Evolutionary Journal of the Linnean Society, 2023, **2**, 1–11
https://doi.org/10.1093/evolinnean/kzad001
Advance access publication 14 February 2023
Original Article

OXFORD

Original Article

Wolbachia and *Spiroplasma* endosymbionts in the *Anurida maritima* (Collembola) species group

MARTIJN J.T.N. TIMMERMANS[1,2*], HARI PRABHA[2], STEPHEN KETT[2]

[1]Natural History Museum, London SW7 5BD, UK
[2]Middlesex University London, London NW4 4BT, UK

[*]Corresponding authors. E-mail: m.timmermans@mdx.ac.uk

A large proportion of arthropods carry maternally-inherited endosymbiotic bacteria with which they have developed close relationships. Some of these endosymbionts are selfish genetic elements and manipulate their hosts' reproduction to their own advantage, for example via cytoplasmic incompatibility (CI) or male killing. Here we report that the sexually reproducing collembolan *Anurida maritima* carries two endosymbionts (*Wolbachia* and *Spiroplasma*) and provide genome sequences for both bacteria. Phylogenomic analyses indicate that the *Wolbachia* belong to the A supergroup and that the *Spiroplasma* are sister to the Citri-Chrysopicola-Mirum lineage. *Anurida maritima* is considered a species group and consists of at least two distinct genetic lineages. We show that both lineages carry both endosymbionts. No homologues of the *Spiroplasma* male-killing gene *SpAID* were observed within our datasets. Homologues of the male-killing associated *wmk* and the CI inducing *cifA* and *cifB* prophage genes were detected in the *Wolbachia* genome. Phylogenetic analyses placed the *cif* genes in the poorly characterized Type V clade. The *cifA* and *cifB* gene sequences of the two *Anurida* lineages are identical. It therefore seems unlikely that the genetic divergence within the species group stems from *cifA* and *cifB* induced CI. Laboratory controlled genetic crosses and sex ratio studies will be needed to reveal any potential effect of the two endosymbionts on *A. maritima*'s reproduction.

ADDITIONAL KEYWORDS: *Anurida bisetosa*; cytoplasmic incompatibility; Hexapoda; male-killing; Oxford Nanopore sequencing; Supergroup A.

INTRODUCTION

Endosymbionts are common in the Hexapoda. A large proportion of taxa within this hyper-diverse subphylum (which includes Insecta, Collembola, Protura and Diplura) have evolved long-standing intra- and intercellular relationships with microorganisms (Weinert *et al.*, 2015). These relationships can be obligate, where both host and microbe are fully dependent on each other, or facultative, where the host does not require the endosymbiont's presence for reproduction or survival (Moran *et al.*, 2009). Both types of interactions have had far reaching eco-evolutionary consequences and have shaped the diversity of the group (Moran *et al.*, 2009; Sudakaran *et al.*, 2017).

Some endosymbionts are exclusively transmitted vertically, via the maternal lineage (Moran and Baumann, 2000). Where vertical transmission is required, obligate endosymbionts are safeguarded by their shared dependence. However, facultative endosymbionts, must promote their own inheritance. To aid their transmission, facultative endosymbionts may provide certain fitness benefits to their hosts. They may, for example, allow their hosts to reside in otherwise unsuitable niches by providing essential dietary nutrients (Cornwallis *et al.*, 2021). Other facultative endosymbionts produce toxins that provide host protection against predators, parasites or fungal pathogens (Brownlie & Johnson, 2009; Eleftherianos *et al.*, 2013; Ballinger and Perlman, 2019; Massey and Newton, 2022). For example, within the genus *Spiroplasma*, which includes species that are extracellular endosymbionts of arthropods, several species produce ribosome-inactivating proteins (RIPs). RIPs are toxins that permanently inhibit normal protein synthesis in invasive organisms and therefore increase fitness of both endosymbiont and host (Ballinger and Perlman, 2017; Ballinger *et al.*, 2018).

However, not all facultative endosymbionts are beneficial. Various species are considered reproductive parasites and enhance their own transmission by manipulating the reproductive biology of their hosts (Stouthamer *et al.*, 1999; Cordaux *et al.*, 2011). This is also observed within the genus *Spiroplasma*, where some species are known to manipulate host reproduction by killing male offspring of infected females

Received September 29 2022; revised February 6 2023; accepted for publication February 10 2023

(Williamson *et al.*, 1999; Martin *et al.*, 2020). The plasmid-encoded SpAID protein is responsible for the male-killing phenotype of *Spiroplasma poulsonii* Williamson *et al.* 1999 and disturbs X-chromosome dosage compensation in embryos (Harumoto and Lemaitre, 2018).

Probably the best-known reproductive manipulators belong to the genus *Wolbachia*. Members of this intracellular genus influence host reproduction to increase frequency of infected individuals over uninfected ones (Stouthamer *et al.*, 1999; Werren *et al.*, 2008). This is achieved by reducing reproductive success of uninfected individuals compared to that of infected ones via mechanisms such as male-killing, parthenogenesis, feminization and cytoplasmic incompatibility (CI) (Correa and Ballard, 2016). Of these four mechanisms, CI is the most prevalent (Stouthamer *et al.*, 1999; Lindsey *et al.*, 2018). In CI, offspring of an infected male and an uninfected female, or a female infected by a different strain, are non-viable (Werren, 1997). CI is controlled by the CI factor genes *cifA* and *cifB*. These two adjacent prophage genes closely interact to define the CI phenotype, with *cifA* reversing *cifB* induced reproductive sterility (LePage *et al.*, 2017). Although the exact molecular mechanisms have not yet been resolved (Wang *et al.*, 2022), laboratory crosses and transgenic experiments have shown that, unless a female that expresses a matching *cifA* is involved, male expression of *cifA* and *cifB* causes embryonic death (LePage *et al.*, 2017; Shropshire *et al.*, 2018). This 'matching-pair' requirement is supported by phylogenetic analyses which have revealed strong congruence between *cifA* and *cifB* gene trees (Martinez *et al.*, 2021). These analyses have also identified five *cif* lineages, which have been named Types I to V (Martinez *et al.*, 2021).

Wolbachia are divided into various supergroups based on sequence divergence and phylogenetic relationships (Lo *et al.*, 2002). Most of the supergroups are restricted to arthropods, with the exception of Supergroups C, D and L whose members infect nematodes, and Supergroup F which has been reported to infect both nematodes and insects (Lefoulon *et al.*, 2016). Of note, the divergent Supergroup E (Vandekerckhove *et al.*, 1999) is predominantly found in asexually reproducing Collembola (springtails) and has been associated specifically with parthenogenesis induction within this order (Ma *et al.*, 2017). For example, in the absence of *Wolbachia* the parthenogenetic collembolan *Folsomia candida* Willem, 1902 will produce non-viable eggs (Pike and Kingcombe, 2009; Timmermans and Ellers, 2009).

The collembolan *Anurida maritima* Guérin-Méneville 1836 is a sexually reproducing species (Dallai *et al.*, 1999) with a cosmopolitan distribution. It is found in the intertidal zone and is notable for its endogenously controlled circatidal rhythms of behaviour (McMeechan *et al.*, 2000). The animals spend a large amount of their life in large aggregations under rocks and in crevices. A fraction of animals will leave these aggregates during low tide to scavenge, but all return 1 hour before high tide to ensure they are protected from the incoming water (Joosse, 1966). The species consists of divergent lineages (Arbea, 2001; Timmermans *et al.*, 2022) and some authors consider it to be a species group (Sun *et al.*, 2018). Others have even suggested the species should be divided into at least two species (*A. maritima* and *Anurida bisetosa* Bagnall, 1949) (Arbea, 2001). What caused the lineages to diverge is not known, but the question arises whether endosymbionts could be involved. *A. maritima* was recently reported to be infected by *Wolbachia* (Gavotte, 2018). In this paper we report that *A. maritima* harbours a second bacterial endosymbiont (*Spiroplasma*). We provide full genome sequences for both bacteria and show they are present in multiple *A. maritima* populations. The genome sequences are used for phylogenomic analyses with two aims: first, to determine if the *Wolbachia* endosymbiont of the sexually reproducing *A. maritima* clusters with other Collembola-infecting and 'parthenogenesis inducing' Supergroup E *Wolbachia*. Second, to get a better understanding of the *Spiroplasma* endosymbiont and its relationship to other insect-infecting and plant-pathogenetic strains. Finally, we look for endosymbiont genes involved in reproductive manipulation. If the endosymbionts are reproductive parasites, they could well have triggered the species group divergence, via CI or other mechanisms.

MATERIAL AND METHODS

Oxford Nanopore sequencing

Genomic DNA was extracted from 30 animals (*A. maritima*) collected in Wells-next-the-Sea (Norfolk, UK) using a Monarch Genomic DNA Purification Kit (New England Biolabs, Ipswich, Massachusetts, USA) following the 'Genomic DNA Purification from Insects' protocol with minor modifications and RNAse A treatment. The genomic DNA (gDNA) was eluted in 50 µL Elution Buffer, and quality and quantity assessed using agarose gel electrophoresis, TapeStation automated electrophoresis (Genomic DNA Screentape) (Agilent, Waldbronn, Germany), Nanodrop spectral analysis (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and Qubit fluorometric quantification (Broad Range Kit) (Thermo Fisher Scientific). Subsequently, 400 ng of gDNA was used to construct a sequencing library using a Rapid Sequencing Kit (Oxford Nanopore [https://nanoporetech.com/], Oxford, UK). The library was sequenced using a MinION and a R9.4.1 flow cell. Base calling of the raw signal data (Fast5 format) was performed using Guppy (guppy_basecaller v.5.0.7 + 2332e8d). Sequence quality control was conducted using the NanoPlot 1.28.0 (de Coster *et al.*, 2018) tool.

Sequence assembly, correction and genome annotation

The Oxford Nanopore sequencing reads were assembled using Flye v.2.8.3-b1767 (Kolmogorov *et al.*, 2019) and Canu v.1.8 (Koren *et al.*, 2017). The Flye assembly used a reduced coverage (asm-coverage 30) and a genome size of 250 million bases (setting: 0.25g) to limit memory consumption. The Canu assembly used the following settings: genomeSize=250m, corMhapFilterThreshold=0.0000000002, corMhapOptions='--threshold 0.80, --num-hashes 512 --num-min-matches 3 --ordered-sketch-size 1000 --ordered-kmer-size 14 --min-olap-length 2000 --repeat-idf-scale 50' mhapMemory=30g mhapBlockSize=500 (see https://canu.readthedocs.io/en/latest/faq.html). The full *Wolbachia* genome was extracted from the Canu assembly and the *Spiroplasma* genome from the Flye assembly (see Results). Both genomes were polished using the medaka_consensus programme which is part of Medaka (Oxford Nanopore Technologies, https://

github.com/nanoporetech/medaka) using the r941_min_fast_g507 model. Further polishing was conducted using DIAMOND and MEGAN as described by Bağcı *et al.* (2021) and Homopolish (Huang *et al.*, 2021). DIAMOND was run in the frameshiftaware alignment mode (Buchfink *et al.*, 2015; Huson *et al.*, 2018) with a penalty of 15, with range culling invoked and only retaining the 10% best scoring alignments (setting: --top 10). Furthermore, the 'query-gencode' flag was used to set the appropriate genetic code (4 for *Spiroplasma* and 11 for *Wolbachia*). The output of DIAMOND was then processed using MEGAN; it was 'meganised' using the daa-meganiser tool and frame-shift corrected data was exported in fasta format using the read-extractor tool. This sequence was then polished using Homopolish, using the R9.4.pkl model and bacteria msh sketch (Huang *et al.*, 2021). To obtain coverage estimates and manually inspect the assembly, all reads were realigned to the two genomes using Minimap2 (Li, 2018), using –N 1 to prevent secondary alignments. Samclip (https://github.com/tseemann/samclip) was used to remove all reads that had more than 100 positions clipped and coverage data was subsequently calculated using Samtools (Li *et al.*, 2009) depth function, with mean values calculated using an AWK statement. Mapped reads were used to visually inspect the quality of the assembly.

Both genomes were circularized using Geneious Prime (https://www.geneious.com/prime/) and annotated using NCBI Prokaryotic Genome Annotation Pipeline (PGAP v.2022-04-14.build6021) (Tatusova *et al.*, 2016). Completeness of the genomes was assessed using BUSCO (Simão *et al.*, 2015) v.5.2.2 with the Rickettsiales and Entomoplasmatales odb10 datasets for *Wolbachia* and *Spiroplasma*, respectively. To compare completeness statistics and overall genome length with those of other published genomes, BUSCO was also run on all full length, 'non-anomalous' *Wolbachia* (n = 51) and Entomoplasmatales (n = 63) genomes available in NCBI's 'Assembly' database.

Phylogenetic analyses

Phylogenomic analyses were based on the GToTree workflow (Lee, 2019), which uses HMMER3 (Eddy, 2011), MUSCLE (Edgar, 2004), TrimAl (Capella-Gutiérrez *et al.*, 2009), TaxonKit (Shen & Xiong, 2019) and GNU Parallel (Tange, 2021) to extract and align sequences. For these analyses the PGAP annotations were first extracted from both endosymbiont genomes and protein translations obtained using Geneious Prime. Full genomes of related species were downloaded from NCBI's Reference Sequence Database (RefSeq). For the analysis involving the *Spiroplasma* genome all available full length Entomoplasmatales, *Acholeplasma* and *Phytoplasma* genomes were retrieved using the following search command: (acholeplasma[ORGN] OR phytoplasma[ORGN] OR entomoplasmatales[ORGN]) AND 'latest refseq'[filter] AND 'complete genome'[filter]. This dataset was supplemented by *Spiroplasma ixodetis* Tully *et al.* 1995, *Spiroplasma poulsonii* and *Spiroplasma platyhelix* Williamson *et al.* 1997 (GCA_017847675.1, GCA_017847655.1, GCF_009866525.1, GCF_012163225.1). The *Acholeplasma* and *Phytoplasma* genomes were included as outgroups. For the analysis involving the *Wolbachia* genome the reference genomes of Scholz *et al.* (2020) were used, supplemented by the genomes of the endosymbionts of *Folsomia candida* Willem, 1902, *Ctenocephalides felis* Bouché, 1835, *Cruorifilaria tuberocauda* Eberhard, Morales and Orihel, 1976 and *Dipetalonema caudispina* Molin, 1858 (GCF_1931755, GCF_12277295, GCF_13365475, GCF_13365495). The Alphaproteobacteria (117 genes) and Tenericutes (99 genes) gene-sets were used within GToTree to extract conserved single copy genes from the *Wolbachia* and *Mollicutes* (Entomoplasmatales, *Acholeplasma* and *Phytoplasma*) genomes, respectively. Both gene retrieval searches used four parallel jobs (Tange, 2021) and a gene-length filtering cut-off threshold of 0.40. Potentially duplicated genes and genomes with less than 50% of the total target genes recovered were removed from further analyses. The obtained amino acid data matrices were used for phylogenetic analyses using IQ-TREE (Nguyen *et al.*, 2015). IQ-TREE was first used to find a suitable partitioning scheme (-m MF+MERGE) (Kalyaanamoorthy *et al.*, 2017) considering the top 10% partitioning schemes only (-rcluster 10). Tree searches were performed using the selected scheme, and branch support was evaluated using 1000 ultrafast bootstrap replicates and the SH-aLRT test (Guindon *et al.*, 2010). Trees were visualized using FigTree v.1.4.4 (https://github.com/rambaut/figtree/).

Infections in additional populations

To investigate the broader prevalence of *Spiroplasma* and *Wolbachia* in the *A. maritima* species group additional populations were analysed using PCR. Animals were collected from Lundy (Bristol Channel, UK), and from Leigh, Maldon and Goldhanger (Essex, UK). The Lundy and Goldhanger populations have previously been identified as *A. bisetosa* and the Maldon and Leigh populations as *A. maritima* J. Arbea (personal communication). For each sample location, DNA was extracted from 10 pooled individuals using a DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany). PCR was performed using PCRBIO Taq Mix Red (PCRBiosystems, London, UK) in 20 µL reaction volumes, using a final concentration of 0.5 µM for each primer. To detect the presence of *Wolbachia*, primers ftsZfl: 5'-GTTGTCGCAAATACCGATGC-3' and ftsZrl: 5'-CTTAAGTA AGCTGGTATATC-3' (Werren *et al.*, 1995) were used. To detect the presence of *Spiroplasma*, newly designed primers ITS_3140F: 5'-GCATCCACTATATGCTCTTTCTAACTT-3' and ITS_3377R: 5'-GATGGATCACCTCCTTTCTATGGA-3' were used. These primers target the Internal Transcribed Spacer (ITS) region (which is present twice in the assembly that we present here; see Results, Table 1). PCR reactions used the following conditions: 94°C for 5 min, 30 cycles of 94°C for 30 sec, annealing at 50°C for 30 sec, 72°C for 30 sec and a final extension at 72°C for 10 min. In addition, a publicly available transcriptome dataset (Misof *et al.*, 2014), generated from *A. maritima* collected on the island of Texel (the Netherlands) was searched for *Wolbachia* and *Spiroplasma* derived sequences using Blastn (Altschul *et al.*, 1997).

Male killing and CI genes

The raw Nanopore sequencing reads and all assembled data (i.e. Canu and Flye assemblies) were searched for homologues of *Spiroplasma* RIPs and the plasmid-encoded male-killing phenotype associated gene *SpAID* (accession number: PTQ58132; locus SMSRO_SFP00290) (Harumoto and

Table 1. Assembly statistics (length, BUSCO metrics and gene count) for the *Wolbachia* and *Spiroplasma* genomes. Both genomes were assembled into a single contig. The *Wolbachia* genome contained one full rRNA set, the *Spiroplasma* genome contained two full rRNA sets

Genome	*Wolbachia*	*Spiroplasma*
Accession number	CP102058	CP102059
Length (bp)	1 375 523	801 388
GC%	34.9	26.7
BUSCO		
Complete	357 (98.1%)	272 (81.9%)
Complete and single copy	354 (97.3%)	272 (81.9%)
Complete and duplicated	3 (0.8%)	0 (0%)
Fragmented	3 (0.8%)	2 (0.6%)
Missing	4 (1.1%)	58 (17.5%)
Total searched	364	332
Genes		
Protein coding	1397	714
rRNA	3 (1 set)	6 (2 sets)
ncRNA	3	2
Regulatory	0	3
tmRNA	1	1
tRNA	34	28
Total	1438	754

Lemaitre, 2018; Vera-Ponce León *et al.*, 2021) using Blastn (Altschul *et al.*, 1997). This was also done for homologues of the *Wolbachia* CI factors A and B (*cifA* and *cifB*) and the *WO-phage mediated killing* (*wmk*) gene (accession number: AE017196.1:611371-612282; WD_0626). For *cifA* and *cifB*, the gene sequences reported by Martinez *et al.* (2021) (Supporting Information; Table S2) were used as query sequences. Homologues of *cifA* and *cifB* were recovered (see Results), which were subsequently extracted from the assembled genomes and aligned to all the complete gene sequences of Martinez *et al.* (2021) using TranslatorX (Abascal *et al.*, 2010), with MAFFT (Katoh *et al.*, 2009) and GBlocks (Castresana, 2000) selection. Maximum Likelihood tree estimation was performed using the web server implementation of IQ-TREE (Trifinopoulos *et al.*, 2016) and used the GTR+G model with four rate categories. Branch support was evaluated using 1000 ultrafast bootstrap replicates and the SH-aLRT test. Trees were visualized using FigTree v.1.4.4.

To determine whether the *Wolbachia* endosymbionts of *A. bisetosa* carry different *cifA* and *cifB* alleles, Long-Range PCR (LR-PCR) was performed on animals collected in the Lundy landing bay area, followed by Nanopore sequencing. LR-PCR was performed on DNA extracted from a pool of 10 individuals. These PCR reactions used LA Taq (Takara Bio, Kusatsu, Japan) and the following PCR programme: 98°C for 1 min, 36 cycles of 98°C for 5 sec, 52°C for 30 sec, 60°C for 15 min and a final extension at 72°C for 10 min. Two different primer pairs were used: cifA-cifB_1F 5'-TACTGAATTCGTTACTTCGTGC-3' with cifA-cifB_1R 5'-AGAACGTTCTTGATTGGAGG-3' and cifA-cifB_2F 5'-CACCTGTTTGAGAATTGAGAACA-3' with cifA-cifB_2R5'-TGAAGACATGCTCTTAGATGTGG-3'. Both fragments covered the two genes as a single fragment. The products were then mixed and 80 µL of combined sample purified using AMPure XP Beads (Beckman Coulter, Brea, USA). Two hundred nanogrammes of purified product were used for library preparation using a Rapid Sequencing Kit (Oxford Nanopore). The library was sequenced on a MinION sequencer using a flongle adapter. Base calling of the raw signal data was performed as described above. Nanopore reads were mapped onto the *Wolbachia* genome using Minimap2 (Li, 2018).

RESULTS

Genome data, assembly and annotation

A total of 2 907 319 sequencing reads were obtained with a mean read length of 2643, a mean Phred quality score of 12.5 and a N50 of 6795. The reads were assembled using Canu and Flye. Inspection of the assembled data revealed the presence of a full-length *Wolbachia* genome in the Canu assembly and a full-length *Spiroplasma* genome in the Flye assembly. Partial *Spiroplasma* and *Wolbachia* genomes were present in the Canu and Flye assemblies, respectively, but only the two full genome sequences were used for further analyses. Completeness was assessed using BUSCO, which at this stage was 65.1% for the *Wolbachia* genome and 57.2% for the *Spiroplasma* genome. The two full genomes were subject to further polishing, which included frame-shift correction by DIAMOND and MEGAN. By adding single Ns (any nucleotide), the MEGAN frame-shift corrected 267 and 85 instances of a missing base in the *Wolbachia* and *Spiroplasma*, genomes respectively. By adding two additional Ns (any nucleotide), the MEGAN frame-shift corrected 332 and 150 instances of a superfluous base in the *Wolbachia* and *Spiroplasma* genomes, respectively. After these steps the *Wolbachia* and *Spiroplasma* genomes had an estimated BUSCO completeness of 94% and 82%, respectively.

These genome sequences were subject to a final round of polishing using Homopolish, which increased the BUSCO completeness estimate to 98.1% for the *Wolbachia* genome (Table 1). Unfortunately, the *Spiroplasma* genome failed this step as no related genomes with less than 5% divergence were available. The final *Wolbachia* and *Spiroplasma* genomes had a length of 1 375 523 bp and 801 388 bp, respectively. These lengths are within the expected range, but the BUSCO metrics are low compared to other full-length genomes available in the NCBI 'Assembly' database (Fig. 1; Supporting Information, Table S1). Read coverage was highly comparable for both genomes and estimated to be 25.0 (SD: 6.1) for the *Wolbachia* and 25.7 (SD: 9.1) for the *Spiroplasma* genome.

A total of 1438 genes were predicted for the *Wolbachia* genome (1397 protein coding genes, three rRNA genes, three non coding RNAs (ncRNAs), one transfer messenger RNA (tmRNA) and 34 tRNAs covering all 20 amino acids) (Table 1). Two CI factor homologues (*cifA* and *cifB*) were not detected correctly and were added manually. A total of 754 genes were predicted for the *Spiroplasma* genome [714 protein coding genes, three regulatory RNAs, two ncRNAs, six rRNA genes (i.e. two complete sets), one tmRNA and 28 tRNAs covering all 20 amino acids].

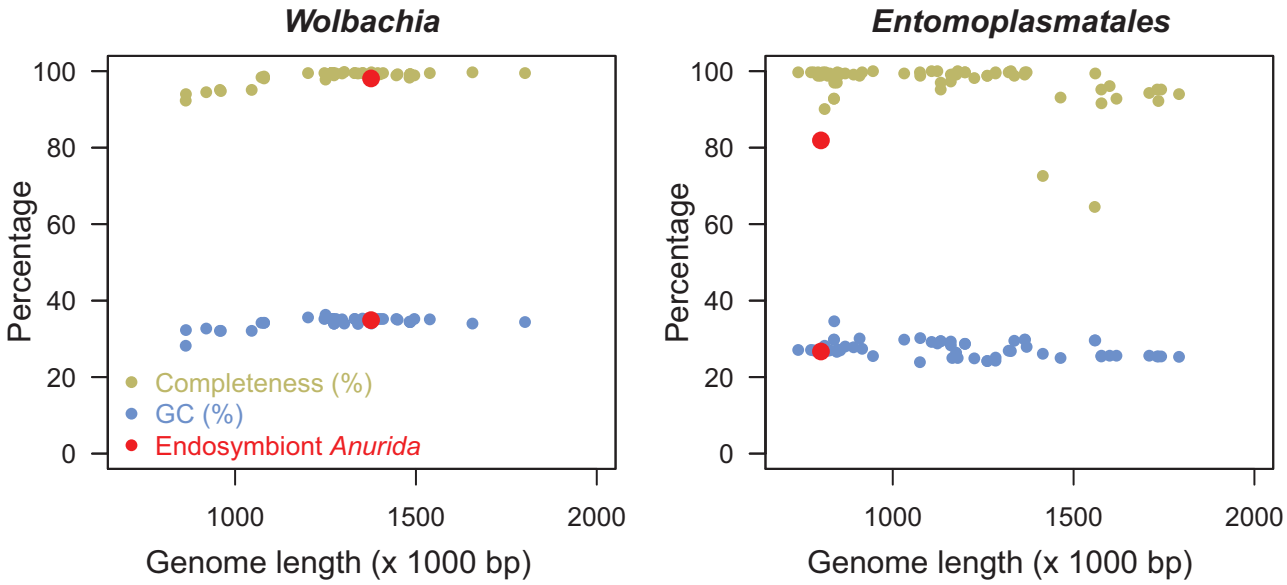

Figure 1. GC percentage (blue) and completeness (BUSCO metric; brown) of *Wolbachia* and Entomoplasmatales genomes plotted against their respective length. Data on the endosymbionts of *A. maritima* (red) are plotted together with data on 51 full length *Wolbachia* (left) and 63 full length Entomoplasmatales (right) genomes available in NCBI's 'Assembly' database. Accession numbers are given in the Supporting Information (Table S1).

Phylogenomic analyses

Wolbachia

A total of 48 *Wolbachia* genomes were selected for phylogenomic analyses, targeting 117 genes using the GToTree Alphaproteobacteria HMM source file. Out of these 117 genes, seven were missing from all the included *Wolbachia* genomes. Out of the remaining 110 genes, 106 (96%) were found in the genome sequence of the *Wolbachia* endosymbiont of *A. maritima*. Four genomes for which fewer than 50% of the targets were retrieved were excluded. The final data matrix consisted of 44 taxa, had a gene completeness of 87% and a length of 21 240 amino acids. The full partition model (i.e. one partition per gene for a total of 110 partitions) was merged into an optimal partition scheme consisting of three partitions (Supporting Information, Table S2), each with its own rate of evolution, and was used for phylogenetic tree reconstruction. The phylogenetic tree was rooted on the endosymbiont of the nematode *Pratylenchus penetrans* Cobb, 1917, which has been reported as the earliest diverging *Wolbachia* (Brown *et al.*, 2016). The endosymbiont of the collembolan *Folsomia candida* was the first to branch off (Supergroup E), followed by that of *Ctenocephalides felis*. Members of the C + D + F supergroups clustered together with high support, with each of these supergroups recovered as monophyletic. The A and B supergroups were both recovered and strongly supported. The endosymbiont of *A. maritima* was found at the most basal position in the A supergroup (Fig. 2).

Spiroplasma

For *Spiroplasma*, a total of 83 genomes were included for phylogenetic analysis, from which 99 genes were extracted using GToTree and the Tenericutes HMM source file. Of the 99 genes, 98 (99%) were recovered from the *Spiroplasma* genome sequence. This is a higher percentage than expected from the BUSCO analyses, which suggested a genome completeness of 82% only. The final data matrix contained all 83 species, had a gene completeness of 93% and a length of 19 519 amino acids. A partition scheme with nine partitions was deemed best fitting (Supporting Information, Table S2). The obtained topology was rooted using *Acholeplasma* and *Phytoplasma* as outgroups. The genus *Spiroplasma* is a paraphyletic genus (Gasparich *et al.*, 2004), and was recovered as such: the Mycoides-Entomoplasmatacea clade was sister to the *Spiroplasma* Apis clade. These two clades were then sister to the *Spiroplasma* Citri-Chrysopicola-Mirum (CCM) clade. The *Spiroplasma* endosymbiont of *A. maritima* occupied the most basal position in this latter clade (Fig. 3). The Ixodites clade was the most basal ingroup lineage.

Broader prevalence of endosymbionts and the *Wolbachia cifA cifB* phylogeny

To get a better understanding of the prevalence of the *Wolbachia* and *Spiroplasma* infections in the *A. maritima* species group, additional populations were screened using PCR (Fig. 4A). All included populations were shown to harbour both bacterial species. In addition, sequence similarity searches confirmed *Wolbachia* and *Spiroplasma* sequences to be present in a publicly available transcriptome dataset (Misof *et al.*, 2014) generated from animals sampled on the island of Texel (the Netherlands), indicating this population is infected by both endosymbionts too. The populations sampled represent two divergent lineages within the *A. maritima* species group: Lundy and Goldhanger populations have previously been identified as *A. bisetosa*, Wells-next-the-Sea and Maldon populations have previously been identified as *A. maritima* J. Arbea (personal communication). To determine if these endosymbionts play a role in the divergence observed, we searched for genes known

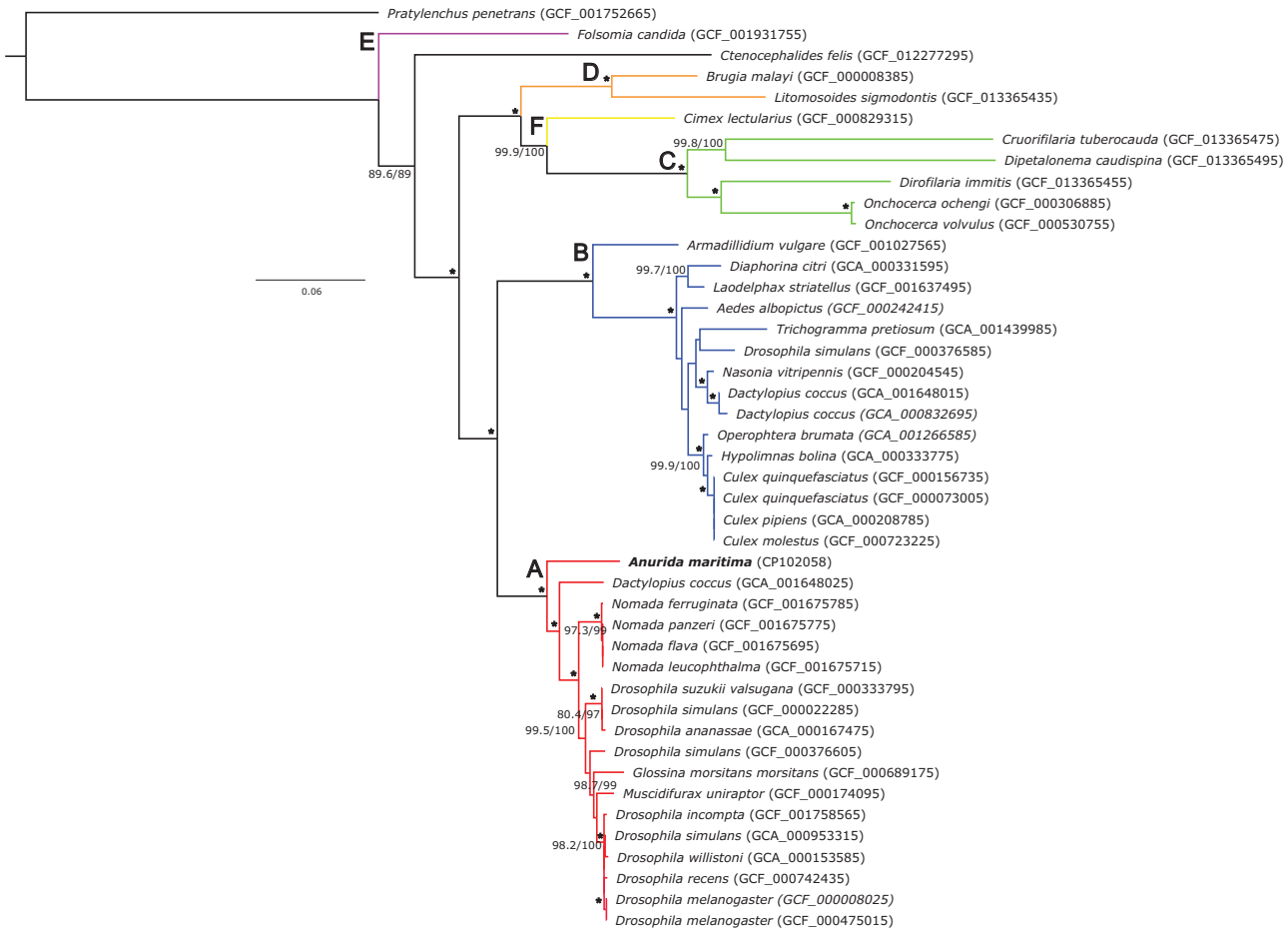

Figure 2. Maximum Likelihood phylogenetic tree showing the relationship among 44 *Wolbachia* endosymbionts, with host names and GenBank accession codes as taxon labels. Supergroups are indicated with letters A to F. The *Wolbachia* endosymbiont of *A. maritima* clusters with Supergroup A *Wolbachia* and is highlighted in bold. Numbers indicate support values (left: SH-aLRT test; right: ultrafast bootstrap). Values are given in bold if one of the two support values is lower than 80. An asterisk is used when both support values are 100.

to affect host reproduction. For *Spiroplasma* we focused on *SpAID*, and for *Wolbachia* on *wmk* and *cifA* and *cifB*. No *SpAID* homologues were detected in the raw and assembled datasets. A *wmk* homologue that showed 73.7% nucleotide sequence identity and 70.6% amino acid sequence identity to *wmk* of wMel (WD_0626) was found in the *Wolbachia* genome. In addition, sequences with high similarity to the *cifA* and *cifB* genes of *Acromyrmex echinatior* Forel, 1899 were recovered. As in other *Wolbachia* genomes, *cifA* is located directly upstream of *cifB* and the genes are coded on the same strand. Phylogenetic analyses confirmed a close relationship of both genes to the Type V *cifA* and *cifB* genes of *A. echinatior* (Fig. 4B).

High coverage Nanopore resequencing of a 5kb region covering the *cifA-cifB* (Fig. 4C) region from the divergent *A. bisetosa* lineage revealed two variants: a single polymorphic, non-synonymous site (P32L) in *cifA* and a single polymorphic site five bp after the stop codon of *cifB*. However, inspection of the whole genome shotgun reads derived from *A. maritima* from Wells-next-the-Sea revealed these sites to be polymorphic in the endosymbiont of the other lineage too. As these analyses were performed on pooled specimens, it remains currently unclear whether this variation is observed within or between individuals.

DISCUSSION

Here we report that the sexually reproducing collembolan *A. maritima* carries bacterial endosymbionts of the genera *Wolbachia* and *Spiroplasma*, and provide full length genome sequences for both microbes. It is, to our knowledge, the first reported *Spiroplasma* infection in the Order Collembola. Collembola have been reported to be infected by *Wolbachia*, and our phylogenomic analyses confirm the observation of Gavotte (2018) that the *Wolbachia* endosymbiont of *A. maritima* is in Supergroup A. This makes it unique in belonging to a supergroup that has not yet been reported for any other species of Collembola. At present, the nature of the host-microbe interactions remains unknown. However, the fact that all screened field populations were shown to be infected, indicates that the relationships are close and potentially biologically relevant.

The presented *Spiroplasma* and *Wolbachia* genomes are within the expected size range and are of acceptable completeness. Genome quality improved after every polishing step. This was specifically the case for the *Wolbachia* genome which went from 65.1% (unpolished), to 98.1% BUSCO completeness. Unfortunately, Homopolish could not be applied on the *Spiroplasma* assembly as there are currently no closely

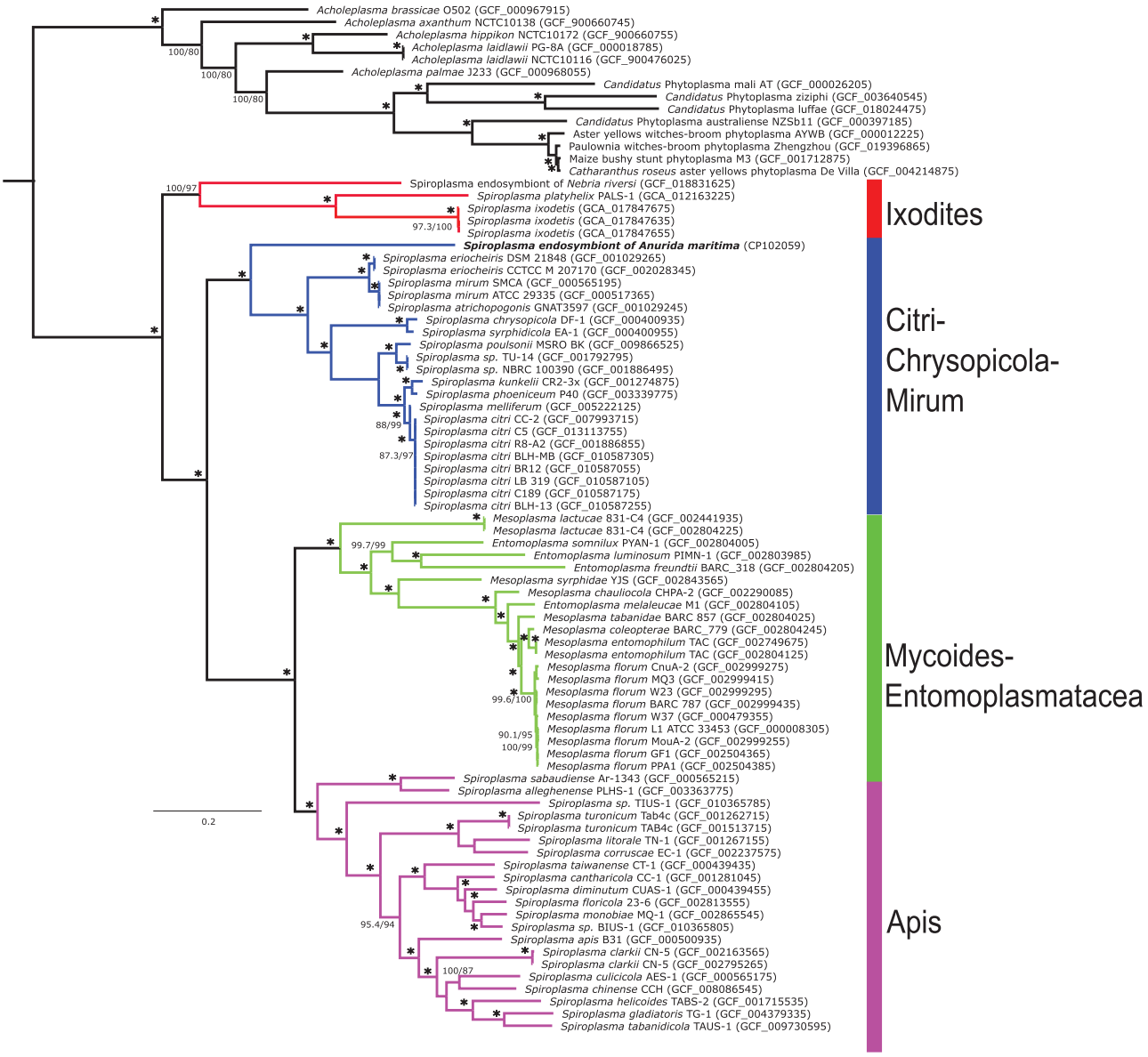

Figure 3. Maximum Likelihood phylogenetic tree showing the relationship among 83 full length Entomoplasmatales, *Achoplasma* and *Phytoplasma* genomes. Taxon labels give species and strain name and GenBank accession codes. Four different *Spiroplasma* clades are highlighted in red (Ixodites), blue (Citri-Chrysopicola-Mirum), green (Mycoides-Entomoplasmatacea) and pink (Apis). The endosymbiont of A. *maritima* groups with the Citri-Chrysopicola-Mirum clade and is highlighted in bold. Numbers indicate support values (left: SH-aLRT test; right: ultrafast bootstrap). Values are not shown if one of the two support values is lower than 80. An asterisk is used when both support values are 100.

related genomes (< 5% divergence) available. Although, the final BUSCO metrics for this latter genome were much lower (81.9% completeness), 98 out of the 99 included loci were retrieved by GToTree (i.e. 99%). Combined, these results suggest that the *Spiroplasma* genome is probably largely complete too, but that sequence errors remain and further polishing is needed.

Wolbachia

The two genomes were used for phylogenomic analyses, placing *A. maritima*'s *Wolbachia* endosymbiont at the base of the A supergroup. All except one species of Collembola-infecting *Wolbachia* belong to Supergroup E (Ma *et al.*, 2017); a supergroup that has been implicated to induce thelytokous parthenogenesis in infected species (Ma *et al.*, 2017; Konecka & Olszanowski, 2021) [but see Konecka *et al.* (2019)]. The prevailing phenotype of Supergroup A *Wolbachia* is CI and indeed homologues of genes known to induce CI were detected in *A. maritima*'s endosymbiont. Only a single pair of these homologues (*cifA* and *cifB*) was observed, with the two genes tandemly arranged and in the expected orientation. Both homologues were free of internal stop codons. They showed the highest similarity to *cif* genes of the leaf-cutter ant *A. echinatior* endosymbiont and phylogenetic analysis placed them in the Type V clade. This divergent clade is the most recent addition to the *cif* gene tree and remains the least studied of the five recognized types (Martinez *et al.*, 2021). It has, however, for at least one species (the *Wolbachia* of

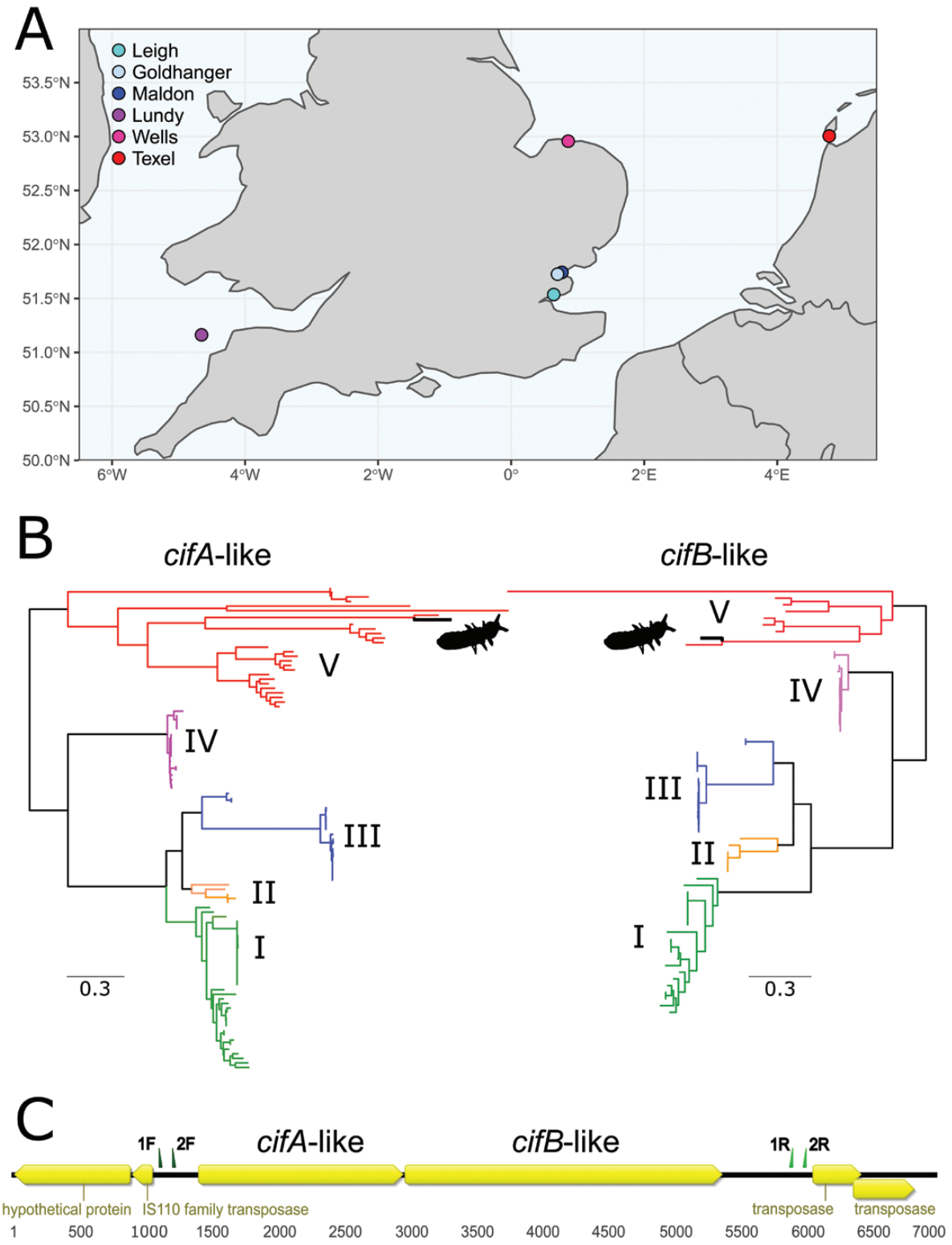

Figure 4. Field populations of *A. maritima* analysed for *cifA* and *cifB* homologues. A, the geographic location of English and Dutch field populations investigated for *Wolbachia* and *Spiroplasma* infections. B, Maximum Likelihood phylogenetic trees depicting relationships among *cifA* (left) and *cifB* (right) homologues. The five clades (I–V) are *sensu* Martinez *et al.* (2021). The *cifA* and *cifB* homologues of the *Wolbachia* endosymbiont of *A. maritima* cluster with clade V. Taxon labels and support values are provided in the Supporting Information (Fig. S1). C, orientation of *cifA* and *cifB* homologues in the genome of the *Wolbachia* endosymbiont of *A. maritima* and positions of primers used for LR-PCR.

Laodelphax striatellus Fallen, 1826) been linked to CI induction (Bing *et al.*, 2020).

Wolbachia infections can have profound evolutionary consequences for their hosts (Hurst and Jiggins, 2005). *Wolbachia* has been reported to have caused elevated (mitogenomic) divergence (Kodandaramaiah *et al.*, 2013; Ritter *et al.*, 2013) and has even been implicated in the process of speciation (Bordenstein, 2003). To determine whether the genetic divergence in the *A. maritima* species group is linked to the *Wolbachia* endosymbionts, we sequenced the 5 kb *cif* locus from the divergent '*A. bisetosa*-lineage'. If the divergence originated from CI induced genetic barriers, differences among the *cifA* and *cifB* alleles of the *Wolbachia* of the two host lineages would be expected. This was not the case; sequencing revealed the loci to be 100% identical, refuting the notion that divergence within the species group was enhanced by *Wolbachia cif* induced CI.

A second gene linked to reproductive manipulation was observed in the genome of *A. maritima*'s *Wolbachia* endosymbiont: a homologue of the *wmk* gene (Perlmutter *et al.*, 2019). *wmk* genes have been associated specifically with male killing. Unfortunately, the phenotypic expression of *wmk* and *wmk-like* genes is complex, with effects being dependent on gene sequence and host background (Perlmutter *et al.*, 2021). Transgenic expression has shown that small changes in gene sequence (including synonymous changes) and alternative start codon usage can abolish male killing or result in the death of both sexes (Perlmutter *et al.*, 2021). Hence, although slightly female biased sex ratios have been reported for *A. maritima* (Joosse, 1966), it is impossible to ascertain from the *wmk* homologue sequence alone whether the *Wolbachia* endosymbiont has any male-killing properties. Further sex ratio studies will be needed to determine whether or not this is the case.

Spiroplasma

Bacteria of the genus *Spiroplasma* associate closely with eukaryotes and infect a broad range of plants and animals (Harne *et al.*, 2020). Here we report for the first time an infection in the Order Collembola. The *Spiroplasma* endosymbiont of *A. maritima* represents a basal lineage of the CCM clade. The CCM clade is a diverse clade that harbours various plant and animal pathogens, some of which cause deadly diseases. For example, *Spiroplasma eriocheiris* causes tremor disease in Chinese mitten crabs (*Eriocheir sinensis* Panning, 1938) (Wang *et al.*, 2004), a disease from which the animals will eventually die (Liang *et al.*, 2009). Lethal CCM *Spiroplasma* have also been found in the crayfish *Procambarus clarkii* (Wang *et al.*, 2005), the shrimp *Penaeus vannamei* (Nunan *et al.*, 2005) and the honey bee *Apis mellifera* (Mouches *et al.*, 1983). The clade also contains the animal pathogen *Spiroplasma mirum*, originally isolated from ticks (Tully *et al.*, 1982). This species is known to cause cataracts in suckling mice (Tully *et al.*, 1982) and has been linked to the neurodegenerative disease transmissible spongiform encephalopathies (Bastian, 2014) [but see Alexeeva *et al.* (2006)]. Within the CCM clade *Spiroplasma poulsonii* (Williamson *et al.*, 1999) (an endosymbiont of *Drosophila* species) kills male embryos via the toxin Spaid. Interestingly, the same endosymbiont uses a different class of toxins (RIPs) to provide its *Drosophila* host protection against parasites, albeit at a cost to total lifespan (Garcia-Arraez *et al.*, 2019). It currently remains unknown whether the *Spiroplasma* endosymbiont has any phenotypic effects on *A. maritima*. Although the bacterium does not cause a deadly disease, it is not unlikely that it has other, less noticeable effects.

Sex ratio studies and controlled genetic crosses are now needed to gain a better understanding of *Anurida*-endosymbiont interactions and to reveal any phenotype and genome level effects. However, this will not be easy, as species belonging to the family Neanuridae are notably difficult to maintain under laboratory conditions (Hoskins *et al.*, 2015).

ACKNOWLEDGEMENTS

We would like to thank Daniel Garrun (Two Tree Island warden) and Elise Heinz for help with field sampling, Rebecca McPherson for commenting on an earlier version of the manuscript, Javier Arbea for help with identifications, and the Landmark Trust, Maldon District Council, Essex Wildlife Trust and Natural England for providing permission to sample at Two Tree Island, Maldon, Goldhanger and Lundy.

Conflict of interest: The authors have no conflicts of interest to declare.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.

Table S1. Accession numbers of *Wolbachia* and *Spiroplasma* genomes used for BUSCO.

Table S2. Partition schemes and models used for phylogenetic analyses.

Figure S1. *cifA* and *cifB* gene trees.

DATA AVAILABILITY

The data underlying this article are available in the GenBank Nucleotide Database at https://ncbi.nlm.nih.gov/ under BioProject PRJNA860916 and genome assemblies can be accessed with accession codes CP102058 and CP102059.

REFERENCES

Abascal F, Zardoya R, Telford MJ. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research* 2010;**38**:W7–13.

Alexeeva I, Elliott EJ, Rollins S, *et al.* Absence of *Spiroplasma* or other bacterial 16S rRNA genes in brain tissue of hamsters with scrapie. *Journal of Clinical Microbiology* 2006; **44**:91–7.

Altschul SF, Madden TL, Schaffer AA, *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 1997;**25**:3389–402.

Arbea J. Las especies del grupo de *Anurida maritima* (Guerin, 1839) (Collembola: Neanuridae) en la peninsula Iberica. *Zapateri: revista aragonesa de entomología* 2001;**9**:37–42.

Bağcı C, Patz S, Huson DH. DIAMOND+MEGAN: fast and easy taxonomic and functional analysis of short and long microbiome sequences. *Current Protocols* 2021;**1**:e59.

Ballinger MJ, Gawryluk RMR, Perlman SJ. Toxin and genome evolution in a *Drosophila* defensive symbiosis. *Genome Biology and Evolution* 2018;**11**:253–262.

Ballinger MJ, Perlman SJ. Generality of toxins in defensive symbiosis: ribosome-inactivating proteins and defense against parasitic wasps in *Drosophila*. *PLoS Pathogens* 2017;**13**:e1006431.

Ballinger MJ, Perlman SJ. The defensive *Spiroplasma*. *Current Opinion in Insect Science* 2019;**32**:36–41.

Bastian FO. The case for involvement of *Spiroplasma* in the pathogenesis of transmissible spongiform encephalopathies. *Journal of Neuropathology & Experimental Neurology* 2014;**73**:104–114.

Bing XL, Zhao DS, Sun JT, *et al.* Genomic analysis of *Wolbachia* from *Laodelphax striatellus* (Delphacidae, Hemiptera) reveals insights into its 'Jekyll and Hyde' mode of infection pattern. *Genome Biology and Evolution* 2020;**12**:3818–31.

Bordenstein S. Symbiosis and the origin of species. In: Bourtzis K, Miller T, (ed.) *Insect symbiosis*. Boca Raton, Florida: CRC Press, 2003, 283–304.

Brown AMV, Wasala SK, Howe DK, *et al.* Genomic evidence for plant-parasitic nematodes as the earliest *Wolbachia* hosts. *Scientific Reports* 2016;**6**:34955.

Brownlie JC, Johnson KN. Symbiont-mediated protection in insect hosts. *Trends in Microbiology* 2009;**17**:348–54.

Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nature Methods* 2015;**12**:59–60.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009;**25**:1972–73.

Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* 2000;**17**:540–552.

Cordaux R, Bouchon D, Grève P. The impact of endosymbionts on the evolution of host sex-determination mechanisms. *Trends in Genetics* 2011;**27**:332–41.

Cornwallis C, van 't Padje A, Ellers J, *et al.* Symbiont-driven niche expansion shaped the adaptive radiation of insects. 2021. https://doi.org/10.21203/rs.3.rs-1063949/v2

Correa CC, Ballard JWO. *Wolbachia* associations with insects: winning or losing against a master manipulator. *Frontiers in Ecology and Evolution* 2016;**3**:153.

Dallai R, Fanciulli PP, Frati F. Chromosome elimination and sex determination in springtails (Insecta, Collembola). *The Journal of Experimental Zoology* 1999;**285**:215–25.

de Coster W, D'Hert S, Schultz DT, *et al.* NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 2018;**34**:2666–2669.

Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Computational Biology* 2018;**7**:e1002195.

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 2004;**32**:1792–97.

Eleftherianos I, Atri J, Accetta J, *et al.* Endosymbiotic bacteria in insects: guardians of the immune system? *Frontiers in Physiology* 2013;**4**:46.

Garcia-Arraez MG, Masson F, Escobar JCP, *et al.* Functional analysis of RIP toxins from the *Drosophila* endosymbiont *Spiroplasma poulsonii*. *BMC Microbiology* 2019;**19**:46.

Gasparich GE, Whitcomb RF, Dodge D, *et al.* The genus *Spiroplasma* and its non-helical descendants: phylogenetic classification, correlation with phenotype and roots of the *Mycoplasma mycoides* clade. *International Journal of Systematic and Evolutionary Microbiology* 2004;**54**:893–918.

Gavotte L. *Origine, evolution et epidemiologie des parasites intracellulares viraux et bacteriens. Microbiologie et parasitologie*. Muséum National d'Histoire Naturelle. UMR ISEM - Institut des Sciences de l'Evolution de Montpellier. 2018. https://hal.science/tel-02303297

Guindon S, Dufayard JF, Lefort V, *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* 2010;**59**:307–21.

Harne S, Gayathri P, Béven L. Exploring *Spiroplasma* biology: opportunities and challenges. *Frontiers in Microbiology* 2020;**11**:589279.

Harumoto T, Lemaitre B. Male-killing toxin in a bacterial symbiont of *Drosophila*. *Nature* 2018;**557**:252–5.

Hoskins JL, Janion-Scheepers C, Chown SL, *et al.* Growth and reproduction of laboratory-reared neanurid Collembola using a novel slime mould diet. *Scientific Reports* 2015;**5**:11957.

Huang YT, Liu PY, Shih PW. Homopolish: a method for the removal of systematic errors in nanopore sequencing by homologous polishing. *Genome Biology* 2021;**22**:95.

Hurst GDD, Jiggins FM. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society B: Biological Sciences* 2005;**272**:1525–34.

Huson DH, Albrecht B, Bağcı C, *et al.* MEGAN-LR: new algorithms allow accurate binning and easy interactive exploration of metagenomic long reads and contigs. *Biology Direct* 2018;**13**:6.

Joosse ENG. Some observations on the biology of *Anurida maritima* (Guérin), (collembola). *Zeitschrift für Morphologie und Ökologie der Tiere* 1966;**57**:320–8.

Kalyaanamoorthy S, Minh BQ, Wong TKF, *et al.* ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 2017;**14**:587–89.

Katoh K, Asimenos G, Toh H. Multiple alignment of DNA sequences with MAFFT. Posada D, (ed.) *Methods in molecular biology, Vol.* **537**. Totowa, New Jersey: Humana Press, 2009:39–64.

Kodandaramaiah U, Simonsen TJ, Bromilow S, *et al.* Deceptive single-locus taxonomy and phylogeography: *Wolbachia*-associated divergence in mitochondrial DNA is not reflected in morphology and nuclear markers in a butterfly species. *Ecology and Evolution* 2013;**3**:5167–76.

Kolmogorov M, Yuan J, Lin Y, *et al.* Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology* 2019;**37**:540–46.

Konecka E, Olszanowski Z. *Wolbachia* Supergroup E found in *Hypochthonius rufulus* (Acari: Oribatida) in Poland. *Infection, Genetics and Evolution* 2021;**91**:104829.

Konecka E, Olszanowski Z, Koczura R. *Wolbachia* of phylogenetic Supergroup E identified in oribatid mite *Gustavia microcephala* (Acari: Oribatida). *Molecular Phylogenetics and Evolution* 2019;**135**:230–5.

Koren S, Walenz BP, Berlin K, *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research* 2017;**27**:722–36.

Lee MD. GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics (Oxford, England)* 2019;**35**:4162–64.

Lefoulon E, Bain O, Makepeace BL, *et al.* Breakdown of coevolution between symbiotic bacteria *Wolbachia* and their filarial hosts. *PeerJ* 2016;**4**:e1840.

LePage DP, Metcalf JA, Bordenstein SR, *et al.* Prophage WO genes recapitulate and enhance *Wolbachia*-induced cytoplasmic incompatibility. *Nature* 2017;**543**:243–7.

Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;**34**:3094–100.

Li H, Handsaker B, Wysoker A, *et al*; 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics (Oxford, England)* 2009;**25**:2078–79.

Liang T, Feng Q, Wu T, *et al.* Use of oxytetracycline for the treatment of tremor disease in the Chinese mitten crab *Eriocheir sinensis*. *Diseases of Aquatic Organisms* 2009;**84**:243–50.

Lindsey ARI, Rice DW, Bordenstein SR, *et al.* Evolutionary genetics of cytoplasmic incompatibility genes *cifA* and *cifB* in prophage WO of *Wolbachia*. *Genome Biology and Evolution* 2018;**10**:434–51.

Lo N, Casiraghi M, Salati E, *et al.* How many *Wolbachia* supergroups exist? *Molecular Biology and Evolution* 2002;**19**:341–6.

Ma Y, Chen WJ, Li ZH, *et al.* Revisiting the phylogeny of *Wolbachia* in Collembola. *Ecology and Evolution* 2017;**7**:2009–17.

Martin SH, Singh KS, Gordon IJ, *et al.* Whole-chromosome hitchhiking driven by a male-killing endosymbiont. *PLoS Biology* 2020;**18**:e3000610.

Martinez J, Klasson L, Welch JJ, *et al.* Life and death of selfish genes: comparative genomics reveals the dynamic evolution of cytoplasmic incompatibility. *Molecular Biology and Evolution* 2021;**38**:2–15.

Massey JH, Newton ILG. Diversity and function of arthropod endosymbiont toxins. *Trends in Microbiology* 2022;**30**:185–98.
McMeechan FK, Manica A, Foster WA. Rhythms of activity and foraging in the intertidal insect *Anurida maritima*: coping with the tide. *Journal of the Marine Biological Association of the United Kingdom* 2000;**80**:189–90.
Misof B, Liu S, Meusemann K, *et al.* Phylogenomics resolves the timing and pattern of insect evolution. *Science* 2014;**346**:763–7.
Moran N, McCutcheon J, Nakabachi A. Genomics and evolution of heritable bacterial symbionts. *Annual Review of Genetics* 2009;**42**:165–90.
Moran NA, Baumann P. Bacterial endosymbionts in animals. *Current Opinion in Microbiology* 2000;**3**:270–5.
Mouches C, Bové JM, Tully JG, *et al. Spiroplasma apis*, a new species from the honey-bee *Apis mellifera*. *Annales de l'Institut Pasteur/ Microbiologie* 1983;**134**:383–97.
Nguyen LT, Schmidt HA, von Haeseler A, *et al.* IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 2015;**32**:268–74.
Nunan LM, Lightner DV, Oduori MA, *et al. Spiroplasma penaei* sp. nov., associated with mortalities in *Penaeus vannamei*, Pacific white shrimp. *International Journal of Systematic and Evolutionary Microbiology* 2005;**55**:2317–22.
Perlmutter JI, Bordenstein SR, Unckless RL, *et al.* The phage gene *wmk* is a candidate for male killing by a bacterial endosymbiont. *PLoS Pathogens* 2019;**15**:e1007936.
Perlmutter JI, Meyers JE, Bordenstein SR. A single synonymous nucleotide change impacts the male-killing phenotype of prophage WO gene *wmk*. *eLife* 2021;**10**:e67686.
Pike N, Kingcombe R. Antibiotic treatment leads to the elimination of *Wolbachia* endosymbionts and sterility in the diplodiploid collembolan *Folsomia candida*. *BMC Biology* 2009;**7**:54.
Ritter S, Michalski SG, Settele J, *et al. Wolbachia* infections mimic cryptic speciation in two parasitic butterfly species, *Phengaris teleius* and *P. nausithous* (Lepidoptera: Lycaenidae). *PLoS One* 2013;**8**:e78107.
Scholz M, Albanese D, Tuohy K, *et al.* Large scale genome reconstructions illuminate *Wolbachia* evolution. *Nature Communications* 2020;**11**:5235.
Shen W, Xiong J. *TaxonKit: a cross-platform and efficient NCBI taxonomy toolkit*. *Bioinformatics* 2019;**48**:844–50.
Shropshire JD, On J, Layton EM, *et al.* One prophage WO gene rescues cytoplasmic incompatibility in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences* 2018;**115**:4987–91.
Simão FA, Waterhouse RM, Ioannidis P, *et al.* BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 2015;**31**:3210–12.
Stouthamer R, Breeuwer JAJ, Hurst GDD. *Wolbachia pipientis*: microbial manipulator of arthropod reproduction. *Annual Review of Microbiology* 1999;**53**:71–102.
Sudakaran S, Kost C, Kaltenpoth M. Symbiont acquisition and replacement as a source of ecological innovation. *Trends in Microbiology* 2017;**25**:375–90.
Sun X, Bedos A, Deharveng L. Unusually low genetic divergence at COI barcode locus between two species of intertidal *Thalassaphorura* (Collembola: Onychiuridae). *PeerJ* 2018;**6**:e5021.
Tange O. GNU Parallel 20210822 ('Kabul'). 2021. https://doi.org/10.5281/zenodo.5233952
Tatusova T, DiCuccio M, Badretdin A, *et al.* NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research* 2016;**44**:6614–24.
Timmermans MJTN, Arbea JI, Campbell G, *et al.* Mitochondrial genome divergence supports an ancient origin of circatidal behaviour in the *Anurida maritima* (Collembola: Neanuridae) species group. *Organisms Diversity & Evolution* 2022;**22**:131–40.
Timmermans MJTN, Ellers J. *Wolbachia* endosymbiont is essential for egg hatching in a parthenogenetic arthropod. *Evolutionary Ecology* 2009;**23**:931–42.
Trifinopoulos J, Nguyen LT, von Haeseler A, *et al.* W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research* 2016;**44**:W232–35.
Tully JG, Whitcomb RF, Rose DL, *et al. Spiroplasma mirum*, a new species from the rabbit tick (*Haemaphysalis leporispalustris*). *International Journal of Systematic Bacteriology* 1982;**32**:92–100.
Vandekerckhove TTM, Watteyne S, Willems A, *et al.* Phylogenetic analysis of the 16S rDNA of the cytoplasmic bacterium *Wolbachia* from the novel host *Folsomia candida* (Hexapoda, Collembola) and its implications for wolbachial taxonomy. *FEMS Microbiology Letters* 1999;**180**:279–86.
Vera-Ponce León A, Dominguez-Mirazo M, Bustamante-Brito R, *et al.* Functional genomics of a *Spiroplasma* associated with the carmine cochineals *Dactylopius coccus* and *Dactylopius opuntiae*. *BMC Genomics* 2021;**22**:240.
Wang W, Cui W, Yang H. Toward an accurate mechanistic understanding of *Wolbachia*-induced cytoplasmic incompatibility. *Environmental Microbiology* 2022;**24**:4519–32.
Wang W, Gu W, Ding Z, *et al.* A novel *Spiroplasma* pathogen causing systemic infection in the crayfish *Procambarus clarkii* (Crustacea: Decapod), in China. *FEMS Microbiology Letters* 2005;**249**:131–7.
Wang W, Wen B, Gasparich GE, *et al.* A spiroplasma associated with tremor disease in the Chinese mitten crab (*Eriocheir sinensis*). *Microbiology* 2004;**150**:3035–40.
Weinert LA, Araujo-Jnr EV, Ahmed MZ, *et al.* The incidence of bacterial endosymbionts in terrestrial arthropods. *Proceedings of the Royal Society B: Biological Sciences* 2015;**282**:20150249.
Werren JH. Biology of *Wolbachia*. *Annual Review of Entomology* 1997;**42**:587–609.
Werren JH, Baldo L, Clark ME. *Wolbachia*: master manipulators of invertebrate biology. *Nature Reviews Microbiology* 2008;**6**: 741–51.
Werren JW, Zhang W, Guo LR. Evolution and phylogeny of *Wolbachia*: reproductive parasites of arthropods. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 1995;**261**:55–63.
Williamson DL, Sakaguchi B, Hackett KJ, *et al. Spiroplasma poulsonii* sp. nov., a new species associated with male-lethality in *Drosophila willistoni*, a neotropical species of fruit fly. *International Journal of Systematic and Evolutionary Microbiology* 1999;**49**:611–8.