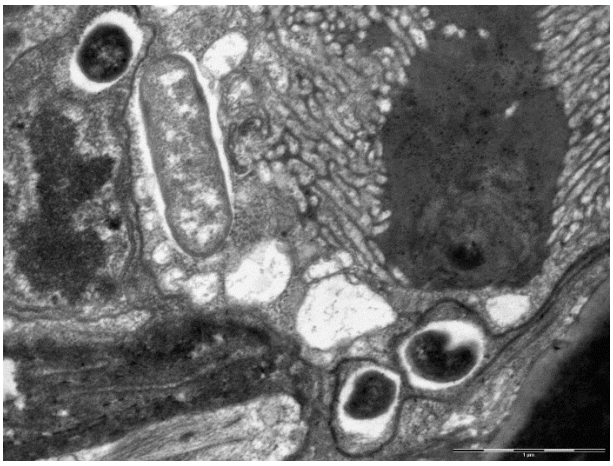
UNIVERSITY OF LIVERPOOL

The prevalence of endosymbiotic bacteria in *Culicoides* biting midges and the distribution of Torix group *Rickettsia* in arthropod hosts



Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

Jack David Rogers Pilgrim

February 2020

The prevalence of endosymbiotic bacteria in *Culicoides* biting midges and the distribution of Torix-group *Rickettsia* in arthropod hosts

Abstract

Culicoides biting midges (Diptera: Ceratopogonidae), are the vectors of several viruses affecting livestock including bluetongue and Schmallenberg viruses. There are no effective control methods of the vectors, and disease control therefore relies on vaccines which, given the rapid emergence/spread of the viruses, are often not available. Thus, there is increasing interest in the heritable bacteria (endosymbionts) of invertebrates as they present novel targets for control initiatives. For example, the endosymbiont *Wolbachia* is capable of inducing a "virus blocking" phenotype in mosquito hosts. Previous studies on biting midges, have revealed infections with the endosymbiotic bacteria, *Wolbachia* and *Cardinium*. However, other common symbionts, such as *Rickettsia*, are underexplored.

I first clarify which *Culicoides* vector species are appropriate for further study of *Cardinium*-midge interactions. Reinvestigation of a previous UK screening study indicates spurious identification of *Cardinium* infection in the vector species *C. pulicaris*, as a result of inappropriate methodology and interpretation. In addition, this chapter establishes associations between mitochondrial haplotypes (mitotypes), used as a phylogeographic marker, and *Cardinium* infection in the globally important vector *C. imicola*. The concordance of mitotypes and *Cardinium* infection in populations of *C. imicola* from different geographic regions suggests a potential confounding of future biodiversity studies which fail to consider the presence of *Cardinium*.

I then describe the results of a targeted screening of *Culicoides* populations for *Rickettsia* symbionts. Through conventional PCR, I demonstrated that *Rickettsia* represent a widespread but previously overlooked association, reaching high frequencies in midge populations and present in over a third of the species tested. Sequence typing clusters the *Rickettsia* within the Limoniae group of the genus, a group known to infect several aquatic and haematophagous taxa. Considering the presence of *Rickettsia* in several vector species, this result highlights the need to establish the impact of this newly-found association on vector competence. Leading from this, I describe the tropism of a *Rickettsia* endosymbiont present in the Scottish Highland midge, *Culicoides impunctatus*. Fluorescence *in-situ* hybridisation (FISH) and transmission electron microscopy (TEM) analysis indicated the presence of *Rickettsia* bacteria in ovarian tissue and the ovarian suspensory ligament suggesting a novel germline targeting strategy. In addition, *Rickettsia* presence in the fat body of larvae indicates potential host fitness effects to be investigated in the future.

In the final study, I investigate inadvertent amplification of *Rickettsia* DNA through the common taxonomic identification technique of DNA barcoding. Through collaboration with the Barcoding of Life Data System (BOLD) curators, I undertook a systematic survey to determine the scale of this phenomenon, as well as investigate its potential to unveil new *Rickettsia*-infected host species. My results determine that unintended *Rickettsia* amplification is common and should be considered when designing future barcoding studies. In addition, a new wealth of host information was uncovered which can inform future directions of investigation pertaining to *Rickettsia.*

Preface

This thesis is based on research carried out at the Institute of Infection and Global Health at the University of Liverpool. The work in this thesis is my own, with the following exceptions:

In chapters 2 and 3, the collection and identification of *Culicoides* from Sweden and France were undertaken by Dr Jan Chirico (National Veterinary institute, Uppsala, Sweden) and Dr Claire Garros (CIRAD, Agricultural Research for Development, Montpellier, France). All other collections were undertaken by the author. In chapter 3 and 5, DNA extracts of *Rickettsia* positive taxa were provided by Panupong Thongprem.

In chapter 4, electron microscopy imaging was undertaken by Alison Beckett (University of Liverpool).

The phylogeographic analyses undertaken in chapter 3 were published in Environmental Microbiology, 19(10): pp.4238-4255. This publication also includes genomic analysis undertaken by Dr Stefanos Siozios (University of Liverpool).

Jack Pilgrim

Acknowledgements

I'm grateful to the Biotechnology and Biological Sciences Research Council Doctoral Training Partnership programme for allowing me to undertake this project. I want to extend my gratitude to my supervisors Professor Matthew Baylis and Professor Greg Hurst. I would particularly like to thank Matthew for giving me opportunities to undertake field work in Africa and the United States; and Greg for his enthusiasm, patience and kind words. Thanks go to the Hurst group who have made all the lab work here so enjoyable. In particular, I would like to thank Dr Stefanos Siozios for his assistance in teaching and guidance over the past 4 years and more importantly his friendship. Many thanks to Matthew Palmer, Charlie Winstanley, Vinnie Keenan and Lukasz Lukomski for agreeing to venture into the Scottish Highlands with me in order to assist with collections of midges. Thank you to Chloé Leseney for her friendship and keeping me awake when I was falling asleep during those long hours in the microscope room. Finally, a special thanks go to my parents for their love and support which has been invaluable.

'*For an instance of insects endued with a spear, I shall, for its peculiarity, pitch upon one of the smallest, if not the very smallest of all of the Gnat-kind…called midges*.'

Physico-Theology: or a Demonstration of the Being and Attributes of God from His Works of Creation. p.191

Reverand W. Derham, 171

List of Abbreviations

%	percentage
°C	degrees Celsius
17KDa	17 kilodalton antigenic protein (*Rickettsia*)
A	adenine
ADP	adenosine diphosphate
AHSV	African horse sickness virus
AKAV	Akabane virus
atpA	adenosine triphosphatase subunit alpha
BIN	barcode index number
BOLD	barcoding of life data system
bp	base pairs
BTV	bluetongue virus
C	cytosine
cCpun	*Cardinium* genome of *Culicoides punctatus*
CDC	Centre for Disease Control and Prevention
CI	cytoplasmic incompatibility
CO_2	carbon dioxide
COI	cytochrome oxidase subunit 1 (mitochondrial)
coxA	cytochrome oxidase subunit 1 (bacterial)
CS	central strain
DAPI	4',6-diamidino-2-phenylindole
ddH_2O	double distilled water
DEFRA	Department for Environment, Food and Rural Affairs
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
EEV	equine encephalosis virus
EHDV	epizootic haemorrhagic disease virus
EIP	extrinsic incubation period
EMB	eastern Mediterranean basin

ExoSAP	exonuclease/shrimp alkaline phosphatase
FISH	fluorescence *in-situ* hybridisation
G	guanine
GIS	geographic information system
gltA	citrate synthase subunit alpha
GyrB	Gyrase B
Hg	inch of mercury (pressure)
ITS	internal transcriber sequences
Ka/Ks	non-synonymous/synonymous mutation ratio
Km	kilometers
kV	kilovolts
LED	light emitting diode
ln	likelihood score
m/s	metres per second
MALDI-TOF-MS	matrix-assisted laser desorption/ionization time-of flight mass spectrometry
min(s)	minute(s)
ml	Microlitres
ML	maximum likelihood
MLST	multi locus sequence typing
mM	micromolar
MST	minimum spanning tree
mtDNA	mitochondrial deoxyribonucleic acid
NaCl	sodium chloride
NCBI	National Center for Biotechnology Information
NUMT	nuclear mitochondrial DNA
OIE	World Organisation for Animal Health
OSL	ovarian suspensory ligament
OsO_4	osmium tetroxide
PBF	post blood feeding
PCR	polymerase chain reaction

pmol	Picomoles
Q	phred score
qPCR	real-time polymerase chain reaction
qRT-PCR	real-time reverse transcription polymerase chain reaction
RiCNE	*Rickettsia* genome of *Culicoides newsteadi* N5
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid
SBV	Schmallenberg virus
SLV	single locus variant
SNP	single nucleotide polymorphism
sp.	species (individual)
spp.	species (plural)
T	Thymine
TEM	transmission electron microscopy
Tris-HCl	tris(hydroxymethyl)aminomethane
UK	United Kingdom
UPGMA	Unweighted Pair Group Method with Arithmetic Mean
USA	United States of America
W	Watts
w/v	weight/volume ratio
WMB	western Mediterranean basin
wsp	*Wolbachia* surface protein
µl	microlitres
µm	micrometres

List of Tables

List of Figures

Table of Contents

Chapter 1: General Introduction

1.1 Vector-borne diseases

1.1.1 Overview

Vector-borne diseases are infections transmitted by blood-feeding arthropods. The most well-known examples of these infections are pathogens spread by mosquitoes, although tick and midge-borne diseases are also of importance (Gubler, 2009). For example, *Anopheles* and *Aedes* mosquito species transmit malaria and flaviviruses (dengue, Zika and chikungunya), causing hundreds of thousands of human deaths every year (World Health Organisation, 2019). Additionally, mosquitoes can transmit filarial nematodes including *Brugia* malayi, which can cause lymphatic filariasis (elephantiasis); a disease leading to disability and social stigma in developing nations (Zeldenryk *et al.*, 2011). Vector borne diseases are also of importance in the veterinary world including pathogens such as tick-borne *Babesia* sp. (Babesiosis in several mammals and birds) and mosquito transmitted *Dirofilaria immitis* (Heart worm in dogs and cats) (Morchón *et al.*, 2012; Dantas-Torres, Alves and Uilenberg, 2016).

Conventional vector control approaches have heavily relied on the removal of breeding sites and insecticides (Flores and O'Neill, 2018). However, these are proving insufficient to cope with host population density increases as a result of urbanisation particularly in tropical regions (Pang, Mak and Gubler, 2017). Additionally, meteorological factors can influence the reproduction and survival of vectors meaning that climatic variables can influence invasive ranges and population sizes (Takken and Knols, 2007). Indeed, approximately a third of emerging diseases are deemed to be vector-borne (Jones *et al.*, 2008), suggesting new health interventions are of pressing need.

One vector control approach aims to reduce vector population numbers. Population suppression approaches include sterile insect technique (SIT) where male vectors are sterilised via irradiation or chemical treatment so females fail to produce offspring after mating (Lees *et al.*, 2015); and the release of transgenic insects carrying a dominant lethal (RIDL) which leads to larval or pupal mortality when the lethal gene is switched on in the wild (Phuc *et al.*, 2007; Wise De Valdez *et al.*, 2011). The shortcomings of these approaches include the limited epidemiological evidence available to suggest effectiveness, the reduced fitness of lab-reared mosquitoes compared to their wild counterparts (particularly irradiated insects) (Bowman, Donegan and McCall, 2016), and the negative public perception of releasing genetically modified insects into nature (Flores and O'Neill, 2018). An approach which aims to overcome some of these obstacles is the use of endosymbionts (bacteria residing within the body or cells of insect hosts).

1.1.2 Endosymbiont-based control of vector-borne diseases

Endosymbiont protection of Dipteran hosts was first observed in *Drosophila* fruit flies infected naturally with *Wolbachia*, which were protected against fungal and viral pathogens (Panteleev *et al.*, 2007; Hedges *et al.*, 2008; Teixeira, Ferreira and Ashburner, 2008). Further experimental work demonstrated the ability to transinfect mosquito species with the *Wolbachia* strain *w*Mel, leading to a blocking effect of dengue virus, chikungunya virus, yellow fever virus, West Nile virus and Zika virus (Bian *et al.*, 2010b; Blagrove *et al.*, 2012; Hussain *et al.*, 2012; van den Hurk *et al.*, 2012; Chouin-Carneiro *et al.*, 2019). The further *Wolbachia* strains *w*AlbA, *w*Au and *w*MelPop also appear to perturb pathogen transmission suggesting multiple routes of disease control investigation (Moreira *et al.*, 2009; Walker *et al.*, 2011; Ant *et al.*, 2018). The potential to deploy these effects in wild populations of mosquitoes is

enabled by *Wolbachia*-induced cytoplasmic incompatibility (CI) (Ant *et al.*, 2019). This is the induction of embryo death in mating between infected males and uninfected females, with all other crosses remaining viable (Werren and O'Neill, 1997). The significance of this effect is the rapid spread of the symbiont through a population before gaining stability at a near 100% prevalence rate. This is contingent on an invasion threshold being reached dependent on the "strength" of CI, vertical transmission, and fitness cost to infected females (Fine, 1978; Stouthamer, Breeuwer and Hurst, 1999). Large initial releases aid in achieving the invasion threshold, and persistence may then occur through the CI phenotype without onward releases. Viral suppression is dependent on both host and *Wolbachia* genetic backgrounds. For example, when the natural strain of *Aedes albopictus* (*w*AlbB) was transferred to the naïve host, *Aedes aegypti*, a stronger inhibitory effect of dengue virus was observed than in the native host (Bian *et al.*, 2010). Furthermore, high endosymbiont densities, and subsequently a greater virus inhibition effect has been observed when *Drosophila* strains of *Wolbachia* (*w*Mel and *w*MelPop) are transinfected into mosquitoes (Blagrove *et al.*, 2012).

Several mechanisms have been proposed for *Wolbachia*-mediated virus blocking. These include immune-effector modulation and competition for resources (Sinkins, 2013; Lindsey *et al.*, 2018). With respect to the latter, the bacterium's relationship with host lipid metabolism has attracted attention. Flaviviruses, with a limited genetic repertoire, rely on host-derived components for maintenance, such as cholesterol (Osuna-Ramos, Reyes-Ruiz and Del Ángel, 2018). During flavivirus infection, host cholesterol levels can modulate cell entry, replication, assemblage and host interferon responses (Rothwell *et al.*, 2009; Upla, Hyypiä and Marjomäki, 2009; Heaton *et al.*, 2010). Furthermore, dengue virus has been suggested to

indirectly increase intracellular cholesterol through reducing the expression of intramembrane proteins related to cholesterol metabolism (Tree *et al.*, 2019). Subsequently, as *Wolbachia* also relies on such host-derived components, cholesterol has been viewed as the target by which the endosymbiont suppresses viral titres through competition (Caragata *et al.*, 2013; Sinkins, 2013; Molloy *et al.*, 2016; Frentiu, 2017). Furthermore, the localisation of *Wolbachia* within cells is mediated by the actin cytoskeleton which is also used to transport virus within endosomes (Sheehan *et al.*, 2016). Subsequently, a pre-existing *Wolbachia* infection within a cell could lead to disruption of such virion trafficking (Geoghegan *et al.*, 2017). Although, direct interactions are likely to explain partially *Wolbachia*-mediated virus blocking, the independent distribution of *w*MelPop and Dengue virus in *Aedes aegypti* (Moreira *et al.*, 2009) suggests other means should also be considered. Priming of the immune system by the endosymbiont has been proposed as such a mechanism due to the upregulation of several immune genes, including components of the Toll-pathway, which are major regulators of flavivirus replication (Xi, Ramirez and Dimopoulos, 2008; Pan *et al.*, 2012).

Encouragingly, the results of field-based CI-inducing symbiont interventions indicate a viable method to control arboviruses (Hoffmann *et al.*, 2014; Schmidt *et al.*, 2017). For example, since the 2011 release of *w*Mel transinfected *Aedes aegypti* in Northern Australia, there has been no local transmission of dengue fever where the bacterium is established (Ritchie, 2018). Importantly, as symbionts are naturally occurring and ubiquitous, public acceptance of these *Wolbachia*-based initiatives may be seen as more acceptable than other genetic modification interventions. Furthermore, this protective phenotype appears to be associated with RNA viruses leading to the potential for symbiont-based biocontrol in major midge-borne

pathogens such as bluetongue virus (BTV), African horse sickness (AHSV) and Schmallenberg virus (SBV). Finally, *Wolbachia* can be used as part of an incompatible insect technique (IIT), which is a modified version of SIT where the endosymbiont is used to effectively sterilise males without the fitness costs encountered by irradiation (Zheng *et al.,* 2019). When males are released in high enough numbers, vector population collapse occurs due to incompatible matings. Despite these promising field advancements, symbionts are often a neglected factor in non-mosquito-borne arbovirus disease dynamics.

1.2 *Culicoides* biting midges

1.2.1 The significance of *Culicoides* as vectors and pests

Culicoides (Diptera: Ceratopogonidae) biting midges are blood-feeding flies which range from 1-3 mm in length (Mellor, Boorman and Baylis, 2000). As well as being a biting nuisance to humans, a number of *Culicodies* species are able to transmit filarial nematodes and protozoan parasite to birds, humans, and other animals (Yates, Lowrie and Eberhard, 1982; Linley, 1985a; Veiga *et al.*, 2018). Although other haematophagous flies from the family Ceratopogonidae are known as vertebrate parasites, the dominant research interest for the genus *Culicoides* are the numerous viruses (Reoviridae and Peribunyaviridae families) they transmit (Mellor, Boorman and Baylis, 2000; Sick *et al.*, 2019). Pathogens affecting ruminants include BTV, SBV, Akabane virus (AKAV) and epizootic haemorrhagic disease virus (EHDV). Other viruses are known to affect equids including AHSV and equine encephalosis virus (EEV), which are of great significance for animal welfare, as well as the horse racing industry (Bachanek-Bankowska *et al.*, 2009).

Apart from the economic and animal health implications of these veterinary viruses, the human pathogen oropouche virus, observed in the Americas (during the 1950s), makes these vectors additionally of human health importance (Anderson *et al.*, 1961). Indeed, a recent review (Sick *et al.*, 2019) suggests *Culicoides* are possibly a neglected vector of human viruses, with the lack of information deriving from a bias towards surveillance of mosquitoes. Considering the Peribunyaviridae contain human pathogens, and that this virus family are largely associated with *Culicoides*, it would be unsurprising to observe future midge-borne outbreaks of human viruses.

SBV is a negative-sense single-stranded RNA member of the Bunyaviridae family and only recently emerged in 2011 in the German/Dutch border region (Elbers *et al.*, 2013), causing disease primarily in sheep along with less frequent cases in cattle. Whilst SBV infections in adult animals are usually subclinical, vertical transmission of SBV from mother to progeny can result in severe and fatal congenital malformations along with still births and abortion (European Food Safety Authority, 2016). Since its first detection, multiple outbreaks of SBV have been reported in a range of European countries including Belgium, United Kingdom, Luxembourg, France, Spain and Italy (Beer, Conraths and Van Der Poel, 2012), and has also been suspected in Africa (Leask, Botha and Bath, 2013). The Palearctic biting midges predominantly considered as vectors are the *C. obsoletus* complex and *C. imicola*. This was confirmed by infection experiments undertaken in field-caught midges of both groups (Pagès *et al.*, 2018). Additionally, *C. punctatus*, *C. imicola* and the *C. pulicaris/C. newsteadi* complexes were also positive for the viral genome by qRT-PCR suggesting they may also contribute to

the epidemiology of SBV (Rasmussen *et al.*, 2012; Larska *et al.*, 2013; Balenghien *et al.*, 2014; Ségard *et al.*, 2018)

BTV is an OIE (World Organisation for Animal Health) list A pathogen, meaning it has the potential for very rapid spread and can cause serious socio-economic consequence due to its impact on the international trade of animals. Clinical signs of BTV in ruminants include depression, pyrexia, serous nasal discharge, hypersalivation, facial oedema, as well as hyperaemia and ulceration of the oral cavity (Maclachlan *et al.*, 2009). Death can occur as a result of disease or euthanasia on animal welfare grounds. Bluetongue was first described in sheep from South Africa in the late 18th century (Callis and Kramer, 1997). However, throughout the 20th century, several incursions of BTV serotype 9 from Africa have occurred in the Mediterranean basin leading to outbreaks in the Iberian Peninsula (1950s and 1960s) (Mellor and Boorman, 1995), as well as Greece, Turkey and Bulgaria (1990s) (Mellor *et al.*, 1985; Boorman, 1986b). These invasions were attributed to the invasion of the Afrotropical vector *C. imicola* into southern Europe. However, the 2006 outbreak of BTV serotype 8 across Europe, including the UK, was spread by the common Palearctic species *C. obsoletus*. This outbreak of BTV caused serious economic and animal health damage to the European livestock industries (Wilson and Mellor, 2009) as it spread through a serologically naïve and highly susceptible ruminant population. The cost of the UK epidemic was estimated to cost up to £485M with losses in livestock productivity, restrictions on UK agricultural exports and employment impacts cited as the main contributors to this figure (Pirbright Institute, 2009).

Like mosquitoes, only female *Culicoides* will search for vertebrate hosts to acquire a blood meal in order to allow for the full development of eggs. If these females feed on viraemic blood, then they have the capacity to contribute to the virus transmission cycle (Boorman, 1974). After blood-feeding, there are many barriers which the virus must overcome in order for successful vertebrate to vertebrate transmission to occur (Fu *et al.*, 1999). First, the virus must be able to invade, and then bypass, the gut epithelium to make its way into the haemolymph and secondary organs, such as the fat body. Second, the pathogen must evade host immune responses before reaching the salivary glands where it proliferates. Finally, the host must live long enough for sustained transmission to occur. With respect to the latter, an important concept to consider is the extrinsic incubation period (EIP)(Mills *et al.*, 2017a), which is the time it takes for a virus to initially enter the female *Culicoides* through feeding, until the time that the virus is able to be transmitted via the salivary glands (Pinheiro, Travassos da Rosa and Travassos da Rosa, 1981; Kedmi *et al.*, 2010). This varies widely with virus and invertebrate host, but in the case of BTV serotype 11 and *C. sonorensis*, the EIP is estimated at between 18-22 days at 21°C (Mullens *et al.*, 1995). The mechanical transmission of viruses has also been noted in *Culicoides* transmitting fowl pox (Fukuda *et al.*, 1979). However, it is unknown if *Culicoides* can mechanically transmit BTV or SBV. More recently, there has been a suggestion that certain viruses can be vertically transmitted through *Culicoides*' eggs with SBV being detected in the eggs of *C. punctatus* (Larska *et al.*, 2013).

1.2.2 Biology of *Culicoides*

Over 1,400 *Culicoides* species are currently recognised within the genus and they occur globally with the exceptions of New Zealand, Patagonia, Hawaii and Antarctica (Mellor, Boorman and Baylis, 2000). *Culicoides* are holometabolous and have four larval stages as well

as a pupal and imago stage. Egg hatching times, as well as larval and pupal development periods are dependent on species as well as ambient temperature (Meiswinkel, 1989) with a full lifecycle taking between less than two weeks to several months (Mellor, Boorman and Baylis, 2000). Cigar-shaped eggs are often laid in bunches and are initially translucent before darkening with maturation. The presence of moisture is one of the most important determinants for suitable larval habitats (Blackwell, Young and Mordue, 1994; Lardeux and Ottenwaelder, 1997; Meiswinkel, 1997), although some species such as *C. imicola* will drown in flooded breeding-sites (Nevill, 1969). Thus, immatures are found in a diverse range of habitats including animal dung, tree-holes, soil, swamps, beaches, rivers, marshes and bogs (Linley and Davies, 1971; Kline and Axtell, 1976; Meiswinkel, 1989; Blackwell, Young and Mordue, 1994). Larvae are considered omnivores and can subsist on detritus and algae (Aussel and Linley, 1994), but some species are predatory, having been observed to feed on nematodes (Kettle, Wild and Elson, 1975; Boorman, 1985). In temperate countries with a prolonged cold winter, fourth stage instars will undergo diapause (Kettle, 1977). The continuation of development into pupa has been linked to temperature and increased day length although exact cues remain unknown (Searle *et al.*, 2014; Lühken *et al.*, 2015; White *et al.*, 2017). Pre-pupae will generally rise to the surface of breeding habitats where development continues in loose debris. The pupal period is much shorter than larval stages, usually lasting for only a few days (Barceló and Miranda 2018).

Most adult *Culicoides* species are crepuscular in activity (Kettle, 1977; Mellor, Boorman and Baylis, 2000) although others are present diurnally (Mercer *et al.*, 2009). The life-span of adults is short with most individuals surviving between 2-3 weeks with exceptions living for

periods of up to 90 days (Mellor, Boorman and Baylis, 2000). Flight activity is limited to behaviours such as feeding, mating, host seeking and ovipositing (Campbell and Kettle, 1976; Bishop *et al.*, 1995) and is determined by light intensity, temperature and humidity (Kettle, 1957; Blackwell, 1997). Adults generally find it difficult to fly in wind speeds greater than 3 m/s and will retreat to shelter in rainy conditions (Carpenter *et al.*, 2008). Although the flight range of adults is generally limited to a few hundred metres from breeding sites, long range dispersal has been described spanning thousands of kilometres. Simulation models suggests that wind streams have facilitated the dispersal of populations from Indonesia to Australia (Eagles *et al.*, 2014) as well as Corsica to Spain (Jacquet *et al.*, 2016). Mating is thought to occur through swarms consisting of a much higher proportion of males than females (1 female: 200 males) (Downes, 1950, 1955; Zimmerman, Barker and Turner, 1982). The seasonality of *Culicoides* is determined by climatic and meteorological variables. In temperate regions, numbers are much reduced during winter but during spring, increased temperatures and day-length leads to a population increase which tends to peak a few weeks after the hottest time of year (Kline and Axtell, 1976; Mellor, Boorman and Baylis, 2000). Additionally, the number of generations per year will affect population peaks; univoltine species (undergoing 1 gonotrophic cycle per year) will have greater numbers at the start of the season compared to multivoltine species, where populations peak later (Mellor, Boorman and Baylis, 2000).

1.2.3 *Culicoides* in the UK

At least 46 species of *Culicoides* have been described in the UK, of which 38 have been associated with virus transmission (Wittmann and Baylis, 2000). Of note are the obsoletus and pulicaris species complexes which are of concern for future UK outbreaks of BTV and SBV.

Apart from *C. chiopterus*, all other members of the obsoletus complex are known to be putative vectors of BTV via qRT-PCR screening of field-caught populations (Meiswinkel *et al.*, 2007; Foxi *et al.*, 2016). Furthermore, *C. scoticus*, *C. chiopterus* and *C. obsoletus* of the obsoletus complex have been implicated as putative vectors of SBV (Elbers *et al.*, 2013). Like other temperate regions, adult *Culicoides* in the UK tend to emerge from spring (April onwards) with numbers reducing towards autumn (November) (Mellor, Boorman and Baylis, 2000). However, phenology varies by species with *C. impunctatus* observed between May and September (Holmes and Boorman, 1987). Aside from seasonality, the number of generations throughout a season are also important in determining timing and spread of midge-borne diseases in the UK, with known species having between 1-3 generations per year (Baylis *et al.*, 1997; Sanders *et al.*, 2011). Of interest is *C. impunctatus*, which appears to be univoltine in the South of England but bivoltine in Scotland (Hill, 1947; Blackwell *et al.*, 1992).

Culicoides impunctatus, commonly known as the "Scottish Highland midge", is a UK species of note due to the large numbers observed in Western Scotland. The comparatively small wing-length (approximately 1.5mm) and unmarked thorax makes *C. impunctatus* easy to distinguish from other species in the same subgenus (Culicoides), such as *C. fagineus* (Campbell and Pelham-Clinton, 1960). *C. impunctatus* is prevalent throughout mostly Northern Europe although they can be found as far South as the Iberian Peninsula (Hill, 1947; Campbell and Pelham-Clinton, 1960; John Boorman, 1986b; Rawlings, 1996). The autogenic lifestyle of *C. impunctatus* (it does not require a blood meal for its initial gonotrophic cycle), and its prolificity in peat bog breeding sites has led to its abundance and is a well-known biting nuisance (Boorman and Goodard, 1970). Indeed, Boorman (1986a) attributed between 70-

95% of *Culicoides* attacks on humans to *C. impunctatus*. This biting nuisance contributes negatively to two of Scotland's major industries, Forestry and Tourism. The effects on the former, are as a result of "midge attacks" disrupting arborists' vision while operating machinery and refuelling chainsaws, with up to 20% of summer working hours being lost per year (Hendry and Godwin, 1988). The economic impact on tourism (worth over £2 billion each year to the Scottish economy), however, is far more difficult to quantify due to the transient nature of visitors (Hendry, 1996). Some impact, however, is clear: a questionnaire study, exploring visitor's attitudes to midges during the summer months suggested roughly half would not visit Scotland at the same time of year again. This economic cost prompted several studies focusing on the breeding-sites and seasonality of *C. impunctatus* in the hope of finding a suitable control intervention (Blackwell *et al.*, 1992; Blackwell, 1997). However, the broad breeding site range was shown to be a major limiting factor for restricting numbers (Blackwell, Young and Mordue, 1994). Subsequently, the problem of *C. impunctatus*, as a pest remains unsolved.

1.2.4 Control of *Culicoides* and their viruses

During outbreaks of BTV, movement restrictions of livestock and their products are implemented to stifle the spread of viruses into new regions (Carpenter *et al.*, 2013). In the case of UK outbreaks, this involves the active testing of animals within a 20 km zone of confirmed cases. Additionally serological surveillance of sentinel ruminants and entomological surveillance is implemented within a 100 km protection zone where movement is only allowed within this zone (Defra, 2014). Movements in and out of these zones is permitted during a "vector-free period", which can be declared if there is evidence of a lack of circulating virus along with a lack of vector activity through surveillance and

temperature thresholds (Defra, 2014). There have also been suggestions for farming practices to accommodate for potential teratogenic effects caused by midge-borne viruses such as SBV, which cause no or minor disease in adults (Wernike, Hoffmann and Beer, 2013). For example, the breeding seasons of sheep can be adjusted to ensure the exposure of young livestock to infected vectors (during peak midge activity) so the dam can acquire immunity before conceiving. However, this is impractical because of year to year differences in peak vector activity, as well as a reluctance to change livestock-management from farmers (Dwyer *et al.*, 2007; Carpenter, Mellor and Torr, 2008; Wernike *et al.*, 2014).

Control strategies for BTV and SBV outside of vector control thus rely on immunity acquired through vaccination (Szmaragd *et al.*, 2010; Wernike, Hoffmann and Beer, 2013; Defra, 2014). The successful vaccination of millions of cattle during the 2006-2009 BTV European outbreak demonstrated that immunisation can play a major role in the control of epidemics. In the UK, the >80% uptake rate of vaccination by farmers was deemed a contributing factor to the lack of BTV infections observed in the UK during 2008 (Szmaragd *et al.*, 2010). However, the presence of 27 circulating serotypes complicate application of vaccine strategies. Furthermore, segmented viruses, such as BTV and AHSV, show reassortment of genome segments in hosts co-infected with more than one serotype (Verwoerd *et al.*, 1972; Roy, Mertens and Casal, 1994; Roy, 2017). This can produce novel strains of the virus with a combination of characteristics of the original parental serotypes (Maan *et al.*, 2011; Zientara and Sánchez-Vizcaíno, 2013). Subsequently, such events create difficulties when producing vaccines, and often outbreaks can only be controlled by vaccines after a time-delay of serotyping and developing an effective vaccine (Calvo-Pinilla *et al.*, 2014; Zulu and Venter,

2014; Nomikou *et al.*, 2015; Harrup, Miranda and Carpenter, 2016). Furthermore, the use of live pentavalent BTV vaccines, which comprise five attenuated strains, has led to an increase in genetic diversity of serotypes circulating in the field (Batten *et al.*, 2008). This perpetuates existing concerns over the reliance on live vaccines as it has been suggested that this could play a role in accelerating the evolution of BTV via horizontal gene transfer (Shaw *et al.*, 2012). Other limitations of reactive control measures include the detection failure of new outbreaks by both farmers and vets (Elbers *et al.*, 2010). Additionally, the unpredictable nature of midge-borne outbreaks can cause problems in anticipating the emergence of outbreaks. An example of this is the disappearance and re-emergence of SBV in the UK, within a few years, which has created problems for both prophylactic and reactive control measures (Stokes, Baylis and Duncan, 2016; Stokes *et al.*, 2018). Further, the historic incursions of the vector species *C. imicola* into continental Europe, as a result of climate change and wind-transport events (Mellor and Boorman, 1995; Purse *et al.*, 2005; Venail *et al.*, 2012; Jacquet *et al.*, 2016), has also highlighted the challenge of mapping *Culicoides* movement.

Considering all the above, proactive measures such as vector control are seen to be more attractive options in controlling the spread of these arboviruses (Harrup, Miranda and Carpenter, 2016). Current interventions targeting adults include insecticide treatments of livestock, moving livestock into sheltered housing during peak activities of biting, and the use of repellents (Carpenter, Mellor and Torr, 2008; Harrup, Miranda and Carpenter, 2016; Snyder *et al.*, 2016). For immature stages, the elimination of breeding sites through insecticides and pathogen application has been trialled with some success (Ansari, Carpenter and Butt, 2010; Harrup, Miranda and Carpenter, 2016). However, due to the broad range of habitats, the

disruption of breeding sites is impractical. Finally, insecticide use can lead to unwanted ecological effects due to their non-specific effects (Carpenter, Mellor and Torr, 2008; Harrup, Miranda and Carpenter, 2016).

1.2.5 Taxonomic classification of *Culicoides*

An issue with advancing *Culicoides* control initiatives is the problem of predicting and monitoring vector species due to an underdeveloped taxonomic classification system. Taxonomic classification of adults has primarily relied on wing-spot patterns for adults and head capsule morphology for larvae with multiple keys having been devised (Meiswinkel, 1994; Rawlings, 1996; Mathieu *et al.*, 2012; Bellis, Halling and Anderson, 2015). Certain species are morphometrically similar such as *C. pulicaris*, *C. lupicaris*, *C. bysta* and *C. newsteadi* which make up what is known as the pulicaris complex (Yildirim *et al.*, 2019). Similarly, the obsoletus complex encompasses several cryptic species including *C. obsoletus*, *C. dewulfi*, *C. chiropterus*, and *C. scoticus* (Downes and Kettle, 1952; Nielsen and Kristensen, 2011). Although identification through morphological assessment is possible, this requires laborious geometric analysis of various structures including sex organs, spermathecae and maxillary measurements (Downes and Kettle, 1952; Campbell and Pelham-Clinton, 1960; Mathieu *et al.*, 2012). This difficulty in classification is problematic because species complexes contain some members which have been established as vectors and other which have not. For example, members of both obsoletus and pulicaris complexes have different levels of susceptibility to the same serotype of BTV (Carpenter *et al.*, 2006).

To overcome these issues in identification, molecular techniques have been developed. This can involve multiplex PCR (Nolan *et al.*, 2007) of DNA marker regions or matrix-assisted laser desorption/ionization time-of flight mass spectrometry (MALDI-TOF-MS) which identifies species based on protein signatures (Uhlmann *et al.*, 2014). Furthermore, Internal transcriber sequences (ITS) of ribosomal DNA are occasionally used, but these often lack resolution to differentiate sister species (Cêtre-Sossah *et al.*, 2004; Gomulski *et al.*, 2006). Latterly, mitochondrial markers such as DNA barcodes, using the sequence of the Cytochrome oxidase subunit 1 gene (COI), have been widely utilised (Pagès *et al.*, 2009; Lassen *et al.*, 2012; Ander, Troell and Chirico, 2013; Nielsen and Kristensen, 2015). Following barcoding, it is possible to allocate the mtDNA reference sequence to other barcodes based on percentage similarity. Traditionally, sequences with >98% homology are deemed the same species, although this presumes that the minimal interspecific mtDNA divergence between sister species is greater than maximum intraspecific divergence i.e. a "barcoding gap" exists (Hebert *et al.*, 2003; Wiemers and Fiedler, 2007). Following barcode allocation, it may be possible to return to the voucher specimen, if tissues of taxonomic interest are retained (e.g. head, terminal abdomen and wings), to identify characters which can be used to identify this species in future studies (Nielsen and Kristensen, 2015). Alternatively, through deposition of the sequences into an online database, it is possible for studies to remove the need for morphological assessment altogether (Ratnasingham and Hebert, 2007). Barcoding is not only useful in identifying cryptic vector species, but can allow for the discovery of putative new *Culicoides* species to be investigated in the future (Sarvašová *et al.*, 2017).

1.2.6 Laboratory rearing of *Culicoides*

In order to investigate the role of *Culicoides* in virus transmission cycles, and to investigate life stages outside of the short adult season, this requires the rearing and maintenance of laboratory colonies (Hill, 1947; Megahed, 1956; Kettle, Wild and Elson, 1975; Boorman, 1985; Linley, 1985b; Mullens and Velten, 1994). However, *Culicoides* research has lagged behind that of many other invertebrate vectors due to logistical problems, such as the midges' reluctance to copulate in confined spaces (Linley, 1968; Nayduch *et al.*, 2014). A further limitation to utilising cultivated *Culicoides* for infection experiments, is the trade-off between temperature dependent life-cycle times and mortality. When *Culicoides* adults and larvae are maintained at higher temperatures their survival rate is lower (Hill, 1947; Wellby *et al.*, 1996; Wittmann, Mellor and Baylis, 2002). However, larval and pupal periods are also shorter with these elevated temperatures allowing for quicker access to study materials.

Initial failed attempts at colonisation of *C. imicola* in South Africa was followed by the first successful establishment in the USA vector species *C. sonorensis*, which could mate without displaying swarming behaviour and had a comparatively quick life cycle (Boorman, 1974). Aside from *C. sonorensis*, success with *C. nubeculosus* rearing has led to extant colonies of both species being maintained in the USA (ARS-USDA; Kansas) and UK (Pirbright Institute; Surrey) (Nayduch *et al.*, 2014; Table 1.1.). As only *C. sonorensis* is a vector of BTV, the major data pertaining to experimental infections have been acquired from this species (Chandler *et al.*, 1985; Nunamaker *et al.*, 1997; Fu *et al.*, 1999). However, an interest in colonisation is not limited to only vector species. For example, the laboratory rearing of pest species, such as *C. impunctatus* (Hill, 1947; Carpenter, 2001), has been suggested to allow for integral studies of aggregation cues to increase trapping efficiency (Bhasin, Mordue and Mordue, 2001).

Furthermore, laboratory investigations into ovipositing and fecundity of this pest species could lead to control interventions at breeding-sites (Carpenter, Mordue and Mordue, 2001).

Maintenance of adults has been achieved using cardboard pill boxes or large mesh cages where midges are fed on sucrose-soaked cotton wool, or via artificial blood-feeding devices if initiation of a gonotrophic cycle is needed (Boorman, 1974; Barceló and Miranda, 2018). After mating and blood-feeding, ovipositing occurs on damp filter paper, although in the case of *C. impunctatus*, *Sphagnum* spp. moss can be used to increase ovipositing success (Carpenter, Mordue and Mordue, 2001). However, after hatching, maintenance of larvae has proven more difficult due to problems in identifying suitable substrates for development. Due to the diverse natural larval habitats of *Culicoides* (see section 1.2.2), it is unlikely that a single larval medium will be appropriate for universal use (Carpenter, 2001). Indeed, it should be noted that both *C. sonorensis* and *C. nubeculosus* develop in heavily manured soils, which may account for their rearing success using an organically polluted substrate (Hribar, 1990). In these cases, larval pans containing manure/soil and water, were optimised to include liver powder, yeast and wheat germ to promote bacterial growth (Megahed, 1956; Boorman, 1974). However, when introducing other important vector species into this system (e.g. *C. obsoletus*), high larval mortality was observed (Boorman, 1974). Another problem in investigating larval rearing conditions is the visualision of immatures through these substrates, which can make it difficult to assess mortality rates (Carpenter, 2001). Furthermore, visualisation problems complicate the monitoring of life stage development times, which is a hindrance to understanding the bionomics of many species (Barceló and Miranda, 2018). To overcome this, the rearing of larvae on agar dishes, with nematodes as a

Species	BTV vector	Colony status	References to maintenance techniques
C. arakawae	No	Discontinued	Sun, 1974
C. brevitarsis	Yes	Failed	Campbell, 1974
C. furens	No	Discontinued	Linley, 1968; Koch and Axtell, 1978
C. hollensis	No	Discontinued	Koch and Axtell, 1978
C. imicola	Yes	Failed	Veronesi *et al.*, 2009
C. impunctatus	No	Failed	Hill, 1947; Carpenter, 2001
C. melleus	Yes	Failed	Koch and Axtell, 1978
C. nubeculosus	No	Extant	Megahed, 1956; Boorman, 1974
C. obsoletus	Yes	Failed	Boorman, 1985
C. oxystoma	Potential	Discontinued	Sun, 1974
C. riethi	No	Discontinued	Boorman, 1974
C. sonorensis	Yes	Extant	Jones, 1957, 1960; Boorman, 1974

Table 1.1. Colonisation attempts of *Culicoides* species. Modified from Nayduch *et al.* (2014).

diet, has been trialled with success (Kettle, Wild and Elson, 1975; Linley, 1979; Aussel and Linley, 1994; Barceló and Miranda, 2018). This has allowed for the small-scale maintenance of populations used for developmental, behavioural and diet studies (Linley, 1985b; Barceló and Miranda, 2018).

1.3 Endosymbionts

1.3.1 Primary and secondary endosymbionts

Insects harbour a diverse range of endosymbionts and their association with hosts range from obligate mutualism to facultative parasitism (Kikuchi, 2009). Obligate endosymbionts tend to have long shared evolutionary history with their hosts and are termed "primary" symbionts (Ferrari and Vavre, 2011). A well-known example is *Buchnera aphidicola*, which provides essential amino acids to its aphid host, *Acyrthosiphon pisum*, with this association occurring for between 160-280 million years (Moran *et al.*, 1993). "Secondary" symbionts, on the other hand, are organisms that have a long-term association with their host but are not required for survival or fundamental physiological processes, although they may induce a range of phenotypic effects (Ferrari and Vavre, 2011). These facultative associations can come in several forms:

1) Symbiont protection to their hosts from natural enemies. The best studied example of this interaction occurs between the symbiont *Hamiltonella defensa* in pea aphids (*Acyrthosiphon pisum*) and a parasitoid wasp (*Aphidius ervi)*. In this case, *A. pisum* harbouring *Hamiltonella* had reduced levels of mortality after attack by *A. ervi* (Oliver *et al.*, 2003). A further example of parasitoid resistance is the protection of *Drosophila hydei* by *Spiroplasma* infection (Xie, Vilchez and Mateos, 2010).

2) Resilience to harsh environments through mechanisms such as thermal tolerance (Corbin *et al.*, 2017). An example being *Anaplasma phagocytophilum*, which induces their tick hosts to express an antifreeze glycoprotein gene to enhance their survival under cold conditions (Neelakanta *et al.*, 2010).
3) Symbionts can also induce reproductive manipulations. These often come in the form of biasing sex ratios to favour female offspring (Werren, Baldo and Clark, 2008; Ferrari and Vavre, 2011). This occurs due to the mostly maternal route by which endosymbionts are vertically transmitted, although paternal transmission has been rarely described in some cases (Moran and Dunbar, 2006; Watanabe *et al.*, 2014). As males are often (but not always) an evolutionary dead end, this leads to symbiont and female host fitness becoming entwined. Sex ratio distorting phenotypes include male-killing (e.g. *Rickettsia* in ladybirds; *Wolbachia* in butterflies), parthenogenesis induction (e.g. *Wolbachia, Rickettsia* and *Cardinium* in parasitoid wasps) and feminisation of males (e.g *Wolbachia* in woodlice) (Juchault *et al.*, 1994; Werren *et al.*, 1994; Dyson, Kamath and Hurst, 2002; Zchori-Fein *et al.*, 2002; Giorgini *et al.*, 2010). Another strategy to improve female host fitness is via cytoplasmic incompatibility (CI) (see section 1.1.2), which has evolved independently at least twice in the symbionts *Wolbachia* and *Cardinium* (Werren and O'Neill, 1997; Gotoh, Noda and Ito, 2007) and is crucial in driving the symbiont into populations for vector-control initiatives.

1.3.2 Endosymbiont detection

Detection of symbionts was initially conducted through Giemsa staining of insect tissues and then later transmission electron microscopy (TEM) (Hertig and Wolbach, 1924; Suitor and Weiss, 1961; Roshdy, 1968). However, a lack of genetic information made classification of

these bacteria difficult with many symbionts being termed “Rickettsia-like microorganisms” up until PCR and sequencing technologies allowed for the gathering of DNA sequence data and phylogenetic information. Subsequently, these advances have allowed for the large-scale targeted screening of symbionts, which has been invaluable in assessing symbiotic associations in time and space (Simões *et al.*, 2011; Weinert *et al.*, 2015). Due to low titres leading to false negatives through conventional PCR, more sensitive assays such as nested and quantitative PCR have been used recently (Simoncini *et al.*, 2001; Wolfgang *et al.*, 2009; Mee *et al.*, 2015). This has been important as many symbiont-induced host phenotypes are titre dependent. For example, low-titre *Wolbachia* infections can lead to a weakened CI phenotype in *Dosophila melanogaster* (Merçot and Charlat, 2004). One problem of relying solely on sensitive molecular techniques, to define endosymbiotic associations, is the possibility of false positives through laboratory or natural contamination of a symbiont from a parasitoid of the targeted host (Wolfgang *et al.*, 2009; Ramage *et al.*, 2017). Therefore, true endosymbiotic associations can only be truly confirmed through symbiont-specific imaging techniques such as fluorescence *in-situ* hybridisation (FISH) of host tissues (Koga, Tsuchida and Fukatsu, 2009). The localisation of symbionts to specific tissues can also give information on transmission strategies of the symbiont as well as provide clues to host effects. For example, the detection of *Blattabacterium*, in the fat body urocytes of cockroaches, is linked to the assistance of nitrogen metabolism within these cells (Sabree, Kambhampati and Moran, 2009). Furthermore, identification of tropisms to the germline are also important in order to confirm symbiont vertical transmission and reproductive manipulations (Werren, Baldo and Clark, 2008).

1.3.3 Endosymbionts of *Culicoides*

Despite an increasing interest in insects and their endosymbionts, there are few studies relating to biting midges and their microbiota. A 16S metagenomic screening of *C. sonorensis* gut samples revealed amplicons allied to several genera, including *Rickettsia* (Campbell *et al.*, 2004), albeit with no phylogenetic or population-based information. Despite this, a subsequent targeted screening attempt (Lewis *et al.*, 2014) failed to detect any *Rickettsia* symbionts. Nakamura *et al.* (2009) initially discovered the symbionts *Cardinium* and *Wolbachia* in Japanese *Culicoides* before subsequent findings in midges from Australasia and Palearctic regions (Morag *et al.*, 2012; Lewis *et al.*, 2014; Mee *et al.*, 2015; Pagès *et al.*, 2017). Despite *Wolbachia* being the more widely researched of the two endosymbionts, interest has focussed on *Cardinium* as its discovery in a blood-feeding insect is the first of its kind (Nakamura *et al.*, 2009). Of the *Culicoides* species known to be infected with *Cardinium*; *C. imicola*, *C. obsoletus*, *C. pulicaris*, *C. punctatus* and *C. brevitarsis* are known vectors of bluetongue, having been directly linked with epidemics of BTV (Muller *et al.*, 1982; Goffredo *et al.*, 2015).

1.3.4 Cardinium

Candidatus Cardinium hertigii (Bacteroides) is a genus of bacteria that has multiple effects on the host biology of numerous arthropods. It is estimated between 6-13% of arthropod species carry *Cardinium* (Weinert *et al.*, 2015), making it the third most common endosymbiont-arthropod interaction documented (behind *Wolbachia* ~52% and *Rickettsia* ~24%). This endosymbiont forms “hot spots” in arachnids such as spiders and mites, but also has an affinity for parasitoid wasps and whiteflies (Zchori-Fein *et al.*, 2004; Gotoh, Noda and Ito, 2007; Duron *et al.*, 2008; Nakamura *et al.*, 2009). Classification of *Cardinium* is based on the

Gyrase B and 16S rRNA genes which suggest the presence of at least four groups designated as A, B, C and D (Nakamura *et al.*, 2009; Edlund *et al.*, 2012). Group A is the largest and encompasses a variety of hosts including diaspidid scale insects, planthoppers and spider mites (Gruwell, Wu and Normark, 2009; Nakamura *et al.*, 2009); B includes parasitic nematodes (Noel and Atibalentja, 2006; Brown *et al.*, 2018); C is so far exclusive to *Culicoides* biting midges (Nakamura *et al.*, 2009; Morag *et al.*, 2012; Mee *et al.*, 2015; Pagès *et al.*, 2017); and D is the most recent addition found in the marine copepod *Nitocra spinipes* (Edlund *et al.*, 2012). Additionally, the recent discovery of a *Cardinium* infected non-marine ostracod (Schön *et al.*, 2019), which doesn't place in any of these groups based on a 16S phylogeny, suggests the host range of this symbiont is more extensive than previously thought.

Cardinium can induce CI in both spider-mites and parasitoid wasps (Hunter, Perlman and Kelly, 2003; Gotoh, Noda and Ito, 2007; Perlman, Kelly and Hunter, 2008; Ros and Breeuwer, 2009), parthenogenesis (Zchori-Fein *et al.*, 2002), as well as inducing feminisation in several spider mite species (Weeks, Marec and Breeuwer, 2001; Groot and Breeuwer, 2006). *Cardinium* also has the capacity to improve host fitness by increasing longevity in the wasp *Encarsia inaron* (White *et al.*, 2011) and fecundity in the mite *Metaseiulus occidentalis* (Weeks and Stouthamer, 2004). B-vitamin provisioning has also been a suggested phenotype, which is of significance for any *Cardinium*-infected haematophagous arthropods, as blood is deficient in such micronutrients (Rio, Attardo and Weiss, 2016). However, this appears not to be the case in *Cardinium*-infected *Culicoides punctatus*, where the recently published genome (*c*Cpun) suggests the symbiont is a biotin sink as opposed to a provider (Siozios *et al.*, 2019).

1.3.5 Rickettsia

The genus *Rickettsia* lies within the alphaproteobacteria, and are members of the Rickettsiaceae family being closely related to *Wolbachia* (Order: Rickettsiales) (Williams, Sobral and Dickerman, 2007). *Rickettsia* was initially identified as the causative agent of the human disease epidemic typhus, spread by the human biting louse *Pediculus humanus* (Gross, 1996). Further severe vertebrate diseases caused by *Rickettsia* were then discovered associated with ticks, including Rocky Mountain spotted fever (Fricks, 1916). Although these early discoveries of *Rickettsia* involved a vertebrate host as part of the transmission cycle, this is atypical of the genus, with subsequent findings of *Rickettsia* groups with no known secondary host (Perlman, Hunter and Zchori-Fein, 2006; Weinert *et al.*, 2009; Figure 1.1.). Subsequently, the known host range of this group now encompasses a wide range of aquatic, terrestrial and blood-feeding taxa identified including water beetles, sand flies, leeches and amoebae (Dyková *et al.*, 2003; Kikuchi and Fukatsu, 2005; Reeves, Kato and Gilchriest, 2008; Küchler, Kehl and Dettner, 2009). The discovery of male-killing *Rickettsia* in *Adalia* ladybirds in the 1990s suggested this genus also contributed to reproductive effects similar to other endosymbionts such as *Wolbachia* (Werren *et al.*, 1994). Indeed, since then, parthenogenesis induction has also been observed in the parasitoid wasps *Pnigalio soemius* and *Neochrysocharis formosa* (Hagimori *et al.*, 2006; Giorgini *et al.*, 2010). Similar to *Wolbachia*, *Rickettsia* symbiont infection appears to be protective against some pathogens; enhancing resistance of aphids (*Acyrthosiphon pisum*) to fungal attack, and whiteflies (*Bemisia tabaci*) to bacterial challenge (Łukasik *et al.*, 2013; Hendry, Hunter and Baltrus, 2014). Multiple vertical and horizontal transmission routes have been established for the *Rickettsia* genus. Like the common endosymbionts *Wolbachia* and *Cardinium*, most members are maternally

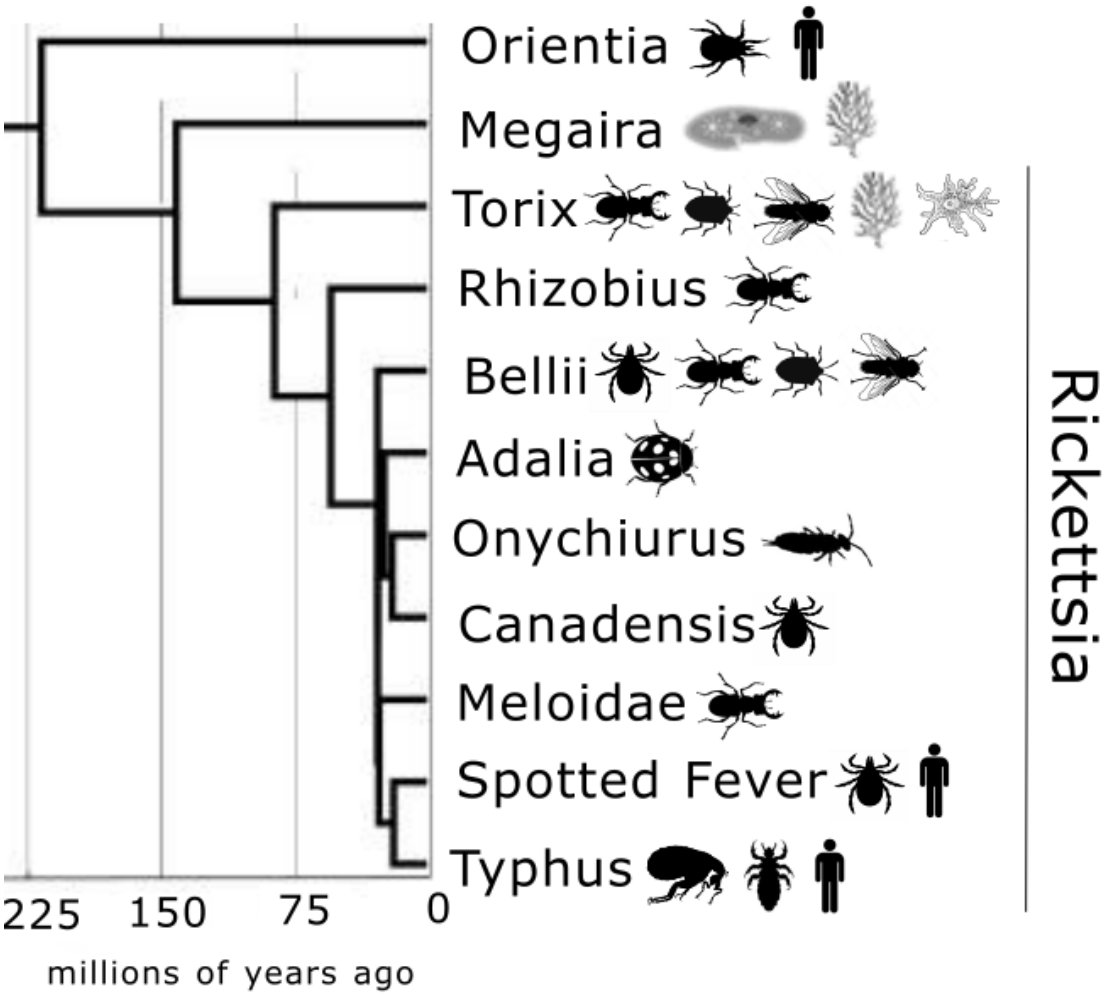


Figure 1.1. A 16S rRNA phylogenetic tree with molecular clock showing the evolutionary relationships among lineages within the *Rickettsiaceae* family of *Rickettsiales*. On the right side are illustrations of typical host organisms. These include mites; ticks; true bugs; beetles; true flies; springtails; lice; fleas; ladybirds; humans/vertebrates; aquatic invertebrates including hydra and leeches (depicted by coral); amoeba; algae/ciliates. Modified from Weinert *et al.* (2009) and Castelli *et al.* (2016).

transmitted (Weinert *et al.*, 2009). However, a peculiarity of note is the paternal transmission observed in the leafhopper *Nephotettix cincticeps* (Watanabe *et al.*, 2014).This additional vertical transmission route has been suggested to allow for the drive of a symbiont into a population when maternal transmission is incomplete. *Rickettsia* has also been suggested to have strategies allowing for host shifts, which include blood feeding where co-feeding on vertebrate hosts allows for the transmission of spotted fevers (Wood and Artsob, 2012). Additionally, there is some evidence to suggest plant-feeders could acquire *Rickettsia* with phloem acting as an intermediary reservoir (Chrostek *et al.*, 2017). This suggestion is supported by the transmission of a Belli group *Rickettsia* from infected whiteflies to non-infected conspecifics feeding sequentially on the same leaves (Caspi-Fluger *et al.*, 2012). In addition, a *Rickettsia* member is the causative agent of the plant disease, papaya bunchy top disease (Luis-Pantoja *et al.*, 2015). Subsequently, *Rickettsia* represents an intriguing endosymbiotic group with numerous transmission routes and ecological anomalies.

1.3.6 Endosymbionts and the confounding of DNA barcoding studies

Although mtDNA barcoding has been used widely for the taxonomic classification of *Culicoides* (see section 1.2.5), the presence of symbionts can often confound studies using this technique. As endosymbionts and mitochondria are generally co-inherited together (via the maternal line), they are often observed in linkage disequilibrium when a symbiont selective sweep occurs in a population. Subsequently, indirect selection on mtDNA can also occur in insects containing maternally inherited endosymbionts (Hurst and Jiggins, 2005). This can lead to the confounding effects of biodiversity studies where one of the assumptions is that the mtDNA marker is under neutral selection. Specific effects can include the fragmenting

or homogenising of mitochondrial lineages. For example, the erasure of mtDNA diversity within a population can occur as a result of selective sweeps driven by a strain conferring CI, which can lead to an underestimation of diversity (Ballard *et al.*, 1996; Jiggins, 2003; Shoemaker *et al.*, 2004; Zhang *et al.*, 2013). Conversely, when more than one symbiont strain enters a population, this can lead to the maintenance of divergent mtDNA haplotypes (Schulenburg *et al.*, 2002; Jiggins, 2003; Kvie *et al.*, 2013). Finally, the conflation of differing species through past mitochondrial introgression can occur when an endosymbiont transfers during a hybridisation event, driving into a population and carrying with it the heterospecific mtDNA (Jiggins, 2003; Whitworth *et al.*, 2007; Raychoudhury *et al.*, 2009; Jäckel, Mora and Dobler, 2013).

Another confounding effect of symbionts is related to the inadvertent amplification of bacteria DNA (Linares *et al.*, 2009; Smith *et al.*, 2012; Ceccarelli, Haddad and Ramírez, 2016). As the hypothetical bacterial ancestor of mitochondria (proto-mitochondria) is thought to be closely related to *Wolbachia* and *Rickettsia* (Wang and Wu, 2015), DNA barcoding can erroneously amplify these symbionts instead of the host target. This phenomenon was first described when the barcoding of lepidopterans retrieved several *Wolbachia COI* sequences (Linares *et al.*, 2009). Since then, a large scale systematic review using the Barcoding Of Life Data System (BOLD), assessing over 2 million sequences, revealed that 0.16% of barcoding attempts of insects led to similar unintended *Wolbachia* amplifications (Smith *et al.*, 2012). More recently, *Rickettsia* has also been observed to produce similar outcomes. For example, the erroneously barcoding amplification of *Rickettsia* has been observed in two families of spider (Anyphaenidae and Dysderidae) (Řezáč *et al.*, 2014; Ceccarelli, Haddad and Ramírez,

2016). Subsequently, this widespread direct and indirect confounding suggests the interpretation of barcoding studies should only be considered in the context of past and present symbiont infections.

1.4 Thesis aims

Overall, this thesis had 3 key aims:

1) To investigate the occurrence and distribution of *Cardinium* and *Rickettsia* in *Culicoides* from Palearctic and Afrotropical regions.

2) To assess potential transmission dynamics and host effects of *Rickettsia* in *Culicoides* through phylogenetic analysis and the localisation of the symbiont to specific tissues.

3) To determine the confounding effects of *Rickettsia* and *Cardinium* on DNA barcoding studies.

The first section of this thesis aimed to investigate the symbiotic interactions of *Cardinium* and *Culicoides* through a targeted screening of biting midge species. Although, *Culicoides* populations from various biogeographic regions were assessed, a specific focus was given to UK species to identify candidates to undertake further host fitness and infection work. Specifically, this relied on rescreening the only two UK vector species identified as being *Cardinium*-infected, *C. pulicaris* and *C. punctatus* (Lewis *et al.*, 2014). It was of importance to assess these species for polymorphic infections as this would provide a suitable population for vector competence experiments. This is because the naturally uninfected individuals from these populations could be used as negative controls without the confounding effects of antibiotic use. This initial investigation also aimed to assess the prevalence and distribution of *Cardinium* infections in *Culicoides* to assess clustering patterns of geography and host

phylogeny. Finally, as symbionts can confound the interpretation of biodiversity studies which use DNA barcoding, the well-known vector species *C. imicola* is used to assess if mtDNA haplotypes reflect *Cardinium* presence rather than host gene flow patterns.

Research on symbionts of *Culicoides* species is in its infancy. Subsequently, it was unsurprising when the common *Rickettsia* endosymbiont was uncovered through the serendipitous shallow whole genome sequencing of *C. newsteadi* by a colleague (Stefanos Siozios, University of Liverpool). With limited opportunities to identify an *in vivo* system to study midge-symbiont interactions arising from the first section of this thesis, a focus was shifted to identify such a system in relation to *Rickettsia*, which required its own targeted screening presented in chapter 3. This chapter not only described the host diversity of *Rickettsia*, but through prevalence and phylogenetic analysis, an aim was to identify ecological dynamics which are imperative in understanding the potential for novel biological intervention involving *Rickettsia* in the future.

These symbiont screens then informed the fourth chapter, which describes the tissue tropisms of *Rickettsia* in the UK pest species *C. impunctatus*. Despite previous difficulties in rearing *C. impunctatus* (Carpenter, 2001), cultivation of immatures for processing were successful, which allowed for tropism information to be gathered for various life stages. Through transmission electron microscopy (TEM) and fluorescence *in-situ* hybridisation (FISH) the cellular and subcellular localisation of *Rickettsia* in both somatic and germline tissues is described. This advances the understanding of *Rickettsia*-midge interactions by contributing knowledge to transmission dynamics as well as potential host effects induced by the symbiont. Due to *Rickettsia*'s potential to increase host fitness (Kikuchi and Fukatsu, 2005),

identification of such effects would be essential for any future control program for this economically important pest species.

Comparing sequences identified in the *Rickettsia* screening of chapter 3 to online databases, it was found that hundreds of supposed mitochondrial (barcode) sequences from Genbank (NCBI; http://www.ncbi.nlm.nih.gov/genbank/) and BOLD (Barcoding of Life Data System; http://www.boldsystems.org/) had a high homology to *Culicoides Rickettsia*. This indicated barcoding studies were often amplifying *Rickettsia* DNA instead of host mtDNA. Subsequently, through a systematic review of *in silico* barcoding sequences, the final section of this thesis investigates how common this phenomenon of inadvertent *Rickettsia* amplification is. Through special access to DNA extracts producing erroneous *Rickettsia* amplifications (provided by BOLD), the screening and sequencing of housekeeping genes used in chapter 3 is used to expand our phylogenetic knowledge of the *Rickettsia* genus. A further aim of this study was to access if DNA barcoding could be used as an adjunctive to targeted *Rickettsia* screens to uncover host new host interactions which would inform future study systems of the underexplored “ancestral” group. Finally, a discussion of the results aims to summarise the findings from all chapters with a focus on the prospects of *Rickettsia*’s contribution to arbovirus transmission dynamics.

Chapter 2: Assessing the Phylogeographic Distribution of *Candidatus Cardinium hertigii* in *Culicoides* with a Focus on Two UK Vector Species

2.1 Abstract

The endosymbiont *Candidatus Cardinium hertigii* (Bacteroidetes), has previously been described in *Culicoides* (Diptera: Ceratopogonidae) populations from Japan, Australia, Israel, Spain and UK, with the latter account suggesting infection of both *C. pulicaris* and *C. punctatus*, putative vectors of bluetongue virus in Europe. Due to the lack of a model UK *Culicodies* species to test *Cardinium*-midge interactions, this chapter aimed to assess the suitability of both species. The attempted replication of the previous UK *Cardinium* screening of *C. pulicaris* and *C. punctatus* showed several errors in the original methodology and interpretation of *Cardinium* infection status, suggesting neither are suitable for future *in vivo* work. In addition, nested PCR screening was undertaken for 338 individuals of 24 species of biting midges from both Palearctic and Afrotropical regions to assess patterns of geography and host phylogeny. Infections were observed in 64 individuals out of 9 species, with prevalence ranging from low and intermediate, to fixation. These data corroborate previous estimates of *Cardinium* prevalence in midges. Phylogenetic analysis based on the Gyrase B gene grouped new isolates within "group C" of the genus, a clade which has so far been exclusive to *Culicoides*, and in which all *Culicoides Cardinium* infections described to date lie. Additionally, the concordance of mitochondrial haplotypes (mitotypes) and *Cardinium* infection in populations of *C. imicola* suggests a potential confounding of phylogeographic studies based on this marker. The biological significance of *Cardinium* in midges, and effects on vector capacity, are still to be elucidated.

2.2 Introduction

Worldwide, biting midges of the genus *Culicoides* (Diptera: Ceratopogonidae) are known to transmit more than 50 arboviruses. Notably, bluetongue virus (BTV) and Schmallenberg virus (SBV) pose a great threat to livestock and wild animal welfare (Mellor, Boorman and Baylis, 2000). Outbreaks of BTV have caused serious economic damage to the European livestock industry (Wilson and Mellor, 2009). For example, a single epidemic of BTV in the UK has been estimated to potentially cost up to £485M with losses in livestock productivity, restrictions on UK agricultural exports and employment impacts cited as the main contributors to this figure (Pirbright Institute, 2009). Current control methods of BTV rely on vaccines which, due to the quick emergence and uncertainty of circulating strains, can be ineffective. The recent introduction of the vector species *Culicoides imicola* into continental Europe, as a result of climate change and wind-transport events (Purse *et al.*, 2005; Venail *et al.*, 2012; Jacquet *et al.*, 2016), has also highlighted the challenge of mapping midge-borne pathogens. Thus, novel vector control measures are seen to be more attractive options in controlling the spread of these arboviruses (Carpenter, Mellor and Torr, 2008).

Intracellular bacteria (endosymbionts) and insects frequently form symbiotic associations, some of which have been implicated in host defence against pathogens. As such, this symbiosis can have profound effects on the biology, ecology and evolutionary dynamics of host-pathogen interactions (Ferrari and Vavre, 2011). Importantly, the endosymbiont *Wolbachia* has demonstrated the ability to transinfect mosquito species leading to a blocking effect of dengue virus, chikungunya virus, yellow fever virus and West Nile virus (Bian *et al.*, 2010; Blagrove *et al.*, 2012; Hussain *et al.*, 2012; van den Hurk *et al.*, 2012). The potential to

extrapolate these effects to wild populations of *Aedes* mosquitoes is enabled by *Wolbachia*-induced cytoplasmic incompatibility (CI); embryo death in mating between infected males and uninfected females (Werren and O'Neill, 1997; Bourtzis *et al.*, 2014). CI is used as mechanism to drive the *Wolbachia* into a population to distribute this virus blocking phenotype into a naïve population (Mouton *et al.*, 2003; Miller, Ehrman and Schneider, 2010; Hoffmann *et al.*, 2011, 2014). In addition, this can allow for the suppression of vector numbers when releasing *Wolbachia*-infected male *Aedes aegypti* (Mains *et al.*, 2019; Zheng *et al.*, 2019). Encouragingly, the results of field-based CI-inducing symbiont interventions indicate a viable method to control arboviruses (Hoffmann *et al.*, 2014). Importantly, this protective phenotype appears to be associated with RNA viruses leading to the potential for symbiont-based biocontrol in major midge-borne pathogens such as BTV and SBV.

Candidatus Cardinium hertigii (Bacteroidetes) is an endosymbiont which has been widely associated with *Culicoides* (Nakamura *et al.*, 2009; Morag *et al.*, 2012; Mee *et al.*, 2015; Pagès *et al.*, 2017). Of the *Culicoides* species known to be infected with *Cardinium; C. imicola, C. obsoletus, C. pulicaris, C. punctatus* and *C. brevitarsis* are known vectors of BTV, having been directly linked with epidemics of bluetongue (Muller *et al.*, 1982; Goffredo *et al.*, 2015). It is estimated up to 13% of arthropod species carry *Cardinium* (Weinert *et al.*, 2015) with induced phenotypes including the reproductive manipulations feminisation (Weeks, Marec and Breeuwer, 2001; Groot and Breeuwer, 2006) and parthenogenesis (Zchori-Fein *et al.*, 2002). CI has also been observed in *Cardinium* in both spider-mites and parasitoid wasps (Hunter, Perlman and Kelly, 2003; Gotoh, Noda and Ito, 2007; Perlman, Kelly and Hunter, 2008; Ros

and Breeuwer, 2009) suggesting this could be a driving mechanism for future introductions into *Culicoides* populations.

Mitochondrial DNA (mtDNA) markers have been widely used in the taxonomic classification of animals due to its high evolutionary rate, suitable effective population size and low recombination rate when compared to nuclear markers (Hurst and Jiggins, 2005). As well as being utilised as a molecular technique to identify arthropods, mtDNA "barcodes" have frequently been used in phylogeographic studies to detect biodiversity and dispersal events (Craft *et al.*, 2010; Papadopoulou *et al.*, 2011; Baselga *et al.*, 2013). Despite this utility, these studies can occasionally be confounded by the linkage disequilibrium of endosymbionts and host mtDNA. As endosymbionts enter a naïve population, selective sweeps of the bacteria can occur and with it the homogenisation of linked mtDNA haplotypes, leading to the apparent absence of biodiversity (Hurst and Jiggins, 2005). Conversely, the introduction of more than one endosymbiont in linkage disequilibrium with an mtDNA haplotype can lead to a perceived genetic population structure when there is a lack of one. Subsequently, inferences from barcoding studies can only be interpreted in the context of recent infection (or lack thereof) with endosymbiotic bacteria. This may therefore confound attempts to reconstruct vector history of invasion and spread.

The motivation for this study stems from the need for a model system to study *Cardinium*-*Culicoides* interactions. First, the phylogeographic distribution of *Cardinium* infections is investigated in *Culicoides* populations from Palearctic and Afrotropical regions through screening using a nested PCR. A focus is placed on *C. pulicaris* and *C. punctatus*, two UK vector species described as being *Cardinium*-infected (Lewis *et al.*, 2014), with the intention of

assessing their suitability for future *in vivo* work. A polymorphically infected population (containing infected and uninfected individuals), with high numbers in the wild, is preferable for such studies due to readily available negative (uninfected) controls. Finally, this chapter aims to establish associations between mitochondrial haplotypes (mitotypes) and *Cardinium* infection in the globally important vector *Culicoides imicola*, as this could have implications for inferring movement of this species into naïve geographic areas.

2.3 Methods

2.3.1 *Culicoides* collection and identification

Overall, 338 specimens of 24 *Culicoides* species were collected using light traps from May 2007 to July 2016 across 17 sites spanning France, South Africa, Sweden and the UK. Populations collected by the author included all midges collected from 2015 onwards from the UK and South Africa. Other specimens were provided by Ken Sherlock (Institute of Infection and Global Health, Liverpool, UK), Dr Jan Chirico (National Veterinary institute, Uppsala, Sweden) and Dr Claire Garros (CIRAD, Agricultural Research for Development, Montpellier, France). Those obtained by the author were collected through the use of both CDC (6-volt motorbike battery) and Onderstepoort light traps (mains) utilising an ultraviolet Light Emitting Diode (LED). Generally, traps were placed above two metres at sites before dusk and collected at dawn the next day. A downward draught from a fan directed insects into a beaker filled with water and a small amount of detergent to break surface tension. After collection, all midge specimens were stored in 70% ethanol for preservation before being sexed and separated morphologically down to the species level using relevant keys (Downes and Kettle 1952; Campbell and Pelham-Clinton 1960; Delécolle 1985; Meiswinkel 1994). Sampled species included both vectors and non-vectors of BTV. Morphological identification of *Cardinium* positive specimens was confirmed by sequencing a fragment of the mitochondrial cytochrome c oxidase subunit 1 (*COI*) barcode (Pagès *et al.*, 2009; Ander, Troell and Chirico, 2013; Nielsen and Kristensen, 2015).

2.3.2 DNA extraction

DNA extractions were prepared based on the protocol of Ander *et al.* (2013). Briefly, whole individual insects were washed in molecular grade water, transferred to blue roll and dried at

room temperature for approximately two minutes. Using a sterile pipette tip with a melted tip, the specimen was transferred to an Eppendorf tube, where 30 µL of PrepMan Ultra (Life Technologies, Inc., Carlsbad, CA, USA) was added. The specimen was homogenized using the same sterile pipette tip until no macroscopic structure was visible and then briefly centrifuged. The preparation was incubated at 100 °C for ten minutes, and samples were then kept on ice for two minutes. Remaining fragments of tissue were removed from the homogenate by centrifugation at 12,000 × g for three minutes, before 20 µL of the supernatant was transferred to a fresh Eppendorf tube. Before the above fresh samples were available, DNA extracts from a previous study by Lewis *et al.* (2014), describing the positive infection status of *C. pulicaris* and *C. punctatus* in the UK, were used to test screening assays.

2.3.3 *COI* and *Cardinium* screening

Amplification of the *COI* gene was initially assessed as a means of quality control by conventional PCR assay using the commonly used primer set HCO1490/LCO2198 (Folmer *et al.*, 1994). For individuals which failed to produce a positive result, a secondary set of primers C1-J-1718/C1-N-2191 were utilised (Dallas *et al.*, 2003). Samples that did not amplify with either primer set were deemed to contain low quality DNA and were removed from further analysis. DNA extracts which passed quality control were then screened for *Cardinium* using conventional and nested primers amplifying partial sequences of the Gyrase B *(GyrB*) gene. PCR assays consisted of a total of 15 µL per well, comprising of 7.5 µL GoTaq® Hot Start Polymerase (Promega), 5.1 µL nuclease free water, 0.45 µL forward and reverse primers (concentration 10 pmol/µL) and 1.5 µL DNA template. PCR products were separated on 1% agarose gels stained with Midori Green Nucleic Acid Staining Solution (Nippon Genetics

Europe). Primer sequence details and PCR cycling conditions can be found in Tables 2.1.1 and 2.1.2.

2.3.4 Nested primer design and sensitivity

Nested primers were designed by aligning known *Cardinium hertigii GyrB* sequences (Table 2.2.) available in GenBank (NCBI) using the ClustalW algorithm in Mega6 (Tamura *et al.*, 2013). In order to identify the correct reading frame, both complementary strands in all three reading frames were translated via the ExPasy translate tool (Gasteiger *et al.*, 2003) into 6 different amino acid sequences. The result that gave the longest amino acid sequence, before a stop codon was encountered, was deemed the correct reading frame. Specific nested primers were then designed based on homology inferred from the multiple alignment of *GyrB* sequences using the online tool PriFi (Fredslund *et al.*, 2005) and its default parameters. The stable anchoring of the primers' 3' termini was ensured by the inclusion of two-fold degenerate sites at the third position of codons to avoid effects of base-wobble pairing. The diagnostic sensitivity of conventional and nested PCR assays was quantified by serially diluting *Cardinium* positive DNA extracts using 10-fold dilutions of nuclease-free water. Dilutions ranged from undiluted to 10^{-6} for both *C. punctatus* (n=5) and *C. imicola* (n=2) DNA templates. Nested PCR is prone to contamination and subsequently false positives. Precautions undertaken to avoid this outcome include repeated disinfecting of surfaces**,** separation of lab space between sample preparation and amplification handling, the use of filter tips and multiple (four) negative controls.

Target	Primer name	Sequence (5'-3')	Tm (°C)	Size (bp)	Reference
Cytochrome oxidase subunit 1 (*COI*)	LCO2198	GGTCAACAAATCATAAAGATATTGG	51	708	Folmer *et al.* 1994
	HCO1490	TAAACTTCAGGGTGACCAAAAAATCA	56		
	C1-J-1718	GGAGGATTTGGAAATTGATTAGT	52	523	Dallas *et al.* 2003
	C1-N-2191	CAGGTAAAATTAAAATATAAACTTCTGG	50		
Gyrase B	gyrB23F	GGAGGATTACATGGYGTGGG	60	1368	Lewis *et al.* 2014
	gyrB1435R	GTAACGCTGTACATACACGGCATC	60		
	gyrBnest212F	AAGGCAACCCTATGCACCAA	59	347	This study
	gyrBnest654R	GGYCTTAGTTTGCCCTTCAAATTG	59		

Table 2.1.1. *COI* and *Gyrase B* gene primer attributes.

Target	Initialisation	Denaturation, annealing, extension	Final extension, hold	Number of cycles
Cytochrome oxidase subunit 1 (COI)	95°C/5min	95°C/30s, 50°C/1min, 72°C/1min	72°C/7min, 15°C/∞	35
Gyrase B (Conventional)	95°C/5min	95°C/30s, 55°C/1min, 72°C/1min30s	72°C/7min, 15°C/∞	35
Gyrase B (Nested)	95°C/5min	95°C/25s, 55°C/30s, 72°C/45s	72°C/7min, 15°C/∞	35

Table 2.1.2 PCR cycling conditions

***Cardinium* host**	**Genbank accession numbers**
Amoebophilus asiaticus (Outgroup)	AB506793
Amphitetranychus quercivorus	KU323553
Culicoides arakawae (Japan)	AB506791
Culicoides brevitarsis (Australia)	KR026924
Culicoides henryi (Australia)	KR026926
Culicoides imicola (South Africa/Kenya/Israel)	This study/KR026927/JN166963
Culicoides multimaculatus (Australia)	KR026928
Culicoides newsteadi N2 (Corsica)	This study
Culicoides ohmorii (Japan)	AB506792
Culicoides oxystoma (Israel)	JN166964
Culicoides pulicaris haplotype 2 (Corsica)	This study
Culicoides punctatus (UK)	This study/HG380244
Culicoides victoriae (Australia)	KR026931
Culicoides williwilli (Australia)	KR026935
Encarsia pergandiella	HE983995
Euides speciosa	AB506788
Heterodera glycines	CP029619
Indozuriel dantur	AB506789
Oligonychus ilicis	AB506783
Pratylenchus penetrans	PRJNA308318
Sogatella furcifera	CP022339
Tetranychus puerariсola	AB506784
Tetranychus suginamensis	AB506782
Tetranychus urticae	KU323550

Table 2.2. Accession numbers of taxa used for *Gyrase B* nested primer design and reference sequences in Figure 2.5.

2.3.5 Mapping *Cardinium* spatial distribution

Collection sites of *Culicoides* populations were plotted against *Cardinium* infection prevalence based on the nested PCR assay and mapped using Quantum GIS v3 software (Quantum GIS, 2016). The world borders shapefile was obtained from Thematicmappings.org under an Attribution-Share Alike License (Bjorn Sandvik, 2009), accessed December 2017.

2.3.6 Sanger sequencing

Amplicons identified by gel electrophoresis were subsequently purified enzymatically (ExoSAP) before being sent for sequencing if in batches of <24 (GATC Biotech AG, Konstanz, Germany) or sequenced using a BigDye® Terminator v3.1 kit (Thermo Scientific, Waltham, USA), and capillary sequenced on a 3500 xL Genetic Analyser (Applied Biosystems, Austin, USA) for larger batch numbers. Amplicons were sequenced through both strands for *GyrB* and through single strands for *COI*. Generated chromatograms were assessed visually for correct base calls with consensus sequences being created in Ugene v1.22 (Okonechnikov *et al.*, 2012).

2.3.7 Phylogenetic analysis

The relatedness of *Cardinium* strains from different host species was analysed using nucleotide sequences from conventional PCR. *GyrB* was chosen for phylogenetic analysis because it has a higher divergence when compared to the conserved 16S rRNA gene, another gene used in *Cardinium* phylogeny reconstruction. *Cardinium GyrB* sequences from groups A, B and C were retrieved from GenBank (NCBI) (Table 2.2.) to improve phylogenetic resolution. Sequences were aligned using the "L-INS-I" algorithm in MAFFT v7 (Katoh and Standley, 2013). A maximum likelihood (ML) phylogeny was inferred with RAxML v8 (Stamatakis, 2014) using 1000 rapid bootstrap replicates and using the GTR + I + G model which was selected with

jModelTest 2 (Darriba *et al.*, 2012) using the Akaike information criterion, with the topology search taking the best of Subtree Pruning and Nearest Neighbour Interchange rearranging. Finally, trees were drawn using the EvolView (He *et al.*, 2016) online tree annotation and visualization tool.

2.3.8 mtDNA haplotype networks of *Cardinium* infected and uninfected *C. imicola*

Phylogenetic relationships between *Cardinium* positive and negative *C. imicola* from South Africa were assessed using a haplotype network of a 517 bp region of the *COI* gene. The network was then extended to include *C. imicola* from France, Corsica (this study) as well as Portugal and Israel (Dallas *et al*. 2003; Accession numbers: AF078098–AF078100, AF080531, AF080532, AF080534, AF080535, AJ549393–AJ549426) where varying *Cardinium* infection status had previously been described. Analysis and visualisation was undertaken using the TCS haplotype network algorithm generated in PopART v1 (Leigh and Bryant, 2015).

2.3.9 *Cardinium* prevalence comparisons of site, species and sex

Differences in *Cardinium* prevalence were tested for each collection site, species and sex using the two tailed Fisher's exact test. More specifically, this was achieved by comparing the same species at different collection sites where at least one population had a *Cardinium* positive. All statistical analyses in this study was undertaken using Rstudio version 1.1.419 (Rstudio Team, 2015).

2.4 Results

2.4.1 Clarification of the *Cardinium* infection status of *C. pulicaris* and *C. punctatus*

The rescreening of DNA extracts from the Lewis study provided contrasting results with those described in the original study. Of the 26 *C. punctatus* DNA extracts screened, 25/26 (96%) were positive for *Cardinium*, all of which were validated for sufficient DNA quality via PCR amplification of the *COI* subunit using the HCO1490/LCO2198 primer set. Despite adequately replicating Lewis' findings for *C. punctatus*, only 12/39 (31%) *C. pulicaris* samples were positive when using the HCO1490/LCO2198 primer pair, as opposed to 39/39 (100%) claimed in the previous study. Incidentally, the 12 *COI* positive samples were the same DNA extracts which came back PCR positive for the *Cardinium GyrB* screen as documented by Lewis. Furthermore, fresh *C. pulicaris* caught in the autumn of 2015 did not show positive signals for amplification of *COI* using HCO1490/LCO2198 primers. This led to the likelihood that there was a polymorphism at one of the forward or reverse primer sites designed to amplify *COI* in *C. pulicaris*. To test this, a new *COI* locus was chosen to observe whether the previous negative samples would become positive. The secondary quality control primers C1-J-1718/C1-N-2191 (Dallas *et al.*, 2003) amplified all the *C. pulicaris* DNA extracts confirming this premise (Table 2.3.2).

Although all of these samples had now been confirmed to have adequate DNA quality, there were still the unresolved issues of how Lewis had detected *COI* positive *C. pulicaris* with the invalid HCO1490/LCO2198 primers, and why the re-screened HCO1490/LCO2198 positives were coincidentally the only samples to come back positive for *Cardinium*. Obtaining the raw data from Lewis' study resolved the first of these issues. Examination of Lewis' gel images

corroborated my re-screening data (Figure 2.1.); 12/39 *C. pulicaris* samples gave positive bands after amplification with HCO1490/LCO2198 primers despite Lewis claiming that all passed the *COI* quality control screen. To explain the matter of why there were any HCO1490/LCO2198 amplified *C. pulicaris* extracts to begin with, there were several possible explanations: the DNA extracts had been cross contaminated with other *Culicoides*' extracts through a pipetting error; the individuals had been misidentified when morphologically distinguishing the *Culicoides* species and were in fact another species*;* or the individuals morphologically appeared as *C. pulicaris* but were an unidentified cryptic species which lack the polymorphism at the HCO1490/LCO2198 site and are infected with *Cardinium*. To distinguish between these possibilities, Sanger sequencing of the purified PCR product of both HCO1490 and C1-N-2191 was undertaken and compared to known mitochondrial DNA barcodes on Genbank. BLASTn alignment identified the HCO1490 product to be 99% homologous to *C. punctatus* COI (Accession number: KX064676), whereas the C1-N-2191 product was 100% identical to *C. pulicaris* COI (Accession number: KJ624116). As both loci overlap with each other, these data implied cross contamination of *C. pulicaris* and *C. punctatus* DNA extracts (Table 2.3.1.).

2.4.2 Nested PCR sensitivity

The serial dilutions of *Cardinium* positive samples saw an increase in sensitivity by between 10- and 100-fold lower detection limit for nested compared to the

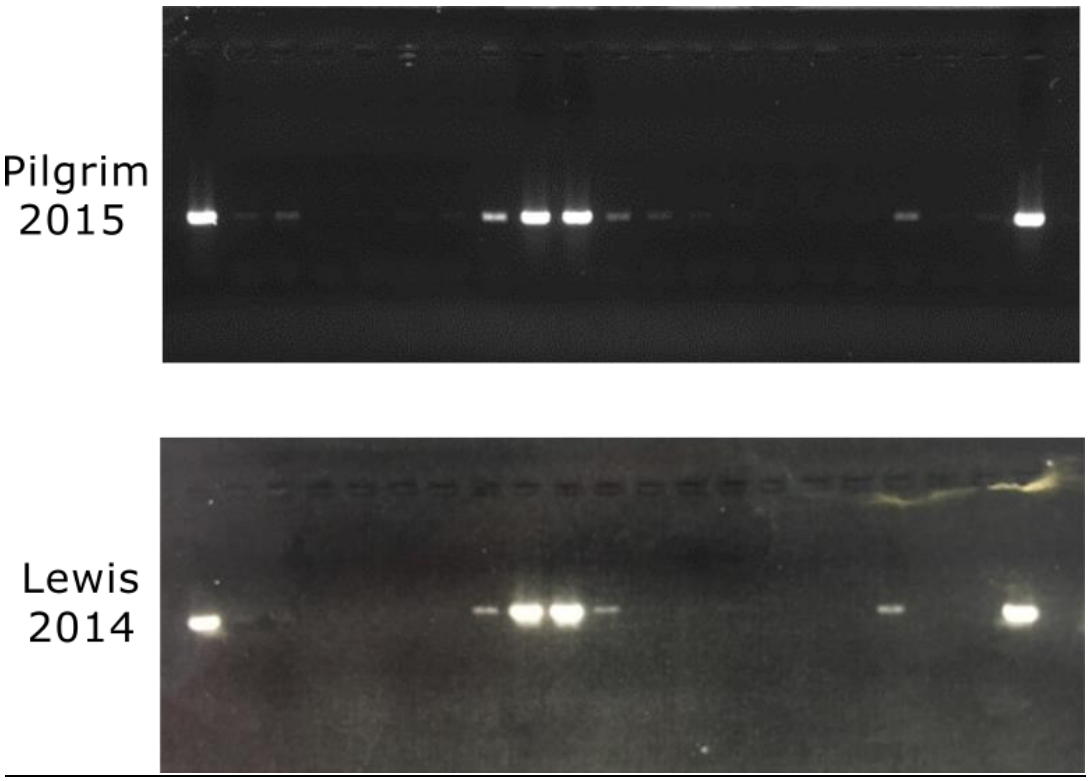


Figure 2.1. Quality control of DNA extracts labelled as *Culicoides pulicaris* from Lewis *et al*.'s study using HCO1490/LCO2198 primers (Folmer *et al.* 1994). Gel images are taken from a screen undertaken separately by the author in this study (2015) and by Lewis *et al.* (2014).

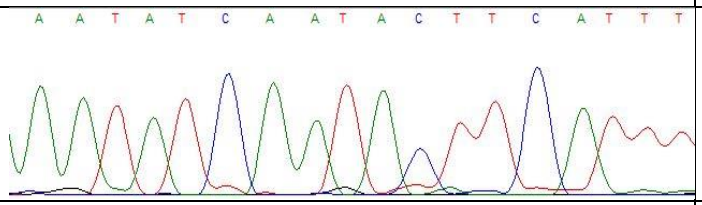
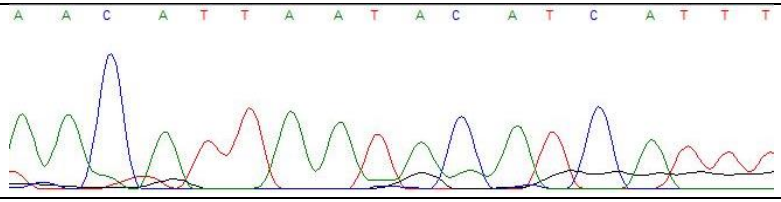
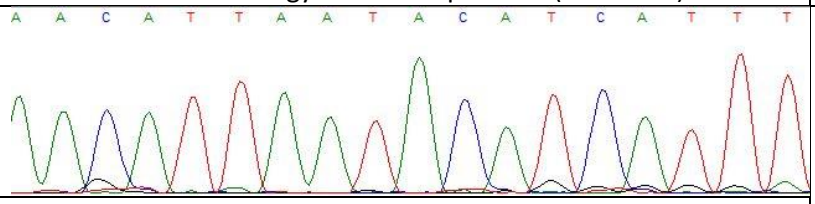
Identifier	LCO1490/HCO2198 (Folmer *et al.* 1994)	C1J-1718/C1N-2191 (Dallas *et al.* 2003)	*Cardinium* infection
Lewis			+
	>99% homology Culicoides punctatus (KX064676)	>99% homology Culicoides pulicaris (KJ624116)	
Pilgrim	N/A		-
		>99% homology Culicoides pulicaris (KJ624116)	

Table 2.3.1 Sequencing of *COI* amplicons from samples initially identified as *Culicoides pulicaris* from Lewis *et al.* (2014) and this study, suggesting cross-contamination with *Culicoides punctatus* in *Cardinium* positives from Lewis' study. N/A=Non-amplifiable.

Culicoides species	LCO1490/HCO2198 (Folmer *et al.* 1994)	C1J-1718/C1N-2191 (Dallas *et al.* 2003)
C. pulicaris P1	-	+
C. bysta	+	+
C. punctatus	+	-
C. obsoletus	+	+
C. impunctatus	+	+

Table 2.3.2 PCR amplification status of *Culicoides* species of interest with two sets of *COI* primers.

conventional PCR assay (Figure 2.2.). Out of the 7 DNA extracts used to test the lower detection limit, all gave positives for neat templates for both conventional and nested assays. However, all templates gave a signal at 10^{-2} dilutions for the nested PCR whereas the lowest detection limit for the conventional PCR were weak bands observed at 10^{-1} dilutions in 4/7 replicates. When screening the samples in this study using the nested assay a substantially higher prevalence of *Cardinium* infection was detected (Figure 2.3. and Table 2.4.) with the prevalence estimate of *Cardinium* infections more than doubling from 0.09 (30/338 individuals; True proportion 95% CI=0.06-0.13) to 0.19 (64/338; True proportion 95% CI=0.15-0.24).

2.4.3. An extended *Cardinium* survey of *Culicoides* species and the detection of low titre infections

Cryptic species of both *C. newsteadi* and *C. pulicaris* were identified after sequencing of the *COI* gene and subsequently named with respect to their nearest mtDNA barcode match as *C. newsteadi* N1,2 and 3, *C. pulicaris* P1 and *C. bysta* (Pagès *et al.*, 2009; Ander, Troell and Chirico, 2013; Sarvašová *et al.*, 2017). The conventional assay detected *Cardinium* in 4 putative vector species of bluetongue virus (BTV); *C. imicola*, *C. newsteadi* N2, *C. bysta* and *C. punctatus* from 30 individuals. In contrast, evidence of low-level *Cardinium* infections was detected in a further 5 species (*C. achrayi*, *C. newsteadi* N1 and N6, *C. pulicaris* P1 and *C. sphagnumensis*) of 64 *Culicoides* when screened with the nested assay (Table 2.4.). Infections were found in the subgenera Avaritia, Culicoides, Beltranmyia and Silvaticulicoides with prevalence within populations ranging from 4%-100%. The lowest and highest prevalence observed

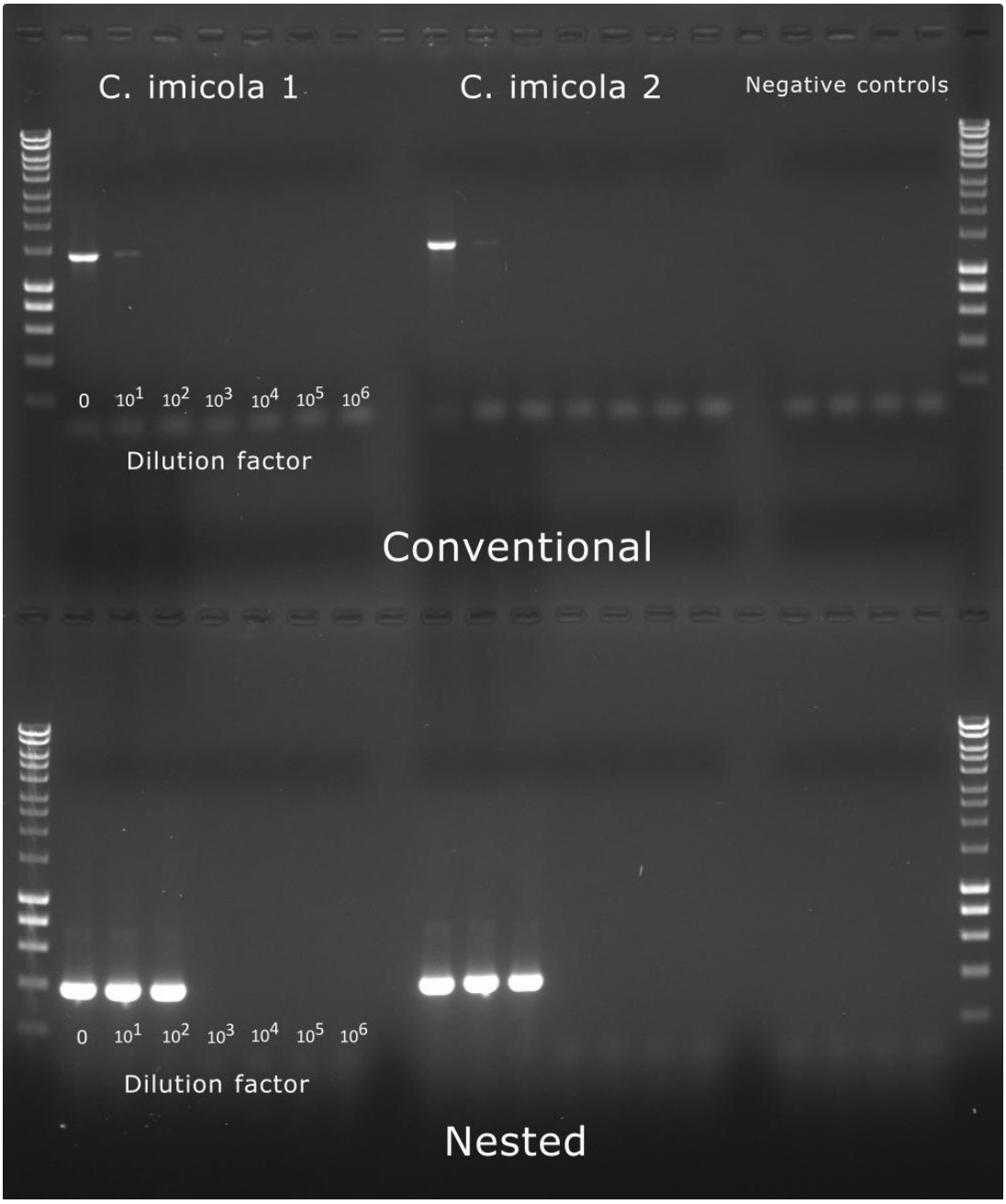


Figure 2.2. Detection limits of *GyrB* conventional and nested PCR. Dilution series of DNA extracts from *Culicoides imicola* were formulated by diluting neat extracts with nuclease-free water.

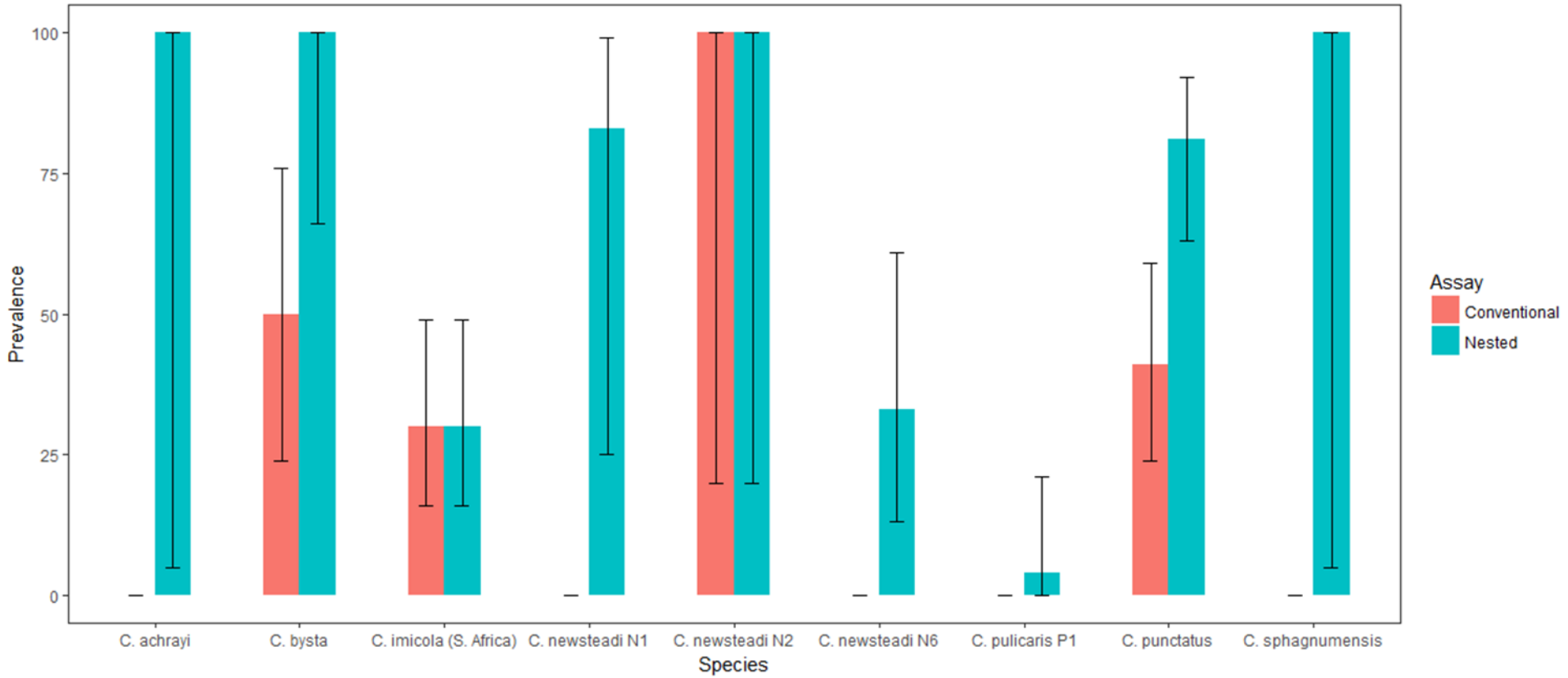


Figure 2.3. The comparison of *Cardinium* prevalence rates in *Culicoides* species detected by both *GyrB* conventional and nested PCR assays. Error bars are 95% confidence intervals of the true proportion of infection.

Subgenus	*Culicoides* Species	Locality (Site name)	Year of collection	Proportion of *Cardinium* positive conventional PCR (n) [95% Confidence interval]	Proportion of *Cardinium* positive nested PCR (n) [95% Confidence interval]
Avaritia	*C. bolitinos*	South Africa (Onderstepoort)	2016	0 (19) [0-0.21]	0 (19) [0-0.21]
	C. huambensis	South Africa (Koeburg)	2007	0 (1) [0-0.95]	0 (1) [0-0.95]
	C. imicola	Corsica (2APL7)	2015	0 (46) [0-0.1]	0 (46) [0-0.1]
		Corsica (2BPL2)		0 (27) [0-0.16]	0 (27) [0-0.16]
		South Africa (Onderstepoort)	**2016**	**0.3 (33) [0.16-0.49]**	**0.3 (33) [0.16-0.49]**
	C. tuttifrutti	South Africa (Chintsa)	2016	0 (10) [0-0.34]	0 (10) [0-0.34]
	C. obsoletus	UK (Neston)	2012-2015	0 (33) [0-0.13]	0 (33) [0-0.13]
Beltranmyia	*C. salinarius*	Sweden (Unknown site)	2009	0 (1) [0-0.95]	0 (1) [0-0.95]
	C. sphagnumensis	**Sweden (Axvalla)**	**2008**	**1 (1) [0.05-1]**	**1 (1) [0.05-1]**
Culicoides	*C. impunctatus*	Sweden (Torsås)	2008	0 (20) [0-0.2]	0 (20) [0-0.2]
		UK (Kielder)	2016	0 (13) [0-0.28]	0 (13) [0-0.28]
	C.magnus	South Africa (Koeburg)	2007	0 (1) [0-0.95]	0 (1) [0-0.95]
	***C. newsteadi* N1[a]**	**Corsica (2APL7)**	**2015**	**0 (4) [0-0.6]**	**0.75 (4) [0.22-0.99]**
		Corsica (2BPL2)	**2015**	**0 (2) [0-0.8]**	**1 (2) [0.2-1]**
	***C. newsteadi* N2[a]**	**Corsica (2BPL2)**	**2015**	**1 (2) [0.2-1]**	**1 (2) [0.2-1]**
	C. newsteadi N3[b]	Sweden (Unknown site)	2008-2010	0 (4) [0-0.6]	0 (4) [0-0.6]
	***C. newsteadi* N6[c]**	**Corsica (2APL7)**	**2015**	**0 (12) [0-0.3]**	**0.42 (12) [0.16-0.71]**

		Corsica (2BPL2)	2015	0 (3) [0-0.69]	0 (3) [0-0.69]
	C. pulicaris P1	UK (Canterbury)	2014	0 (3) [0-0.69]	0 (3) [0-0.69]
		UK (Hereford)	2014	0 (1) [0-0.95]	0 (1) [0-0.95]
		UK (Luton)	2014	0 (1) [0-0.95]	0 (1) [0-0.95]
		UK (Wolverhampton)	**2013**	**0 (15) [0-0.25]**	**0.07 (15) [0-0.34]**
		UK (Worcester)	2014	0 (7) [0-0.44]	0 (7) [0-0.44]
	C. bysta	**Corsica (2BPL2)**	**2015**	**0.5 (10) [0.24-0.76]**	**1 (10) [0.66-1]**
	C. punctatus	**Sweden (Torsås)**	**2008**	**0.44 (25) [0.25-0.65]**	**0.8 (25) [0.59-0.92]**
		UK (Luton)	**2014**	**0 (2) [0-0.8]**	**1 (2) [0.2-1]**
		UK (Wolverhampton)	**2014**	**0.4 (5) [0.3-0.99]**	**1 (5) [0.46-1]**
Meijerehelea	*C. leucostictus*	South Africa (Kuleni)	2014	0 (1) [0-0.95]	0 (1) [0-0.95]
Monoculicoides	*C. stigma*	Sweden (Unknown site)	2008	0 (3) [0-0.69]	0 (3) [0-0.69]
Oecacta	*C. clastrieri*	Sweden (Unknown site)	2009	0 (2) [0-0.8]	0 (2) [0-0.8]
	C. duddingstoni	Sweden (Bara)	2008	0 (4) [0-0.6]	0 (4) [0-0.6]
Silvaticulicoides	***C. achrayi***	**Sweden (Axvalla)**	**2009**	**0 (1) [0-0.95]**	**1 (1) [0.05-1]**
	C. subfascipennis	Sweden (Romakloster)	2008	0 (12) [0-0.3]	0 (12) [0-0.3]
Silvicola	*C. grisescens*	Sweden (Torsås)	2008	0 (6) [0-0.48]	0 (6) [0-0.48]
Synhelea	*C. similis*	South Africa (Alexandria)	2016	0 (6) [0-0.48]	0 (6) [0-0.48]
Wirthomyia	*C. reconditus*	Sweden (unknown site)	2008	0 (1) [0-0.95]	0 (1) [0-0.95]

Table 2.4. *GyrB* conventional and nested PCR assay results. *Culicoides newsteadi* haplotypes are designated by [a]Pagès *et al.*, 2009, [b]Ander *et al.*, 2013, [c]*Culicoides newsteadi* N6 previously undesignated. Bold entries are of species identified as being infected with *Cardinium.*

from populations with more than a few individuals were from *C. pulicaris P1* (Prevalence proportion=0.04; True proportion 95% CI=0.01-0.21) and *C. bysta* (Prevalence proportion=1; True proportion 95% CI=0.66-1) respectively.

2.4.4. *Cardinium* prevalence comparisons of site, species and sex

Where there were multiple populations of the same species at different geographical locations, most showed no evidence of heterogeneity in the fraction infected (*C. newsteadi*, *C. pulicaris*, *C. punctatus*) apart from *C. imicola* populations from South Africa and Corsica (Fisher's exact, $P<0.01$) (Table 2.5.1). Although infected *C. imicola* were observed at an intermediate prevalence in Onderstepoort, South Africa (Prevalence proportion=0.33; True proportion 95% CI=0.16-0.49), no infection was detected in the same species in Corsica (True proportion 95% CI=0-0.06). Sex ratio differences were not observed within infected species of *C. punctatus* (Fisher's exact, $p=0.76$) or *C. imicola* (Fisher's exact, $p=0.68$) (Table 2.5.2). No clear pattern of spatial distribution was observed between the three geographical regions of the study (Figure 2.4.1 and 2.4.2.).

2.4.5. Phylogenetic analysis

Phylogenetic analysis based on a 1300 bp region of the *GyrB* gene confirmed a monophyletic clade; grouping all *Culicoides Cardinium* isolates in group C of the genus where all previously described sequences of *Culicoides* clustered (Figure 2.5.). Sequences isolated from this study (*C. newsteadi* N2, *C. bysta*, *C. imicola* and *C. punctatus*) grouped neither based on subgenus nor geography. *Culicoides newsteadi* N2 and *C. bysta* from the subgenus Culicoides both formed separate clades with *C. ohmorii* (subgenus Hoffmania) and *C. imicola*/*C. oxystoma* (Subgenus Avaritia/Remmia) respectively. Sequence divergence between C. *newsteadi* N2

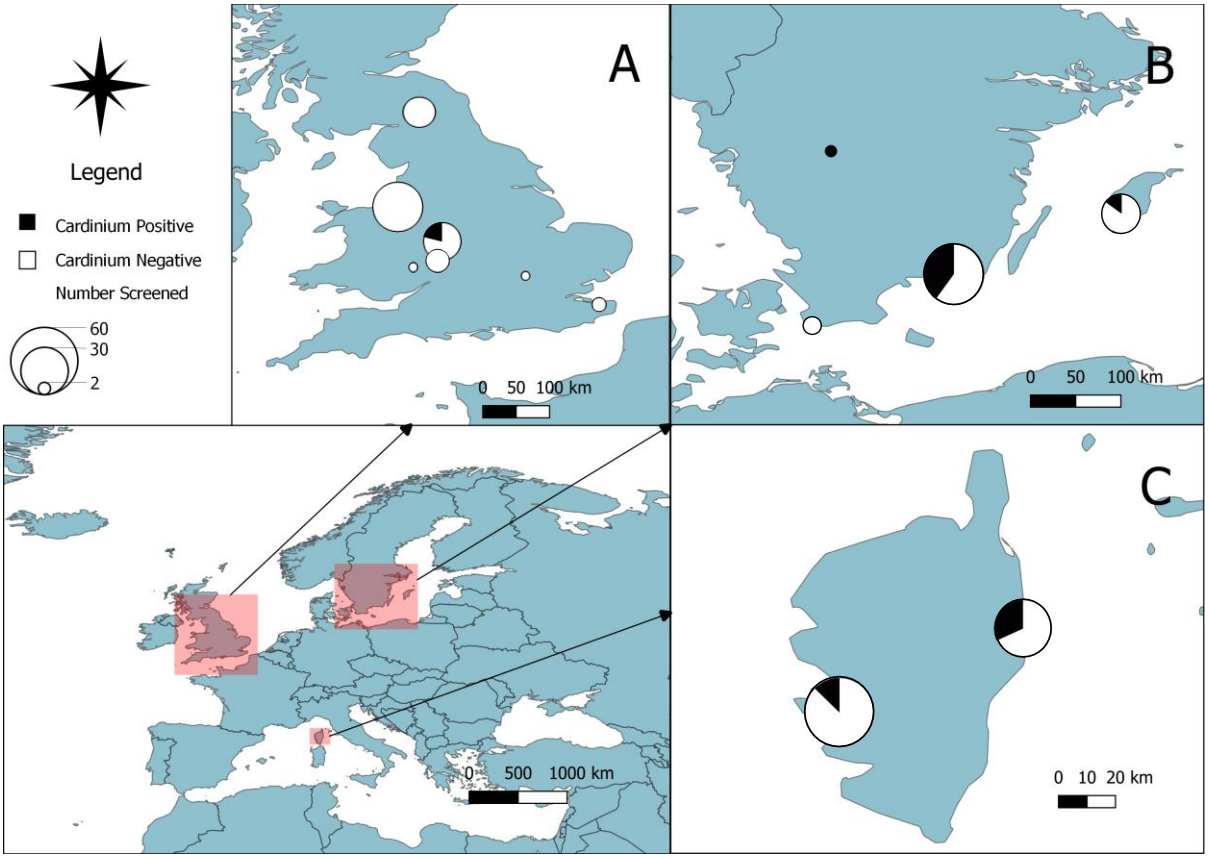


Figure 2.4.1. Palearctic *Culicoides* trapping and endosymbiont distribution. A QGIS map depicting collection sites across the UK (A), Sweden (B), and France, Corsica (C). The size of pie charts represents the number of individuals screened, with the proportion of *Cardinium*-infected designated in black. White circles indicate individuals where no *Cardinium* was detected by *GyrB* conventional or nested PCR.

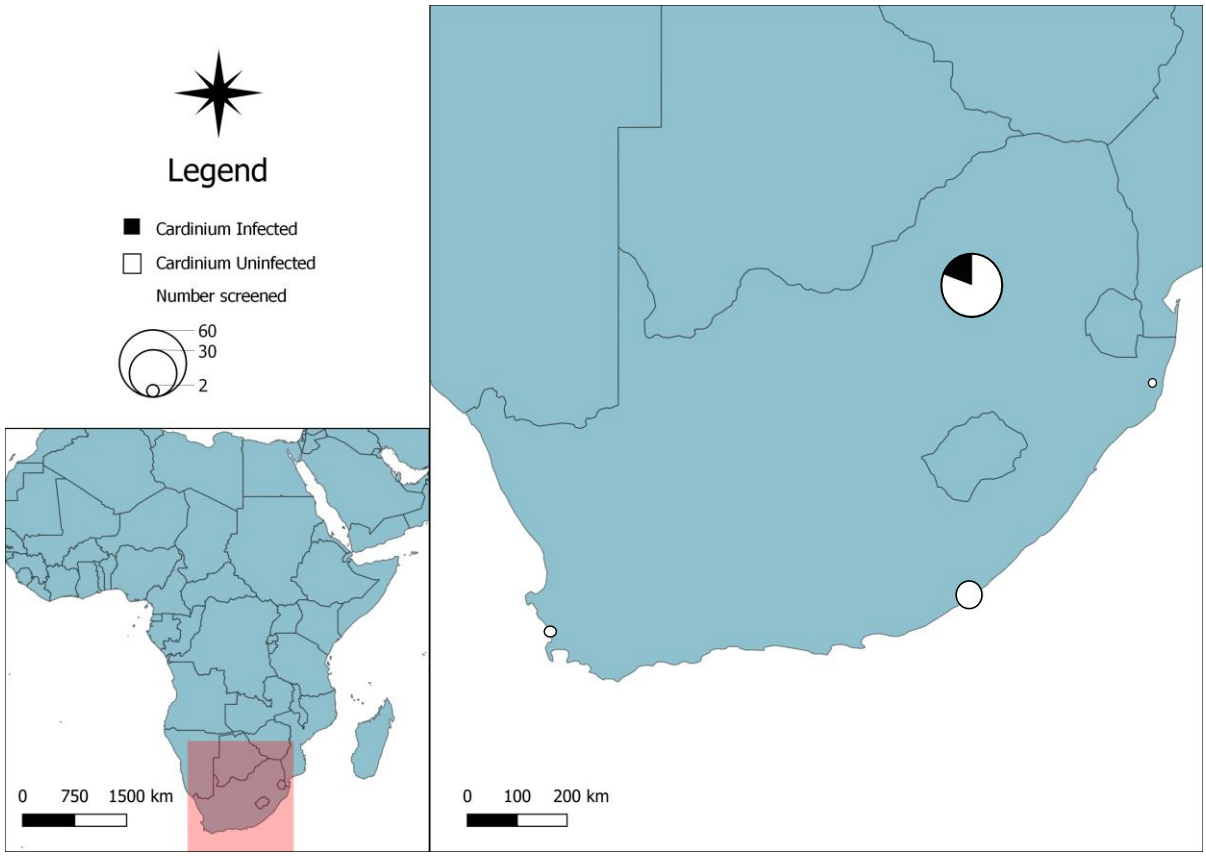


Figure 2.4.2. Southern Afrotropical *Culicoides* trapping and endosymbiont distribution. A QGIS map depicting collection sites across four sites in South Africa. The size of pie charts represents the number of individuals screened, with the proportion of *Cardinium*-infected designated in black. White circles indicate individuals where no *Cardinium* was detected by *GyrB* conventional or nested PCR.

Site 1	Site 2	*Culicoides* species	*p* value
South Africa (Onderstepoort)	Corsica (2APL7)	*C. imicola*	<0.01
South Africa (Onderstepoort)	Corsica (2BPL2)	*C. imicola*	<0.01
Corsica (2BPL2)	Corsica (2APL7)	*C. newsteadi* N1	1
Sweden (Torsås)	UK (Luton)	*C. punctatus*	1
Sweden (Torsås)	UK (Wolverhampton)	*C. punctatus*	1
UK (Luton)	UK (Wolverhampton)	*C. punctatus*	1
Corsica (2BPL2)	Corsica (2APL7)	*C. newsteadi* N6	0.5

Table 2.5.1. Fisher's exact results for *Cardinium* infection comparisons of populations of the same species but at different sites.

Culicoides species	Site	*p* value
C. imicola	Onderstepoort	0.68
C. punctatus	Torsas	0.76

Table 2.5.2. Fisher's exact results for *Cardinium* infection comparisons of females and males of the same species at the same site.

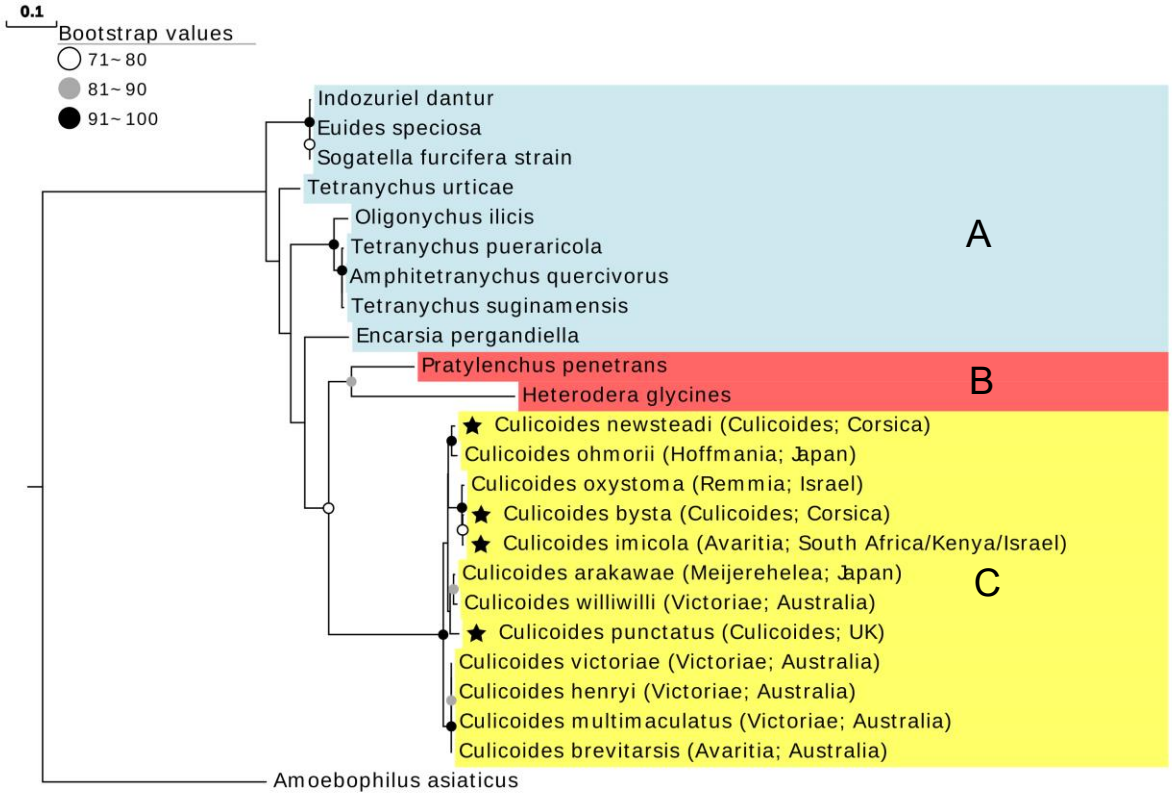


Figure 2.5. A maximum likelihood phylogram based on a 1300 bp region of the *GyrB* gene generated in RAxML using 1000 rapid bootstrap replicates and using the GTR + I + G substitution model. Groups A (blue) B (red) and C (yellow) are designated based on *Cardinium* taxonomic convention. Stars represent sequences generated in this study. Brackets indicate subgenus and country of origin. Accession numbers of taxa can be found in Table 2.2.

and *C. ohmorii* was 1.6% (21 nucleotide differences), whereas divergence between *C. bysta* and *C. imicola* was just 0.2% (3 nucleotide differences). This was in comparison to the average *GyrB* divergence of 3.1% across *Cardinium* group C. The largest distance between sequences came between *C. multimaculatus* (Australia) and both of *C. imicola* and *C. bysta* (4.8%; 62 nucleotide substitutions). The *C. imicola GyrB* sequence obtained from South Africa in this study was identical to both of those obtained from Kenya (Accession number: KR026927) and Israel (Accession number: JN166963). Likewise, the *Cardinium* sequences obtained from *C. punctatus* from Wolverhampton, UK and Torsås, Sweden had 100% identity to the sequence reported in the same species by Lewis *et al*. (Accession number: HG380244).

2.4.6 *C. imicola* haplotype network analysis

COI sequences were successfully obtained for *C. imicola* populations from two sites in Corsica (2APL7 and 2BPL2; *Cardinium* negative) and South Africa (Onderstepoort; *Cardinium* positive). To confirm these were not cryptic species, a distance estimation of *C. imicola* barcodes was assessed giving a minimum identity of >98%, consistent with within-species polymorphism (Lassen *et al.*, 2012; Nielsen and Kristensen, 2015). A *C. imicola* mitochondrial haplotype (mitotype) network (Figure 2.6.) of the single South African population, containing both *Cardinium*-infected (n=10) and uninfected (n=9) individuals, showed the presence of 7 mtDNA haplotypes split into two broad groups separated by 3 SNPs. Infection status and mitotype were shown to be associated (Fisher's exact test, $P \leq 0.001$). The *COI* divergence range between infected individuals was 0-0.39% (0-2 nucleotides) and 0-0.98% (0-5 nucleotides) in uninfected insects. Additionally, the mean variance between infected individuals was 0.077% compared to 0.37% in uninfected individuals, showing a significant difference (Wilcoxon Unpaired Two-Sample Test, $P \leq 0.001$). The 3 infected haplotypes consisted of a single major

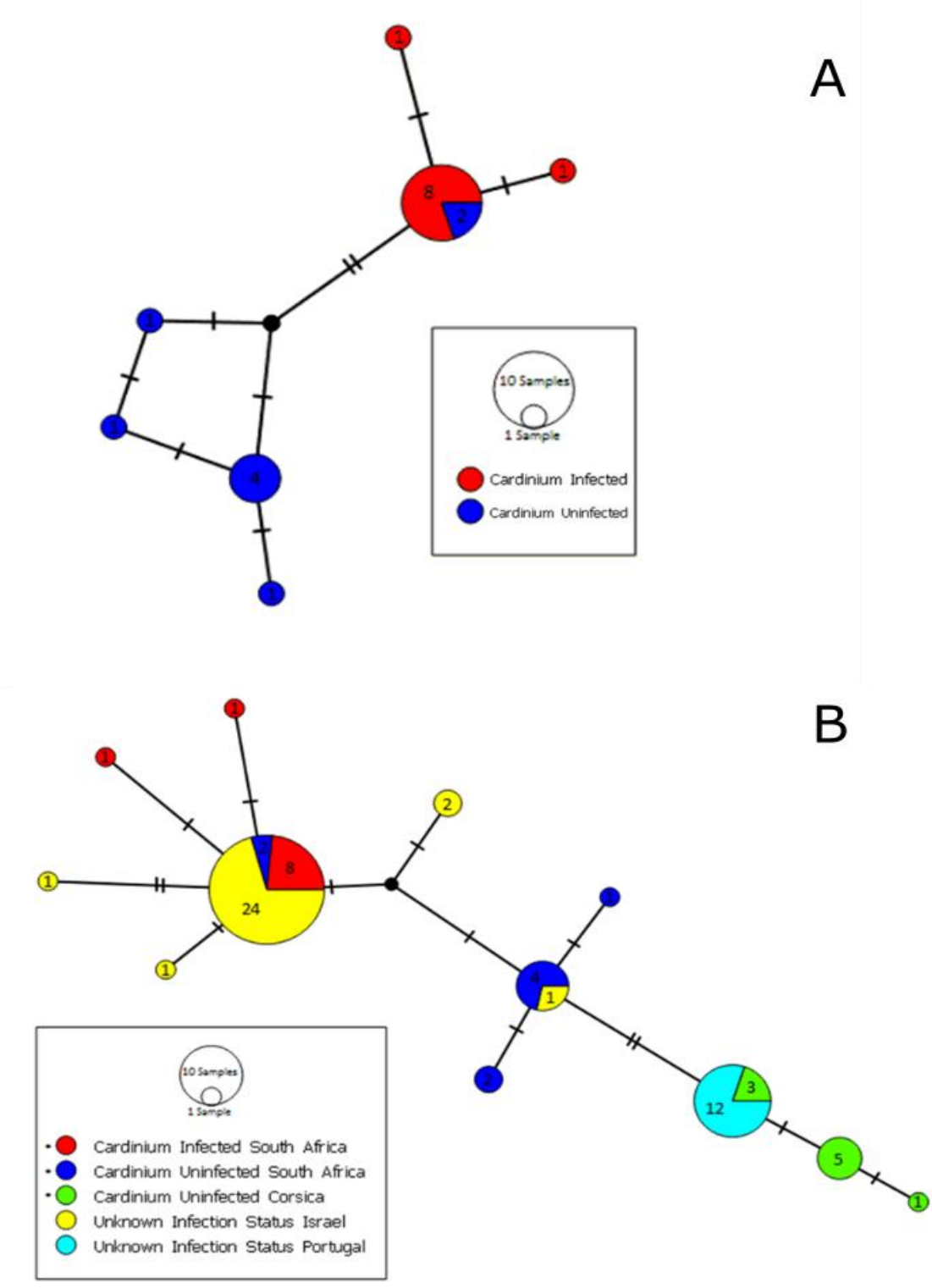


Figure 2.6. A) An mtDNA haplotype network of *Cardinium* infected and uninfected *Culicoides imicola* from a single site at ARC-OVI, Onderstepoort, South Africa; based on a 517 bp *COI* sequence. **B)** An mtDNA haplotype network of *Culicoides imicola* from sites spanning South Africa and the Mediterranean basin; based on a 451 bp *COI* sequence. Haplotype networks were generated using the TCS network algorithm in PopART v1.7 (Leigh & Bryant 2015). Numbers within circles represent the numbers of individuals designated to each haplotype. The numbers of substitutions separating haplotypes are indicated by dashes. *Indicates mtDNA haplotypes generated in this study.

group (8 individuals) and 2 haplotypes containing singletons. After extending the haplotype network to include *C. imicola* populations from the western Mediterranean basin (WMB) (Corsica and Portugal) and the eastern Mediterranean basin (EMB) (Israel), 12 haplotypes were observed with 0-1.77% range of divergence over a 451 bp region. A majority of the available *COI* sequences (24/29) from Israel clustered in the main haplotype associated with *Cardinium* infection in South African individuals (Figure 2.6.). Another association was identified between *COI* sequences from Portugal and Corsica which shared 3 haplotypes. The two predominant haplotypes from EMB and WMB were separated by two SNPs either side of the main *Cardinium* uninfected haplotype from South Africa.

2.5 Discussion

A main aim of this study was to explore the potential of *C. pulicaris* and *C. punctatus* as suitable model organisms to undertake future UK-based work pertaining to *Culicoides*-*Cardinium* interactions. Overall, the erroneous methodology and interpretation of a previous study, assessing these two species (Lewis *et al.*, 2014), has led to a false estimation of *Cardinium* prevalence in *C. pulicaris* which required correction. First, the majority of *C. pulicaris* extracts from Lewis were incorrectly interpreted as passing quality control when, in fact, positive PCR bands were not observed (Figure 2.1.). Usually, negative PCR results are interpreted as poor DNA quality, containing inhibitors, or containing a polymorphism at a primer binding site. The latter appears to be true in this case, with inclusion of a secondary set of primers (Dallas *et al.*, 2003) amplifying the *COI* gene in all *C. pulicaris* extracts.

Apart from the poor interpretation of quality control results by Lewis, the author's screening of UK *C. pulicaris* from archived samples (2014) has identified a 7% prevalence through nested PCR screening compared to the 26% prevalence inferred by Lewis using a conventional assay. This higher prevalence is likely a result of cross-contamination from *Cardinium* positive *C. punctatus* samples which were stored in the same cryobox. This is supported by the successful sequencing of both *C. pulicaris* and *C. punctatus COI* sequences from the same extracts labelled as *C. pulicaris* by Lewis (Table 2.3.1). The successful replication of raw screening data from both Lewis and the author (Figure 2.1.) suggests any cross-contamination occurred at the time of the original study in 2014. The renewed low *Cardinium* prevalence recorded for *C. pulicaris* in the UK has implications for potential future *in vivo* work. As the more abundant of the two species from UK collections, *C. pulicaris* (as opposed to *C. punctatus*) would be the preferred candidate for any future work examining the vector competence or other

phenotypic effects relating to *Cardinium*. The initial intermediate prevalence suggested by Lewis could have opened the possibility of case-control studies investigating the effects of *Cardinium* on wild *C. pulicaris* populations. However, the single low-titre individual (1/27) found to be infected suggests the numbers required to do such a study would be too large.

Despite the low titre and prevalence of infection observed here, a cryptic species of the pulicaris complex, *C. bysta*, was found to be infected at fixation in Corsica, France. If *Cardinium* does influence vectorial capacity, this result highlights the importance of using molecular markers alongside morphological data to decipher cryptic species which vary in their endosymbiont infection status. The high prevalence of *Cardinium* in the only other infected UK species, *C. punctatus*, would require antibiotics to cure infection in order to obtain a control group for experimental analysis. More importantly, due to the limited numbers of wild *C. punctatus*, finding material to initiate a colony is the major issue of using this species as a model organism.

The extended screening of midge populations in both Palearctic and Afrotropical regions produced varying prevalence ratios of *Cardinium*. Most populations did not show any signs of *Cardinium* infection (22/33 populations), 5 populations showed signs of low-intermediate prevalence, and 6 were at fixation for *Cardinium* although a few of these had low sample sizes. This varying prevalence is similar to the study by Mee *et al.* (2015), in which 3/26 positive species had a fixed infection, and 6 carried *Cardinium* in under 20% of individuals sampled, with the remainder being of variable intermediate prevalence. However, Mee's study (2015) also detected at least one *Cardinium* positive individual in each population screened which is at odds to the author's findings and another study by Pagès *et al.* (2017),

investigating *Cardinium* distribution in *Culicoides* from Spain. It is possible that the apparent "hotspot" in Australia is as a result of a more sensitive assay, as qPCR is often preferred to nested PCR in screening (Ikewaki *et al.*, 2003; Hyong *et al.*, 2008). Alternatively, a temperature effect could explain the discrepancies observed between these bioclimatic zones. Morag *et al*. (2012) found a positive correlation between land surface temperature and *Cardinium* prevalence in *Culicoides*. Similarly, a recent meta-analysis (Charlesworth *et al.*, 2018) found a a clear positive effect of temperature between arthropods and *Cardinium* infection rates when assessing arthropods in 137 countries. Further climatic effects have also been implicated in endosymbiont titre fluctuation (Hurst *et al.*, 2000; Corbin *et al.*, 2017) and subsequently detection rate. Therefore, it is possible that both the varying *Cardinium* titres and prevalence observed in this study, could be accounted for by temperature. An alternative hypothesis is that invasive ranges may play a role in prevalence heterogeneity between populations. For example, the loss of *Wolbachia* has been observed in invasive populations of the thrip *Pezothrips kellyanus*, possibly due to selection pressures that vary throughout host ranges (Nguyen, Spooner-Hart and Riegler, 2016).

A species which has come under scrutiny in the past due to its incursions into southern Europe, leading to outbreaks of BTV in Europe, is *C. imicola* (Mellor *et al.*, 1985; Boorman, 1986b; Jacquet *et al.*, 2015). The absence of *Cardinium* in *C. imicola* populations from Corsica, in this chapter, was recapitulated in a recent study from Spain (Pagès *et al.*, 2017), where lack of infection was also observed. No difference in prevalence was seen with the additional nested screening indicating a lack of low-titre infections. However, the presence of *Cardinium* in Israel (Morag *et al.*, 2012) suggests infection heterogeneity exists across the Mediterranean

basin. To investigate this further, mtDNA haplotypes and the infection status of *C. imicola* individuals within and between populations were compared. Endosymbionts can lead to selective sweeps that result in a reduction in diversity of mtDNA markers. This has been shown for *Tetranychus truncatus* mites, where haplotype diversity was greater in *Wolbachia* uninfected individuals compared to those infected (Zhang *et al.*, 2013). The haplotype networks produced from this study (Figure 2.6.), suggests *Cardinium* infection in *Culicoides* is another example of a selective sweep of the symbiont which facilitates the spread of specific mtDNA haplotypes. The low diversity observed between *C. imicola* infected mtDNA haplotypes when compared to those uninfected suggests a historic selective sweep (homogenising mtDNA haplotypes) followed by the gradual accumulation of mutations and subsequently new haplotypes. This interpretation is supported by the two minor infected haplotypes differing by one nucleotide difference. Alternative explanations for these data would be a bottleneck or founder event, although low mtDNA diversity due to endosymbiont effects in *C. imicola* is more parsimonious (Hurst and Jiggins, 2005). Although there is a structure observed between infection status and mtDNA haplotype, the presence of 2 uninfected individuals in the main infected haplotype in *C. imicola* from South Africa suggests imperfect vertical endosymbiont transmission.

Infection heterogeneity between *C. imicola* populations from the EMB and WMB could be explained by the dispersal of both infected and uninfected founding populations to both areas. Indeed, a recent study (Jacquet *et al.*, 2016) suggests the dispersal of *C. imicola* to the WMB was as a result of wind-dispersal of populations from Corsica, suggesting a lack of *Cardinium* infection was as a result of absence in dispersed populations. Further to this, Israeli

(EMB) haplotypes clustered within the main *Cardinium* infected mitotype observed in South Africa, whereas the Portuguese (WMB) haplotypes all clustered within an uninfected Corsican mitotype. Although, the infection statuses of these Portuguese and Israeli individuals are unknown, the recent Pagès (2017) screening for *Cardinium* found no indication of *C. imicola* infection in the WMB (Spain) but others (Morag *et al.*, 2012) have observed the symbiont to be present in the EMB (Israel). The matrilineal subdivision in *Cardinium* infected South African *C. imicola* and uninfected Corsican *C. imicola* (Figure 2.6.) is similar to observations by several studies (Dallas *et al.*, 2003; Nolan *et al.*, 2008; Calvo *et al.*, 2009) which all noted a remarkable divergence between *C. imicola* mitotypes from the EMB and WMB. This suggests that the linkage disequilibrium of *Cardinium* and mitochondria offers insights into symbiont gene flow within the Mediterranean basin but may not assist in elucidating host gene flow. For example, if dispersal events occur involving *Cardinium*-infected *C. imicola*, selective sweeps of the symbiont may erase any previously detectable biodiversity. This is reminiscent of similar patterns observed in other insects with endosymbiont infections: the mosquito vector, *Culex pipiens* (Rasgon, Cornel and Scott, 2006), the parasitoid wasp, *Nasonia vitripennis* (Raychoudhury *et al.*, 2010) and the ladybird *Adalia bipunctata* (Schulenburg *et al.*, 2002). In all these cases, the frequency of mtDNA haplotypes are more closely associated to the endosymbionts in a population rather than geography. Cryptic species can also offer explanations for endosymbiont infection heterogeneity. However, there was no evidence of cryptic species between *C. imicola* populations based on *COI* sequence divergence ranges of <2% (Lassen *et al.*, 2012; Nielsen and Kristensen, 2015). Overall, this suggests that a non-biased marker is needed for biogeographic studies aiming to evaluate the population structure of *C. imicola*.

Analysis of the relatedness of strains using the *GyrB* locus in this study (Figure 2.5.) corroborate a recent study by Siozios *et al.* (2019) placing the *Culicoides Cardinium* as a monophyletic sister clade to *Cardinium* group B from parasitic plant nematodes. Mee *et al*. (2015) has previously suggested that *Cardinium* strains cluster by geography, however, we find a sporadic distribution of strains pertaining to location; the two strains identified from Corsica, group with strains from Japan, Israel and Southern Africa instead of each other. Although symbionts have been known to cluster based on location due to geographic barriers (Russell *et al.*, 2009), the ability of *Culicoides* to disperse over long ranges (Burgin *et al.*, 2013; Jacquet *et al.*, 2016) can explain why we do not see a similar scenario in *Cardinium*-infected midges.

It is curious to note that although *Cardinium* was detected in five species in this study, all of which are vectors of BTV (*C. newsteadi, C. pulicaris, C. punctatus, C. bysta and C. imicola*), *C. obsoletus* (the major Palearctic vector of BTV) in the UK appears to be infection free**.** Pagès *et al*. (2017), however, have recently shown very low infection rates in *C. obsoletus* from Spain (2/304 individuals). Despite this, such low prevalence indicates *Cardinium* is most likely not relevant to vector biology at a population level. There is a particular interest in varying endosymbiont titres of vector species as they could influence potential biocontrol effects including viral blockage and host longevity (Hoffmann, Ross and Rašić, 2015). Aside from environmental effects, endosymbiont density can be determined by a range of biotic factors including bacterial strain and sex (Duron, Fort and Weill, 2007; Tortosa *et al.*, 2010).

Considering the discovery of SBV transovarial transmission in *C. punctatus* (Larska *et al.*, 2013), and *Cardinium* presence in ovaries (Nakamura *et al.*, 2009; Morag *et al.*, 2012), direct

interactions between *Cardinium* and SBV could alter rates of viral vertical transmission via the production of antiviral metabolites (Joyce *et al.*, 2011) or resource competition (Caragata *et al.*, 2013; Frentiu, 2017). However, in contrast with other symbiont-insect associations, the biological significance of this *Cardinium* clade in *Culicoides* remains unknown and requires further research. Despite a likely vertical transmission, the lack of sex-ratio distortion observed in this study in the two species which contained both males and females (*C. imicola* and *C. punctatus;* Table 2.5.2), suggests the induction of parthenogenesis, feminisation or male-killing is unlikely to be associated with *Cardinium* in this instance.

Other *Cardinium* strains have been implicated in cytoplasmic incompatibility (CI) in the wasp *Encarsia pergandiella* (Hunter, Perlman and Kelly, 2003), the planthopper *Sogatella furcifera* (Zhang, Zhao and Hong, 2012) and spider mites (Gotoh, Noda and Ito, 2007). The recent publication of the *C. punctatus Cardinium* genome by Siozios *et al*. (2019) has suggested possible unique genes related to CI. These are a family of endonucleases (DUF1703) whose paralogs are associated with the induction of CI in *Wolbachia* (Beckmann, Ronau and Hochstrasser, 2017). The presence of low *Wolbachia* densities have been suggested to lead to a weak sperm modification ability and a reduced CI effect (Kondo *et al.*, 2002). If indeed midge *Cardinium* does induce CI, low titres described in this study and elsewhere could suggest a weak CI effect in some host species. The low-density infections observed here along with the geographical variability in *Cardinium* infection status, has important implications for any future *Cardinium*-based vector control initiatives through effects on population dynamics. This is especially relevant if the strategy is reliant on the driving-mechanism of CI. Certain symbionts of haematophagus insects have also been shown to enable the synthesis of B vitamins due to their scarce availability in a blood diet (Rio *et al*. 2016). However, there is no

indication of this from genomic data, with no biotin or other B vitamin biosynthetic pathways present; indeed, a biotin transport system indicates *Cardinium* is a B vitamin sink rather than a provider (Siozios *et al.*, 2019).

To conclude, the *Cardinium* infection status of *Culicoides pulicaris* has been reassessed to suggest this species is not suitable for further *in vivo* work relating to the symbiont in the UK. The difficulties of culturing and maintaining *Culicoides* colonies (Mullen *et al.*, 1985; Jennings and Mellor, 1988; Nayduch *et al.*, 2014) has so far stunted investigations into the role of *Cardinium*. However, *Cardinium* has been shown to disrupt immune gene expression in cell culture models (Nakamura *et al.*, 2011). Subsequently, cell-lines could potentially be used as an alternative to whole-midge assays in order to assess future host phenotypic effects of *Cardinium*. Finally, the linkage disequilibrium between *Cardinium* and mitochondria suggests a potential to confound dispersal studies utilising the *COI* gene as a phylogeographic marker. Subsequently, due to the importance of *C. imicola* as a vector in Africa and Europe, the utilisation of mtDNA markers to assess *C. imicola* population structure should be met with caution.

Chapter 3: Limoniae Group *Rickettsia* are Widespread in *Culicoides* and Reach High Frequency

3.1 Abstract

There is increasing interest in the heritable bacteria of invertebrate vectors of disease as they present a novel target for control initiatives. Previous studies on biting midges (*Culicoides* spp.), known to transmit several RNA viruses of veterinary importance, have revealed infections with the endosymbiotic bacteria, *Wolbachia* and *Cardinium*. However, rickettsial symbionts in these vectors are underexplored. Screening of 493 *Culicoides* individuals from 30 Palearctic, Nearctic or Afrotropical species revealed *Rickettsia* represent a widespread but previously overlooked association, reaching high frequencies in midge populations and present in 37% of the species tested. Phylogenetic analysis of a multigene dataset clusters the *Rickettsia* within the Limoniae clade of the genus, a group known to infect several aquatic and hematophagous taxa. Given the importance of biting midges as vectors, a key area of future research is to establish the impact of this endosymbiont on vector competence.

3.2 Introduction

Heritable bacteria represent an important component of the biology of many arthropods. Carried by over half of all species (Weinert *et al.*, 2015), many vertically transmitted microbes contribute to host function. This contribution is most commonly through specific services, such as nutrient provisioning or protection (Douglas, 2009; Oliver *et al.*, 2009; Jaenike *et al.*, 2010). Conversely, their maternal-inheritance has led symbionts to favour production of daughters by their host, leading to the evolution of systems biasing the offspring sex ratio towards females (reproductive parasitism) (Hurst and Frost 2015). The strength of symbiont impact on individual biology, combined with the high frequency with which arthropod species are infected with symbionts, has led to intense study.

Particular attention has been focussed on symbiont/host interactions in vector species. Through the induction of cytoplasmic incompatibility, the endosymbiont *Wolbachia* prevents the formation of viable progeny between infected males and uninfected females in various dipterans including *Drosophila* spp. and *Aedes* spp. (Werren, Baldo and Clark, 2008). With respect to the latter, not only can this incompatibility enable vector population suppression but, through unknown mechanisms, a strong RNA virus resistance phenotype (Moreira *et al.*, 2009; Bian *et al.*, 2010; Blagrove *et al.*, 2012; van den Hurk *et al.*, 2012). Furthermore, experimental evidences show that both *Wolbachia* and another proteobacteria, *Wigglesworthia,* can act as obligate (required) symbionts, provisioning blood sucking vector hosts with B vitamins that are lacking in a blood-diet (Rio *et al*. 2016). This provisioning has evolved independently in bed bugs (*Cimex lectularius*; Nikoh *et al*. 2014), tsetse flies (*Glossina* sp.; Akman *et al*. 2002; Snyder *et al*. 2010; Rio *et al*. 2012) and lice (*Pediculus schaeffi*; Boyd

et al. 2014). Additional genomic surveys suggest that other proteobacterial symbionts including *Coxiella* are involved in metabolic homeostasis (Zhong, Jasinskas and Barbour, 2007; Manzano-Marín *et al.*, 2015; Smith *et al.*, 2015). As such, these symbioses can have profound effects on the biology, ecology and evolutionary dynamics of vector-pathogen interactions.

Rickettsia (Class: Alphaproteobacteria; Order: Rickettsiales) symbionts are obligate intracellular bacteria most notable for containing species pathogenic to vertebrates, such as *Rickettsia prowazekii*, the causative agent of louse-borne Typhus fever, *Rickettsia rickettsii* (Rocky Mountain spotted fever) and *Rickettsia conorii* (Boutonneuse or Mediterranean spotted fever). Despite this, vertebrate disease-causing *Rickettsia* are atypical of the genus as a whole (Perlman *et al*. 2006; Weinert *et al*. 2009), and many *Rickettsia* are maintained without infectious transfer. Members are known to induce a variety of reproductive manipulations, including male-killing in ladybird beetles (*Adalia bipunctata*) (Werren *et al.*, 1994; Hurst *et al.*, 1999; Majerus *et al.*, 1999) and parthenogenesis induction in parasitoid wasps (*Pnigalio soemius; Neochrysocharis formosa*) (Hagimori *et al.*, 2006; Giorgini *et al.*, 2010). *Rickettsia* symbiont infection can also be protective, enhancing resistance of aphids (*Acyrthosiphon pisum*) to fungal attack, and whiteflies (*Bemisia tabaci*) to bacterial challenge (Łukasik *et al.*, 2013; Hendry, Hunter and Baltrus, 2014). Of significance to the study of vectors, *Rickettsia* are also known to increase the competence of *Bemisia* whiteflies for transmission of tomato leaf curl virus (Kliot *et al.*, 2014). Members of the genus can also be insect-vectored plant pathogens in their own right, for example, underlying papaya bunchy top disease (Luis-Pantoja *et al.*, 2015). As such, symbiosis with *Rickettsia* is biologically

important at the individual and population level, and both as vectored disease agents in themselves and as a symbiont facilitating the spread of other diseases.

Previous studies of *Culicoides* symbionts have screened extensively for *Cardinium* and *Wolbachia* infections (Nakamura *et al.*, 2009; Morag *et al.*, 2012; Lewis *et al.*, 2014; Mee *et al.*, 2015; Pagès *et al.*, 2017), with most failing to report presence of *Rickettsia*. However, a 16S metagenomic screening project in *C. sonorensis* gut samples revealed amplicons allied to *Rickettsia* (Campbell *et al.*, 2004), albeit with no phylogenetic or population-based information. Further, when a colleague (Dr Stefanos Siozios, University of Liverpool) performed a shallow whole-genome sequencing of *C. newsteadi* N5, he recovered a near complete genome of an uncharacterized *Rickettsia* species related to the Limoniae clade of *Rickettsia,* a group known to infect several aquatic and hematophagous taxa.

Due to the genetic diversity and various phenotypes associated with endosymbionts, systems of strain typing have been developed. For *Wolbachia*, the surface protein gene *wsp* was initially used as a genetic marker (Zhou, Rousset and O'Neil, 1998). However, evidence of diversifying selection (Jiggins, Hurst and Yang, 2002) and recombination (Baldo, Lo and Werren, 2005) made this gene unsuitable for isolate characterization if used alone. Further to this, horizontal gene transfer events between genes (intergenic recombination) detected in *Wolbachia* suggested a single-locus approach to strain typing was inappropriate (Baldo *et al*. 2006). Subsequently, multigenic approaches have been seen as a superior tool to assess genetic relatedness between endosymbionts. A commonly used method, multi locus sequence typing (MLST), defines the haplotypes of individual strains based on house-keeping

genes and was initially used as a surveillance tool for pathogenic bacteria (Maiden *et al.*, 1998). However, more recently, this approach has included non-pathogenic bacteria (Baldo *et al*. 2006; Cai *et al*. 2007) for population and evolutionary analyses, as well as the association of particular strains with phenotypes. The conserved nature of housekeeping genes provides enough discrimination over an evolutionary timescale to be useful in typing isolates, as well as being less likely to undergo diversifying selection and intragenic recombination, which can disguise relationships among strains. For each locus, sequences are designated unique alleles and by combining these from the same isolate, an allelic profile is created. No weighting is given to single nucleotide changes or recombination exchanges meaning strains can be compared even when recombination is frequent in a genus. Thus, this system can allow for the accurate tracing of endosymbionts between geographic locations and insect communities. However, the comparison of allelic profiles can only give limited information related to evolutionary history and fails to distinguish recently diverged strains meaning concatenated gene phylogenies are seen as a suitable alternative approach (Bleidorn and Gerth 2017).

Subsequently, this chapter aimed to initially identify *Rickettsia*'s occurrence and distribution through a targeted *Culicoides* screen. Isolates were then assessed via a strain typing system similar to MLST, before a concatenated gene phylogeny was constructed to give better phylogenetic resolution after the investigation of any potential confounding recombination events.

3.3 Methods

3.3.1 *Culicoides* collection identification and DNA extractions

A majority of collection, identification and DNA extractions of *Culicoides* details can be found in chapter 2. Slight differences include the inclusion of *C. sonorensis* populations of midges for screening from California, USA (collected February/March 2018); a major vector of BTV in North America. Collection of *C. sonorensis* was undertaken using carbon-dioxide baited traps and confirmation of identification was performed with the help of Dr Alec Gerry and Xinmi Zhang (UC Riverside, USA) using wing morphology. DNA extractions of 29 individual *C. sonorensis* and 10 pools of 5 individuals were prepared using Qiagen DNEasy kits (Qiagen. Inc., Valencia, CA, USA) following the manufacturer's instructions for insect tissue (Qiagen, 2006). This brings the total number of specimens in this study to 493 specimens of 30 *Culicoides* species spanning Palearctic, Nearctic and Afrotropical ecozones.

3.3.2. PCR screening

Presence of *Rickettsia* was initially assessed by PCR assay using *Rickettsia*-specific primers designed to amplify a 320-bp region of the 17kDa surface antigen precursor. Cycling conditions were as follows: initial denaturation at 95°C for 5 min, followed by 35 cycles of denaturation (94°C, 30 sec), annealing (54°C, 30 sec), extension (72°C, 120 sec), and a final extension at 72°C for 7 min. Amplicons identified by gel electrophoresis were subsequently purified enzymatically (ExoSAP) and Sanger sequenced as described in chapter 2.

3.3.3 Multilocus typing gene choice

Based on previous studies (Fournier *et al*. 2003; Weinert *et al*. 2009; Li *et al*. 2010; Machtelinckx *et al*. 2012; Santibáñez *et al*. 2013) that profile *Rickettsia* diversity, the 16S *rRNA*, g*ltA* (citrate synthase), c*oxA* (cytochrome oxidase) and a*tpA* (ATP synthase) genes were chosen as indicators of genetic relatedness between isolates. With the consideration that these housekeeping loci may be too conserved to resolve recently diverged strains, the *17KDa* gene, was included to allow for higher resolution in typing, alongside an inference of selection pressure due to the divergent nature of antigen genes compared to housekeeping genes. Briefly, *16S* has a structural role in the ribosomal protein scaffold; *gltA* encodes a subunit of the Citrate Synthase enzyme utilised in the Kreb's cycle; *coxA* encodes a catalytic subunit of the respiratory electron transport chain; *atpA* is a subunit of the enzyme ATP synthase assisting in the conversion of Adenosine diphosphate (ADP) to Adenosine triphosphate (ATP); and finally, *17KDa* encodes an antigen of the outer membrane protein, thought to assist in the adhesion of *Rickettsia* to host cells. *Culicoides sonorensis* individuals and pools were screened using all 5 primers in contrast to other populations where the *17KDa* primers were used alone. Insects from positive pooled *C. sonorensis* were rescreened to identify infected individuals.

Primers to amplify these loci (Table 3.1.) were designed in Geneious (Kearse *et al.*, 2012) based on the *C. newsteadi N5* (*RiCNE*) genome (BioProject accession number PRJNA376033; Dr Stefanos Siozios) and available conserved gene sequences so that they could amplify across several *Rickettsia* groups but would not cross-amplify any alpha-proteobacteria outgroups. All PCR amplifications were performed under the same *17KDa* assay conditions described above. Sequencing through both strands allowed for the clarification of ambiguous base calls

as well as giving greater sequence coverage at individual loci. Raw sequences were edited in UGENE (Okonechnikov *et al.*, 2012) and alignments for each locus were generated in MEGA6 using the ClustalW algorithm (Tamura *et al.*, 2013).

A profile of each locus was constructed by calculating GC content, nucleotide diversity per site (π) and the percentage of variable sites using DNAsp v5 (Librado and Rozas 2009). As phylogenetic inferences can be complicated by recombination, the presence of intragenic and intergenic recombination was investigated. The program RDP v4 (Martin *et al.*, 2015) was used to test for intragenic recombination using the MaxChi, RDP, Chimaera, Bootscan and GENECONV algorithms with the following criteria to assess a true recombination positive: a *p*-value of <0.001, sequences were considered linear with 1000 permutations being performed. As these algorithms are prone to false positives, a recombination event was considered likely if more than one algorithm gave a significant *p*-value, as well as being confirmed by visual inspection. To this end, maximum likelihood trees either side of putative recombination breaking points were generated for inspection of congruency. For any intragenic recombination events detected, the locus for that strain was removed from concatenation. Intergenic recombination was also assessed by comparing single-gene trees. Additionally, all protein coding loci (*17KDa*, *atpA*, *coxA* and *gltA*) were investigated for evidence of diversifying selection at the gene level via pairwise non-synonymous/synonymous rate ratio analysis (Ka/Ks ratio).

Gene	Product	Primer name	Sequence (5'-3')	Gene length (bp)	MLST fragment size (bp)
atpA	ATP synthase subunit alpha	RiAtpA327_F	GTCGGTAAAGCATTGCTTGGT	1560	977
		RiAtpA1309_R	ATTGATCCTGCTTCAATA		
coxA	Cytochrome c oxidase, subunit I	RiCoxA317_F	ATAGGTGCACCGGATATGGC	1569	1021
		RiCoxA1409_R	CCGATAGATGATACCATATTCCA		
gltA	Citrate synthase	RiGltA405_F	GATCATCCTATGGCA	1287	786
		RiGltA1193_R	TCTTTCCATTGCCCC		
17KDa	17KDa antigenic protein precursor	Ri17kD_F	TCTGGCATGAATAAACAAGG	471	319
		Ri17kD_R	ACTCACGACAATATTGCCC		
16S	16S ribosomal RNA	Ri170_F	GGGCTTGCTCTAAATTAGTTAGT	1511	1170
		Ri1500_R	ACGTTAGCTCACCACCTTCAGG		

Table 3.1 Housekeeping and *17KDa* gene primer attributes.

3.3.4 Multigenic strain typing

Following the convention of multilocus sequence typing (MLST), all unique genotypes were designated allele numbers (used as a unique identifier) which, when combined at all loci, produce an allelic profile (Maiden *et al.*, 1998). Aside from identifying specific isolates, this multigenic approach also allows for clonal complexes to be identified (conventionally allelic profiles which are identical at three or more loci). Allelic profiles and complexes were designated based on unweighted pair group method with arithmetic mean (UPGMA) cluster analysis and visualised as a minimum spanning tree (MST) implemented by Bionumerics v7 (Applied Maths 2012). Quantum GIS software v3 (Quantum GIS Development Team, 2016) was then used to map strains to investigate geographic patterns of relatedness.

3.3.5 Phylogenetic analysis

As UPGMA analysis does not take into consideration substitution models of character changes, the addition of other *Rickettsia*-positive taxa, provided by Panupong Thongprem (University of Liverpool) (Appendix 1), were included in a further phylogenetic analysis. Sanger sequence visualisation and analysis was performed by the author for all loci except *gltA* which was used in the initial screening for Thongprem's screen. After checking for recombination events likely to affect phylogenetic inferences, the relatedness of midge *Rickettsia* within the *Rickettsiaceae* was assessed using the concatenation of the 5 individual genes. First, sequences of *16S rRNA, 17KDa, atpA, coxA and gltA* from selected *Rickettsia* groups were extracted and concatenated using SequenceMatrix v1.8 (Vaidya, Lohman and Meier, 2011). Sequences were then combined with others selected from across the genus (Weinert *et al*. 2009) obtained from GenBank (NCBI). Strains included both the ancestral "Hydra" (Megaira) group as well as derived pathogenic isolates (Spotted fever and

Typhus). For some of the taxa included, only *16S* or *gltA* sequences were available, and so missing data was allowed in the alignment where a gene was absent. According to Wiens (2006), missing character data should not affect phylogenetic resolution, and is only likely to be an issue for incomplete data if the number of characters is very low. The sequences for each locus were aligned using the "L-INS-I" algorithm in MAFFT v7.4 (Katoh and Standley 2013) before a partition scheme was devised in order to allow for an optimised substitution model for each gene using ModelFinder (Kalyaanamoorthy *et al.*, 2017). Model selection included free rate as well as traditional gamma distributed rate categories which may "underfit" multi-gene data. Edge-proportional models were included where varying branch lengths were allowed in different partitions before proportionally rescaling when devising the multi-gene tree. The selected models were as follows: *16S* and GTR+F+R4; *17KDa* and TVM+F+I+G4; *atpA* and GTR+F+I+G4; *coxA* and TVM+F+I+G4; *gltA* and GTR+F+R3. A maximum likelihood (ML) phylogeny was estimated with IQTree (Nguyen *et al.*, 2015) using the concatenated alignment of 4403 nucleotides and 5000 ultrafast bootstraps. Various constrained tree searches alongside the Shimodaira-Hasegawa (SH) test (Significance threshold $p<0.001$) were implemented in IQTree to test alternative hypotheses of monophyly with respect to host taxa. The hypothesis test statistic, for a comparison of alternative topologies, was achieved by obtaining a maximised likelihood score (-ln) for each comparison. Finally, trees were drawn and annotated using the EvolView (He *et al.*, 2016) online tree annotation and visualization tools.

3.3.6 Nucleotide sequence accession numbers

Culicoides' *COI* barcodes and sequences generated for individual *Rickettsia* loci in this study were deposited in GenBank under deposition numbers KY765346-KY765408 and KY778697-KY778698.

3.4 Results

3.4.1 Prevalence of *Rickettsia* in biting midges

Screening of field collected midge specimens using the *17KDa* gene revealed *Rickettsia* infections in 155 of 493 (31%) individuals and 11 of 30 (37%) *Culicoides* species sampled (Tables 3.2.1 and 3.2.2). *Rickettsia*-positive species of biting midge were recorded across *Culicoides* subgenera. Infection was identified across the subgenera Beltranmyia (1/3 species), Culicoides (7/11 species), Monoculicoides (2/3) and Oecacta (1/4 species) [as determined by Borkent (2016)]. There was no apparent host sex bias in the presence of *Rickettsia* for either *C. pulicaris P1* (UK) (Fisher's two-tailed test; p=1) or *C. impunctatus* (Fisher's two-tailed test; p=0.36), the only infected species with both host sexes available to compare.

Rickettsia was found in all individuals in 16 of the 20 positive populations screened, with prevalence being at low or intermediate prevalence in the remaining 4 (1 *C. newsteadi* N1 population and 3 *C. impunctatus* populations) (Table 3.2.1. and 3.2.2; Figure 3.1.). Where multiple samples of particular species were tested, there was no significant difference in the fraction infected (*C. impunctatus* populations from Bala vs Kielder in the UK, N1=31, N2=23, Fisher's two-tailed test; p=0.37). Mitochondrial DNA barcoding of infected (KY765353) and uninfected (KY765354) individuals of *C. impunctatus* confirmed these individuals shared a barcode, consistent with infection showing within-species polymorphism.

Subgenus	*Culicoides* Species	Locality (Site name)	Year of Collection	Proportion of Female *Rickettsia* positive conventional PCR (n) [95% Confidence interval]	Proportion of male *Rickettsia* positive conventional PCR (n) [95% Confidence interval]
Beltranmyia	*C. salinarius*	Sweden (Unknown site)	2009	**1** (2) [0.2-1]	
Culicoides	*C. impunctatus*	Sweden (Torsås)	2008	**0.3** (20) [0.13-0.54]	
		UK (Bala)	2012	**0.81** (17) [0.5-0.92]	**0.5** (14) [0.27-0.73]
		UK (Kielder)	2016	**0.75** (16) [0.47-0.92]	**0.86** (7) [0.42-0.99]
	C. newsteadi N1[a]	Corsica (2BPL2)	2015	**0.5** (2) [0.1-0.91]	
	C. newsteadi N2[a]	Corsica (2BPL2)	2015	**1** (2) [0.2-1]	
	C. newsteadi N3[b]	Sweden (Unknown site)	2008-2010	**1** (6) [0.52-1]	
	C. newsteadi N5[b]	UK (Neston)	2015	**1** (13) [0.72-1]	
	C. pulicaris P1	UK (Canterbury)	2014	**1** (2) [0.2-1]	
		UK (Hereford)	2014	**1** (1) [0.05-1]	
		UK (Luton)	2014	**1** (1) [0.05-1]	
		Sweden (Unknown site)	2008-2010	**1** (6) [0.52-1]	
		UK (Neston)	2015	**1** (32) [0.87-1]	
		UK (Wolverhampton)	2013	**1** (11) [0.68-1]	**1** (4) [0.4-1]
		UK (Worcester)	2014	**1** (6) [0.52-1]	
	C. bysta[*]	Corsica (2BPL2)	2015	**1** (13) [0.72-1]	
Monoculicoides	*C. riethi*	Sweden (Ljungbyholm)	2010	**1** (1) [0.05-1]	
	C. stigma	Sweden (Unknown site)	2008	**1** (3) [0.31-1]	
Oecacta	*C. duddingstoni*	Sweden (Bara)	2008	**1** (4) [0.4-1]	
		Sweden (Unknown site)	2008-2010	**1** (3) [0.31-1]	

Table 3.2.1. *17KDa* conventional PCR assay results for *Rickettsia*-positive *Culicoides* species. *Culicoides newsteadi* haplotypes are designated by [a]Pagès *et al.*, 2009 and [b]Ander *et al.*, 2013.

Subgenus	*Culicoides* species	Location	Year of Collection	Proportion of Female *Rickettsia* positive conventional PCR (n) [95% Confidence interval]	Proportion of male *Rickettsia* positive conventional PCR (n) [95% Confidence interval]
Avaritia	*C. bolitinos*	South Africa (Onderstepoort)	2016		**0** (2) [0-0.8]
		South Africa (Port Elizabeth)	2014-2016	**0** (18) [0-0.22]	
	C. imicola	South Africa (Onderstepoort)	2016	**0** (23) [0-0.18]	**0** (9) [0-0.37]
		Corsica (2APL7)	2016	**0** (24) [0-0.17]	
		Corsica (2BPL2)	2016	**0** (16) [0-0.24]	
	C. obsoletus	Sweden (Bara)	2008		**0** (8) [0-0.8]
		UK (Neston)	2012-2015	**0** (14) [0-0.27]	
Beltranmyia	*C. circumscriptus*	Sweden (Unknown site)	2009	**0** (1) [0-0.95]	
	C. sphagnumensis	Sweden (Axvalla)	2009	**0** (1) [0-0.95]	
Culicoides	*C. brucei*	South Africa (Onderstepoort)	2016	**0** (1) [0-0.95]	
	C. grisescens	Sweden (Torsås)	2008	**0** (4) [0-0.6]	
	C. newsteadi N1a	Corsica (2APL7)	2015	**0** (4) [0-0.6]	
	C. newsteadi N6[b]	UK (Wolverhampton)	2014		**0** (1) [0-0.95]
		Corsica (2APL7)	2015	**0** (12) [0-0.30]	
		Corsica (2BPL2)	2015	**0** (3) [0-0.69]	
	C. punctatus	UK (Wolverhampton)	2014	**0** (4) [0-0.6]	
		UK (Neston)	2015	**0** (23) [0-0.18]	
		Sweden (Torsås)	2008	**0** (14) [0-0.27]	
		Sweden (Unknown site)	2008		**0** (9) [0-0.37]

Monoculicoides	*C. sonorensis*[$]	USA (Riverside, CA)	2018	**0** (79) [0-0.06]	
Oecacta	*C. clastrieri*	Sweden (Unknown site)	2009		**0** (1) [0-0.95]
	C. festivipennis	Sweden (Uppsalla)	2009	**0** (1) [0-0.95]	
	C. truncorum	Sweden (Torsås)	2008-2009	**0** (3) [0-0.69]	**0** (2) [0-0.8]
Remmia	*C. subshultzei*	South Africa (Onderstepoort)	2016		**0** (5) [0-0.54]
Silvaticulicoides	*C. achrayi*	Sweden (Axvalla)	2008	**0** (8) [0-0.4]	
	C. subfascipennis	Sweden (Romakloster)	2008	**0** (8) [0-0.4]	
Synhelea	*C. bedfordi*	South Africa (Onderstepoort)	2016	**0** (1) [0-0.95]	
Wirthomyia	*C. reconditus*	Sweden (Unknown site)	2007-2010	**0** (3) [0-0.69]	
	C. segnis	Sweden (Romakloster)	2008	**0** (11) [0-0.32]	

Table 3.2.2. *17KDa* conventional PCR assay results for *Rickettsia*-negative *Culicoides* species. *Culicoides newsteadi* haplotype N1[a] designated by Pagès *et al*. (2009), *Culicoides newsteadi* N6[b] previously undesignated. [$]*C. sonorensis* were screened both as individuals and pooled individuals of 5.

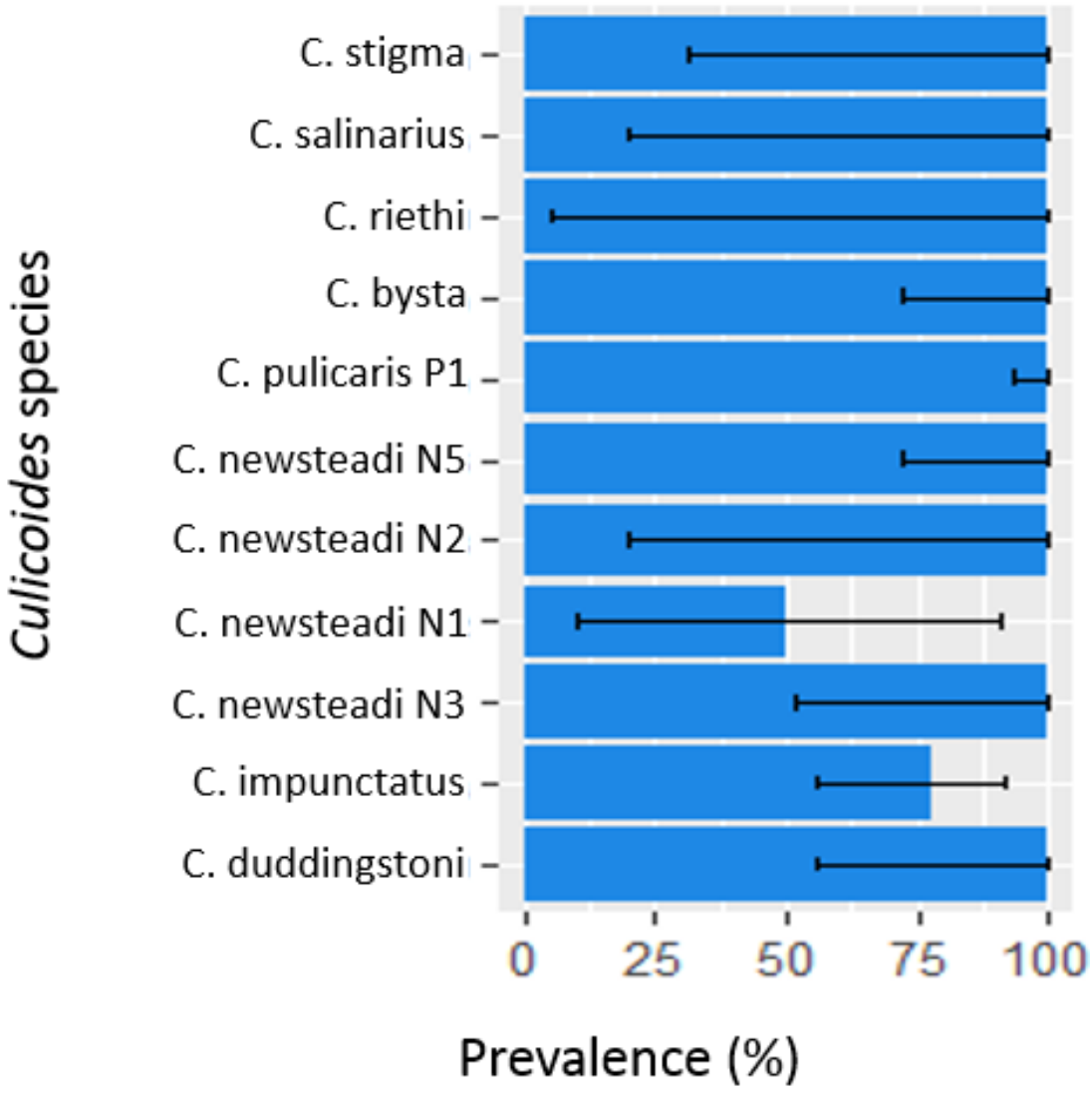


Figure 3.1. Prevalence of *Rickettsia* positive *Culicoides* species. Error bars are 95% confidence intervals of the true infected proportion.

3.4.2 The detection of *Rickettsia* coinfections

The level of 16S rRNA divergence within the *Culicoides Rickettsia* was low (0.9% segregating sites, π=0.002) (Table 3.3), such that the strains would all be considered as belonging to a single species in classic bacteriological nomenclature (Stackebrandt and Goebel 1994). To resolve patterns of relatedness more fully, we obtained the sequence of three further housekeeping loci as well as the *17KDa* gene, for each of the specimens. Housekeeping gene PCR amplification was successful for 13 individuals; however, on visualisation of *C. salinarius* and *C. duddingstoni* (Unknown site, Sweden) chromatograms, double peaks were identified at the third codon positions of protein-coding genes. The possibility of DNA extract cross-contamination (as seen in chapter 2) was eliminated as the respective chromatograms for host identification, using the *COI* gene, gave clear base-calls. Furthermore, similar patterns of multiple peaks had been observed in *Rickettsia*-infected *Sympetrum* sp. (Odonata, Libellulibidae) screened by a colleague (Panupong Thongprem, University of Liverpool). Subsequently, it was inferred that these individuals were likely coinfected with two strains of *Rickettsia* and were removed from subsequent genetic comparison with other isolates, as the phase of the variants could not be known.

3.4.3 Individual allele information

Of the 11 remaining strain typings (Table 3.4), the *C. pulicaris* strain (G) from the UK failed to amplify with the *coxA* primers after more than one attempt. An exclusive allele was designated to this locus, because non-amplification implies the genotype of this strain is unique at the priming site, as failure to amplify occurred on a background of successful amplification for other loci in these specimens. All gene sequences, including the non-housekeeping gene *17KDa*, maintained an intact coding frame, consistent with their presence in a symbiont genome, rather than a nuclear insertion of a *Rickettsia* gene.

Locus	Number of Alleles	Nucleotide diversity/site (π)	Variable sites (%)	G+C content	K_a/K_s average	Recombination (MaxChi and Chimaera *p*<0.001)
atpA	8	0.048	9.6	0.399	0.059	Yes
coxA	6	0.008	2.6	0.372	0.090	No
gltA	6	0.018	4.4	0.369	0.022	No
16S	7	0.002	0.9	0.516	N/A	No
17KDa	6	0.021	10.4	0.426	0.299	No

Table 3.3. Genetic characteristics of housekeeping and *17KDa* alleles.

Species	Strain identity	Clonal complex	*atpA*	*coxA*	*gltA*	*16S*	*17KDa*
Culicoides stigma	A	1	1	1	1	1	1
Culicoides newsteadi N3	A	1	1	1	1	1	1
Culicoides riethi	B	1	1	1	1	3	1
Culicoides newsteadi N1	C	2	2	2	2	2	2
Culicoides duddingstoni (Bara, Sweden)	C	2	2	2	2	2	2
Culicoides pulicaris haplotype 1 (Sweden)	D	2	3	2	2	2	2
Culicoides bysta	E	2	4	2	2	2	2
Culicoides newsteadi N2	F	3	5	3	3	4	3
Culicoides pulicaris haplotype 1 (UK)	G	4	6	4 (NA)	4	5	4
Culicoides newsteadi N5	H	5	7	5	5	6	5
Culicoides impunctatus	I	6	8	6	6	7	6

Table 3.4. *Rickettsia* strains recovered from *Culicoides* midges, with allelic profiles; strains sharing the same allelic profiles at all five loci were designated as a single strain. NA=non amplifiable.

The most polymorphic housekeeping locus was a*tpA*, with 9.6% variable sites and the highest level of nucleotide diversity per site (π=0.048) (Table 3.3.). All genes showed average K_a/K_s of less than 1, indicating that the genes were subject to purifying selection, conforming to the general requirements for reliable indicators of genetic relatedness between bacterial isolates. Predictably, as an antigenic protein with less intense purifying selection and potential episodes of positive selection, *17KDa* had a greater average K_a/K_s (0.299) than the other loci; although no signs of positive selection were observed at the gene-level.

Whilst there was evidence the strains found within *Culicoides* were closely related, some loci demonstrated 100% sequence identity with *Rickettsia* strains from other taxa. These included the partial *gltA* sequences of *C. impunctatus*, which was identical to the *Rickettsia* symbionts of the beetle *Deronectes platynotus* (Dytiscidae; FM177878) (Küchler *et al*., 2009), the Dipteran fly *Chrysotimus flaviventris* (Dolichopodidae; JQ925578) (Martin *et al*., 2013) and the spider *Pityohyphantes phrygianus* (Linyphiidae; DQ 231491) (Goodacre *et al*., 2006), and the partial *16S* sequences of clonal complex 2 strains (*C. duddingstoni*; Bara, Sweden), *C. pulicaris* haplotype 1 (Sweden), *C. newsteadi* N1, *C. bysta* which were identical to the *16S* sequence of the *Rickettsia* in the cranefly *Limonia chorea* (Limoniidae; AF322443). Furthermore, a *coxA* 995 bp region of the Hemipteran bug *Macrolophus* sp. *Rickettsia* 1 (Miridae; HE583223) (Machtelinckx *et al*., 2012) was >99% similar to all *Culicoides*' strains except for *C. impunctatus*.

3.4.4 Multilocus strain typing and geographic comparison

I next examined the relationship of the *Rickettsia* strains from different *Culicoides*' species using allelic profiles across loci (Figure 3.2.). The number of alleles per locus ranged from 6 to 8, with most allelic profiles obtained from different host populations (9/11) being unique.

Furthermore, of these 9 unique allelic profiles, 4 (F, G, H and I) shared no alleles with other strains (Table 3.4.). Allelic profiles which were shared by more than one host species were designated as central strains (CSs); whereas isolates which varied at one locus to these CSs were termed single locus variants (SLVs). Together the CSs and SLVs form clonal complexes, as they are presumed to be closely related. Two clonal complexes were identified in this study (Figure 3.2.); the central strain A from *C. stigma* and *C. newsteadi* N3 formed clonal complex 1 with the SLV strain from *C. riethi* (B); whereas the central strain C from *C. newsteadi* N1 and *C. duddingstoni* (Bara, Sweden) formed clonal complex 2 with the SLV strains from *C. pulicaris P1* (Sweden) (D) and *C. bysta* (E).

When strains were mapped by geographic location (Figure 3.2.), no clear pattern between host, location and isolate was observed. Strain C was observed in both Corsica, France and Sweden in *C. newsteadi N1* and *C. duddingstoni* respectively; *Culicoides* species of differing subgenera. However, *C. impunctatus* contained the same strain in 3 populations in the UK (Bala and Kielder), as well as Sweden (Torsås).

3.4.5 Phylogenetic analysis

Aside from biting midges, sequences of all 5 strain-typing genes were generated in the following host taxa; *Hilara interstincta* (Diptera: Empididae), *Zavrelimyia sp*. (Diptera: Chironomidae), *Glyptotendipes sp.* (Diptera: Chironomidae), *Macrolophus pygmaeus* (Hemiptera: Miridae), *Pyrrhosoma nymphula* (Odonata: Coenagrionidae), *Enallagma cyathingerum* (Odonata: Coenagrionidae), *Coenagrion mercuraile* and *Coenagrion puella* (Odonata: Coenagrionidae). Further taxa which gave incomplete profiles were an unidentified *Chironomidae* sp. and the black fly *Simulium aureum* (Appendix 1). The *atpA* gene was the

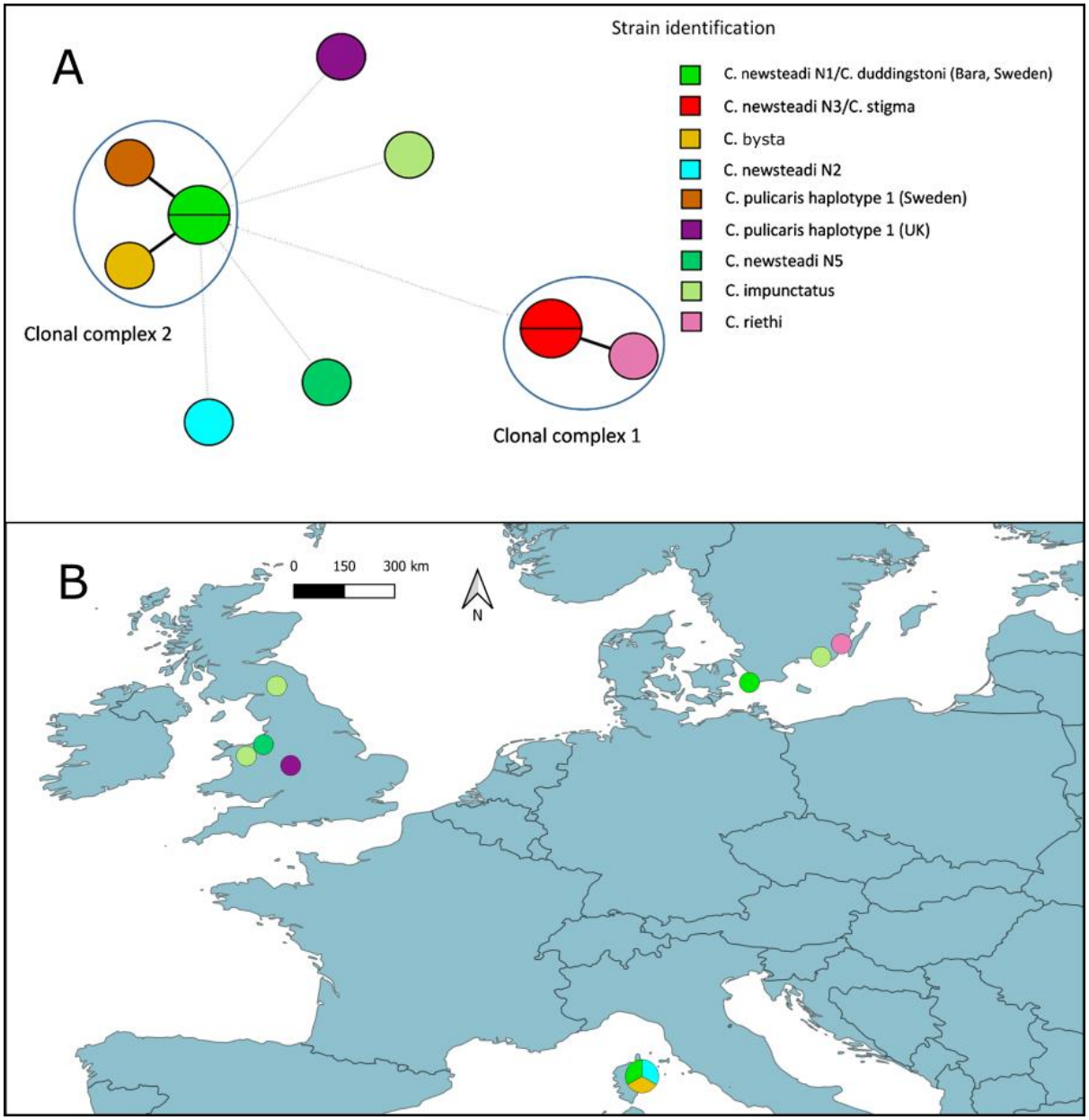


Figure 3.2. A) Minimum spanning tree using unweighted pair group method with arithmetic mean (UPGMA) cluster analysis of isolates. Allelic profiles that are shared by more than one host species are designated as central strains (CSs). Strains differing at one locus (SLVs) are connected by a solid line, and unique strains sharing no allele identity are connected by a faded line. **B)** A map depicting *Rickettsia* strain distribution across the UK, Sweden and Corsica, France. Some strains are absent as collection site information was unavailable.

only locus to exhibit signs of intragenic recombination through detection by RDP v4 (Martin *et al.*, 2015), with both the MaxChi (p=1.78x10^{-6}) and Chimaera (p=1.25x10^{-7}) algorithms detecting a breakpoint at position 789 of the gene (Table 3.6.).Visualisation of ML trees either side of the breakpoint demonstrated *Culicoides impunctatus* and *Hilara interstincta* as a likely recombinant sequence (Figure 3.3.). Subsequently, these loci were removed from further phylogenetic analysis. Comparisons of individual gene trees showed no clear signs of intergenic recombination between *Rickettsia* groups (Figures 3.4.1-3.4.5).

The concatenated phylogeny of housekeeping genes and a single antigenic protein gene placed the midge strains in the Torix group (Figure 3.5.), a basal *Rickettsia* clade known to infect both haematophagous and non-blood feeding arthropods. Group Torix has previously been split into two subclades; “Leech” and “Limoniae” with all isolates from this study placing in the latter except for the *Simulium aureum* strain which grouped in the Leech group. Multiple taxa such as *Hilara interstincta*, *Deronectes* water beetles and Araneae (spider) strains were observed in both subclades.

Constrained trees to test alternative monophyletic hypotheses with respect to host taxa, were significantly different (p<0.001) for all comparisons of *Culicoides*, Odonata, *Deronectes* and Empididae but not for Glossiphoniidae (Leeches) (p=0.145) (Table 3.6.). Furthermore, there were near identical strains of both *Enallagma cyathingerum* and *Coenagrion mercuriale* (odonates) to the major isolate of *Culicoides* clonal complex 1 (*Culicoides stigma* and *Culicoides newsteadi N3;* strain A); just 1 nucleotide out of 3896 (0.03%) was different over the 5 loci, both at the *atpA* gene.

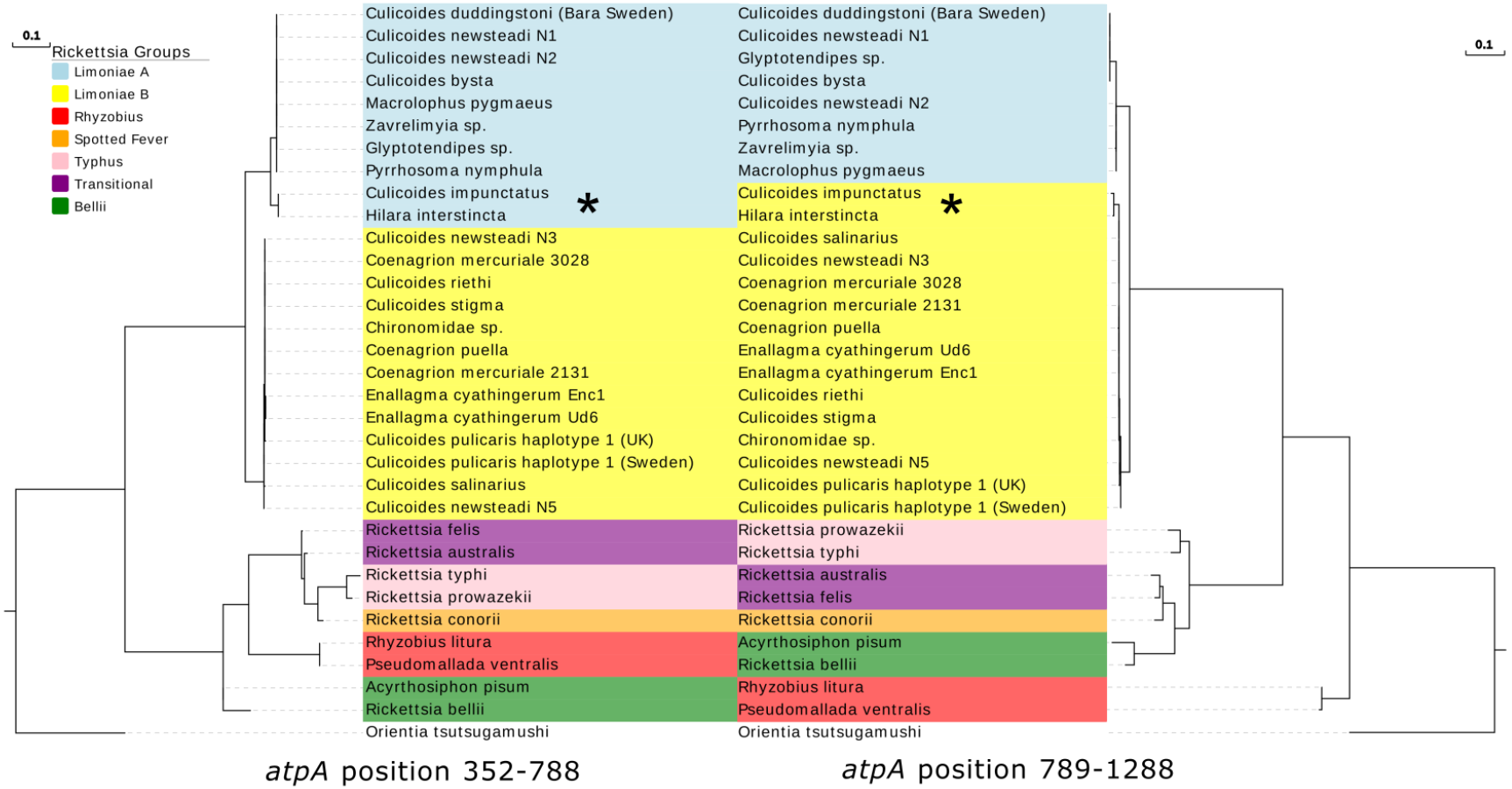


Figure 3.3. Mirrored maximum likelihood trees of the *atpA* gene either side of a recombination breaking point detected by RDP v4. Asterisks depict putative recombinant sequences.

Putative recombinant sequence	Putative minor parental sequence	Putative major parental sequence	RDP *p* value	Maxchi *p* value	Chimaera *p* value
Culicoides impunctatus/Hilara interstincta	*Coenagrion puella*	*Glyptotendipes* sp.	1.50E-02	1.78E-06	1.25E-07

Table 3.5. Evidence of *atpA* intragenic recombination in *Hilara interstincta* and *Culicoides impunctatus* isolates as determined with RDP4. Minor parent = Parent contributing the smaller fraction of sequence. Major parent = Parent contributing the larger fraction of sequence.

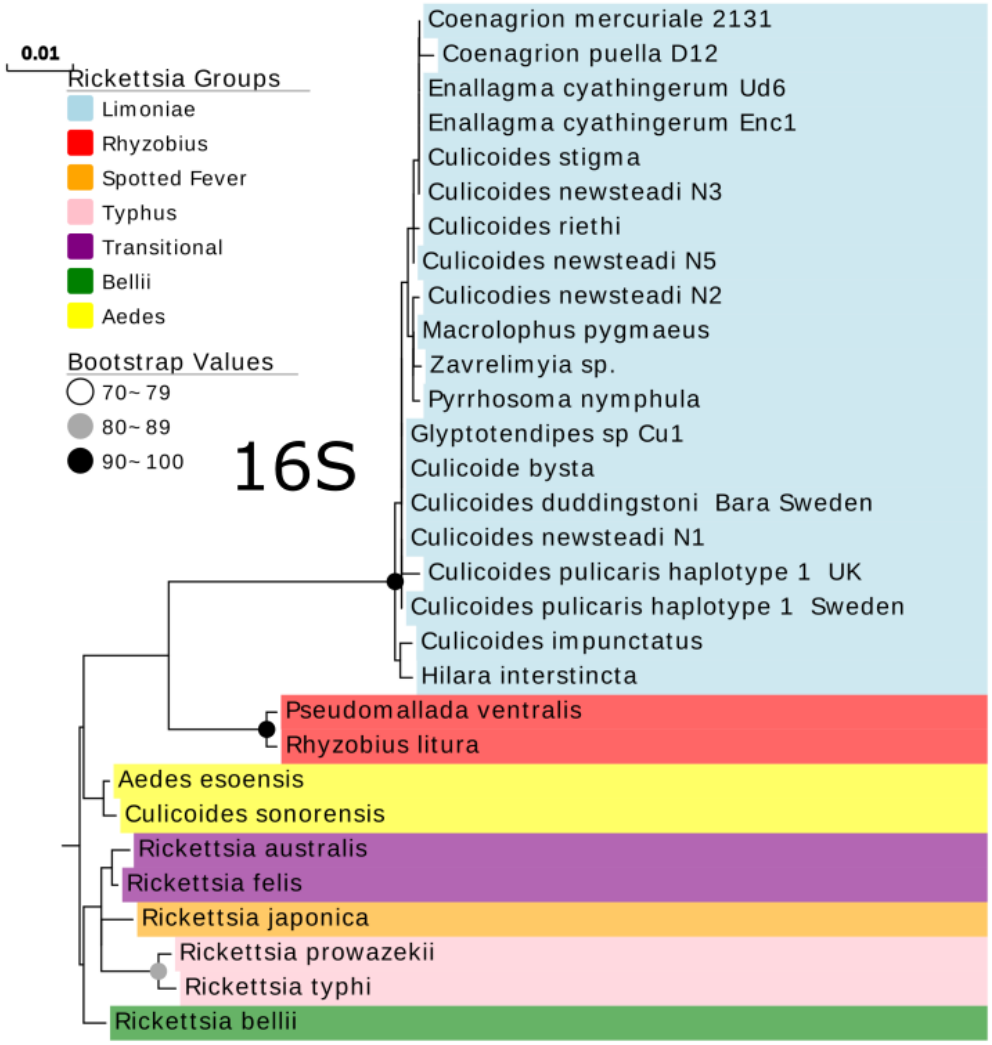


Figure 3.4.1 *16S* maximum likelihood phylogeny inferred using IQTree under the GTR+F+R4 model and 100 bootstraps.

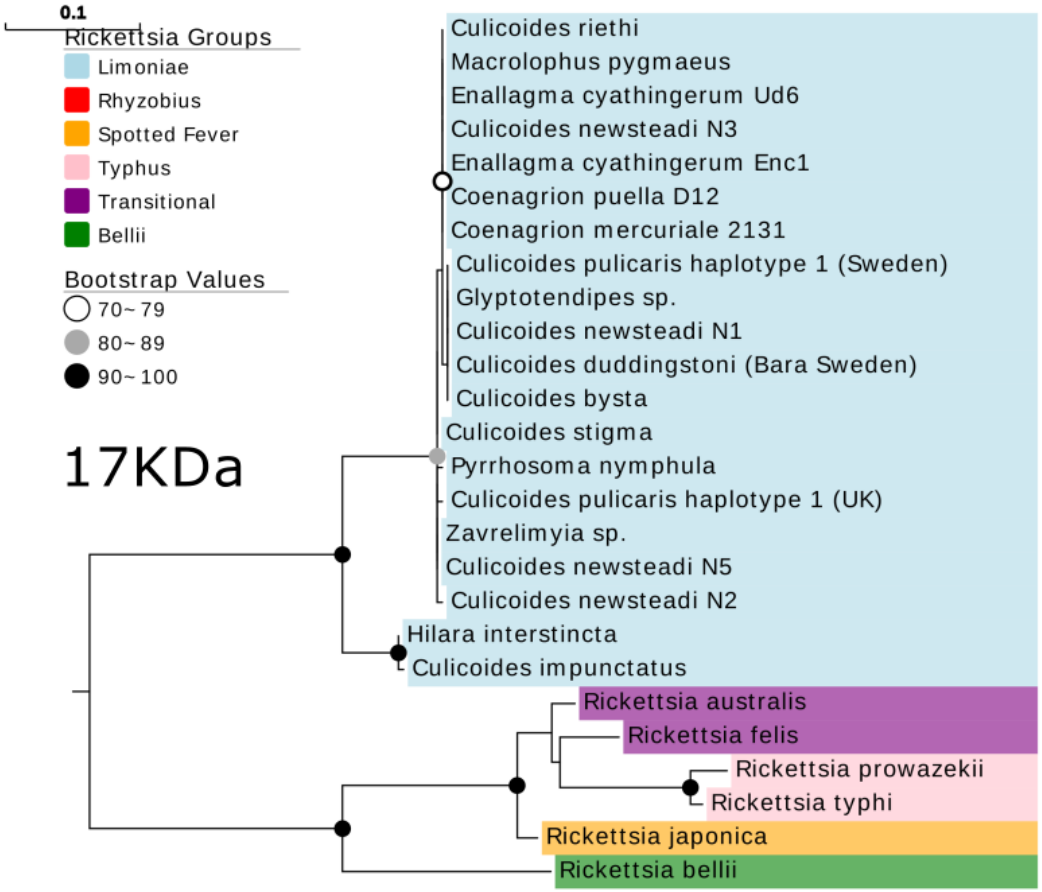


Figure 3.4.2 *17KDa* maximum likelihood phylogeny inferred using IQTree under the TVM+F+I+G4 model and 100 bootstraps.

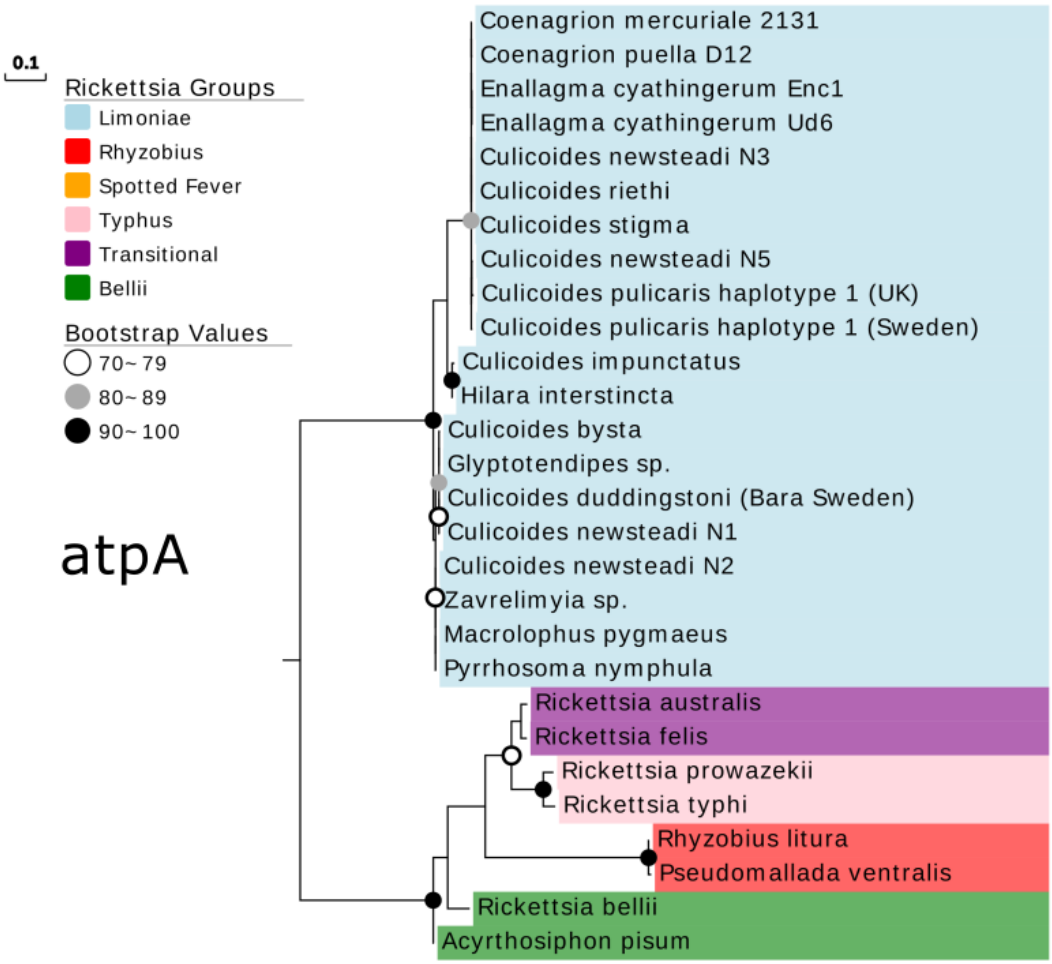


Figure 3.4.3 *atpA* maximum likelihood phylogeny inferred using IQTree under the GTR+F+I+G4 model and 100 bootstraps.

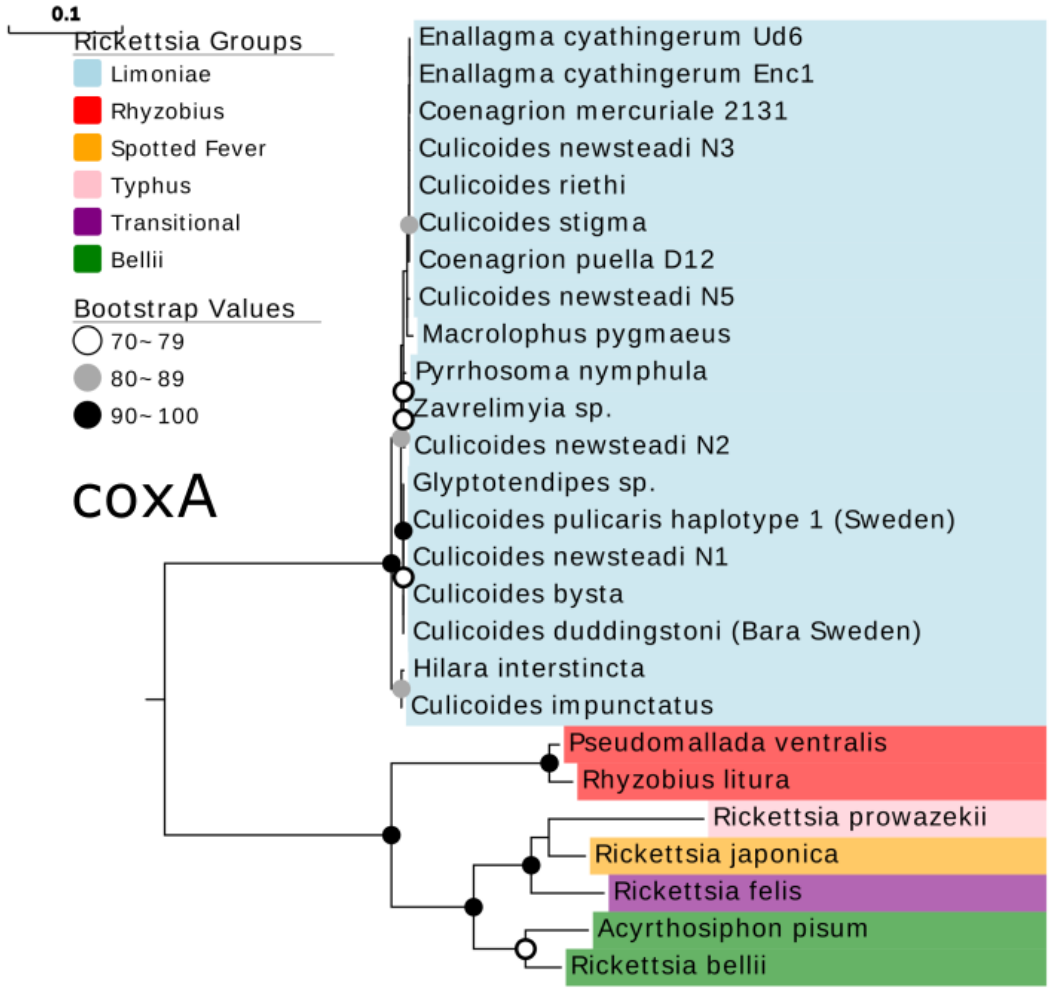


Figure 3.4.4 *coxA* maximum likelihood phylogeny inferred using IQTree under the TVM+F+I+G4 model and 100 bootstraps.

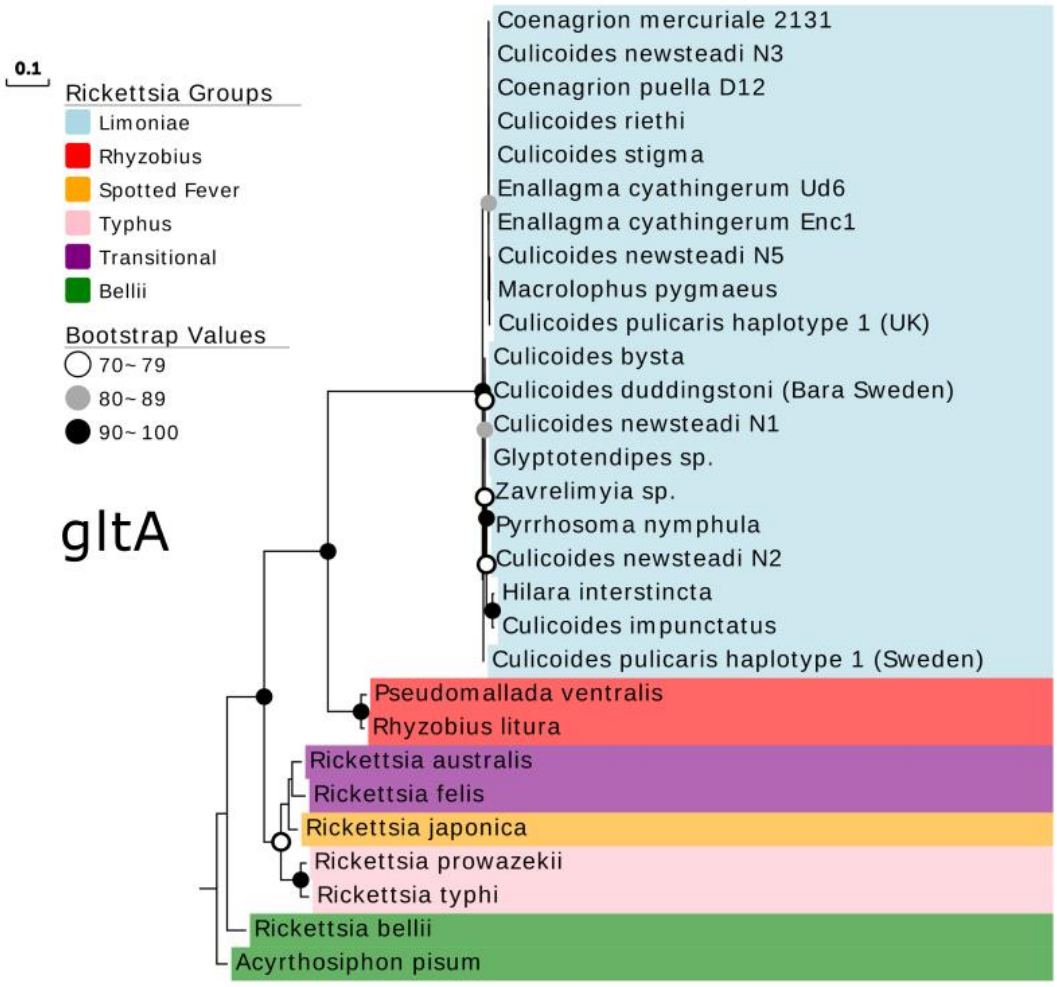


Figure 3.4.5 *gltA* maximum likelihood phylogeny inferred using IQTree under the GTR+F+R3 model and 100 bootstraps.

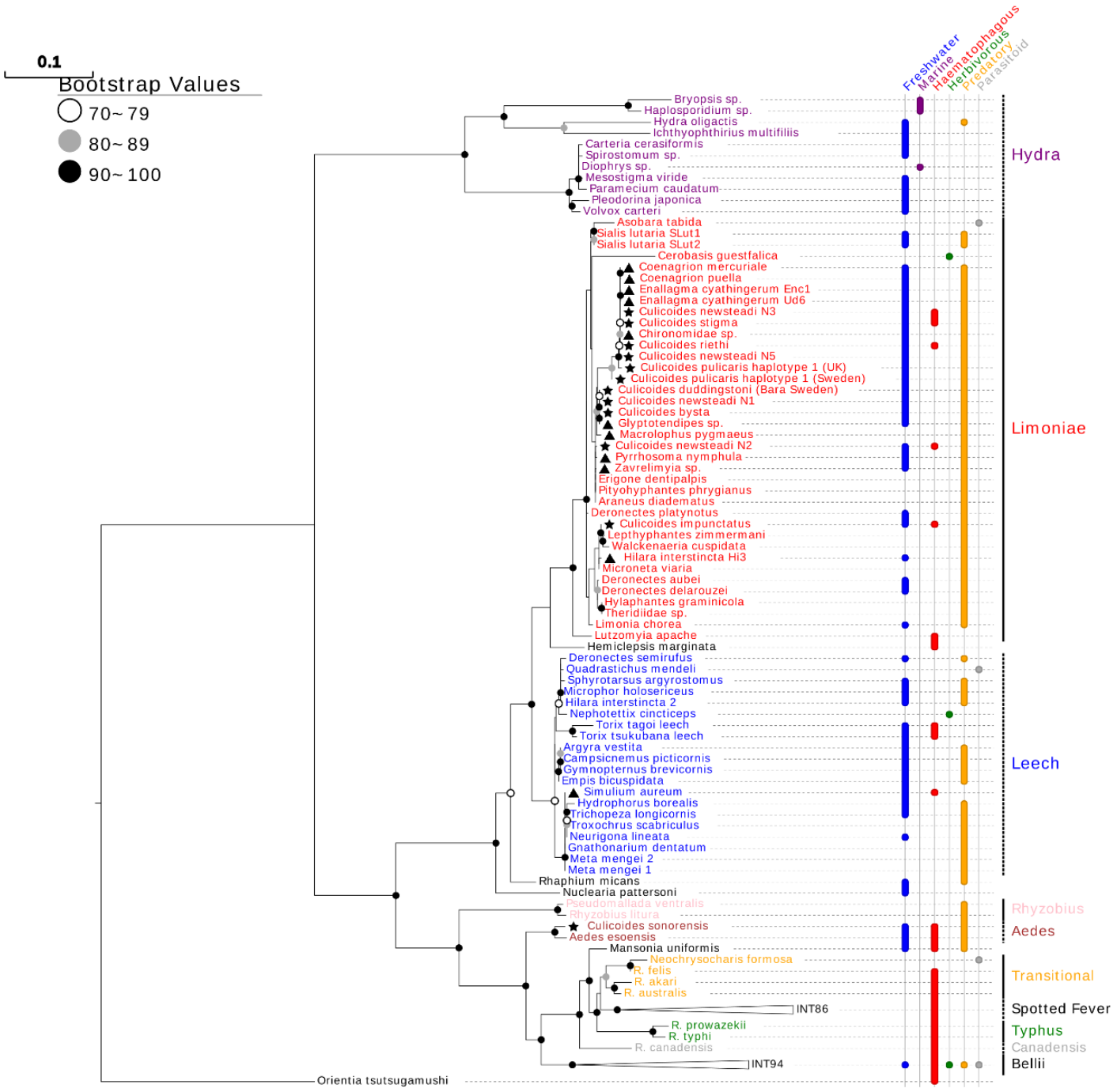


Figure 3.5. Phylogenetic placement of the *Rickettsia* symbionts of *Culicoides* midges based on a 4403 nucleotide concatenated gene set of *16S*, *17KDa*, *atpA*, *coxA* and *gltA* genes. Previously characterized *Rickettsia* groups including the basal group of Hydra are also presented. Host names are used in the absence of official *Rickettsia* species names in most cases. The tree topology was inferred using Maximum Likelihood (ML) analysis in IQTree under a partition model and 5000 ultrafast bootstraps. Black stars depict *Culicoides*' *Rickettsia* and black triangles depict other isolates generated in this study. Sequence accession numbers are available in Appendix 1. The habitats and lifestyles of the host are given to the right of the phylogeny.

Enforced host monophyly	Tree topology likelihood (-ln)	Likelihood difference (Δln)	SH-test *p* value
Unconstrained	24042.82	0	1
Culicoides spp. (Diptera: Ceratopogonidae)	24804.24	761.42	<0.0001
Chironomidae (Diptera)	24545.04	502.22	<0.0001
Deronectes spp. (Coleoptera: Dytiscidae)	24257.21	214.39	<0.0001
Coenagrionidae (Odonata)	24501.13	458.31	<0.0001
Empididae (Diptera)	24272.97	230.15	<0.0001
Glossiphoniidae (Hirudinea)	24104.48	61.66	0.145

Table 3.6. Comparison of the tree topologies of the unconstrained concatenated dataset against topologies where host monophyly was enforced using the Shimodaira-Hasegawa (SH) test.

In contrast, strains from *Culicoides impunctatus*, which also placed in the Limoniae group, showed a divergence of 3.2% when compared to the same strain from clonal complex 1.

Despite all midge strains from Palearctic regions clustering in the Limoniae clade, an isolate was detected from a single *Culicoides sonorensis* which was placed with an isolate discovered recently in the mosquito *Aedes esoensis* from the demilitarized zone in Korea (Maina *et al.*, 2017) (Figure 3.5.). Only the *16S* locus amplified with this sample, with all other primers failing to generate a band for sequencing. Despite this, the clade described in this study as the "Aedes" group after the initial host taxa it was discovered (as is convention for the genus), is well supported and places basal to derived pathogenic groups but is a derivative of Torix and Rhizobius clades.

3.5 Discussion

In this study, we show Limoniae group *Rickettsia* are common in biting midges, and thus represent a previously unrecognized component of the biology of this important vector group. Previous work on *Culicoides*, using conventional PCR to establish the presence of the heritable symbiont *Cardinium*, revealed interspecies infection rates ranging from 16% to 29% (Nakamura *et al.*, 2009; Lewis *et al.*, 2014; Mee *et al.*, 2015). Thus, our PCR screen suggests that *Rickettsia* is the most common symbiont of *Culicoides* identified so far, being present in 11 of 30 species tested (37%) and in 100% of specimens examined in 9 of the *Rickettsia* positive species. Hence, this *Rickettsia* clade represents an important associate found widely in *Culicoides* midges. It is noteworthy that our assessment of incidence is conservative, being based on a conventional PCR assay which will likely report false negatives for low titre infections.

A *Rickettsia* with 99% identity to Torix Leech isolates has previously been described in *Culicoides sonorensis* by Campbell *et al*. (Campbell *et al.*, 2004). Although the sequence of the strain from this midge is not available, due to the similar homology found in this study between our *Rickettsia* and the same leeches, it can be presumed that these are of the same clade. Subsequently, a focussed screening of *Culicoides sonorensis* from California, USA was undertaken due primarily to its significance as a major vector of BTV in North America. A single individual from a population of *Culicoides sonorensis* (0/29 extracted individuals and 1/10 pooled samples) was found to be PCR positive. This individual, identified from a pool of 5 midges, was positive through *16S* PCR screening which was later confirmed as *Rickettsia* by Sanger sequencing. Other housekeeping gene markers and *17KDa* screening gave negative results, indicating a divergent isolate compared to the Torix *Rickettsia* found in other

Culicoides species. Furthermore, through phylogenetic analysis (Figure 3.5.), the isolate was shown to cluster in a sister group of *Rickettsia* clades occasionally known to induce pathogenicity in vertebrates. A similar isolate was also recently reported in *Aedes* mosquitoes from the demilitarised zone of Korea. The placement as a sister clade of derived pathogenic isolates, alongside the presence in two blood feeders and its low prevalence, suggests a possible haematophagous transmission route of this *Rickettsia*. Subsequently, the *Rickettsia* isolate reported here may play a role in animal and/or human health which requires further investigation.

Of note is the absence of Limoniae *Rickettsia* in this *C. sonorensis* Californian population. Campbell *et al*. (2004) found Torix *Rickettsia* in 3/10 individuals from Colorado suggesting it would be likely to be detected in the author's screen considering the higher sample size. The lack of detection could be due to geographical barriers, such as the Rocky Mountains, between the two populations. Regardless, the heterogeneous infection status of *C. sonorensis* in the USA may have an impact on spatiotemporal infectious disease dynamics of BTV, if indeed the presence of this endosymbiont affects vectorial capacity. Both *C. pulicaris* and *C. newsteadi* complexes are thought to be vectors of BTV, with this study demonstrating them both to be *Rickettsia* infected. Despite this, the major vectors of BTV in Europe and Africa, *C. obsoletus* and *C. imicola* respectively, are shown to be uninfected.

The Torix group of *Rickettsia* has been recorded previously in an array of invertebrate species. Many of these species share ecological characteristics including an aquatic phase and predatory larval stages (e.g. biting midges, diving beetles, leeches, crane flies) (Figure 3.5.). Others are notable for haematophagy (e.g. biting midges, leeches, sandflies). Moreover, no

secondary associations with vertebrate hosts or pathogenicity have been associated so far with this *Rickettsia* group. Due to the initial lack of available Torix *Rickettsia* homologues of the different loci amplified in this study, a multi-locus typing scheme was used to assess relatedness between midge *Rickettsia* strains. This revealed 2 clonal complexes of close relatedness. The presence of 2 identical strains in differing hosts, as is seen in *C. stigma*/*C. newsteadi N3* and *C. duddingstoni*/*C. newsteadi N1*, indicate a lack of host specificity and suggests frequent host shifts. A similar scenario has previously been observed in the Panamanian leaf-cutting ants *Acromyrmex echinatior* and *Acromyrmex insinuator*, where shared *Wolbachia* isolates were observed in the cohabiting species (Van Borm *et al.*, 2003).

The further inclusion of additional sequences from other Torix-infected taxa, showed near identical strains between a *Culicoides* isolate in *C. stigma*/*C. newsteadi N3*, and the odonates *Enallagma cyathingerum* and *Coenagrion mercurial* (Figure 3.5.). This identity suggests recent host shifts are not limited to closely related taxa but can extend to horizontal transmission of distantly related taxa. The pattern contrasts with the midge *Cardinium* (group C, chapter 2) which is apparently restricted in its incidence. Host shift events have been created under experimental conditions with *Wolbachia* from *Drosophila simulans* being transmitted to the parasitic wasp, *Leptopilina*, before undergoing vertical transmission in this novel host (Heath *et al.*, 1999). Due to the predaceous nature of Torix hosts, it is possible that contamination from ingested prey is accounting for the observed similarity in strains. However, the observation of multiple populations of *C. impunctatus* from different locations containing identical strains (Figure 3.2.), indicates this is unlikely as it is improbable that multiple populations acquired identical *Rickettsia* isolates from their environment by chance alone.

Despite this, imaging techniques such as electron microscopy or Fluorescence *in-situ* hybridisation (FISH) is needed in order to confirm a true endosymbiosis (see chapter 4).

The sporadic patterns of *Rickettsia* across a range of both related and unrelated host taxa raises the question of how this *Rickettsia* endosymbiont is maintained in populations. Endosymbiotic bacteria maintain high prevalence rates in populations via high fidelity maternal transmission often leading to co-cladogenesis between symbiont and host (Chen, Li and Aksoy, 1999; Clark *et al.*, 2000). However, incongruencies between host and symbiont phylogenies in *Wolbachia* suggests host shifts also occur as a strategy to explore new niches (Turelli *et al.*, 2018). The significantly worse topologies produced when host monophyly is enforced in this study (Table 3.6.), suggests this lack of co-speciation is also the case with Torix *Rickettsia*. Therefore, the high prevalence and close relatedness of midge *Rickettsia* strains in distantly related hosts indicate a mixture of vertical and horizontal transmission.

Previous reports have also suggested mixed transmission strategies for *Rickettsia* (Perlman, Hunter and Zchori-Fein, 2006; Caspi-Fluger *et al.*, 2012). Due to the lifestyles of haematophagous hosts which spread pathogenic *Rickettsia* strains to vertebrates, it has been assumed that blood-feeding on common infected hosts has led to host shifts in these groups. This is supported by other Rickettsiales families, *Ehrlichia*, *Anaplasma and Orientia,* all sustaining infection through blood-feeding in experimental settings (Saito and Walker 2015; Ogden *et al*. 2003; Frances *et al*. 2000). Although this may be the case for some Rickettsiales, the presence of non-blood feeders in the Torix group, suggests alternative routes of transmission may be occurring.

As host shifts of endosymbionts necessitate intimate interactions between current and putative hosts, predation and parasitism have been proposed to account for horizontal transfer events (Heath *et al.*, 1999; Noda *et al.*, 2001; Ahmed *et al.*, 2015; Mascarenhas *et al.*, 2016). All of the dipteran hosts so far identified in the Limoniae subclade of Torix have larval stages which are predatory. In addition, the adults of spiders, damselflies, predatory bugs and dragonflies are also predators. Despite this, the isolate from the black fly *Simulium aureum,* which was placed in the Torix subclade Leech, has no predatory life stages (although it is possible filter feeders could acquire rickettsial symbionts from their diet due to symbionts in protozoans). Another connection of note is the aquatic lifestyle of a majority of taxa in the Torix group. Weinert *et al*. (2009) have previously suggested host switching could be occurring between *Rickettsia*-infected freshwater taxa and arthropods, specifically via protists. This is on account of the ancestral *Rickettsia* Hydra group (recently renamed *Megaira*), which contains aquatic ciliate hosts. Additionally, the freshwater amoeboid parasite of fish, *Nuclearia pattersoni*, has been shown to belong to the Torix group (Dyková *et al.*, 2003). An alternative hypothesis to explain host-shifts includes the transmission via endoparasitoids. Torix-infected parasitoid wasps *Asobara tabida* and *Quadrastichus mendeli* should be considered for such events although the latter was shown to contain a different strain of *Rickettsia* to its gall wasp host *Leptocybe invasa* (Gualtieri *et al.*, 2017). The transmission strategies of Torix *Rickettsia* are discussed further in chapters 4 and 5.

The segregation of the Torix group into Leech and Limoniae groups had originally been suggested by Perotti *et al*. (2006), although only limited taxa and sequence data were available. The inclusion of the genes *atpA*, *coxA* and *17KDa* in this project allowed for the confident placement of sequences from previous studies which had relied primarily on just

one of 16S or *gltA* genes. Common taxa were observed in both Limoniae and Leech subclades including, spiders, empidid dance flies and *Deronectes* water beetles. Despite this pattern suggesting a common transmission route of the endosymbiont as mentioned above, the divergence into two subclades could offer different effects on host biology. So far, the only genome available comes from *Culicoides newsteadi N5* (RiCNE) in the Limoniae group (Pilgrim *et al.*, 2017). However, comparative analysis with a Leech counterpart may offer insights into differences between accessory genomes and subsequently specific effects on arthropod hosts.

A further intriguing finding was the presence of coinfections of *Rickettsia* in *Culicoides* with multiple strains being detected in *C. salinarius* and *C. duddingstoni*. Coinfections of different *Wolbachia* strains have been attributed to exogenous acquirement through wounds of cohabiting insects, as well as prey-predator or parasitoid interactions (Rigaud and Juchault 1995; Werren *et al*. 1995). Although coinfections can lead to competition for resources and space, there may be a selection advantage for multiple infections as a result of an increased potential to expand genetic repertoire through horizontal gene transfer. The restrictive lifestyle of intracellular bacteria can often lead to limited opportunities for bacterial genome evolution. However, through mobile genetic elements facilitated by plasmids and bacteriophages, this can lead to the novel exchange of genes between symbionts infecting the same cell (Frost *et al.*, 2005; Kent *et al.*, 2011). Furthermore, theory predicts coinfections persist if a coinfected female produces more coinfected daughters than a single infected female produces singly infected daughters. A specific evolutionary advantage for coinfections of maternally inherited endosymbionts is the induction of bi-directional cytoplasmic incompatibility (CI); the failure of reproduction in male and females containing different

Wolbachia strains. However, CI has not been observed in *Rickettsia* and so multiple infections may provide a fitness advantage for infected individuals in a different way (discussed further in chapter 5).

At least one recombination event, such as those facilitated by coinfections, was observed for the *atpA* gene in *Hilara intersincta* and *Culicoides impunctatus* (Table 3.5. and Figure 3.3.). As recombination can confound the true relationship between isolates, the loci of both strains were removed from phylogenetic analysis. This is the first time recombination has been detected in the Torix group of *Rickettsia*, although similar genetic reshuffling has also been observed in Transitional, Rhizobius and Adalia groups, in the genes *coxA* and *gltA* (Weinert *et al*. 2009). *Rickettsia* plasmids have been described several times across the genus, implicating them as facilitators for such processes (Ogata *et al.*, 2005; Baldridge *et al.*, 2007; Blanc *et al.*, 2007). However, although the Torix genome of *Culicoides newsteadi N5* (RiCNE) identified a *tra* conjugative DNA transfer element, it was not clear if this was present as part of a chromosome or a low-copy number plasmid (Pilgrim *et al.*, 2017). The common recombination across the genus (Ogata *et al.*, 2005; Blanc *et al.*, 2007; El Karkouri *et al.*, 2016) led to the comparison of single gene trees to uncover any intergenic recombination but topologies appeared similar (Figures 3.4.1-3.4.5).

The impact of this *Rickettsia* on host biology is uncertain. *Rickettsia* infections are known to be associated with a variety of reproductive manipulations of their host (reproductive parasitisms), including male-killing in ladybird beetles (Werren *et al.*, 1994) and parthenogenesis induction in parasitoids (Hagimori *et al.*, 2006; Giorgini *et al.*, 2010). However, equal likelihood of male and female midges being infected indicates sex ratio

distortion is unlikely to be a phenotype for the *Rickettsia* in midges. Further to this, *Rickettsia felis* represents an obligate symbiont in book lice (*Liposcelis bostrychophila*) required for egg production (Perotti *et al.*, 2006). However, the sporadic distribution of *Rickettsia* across diverse taxa suggests a lack of co-speciation making it unlikely that the host requires symbiont presence for its function.

Overall, the data suggests the Torix group *Rickettsia* identified in this study may have some facultative (ecologically contingent) benefit to their host. Indeed, *Rickettsia* from this clade have been linked with a fitness benefit (increased body size) in leeches (Kikuchi and Fukatsu 2005). Endosymbionts of haematophagous arthropods occasionally have the capacity to synthesise B vitamins to supplement their host's lacking diets. This has previously been recorded for *Wigglesworthia* symbionts in tsetse flies and *Wolbachia* in *Cimex* bedbugs (Snyder *et al.*, 2010; Nikoh *et al.*, 2014). However, with the exception of a reduced pathway for folate biosynthesis (also found in other *Rickettsia*), RiCNE lacks known pathways for the biosynthesis of cofactors and B-vitamins (Pilgrim *et al.*, 2017).

A peculiarity of note is the detection of coexisting infected and uninfected individuals in *C. impunctatus* populations, a scenario contrary to the more common fixed infections observed in this study. However, low titre infections in those testing negative for *Rickettsia* cannot be ruled out (chapter 2 and Mee *et al*. 2015). An alternative explanation for this difference is that the strain in *C. impunctatus* has a different role in its host in comparison to the other isolates at fixation in midges. *Rickettsia* surface antigens, like 17KDa, have previously been identified to be evolving under positive selection and may have key roles in host adherence and infiltration (Blanc *et al.*, 2005), although Ka/Ks ratios suggest this might not be the case here.

A major research effort for the future lies in identifying the impact of *Rickettsia* on host biology.

In conclusion, this study has identified a common but neglected association between *Rickettsia* and biting midges. Given the importance of biting midges as vectors, two key areas of future research are to establish the impact of *Rickettsia* presence on vector competence. Symbionts may reduce vector competence (as in *Wolbachia* in *Aedes aegypti*), increase it (as for *Rickettsia* in *Bemisia tabaci*) or have no impact. *Rickettsia* infections are also known to affect host dispersal tendencies, with Torix *Rickettsia*-infected spiders showing lower motivation for dispersal (Goodacre *et al.*, 2009). Symbiont impact on either of these characteristics would significantly alter the local and spatial spread of vector-borne infections, and thus pressingly deserve attention.

Chapter 4: The Tropism and Transtadial Transmission of a *Rickettsia* Endosymbiont in *Culicoides impunctatus*

4.1 Abstract

Previous work in this thesis has shown *Rickettsia* (group Limoniae) is widespread in midge populations with infection often reaching fixation. The maintenance of *Rickettsia* endosymbionts in insect populations is achieved through both vertical and horizontal transmission routes. For example, the presence of the symbiont in the follicle cells and salivary glands of *Bemisia* whiteflies allows *Rickettsia* transmission via the germline and plants respectively. Through fluorescence *in-situ* hybridisation (FISH) and transmission electron microscopy screening, this study describes the pattern of *Rickettsia* tissue tropisms observed in the Scottish highland midge, *Culicoides impunctatus*. Of note, is the infection of the ovarian suspensory ligament, indicating a novel germline targeting strategy. Additionally, localisation in ovaries and larval fat bodies suggests transtadial transmission is a major route of ensuring maintenance of *Rickettsia* within *C. impunctatus* populations. Aside from providing insights into transmission strategies, *Rickettsia* presence in the fat body of larvae indicates potential host fitness and vector effects to be investigated in the future.

4.2 Introduction

Endosymbiotic bacteria of arthropods primarily maintain themselves within a population through maternal transmission to offspring. In insects there are three distinct methods of ensuring such vertical transmission occurs (Russell, Chappell and Sullivan, 2019).

1) Soma-soma interactions; a rare occurrence in arthropods, although observed in the viviparous tsetse fly where developing embryos gain *Wigglesworthia* infections via milk gland secretions (Rio *et al.*, 2012).

2) Germline-germline transmission; where certain *Wolbachia* strains of *Drosophila melanogaster* localise in the germline continuously throughout development (Veneti *et al.*, 2004; Serbus and Sullivan, 2007; Toomey *et al.*, 2013).

3) Soma-germline transmission; the best-known example of which involves the deliverance of endosymbionts from specialised cells (bacteriocytes) to gametes or developing embryos.

The latter is the most commonly described and is hypothesised to be as a result of evolutionary histories (endosymbiotic events pre-date germline appearance) and selection pressures (restriction to somatic tissues prevents detrimental local effects during gamete development) (Russell, Chappell and Sullivan, 2019). Furthermore, the sequestering of endosymbionts by the soma offers a large selection pool for occupation by the germline (Mira and Moran, 2002; Ikuta *et al.*, 2016). Subsequently, the success of vertical propagation is often determined by both germline and somatic bacteria populations.

Although bacteriocytes primarily transmit microbes upon which their host is dependent for survival (primary endosymbionts), facultative (secondary) endosymbionts also utilise somatic

tissues to target the germline. However, due to differing evolutionary histories, varying routes of inheritance occur. For example, *Buchnera aphidicola*, a primary endosymbiont of the pea aphid which provides essential amino acids to its host, co-exists with the thermal and parasite protective secondary endosymbiont *Serratia symbiotica* (Douglas, 2002; Oliver *et al.*, 2003; Burke, Fiehn and Moran, 2010). Although *Buchnera* utilises the exocytosis of bacteriocytes, *Serratia* co-opts endocytic mechanisms via the haemolymph with subsequent segregation of both endosymbionts after the formation of embryonic bacteriocytes (Koga *et al.*, 2012).

Many endosymbionts exhibit evidence of both vertical and horizontal (mixed) transmission. This is exemplified in *D. melanogaster*, where microinjection of *Wolbachia* led to infection of the germline after localisation in the haemolymph and somatic stem cells (Frydman *et al.*, 2006). Thus, symbionts newly infecting somatic tissues can transfer into the germ line, and this provides a pathway for host shifts to occur. Furthermore, soma infections are often linked to host fitness effects. For example, close association with uric acid-containing cells (urocytes) by *Blattabacterium* in cockroaches is associated with nitrogen recycling into amino acids (Sacchi *et al.*, 1998; Sabree, Kambhampati and Moran, 2009), whilst the presence of *Wolbachia* in *Drosophila* muscles and nervous tissue is correlated with olfactory-cued locomotion (Peng *et al.*, 2008). Although these associations are well-characterised in some symbioses, many recently identified insect-bacteria associations are yet to be investigated. Thus, the localisation of symbionts to both somatic and germline tissues can offer insights into these intimate interactions.

Rickettsia are an intracellular group of alphaproteobacteria which demonstrate mixed modes of transmission. The horizontal and vertical transmission strategies of *Rickettsia* have allowed for the infection of a diverse range of organisms including protists, arthropods, plant and vertebrates (Weinert *et al.*, 2009; Zchori-Fein, 2011). Haematophagous spread of pathogenic *Rickettsia* to vertebrates has been well documented, and phloem-feeding insects have been implicated in transmission of *Rickettsia* to plants (Brumin, Levy and Ghanim, 2012; Caspi-Fluger *et al.*, 2012). *Rickettsia* also has the unusual ability to infect sperm head nuclei, allowing for paternal inheritance, which can combine with maternal transmission to drive a costly symbiont into the population (Watanabe *et al.*, 2014). Subsequently, the multiple transmission pathways of *Rickettsia* makes this symbiont a model organism for providing insights into transitions between mutualism and pathogenicity (Merhej and Raoult, 2011).

Previous work in this thesis (chapter 3) has demonstrated the presence of Limoniae *Rickettsia* in several species of *Culicoides* although host effects and transmission routes are yet to be determined. The absence of sex-ratio distortion and obligate mutualism (not all individuals were infected within a population) suggests the lack of a reproductive parasitism. By exclusion, these indicate the drive of this endosymbiont is related to a facultative benefit. Indeed, a Torix group *Rickettsia* in leeches is positively correlated with body size (Kikuchi and Fukatsu, 2005). Furthermore, as *Culicoides* transmit several veterinary viruses there is particular interest in the effects of *Rickettsia* on vectorial capacity. For example, *Rickettsia* infection of *Bemisia tabaci* has been shown to increase the transmission of tomato yellow leaf curl virus (Kliot *et al.*, 2014).

The biting midge *Culicoides impunctatus* is prevalent across Northern Europe but is most abundant in the Highlands of Scotland. Here they are a biting nuisance with "midge attacks" accountable for significant economic impact through losses in tourist and forestry industries (Hendry and Godwin, 1988; Hendry, 1996). The abundance of *C. impunctatus* in Scotland is likely due to a large number of acidic boggy breeding sites. Further, their ability to reproduce once in the absence of a blood meal (autogeny) means huge numbers can develop even where vertebrate hosts are not available (Boorman and Goodard, 1970; Blackwell *et al.*, 1992; Blackwell, Young and Mordue, 1994).

Despite their commonness in the field, attempts at laboratory-rearing *C. impunctatus* have proven largely unsuccessful (Hill, 1947; Carpenter, 2001). A main barrier to cultivation of this species is the lack of identification of a suitable diet. The two extant *Culicoides* species colonised (*C. sonorensis* and *C. nubeculosus*) rely on substrates, such as liver powder and wheat germ, which encourage microbial growth in larval pans (Boorman, 1974). However, previous attempts utilising this method in *C. impunctatus* have been unsuccessful (Carpenter, 2001). In addition, developmental stages are difficult to observe with these methods. Subsequently, the investigation of *Culicoides* bionomics often utilises agar dishes to gain basic developmental information which can inform future cultivation attempts (Barceló and Miranda, 2018). Additionally, this technique can allow for the investigation of life stage-specific traits.

In order to investigate potential phenotypic effects on the host as well as symbiont transmission strategies, both the tissue and subcellular tropisms of *Rickettsia* must be

elucidated. Subsequently, this study assesses both germline and somatic tissues of *C. impunctatus* for *Rickettsia* infection via transmission electron microscopy (TEM) and fluorescence *in-situ* hybridisation (FISH).

4.3 Methods

4.3.1 The collection of *Culicoides impunctatus*

Collections of *Culicoides impunctatus* for imaging of adult tissues were carried out at multiple sites in the United Kingdom; Dumfries, Kielder forest, Loch Lomond and Fort William between June 2017 and September 2018. For midges not used for ovipositing, individual *Culicoides* were allowed to rest on the author's arm before aspiration into 1.5 ml Eppendorf tubes with 10% sucrose-soaked cotton wool placed at the bottom and damp sphagnum moss filled in the lid to maintain a high humidity. Tubes were then placed horizontally on sticky tape in Tupperware boxes before incubation at 23 °C. For ovipositing *Culicoides*, female midges were collected by allowing feeding on the author and field assistants' forearms from the Loch Lomond site (May 2018). When midges were observed to be replete (approximately after 5 minutes), or the insect released mouth parts from the skin, they were aspirated and stored in a plastic container at ambient temperature (13-17°C) until the end of the collection session.

Ovipositing containers were assembled based on previous studies undertaken by Carpenter (2001). Briefly, approximately 50 *Culicoides* were transferred to containers consisting of cylindrical pill boxes (Watkins and Doncaster, UK; 64 mm diameter x 60mm depth), with cotton wool soaked in 10% w/v sucrose solution (replaced every two days) placed on top of a fine net meshing which covered the tops of the pillboxes. For ovipositing areas, 50 ml Falcon tube lids were filled with damp sphagnum moss (B and Q, UK) by soaking in 1% nipagin dissolved in distilled water and squeezing until drops could be counted, before being placed on top of damp filter paper. This was secured by cutting a cylindrical hole in the bottom of the pill box. These were then transferred to 19 x 12 x 8 cm Tupperware boxes containing a

50 ml beaker of saturated sodium sulphate (Na_2SO_4) which has previously been described to maintain humidity at >90% when between 20-25°C (Winston and Bates, 1960). These were then transported to a laboratory in Liverpool before the Tupperware boxes were placed in an incubator where they were maintained at 23°C with a photoperiod of 12 Light: 12 Dark hours.

4.3.2 Larval rearing

Due to *C. impunctatus'* small size and vulnerability to drowning on surface films under laboratory conditions it is unlikely that this species oviposit in open water (Carpenter, 2001). This alongside the observation by Carpenter that L1 instars of *C. impunctatus* get trapped in agar solutions of >1% led to the rearing on 0.5% agar; petri dishes used were 100 mm in diameter and x 15 mm in depth with agar made up from deionised water poured to a depth of 5 mm. Eggs oviposited onto the sphagnum moss substrate were picked individually with a fine paintbrush or the damp edge of a sharpened tungsten needle and placed onto agar dishes (100-180 eggs per dish; n=3 dishes) and spaced evenly apart. *Culicoides impunctatus* identification was confirmed by the distinctive brown heads of larvae (Kettle and Lawson, 1952) as well as identification of ovipositing adults in pill boxes. Two larval diets were briefly trialled for suitability; Banana worms (*Panagrellus nepenthicola*), and the ciliate, *Volvox carteri.* Subsequently, Banana worms were chosen because larvae were seen to actively attack the worms once introduced, whereas the surface tension from *Volvox* prevented larvae from moving freely in the agar. Banana worms were cultured using the manufacturer's instructions (Ron's worms; Amazon UK) before a fine paintbrush was used to place the nematodes in 2-3 ml of deionised water and spread evenly across each agar dish. This was performed daily, providing nutrition and ensuring the agar did not desiccate. Larvae were then stored under the same temperature and photoperiods as adults previously mentioned.

Allocation of larval instars were designated by head capsule length measurements as described by Kettle and Lawson (1952). A Kaplan-Meier estimator was used to measure cumulative survival over time as well as larval periods. Survival curve censoring of individuals was undertaken for larvae used for imaging assays, which were retrieved from the agar dishes using a coiled piece of tungsten wire.

4.3.3. Dissections of adults

Post blood-feeding in the wild, *Culicoides* were processed after carefully timed transportation to a laboratory in Liverpool. Individuals were sacrificed at various time points post blood feeding (PBF); 0 hours (non-blood fed), 12 hours PBF, 48 hours PBF and 120 hours PBF. First, *Culicoides* were chilled in the freezer for immobilisation before being placed in a drop of Phosphate-buffered saline on a petri dish. Midges were then killed by piercing the thorax with a sharp tungsten needle before being confirmed as *C. impunctatus*. Ovaries were exposed in various ways through dissection under a stereoscopic microscope:

1) By a midline incision of the ventral abdomen

2) By shelling away cuticle from abdominal segment 8 and repeating proximally

3) By quick evisceration through holding the dorsal thorax and pulling the posterior abdomen with forceps. Any other tissues which could also be recovered were then stored separately for further investigation of somatic tropism.

Time points for sacrifice were chosen as a result of a study by Carpenter (2001) which identified developmental stages of forming eggs as a function of time after blood-feeding. Stages were confirmed by using a system developed by Linley (1965) which was subsequently modified by Campbell and Kettle (1975):

stage 1) No observation of yolk within the oocyte;

stage 2) Yolk can be identified within the oocyte;

stage 3) Yolk proteins occupy up to three-quarters of the oocyte;

stage 4) The oocyte is elongated and no longer oval, resembling the mature egg;

stage 5) Egg fully mature, with chorion visible.

4.3.4 Tissue preparation and fluorescence *in-situ* hybridisation (FISH)

Tissue samples were initially stored in Carnoy's solution (chloroform:ethanol:glacial acetic acid, 6:3:1) overnight in 100 µl wells covered in Parafilm, but were difficult to retrieve due to their size and tendency to stick to the sides of wells. Subsequently, preparation was changed to direct fixing onto poly-L-lysine covered slides which allowed for adherence of the tissues on the slide without problems of disruption when undergoing subsequent washes. Tissues examined included eggs of different developmental stage, Malpighian tubules, midgut, foregut, hindgut, fat body and salivary glands. Additionally, crushed spermathecae were prepared for visualisation of the spermatophore and spermatids contained within. This was achieved by separating paired spermathecae from ovaries with diligent attempts to prevent any contamination from any ovaries that might be *Rickettsia*-infected. Spermathecae were then suspended in Phosphate-buffered saline and allowed to dry before pressing a coverslip over the slide to break open the tissue.

For fluorescence *in-situ* hybridisation (FISH) imaging, the above tissues were fixed for one hour in Carnoy's solution and tissues cleared by treating with 6% H_2O_2 in ethanol for 2 hours.

Two pre-hybridisation washes were undertaken using wash buffer (20 mM Tris-HCl, pH 8.0, 50mM NaCl, 0.01% sodium dodecyl sulphate, 5 mM EDTA). Hybridisation was performed overnight in hybridisation buffer (20 mM Tris-HCl, pH 8.0, 90mM NaCl, 0.01% sodium dodecyl sulphate, 30% formamide) containing 10 pmol/ml of the *Rickettsia* specific probe [5'-CCATCATCCCCTACTACA-(ATTO 633)-3'] adapted from Perotti *et al*. (2006) which were checked for specificity against the Torix 16S gene of *Culicoides newsteadi* (RiCNE; Bioproject accession number PRJNA376033). After hybridisation, the samples were thoroughly washed twice in wash buffer and slide mounted in Vectashield with DAPI (Vector Laboratories) and viewed under a Zeiss LSM 880 BioAFM confocal microscope. The specificity of the detection and any autofluorescent properties of midge tissue was assessed using *Rickettsia*-free midges (*Culicoides nubeculosus*; Pirbright Institute) as negative controls. Further negative controls were used when fresh *C. nubeculosus* were not available by using probe-free samples. For each tissue, at least 5 specimens were viewed under the microscope to confirm reproducibility. Optical sections (0.7µm thick) were prepared from each specimen to create a Z-stack image to be processed in ImageJ. All FISH imaging equipment and technical assistance was provided by the Liverpool Centre for Cell Imaging (University of Liverpool).

4.3.5. Transmission electron microscopy

As host-seeking *Culicoides* mate prior to blood feeding it was possible to examine spermatids from female spermathecae. Ovaries and spermathecae were prepared for transmission electron microscopy (TEM) as follows. Tissues were dissected into 2% (w/v) paraformaldehyde + 2.5% (w/v) glutaraldehyde in 0.1M phosphate buffer (pH 7.4). Fixative was then changed for 2.5% (w/v) glutaraldehyde in 0.1M phosphate buffer (pH 7.4). Heavy metal staining consisted of 2% (w/v) OsO4 in ddH_2O, followed by 1% (w/v) Tannic acid in

ddH_2O and then 1% (w/v) aqueous uranyl acetate. To prevent precipitation artefacts the tissue was washed copiously with ddH_2O between each staining step. Fixation and staining steps were performed in a Pelco Biowave®Pro (Ted Pella Inc.Redding California, USA) at 100W 20Hg, for 3 mins and 1min respectively. Dehydration was in a graded ethanol series before filtration and embedding in medium premix resin (TAAB, Reading, UK). For TEM, 70-74 nm serial sections were cut using a UC6 ultra microtome (Leica Microsystems, Wetzlar, Germany) and collected on Formvar (0.25% (w/v) in chloroform, TAAB, Reading, UK) coated Gilder 200 mesh copper grids (GG017/C, TAAB, Reading, UK). Images were acquired on a 120 kV Tecnai G2 Spirit BioTWIN (FEI, Hillsboro, Oregon, USA) using a MegaView III camera and analySIS software (Olympus, Germany). Dissections and initial fixation were performed by the author with subsequent staining and imaging undertaken with assistance from Alison Beckett and the Electron microscopy unit (Institute of Translational Medicine; University of Liverpool).

4.4 Results

4.4.1 *Rickettsia* infection during oogenesis

Underdeveloped (stage 1) eggs contained clusters of *Rickettsia* predominantly in oocytes with no presence of yolk deposition seen (Figures 4.1. and 4.2.). Signal was strongest in the oocyte with bacteria also being observed within nurse cells. Electron microscopy images suggest the *Rickettsia* are perinuclear rather than within the nurse cell nuclei themselves (Figure 4.1.). The follicular epithelium is also infected with bacteria seen in transit between follicle cells and the oocyte. No signal was observed in *Rickettsia*-free *Culicoides nubeculosus* or probe-free negative controls. After 48 hours post-blood feeding, yolk deposition is seen as a clouding in the (stage 2) oocyte although individual yolk granules are not able to be seen. Clusters of bacteria no longer occupy nurse cells but still predominantly fill the oocyte (Figure 4.2.). In stage 4 eggs, yolk granules become visible and appear to harbour sparse numbers of bacteria with a predominant localisation in follicle cells (Figure 4.2.). However, in certain focal planes *Rickettsia* are seen to remain in the oocyte cytoplasm but presents only at the periphery of the oocyte.

4.4.2 Germline targeting of *Rickettsia* via the ovarian suspensory ligament

During dissections of adult female *C. impunctatus* it was often possible to retrieve most of the gut, Malphigian tubules and paired ovaries. This allowed for comparing localisation patterns in the somatic tissues as well as the germline. Initial examination of adult specimens at low magnification gave a consistent pattern of strong localised signal at the anterior/posterior

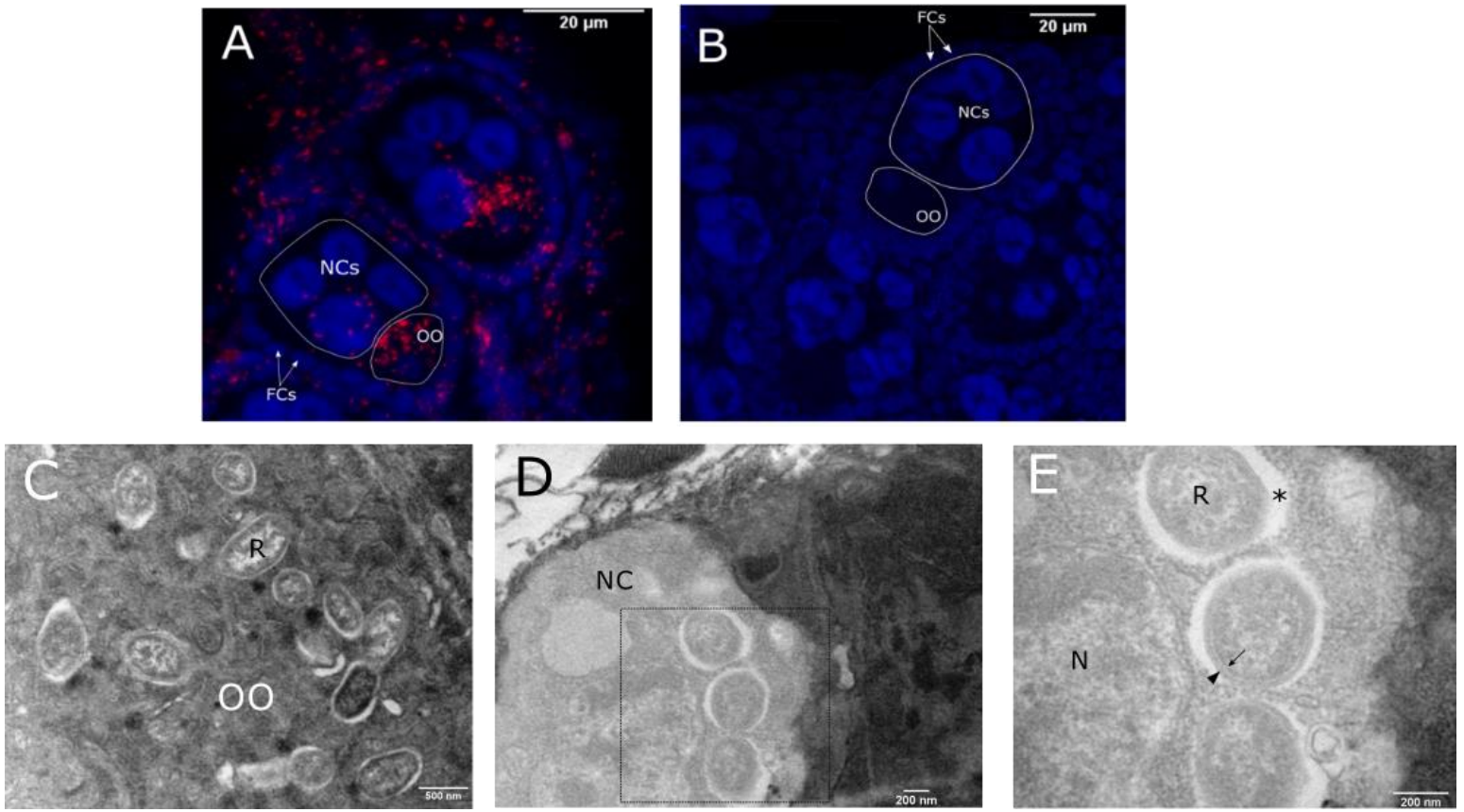


Figure 4.1. FISH and TEM images of stage 1 *C. impunctatus* and *C. nubeculosus* eggs. *Rickettsia*-specific probe = red; DAPI-staining = blue. **A)** FISH image of an infected *C. impunctatus* stage 1 egg. **B)** FISH image of an uninfected (negative control) *C. nubeculosus* egg. OO = oocyte; NCs = nurse cells; FCs = follicle cells. **C)** TEM section of a *C. impunctatus* egg demonstrating clusters of *Rickettsia* in the oocyte cytoplasm (OO). **D)** TEM section of a *C. impunctatus* egg demonstrating *Rickettsia* presence in a nurse cell (NC). **E)** Magnified details of the box in D demonstrating perinuclear *Rickettsia* (R). The *Rickettsiae* have a distinctive cell wall (arrowhead) and cell membrane (arrow) separated by a periplasmic space. * = radiolucent halo/slime layer; N= nucleus.

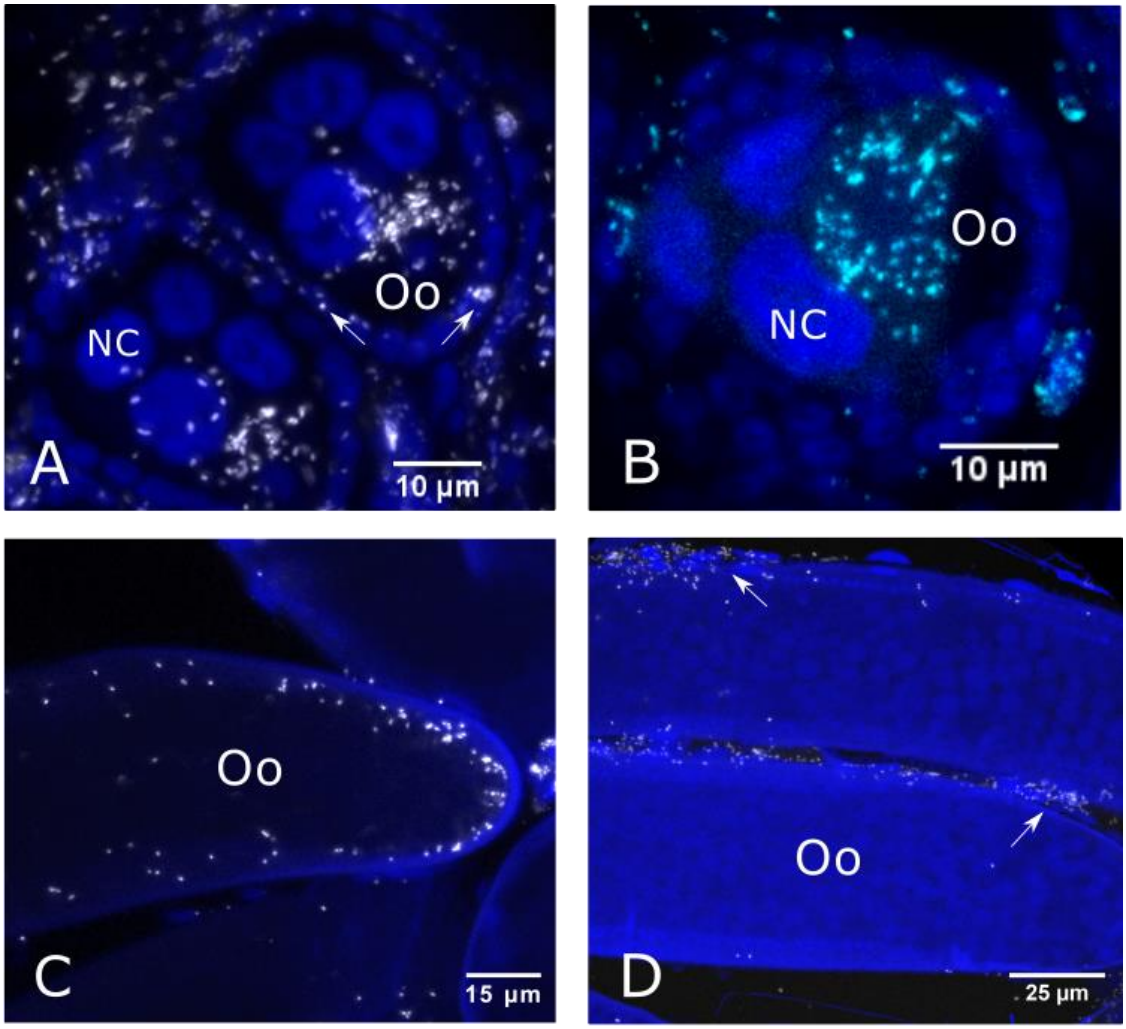


Figure 4.2. FISH images of *C. impunctatus* egg chambers at different developmental stages of oogenesis. *Rickettsia*-specific probe = white; DAPI-staining = blue. **A)** *Rickettsia* infection of stage 1 eggs (0 hours post blood-feeding) with predominant localisation in the Oocyte (Oo), Nurse cells (NC) and follicle cells (arrows). **B)** *Rickettsia* infection of stage 2 eggs (12 hours post blood-feeding) with the accumulation of cloudy yolk deposits in the oocyte (Oo). Infection is still primarily observed in the Oocyte (Oo) although infection of Nurse cells (NC) is now absent. **C)** A focal plane of stage 4 eggs (120 hours post blood-feeding) showing localisation at the periphery of the oocyte (Oo). **D)** A focal plane of stage 4 eggs (120 hours post blood-feeding) showing infection of follicle cells (arrows).

midgut junction but not throughout the majority of the alimentary tract (Figure 4.3.). Additionally, on examination at a higher magnification, it was difficult to place the signal in any of the midgut epithelial cells. Examination of further individuals led to the discovery that both the midgut junction and ovaries were connected by a fibrous piece of tissue which was to be later identified as the ovarian suspensory ligament. This structure, otherwise known as the "median ligament", was seen to pair off and loop down from the midgut attachment site before attaching at the apex of the ovary.

Infection of the suspensory ligament apparatus was strongest at the anterior-posterior midgut junction (Figure 4.3.) although this could be attributed to connective tissue density. It was possible to follow the signal down the suspensory ligament where the structure became continuous with the terminal filaments of ovarioles and the ovarian epithelial sheath; the structure separating ovarioles from one another. Strength of infection was consistent in the ovarian epithelial sheath over the long axis of the ovary with neighbouring immature egg chambers seen to be heavily infected. Bacteria could be seen migrating from this densely populated structure into neighbouring follicle cells and further into oocytes themselves, which again, had the distinct pattern of infection both in undifferentiated oocyte cytoplasm as well as nurse cells. Within the ovary, the germarium appeared to be no more infected than the rest of the ovarian tissue. In one individual, the attachment of the suspensory ligament to a sparsely cellular but strongly signalled structure is thought to be part of a lobe of the fat body (Figure 4.3.) although the ethanol based (Carnoy's) fixative diminishes lipids leading to ambiguity of identification when observed under transmission light. Signal was observed in

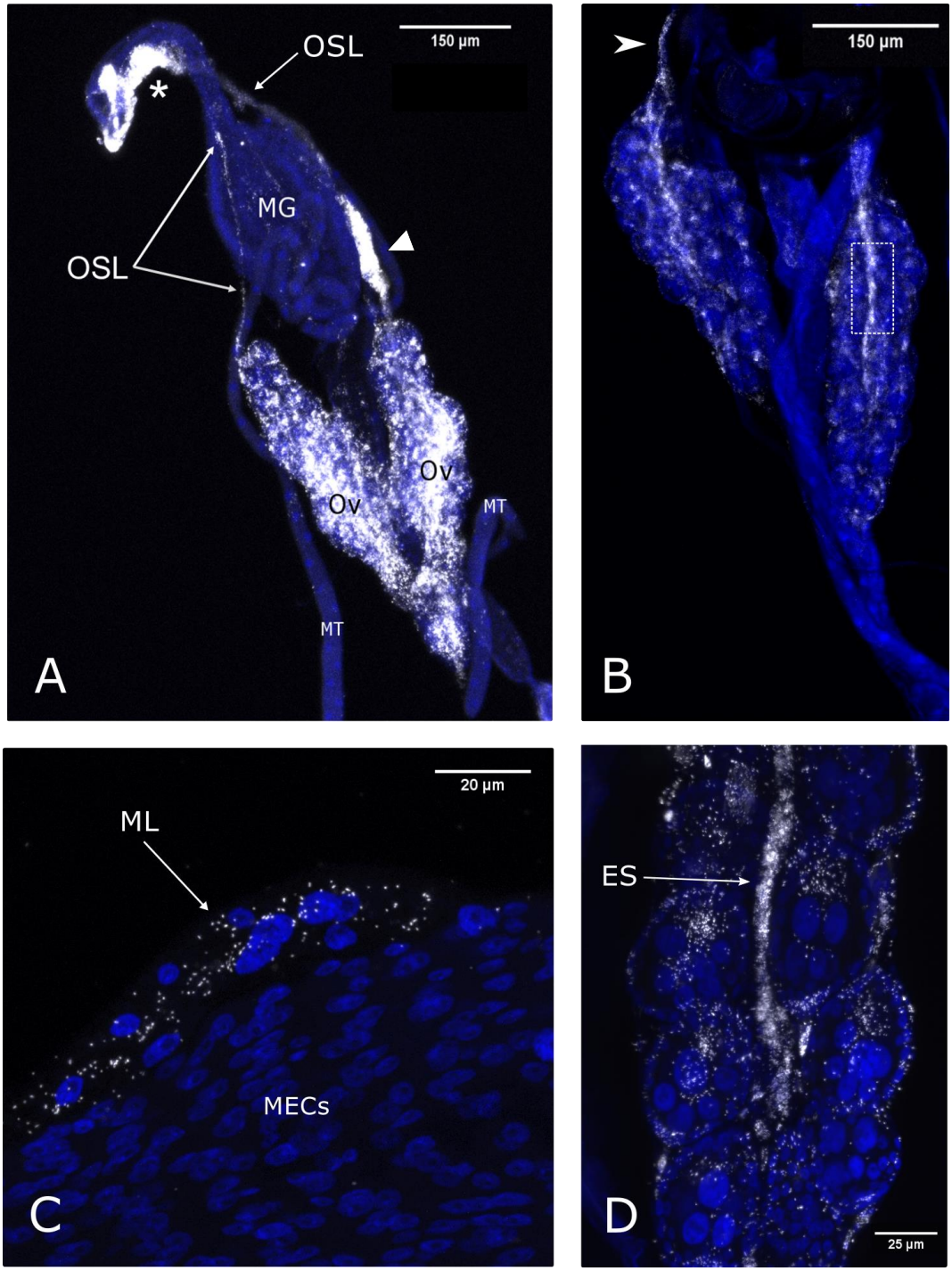


Figure 4.3. *Rickettsia* localisation in *C. impunctatus* adult connective tissues associated with both the midgut and ovaries via FISH imaging. *Rickettsia*-specific probe = white; DAPI-staining = blue. **A)** Strong *Rickettsia* signals identified at the anterior-posterior midgut junction (*) as well as the paired ovaries (Ov). These two areas are connected via the ovarian suspensory ligament (OSL) which runs from the midgut junction to the apex of the ovary. White triangle=putative fat body lobe; MG=midgut; MT=Malphigian tubules. **B)** A focal plane of the paired ovaries demonstrating the continuation of the suspensory ligament attachment site at the ovary apex (arrowhead) into the ovary. **C)** *Rickettsia* localisation at the median ligament (ML); the fusion of the ovarian suspensory ligaments at the attachment site at the anterior-midgut junction. Lack of infection is observed in midgut epithelial cells (MECs). **D)** Higher magnification of the box in B. The continuation of the ovarian suspensory ligament with the ovarian epithelial sheath (ES) allows for the delivery of *Rickettsia* into neighbouring egg chambers.

the crop of the foregut in one sample although it is possible that this is part of the connective tissue matrix continuous with the median ligament. Fragility of tissues when dissecting often impaired the recovery of the completely intact connective tissue; some samples only showed signal at the attachment sites of the ovarian ligament suggesting the snapping of the structure during processing. No signal was observed in Malpighian tubules, heads, salivary glands or other alimentary structures except for the 1 individual which demonstrated infection in the crop.

4.4.3 FISH and transmission electron microscopy of spermathecae

Rickettsia was further detected in crushed spermathecae from fertilised females (Figure 4.4). DAPI staining gave an unspecific signal making it difficult to distinguish spermatids from the rest of the acellular spermatophore. However, *Rickettsia* signal was detected diffusely in samples. This observation suggested possible male horizontal transmission (to mates) and/or vertical transmission (to offspring). Alternatively, this signal could reflect presence in maternally-derived spermathecal structures. Subsequently, TEM sections of spermathecae were assessed to clarify the nature of this signal. Overall, infected spermathecae showed no sign of bacteria in sperm heads or tails of spermatids, nor in the acellular matrix (Figure 4.4.). However, *Rickettsia* was identified in the maternally derived spermathecal epithelium, occasionally associated with actin-like filaments. Unfortunately, due to the difficulties of lab maintenance of *Culicoides*, a crossing-system to definitively rule out paternal transmission was not possible.

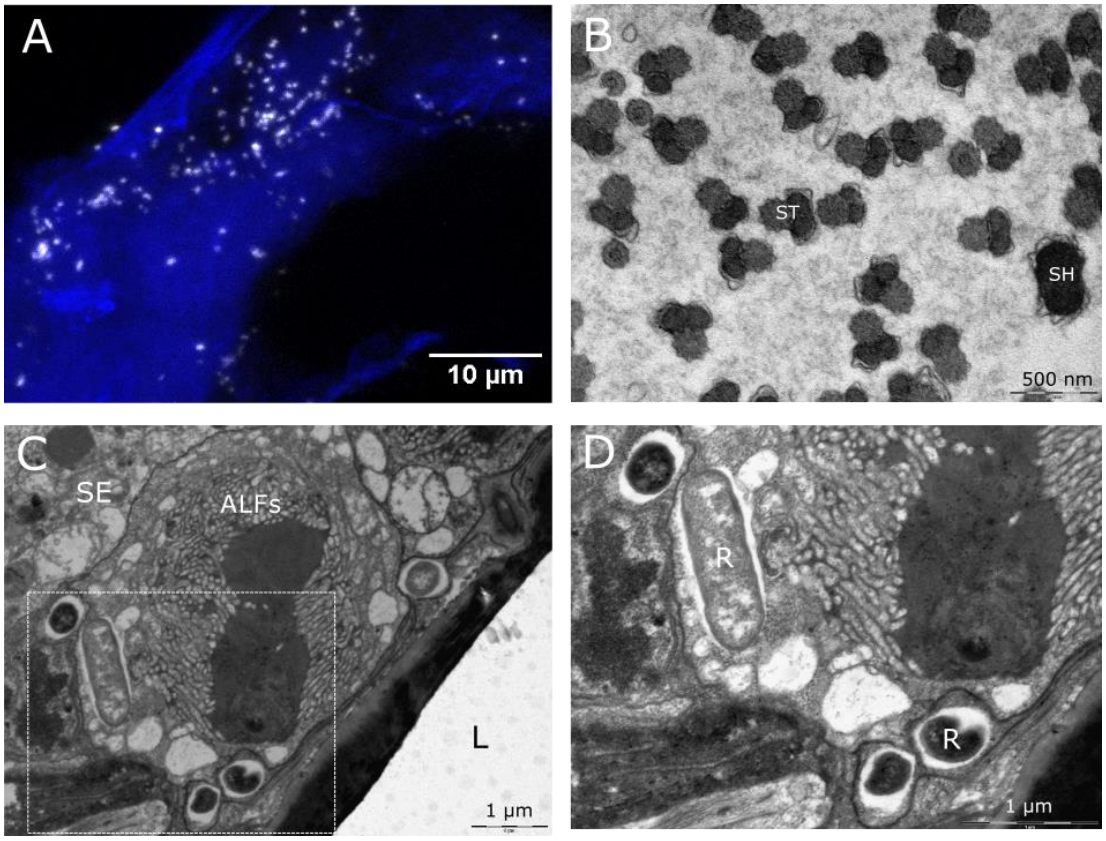


Figure 4.4. FISH and TEM analysis of *C. impunctatus* spermathecae. **A)** FISH image of *Rickettsia* presence in a crushed spermatheca. *Rickettsia*-specific probe = white; DAPI-staining = blue. **B)** TEM section of *Rickettsia* uninfected sperm heads (SH) and sperm tails (ST). **C)** TEM section of junction between the spermathecal lumen (L) and spermathecal epithelium (SE). ALFs=Actin-like filaments. **D)** Higher magnification details of the box in C, demonstrating longitudinal and cross-sectioned *Rickettsia* (R) residing in the spermathecal epithelium.

4.4.4 Subcellular location and associations

Transmission electron microscopy of spermathecae and immature eggs revealed coccobacilli *Rickettsia* free in the cytoplasm of oocytes and spermathecal epithelial cells (Figures 4.1. and 4.4.). Sections of bacteria ranged up to 1.35 µm in length and were seen together either in clumps, likely a result of recent division, or diffusely in tissues. The ultrastructure of *Rickettsia* demonstrated distinctive characteristics typical of the genus; a slime layer/radiolucent halo, and an outer trilaminar cell wall followed internally by a periplasmic space and membrane (Figure 4.1.).

4.4.5 The rearing of *Culicoides impunctatus*

183 larvae were monitored for survival rates and instar development (Figure 4.5.). All larvae were positively identified as *Culicoides impunctatus* and observed predating *Panagrellus* nematodes. However, L1 instars which were a similar size to the worms (<1mm), were frequently seen attacking conspecifics. Often, several individuals were seen in a ball-like structure with mouth parts attached to one another. After a couple of days, several degenerative remains of larvae were observed indicating many did not survive this cannibalism. Indeed, over 30% (95% CI intervals: 24-38%) of the initial starting population died within the first week (Figure 4.5.). After the first appearance of L2s (Day 15) no further cannibalistic behaviour was observed. Development into L3s was first observed at day 27. However, the burrowing behaviour of mature larvae made for difficult retrieval and head measurements; meaning a formal identification of the first appearance of L4 instars was not achieved. Between days 57 and day 67, 5 prepupae were transferred from the surface of the agar onto damp filter paper in pillboxes used to previously sustain gravid adults (Figure 4.6.). Although movement of pupae was observed several days after transfer, only 1 individual

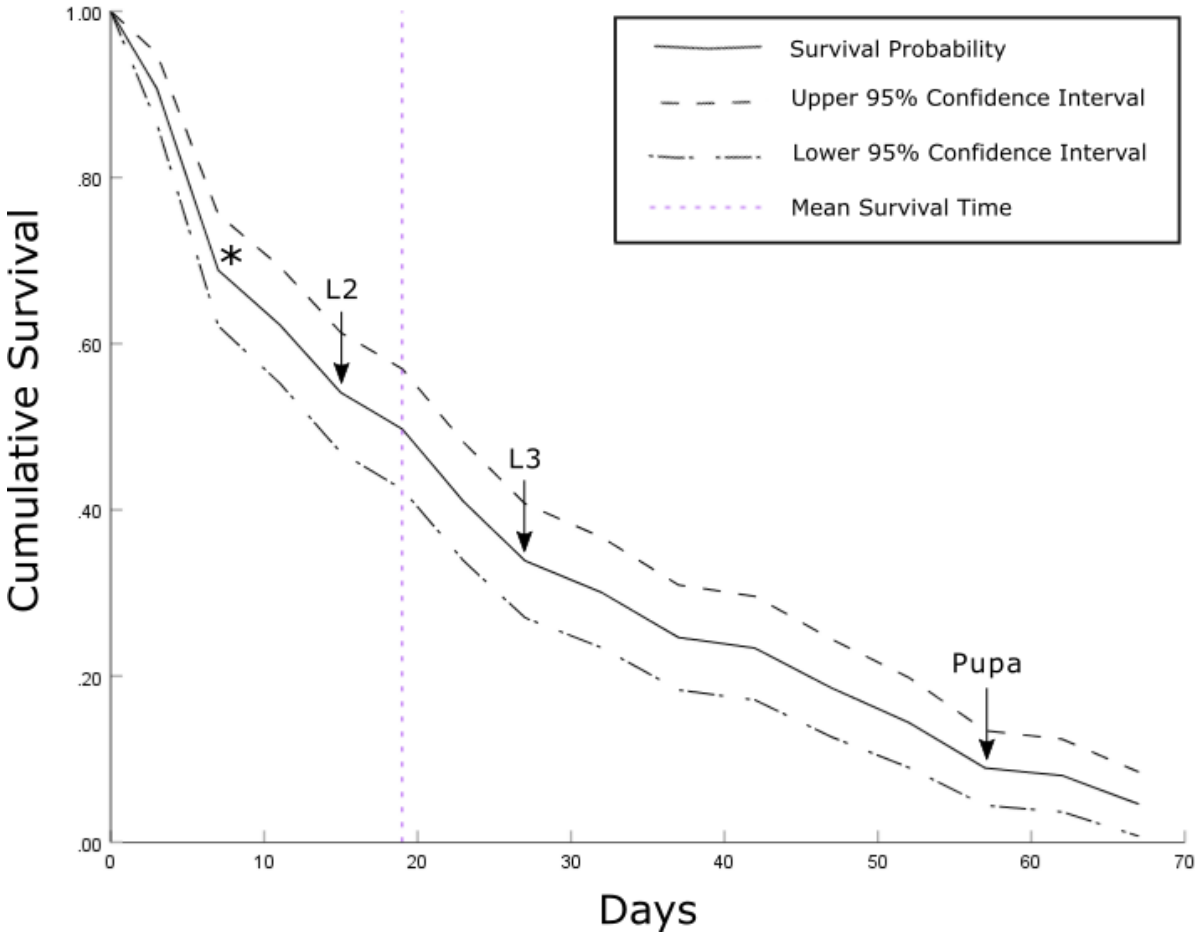


Figure 4.5. Kaplan-Meier survival curve monitoring cumulative survival of *C. impunctatus* larvae (n=183) over time. Arrows demonstrate the first appearance of different instars. The asterisk is the point at which cannibalism ceased to be observed in L1s.

melanised with none eventually emerging as adults. Development into pupae took a minimum of 56 days and the mean survival time of larvae was 19 days.

4.4.6. Larval tissue localisation

Due to difficulties in the manipulation and dissection of such small larvae, whole insect FISH assays were performed. Out of 10 L3 larvae assessed of unknown infection status, 4 showed a positive signal in the terminal abdominal (anal) segment fat body (Figure 4.7.) with 3 of these also demonstrating infection in the heads. One individual of the 4 positives showed sporadic multifocal infections across the rest of the length of the body. Focal "Burst" patterns of signal in the terminal abdominal segment suggest infections of globe-like structures such as cells or lipid droplets of the fat body. Of the infections in the head, although the exact tissue could not definitively be deduced, *Rickettsia* was detected closely associated with the head body wall (within 0-3 µm proximity of the autofluorescent cuticle of each focal plane). Concurrent presence in the large fat bodied anal segment (Figure 4.6.) of the larvae alongside the fat body's frequent attachment to the body wall suggests the pericerebral fat body is the most likely tissue affected, although further work is needed to confirm this. DAPI staining was poor in the heads of larvae making it difficult to ascertain the positions of nuclei. However, infected regions often surrounded circular areas (3-5 µm in length) lacking infection which are potentially uninfected nuclei (Figure 4.7). Instead of imaging pupae, attempts were made to confirm transgenerational transmission by monitoring and screening newly emerged adults. Unfortunately, none of the 5 reaching pupation emerged as adults.

Figure 4.6. Transmitted light microscope images of different life stages of *C. impunctatus.* **A)** Stage 4 eggs. **B**) L3 larva with lateral fat bodies terminating in the terminal abdominal segment (Box); * = *Panagrellus nephenticola* nematode. **C)** Pupa.

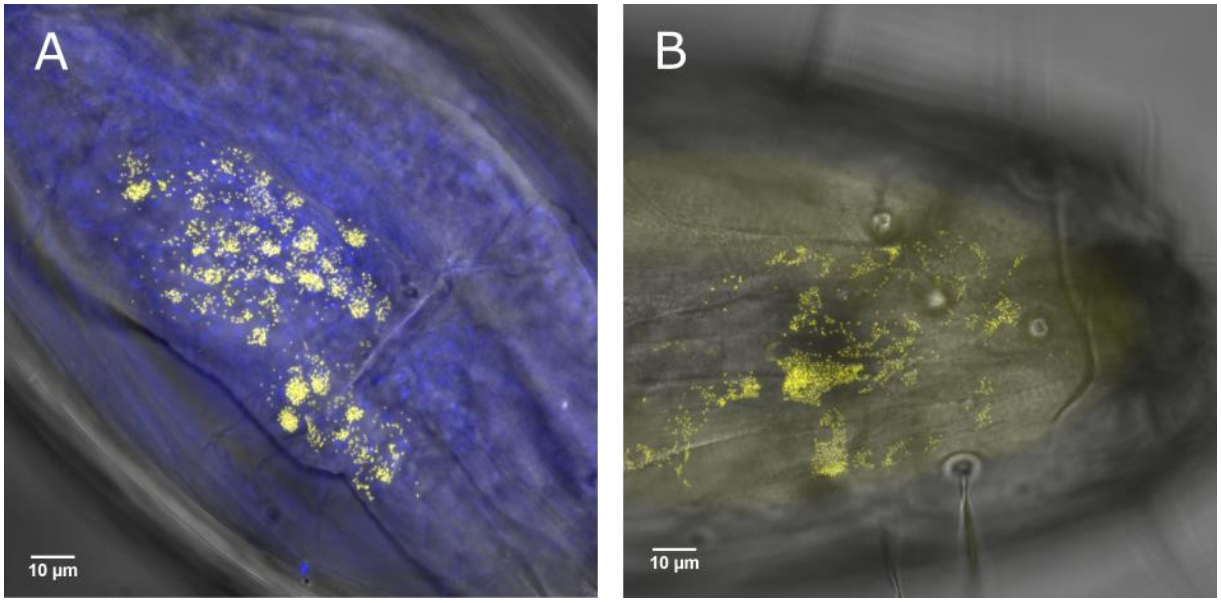


Figure 4.7. FISH imaging analysis of an L3 *C. impunctatus* larva. *Rickettsia*-specific probe = yellow; DAPI-staining = blue. **A)** Fat body *Rickettsia* infection of the terminal abdominal segment. **B)** *Rickettsia* larval head infection.

4.5 Discussion

The transmission routes of *Rickettsia* endosymbionts and their effects on biting midges are unknown but can be informed by symbiont tissue localisation. A major finding of this study is the *Rickettsia* infection of the ovarian suspensory ligament indicating a novel means of endosymbiont germline targeting (Figures 4.3.). The ovarian suspensory ligament of insects is a connective tissue which fuses the terminal filaments of ovarioles together at the anterior apex of the ovary. Commonly, the ligaments of each ovary join to form what is known as the "median ligament" which can attach to several tissues including the dorsal diaphragm, body wall, heart or fat body (Storto, 1994; Szklarzewicz *et al.*, 2007). In the case of *Culicoides impunctatus*, my findings suggest the median ligament fuses at the anterior posterior midgut junction, although it is possible that the ligament is part of a more extensive matrix with several attachment points: the fragile and thin nature of the ovarian suspensory ligament make it likely that this can be excluded in dissected or sectioned whole insects.

Endosymbionts which undergo mixed modes of transmission, such as *Rickettsia*, require mechanisms of cell-to-cell transfer to ensure migration from the soma to the germline (Allen *et al.*, 2007; Landmann *et al.*, 2010; Stoll *et al.*, 2010). The ligament itself is thought to consist of muscle and fibrous tissue, with our data showing the presence of both small and large nuclei (Figure 4.3.) consistent with Curtin and Jones' (1961) description of the *Aedes aegypti* suspensory ligament. As *Rickettsia* have been noted to utilise actin for movement within and between cells, this offers a possible mechanism for how the bacteria are able to pass through this structure (Teysseire, Chiche-Portiche and Raoult, 1992; Heinzen *et al.*, 1993; Gouin *et al.*, 1999; Reed *et al.*, 2014).

Aside from host-derived bacteriocytes transferring primary endosymbionts to ovaries, facultative bacteria possess innate qualities allowing them to target the germline (Russell and Moran, 2005; Frydman *et al.*, 2006; Weiss *et al.*, 2006; Nakayama *et al.*, 2015). Thus, determining these mechanisms of germline infection is of interest as it allows for insights into the initiation of novel insect-endosymbiont interactions. The continuation of the suspensory ligament with the ovarian epithelial sheath (Figure 4.3.) has previously been described in mosquitoes (Clements, 1992) and suggests an efficient route for ensuring transport of *Rickettsia* to egg chambers via follicle cells. The passage of *Rickettsia* through the follicle cells of ovarioles has been previously described in the ladybird *Adalia bipunctata* and the whitefly *Bemisia tabaci* (Sokolova, Zinkevich and Zakharov, 2002; Brumin, Levy and Ghanim, 2012). In these instances, *Rickettsia* is thought to migrate through follicle cells due to contact with infected haemocytes (Chen, Campbell and Purcell, 1996; Hurst, Walker and Majerus, 1996). Other studies have suggested bacteriocytes could be a means of transovarial transmission for *Rickettsia* (Gottlieb *et al.*, 2006, 2008), although further investigation suggested this was not a primary route of ovary-targeting in all cases (Brumin, Levy and Ghanim, 2012). Subsequently, infection of connective tissue directly linking the germline, offers a novel alternative route for the infection of developing eggs.

Multiple means of germline targeting have been previously observed in *Wolbachia*, where the endosymbiont utilises different germ line and somatic tissues depending on strain and host. In most Drosophila species, infections tend to occur in both somatic stem cell niches and germline stem cells, with both niches allowing for germline entry (Miller and Riegler, 2006; Toomey *et al.*, 2013). However, a select few *Wolbachia* strains appear to exhibit strict tropism

to somatic stem cells. The utilisation of somatic tissues as an indirect route of *Wolbachia* germline colonisation appears to also occur in filarial nematodes (Onchocercidae family) (Fischer *et al.*, 2011; Landmann *et al.*, 2012).

Our findings of *Rickettsia* infection in follicle cells but with limited infection of mature oocytes (Figure 4.2.), corroborates previous observations in the whitefly *Bemisia tabaci* (Brumin, Levy and Ghanim, 2012). In this case, heavy *Rickettsia* infection in immature oocytes, but not mature stages, was attributed to younger eggs being more permeable than their mature counterparts. The *Rickettsia* infection and close contact of the epithelial sheath to the follicle cells offers a suitable mechanism for ensuring persistent follicular infection (Figure 4.3.). Furthermore, as only a few bacteria cells are required to ensure subsequent infection of life stages (Watanabe *et al.*, 2014), remnant *Rickettsia* in the oocyte periphery of mature eggs (Figure 4.2.) appears to be sufficient for transtadial transmission (Figure 4.7.). The co-option of yolk granules by endosymbionts to gain entry into egg chambers via endocytosis (Herren *et al.*, 2013) can be dismissed in this case as *Rickettsia* are seen present in stage 1 oocytes before the accumulation of yolk (Figures 4.1. and 4.2.).

While maternal transmission is the primary transmission route for endosymbionts, horizontal transfer can lead to increased gene flow between co-habiting symbionts (Russell, Corbett-Detig and Cavanaugh, 2017), expand host-range (Stahlhut *et al.*, 2010; Morrow *et al.*, 2014) and negate the effects of detrimental mutations (Muller's ratchet) which often occur in intracellular bacteria (Moran, 1996). Horizontal transmission to vertebrate or invertebrate hosts, via haematophagy, is unlikely due to a lack of observed signal in salivary glands, and

the parity of prevalence in male and female midges. However, *Rickettsia* have also been known to establish intranuclear infections allowing for paternal (horizontal) transmission via sperm heads (Watanabe *et al.*, 2014). Subsequently, this previous study led to the investigation of intranuclear tropisms in both male and female tissues. Bacteria initially appeared to be residing within nurse cell nuclei on examination of FISH assays. However, TEM sections demonstrated the presence of perinuclear *Rickettsia* only (Figure 4.1.). Similarly, despite signal detection in spermathecae, microbes were only observed in maternally-derived epithelia as opposed to sperm heads. The absence of an intranuclear tropism in eggs as well as sperm (Figures 4.1. and 4.4.) suggests this characteristic may not be representative of the entire genus or be specific to particular host species. Further work is needed to ascertain the role of intranuclear *Rickettsia* in the biology of its hosts.

Many factors influence the ability of arthropods to sustain symbiont infections within and between populations. These include fitness effects and the genetic background of the host, the length of association between the host and endosymbiont, and the capabilities of the bacteria to persist in novel host cells (Moran, McCutcheon and Nakabachi, 2008; Hughes and Rasgon, 2014; Chrostek *et al.*, 2017; Tedeschi and Bertaccini, 2019). Certainly, the bias towards infections of aquatic taxa (Weinert *et al.*, 2009; chapter 3) can explain partially the distribution of Torix *Rickettsia* although it is still unclear why this bias exists. However, the somatic infections observed in this study can offer insights into past initiations of novel host-bacteria interactions. For example, *D. melanogaster Wolbachia* microinjections of the soma surrounding the germline have been proposed to offer a site of stable replication and proliferation for the native *w*Mel before targeting oocytes (Frydman *et al.*, 2006). However,

it should be noted that this did not work for non-native strains of *Wolbachia*. The proximity of the OSL to the midgut (Figure 4.3.) suggests predation may be a viable route of horizontal transmission. However, there are harsh environmental conditions needed to initiate infections within the gut epithelium including high pH, reactive oxygen species and peristaltic movements (Vallet-Gely, Lemaitre and Boccard, 2008). Subsequently, infection via endoparasitism of the major fat body tissues of larvae (or pupae), can offer an alternative route due to the fat body's association with the ovarian suspensory ligament (Storto, 1994). Thus, such soma-germline infection dynamics could offer efficient strategies for both horizontal and vertical transmission. This pattern corroborates findings in chapter 3 and chapter 5 where the coalescing of *Rickettsia* strains in diverse host taxa suggests frequent host shifts.

Aside from predatory and parasitoid interactions, it would be of particular interest to see whether the various physiologies of the suspensory ligament in different hosts can predict rickettsial infection. For example, the robber fly *Asilius crabroniformis*' ovarian suspensory ligament does not form a median ligament, but attaches to the body wall via the outer peritoneal sheath of ovaries instead of the terminal filaments (Owsley, 1946). Similar physiological barriers of infection have been proposed in the *Spiroplasma* endosymbiont of *Drosophila melanogaster*, where the inability to host shift has been suggested to be due to the divergence of yolk proteins between hosts (Nakayama *et al.*, 2015).

Previous examples of endosymbiont fat body infections of *Culicoides* includes the observation of "Rickettsia-like" organisms by Wolbach and Hertig (1924) and "twinkling" symbionts under

polarised light by Lawson (1951). Most of the larval fat body comprises of two bands extending down the lateral abdomen terminating in large lobes of the terminal segment (Figure 4.6. and Lawson, 1951). *Rickettsia* appear to be one of few known endosymbionts to reside in the fat body; the others being *Blattabacterium* of cockroaches (Brooks, 1970) and *Wolbachia* in a variety of insects (Cheng *et al.*, 2000; Clark *et al.*, 2005; Hughes *et al.*, 2011). The presence of *Wolbachia* in the fat body, an important endocrine tissue, has been associated with effects on host glucose and glycogen metabolism via altered enzyme activity and insulin signalling (Ikeya *et al.*, 2009; Voronin *et al.*, 2016). Not only is the fat body metabolically active, allowing for the bacterial sequestering of metabolite precursors, but fat cells during larval and pupal development are refractive to degradation (Nelliot, Bond and Hoshizaki, 2006; Aguila *et al.*, 2007). Thus, this tissue offers a suitable niche in arthropods for stable division and proliferation of endosymbionts.

Aside from *Rickettsia* observed in the abdominal fat bodies of larvae, another heavily infected area was the head (Figure 4.7.). Although it is not clear exactly which tissues are affected by the *Rickettsia*, the pericerebral fat body and cerebral ganglia (brain) are two candidates for future consideration. Endosymbiont brain infections have been proposed to lead to behavioural modifications (Albertson *et al.*, 2013; Strunov *et al.*, 2017) and early death in *Drosophila* fruit flies (Min and Benzer, 1997). Additionally, the fat body surrounding the brain has been demonstrated to play a different physiological role to that in the abdomen. Indeed, unique insulin signalling pathways have been observed in the head fat bodies of *D. melanogaster* leading to increased longevity as a result of inhibited senescence (Hwangbo *et al.*, 2004).

The presence of *Rickettsia* in the fat body of *Culicoides*, a vector of several veterinary viruses, raises questions on their effects on vector competence. Bluetongue virus (BTV) and epizootic haemorrhagic disease (EHDV) of ruminants replicate in the fat body of midges (Fu *et al.*, 1999; Mills *et al.*, 2017b) before travelling to the salivary glands, suggesting direct interactions between *Rickettsia* and the virus could be occurring. For example, competition for lipids between bacteria and virus have been suggested to influence viral titres (Caragata *et al.*, 2013; Sinkins, 2013; Geoghegan *et al.*, 2017; Schultz *et al.*, 2017). Additionally, antimicrobial peptides are synthesised in the fat body (Arrese and Soulages, 2010), again suggesting *Rickettsia* effects on vectorial capacity warrants further investigation. Furthermore, *Rickettsia* localisation to ovaries can indicate a possible role in autogeny (the development of eggs without a blood meal) in *C. impunctatus* (Boorman and Goddard, 1970). Indeed, *Rickettsia felis* is necessary for egg development in the booklouse *Liposcelis bostrychophila*. As autogeny is responsible for the pest burden of *C. impunctatus*, this could offer a target for population suppression in the future.

Although several *Culicoides* species have previously been reared on a *Panagrellus* nematode diet (Linley, 1979, 1985b; Mullens and Velten, 1994; Barceló and Miranda, 2018), the results in this chapter represent the first documented case in *C. impunctatus*, and provides a first step to making this a tractable system for investigating the effect of the symbiont on the host. Barceló *et al*. (2018) used such a diet recently with successful rearing and emergence of 5/6 species within a month (mean time 28.2 days). However, in the current study, development rates were comparatively slow, with the first observations of pupa at day 56 (Figure 4.5.). Furthermore, only 5/183 larvae pupated with none subsequently emerging. The emergence

of *C. impunctatus* in a natural soil habitat at 16-19°C by Hill (1947) took up to 5 months but with no information of developmental stages; replicate breeding pots at 23-24°C showed no emergence. Subsequently, the temperature used in this study may have been inappropriate for maintaining high levels of survival in *C. impunctatus*. However, thermal environment may not completely account for difficulties in rearing *C. impunctatus*, with insufficient aeration of culturing media and inappropriate di*et* also being proposed (Carpenter, 2001). With respect to the latter, although all instars were seen attacking nematodes, cannibalism was observed only in L1 instars (Figure 4.5.) indicating possible difficulties in predation of *Panagrellus* due to size. Alternatively, cannibalism is a natural behaviour of L1s which may be an inherent problem in rearing *C. impunctatus* (Hill, 1947). An interesting hypothesis to test in the future is whether cannibalism can facilitate the spread of *Rickettsia* in naïve populations. Initial infections of a few individuals could lead to this outcome if several larvae predate on a single (infected) conspecific as observed in this study. Furthermore, experimental evidence suggests the cannibalism of isopods can lead to the initiation of novel *Wolbachia* infections in naïve hosts (Le Clec'h *et al.*, 2013). Overall, the reasons for the slow development and poor survival observed here in comparison with previous attempts is unknown. Further optimisation of rearing was planned to be investigated further, but the unusually hot and dry summer of 2018 prevented sufficient numbers being collected.

In conclusion, this study has identified several somatic and germline infections of the endosymbiont *Rickettsia* in larval and adult life stages. Infection of the ovarian suspensory ligament, a continuation of the ovarian epithelial sheath, has identified a potential novel means of endosymbiont germline targeting. Further, somatic tissue infections of note include

the fat body of larvae and spermathecal epithelium of adult females which could have implications for host effects, as well as symbiont and arbovirus transmission dynamics.

Chapter 5: The Investigation of *Rickettsia* Host and Bacterial Diversity through DNA Barcoding

5.1. Abstract

The Barcode of Life Data System (BOLD) is a sequence database specifically designed for the curation of DNA barcodes, a widely used tool for species identification. Despite its utility, DNA barcoding can be confounded by the unintended amplification of endosymbionts. Previous studies have focussed primarily on the inadvertent amplification of *Wolbachia* due to its commonness in arthropods. However, closely related endosymbionts such as *Rickettsia* remain unexplored. Through the serendipitous discovery of *Rickettsia* amplicons in BOLD, this chapter describes the host and bacterial diversity revealed though a systematic review of >185,000 barcodes. *Rickettsia* is observed in approximately 0.4% of barcode submissions and is over twice as likely to be amplified as *Wolbachia* (0.165%). Through phylogenetic analysis of BOLD DNA extracts, the Torix group of *Rickettsia* are shown to account for 95% of all unintended amplifications from the genus. Furthermore, re-barcoding of extracts uncovers a new and substantial host diversity associated with *Rickettsia,* including phloem-feeding bugs, parasitoid wasps, blood-feeding flies and several detritivores. This suggests barcoding can be used as an adjunctive to targeted screens in discovering novel *Rickettsia*-host interactions. The inadvertent amplification of *Rickettsia* should be considered during the design of barcoding studies to avoid erroneous interpretation of species identification and population biodiversity.

5.2 Introduction

DNA barcoding is a technique which uses a mitochondrial DNA (mtDNA) reference sequence to identify existing species and aid in the discovery of new species (Hebert *et al.*, 2003). As well as being utilised as an identification system of individuals, mtDNA barcodes have also been used in phylogeographic studies to detect biodiversity and dispersal events of ancestors, including those of *Culicoides* biting midges (Craft *et al.*, 2010; Papadopoulou *et al.*, 2011; Baselga *et al.*, 2013; Jacquet *et al.*, 2015). The success of barcoding can be attributed to the "universal" nature of DNA barcoding primers used, making it a simple and standardised method. Barcoding has been particularly useful in the designation of arthropod species which can be difficult to assign taxonomic status to, based on morphological assessment alone. Ambiguities can arise from the presence of cryptic or new species within populations or because certain life stages of arthropods are problematic to classify (Kekkonen and Hebert 2014; Ashfaq *et al*. 2016; Horecka *et al*. 2017; Kanturski *et al*. 2018). In addition, the advent of sequencing technology has allowed for "metabarcoding" of environmental samples, which can allow for environmental monitoring as well as diet analysis studies (Xiong *et al.*, 2017; Kimmerling *et al.*, 2018; Robeson *et al.*, 2018; Yan *et al.*, 2018). The widespread uptake of DNA barcoding as a technique has led to the deposition and curation of millions of barcode sequences in databases such as the Barcode of Life Data System (BOLD; http://www.boldsystems.org/).

Like most methods, there are practical limitations to DNA barcoding. The *COI* (Cytochrome oxidase subunit I) gene used in DNA barcoding was initially chosen due to its high evolutionary rate, suitable effective population size and low recombination rate when compared to nuclear markers (Hurst and Jiggins 2005). A further important assumption was that *COI* was a neutral

genetic marker i.e. little selection was occurring at this gene. However, by virtue of *COI* being a functional gene, both purifying and positive selection have been observed at this locus (Ward and Holmes 2007; Castoe *et al*. 2008). Significantly, indirect selection on mtDNA can also occur in insects containing maternally inherited endosymbionts (Hurst and Jiggins 2005). This event arises as a result of the coinheritance of symbiont and mitochondria and can have several confounding effects on perceived population structures.

1) Through the erasure of mtDNA diversity within a population as a result of selective sweeps driven by an endosymbiont reproductive parasitism, such as cytoplasmic incompatibility (CI), which can lead to an underestimation of diversity (Ballard *et al.*, 1996; Jiggins, 2003; Shoemaker *et al.*, 2004; Zhang *et al.*, 2013).

2) Via the mistaken delineation of insect species occurring when isolating mechanisms (e.g. CI) of more than one endosymbiont strain is present within a population, leading to the segregation of mtDNA haplotypes (Schulenburg *et al.*, 2002; Jiggins, 2003; Kvie *et al.*, 2013).

3) Though the conflation of interspecific lineages through past mitochondrial introgression of hybridising species (Jiggins, 2003; Whitworth *et al.*, 2007; Raychoudhury *et al.*, 2009, 2010; Jäckel, Mora and Dobler, 2013).

The above confounding effects have been limited to *Wolbachia* and *Cardinium,* until a previous example of identical barcodes in two damselfly species (*Coenagrion puella* and *Coenagrion pulchellum;* Freeland and Conrad, 2002) was judged to be as a result of mtDNA interspecific introgression linked to *Rickettsia* infection (Panupong Thongprem, University of Liverpool; unpublished data). Consequently, *Rickettsia* infections also appear to have the potential to confound barcoding studies when interpreting mtDNA divergence.

A further problem associated with DNA barcoding is the accidental amplification of an unintended *COI* sequence instead of a target barcode. The universal nature of barcoding primers can lead to unintended amplifications through contamination from human handling (Ratnasingham and Hebert 2007), as well as organisms associated with the intended host target, such as parasitoids (Ramage *et al.*, 2017). These occur despite the attempted development of taxa-specific barcoding primers (Hebert *et al.*, 2004). Further, there may be amplification of mitochondrial gene copies that are present in the nuclear genome (NUMTs), and these may render barcodes unreadable using the Sanger method (Song *et al.*, 2008).

It is also possible to experience the amplification of bacterial *COI* from DNA barcoding of arthropods (Linares *et al.*, 2009; Smith *et al.*, 2012; Řezáč *et al.*, 2014; Ceccarelli, Haddad and Ramírez, 2016). In the largest investigation into such microbial amplification, Smith *et al*. (2012) found numerous *Wolbachia COI* hits when exploring barcodes derived from BOLD. Although it was concluded that *Wolbachia* presence would rarely interfere with intended target *COI* amplification, it was noted that DNA barcoding could inadvertently reveal previously unknown *Wolbachia*-insect associations, as well as providing additional bacterial phylogenetic information.

The only other microbe demonstrated to cause erroneous amplification is *Rickettsia,* which is unsurprising considering it belong to the Rickettisales order of bacteria alongside *Wolbachia*. For example, erroneous amplification of a *Rickettsia* endosymbiont has been detected in the Anyphaenidae and Dysderidae families of spider (Řezáč *et al.*, 2014; Ceccarelli, Haddad and Ramírez, 2016). *Amaurobioides africana*, an intertidal spider of the southern coast of Africa,

was noted to amplify *Rickettsia* in 6/19 individuals via a barcoding primer set containing several degenerate sites. Although the inclusion of degenerate sites in primer design is often used to increase sensitivity, this is likely to reduce specificity and subsequently increase chances of microbe *COI* amplification. Furthermore, the multilocus *Rickettsia* designation system devised in chapter 3 included *CoxA* (the bacterial equivalent of *COI*) and, when compared to BOLD and NCBI databases, returned several hundred unintended database entries that were the result of barcoding projects. These observations indicate that inadvertent *Rickettsia* amplification warrants further investigation.

Rickettsia are most commonly regarded as human pathogens vectored by haematophagous arthropods, being responsible for epidemic typhus and various "spotted fevers" (Cowan, 2000; Parola *et al.*, 2013). Therefore, the initial research focus on *Rickettsia* pertained to disease treatment and vector control. However, in subsequent years, knowledge of the genus has greatly expanded and has been found to include hosts which include non-blood feeding arthropods, as well as amoebae and leeches (Dyková *et al.*, 2003; Perlman, Hunter and Zchori-Fein, 2006; Weinert *et al.*, 2009) and it is likely that vertebrate disease-causing *Rickettsia* are atypical of the genus as a whole. Although some members are known to induce male-killing in ladybirds, (Werren *et al.*, 1994; Hurst *et al.*, 1999; Majerus *et al.*, 1999) and parthenogenesis in parasitoid wasps (Hagimori *et al.*, 2006; Giorgini *et al.*, 2010), the interactions between a majority of new strains and their hosts remain unexplored. This includes, specifically, the Torix group of *Rickettsia* which encompasses a variety of hosts that have mostly been associated with a semi-aquatic habitat (chapter 3). Further work is needed to explore Torix interactions, especially as many hosts are vectors of vertebrate diseases.

These include: *Simulium* sp. (black flies); *Phlebotomus chinensis* and *Lutzomyia apache* (sand flies); *Nosopsyllus laeviceps* (fleas) and *Culicoides* spp. (biting midges) (Reeves *et al*. 2008; Li *et al*. 2016; Song *et al*. 2018; chapter 3). Additionally, many of these insects are difficult to maintain in lab culture, leading to the pressing need to find model systems to elucidate *Rickettsia* effects on vectorial capacity.

Considering the above, and with special permission to access BOLD trace files, this chapter intended to uncover the extent of inadvertent *Rickettsia COI* amplification through barcoding. This study specifically aimed to address the questions:

1) Is this phenomenon common and, if so, is *Rickettsia* likely to affect the integrity of future barcoding studies?

2) Is *Rickettsia* amplification a result of an affinity to frequently used barcoding primers?

3) Can barcoding be used as an adjunctive to targeted *Rickettsia* screens to discover new host taxa?

4) What phylogenetic information can be gathered from non-target *Rickettsia* COI?

5.3 Methods

5.3.1 The erroneous amplification of *Rickettsia* through barcoding

As part of the multilocus analysis from chapter 3 describing the analysis of several *Rickettsia* housekeeping genes, it was noted that *Rickettsia coxA* (the bacterial equivalent of *COI*) was found to match to several sequences on Genbank deposited as invertebrate mitochondrial sequences (NCBI; http://www.ncbi.nlm.nih.gov/genbank/; accessed January 2018) (Table 5.1.). This match indicated barcoding studies were occasionally amplifying *Rickettsia* DNA instead of host mtDNA. Further to this, searching BOLD, a depository of >6 million *COI* mtDNA sequences, hundreds of hits were observed with high homology to both Limoniae and Leech *Rickettsia* subclades. Subsequently, as the record details of these matches were not publicly available, Professor Alex Smith (University of Guelph, Canada), who had previously worked on non-target *Wolbachia* amplification associated with BOLD (Smith *et al.*, 2012) was contacted for further investigation.

5.3.2 BOLD datasets acquisition

As BOLD archives all Sanger sequence traces, non-target sequences from the database were requested. Subsequently, *COI* barcoding data deriving from a screening project, totalling >185K arthropod specimens from 21 countries, were provided. *COI* sequences provided by BOLD were generally from somatic tissues (legs are often used in order to retain most of the specimen for further analyses if necessary), but also rarely included abdominal tissues. The first dataset made available included 3,829 sequences deemed as contaminant sequences as a result of not matching initial morphotaxa assignment. The second dataset included 55,271 specimens judged to not contain erroneous amplicons. A remaining 126K specimens were not

NCBI Host identification	Accession number	BLAST hit
Calopteryx maculata (damselfly)	KM383849	100% identity to *Culicoides bysta Rickettsia* (KY765405)
Eucharitidae sp. (parasitoid wasp)	KC182318	94% identity to *Culicoides bysta Rickettsia* (KY765405)
Pimelia sp. (darkling beetle)	MH158030	92% identity to *Culicoides bysta Rickettsia* (KY765405)
Paracalliope fluviatilis (Freshwater amphipod)	KR336946	92% identity to *Culicoides impunctatus Rickettsia* (KY765403)
Formicidae sp. (ant)	KP421783	92% identity to *Culicoides newsteadi* N1 *Rickettsia* (KY765400)
Flavina sp. (Issidae planthopper)	HM452248	98% identity to *Amaurobioides africana Rickettsia* (KU600823)
Mycetophila lunata (fungus gnat)	KM679400	98% identity to *Amaurobioides africana Rickettsia* (KU600823)

Table 5.1. NCBI matches mistaken for true mtDNA barcodes and their homology to *Rickettsia COI* (Accessed January 2018).

made available meaning the 55K subsample was used as a representative sample from which the contaminants had originated (Figure 5.1.).

The highest percentage of specimens from the project (33%) came from Canada, likely due to the location of BOLD database curation (Guelph, Canada), as well as the large-scale barcoding projects that have been undertaken in the country over the past decade (Hebert *et al.*, 2016). Other countries providing specimens included Australia, Argentina, Bulgaria, Brazil, China, Costa Rica, Egypt, Finland, Germany, Greenland, Honduras, Malaysia, Madagascar, Mexico, New Zealand, Norway, Pakistan, Saudi Arabia, South Africa and USA. Collections took place between 2010 and 2014 using predominantly malaise traps although canopy, intercept and pitfall traps, alongside sweep and dip nets, were also utilised. Aside from collection date, method and location, further metadata assigned taxa identification usually to the order level, as well as a unique identifier code with the prefix "BIOUG". Conforming to BOLD workflow protocols (Ratnasingham and Hebert 2007), specimens had been photographed before dissections of legs (or occasionally abdomens), genomic DNA extraction and amplification of the *COI* gene. *COI* amplicon sequences generated through Sanger sequencing were then deposited in the BOLD database alongside associated metadata before being allocated a BIN (Barcode Index Number), assisted by a clustering algorithm designed with the intention to group barcodes to specific species.

5.3.3 Assessment of non-target microbe amplicons

Of the contaminant sequences obtained, bacterial and non-bacterial amplicons were separated by BOLD's filtering system where translated amino acid sequences are compared to the *COI* protein to confirm correct gene amplification (Ratnasingham and Hebert 2007).

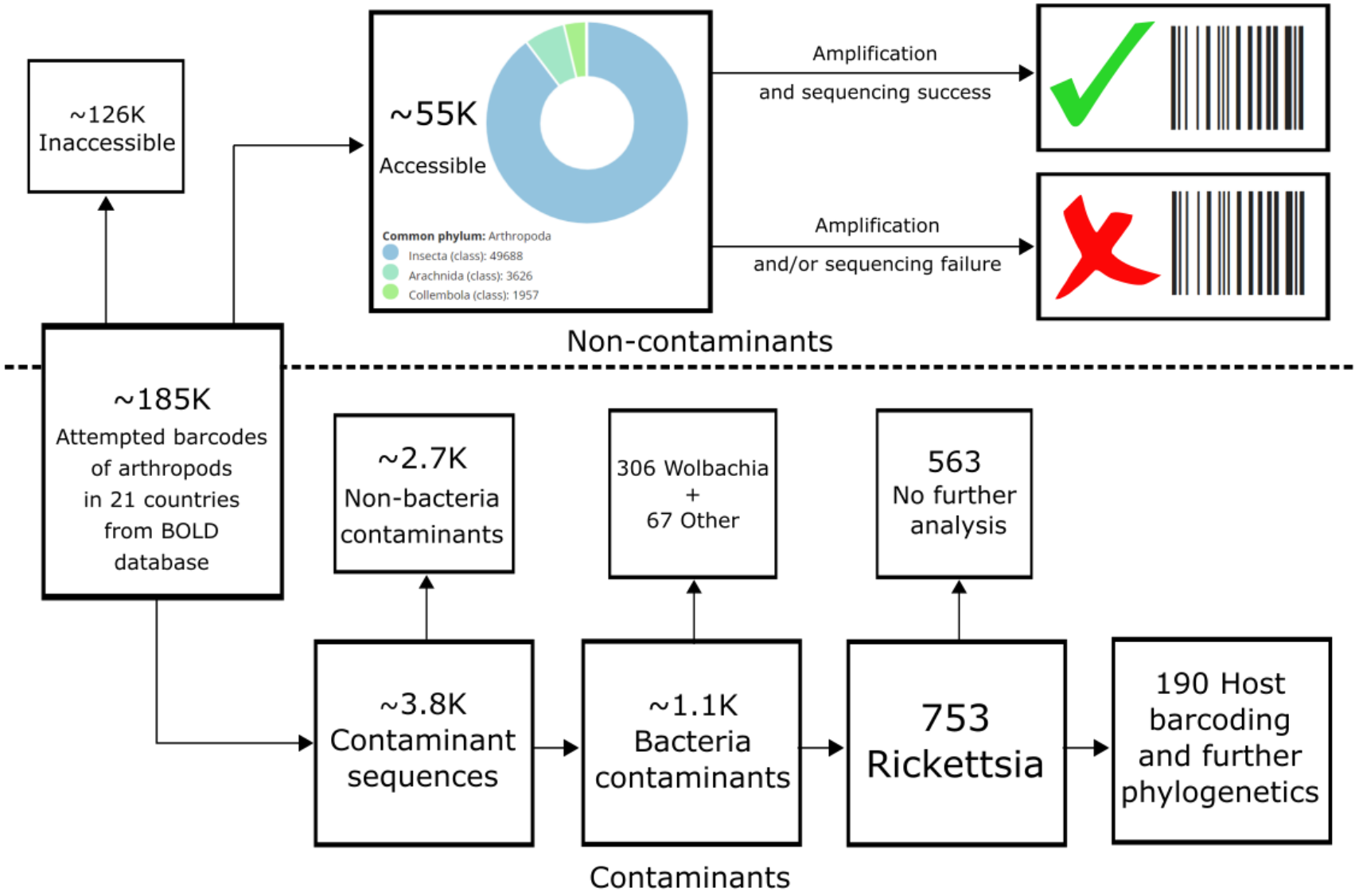


Figure 5.1. Workflow of the project demonstrating the acquisition and fates of contaminant and non-contaminant *COI* barcoding sequences.

Sequences were then compared against a variety of potential contaminants including *Wolbachia* and human homologues. Bacterial identity was also confirmed by measuring GC content, as microbes do not contain the distinctive arthropod mtDNA AT bias (Smith *et al.*, 2012). The confirmed microbial sequences were then aligned using the "L-INS-I" algorithm in MAFFT v7.4 (Katoh and Standley 2013) before using Gblocks (Castresana, 2000) to exclude areas of the alignment with excessive gaps or poor alignment; a less stringent selection was chosen through parameters allowing smaller final conserved blocks and gap positions.ModelFinder (Kalyaanamoorthy *et al.*, 2017) then determined the TIM3+F+G4 model to be used after selection based on default "auto" parameters using the Bayesian information criteria. A maximum likelihood (ML) phylogeny was then estimated with IQTree (Nguyen *et al.*, 2015) using an alignment of 559 nucleotides and 1000 standard bootstraps. The Rickettsiales genera *Ehrlichia*, *Anaplasma*, *Neorickettsia*, *Orientia*, *Wolbachia* (Supergroups A, B, E, F, H) and *Rickettsia* were included in the analysis as references. Quantum GIS software v3 (Quantum GIS Development Team, 2016) was then used to map BOLD-associated *Rickettsia* with their collection site. Finally, trees were drawn and annotated based on host taxa (order) using the EvolView (He *et al.*, 2016) online tree annotation and visualisation tools.

5.3.4 Phylogenetic analysis

Since original BOLD genomic DNA extracts are archived at the Biodiversity institute of Ontario, it was possible to request *Rickettsia*-containing samples of interest. This allowed for both validation and further phylogenetic analysis based on more housekeeping genes other than *COI*. As Torix *Rickettsia* are known to be at high prevalence within populations (chapter 3), it was likely that most of the genetically similar *Rickettsia* from the same taxa (order) and

collection sites would be from the same host species. Subsequently, aside from selecting samples spread across the whole of the phylogeny, a mixture of 190 DNA extracts were chosen based on varied collection site and host order.

The details of multilocus PCR screening and phylogenetic analysis of *Rickettsia* can be found in chapter 3. However, slight variations include the exclusion of the *atpA* gene due to observed recombination at this locus. Furthermore, the amplification conditions for the *17KDa* locus was changed because the single Leech reference DNA extract (Host: *Simulium aureum*) failed to amplify with the primer set Ri_17KD_F/ Ri_17KD_R from chapter 3. Subsequently, a *17KDa* alignment of *Rickettsia* groups originating from the genomes of Spotted fever, Typhus, Transitional, Belli, Limoniae and the newly available Megaira group (provided by Prof. R. S. Coyne, Dr T. Doak, and Prof. M. Lynch) was generated to design a new set of primers using the online tool PriFi (Fredslund *et al.*, 2005). The chosen locus was ensured to overlap with already characterised regions of the gene assessed in chapter 3. Once multilocus profiles of the *Rickettsia* had been established, samples amplifying at least 3 out of 4 genes (*16S*, *17KDa*, *coxA* and *gltA*) were concatenated before being analysed for recombination within and between genes using RDP v4 (Martin *et al.*, 2015) (details chapter 3). These sequences were then assessed again through ML phylogenetic analysis to differentiate samples into *Rickettsia* subclades. The selected models used in the concatenated partition scheme were as follows: *16S* and GTR+F+R2; *17KDa* and GTR+F+G4; *coxA* and TVM+F+I+G4; *gltA* and TVM+F+I+G4. Accession numbers for all reference *Rickettsia* and outgroup sequences can be found in Appendix 2.

5.3.5 Host barcoding

Aside from phylogenetic placement of these *Rickettsia*-containing samples, attempts were made to extract an mtDNA barcode in order to uncover the host diversity of infected specimens. Previous erroneous amplification of *Rickettsia* through DNA barcoding had occurred in the bed bug *Cimex lectularius*, with a recovery of the true barcode after using the primer set C1-J-1718/HCO1490 (Panupong Thongprem, University of Liverpool; unpublished data), which amplifies a shortened 455 bp sequence within the *COI* locus. Subsequently, all samples were screened using these primers or a further set of secondary *COI* primers (LCOt_1490/ MLepR1) if the first failed to give an adequate host barcode. All *COI* and *Rickettsia* multilocus screening primers, including references, are available in Table 5.2.

Cycling conditions for *COI* PCRs were as follows: initial denaturation at 95°C for 5 min, followed by 35 cycles of denaturation (94°C, 30 sec), annealing (50°C, 60 sec), extension (72°C, 90 sec), and a final extension at 72°C for 7 min. *Rickettsia* and host amplicons identified by gel electrophoresis were subsequently purified enzymatically (ExoSAP) and Sanger sequenced through both strands as described in chapter 2. Forward and reverse reads were assessed in UGENE (Okonechnikov *et al.*, 2012) to create a consensus sequence by eye with a cut-off phred (Q) score (Ewing and Green 1998) of 20; any ambiguous base calls below this threshold were attributed as “N” according to IUPAC nucleotide nomenclature . Primer regions were trimmed from barcodes before being matched to Genbank and BOLD databases by BLAST based on default parameters and an e-value threshold of <1e-85. Host taxonomy was determined by a barcode-based assignment of the closest BLAST hit, under the following criteria modified from Ramage *et al.* (2017):

Gene	Product	Primer name	Sequence (5'-3')	Product length (bp)	Reference
COI	Cytochrome c oxidase, subunit I	C1-J-1718	GGAGGATTTGGAAATTGATTAGT	455	Folmer *et al.* 1994; Dallas *et al.* 2003
		HCO1490	TAAACTTCAGGGTGACCAAAAAATCA		
		LCOt_1490	TGTAAAACGACGGCCAGTGGTCAACAAATCATAAAGATATTGG	310	Simon *et al.* 1994; Hajibabaei *et al.* 2006
		MLepR1	CCTGTTCCAGCTCCATTTTC		
		LepF1	ATTCAACCAATCATAAAGATAT	320	Hebert *et al.* 2004; Fisher and Smith 2008
		C_ANTMR1D_RonIIdeg_R	GGRGGRTARAYAGTTCATCCWGTWCC		
gltA	Citrate synthase	RiGltA405_F	GATCATCCTATGGCA	786	Pilgrim *et al.* 2017
		RiGltA1193_R	TCTTTCCATTGCCCC		
17KDa	17KDa antigenic protein precursor	Ri_Meg17kD_F	TGGYATGAATAARCAAGGTGG	319	This study
		Ri_Meg17kD_R	ATACTCACGACAATAYTGCCC		
16S	16S ribosomal RNA	Ri170_F	GGGCTTGCTCTAAATTAGTTAGT	1170	Pilgrim *et al.* 2017
		Ri1500_R	ACGTTAGCTCACCACCTTCAGG		

Table 5.2. Mitochondrial *COI* and bacterial gene primers used for re-barcoding and multilocus phylogenetic analysis.

1) Species level designation for at least 98% sequence identity over at least 90% of the barcode.

2) Genus level designation for at least 95% sequence identity and 90% overlap.

3) Family level designation for at least 80% sequence identity and 90% overlap.

Additionally, all sequences were required to be at least >200 bp in length. Newly identified hosts which amplified with at least 2 out of 4 *Rickettsia* genes were then placed phylogenetically using the same methods as above before being mapped by lifestyle and diet.

5.3.6 Assessment of barcoding success

One of the factors determining a successful *COI* bacterial amplification is the initial failure of an extract to amplify mtDNA. Subsequently, to determine the likelihood of this event within taxa, we used the 55K representative dataset to evaluate failure rates. To this end, all taxa at the order level, which gave at least one non-target bacteria *COI* hit were assessed. The barcoding success rate was determined as the proportion of specimens which matched initial morphotaxa assignment and were not removed after BOLD quality control (Ratnasingham and Hebert 2007). With respect to the latter, chromatograms were allocated a Q score for each nucleotide position as well as a mean value for the entire barcode. Generally poor and failed trace files were defined as mean Q scores of <30, although designation of a submitted BOLD sequence was still ultimately determined by the submitter. As the total *Rickettsia* count was from a larger dataset than the one made available, an adjusted prevalence for each taxa was calculated based on the 55K representative dataset.

5.3.7 Assessment of primer bias

A further determining factor for erroneous amplification of bacteria is primer site matching to respective microbe sequences. Subsequently, pairwise homology of the primer set predominantly used for BOLD barcode screening and *Rickettsia* or *Wolbachia* was assessed using Mega v6 (Tamura *et al.*, 2013).

5.4 Results

5.4.1 *Rickettsia* is the most common bacterial contaminant following barcoding

Out of 3,829 sequences considered contaminants, 1,126 of these were deemed by BOLD to be bacterial in origin (Figure 5.1. and 5.2.). The higher 37.16% mean GC content compared to 30.82% in non-bacterial contaminants supported the correct designation of microbes. The dominating genus was *Rickettsia* with 753 (66.9%) amplifications compared to *Wolbachia* with 306 (27.2%). When considering the 185K specimens in the total project, this gave an overall *Rickettsia* and *Wolbachia* prevalence of 0.406% and 0.165% respectively. Through later access to the 55K representative dataset from where the contaminants originated, further unique bacteria contaminants were also detected (possibly missed by BOLD's automated contaminant filtering system). This indicates that the prevalences estimated here are likely to be an underestimate of the overall project. Of the remaining 67 non-target sequences, 16 formed a monophyletic group with other Anaplasmataceae whereas 51 were undesignated proteobacteria. A majority of *Wolbachia* clustered with supergroup A (215/306) and were dominated by Hymenoptera (92%), whereas the second highest representation was found in supergroup B (42/306), which were predominantly hemipteran (81%) (Figure 5.2.).

5.4.2 *COI* primer attributes

Two sets of *COI* primers were responsible for 99% of *Rickettsia* amplifications (Table 5.3.) with a majority (90%) amplifying with the primer cocktails of CLepFolF/ClepFolR (Folmer *et al*. 1994; Hebert *et al*. 2004). When comparing CLepFolF/ClepFolR primer homology to the equivalent priming sites of *Wolbachia* and *Rickettsia* groups, 3-6 pairwise nucleotide

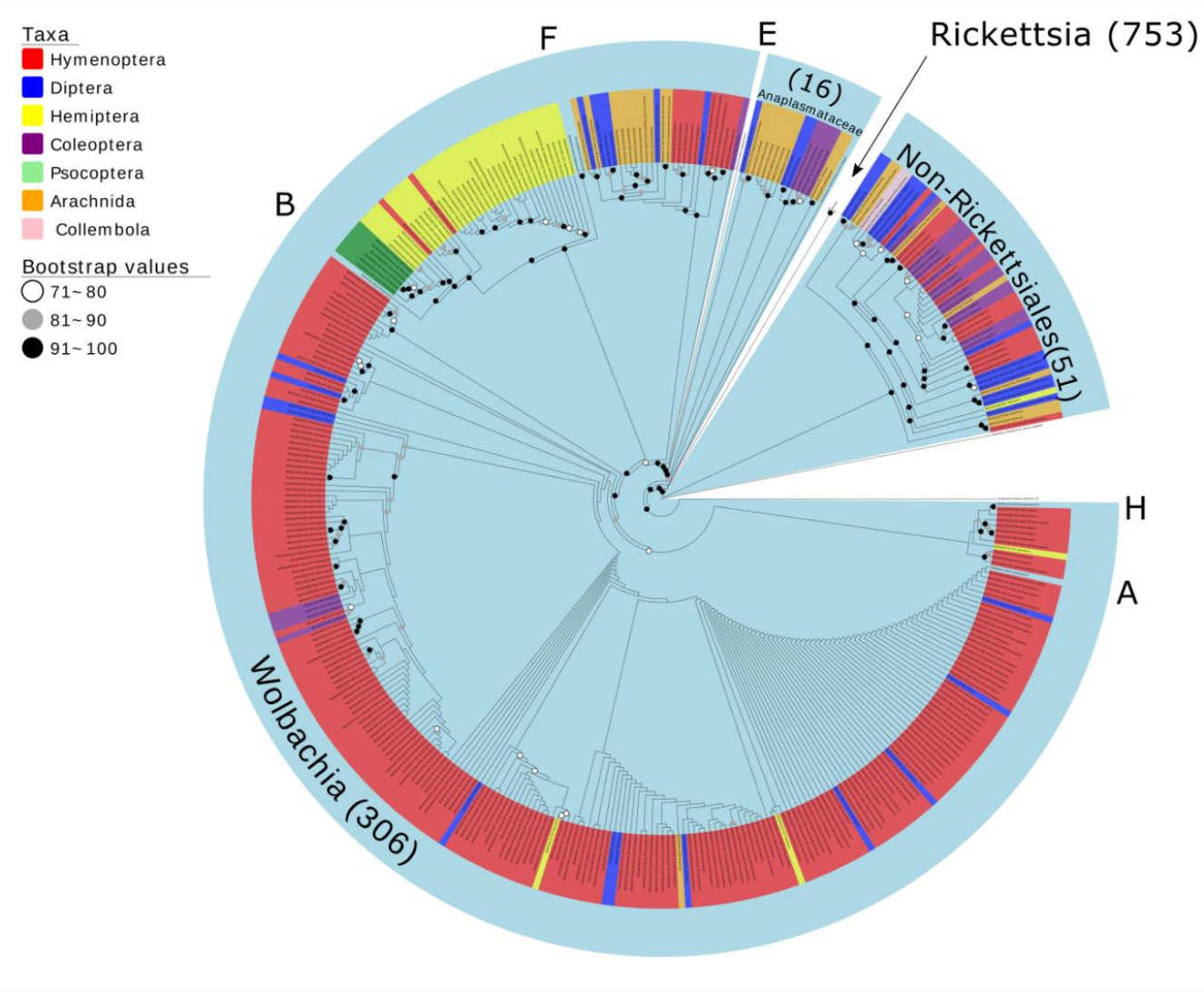


Figure 5.2. Cladogram of the maximum likelihood (ML) tree of 1,126 alphaproteobacteria *COI* contaminants retrieved from BOLD projects incorporating >185,000 arthropod specimens. The tree is based on 559bp and is rooted with the free-living alphaproteobacteria *Pelagibacter ubique*. Parantheses indicate the number of BOLD contaminants present in each group. Tips are labelled by BOLD processing ID and host arthropod taxonomy: Red=Hymenoptera; Blue=Diptera; Yellow=Hemiptera; Purple=Coleoptera; Green=Psocoptera; Orange=Arachnida; Pink=Collembola; No colour=Non-BOLD Reference. The Rickettsiales sequences of *Ehrlichia*, *Anaplasma*, *Neorickettsia*, *Orientia*, *Wolbachia* supergroups (A, B, E, F and H) are included as references.

Primer pair	Sequences (5'-3')	Rickettsia amplification	Reference
C_LepFolF/ C_lepFolR	F: RKTCAACMAATCATAAAGATATTGG	89%	Hernández-Triana *et al.*, 2014
	R: TAAACTTCWGGRTGWCCAAAAAATCA		
LepF2_t1 / LepR1	F: TGTAAAACGACGGCCAGTAATCATAARGATATYGG	10%	Hebert *et al.*, 2004; Park *et al.*, 2011
	R: TAAACTTCTGGATGTCCAAAAA		
LepF1/ LepR1	F: ATTCAACCAATCATAAAGATAT	0.5%	Hebert *et al.*, 2004
	R: TAAACTTCTGGATGTCCAAAAA		
LepF2_t1 / MHemR	F: TGTAAAACGACGGCCAGTAATCATAARGATATYGG	0.5%	Park *et al.*, 2011
	R: GGTGGATAAACTGTTCAWCC		

Table 5.3. Primer pairs involved in the unintended amplification of 753 *Rickettsia COI* from BOLD projects.

differences were observed across the 26 nucleotide C_LepFolF and 1-4 nucleotides across C_LepFolR (Tables 5.4.1 and 5.4.2.). Although all *Rickettsia* and *Wolbachia* contained a SNP at the 3' priming site (position 24) of CLepFolR, Torix *Rickettsia* (*Culicoides newsteadi N5*; Pilgrim *et al.*, 2017) was the only sequence to not contain a similar SNP at the 3' priming end of CLepFolF (position 23).

5.4.3 The diversity of non-target *Rickettsia*

BOLD *Rickettsia* contaminants were dominated by the Torix group with 716/753 (95.1%) forming a clade with Limoniae and Leech reference sequences generated in chapter 3, which overlapped with the BOLD *COI* barcode locus by approximately 350 bp (Figure 5.3.1.). The remaining 37 *Rickettsia* clustered with groups Transitional (n=15), Belli (n=9), Rhyzobius (n=1) and 12 were undesignated. Breaking this down at a host taxa level: 294 (39%) were present in Hymenoptera; 187 (25%) in Diptera; 177 in Hemiptera (24%); 40 each in Coleoptera and Psocodea (5% each); 8 in Arachnida (1%); 4 in Trichoptera (0.5%); and single cases of Thysanoptera, Diplopoda and Dermaptera (0.3% each).

Using the reference sequences for both Limoniae and Leech subclades of the Torix group it was possible to tentatively allocate the *Rickettsia* sequences to either subclade (Figure 5.3.2.). Hymenoptera and Diptera were present in both subclades, while 168/170 Hemiptera were allocated to the Leech group. Conversely, a single coleopteran out of 33 was allocated in the Leech group with the remainder placed in the Limoniae subclade. A majority of Diptera were not able to be placed in either subclade with strong support Mapping the 753 *Rickettsia* to

Sequence 5-3' (accession number)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**
CLepFolF (primer)	R	K	T	C	A	A	C	M	A	A	T	C	A	T	A	A	A	G	A	T	A	T	T	G	G
Torix (PRJNA376033)	T	C	.	.	T	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.
Bellii (CP000087)	T	.	.	.	T	.	.	.	.	.	C	.	.	G	.	.	G	.	.	.	.	.	C	.	.
Spotted fever (CP000848)	T	.	.	.	T		.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	C	.	.
Typhus (CP004888)	T	.	.	.	T	.	.	T	.	.	.	.	.	C	.	.	.	.	.	.	.	.	C	.	.
Transitional (CP000053)	T	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.
Megaira (unpublished)	T	.	.	.	T	.	.	T	.	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	.
Wolbachia (AE017196)	T	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.

Table 5.4.1. Homology of *Rickettsia* groups and *Wolbachia* to the most common forward primer (CLepFolF) attributed to bacterial *COI* amplification from arthropod DNA extracts.

Sequence 5-3' (accession number)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**
CLepFolR (primer)	T	A	A	A	C	T	T	C	W	G	G	R	T	G	W	C	C	A	A	A	A	A	A	T	C	A
Torix (PRJNA376033)	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.
Bellii (CP000087)	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.
Spotted fever (CP000848)	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	.	.
Typhus (CP004888)	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.
Transitional (CP000053)	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	.	.
Megaira (unpublished)	.	.	.	.	.	C	.	.	.	.	.	.	.	.	C	.	.	.	.	.	C	.	.	C	.	.
Wolbachia (AE017196)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.

Table 5.4.2. Homology of *Rickettsia* groups and *Wolbachia* to the most common reverse primer (CLepFolR) attributed to bacterial *COI* amplification from arthropod DNA extracts.

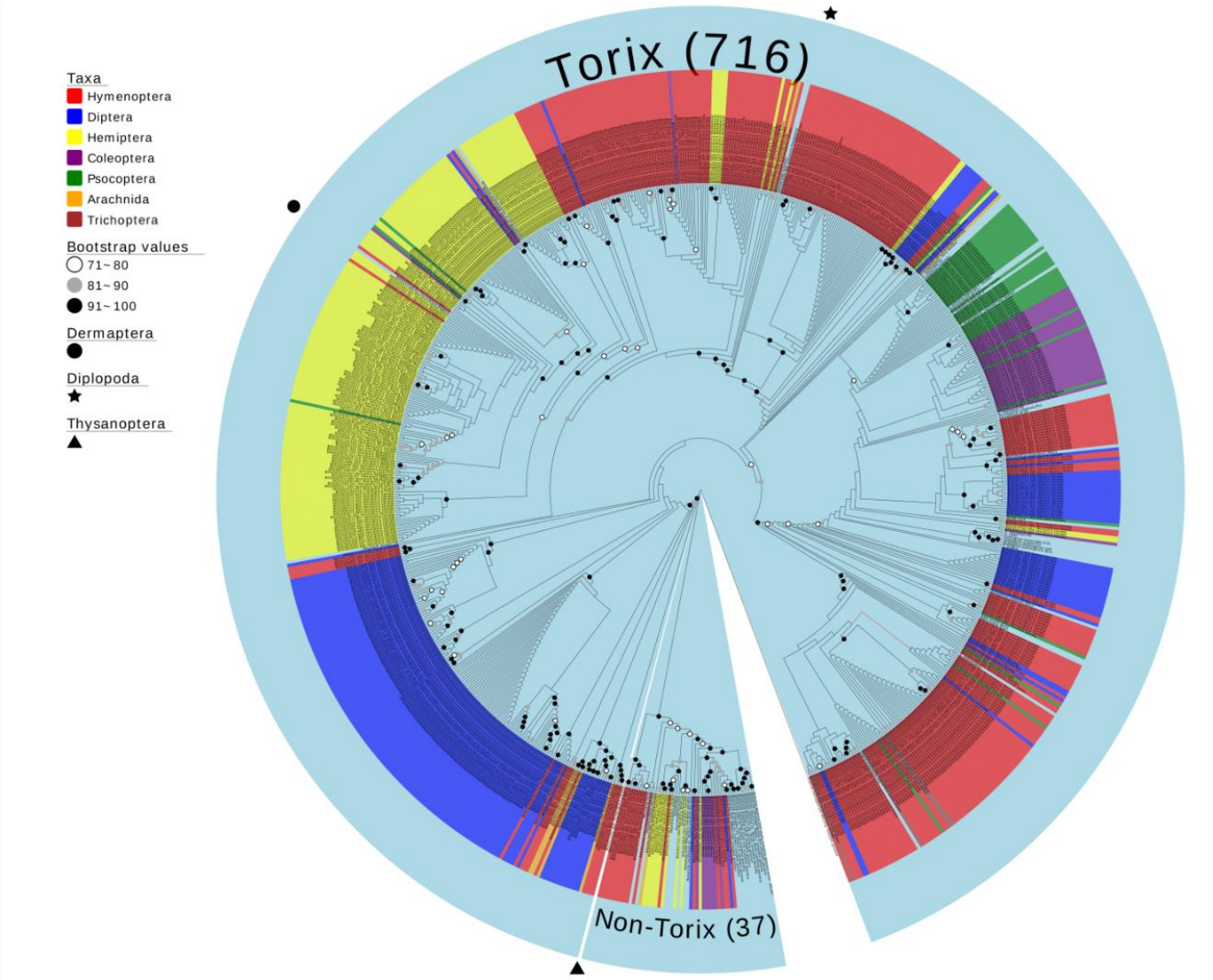


Figure 5.3.1 Cladogram of a maximum likelihood (ML) tree of 753 *COI Rickettsia* contaminants retrieved from BOLD projects incorporating >185,000 arthropod specimens. The tree is based on 559bp with an overlap of 349bp for *Rickettsia* reference sequences and is rooted by *Orientia tsutsugamushi*. Parantheses indicate the number of BOLD contaminants present in Torix and non-Torix *Rickettsia* groups. Tips are labelled by BOLD processing ID and host arthropod taxonomy: Red=Hymenoptera; Blue=Diptera; Yellow=Hemiptera; Purple=Coleoptera; Green=Psocoptera; Orange=Arachnida; Brown=Trichoptera; No colour=Non-BOLD reference sequence unless designated by a circle (Dermaptera), star (Diplopoda), triangle (Thysanoptera). The *Rickettsia* groups: Spotted fever, Transitional, Bellii, Typhus and Torix (Leech/Limoniae) are included as references.

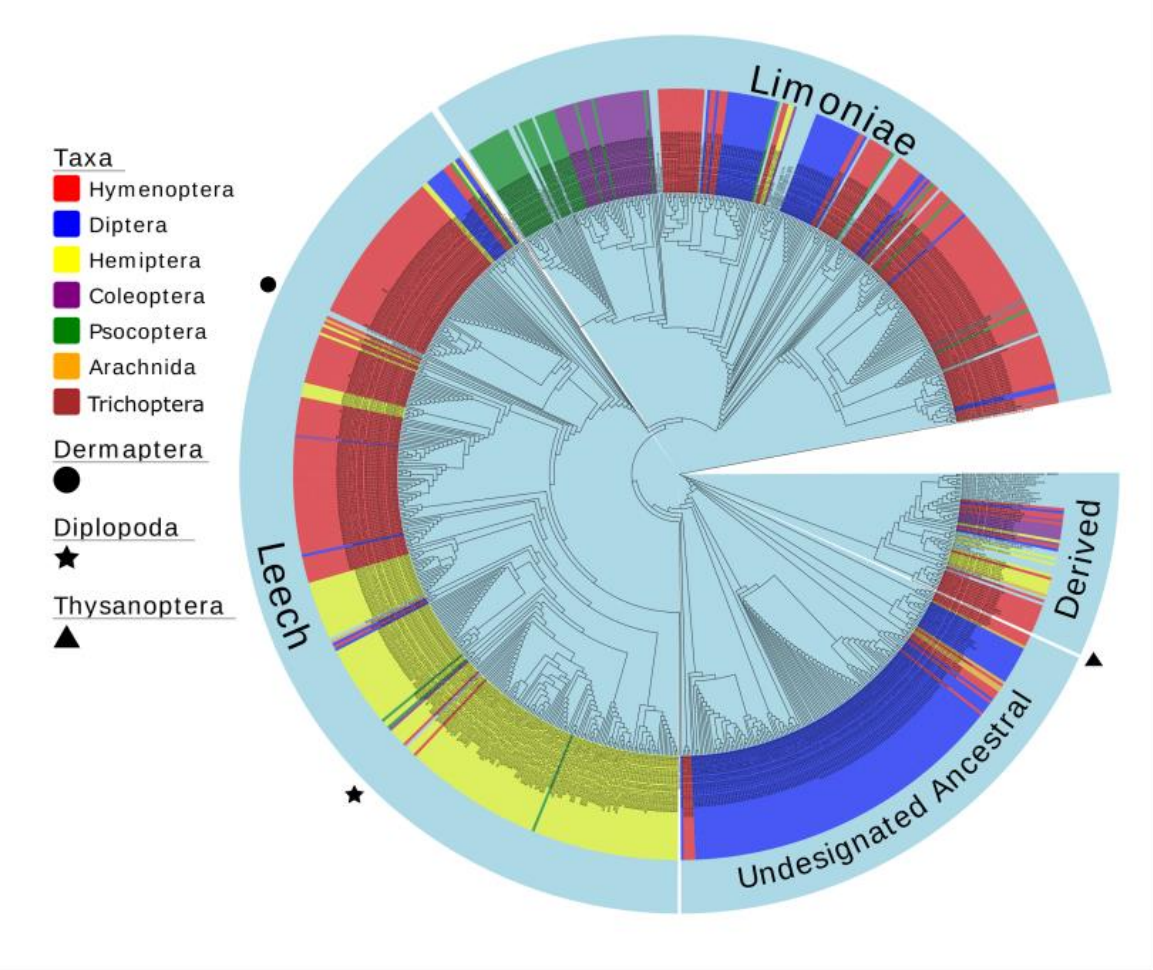


Figure 5.3.2. Identical cladogram tree to Figure 5.3.1. but taxa are separated into the Leech and Limoniae subclades of the Torix *Rickettsia* group. Torix sequences which don't form a clade with either group are undesignated.

collection site (Figure 5.4.) revealed arthropod infections predominantly from Canada with other locations in South/Central America, Europe, Africa and Asia.

5.4.4. *Rickettsia* multilocus phylogenetic analysis

To confirm the allocation of these sequences and to better place the undesignated Torix *Rickettsia*, a subsample of 190 *Rickettsia*-containing taxa were screened with 2 further housekeeping genes (*16S, gltA*) and the antigenic 17KDa protein gene. 120 extracts successfully amplified and gave a high-quality sequence for at least one gene. No intragenic or intergenic recombination was detected for any of the genes. A phylogram including 106 taxa containing a concatenation of at least 3 of the 4 *Rickettsia* genes of interest (including *COI*) confirmed the initial designation of Leech and Limoniae *Rickettsia*, while placing the undesignated, largely Dipteran group, as an early branch of the Limoniae subclade (Figure 5.5.). Through inclusion of multilocus profiles from Torix *Rickettsia* not associated with BOLD samples (Appendix 2), it became clear that both Leech and Limoniae subclades were paraphyletic with respect to host taxa. For example, Odonata were distributed sporadically across the Limoniae clade, while specific families (Hemiptera: Psyliidae and Hymenoptera: Diapriidae) were found present in both Leech and Limoniae groups. A full list of multilocus profiles and *Rickettsia* group designation can be found in Appendix 3.

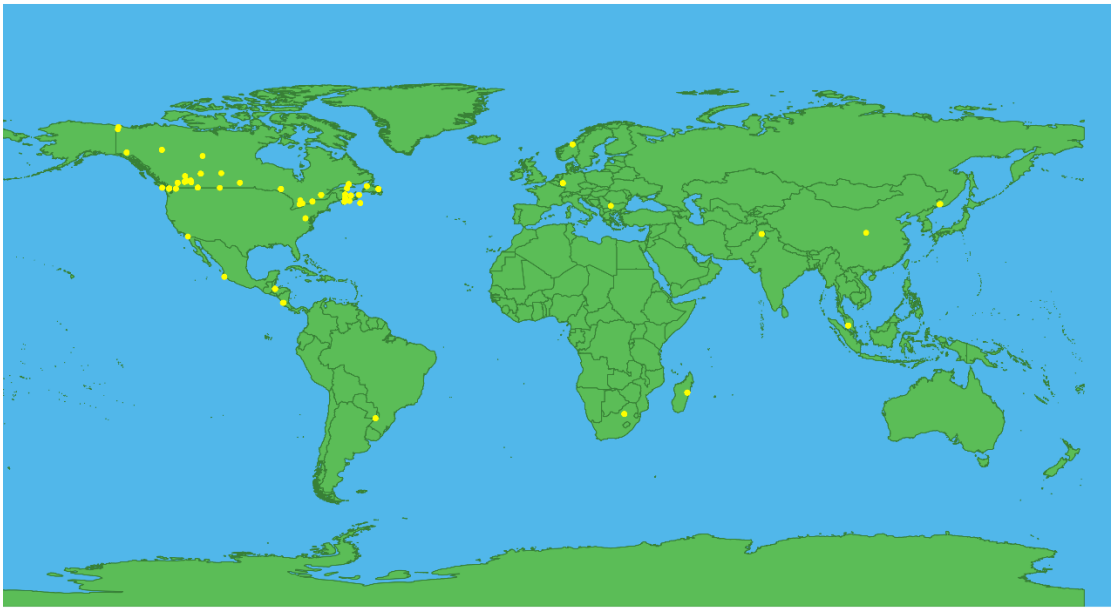


Figure 5.4. Collection sites of the 753 *COI Rickettsia* contaminants retrieved from BOLD projects.

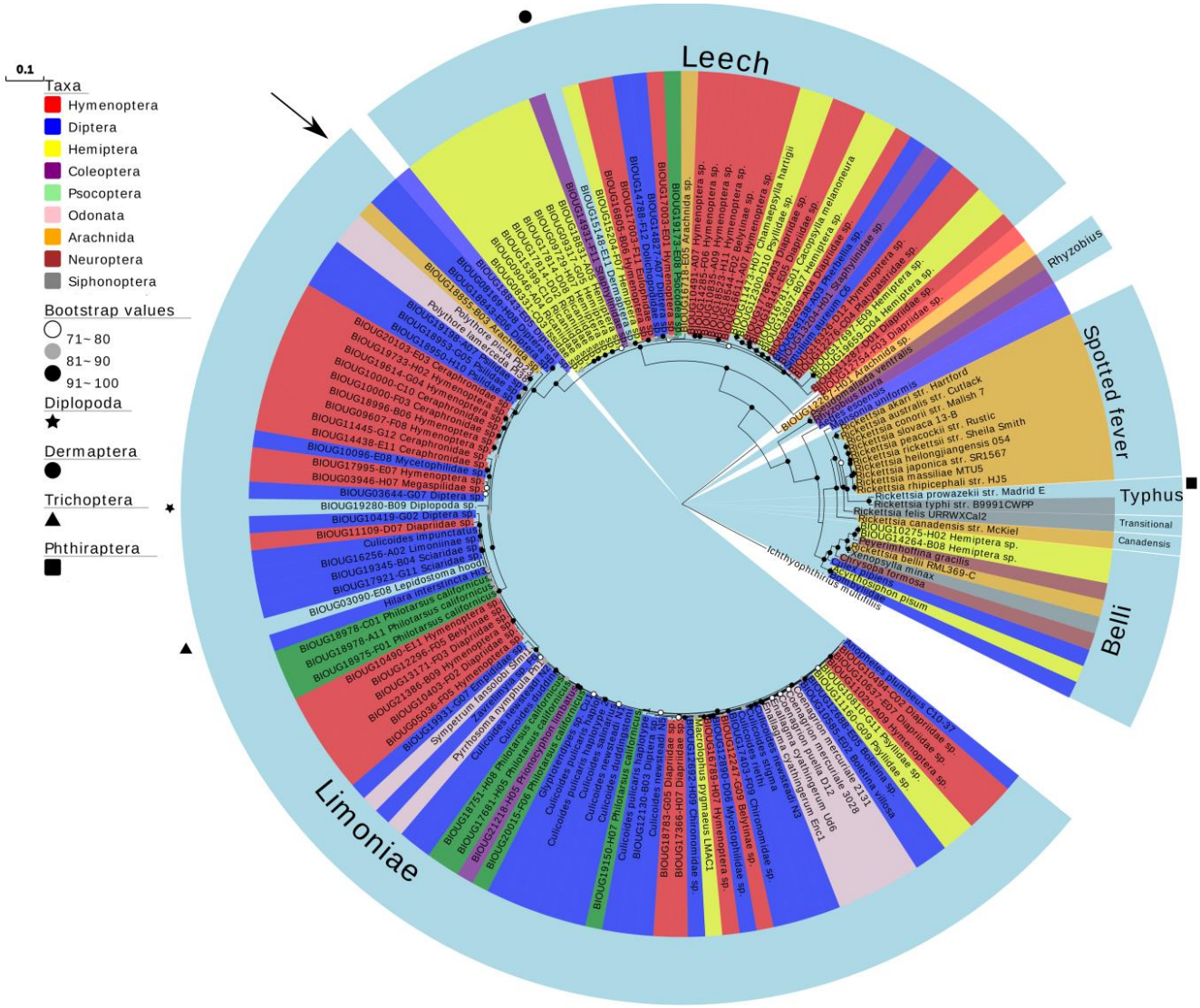


Figure 5.5. Phylogram of the maximum likelihood (ML) tree of 106 COI Rickettsia contaminants (prefix "BIOUG") used for further phylogenetic analysis and 53 Non-BOLD reference profiles. The tree is based on the concatenation of 4 loci; *16S*, *17KDa*, *gltA* and *COI* under a partition model, with profiles containing at least 3 out of 4 sites included in the tree (2,834bp total) and is rooted by the *Rickettsia* (group Megaira) endosymbiont of *Ichthyophthirius multifiliis*. Tips are labelled by host arthropod taxonomy: Red=Hymenoptera; Blue=Diptera; Yellow=Hemiptera; Purple=Coleoptera; Green=Psocoptera; Orange=Arachnida; Pink=Odonata; Brown=Neuroptera; Grey=Siphonoptera. The *Rickettsia* groups Spotted fever, Transitional, Bellii, Rhizobius, Canadensis and Typhus are included as references (Accession numbers: Appendix 2). Arrow= Dipteran *Rickettsia* strains previously undesignated from single locus (*COI*) phylogenetic tree in Figure 5.3.2.

5.4.5 The hidden host diversity of Torix *Rickettsia*

Through the discovery of *Rickettsia* NCBI hits labelled as arthropod mtDNA (Table 5.1.), it became clear that host diversity of Torix *Rickettsia* was wider than previously thought. Identified erroneous amplifications of *Rickettsia* from previous studies include *Calopteryx maculata* (Damselflies); Eucharitidae (Parasitoid wasps); *Pimelia* sp. (Darkling beetles); Formicidae (Ants); *Flavina* sp. (planthoppers); *Mycetophila lunata* (Fungus gnats) and *Paracalliope fluviatilis*, a freshwater amphipod. Although taxonomic classification for most BOLD IDs are initially given through morphological assessment, this is usually only down to the order level. Subsequently, in order to refine the taxonomy further of the *Rickettsia*-containing specimens, attempts were made to re-barcode DNA extracts using a different set of primers to the ones used originally in the initial barcoding project. Successful barcodes were obtained from 84 of the 190-subsample set using the primer pair C1-J-1718/HCO1490 (Table 5.2.). A further 5 Diptera returned a barcode after a secondary amplification with the primer pair LCOt_1490/MLepR1 and a single hymenopteran produced a barcode with primers LepF1/C_ANTMR1D.

Out of 28 Diptera producing successful barcodes, those conforming to the requirements of a species, genus or family allocation (see methods for details) and representing novel *Rickettsia*-hosts included the Fungus gnat *Boletina villosa* and the Marsh fly *Pherbellia tenuipes*; while a single specimen was classified as belonging to the genus *Cerodontha* (Leaf-miner flies). Several Dipteran families were also identified including: Sciaridae (Dark-winged fungus-gnat); Cecidomyiidae (Gall-midges); Mycetophilidae (Fungus-gnats); Chironomidae (Chironomids); Dolichopodidae (Long-legged flies); Limoniidae (Crane flies); Empididae (Balloon flies); Psilidae (Rust flies); Hybotidae (Dance flies) and Tabanidae (Horseflies). From

27 Hymenoptera barcodes several parasitoid wasp families were revealed including: Megaspilidae; Ceraphronidae; Diapriidae; Platygastridae; Eulophidae; Eurytomidae and Mymaridae (Fairy wasps). Of the 15 Hemipteran barcodes, the planthopper families: Issidae, Ricaniidae, Achilidae and Flatidae (subfamily: Aurrenchorhyna) were observed. At the species level, the phloem-feeding Psyllidae *Chamaepsylla hartigii* and *Cacopsylla melanoneura* were identified. Coleopteran families included: Staphylinidae (Rove beetles); Dystiscidae (water beetles) and Elateridae (Click beetles) with a single barcode indicating the Marsh beetle *Prionocyphon limbatus* as a host species. Out of 11 screened Psocodea, 10 from Canada gave an identity of >99% to the barklouse *Philotarsus californicus* while the remaining individual from Honduras was classified to the barklouse family Myopscocidae. Single barcodes were recovered for the orders Thysanoptera, Trombidiformes and Trichopteran which were refined down to the families Thripidae (Thrips), Calyptostomatidae (Water mites) and the species *Lepidostoma hoodi* (Caddisfly) respectively. Notably, the single specimens of Dermaptera and Diplopoda failed to barcode; full details of barcode success of BOLD *Rickettsia* extracts including BLAST hits can be found in Appendix 3. Noteworthy lifestyles of hosts were then mapped on a phylogram (Figure 5.6.) of the *Rickettsia* genus to assess any significant patterns. Similar to when taxa were mapped across the phylogeny (Figure 5.5.), feeding habits such as herbivory, predation and haematophagy were not correlated with any particular *Rickettsia* clade. Furthermore, parasitoidism and aquatic lifestyles were seen across the phylogeny.

5.4.6 *Rickettsia* co-infections

In concordance with observations in chapter 3, double peaks were occasionally found at third codon sites in chromatogram trace files from protein coding genes. This suggests co-infecting

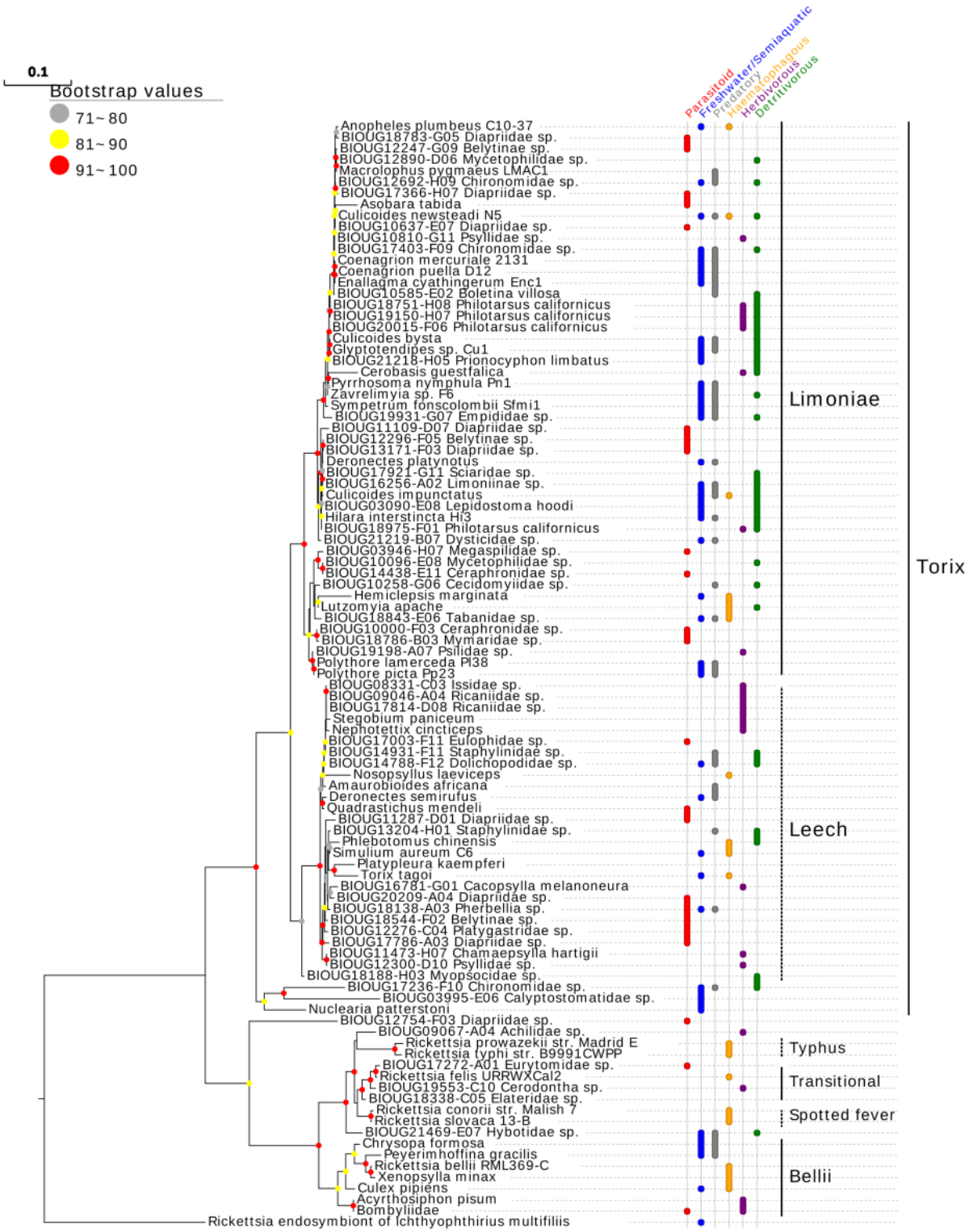


Figure 5.6. Phylogram of a maximum likelihood (ML) tree of 55 *COI Rickettsia* contaminants (prefix "BIOUG") giving a host barcode and 43 Non-BOLD reference profiles. The tree is based on 4 loci; *16S*, *17KDa*, *gltA* and *CoxA* under a partition model with profiles containing at least 2 out of 4 sites included in the tree (2,834bp total) and is rooted by the *Rickettsia* (group Megaira) endosymbiont of *Ichthyophthirius multifiliis*. The habitats and lifestyles of the host are given to the right of the phylogeny.

Rickettsia strains in hosts is a widespread phenomenon of the Torix group. To ensure cross-contamination between DNA extracts had not occurred, the original trace files from attempted BOLD *COI* amplifications were visualised to confirm double peaks had also occurred at the time of initial processing. Co-infections were observed in 6/10 *Philotarsus californicus* as well as a minority of Psilidae (1/3), Sciaridae (1/4), Chironomidae (1/4) and Diapriidae (1/18) (Appendix 3).

5.4.7 Barcoding success of taxa

An available BOLD project of attempted barcodes associated with the contaminants contained 55,271 out of 185K arthropods originally used in the overall study. The 3 classes of Insecta (n=49,688), Arachnida (n=3,626) and Collembola (n=1,957) accounted for >99.8% of total specimens (Figure 5.1.). Successful amplification and sequencing of *COI* was achieved in 43,246 specimens (78.1%) of the genomic extracts, but when assessed at the order level success rates varied (Table 5.5.). The likely explanation for this variation is taxa-specific divergence of sequences at priming sites.

The number of each taxonomic order giving at least 1 *Rickettsia* amplification was then calculated and adjusted based on the total number of specimens in the project to allow for a prevalence estimate. Overall, Hymenoptera, Diptera and Hemiptera were the three most represented taxa associated with *Rickettsia* COI amplification (87.4%). Similarly, on assessment of a subsample from the project where the contaminants originated, a majority (77.9%) of the dataset were also accounted for by these three orders. After adjusting the prevalence to take into account the number of inaccessible specimens, Trichoptera (2.27%),

Subsample of screened taxa (N=51,475) giving at least one *Rickettsia COI* sequence			***Rickettsia* contaminants from total screened taxa (N=185,250)**		
Taxa	Specimens (n)	Barcoding success rate	Adjusted total specimens (n)	*Rickettsia* (n)	Adjusted *Rickettsia* prevalence
Hymenoptera	23,873	74.60%	85,915	294	0.34%
Diptera	10,062	93.30%	36,211	187	0.52%
Hemiptera	9,098	68.00%	32,742	177	0.54%
Arachnida	3,626	67.20%	13,049	8	0.06%
Coleoptera	3,544	83.80%	12,754	40	0.31%
Psocodea	736	88.60%	2,649	40	1.51%
Thysanoptera	471	79.40%	1,695	1	0.06%
Trichoptera	49	91.80%	176	4	2.27%
Dermaptera	16	56.30%	58	1	1.72%

Table 5.5. The barcoding success rate of taxa which gave at least one bacteria *COI* inadvertent amplification (N=51,475 accessible specimens) with an adjusted *Rickettsia* prevalence based on the total number of arthropods in project (N=187,250 accessible + inaccessible).

Dermaptera (1.72%) and Psocodea (1.51%) were the most likely taxa to give an inadvertent *Rickettsia* amplification. Despite Hemiptera and Diptera having a similar estimated prevalence (0.54% and 0.52%), Hemiptera were much more likely to fail to barcode (67.2% vs 93.3%) indicating the true Dipteran prevalence is likely to be higher, as a barcoding failure is necessary to amplify non-target bacteria *COI*.

5.5 Discussion

The utility of DNA barcoding as a technique is highlighted by its widespread use in taxonomic and other biodiversity studies (Goldstein and DeSalle 2019). However, this study demonstrates that the integrity of this approach can be compromised through the erroneous amplification of *Rickettsia* DNA. For such a non-target amplification to occur, two criteria must be fulfilled:

1) An unintended target must complement barcoding primer sites and

2) The unintended target must hybridise in preference to the intended target during PCR cycling.

With respect to the first, the few bacteria likely to share homology with mitochondria are the Rickettsiales order of alphaproteobacteria. The hypothetical bacterial ancestor of mitochondria (proto-mitochondria) is thought to be phylogenetically embedded within the Rickettsiales (Sassera *et al*. 2011; Ferla *et al*. 2013; Wang and Wu 2015), although its evolutionary origin is still debated (Martijn *et al.*, 2018; Muñoz-Gómez *et al.*, 2019). Regarding the second, if both target and non-target *COI* can hybridise during amplification, then annealing steps early on in PCR cycling are critical in subsequent preferential amplification. It is likely that this amplification preference is determined by gene copy number (Linares *et al.*, 2009) but it has also been suggested to be affected by DNA extraction protocols or variety of *Taq* polymerase utilised (Smith *et al.*, 2012). Despite this, the amplification of more than one *COI* gene often results in “messy” chromatograms during Sanger sequencing. Subsequently, the clean traces of non-target *COI*, as observed in this study, are likely to be the result of initial primer mismatching with the intended host target.

Due to its abundance across arthropod taxa (Hilgenboecker *et al*. 2008; Zug and Hammerstein 2012), many previous reports have proposed *Wolbachia* (Rickettsiales: Anaplasmataceae) to be the main bacterial contaminant associated with barcoding. Smith *et al*. (2012) demonstrated *Wolbachia* was present in multiple BOLD *COI* traces related to insects although, more often than not, endosymbiont presence did not interfere with barcoding. However, other Rickettsiales, namely from the family Rickettsiaceae, have not been considered as significant by-products until recently (Řezáč *et al.*, 2014; Ceccarelli, Haddad and Ramírez, 2016). In a demonstration of how unintended *Rickettsia* amplifications can affect phylogeographic studies relying on DNA barcoding, a *Rickettsia COI* was conflated with a species of freshwater amphipod, *Paracalliope fluvitalis* (Lagrue *et al.*, 2016). Subsequently, supposed unique mtDNA haplotypes were allocated to a particular collection site, whereas this merely demonstrated the presence of *Rickettsia* in the lake. A consequence of this misamplification is that the large-scale use of DNA barcoding can allow for the uncovering of a large, and previously unidentified, host range of *Rickettsia*. Therefore, the *Rickettsia* hits identified from BOLD in this study has led to a unique opportunity to further investigate both host and bacteria diversity.

In the current study, *Wolbachia* and *Rickettsia* accounted for most bacteria contaminants resulting from DNA barcoding (>94%). Notably, *Wolbachia* from supergroup A was the main subgroup of the genus identified in accordance with the previous study by Smith *et al*. (2012). Additionally, our 0.165% prevalence estimate of barcoded *Wolbachia* is the same as the frequency estimated in the same study. However, when comparing the two alphaproteobacteria, there were over twice as many *Rickettsia* contaminants (n=753) than

Wolbachia (n=306). The global estimates of arthropod *Wolbachia* and *Rickettsia* host species when compared in the same analysis (Weinert *et al.*, 2015) are given as between 48-57% and 20-42% respectively, leading to the question: why is *Rickettsia* over-represented through barcoding? There are several possibilities:

1) The projects from where the contaminants originated contained more arthropods harbouring *Rickettsia* than *Wolbachia*;

2) *Rickettsia*-containing taxa are more likely to fail to barcode leading to an increased likelihood of bacteria-*COI* amplification;

3) *Rickettsia* strains are a better match to barcoding priming sites than *Wolbachia*;

4) Tissues used for DNA barcoding generally contain more *Rickettsia*.

Unfortunately, an accurate assessment of 1 and 2 cannot be made without the calculation of a true endosymbiont prevalence through a targeted screen. However, neither of these taxa biases appear likely: despite there being fewer *Wolbachia COI* sequences from Hymenoptera (the most represented insect order) than *Rickettsia* detected in this study, *Wolbachia*-infection is deemed to be much more common in Hymenoptera overall (Ahmed *et al.*, 2013; Weinert *et al.*, 2015).

Subsequently, primer and tissue biases are both possible explanations for the large number of *Rickettsia* observed. When assessing primer site homology with *Rickettsia* and *Wolbachia* genomes, the Torix *Rickettsia* from *Culicoides newsteadi* N5 (Pilgrim *et al.*, 2017) was the only sequence to not contain a SNP at the 3' binding site of the forward primer, from the primer set CLepFolF/CLepFolR, responsible for a majority of attempted barcodes in the project (Table

5.4.1.). As the 3' priming site is responsible for efficient primer annealing, and most of the *Rickettsia* amplified belonged to the Torix group, then this could account, at least partially, for why *Rickettsia* is overrepresented in DNA barcoding. Furthermore, as arthropod legs are generally used in barcoding, another explanation is that *Rickettsia* is more common in somatic tissues when compared to *Wolbachia*. There are multiple examples of both *Wolbachia* and *Rickettsia* tropisms to various somatic tissues (Hurst *et al*. 1996; Dobson *et al*. 1999; Brumin, Levy and Ghanim, 2012; Pietri *et al*. 2016; chapter 4). However, there are no recorded examples of *Rickettsia* exclusively residing in reproductive tissues; in contrast, some *Wolbachia* strains have been known to exclusively associate with ovaries (Cheng *et al.*, 2000; Zouache *et al.*, 2009). Regardless, as titres are likely to be higher for both endosymbionts in the abdomen, DNA barcoding is likely to underestimate infection rates for both bacteria.

Just as there is a bias for *Rickettsia* amplification over *Wolbachia*, there is an overrepresentation of Torix group *Rickettsia* within the genus (Figure 5.3.1.). Again, both primer and taxa biases could account for this. Indeed, the same SNP suggested to hinder primer annealing in *Wolbachia,* above, is present in all the major *Rickettsia* groups except Torix. An alternative explanation is that blood-feeding taxa, strongly associated with derived *Rickettsia* clades, are underrepresented in the project although the Belli and Transitional groups containing non-blood feeders are also in a minority. It is also possible that ancestral group *Rickettsia*, such as Torix, are simply more common than these derived clades. Certainly, a focus on derived *Rickettsia* clades was initially driven by medically important strains (e.g. *Rickettsia prowazekii*), with subsequent studies showing that the Torix group had been neglected (Perlman, Hunter and Zchori-Fein, 2006; Weinert *et al.*, 2009; Pilgrim *et al.*, 2017).

Overall, Hymenoptera, Diptera and Hemiptera were the three most represented taxa associated with *Rickettsia COI* amplification (87.4%); with also a majority (77.9%) of the project being accounted for by these three orders. Subsequently, it is unlikely that there is a strong inherent bias for these host taxa in comparison to other insects identified as hosts in this study. Indeed, when adjusted for relative prevalence, Psocodea and Trichoptera were estimated to be the most likely taxa to harbour *Rickettsia* (Table 5.5.). Furthermore, by calculating barcode success rate at an order level, Hemiptera were deemed to fail barcoding (either lack of amplification and/or quality sequence) more so than Diptera despite having a similar adjusted prevalence. As an increased barcoding failure rate is associated with non-target *COI* amplification, it is likely there are more Torix *Rickettsia*-associated Diptera than Hemiptera in the BOLD database. This is in accordance with targeted *Rickettsia* screens which have demonstrated mostly Dipteran hosts in the Torix group (Vandekerckhove *et al*. 2005; Reeves *et al*. 2008; Weinert *et al*. 2009; Martin *et al*. 2013; Pilgrim *et al*. 2017).

Although *Rickettsia* will only interfere with barcoding in a minority of cases (~0.4%), it is likely that the redesign of screening primers for some taxa harbouring the endosymbiont will need to be considered. For example, a colleague (Panupong Thongprem, University of Liverpool) demonstrated a case of *COI* primers (LCO1490/HCO2198; Folmer *et al*. 1994) amplifying *Rickettsia* in a bed bug *Cimex lectularius* population, which is polymorphically infected. Based on gel visualisation of amplicons, no *COI* amplification was detected in *Rickettsia* uninfected individuals and a weak band was observed in *Rickettsia* infected individuals. On sequencing this was confirmed as Leech group *Rickettsia*. It was then possible to rescue mtDNA amplification by using the alternative set of primers used in this study (C1-J-1718/HCO1490;

Table 5.2.) to re-barcode "contaminant" specimens. In a demonstration of the challenges faced when redesigning primers to be taxa-specific, Hebert *et al*. (2004) established the LepF1/LepR1 primer set to assist in increasing the barcoding specificity of targeting Lepidoptera. Despite this, not only have this primer set since been used widely for a variety of arthropod taxa (Smith *et al.*, 2012), subsequent *Wolbachia* amplification has often been described (Linares *et al.*, 2009; Smith *et al.*, 2012).

As the *in silico* and empirical evidence suggests *Rickettsia COI* amplification is not uncommon, why has this phenomenon not been described more widely before? The conduction of a previous large-scale non-target *COI* study using BOLD specimens (Smith *et al.*, 2012), revealed only *Wolbachia* hits through comparison with a *Wolbachia*-specific reference library and was thus likely to miss *Rickettsia*. Additionally, there has been a lack of Torix *Rickettsia COI* homologues to compare barcodes to until recently, where a multilocus identification system, including *COI* was described (chapter 3). Indeed, out of the contaminant dataset received in this study, some of the *Rickettsia* contaminants were tentatively described by BOLD as *Wolbachia* probably due to the absence of publicly available *Rickettsia COI* to compare.

A further important finding from this study is the uncovered host diversity of *Rickettsia* through DNA barcoding. As the host effects, especially of Torix group *Rickettsia*, are unknown, this leads to a great opportunity to identify model organisms to assess the effects of the bacteria on arthropods and vice-versa. At the order level, previously unidentified Torix *Rickettsia*-associated taxa detected from the BOLD database included Trichoptera (caddis flies), Trombidiformes (water mites), Thysanoptera (thrips), Diplopoda (millipedes) and

Dermaptera (earwigs). Aside from insects and arachnids, the identification of a *Rickettsia COI* from the freshwater amphipod, *Paracalliope fluvitalis* in Genbank (Lagrue *et al.*, 2016) suggests the Malacostraca should also be further investigated for their role in *Rickettsia* ecology. Caution needs to be taken when interpreting what these newly found associations mean, as mere presence of *Rickettsia* DNA does not definitively indicate an endosymbiotic association. Indeed, parasitism or ingestion of symbiont-infected biota can also result in PCR detection (Plantard *et al.*, 2012; Le Clec'h *et al.*, 2013; Ramage *et al.*, 2017).

The above taxa join the previously known Torix *Rickettsia*-infected arthropods of Diptera, Hemiptera, Hymenoptera, Coleoptera, Psocodea and Araneae (Goodacre *et al*. 2006; Perotti *et al*. 2006; Küchler *et al*. 2009; Machtelinckx *et al*. 2012; Martin *et al*. 2013; Gualtieri *et al*. 2017). However, the ability to re-barcode some of the DNA extracts responsible for *Rickettsia COI* amplification made it possible to refine the taxonomy of hosts even further. Taxa of note included detritivores residing in woodlands, such as Sciaridae (Dark-winged fungus-gnat), Mycetophilidae (Fungus-gnats), *Philotarsus californicus* (Barklice), and the tree-hole dwelling beetle *Prionocyphon limbatus* (Figure 5.6. and Appendix 3) indicating a novel ecological niche for Torix-*Rickettsia* which has primarily been associated with taxa residing near rivers. Furthermore, the discovery of new terrestrial hosts such as, Aurrenchoryna (Hemiptera) and parasitoid wasps confirm Torix *Rickettsia* to be an ecological component to be investigated in both terrestrial and aquatic ecosystems.

The varied ecological niches of Torix *Rickettsia*, alongside sporadic distribution of the host taxa across the *Rickettsia* phylogeny (indicating frequent host shifts), suggests environmental

sources of this endosymbiont should be considered. Many parasitoid wasps have been associated with facilitating such novel bacteria-host associations (Vavre *et al.*, 1999; Huigens *et al.*, 2004; Johannesen, 2017). Specifically, parasitoids have been considered as phoretic vectors of endosymbionts via sequential stabbing of an infected host and then an uninfected host (Gehrer and Vorburger 2012; Ahmed *et al*. 2015), although sustained transmission from host to wasp has also been suggested (Vavre *et al.*, 1999; Chiel *et al.*, 2009; Johannesen, 2017). Thus, it is unsurprising to find many *Rickettsia*-associated parasitoids present in this study, which may be acting as an intermediary for the endosymbiont's promiscuity. The most common hymenopteran identified from the targeted re-barcoding is from the Diapriidae wasp family known to primarily infect various Dipterans. Several Torix-associated families of the true flies, which have been known to act as hosts to these wasps include the Ceratopogonidae (e.g. *Culicoides* spp.), Tabanidae, Mycetophilidae, Sciaridae, Chironomidae and Psilidae (Wright, Geering and Ashby, 1947; Nixon, 1957, 1980; Wild, 1972; Hellqvist, 1994). Notably, one of the few taxa identified as an alternative host to flies is the rove beetle (Staphylinidae) (Nixon, 1980), another *Rickettsia*-infected taxon recognised in this study. The *Rickettsia* presence in the Platygastridae family of parasitoid wasps is also intriguing because the major subfamily, Platygastrinae, are often coupled with Torix-associated gall midges (Cecidomyiidae) (Austin, Johnson and Dowton, 2005). Finally, nearly half of the hosts of another *Rickettsia*-associated parasitoid family, Mymaridae, parasitise Auchenorrhyncha hemipterans (Huber, 1986), which are over-represented in this study.

Despite endoparastioids clearly being a hotspot for *Rickettsia* association, it is also possible for alternative routes such as predation (Le Clec'h *et al.*, 2013) and phloem-feeding (Caspi-

Fluger *et al.*, 2012; Li *et al.*, 2017) to allow for horizontal transmission. Phloem-feeding Psyllids, identified as *Rickettsia*-infected in this study, are a candidate for such future investigations. Recently, there has been increasing interest in the role plants play in endosymbiont transmission (Reviewed by Chrostek *et al*. 2017). Circumstantial evidence of herbivorous insects sharing common plants and symbionts (Sintupachee *et al.*, 2006; Stahlhut *et al.*, 2010; Morrow *et al.*, 2014) has been superseded by direct evidence of sustained *Wolbachia* and *Serratia* transmission in the adult progeny of *Bemisia* whiteflies and *Aphis fabae* aphids infected by phloem-feeding (Li *et al.*, 2017; Pons *et al.*, 2019). Additionally, other endosymbionts, including *Rickettsia*, *Cardinium* and *Arsenophonus*, have been found in phloem (Bressan, Terlizzi and Credi, 2012; Caspi-Fluger *et al.*, 2012; Gonella *et al.*, 2015) increasing the likelihood of widespread symbiont transmission between phloem-feeders.

The establishment of a new arthropod-endosymbiont interaction is easier if the donor and recipient share a similar host genetic background (Łukasik *et al.*, 2015). Subsequently, it is likely that the development of several horizontal transmission strategies in different ecosystems, as discussed above, increases the chances of overcoming such host shift barriers. Indeed, if multiple horizontal transmission paths do exist, this could account for the diverse plethora of infected taxa, as well as arthropods harbouring more than one strain of symbiont (Vavre *et al.*, 1999; Morrow *et al.*, 2014). In addition, shared host-shifting mechanisms support observations in this study and chapter 3, where common taxa were observed in both subclades of the Torix group; Leech and Limoniae (Figures 5.5 and 5.6). Further work is needed to assess if these subclades should be distinguished based on functionality, or if they should be classed as a single group for ease of study.

Finally, the discovery of a tabanid fly harbouring *Rickettsia* adds to the existing haematophagous taxa previously identified. Other Torix-associated blood-feeders include vectors of disease such as *Simulium* sp. (chapter 3), *Anopheles plumbeus* (unpublished data), *Phlebotomus chinensis* (Li *et al.*, 2016)*, Lutzomyia apache* (Reeves, Kato and Gilchriest, 2008), *Nosopsyllus laeviceps* (Song *et al.*, 2018) and *Culicoides* spp. (chapter 3). Additionally, this study has uncovered a *Rickettsia*-infected psyllid (*Cacopsylla melanoneura*), a major vector of phytoplasma. Subsequently, the question of *Rickettsia* vector-competence effects proposed in previous chapters of this thesis are clearly of widespread relevance.

To conclude, this study has shown that the widely-used technique of DNA barcoding can be confounded by the inadvertent amplification of *Rickettsia,* particularly of the Torix group. Not only does this have implications for the design of barcoding studies incorporating *Rickettsia*-infected taxa, but due to the unknown host effects and transmission strategies of these endosymbionts, barcoding has been shown to act as an important tool in informing future directions of investigation involving the understudied Torix *Rickettsia*.

Chapter 6: General Discussion and Future Directions

6.1 Overview

Diseases transmitted by *Culicoides* are of both animal and public health importance (Mellor, Boorman and Baylis, 2000; Sick *et al.*, 2019). Furthermore, the economic significance in both vector and pest midge species, has led to targeted research on these tiny flies. Predictive climate models indicate that continued emergence of outbreaks in previously naïve areas is likely (Kilpatrick *et al.*, 2008; Gould and Higgs, 2009; Weaver and Reisen, 2010). Additionally, there is an increasing concern of the surfacing of novel diseases, transmitted by *Culicoides*, which could affect humans (Sick *et al.*, 2019). Currently, vaccination and vector control have been the focus of such interventions. In the case of immunisation, vaccination schemes have proven to be effective in some cases, especially in the only known BTV outbreak in the UK in 2007 (Szmaragd *et al.*, 2010). However, a reliance on the knowledge of circulating virus strains, and lengthy vaccine stockpiling times are needed to ensure this strategy is effective. Vector control, on the other hand has relied on the removal of breeding sites and moving cattle into screened housing, which have logistical and farmer compliance issues, and has proven ineffective (Dwyer *et al.*, 2007; Carpenter, Mellor and Torr, 2008). Subsequently, the need for novel control interventions is necessary.

Vector control has advanced considerably over the past decade, with the release of transgenic and sterilised mosquitoes garnering much attention (Flores and O'Neill, 2018). Success stories in field trials also include endosymbiont-based strategies. After the introduction of *Aedes aegypti* mosquitoes containing the *w*Mel strain of *Wolbachia*, no new cases of dengue have been observed in Northern Australia (Ritchie, 2018). Since this initial program, concurrent efforts have been expanded to 21 countries spanning Asia and the Americas (World Mosquito

Program, 2019). Importantly, this includes a randomised control trial in Indonesia, where the large-scale release of *Wolbachia*-infected *Ae. aegypti* is being undertaken, with an aim to measure dengue fever incidence reduction in treated versus untreated areas (Anders *et al.*, 2018). This contrasts with the original Australian initiative where no controls were utilised (O'Neill, 2012). Importantly, as symbionts are naturally occurring and ubiquitous, public acceptance of these *Wolbachia*-based initiatives are generally seen as more acceptable than other genetic modification interventions (Flores and O'Neill, 2018).

Considering the important contribution of symbionts to the control of vector competence, there is a lack of work relating to the symbionts of biting midges. Previous studies have uncovered *Wolbachia* and *Cardinium* in populations from around the world, but host effects are unknown and studies have been limited by the difficult initiation and maintenance of midge colonies (Nakamura *et al.*, 2009; Morag *et al.*, 2012; Lewis *et al.*, 2014; Mee *et al.*, 2015; Pagès *et al.*, 2017). Subsequently, this thesis aimed to develop our understanding of known midge symbionts and their potential influences on host biology. Listed below is a summary of findings and their significance, followed by suggested future research directions.

6.2 The re-examination of the *Cardinium* infection status of *Culicoides*

The presence of low-titre *Cardinium* infections in chapter 2 highlights the need for additional studies to assess potential host effects conferred by *Cardinium*. For example, low-titre *Wolbachia* infections in *Drosophila paulistorum* can lead to bi-directional cytoplasmic incompatibility (CI), suggesting these symbiont interactions could be missed by non-sensitive assays such as conventional PCR (Miller, Ehrman and Schneider, 2010). It is conceivable that

a CI phenotype involving *Cardinium* will also be determined by symbiont load. Subsequently, when *Cardinium* effects on their midge hosts are formally identified, research may move away from investigating only infections with high symbiont titres. Therefore, the utilisation of nested PCR provides an efficient methodology to profile such low-tire infections in a range of *Culicoides* species.

The clarification of a low *Cardinium* infection rate in *C. pulicaris* is important due to this species being previously considered as a candidate for future *Cardinium* work as a result of erroneously inferred high *Cardinium* prevalence (Lewis *et al.*, 2014). Furthermore, due to the limited numbers of wild *C. punctatus* (infected at high prevalence) inside and outside of the UK, advancements in ascertaining the endosymbiont's effects on *Culicoides* biology are likely to derive from more common vector species infected with *Cardinium*, such as *C. imicola*. Crucially, for further investigations into *Cardinium* interactions, problems in overcoming the rearing and maintaining of colonies must also be achieved (Nayduch *et al.*, 2014).

6.3 The role of *Rickettsia* in *Culicoides* biology

A targeted screening revealed Limoniae *Rickettsia* reached high frequencies in midge populations and was present in 37% of the species tested. Importantly, infections were observed in the BTV vector species *C. pulicaris* and *C. newsteadi* complexes. Shortly after the publication of this data (Pilgrim *et al.*, 2017), *Rickettsia* was identified in the Australasian vector *C. brevitarsis*, during a metagenomic screen of guts (Mee, 2017). Furthermore, in a separate study, *Rickettsia* was identified in the obsoletus complex and *C. sonorensis* (Möhlmann, 2019). Overall, the high frequency of Limoniae *Rickettsia* infections identified in

these studies and chapter 3, suggest this is an important, but previously unrecognized, component of the biology of *Culicoides*. Although *Rickettsia* are known to have both reproductive and fitness benefits (Hagimori *et al.*, 2006; Giorgini *et al.*, 2010; Himler *et al.*, 2011), the impact of Limoniae *Rickettsia* on host biology is uncertain. Of relevance to *Culicoides* are potential effects relating to B vitamin provisioning, vectorial capacity, dispersal behaviour and reproduction manipulation. These are now discussed below:

Symbiont provisioning of B vitamins has been recorded in several blood feeding insects including *Wigglesworthia* in Tsetse flies and *Wolbachia* in bed bugs (Snyder *et al.*, 2010; Nikoh *et al.*, 2014). However, this appears to be unlikely for *Rickettsia* and *Culicoides*; with the exception of a reduced pathway for folate biosynthesis, no known pathways for the biosynthesis of cofactors and B-vitamins have been observed through genomic analysis (Pilgrim *et al.*, 2017). Furthermore, as only females require a blood meal to reproduce, it is likely that B vitamin supplementation may occur at different life stages through other diet components, such as from nectar (adults) or detritus/small arthropods (larvae).

Symbiont influences on vector competence have focussed predominantly on *Wolbachia* and mosquitoes due to their virus blocking effect (Bian *et al.*, 2010; Blagrove *et al.*, 2012; Hussain *et al.*, 2012; van den Hurk *et al.*, 2012; Chouin-Carneiro *et al.*, 2019). However, other endosymbionts are likely to have similar modifying effects. Indeed, the presence of *Rickettsia* in the fat body of *Culicoides* is of potential relevance to vector competence as BTV and epizootic haemorrhagic disease (EHDV) of ruminants replicate in this tissue (Fu *et al.*, 1999;

Mills *et al.*, 2017b) before travelling to the salivary glands. This suggests that direct interactions between *Rickettsia* and the virus could be occurring.

The natural *Rickettsia* infections observed in *Culicoides* vector species (*C. obsoletus, C. sonorensis, C. pulicaris, C. newsteadi*) contrasts with *Aedes aegypti* populations, where natural *Wolbachia* infections are absent or present at low prevalence (Gloria-Soria, Chiodo and Powell, 2018; Kulkarni *et al.*, 2019). Endosymbiont transinfections usually leads to a higher virus blocking impact compared to natural infections possibly due to the evolution of immune tolerance to the naturally occurring symbionts (Sinkins, 2013). For example, the transinfection of *w*Mel in *Aedes albopictus* gave a reduced virus blocking effect when compared to the same artificial infection created in *Aedes aegypti* (Blagrove *et al.*, 2012) and was deemed to be as a result of *Aedes albopictus'* immunotolerance to *Wolbachia* (*Ae. albopictus* harbours a natural *Wolbachia* double infection). Despite this, immune priming has been shown to only partially account for virus blocking with perturbation of intracellular trafficking and competition for resources also being hypothesised to account for virus load decreases (Caragata *et al.*, 2013; Geoghegan *et al.*, 2017; Schultz *et al.*, 2017). Overall, this implies that natural infections have the potential to alter arbovirus transmission dynamics.

Conversely, *Wolbachia* infections have also been shown to lead to a virus potentiating effects in insects (Graham *et al.*, 2012; Dodson *et al.*, 2014). Although the mechanism of this virus titre alteration is unknown, natural *Rickettsia* infections appear to also lead to a similar effect with an increase in vector competence observed in *Bemisia* white flies transmitting tomato yellow leaf curl virus (Kliot *et al.*, 2014). Whether there is a virus blocking or potentiating

effect induced by *Rickettsia* in *Culicoides*, both are of interest for predicting outcomes of future midge-borne diseases.

The presence of *Rickettsia* in the ovaries of *C. impunctatus* suggests potential reproductive effects could be occurring. *Rickettsia* infections are known to be associated with a variety of reproductive manipulations of their host, including male-killing in ladybird beetles (Werren *et al.*, 1994) and parthenogenesis induction in parasitoid wasps (Hagimori *et al.*, 2006; Giorgini *et al.*, 2010). However, equal likelihood of male and female midges being infected indicates sex ratio distortion is unlikely to be a phenotype for the *Rickettsia* in midges. The autogenous nature of some *Culicoides* species including *C. impunctatus* (Boorman and Goodard, 1970), suggests *Rickettsia* could be facilitating the initial gonotrophic cycle of these species. Blood-feeding is a costly behaviour due to the dangers that midges encounter when straying from sheltered habitats. Subsequently, facilitation of oogenesis without a blood meal would increase host fitness (Bradshaw *et al.*, 2018).

6.4 *Rickettsia* transmission strategies

As well as informing host phenotypic effects, investigating *Rickettsia* localisation to tissues can help elucidate transmission strategies of the bacteria. The discovery of somatic and germline infections corroborate previous studies of *Rickettsia* (Gottlieb *et al.*, 2006; Perotti *et al.*, 2006; Küchler, Kehl and Dettner, 2009; Brumin, Levy and Ghanim, 2012). Presence in the ovaries and larvae suggests high fidelity maternal transmission and substantiates the findings of high prevalence in *Culicoides* populations from chapter 3. In the leaf hopper *Nephotettix*

cincticeps (Watanabe *et al.*, 2014), paternal transmission is thought to assist in driving a costly symbiont into a new population. However, there was no intrasperm infection observed in *C. impunctatus* spermatozoa through TEM. To truly confirm a lack of paternal transmission, crossing experiments are needed, although the absence of an available *C. impunctatus* colony makes this problematic. However, a recent crossing study in the bed bug *Cimex lectularius* infected with Leech group *Rickettsia* revealed a lack of paternal transmission (Panupong Thongprem, University of Liverpool; unpublished data). This suggests paternal transmission of Torix group *Rickettsia* are likely to be atypical of the clade.

The infection of the ovarian suspensory ligament (OSL) in *C. impuncatus* is intriguing, as it offers a novel route of germline targeting. As a connective tissue attaching the ovaries to various insect tissues including the body wall and fat body (Storto, 1994; Szklarzewicz *et al.*, 2007), the OSL is an obvious target for symbionts requiring localisation to the ovaries. The OSL's continuation with the ovarian epithelial sheath suggests this is an efficient route of targeting all developing egg chambers. Despite this, a time series is needed to definitively confirm the infection of the OSL precedes the infection of ovaries. The follicular epithelium infection observed in *C. impunctatus* is reminiscent of *Adalia* ladybird and *Bemisia* whitefly infections (Sokolova, Zinkevich and Zakharov, 2002; Brumin, Levy and Ghanim, 2012), although in these cases it is suggested to be as a result of infected haemocytes transferring the symbiont from haemolymph to follicular cells (Chen, Campbell and Purcell, 1996; Hurst, Walker and Majerus, 1996). Some studies have also proposed bacteriocytes as a means of transovarial transmission for *Rickettsia* (Gottlieb *et al.*, 2006, 2008) in *Bemisia*, although further investigation suggested this was not a primary route of ovary-targeting in all cases

(Brumin, Levy and Ghanim, 2012). Infection of connective tissue directly linking the germline, as observed in this thesis, suggests there may be multiple routes of germline targeting dependant on *Rickettsia* and host genetic background. This is certainly the case in *Wolbachia* where access to ovaries can be via somatic or germline stem cells depending on symbiont strain (Frydman *et al.*, 2006; Serbus and Sullivan, 2007; Toomey *et al.,* 2013).

The infection of somatic tissues can also offer insights into the evolutionary history of the symbiont's past transmission routes. For example, *Wolbachia* infections of *Drosophila melanogaster* in somatic tissues can replicate after microinjection (Frydman *et al.*, 2006). This has been proposed to offer a site of stable replication and proliferation for a newly introduced symbiont in a naïve host, before targeting oocytes. In the case of *C. impunctatus* the attachment site is the anterior/posterior midgut junction which suggests possible past ingestion of *Rickettsia*-infected taxa could be followed by the passage into the suspensory ligament and the germline. Another possible route of host shifting is via entry into the fat body through parasitism. Although fat body infection was observed in larval stages, this was difficult to assess in adults; only parous individuals were available for evaluation in the current study which have depleted fat bodies. Subsequently, nulliparous females will be needed in the future to confirm adult fat body infection.

Moreover, insights into transmission dynamics can be provided by phylogenetic information. Obligate symbionts, which share a long evolutionary history with their hosts, will have congruent phylogenies as a result of co-cladogenesis (Ferrari and Vavre, 2011). On the other hand, where bacteria and host phylogenies do not mirror each other, this suggests frequent

host shifts are likely to occur (Turelli *et al.*, 2018). The phylogenies of Torix *Rickettsia* generated in chapters 3 and 5 show a clear lack of clustering between strains and host taxonomy, indicating host shifts are rampant. There are several possible environmental sources to initiate novel *Rickettsia*-host interactions, with the various lifestyles of Torix hosts suggesting multiple horizontal transmission paths could exist. These include parasitic wasps, or routes via ingestion of phloem or *Rickettsia*-infected prey. Parasitoids have been considered as phoretic vectors of endosymbionts via sequential stabbing of an infected host and then an uninfected host (Gehrer and Vorburger, 2012; Ahmed *et al.*, 2015). The uncovering of *Rickettsia* associations in several wasp families, and their hosts, suggests this is a viable route. However, it has been noted previously, that because many successful parasitisms lead to the death of the host, this is likely to hinder a symbiont host shift (Vavre *et al.*, 1999). This project has also unveiled phloem-feeders (Hemiptera; Psyllidae) associated with Torix-*Rickettsia* infection, suggesting a further potential horizontal transmission route. A Belli group *Rickettsia* has previously been transmitted from infected *Bemisia* white flies to uninfected individuals via phloem (Caspi-Fluger *et al.*, 2012) increasing the likelihood that this could be common across the genus.

6.5 *Rickettsia* and *Cardinium* confounding of DNA barcoding

Endosymbionts can impact the interpretation of DNA barcoding studies in two ways:

1) Through the co-inheritance of mtDNA and symbionts leading to difficulties in interpreting mtDNA divergence patterns (Hurst and Jiggins, 2005).

2) Via the inadvertent amplification of endosymbionts instead of the targeted host arthropod mitochondria (Smith *et al.*, 2012; Ceccarelli, Haddad and Ramírez, 2016).

The first of these is observed in chapter 2, relating to *Cardinium*-infected *Culicoides imicola*, the primary vector of BTV in Afrotropical regions. The *Cardinium* infection status of this midge species is intriguing due to its presence and absence across the Mediterranean basin (Morag *et al.*, 2012; Pagès *et al.*, 2017). The linkage disequilibrium of *Cardinium* and mitochondria suggests the dispersal of *Cardinium*-infected *C. imicola* into the Eastern Mediterranean basin (EMB) but not the Western basin (WMB) from Africa. Although this makes mtDNA useful for understanding symbiont gene flow, the selective sweeps of symbionts mean interpretation of host gene flow is difficult. For example, if a *Cardinium* sweep homogenises the mtDNA diversity in a region where a recent incursion has occurred, then it wipes out previous gene flow patterns which can no longer be interpreted (Hurst and Jiggins, 2005). This suggests that a non-biased marker is needed for biogeographic studies aiming to evaluate the population structure of *C. imicola*. This is important for understanding *C. imicola* dispersal, which has come under scrutiny previously due to their incursions into Europe leading to the spread of BTV in serologically naïve animals (Mellor *et al.*, 1985; J. Boorman, 1986b).

The inadvertent amplification of *Rickettsia* DNA, as opposed to host DNA, was observed in Chapter 5. Previously, *Wolbachia* has been deemed the most important bacterial contaminant confounding this method (Smith *et al.*, 2012). However, the demonstration that *Rickettsia* is over twice as likely to be amplified than *Wolbachia* suggests the former is more important to consider when designing barcoding studies. Related to this, the overrepresentation of erroneous Torix *Rickettsia* amplification could be rectified by using

specifically designed primers as opposed to commonly used universal primers. The use of the primer set C1-J-1718/HCO1490 managed to rescue the amplification of mtDNA in problematic host taxa and so should be considered as candidate primers for future studies. Despite this, the inadvertent amplification of Torix *Rickettsia* has had the unintended benefit of expanding our knowledge of this clade of the genus. For example, the uncovering of tabanid flies and psyllids add to existing knowledge of Torix-*Rickettsia* associated vectors including: *Nosopsyllus* fleas (Song *et al.*, 2018); *Lutzomyia* sand flies (Reeves, Kato and Gilchriest, 2008); *Simulium* black flies (Panupong Thongprem, University of Liverpool; unpublished data) and *Culicoides* biting midges. In addition, through a recent targeted screening undertaken by the author, the mosquito *Anopheles plumbeus* has been identified as harbouring Torix *Rickettsia*.

6.6 Future directions

The widespread presence of Limoniae *Rickettsia* in *Culicoides* is of great significance due to their potential to affect vectorial capacity. Considering this, the identification of a suitable model organism and system is a priority. The lack of available biting midge colonies is a clear hinderance in this instance, with only *C. sonorensis* and *C. nubeculosus* publicly available (Nayduch *et al.*, 2014); only the former of which is a BTV vector. This study failed to find Limoniae *Rickettsia* in *C. sonorensis* wild populations and the extant "AA" colony deriving from Californian populations (held at the Pirbright institute; Surrey, UK). However, *Rickettsia* have been identified in a colony originating from *C. sonorensis* in Colorado (held at the ARS-USDA; Kansas, USA) (Möhlmann, 2019). This colony is polymorphically infected (40% positive), meaning that curing via antimicrobials is not necessary to obtain negative controls for any future studies. In addition, *C. sonorensis* is the gold standard for work investigating BTV infection dynamics (Nayduch *et al.*, 2014), making this the most promising avenue for future

work pertaining to vector competence involving *Rickettsia*. At the time of writing (January 2020), the author is in contact with the USDA to negotiate working with their *Rickettsia* infected *C. sonorensis* colony.

A future focus on both indirect and direct effects on vector competence should be considered. A comparative transcriptomic study of *Rickettsia* positive and negative *C. sonorensis* would allow for the identification of immune genes possibly responsible for influencing viral titres. The presence of *Rickettsia* in fat bodies, and the perceived relevance of virus maintenance as a result of intracellular cholesterol modification, suggests direct resource competition may be occurring between symbiont and virus (Caragata *et al.*, 2013; Sinkins, 2013; Frentiu, 2017; Geoghegan *et al.*, 2017). Therefore, cholesterol quantification assays, in a *Rickettsia* infected cell line, would allow for the assessment of lipid utilisation by the symbiont (Geoghegan *et al.*, 2017). *Rickettsia-Culicoides* interactions should not be limited to investigations solely relating to vector competence. Cytoplasmic incompatibility, which has evolved independently at least twice in the symbionts *Cardinium* and *Wolbachia* (Werren and O'Neill, 1997; Gotoh, Noda and Ito, 2007), should be investigated in Limoniae *Rickettsia*. This is relevant to both vector and pest species as a population suppression technique (through the release of infected males; Mains *et al.*, 2019) as well as a mechanism to drive the symbiont into an uninfected target population (through the release of infected males and females; Hoffmann *et al.*, 2011).

The interactions between gut microbiota and *Rickettsia* should also be considered in futures studies due to the numerous effects gut bacteria can have on host biology, including the

protection against pathogens (Cirimotich, Ramirez and Dimopoulos, 2011). However, studies (Wong *et al.*, 2011; Rottschaefer and Lazzaro, 2012) have mostly focussed on endosymbiont effects on pathogenic bacteria challenges, as opposed to already established gut flora. Ye *et al.* (2017) identified that *Drosophila* gut microbiota diversity can be reduced in the presence of *Wolbachia* although the causative mechanism has still not been elucidated (Simhadri *et al.*, 2017). Specifically, relating to *Culicoides* endosymbionts, *Cardinium* does not appear to correlate with any broad changes in the taxonomic diversity of gut bacteria (Mee, 2017). Despite this, *Rickettsia*'s effect on midge gut microbiota are still to be elucidated and should be considered once a suitable system of investigation is in place.

The infection rates of both *Rickettsia* and *Cardinium* described in this thesis should be met with caution as facultative endosymbiont prevalences are known to fluctuate with time. For example, *Bemisia* white fly *Rickettsia* infections in the USA were observed to reach fixation from 1% prevalence from 2000 to 2006, before recently being observed at 37% prevalence in 2017 (Bockoven *et al.*, 2019). This endosymbiont fluctuation can be attributed to abiotic factors such as temperature tolerance, as well as host-specific factors or parasitoid pressure (Bordenstein and Bordenstein, 2011; Smith *et al.*, 2015; Bockoven *et al.*, 2019). Therefore, the monitoring of both symbiont prevalence and titre fluctuation over time should continue in field *Culicoides*. Additionally, studies should also consider *Culicoides*' endosymbiont co-infections, which can lead to competition and reduction/clearance of one competitor (Jaenike, 2009; Kriesner *et al.*, 2013; Rossi *et al.*, 2015). Subsequently, *C. bysta* and *C. newsteadi* N2 are both candidates for the monitoring of such symbiont dynamics, due to their harbouring of both *Rickettsia* and *Cardinium*.

Finally, a focus should be put on the economically important pest species *C. impunctatus*, with *Rickettsia* presence likely to be of great interest to the Scottish tourism and forestry industries (Hendry and Godwin, 1988; Hendry, 1996). Unfortunately, the two past attempts to investigate *C. impunctatus* through initiating colonies have failed (Hill, 1947; Carpenter, 2001). In both previous cases, rearing to pupation either failed or life stages were not visible to assess development. Work undertaken in this thesis appears to have overcome the problem of provisioning larvae with a suitable diet to achieve pupation. Although only a minority developed to pupation, this indicates progress towards the goal of initiating a sustainable colony. The establishment of a future *C. impunctatus* colony will allow for the investigation of *Rickettsia* effects on reproductive parameters. Specifically, future work should focus on the potential for *Rickettsia* to influence fecundity and autogeny. Indeed, *Rickettsia felis* has been shown necessary for oogenesis in the booklouse *Liposcelis bostrychophila* (Perotti *et al.*, 2006).

Bibliography

Aguila, J.R. *et al*., 2007. The role of larval fat cells in adult *Drosophila melanogaster*. *Journal of Experimental Biology*, 210(6), pp.956–963.

Ahmed, M.Z. *et al*., 2013. Consistently high incidence of *Wolbachia* in global fig wasp communities. *Ecological Entomology*, 38(2), pp.147–154.

Ahmed, M.Z. *et al*., 2015. The intracellular bacterium *Wolbachia* uses parasitoid wasps as phoretic vectors for efficient horizontal transmission. *PLoS Pathogens*, 11(2): e1004672.

Akman, L. *et al*., 2002. Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. *Nature genetics*, 32(3), pp.402–407.

Albertson, R. *et al*., 2013. Mapping *Wolbachia* distributions in the adult *Drosophila* brain. *Cellular Microbiology*, 15(9), pp.1527–1544.

Allen, J.M. *et al*., 2007. Evolutionary relationships of "*Candidatus* Riesia spp.," endosymbiotic Enterobacteriaceae living within hematophagous primate lice. *Applied and Environmental Microbiology*, 73(5), pp.1659–1664.

Ander, M., Troell, K. and Chirico, J., 2013. Barcoding of biting midges in the genus *Culicoides*: A tool for species determination. *Medical and Veterinary Entomology*, 27(3), pp.323–331.

Anders, K.L. *et al*., 2018. The AWED trial (Applying *Wolbachia* to Eliminate Dengue) to assess the efficacy of *Wolbachia*-infected mosquito deployments to reduce dengue incidence in Yogyakarta, Indonesia: Study protocol for a cluster randomised controlled trial. *Trials*, 19(1).

Anderson, C.R. *et al*., 1961. Oropouche virus: a new human disease agent from Trinidad, West Indies. *The American journal of tropical medicine and hygiene*, 10, pp.574–578.

Ansari, M.A., Carpenter, S. and Butt, T.M., 2010. Susceptibility of *Culicoides* biting midge larvae to the insect-pathogenic fungus, *Metarhizium anisopliae*: Prospects for bluetongue vector control. *Acta Tropica*, 113(1), pp.1–6.

Ant, T. H. *et al.,* 2018. The *Wolbachia* strain *w*Au provides highly efficient virus transmission blocking in *Aedes aegypti*. *PLoS Pathogens,* 14(1): e1006815.

Ant, T.H. *et al*., 2020. *Wolbachia* transinfections in *Culex quinquefasciatus* generate cytoplasmic incompatibility. *Insect Molecular Biology*, 29(1), pp.1-8.

Applied Maths NV, 2012. Bionumerics. Available at: http://www.applied-maths.com [Accessed 25 August, 2016].

Arrese, E.L. and Soulages, J.L., 2010. Insect Fat Body: Energy, Metabolism, and Regulation. *Annual Review of Entomology*, 55(1), pp.207–225.

Ashfaq, M., Hebert, P.D.N. and Naaum, A., 2016. DNA barcodes for bio-surveillance: Regulated and economically important arthropod plant pests. *Genome*, 59(11), pp.933–945.

Aussel, J.P. and Linley, J.R., 1994. Natural food and feeding behavior of *Culicoides furens* larvae (Diptera, Ceratopogonidae). *Journal of Medical Entomology*, 31(1), pp.99–104.

Austin, A.D., Johnson, N.F. and Dowton, M., 2005. Systematics, evolution, and biology of scelionid and platygastrid wasps. *Annual Review of Entomology*, 50(1), pp.553–582.

Bachanek-Bankowska, K. *et al*., 2009. Une avancée dans la construction d'une base de données moléculaire sur le virus de la peste équine africaine. *Revue d'élevage et de médecine vétérinaire des pays tropicaux*, 62(2–4): 174.

Baldo, L., Hotopp, J.C.D., *et al*., 2006. Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. *Applied and Environmental Microbiology*, 72(11), pp.7098–7110.

Baldo, L., Bordenstein, S., *et al*., 2006. Widespread recombination throughout *Wolbachia* genomes. *Molecular Biology and Evolution*, 23(2), pp.437–449.

Baldo, L., Lo, N. and Werren, J.H., 2005. Mosaic nature of the *Wolbachia* surface protein. *Journal of Bacteriology*, 5406–5418(187): 15.

Baldridge, G.D. *et al*., 2007. Transposon insertion reveals pRM, a plasmid of *Rickettsia monacensis*. *Applied and Environmental Microbiology*, 73(15), pp.4984–4995.

Balenghien, T. *et al*., 2014. The emergence of Schmallenberg virus across *Culicoides* communities and ecosystems in Europe. *Preventive Veterinary Medicine*, 116(4), pp.360–369.

Ballard, J.W.O. *et al*., 1996. Reduced variation in *Drosophila simulans* mitochondrial DNA. *Genetics,* 144(4), pp.1519–1528.

Barceló, C. and Miranda, M.A., 2018. Bionomics of livestock-associated *Culicoides* (biting midge) bluetongue virus vectors under laboratory conditions. *Medical and Veterinary Entomology*, 32(2), pp.216–225.

Baselga, A. *et al*., 2013. Whole-community DNA barcoding reveals a spatio-temporal continuum of biodiversity at species and genetic levels. *Nature Communications*, 4: 1892.

Batten, C.A. *et al*., 2008. A European field strain of bluetongue virus derived from two parental vaccine strains by genome segment reassortment. *Virus Research*, 137(1), pp.56–63.

Baylis, M. *et al*., 1997. The spatial and seasonal distribution of African horse sickness and its potential *Culicoides* vectors in Morocco. *Medical and Veterinary Entomology*, 203–212(11): 3.

Beckmann, J.F., Ronau, J.A. and Hochstrasser, M., 2017. A *Wolbachia* deubiquitylating enzyme induces cytoplasmic incompatibility. *Nature Microbiology*, 1:17007.

Beer, M., Conraths, F.J. and Van Der Poel, W.H.M., 2012. 'Schmallenberg virus' – a novel orthobunyavirus emerging in Europe. *Epidemiology and Infection*, 141(1), pp.1–8.

Bellis, G.A., Halling, L. and Anderson, S.J., 2015. Pictorial key to adult female Culicoides Latreille, 1809 (Diptera: Ceratopogonidae) from the Northern Territory, Western Australia and South Australia. *Austral Entomology*, 54(1), pp.28–59.

Bhasin, A., Mordue, A.J. and Mordue, W., 2001. Field studies on efficacy of host odour baits for the biting midge *Culicoides impunctatus* in Scotland. *Medical and Veterinary Entomology*, 15(2), pp.147–156.

Bian, G. *et al*., 2010. The endosymbiotic bacterium *Wolbachia* induces resistance to dengue virus in *Aedes aegypti*. *PLoS pathogens*, 6(4): e1000833.

Birley, M.H. and Boorman, J.P.T., 1982. Estimating the survival and biting rates of haematophagous Insects, with particular reference to the *Culicoides obsoletus* group

(Diptera, Ceratopogonidae) in Southern England. *The Journal of Animal Ecology*, 51(1): 135.

Bishop, A.L. *et al*., 1995. The effects of habitat on the distribution of *Culicoides brevitarsis* kieffer (diptera: Ceratopogonidae) during its resting phase. *Australian Journal of Zoology*, 43(5), pp.531–539.

Bjorn Sandvik, 2009. World Borders Shapefile. Available at: http://thematicmapping.org/downloads/world_borders.php [Accessed 4 April, 2019].

Blackwell, A., Mordue, A.J., Young, M.R., *et al*., 1992. Bivoltinism, survival rates and reproductive characteristics of the Scottish biting midge, *Culicoides impunctatus* (Diptera, Ceratopogonidae) in Scotland. *Bulletin of Entomological Research*, 82(3), pp.299–306.

Blackwell, A., 1997. Diel flight periodicity of the biting midge *Culicoides impunctatus* and the effects of meteorological conditions. *Medical and Veterinary Entomology*, 11(4), pp.361–367.

Blackwell, A., Mordue (Luntz), A.J., Young, M.R., *et al*., 1992. The swarming behaviour of the Scottish biting midge, *Culicoides impunctatus* (Diptera: Ceratopogonidae). *Ecological Entomology*, 17(4), pp.319–325.

Blackwell, A., Young, M.R. and Mordue, W., 1994. The microhabitat of *Culicoides impunctatus* (Diptera: Ceratopogonidae) larvae in Scotland. *Bulletin of Entomological Research*, 84(3), pp.295–301.

Blagrove, M.S.C. *et al*., 2012. *Wolbachia* strain *w*Mel induces cytoplasmic incompatibility and blocks dengue transmission in *Aedes albopictus*. *Proceedings of the National Academy of Sciences of the United States of America*, 109(1), pp.255–260.

Blanc, G. *et al*., 2007. Lateral gene transfer between obligate intracellular bacteria: Evidence from the *Rickettsia massiliae* genome. *Genome Research*, 17(11), pp.1657–1664.

Blanc, G. *et al*., 2005. Molecular evolution of *Rickettsia* surface antigens: evidence of positive selection. *Molecular biology and evolution*, 22(10), pp.2073–2083.

Bleidorn, C. and Gerth, M., 2018. A critical re-evaluation of multilocus sequence typing

(MLST) efforts in *Wolbachia*. *FEMS Microbiology Ecology*, 94(1): fix163.

Bockoven, A. A. *et al.,* 2019. What goes up might come down: the spectacular spread of an endosymbiont is followed by its decline a decade later. *Microbial Ecology,* [Epub ahead of print]. doi: 10.1007/s00248-019-01417-4.

Boorman, J. and Goodard, P., 1970. Observations on the biology of *Culicoides impunctatus* Goetgh. (Dipt., Ceratopogonidae) in southern England. *Bulletin of entomological research*, 60(2), pp.189–98.

Boorman, J., 1974. The maintenance of laboratory colonies of *Culicoides variipennis* (Coq.), *C. nubeculosus* (Mg.) and *C. riethi* Kieff. (Diptera, Ceratopogonidae). *Bulletin of Entomological Research*, 64: 371.

Boorman, J., 1986. British *Culicoides* (Diptera: Ceratopogonidae): Notes on distribution and biology. *Entomologist's Gazette*, 37, pp.253–266.

Boorman, J., 1986. Presence of bluetongue virus vectors on Rhodes. *The Veterinary record*, 118(1): 21.

Boorman, J., 1985. Rearing *Culicoides obsoletus* (Diptera, Ceratopogonidae) on agar cultures of nematodes. *Progress in clinical and biological research*, 178, pp.229-31.

Bordenstein, Sarah R. and Bordenstein, Seth R., 2011. Temperature affects the tripartite interactions between bacteriophage WO, *Wolbachia*, and cytoplasmic incompatibility. *PLoS ONE,* 6(12): e29106.

Van Borm, S. *et al*., 2003. Cloning and sequencing of wsp encoding gene fragments reveals a diversity of co-infecting *Wolbachia* strains in *Acromyrmex leafcutter* ants. *Molecular Phylogenetics and Evolution*, 26(1), pp.102–109.

Bourtzis, K. *et al*., 2014. Harnessing mosquito-*Wolbachia* symbiosis for vector and disease control. *Acta Tropica*, 132(1), pp.150-163 .

Bowman, L. R., Donegan, S. and McCall, P. J., 2016. Is dengue vector control deficient in effectiveness or evidence?: Systematic review and meta-analysis. *PLOS Neglected Tropical Diseases*, 10(3): e0004551.

Boyd, B.M. *et al*., 2014. Genome Sequence of *Candidatus* Riesia pediculischaeffi,

endosymbiont of chimpanzee lice, and genomic comparison of recently acquired endosymbionts from human and chimpanzee lice. *G3: Genes, Genomes, Genetics*, 4(11), pp.2189–2195.

Bradshaw, W.E. *et al*., 2018. Evolutionary transition from blood feeding to obligate nonbiting in a mosquito. *Proceedings of the National Academy of Sciences of the United States of America*, 115(5), pp.1009–1014.

Bressan, A., Terlizzi, F. and Credi, R., 2012. Independent origins of vectored plant pathogenic bacteria from arthropod-associated *Arsenophonus* Endosymbionts. *Microbial Ecology*, 63(3), pp.628–638.

Brooks, M.A., 1970. Comments on the classification of intracellular symbiotes of cockroaches and a description of the species. *Journal of Invertebrate Pathology*, 16(2), pp.249–258.

Brown, A.M.V. *et al*., 2018. Comparative genomics of *Wolbachia-Cardinium* dual endosymbiosis in a plant-parasitic nematode. *Frontiers in Microbiology*, 9,: 2482.

Brumin, M., Levy, M. and Ghanim, M., 2012. Transovarial transmission of *Rickettsia* spp. and organ-specific infection of the whitefly *Bemisia tabaci*. *Applied and Environmental Microbiology*, 78(16), pp.5565–5574.

Burgin, L.E. *et al*., 2013. Investigating incursions of bluetongue virus using a model of long-distance *Culicoides* biting midge dispersal. *Transboundary and Emerging Diseases*, 60(3), pp.263–272.

Burke, G., Fiehn, O. and Moran, N., 2010. Effects of facultative symbionts and heat stress on the metabolome of pea aphids. *ISME Journal*, 4(2), pp.242–252.

Cai, H. *et al*., 2007. Genotypic and phenotypic characterization of *Lactobacillus casei* strains isolated from different ecological niches suggests frequent recombination and niche specificity. *Microbiology*, 153(8), pp.2655–2665.

Callis, J. and Kramer, T., 1997. Infectious disease of livestock with special reference to Southern Africa. *Veterinary Immunology and Immunopathology*, 58(2): 194.

Calvo-Pinilla, E. *et al*., 2014. Recombinant vaccines against bluetongue virus. *Virus Research*,

182, pp.78–86.

Calvo, J.H. *et al*., 2009. Variations in the mitochondrial cytochrome c oxidase subunit I gene indicate northward expanding populations of *Culicoides imicola* in Spain. *Bulletin of entomological research*, 99(6), pp.583–591.

Campbell, C.L. *et al*., 2004. Culture-independent analysis of midgut microbiota in the arbovirus vector *Culicoides sonorensis* (Diptera: Ceratopogonidae). *Journal of Medical Entomology*, 41(3), pp.340–348.

Campbell, J.A. and Pelham-Clinton, E.C., 1960. A taxonomic review of the British species of *Culicoides* Latreille (Diptera, Ceratopogonidæ). *Proceedings of the Royal Society of Edinburgh (B)*, 67(3), pp.181–302.

Campbell, M.M. and Kettle, D.S., 1976. Number of adult *Culicoides brevitarsis* kieffer (Diptera: Ceratopogonidae) emerging from bovine dung exposed under different conditions in the field. *Australian Journal of Zoology*, 24(1), pp.1–6.

Campbell, M.M. and Kettle, D.S., 1975. Oogenesis in *Culicoides brevitarsis* kieffer (Diptera: Ceratopogonidae) and the development of a plastron-like layer on the egg. *Australian Journal of Zoology*, 23(2), pp.151–157.

Caragata, E.P. *et al*., 2013. Dietary cholesterol modulates pathogen blocking by *Wolbachia*. *PLoS Pathogens*, 9(6): e1003459.

Carpenter, S., Szmaragd, C., *et al*., 2008. An assessment of *Culicoides* surveillance techniques in northern Europe: Have we underestimated a potential bluetongue virus vector? *Journal of Applied Ecology*, 45(4), pp.1237–1245.

Carpenter, S., 2001. Colonisation and dispersal studies of the Scottish biting midge, *Culicoides impunctatus* Goetghebuer. University of Aberdeen.

Carpenter, S., Mordue, A.J. and Mordue, W., 2001. Oviposition in *Culicoides impunctatus* under laboratory conditions. *Entomologia Experimentalis et Applicata*, 101(2), pp.123–129.

Carpenter, S. *et al*., 2006. Oral susceptibility to bluetongue virus of *Culicoides* (Diptera: Ceratopogonidae) from the United Kingdom. *Journal of Medical Entomology*, 43(1),

pp.73–78.

Carpenter, S., Mellor, P.S. and Torr, S.J., 2008. Control techniques for *Culicoides* biting midges and their application in the U.K. and northwestern Palaearctic. *Medical and Veterinary Entomology*, 122(3), pp.175–187.

Carpenter, S., Mordue, W. and Mordue, J., 2008. Selection of resting areas by emerging *Culicoides impunctatus* (Diptera: Ceratopogonidae) on downy birch (Betula pubescens). *International Journal of Pest Management*, 54(1), pp.39–42.

Carpenter, S. *et al.* 2008. An assessment of *Culicoides* surveillance techniques in northern Europe: Have we underestimated a potential bluetongue virus vector? *Journal of Applied Ecology,* 45(4), pp.1237 - 1245.

Carpenter, S. *et al*., 2013. *Culicoides* biting midges, arboviruses and public health in Europe. *Antiviral Research*, 100(1), pp.102–13.

Carpenter, S. *et al*., 2015. Vector competence of *Culicoides* for arboviruses: Three major periods of research, their influence on current studies and future directions. *OIE Revue Scientifique et Technique*, 34(1), pp.97–112.

Caspi-Fluger, A. *et al*., 2012. Horizontal transmission of the insect symbiont *Rickettsia* is plant-mediated. *Proceedings of the Royal Society B: Biological Sciences*, 279(1734), pp.1791–1796.

Castelli, M., Sassera, D. and Petroni, G., 2016. Biodiversity of “non-model” Rickettsiales and their association with aquatic organisms. [Book] "Rickettsiales: Biology, Molecular Biology, Epidemiology, and Vaccine Development." *Springer*, pp. 59–91.

Castoe, T.A. *et al*., 2008. Adaptive evolution and functional redesign of core metabolic proteins in snakes. *PLoS ONE*, 3(5): e2201.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, 17(4):540-552.

Ceccarelli, F.S., Haddad, C.R. and Ramírez, M.J., 2016. Endosymbiotic Rickettsiales (Alphaproteobacteria) from the spider genus *Amaurobioides* (Araneae: Anyphaenidae). *Journal of Arachnology*, 44(2), pp.251–253.

Cêtre-Sossah, C. *et al*., 2004. Molecular detection of *Culicoides* spp. and *Culicoides imicola*, the principal vector of bluetongue (BT) and African horse sickness (AHS) in Africa and Europe. *Veterinary Research*, 35(3), pp.325–337.

Chandler, L.J. *et al*., 1985. The virogenesis of bluetongue virus in *Culicoides variipennis*. *Progress in clinical and biological research*, 178, pp.245–53.

Charlesworth, J. *et al*., 2019. *Wolbachia*, *Cardinium* and climate: an analysis of global data. *Biology Letters*, 15(8): 20190273.

Chen, D.Q., Campbell, B.C. and Purcell, A.H., 1996. A new *Rickettsia* from a herbivorous insect, the pea aphid *Acyrthosiphon pisum* (Harris). *Current Microbiology*, 33(2), pp.123–128.

Chen, X., Li, S. and Aksoy, S., 1999. Concordant evolution of a symbiont with its host insect species: Molecular phylogeny of genus *Glossina* and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia*. *Journal of Molecular Evolution*, 48(1), pp.49-58.

Cheng, Q. *et al*., 2000. Tissue distribution and prevalence of *Wolbachia* infections in tsetse flies, Glossina spp. *Medical and Veterinary Entomology*, 14(1), pp.44–50.

Chiel, E. *et al*., 2009. Almost there: Transmission routes of bacterial symbionts between trophic levels. *PLoS ONE*, 4(3): e4767.

Chouin-Carneiro, T. *et al*., 2020. *Wolbachia* strain *w*AlbA blocks Zika virus transmission in *Aedes aegypti*. *Medical and Veterinary Entomology*, 34(1), pp.116-119.

Chrostek, E. *et al*., 2017. Horizontal transmission of intracellular insect symbionts via plants. *Frontiers in Microbiology*, 8: 2237.

Cirimotich, C. M., Ramirez, J. L. and Dimopoulos, G., 2011. Native microbiota shape insect vector competence for human pathogens. *Cell Host and Microbe*, 10(4), pp.307–310.

Clark, M.A. *et al*., 2000. Cospeciation between bacterial endosymbionts (*Buchnera*) and a recent radiation of aphids (*Uroleucon*) and pitfalls of testing for phylogenetic congruence. *Evolution*, 54(2), pp.517–525.

Clark, M.E. *et al*., 2005. Widespread prevalence of *Wolbachia* in laboratory stocks and the

implications for *Drosophila* research. *Genetics*, 170(4), pp.1667–1675.

Le Clec'h, W. *et al*., 2013. Cannibalism and predation as paths for horizontal passage of *Wolbachia* between terrestrial Isopods. *PLoS ONE*, 8(4): e60232.

Clements, A., 1992. [Book] The biology of mosquitoes: v.1: Development, nutrition and reproduction. *Chapman and Hall*, p. 509.

Corbin, C. *et al*., 2017. Heritable symbionts in a world of varying temperature. *Heredity*, 118, pp.10–20.

Cowan, G., 2000. Rickettsial diseases: The typhus group of fevers - A review. *Postgraduate Medical Journal*, 76(895), pp.269–272.

Craft, K.J. *et al*., 2010. Population genetics of ecological communities with DNA barcodes: An example from New Guinea Lepidoptera. *Proceedings of the National Academy of Sciences*, 107(11), pp.5041–5046.

Curtin, T.J. and Jones, J.C., 1961. The mechanism of ovulation and oviposition in *Aedes aegypti*. *Annals of the Entomological Society of America*, 54, p.295.

Dallas, J.F. *et al*., 2003. Phylogenetic status and matrilineal structure of the biting midge, *Culicoides imicola*, in Portugal, Rhodes and Israel. *Medical and Veterinary Entomology*, 17(4), pp.379–387.

Dantas-Torres, F., Alves, L. C. and Uilenberg, G., 2016. [Book chapter] Babesiosis. Arthropod Borne Diseases. *Springer,* pp. 347-354.

Darriba, D. *et al*., 2012. JModelTest 2: More models, new heuristics and parallel computing. *Nature Methods*, 9: 772.

DEFRA, 2014. GB bluetongue virus disease control strategy, Available at: www.gov.uk/defra [Accessed 20 November, 2015].

Delécolle, J.C., 1985. Nouvelle contribution à l'étude systématique et iconographique des espéces du genre *Culicoides* (Diptera: Ceratopogonidae) du Nord-Est de la France. Université Louis Pasteur du Strasbourg.

Dobson, S.L. *et al*., 1999. *Wolbachia* infections are distributed throughout insect somatic

and germ line tissues. *Insect Biochemistry and Molecular Biology*, 29(2), pp.153–160.

Dodson, B. L. *et al.,* 2014. *Wolbachia* Enhances West Nile Virus (WNV) Infection in the Mosquito *Culex tarsalis*. *PLoS Neglected Tropical Diseases,* 8(7): e2965.

Douglas, A.E., 2002. Nutritional interactions in insect-microbial symbioses: Aphids and their symbiotic bacteria *Buchnera*. *Annual Review of Entomology*, 43(1), pp.17–37.

Douglas, A.E., 2009. The microbial dimension in insect nutritional ecology. *Functional Ecology*, 23(1), pp.38–47.

Downes, J.A., 1950. Habits and life-cycle of *Culicoides nubeculosus* mg. *Nature*, 166(4221), pp.510–511.

Downes, J.A., 1955. Observations on the swarming flight and mating of *Culicodies* (Diptera: Ceratopogonidae). *Transactions of the Royal Entomological Society of London*, 106(5), pp.213–236.

Downes, J.A. and Kettle, D.S., 1952. Descriptions of three species of *Culicoides* Latreille (Diptera: Ceratopogonidae) new to science, together with notes on, and a revised key to the British species of the pulicaris and obsoletus groups. *Proceedings of the Royal Entomological Society of London B*, 21, pp.3–78.

Duron, O. *et al*., 2008. High incidence of the maternally inherited bacterium *Cardinium* in spiders. *Molecular Ecology*, 17(6), pp.1427–1437.

Duron, O., Fort, P. and Weill, M., 2007. Influence of aging on cytoplasmic incompatibility, sperm modification and *Wolbachia* density in *Culex pipiens* mosquitoes. *Heredity*, 98, pp.368–374.

Dwyer, J. *et al*., 2007. *Understanding and influencing positive behaviour change in farmers and land managers - a project for Defra*. *DEFRA Report.* Available at: https://www.academia.edu/2604701/Understanding_and_influencing_positive_behaviour_change_in_farmers_and_land_managers [Accessed 20 November, 2015].

Dyková, I. *et al*., 2003. *Nuclearia pattersoni* sp. n. (Filosea), a new species of amphizoic amoeba isolated from gills of roach (*Rutilus rutilus*), and its rickettsial endosymbiont. *Folia Parasitologica*, 50(3), pp.161–170.

Dyson, E.A., Kamath, M.K. and Hurst, G.D.D., 2002. *Wolbachia* infection associated with all-female broods in *Hypolimnas bolina* (Lepidoptera: Nymphalidae): Evidence for horizontal transmission of a butterfly male killer. *Heredity*, 88(3), pp.166–171.

Eagles, D. *et al*., 2014. Long-distance aerial dispersal modelling of *Culicoides* biting midges: Case studies of incursions into Australia. *BMC Veterinary Research*, 10: 135.

Edlund, A. *et al*., 2012. Antibiotic-induced change of bacterial communities associated with the copepod *Nitocra spinipes*. *PLoS ONE*, 7(3): e33107.

Elbers, A.R. *et al*., 2010. To report or not to report: a psychosocial investigation aimed at improving early detection of avian influenza outbreaks. *Rev Sci Tech*, 29(3), pp.435–449.

Elbers, A.R.W. *et al*., 2013. Schmallenberg virus in *Culicoides* spp. biting midges, the Netherlands, 2011. *Emerging Infectious Diseases*, 19(1), pp.106–109.

European Food safety Authority, 2016. Schmallenberg virus: State of Art. *EFSA Journal*, 12(5): 3681.

Ewing, B. and Green, P., 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research*, (8)3, pp.186-194.

Ferla, M.P. *et al*., 2013. New rRNA gene-based phylogenies of the Alphaproteobacteria provide perspective on major groups, mitochondrial ancestry and phylogenetic instability. *PLoS ONE*, 8(12): e83383.

Ferrari, J. and Vavre, F., 2011. Bacterial symbionts in insects or the story of communities affecting communities. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 366(1569), pp.1389–1400.

Fine, P.E.M., 1978. On the dynamics of symbiote-dependent cytoplasmic incompatibility in culicine mosquitoes. *Journal of Invertebrate Pathology*, 31(1), pp.8–10.

Fisher, B. and Smith, A.M., 2008. A revision of Malagasy species of *Anochetus mayr* and *Odontomachus latreille* (hymenoptera: formicidae), *PLOS ONE*, 3(5): e1787.

Flores, H.A. and O'Neill, S.L., 2018. Controlling vector-borne diseases by releasing modified mosquitoes. *Nature Reviews Microbiology*, 16(8), pp.508–518.

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.

Fournier, P.E. *et al*., 2003. Gene sequence-based criteria for identification of new *Rickettsia* isolates and description of *Rickettsia heilongjiangensis* sp. nov. *Journal of Clinical Microbiology*, 41(12), pp.5456–5465.

Foxi, C. *et al*., 2016. Role of different *Culicoides* vectors (Diptera: Ceratopogonidae) in bluetongue virus transmission and overwintering in Sardinia (Italy). *Parasites and Vectors*, 9(1): 440.

Frances, S.P. *et al*., 2000. Transmission of *Orientia tsutsugamushi*, the aetiological agent for scrub typhus, to co-feeding mites. *Parasitology*, 6, pp.601–607.

Fraune, S. and Bosch, T.C.G., 2007. Long-term maintenance of species-specific bacterial microbiota in the basal metazoan Hydra. *Proceedings of the National Academy of Sciences of the United States of America*, 104(32), pp.13146–13151.

Fredslund, J. *et al*., 2005. PriFi: Using a multiple alignment of related sequences to find primers for amplification of homologs. *Nucleic Acids Research*, 1(33), pp.516–20.

Freeland, J.R. and Conrad, K.F., 2002. Genetic similarity within and among populations of the variable and Azure damelflies (*Coenagrion pulchellum* and *C. puella*). *Hydrobiologia*, 479(1–3), pp.69–73.

Frentiu, F.D., 2017. Lipids and Pathogen Blocking by *Wolbachia*. *Trends in Parasitology*, 33(12), pp.916–917.

Fricks, L.D., 1915. Rocky Mountain Spotted Fever: A report of its investigation and the measures undertaken for its eradication during 1914. *Public Health Reports (1896-1970)*, 30(3), pp.148–165.

Frost, L.S. *et al*., 2005. Mobile genetic elements: The agents of open source evolution. *Nature Reviews Microbiology*, 3(9), pp.722–732.

Frydman, H.M. *et al*., 2006. Somatic stem cell niche tropism in *Wolbachia*. *Nature*, 441(7092), pp.509–512.

Fu, H. *et al*., 1999. The barriers to bluetongue virus infection, dissemination and transmission in the vector, *Culicoides variipennis* (Diptera: Ceratopogonidae). *Archives of Virology*, 144(4), pp.747–761.

Fukuda, T. *et al*., 1979. Experimental transmission of fowl pox by *Culicoides arakawae*. *National Institute of Animal Health quarterly*, 19(3), pp.104–105.

Gasteiger, E. *et al*., 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, 31(13), pp.3784–3788.

Gehrer, L. and Vorburger, C., 2012. Parasitoids as vectors of facultative bacterial endosymbionts in aphids. *Biology Letters*, 8(4), pp.613–615.

Geoghegan, V. *et al*., 2017. Perturbed cholesterol and vesicular trafficking associated with dengue blocking in *Wolbachia*-infected *Aedes aegypti* cells. *Nature Communications*, 8(1): 526.

Gerth, M. *et al*., 2017. Green lacewings (Neuroptera: Chrysopidae) are commonly associated with a diversity of rickettsial endosymbionts. *Zoological Letters*, 3: 12.

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.

Gloria-Soria, A., Chiodo, T.G. and Powell, J.R., 2018. Lack of evidence for natural *Wolbachia* infections in *Aedes aegypti* (Diptera: Culicidae). *Journal of medical entomology*, 55(5), pp.1354–1356.

Goffredo, M. *et al*., 2015. Vector species of *Culicoides* midges implicated in the 2012‑2014 Bluetongue epidemics in Italy. *Veterinaria italiana*, 51(2), pp.131–138.

Goldstein, P. and DeSalle, R., 2019. Review and Interpretation of Trends in DNA Barcoding. *Frontiers in Ecology and Evolution*, 7: 302.

Gomulski, L.M. *et al*., 2006. Phylogeny of the subgenus *Culicoides* and related species in Italy, inferred from internal transcribed spacer 2 ribosomal DNA sequences. *Medical and Veterinary Entomology*, 20(2), pp.229–238.

Gonella, E. *et al*., 2015. Plant-mediated interspecific horizontal transmission of an

intracellular symbiont in insects. *Scientific Reports*, 13(5): 15811.

González, M. *et al*., 2013. A survey of *Culicoides* developmental sites on a farm in northern Spain, with a brief review of immature habitats of European species. *Veterinary Parasitology*, 191(1–2), pp.81–93.

Goodacre, S.L. *et al*., 2009. Microbial modification of host long-distance dispersal capacity. *BMC Biology*, 7: 32.

Goodacre, S.L. *et al*., 2006. *Wolbachia* and other endosymbiont infections in spiders. *Molecular Ecology*, 15(2), pp.517–527.

Gotoh, T., Noda, H. and Ito, S., 2007a. *Cardinium* symbionts cause cytoplasmic incompatibility in spider mites. *Heredity*, 98(1), pp.13–20.

Gottlieb, Y. *et al*., 2006. Identification and localization of a *Rickettsia* sp. in *Bemisia tabaci* (Homoptera: Aleyrodidae). *Applied and Environmental Microbiology*, 5(72), pp.3646–3652.

Gottlieb, Y. *et al*., 2008. Inherited intracellular ecosystem: symbiotic bacteria share bacteriocytes in whiteflies. *The FASEB Journal*, 22(7), pp.2591–2599.

Gouin, E. *et al*., 1999. A comparative study of the actin-based motilities of the pathogenic bacteria *Listeria monocytogenes*, *Shigella flexneri* and *Rickettsia conorii*. *Journal of cell science*, 112(11), pp.1697–1708.

Gould, E.A. and Higgs, S., 2009. Impact of climate change and other factors on emerging arbovirus diseases. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 103(2), pp.109–121.

Graham, R. I. *et al.*, 2012. *Wolbachia* in a major African crop pest increases susceptibility to viral disease rather than protects. *Ecology Letters,* 15(9); pp. 993-1000.

Groot, T.V.M. and Breeuwer, J.A.J., 2006. *Cardinium* symbionts induce haploid thelytoky in most clones of three closely related *Brevipalpus* species. *Experimental and Applied Acarology*, 39(3–4), pp.257–271.

Gross, L., 1996. How Charles Nicolle of the Pasteur Institute discovered that epidemic typhus is transmitted by lice: reminiscences from my years at the Pasteur Institute in

Paris. *Proceedings of the National Academy of Sciences of the United States of America*, 93(20), pp.10539–10540.

Gruwell, M.E., Wu, J. and Normark, B.B., 2009. Diversity and phylogeny of *Cardinium* (Bacteroidetes) in armored scale insects (Hemiptera: Diaspididae). *Annals of the Entomological Society of America*, 102(6), pp.1050–1061.

Gualtieri, L. *et al*., 2017. Life inside a gall: closeness does not favour horizontal transmission of *Rickettsia* between a gall wasp and its parasitoid. *FEMS microbiology ecology*, 93(7).

Gubler, D. J., 2009. Vector-borne diseases. *Revue Scientifique et Technique de l'OIE*, 28(2), pp. 583–588.

Hagimori, T. *et al*., 2006. The first finding of a *Rickettsia* Bacterium associated with parthenogenesis induction among insects. *Current Microbiology*, 52(2), pp.97–101.

Hajibabaei, M. *et al.,* 2006. DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences of the United States of America,* 103(4), pp.968-971.

Haqshenas, G. *et al*., 2019. A role for the insulin receptor in the suppression of dengue virus and zika virus in *Wolbachia*-infected mosquito cells. *Cell Reports*, 26(3), pp.529-535.

Harrup, L.E. *et al*., 2013. Larval development and emergence sites of farm-associated *Culicoides* in the United Kingdom. *Medical and Veterinary Entomology*, 27(4), pp.441–449.

Harrup, L.E., Miranda, M.A. and Carpenter, S., 2016. Advances in control techniques for *Culicoides* and future prospects. *Veterinaria Italiana*, 52(3–4), pp.247–264.

He, Z. *et al*., 2016. Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Research*, 44(1), pp.236–241.

Heath, B.D. *et al*., 1999. Horizontal transfer of *Wolbachia* between phylogenetically distant insect species by a naturally occurring mechanism. *Current Biology*, 9(6), pp.313–316.

Heaton, N.S. *et al*., 2010. Dengue virus nonstructural protein 3 redistributes fatty acid synthase to sites of viral replication and increases cellular fatty acid synthesis.

Proceedings of the National Academy of Sciences of the United States of America, 107(40), pp.17345–17350.

Hebert, P.D.N. *et al*., 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, 270(1512), pp.313–321.

Hebert, P.D.N. *et al*., 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, 101(41), pp.14812–14817.

Hebert, P.D.N. *et al*., 2016. Counting animal species with DNA barcodes: Canadian insects. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1702).

Hedges, L.M. *et al*., 2008. *Wolbachia* and virus protection in insects. *Science*, 322(5902): 702.

Heinzen, R.A. *et al*., 1993. Directional actin polymerization associated with spotted fever group *Rickettsia* infection of Vero cells. *Infection and Immunity*, 61(5), pp.1926–1935.

Hellqvist, S., 1994. Biology of *Synacra* sp. (Hym., Diapriidae), a parasitoid of *Bradysia paupera* (Dipt., Sciaridae) in Swedish greenhouses. *Journal of Applied Entomology*, 117(1–5), pp.491–497.

Hendry, G.A.F., 1996. Midges in Scotland, Edinburgh: Mercat Press.

Hendry, G.A.F. and Godwin, G., 1988. Biting midges in Scottish forestry: a costly irritant or a trivial nuisance? *Scottish Forestry,* 42, pp.113-119.

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.

Hernández-Triana, L.M., *et al.* 2014. Recovery of DNA barcodes from blackfly museum specimens (Diptera: Simuliidae) using primer sets that target a variety of sequence lengths. *Molecular Ecology Resources*, 14(3), pp.508-518.

Herren, J.K. *et al*., 2013. Vertical transmission of a *Drosophila* endosymbiont via cooption of the yolk transport and internalization machinery. *mBio*, 4(2), pp.e00532-00512.

Hertig, M. and Wolbach, S.B., 1924. Studies on Rickettsia-Like Micro-Organisms in Insects. *The Journal of medical research*, 44(3), pp.329–374.

Hilgenboecker, K. *et al*., 2008. How many species are infected with *Wolbachia*? - A statistical analysis of current data. *FEMS Microbiology Letters*, 281(2), pp.215–220.

Hill, M.A., 1947. The life-cycle and habits of *Culicoides impunctatus* goetghebuer and *Culicoides obsoletus* meigen, together with some observations on the life-cycle of *Culicoides odibilis* austen, *Culicoides pallidicornis* kieffer, *Culicoides cubitalis* edwards and *Culicoides*. *Annals of Tropical Medicine and Parasitology*, 41(1), pp.55–115.

Himler, A. G. *et al.,* 2011. Rapid spread of a bacterial symbiont in an invasive whitefly is driven by fitness benefits and female bias. *Science,* 332(6026), pp.254-256.

Hoffmann, A. A. *et al.*, 2011. Successful establishment of *Wolbachia* in *Aedes* populations to suppress dengue transmission. *Nature*, 476(7361), pp.454–457.

Hoffmann, A.A. *et al*., 2014. Stability of the *w*Mel *Wolbachia* infection following invasion into *Aedes aegypti* populations. *PLoS Neglected Tropical Diseases*, 8(9): e3115.

Hoffmann, A.A., Ross, P.A. and Rašić, G., 2015. *Wolbachia* strains for disease control: ecological and evolutionary considerations. *Evolutionary Applications*, 8(8), pp.751–768.

Holmes, P.R. and Boorman, J.P.T., 1987. Light and suction trap catches of *Culicoides* midges in southern England. *Medical and Veterinary Entomology*, 1(4), pp.349–359.

Horecka, B. *et al*., 2017. Usefulness of DNA barcoding for identification of preimaginal stages species of the Mecinini: Genus *Miarus* (Schoenherr, 1826) and *Cleopomiarus* (Pierce, 1919) (Coleoptera: Curculionidae). *Folia Biologica (Poland)*, 65(1).

Hribar, L.J., 1990. A review of methods for recovering biting midge larvae (Diptera: Ceratopogonidae) from substrate samples. *Journal of Agricultural Entomology*, 71, pp.71–77.

Huber, J., 1986. Systematics, biology, and hosts of the Mymaridae and Mymarommatidae (Insecta: Hymenoptera): 1758-1984. *Entomography*, 4, pp.185–243.

Hughes, G.L. *et al*., 2011. *Wolbachia* infections are virulent and inhibit the human malaria

parasite *Plasmodium falciparum* in *Anopheles gambiae*. *PLoS Pathogens*, 7(5): e1002043.

Hughes, G.L. and Rasgon, J.L., 2014. Transinfection: A method to investigate *Wolbachia*-host interactions and control arthropod-borne disease. *Insect Molecular Biology*, 23(2), pp.141–151.

Huigens, M.E. *et al*., 2004. Natural interspecific and intraspecific horizontal transfer of parthenogenesis-inducing *Wolbachia* in *Trichogramma* wasps. *Proceedings of the Royal Society B: Biological Sciences*, 271(1538), pp.509–515.

Hunter, M.S., Perlman, S.J. and Kelly, S.E., 2003. A bacterial symbiont in the Bacteroidetes induces cytoplasmic incompatibility in the parasitoid wasp *Encarsia pergandiella*. *Proceedings of the Royal Society B: Biological Sciences*, 270(1529), pp.2185–2190.

van den Hurk, A.F. *et al*., 2012. Impact of *Wolbachia* on infection with chikungunya and yellow fever viruses in the mosquito vector *Aedes aegypti*. *PLoS Neglected Tropical Diseases*, 6(11): e1892.

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

Hurst, G.D.D. *et al*., 1999. Invasion of one insect species, *Adalia bipunctata*, by two different male-killing bacteria. *Insect Molecular Biology*, 8(1), pp.133–139.

Hurst, G.D.D. *et al*., 2000. Male killing *Wolbachia* in *Drosophila*: a temperature sensitive trait with a threshold density. *Genetics*, 156(2), pp.699–709.

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

Hurst, G.D.D. and Jiggins, F.M., 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings Biological sciences / The Royal Society*, 272(1572), pp.1525–1534.

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

Hussain, M. *et al*., 2012. Effect of *Wolbachia* on replication of West Nile Virus in mosquito cell line and adult mosquitoes. *Journal of Virology*, 87(2), pp.851–858.

Hwangbo, D.S. *et al*., 2004. *Drosophila* dFOXO controls lifespan and regulates insulin signalling in brain and fat body. *Nature*, 429(6991), pp.562–566.

Hyong, S.K. *et al*., 2008. Comparison of conventional, nested, and real-time PCR assays for rapid and accurate detection of *Vibrio vulnificus*. *Journal of Clinical Microbiology*, 46(9), pp.2992–2998.

Ikewaki, J. *et al*., 2003. Real-time PCR assay compared to nested PCR and antigenemia assays for detecting cytomegalovirus reactivation in adult T-cell leukemia-lymphoma patients. *Journal of Clinical Microbiology*, 41(9), pp.4382–4387.

Ikeya, T. *et al*., 2009. The endosymbiont *Wolbachia* increases insulin/IGF-like signalling in *Drosophila*. *Proceedings of the Royal Society B: Biological Sciences*, 276(1674), pp.3799–3807.

Ikuta, T. *et al*., 2016. Surfing the vegetal pole in a small population: Extracellular vertical transmission of an "intracellular" deep-sea clam symbiont. *Royal Society Open Science*, 3(5): 160130.

Iturbe-Ormaetxe, I., Walker, T. and O' Neill, S.L., 2011. *Wolbachia* and the biological control of mosquito-borne disease. *EMBO reports*, 12(6), pp.508–518.

Jäckel, R., Mora, D. and Dobler, S., 2013. Evidence for selective sweeps by *Wolbachia* infections: Phylogeny of *Altica* leaf beetles and their reproductive parasites. *Molecular Ecology*, 22(16), pp.4241–4255.

Jacquet, S. *et al*., 2015. Colonization of the Mediterranean basin by the vector biting midge species *Culicoides imicola*: An old story. *Molecular Ecology*, 24(22), pp.5707–5725.

Jacquet, S. *et al*., 2016. Range expansion of the bluetongue vector, *Culicoides imicola*, in

continental France likely due to rare wind-transport events. *Scientific Reports*, 6: 27247.

Jaenike, J. 2009. Coupled population dynamics of endosymbionts within and between hosts. *Oikos,* 118(3), pp.353-362.

Jaenike, J. *et al*., 2010. Adaptation via symbiosis: Recent spread of a *Drosophila* defensive symbiont. *Science*, 329(5988), pp.212–215.

Jennings, D.M. and Mellor, P.S., 1988. The vector potential of British *Culicoides* species for bluetongue virus. *Veterinary Microbiology*, 17(1), pp.1–10.

Jiggins, F.M., 2003. Male-killing *Wolbachia* and mitochondrial DNA: Selective sweeps, hybrid introgression and parasite population dynamics. *Genetics*, 164(1), pp.5–12.

Jiggins, F.M., Hurst, G.D.D. and Yang, Z., 2002. Host-symbiont conflicts: positive selection on an outer membrane protein of parasitic but not mutualistic Rickettsiaceae. *Molecular Biology and Evolution*, 19(8), pp.1341–1349.

Johannesen, J., 2017. Tracing the history and ecological context of *Wolbachia* double infection in a specialist host (*Urophora cardui*)—parasitoid (*Eurytoma serratulae*) system. *Ecology and Evolution*, 7(3), pp.986–996.

Jones, R. H., 1957. The laboratory colonization of *Culicoides variipennis* (Coq.)1. *Journal of Economic Entomology*, 50(1), pp.107–108.

Jones, R. H., 1960. Mass-production methods for the colonization of *Culicoides variipennis sonorensis*. *Journal of Economic Entomology*, 53(5), pp.731–735.

Jones, K. E. *et al*. 2008. Global trends in emerging infectious diseases, *Nature*, 451(7181), pp.990–993.

Joyce, J.D. *et al*., 2011. Interactions between La Crosse virus and bacteria isolated from the digestive tract of *Aedes albopictus* (Diptera: Culicidae). *Journal of Medical Entomology*, 48(2), pp.389–394.

Juchault, P. *et al*., 1994. New evidence for feminizing bacteria in terrestrial isopods: Evolutionary implications. *Comptes Rendus de l'Academie des Sciences - Serie III*, 317(3), pp.225–230.

Kalyaanamoorthy, S. *et al*., 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6), pp.587–589.

Kanturski, M. *et al*., 2018. DNA barcoding and a precise morphological comparison revealed a cryptic species in the *Nippolachnus piri* complex (Hemiptera: Aphididae: Lachninae). *Scientific Reports*, 8(1): 8998.

El Karkouri, K. *et al*., 2016. Origin and evolution of rickettsial plasmids. *PLoS ONE*, 11(2): e0147492.

Katoh, K. and Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), pp.772–780.

Kawafune, K. *et al*., 2015. Two different rickettsial bacteria invading *Volvox carteri*. *PLoS ONE*, 10(2): e0116192.

Kearse, M. *et al*., 2012. Geneious. *Bioinformatics (Oxford, England)*.

Kedmi, M. *et al*., 2010. Assessment of the productivity effects associated with epizootic hemorrhagic disease in dairy herds. *Journal of Dairy Science*, 93(6), pp.2486–2495.

Kekkonen, M. and Hebert, P.D.N., 2014. DNA barcode-based delineation of putative species: Efficient start for taxonomic workflows. *Molecular Ecology Resources*, 14(4), pp.706–715.

Kent, B.N. *et al*., 2011. Complete bacteriophage transfer in a bacterial endosymbiont (Wolbachia) determined by targeted genome capture. *Genome Biology and Evolution*, 3, pp.209–218.

Kettle, D.S., 1977. Biology and Bionomics of Bloodsucking Ceratopogonids. *Annual Review of Entomology*, 22, pp.33–51.

Kettle, D.S., 1957. Preliminary observations on weather conditions and the activity of biting flies. *Proceedings of the Royal Entomological Society of London. Series A, General Entomology*, 32(103), pp.13-20.

Kettle, D.S. and Lawson, J.W.H., 1952. The early stages of British biting midges *Culicoides* Latreille (Diptera: Ceratopogonidae) and allied Genera. *Bulletin of Entomological*

Research, 43(3), pp.421–467.

Kettle, D.S., Wild, C.H. and Elson, M.M., 1975. A new technique for rearing individual *Culicoides* larvae (Diptera: Ceratopogonidae). *Journal of Medical Entomology*, 12(2), pp.263–264.

Kikuchi, Y. and Fukatsu, T., 2005. *Rickettsia* infection in natural leech populations. *Microbial Ecology*, 49(2), pp.265–271.

Kikuchi, Y., 2009. Endosymbiotic bacteria in insects: Their diversity and culturability. *Microbes and Environments*, 24(3), pp.195–204.

Kikuchi, Y. *et al*., 2002. Novel clade of *Rickettsia* spp. from leeches. *Applied and Environmental Microbiology*, 68(2), pp.999–1004.

Kilpatrick, A.M. *et al*., 2008. Temperature, viral genetics, and the transmission of West Nile virus by *Culex pipiens* mosquitoes. *PLoS Pathogens*, 4(6): e1000092.

Kimmerling, N. *et al*., 2018. Quantitative species-level ecology of reef fish larvae via metabarcoding. *Nature Ecology and Evolution*, 2(2), pp.306–316.

Kline, D.L. and Axtell, R.C., 1976. Salt marsh *Culicoides* (Diptera: ceratopogonidae): Species, seasonal abundance, and comparison of trapping methods. *Mosquito News*, 36(1), pp.1–10.

Kliot, A. *et al*., 2014. Implication of the bacterial endosymbiont *Rickettsia* spp. in interactions of the whitefly *Bemisia tabaci* with Tomato yellow leaf curl virus. *Journal of Virology*, 88(10), pp.5652–5660.

Koch, H. and Axtell, R., 1978. Autogeny and rearing of *Culicoides furens, C. hollensis* and *C. melleus* (Diptera: Ceratopogonidae) from coastal North Carolina. *Mosquito News,* 38(2), pp.240-244.

Koga, R. *et al*., 2012. Cellular mechanism for selective vertical transmission of an obligate insect symbiont at the bacteriocyte-embryo interface. *Proceedings of the National Academy of Sciences*, 109(20), pp.1230–1237.

Koga, R., Tsuchida, T. and Fukatsu, T., 2009. Quenching autofluorescence of insect tissues for in situ detection of endosymbionts. *Applied Entomology and Zoology*, 44(2): 281.

Kondo, N. *et al*., 2002. Prevailing triple infection with *Wolbachia* in *Callosobruchus chinensis* (Coleoptera: Bruchidae). *Molecular Ecology*, 11(2), pp.167–180.

Kriesner, P. *et al.,* 2013. Rapid sequential spread of two *Wolbachia* variants in *Drosophila simulans*. *PLoS Pathogens,* 9(9): e1003607.

Küchler, S.M., Kehl, S. and Dettner, K., 2009. Characterization and localization of *Rickettsia* sp. in water beetles of genus *Deronectes* (Coleoptera: Dytiscidae). *FEMS Microbiology Ecology*, 68(2), pp.201–211.

Kulkarni, A. *et al*., 2019. *Wolbachia pipientis* occurs in *Aedes aegypti* populations in New Mexico and Florida, USA. *Ecology and Evolution*, 9(1), pp.6148–6156.

Kvie, K.S. *et al*., 2013. Deep sympatric mtDNA divergence in the autumnal moth (Epirrita autumnata). *Ecology and Evolution*, 1(3), pp.126–144.

Lagrue, C. *et al*., 2016. Genetic structure and host-parasite co-divergence: Evidence for trait-specific local adaptation. *Biological Journal of the Linnean Society*, 118(2), pp.344–358.

Landmann, F. *et al*., 2010. Asymmetric *Wolbachia* segregation during Early *Brugia malayi* embryogenesis determines its distribution in adult host tissues. *PLoS Neglected Tropical Diseases*, 4(7): e758.

Landmann, F. *et al.*, 2012. Both asymmetric mitotic segregation and cell-to-cell invasion are required for stable germline transmission of *Wolbachia* in filarial nematodes. *Biology Open,* 1(6): 536–547.

Lardeux, F.J.R. and Ottenwaelder, T., 1997. Density of larval *Culicoides belkini* (Diptera: Ceratopogonidae) in relation to physicochemical variables in different habitats. *Journal of Medical Entomology*, 34(4), pp.387–395.

Larska, M. *et al*., 2013. Detection of the Schmallenberg virus in nulliparous *Culicoides obsoletus*/*scoticus* complex and *C. punctatus*-The possibility of transovarial virus transmission in the midge population and of a new vector. *Veterinary Microbiology*, 166(3–4), pp.467–473.

Lassen, S.B. *et al*., 2012. Molecular differentiation of *Culicoides* biting midges (Diptera: Ceratopogonidae) from the subgenus *Culicoides* Latreille in Denmark. *Parasitology*

Research, 110(5), pp.1765–1771.

Lawson, J.W.H., 1951. The anatomy and morphology of the earky stages of *Culicoides nubeculosus* Meigen (Diptera: Ceratopogonidae = Heleidae). *Transactions of the Royal Entomological Society of London*, 102(9), pp.511–574.

Leask, R., Botha, A.M. and Bath, G.F., 2013. Schmallenberg virus - Is it present in South Africa? *Journal of the South African Veterinary Association*, 84(1), pp.1–4.

Lees, R. S. *et al.*, 2015. Back to the future: the sterile insect technique against mosquito disease vectors, *Current Opinion in Insect Science*, 10, pp.156–162.

Leigh, J.W. and Bryant, D., 2015. Popart: Full-Feature Software for Haplotype Network Construction. *Methods in Ecology and Evolution*, 6(9), pp.1110–1116.

Lewis, S.E. *et al*., 2014. First detection of endosymbiotic bacteria in biting midges *Culicoides pulicaris* and *Culicoides punctatus*, important Palaearctic vectors of bluetongue virus. *Medical and Veterinary Entomology*, 28(4), pp.453–456.

Li, A.Y. *et al*., 2010. High prevalence of *Rickettsia gravesii* sp. nov. in *Amblyomma triguttatum* collected from feral pigs. *Veterinary Microbiology*, 146(1–2), pp.59–62.

Li, K. *et al*., 2016. Diversity of bacteriome associated with *Phlebotomus chinensis* (Diptera: Psychodidae) sand flies in two wild populations from China. *Scientific Reports*, 6: 36406.

Li, S.J. *et al*., 2017. Plant-mediated horizontal transmission of *Wolbachia* between whiteflies. *ISME Journal*, 11(4), pp.1019–1028.

Librado, P. and Rozas, J., 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25(11), pp.1451–1452.

Linares, M.C. *et al*., 2009. High mitochondrial diversity in geographically widespread butterflies of Madagascar: A test of the DNA barcoding approach. *Molecular Phylogenetics and Evolution*, 50(3), pp.485–495.

Lindsey, A.R.I. *et al*., 2018. Conflict in the intracellular lives of endosymbionts and viruses: A mechanistic look at *Wolbachia*-mediated pathogen-blocking. *Viruses*, 10(4): 141.

Linley, J.R., 1985a. Biting midges (Diptera: Ceratopogonidae) as vectors of nonviral animal

pathogens. *Journal of Medical Entomology*, 22(6), pp.589–599.

Linley, J.R., 1968. Colonization of *Culicoides furens*. *Annals of the Entomological Society of America*, 61(6), pp.1486–1490.

Linley, J.R., 1985b. Growth and survival of *Culicoides melleus* larvae (Diptera: Ceratopogonidae) on four prey organisms. *Journal of medical entomology*, 22(2), pp.178–189.

Linley, J.R., 1979. Nutrient agar mixtures for culture of food organisms for predaceous *Culicoides* larvae (Diptera: Ceratopogonidae). *Journal of Medical Entomology*, 16(2), pp.171–172.

Linley, J.R., 1965. The ovarian cycle and egg stage in *Leptoconops* (*Holoconops*) *becquaerti* (Kieff.) (Diptera, Ceratopogonidae). *Bulletin of Entomological Research*, 56(1), pp.37–56.

Linley, J.R. and Davies, J.B., 1971. Sandflies and Tourism in Florida and the Bahamas and Caribbean Area. *Journal of Economic Entomology*, 64(1), pp.264–278.

Lühken, R. *et al*., 2015. Effects of temperature and photoperiod on the development of overwintering immature *Culicoides chiopterus* and *C. dewulfi*. *Veterinary Parasitology*, 214(1–2), pp.195–199.

Luis-Pantoja, M. *et al*., 2015. *Rickettsia*-related bacteria associated with papaya plants showing bunchy top disease in Cuba. *Journal of General Plant Pathology*, 81(2), pp.166–168.

Łukasik, P. *et al*., 2015. Horizontal transfer of facultative endosymbionts is limited by host relatedness. *Evolution*, 69(1), pp.2757–2766.

Łukasik, P. *et al*., 2013. Protection against a fungal pathogen conferred by the aphid facultative endosymbionts *Rickettsia* and *Spiroplasma* is expressed in multiple host genotypes and species and is not influenced by co-infection with another symbiont. *Journal of Evolutionary Biology*, 26(12), pp.2654–2661.

Maan, S. *et al*., 2011. Complete genome characterisation of a novel 26th bluetongue virus serotype from Kuwait. *PLoS ONE*, 6(10): e26147.

Machtelinckx, T. *et al*., 2012. Microbial community of predatory bugs of the genus *Macrolophus* (Hemiptera: Miridae). *BMC Microbiology*, 12(1): 9.

Maclachlan, N.J. *et al*., 2009. The Pathology and Pathogenesis of Bluetongue. *Journal of Comparative Pathology*, 141(1), pp.1–16.

Maiden, M.C. *et al*., 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences of the United States of America*, 95(6), pp.3140–5.

Maina, A.N. *et al*., 2017. Molecular characterization of novel mosquito-borne *Rickettsia* spp. from mosquitoes collected at the Demilitarized Zone of the Republic of Korea. *PLoS ONE*, 12(11): e0188327.

Mains, J.W. *et al*., 2019. Localized control of *Aedes aegypti* (Diptera: Culicidae) in Miami, FL, via inundative releases of *Wolbachia*-infected male mosquitoes. *Journal of Medical Entomology*, 56(5), pp.1296–1303.

Majerus, T.M.O. *et al*., 1999. Molecular identification of a male-killing agent in the ladybird *Harmonia axyridis* (Pallas) (Coleoptera: Coccinellidae). *Insect Molecular Biology*, 8(4), pp.551–555.

Manzano-Marín, A. *et al*., 2015. Solving a bloody mess: B-vitamin independent metabolic convergence among gammaproteobacterial obligate endosymbionts from blood-feeding arthropods and the Leech *haementeria officinalis*. *Genome Biology and Evolution*, 7(10), pp.2871–2884.

Martijn, J. *et al*., 2018. Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature*, 557(7703), pp.101–105.

Martijn, J. *et al*., 2015. Single-cell genomics of a rare environmental alphaproteobacterium provides unique insights into Rickettsiaceae evolution. *ISME Journal*, 9(11), pp.2373–2385.

Martin, D.P. *et al*., 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution*, 1(1), pp.1–5.

Martin, O.Y. *et al*., 2013. Infections with the microbe *Cardinium* in the Dolichopodidae and

other Empidoidea. *Journal of insect science (Online)*, 13: 47.

Martin, O.Y. *et al*., 2013. Infections with *Wolbachia*, *Spiroplasma*, and *Rickettsia* in the Dolichopodidae and other Empidoidea. *Infection, Genetics and Evolution*, 13, pp.317–330.

Mascarenhas, R.O. *et al*., 2016. *Wolbachia* in guilds of *Anastrepha* fruit flies (Tephritidae) and parasitoid wasps (Braconidae). *Genetics and Molecular Biology*, 39(4), pp.600–610.

Mathieu, B. *et al*., 2012. Development and validation of IIKC: An interactive identification key for *Culicoides* (Diptera: Ceratopogonidae) females from the Western Palaearctic region. *Parasites and Vectors*, 5: 137.

McDermott, E.G. *et al*., 2015. Bluetongue virus infection creates light averse *Culicoides* vectors and serious errors in transmission risk estimates. *Parasites and Vectors*, 8: 460.

Mee, P.T. *et al*., 2015. Detection of low-level *Cardinium* and *Wolbachia* infections in *Culicoides*. *Applied and Environmental Microbiology*, 81(18), pp.6177–6188.

Mee, P.T., 2017. [Thesis] Profiling the occurrence of bacterial endosymbionts in Australian Culicoides species and an investigation of possible interactions with the host insect. *University of Melbourne.*

Megahed, M.M., 1956. A culture method for *Culicoides nubeculosus* (meigen) (diptera: Ceratopogonidae) in the laboratory, with notes on the biology. *Bulletin of Entomological Research*, 47(1), pp.107–114.

Meiswinkel, R., 1989. Afrotropical *Culicoides*: a redescription of *C*. (Avaritia) *imicola* Kieffer, 1913 (Diptera: Ceratopogonidae) with description of the closely allied *C*. (A.) *bolitinos* sp. nov. reared from the dung of the African buffalo, blue wildebeest and cattle in South Africa. *The Onderstepoort journal of veterinary research*, 56(1), pp.23–39.

Meiswinkel, R., 1997. Discovery of a *Culicoides* imicola-free zone in South Africa: Preliminary notes and potential significance. *Onderstepoort Journal of Veterinary Research*, 64(1), pp.81–86.

Meiswinkel, R. *et al*., 2007. Potential new *Culicoides* vector of bluetongue virus in northern Europe. *Veterinary Record*, 161(16), pp.564–565.

Meiswinkel, R., 1994. Preliminary Wing Picture Atlas of Afrotropical Culicoides (119 Species), Pretoria, South Africa.

Meiswinkel, R. and Elbers, A.R.W., 2016. The dying of the light: Crepuscular activity in *Culicoides* and impact on light trap efficacy at temperate latitudes. *Medical and Veterinary Entomology*, 30(1), pp.53–63.

Mellor, P.S. *et al*., 1985. *Culicoides imicola*: a bluetongue virus vector in Spain and Portugal. *Veterinary Record*, 116(22), pp.589–590.

Mellor, P.S. and Boorman, J., 1995. The transmission and geographical spread of African horse sickness and bluetongue viruses. *Annals of Tropical Medicine* and *Parasitology*, 89(1), pp.1–15.

Mellor, P.S., Boorman, J. and Baylis, M., 2000. *Culicoides* biting midges: their role as arbovirus vectors. *Annual review of entomology*, 45(1), pp.307–340.

Mercer, D.R. *et al*., 2003. Biting Rates and Developmental Substrates for Biting Midges (Diptera: Ceratopogonidae) in Iquitos, Peru. *Journal of Medical Entomology*, 40(6), pp.807–812.

Merçot, H. and Charlat, S., 2004. *Wolbachia* infections in *Drosophila melanogaster* and *D. simulans*: Polymorphism and levels of cytoplasmic incompatibility. In *Genetica*. pp. 51–59.

Merhej, V. and Raoult, D., 2011. Rickettsial evolution in the light of comparative genomics. *Biological Reviews*, 86(2), pp.379–405.

Miller, W.J. and Riegler, M., 2006. Evolutionary dynamics of *w*Au-like *Wolbachia* variants in neotropical *Drosophila* spp. *Applied and Environmental Microbiology*, 72(1), pp.826–835.

Miller, W.J., Ehrman, L. and Schneider, D., 2010. Infectious speciation revisited: Impact of symbiont-depletion on female fitness and mating behavior of *Drosophila paulistorum*. *PLoS Pathogens*, 6(12): e1001214.

Mills, M.K. *et al*., 2017a. *Culicoides*–virus interactions: infection barriers and possible factors underlying vector competence. *Current Opinion in Insect Science*, 22, pp.7–15.

Mills, M.K. *et al*., 2017b. Dynamics of epizootic hemorrhagic disease virus infection within the vector, *Culicoides sonorensis* (Diptera: Ceratopogonidae) B. J. Mans, ed. *PLOS ONE*, 12(11): e0188865.

Min, K.-T. and Benzer, S., 1997. Wolbachia, normally a symbiont of *Drosophila*, can be virulent, causing degeneration and early death. *Proceedings of the National Academy of Sciences*, 94(20), pp.10792–10796.

Mira, A. and Moran, N.A., 2002. Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. *Microbial Ecology*, 44(2), pp.137–143.

Möhlmann, T., 2019. [Thesis] Mosquitoes, midges and microbiota: European vector diversity and the spread of pathogens. *Wageningen University.*

Molloy, J.C. *et al*., 2016. *Wolbachia* modulates lipid metabolism in *Aedes albopictus* mosquito cells. *Applied and Environmental Microbiology*, 82(10), pp.3109–3120.

Morag, N. *et al*., 2012. Prevalence of the symbiont *Cardinium* in *Culicoides* (Diptera: Ceratopogonidae) vector species is associated with land surface temperature. *FASEB Journal*, 26(10), pp.4025–4034.

Moran, N.A. *et al*., 1993. A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 253(1337), pp.167–171.

Moran, N.A., 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 93(7), pp.2873–2878.

Moran, N.A. and Dunbar, H.E., 2006. Sexual acquisition of beneficial symbionts in aphids. *Proceedings of the National Academy of Sciences*, 103(34), pp.12803–12806.

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.

Morchón, R. *et al.*, 2012. Heartworm disease (*Dirofilaria immitis*) and their vectors in Europe - new distribution trends. *Frontiers in Physiology*, 3: 196.

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.

Morrow, J.L. *et al*., 2014. Tropical tephritid fruit fly community with high incidence of shared *Wolbachia* strains as platform for horizontal transmission of endosymbionts. *Environmental Microbiology*, 16(12), pp.3622–3637.

Mouton, L. *et al*., 2003. Strain-specific regulation of intracellular *Wolbachia* density in multiply infected insects. *Molecular Ecology*, 12(12), pp.3459–3465.

Mullen, G.R. *et al*., 1985. Laboratory infections of *Culicoides debilipalpis* and *C. stellifer* (Diptera: Ceratopogonidae) with bluetongue virus. *Prog Clin Biol Res*, 178, pp.239–243.

Mullens, B.A. and Velten, R.K., 1994. Rearing *Culicoides variipennis sonorensis* (Diptera: Ceratopogonidae) on agar and nematodes. *Journal of Medical Entomology*, 31(1), pp.175–177.

Mullens, B.A. *et al*., 1995. Effects of temperature on virogenesis of bluetongue virus serotype 11 in *Culicoides variipennis sonorensis*. *Medical and Veterinary Entomology*, 9(1), pp.71–76.

Mullens, B.A., Sarto Monteys, V. and Przhiboro, A.A., 2008. Mermithid parasitism in Ceratopogonidae: A literature review and critical assessment of host impact and potential for biological control of *Culicoides*. *Russian Entomology Journal*, 17(1), pp.87–113.

Muller, M.J. *et al*., 1982. [Conference material] *Culicoides brevitarsis* (Diptera: Ceratopogonidae) as a vector of arboviruses in Australia. *Proceedings of the 3rd Symposium on Arbovirus Research in Australia.*

Muñoz-Gómez, S.A. *et al*., 2019. An updated phylogeny of the Alphaproteobacteria reveals that the parasitic Rickettsiales and Holosporales have independent origins. *eLife*, 8: e42535.

Nakamura, Y. *et al*., 2011. Differentially expressed genes in silkworm cell cultures in response to infection by *Wolbachia* and *Cardinium* endosymbionts. *Insect Molecular Biology*, 20(3), pp.279–289.

Nakamura, Y. *et al*., 2009. Prevalence of *Cardinium* bacteria in planthoppers and spider

mites and taxonomic revision of “Candidatus Cardinium hertigii” based on detection of a new *Cardinium* group from biting midges. *Applied and Environmental Microbiology*, 75(21), pp.6757–6763.

Nakayama, S. *et al*., 2015. Can maternally inherited endosymbionts adapt to a novel host? Direct costs of *Spiroplasma* infection, but not vertical transmission efficiency, evolve rapidly after horizontal transfer into *D. melanogaster*. *Heredity*, 114(6), pp.539–543.

Nayduch, D. *et al*., 2014. Studying *Culicoides* vectors of BTV in the post-genomic era: Resources, bottlenecks to progress and future directions. *Virus Research*, 182, pp.43–49.

Neelakanta, G. *et al*., 2010. *Anaplasma phagocytophilum* induces *Ixodes scapularis* ticks to express an antifreeze glycoprotein gene that enhances their survival in the cold. *Journal of Clinical Investigation*, 120(9), pp.3179–3190.

Nelliot, A., Bond, N. and Hoshizaki, D.K., 2006. Fat-body remodeling in *Drosophila melanogaster*. *genesis*, 44(8), pp.396–400.

Nevill, E.M., 1969. The morphology of the immature stages of some South African *Culicoides* species (Diptera: Ceratopogonidae). *Onderstepoort Journal of Veterinary Research*, 36(2), pp.265–283.

Nguyen, D.T., Spooner-Hart, R.N. and Riegler, M., 2016. Loss of *Wolbachi*a but not *Cardinium* in the invasive range of the Australian thrips species, *Pezothrips kellyanus*. *Biological Invasions*, 18(1), pp.197–214.

Nguyen, L.T. *et al*., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1), pp.268–274.

Nielsen, S.A. and Kristensen, M., 2015. Delineation of *Culicoides* species by morphology and barcode exemplified by three new species of the subgenus *Culicoides* (Diptera: Ceratopogonidae) from Scandinavia. *Parasites* and *Vectors*, 8(1): 151.

Nielsen, S.A. and Kristensen, M., 2011. Morphological and molecular identification of species of the Obsoletus group (Diptera: Ceratopogonidae) in Scandinavia. *Parasitology Research*, 109(4), pp.1133–1141.

Nikoh, N. *et al*., 2014. Evolutionary origin of insect-*Wolbachia* nutritional mutualism. *Proceedings of the National Academy of Sciences*, 111(28), pp.10257–10262.

Nixon, G., 1980. Handbooks for the Identification of British Insects: Diapriidae (Diapriinae), Royal Entomological Society of London.

Nixon, G., 1957. Hymenoptera, Proctotrupoidea, Diapriidae, subfamily Belytinae. Handbooks for the Identification of British Insects no. 8(3dii), Royal Entomological Society of London.

Noda, H. *et al*., 2001. *Wolbachia* infection shared among planthoppers (Homoptera: Delphacidae) and their endoparasite (Strepsiptera: Elenchidae): a probable case of interspecies transmission. *Molecular Ecology*, 10(8), pp.2101–2106.

Noel, G.R., 2006. "Candidatus Paenicardinium endonii", an endosymbiont of the plant-parasitic nematode *Heterodera glycines* (Nemata: Tylenchida), affiliated to the phylum Bacteroidetes. *International Journal of Systematic and Evolutionary Microbiology*, 56(7), pp.1697–1702.

Nolan, D. V. *et al*., 2008. Incursion and range expansion in the bluetongue vector *Culicoides imicola* in the Mediterranean basin: A phylogeographic analysis. *Medical and Veterinary Entomology*, 22(4), pp.340–351.

Nolan, D. V. *et al*., 2007. Rapid diagnostic PCR assays for members of the *Culicoides obsoletus* and *Culicoides pulicaris* species complexes, implicated vectors of bluetongue virus in Europe. *Veterinary Microbiology*, 124(1–2), pp.82–94.

Nomikou, K. *et al*., 2015. Widespread reassortment shapes the evolution and epidemiology of bluetongue virus following European invasion G. Tiao, ed. *PLOS Pathogens*, 11(8): e1005056.

Nunamaker, R.A. *et al*., 1997. bluetongue virus in laboratory-reared *Culicoides variipennis sonorensis*: Applications of dot-blot, ELISA, and immunoelectron microscopy. *Journal of Medical Entomology*, 34(1), pp.18–23.

O'Neill, S., 2012. [Newsletter] Eliminate dengue our challenge. *Program Newsletter*. Available at: http://www.eliminatedengue.com/library/media/document/edp-

program_overview.pdf [Accessed 13 September, 2019].

Ogata, H. *et al*., 2005. The genome sequence of *Rickettsia felis* identifies the first putative conjugative plasmid in an obligate intracellular parasite. *PLoS Biology*, 3(8): e248.

Ogden, N.H. *et al*., 2003. Transmission of *Anaplasma phagocytophilum* to *Ixodes ricinus* ticks from sheep in the acute and post-acute phases of infection. *Infection and Immunity*, 71(4), pp.2071–2078.

Okonechnikov, K. *et al*., 2012. Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics*, 28(8), pp.1166–1167.

Oliver, K.M. *et al*., 2003. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proceedings of the National Academy of Sciences*, 100(4), pp.1803–1807.

Oliver, K.M. *et al*., 2010. Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. *Annual Review of Entomology*, 55(1), pp.247–266.

Osuna-Ramos, J.F., Reyes-Ruiz, J.M. and del Ángel, R.M., 2018. The role of host cholesterol during Flavivirus infection. *Frontiers in Cellular and Infection Microbiology*, 8: 388.

Owsley, W.B., 1946. The comparative morphology of internal structures of the Asilidae (Diptera). *Annals of the Entomological Society of America*, 39, p.52.

Pagès, N. *et al*., 2009. Identification of cryptic species of *Culicoides* (Diptera: Ceratopogonidae) in the subgenus Culicoides and development of species-specific PCR assays based on barcode regions. *Veterinary Parasitology*, 165(3–4), pp.298–310.

Pagès, N. *et al*., 2017. First detection of *Wolbachia*-infected *Culicoides* (Diptera: Ceratopogonidae) in Europe: *Wolbachia* and *Cardinium* infection across *Culicoides* communities revealed in Spain. *Parasites* and *Vectors*, 10(1), p.582.

Pagès, N. *et al*., 2018. Schmallenberg virus detection in *Culicoides* biting midges in Spain: First laboratory evidence for highly efficient infection of *Culicoides* of the Obsoletus complex and *Culicoides imicola*. *Transboundary and Emerging Diseases*, 65(1), pp.e1–e6.

Pan, X. *et al*., 2012. *Wolbachia* induces reactive oxygen species (ROS)-dependent activation

of the Toll pathway to control dengue virus in the mosquito *Aedes aegypti*. *Proceedings of the National Academy of Sciences*, 109(1), pp.e23–e31.

Pang, T., Mak, T. K. and Gubler, D. J. 2017 Prevention and control of dengue—the light at the end of the tunnel. *The Lancet Infectious Diseases*, 17(3), pp.e79–e87.

Panteleev, D.. Y. *et al*., 2007. The endosymbiotic bacterium *Wolbachia* enhances the nonspecific resistance to insect pathogens and alters behavior of *Drosophila melanogaster*. *Russian Journal of Genetics*, 43(9), pp.1066–1069.

Papadopoulou, A. *et al*., 2011. Testing the species–genetic diversity correlation in the Aegean archipelago: Toward a haplotype-based macroecology? *The American Naturalist*, 178(2), pp.241–255.

Park, D. S. *et al.*, 2011. Barcoding bugs: Dna-based identification of the true bugs (insecta: Hemiptera: Heteroptera). *PLoS ONE,* 6(4): e18749.

Parola, P. *et al*., 2013. Update on tick-borne rickettsioses around the world: a geographic approach. *Clinical Microbiology Reviews*, 26(4), pp.657–702.

Peng, Y. *et al*., 2008. *Wolbachia* infection alters olfactory-cued locomotion in *Drosophila* spp. *Applied and Environmental Microbiology*, 74(13), pp.3943–3948.

Penz, T. *et al*., 2012. Comparative genomics suggests an independent origin of cytoplasmic incompatibility in *Cardinium hertigii* N. A. Moran, ed. *PLoS Genetics*, 8(10): e1003012.

Perlman, S.J., Hunter, M.S. and Zchori-Fein, E., 2006. The emerging diversity of *Rickettsia*. *Proceedings of the Royal Society B*, 273(1598), pp.2097–2106.

Perlman, S.J., Kelly, S.E. and Hunter, M.S., 2008. Population biology of cytoplasmic incompatibility: maintenance and spread of *Cardinium* symbionts in a parasitic wasp. *Genetics*, 178(2), pp.1003–1011.

Perotti, M.A. *et al*., 2006. *Rickettsia* as obligate and mycetomic bacteria. *The FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 20(13), pp.2372–2374.

Phuc, H. *et al.* 2007. Late-acting dominant lethal genetic systems and mosquito control. *BMC Biology,* 5: 11.

Pietri, J.E., DeBruhl, H. and Sullivan, W., 2016. The rich somatic life of *Wolbachia*. *MicrobiologyOpen*, 5(6), pp.923–936.

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.

Pinheiro, F.P., Travassos da Rosa, A.P.A. and Travassos da Rosa, J.F.S., 1981. Oropouche virus. I. A review of clinical, epidemiological, and ecological findings. *American Journal of Tropical Medicine and Hygiene*, 30(1), pp.149–160.

Pirbright Institute, 2009. The economic and social impact of the Institute for Animal Health's work on bluetongue disease (BTV-8). Available at: http://www.pirbright.ac.uk/ecosoc/docs/Blue-Tongue-case-study.pdf [Accessed December 8, 2016].

Plantard, O. *et al*., 2012. Detection of *Wolbachia* in the tick *Ixodes ricinus* is due to the presence of the Hymenoptera endoparasitoid *Ixodiphagus hookeri*. *PLoS ONE*, 7(1): e30692.

Pons, I. *et al*., 2019. Circulation of the cultivable symbiont *Serratia symbiotica* in aphids is mediated by plants. *Frontiers in Microbiology*, 10, p.764.

Purse, B. V. *et al*., 2005. Climate change and the recent emergence of bluetongue in Europe. *Nature Reviews Microbiology*, 3(2), pp.171–181.

Qiagen, 2006. Purification of total DNA from insects using the DNeasy® Blood and Tissue Kit Title. Available at: https://www.qiagen.com/gb/resources/resourcedetail?id=cabd47a4-cb5a-4327-b10d-d90b8542421eandlang=en [Accessed February 5, 2018].

Quantum GIS Development Team, 2016. Quantum GIS Geographic Information System. *Open Source Geospatial Foundation Project*.

Ramage, T. *et al*., 2017. A DNA barcode-based survey of terrestrial arthropods in the Society Islands of French Polynesia: host diversity within the SymbioCode Project. *European Journal of Taxonomy*, 272, pp. 1-13.

Rasgon, J.L., Cornel, A.J. and Scott, T.W., 2006. Evolutionary history of a mosquito endosymbiont revealed through mitochondrial hitchhiking. *Proceedings of the Royal Society B: Biological Sciences*, 273(1594), pp.1603–1611.

Rasmussen, L.D. *et al*., 2012. Culicoids as vectors of Schmallenberg virus. *Emerging Infectious Diseases*, 18(7): pp.1204–1206.

Ratnasingham, S. and Hebert, P.D.N., 2007. BOLD: The Barcode of Life Data System: Barcoding. *Molecular Ecology Notes*, 7(3), pp.355–364.

Rawlings, P., 1996. A key, based on wing patterns of biting midges (genus *Culicoides* Latreille - Diptera: Ceratopogonidae) in the Iberian Peninsula, for use in epidemiological studies. *Graellsia*, 52, pp.57–71.

Raychoudhury, R. *et al*., 2009. Modes of acquisition of *Wolbachia*: Horizontal transfer, hybrid introgression, and codivergence in the *Nasonia* species complex. *Evolution*, 63(1), pp.165–183.

Raychoudhury, R. *et al*., 2010. Phylogeography of *Nasonia vitripennis* (Hymenoptera) indicates a mitochondrial–*Wolbachia* sweep in North America. *Heredity*, 104(3), pp.318–326.

Reed, S.C.O. *et al*., 2014. *Rickettsia* actin-based motility occurs in distinct phases mediated by different actin nucleators. *Current Biology*, 24(1), pp.98–103.

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

Řezáč, M. *et al*., 2014a. Integrative taxonomy and evolutionary history of a newly revealed spider *Dysdera ninnii* complex (Araneae: Dysderidae). *Zoological Journal of the Linnean Society*, 172(2), pp.451–474.

Rigaud, T. and Juchault, P., 1995. Success and failure of horizontal transfers of feminizing *Wolbachia* endosymbionts in woodlice. *Journal of Evolutionary Biology*, 8(2), pp.249–255.

Rio, R.V.M. *et al*., 2012. Insight into the transmission biology and species-specific functional

capabilities of Tsetse (Diptera: Glossinidae) obligate symbiont *Wigglesworthia*. *mBio*, 3(1), pp.1–13.

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

Ritchie, S.A., 2018. *Wolbachia* and the near cessation of dengue outbreaks in Northern Australia despite continued dengue importations via travellers. *Journal of Travel Medicine*, 25(1): tay084.

Robeson, M.S. *et al*., 2018. Assessing the utility of metabarcoding for diet analyses of the omnivorous wild pig (*Sus scrofa*). *Ecology and Evolution*, 8(1), pp.185–196.

Ros, V.I.D. and Breeuwer, J.A.J., 2009. The effects of, and interactions between, *Cardinium* and *Wolbachia* in the doubly infected spider mite *Bryobia sarothamni*. *Heredity*, 102(4), pp.413–422.

Roshdy, M.A., 1968. A rickettsia-like microorganism in the tick *Ornithodoros savignyi*; Observations on its structure and distribution in the tissues of the tick. *Journal of Invertebrate Pathology*, 11(2), pp.155–169.

Rossi, P. *et al.*, 2015. Mutual exclusion of *Asaia* and *Wolbachia* in the reproductive organs of mosquito vectors. *Parasites and Vectors,* 8:278.

Rothwell, C. *et al*., 2009. Cholesterol biosynthesis modulation regulates dengue viral replication. *Virology*, 389(1–2), pp.8–19.

Roy, P., 2017. Bluetongue virus structure and assembly. *Current Opinion in Virology*, 24, pp.115–123.

Roy, P., P.C. Mertens, P. and Casal, I., 1994. African horse sickness virus structure. *Comparative Immunology, Microbiology and Infectious Diseases*, 17(3–4), pp.243–273.

RStudio, Rs., 2015. RStudio: Integrated Development for R. *[Online] RStudio, Inc., Boston, MA*.

Russell, J.A. *et al*., 2009. Specialization and geographic isolation among *Wolbachia* symbionts from ants and lycaenid butterflies. *Evolution*, 63(3), pp.624–640.

Russell, J.A. and Moran, N.A., 2005. Horizontal transfer of bacterial symbionts: Heritability and fitness effects in a novel aphid host. *Applied and Environmental Microbiology*, 71(12), pp.7987–7994.

Russell, S.L., Chappell, L. and Sullivan, W., 2019. A symbiont's guide to the germline. In *Current Topics in Developmental Biology*. pp. 315–351.

Russell, S.L., Corbett-Detig, R.B. and Cavanaugh, C.M., 2017. Mixed transmission modes and dynamic genome evolution in an obligate animal–bacterial symbiosis. *The ISME Journal*, 11(6), pp.1359–1371.

Sabree, Z.L., Kambhampati, S. and Moran, N.A., 2009a. Nitrogen recycling and nutritional provisioning by *Blattabacterium*, the cockroach endosymbiont. *Proceedings of the National Academy of Sciences*, 106(46), pp.19521–19526.

Sacchi, L. *et al*., 1998. Ultrastructural studies of the fat body and bacterial endosymbionts of *Cryptocercus punctulatus* Scudder (Blattaria: Cryptocercidae). *Symbiosis*, 25(1–3), pp.251–269.

Saito, T.B. and Walker, D.H., 2015. A tick vector transmission model of monocytotropic Ehrlichiosis. *Journal of Infectious Diseases*, 212(6), pp.968–977.

Sanders, C.J. *et al*., 2011. Influence of season and meteorological parameters on flight activity of *Culicoides* biting midges. *Journal of Applied Ecology*, 48(6), pp.1355–1364.

Santibáñez, S. *et al*., 2013. Usefulness of rickettsial PCR assays for the molecular diagnosis of human Rickettsioses. *Enfermedades Infecciosas y Microbiologia Clinica*, 31(5), pp.283–288.

Sarvašová, A. *et al*., 2017. Description of *Culicoides* (Culicoides) bysta n. sp., a new member of the Pulicaris group (Diptera: Ceratopogonidae) from Slovakia. *Parasites* and *Vectors*, 10(1): 279.

Sassera, D. *et al*., 2011. Phylogenomic evidence for the presence of a flagellum and cbb 3 oxidase in the free-living mitochondrial ancestor. *Molecular Biology and Evolution*, 28(12), pp.3285–3296.

Schmidt, T.L. *et al*., 2017. Local introduction and heterogeneous spatial spread of dengue-

suppressing *Wolbachia* through an urban population of *Aedes aegypti*. *PLOS Biology*, 15(5): e2001894.

Schön, I. *et al*., 2019. Novel *Cardinium* strains in non-marine ostracod (Crustacea) hosts from natural populations. *Molecular Phylogenetics and Evolution*, 130, pp.406–415.

Schrallhammer, M. *et al*., 2013. 'Candidatus Megaira polyxenophila' gen. nov., sp. nov.: Considerations on evolutionary history, host range and shift of early divergent Rickettsiae S. A. Ralph, ed. *PLoS ONE*, 8(8): e72581.

Schulenburg, J.H.G.V.D. *et al*., 2002. History of infection with different male-killing bacteria in the two-spot ladybird beetle *Adalia bipunctata* revealed through mitochondrial DNA sequence analysis. *Genetics*, 10(3), pp.1075–1086.

Schultz, M.J. *et al*., 2017. Variable inhibition of zika virus replication by different *Wolbachia* strains in mosquito cell cultures. *Journal of Virology*, 91(14).

Searle, K.R. *et al*., 2014. Environmental drivers of *Culicoides* phenology: How important is species-specific variation when determining disease policy? *PLoS ONE*, 9(11): e111876.

Ségard, A. *et al*., 2018. Schmallenberg virus in *Culicoides* Latreille (Diptera: Ceratopogonidae) populations in France during 2011-2012 outbreak. *Transboundary and Emerging Diseases*, 65(1), pp.e94–e103.

Serbus, L.R. and Sullivan, W., 2007. A cellular basis for *Wolbachia* recruitment to the host germline. *PLoS Pathogens*, 3(12): e190.

Service, M.W., 1977. A critical review of procedures for sampling populations of adult mosquitoes. *Bulletin of Entomological Research*, 67(3), pp.343–382.

Shaw, A.E. *et al*., 2013. Reassortment between two serologically unrelated bluetongue virus strains is flexible and can involve any genome segment. *Journal of Virology*, 87(1), pp.543–557.

Sheehan, K.B. *et al*., 2016. Identification and characterization of a candidate *Wolbachia* pipientis type IV effector that interacts with the actin cytoskeleton. *mBio*, 7(4): e00622-16.

Shoemaker, D.D. *et al*., 2004. Decreased diversity but increased substitution rate in host

mtDNA as a consequence of *Wolbachia* endosymbiont infection. *Genetics*, 168(4), pp.2049–2058.

Sick, F. *et al*., 2019. *Culicoides* biting midges—underestimated vectors for arboviruses of public health and veterinary importance. *Viruses*, 11(4): 376.

Simhadri, R. K. *et al.* 2017. The gut commensal microbiome of *Drosophila melanogaster* is modified by the endosymbiont *Wolbachia*. *mSphere*, 2(5), pp.e00287-17.

Simões, P.M. *et al*., 2011. *Wolbachia* detection: An assessment of standard PCR Protocols. *Molecular Ecology Resources*, 11(3), pp.567–572.

Simoncini, L. *et al*., 2001. Real-time PCR for quantification of the bacterial endosymbionts (*Wolbachia*) of filarial nematodes. *Parassitologia,* 43(4), pp.173-178.

Sinkins, S.P., 2013. *Wolbachia* and arbovirus inhibition in mosquitoes. *Future Microbiology*, 8(10), pp.1249–1256.

Sintupachee, S. *et al*., 2006. Closely related *Wolbachia* strains within the pumpkin arthropod community and the potential for horizontal transmission via the plant. *Microbial Ecology*, 51(3), pp.294–301.

Siozios, S. *et al*., 2019a. The draft genome of strain *c*Cpun from biting midges confirms insect *Cardinium* are not a monophyletic group and reveals a novel gene family expansion in a symbiont. *PeerJ*, 7: e6448.

Smith, A. H. *et al.,* 2015. Patterns, causes and consequences of defensive microbiome dynamics across multiple scales. *Molecular Ecology*, 24(5), pp.1135-1149.

Smith, M.A. *et al*., 2012. *Wolbachia* and DNA barcoding insects: Patterns, potential, and problems J. H. Badger, ed. *PLoS ONE*, 7(5): e36514.

Smith, T.A. *et al*., 2015. A Coxiella-like endosymbiontis a potential vitamin source for the lone star tick. *Genome Biology and Evolution*, 7(3), pp.831–838.

Snyder, A.K. *et al*., 2010. Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts. *Proceedings. Biological sciences / The Royal Society*, 277(1692), pp.2389–2397.

Snyder, D. *et al*., 2016. Insecticidal sugar baits for adult biting midges. *Medical and Veterinary Entomology*, 30(2), pp.209–217.

Sokolova, M.I., Zinkevich, N.S. and Zakharov, I.A., 2002. Bacteria in ovarioles of females from maleless families of ladybird beetles *Adalia bipunctata* L. (Coleoptera: Coccinellidae) naturally infected with *Rickettsia*, *Wolbachia*, and *Spiroplasma*. *Journal of Invertebrate Pathology*, 79(2), pp.72–79.

Song, H. *et al*., 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences*, 105(36), pp.13486–13491.

Song, S. *et al*., 2018. Diversity of *Rickettsia* species in border regions of northwestern China. *Parasites* and *Vectors*, 11(1): 634.

Stalhut, J.K. *et al*., 2010. The mushroom habitat as an ecological arena for global exchange of *Wolbachia*. *Molecular Ecology*, 19(9), pp.1940–1952.

Stamatakis, A., 2014. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics*, 30(9): 1312–1313.

Steinke, S., Lühken, R. and Kiel, E., 2014. Assessment of the abundance of *Culicoides chiopterus* and *Culicoides dewulfi* in bovine dung: A comparison of larvae extraction techniques and emergence traps. *Veterinary Parasitology*, 205(1–2), pp.255–262.

Stokes, J.E. *et al*., 2018. Survey to determine the farm-level impact of Schmallenberg virus during the 2016–2017 United Kingdom lambing season. *Veterinary Record*, 183(22), pp.690–690.

Stokes, J.E., Baylis, M. and Duncan, J.S., 2016. A freedom from disease study: Schmallenberg virus in the south of England in 2015. *Veterinary Record*, 179(17): 435.

Stoll, S. *et al*., 2010. Bacteriocyte dynamics during development of a holometabolous insect, the carpenter ant *Camponotus floridanus*. *BMC Microbiology*, 10(1): 308.

Storto, P.D., 1995. The Insect Ovary: Ultrastructure, previtellogenic growth and evolution. *Invertebrate Reproduction* and *Development*, 28(3), pp.217–218.

Strunov, A. *et al.*, 2017. Restricted distribution and lateralization of mutualistic *Wolbachia* in

the *Drosophila* brain. *Cellular Microbiology*, 19(1): e12639.

Suitor, E.C. and Weiss, E., 1961. Isolation af a rickettsia-like microorganism (*Wolbachia persica*, n. sp.) From *Argas persicus* (Oken). *The Journal of Infectious Diseases*, 108(1), pp.95–106.

Sun, W. K. C., 1974. Laboratory colonization of biting midges (Diptera: Ceratopogonidae). *Journal of Medical Entomology*, 11(1), pp. 71–73.

Szklarzewicz, T. *et al*., 2007. Structure of the ovaries and oogenesis in *Cixius nervosus* (Cixiidae), *Javesella pellucida* and *Conomelus anceps* (Delphacidae) (Insecta, Hemiptera, Fulgoromorpha). *Arthropod Structure* and *Development*, 36(2), pp.199–207.

Szmaragd, C. *et al*., 2010. The spread of bluetongue virus serotype 8 in Great Britain and its control by vaccination. *PLoS ONE*, 5(2): e9353.

Takken, W. and Knols, B. G. J. 2007. [Book] Emerging pests and vector-borne diseases in Europe: ecology and control of vector-borne diseases, *Wageningen Academic Publishers*, pp. 97-185.

Tamura, K. *et al*., 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30(12), pp.2725–2729.

Tedeschi, R. and Bertaccini, A., 2019. Transovariol transmission in insect vectors. In *Phytoplasmas: Plant Pathogenic Bacteria - II: Transmission and Management of Phytoplasma - Associated Diseases*. Springer, p. 116.

Teixeira, L., Ferreira, Á. and Ashburner, M., 2008. The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. *PLoS Biology*, 6(12): e1000002.

Teysseire, N., Chiche-Portiche, C. and Raoult, D., 1992. Intracellular movements of *Rickettsia conorii* and *R. typhi* based on actin polymerization. *Research in Microbiology*, 143(9), pp.821–829.

Thompson, G.M., Jess, S. and Murchie, A.K., 2013. Differential emergence of *Culicoides* (Diptera: Ceratopogonidae) from on-farm breeding substrates in Northern Ireland. *Parasitology*, 140(6), pp.699–708.

Toomey, M.E. *et al*., 2013. Evolutionarily conserved *Wolbachia*-encoded factors control pattern of stem-cell niche tropism in *Drosophila* ovaries and favor infection. *Proceedings of the National Academy of Sciences,* 110(26), pp.10788–10793.

Tortosa, P. *et al*., 2010. Wolbachia age-sex-specific density in *Aedes albopictus*: A host evolutionary response to Cytoplasmic Incompatibility? A. K. Chippindale, ed. *PLoS ONE*, 5(3): e9700.

Tree, M.O. *et al*., 2019. Dengue virus reduces expression of low-density lipoprotein receptor-related protein 1 to facilitate replication in *Aedes aegypti*. *Scientific Reports*, 9(1): 6352.

Turelli, M. *et al*., 2018. Rapid global spread of *w*Ri-like *Wolbachia* across multiple Drosophila. *Current Biology*, 28(6), pp.963–971.

Uhlmann, K.R. *et al*., 2014. Species determination of *Culicoides* biting midges via peptide profiling using matrix-assisted laser desorption ionization mass spectrometry. *Parasites* and *Vectors*, 7(1): 392.

Upla, P., Hyypiä, T. and Marjomäki, V., 2009. Role of lipid rafts in virus infection. *Future Virology*, 4(5), pp.487–500.

Vaidya, G., Lohman, D.J. and Meier, R., 2011. SequenceMatrix: Concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics,* 27(2), pp.171-180.

Vallet-Gely, I., Lemaitre, B. and Boccard, F., 2008. Bacterial strategies to overcome insect defences. *Nature Reviews Microbiology,* 6(4)*,* pp.302–313.

Vandekerckhove, T.T; Swings, J.G; Mertens, J; Gillis, M., 2005. *Rickettsia limoniae* sp. nov., intracellular symbiont of the cranefly Limonia chorea (Hexapoda, Diptera, Limoniidae) and so far the earliest diverging species in *Rickettsia* phylogeny. *unpublished*. Available at: https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Infoandid=172566 [Accessed 20 October, 2016].

Vannini, C. *et al*., 2005. A bacterium belonging to the Rickettsiaceae family inhabits the

cytoplasm of the marine ciliate *Diophrys appendiculata* (Ciliophora, Hypotrichia). *Microbial Ecology*, 49(3), pp.434–442.

Vavre, F. *et al*., 1999. Phylogenetic evidence for horizontal transmission of *Wolbachia* in host- parasitoid associations. *Molecular Biology and Evolution*, 16(12), pp.1711–1723.

Veiga, J. *et al*., 2018. *Culicoides paolae* and *C. circumscriptus* as potential vectors of avian haemosporidians in an arid ecosystem. *Parasites* and *Vectors*, 11(1): 524.

Veronesi, E. *et al.*, 2009. Life-history parameters of *Culicoides* (Avaritia) *imicola* Kieffer in the laboratory at different rearing temperatures. *Veterinary Parasitology,* 163(4), pp.370-373.

Venail, R. *et al*., 2012. Assessing diversity and abundance of vector populations at a national scale: Example of *Culicoides* surveillance in France after bluetongue virus emergence. In *Arthropods as Vectors of Emerging Diseases*. pp. 77–102.

Veneti, Z. *et al*., 2004. Heads or tails: Host-parasite interactions in the *Drosophila*-*Wolbachia* system. *Applied and Environmental Microbiology*, 70(9), pp.5366–5372.

Verwoerd, D.W. *et al*., 1972. Structure of the bluetongue virus capsid. *J. Virol.*, 10(4), pp.783–794.

Viennet, E. *et al*., 2011. Assessment of vector/host contact: comparison of animal-baited traps and UV-light/suction trap for collecting *Culicoides* biting midges (Diptera: Ceratopogonidae), vectors of Orbiviruses. *Parasites* and *Vectors*, 4(1): 119.

Voronin, D. *et al*., 2016. Glucose and glycogen metabolism in *Brugia malayi* is associated with *Wolbachia* symbiont fitness C. K. S. Carlow, ed. *PLOS ONE*, 11(4): e0153812.

Walker, T. *et al.*, 2011. The *w*Mel *Wolbachia* strain blocks dengue and invades caged *Aedes aegypti* populations. *Nature*, 476(7361), pp. 450–453.

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

Ward, R.D. and Holmes, B.H., 2007. An analysis of nucleotide and amino acid variability in the barcode region of cytochrome c oxidase I (cox1) in fishes. *Molecular Ecology Notes*, 7(6), pp.899–907.

Watanabe, K. *et al*., 2014. Intrasperm vertical symbiont transmission. *Proceedings of the National Academy of Sciences*, 111(20), pp.7433–7437.

Weaver, S.C. and Reisen, W.K., 2010. Present and future arboviral threats. *Antiviral Research*, 85(2), pp.328–345.

Weeks, A.R., Marec, F. and Breeuwer, J.A.J., 2001. A mite species that consists entirely of haploid females. *Science,* 292(5526), pp.2479-82.

Weeks, A.R. and Stouthamer, R., 2004. Increased fecundity associated with infection by a Cytophaga –like intracellular bacterium in the predatory mite, *Metaseiulus occidentalis*. *Proceedings of the Royal Society of London. Series B: Biological Sciences,* 271(4): S193–S195.

Weinert, L.A. *et al*., 2009. Evolution and diversity of *Rickettsia* bacteria. *BMC Biology*, 7(1): 6.

Weinert, L.A. *et al*., 2015. The incidence of bacterial endosymbionts in terrestrial arthropods. *Proceedings of the Royal Society B: Biological Sciences*, 282(1807): 20150249.

Weiss, B.L. *et al*., 2006. Interspecific transfer of bacterial endosymbionts between tsetse fly species: Infection establishment and effect on host fitness. *Applied and Environmental Microbiology*, 72(11), pp.7013–7021.

Wellby, M.P. *et al*., 1996. Effect of temperature on survival and rate of virogenesis of African horse sickness virus in *Culicoides variipennis sonorensis* (Diptera: Ceratopogonidae) and its significance in relation to the epidemiology of the disease. *Bulletin of Entomological Research*, 86(6), pp.715–720.

Wernike, K. *et al*., 2014. Schmallenberg virus—Two years of experiences. *Preventive Veterinary Medicine*, 116(4), pp.423–434.

Wernike, K., Hoffmann, B. and Beer, M., 2013. Schmallenberg Virus. In *Developments in Biologicals*. pp. 175–182.

Werren, J.H. *et al*., 1994. Rickettsial relative associated with male killing in the ladybird beetle (*Adalia bipunctata*). *Journal of Bacteriology*, 176(2), pp.388–394.

Werren, J.H., Baldo, L. and Clark, M.E., 2008. *Wolbachia*: master manipulators of

invertebrate biology. *Nature Reviews Microbiology*, 6(10), pp.741–751.

Werren, J.H. and O'Neill, S.L., 1997. The evolution of heritable symbionts. In *Influential Passengers: Inherited Microorganisms and Arthropod Reproduction*. pp. 3–41.

Werren, J.H., Zhang, W. and Guo, L.R., 1995. Evolution and phylogeny of *Wolbachia*: reproductive parasites of arthropods. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 261(1360), pp.55–63.

White, J.A. *et al*., 2011. Endosymbiont costs and benefits in a parasitoid infected with both *Wolbachia* and *Cardinium*. *Heredity*, 106(4), pp.585–591.

White, S.M. *et al*., 2017. Mechanistic model for predicting the seasonal abundance of *Culicoides* biting midges and the impacts of insecticide control. *Parasites* and *Vectors*, 10(1): 162.

Whitworth, T.. *et al*., 2007. DNA barcoding cannot reliably identify species of the blowfly genus *Protocalliphora* (Diptera: Calliphoridae). *Proceedings of the Royal Society B: Biological Sciences*, 274(1619), pp.1731–1739.

Wiemers, M. and Fiedler, K., 2007. Does the DNA barcoding gap exist? – a case study in blue butterflies (Lepidoptera: Lycaenidae). *Frontiers in Zoology*, 4(1): 8.

Wiens, J.J., 2006. Missing data and the design of phylogenetic analyses. *Journal of Biomedical Informatics*, 39(1), pp.34–42.

Wild, C., 1972. A study of the inquilines of some Victorian social insects. La Trobe University, Melbourne.

Williams, K.P., Sobral, B.W. and Dickerman, A.W., 2007. A robust species tree for the alphaproteobacteria. *Journal of Bacteriology*, 189(13), pp.4578–4586.

Wilson, A.J. and Mellor, P.S., 2009. Bluetongue in Europe: past, present and future. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 364(1530), pp.2669–2681.

Winston, P.W. and Bates, D.H., 1960. Saturated solutions for the control of humidity in biological research. *Ecology*, 41(1), pp.232–237.

Wise De Valdez, M. R. *et al.,* 2011. Genetic elimination of dengue vector mosquitoes. *Proceedings of the National Academy of Sciences of the United States of America*, 108(12), pp.4772–4775.

Wittmann, E.J. and Baylis, M., 2000. Climate change: Effects on *Culicoides*-transmitted viruses and implications for the UK. *The Veterinary Journal*, 160(2), pp.107–117.

Wittmann, E.J., Mellor, P.S. and Baylis, M., 2002. Effect of temperature on the transmission of orbiviruses by the biting midge, *Culicoides sonorensis*. *Medical and Veterinary Entomology*, 16(2), pp.147–156.

Wolfgang, A. *et al*., 2009. Evidence for low-titre infections in insect symbiosis: *Wolbachia* in the bark beetle *Pityogenes chalcographus* (Coleoptera, Scolytinae). *Environmental Microbiology,* 11(8), pp.1923-33.

Wood, H. and Artsob, H., 2012. Spotted fever group rickettsiae: A brief review and a canadian perspective. *Zoonoses and Public Health*, 59, pp.65–79.

World Mosquito Program, 2019. The World Mosquito Program's project sites around the world. *World Mosquito Program*. Available at: https://www.worldmosquitoprogram.org/ [Accessed August 30, 2019].

Wright, D.W., Geering, Q.A. and Ashby, D.G., 1947. The insect parasites of the Carrot fly, *Psila rosae* , fab. *Bulletin of Entomological Research*, 37(4), pp.507–529.

Xi, Z., Ramirez, J.L. and Dimopoulos, G., 2008. The *Aedes aegypti* toll pathway controls dengue virus infection. *PLoS Pathogens*, 4(7): e1000098.

Xie, J., Vilchez, I. and Mateos, M., 2010. *Spiroplasma* bacteria enhance survival of *Drosophila hydei* attacked by the parasitic wasp *Leptopilina heterotoma*. *PLoS ONE*, 5(8): e12149.

Xiong, M. *et al*., 2017. Molecular dietary analysis of two sympatric felids in the Mountains of Southwest China biodiversity hotspot and conservation implications. *Scientific Reports*, 7(1): 41909.

Yan, D.F. *et al*., 2018. High-throughput eDNA monitoring of fungi to track functional recovery in ecological restoration. *Biological Conservation*, 217, pp. 113-120.

Yang, A., Narechania, A. and Kim, E., 2016. Rickettsial endosymbiont in the "early-diverging"

streptophyte green alga *Mesostigma viride*. *Journal of Phycology*, 52(2), pp.219–229.

Yates, J.A., Lowrie, R.C. and Eberhard, M.L., 1982. Development of *Tetrapetalonema llewellyni* to the infective stage in *Culicoides hollensis*. *The Journal of Parasitology*, 68(2): 293.

Ye, Y. H. *et al.*, 2017. Gut microbiota in *Drosophila melanogaster* interacts with *Wolbachia* but does not contribute to *Wolbachia*-mediated antiviral protection. *Journal of Invertebrate Pathology*, 143, pp.18-25.

Yildirim, A. *et al*., 2019. Genetic diversity of *Culicoides* species within the Pulicaris complex (Diptera: Ceratopogonidae) in Turkey inferred from mitochondrial *COI* gene sequences. *Acta Tropica*, 190, pp.380–388.

Zchori-Fein, E. *et al*., 2001. A newly discovered bacterium associated with parthenogenesis and a change in host selection behavior in parasitoid wasps. *Proceedings of the National Academy of Sciences*, 98(22), pp.12555–12560.

Zchori-Fein, E., 2004. Characterization of a "Bacteroidetes" symbiont in *Encarsia* wasps (Hymenoptera: Aphelinidae): proposal of "Candidatus Cardinium hertigii." *International Journal of Systematic and Evolutionary Microbiology*, 54(3), pp.961–968.

Zchori-Fein, E., 2011. *Manipulative Tenants*, CRC Press.

Zeldenryk, L. M. *et al.*, 2011. The emerging story of disability associated with lymphatic filariasis: A critical review. *PLoS Neglected Tropical Diseases*, 5(12): e1366.

Zhang, X.-F., Zhao, D.-X. and Hong, X.-Y., 2012. *Cardinium* —the leading factor of cytoplasmic incompatibility in the planthopper *Sogatella furcifera* doubly infected with *Wolbachia* and *Cardinium*. *Environmental Entomology*, 41(4), pp.833–840.

Zhang, Y.-K. *et al*., 2013. *Wolbachia* play an important role in affecting mtDNA variation of *Tetranychus truncatus* (Trombidiformes: Tetranychidae). *Environmental Entomology*, 42(6), pp.1240–1245.

Zheng, X. *et al.,* 2019. Incompatible and sterile insect techniques combined eliminate mosquitoes. *Nature,* 572, pp.56–61.

Zhong, J., Jasinskas, A. and Barbour, A.G., 2007. Antibiotic treatment of the tick vector

Amblyomma americanum reduced reproductive fitness. *PLOS ONE*, 2(5): e405.

Zhou, W., Rousset, F. and O'Neil, S., 1998. Phylogeny and PCR-based classification of *Wolbachia* strains using *wsp* gene sequences. *Proceedings. Biological sciences / The Royal Society*, 265(1395), pp.509–515.

Zientara, S. and Sánchez-Vizcaíno, J.M., 2013. Control of bluetongue in Europe. *Veterinary Microbiology*, 165(1–2), pp.33–37.

Zimmerman, R.H., Barker, S.J. and Turner, E.C., 1982. Swarming and mating behavior of a natural population of *Culicoides variipennis* (Diptera: Ceratopogonidae). *Journal of Medical Entomology*, 19(2), pp.151–156.

Zouache, K. *et al*., 2009. Persistent *Wolbachia* and cultivable bacteria infection in the reproductive and somatic tissues of the mosquito vector *Aedes albopictus*. *PLoS ONE*, 4(7): e6388.

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

Zulu, G.B. and Venter, E.H., 2014. Evaluation of cross-protection of bluetongue virus serotype 4 with other serotypes in sheep. *Journal of the South African Veterinary Association*, 85(1): 1041.

Appendix 1. Accession numbers of loci used in the concatenated gene phylogeny in Figure 3.5. Bold taxa are hosts from which novel sequences were generated in chapter 3. Taxa containing positives without accession numbers were provided by Panupong Thongprem (University of Liverpool).

Host species	***16S***	***gltA***	***atpA***	***coxA***	***17KDa***	**Group**
Acyrthosiphon pisum	FJ609391	FJ666756	FJ666799	FJ666777	-	Belli
Aedes esoensis	KY799072	MF590070	-	-	KY799073	Aedes
Araneus diadematus	-	DQ231490	-	-	-	Limoniae
Argyra vestita	-	JQ925587	-	-	-	Leech
Asobara tabida	FJ603467	-	-	-	-	Limoniae
Bombyliidae sp.	FJ609390	FJ666755	FJ666798	FJ666776	-	Belli
Brachys tessellatus	FJ609393	FJ666758	-	-	-	Belli
Bryopsis	HE648945	-	-	-	-	Hydra
Campsicnemus picticornis	-	JQ925555	-	-	-	Leech
Carteria cerasiformis	AB688628	-	-	-	-	Hydra
Cerobasis guestfalica	DQ652596	-	-	-	-	Limoniae
***Chironomidae* sp.**	-	+	+	-	+	Limoniae
Chrysopa formosa	MF156612	MF156674	MF156637	MF156654	-	Belli
Coegnarion mercuriale	+	+	+	+	+	Limoniae
Coegnarion puella	+	+	+	+	+	Limoniae
Culex pipiens	KY799068	KY799071	-	-	KY799069	Belli
Culicoides bysta	KY777727	KY765382	KY765370	KY765405	KY765391	Limoniae
Culicoides duddinsgtoni (Bara, Sweden)	KY777731	KY765377	-	KY765401	KY765389	Limoniae
Culicoides impunctatus	KY777729	KY765379	-	KY765403	KY765391	Limoniae
***Culicoides newsteadi* N1**	KY777732	KY765374	KY765362	KY765398	KY765386	Limoniae
***Culicoides newsteadi* N2**	KY777733	KY765375	KY765363	KY765399	KY765387	Limoniae
***Culicoides newsteadi* N3**	KY777722	KY765376	KY765364	KY765400	KY765388	Limoniae
***Culicoides newsteadi* N5**	PRJNA376033	PRJNA376033	PRJNA376033	PRJNA376033	PRJNA376033	Limoniae
***Culicoides pulicaris* (Sweden)**	KY777730	KY765380	KY765368	KY765404	KY765392	Limoniae

***Culicoides pulicaris* (UK)**	KY777728	KY765381	KY765369	-	KY765393	Limoniae
Culicoides riethi	KY777726	KY765383	KY765371	KY765406	KY765395	Limoniae
Culicoides stigma	KY777724	KY765385	KY765373	KY765408	KY765397	Limoniae
Culicoides sonorensis	+	-	-	-	-	Aedes
Deronectes aubei	FM955310	FM955315	-	-	-	Limoniae
Deronectes delarouzei	FM955312	FM955313	-	-	-	Limoniae
Deronectes platynotus	FM177877	FM177878	-	-	-	Limoniae
Deronectes semirufus	FM955311	FM955314	-	-	-	Leech
Diophrys	AJ630204	-	-	-	-	Hydra
Empis bicuspidata	-	JQ925616	-	-	-	Leech
***Enallagma cyathigerum* Enc1**	+	+	+	+	+	Limoniae
***Enallagma cyathigerum* Ud6**	+	+	+	+	+	Limoniae
Erigone dentipalpis	-	DQ231492	-	-	-	Limoniae
***Glyptotendipes* sp.**	+	+	+	+	+	Limoniae
Gnathonarium dentatum	-	DQ231484	-	-	-	Leech
Gymnopternus brevicornis	-	JQ925600	-	-	-	Leech
Haplosporidium sp.	AJ319724	-	-	-	-	Hydra
Hemiclepsis marginata	AB113215	-	-	-	-	N/A
Hilara interstincta 2	-	JQ925614	-	-	-	Leech
***Hilara interstincta* Hi3**	+	+	-	+	+	Limoniae
Hydra oligactis	EF667896	-	-	-	-	Hydra
Hydrophorus borealis	-	JQ925566	-	-	-	Leech
Hylaphantes graminicola	-	DQ231487	-	-	-	Limoniae
Ichthyophthirius multifiliis	GQ870455	-	-	-	-	Hydra
Lepthyphantes zimmermani	-	DQ231488	-	-	-	Limoniae
Limonia chorea	AF322443	-	-	-	-	Limoniae
Lutzomyia apache	EU223247	EU368001	-	-	-	Limoniae
Macrolophus pygmaeus	+	+	+	+	+	Limoniae
Mansonia uniformis	KY799063	KY799066	-	-	KY799064	N/A
Mesostigma viride	KJ808701	-	-	-	-	Hydra
Meta mengei 1	-	DQ231482	-	-	-	Leech
Meta mengei 2	-	DQ231483	-	-	-	Leech
Microneta viaria	-	DQ231493	-	-	-	Limoniae

Microphor holosericeus	-	JQ925617	-	-	-	Leech
Neochrysocharis formosa	AB231472	-	-	-	-	Transitional
Nephotettix cincticeps	AB702995	-	-	-	-	Leech
Nesidiocoris tenuis	KF646705	KF646706	-	-	-	Belli
Neurigona lineata	-	JQ925596	-	-	-	Leech
Nuclearia pattersoni	AY364636	-	-	-	-	Leech
Orientia tsutsugamushi	AM494475	AM494475	AM494477	AM494478	-	Outgroup
Paramecium caudatum	FR822997	-	-	-	-	Hydra
Peyerimhoffina gracilis	MF156627	MF156685	MF156648	MF156666	-	Belli
Pityohyphantes phrygianus	-	DQ231491	-	-	-	Limoniae
Pleodorina japonica	AB688629	-	-	-	-	Hydra
Pseudomallada ventralis	MF156633	MF156688	MF156652	MF156669	-	Rhizobius
Pyrrhosoma nymphula	+	+	+	+	+	Limoniae
Quadrastichus mendeli	KX592502	KX673390	-	-	-	N/A
Raphium micans	-	JQ925570	-	-	-	N/A
Rhizobius litura	FJ609388	FJ666753	FJ666796	FJ666774	-	Rhizobius
Rickettsia akari	CP000847	CP000847	CP000847	CP000847	CP000847	Transitional
Rickettsia australis	CP003338	CP003338	CP003338	CP003338	CP003338	Transitional
Rickettsia bellii	CP000087	AY375161	CP000087	CP000087	CP000087	Belli
Rickettsia canadensis	CP000409	AB297809	CP000409	CP000409	CP000409	Canadensis
Rickettsia conorii	AE006914	AE006914	AE006914	AE006914	-	Spotted Fever
Rickettsia felis	CP000053	CP000054	CP000056	CP000055	CP000053	Transitional
Rickettsia heilongjiangensis	CP002912	NC015866	CP002912	CP002912	CP002912	Spotted Fever
Rickettsia japonica	AP017601	AP017601	AP017601	AP017601	AP017601	Spotted Fever
Rickettsia massiliae	CP000683	CP000683	CP000683	CP000683	CP000683	Spotted Fever
Rickettsia peacockii	CP001227	DQ100162	CP001227	CP001227	CP001227	Spotted Fever
Rickettsia prowazekii	AJ235272	AJ235272	AJ235272	AJ235272	AJ235272	Typhus
Rickettsia rhipicephali	CP013133	CP013133	CP013133	CP013133	CP013133	Spotted Fever

Rickettsia rickettsii	CP000848	CP000848	CP000848	CP000848	CP000848	Spotted Fever
Rickettsia slovaca	CP002428	CP002428	CP002428	CP002428	CP002428	Spotted Fever
Rickettsia typhi	CP003398	CP003398	CP003398	CP003398	CP003398	Typhus
Sialis lutaria	MF156635	-	-	MF156671	-	Limoniae
Sialis lutaria	MF156636	-	-	MF156672	-	Limoniae
Simulium aureum	+	+	-	+	-	Leech
Sphyrotarsus argyrostomus	-	JQ925623	-	-	-	Leech
Spirostomum	FR822999	-	-	-	-	Hydra
Theridiidae sp.	-	DQ231486	-	-	-	Limoniae
Torix tagoi	AB066351	-	-	-	-	Leech
Torix tsukubana	AB113214	-	-	-	-	Leech
Trichopeza longicornis	-	JQ925611	-	-	-	Leech
Troxochrus scabriculus	-	DQ231485	-	-	-	Leech
Volvox carteri	AB861537	-	-	-	-	Hydra
Walckenaeria cuspidata	-	DQ231489	-	-	-	Limoniae
Xenopsylla minax	KX254161	KX254162	-	-	KX254163	Belli
***Zavrelimyia* sp.**	+	+	+	+	+	Limoniae

Appendix 2. Non-BOLD derived housekeeping gene sequences used for phylogenetic analysis in chapter 5 (Figures 5.2, 5.3, 5.5 and 5.6). "+" indicates sequences generated in this thesis (except for *Culicoides* accession numbers).

Host species	***16S***	***gltA***	***coxA***	***17KDa***	**Group**
Acyrthosiphon pisum	FJ609391	FJ666756	FJ666777	-	Belli
Amaurobioides africana	-	-	KU600824	-	Leech
Aedes esoensis	KY799072	MF590070	-	KY799073	Aedes
Anopheles Plumbeus	+	+	+	+	Limoniae
Asobara tabida	FJ603467	-	-	-	Limoniae
Bombyliidae sp.	FJ609390	FJ666755	FJ666776	-	Belli
Cerobasis guestfalica	DQ652596	-	-	-	Limoniae
Chrysopa formosa	MF156612	MF156674	MF156654	-	Belli
Coenagrion mercuriale	+	+	+	+	Limoniae
Coenagrion puella	+	+	+	+	Limoniae
Culex pipiens	KY799068	KY799071	-	KY799069	Belli
Culicoides bysta	KY777727	KY765382	KY765405	KY765391	Limoniae
Culicoides duddinsgtoni (Bara, Sweden)	KY777731	KY765377	KY765401	KY765389	Limoniae
Culicoides duddinsgtoni (Unknown site, Sweden)	KY777723	KY765378	KY765402	KY765390	Limoniae
Culicoides newsteadi N1	KY777732	KY765374	KY765398	KY765386	Limoniae
Culicoides newsteadi N2	KY777733	KY765375	KY765399	KY765387	Limoniae
Culicoides newsteadi N3	KY777722	KY765376	KY765400	KY765388	Limoniae
Culicoides newsteadi N5	PRJNA376033	PRJNA376033	PRJNA376033	PRJNA376033	Limoniae
Culicoides pulicaris (UK)	KY777728	KY765381	-	KY765393	Limoniae
Culicoides riethi	KY777726	KY765383	KY765406	KY765395	Limoniae
Culicoides stigma	KY777724	KY765385	KY765408	KY765397	Limoniae
Deronectes platynotus	FM177877	FM177878	-	-	Limoniae
Deronectes semirufus	FM955311	FM955314	-	-	Leech

Enallagma cyathigerum Enc1	+	+	+	+	Limoniae
Enallagma cyathigerum Ud6	+	+	+	+	Limoniae
Glyptotendipes sp.	+	+	+	+	Limoniae
Hemiclepsis marginata	AB113215	-	-	-	Limoniae
Hilara interstincta Hi3	+	+	+	+	Limoniae
Ichthyophthirius multifiliis (Outgroup Figures 5.5. and 5.6.)	+	+	+	+	Megaira
Lutzomyia apache	EU223247	EU368001	-	-	Limoniae
Macrolophus pygmaeus	+	+	+	+	Limoniae
Mansonia uniformis	KY799063	KY799066	-	KY799064	N/A
Nephotettix cincticeps	AB702995	-	-	-	Leech
Nosopsyllus laeviceps	KX457949	KX457954	-	-	Leech
Nuclearia pattersoni	AY364636	-	-	-	Leech
Orientia tsutsugamushi (Outgroup Figures 5.3.1. and 5.3.2.)	-	-	AM494475	-	Orientia
Peyerimhoffina gracilis	MF156627	MF156685	MF156666	-	Belli
Pelagibacter ubique (Outgroup Figure 5.2.)	JNIU01000001	JNIU01000001	JNIU01000001	JNIU01000001	N/A
Phlebotomus chinensis	KX363668	-	-	-	Leech
Platypleura kaempferi	KR911839	-	-	-	Leech
Polythore lamerceda	+	+	+	+	Limoniae
Polythore picta	+	+	+	+	Limoniae
Pseudomallada ventralis	MF156633	MF156688	MF156669	-	Rhizobius
Pyrrhosoma nymphula	+	+	+	+	Limoniae
Quadrastichus mendeli	KX592502	KX673390	-	-	Leech
Rhizobius litura	FJ609388	FJ666753	FJ666774	-	Rhizobius
Rickettsia akari	CP000847	CP000847	CP000847	CP000847	Transitional

Rickettsia australis	CP003338	CP003338	CP003338	CP003338	Transitional
Rickettsia bellii	CP000087	AY375161	CP000087	CP000087	Belli
Rickettsia canadensis	CP000409	AB297809	CP000409	CP000409	Canadensis
Rickettsia conorii	AE006914	AE006914	AE006914	-	Spotted Fever
Rickettsia felis	CP000053	CP000054	CP000055	CP000053	Transitional
Rickettsia heilongjiangensis	CP002912	NC015866	CP002912	CP002912	Spotted Fever
Rickettsia japonica	AP017601	AP017601	AP017601	AP017601	Spotted Fever
Rickettsia massiliae	CP000683	CP000683	CP000683	CP000683	Spotted Fever
Rickettsia peacockii	CP001227	DQ100162	CP001227	CP001227	Spotted Fever
Rickettsia prowazekii	AJ235272	AJ235272	AJ235272	AJ235272	Typhus
Rickettsia rhipicephali	CP013133	CP013133	CP013133	CP013133	Spotted Fever
Rickettsia rickettsii	CP000848	CP000848	CP000848	CP000848	Spotted Fever
Rickettsia slovaca	CP002428	CP002428	CP002428	CP002428	Spotted Fever
Rickettsia typhi	CP003398	CP003398	CP003398	CP003398	Typhus
Simulium aureum	+	+	+	-	Leech
Sympetrum fonscolombii	+	+	+	-	Limoniae
Stegobium paniceum	JQ805029	-	-	-	Leech
Torix tagoi	AB066351	-	-	-	Leech
Xenopsylla minax	KX254161	KX254162	-	KX254163	Belli
Zavrelimyia sp.	+	+	+	+	Limoniae

Appendix 3. Re-barcoding status and nearest BLAST hit (NCBI) of mtDNA *COI* arthropod DNA extracts accessed for further analysis in chapter 5, along with the success of multilocus *Rickettsia* profiles with allocated *Rickettsia* group (based on phylogenetic analysis) and co-infection status.

BOLD ID	Host taxonomy	Nearest host barcoding BLAST hit	*gltA*	*16S*	*17k Da*	Co-infection	*Rickettsia* group
BIOUG03090-E08	Trichoptera	100% identity to *Lepidostoma hoodi* KM533759 Caddis fly	+	-	+	-	Limoniae
BIOUG03644-G07	Diptera	-	+	+	-	-	Limoniae
BIOUG03946-H07	Hymenoptera	86% identity to Megaspilidae sp. KP693295 Parasitoid wasp	+	+	+	-	Limoniae
BIOUG03995-E06	Arachnida	99% identity to Calyptostomatidae sp. MG317701 Water mite	+	-	-	-	Torix
BIOUG05036-F05	Hymenoptera	-	+	+	-	-	Limoniae
BIOUG05952-E04	Diptera	99% identity to Sciaridae sp. KR435639 Dark winged fungus gnat	+	-	-	-	Torix
BIOUG05952-F01	Diptera	99% identity to Sciaridae sp. KR435639 Dark winged fungus gnat	-	-	-	-	Torix
BIOUG05952-F09	Diptera	99% identity to Sciaridae sp. KR435639 Dark winged fungus gnat	-	-	-	-	Torix

BIOUG06945-H09	Psocodea	99% identity to *Philotarsus californicus* MF750052 Loving barklouse	-	-	-	-	Limoniae
BIOUG08169-H08	Diptera	-	+	-	+	-	Limoniae
BIOUG08331-C03	Hemiptera	81% identity to Hemisphaerius palaemon (Issidae) KX761497 Planthopper	+	+	+	-	Leech
BIOUG09046-A04	Hemiptera	86% identity to Ricaniidae sp. KY841343 Planthopper	+	+	+	-	Leech
BIOUG09067-A04	Hemiptera	83% identity to *Synecdoche impunctata* KR042775 Planthopper	-	+	-	-	N/A
BIOUG09317-G05	Hemiptera	-	+	+	-	-	Leech
BIOUG09528-H02	Diptera	99% identity to Cecidomyiidae sp. KR225720 Gall midge	-	+	-	-	Torix
BIOUG09607-F08	Hymenoptera	-	+	+	+	-	Limoniae
BIOUG09729-H07	Hemiptera	-	+	+	-	-	Leech
BIOUG10000-C10	Hymenoptera	89% identity to *Ceraphron* sp. KJ570858 Parasitoid wasp	+	+	+	-	Limoniae
BIOUG10000-F03	Hymenoptera	89% identity to *Ceraphron* sp. KJ570858 Parasitoid wasp	+	+	+	-	Limoniae

BIOUG10096-E08	Diptera	88% identity to Mycetophilidae sp. KR433437 Fungus gnat	+	+	+	-	Limoniae
BIOUG10258-G06	Diptera	88% identity to Cecidomyiidae sp. KM958360 Gall midge	+	+	+	-	Limoniae
BIOUG10275-H02	Hemiptera	-	-	+	+	-	Belli
BIOUG10403-F02	Hymenoptera	98% identity to Diapriidae sp. MG445632 Parasitoid wasp	+	+	+	-	Limoniae
BIOUG10419-G02	Diptera	-	+	+	+	-	Limoniae
BIOUG10490-E11	Hymenoptera	-	+	+	+	-	Limoniae
BIOUG10491-A07	Hymenoptera	-	+	+	+	-	Leech
BIOUG10494-C02	Hymenoptera	99% identity to Diapriidae sp. KR792549 Parasitoid wasp	+	+	+	-	Limoniae
BIOUG10585-E02	Diptera	100% identity to *Boletina villosa* HQ230432 Fungus gnat	+	-	+	-	Limoniae
BIOUG10637-E07	Hymenoptera	99% identity to Diapriidae sp. MG445632 Parasitoid wasp	+	+	+	-	Limoniae
BIOUG10694-H11	Diptera	99% identity to Cecidomyiidae sp. KR953731 Gall midge	-	-	-	-	Torix
BIOUG10810-G11	Hemiptera	99% identity to Psyllidae sp. MG398967 Psyllid	+	+	+	-	Limoniae
BIOUG10835-A09	Hymenoptera	-	+	+	+	-	Leech

BIOUG11020-A09	Hymenoptera	-	+	+	+	-	Limoniae
BIOUG11077-G04	Hymenoptera	94% identity to Diapriidae sp. KR889291 Parasitoid wasp	-	+	-	-	Transitional
BIOUG11109-D07	Hymenoptera	97% identity to Diapriidae sp. MG481766 Parasitoid wasp	+	+	+	-	Limoniae
BIOUG11160-G09	Hemiptera	99% identity to Psyllidae sp. MG398967 Psyllid	+	+	+	-	Limoniae
BIOUG11193-E04	Hemiptera	97% identity to *Trioza urticae* KY011119 Psyllid	-	-	-	-	Belli
BIOUG11287-D01	Hymenoptera	96% identity to Diapriidae sp. MG483535 Parasitoid wasp	+	+	+	-	Leech
BIOUG11445-G12	Hymenoptera	97% identity to Ceraphronidae sp. MG483592 Parasitoid wasp	+	+	+	-	Limoniae
BIOUG11473-H07	Hemiptera	99% identity to *Chamaepsylla hartigii* MG988705 Psyllid	+	+	-	-	Leech
BIOUG12130-B03	Diptera	-	+	+	+	-	Limoniae
BIOUG12247-G09	Hymenoptera	98% identity to Belytinae sp. FJ413688 Parasitoid wasp	+	+	+	-	Limoniae
BIOUG12267-A05	Thysanoptera	92% identity to Panchaetothripinae MG334457 Thrip	-	-	-	-	N/A

BIOUG12276-C04	Hymenoptera	96% identity to Platygastridae sp. NB exclusive parasitoid wasp of gall midges	+	+	-	-	Leech
BIOUG12267-H01	Arachnida	-	+	-	+	-	N/A
BIOUG12276-C04	Hymenoptera	-	+	+	+	-	Leech
BIOUG12296-F05	Hymenoptera	91% identity to Belytinae sp. MG442688 Parasitoid wasp	+	+	+	-	Limoniae
BIOUG12300-D10	Hemiptera	99% identity to Psyllidae sp. KR575644 Psyllid	+	+	-	-	Leech
BIOUG12559-D05	Diptera	88% identity to Cecidomyiidae sp. KM958360 Gall midge	-	-	-	-	Torix
BIOUG12651-B03	Hymenoptera	96% identity to Diapriidae sp. MF900305 Parasitoid wasp	-	-	-	-	Leech
BIOUG12657-H04	Diptera	98% identity to *Boletina villosa* HQ230432 Fungus gnat	-	-	+	-	Limoniae
BIOUG12688-E05	Diptera	98% identity to *Boletina villosa* HQ230432 Fungus gnat	+	+	+	-	Limoniae
BIOUG12692-H09	Diptera	98% identity to Chironomidae sp. KR596545 Chironomid	+	+	+	+	Limoniae
BIOUG12754-B10	Hymenoptera	92% identity to Diapriidae sp. parasitoid wasp	+	-	-	-	N/A
BIOUG12754-F03	Hymenoptera	-	-	+	+	-	N/A

BIOUG12890-D06	Diptera	88% identity to Mycetophilidae sp. KR433437 Fungus gnat	+	+	+	-	Limoniae
BIOUG13171-F03	Hymenoptera	96% identity to Diapriidae sp. MG481766 Parasitoid wasp	+	+	+	+	Limoniae
BIOUG13204-H01	Coleoptera	99% identity to Staphylinidae sp. MF636000 Rove beetle	+	+	+	-	Leech
BIOUG13697-B07	Hemiptera	-	+	-	+	-	Leech
BIOUG14264-B08	Hemiptera	-	-	+	+	-	Belli
BIOUG14285-F06	Hymenoptera	-	+	+	+	-	Leech
BIOUG14438-E11	Hymenoptera	98% identity to Ceraphronidae sp. MG483592 Parasitoid wasp	+	+	+	-	Limoniae
BIOUG14788-F12	Diptera	86% identity to Dolichopodidae KY836029 Long-legged flies	+	+	-	-	Leech
BIOUG14827-A07	Diptera	-	+	+	-	-	Leech
BIOUG14931-F11	Coleoptera	85% identity to Staphylinidae sp. MK082532 Rove beetle	+	+	+	-	Leech
BIOUG15148-E11	Dermaptera	-	+	+	-	-	Leech
BIOUG15204-F07	Hemiptera	-	+	+	-	-	Leech
BIOUG15399-C10	Hemiptera	100% identity to Ricaniidae sp. KY839843 Planthopper	+	+	-	-	Leech
BIOUG16256-A02	Diptera	88% identity to *Ormosia* sp.	+	-	+	-	Limoniae

		KR969558.1 Crane fly					
BIOUG16326-D12	Hymenoptera	-	+	+	+	-	Leech
BIOUG16611-A07	Coleoptera	-	+	+	+	-	Leech
BIOUG16718-E05	Arachnida	-	+	+	+	-	Leech
BIOUG16781-G01	Hemiptera	100% identity *Cacopsylla melanoneura* FJ648816.1 Psyllid vector of phytoplasma	+	+	+	-	Leech
BIOUG16789-H07	Hymenoptera	-	+	-	+	-	Limoniae
BIOUG16805-B06	Hymenoptera	-	+	+	-	-	Leech
BIOUG17003-E01	Hymenoptera	-	+	+	-	-	Leech
BIOUG17003-F11	Hymenoptera	80% identity to Eulophidae sp. KY832534.1 Parasitoid wasp	+	+	-	-	Leech
BIOUG17121-C06	Hemiptera	85% identity to *Metcalfa pruinosa* KX761467 Planthopper	-	+	-	-	Belli
BIOUG17236-F10	Diptera	99% identity to Orthocladiinae sp. KT700901 Chironomid	-	+	-	-	Torix
BIOUG17263-B09	Hymenoptera	85% identity to Diapriidae sp. MG488103 Parasitoid wasp	-	-	-	-	Torix
BIOUG17272-A01	Hymenoptera	93% identity to Eurytomidae sp. MG485494 Parasitoid wasp	-	+	-	-	Transitional

BIOUG17366-H07	Hymenoptera	99% identity to Diapriidae sp. JN300282 Parasitoid wasp	+	+	+	-	Limoniae
BIOUG17403-F09	Diptera	100% identity to Orthocladiinae sp. KT700901 Chironomid	+	+	-	-	Limoniae
BIOUG17681-H05	Psocodea	99% identity to *Philotarsus californicus* MF750052 Loving barklouse	+	-	+	+	Limoniae
BIOUG17697-E09	Hemiptera	-	+	+	+	-	Leech
BIOUG17786-A03	Hymenoptera	99% identity to Diapriidae sp. MG479783 Parasitoid wasp	+	+	+	-	Leech
BIOUG17814-D02	Hemiptera	100% identity to Ricaniidae sp. KY839843 Planthopper	+	+	-	-	Leech
BIOUG17814-D08	Hemiptera	100% identity to Ricaniidae sp. KY839843 Planthopper	+	+	-	-	Leech
BIOUG17921-G11	Diptera	99% identity to Sciaridae sp. KR435639 Dark winged fungus gnat	+	+	+	+	Limoniae
BIOUG17995-E07	Hymenoptera	-	+	+	+	-	Limoniae
BIOUG18138-A03	Diptera	98% identity to *Pherbellia tenuipes* KM941178.1 Marsh fly	+	+	+	-	Leech
BIOUG18141-E05	Hymenoptera	99% identity to Diapriidae sp.	+	+	-	-	Leech

		MG479783 Parasitoid wasp					
BIOUG18188-H03	Psocodea	85% identity to *Myopsocus* sp. KX054004 Barklouse	-	-	+	-	Torix
BIOUG18338-C05	Coleoptera	83% to *Homotechnes motschulskyi* KM612696 Click beetle	-	-	-	-	Transitional
BIOUG18523-H11	Hymenoptera	-	+	-	+	-	Leech
BIOUG18544-F02	Hymenoptera	90% identity to Belytinae sp. MG442688 Parasitoid wasp	+	+	+	-	Leech
BIOUG18751-H08	Psocodea	99% identity to *Philotarsus californicus* MF750052 Loving barklouse	+	+	+	+	Limoniae
BIOUG18783-G05	Hymenoptera	100% identity to Diapriidae sp. KR796688 Parasitoid wasp	+	+	+	-	Limoniae
BIOUG18786-B03	Hymenoptera	93% identity to Mymaridae sp. KR421238 Fairy wasp	-	-	+	-	Limoniae
BIOUG18831-A05	Coleoptera	-	+	+	+	-	Leech
BIOUG18843-E05	Diptera	-	+	+	+	-	N/A
BIOUG18843-E06	Diptera	88% identity to *Phaeotabanus cajennensis* KY777207 Tabanid fly	+	+	+	-	Limoniae
BIOUG18855-B03	Arachnida	-	+	+	+	-	Limoniae

BIOUG18682-A09	Diptera	88% identity to Chironomidae sp. KM634101 Chironomid	-	-	-	-	Torix
BIOUG18950-H10	Diptera	100% identity to Psilidae sp. KP040464 Rust flies	+	+	+	+	Leech
BIOUG18953-G05	Diptera	99%% identity to Psilidae sp. KP040464 Rust flies	+	+	+	-	Limoniae
BIOUG18975-F01	Psocodea	99% identity to *Philotarsus californicus* MF750052 Loving barklouse	+	+	+	+	Limoniae
BIOUG18978-A11	Psocodea	100% identity to *Philotarsus californicus* MF750052 Loving barklouse	+	+	+	+	Limoniae
BIOUG18978-C01	Psocodea	100% identity to *Philotarsus californicus* MF750052 Loving barklouse	+	+	+	+	Limoniae
BIOUG18996-B06	Hymenoptera	-	+	-	+	-	Limoniae
BIOUG19150-H07	Psocodea	99% identity to *Philotarsus californicus* MF750052 Loving barklouse	+	-	+	-	Limoniae
BIOUG19173-E08	Psocodea	-	+	+	+	-	Leech
BIOUG19198-A07	Diptera	100% identity to Psilidae sp.	+	+	+	-	Limoniae

		KP040464 Rust fly					
BIOUG19280-B09	Diplopoda	-	+	+	+	-	Limoniae
BIOUG19345-B04	Diptera	99% identity Sciaridae sp. KR435639 Dark winged fungus gnats	+	-	+	-	Leech
BIOUG19553-C10	Diptera	95% identity to *Cerodontha luctuosa* JF875118 Leaf miner fly	-	+	-	-	Transitional
BIOUG19614-G04	Hymenoptera	-	+	+	+	-	Limoniae
BIOUG19659-D04	Hemiptera	-	-	+	+	-	Leech
BIOUG19733-H02	Hymenoptera	-	+	-	+	-	Limoniae
BIOUG19931-G07	Diptera	85% identity to Empididae sp. KP697561 Dance fly	+	+	+	-	Limoniae
BIOUG20015-F06	Psocodea	99% identity to *Philotarsus californicus* MF750052 Loving barklouse	+	-	+	+	Limoniae
BIOUG20052-H08	Hemiptera	85% identity to *Nasatus davidouvrardi* KX702952 Planthopper	+	+	-	-	Leech
BIOUG20053-D06	Hemiptera	86% identity to *Nasatus davidouvrardi* KX702952 Planthopper	+	+	-	-	Leech
BIOUG20103-E03	Hymenoptera	88% identity to Ceraphronidae sp.	+	+	+	-	Leech

		MG443347 Parasitoid wasp					
BIOUG20209-A04	Hymenoptera	97% identity to Diapriidae sp. MG488194 Parasitoid wasp	+	-	+	-	Leech
BIOUG20920-G10	Psocodea	99% identity to *Philotarsus californicus* MF750052 Loving barklouse	+	+	+	-	Limoniae
BIOUG20920-H03	Psocodea	100% identity to *Philotarsus californicus* MF750052 Loving barklouse	+	+	+	-	Limoniae
BIOUG21218-H05	Coleoptera	100% identity to *Prionocyphon limbatus* KR485242 Marsh beetle	+	-	+	-	Limoniae
BIOUG21219-B07	Coleoptera	86% identity to *Laccophilus biguttatus* KJ203140 Water beetle	-	-	+	-	Limoniae
BIOUG21386-B09	Hymenoptera	-	-	+	+	-	Limoniae
BIOUG21469-E07	Diptera	100% identity to Hybotidae sp MG083504 Dance fly	+	+	-	-	N/A