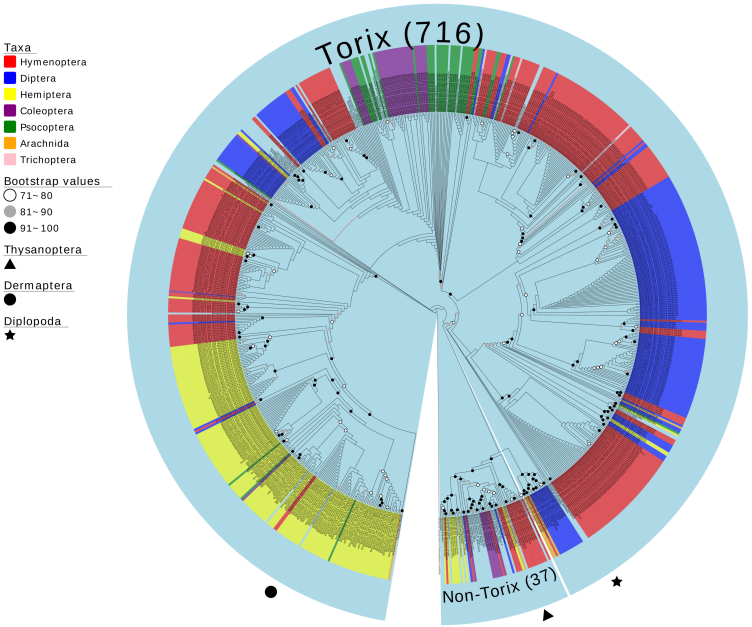
Table S7. Re-barcoding status and nearest BLAST hit (NCBI) of mtDNA *COI* arthropod DNA extracts accessed for further analysis, along with the success of multilocus *Rickettsia* profiles with allocated *Rickettsia* group (based on phylogenetic analysis) and co-infection status.

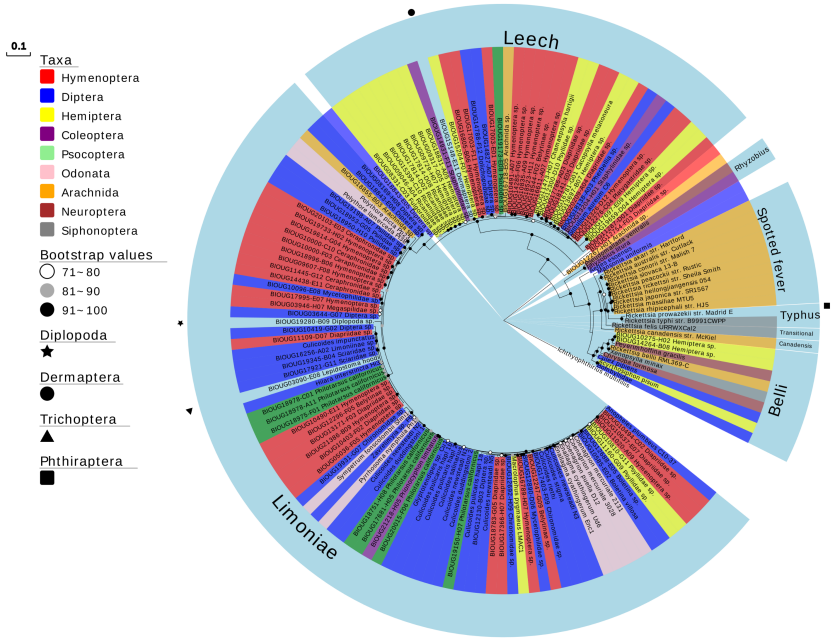
Table S8. The barcoding success rate of taxa which gave at least one bacteria *COI* inadvertent amplification (N=51,475 accessible specimens) with an adjusted *Rickettsia* prevalence based on an estimated total number of arthropods to account for inaccessible specimens (N=184,585).

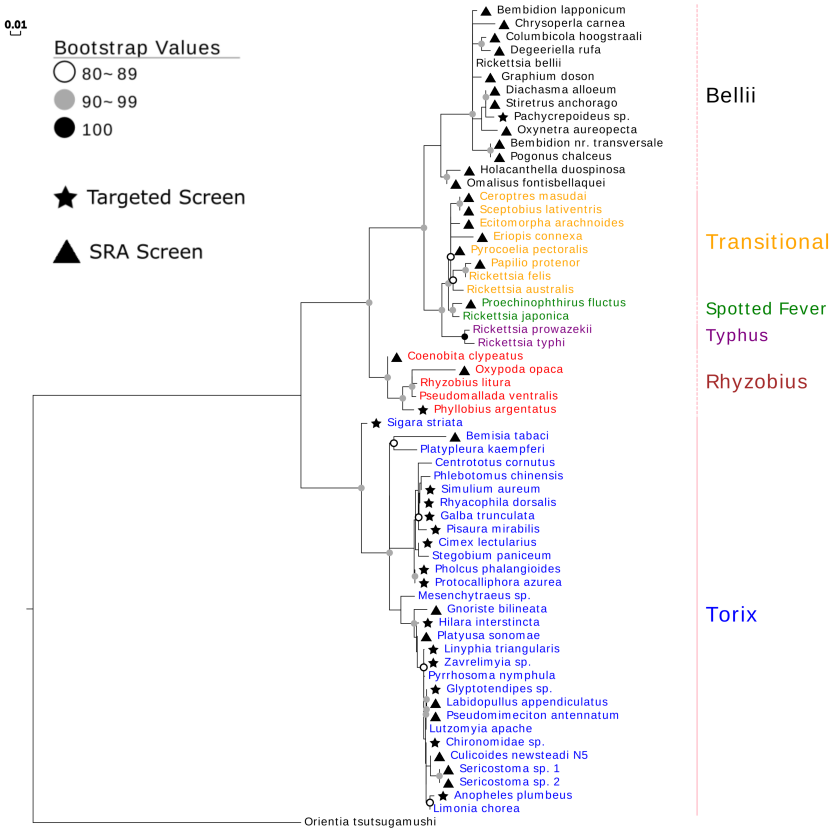
Table S9. NCBI matches mistaken for true mtDNA barcodes and their homology to *Rickettsia COI* (Accessed 29th June 2020).

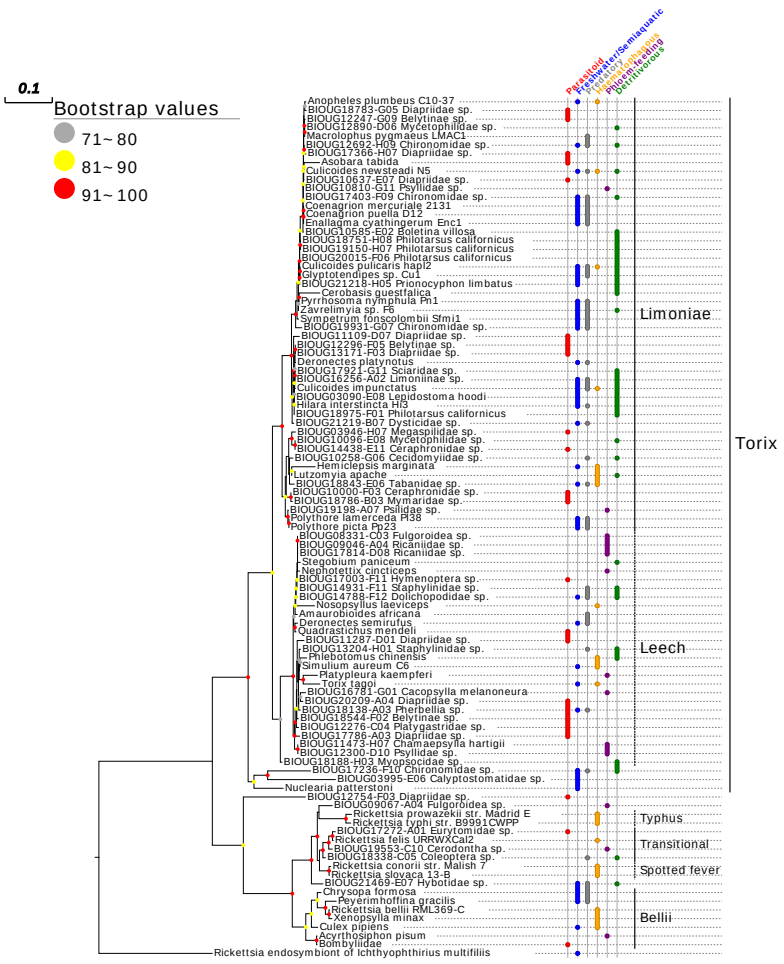












Hosted file

Table 1.docx available at https://authorea.com/users/344783/articles/471183-torix-rickettsia-are-widespread-in-arthropods-and-reflect-a-neglected-symbiosis

Hosted file

Table 2.docx available at https://authorea.com/users/344783/articles/471183-torix-rickettsia-are-widespread-in-arthropods-and-reflect-a-neglected-symbiosis