'*Candidatus* Tisiphia' is a widespread Rickettsiaceae symbiont in the mosquito *Anopheles plumbeus* (Diptera: Culicidae)

Authors

Helen R. Davison[1], Jessica Crozier[1], Stacy Pirro[2], Doreen Werner[3], Gregory D.D. Hurst[1]

Affiliations

1. Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Crown Street Liverpool L69 7ZB UK
2. Iridian Genomes, Bethesda, MD, USA
3. Leibniz Centre for Agricultural Landscape Research (ZALF), 15374 Müncheberg, Germany

Corresponding Author

Helen R. Davison - hlhdavi5@liverpool.ac.uk

Author contributions

Conceptualisation, supervision, methodology and funding acquisition – HRD and GDDH
Resources – HRD, DW, and SP
Investigation – JC and HRD
Analysis, Visualisation, and Data Curation – HRD
Original Draft Preparation – HRD, GDDH

Data availability statement

Genome bioproject accessions: PRJNA694375 and PRJNA901697. PCR sequences are deposited in accessions OQ512853-OQ512860

Supplementary data is available online at: https://doi.org/10.6084/m9.figshare.21865416.v1

Acknowledgments

Funding statement

NERC ACCE Doctoral Training Studentship NE/L002450/1

Genome sequencing was supported by Iridian Genomes, grant# IRGEN_RG_2021-1345
Genomic Studies of Eukaryotic Taxa

Ethical approval statement

All samples were collected in accordance with ethical standards for insect collection.

Conflicts of interest

The authors declare no conflicts of interest.

Abstract

Symbiotic bacteria alter host biology in numerous ways, including the ability to reproduce or vector disease. Deployment of symbiont control of vector borne disease has focused on *Wolbachia* interactions with *Aedes* and is hampered in *Anopheles* by a lack of compatible symbioses. Previous screening found the symbiont '*Ca.* Tisiphia' in *Anopheles plumbeus*, an aggressive biter and potential secondary vector of malaria parasites and West Nile virus. We screen *An. plumbeus* samples collected over a ten-year period across Germany and use climate databases to assess environmental influence on incidence. We find a 95% infection rate that does not fluctuate with broad environmental factors. Microscopy suggests the infection is maternally inherited based on strong localisation throughout the ovaries. Finally, we assemble a high-quality draft genome of '*Ca.* Tisiphia' to explore its phylogeny and potential metabolism. This strain is closely related to those found in *Culicoides* midges and shows similar patterns of metabolic potential. *An. plumbeus* provides a viable avenue of symbiosis research in anopheline mosquitoes, which to date have one other proven infection of a heritable symbiont. Additionally, it provides future opportunity to study the impacts of '*Ca.* Tisiphia' on natural and transinfected hosts, especially in relation to reproductive fitness and vector efficiency.

Introduction

Bacterial symbionts in insects form vital components of their host's biology, ecology, and evolution. They are known to influence how insects reproduce, how they respond to environmental stress, and they interact with pathogens (Dunbar *et al.*, 2007; Himler *et al.*, 2011; Vega, Arribére and Castro-Vazquez, 2012; Hendry, Hunter and Baltrus, 2014; Xie *et al.*, 2014; Hayashi *et al.*, 2016). Several species of symbionts are vertically inherited from one generation to the next, usually through the maternal germline, and may become intrinsically linked with their host physiology, metabolism, and development (Buchner, 1965; Zchori-Fein, Borad and Harari, 2006; Moran, McCutcheon and Nakabachi, 2008; Kremer *et al.*, 2009; Giorgini *et al.*, 2010). Most importantly, symbionts have been deployed in the control of vector populations and vector competence (Hoffmann *et al.*, 2011).

Success in symbiont-mediated disease control has been restricted to species from the genus *Aedes*. Transinfection with *Wolbachia* from a drosophilid fly has been successfully used to alter vector competence and lower risk of catching dengue fever from *Aedes aegypti* (Linnaeus, 1762) in endemic areas (Hoffmann *et al.*, 2011; Walker *et al.*, 2011; Pereira *et al.*, 2018). However, important vectors like *Anopheles* are rarely naturally infected with *Wolbachia*, and species within the group are commonly unreceptive to artificial *Wolbachia* infections (Hughes *et al.*, 2014). In *Anopheles* mosquitoes, for instance, there is a single well-established case of natural *Wolbachia* infection (Walker *et al.*, 2021). Therefore, it is desirable to find potential alternatives that are either more capable of surviving transinfection or alter vector competence in the native host species.

We previously detected the symbiont '*Ca.* Tisiphia' (= Torix group *Rickettsia*) in *Anopheles plumbeus* (Stephens, 1828) in the UK (Pilgrim *et al.*, 2021). *An. plumbeus* is broadly distributed across Europe and is an indiscriminate biter. It is also capable of transmitting West Nile virus and malaria parasites, although these pathogens do not natively occur in the majority of the mosquito species´ known distribution range, and vector competence has only been tested in the laboratory setting (Bueno-Marí and Jiménez-Peydró, 2011; Dekoninck *et al.*, 2011; Schaffner *et al.*, 2012). It has been highlighted as a species that could act as a secondary vector of "tropical" disease agents as changing climate causes their northward spread and their associated primary hosts like *Aedes albopictus* (Skuse, 1894) (Schaffner *et al.*, 2012; Heym *et al.*, 2017).

'*Candidatus* Tisiphia' (= Torix group *Rickettsia*) appear to be particularly associated with hosts deriving from wet or aquatic environments and may originate from symbionts of freshwater ciliates (Driscoll *et al.*, 2013; Schrallhammer *et al.*, 2013; Kang *et al.*, 2014). Infection with this symbiont occur in a broad range of invertebrates, from annelids to gastropods to arthropods (Pilgrim *et al.*, 2021), as well as in algae (Hollants *et al.*, 2013) and amoebae (Dyková *et al.*, 2013). Their relatives in *Rickettsia* are capable of nutritional symbioses, protecting against fungal infections and reproductive manipulation (Hurst *et al.*, 1994; Giorgini *et al.*, 2010; Hendry, Hunter and Baltrus, 2014; Bodnar *et al.*, 2018). However, the known effects of '*Ca.* Tisiphia' are limited to an association with increased host size in *Torix* leeches, and weak impacts on fecundity in *Cimex lectularius* bedbugs (Kikuchi and Fukatsu, 2005; Thongprem *et al.*, 2020). There is no observed congruence of host and symbiont phylogeny, indicating that host shifts occur commonly and that long-standing associations with species are rare. External influence such as temperature or natural enemies can also influence the prevalence of symbionts in host populations (Cass *et al.*, 2016; Corbin *et al.*, 2017; Leclair *et al.*, 2017).

Here we use PCR assays to establish the extent of '*Ca.* Tisiphia' infection in *An. plumbeus* mosquitoes across Germany collected by DW and through citizen science initiatives, and assess associations with temperature, precipitation, and forest type. We also sequence and assemble a high-quality draft genome for the symbiont '*Ca.* Tisiphia' and provide evidence of vertical transmission of the symbiont through the maternal germline through FISH imaging. The symbiont genome is examined through bioinformatics approaches to establish potential nutritional or protective symbioses.

Experimental Procedures

Collection of *Anopheles plumbeus* (Stephens, 1828)

Two hundred and fifty five *An. plumbeus* specimens were collected from 2012-2021 across Germany by DW and citizen volunteers as part of the mosquito atlas (Mückenatlas) project (Werner *et al.*, 2014). These were stored in 70% ethanol or dry (see supplementary materials for storage and exact geographic information). *Post hoc* analysis indicated storage method did not affect detection of symbionts by PCR assay.

Specimens were also collected by HD as larvae and raised to adults in water collected from their larval pools. These specimens were either killed by flash freezing in liquid nitrogen prior

to genomic DNA extraction, or in 4% paraformaldehyde solution prior for fluorescence imaging.

DNA extraction and PCR screening of *Anopheles plumbeus* for '*Ca*. Tisiphia'

Promega Wizard® Genomic DNA Purification kit was used for DNA preparation. DNA quality was then checked with a combination of HCO/C1J primers HCO_2198 (5'-TAA ACT TCA GGG TGA CCA AAA AAT CA-3')/CIJ_1718 (5'-GGA GGA TTT GGA AAT TGA TTA GT-3') (Folmer *et al.*, 1994; Hajibabaei *et al.*, 2005; Siozios *et al.*, 2020). *Ca.* Tisiphia presence was assessed with a PCR assay amplifying the 320-bp region of the 17 kDa OMP gene Ri17kD_F (5'-TCTGGCATGAATAAACAAGG-3')/Ri17kD_R (5'-ACTCACGACAATATTGCCC-3') (Pilgrim *et al.*, 2017). PCR conditions used were as follows: 95 °C for 5 min, followed by 35 cycles of denaturation (94 °C, 30 s), annealing (54 °C, 30 s) and extension (72 °C, 120 s).

A selection of '*Ca.* Tisiphia' amplicons from positive samples across time and space were Sanger-sequenced through Eurofins barcode service and identity confirmed by comparing the sequences to the NCBI database via BLAST homology searches. These sequences are deposited in accessions OQ512853-OQ512860.

Association of symbiont prevalence with geographic and climatic information

Annual average monthly temperature and precipitation data were retrieved for each sample's coordinate and year from TerraClim (Abatzoglou *et al.*, 2018) which has a spatial resolution of ~4-km (1/24th degree). Forest cover data was retrieved from Copernicus land datasets for 2018 (European Union, 2018) and raster data for forest type extracted in QGIS 3.16 (QGIS.org, 2020) within a 3km radius of each sample location. *Anopheles plumbeus* has historically been recorded to have a maximum flight range of up to 13km (Becker *et al.*, 2010). However, this is based on one single study from 1925 and is not verified by other sources. As such we chose an estimated range of 3km based on the average flight ranges other anopheline mosquito species (Becker *et al.*, 2010; Verdonschot & Besse-Lototskaya, 2014). Scikitlearn's standard scaler (Pedregosa *et al.*, 2011) was applied to data before performing a generalised linear model with a binomial logit link function on data with the following formula:

$$Infected \sim tasmin + tasmax + precip + forest\_ratio$$

All statistics and geographic inferences were carried out in Python with the packages Statsmodel and Scikit-learn (Rossum and Drake, 2009; Seabold and Perktold, 2010; Pedregosa

et al., 2011). QGIS 3.16 was used to produce maps and extract raster data for forest types before passing it to Python for analysis (QGIS.org, 2020). All other figures were produced with Matplotlib and Seaborn (Hunter, 2007; Waskom and Seaborn development team, 2020).

Fluorescence in situ microscopy (FISH)

Reproductive organs of a single female and a single male were dissected and incubated in cold 4% paraformaldehyde for 3 hours, agitated gently every 30 minutes, then washed with cold PBS for 5 minutes two times. Tissue was stained with Hoescht 33342 (that binds double-stranded DNA) for 30 minutes at room temperature, then hybridised overnight at room temperature with hybridisation buffer (5X SSC, 0.01% SDS, 30% formamide) and 5'-CCATCATCCCCTACTACA-(ATTO 633)-3' oligonucleotide probe specific to '*Ca*. Tisiphia' 16S rRNA (Pilgrim *et al.*, 2017). Hybridised tissue was washed in wash buffer (5X SSC, 0.01% SDS) at 48°C for 60 minutes with gentle shaking every 20 minutes. Samples were then mounted in Vectashield. Images were taken with a ZEIS LSM 880 confocal microscope through ZEIS Zen black, and final images were annotated in Inkscape Ver 1.2 (Inkscape Project, 2020).

De novo sequencing, assembly, and annotation.

A combination of short and long read sequencing was used to construct scaffolds for the '*Ca.* Tisiphia' genome. For short reads, Iridian Genomes extracted and processed DNA of one male for Illumina sequencing deposited under bioproject accession PRJNA694375. The short reads of *An. plumbeus* were cleaned with Trimmomatic 0.36 (Bolger, Lohse and Usadel, 2014) and quality checked with FASTQ (Babraham Bioinformatics, 2019). For long reads, genomic DNA from one male was extracted with Qiagen Genomic-tip for ultra-low PacBio sequencing carried out by the Centre for Genomic Research, University of Liverpool. Long read sequences are deposited under bioproject accession number PRJNA901697.

A combination of long and short reads were used to assemble as complete a genome for '*Ca.* Tisiphia' as possible. First, the '*Ca*. Tisiphia' genome was identified in the Illumina short reads and assembled through Minmap2, MEGAHIT and MetaBAT2 as per the pipeline used in Davison *et al.* (2022). Second, PacBio HiFi long read sequences were assembled using Flye 2.9.1-b1780 with the '-meta' flag to improve sensitivity for low coverage reads. Third, the long read assembly was queried against a local blast database of *Rickettsia* and '*Ca*. Tisiphia' genomes (including the Illumina assembly from the first step) to identify sequences belonging to this strain of '*Ca*. Tisiphia'. Lastly, the long read assembly was polished with the Illumina

reads using Polypolish v0.5.0 (Wick and Holt, 2022) with default settings to give 23 final scaffolds.

Phylogeny and metabolic predictions

Annotation of the final assembly was carried out with InterProScan v5 (Jones *et al.*, 2014). Metabolic pathway prediction for presence and completion was carried out through Anvi'o 7 using KEGG kofams and COG20 (Aramaki *et al.*, 2020; Eren *et al.*, 2021; Galperin *et al.*, 2021). NRPS pathways were investigated with AntiSMASH 6.0 (Blin *et al.*, 2021).

The '*Ca*. Tisiphia' strain for *An. plumbeus* was compared to the other existing '*Ca*. Tisiphia' genomes through Anvi'o 7. A core genome consisting of 205 gene clusters that contain a total of 3690 genes was found through Anvio-7. Phylogenies were estimated from single copy gene clusters with IQTREE 2.2.0.3 using Model Finder Plus and with 1000 ultrafast bootstraps (Kalyaanamoorthy *et al.*, 2017; Hoang *et al.*, 2018; Minh *et al.*, 2020). The model selected by Model Finder Plus is Q.plant+F+R4. A supporting phylogeny to confirm the identity of 17 kDa OMP genes was produced with the model TIM2+F.

Results and Discussion

Distribution and predicted environment

'Ca. Tisiphia' was observed to infect *An. plumbeus* across all sites examined in Germany, with 95% specimens positive on PCR assay. The few negative specimens were found to mostly occur in the southeast of the country (Figure 1 and Supplementary Figure 1). The infection seems to be stable, and there is no evidence of frequency change over time, with samples from all years spanning 2012 to 2022 displaying similar rates of infection (Supplementary Figure 1, Supplementary Table S1, Supplementary Figure 2).

There is no clear evidence of an influence on '*Ca*. Tisiphia' infection rates in *An. plumbeus* caused by average minimum or maximum temperature, precipitation or forest types (Figure 2 and Supplementary Figure 2 and 3). While there appears to be a significant effect of precipitation on the number of uninfected individuals, this could be an artifact of increased water availability leading to more mosquitoes and thus a higher chance of detecting rarer uninfected individuals (Supplementary Figure 2). No variation is unsurprising as it appears to be a very high prevalence infection. We also acknowledge that using climate databases to retroactively find data is not as accurate as field measurements. However, results agree with

previous field observations of *Rickettsia* infection in *Acyrthosiphon pisum* in Japan (Tsuchida *et al.*, 2002). We chose to use the high resolution TerraClim database, but this may still mask small differences in microenvironments are limited to mostly abiotic data. We encourage future symbiosis research to consider environmental measurements to describe the ecology of these organisms more comprehensively.

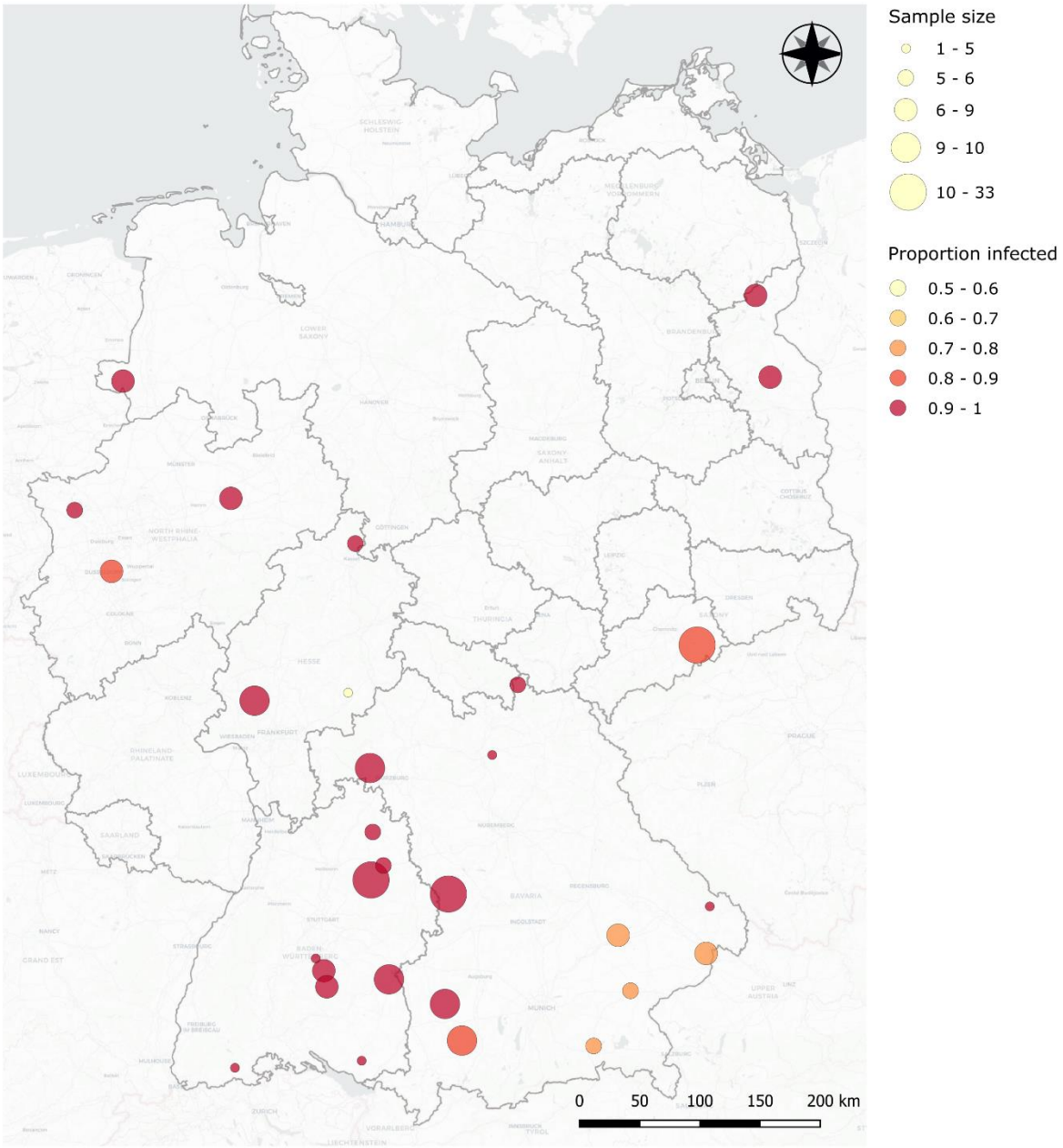
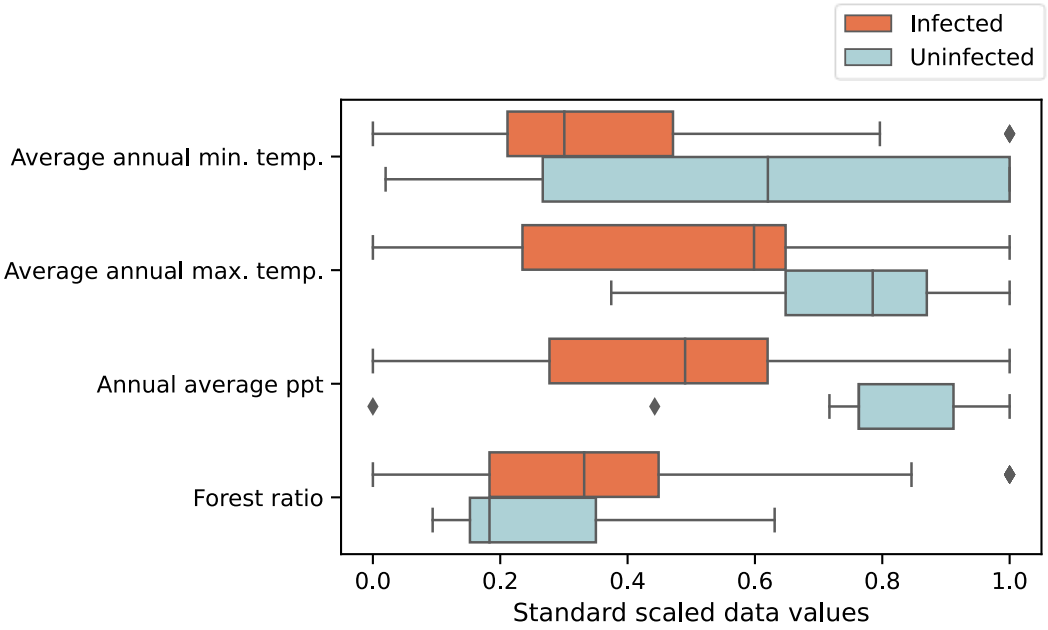


Figure 1. Map of '*Ca*. Tisiphia' infection rates across Germany where the size of the circle represents the number of individuals sampled and the colour indicates the proportion of '*Ca*. Tisiphia' infected individuals. Red = 90-100% infection to light yellow = 50-60% infection. Source data can be found in Supplementary Table S1.



211

212 **Figure 2. Standardised and scaled environmental data** comparing Uninfected (N=13) and Infected (N=237) by
213 environmental variable. Source data can be found in Supplementary Table S1.

214 *Phylogeny and metabolism*

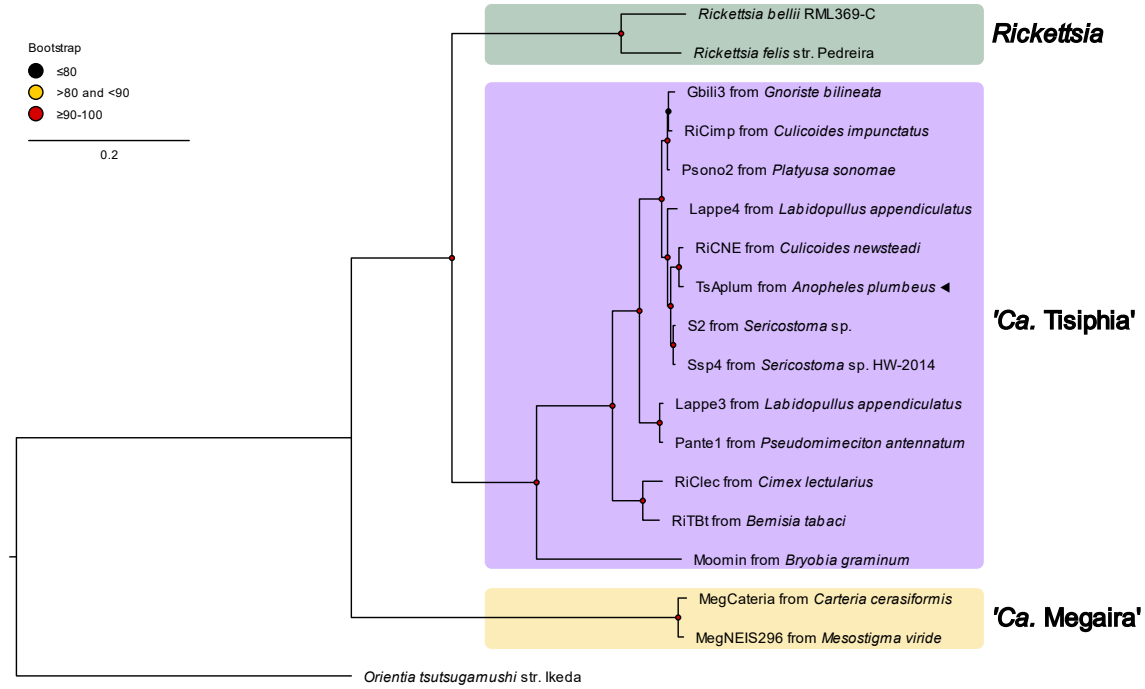
215 The bacteria sequenced from *An. plumbeus* is most closely related to a '*Ca*. Tisiphia' found in
216 the biting midge *Culicoides newsteadi* (Figure 3). All infections tested are the same strain of
217 '*Ca.* Tisiphia' (Supplementary Figure 4). General features of both genomes are consistent with
218 other '*Ca*. Tisiphia' (Table 1 and Supplementary Data S2-S4); TsAplum has a single full set of
219 rRNAs (16S, 5S and 23S), and GC content is ~33%. It also has several repeat domains (Table 1)
220 which are associated with protein-protein interactions and are prevalent in *Wolbachia*
221 symbionts (Siozios *et al.*, 2013; Rice, Sheehan and Newton, 2017).

222

223 *Table 1. Summary of the genome assembly for TsAplum.*

Strain Name	TsAplum	224
Symbiont genome accession	SAMN31737641	
Host	*Anopheles plumbeus*	
Raw reads accession	Pacbio SRR22298143, Illumina SRR13516402	
Total nucleotides	1,622,210	
Contigs	31	
GC content	32.82%	
N50	62798	
Number of CDS	1701	
Avg. CDS length (bp)	788	
Coding density	82.57%	
rRNAs	1 x 5S, 1 x 16S, 1 x 23S	
tRNAs	31	
ORFs with Ankyrin repeat domains	3	
ORFs with Leucine rich repeats	1	
ORFs with Tetratricopeptide repeats	6	

225



226 **Figure 3. Genome wide phylogeny of '*Ca*. Tisiphia' and '*Ca*. Megaira'.** Maximum likelihood (ML) phylogeny
227 constructed from 205 single copy gene clusters that contain a total of 3690 genes. New genomes are
228 indicated by ◄ and bootstrap values based on 1000 replicates are indicated with coloured circles (red = 91-
229 100, yellow = 81-90, black <= 80). Accession numbers for each genome are available in Supplementary Table
230 S2.

Overall '*Ca*. Tisiphia' found in *An. plumbeus* mirrors the metabolic potential found in other members of its genus (Figure 4 and Figure 5). It does not have any obvious metabolic pathway that would contribute to nutritional symbiosis such as B vitamin production nor any NRPS/PKS system for small molecule synthesis (Supplementary Tables S3 and S4). It does have several toxin/anti-toxin systems as well as secretion pathways Tat, Sec, VirB (Type IV), all of which are essential in various symbiont-host interactions (Masui, Sasaki and Ishikawa, 2000; Meloni *et al.*, 2003; Wu *et al.*, 2004; Dale and Moran, 2006; Tseng, Tyler and Setubal, 2009). Additionally, it has a number of ORFs containing ankyrin- and leucine-rich repeats which are thought to be important in interactions with cognate eukaryotic proteins (Siozios *et al.*, 2013; Rice, Sheehan and Newton, 2017). Thus, the genome itself, whilst firmly placing the symbiont in the context of the genus and identifying relatedness to other strains, does not raise obvious hypotheses about the impact of infection on the host. Phenotype studies are required to properly assess the influence of this bacteria on its host. Key studies would address the factors driving the spread of the symbiont into the population (testing for beneficial aspects of infection, cytoplasmic incompatibility, and paternal inheritance) and impacts on viral infection and transmission outcomes.

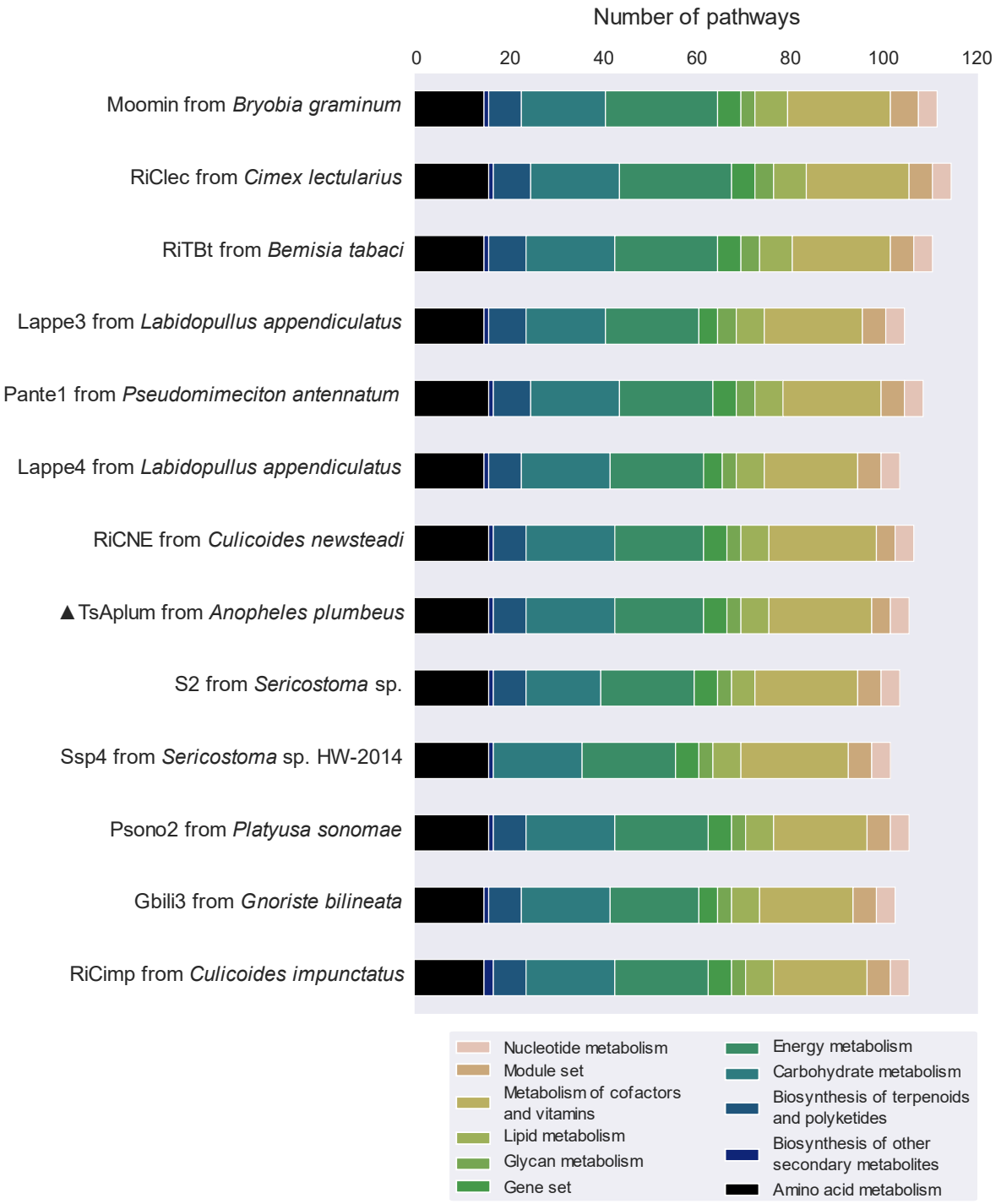


Figure 4. KEGG module distribution in '*Ca*. Tisiphia'. The number of pathways found per genome annotated by KEGG module category for '*Ca*. Tisiphia'. Full metadata can be found in Supplementary Data 1. ▲ indicates the genome assembled in this study. Full metadata can be found in Supplementary Tables S3 and S4.

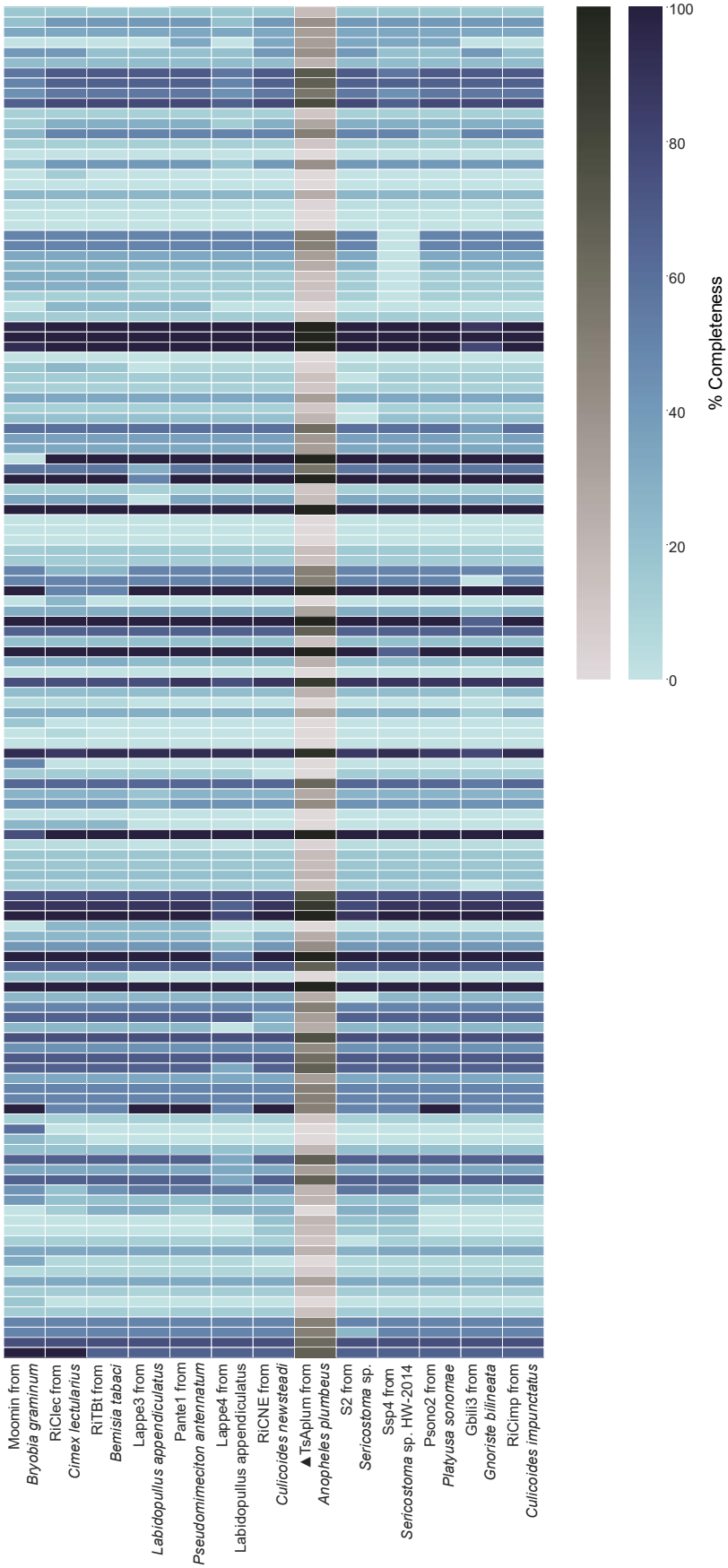


Figure 5. Predicted completeness of KEGG kofam metabolic pathways across 'Ca. Tisiphia'. The genome assembled in this chapter is coloured grey and indicated with ▲. Full metadata can be found in Supplementary Tables S3 and S4.

FISH imaging

'Ca. Tisiphia' is observed in oocytes and oviduct branches but was not detected in testes (Figure 6 versus Supplementary Figure 5). Localisation and clear polarity of the infection in ovaries strongly suggest that this is a maternally inherited infection (Figure 6). The bacteria cluster around the oocyte of the primary follicles as well as in the lateral ducts and secondary follicles. In *Drosophila melanogaster*, *Wolbachia* is similarly polarised to one end of the primary follicles to the oocyte (Ferree *et al.*, 2005), and in *Proechinophthirus fluctus*, their endosymbionts *Sodalis* appears to use the lateral oviducts to access the ovaries (Boyd *et al.*, 2016).

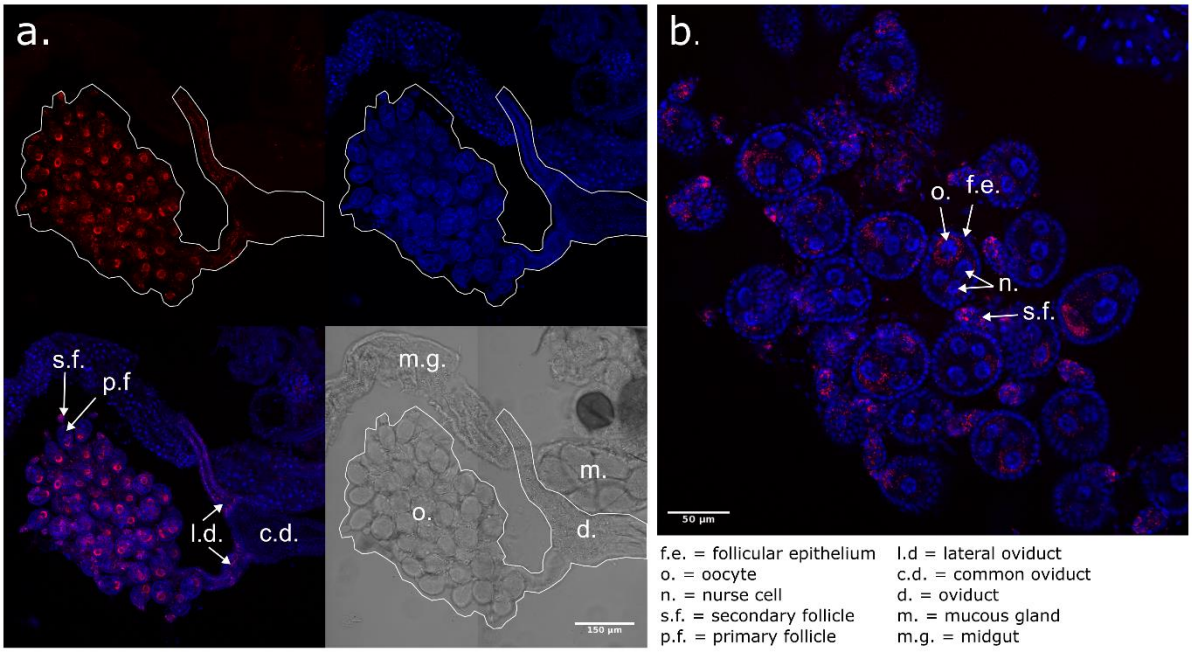


Figure 6. Fluorescence in situ microscopy images of *Anopheles plumbeus* ovaries infected with '*Ca*. Tisiphia'. Red shows '*Ca*. Tisiphia' stained with ATTO-633, blue are host nuclei stained with Hoechst. Panels show a) the whole female reproductive organ outlined in white and a breakdown of each light channel and b) a close up of the ovaries showing localised infection within the primary and secondary follicles. White bars indicate a) 150 micrometres and b) 50 micrometres.

Final conclusions

Anopheles plumbeus and its '*Ca.* Tisiphia' make a good potential model for symbioses in *Anopheles* mosquitoes as well as for '*Ca*. Tisiphia' infection more generally. Outside of *An. plumbeus*, there is only a single well substantiated case of *Wolbachia* and no other symbiont in anopheline mosquitoes (Walker *et al.*, 2021). The infection in *An. plumbeus* is clearly evidenced, likely heritable, and occurs in a species that has seen previous success as a laboratory colony (Kotter, 2005). Beyond this, *An. plumbeus* is a species of interest with the capability to carry West Nile virus and *Plasmodium* parasites (Dekoninck *et al.*, 2011;

Schaffner *et al.*, 2012). '*Ca*. Tisiphia' in *An. plumbeus* provides a viable avenue for symbiont-mediated vector modification and control to be tested in anopheline species. It is also an example of a temporally and spatially stable infection of non-pathogenic Rickettsiaceae and a good foil to fluctuating systems like Belli *Rickettsia* in *Bemisia tabaci* (Bockoven *et al.*, 2020).

Future work should also establish the effects of this symbiont in transinfection in alternative hosts alongside the native *An. plumbeus* host. Other symbionts like *Wolbachia* are known to produce functionally interesting phenotypes related to vector competence when transferred from the original host into other, naïve species (Moreira *et al.*, 2009). Alongside this, impacts on host function and physiology, and potential means of spread into natural populations would need to be assessed. A first step to establishing transinfection would be to isolate the '*Ca.* Tisiphia' infection into cell culture, which would also represent an important community resource for onward study.

In summary, '*Ca*. Tisiphia' is found in 95% of *An. plumbeus* individuals from Germany and forms a well-established, stable, and heritable infection that persists across space and time. Metabolic potential is typical of similar symbiotic bacteria species, and we find no evidence of large-scale environmental factors influencing rates of infection. However, '*Ca.* Tisiphia' and *An. plumbeus* provide a unique opportunity to study the effects of a native symbiont infection in anopheline mosquitoes, as well as explore its potential use for disease mitigation in other species that cannot be infected with currently used symbionts.

References

Abatzoglou, J.T. *et al.* (2018) 'TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015', *Scientific Data*, 5(1), p. 170191. Available at: https://doi.org/10.1038/sdata.2017.191.

Aramaki, T. *et al.* (2020) 'KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold', *Bioinformatics*. Edited by A. Valencia, 36(7), pp. 2251–2252. Available at: https://doi.org/10.1093/bioinformatics/btz859.

Babraham Bioinformatics (2019) 'Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data'. Babraham Bioinformatics. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (Accessed: 20 December 2022).

Blin, K. *et al.* (2021) 'antiSMASH 6.0: improving cluster detection and comparison capabilities', *Nucleic Acids Research*, 49(W1), pp. W29–W35. Available at: https://doi.org/10.1093/nar/gkab335.

Bockoven, A.A. *et al.* (2020) 'What goes up might come down: the spectacular spread of an endosymbiont is followed by its decline a decade later', *Microbial Ecology*, 79(2), pp. 482–494. Available at: https://doi.org/10.1007/s00248-019-01417-4.

Bodnar, J.L. *et al.* (2018) 'The folA gene from the *Rickettsia* endosymbiont of *Ixodes pacificus* encodes a functional dihydrofolate reductase enzyme', *Ticks and Tick-borne Diseases*, 9(3), pp. 443–449. Available at: https://doi.org/10.1016/j.ttbdis.2017.12.013.

Bolger, A.M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: a flexible trimmer for Illumina sequence data', *Bioinformatics (Oxford, England)*, 30(15), pp. 2114–2120. Available at: https://doi.org/10.1093/bioinformatics/btu170.

Boyd, B.M. *et al.* (2016) 'Two bacterial genera, *Sodalis* and *Rickettsia*, associated with the seal louse *Proechinophthirus fluctus* (Phthiraptera: Anoplura)', *Applied and Environmental Microbiology*, 82(11), pp. 3185–3197. Available at: https://doi.org/10.1128/AEM.00282-16/SUPPL_FILE/ZAM999117155SO1.PDF.

Buchner, P. (1965) *Endosymbiosis of animals with plant microorganisms*. New York: Interscience Publishers.

Bueno-Marí, R. and Jiménez-Peydró, R. (2011) 'Anopheles plumbeus Stephens, 1828: a neglected malaria vector in Europe', *Malaria Reports*, 1(1), p. 2. Available at: https://doi.org/10.4081/malaria.2011.e2.

Cass, B.N. *et al.* (2016) 'Conditional fitness benefits of the Rickettsia bacterial symbiont in an insect pest', *Oecologia*, 180(1), pp. 169–179. Available at: https://doi.org/10.1007/s00442-015-3436-x.

Corbin, C. *et al.* (2017) 'Heritable symbionts in a world of varying temperature', *Heredity*, 118(1), pp. 10–20. Available at: https://doi.org/10.1038/hdy.2016.71.

Dale, C. and Moran, N.A. (2006) 'Molecular Interactions between Bacterial Symbionts and Their Hosts', *Cell*, 126(3), pp. 453–465. Available at: https://doi.org/10.1016/j.cell.2006.07.014.

Dekoninck, W. *et al.* (2011) 'Human-Induced Expanded Distribution of Anopheles plumbeus, Experimental Vector of West Nile Virus and a Potential Vector of Human Malaria in Belgium', *Journal of Medical Entomology*, 48(4), pp. 924–928. Available at: https://doi.org/10.1603/ME10235.

Driscoll, T. *et al.* (2013) 'Bacterial DNA sifted from the *Trichoplax adhaerens* (Animalia: Placozoa) genome project reveals a putative rickettsial endosymbiont', *Genome Biology and Evolution*, 5(4), pp. 621–645. Available at: https://doi.org/10.1093/GBE/EVT036.

Dunbar, H.E. *et al.* (2007) 'Aphid thermal tolerance is governed by a point mutation in bacterial symbionts', *PLoS Biology*, 5(5), pp. 1006–1015. Available at: https://doi.org/10.1371/journal.pbio.0050096.

Dyková, I. *et al.* (2013) 'Nuclearia pattersoni sp. n. (Filosea), a new species of amphizoic amoeba isolated from gills of roach (Rutilus rutilus), and its rickettsial endosymbiont', *Folia Parasitologica*, 50(3), pp. 161–170. Available at: https://doi.org/10.14411/fp.2003.030.

Eren, A.M. *et al.* (2021) 'Community-led, integrated, reproducible multi-omics with anvi'o', *Nature Microbiology*, 6(1), pp. 3–6. Available at: https://doi.org/10.1038/s41564-020-00834-3.

European Union (2018) *Copernicus Land Monitoring Service*, *European Environment Agency (EEA)*. Available at: https://land.copernicus.eu/pan-european/high-resolution-layers/forests/forest-type-1/status-maps/forest-type-2018 (Accessed: 31 August 2022).

Ferree, P.M. *et al.* (2005) 'Wolbachia Utilizes Host Microtubules and Dynein for Anterior Localization in the Drosophila Oocyte', *PLOS Pathogens*, 1(2), p. e14. Available at: https://doi.org/10.1371/journal.ppat.0010014.

Folmer, O. *et al.* (1994) 'DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates', *Molecular Marine Biology and Biotechnology*, 3(5), pp. 294–299.

Galperin, M.Y. *et al.* (2021) 'COG database update: focus on microbial diversity, model organisms, and widespread pathogens', *Nucleic Acids Research*, 49(D1), pp. D274–D281. Available at: https://doi.org/10.1093/nar/gkaa1018.

Giorgini, M. *et al.* (2010) '*Rickettsia* symbionts cause parthenogenetic reproduction in the parasitoid wasp *Pnigalio soemius* (hymenoptera: Eulophidae)', *Applied and Environmental Microbiology*, 76(8), pp. 2589–2599. Available at: https://doi.org/10.1128/AEM.03154-09.

Hajibabaei, M. *et al.* (2005) 'Critical factors for assembling a high volume of DNA barcodes', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1462), pp. 1959–1967. Available at: https://doi.org/10.1098/rstb.2005.1727.

Hayashi, M. *et al.* (2016) 'A Nightmare for Males? A Maternally Transmitted Male-Killing Bacterium and Strong Female Bias in a Green Lacewing Population', *PLOS ONE*. Edited by K. Bourtzis, 11(6), p. e0155794. Available at: https://doi.org/10.1371/journal.pone.0155794.

Hendry, T.A., Hunter, M.S. and Baltrus, D.A. (2014) 'The facultative symbiont *Rickettsia* protects an invasive whitefly against entomopathogenic *Pseudomonas syringae* strains', *Applied and Environmental Microbiology*, 80(23), pp. 7161–7168. Available at: https://doi.org/10.1128/AEM.02447-14.

Heym, E.C. *et al.* (2017) 'Anopheles plumbeus (Diptera: Culicidae) in Germany: updated geographic distribution and public health impact of a nuisance and vector mosquito', *Tropical Medicine & International Health*, 22(1), pp. 103–112. Available at: https://doi.org/10.1111/tmi.12805.

372 Himler, A.G. *et al.* (2011) 'Rapid Spread of a Bacterial Symbiont in an Invasive Whitefly Is Driven by
373 Fitness Benefits and Female Bias', *Science*, 332(6026), pp. 254–256. Available at:
374 https://doi.org/10.1126/science.1199410.

375 Hoang, D.T. *et al.* (2018) 'UFBoot2: Improving the Ultrafast Bootstrap approximation', *Molecular*
376 *Biology and Evolution*, 35(2), pp. 518–522. Available at: https://doi.org/10.5281/zenodo.854445.

377 Hoffmann, A.A. *et al.* (2011) 'Successful establishment of Wolbachia in Aedes populations to
378 suppress dengue transmission', *Nature*, 476(7361), pp. 454–457. Available at:
379 https://doi.org/10.1038/nature10356.

380 Hollants, J. *et al.* (2013) 'Permanent residents or temporary lodgers: Characterizing intracellular
381 bacterial communities in the siphonous green alga bryopsis', *Proceedings of the Royal Society B:*
382 *Biological Sciences*, 280(1754), pp. 1–8. Available at: https://doi.org/10.1098/rspb.2012.2659.

383 Hughes, G.L. *et al.* (2014) 'Native microbiome impedes vertical transmission of *Wolbachia* in
384 *Anopheles* mosquitoes', *Proceedings of the National Academy of Sciences*, 111(34), pp. 12498–
385 12503. Available at: https://doi.org/10.1073/pnas.1408888111.

386 Hunter, J.D. (2007) 'Matplotlib: A 2D Graphics Environment', *Computing in Science & Engineering*,
387 9(3), pp. 90–95. Available at: https://doi.org/10.1109/MCSE.2007.55.

388 Hurst, G.D.D. *et al.* (1994) 'The effect of infection with male-killing *Rickettsia* on the demography of
389 female *Adalia bipunctata* L. (two spot ladybird)', *Heredity*, 73(3), pp. 309–316. Available at:
390 https://doi.org/10.1038/hdy.1994.138.

391 Inkscape Project (2020) 'Inkscape'. Available at: https://inkscape.org.

392 Jones, P. *et al.* (2014) 'InterProScan 5: genome-scale protein function classification', *Bioinformatics*,
393 30(9), pp. 1236–1240. Available at: https://doi.org/10.1093/bioinformatics/btu031.

394 Kalyaanamoorthy, S. *et al.* (2017) 'ModelFinder: Fast model selection for accurate phylogenetic
395 estimates', *Nature Methods*, 14(6), pp. 587–589. Available at: https://doi.org/10.1038/nmeth.4285.

396 Kang, Y.J. *et al.* (2014) 'Extensive diversity of *Rickettsiales* bacteria in two species of ticks from China
397 and the evolution of the *Rickettsiales*', *BMC Evolutionary Biology*, 14(1), pp. 1–12. Available at:
398 https://doi.org/10.1186/S12862-014-0167-2/FIGURES/4.

399 Kikuchi, Y. and Fukatsu, T. (2005) '*Rickettsia* Infection in Natural Leech Populations', *Microbial*
400 *Ecology*, 49(2), pp. 265–271. Available at: https://doi.org/10.1007/s00248-004-0140-5.

401 Kotter, H. (2005) *Bionomie und Verbreitung der autochthonen Fiebermücke Anopheles plumbeus*
402 *(Culicidae) und ihrer Vektorkompetenz für Plasmodium falciparum, Erreger der Malaria tropica*.
403 Available at: https://doi.org/10.11588/heidok.00006104.

404 Kremer, N. *et al.* (2009) 'Wolbachia Interferes with Ferritin Expression and Iron Metabolism in
405 Insects', *PLoS Pathogens*, 5(10), p. e1000630. Available at:
406 https://doi.org/10.1371/journal.ppat.1000630.

407 Leclair, M. *et al.* (2017) 'Consequences of coinfection with protective symbionts on the host
408 phenotype and symbiont titres in the pea aphid system', *Insect Science*, 24(5), pp. 798–808.
409 Available at: https://doi.org/10.1111/1744-7917.12380.

Masui, S., Sasaki, T. and Ishikawa, H. (2000) 'Genes for the Type IV Secretion System in an Intracellular Symbiont, Wolbachia, a Causative Agent of Various Sexual Alterations in Arthropods', *Journal of Bacteriology*, 182(22), pp. 6529–6531.

Meloni, S. *et al.* (2003) 'The twin-arginine translocation (Tat) system is essential for Rhizobium–legume symbiosis', *Molecular Microbiology*, 48(5), pp. 1195–1207. Available at: https://doi.org/10.1046/j.1365-2958.2003.03510.x.

Minh, B.Q. *et al.* (2020) 'IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era', *Molecular Biology and Evolution*. Edited by E. Teeling, 37(5), pp. 1530–1534. Available at: https://doi.org/10.1093/molbev/msaa015.

Moran, N.A., McCutcheon, J.P. and Nakabachi, A. (2008) 'Genomics and Evolution of Heritable Bacterial Symbionts', *Annual Review of Genetics*, 42(1), pp. 165–190. Available at: https://doi.org/10.1146/annurev.genet.41.110306.130119.

Moreira, L.A. *et al.* (2009) 'A Wolbachia Symbiont in Aedes aegypti Limits Infection with Dengue, Chikungunya, and Plasmodium', *Cell*, 139(7), pp. 1268–1278. Available at: https://doi.org/10.1016/j.cell.2009.11.042.

Pedregosa, F. *et al.* (2011) 'Scikit-learn: Machine Learning in Python', *Journal of Machine Learning Research*, 12(85), pp. 2825–2830.

Pereira, T.N. *et al.* (2018) 'Wolbachia significantly impacts the vector competence of Aedes aegypti for Mayaro virus', *Scientific Reports*, 8(1), pp. 1–9. Available at: https://doi.org/10.1038/s41598-018-25236-8.

Pilgrim, J. *et al.* (2017) 'Torix group *Rickettsia* are widespread in Culicoides biting midges (Diptera: Ceratopogonidae), reach high frequency and carry unique genomic features', *Environmental Microbiology*, 19(10), pp. 4238–4255. Available at: https://doi.org/10.1111/1462-2920.13887.

Pilgrim, J. *et al.* (2021) 'Torix *Rickettsia* are widespread in arthropods and reflect a neglected symbiosis', *GigaScience*, 10(3), pp. 1–19. Available at: https://doi.org/10.1093/gigascience/giab021.

QGIS.org (2020) 'QGIS 3.16'. Hannover: QGIS Association. Available at: http://www.qgis.org/.

Rice, D.W., Sheehan, K.B. and Newton, I.L.G. (2017) 'Large-Scale Identification of *Wolbachia pipientis* Effectors', *Genome Biology and Evolution*, 9(7), pp. 1925–1937. Available at: https://doi.org/10.1093/gbe/evx139.

Rossum, G.V. and Drake, F.L. (2009) *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.

Schaffner, F. *et al.* (2012) 'Anopheles plumbeus (Diptera: Culicidae) in Europe: A mere nuisance mosquito or potential malaria vector?', *Malaria Journal*, 11, p. 393. Available at: https://doi.org/10.1186/1475-2875-11-393.

Schrallhammer, M. *et al.* (2013) '"*Candidatus* Megaira polyxenophila" gen. nov., sp. nov.: Considerations on Evolutionary History, Host Range and Shift of Early Divergent *Rickettsiae*', *PLoS ONE*. Edited by S.A. Ralph, 8(8), p. e72581. Available at: https://doi.org/10.1371/journal.pone.0072581.

Seabold, S. and Perktold, J. (2010) 'Statsmodels: Econometric and Statistical Modeling with Python', *PROC. OF THE 9th PYTHON IN SCIENCE CONF*. Available at: http://statsmodels.sourceforge.net/ (Accessed: 11 February 2021).

Siozios, S. *et al.* (2013) 'The diversity and evolution of *Wolbachia* ankyrin repeat domain genes', *PLoS ONE*, 8(2), p. e55390. Available at: https://doi.org/10.1371/journal.pone.0055390.

Siozios, S. *et al.* (2020) 'DNA barcoding reveals incorrect labelling of insects sold as food in the UK', *PeerJ*, 8, p. e8496. Available at: https://doi.org/10.7717/peerj.8496.

Thongprem, P. *et al.* (2020) 'Transmission, tropism, and biological impacts of torix *Rickettsia* in the common bed bug *Cimex lectularius* (Hemiptera: Cimicidae)', *Frontiers in Microbiology*, 11. Available at: https://doi.org/10.3389/fmicb.2020.608763.

Tseng, T.-T., Tyler, B.M. and Setubal, J.C. (2009) 'Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology', *BMC Microbiology*, 9(1), p. S2. Available at: https://doi.org/10.1186/1471-2180-9-S1-S2.

Tsuchida, T. *et al.* (2002) 'Diversity and geographic distribution of secondary endosymbiotic bacteria in natural populations of the pea aphid, Acyrthosiphon pisum', *Molecular Ecology*, 11(10), pp. 2123–2135. Available at: https://doi.org/10.1046/j.1365-294X.2002.01606.x.

Vega, I.A., Arribére, M.A. and Castro-Vazquez, A. (2012) 'Apple snails and their endosymbionts bioconcentrate heavy metals and uranium from contaminated drinking water', *Environmental Science and Pollution Research*, 19, pp. 3307–3316. Available at: https://doi.org/10.1007/s11356-012-0848-6.

Walker, T. *et al.* (2011) 'The wMel Wolbachia strain blocks dengue and invades caged Aedes aegypti populations', *Nature*, 476(7361), pp. 450–453. Available at: https://doi.org/10.1038/nature10355.

Walker, T. *et al.* (2021) 'Stable high-density and maternally inherited Wolbachia infections in Anopheles moucheti and Anopheles demeilloni mosquitoes', *Current Biology*, 31(11), pp. 2310-2320.e5. Available at: https://doi.org/10.1016/J.CUB.2021.03.056.

Waskom, M. and Seaborn development team (2020) 'mwaskom/seaborn'. Zenodo. Available at: https://doi.org/10.5281/zenodo.592845.

Werner, D. *et al.* (2014) 'The citizen science project "Mückenatlas" supports mosquito (Diptera, Culicidae) monitoring in Germany.', *Proceedings of the 8th International Conference on Urban Pests, 20-23 July 2014, Zurich, Switzerland*, pp. 119–124.

Wick, R.R. and Holt, K.E. (2022) 'Polypolish: Short-read polishing of long-read bacterial genome assemblies', *PLoS computational biology*, 18(1), p. e1009802. Available at: https://doi.org/10.1371/journal.pcbi.1009802.

Wu, M. *et al.* (2004) 'Phylogenomics of the Reproductive Parasite Wolbachia pipientis wMel: A Streamlined Genome Overrun by Mobile Genetic Elements', *PLoS Biology*, 2(3), p. e69. Available at: https://doi.org/10.1371/journal.pbio.0020069.

Xie, J. *et al.* (2014) 'Male killing Spiroplasma protects Drosophila melanogaster against two parasitoid wasps', *Heredity*, 112(4), pp. 399–408. Available at: https://doi.org/10.1038/hdy.2013.118.

Zchori-Fein, E., Borad, C. and Harari, A.R. (2006) 'Oogenesis in the date stone beetle, Coccotrypes dactyliperda, depends on symbiotic bacteria', *Physiological Entomology*, 31(2), pp. 164–169. Available at: https://doi.org/10.1111/j.1365-3032.2006.00504.x.

Table and Figure legends

Figure 1. Map of '*Ca.* Tisiphia' infection rates across Germany where the size of the circle represents the number of individuals sampled and the colour indicates the proportion of '*Ca.* Tisiphia' infected individuals. Red = 90-100% infection to light yellow = 50-60% infection. Source data can be found in Supplementary Table S1.

Figure 2. Standardised and scaled environmental data comparing Uninfected (N=13) and Infected (N=237) by environmental variable. Source data can be found in Supplementary Table S1.

Figure 3. Genome wide phylogeny of '*Ca*. Tisiphia' and '*Ca*. Megaira'. Maximum likelihood (ML) phylogeny constructed from 205 single copy gene clusters that contain a total of 3690 genes. New genomes are indicated by ◀ and bootstrap values based on 1000 replicates are indicated with coloured circles (red = 91-100, yellow = 81-90, black <= 80). Accession numbers for each genome are available in Supplementary Table S2.

Figure 4. KEGG module distribution in '*Ca*. Tisiphia'. The number of pathways found per genome annotated by KEGG module category for '*Ca*. Tisiphia'. Full metadata can be found in Supplementary data. ▲ indicates the genome assembled in this study. Full metadata can be found in Supplementary Tables S3 and S4.

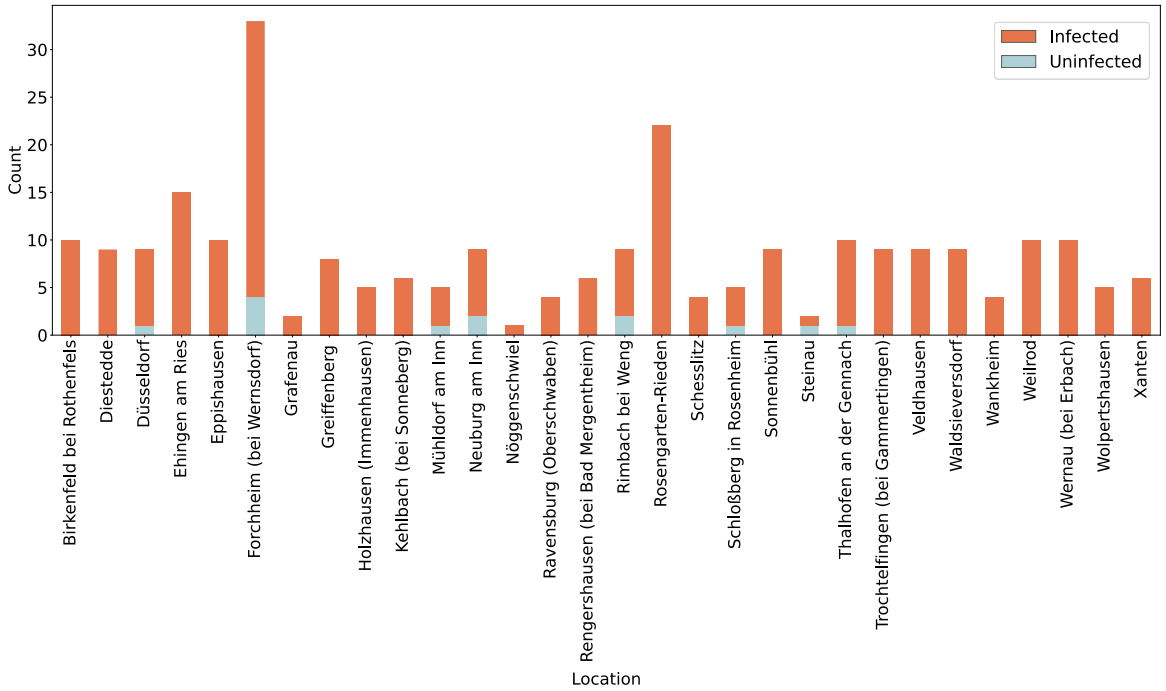
Figure 5. Predicted completeness of KEGG kofam metabolic pathways across '*Ca*. Tisiphia'. The genome assembled in this chapter is coloured grey and indicated with ▲. Full metadata can be found in Supplementary Tables S3 and S4.

Figure 6. Fluorescence in situ microscopy images of *Anopheles plumbeus* ovaries infected with '*Ca*. Tisiphia'. Red shows '*Ca*. Tisiphia' stained with ATTO-633, blue are host nuclei stained with Hoechst. Panels show a) the whole female reproductive organ outlined in white and a breakdown of each light channel and b) a close up of the ovaries showing localised infection within the primary and secondary follicles. White bars indicate a) 150 micrometres and b) 50 micrometres.

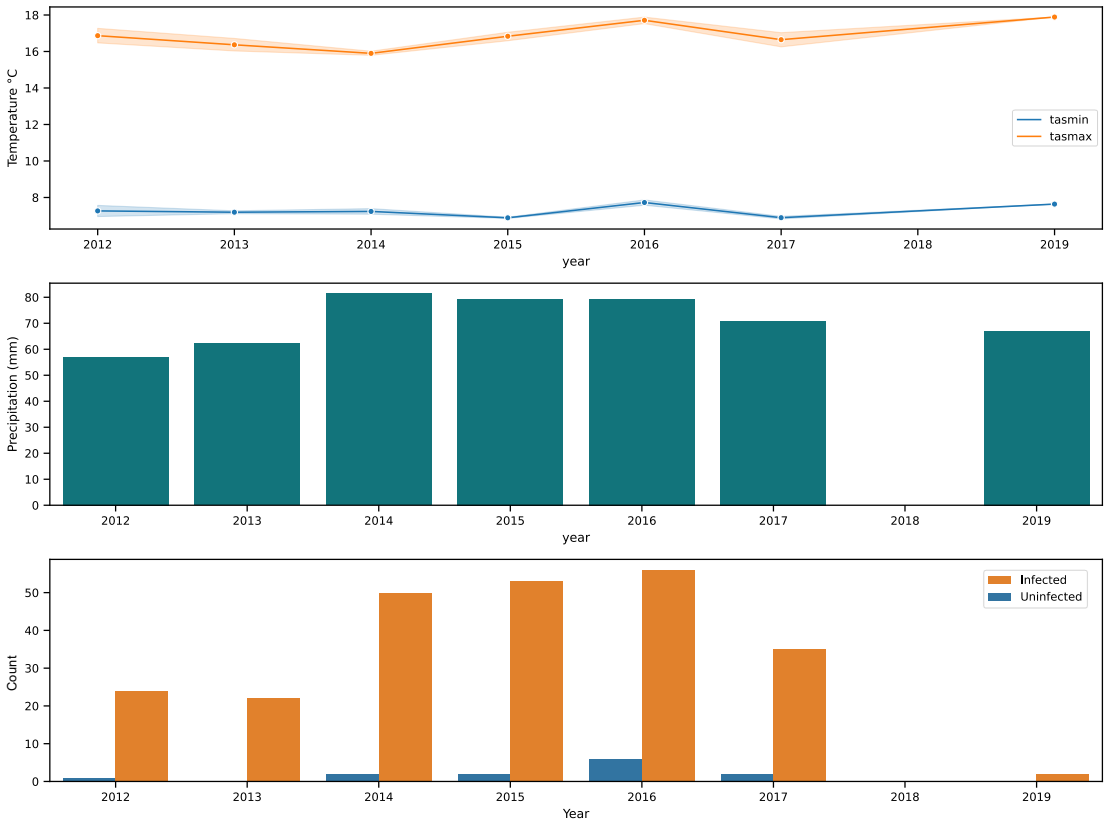
Figure 7. Fluorescence in situ microscopy images of *Anopheles plumbeus* testes. Blue are host nuclei stained with Hoechst 33342, Red is ATTO-633 auto-fluorescence in the testes not '*Ca.* Tisiphia' staining. White bars indicate a) 150 micrometres and b) 50 micrometres.

Table 1. Summary of the genome assembly for TsAplum.

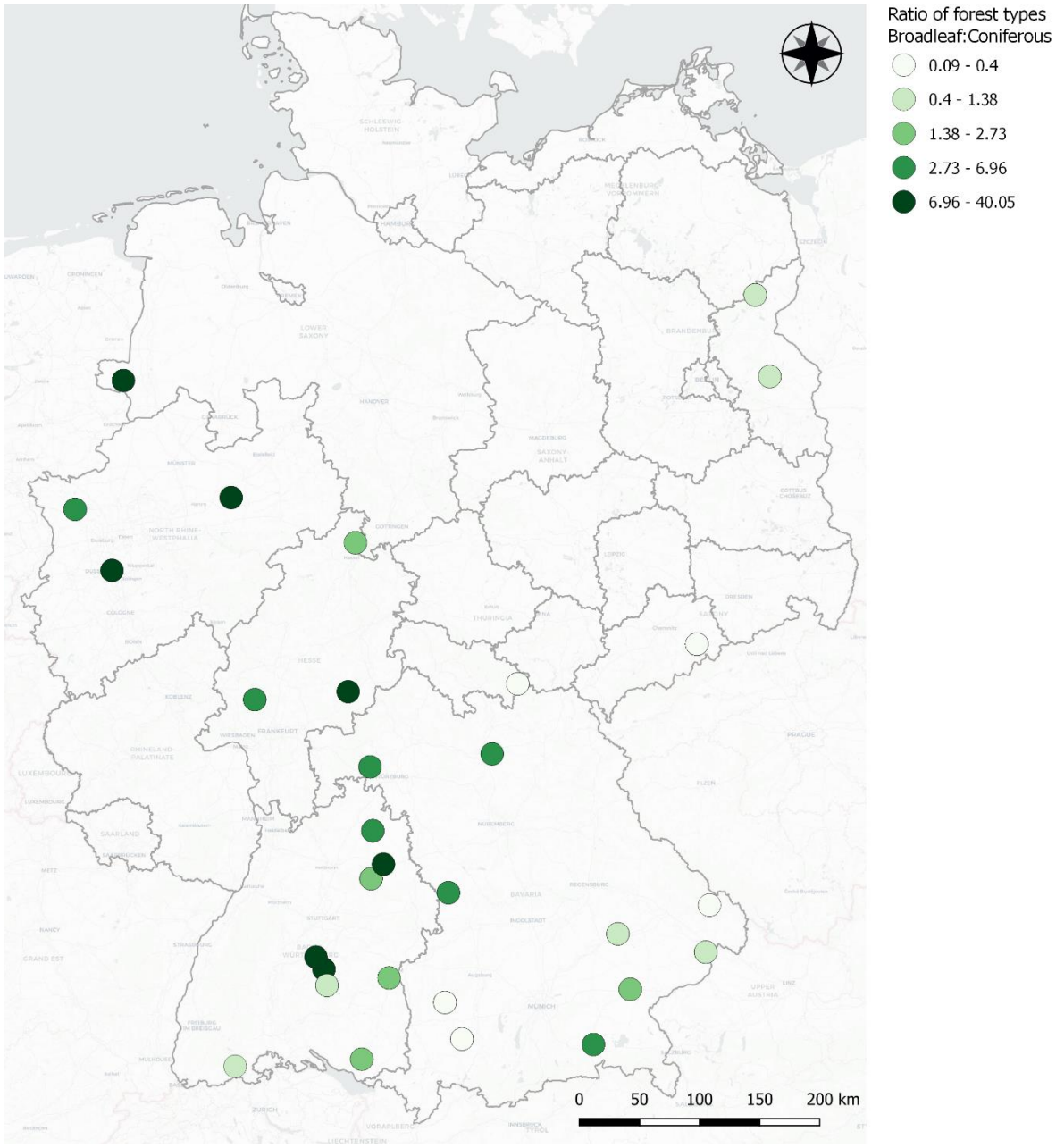
Supplementary Figures



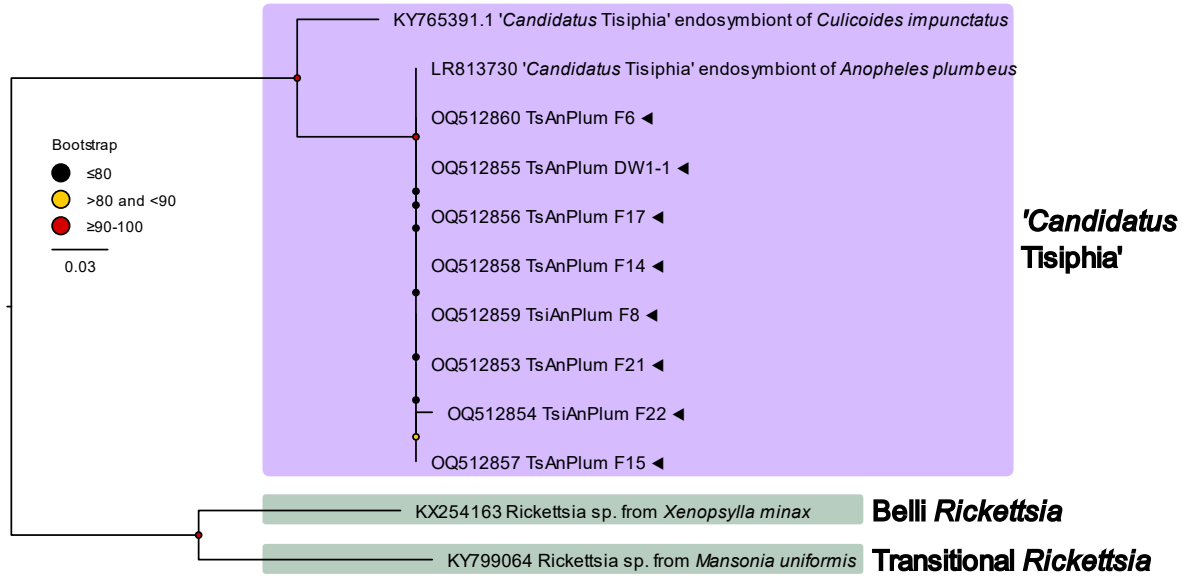
Supplementary Figure 1. '*Ca*. Tisiphia' infection rates by site'. Infected samples are shown in orange, uninfected are shown in light blue. Source data can be found in Supplementary Table S1.



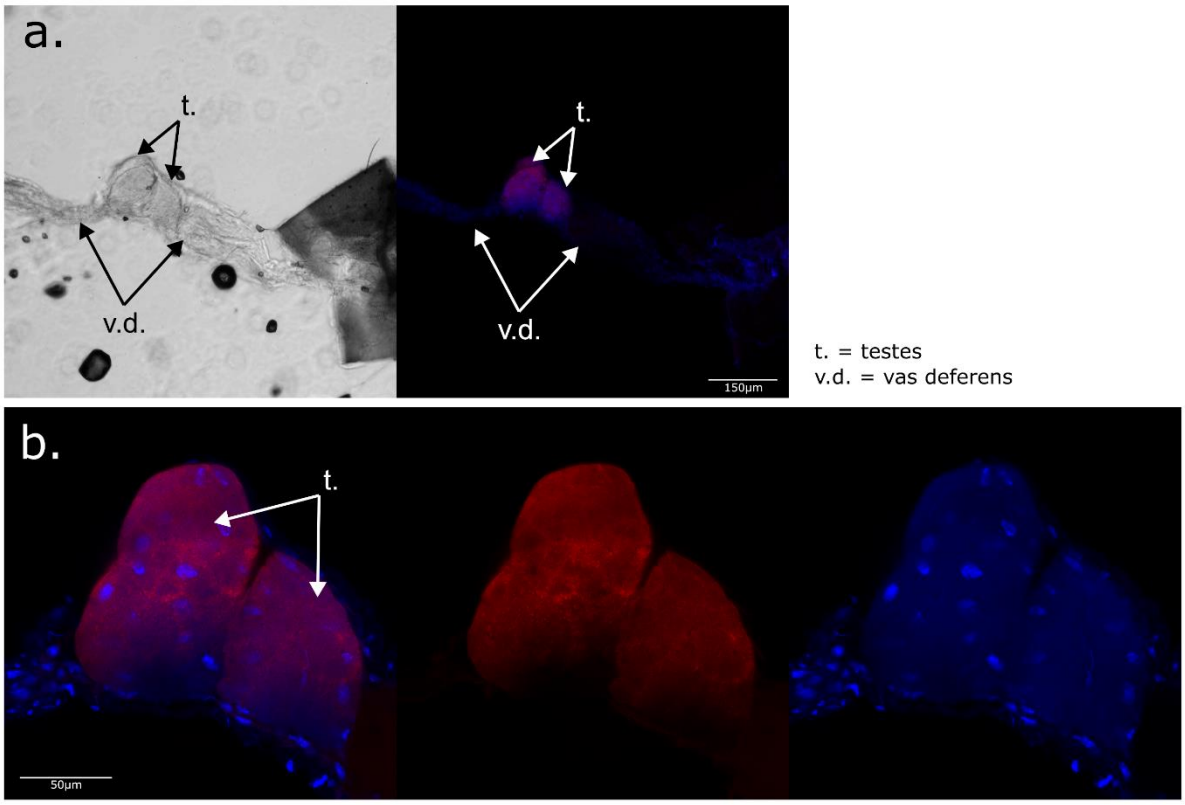
Supplementary Figure 2. Environmental data for *Anopheles plumbeus* collection sites across Germany extracted from the TerraClim database. (Top) average annual minimum and maximum temperature across all *An. plumbeus* collection sites in Germany. (Middle) average annual precipitation across all sites. (Bottom) counts of infected and uninfected individuals across all sites where dark blue = infected and orange = uninfected.



Supplementary Figure 3. The ratio of broadleaf to coniferous forest in a 3km radius of each collection site. Darker green indicates more broadleaf, lighter green indicates closer to equal proportions. Source data can be found in Supplementary Table S1.



Supplementary Figure 4. Maximum likelihood tree for 17 kDa surface antigen (omp) for '*Candidatus* Tisiphia' extracted from *An. plumbeus*. New genomes are indicated by ◄ and bootstrap values based on 1000 replicates are indicated with coloured circles (red = 91-100, yellow = 81-90, black <= 80).



Supplementary Figure 5. Fluorescence in situ microscopy images of *Anopheles plumbeus* testes. Blue are host nuclei stained with Hoechst 33342, Red is ATTO-633 auto-fluorescence in the testes not '*Ca*. Tisiphia' staining. White bars indicate a) 150 micrometres and b) 50 micrometres.