



OPEN

Widespread *Torix Rickettsia* in New Zealand amphipods and the use of blocking primers to rescue host COI sequences

Eunji Park & Robert Poulin

Endosymbionts and intracellular parasites are common in arthropod hosts. As a consequence, (co) amplification of untargeted bacterial sequences has been occasionally reported as a common problem in DNA barcoding. While identifying amphipod species with universal COI primers, we unexpectedly detected rickettsial endosymbionts belonging to the *Torix* group. To map the distribution and diversity of *Rickettsia* species among amphipod hosts, we conducted a nationwide molecular screening of seven families of New Zealand freshwater amphipods. In addition to uncovering a diversity of *Torix Rickettsia* species across multiple amphipod populations from three different families, our research indicates that: (1) detecting *Torix Rickettsia* with universal primers is not uncommon, (2) obtaining 'Rickettsia COI sequences' from many host individuals is highly likely when a population is infected, and (3) obtaining 'host COI' may not be possible with a conventional PCR if an individual is infected. Because *Rickettsia* COI is highly conserved across diverse host taxa, we were able to design blocking primers that can be used in a wide range of host species infected with *Torix Rickettsia*. We propose the use of blocking primers to circumvent problems caused by unwanted amplification of *Rickettsia* and to obtain targeted host COI sequences for DNA barcoding, population genetics, and phylogeographic studies.

The cytochrome c oxidase subunit 1 gene (COI), a partial fragment of mitochondrial DNA, is the marker of choice for DNA barcoding, and is also widely used for population genetics and phylogeographic studies^{1–3}. A variable region is flanked by highly conserved regions; this allowed for the design of a pair of universal primers and their application to various organisms^{4,5}. With the advancement of fast and cost-effective next-generation sequencing technologies, which enables metabarcoding^{6,7}, the number of COI sequences is increasing rapidly in public databases such as GenBank and The Barcode of Life DataSystems (BOLD)⁸. However, quality control is often an issue due to the presence of questionable “COI-like” sequences⁹ or nuclear mitochondrial pseudogenes (numts) that are often coamplified with orthologous mtDNA¹⁰. Bacterial sequences are also often coamplified with universal primers. Indeed, there have been reports of the amplification of untargeted sequences of endosymbiotic bacteria such as *Wolbachia* and *Aeromonas* during DNA barcoding with universal primers and their misidentification as those of invertebrate hosts during deposition in databases^{11,12}.

The bacterial genus *Rickettsia* is another of these endosymbiotic taxa. This genus belongs to the order Rickettsiales along with *Wolbachia*, and comprises diverse pathogenic species that can cause vector-borne diseases in birds and mammals including humans as well as non-pathogenic species. Some rickettsioses with severe symptoms are well known, and include Rocky Mountain spotted fever, Queensland tick typhus, rickettsial pox, murine typhus and epidemic typhus^{13–15}. To date, at least 13 groups are known within the genus *Rickettsia*: *Adalia*, *Bellii*, *Canadensis*, *Guiana*, *Helvetica*, *Meloidae*, *Mendelii*, *Rhyzobious*, *Spotted fever*, *Scapularis*, *Torix*, *Transitional*, and *Typhus*^{16–18}. All these groups except the *Torix* group are exclusively associated with arthropod hosts, such as mites, fleas, ticks, and spiders. The *Torix* group, which is sister to all other groups, is the only group that includes non-arthropod hosts such as amoeba and leeches^{19–21}. In addition to these freshwater hosts, the *Torix* group occurs in diverse arthropod groups that spend part of their life cycle in the aquatic environment (e.g. Coleoptera and Diptera)^{16,22,23}.

Department of Zoology, University of Otago, 340 Great King Street, Dunedin 9016, New Zealand. ✉email: eunjisea@gmail.com

Although *Rickettsia* species are known as common pathogens or endosymbionts in arthropod hosts, these agents have never been reported in crustaceans. *Rickettsia*-like organisms (RLO) have been reported in some groups of crustaceans including crabs, crayfish, lobsters, shrimps, and amphipods²⁴. However, most reports of these RLOs were based on morphological similarity with *Rickettsia* and were rarely confirmed by molecular data. In amphipods, RLOs were reported in several species of gammarids, as well as other taxa (e.g., *Crangonyx floridanus* and *Diporeia* sp.)^{25–28}. 16S rRNA (*rrs*) sequences of RLOs are available for *Diporeia* sp. and some gammarids, but none of them belong to the genus *Rickettsia*^{28,29}.

While identifying amphipod species with DNA barcoding for cophylogenetic analyses between microsporidians and their amphipod hosts³⁰, we obtained a suspicious COI sequence from *Paracalliope fluviatilis*, the most common freshwater amphipod species in New Zealand. According to a blast search in GenBank, this sequence obtained from a stream in the Southland region was ~99.5% identical to 12 sequences of the same ‘amphipod’ species available in GenBank from a previous study conducted in our lab³¹. These sequences were obtained from two different locations (Waikouaiti River and Waitaki River) in the Otago region. However, all sequences were highly divergent from that obtained from other populations (~57%) of the same host group (*Paracalliope* species complex). DNA from the individual amphipod was extracted from its legs (i.e., low chance of contamination due to gut contents). We obtained a clear chromatogram with no ambiguous peaks. Furthermore, these sequences were similar to other COI sequences obtained from diverse insects (Coleoptera, Diptera, Hemiptera, Hymenoptera, Odonata) in GenBank with sequence similarity ranging from 80 to 99%. However, these sequences were also similar (~92%) to sequences of rickettsial endosymbionts of insects and spiders that have been recently registered in GenBank. Because such highly conserved COI sequences among distantly related arthropod groups are unlikely, we assumed that these sequences were actually obtained from their rickettsial endosymbionts. We independently confirmed the presence of *Rickettsia* in our amphipod hosts using three genetic markers that were designed to be specific to *Rickettsia*.

In fact, this phenomenon of the amplification of non-targeted COI sequences with widely used DNA barcoding primers has already been reported. Řezáč et al. (2014)³² obtained a rickettsial COI sequence from a spider species in a study with a different purpose. Ceccarelli et al. (2016)³³ obtained COI sequences of *Torix Rickettsia* from six individuals of a spider species while conducting DNA barcoding, and these authors formally discussed the presence of misidentified COI sequences in GenBank. However, despite these early reports, the deposition of misidentified sequences to GenBank has continued until recently. A very recent survey on BOLD reported that 0.41% of the barcode submission in BOLD are actually from *Rickettsia*, which is higher than that from *Wolbachia* (0.17%)³⁴.

Because *Rickettsia* species are endosymbionts within host cells, DNA extracts from infected host tissue will inevitably include DNA of endosymbionts as well. If binding sites for ‘universal primers’ are conserved in both hosts and their endosymbionts, PCR products obtained from these mixed templates may result in mixed signals in chromatograms, or in the amplification of endosymbiont instead of host sequences³⁵. Using primers that are designed to bind uniquely and specifically to host templates would reduce this problem. However, designing group-specific primers is not always possible, especially when reference sequences are scarce or not available. Also, finding conserved regions across a given taxonomic group may not be achievable. Alternatively, blocking primers can be used to prevent the amplification of unwanted or dominant sequences among DNA templates³⁵. For example, this method has been successfully applied to identify prey items (by suppressing the amplification of predator DNA in gut contents), or to obtain rare mammal sequences from ancient DNA (by blocking the amplification of human DNA)^{36,37}. Because COI sequences of *Torix Rickettsia* are highly conserved in diverse host groups, we were able to design blocking primers that are intended to specifically block the amplification of *Torix Rickettsia* but allow amplification of the COI region of (any) host mtDNA.

In this study, we first screened rickettsial infections in diverse amphipods collected throughout New Zealand to determine their prevalence and distribution. Secondly, we characterized the genetic diversity of the newly found *Rickettsia* species in relation to other *Torix Rickettsia* using 4 distinct markers, namely *rrs*, *gltA*, *atpA*, and COI, and expand current understanding of *Rickettsia* phylogeny. Thirdly, we demonstrate that unwanted amplification of rickettsial COI sequences during DNA barcoding is a common problem, and that such sequences have been frequently reported and misidentified in GenBank. Fourthly, we suggest that using blocking primers in addition to universal primers for PCR is an effective solution to obtain targeted host COI sequences. Finally, we discuss the implications of these pseudo-sequences in public databases, ways to reduce this problem, and possible applications of blocking primers for similar problems.

Results

Distribution of *Rickettsia* in amphipod hosts in New Zealand. *Rickettsia* was detected in 26 of 69 locations (37.7%) from 3 families of amphipods: *Paracalliope* species complex (24/59 populations; 40.7%), *Paraleptamphopus* sp. (3/5 populations; 60%), and one undescribed family of Senticaudata sp. (1/1 population; 100%) (Fig. 1 and Supplementary Table 1). Because pooled samples were used, accurate prevalence in each population could not be obtained. However, a relative comparison was possible among the populations in which the same number of individuals per sample and the same total number was used (i.e. populations with a total of 48 individuals, with 12 samples each containing 4 individuals) (Supplementary Table 1). With parsimonious interpretation, among 18 populations, 12 populations showed at least 10% prevalence (> 5 positive tubes/12 tubes tested). Seven populations showed at least 20% prevalence (> 10 positive tubes/12 tubes tested). And three populations showed 100% positive tubes (12/12), with prevalence thus possibly ranging from 25–100%. Although *Rickettsia* was detected in both the North and South Islands, its distribution was confined to the southern parts of both islands (Fig. 1).

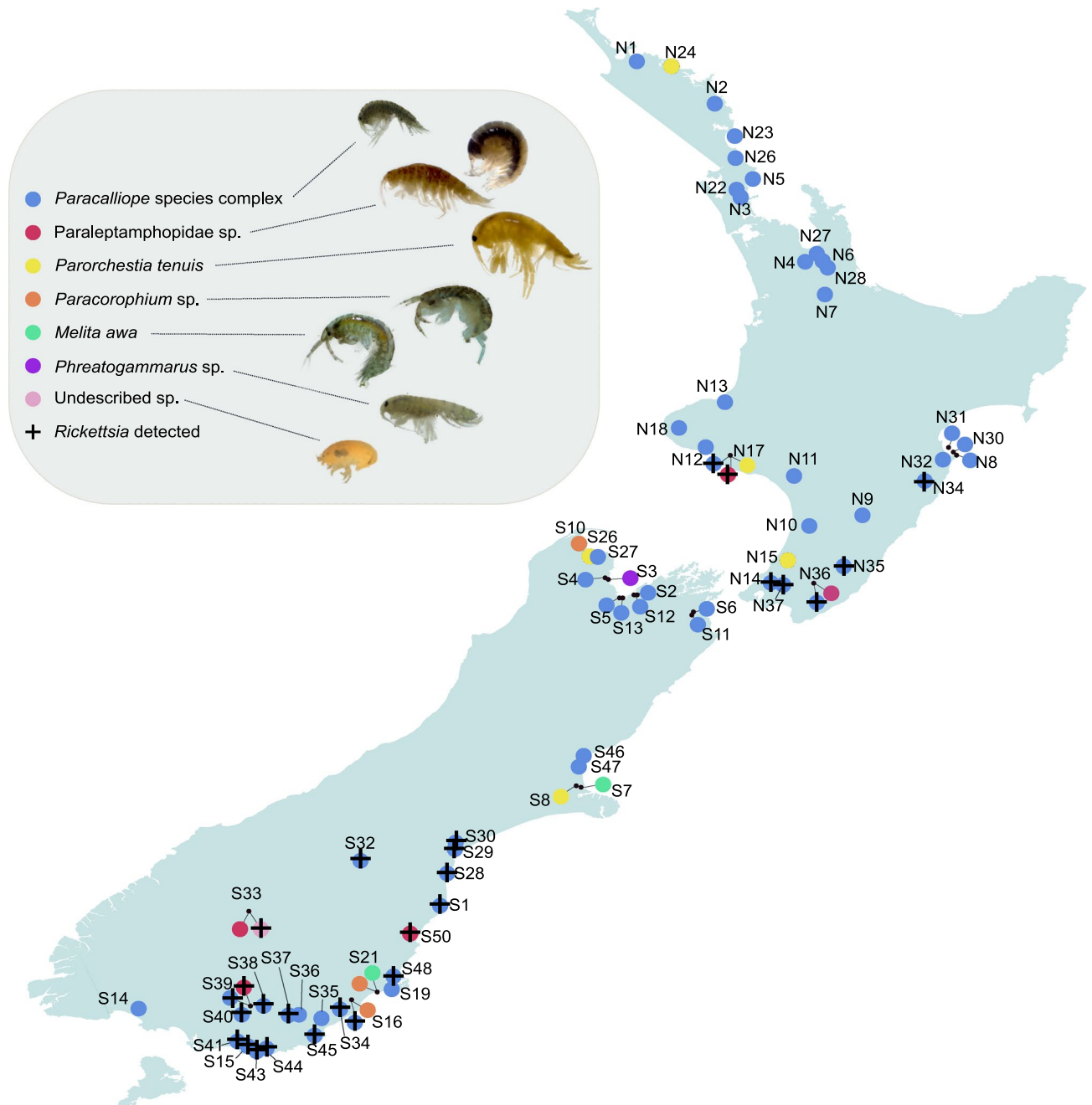


Figure 1. Map of sampling sites. Map of New Zealand showing the sixty-nine sampling sites with circles. Seven different families of amphipods are marked with different colours in the circles. The sites where *Rickettsia* was detected are marked with +. Site codes correspond to those in Supplementary Table 1.

Genetic characterization of *Rickettsia* sequences. At least one *rrs*, one *gltA*, or one *atpA* sequence was obtained from each of the population/species that were positive in the initial molecular screening (Supplementary Table 1). Specifically, 24 sequences of *rrs*, 14 sequences of *gltA*, and 19 sequences of *atpA* were obtained. Also, 8 sequences of COI were added to our dataset (Supplementary Table 2). Fourteen genotypes were identified using *rrs* sequences. All *rrs* showed higher similarity to each other (>99.4%). All *gltA*, *atpA*, and COI sequences of Torix *Rickettsia* from amphipods showed high similarity to each other: >95%, >94%, >95%, respectively.

Compiling molecular data on Torix *Rickettsia* from GenBank. A total of 183 nucleotide sequences of Torix *Rickettsia* were obtained from GenBank (Supplementary Tables 3–5). Specifically, 51 *rrs* sequences were available from Amoeba, Annelida, Arachnida, Coleoptera, Diptera, Hemiptera, Hymenoptera, Psocoptera, Megaloptera, and an environmental sample representing 18 studies. A total of 68 sequences of *gltA* were obtained from Arachnida, Coleoptera, Diptera, Hemiptera, Hymenoptera, and Siphonaptera representing 12 studies. A

total of 64 COI sequences were available from Amphipoda, Arachnida, Coleoptera, Diptera, Hemiptera, Hymenoptera, Megaloptera, and Odonata, representing 17 studies. Among these COI sequences, 42 sequences from 11 studies assigned rickettsial COI sequences to their invertebrate hosts. Since the very first misassignment in 2013, these mislabeled sequences have been deposited every year. Eight of these studies (representing 26 sequences) were involved with DNA barcoding and therefore voucher specimens (Supplementary Table 5).

Phylogeny of *Rickettsia*. All three trees inferred by *rrs*, *gltA*, and COI sequences clearly show two lineages within the genus *Rickettsia*: one clade of Torix *Rickettsia* and the other clade including all other 12 recognized groups within *Rickettsia* (Figs. 2, 3, 4). A Bayesian tree based on *rrs* sequences (Fig. 2) shows that all sequences obtained from New Zealand amphipods belong to the Torix group of *Rickettsia*. Even when the *rrs* conserved marker was used, all sequences obtained from amphipods (except S39_1542), were grouped in the same clade and distinct from other sequences, although this clade was not strongly supported (PP = 0.84). Several subgroups were identified in the *gltA* tree (Fig. 3). Most *Rickettsia* from amphipods were grouped within the same clade, similar to that revealed in the *rrs* tree. Also, a Bayesian tree inferred from COI sequences (Fig. 4) shows that *Rickettsia* from amphipods and some insects obtained in New Zealand are closely related, and the clade containing them is strongly supported (PP = 0.96).

Testing and validating blocking primers. The ratio of the amplicons of host COI to *Rickettsia* COI (based on their expected fragment sizes) was calculated to compare the efficiency of each primer under different conditions (Table 1). Not all fragment analyses were successful, but we were able to compare some effects. Except for Bloc_R, which always resulted in amplifying an excess of *Rickettsia* COI (ratio 0.06 ~ 0.57), all primers showed some blocking effects. Specifically, Bloc_F, Bloc_F2, and HCO_DPO primers showed increased blocking effects (higher ratio of hostCOI:RickettsiaCOI) when more blocking primers were added. An increased number of PCR cycles resulted in a high number of host COI fragments. However, this increased somewhat the amplification of *Rickettsia* COI as well. Overall, HCO_DPO showed the highest efficiency among tested primers by preventing the amplification of *Rickettsia* COI, even at low concentrations.

Discussion

Our nationwide molecular screening results show that the Torix group of *Rickettsia* is widespread in freshwater amphipod hosts in New Zealand, and is to our knowledge the first report of *Rickettsia* in crustacean hosts. Because of the lack of information on *Rickettsia* infections in other groups of amphipods in other parts of the world, when and how these bacteria colonised and spread among New Zealand amphipods remain in question. Because freshwater amphipods have limited dispersal abilities³⁸, the widespread distribution of Torix *Rickettsia* in New Zealand amphipods may be explained by an ancient acquisition followed by vertical transmission, or by many independent events of recent horizontal transmission from other organisms. Several lines of evidence indicate that both horizontal and vertical transmission may have played roles in spreading and maintaining these bacteria in New Zealand amphipod hosts. The monophyletic relationships of most *Rickettsia* from New Zealand amphipod populations inferred by *rrs* and *gltA* sequences (Figs. 2, 3) seem to support the ancient acquisition scenario. Also, genetically closely related *Paracalliope* populations harbored the same genotype of *Rickettsia*, which also strongly supports their long-lasting relationship probably maintained by vertical transmission. Meanwhile, sharing of *Rickettsia* genotypes between sympatric amphipod species of different families suggests host shifts among genetically distant host species within the same order. Such a complex evolutionary history involving both vertical transmission and horizontal transfers has been reported for other insect/endosymbiotic systems³⁹. The Bayesian tree obtained with COI sequences provides some hints for horizontal transmission among amphipods and other arthropods (Fig. 4). *Rickettsia* sequences from darkling beetles (*Pimelia* sp.) obtained in Europe were highly similar to the sequences identified in New Zealand amphipods (96 ~ 98% similarity)⁴⁰. Moreover, *Rickettsia* sequences obtained from several unspecified arthropod species (*Mandibulata* sp., *Endopterygota* sp., and *Formicidae* sp.) in New Zealand (although these were originally identified as invertebrate COI sequences)⁴¹ are closely related to those of New Zealand amphipods (96 ~ 99% similarity), providing strong evidence of recent horizontal transmission among them. Unfortunately, details regarding the host specimens and local origins of these sequences in New Zealand are not available. Direct detection of *Rickettsia* from these arthropod species and multi-gene analyses will be necessary to elucidate their transmission routes. Interestingly, with a larger dataset with multigene data, Pilgrim et al. (2020)³⁴ have recently inferred frequent horizontal transmissions of Torix *Rickettsia* among distantly related hosts. This supports the recent horizontal transmission of *Rickettsia* among amphipods and insects in the shared habitat, warranting further investigation.

Our findings support the early observation that the Torix group of *Rickettsia* may be highly associated with aquatic and damp environments¹⁶, which was based on the detection of this group in leeches, amoeba, Diptera, and Coleoptera^{14,19,22,42}. Pilgrim et al. (2017) provided support for this view by detecting *Rickettsia* from 38% of *Culicoides* species tested and hypothesized that Torix *Rickettsia* may be dominant in insects with aquatic larval stages, and the 'aquatic hotspot' hypothesis has recently been strongly supported by a comparison of the incidence of Torix *Rickettsia* between terrestrial and aquatic hosts³⁴. Horizontal transmission of Torix *Rickettsia* among genetically distantly related but spatially co-occurring species may have occurred frequently¹⁶. The high prevalence of Torix *Rickettsia* and their stable association with their hosts suggest negligible pathogenic effects of this group^{19,22,42,43}. Some Torix *Rickettsia* may even be beneficial for their hosts. For example, infected leeches can have larger body sizes than uninfected individuals, although the possibility that larger individuals are more likely to acquire *Rickettsia* via horizontal transmission cannot be ruled out^{14,20}. Ecological impacts of Torix *Rickettsia* on their hosts, and direct evidence of horizontal transmission among aquatic host groups, could be better answered with targeted community-level studies.

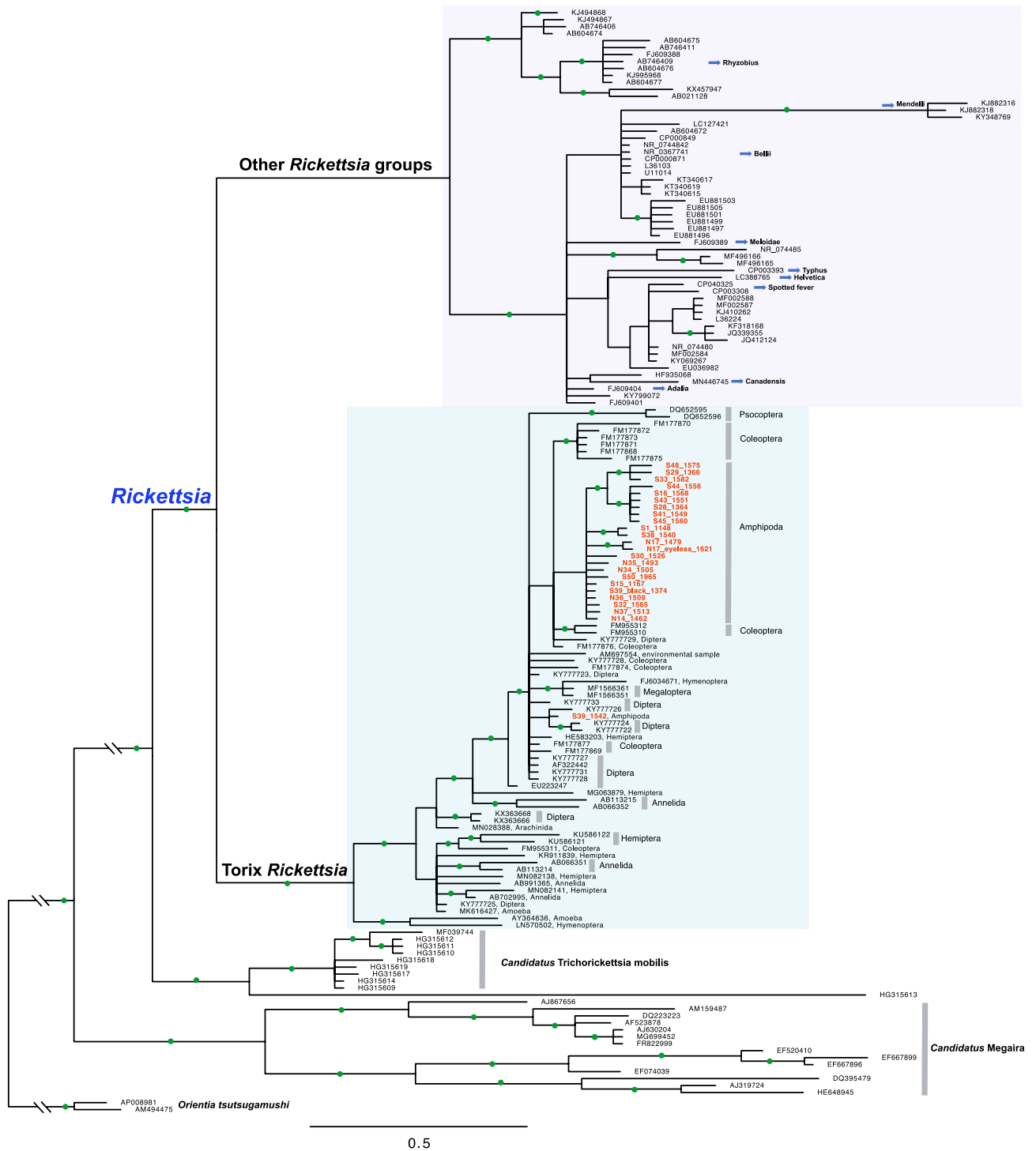


Figure 2. Bayesian tree of the genus *Rickettsia* based on 16S rRNA sequences. An alignment of 1198 bp of 158 taxa was used. Two well-supported clades are shown within the genus *Rickettsia*. One is the Torix group which includes endosymbionts of diverse hosts (host taxa indicated on the right), and the other clade includes all other 12 recognized groups of *Rickettsia*. Nodes with a posterior probability higher than 0.9 are shown with green circles. Sequences obtained in this study are highlighted in orange colour.

With the advancement of molecular techniques, our knowledge of the diversity of the sister groups of *Rickettsia* is also increasing and changing rapidly. Earlier studies focused mainly on the pathogenic and medically important species in arthropod hosts^{44,45}. Until 2005, only two genera, *Rickettsia* and *Orientia*, were known within the family *Rickettsiaceae*, which now contains seven more genus-level taxa^{46,47}. All these new genera are exclusively found in aquatic environments, mostly within ciliate hosts. It seems that adaptation to the use of arthropod hosts occurred several times independently within the family *Rickettsiaceae*⁴⁶. In addition, the phylogenetic status and

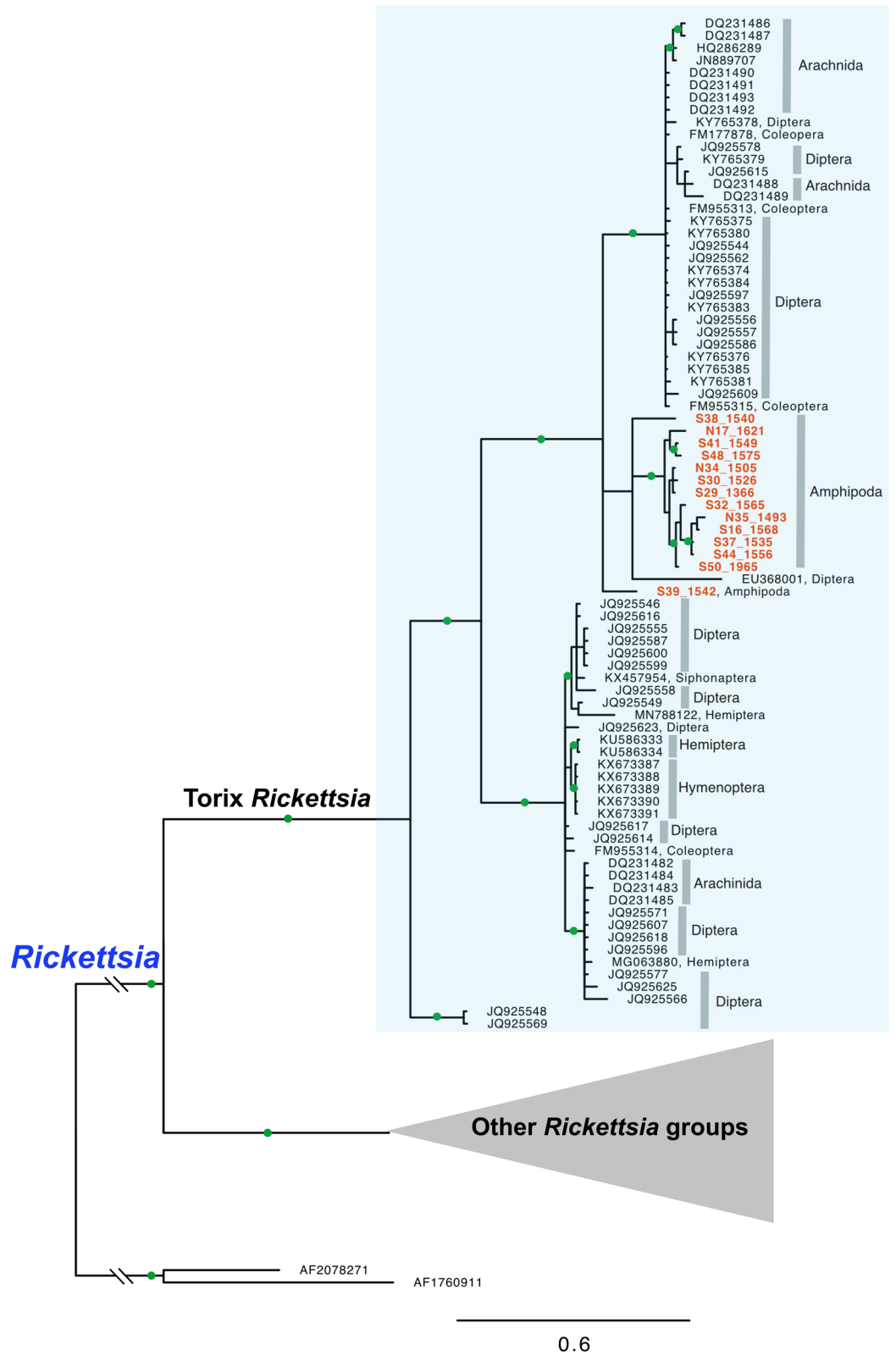


Figure 3. Bayesian tree of the genus *Rickettsia* based on an alignment of 765 bp of *gltA* sequences of 130 taxa (host taxa indicated on the right). Nodes with a posterior probability higher than 0.9 are shown with green circles. Sequences obtained in this study are highlighted in orange colour.

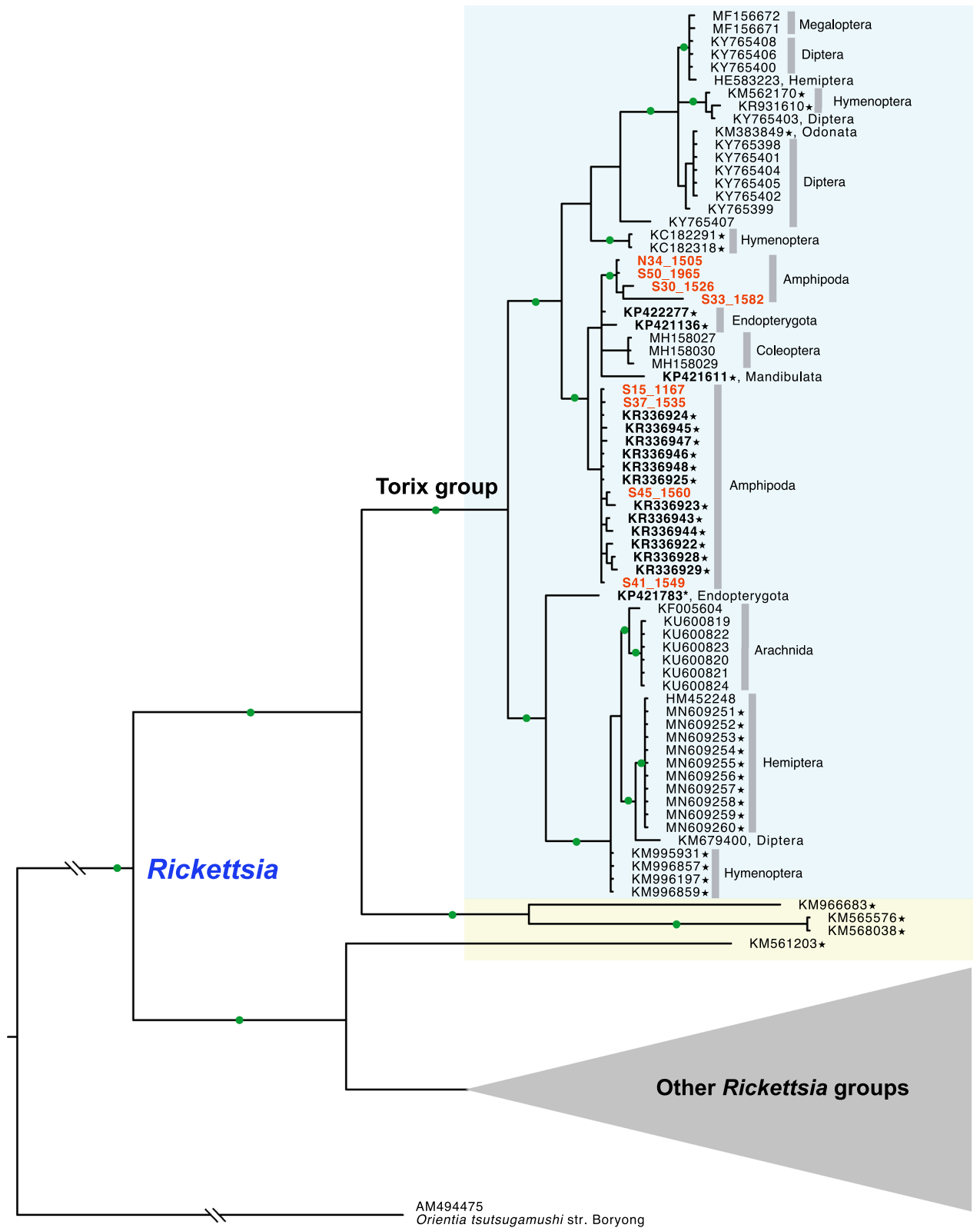


Figure 4. Bayesian tree of the genus *Rickettsia* based on an alignment of 559 bp of COI sequences of 114 taxa (host taxa indicated on the right). Nodes with a posterior probability higher than 0.9 are shown with green circles. Sequences that were misidentified as COI from invertebrate hosts, and initially not as rickettsial endosymbionts, in GenBank are highlighted with ★. Sequences obtained in this study are highlighted in orange colour. Sequences obtained in New Zealand are highlighted with bold text. Sequence similarity among Torix *Rickettsia* in the blue box is above 89%. Sequences that are similar to the other Torix *Rickettsia* but with lower similarity (80 ~ 82%), are highlighted within the yellow box.

	Condition 1				Condition 2				Condition 3			Condition 4
Annealing temperature	44 °C				48 °C				48 °C			48 °C
DNA templates	DNA extract 1				DNA extract 1				DNA extract 2			DNA extract 1
Number of cycles	38				38				38			48
Primer ratio (universal:blocking)	0	2	5	10	0	2	4	8	1	2	3	2
-	NA	-	-	-	448/7010 (0.0601)	-	-	-	NA	-	-	-
COI_BlocF	NA	7126/3426 (0.6753)	9257/1669 (0.8472)	1809/303 (0.8565)	NA	NA	NA	959/4955 (0.1622)	-	-	-	-
COI_BlocR	NA	7428/5660 (0.56754)	NA	1751/1450 (0.5470)	NA	NA	406/5975 (0.0636)	NA	-	-	-	-
COI_BlocF_2	-	-	-	-	NA	404/2028 (0.1783)	631/772 (0.4498)	618/332 (0.6505)	NA	779/2676 (0.2255)	NA	-
HCO_DPO	-	-	-	-	NA	407/0 (1)	156/0 (1)	NA	445/144 (0.7555)	463/91 (0.8357)	175/0 (1)	18488/196 (0.9895)

Table 1. PCR conditions and results of fragment analyses. PCR products obtained under different PCR conditions (different primers with different concentrations, annealing temperature, different DNA templates, number of PCR cycles) were run on capillary electrophoresis. The number of amplicons of host COI is followed by the number of amplicons of *Rickettsia* COI separated by (/), and their ratio is shown in parenthesis. The effects of blocking primers are highlighted in the grey boxes. Universal to blocking primer ratios that were higher than 0.8 are highlighted in red. NA: fragment analysis was conducted but the result was not available.

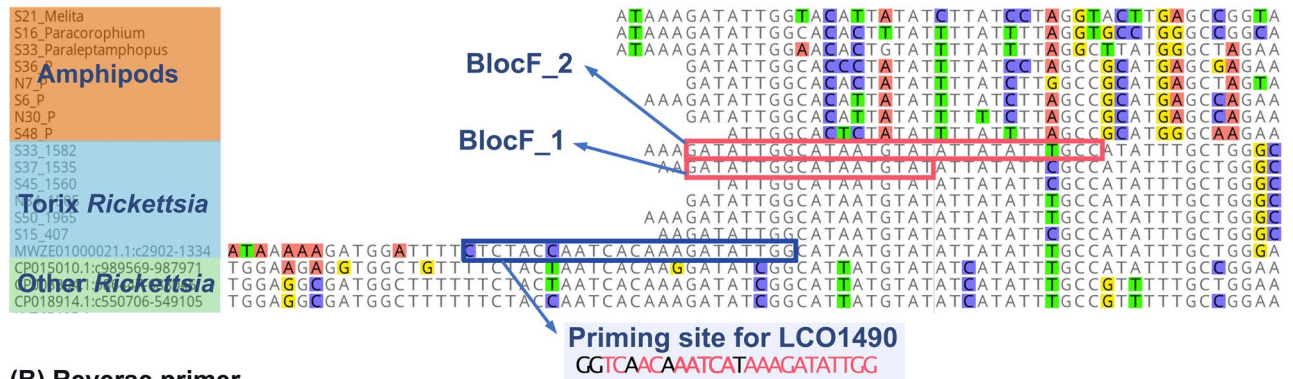
relationships among *Rickettsia* groups are also changing. The Hydra group, which was once considered to be an ancient group within the genus *Rickettsia*, along with the Torix and Belli groups, is now regarded as a separate genus: *Candidatus Megaira*⁴⁸. Another recently recognized group, *Candidatus Trichorickettsia*, is now believed to be a sister clade to the genus *Rickettsia*⁴⁷.

The Torix group is largely different from the other groups of *Rickettsia* in many respects, including host range and habitat. The Torix group includes not only endosymbionts of diverse aquatic invertebrates (that are more complex than ciliates), but also diverse terrestrial arthropod hosts. Also, the Torix group is genetically distinct from other groups of *Rickettsia*, which all are sister to Torix *Rickettsia*. Specifically, the genetic divergences between the Torix and the Belli groups are 96% in *rrs*, 78% in *gltA*, and 76% in COI sequences. The delimitation criteria we used for the Torix group in this study were 98.1% in *rrs*, 87.6% in *gltA*, and 89% in COI (broadly 80%; see Fig. 4). Two genome sequences of Torix *Rickettsia* recently became available^{23,43}. These genomes have the typical characteristics of *Rickettsia* (e.g. reduced genome size, and biosynthetic and catabolic capacity) but also have unique characteristics different from the other groups of *Rickettsia* (e.g. the presence of non-oxidative PPP, methionine salvage pathway, and glycolysis). It would be interesting to see how these two sister lineages, one mainly pathogenic and the other nonpathogenic, evolved and diverged from their common ancestor.

Interestingly, COI sequences revealed a diversity of Torix *Rickettsia*, as much as other popular markers for Rickettsiales such as *rrs* and *gltA*, even though COI has rarely been used as a marker of choice. Among 17 studies that have generated COI sequences of *Rickettsia*, only 3 were specifically intended to obtain COI sequences from *Rickettsia*^{23,49,50}. The remaining 14 studies obtained *Rickettsia* COI sequences as a byproduct of other research objectives (i.e. host identification, population genetic studies, or DNA barcoding) with PCR using universal primers. *rrs* is a widely used marker but may be too conserved to resolve phylogenetic relationships among closely related species, while the *gltA* gene shows more variability. Only 6 studies (including the present one) produced both *rrs* and *gltA* sequences^{22,23,43,51,52}. *gltA* sequences are not available for most Torix *Rickettsia* including endosymbionts of leeches and amoeba, as the reports of Torix *Rickettsia* from these groups precede the first use of *gltA* sequences to study Torix *Rickettsia*^{19,20,42,53}. Conversely, only *gltA* sequences are available for some species found in spiders and dipterans, which made it difficult to resolve the phylogenetic position of these rickettsial endosymbionts along with other *Rickettsia*^{14,53}. ‘Limoniae’ and ‘Leech’ groups were used within Torix *Rickettsia* in some studies based on the *gltA* gene and concatenated sequences of *gltA* and *rrs* genes, although the ‘Leech’ group was found not to be monophyletic^{22,54}. Our *gltA* tree showed two main lineages which correspond to the clades found in previous studies. Similarly, two main lineages were identified in the Bayesian tree inferred from COI sequences. However, whether the clade containing all endosymbionts from spiders represents the ‘Leech’ group could not be confirmed without direct multi-gene data from the same group. It is likely that there are many more subgroups, given the limited number of targeted studies available to date.

Endosymbionts and vertically transmitted intracellular parasites are common in arthropod hosts^{55,56}. In the context of the growing recognition of the ‘Holobiont’ concept^{57,58}, obtaining bacterial sequences from DNA extracts from host tissue is not surprising. Most bacterial ‘contaminations’ are filtered out during processing of metabarcoding data^{59,60}. The frequent recovery of COI sequences from Torix *Rickettsia* can be partly explained by their nucleotide sequence similarity with mitochondria. The Proto-mitochondrion (the hypothetical common ancestor of all mitochondria) is often recovered as a sister to the Rickettsiales or within the Rickettsiales⁶¹. The alignment of COI sequences among several lineages of *Rickettsia* shows the high similarity between priming sites and the sequences used for universal primers (Fig. 5). The priming site for the forward primer is 80% (20/25 nucleotides) identical to the LCO1490 sequences, and the priming site for the reverse primer is 84.6% (22/26 nucleotide) identical to the HCO2198 sequences. In addition, priming sites for universal primers are not conserved in many groups, which necessitates the need for group-specific or degenerate markers^{62,63}. However,

(A) Forward primer



(B) Reverse primer

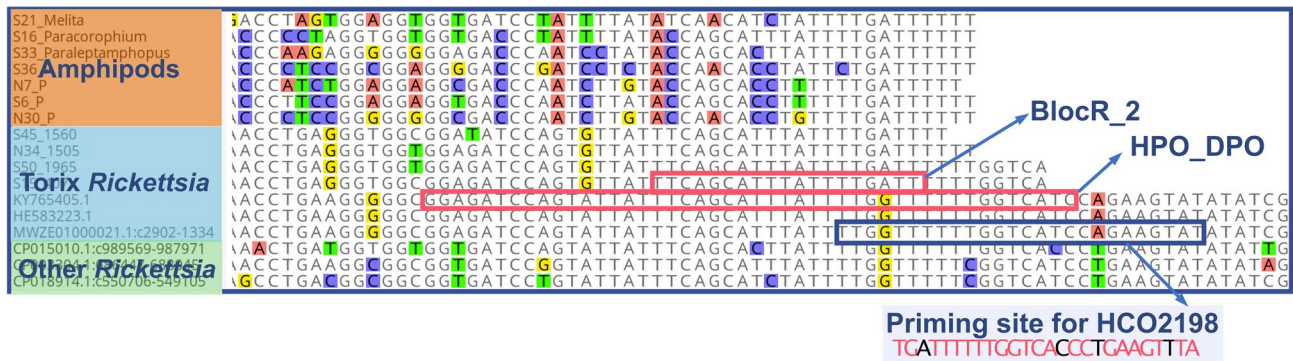


Figure 5. COI sequence alignments showing priming sites for both forward and reverse primers. Conserved regions to which universal primers bind are highlighted with blue squares. Positions at which nucleotides are the same as in universal primers are highlighted with pink texts in the primer sequences. COI sequences are highly divergent among amphipods whereas COI sequences from *Torix Rickettsia* from diverse host groups are highly conserved, which allowed the blocking primers (binding regions are highlighted with pink squares).

this does not explain the more frequent reports of *Torix* COI in GenBank, because priming sites are also highly conserved in other groups of *Rickettsia* (Fig. 5). Pilgrim et al. (2020)³⁴ proposed that the lack of SNP near the 3' end of the priming site of *Torix Rickettsia* may be responsible for this bias. Additionally, we hypothesize that overall high prevalence of *Torix Rickettsia* compared to other groups of *Rickettsia* in an infected host population can partly explain the bias, even though a formal comparison of prevalence between *Torix* and non-*Torix* groups was not made (but see Weinert et al., 2015¹⁵ for prevalence of diverse groups of *Rickettsia* in host populations). Therefore, several individuals from a given population may all be infected and could yield rickettsial COI, as illustrated in some previous studies^{31,33}, and in the current study.

These problems can be managed, as they are with *Wolbachia*¹¹. As mentioned earlier, *Torix* COI is highly conserved across diverse hosts. Therefore, comparing newly obtained (and suspicious) COI sequences with known *Torix Rickettsia* COI sequences can be easily done to distinguish *Torix Rickettsia*. Comparing sequences from the same taxon or genetically closely related groups could be useful. Checking for the presence of stop codons could largely decrease this problem, as for numts¹⁰. Bacterial sequences will show stop codons with the translation table for invertebrates, but will be in an open reading frame with the translation table for Bacteria. In addition to the high prevalence of *Torix Rickettsia* in many populations, high copy numbers of *Rickettsia* in host cells also make it difficult to obtain genuine host COI sequences, once a population is infected. In this case, applying blocking primers is a practical solution. Unfortunately, using blocking primers for *Rickettsia* does not always guarantee the amplification of host COI because other symbionts or parasites might still be amplified. Nevertheless, blocking primers can be widely used for any host groups that are infected by *Rickettsia*, and for both next-generation-sequencing as well as Sanger sequencing. These sequences should not be confounded with those of hosts, yet these 'unwanted sequences' or 'contaminations' can provide useful information about their endosymbionts and parasites. Our current study provides an example, as we confirm the presence of *Torix Rickettsia* after discovering contaminated sequences from a previous study³¹. Similarly, the COI sequence obtained from a damselfly suggested the presence of *Rickettsia* in this host group (order Odonata); this has recently been confirmed by a study with a targeted screening⁶⁴. Targeted studies are likely to uncover a huge but under-detected diversity of *Torix Rickettsia*, and with more data, we will be able to answer questions regarding transmission, host switching, and the evolution of pathogenicity. Furthermore, detailed research on a finer scale is needed to elucidate the impact of these widespread endosymbionts on their diverse hosts.

Gene	Primer	Sequence (5'-3')	References
<i>16S rRNA</i>	Ri170_F	GGGCTTGCTCTAAATTAGTTAGT	Kuchler et al. (2009)
	Ri1500_R	ACGTTAGCTCACCCACCTTCAGG	Kuchler et al. (2009)
<i>gltA</i>	RiGltA405_F	GATCATCCTATGGCA	Pilgrim et al. ²³
	RiGltA1193_R	TCTTTCCATTGCCCC	Pilgrim et al. ²³
<i>atpA</i>	RiAtpA327_F	GTCGGTAAAGCATTGCTTGGT	Pilgrim et al. ²³
	RiAtpA1309_R	ATTGATCCTGCTTCAATA	Pilgrim et al. ²³
<i>coxA</i>	LCO1490	GGTCAACAAATCATAAAGATATTGG	Vrijenhoek ⁴
	HCO2198	TAAACTTCAGGGTGACCAAAAAATCA	Vrijenhoek ⁴
<i>coxA</i> (Blocking primers)	COI_BlocF	GATATTGGCRTAATGTATAT	This study
	COI_BlocR	AAATCAAAATAAATGCTGA	This study
	BlocF_2	TTGGCRTAATGTATATTATATTYGCC	This study
	HCO_DPO	GGRTGACCAAAAAAYCAAAATAAATGCTGRIIIIIYACTGGATCTCC	This study

Table 2. List of primers used in this study. All blocking primers were modified with C3 spacer at the 3' end. Five deoxyinosines (I) were added in the middle of HCO_DPO primer as a linker.

Methods

Confirming *Rickettsia* infections by PCR. The presence of *Rickettsia* was confirmed by amplification of three different markers (*rrs*, *gltA*, and *atpA*) using *Rickettsia*-specific primer pairs^{22,23} (Table 2). The DNA sample (S15_470) from which we obtained presumably 'rickettsial COI' was used as DNA templates. Seven additional DNA samples from the same population were also included for PCR detection to compare the efficiency of the primer sets in order to select the best marker for molecular screening. For PCR reactions, 12.3 µl of distilled water, 4 µl of reaction buffer, 0.8 µl of each forward and reverse primers, 0.1 µl of MyTaq (Bioline), and 2 µl of DNA were used. PCR reactions were conducted under the following conditions: 95 °C initial denaturation for 5 min, 35 cycles of 94 °C for 30 s, 54 °C for 30 s, 72 °C for 120 s, final extension for 7 min at 72 °C. Then, 2 µl of PCR product from each PCR reaction was run on a 1.5% agarose gel.

Molecular screening of *Rickettsia* in New Zealand amphipods. We obtained extracted DNA samples of diverse amphipod specimens from Park et al. (2020)³⁰, in which the authors investigated the diversity of microsporidian parasites, a group of obligate intracellular eukaryotic parasites of amphipod hosts. Seven families of amphipods (Melitidae, Paracalliopidae, Paraleptamphopidae, Phreatogammaridae, Talitridae, Paracorophiidae, and an undescribed family of Senticaudata) were collected from 69 locations throughout both the South and North Islands (Fig. 1 and Supplementary Table 1). A total of 724 pooled DNA samples obtained from 2,670 individuals (mostly 4 individual amphipods per pool) were screened for *Rickettsia* by PCR by amplifying the *rrs* region under the PCR conditions and procedures described above.

Sequencing. *rrs*, *gltA*, and *atpA* sequences were obtained from populations that had been found positive for *Rickettsia* infections by PCR screening (Supplementary Table 1). PCR products were purified with MEGAquick-spin™ Total Fragment DNA Purification Kit (iNtRON Biotechnology) following the manufacturer's instructions. Purified PCR products were sent to Macrogen, Korea, for Sanger sequencing. Raw sequences were aligned in Geneious Prime 2019.0.4 (<https://www.geneious.com>) and ambiguous sites were carefully examined by eye. Haplotypes were identified by using the package *pegas*⁶⁵ in R version 3.5.2⁶⁶. We obtained some rickettsial COI sequences as a byproduct during host identification of amphipods with universal primers. We included these COI sequences along with nucleotide sequences from other genes for further analyses.

BLAST search. Blast search was done on GenBank with the *rrs*, *gltA*, *atpA*, and COI sequences obtained in this study. Based on the result of BLAST searches, all sequences that were considered as the Torix group (sequences with similarity to the query sequence higher than that between the query sequence and the Bellii group *Rickettsia*) were downloaded from GenBank (see Supplementary Tables 3–5) for further phylogenetic tree reconstructions. *rrs* and *gltA* sequences from other *Rickettsia* groups were also obtained and included for tree reconstruction (Supplementary Table 6). In addition, *rrs* sequences from recently discovered close relatives to *Rickettsia* (*Candidatus* Trichorickettsia, *Candidatus* Gigarickettsia, and *Candidatus* Megaira), and *Orientia tsutsugamushi* were included as outgroups.

Phylogenetic analyses. For each gene set, all sequences were aligned in Geneious Prime with the MUSCLE algorithm. Ambiguous sites were then eliminated in Gblocks with the least restrictive setting⁶⁷. The best-fitting model of nucleotide evolution for each dataset was determined based on the corrected Akaike information criterion (AICc) using jModelTest v2.1.6⁶⁸, which was conducted through the CIPRES Science Gateway v3.3⁶⁹. For all datasets, the General Time Reversible (GTR) model of nucleotide substitution along with Gamma distributed rate variation across sites (G) and the proportion of invariable sites (I) were chosen as the best model. Bayesian trees were inferred in MrBayes 3.2.7a⁷⁰. For each dataset, two independent runs, which consisted of four chains each, were simultaneously conducted for 10,000,000 generations with a sampling frequency of 1000.

The initial 25% of samples were discarded. The resulting trees were visualized in FigTree v1.4.4 (<https://tree.bio.ed.ac.uk/software/figtree/>).

Design of blocking primers. Blocking primers were designed following the guidelines of Vestheim et al. (2011)³⁶. In order to design blocking primers, COI sequences of the Torix group and some species belonging to other groups of *Rickettsia*, and COI sequences of New Zealand amphipods were aligned in Geneious Prime (Fig. 5). We designed four different annealing inhibiting blocking oligos which were intended to compete with universal primers. All met the following criteria: First, the blocking primers should overlap with one of the universal primers. Second, the blocking primers should specifically bind to the unwanted DNA templates (i.e. *Rickettsia*) but not to our target DNA templates (i.e. amphipod hosts). Third, 3'-end was modified so that it does not prime amplification (here, all with C3 spacer). Initially, two primers were designed: Bloc_F and Bloc_R (Table 2). However, GC contents of these primers were too low (27.5% and 21%, respectively), which resulted in a low expected melting temperature (T_m) of 43.2 °C and 42.5 °C, respectively. Ideally, T_m of a blocking primer should be higher than that of the competing primers³⁶. We, therefore, designed a longer primer, BlocF_2, to increase T_m to 51.7 °C (Table 2). A fourth primer was designed with a dual priming oligonucleotide (DPO): HCO_DPO (Table 2). A DPO can be used when it is impossible to find an appropriate binding site for a blocking primer adjacent to a binding site of a universal primer³⁵. A DPO primer consists of two separate segments connected with five deoxyinosines, and with C3 spacer modification at the 3' end. The total length of a typical DPO primer is long but it does not suffer from high T_m because a deoxyinosine linker, which assumes a bubble-like structure, allows the two segments to act independently⁷¹. All synthesized primers were purified with polyacrylamide gel purification (PAGE) to increase binding specificity by removing under-synthesized oligos.

Validation of blocking primers. We applied fragment analysis to test and compare the effectiveness of our blocking primers³⁵. Fragment analysis of fluorescently labeled PCR products on capillary electrophoresis can separate fragments in different sizes and can be used as a semi-quantitative method. When amplified with the universal LCO1490 and HCO2198 primers, the expected lengths of PCR products were different for amphipod hosts and *Rickettsia* COI, because *Rickettsia* COI is 6 bp longer. The FAM dye was attached to the 5' end of the LCO1490 primer. This fluorescently labeled forward primer was added to the PCR mixture instead of the normal (unlabeled) LCO1490 primer. Various factors can affect PCR success with blocking primers: T_m of primers, the concentration of primers (relative ratio between blocking primer and regular primer), the amount of the DNA templates in a PCR mixture (concentration of DNA), and the number of PCR cycles³⁵. To optimize PCR conditions, PCR reactions were conducted under several different PCR conditions (Table 1). Fragment analyses were carried out with a 1,200 LIZ size marker on an ABI 3730xl System (Applied Biosystems) at Macrogen (Korea). Results were analyzed with Peak Scanner Software 1.0 (Applied Biosystems; <https://www.thermofisher.com/>).

Data availability

All sequences generated in this study were deposited in GenBank (Accession ID: MT507651- MT507674; MT515460-MT515486; MT524989-MT525002).

Received: 3 June 2020; Accepted: 16 September 2020

Published online: 08 October 2020

References

- Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H. & Hallwachs, W. T. species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptus fuligator*. *Proc. Natl. Acad. Sci. USA* **101**, 14812–14817 (2004).
- Hajibabaei, M., Singer, G. A. C., Hebert, P. D. N. & Hickey, D. A. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends Genet.* **23**, 167–172 (2007).
- Bucklin, A., Steinke, D. & Blanco-Bercial, L. DNA Barcoding of Marine Metazoa. *Ann. Rev. Mar. Sci.* **3**, 471–508 (2011).
- Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* **3**, 294–299 (1994).
- Hebert, P. D. N., Cywinska, A., Ball, S. L. & DeWaard, J. R. Biological identifications through DNA barcodes. *Proc. R. Soc. B Biol. Sci.* **270**, 313–321 (2003).
- Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* **21**, 2045–2050 (2012).
- Elbrecht, V. & Leese, F. Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass-sequence relationships with an innovative metabarcoding protocol. *PLoS ONE* **10**, e0130324 (2015).
- Porter, T. M. & Hajibabaei, M. Over 25 million COI sequences in GenBank and growing. *PLoS ONE* **13**, e0200177 (2018).
- Buhay, J. E. “COI-like” sequences are becoming problematic in molecular systematic and DNA barcoding studies. *J. Crustac. Biol.* **29**, 96–110 (2009).
- Song, H., Buhay, J. E., Whiting, M. F. & Crandall, K. A. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proc. Natl. Acad. Sci. USA* **105**, 13486–13491 (2008).
- Smith, M. A. et al. *Wolbachia* and DNA barcoding insects: Patterns, potential, and problems. *PLoS ONE* **7**, e36514 (2012).
- Mioduchowska, M., Czyz, M. J., Goldyn, B., Kur, J. & Sell, J. Instances of erroneous DNA barcoding of metazoan invertebrates: Are universal cox1 gene primers too “universal”? *PLoS ONE* **13**, e0199609 (2018).
- Parola, P., Paddock, C. D. & Raoult, D. Tick-borne rickettsioses around the world: emerging diseases challenging old concepts. *Clin. Microbiol. Rev.* **18**, 719–756 (2005).
- Perlman, S. J., Hunter, M. S. & Zchori-Fein, E. The emerging diversity of *Rickettsia*. *Proc. R. Soc. B Biol. Sci.* **273**, 2097–2106 (2006).
- Weinert, L. A. The diversity and phylogeny of *Rickettsia*. in *Parasite diversity and diversification: evolutionary ecology meets phylogenetics* (eds. Morand, S., Krasnov, B. R. & Littlewood, D. T. J.) 150–181 (Cambridge University Press, 2015). doi:<https://doi.org/10.1017/CBO9781139794749.010>.

16. Weinert, L. A., Werren, J. H., Aebi, A., Stone, G. N. & Jiggins, F. M. Evolution and diversity of *Rickettsia* bacteria. *BMC Biol.* **7**, 1–15 (2009).
17. Hajduskova, E. *et al.* “*Candidatus Rickettsia mendelii*”, a novel basal group rickettsia detected in Ixodes ricinus ticks in the Czech Republic. *Ticks Tick. Borne. Dis.* **7**, 482–486 (2016).
18. Binetruy, F., Buysse, M., Barosi, R. & Duron, O. Novel *Rickettsia* genotypes in ticks in French Guiana. *South America. Sci. Rep.* **10**, 1 (2020).
19. Kikuchi, Y., Sameshima, S., Kitade, O., Kojima, J. & Fukatsu, T. Novel clade of *Rickettsia* spp. from leeches. *Appl. Environ. Microbiol.* **68**, 999–1004 (2002).
20. Kikuchi, Y. & Fukatsu, T. *Rickettsia* infection in natural leech populations. *Microb. Ecol.* **49**, 265–271 (2005).
21. Galindo, L. J. *et al.* Combined cultivation and single-cell approaches to the phylogenomics of nuclearioid amoebae, close relatives of fungi. *Philos Trans R Soc B Biol Sci* **374**, 2019 (2019).
22. Küchler, S. M., Kehl, S. & Dettner, K. Characterization and localization of *Rickettsia* sp in water beetles of genus *Deronectes* (Coleoptera: Dytiscidae). *FEMS Microbiol. Ecol.* **68**, 201–211 (2009).
23. Pilgrim, J. *et al.* Torix group *Rickettsia* are widespread in *Culicoides* biting midges (Diptera: Ceratopogonidae), reach high frequency and carry unique genomic features. *Environ. Microbiol.* **19**, 4238–4255 (2017).
24. Gollas-Galván, T., Avila-Villa, L. A., Martínez-Porchas, M. & Hernandez-Lopez, J. *Rickettsia*-like organisms from cultured aquatic organisms, with emphasis on necrotizing hepatopancreatitis bacterium affecting penaeid shrimp: An overview on an emergent concern. *Rev. Aquac.* **6**, 2014 (2014).
25. Larsson, R. A rickettsial pathogen of the amphipod *Rivulogammarus pulex*. *J. Invertebr. Pathol.* **40**, 28–35 (1982).
26. Graf, F. Presence of bacteria in the posterior caecum in the intestinal lumen of the hypogean Crustacean *Niphargus virei* (Gammaridae: Amphipoda). *Can. J. Zool.* **62**, 1829–1833 (1984).
27. Messick, G. A., Overstreet, R. M., Nalepa, T. F. & Tyler, S. Prevalence of parasites in amphipods *Diporeia* spp from Lakes Michigan and Huron. *USA. Dis. Aquat. Organ.* **59**, 159–170 (2004).
28. Winters, A. D., Marsh, T. L., Brenden, T. O. & Faisal, M. Analysis of bacterial communities associated with the benthic amphipod *Diporeia* in the Laurentian great lakes basin. *Can. J. Microbiol.* **61**, 72–81 (2015).
29. Bojko, J. *et al.* Parasites, pathogens and commensals in the “low-impact” non-native amphipod host *Gammarus roeselii*. *Paras. Vect.* **10**, 193 (2017).
30. Park, E., Jorge, F. & Poulin, R. Shared geographic histories and dispersal contribute to congruent phylogenies between amphipods and their microsporidian parasites at regional and global scales. *Mol. Ecol.* <https://doi.org/10.1111/mec.15562> (2020).
31. Lagrue, C., Joannes, A., Poulin, R. & Blasco-Costa, I. Genetic structure and host-parasite co-divergence: Evidence for trait-specific local adaptation. *Biol. J. Linn. Soc.* **118**, 344–358 (2016).
32. Řezáč, M., Gasparo, F., Král, J. & Heneberg, P. Integrative taxonomy and evolutionary history of a newly revealed spider *Dysdera ninnii* complex (Araneae: Dysderidae). *Zool. J. Linn. Soc.* **172**, 764 (2014).
33. Ceccarelli, F. S., Haddad, C. R. & Ramirez, M. J. Endosymbiotic Rickettsiales (Alphaproteobacteria) from the spider genus *Amaurobioides* (Araneae: Anyphaenidae). *J. Arachnol.* <https://doi.org/10.1636/joa-s-15-009> (2016).
34. Pilgrim, J. *et al.* Torix *Rickettsia* are widespread in arthropods and reflect a neglected symbiosis. *Authorea* <https://doi.org/10.22541/au.159534851.19125003> (2020).
35. Vestheim, H. & Jarman, S. N. Blocking primers to enhance PCR amplification of rare sequences in mixed samples: a case study on prey DNA in Antarctic krill stomachs. *Front. Zool.* **5**, 12 (2008).
36. Vestheim, H., Deagle, B. E. & Jarman, S. N. Application of blocking oligonucleotides to improve signal-to-noise ratio in a PCR. *Methods Mol. Biol.* **687**, 265–274 (2011).
37. Boessenkool, S. *et al.* Blocking human contaminant DNA during PCR allows amplification of rare mammal species from sedimentary ancient DNA. *Mol. Ecol.* **21**, 1806–1815 (2012).
38. Myers, A. A. Dispersal and endemism in gammaridean Amphipoda. *J. Nat. Hist.* **27**, 901–908 (1993).
39. Duron, O. *et al.* Origin, acquisition and diversification of heritable bacterial endosymbionts in louse flies and bat flies. *Mol. Ecol.* **23**, 2105–2117 (2014).
40. López, J. Á.R., Husemann, M., Schmitt, T., Kramp, K. & Habel, J. C. Mountain barriers and trans-Saharan connections shape the genetic structure of *Pimelia* darkling beetles (Coleoptera: Tenebrionidae). *Biol. J. Linn. Soc.* <https://doi.org/10.1093/biolinnean/bly053> (2018).
41. Drummond, A. J. *et al.* Evaluating a multigene environmental DNA approach for biodiversity assessment. *Gigascience* **4**, 2015 (2015).
42. Dyková, I., Veverková, M., Fiala, I., Macháčková, B. & Pecková, H. *Nuclearia pattersoni* sp. n. (Filosea), a new species of amphizoic amoeba isolated from gills of roach (*Rutilus rutilus*), and its rickettsial endosymbiont. *Folia Parasitol. (Praha)* **50**, 161–170 (2003).
43. Wang, H. L. *et al.* A newly recorded *Rickettsia* of the Torix group is a recent intruder and an endosymbiont in the whitefly *Bemisia tabaci*. *Environ. Microbiol.* <https://doi.org/10.1111/1462-2920.14927> (2020).
44. Azad, A. F. & Beard, C. B. Rickettsial pathogens and their arthropod vectors. *Emerg. Infect. Dis.* **4**, 179–186 (1998).
45. Raoult, D. *et al.* A flea-associated *Rickettsia* pathogenic for humans. *Emerg. Infect. Dis.* **7**, 73–81 (2001).
46. Castelli, M., Sasser, D. & Petroni, G. Biodiversity of “non-model” Rickettsiales and their association with aquatic organisms. in *Rickettsiales: Biology, Molecular Biology, Epidemiology, and Vaccine Development* (ed. Thomas, S.) 59–91 (Springer International Publishing, 2016). doi:https://doi.org/10.1007/978-3-319-46859-4_3.
47. Sabaneyeva, E. *et al.* Host and symbiont intraspecific variability: the case of *Paramecium calkinsi* and “*Candidatus Trichorickettsia mobilis*”. *Eur. J. Protistol.* **62**, 79–94 (2018).
48. Schrällhammer, M. *et al.* “*Candidatus Megaira polyxenophila*” gen. nov., sp. Nov.: considerations on evolutionary history, host range and shift of early divergent rickettsiae. *PLoS ONE* **8**, 2013 (2013).
49. MacHtelinckx, T. *et al.* Microbial community of predatory bugs of the genus *Macrolophus* (Hemiptera: Miridae). *BMC Microbiol.* **12**, S9 (2012).
50. Gerth, M. *et al.* Green lacewings (Neuroptera: Chrysopidae) are commonly associated with a diversity of rickettsial endosymbionts. *Zool. Lett.* **3**, 51 (2017).
51. Reeves, W. K., Kato, C. Y. & Gilchrist, T. Pathogen screening and bionomics of *Lutzomyia apache* (Diptera: Psychodidae) in Wyoming, USA. *J. Am. Mosq. Control Assoc.* **24**, 444–447 (2008).
52. Noda, H. *et al.* Bacteriome-associated endosymbionts of the green rice leafhopper *Nephotettix cincticeps* (Hemiptera: Cicadellidae). *Appl. Entomol. Zool.* **47**, 217–225 (2012).
53. Goodacre, S. L., Martin, O. Y., Thomas, C. F. G. & Hewitt, G. M. *Wolbachia* and other endosymbiont infections in spiders. *Mol. Ecol.* **15**, 517–527 (2006).
54. Gualtieri, L., Nuges, F., Nappo, A. G., Gebiola, M. & Bernardo, U. Life inside a gall: closeness does not favour horizontal transmission of *Rickettsia* between a gall wasp and its parasitoid. *FEMS Microbiol. Ecol.* **93**, 2017 (2017).
55. Rousset, F., Bouchon, D., Pintureau, B., Juchault, P. & Solignac, M. *Wolbachia* endosymbionts responsible for various alterations of sexuality in arthropods. *Proc. R. Soc. B Biol. Sci.* **250**, 91–98 (1992).
56. White, J. A., Giorgini, M., Strand, M. R. & Pennacchio, F. Arthropod endosymbiosis and evolution. in *Arthropod biology and evolution: molecules, development, morphology* (eds. Minelli, A., Boxshall, G. & Fusco, G.) 441–477 (Springer, Berlin, Heidelberg, 2013). doi:https://doi.org/10.1007/978-3-642-36160-9_17.

57. Minard, G., Mavingui, P. & Moro, C. V. Diversity and function of bacterial microbiota in the mosquito holobiont. *Paras. Vect.* **6**, 146 (2013).
58. Thompson, J. R., Rivera, H. E., Closek, C. J. & Medina, M. Microbes in the coral holobiont: Partners through evolution, development, and ecological interactions. *Front. Cell. Infect. Microbiol.* **4**, 176 (2014).
59. Siddall, M. E., Fontanella, F. M., Watson, S. C., Kvist, S. & Erős, C. Barcoding bamboozled by bacteria: Convergence to metazoan mitochondrial primer targets by marine microbes. *Syst. Biol.* **58**, 445–451 (2009).
60. Leray, M. *et al.* A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Front. Zool.* **10**, 1 (2013).
61. Roger, A. J., Muñoz-Gómez, S. A. & Kamikawa, R. The origin and diversification of mitochondria. *Curr. Biol.* **27**, 1177–1192 (2017).
62. Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R. & Hebert, P. D. N. DNA barcoding Australia's fish species. *Philos. Trans. R. Soc. B Biol. Sci.* **360**, 1847–1857 (2005).
63. Geller, J., Meyer, C., Parker, M. & Hawk, H. Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Mol. Ecol. Resour.* **13**, 851–861 (2013).
64. Thongprem, P., Davison, H. R., Thompson, D. J., Lorenzo-Carballa, M. O. & Hurst, G. D. D. Incidence and diversity of torix *Rickettsia*-odonata symbioses. *Microb. Ecol.* <https://doi.org/10.1007/s00248-020-01568-9> (2020).
65. Paradis, E. *pegas*: an R package for population genetics with an integrated–modular approach. *Bioinformatics* **26**, 419–420 (2010).
66. R Development Core Team, R. R: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing (2011). doi:<https://doi.org/10.1007/978-3-540-74686-7>.
67. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
68. Darrriba, D., Taboada, G., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9**, 772 (2012).
69. Miller, M. A., Pfeiffer, W. & Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. in *2010 gateway computing environments workshop (GCE)* 1–8 (IEEE, 2010). doi:<https://doi.org/10.1109/GCE.2010.5676129>.
70. Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
71. Chun, J. Y. *et al.* Dual priming oligonucleotide system for the multiplex detection of respiratory viruses and SNP genotyping of CYP2C19 gene. *Nucleic Acids Res.* **35**, e40 (2007).

Acknowledgements

The authors gratefully thank Sophie Courjal, Eleanor Wainwright, Colin Décout, Brandon Ruehle, Jean-François Doherty, and Heloise Pavanato for their assistance in the field and/or laboratory. We appreciate Dr. Jim Lowry's help for identifying amphipods. Also, we thank Dr. Tania King for introducing blocking primers and all practical help in the laboratory. We appreciate helpful comments provided by Dr. Pablo Tortosa and Jean-François Doherty. All the figures were drawn by EP. Figure 1 contains data (map) sourced from the LINZ Data Services (<https://data.www.linz.govt.nz>) licensed for reuse under CC BY 4.0. All the animal images used in Fig. 1 were taken by EP. This research was supported by a University of Otago Doctoral Scholarship to EP.

Author contributions

E.P. and R.P. designed the study. E.P. generated and analyzed the data, and wrote the manuscript. R.P. provided critical feedback on the manuscript. Both authors read and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-73986-1>.

Correspondence and requests for materials should be addressed to E.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020