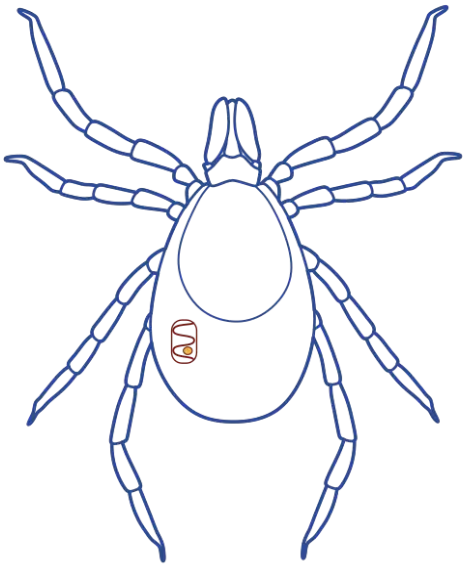
UNIVERSITÀ
DI PAVIA

Dipartimento di Biologia e Biotecnologie "Lazzaro Spallanzani"

Genomic and evolutionary characterisation of symbiotic bacteria in ticks



Tiago Nardi

Dottorato di Ricerca in
Genetica, Biologia Molecolare e Cellulare
Ciclo XXXV – A.A. 2019-2022

Abstract

Obligate symbionts allow many arthropod lineages to thrive on a restricted diet. An interesting model for the study of these symbiosis are ticks, as they are dependent on - mostly - maternally inherited bacteria to complement their blood diet.
For my PhD I studied the bacteria association in three different ticks: *Coxiella* endosymbiont with *Amblyomma nuttalli, Midichloria* and *Francisella* endosymbiont with *Hyalomma marginatum*, *Spiroplasma* and *Midichloria* with *Ixodes frontalis*.
In the first project, I characterised the bacterial symbiont present in the tick, and then did a comparative genomic and phylogenomic analysis of the *Coxiella* endosymbionts, comparing them to the pathogen *Coxiella burnetii* to infer an evolutionary scenario, including the *Coxiella* common ancestor traits, to clarify if the symbionts are deriving from a pathogen, or pathogenicity is a specific novel trait of the *C. burnetii* lineage.
For the second project, I studied the presence of a multi-partite symbiosis that involves the tick *H. marginatum* with two different bacteria for a complete provision of nutrients. Our analysis shows that the *Francisella* endosymbiont in *H. marginatum* has lost part of the pathway for biotin, and the co-presence with *Midichloria* restores the genomic capability for biotin synthesis. At same time the biosynthethic capabilities of *Midichloria* are not sufficient to replace *Francisella*, making the presence of both bacteria necessary for the host's survival.
For the third, I characterise the symbiont population in *I. frontalis*, considering also if there is relation of the two different sublinages of the tick with a specific symbiont. We observed the presence of *Midichloria* and *Spiroplasma* in both sublineages, but with a limited presence for *Midichloria* in one lineage.
Then, for the last project I studied the deep phylogeny of the order *Rickettsiales* which includes some of the tick pathogens and mutualists, but also many other bacteria with a wide range of hosts. According to the newly inferred scenario, *Rickettsiales* evolved intracellularity parallelly in different lineages, developing a dependance on their host through horizontal transfers of transporter genes.

The results presented here show how similar selective pressures, due the symbiotic relationship with ticks, in phylogenetically close systems can results in different outcomes: *Coxiella* is present as a single symbiont derived after convergent reduction from a pathogen, in *H. marginatum* we have a dual bacterial symbiosis with interlocking genomic capabilities, and in *I. frontalis* we have two very different symbionts alternating their presence to provide the same metabolic capabilities. This diversity of outcomes becomes even starker when we enlarge our analysis to the whole *Rickettsiales* order, where we have different cellular localisations, metabolic capabilities, host species preference and relation with the host.

Abbreviations

A. phagocytophilum: *Anaplasma phagocytophilum*
B. aphidicola: Buchnera aphidicola
B. burgdorferi: *Borrelia burgdorferi*
CDS: Coding Sequence
CE: *Coxiella* endosymbiont
CoA: Coenzyme A
COG: Clusters of Orthologous Genes
COI: Cytochrome c oxidase I
DDG-clade*:* clade including the three families *Diomedesiaceae*, *Jistubacteraceae*, *Arkhamiaceae*
ECDC: European Centre for Disease Prevention and Control
FE: *Francisella* endosymbiont
FECH: ferrochelatase
F. tularensis: *Francisella tularensis*
HGT: horizontal gene transfer
H. marginatum: Hyalomma marginatum
I. frontalis: *Ixodes frontalis*
I. ricinus: *Ixodes ricinus*
Ma: million years (ago)
ML: Maximum Likelihood
M. mitochondrii: *Midichloria mitochondrii*
PAMP: pathogen associated molecular patterns
PG: peptidoglycan
R. aeschlimannii: Rickettsia aeschlimannii
SCO: Single Copy Orthologs
SOD: superoxide dismutase
S. ixodetis: Spiroplasma ixodetis
TBE: tick-borne encephalitis
TBEV: tick-borne encephalitis virus

I Introduction

1.1 Bacterial symbionts and their evolution

Symbiosis has been originally described by Anton de Bary's as "the living together of two differently named organisms" (Bary, 1879), independently of the effects on the organisms involved, so including in the definition mutualism, parasitism, and commensalism.
Among a plethora of different symbiosis, mutualistic associations between bacteria and eukaryotes are widespread, enabling adaptation to different situations, often even allowing the colonisation of new ecological niches (Bennett & Moran, 2015).

There are two routes of transmission and the maintenance of the symbiotic bacteria in the host species: horizontal and vertical (Bright & Bulgheresi, 2010). Horizontal transfer is the acquisition of a symbiont from an environment, often a free-living symbiont, otherwise one still capable of surviving in the environment. Often the host has a life phase aposymbiotic, without the presence of the symbiont, that is acquired in a specific stage of development. That is the case of Hawaiian bobtail squid *Euprymna scolopes*, that acquires its symbiont *Vibrio fischeri* filtering it from the sea, selecting the correct species between all the possible environmental bacteria, and hosts it in its light organ (Nyholm & McFall-Ngai, 2004). An environment shared by different generations and life stages of the host species can act a s multigenerational bacterial reservoir, such as *Acetobacter thailandicus* that is residing in the gut *Drosophila melanogaster*: *A. thailandicus* is shed in the environment, and from which *Drosophila* can acquire it (Pais *et al*, 2018)
Vertical transfer is the inheritance of the symbiont from a parent, usually the mother (with transovarial route), or rarely from both parents. This is typical of permanent symbiotic relationships, such as "*Candidatus* Riesia pediculicola", the bacterial symbiont of *Pediculus humanus*, the human head/body louse (Allen *et al*, 2007).

These associations can be facultative, offering advantages to their host, but not required for its survival or reproduction. And at the same time, these associations can be facultative for the symbiotic bacteria that remain able to live and reproduce outside their host. However, the optional nature of the relation can be true only for one partner, as the term "facultative symbiont" is often used regarding the host that can survive without the bacteria, whether the bacteria is able to survive and reproduce without a host (Husnik & Keeling, 2021).
One example is *Serratia symbiotica*, that is an *Enterobacterales*, that, depending on the strain, can be a facultative or an obligate symbiont for the host, be able to survive or not to outside its host (Foray *et al*, 2014). In *Acyrthosiphon pisum* (the pea aphid) *S. symbiotica* is a facultative symbiont, conferring resistance to *Aphidius ervi*, a parasitoid wasp, by causing high mortality of the developing parasitoid larvae (Oliver *et al*, 2003).

An alternative situation is that of obligate symbiosis, in which both the host and the symbiont require each other for their survival, to the point that they present a congruent co-phylogeny (Moran *et al*, 2008). These are typically the results of long term associations, in which facultative symbionts lose their capability to survive in the environment: living within an host some functions are no more necessary, such as the complete biosynthesis capabilities, or thicker membranes, so there are no selective pressures on the conservation of these capabilities. Such conditions have been described especially for nutritional symbionts that are particularly common in arthropods. Metazoa have lost the capabilities to synthesise different amino acids early in the evolution (Costa *et al*, 2014),but this is not an issue when the eukaryotes can retrieve them from the environment. However this can be difficult when there are limited choices of sources, like in a restricted diet. A way to overcome this restriction is through intracellular bacterial mutualists that provide essential nutrients that are absent in the host's diet (Douglas, 1998). That is the case of insects that have a diet based on sap, rich on sugars but poor in amino acids, and tackle the need through their bacterial symbionts, usually maternally transmitted (Sandström & Moran, 1999). Blood feeding arthropods are an analogous case, as they are dependent on provision of nutrients absent in blood. For example tsetse flies that are dependent on B-group vitamin provision by their symbiont

(Nogge, 1976), *Wigglesworthia glossinidia* (Aksoy, 1995).

The long lasting symbionts typically are considered to have evolved from formerly free-living bacteria, but they differ from them for peculiar traits. Typically long term symbiont present genome reduction, limited metabolism to supply their host with essential nutrients, change in the cellular shape and compositional bias with - usually - an increase of the AT content.
One example can be the aforementioned case of the tsetse fly: its obligate symbiont, *W. glossinidia*, has had a long association with its host of at least 50 millions years and has a streamlined genome of 700 kb. Instead, *Sodalis glossinidius*, another symbiont in the tse-tse fly, is facultative, has a genome of 4.2 Mb, can live extracellularly, and is capable of growth in culture (Pais *et al*, 2008).

In the nutritional symbionts the genome modifications undergo different phases (McCutcheon, 2012): the early stages have an increase in mobile elements, with a very "plastic" genome, with a high number of rearrangement, formation of pseudogenes and loss of some fragments. These characteristics lead to the loss of the genes necessary to survive freely, and the symbiont can thus become host-dependent. With time there is a reduction of mobile elements, with stabler and smaller genomes after having lost some of the pseudogenes, till arriving to extremely limited "tiny" genomes.
This genome evolution pattern is the consequence of the vertical transfer of the symbiont between generations that can lead to what has been defined as a potential "Evolutionary Rabbit-Hole" (Bennett & Moran, 2015). Each host generation transfers a limited population of bacterial symbionts to the offspring, leading to a series of population bottlenecks. Also often the transmission is matrilinear, resulting in a clonal population of symbionts. So the symbionts, being shielded by their host, are no longer under strict purifying selection, and through time, slightly deleterious mutations can be accumulated in the bacteria with limited costs, without the possibility to recombine, leading to high rate of fixation of deleterious mutations (Moran 1996). Some losses have limited effect, such as genes involved in survival outside the host, like osmotic resistance. However, even without deletions, there is accumulation of point mutations, with a bias to AT bases, in the whole genome, reducing the codon and folding efficiency in all genes (RCHJ van Ham, et al 2003. The folding efficiency loss is partially rescued by conservation and high expression of chaperones proteins, even if the process is metabolically costly (Kupper *et al*, 2014).
However, a long-time genome erosion can lead to a reduction of the symbionts' beneficial role for their host. For example the symbiont provision capabilities could decrease, narrowing the host ecological range. This symbiont decay pressures the host to compensate (e.g. through host expression of symbiont proteins lost or inefficiently expressed), increasing the permissibility for further symbiont decays, leading the lineage into a symbiosis rabbit hole of genetic drift and host compensation (Bennett & Moran, 2015).
One way to escape this situation is the acquisition of a new symbiotic bacterium, with a more 'apt' genome, which can complement or in even replace the previous endosymbiont.
In the case of the co-presence of symbionts, it is common for the new symbiont to have redundant biosynthetic capabilities, already present in the old symbiont. Through time this can result in the loss of the supernumerary genes, potentially locking both the old and the new bacteria as obligate symbionts, giving rise to a tripartite system. For example, the mealybug *Planococcus citri* is dependent on symbionts for aminoacid supplementation, but instead of interacting with one single species of symbiont, it has *Tremblaya princeps*, a betaproteobacterium, that itself is nested within the cells of *Moranella endobia*, a gammaproteobacterium. The biosynthetic pathways required by the host are composed by a patchwork of genes from both symbionts and even from the host itself (McCutcheon, 2012).
Different examples of these dynamics can be found in aphids: usually one symbiont is the gammaproteobacterium *Buchnera aphidicola,* providing amino acids and vitamins absent in the diet (Manzano-Marín *et al*, 2016). This relation is very ancient but, especially taking into consideration aphids outside the Aphidinae, is less stable than initially thought (Manzano-Marín *et al*, 2022) . In some cases *Buchnera* has been replaced by other symbionts capable of providing the needed nutrients, in one case by a yeast (Vogel & Moran, 2013) and in another by a *Bacteroidetes* representative (Chong & Moran, 2018).
Moreover, complementation by a second symbiont is relatively common (~10% the species) in aphids. This usually happens without replacement, with the second symbiont often providing for the biosynthesis of biotin and riboflavin, less usually providing for tryptophan and histidine (Manzano-Marín *et al*, 2022) . In the Lachninae subfamily for example the *Buchnera* symbiosis is often complemented by a second

symbiont usually represented by *Serratia symbiotica*. In some cases there had also been a further replacement of *S. symbiotica* with a new other bacterium, keeping the "old" *B. aphidicola*. Interestingly, these replacements are not associated with any change in the ecological niche, so the new acquisitions are more probably dependent on genetic drift and degradation of the already present bacteria more than by new opportunities for the host offered by the new bacteria (Meseguer *et al*, 2017). Instead, it is the *Buchnera* lifestyle in the evolutionary rabbit hole of the symbiosis that opens new ecological niches for the new symbionts. This process does not occur through pervasive events and big genome rearrangements in the *Buchnera* genome, but just through small-scale mutations such as inactivation of a few genes (Manzano-Marín *et al*, 2022) .
One of the few ways to escape the evolutionary rabbit-hole is to shelter the genes needed to support the symbiosis from the genome erosion, that means having these genes integrate in the host genome. That can happen with a direct horizontal gene transfer from the bacteria to the host, or otherwise acquiring the needed genes through a third source. In time this will lead to the stabilisation of the symbiont as a proto-organelle, likely what happened to plastids and mitochondria (Husnik & Keeling, 2021).

The adaptation in long term partnerships also affects the host: keeping beneficial bacteria and passing them down to the next generation increases the host fitness. One important step for the host is to control the colonisation of the symbiont, keeping a useful bacterial load, avoiding a more profound colonisation. In that regard, the regulation of the immune system is fundamental: if fully active it could lead to an immune targeting of symbionts, but an excessive inactivation would make the host unable to fight off infections (Zug & Hammerstein, 2015). Another layer to be considered is that a chronic activation of the immune system can be damaging to the host by itself (He *et al*, 2013).

Three different modalities are currently described (Ferrarini *et al*, 2022) to allow the host to have a functional balance between sufficient immune response and avoiding the loss of the symbionts:

- Compartmentalization of the symbionts in specialised niches
- Modifications in the bacterial symbionts promoting immune tolerance
- Host capability to differentiate between the microbe-associated molecular patterns in the symbionts from the ones present in pathogens

The compartmentalisation not only avoids unnecessary immune activation, but has other advantages such as keeping the bacteria in organs in which the symbiont has a functional role (ex. digestive organs), and transfering to the next generations (e. g. localisation in the ovaries).
Some hosts have specialised cells that host bacteria, namely bacteriocytes, in particular insects such as aphids (Buchner *et al*, 1965), planthoppers (Wang *et al*, 2021b), cicadas (Wang *et al*, 2021a) and beetles (Pierantoni, 1927).
One example of the role of bacteriocytes and how compartmentalisation works is the beetle *Sitophilus oryzae*, a rice pest: this insect has an intracellular symbiont, *Sodalis pierantonius,* an *Enterobacterales* which is responsible for the provision of amino acids like tyrosine and phenylalanine, that are needed for cuticle synthesis (Vigneron *et al*, 2014). The bacteria are kept in specialised cells, the bacteriocytes, that are localised in the foregut-midgut junction during larval stages (Heddi *et al*, 1999).
The host immune system of *S. oryzae* does not discriminate *S. pierantonius* from pathogenic bacteria, and is active against it outside the bacteriocyte, as the symbiont presence in the hemolymph causes the production of antimicrobial peptides (Anselme *et al*, 2008). So, the control of the immune reaction relies upon the bacteriomes: they are able to physically separate the bacteria from the antimicrobial peptides, controlling the localisation of the bacteria and the antimicrobial compounds (Ferrarini *et al*, 2022).

An important part of the response of the innate immune system in metazoa, both in vertebrates and invertebrates, is the reaction to pathogen associated molecular patterns (PAMPs). One of the most important PAMP is peptidoglycan (PG), as it is present in most bacteria, and absent from almost every eukaryote (Björn, 2020). In the genome reduction of intracellular bacteria, including mutualistic symbionts and pathogens, it is common to lose components of the peptidoglycan biosynthesis, making the bacteria less targeted by the immune system. These losses are possible thanks to their localisation inside the cell in an isotonic environment that makes them less susceptible to osmotic shocks, and also in

proximity of the cell immune receptors. So, losing components of the PG biosynthesis has a limited cost, and more advantages, compared to free living bacteria, and occurs multiple times in different organisms, resulting in no PG, or a partial reduced synthesis, described as PG-intermediate. Intracellular *Chlamydia* and *Coxiella* endosymbionts are examples of such cases (Otten *et al*, 2018).

Differentiating between symbionts and other potentially pathogenic bacteria can be difficult, and in some cases the host is so dependent on its symbiont that it is more favourable to have a generally less efficient immune system than to risk losing the symbiont. For example, aphids lack genes like defensins, peptidoglycan receptors, immune deficiency signalling pathway, that are present in other insects, and have a role to contain gram-negative bacteria infections (Gerardo *et al*, 2010). This is a sort of specular situation, even if limited, to the genome reduction in the bacterial symbionts, in which the symbiosis is taking a toll on the genomic content of the host (Bennett & Moran, 2015).

1.2 Ticks

An important system to study symbiosis, and the most prominent in my PhD work, is the tick (Ixodida).

The most ancient fossil of a tick, *Carios jerseyi*, dates back to 90-94 Ma ago (Klompen & Grimaldi, 2001), and had been proposed (Walter & Proctor, 1998) that ticks originated from an ancestral mite scavenger that feeded on lymph fluids and only later on blood.
One of the first historical reports about ticks is attributed to Homer in the Odyssey (Philip, 1953), and then by Aristoteles and Pliny the Elder. The first possible representation dates to the 1500 BCE, in the tomb No. 155, Dra Ahn el-Nago, Western Thebes, where a hyena head infested by ticks is represented (Arthur, 1965).
Among blood-feeding arthropods, ticks are extensively studied due to their prominent role as disease vectors for both humans and domestic animals (Dantas-Torres *et al*, 2012). The reported cases of tick-borne disease have been increasing, with a doubling of reported cases between 2004–2016 in the USA, mostly Lyme disease infections. Over 30 000 cases of Lyme were reported in 2016, but this is considered a gross underestimation due to the chronic and aspecific nature of the disease and to the difficulties in obtaining a reliable diagnosis. The true numbers are estimated to be 10 times higher (Rosenberg *et al*, 2018). The increase of cases is influenced by the variation of population size and the geographic range of both the reservoir hosts and their ticks (Paules *et al*, 2018).

There is a high diversity in the pathogens vectored by ticks, which include bacteria, viruses and protozoa, affecting wild and domestic animals, and humans, that are often dead-end hosts (Kernif *et al*, 2016).
The most important and prevalent disease is Lyme disease, caused by *Borrelia burgdorferi* sensu lato, a complex that includes at least 20 different bacterial species. In Europe the principal etiological agents are *Borrelia afzelii* and *Borrelia garinii*, while in North America is *Borrelia burgdorferi* sensu stricto (Stanek & Strle, 2018).
Among the other bacterial pathogens there are various members of the *Rickettsia* genus, causing spotted fevers, various *Anaplasmataceae* including *Ehrlichia*, causing ehrlichiosis, or *Anaplasma*, *Coxiella burnetii*, agent of the Q-fever, and *Francisella tularensis* that causes tularemia (Madison-Antenucci *et al*, 2020).
Between these bacteria, *C. burnetii* has received more attention, as it is capable of causing large outbreaks such as in the Netherlands, in which more than 4000 people were infected between 2007 to 2010 (Delsing *et al*, 2010). That can happen as *C. burnetii* is capable of surviving in the environment so Q-fever cases in humans are usually acquired from domestic animals (cattle, sheeps and goats) through contaminated aerosol, and not directly from the tick bite (Eldin *et al*, 2017).
Babesia is a protozoa transmitted by ticks, mostly by Ixodidae (Yabsley & Shock, 2013), and it is a pathogenic agent of increasing importance, affecting both human and veterinary health (Gray *et al*, 2010). Human babesiosis has a wide spectrum of effects, depending on the *Babesia* species, and the patient's immunocompetence: *Babesia divergens*, common in Europe, often causes an acute disease with severe symptoms, while B. *microti*, the most common species in North America, in one third of the cases is asymptomatic, but can also lead to severe disease especially in immunocompromised or splenectomised patients (Krause *et al*, 2003).

Viruses include in particular the group of tick-borne encephalitis viruses, such as the tick-borne encephalitis virus (TBEV) in Asia and Europe (Süss, 2011) and the Powassan virus (POWV), the only known North American tick-borne encephalitis-causing flavivirus (Krow-Lucal *et al*, 2018), and viral hemorrhagic fevers like the Alkhurma hemorrhagic fever (Al-Tawfiq & Memish, 2017).
Of particular importance is the TBEV, a flavivirus divided in three subtypes: Far-Eastern, Siberian and European. It is transmitted by ticks of *Ixodes* genus - or rarely consuming unpasteurized milk - , in Europe mostly by *Ixodes ricinus,* in part of eastern Europe, Russia, and the far-east Asia by *Ixodes persulcatus*, and in Japan by *Ixodes ovatus*. TBEV affects many European countries, especially Baltic and Central European countries, with a peak in the warm season, and since 2017 there is a gradual increase of TBE cases reported by the European Centre for Disease Prevention and Control (ECDC 2022). Most of the infections are asymptomatic, but can result in non-specific symptoms, or meningitis, meningoencephalitis and meningoencephalomyelitis with possible post-encephalitic syndrome (Bogovic & Strle, 2015).

Differently from many other metazoa, which usually have digestive tracts colonised by a wide and numerous range of bacteria, ticks midguts usually have limited bacterial presence, which has been reported to be unstable, without a common bacterial core (Ross *et al* 2018, Guizzo *et al*, 2020). Instead, other organs have a stable, abundant and simple bacterial community with a single or very few bacteria in high numbers: in the ovaries we have *Midichloria* in *I. ricinus* (Sassera *et al*, 2008) and *Coxiella* endosymbionts in the *R. microplus* (98,2% of the 16S rDNA) of engorged females (Andreotti *et al*, 2011).
Thus, most tick species tend to maintain a single species of bacteria as principal endosymbiont, in high numbers, usually present in the ovaries and vertically transmitted. These bacteria, being maternal inherited, form long-lasting bonds with their hosts, showing partial co-phylogeny. However, compared to what happens in other arthropods with restricted diet, tick symbiotic communities present relatively fluid structures. Other bacteria can be acquired through horizontal transfer, for example through tick co-feeding on the same vertebrate, and a novel symbiont can replace the original one. Also, closely related tick species, and even different individuals of the same species, may harbour different sets of bacterial symbionts, having acquired them through vertical or horizontal transfer (Duron *et al*, 2017).

Some species of bacteria show a partial exclusion pattern, namely the presence of two different bacteria in the same individual host less often than expected by chance. For example, individual ticks seldomly have both *Coxiella* and *Francisella* endosymbionts or *Coxiella* and *Rickettsia* endosymbionts (Duron *et al*, 2017). This has been proposed to be a way for the tick to control the costs of harbouring these bacteria: a single species would be sufficient to provide what the host needs, and could be generally better than maintaining multiple infections that would provide limited advantages for the host, but increased costs (Oliver *et al*, 2006; Vautrin & Vavre, 2009).

The transmission of symbionts is dependent on three transmission routes: through transovarial transmission (maternal inheritance) (Lalzar *et al*, 2014), transstadial transmission - the maintenance from a tick life stage to the subsequent- (Kalmár *et al*, 2015), and horizontal transfer (e.g. through co-feeding) (Randolph *et al*, 1996).
In relation with the transmission mode, specific tick organs are colonised by the bacteria: vertically transferred symbionts usually localise inside ovaries. Instead, localising in the salivary glands is typical of tick-borne pathogens (such as *Anaplasma* or *Borrelia*) as through tick bites they can reach their next host. However, also nutritional symbionts can localise in the salivary glands, so that they can be transferred horizontally to other ticks, probably mainly through co-feeding. Is not unusual that nutritional symbionts localise in both ovaries and salivary glands, so that they can be transferred in both ways.

1.3 Bacterial mutualists in ticks

The most common, and more deeply investigated, tick bacterial symbionts are: the *Coxiella* endosymbionts (CE), *Candidatus* "Midichloria" symbionts, the *Francisella* endosymbionts (FLE), the *Rickettsia* symbionts, and more sporadically, *Rickettsiella* and *Spiroplasma* (Duron *et al*, 2017).

1.3.1 *Coxiella* endosymbionts (CEs)

CEs belong to the *Coxiella* genus, which is part of the *Legionellales* order. This order, which encompasses a vast number of symbionts and pathogens of eukaryotes, is divided in two families, the *Legionellaceae*, which include *Legionella pneumoniae*, the principal agent of Legionnaires' disease, and *Coxiellaceae*, encompassing the agent of Q fever, *Coxiella burnetii*. All described members of the order are capable of invading eukaryotic cells, establishing replication vacuoles, with mammal pathogens often infecting alveolar macrophages (Duron *et al*, 2018a).

CEs are the most common tick symbionts (Duron *et al*, 2017), widely distributed in various tick genera including *Rhipicephalus*, *Amblyomma*, *Haemaphysalis*, *Ornithodoros*, *Argas* and *Carios* (Guizzo *et al*, 2017).

CEs genomes present genes for the provision of vitamins and co-factors, including biotin, CoA, FAD, riboflavin (B2), pyridoxine (B6), folic acid (B9) and pantothenate (B5), as well as, depending on the strain, thiamine (B1) and nicotinate (B3) (Guizzo *et al*, 2017).

Through Fluorescence fluorescence in situ hybridization (FISH) and TEM, it has been shown that CEs localise in the ovaries, salivary tissues, midgut, and Malpighian tubules (Klyachko *et al*, 2007). As the ovarian localisation suggests, the vertical transovarial transmission of CE was confirmed by the presence in eggs and larvae. Horizontal transmission should be also possible, considering the CEs presence in the granular acini of the salivary glands (Klyachko *et al*, 2007). The presence in Malpighian tubules and midgut supports the CEs role in provision of supplementary nutrients and in nitrogen metabolism. The administration of tetracycline, both through injection into the tick's hemocoel and through treatment of the vertebrate host, reduces the presence of CEs, causing the interruption of tick development at the metanymph stage, and, when treating adult females, inhibition of oviposition and hatching, demonstrating the necessity of the CE presence (Guizzo *et al*, 2017).

The CEs phylogeny shows a partial pattern of co-cladogenesis, but does not evidence a strict co-evolution with their tick hosts. Closely related CEs can be associated with very divergent host species, and also inside some species multiple CEs strains with different phylogenetic localisation are present, possibly caused by multiple events of horizontal transfer (Machado-Ferreira, 2016).

1.3.2 "*Candidatus* Midichloria" bacteria

"*Candidatus* Midichloria", hereafter *Midichloria*, are bacteria from the order *Rickettsiales*, belonging to the family "*Candidatus* Midichloriaceae" (Montagna *et al*, 2013). The namesake of the species derives from the unique trait of certain strains (Floriano *et al*, 2022) to colonise the intermembrane space of mitochondria of its host cell. With the exception of a closely related bacterium that arguably causes the Red Mark Syndrome in farmed rainbow trout (Galeotti *et al*, 2021), hosts of *Midichloria* species are ticks. In particular, "*Candidatus* Midichloria mitochondrii" is hosted by the castor bean tick *Ixodes ricinus*, widespread and common in the whole Europe and North Africa, and an important vector of various pathogens of medical and veterinary importance, in particular *Borrelia burgdorferi sensu lato*, *Rickettsia, Anaplasma*, *Francisella tularensis, Babesia*, and tick-borne encephalitis virus. The prevalence of "*Candidatus* Midichloria mitochondrii" in females and immatures of *I. ricinus* is 100%, but the presence in males is limited to less than half of the individuals, usually with a lower load, probably due to the loss of the symbiont during development (Lo *et al*, 2006). *Midichloria* spp. are also present in other tick species, in particular other of the *Ixodes* genus (Duron *et al*, 2017).

Through the use of FISH, it has been shown that *Midichloria* is localised primary in the ovaries (Sacchi *et al*, 2004), and using other techniques, such as whole-mount *in situ* hybridization (Epis 2013), and mass spectrometry (Di Venere *et al*, 2015)) was shown the presence also in the salivary glands.

The ovary localisation is typical for vertically transmitted bacteria, as *Midichloria* is. The presence in the salivary glands is instead typical for horizontally transferred bacteria and tick-borne pathogens, as the bacteria could be injected during the blood meal, and infect the bitten vertebrate host or other ticks

co-feeding on the same host. Indeed, it has been shown that more than half of the people bitten by *I. ricinus* have anti-*Midichloria* antibodies, indicating at least a contact with *Midichloria* proteins (Mariconti *et al*, 2012).
Concerning the role of *Midichloria* for the tick, based on genome predictions it has the capability to synthesise biotin and B9 vitamin, and considering also the presence of *Midichloria* in the Malpighian tubules, this suggests a metabolic mutualism (Sassera, 2011). Also, a role in energy metabolism and oxidative stress response is possible, as genes are present, such as superoxide dismutase (SOD), ferrochelatase (FECH), haem exporter proteins, and transporters involved in ions translocation (Olivieri *et al*, 2019).

1.3.3 *Francisella* endosymbionts

Francisella endosymbionts (FEs) belong to the *Francisella* genus in the *Thiotrichales* order. The genus is composed by pathogens with a broad range of hosts: *F. tularensis*, the agent of tularemia, infects mammals - including humans -, *F. noatunensis* infects fishes but a subspecies infects the marine ciliate *Euplotes raikovi* (Schrallhammer *et al*, 2011).

FEs have been reported in multiple tick species, mostly in the genus *Hyalomma, Amblyomma and Dermacentor* (Duron *et al*, 2017).
FEs were firstly described as coccoid bacteria in *Ornithodoros moubata* in the Malpighian tubules and ovaries (Reinhardt *et al*, 1972). FEs is thought to have superseded the original *Coxiella* symbiont in various tick species, and it has been firstly sequenced and characterised in *Amblyomma maculatum*, even if the most of *Amblyomma* genus present *Coxiella* endosymbionts. The replacement of *Coxiella* has been hypothesised to be possible thanks to FEs having all the required metabolic capabilities of the "old" symbiont, and with more amino acid provision capabilities (Gerhart *et al*, 2016).
FEs are phylogenetic close to *F. tularensis,* with a slightly reduced genome - around 80% of the size of the pathogen - but with less signs of reductive genome evolution compared to symbionts with longer associations, and differently from other symbionts they present no sign of AT skewness (Gerhart *et al*, 2016). A third of the genes shows pseudogenisation, including various virulence genes. So, all these elements suggest that the FEs have been recently acquired, that their ancestor could have been a pathogen similar to *F. tularensis,* having lost through time its virulence genes. The process could be described as a domestication of a vector-borne pathogen, that was already adapted to live inside the tick host, with limited activation of the immune system and limited damage to the host, that was capable of replacing the previous *Coxiella* symbiont present, that had a genome already reduced by Muller's ratchet (Gerhart *et al*, 2018).
Coherent with the FEs presence in ovaries and salivary glands, both horizontal and vertical transfer play a role in maintenance of FEs in the tick hosts. Usually FEs of the same tick species are phylogenetically close, but there is only a partial co-cladogenesis comparing the phylogeny of FEs from different host species with the host phylogeny. Strains that are phylogenetically closer usually share the geographic location, or their tick hosts feed on the same animals. Therefore, this is likely the result of multiple horizontal transfers between species through co-feeding of their hosts (Buysse *et al*, 2022).

1.3.4 *Rickettsia* endosymbionts

The *Rickettsia* genus is composed by intracellular bacteria, which exhibit different roles, from vector-borne human pathogens to symbionts of diverse range of hosts, including non-hematophagous arthropods, and have been divided in groups: Rhizobius, Meloidae, Belli, Adalia, Canadensis, Helvetica, Scapularis, Transitional, Typhus, and Spotted Fever (Weinert *et al*, 2009; Davison *et al*, 2022).
Initially the genus was considered composed only of pathogens. One of the first identified symbionts was *Rickettsia peacockii* in the ovaries of the tick *Dermacentor andersoni* and was transmitted transstadially and transovarially (Niebylski *et al*, 1997). Interestingly, the presence of the tick symbiont was associated with a reduced diffusion of spotted fever caused by *Rickettsia rickettsii* (Burgdorfer *et al*, 1981), as the symbiont presence was interfering with the presence of the pathogen in the tick. A comparison of the genomes of *R. peacockii* and *R. rickettsii* showed a loss of virulence genes in the former, inactivated through a diffusion of transposons (Felsheim *et al*, 2009).
Various *Rickettsia* species are maternally transmitted, with no evidence of pathogenicity in vertebrates,

and were found in various tick species, such as *R. buchneri* in *Ixodes scapularis* (Kurtti *et al*, 2015) and *Rickettsia* species phylotype G021 In *Ixodes pacificus*, that is capable of providing folates (Bodnar *et al*, 2018), and together they form a clade of *Rickettsia* species that acts a nutritional symbiont in ticks.

1.3.4 *Rickettsiella*

Rickettsiella are bacterial symbionts with a wide variety of arthropod hosts, including ticks (Anstead & Chilton, 2014), but also insects, other arachnids, and crustaceans (Cordaux *et al*, 2007). Bacterial replication takes place in cell vacuoles in the fat body, the hepatopancreas, and other organs. As evident from the name, it was initially considered affiliated with the *Rickettsiales* order, due to its association with arthropods and intravacuolar localisation. However, with the first 16S rRNA gene sequence from *Rickettsiella grylli*, a parasite of the cricket *Gryllus bimaculatus*, it was identified as a member of the *Legionellales* (Roux *et al*, 1997).
Rickettsiella has different roles in the different hosts. In some cases it is a pathogen (Cordaux *et al*, 2007), in others it is a nutritional symbiont, for example the symbiont of poultry red mite has the genomic capabilities to synthesise B vitamins and co-factors (Price *et al*, 2021), and in the spider *Mermessus fradeorum* it is a reproductive manipulator, causing cytoplasmic incompatibility (Rosenwald *et al*, 2020). In ticks, *Rickettsiella* does not show a pathogenic role, nor an effect on the fertility(Duron *et al*, 2016).
Rickettsiella is present in various *Ixodes* ticks, in other hard tick genera such *Amblyomma* and *Haemaphysalis* and also in soft ticks, with a wide geographic localisation, from North America, to Australia and Europe (Garcia-Vozmediano *et al*, 2022). *Rickettsiella* presence shows an exclusion pattern, as no individuals were identified with a dual infection of *Rickettsiella* and other endosymbionts (Duron *et al*, 2017). The rate of infection is variable, with the highest rate in western Europe, and lower rates for northern and southern latitudes. This distribution in the various ticks is probably dependent both on the horizontal and vertical transfer (Garcia-Vozmediano *et al*, 2022). The *Rickettsiella* strains are often phylogenetically closer to strains from other arthropods than to other tick strains, suggesting frequent horizontal transfers between different arthropods (Duron *et al*, 2016).

1.3.5 *Spiroplasma ixodetis*

Spiroplasma is a gram-positive bacterium characterised by the absence of a cell wall. The genus includes bacteria that exhibit different behaviours, including pathogens, mutualists, and commensals of various species of arthropods and plants (Gasparich, 2002). Some *Spiroplasma* in insects are reproductive manipulators with male killing capabilities (Anbutsu & Fukatsu, 2006), a behaviour never observed in ticks (Binetruy 2019). In the same insect hosts, they can also confer protection to parasites through the production of ribosome-inactivating proteins (RIPs) (Hamilton *et al*, 2016).
Concerning the phylogeny, they are divided in three clades, based on 16S rRNA gene sequences: Apis, Citri-Chrysopicola and Ixodetis, with symbiont of tick in the Apis and Ixodetis clade (Ogata *et al*, 2021). The first *Spiroplasma* in ticks was isolated in a *I. pacificus* (Tully et al 1981), and this bacterium was later found in various hard ticks, but never in soft ticks (Duron *et al*, 2017),with prevalences depending on tick species, but also on the localisation (Klubal *et al*, 2016).
In different tick species the *Spiroplasma* symbionts are genetically quite similar, probably due to a vertical transmission. However, there is not a single monophyletic clade of symbionts of ticks, with some strains that are closer to *Spiroplasma* hosted in other arthropods than to symbionts of other tick species, indicating possibly a multiple horizontal transfer events to or from ticks, and limited horizontal transfer between different tick species (Ogata *et al*, 2021).
The potential pathogenicity of tick-associated strains to humans is unclear, as there are limited direct proofs. However some tick-borne *Spiroplasma* strains are very similar to *S. mirum* that cause brain damage in mice (Clark, 1974), and can cause neurological damage in ruminants (Nakao *et al*, 2021).

1.4 The role of bacteria on tick physiology

The mutualistic role of bacterial symbionts in tick physiology has experimental evidence: their depletion resulted in impaired growth and reproduction in multiple tick species belonging to different genera (Zhong *et al*, 2007; Guizzo *et al*, 2017; Ben-Yosef *et al*, 2020).

The symbiont localisation is informative also to understand the role in the tick physiology, as nutritional symbionts usually localise in Malpighian tubules, which are organs involved in tick osmoregulation and excretion. Therein, metabolic waste is usually concentrated (Sonenshine & Roe, 2013), and can be used by the endosymbiotic bacteria to produce the required nutrients, detoxificate the compounds in blood and more generally participate in the adaptation of the tick to its specialised diet.
Supplementation of nutrients to overcome the limits of the restricted diet is a common role for symbionts of blood-feeding arthropods. The genome of ticks appears to present some adaptation, as it contains genes involved in various physiological systems, such as in osmotic homeostasis or haemoglobin digestion (Jia, 2020), but many functions are "outsourced" in the genome of their bacteria.

One important task for the tick is the management of its sole food: vertebrate blood: has high quantity of proteins (90% of the dry weight), in particular haemoglobin and albumin, and is ingested in high volumes, up to one hundred times the weight of the tick before the blood-meal, requiring adaptation to scarce vitamin and lipid presence, haem oxidative stress and osmotic stress (Sterkel *et al*, 2017).

1.4.1 Vitamin and co-factors provision

The provision of vitamins and co-factors not available in blood is a fundamental contribution of bacterial symbionts to the adaptation to blood-feeding lifestyle of ticks. Almost all symbionts of blood-feeding arthropods have the genes to provide a core group of vitamins and co-factors: biotin, B2 (riboflavin) and B9 (folate) (Duron, 2020). Biotin is required for various carboxylases, including enzymes involved in fatty acid synthesis and degradation, and gluconeogenesis. B2 is required in flavoproteins that are necessary for electron transport chain, citric acid cycle and β-oxidation of fatty acids. Folate is required in nucleic acids and amino acid metabolism (Douglas, 2017).
The bacterial mutualists could have acquired these pathways in different ways, through different ecological constraints: the biosynthetic capability could have been present in an ancestor of the current mutualist, already adapted to a mutualistic lifestyle with another eukaryotic partner, possibly in a facultative symbiosis, as it happened various time in the *Enterobacteriaceae* lineage (Duron, 2020). In some case the provision of vitamins is an exaptation of pathways that had other functions: in *Francisella* pathogens the synthesis of biotin is required for virulence (Feng *et al*, 2014), but in non-pathogenic mutualistic FEs the bacteria provide vitamins to their tick host (Gerhart *et al*, 2018). There are also more complex situations: in the *Rickettsiales* ancestor the biosynthetic pathway for B9 (folate) was present, then it was lost in several sublineages, but still kept in the mutualist taxa (Sassera, 2011).
Another possibility is to acquire these genes through horizontal transfer, such as the case of *Legionella polyplacis*, symbiont of the louse *Polyplax serrata*: it acquired the full biotin operon from another distant bacteria, enabling the mutualistic lifestyle (Říhová, 2017).
The role of B vitamins in ticks had been demonstrated in one case: the removal of the FEs through antibiotics in the soft tick *Ornithodoros moubata* reduced strongly - completely in females - the adult emergency, and reduced the size of the survivors. Moreover, treating by providing B vitamins rescued the effects of the antibiotic intervention (Duron *et al*, 2018b).

1.4.2 Energy provision

Another possible role is the provision of ATP: *M. mitochondrii* has the gene c*bb*$_3$, an oxidase which reduces O_2 with lower efficiency but higher affinity, working also in limited oxygen conditions. During the oogenesis there is an increase of metabolic and biosynthetic activity causing a high oxygen and energy use (Aboul-Nasr & Bassal, 1972). So, it is possible that *Midichloria* in ovaries provides additional ATP in this phase (Sassera, 2011).

1.4.3 Haem

Haem is a fundamental prosthetic group in all aerobic life forms, and the biosynthesis pathway is highly conserved in the whole tree of life, with the exception of a few parasites (Panek & O'Brian, 2002). These include ticks, which are dependent on the haem they get through blood-feeding to generate viable eggs (Perner, 2016). Ticks' genomes lost most of the haem biosynthesis and degradation genes, keeping only the last three steps of the pathway (Jia, 2020).

Some of the symbionts possess the genomic capabilities for haem biosynthesis: *Midichloria mitochondrii* (Sassera, 2011), *Francisella* endosymbionts (Gerhart *et al*, 2016), but not the *Coxiella* endosymbionts, that have pseudogenised forms of most the genes (Gottlieb 2015). That would suggest that the synthesis from the endosymbionts is not required by all ticks, and when present is still not sufficient to cover the host needs: otherwise *M. mitochondrii* in *I. ricinus* would be able to rescue the haem starvation in (Perner, 2016).

A peculiarity of the tick haemoglobin digestion has been shown in *Rhipicephalus microplus*, in which the haem is digested within midgut cells, where is precipitated inside a specialised membrane-delimited organelle, called hemosome (Lara *et al*, 2003). The haem detoxification in the hemosome requires adequate redox balance, as inhibition of catalase reduces the tick life-span and the rate of egg-laying, impairing haem aggregation (Citelli *et al*, 2007).

These functions could be supplemented by the bacterial symbionts: *M. mitochondrii* possesses various genes involved in anti-oxidative response, including superoxide dismutase, peroxiredoxins, lipoic acid and thioredoxins, and genes involved in the iron metabolism such as protohaem ferro-lyase and haem transporter proteins. These genes are also present in other symbionts, suggesting a common role conserved in the tick symbionts, though with small differences, as protohaem ferro-lyase is absent in CEs, and genes encoding for haem exporter proteins are absent in FEs (Olivieri *et al*, 2019).

1.4.4 Ammonia detoxification

Blood is rich in proteins, and protein degradation produces ammonia (Sterkel *et al*, 2017). One way to avoid ammonia toxicity is through storage in non-toxic form, such as proline (Pennington *et al*, 2003). Indeed, the flux-balance metabolic analysis of *Coxiella* endosymbionts (CEs) of *Rhipicephalus* has shown an excess production of L-proline, paired with multiple copies of proline/betaine transporter (Tsementzi *et al*, 2018). A further confirmation comes from a proteomic analysis of CEs (Cibichakravarthy *et al,* 2022): the L-proline synthesising enzyme ornithine cyclodeaminase is present in Malpighian tubules. And at the same time in ovaries there is the abundant production of proline dehydrogenase PutA, that converts proline to glutamate, that can be used as a source of carbon and nitrogen (Tanner, 2008). Thus, the tick symbionts possibly can control the excess of ammonia, and help with the long-term storage and use of nutrients after the blood meal, in particular for the egg development and symbiont maintenance in the next generations.

1.4.5 Behaviour

The symbionts can also affect the behaviour of their host, modifying the feeding capability.
The *Coxiella* endosymbiont of *Haemaphysalis longicornis* possess genes for production of chorismate, a precursor of tryptophan. The chorismate made by *Coxiella* endosymbionts stimulates the biosynthesis of 5-hydroxytryptamine (serotonin) in the tick, inducing the expression of aromatic amino acid decarboxylase, required for catalysis decarboxylation of 5-hydroxytryptophan to serotonin.
The presence of serotonin in tick synganglion and midgut promotes tick feeding activity, and conversely inhibiting the chorismate pathway using glyphosate reduces the tick feeding behaviour. A similar result is obtained using tetracycline, which removes the bacterial community. Administering chorismate, tryptophan or serotonin to the antibiotic treated population restores the feeding behaviour (Zhong *et al*, 2021).

Host manipulation is a common strategy of parasites, including vector-borne, to increase their diffusion, increasing the number of bitten vertebrate hosts (Lefèvre & Thomas, 2008). Even if it is not possible to change the number of hosts bitten, the tick behaviour can also be affected by some of its tick-borne pathogens with other means: *Borrelia burgdorferi* causes an increase of lifespan of *I. ricinus*, and the time spent questing, increasing the probability of the transmission (Herrmann & Gern, 2015). This is achieved by different actions: ticks infected by *Borrelia* are more resistant to heat and desiccation (Herrmann & Gern, 2010), tend to move less, staying more often immobile and thus reducing water consumption (Herrmann & Gern, 2012), and also have higher fat reserves (Herrmann & Gern, 2013). Thus, the infected

ticks have slower metabolism, and higher energy reserves that they use to spend more time questing, increasing the chance of pathogen diffusion.

Tick questing speed is affected also by the infection of *Anaplasma phagocytophilum*, another tick-borne pathogen. Infected ticks change the expression of the proteins Hsp70 and Hsp90, that cause an increase of the questing speed during high temperatures (Busby *et al*, 2012). *A. phagocytophilum* also has a protective role, as it causes the expression of the antifreeze glycoprotein IAFGP in its host *I. scapularis.* This protein improves the cold tolerance, increasing the tick survival in cold environments (Neelakanta *et al*, 2010).

1.5 Interaction between mutualists and pathogens

The bacteria present in the same host, including mutualists or pathogens, may interact with each other, with different possible effects. There can be a direct competition for the resources available in the same niches, or there can be indirect effects as the symbiont can modify the host environment for their needs or act on the immune system.

In some cases the presence, and the load, of the mutualist can have a protective role: *Coxiella*-endosymbionts of *Rhipicephalus haemaphysaloides* antagonise the transmission of *Babesia microti* (Li *et al*, 2018). *Rickettsia* symbionts limit pathogenic *Anaplasma marginale* infection in *Dermacentor andersoni*, possibly through competition for host resources or through a direct inhibition of the competitor (Gall *et al*, 2016). This case could be similar to the competition between the symbiont *Rickettsia rickettsii* and *Rickettsia* causing spotted fever (Burgdorfer *et al*, 1981). However, the opposition effect is not true for all combinations of mutualists and pathogens: reduction in the abundance of *Francisella* symbiont in the *D. andersoni* is associated with a decreasing infection level of the pathogenic *Francisella novicida* (Gall *et al*, 2016).

One case of the multiple effects of host environment modification is the peritrophic matrix (PM): PM covers the tick gut and protects the epithelium from mechanical damage, and acts as a physical barrier to infections. *A. phagocytophilum*, as previously reported, upregulates the expression of the *iafgp* gene in the tick host. In the gut IAFGP binds to bacterial peptidoglycan, inhibiting biofilm formation. This causes an alteration in the gut microbiome, the dysbiosis damages the PM matrix, helping the *A. phagocytophilum* infection (Abraham *et al*, 2017). At the same time, a compromission of the PM structure reduces the colonisation of the tick by *B. burgdorferi* (Narasimhan *et al*, 2014), showing that the interplay between different bacteria can be complex and with multiple effects.

The observation patterns of co-presence / exclusion of mutualists and pathogens without experimental settings does not necessarily imply an interaction between them, as it can be caused by ecological variables (Krawczyk *et al*, 2022): In *I. ricinus* nymphs *M. mitochondrii* is associated with the pathogens *B. burgdorferi* s.l. and *Neoehrlichia mikurensis*, and negatively associated with the pathogen *A. phagocytophilum.* The associations can be explained as *M. mitochondrii* is more common in female ticks (Lo *et al*, 2006), which take larger blood-meals from their host (Daveu *et al*, 2021), that could expose them more to horizontally transmitted pathogens. The negative association with *A. phagocytophilum* can be dependent on its host (the deer), that is predated by nymphs and not larvae, differently from the hosts (small rodents) of the two other pathogens (Krawczyk *et al*, 2022).

Aim

The purpose of my PhD was to provide novel insight towards the characterization of the evolution of symbiotic bacteria in ticks. In particular, I aimed at understanding how the selective pressures of symbiotic lifestyle affects these bacteria, reconstructing how the relationship with their host formed and evolved, through the analysis of phylogenomics and genomics data.

My work is divided into four different chapters:

- **Chapter II:** The sequencing the *Coxiella symbiont* of *Amblyomma americanum*, and performing a comparative genomic analysis of the order to clarify the relationship of the symbionts with the pathogen *Coxiella burnetii*
- **Chapter III:** The characterisation of a multipartite symbiosis system in *Hyalomma marginatum* that includes *Midichloria* and *Francisella*
- **Chapter IV:** The analysis of the population structure of symbionts of *Ixodes frontalis, Midichloria* and *Spiroplasma*, and how that relate with the two sub-clades of the host
- **Chapter V:** The analysis of the deep phylogeny of the *Rickettsiales* order, that is composed only by host-dependent *professional* symbionts. The order includes various tick nutritional symbionts (*Midichloria* and symbiotic *Rickettsia)* and tick-borne pathogens (e.g. *Ehrlichia*, *Anaplasma, Rickettsia* spp.*)*, but encompass a much larger variability, with a diverse range of hosts and roles

II *Coxiella* endosymbionts, acquired or ancestral pathogenicity?

Published as: **Tiago Nardi**, Emanuela Olivieri, Edward Kariuki, Davide Sassera, Michele Castelli, Sequence of a *Coxiella* Endosymbiont of the Tick *Amblyomma nuttalli* Suggests a Pattern of Convergent Genome Reduction in the *Coxiella* Genus, *Genome Biology and Evolution*, Volume 13, Issue 1, January 2021, evaa253, https://doi.org/10.1093/gbe/evaa253

2.1 Introduction

Differently from the scenario inferred for the *Francisella* endosymbionts (Gerhart *et al*, 2016) , the origin of the *Coxiella* endosymbionts is less clear. Their mutualistic association with ticks is assumed to be of ancient origin, dating back to early tick evolution (Douglas, 2017). The currently known members of the genus, other than the *Coxiella* endosymbionts (CEs), are the pathogen *Coxiella burnetii,* causative agent of the Q-fever, and the scantly described *Coxiella cheraxi*, pathogen of the redclaw crayfish (*Cherax quadricarinatus*) (Elliman & Owens, 2020), the only member of the genus that does not have ticks as primary hosts.

Currently four phylogenetic clades (A-D) are identified in the genus *Coxiella*, exhibiting only partial congruence with their hosts' phylogeny (partial co-cladogenesis). All *C. burnetii* lie in clade A. Clade C - also called *Coxiella mudrowiae* - displays a good degree of co-cladogenesis with *Rhipicephalus* tick hosts, while CEs of two unrelated clades (B and D) were found in *Amblyomma* tick hosts (Duron et al. 2015). Interestingly, the *Amblyomma* hosts of clade B symbionts came from the African continent, while clade D hosts are American (Duron *et al*, 2015; Binetruy *et al*, 2020).

We can observe a wide distribution of *Coxiella* symbionts in ticks, across the various species, genera and families, indicating a long shared evolutionary history. At the same time, all *C. burnetii* isolates are nested within a single *Coxiella* clade, with a more limited diversity that could be compatible with a relatively recent specialisation. So, it has been proposed that *C. burnetii* arose from a maternal inherited endosymbiont, acquiring pathogenic capabilities and becoming a vertebrate parasite (Duron *et al*, 2015), an opposite scenario of what happened more recently to *Francisella*.
At the same time, pathogenic capabilities have been reported for some CEs, in particular from the clade C, specifically in birds (Needle *et al*, 2020), but also in humans (Angelakis *et al*, 2016). Pathogenic potential has been reported also for other related bacteria: *C. cheraxi*, assumed to be phylogenetically basal to the genus, outside all the tick associated *Coxiella*, is a pathogen (Elliman & Owens, 2020). Pathogenicity is considered an ancestral trait in the *Legionellales* order, with different independent adaptations to symbiotic lifestyle such as (*Legionella*) and *Rickettsiella* (Hugoson *et al*, 2022). However, a "pathogenic origin" scenario, would assume a series of independent transitions to a (mostly) mutualistic lifestyle, in all 4 clades of CEs.

A better understanding of the origin and evolution of *C. burnetii* would describe a scenario of symbiont/host adaptation from or to pathogenesis, understanding how this relationship started, potentially providing insights on the origin and evolution of pathogens capable of causing human diseases.
Currently, genomes of representatives of three of the four *Coxiella* clades have been sequenced: three genomes from clade D, four from clade C and *C. burnetii* from clade A. Clade B is the only one without a sequenced member.
Thus, to have a complete picture of the CEs variability I explored a currently neglected clade of *Coxiella* endosymbionts. Good candidates to get a genome from the clade B CEs are the *Amblyomma* species of African origin, so we collected an *Amblyomma nuttalli* tick from Kenya, and sequenced its CE.

2.3 Aim

The aim of this study is to assemble a genome from *Coxiella* clade B, annotate it, analyse its characteristics, and then use it for comparative genomics and phylogenetic analysis with other bacteria from the *Coxiella* genus.
Specifically, I aim to understand the phylogenetic relationship between the different symbionts, if there are differences in their metabolic capabilities and if these differences could highlight specific traits of their hosts.
Then, I plan to use this information to trace the evolution of these symbiotic relationships including the development or loss of pathogenicity in the symbionts, that could be used to clarify if the origin of CEs lies on pathogens or in symbiotic bacteria.

2.4 Methods

2.4.1 Sequencing and assembly

An adult female of *Amblyomma nuttalli* was collected from a white rhinoceros (*Ceratotherium simum*) in the Masai Mara National Reserve, Kenya in February 2016. The tick was morphologically identified following standard taxonomic keys (Theiler & Salisbury, 1959)(Theiler & Salisbury 1959) and subjected to DNA extraction, using NucleoSpin® Tissue Kit (Macherey Nagel, Duren, Germany), according to the manufacturer's instructions. DNA was subjected to Illumina HiSeq X by Admera Health (South Plainfield, NJ, USA) using a Nextera XT library, obtaining 27,3511,224 150-nt paired-end reads.
The reads were assembled using SPAdes (3.6.0) (Bankevich *et al*, 2012), and subjected to a modified version of the blobology pipeline (Kumar *et al*, 2013), in order to select only the symbiont sequences (Figure 1). After excluding contigs with eukaryotic hits, we selected contigs with a $\log_{10}$ coverage higher than 2.5, extracted and reassembled separately the reads mapping on those contigs(Langmead & Salzberg, 2012), and revised the results manually. Contigs belonging to the symbiont were initially identified by its rRNA genes, then, other putative contigs were added using the assembly-derived connections between contig ends and inspected with the Bandage software (Wick *et al*, 2015).
Finally, in order to close the assembly, PCR reactions were performed with primers designed in proximity of contig ends, product were sequenced and results processed, as described previously (Castelli *et al*, 2019).

2.4.2 Genomic and comparative analysis

The completeness level of the genome was confronted with all published CE genomes using Busco 4.0.2 with lineage "gammaproteobacteria_odb10 (2019-04-24)" (Seppey *et al*, 2019).

Genome annotation was performed using Prokka 1.11 (Seemann, 2014) and manually curated by inspecting blastp hits of predicted ORFs on NCBI nr, Uniprot, and *Legionellales* sequences.

ISEScan (Xie & Tang, 2017) and ISfinder (Siguier *et al*, 2006) were used to identify insertion sequences, and PHASTER (Arndt *et al*, 2016) for prophages. Pseudogenes were predicted in the novel genome, as well as in all published CE genomes and representative *C. burnetii* genomes, by using Pseudo-finder (Syberg-Olsen *et al*, 2020).
COGs were predicted using the NCBI pipeline (Galperin *et al*, 2015) on validated genes (i.e. ORFs excluding predicted pseudogenes). COG repertoires were used for comparative analyses. Metabolic pathways were manually reconstructed employing the BioCyc database reference (Karp *et al*, 2019) and KEGG (Kanehisa & Goto, 2000).
The whole bioinformatic workflow is summarised in Figure 2A and figure 2B.

2.4.3 Phylogenetic inference

To assess the phylogenetic relationship between the various CEs I obtained two different phylogenetic trees using two different datasets, using for both maximum likelihood (ML) and bayesian inference (BI) methods. The first dataset is based on a wide number of different CEs, but with few genes, the second dataset uses only species with a complete genome available, thus with an increased number of genes and phylogenetic information per organism.

The first dataset consists on the multi locus sequence typing (MLST) of five housekeeping genes, and is based on a previous dataset (Duron *et al*, 2015) that included a wide sample of different CEs, enlarged with CE *A. nuttalli* and all the CEs with an available genome, thus employing five genes (*16S, 23S, dnaK, rpoB, groEL*) from 96 organisms.

For the second dataset, namely the phylogenomic tree, we included all the previously publishedCE genomes, a selection of *C. burnetii*, a representative selection of *Coxiellaceae*, including 1 MAG (metagenome assembled genome), and two other *Legionellales* as outgroup. Using OrthoFinder (2.3.3) (Emms & Kelly, 2019) we identified single copy conserved orthologs that were used for the analysis.
For both sets, respectively on the nucleotide and protein sequences, each single gene was aligned separately using Muscle (Edgar, 2004), polished with Gblocks (Talavera & Castresana, 2007), and concatenated with an in-house script.
For each set, I inferred the best model (GTR+I+G and LG+I+G, respectively) using modeltest-ng 0.1.3 (Darriba *et al*, 2019), built a maximum likelihood tree with RAxML 8.2.4 (Stamatakis, 2014) with 1000 bootstrap pseudo-replicates, and a Bayesian inference tree with MrBayes (Ronquist *et al*, 2012) using three independent runs for 1 million and 250,000 generations, respectively, with a burn-in of 25%. The whole workflow for the phylogenetic analyses is summarised in Figure 7.

Table 1: List of Genomes Included in the Analyses, with Their Accession Numbers, and for Coxiella *Total Genome Size, Coding DNA Size, and GC Content*

Organism	**Accession**	**Genome Length (bp)**	**Coding DNA (bp)**	**CG Content**
Coxiella burnetii RSA 493	GCA_000007765.2	2,032,807	1601033	42.6%
Coxiella burnetii Dugway 5J108-111	GCA_000017105.1	2,212,937	1769726	42.4%
Coxiella* endosymbiont of *Amblyomma nuttalli	GCA_018107685.1	1,003,026	660534	35.9%
Coxiella endosymbiont of *Amblyomma americanum* 1	GCA_000815025.1	656,901	568842	34.6%
Coxiella endosymbiont of *Amblyomma americanum* 2	GCA_002850495.1	656,933	566434	34.6%
Coxiella endosymbiont of *Amblyomma sculptum*	GCA_009883795.1	622,921	539589	38.1%
Coxiella endosymbiont of *Rhipicephalus microplus* 1	GCA_002871095.1	1,194,772	683985	32.6%
Coxiella endosymbiont of *Rhipicephalus microplus* 2	GCA_002930125.1	1,296,467	623102	31.7%
Coxiella endosymbiont of *Rhipicephalus sanguineus*	GCA_002804145.1	1,715,759	907053	38.0%
Coxiella endosymbiont of *Rhipicephalus turanicus*	GCA_001077715.1	1,733,840	917489	38.2%
Aquicella siphonis	GCA_902459485.1			
"*Candidatus* Berkiella cookevillensis"	GCA_001431315.1			
Legionella pneumophila	GCA_000008485.1			
Environmental *Coxiellaceae* bacterium	GCA_001795425.1			
Rickettsiella grylli	GCA_000168295.1			
Rickettsiella isopodorum	GCA_001881495.1			
Rickettsiella viridis	GCA_003966755.1			
Tatlockia micdadei	GCA_000953635.1			

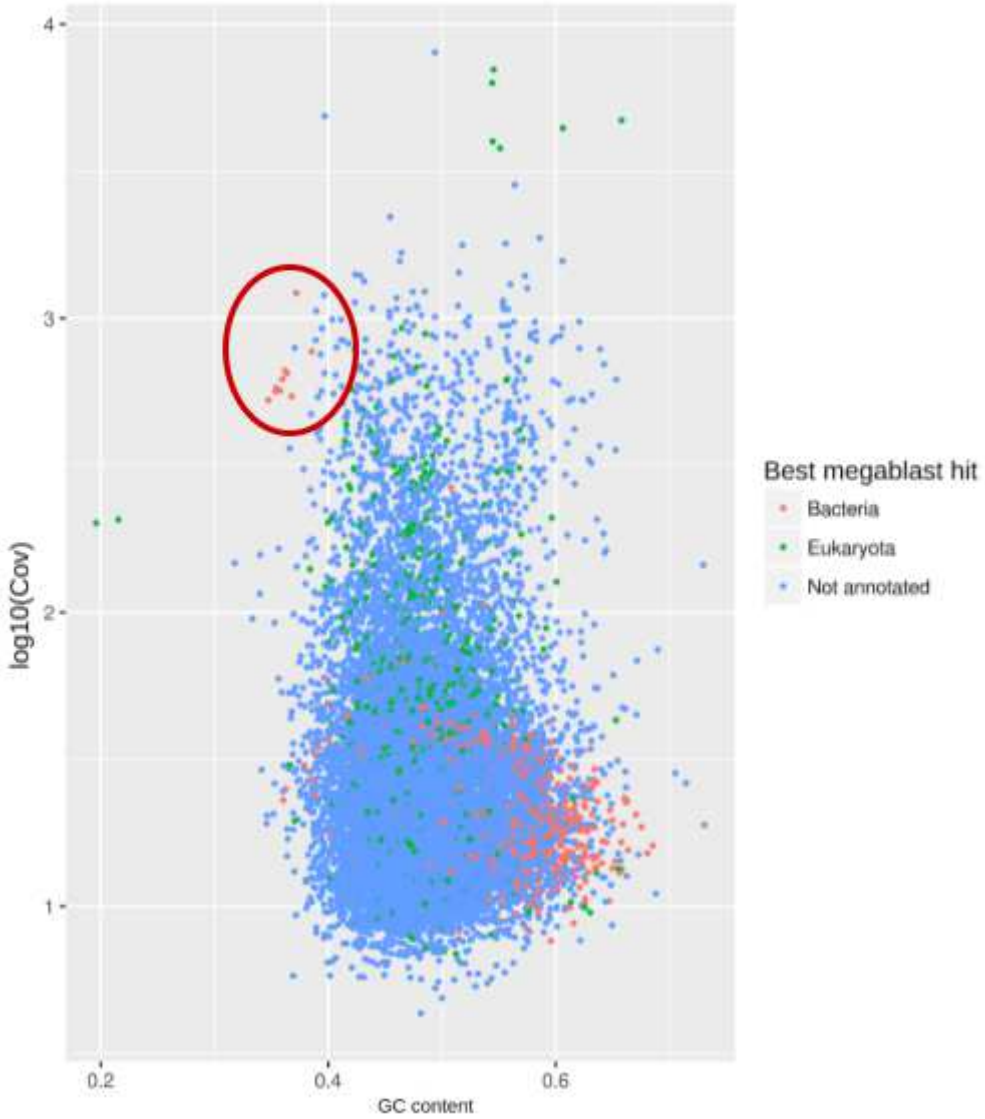


Figure 1: Plot from the blobology pipeline, representing contigs from the preliminary assembly and their GC content and coverage.

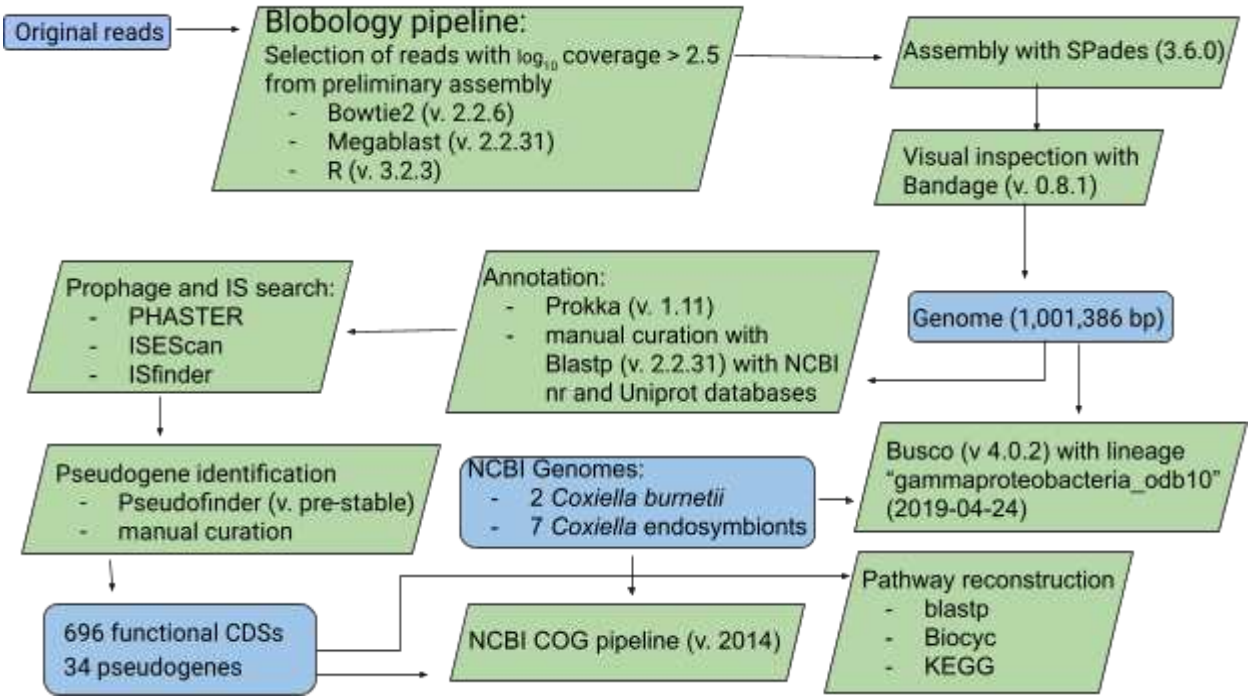


Figure 2A: *Workflow of bioinformatic analyses to obtain CE A. nuttalli genome and comparative analyses.*

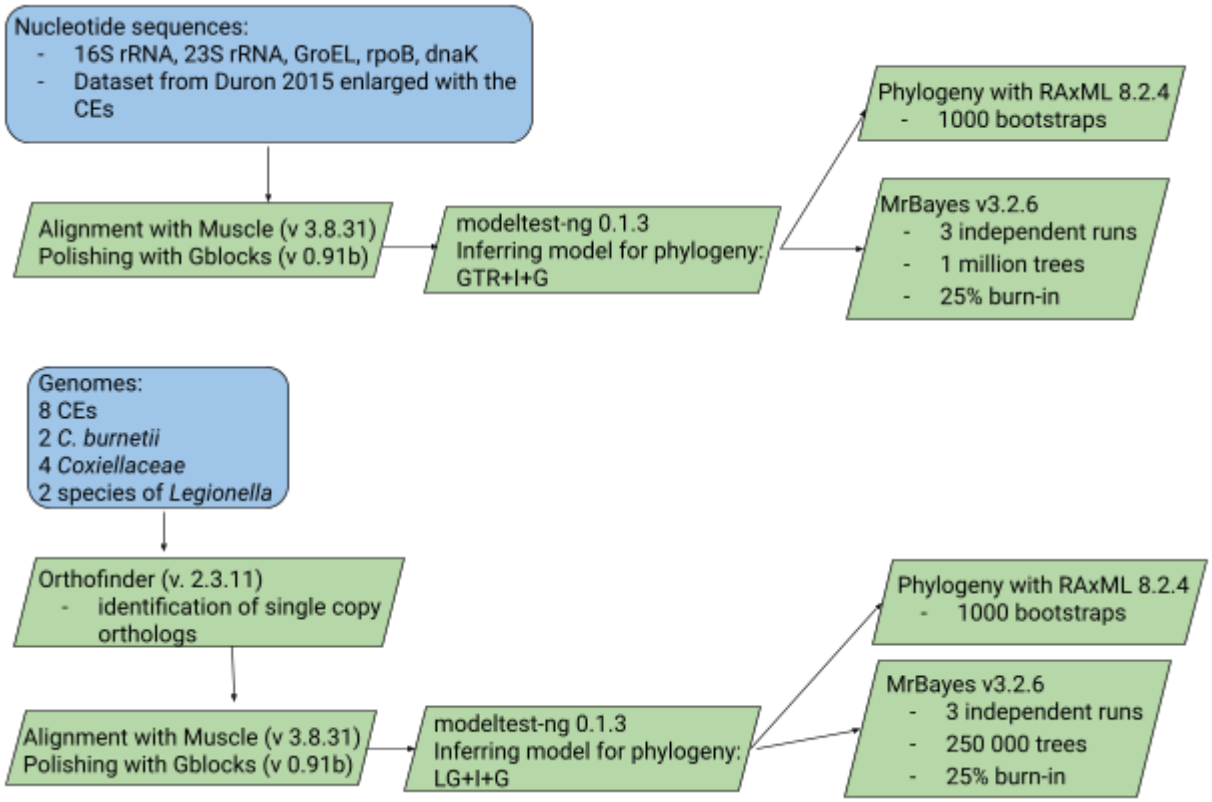


Figure 2B: Workflow to obtain the two different phylogenetic trees from the genetic and genomic data.

2.5 Results and discussion

The genome of the *Coxiella* endosymbiont of *A. nuttalli* was sequenced using Illumina technology and assembled using the blobology pipeline to separate the symbiont and host reads.
The final assembly has a genome length of 1,003,026 bp in a single circular contig, with a BUSCO completeness score of 79,2% (*Figure 3)* similar to other *Coxiella* endosymbionts.

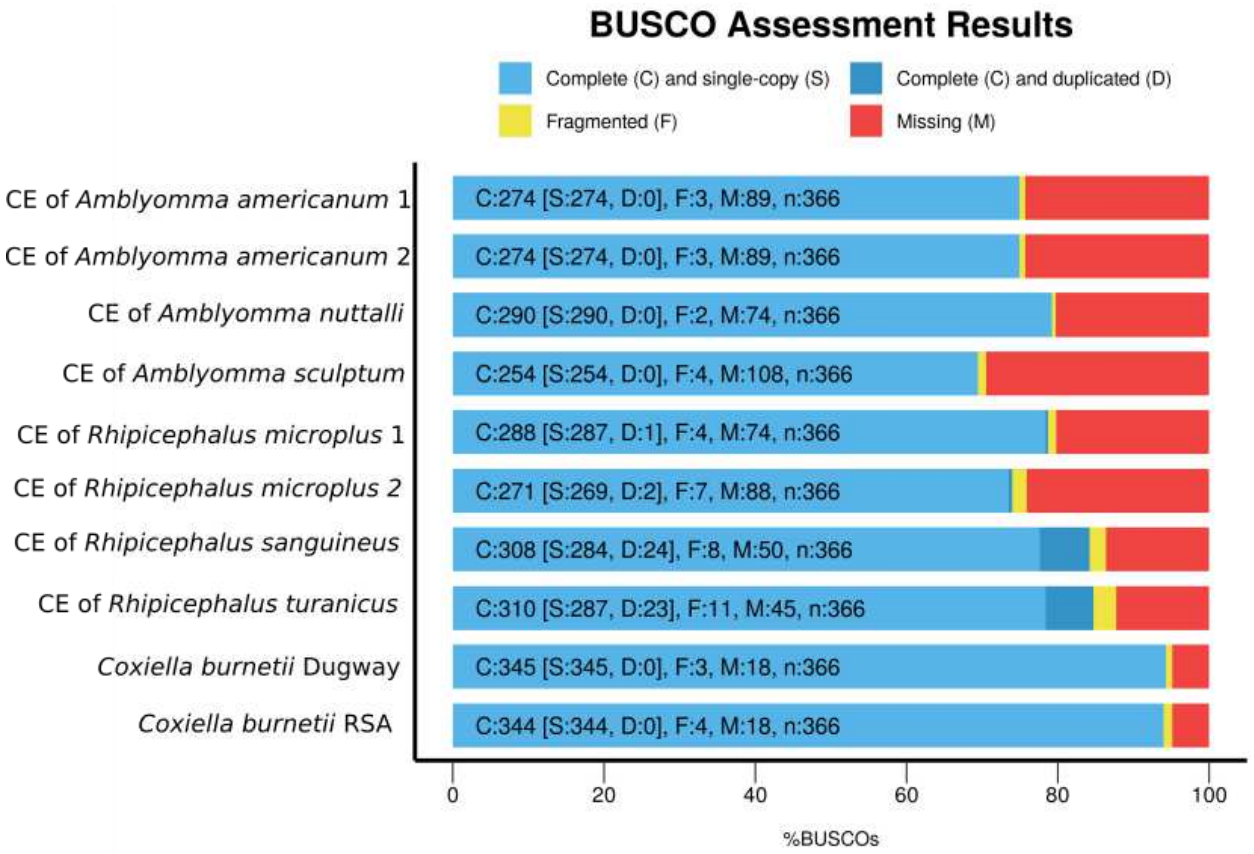


Figure 3: Busco completeness score. Coxiella *endosymbiont of* Amblyomma nuttalli *is comparable to* CE *of R. microplus 1*

A total of 764 genes were found, including 44 RNA genes (including 38 tRNAs and 3 rRNAs). Among these, we identified 702 functional CDSs and 18 pseudogenes, accounting for a total of 660,777 bp (65.8%) coding (including structural RNA genes). Neither prophages nor ISs were found.

MLST phylogeny (Figure 4) provided an overall consistent topology with most previous studies (Duron et al. 2015; Gottlieb et al. 2015), in particular for the major *Coxiella* clades and their relationships (clade A earliest divergent, clade B sister group of clades C + D), with moderate to high support (Figure 4). Consistently with what could be expected based on geographical origin, the CE of *A. nuttalli* lies in the clade B, composed by CEs of African ticks.

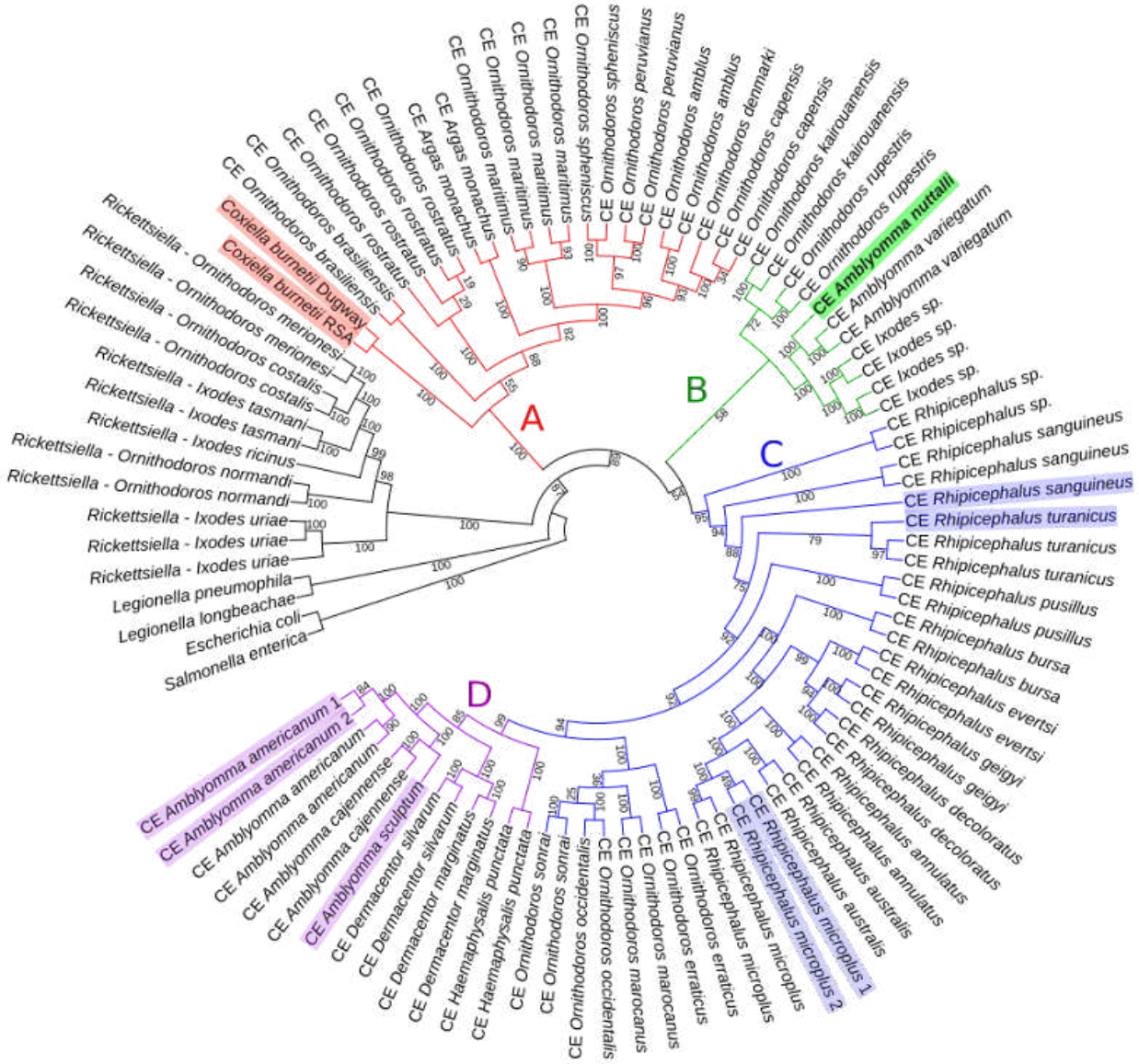


Figure 4: Maximum Likelihood MLST phylogenetic tree of the Coxiella *Endosymbionts and* C. burnetii *using as outgroup* Rickettsiella *and* Enterobacterales*. The same topology was obtained with Bayesian Inference. The four clades are indicated with the letters and colours, the highlighted names are the* Coxiella *bacteria with an available genome.*

The tree, with the same topology in ML and BI (Figure 5), confirmed the relationships observed in the MLST tree. Interestingly, branch lengths are proportional to the degree of genome reduction within the *Coxiella* genus, consistently with previous analyses (Duron *et al*, 2015; Gottlieb *et al*, 2015). This would indicate a higher evolutionary rate in smaller genomes, as predicted for obligate symbionts by genome reduction models (McCutcheon, 2012).

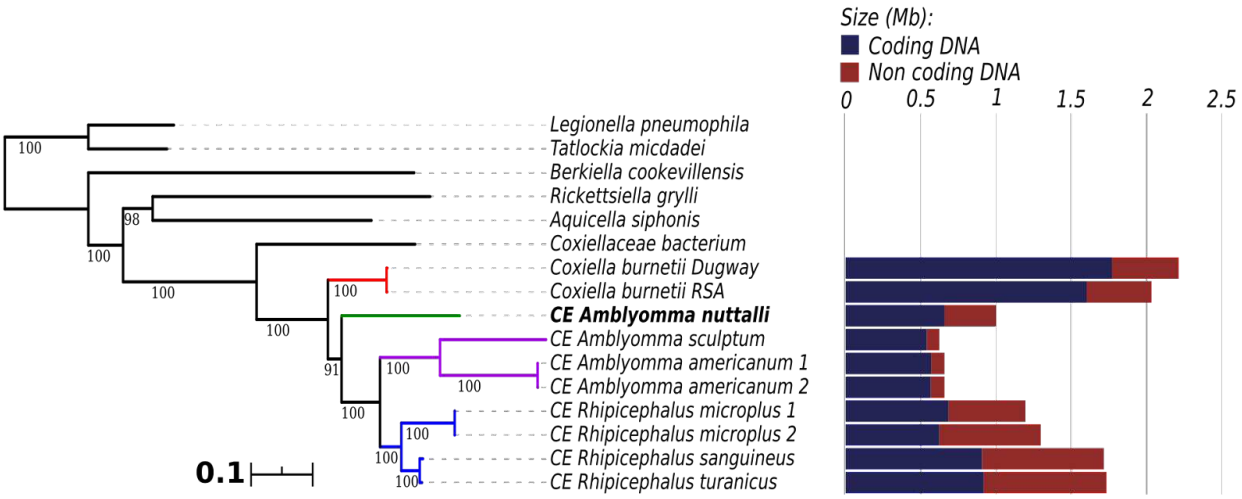


Figure 5: Phylogenomic tree of Coxiellaceae. *The length of the genomes are indicated in the barplot.*

2.5.1 Genome comparison

The CE of *A. nuttalli* has traits of a relatively long-term obligate symbiont: compared to free living bacteria it presents a much smaller genome, has no predicted mobile elements, presents a high number of pseudogenes, albeit conserving pathways potentially useful for the host, such as biotin biosynthesis.
Compared to the other *Coxiella* genomes, both genome length (1 Mb) and density (0.66% of the genome is coding) are at an intermediate level: *C. burnetii*, a pathogen capable of living in different environments and hosts (Angelakis and Raoult 2010), presents the largest genome (2 Mb) and, with a high coding density (79.9%), the highest amount of coding DNA (1.8 Mb). All the host-restricted CEs have smaller genomes, with similar sizes within each clade: 1.2–1.7 Mb for CEs of *Rhipicephalus* (clade C), 1.0 Mb for the novel CE of *A. nuttalli* (clade B), and 0.6–0.7 Mb for CEs of other *Amblyomma* species (clade D). However, the degree of genome reduction does not correlate with the phylogenetic branching pattern between clades, in particular the CEs with more reduced genomes (clades B, D) do not form a single monophyletic group (Figure 5). Accordingly, considering also the novel CE of *A. nuttalli*, a scenario with parallel independent genome reduction in genus *Coxiella* (at least for monophyletic CEs of B–D clades together) appears plausible.
Interestingly, most size variation resides in the noncoding genome (from 88 kb to 816 kb), whereas the length of the functional coding genome is overall less variable in CEs, ranging from 540 kb of CE of *Amblyomma sculptum* (Clade D), to 917 kb of CE of *Rhipicephalus turanicus* (Clade C). These features as well are consistent with a recent and still ongoing parallel genome reduction of CEs under similar constraints, possibly due to an equivalent role for the host. Accordingly, the CE of *A. nuttalli* has traits of a relatively long-term obligate symbiont, at an intermediate stage among CEs for its genome size and coding density (table 1), and having no predicted mobile elements.

2.5.1.1 COG analysis

In the COG analysis (Figure 4) the pattern of genome reduction is clearly visible. The pathogen *C. burnetii*, capable of living in different environments, has the highest number of functional COGs, the highest density of genes, and the largest genome (2 Mb). Instead, all CEs present lower numbers of COGs, proportionally to their genome reduction: among CEs, the symbionts of *R. turanicus* and *R. sanguineus* have a high number of COGs, while the CEs of clade D have the most reduced genomes, and the others, including the novel CE of *A. nuttalli*, have an intermediate situation.

Some categories are conserved even in highly reduced genomes, such as translation machinery (J category), coenzyme (H) and nucleotide (F) metabolism, energy production (C), protein modification and chaperones (O), lipid synthesis (I), cell cycle regulation (D). This is consistent with their expected major role for bacterial survival and/or host-support (coenzymes). On the other side, all CEs are more

pronouncedly depleted in accessory and regulative functions, including poorly characterised ones (R and S), signal transduction (T), secondary metabolite metabolism (Q), cell motility (N), secretion systems (U), extracellular and defence structures (W and V). Such functionalities are probably less important in strictly host-associated bacteria as the CEs. Notable is the case of type IV secretion system, probably ancestral in *Legionellales* (Hugoson *et al*, 2022), and an important virulence factor in *C. burnetii* (Luedtke *et al*, 2017), but absent in all CEs. Some functions display gradients of conservation along the genome size, for example, membrane structure biogenesis (M), which correlates with decrease in lipopolysaccharide complexity, whereas peptidoglycan synthesis is conserved.

Such a scenario is reflected at the level of single COGs (Figure 7), as *C. burnetii* (clade A) presents the highest number of unique COGs. CEs of all clades are a substantial subset of *C. burnetii*, which lacks only 60 of the 1,207 total COGs in the data set. Similar observations can be drawn among progressively more reduced CE clades, with the CE of *A. nuttalli* at an intermediate level between clade B and D CEs.

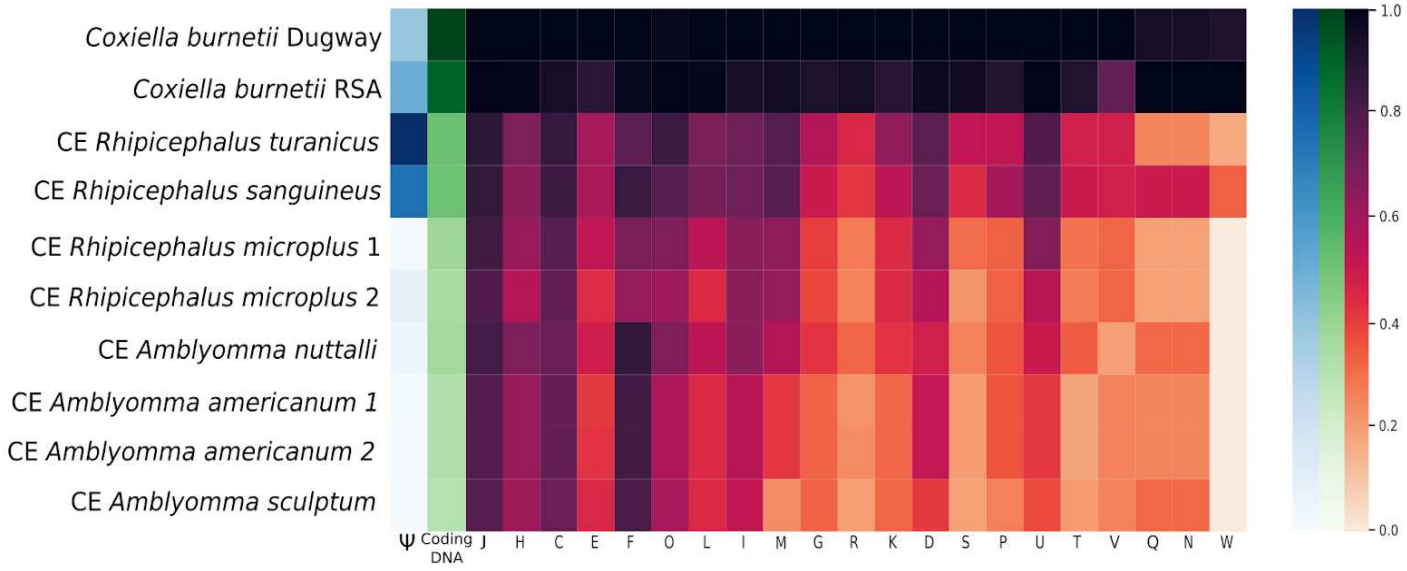


Figure: 6: Heatmap showing variations in number of pseudogenes (blue, indicated with the Ψ), the size of coding DNA (green) and COG (Clusters of Orthologous Groups) repertoire for each functional category (orange-purple) in the Coxiella *genus.*

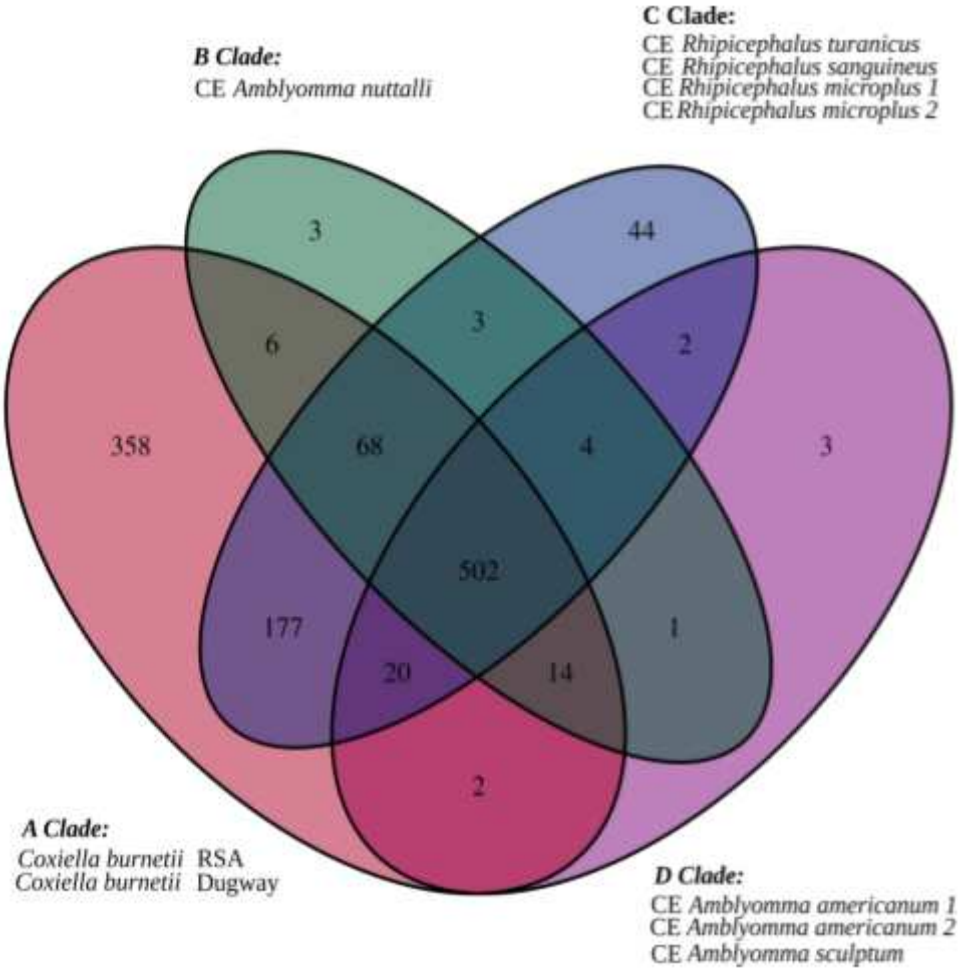


Figure 7: Venn diagram representing COG distribution on Coxiella *clades. For each clade, COGs identified in at least one of the listed genomes are counted.*

2.5.1.2 Metabolic capabilities

The results from COG functional category analyses were confirmed at the single gene level. For example, all CEs lack the biosynthetic pathways for most of the amino acids - namely phenylalanine, tyrosine, aspartate, cysteine, serine, threonine, isoleucine, leucine, alanine, methionine, arginine, asparagine, glutamate, histidine and proline.
However, as expected, all CEs have maintained the pathways for the biosynthesis of vitamins and co-factors, such as biotin, and vitamins B2 (riboflavin), B5 (pantothenate), B6 (pyridoxine), B9.
For biotin (and lipid) synthesis, missing *fabI* functionality is possibly replaced by *FabV* (Massengo-Tiassé & Cronan, 2008).
Similar metabolic capabilities are present in other tick endosymbionts, such as *Francisella* endosymbionts, which are devoid of amino acid synthesis capabilities but have conserved the biosynthesis of biotin (Gerhart *et al*, 2018).
The CEs of *Amblyomma* (clade D), despite having the smallest genomes of all CEs, display an almost complete biosynthetic pathway for thiamine (B1) and NAD (B3). This is also true for the CE of *A. nuttalli*, even if not phylogenetically close to the other *Amblyomma* CE. Instead, symbionts of *Rhipicephalus* retain only partial pathways. Such differences could be explained by the presence of not yet identified transporters or non-canonical enzymes, or, else, could be the effect of different metabolic requirements of the tick hosts, with *Amblyomma* species requiring a full pathway while *Rhipicephalus* species being permissive for the loss of some genes.

CEs are depleted in DNA repair abilities (COG category L), as expected in the symbiotic condition (McCutcheon, 2012), with lineage-specific features.
The loss of these genes is a driver of high mutation rate, pseudogenisation and genome reduction. Noteworthy seems to be the loss of the MutSL pathway, which is absent in all the CEs of *Amblyomma*, in both B and D clades, that are the ones with smaller genomes. The pathway is also possibly incomplete in the CE of *R. microplus* 2, which has a truncated copy of the *mutS* - possibly with partial or none functionality. This CE is characterised by a high genome reduction compared to the other CE of *R. microplus*. Considering the phylogenetic trees we can identify parsimoniously at least three independent losses of the MutSL pathway, complete in the clades B and D, and still ongoing within clade C.
The loss of the *RecFOR* pathway and *RecA*, involved in homologous recombination (Kuzminov, 1999) , happened in the CEs of *R. microplus*, and this loss can explain their medium sized genome, smaller than other *Rhipicephalus* CEs.
Another repair pathway that could be involved in genome reduction is addAB: it is a functional equivalent of RecBCD involved in the recombinational repair system. It is present in many Gram-positive bacteria and also in *C. burnetii* (Mertens *et al*, 2008). Both genes are present in CEs, excluding all CEs of clade D, namely other CEs of *Amblyomma* from the new world, that are characterised by the smallest genomes. Thus, the larger genome CE *Amblyomma nuttalli*, compared to other *Amblyomma* CEs, could be explained by the conservation of addAB, and/or by a more recent loss of MutSL.

Concerning the presence of pathogenic capabilities in the order, it has been hypothesised that the last *Legionellales* common ancestor existed 1.9 Ga, and possessed already some components of the infection machinery, in particular a type IV secretion system, two effectors, and the ability to resist digestion (Hugoson et al. 2022). Thus the whole lineage would be already equipped to interact and use eukaryotic cells, leading to common patterns of evolution to become pathogens (such as *Coxiella burnetii* and *Legionella pneumophila)*. In some cases the machinery could be exapted to a symbiotic relationship, as the case of the *Legionellacae* louse symbiont *Polyplax serrata* (Říhová, 2017). This could be the situation of the CEs, where multiple adaptations could have taken place.

2.6 Conclusions

From the comparative genomic and phylogenetic analyses it can be concluded that the CEs are currently undergoing parallel patterns of genome reduction, with a progressive loss of DNA repair genes, happening independently in the different clades. The constraints seem similar (i.e. keeping the genes for

vitamin provision useful for their host), even if it is possible that some hosts can be more permissive to certain losses (ex. *Amblyomma* genus for B1 biosynthesis).

Concerning the evolution of *C. burnetii* pathogenicity we do not have a "smoking gun" in this situation: notwithstanding the high variability of the CEs clades, according to the COG analysis all CEs have only a subset of *C. burnetii* COGs, and more specifically, slightly lesser metabolic capabilities. As we are observing a convergent genomic reduction, it is possible that the *Coxiella* ancestor had pathogenic capabilities retained only in *C. burnetii,* as it is under different selective pressures. Considering that we see some pathogenic capabilities in CEs of clade C, in *C. cheraxi*, and more generally in the order, it is plausible that the ancestor had pathogenic capability. The multiple adaptations to symbiosis can be explained also as the whole *Legionellales* order is considered composed by "professional symbionts" they have the potential to exapt the various adaptations to survive as a pathogen for creating a symbiotic relationship.

After the publication of this work another research (Brenner *et al*, 2021) analysed the genome of a CE of the soft tick *Ornithodoros amblus*, from the clade A, supporting multiple origins of the CEs from pathogens, and the presence of pathogenic capabilities in the last common ancestor.
Further insight on the evolution of the relationship between the ticks and CEs, and the development of pathogenic capabilities in *C. burnetii* could come from the analysis of other *Coxiella* with hosts different from ticks such as *C. cheraxi* or other undiscovered lineages.

III A multiple partner symbiosis in *Hyalomma marginatum*

Published as: Marie Buysse, Anna Maria Floriano, Yuval Gottlieb, **Tiago Nardi**, Francesco Comandatore, Emanuela Olivieri, Alessia Giannetto, Ana M Palomar, Benjamin L Makepeace, Chiara Bazzocchi, Alessandra Cafiso, Davide Sassera, Olivier Duron (2021) A dual endosymbiosis supports nutritional adaptation to hematophagy in the invasive tick *Hyalomma marginatum* eLife 10:e72747

3.1 Introduction

In the previous chapter I investigated a symbiotic relation between a single bacterium and an eukaryotic host, but it is also possible to have more than one symbiotic bacteria. In these situations the interplay between different symbionts in a single host can have different results: there can be competition between the symbionts, in particular with the "newer" bacterium, not yet affected by the Muller's ratchet, outcompeting the old symbiont, but the cooperative coexistence of different bacteria is also possible.
As an example of symbiont replacement in ticks, we previously described as an example in tick *Amblyomma maculatum*, where the *Francisella* endosymbiont (FE), a more recent symbiont with a less pseudogenised genome, is replacing the ancestral *Coxiella* endosymbiont (Gerhart *et al*, 2016).
FE is present also in another tick species, *Hyalomma marginatum*, but the structure of the symbiont community present in this species was, before this work, unclear.

Hyalomma marginatum is a tick that feeds on humans, various other mammals, and birds. It is usually found in the Mediterranean climatic zone, - i.e. Albania, Algeria, Croatia, Greece, Italy, Morocco, Spain, Tunisia, Turkey (EFSA Panel on Animal Health and Welfare (AHAW), 2010), but has recently expanded its range to the north, being isolated in Western Europe (Vial, 2016). *H. marginatum* is a vector for various diseases, as it can transmit Crimean-Congo hemorrhagic fever virus, protozoan parasites, and intracellular bacterial pathogens such as *Rickettsia aeschlimannii* (Vial, 2016).

Previous studies on the symbionts' presence in *H. marginatum* focused on searching a single species each time, for example detecting only FE (Azagi *et al*, 2017). The only study investigating both symbionts shows co-infection, but only in a single examined individual (Di Lecce *et al*, 2018). Thus, to understand the symbiont population in *H. marginatum*, different tick individuals were collected from different geographical locations, and tested for the presence of symbionts, seeing the prevalence of the different bacteria, if there is a pattern of exclusion with competition between them, or other scenarios, including the co-symbiosis.

3.2 Aims

In this specific work my objective was to investigate the symbiosis of *Midichloria* and *Francisella* endosymbionts in *Hyalomma marginatum* and to identify the nature of this tripartite association.

3.3 Methods

3.3.1 Tick collection and DNA extraction

Adult females of *H. marginatum* were collected from domestic animals from three countries (Italy, Spain, and Israel). For the sequencing, the whole body of one female from Italy was analysed , while ovaries from females collected in Spain and Israel were used for genome sequencing (and for qPCR in the case of

samples from Spain). For each sample, total DNA was extracted using either the DNeasy Blood and Tissue Kit (QIAGEN) or the NucleoSpin Plant II Kit (Macherey-Nagel).

3.3.2 qPCR assays

A molecular screening was performed by qPCR (iQ5 system, Bio-Rad) on Spanish samples to evaluate the bacterial load of *Francisella* and *Midichloria* symbionts in ovaries. For each specimen, three SYBR Green-based qPCR (Bio-Rad) assays were performed:

1. *cal* gene for the *H. marginatum*
2. *rpoB* gene for the FLE
3. *gyrB* gene for the *Midichloria*

The following thermal protocols were used:

- For the *cal gene*: primers concentration 150 nM; thermal profile: 95 °C for 3 min, and 40 cycles at 95 °C for 15 sec, 55 °C for 20 sec, 72 °C for 10 sec
- For *rpoB:* primers concentration 250 nM; thermal profile: 95 °C for 3 min, and 40 cycles at 95 °C for 15 sec, 58 °C for 30 sec;
- For *gyrB*: primers concentration 250 nM; thermal profile: 95 °C for 3 min, and 40 cycles at 95 °C for 15 sec, 58°C for 30 sec

3.3.3 Genome sequencing and assembly, and annotation

Metagenomes were sequenced on Illumina HiSeq 2500, after a library preparation performed with Nextera XT.
For the Italian and the Spanish samples, the raw paired-end reads quality was evaluated with FastQC.. Paired-end reads were trimmed via Trimmomatic (v0.40) (Bolger, 2014) and then assembled using SPAdes (v3.10.0) (Bankevich *et al*, 2012) following the Blobology pipeline (Kumar *et al*, 2013) (thus separating reads according to G+C content, coverage, and taxonomic annotation). Discarding of contaminating sequences and refining of the assemblies were performed manually through the analysis of the assemblies' graphs with Bandage (v0.8.1) (Wick *et al*, 2015). Bacterial reads were separated from the tick reads, allowing the assembling of different bacterial genomes ('filtered' assemblies).
For the Israeli sample, reads were trimmed using Cutadapt (v1.13) (Martin, 2011) and contaminants' reads were removed through a mapping against Coliphage phi-X174 (GenBank: NC_001422) and *Staphylococcus aureus* (NC_010079) genomes with BWA (v0.7.17-r1188) (option mem) and SAMtools (v1.10)(Li, 2009). Remaining reads were further assembled using SPAdes (v3.11.1) (Bankevich *et al*, 2012) with default parameters. These 'raw' assembled scaffolds were compared to different reference genomes separately (FE of *O. moubata* FLEOm, *M. mitochondrii*, and *R. aeschlimannii*; table 3) to discard unspecific scaffolds using RaGOO (v1.02) (Alonge, 2019), and, for each reference genome separately, the assignment of the remaining scaffolds ('filtered' assembly) was confirmed using the NCBI BLAST tool. The quality assessment and the completeness of the 'filtered' assemblies from all *H. marginatum* metagenomes were estimated using QUAST (v4.6.3) (Gurevich, 2013) and miComplete (v1.1.1) (Hugoson, 2020) with the --hmms Bact105 setting.

3.3.4 Genomic annotation

The newly obtained genomes were annotated using Prokka (v1.13.1) (Seemann, 2014) with default parameters. The new *Midichloria* genomes were annotated using a manually revised annotation of *M. mitochondrii* as a reference.
Genes involved in biosynthesis pathways (B vitamins and heme), *Francisella* pathogenicity island (FPI), and mismatch repair systems (*mutSL* and *uvrABCD*) were retrieved using BLASTn, BLASTp, and tBLASTx on both newly obtained and reference FE and *Midichloria* genomes.
To identify mobile elements, we used ISEScan (Xie & Tang, 2017) to predict insertion sequences and PhySpy (Akhter, 2012) for prophage detection. Pseudogenes prediction was performed using Pseudofinder (version downloaded in date April 1, 2020) (Syberg-Olsen *et al*, 2020) on each genome,

including newly obtained and reference genomes of FE and *Midichloria*. The FE and *Midichloria* data sets were filtered excluding pseudogenes from the subsequent analyses. Clusters of orthologous groups of proteins (COGs) were predicted using the NCBI pipeline (Galperin *et al*, 2015) and COGs presence was then plotted with an in-house R script. Then, two different data sets were created: (i) a FE data set including the three newly sequenced FLE-Hmar genomes and other published FE genomes; and (ii) a *Midichloria* data set including the three newly sequenced Midi-Hmar genomes and *M. mitochondrii*. For each data set, Single Copy Orthologs (SCOs) were identified using OrthoFinder (v2.3.12) (Emms & Kelly, 2019) , then SCO lists were retrieved on a R environment (v3.6.2) to build Venn diagrams using the 'VennDiagram' R package (Chen, 2011), and UpSetR packages to build Upset plots (Conway *et al*, 2017).

The FE and *Midichloria* whole-genome alignments were obtained separately using Mauve (Darling *et al*, 2004). Graphical representations of newly sequenced genomes were produced using CGView (v1.5) (Stothard, 2004).

To check for the presence of B-vitamin biosynthesis genes of other origin than FE, *Midichloria,* and *Rickettsia*, we used the following procedure: we predicted the genes present in the preliminary assemblies of the three tick metagenomes using prodigal (using the setting '-p meta'), and annotated them with GhostKOALA (Kanehisa, 2015). For biotin, we checked for the presence of the *bioA*, *bioD*, *bioC*, *bioH*, *bioF*, and *bioB* genes searching for 'K00833, K01935, K02169, K02170, K00652, and K01012' sequences in the annotations, setting a cutoff to exclude contigs shorter than 500 bp or with a kmer coverage below 4. The hits above cutoff were manually examined with BLAST for functional and taxonomic classification.

3.3.5 Phylogenomics

The FE, *Midichloria,* and *Rickettsia* data sets were created using the predicted genes from the genomes assembled in this study, filtered excluding the pseudogenes, and each one were added to the gene predicted from a taxonomically representative dataset of relative, and were then used for the phylogenetic analyses. For each symbiont data set, SCOs were identified using OrthoFinder (v2.3.12) (Emms & Kelly, 2019) and then aligned with MUSCLE (v3.8.31) (Edgar, 2004). Poorly aligned positions and divergent regions were removed by Gblocks (v0.91b) (Talavera & Castresana, 2007) prior to concatenation of individual alignments using an in-house script. Finally, the most suitable evolutionary model was determined using modeltest-ng (v0.1.5) (Darriba *et al*, 2019) according to the Akaike information criterion, and maximum likelihood (ML) trees were inferred using RAxML (v8.2.9) (Stamatakis, 2014) with 1000 bootstrap pseudo-replicates.

3.3.6 Biotin operon phylogeny and structure

Nucleic and protein sequences of the six genes of the biotin biosynthesis pathway were retrieved using BLASTn, BLASTp, and tBLASTx on all *Midichloria* genomes. Sequences from biotin genes from other symbionts (retrieved from GenBank) were included in the data set. All protein sequences of the six biotin genes were manually concatenated and aligned with MAFFT (v7.450)(Katoh, 2002) and ambiguous positions were removed using trimAl (v1.2rev59) (Capella-Gutierrez, 2009). The appropriate substitution model, and then ML-tree, were determined as described above. Meanwhile, 16S rRNA sequences from all organisms included in this data set were retrieved from GenBank and aligned with MEGA software. The Gblocks program with default parameters was used to remove uninformative sites in the alignments. Phylogenetic analyses were performed with MEGA software, by determination of the appropriate substitution model using Akaike information criterion, ML-tree inference. Clade robustness was assessed by bootstrap analysis using 1000 pseudo-replicates.

Contigs carrying the streamlined biotin operon were retrieved from newly obtained and reference genomes of different intracellular bacteria. Contigs were annotated using Prokka (v1.13.1) (Seemann, 2014) with default parameters. Prokka GenBank outputs were imported in R (v3.6.2) and analysed using the 'genoplotR' R package (Guy, 2010). Sequences of *ribH* and *folD* genes from bacteria carrying the streamlined biotin operon and from other selected bacteria were secondarily separately analysed.

Phylogenetic analyses on these two genes were performed as described above for biotin genes.

3.3.7 Bibliographic survey and phylogeny of *Hyalomma*

To assess the associations between *Hyalomma* species with FE and *Midichloria*, we performed an extensive bibliographic survey of published studies. For *Hyalomma* species for which presence/absence of FE and *Midichloria* is documented, we retrieved cytochrome c oxidase I (*cox1*) gene sequences from GenBank and further performed a ML phylogenetic analysis with MEGA software as described above.

3.3.8 Statistical analyses

Statistical analyses were carried out using the R statistical package (v3.6.2) and the 'lme4' R package (Bates, 2015). The FE load was analysed using linear mixed-effects models on a data set containing loads measured in ovaries from adult engorged ticks. The normality of the data distribution was tested through a Shapiro-Wilk test with a 95% confidence interval. Outliers were removed from data sets. To build models, *Midichloria* load was fitted as a fixed explanatory variable. The best-fitted model was determined with the following procedure: the maximal model, including all higher-order interactions, was simplified by sequentially eliminating interactions and non-significant terms to establish a minimal model. A likelihood ratio test (using a chi-square distribution and a p-value cutoff of 0.05) was performed to establish the difference between sequential models. The comparison between the densities of FE and *Midichloria* of *H. marginatum* was tested by a paired t-test.

3.4 Results

A total of 46 *H. marginatum* samples were screened for the presence of symbionts. Among those, 44 were positive for *Midichloria*, 41 for FEs, and 34 for *Rickettsia*.
Nine samples resulted infected by *Midichloria* and FLE, but not *Rickettsia*; two samples by *Midichloria* and *Rickettsia*, but not FLE; one by FE and *Rickettsia*, but not *Midichloria*; and 31 were infected by *Midichloria*, FE and *Rickettsia*. So most ticks had a dual or triple infection, with no signs of exclusion patterns between *Midichloria* and FEs.
The nine samples of dissected ovaries infected by both *Midichloria* and FE were quantified by qPCR assays for both symbionts. The results (FIGURE 8A) indicated that FE and *Midichloria* target genes are abundant in the tick ovaries, with higher quantities for FE (686.09 ± 206.89 FE rpoB gene copies per tick cal gene copies) than *Midichloria* (88.72 ± 28.41 *Midichloria gyrB* gene copies per tick cal gene copies).
The FE and *Midichloria* densities were positively correlated in individual ticks ($\chi^2 = 23.937$, $p = 0.003$). This positive association indicates that the increase of one of them does not impair the growth of the other (FIGURE 8B), excluding a competitive interaction. This association could indicate a positive effect of one symbiont on the other, however a simple explanation could also be that the advancement of the ovary development leads to the growth of the population of both symbionts.

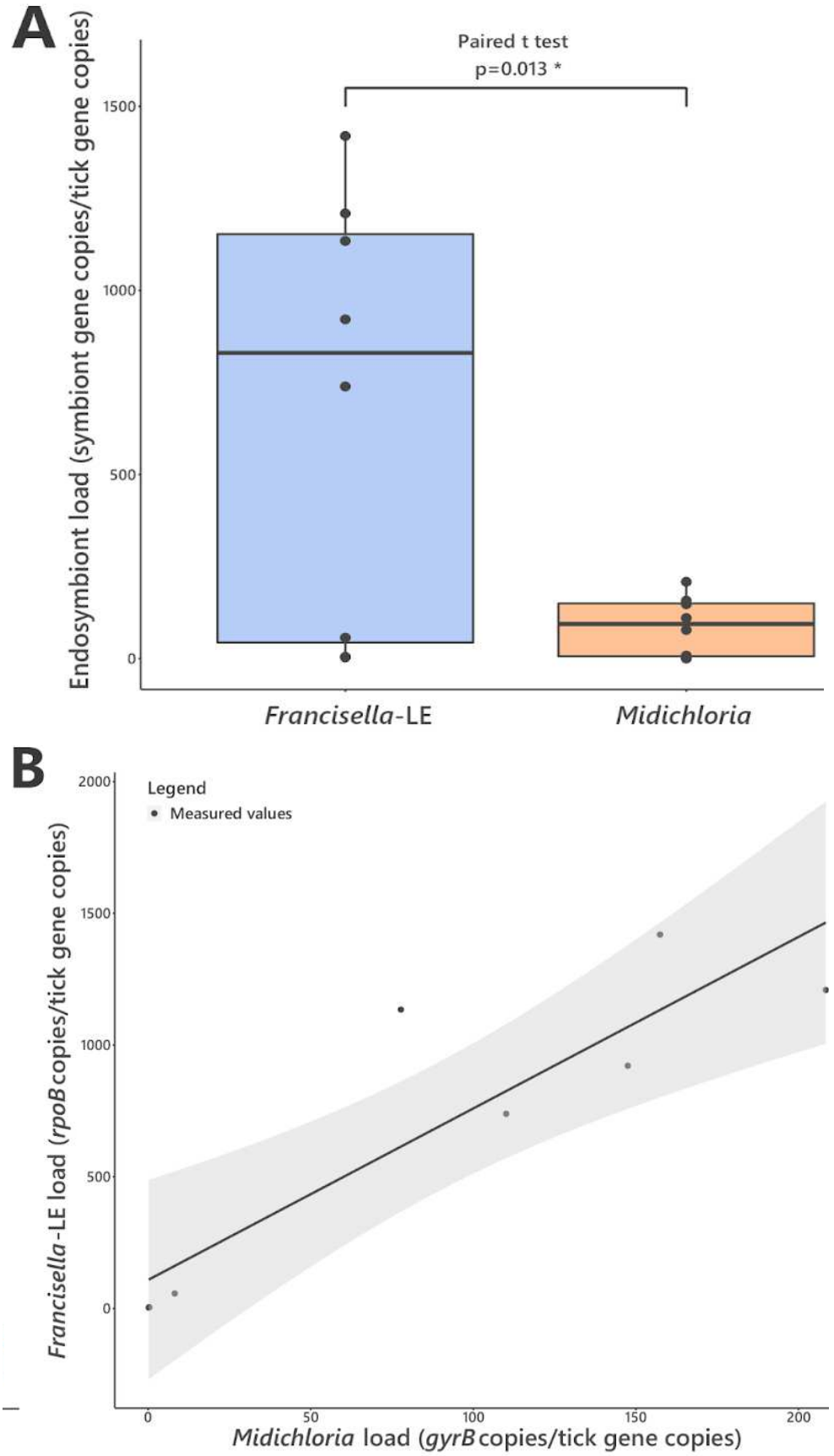


FIGURE 8: 8A quantities of the bacteria detected by qPCR. 8B covariance of the load of the two symbionts.

3.4.1 Genomics

We sequenced three metagenomes from three single *H. marginatum*, collected in Italy, Spain and Israel, respectively named Hmar-IT, Hmar-ES, Hmar-IS.
We obtained 68,515,944 paired-end reads from Hmar-ITA, 57,592,510 paired-end reads from the Hmar-ESP, and 10,807,740 paired-end reads from Hmar-IS.
With the blobology pipeline (Kumar *et al*, 2013) we were able to extract a complete genome of *Francisella* (FLE-Hmar-IT, FLE-Hmar-ES, FLE-Hmar-IS), with a >99% estimated completeness from each metagenome, and *Midichloria* (Midi-Hmar-IT, Midi-Hmar-ES, Midi-Hmar-IS) with >95% estimated completeness from each isolate. We detected the presence of the pathogen *Rickettsia aeschlimannii* in Hmar-IT and Hmar-IS, and we were able to obtain a complete genome (RAES-Hmar-IT), with a 94% estimated completeness, and a partial genome for a second one (RAES-Hmar-IL).
The data about the assembled genomes are in Table 2.

Genome	Isolate	Origin	Genome size (bp)	Contigs number	N50	L50	Completeness (%)	GC content (%)	Genes	Pseudogenes	Coding density (%)
Francisella Endosymbiont of *H. marginatum*	FLE-Hmar-ES	Spain	1509603	13	254696	3	0,9905	31,1	1732	618	56
Francisella Endosymbiont of *H. marginatum*	FLE-Hmar-IT	Italy	1509840	11	382079	2	0,9905	31,1	1725	336	60
Francisella Endosymbiont of *H. marginatum*	FLE-Hmar-IS	Israel	1536710	25	106485	5	0,9905	31,22	1762	821	53
Midichloria of *H. marginatum*	Midi-Hmar-ES	Spain	1161684	137	19210	20	0,981	35,02	1272	164	74
Midichloria of *H. marginatum*	Midi-Hmar-IT	Italy	1116804	272	9983	32	0,9524	35,25	1234	158	71
Midichloria of *H. marginatum*	Midi-Hmar-IS	Israel	1144281	123	12627	30	0,981	35,03	1243	198	70
Rickettsia aeschlimannii of *H. marginatum*	RAES-Hmar-IT	Italy	1344438	71	67640	4	0,9429	32,28	1592	446	70

Table 2: Genomics data of the genomes obtained in this study

Genome	Isolate	Origin	Accession	Reference	Genome size (bp)	Contigs number	GC content (%)
Francisella Endosymbiont of *Ornithodoros moubata*	FLE-Om	Czechia(Lab colony)	LVCE00000000	(Gerhart *et al*, 2018)	1564190	8	31,7
Midichloria mitochondrii	IricVA	Italy	CP002130.1	Sassera et al., 2011	1183732	1	36,55
Rickettsia aeschlimannii	RAES	France	GCA_001051325.1	-	1312196	16	32,1

Table 3: Reference genomes used in this study.

3.4.1.1 Francisella Endosymbionts

The three FE genomes are similar in size, (1.51–1.536 Mb), protein coding genes (937–961 genes), degree of pseudogenisation (789–798 pseudogenes), GC content (31.11–31.23%), and absence of prophages.

For the phylogenomic analysis of the FEs we obtained 436 SCOs, which we used to generate the concatenated alignment. In the resulting tree (Figure 9) the novel FEs cluster all together, in a close relationship with other tick FEs, in particular the symbionts of *Hyalomma asiaticum,* and are more distant from FEs of other genera of ticks, such as *Amblyomma*, *Argas* or *Ornithodoros*.

The FE clade is embedded in the *Francisella* genus and shares a common origin with pathogenic species, including the tularemia agent, *F. tularensis*, and an opportunistic human pathogen, *F. opportunistica*.

The FLE-Hmar genomes are similar in genome structure, with an almost preserved synteny, and overall content, apart from a 127,168 bp inversion in FLE-Hmar-ES and a 4165 bp inversion in FLE-Hmar-IL (Figure 10). Compared to other FE genomes (i.e., FLEOm and FLEAm), they are slightly reduced in size and exhibit substantial rearrangements (Figure 2B).
Based on gene orthologs, the core-genome of all FEs shares 743 genes, while the core genome of the FE of *H. marginatum* contains 796 genes because the FLE-Hmar genomes share 53 genes that are not present in other FE genomes. Interestingly, pseudogenes are highly abundant in the three genomes of FEs of *H. marginatum,* compared with other FE genomes, and most functional categories are impacted by this advanced pseudogenisation process. Most of the genes associated with virulence in pathogenic *Francisella* species, including the *Francisella* pathogenicity island (FPI) and type VI Secretion System, are pseudogenised or completely missing in FLE-Hmar genomes, consistent with previous observations in other FEs. Furthermore, FLE-Hmar genomes contain fewer genes involved in replication, recombination, and repair than any other FEs analysed (Figure 11). The nucleotide excision repair system (*uvrABCD*) in FLE-Hmar genomes is conserved; however, the DNA mismatch repair system (*mutSL*) is fully pseudogenised. In other FEs, these systems are conserved, with the exception of the *mutSL* system in the FE of *Ornithodoros moubata* and *H. asiaticum* genomes (Figure 16).

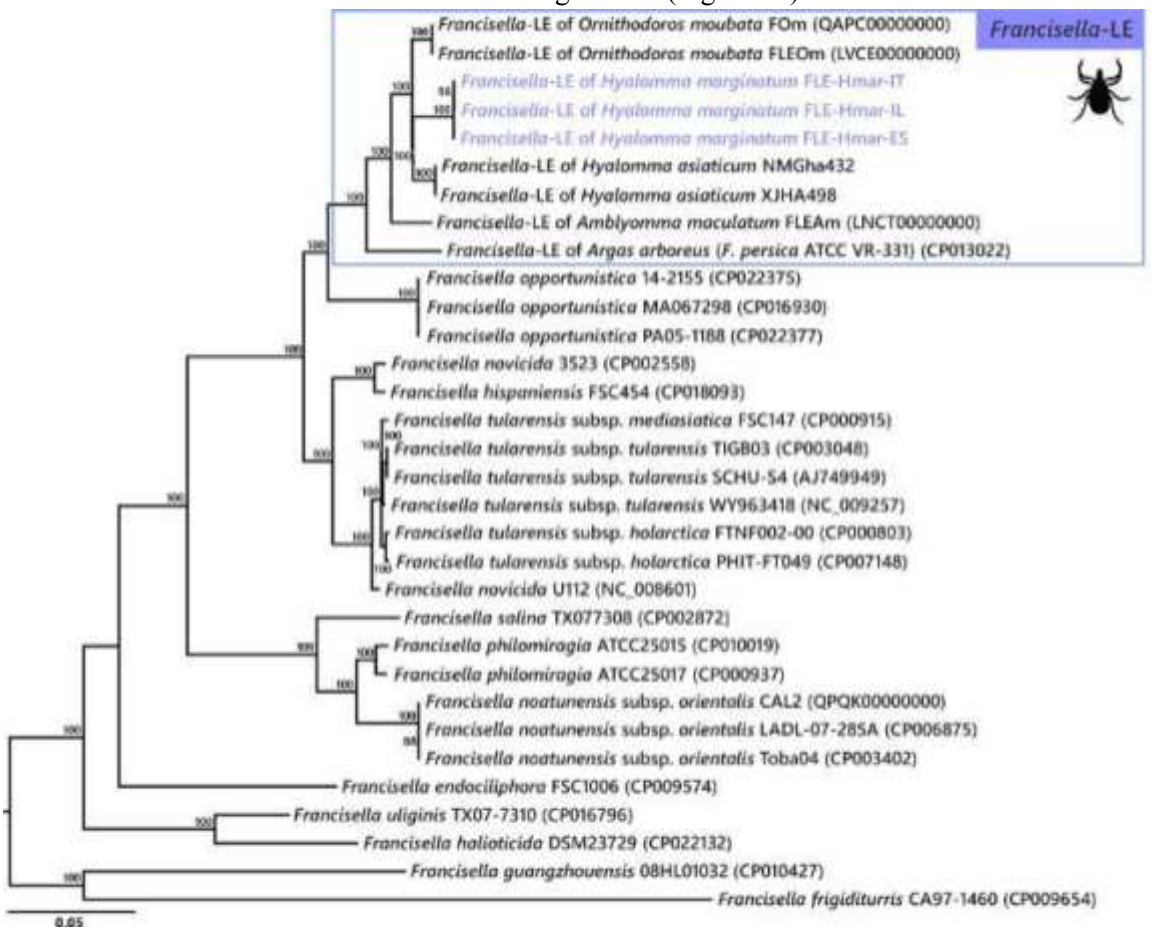


Figure 9: Phylogenomic tree of the Francisella *genus, including FEs. The three coloured leaves are the FEs of this study.*

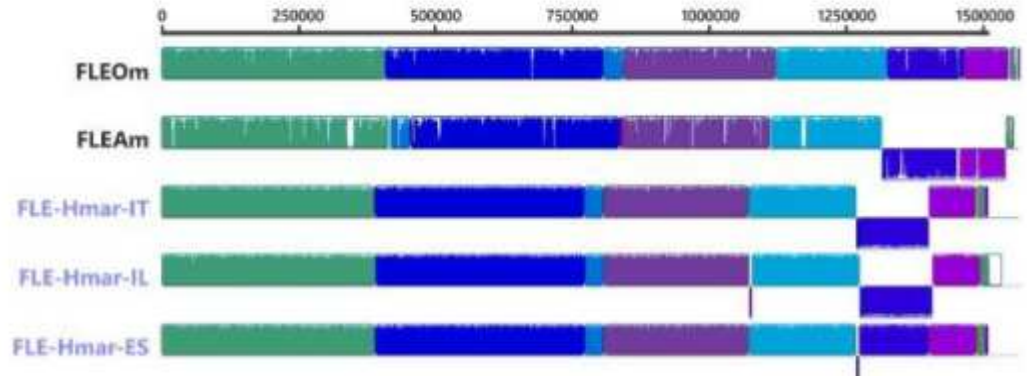


Figure 10: Genome alignments of the FEs. The three coloured samples are the one of this study.

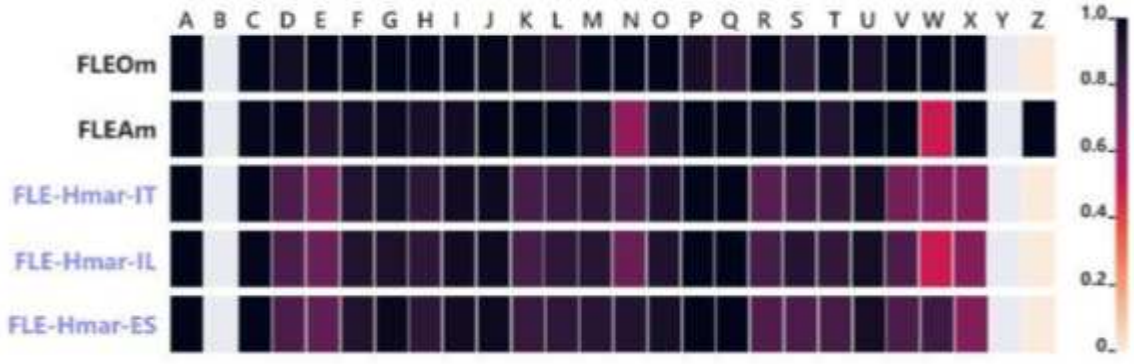


Figure 11: Heatmap with COGs identified in each genome. The presence is represented as the percentage of the highest number present in the sample for each column, with black being the highest value.

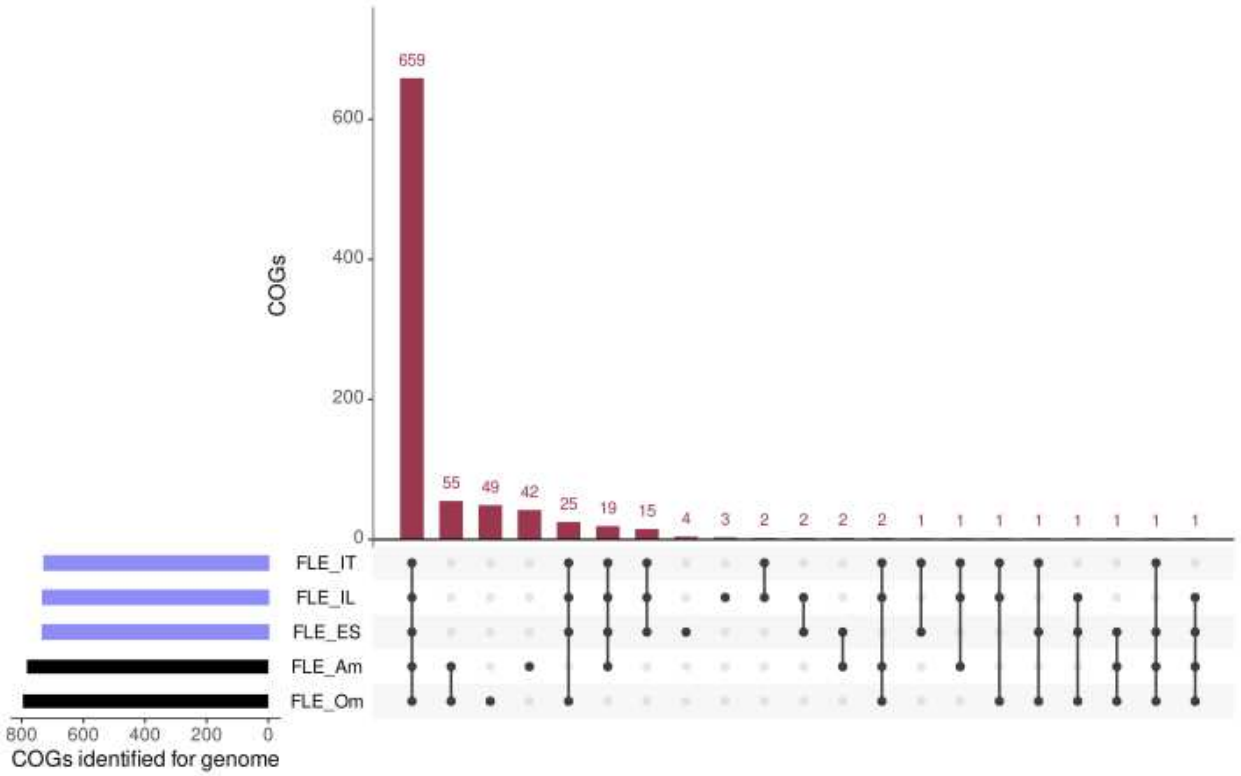


Figure 12: Shared COGs in the Francisella *endosymbiont genomes. In the UpSet the bars represent the COGs in common between each set, the dot matrix represents the sets of intersections considered for each bar.*

This loss could explain the slightly smaller genomes of FEs of this study, with a higher number of pseudogenes.

3.4.1.2 *Midichloria*

The three *Midichloria* genomes Midi-Hmar-ES, Midi-Hmar-IT, and Midi-Hmar-IL, were assembled in 179, 359, and 123 contigs, respectively. The genomes size are between 1.13–1.16 Mb, with GC content of 35.03– 35.23%, 1032–1071 protein coding genes, 190–192 pseudogenes, no prophages, 10 insertion sequences. These characteristics make them consistent between them and to the reference from *I. ricinus*. Their average coding density (76%) is higher than the FEs of *H. marginatum* (54%) and to the other FEs (57%-69%).

In total 242 SCOs are conserved between *Midichloria* and the other *Rickettsiales* sequences used in the dataset, and in the consequent phylogenetic tree (Figure 13) Midi-Hmar cluster in a highly supported monophyletic group, sister to the reference *Midichloria* strain IricVA (IricVA hereafter).

The comparison of genome structure was not possible due to the fragmentation of the genomes. Based on gene orthologs, the pan-genome of *Midichloria* shared 745 genes, and Midi-Hmar genomes shared 181 genes that are not present in IricVA, showing a strong similarity between the Midi-Hmar genomes. Pseudogenisation only impacted a few functional categories and genes of the mobilome, across Midi-Hmar genomes.

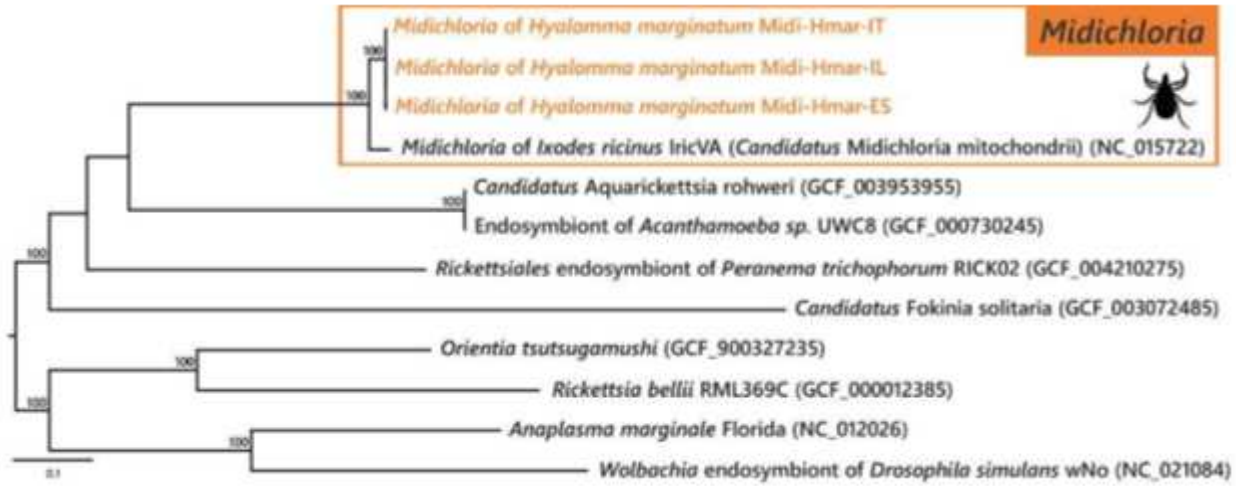


Figure 13: Phylogenomic tree of Midichloria*, including as outgroup other* Rickettsiales*. The three coloured leaves are the* Midichloria *of this study.*

Figure 14: Heatmap with COGs identified in each genomes. The presence is represented as the percentage of the highest number present in the sample for each column, with black being the highest value.

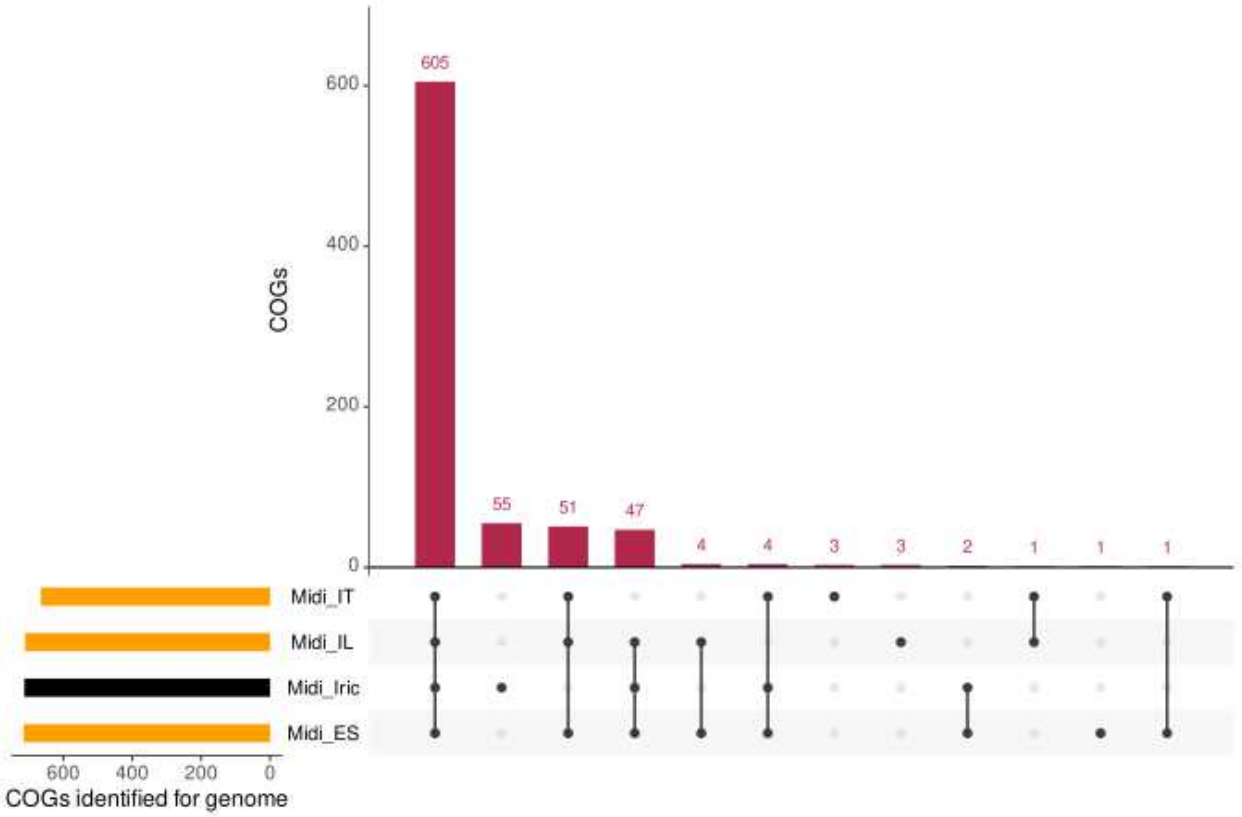


Figure 15: Single copy orthologs in the Midichloria *genomes. In the UpSet the bars represent the COGs in common between each set, the dot matrix represents the sets of intersections considered for each bar.*

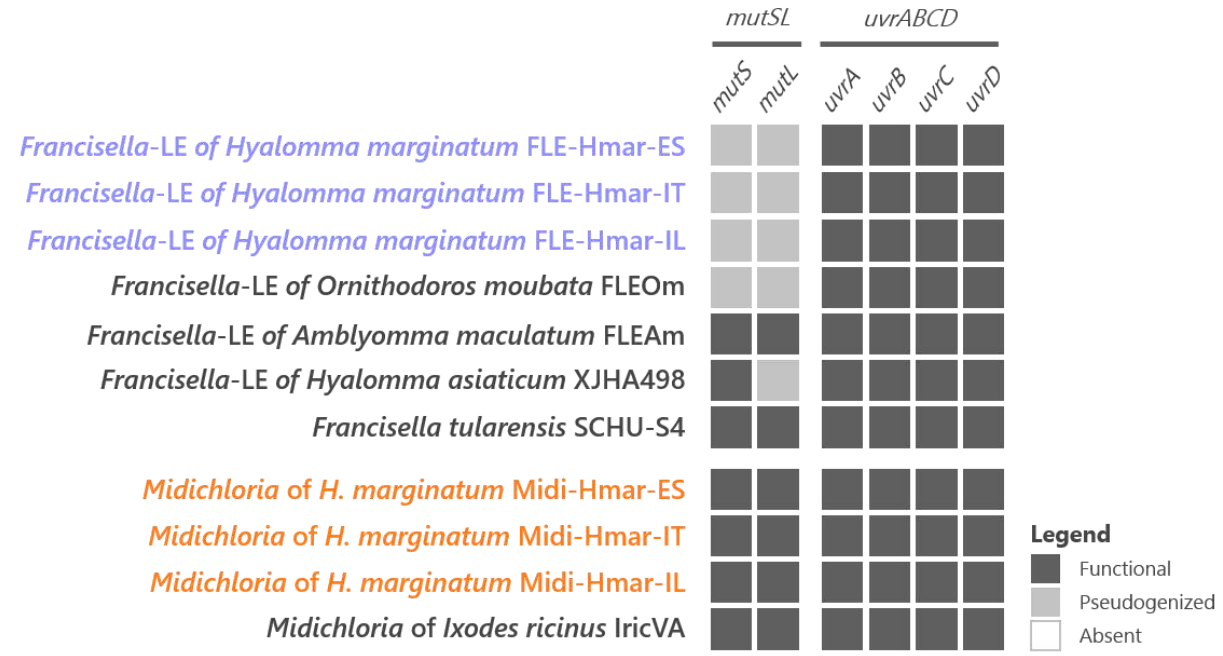


FIGURE 16: Repair systems in the genomes of the studied.

3.4.1.3 Rickettsia aeschlimannii

The recovered genome RAES-Hmar-IT of *Rickettsia aeschlimannii* is 1.35 Mb long, with 1122 genes, 472 pseudogenes and a GC content of 32%, and for RAES-Hmar-IL we were able to retrieve only 0.215 Mb.

The only sample from which we recovered the full genome of *R. aeschlimannii* was a full tick, while the

other two were ovaries. This could indicate that the main population of this *Rickettsia* is located in other organs, with only low density - if any - in the ovaries. The putative absence from the ovaries, together with the low abundance, and its phylogenetic positioning in the spotted fever *Rickettsia* group, suggest that *R. aeschlimannii* could be a tick-borne pathogen, rather than a nutritional symbiont.

3.4.2 Biosynthetic pathways

While all previously sequenced FE genomes contain complete pathways for biotin, riboflavin, and folic acid, the FLE-Hmar genomes only retain part of them. The pathways for riboflavin and folate are fully conserved, but not that of biotin: out of 6 genes, the same 4 contain frameshifts in all FLE-Hmar, so are deemed pseudogenes, and one, bioA, is fully absent.

While *bioA*, *bioD*, *bioC,* and *bioF* contain multiple indels or mutations (introducing premature stop codon, or missing start codon), *bioB* shows a low level of pseudogenisation with a unique indel Specifically, a stop codon results from a reading frame shift due to the insertion of one adenine base on a poly(A) zone. Since infidelity at poly(A) tracks can rescue pseudogenes in endosymbionts (Tamas, 2008), the *bioB* gene could be thus functional, but leaving the pathway incomplete anyway.

Concerning the other B vitamin pathways, all FE genomes have no genes for the biosynthesis of B1, and only partial pathways for B3, B5, B6.

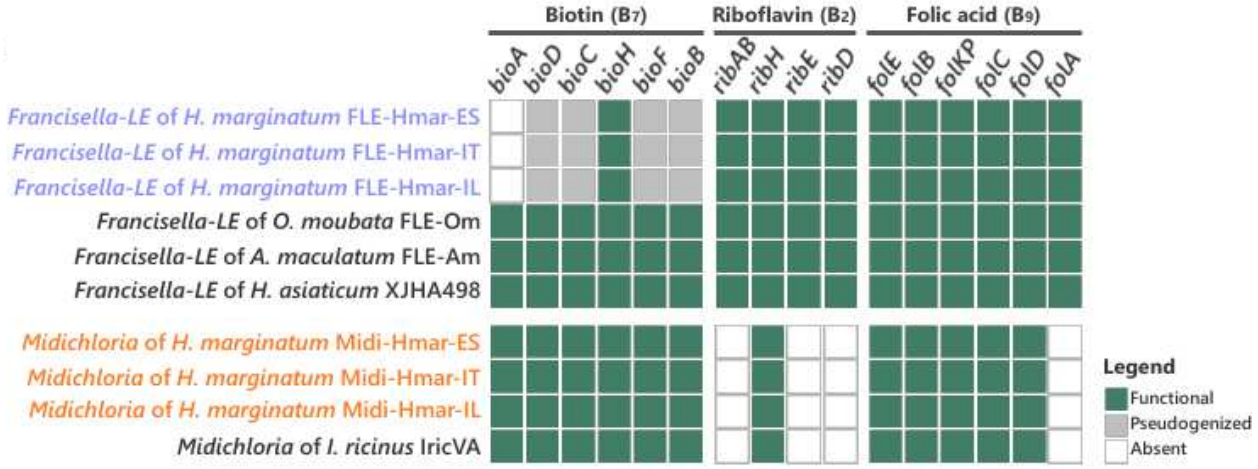


Figure 17: Complemented biosynthetic pathways (Biotin, Riboflavin and Folic Acid)

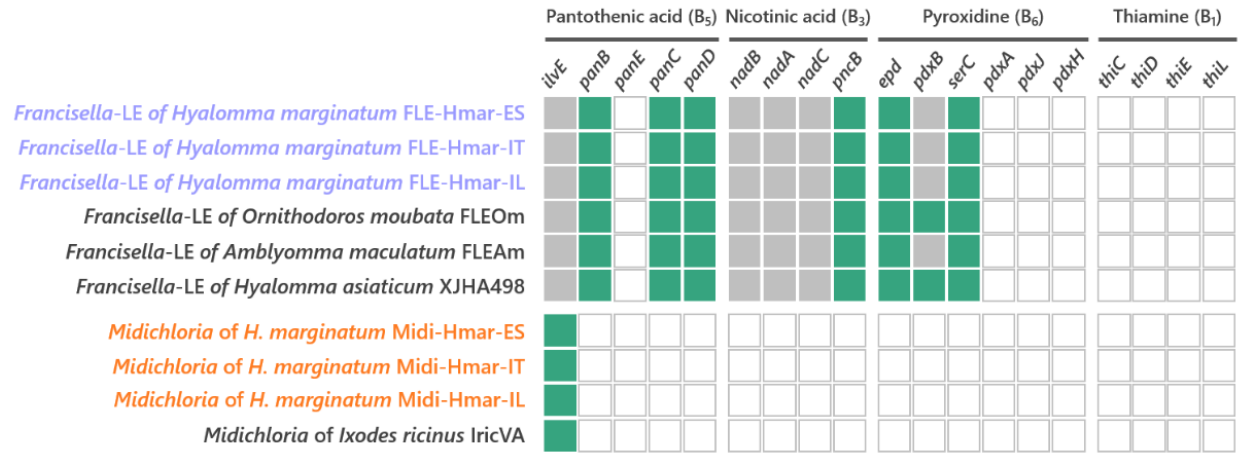


Figure 18: Other partial or absent biosynthetic pathways for B vitamins (B2, B3, B6,B1)

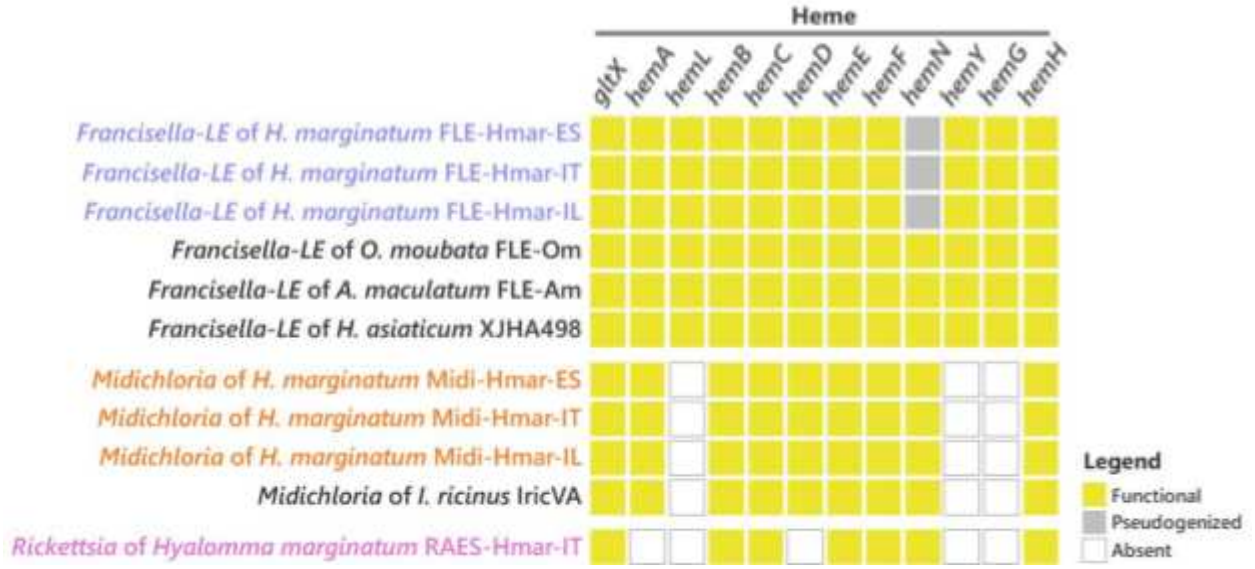


Figure 19: Haem biosynthesis pathway

The B-vitamin pathways in Midi-Hmar genomes are similar to those of *M. mitochondrii*. A complete pathway of biotin was consistently recovered in *Midichloria* genomes (Figure 17). For folate, five of the six biosynthesis genes (*folE*, *folB*, *folKP*, *folC,* and *folD*) were present but the sixth gene, *folA* (encoding for a dihydrofolate reductase), was not. However, we detected an alternative dihydrofolate reductase gene, *folM* (also known as *ydgB*), that was shown in other bacteria to provide the same metabolic function as *folA* (Giladi, 2003) suggesting that the folate pathway is functional in Midi-Hmar genomes. Only partial pathways were detected for riboflavin and pantothenic acid with three and four missing genes, respectively.

No B vitamin pathway seems complete in the RAES-Hmar genome. In the RAES-Hmar-IT genome, partial pathways of biotin and folate were detected with three missing genes each (Figure 17). For pantothenic acid, only one of the five biosynthesis genes is present (Figure 18). No genes for other B vitamin pathways were found in the RAES-Hmar-IT genome. So, we assume that RAES-Hmar is not providing B-vitamin to its host.

We further searched for B vitamin genes of other origins than FLE, *Midichloria,* and *Rickettsia* genomes in tick metagenomes. In the Hmar-ES and Hmar-IL metagenomes (obtained from ovary samples), we found no significant hits, indicating that neither an additional B vitamin provisioning microorganism was present, nor did the tick genome itself encode B vitamin pathways. In the Hmar-IT metagenome (obtained from a whole tick), we found four biotin genes (*bioA*, *bioD*, *bioF,* and *bioB*) in low abundance contigs of *Staphylococcus* bacteria. Previous studies reported the presence of *Staphylococcus*, a common skin-associated bacteria, on tick exoskeletons and midguts and hypothesised this to be due to contamination, or acquisition during feeding (Greay, 2018).

Ticks differ from other eukaryotes in that they lack most of the genes encoding proteins required to produce and degrade haem (Gulia-Nuss, 2016; Jia, 2020). We thus searched for relevant pathways in the symbiont genomes and found that the FLE-Hmar and Midi-Hmar genomes contain genes involved in haem and iron homeostasis. The haem biosynthetic pathway was consistently detected, as well as a gene encoding haem oxygenase that catalyses the degradation of haem and the subsequent release of iron (Figure 19). In these genomes, we detected similar genetic patterns as those observed in the other FEs and *Midichloria* genomes. However, the haem pathway is partially degraded in FLE-Hmar genomes since their *hemN* gene is pseudogenised, a pattern not observed in other FE genomes, suggesting an early stage of degradation for the haem pathway. This gene is present and not pseudogenised in the Midi-Hmar genomes (Figure 19), suggesting these symbionts may compensate for the FEs haem pathway. The three other genes, *hemL*, *hemY,* and *hemG* as expected were not found in *Midichloria* genomes, as they follow the Shemin's pathway as other alphaproteobacteria.

3.5 Discussion

In this study, we show that the tick *H. marginatum* is a complex symbiotic system, combining genes of diverse phylogenetic origins to generate nutritional adaptations to obligate hematophagy. This tick harbours a microbiome functionally analogous to those observed in other tick species, dominated by intracellular bacterial symbionts producing the core set of B vitamins (biotin, riboflavin, and folate), essential for tick growth and survival. However, in other tick species, production of all three core B vitamins is performed by a single nutritional symbiont. Instead, *H. marginatum* harbours two bacterial partners whose core B vitamins biosynthetic capacities depend on a combination of genes. This multi-partner nutritional symbiosis is stably maintained at least at the tick species level, as supported by our founding of genetically similar FE and *Midichloria* present in *H. marginatum* populations from geographically distant regions. Up to now, a similar situation of interlocking biosynthetic pathways has never been described in ticks, but it could be an explanation for the presence of coinfections in some tick species.

Genome reconstructions suggest the key role of symbionts in *H. marginatum* nutrition, but contrary to other tick species, no single symbiont has the genomic potential to compensate for all nutritional needs. The FLE-Hmar genomes contain the genes to synthesise folate and riboflavin, but its biotin synthesis pathway is degraded, and thus non-functional. Genome reduction is a common trait in all FEs, but this process has extended in the FEs of *H. marginatum,* with an additional loss of genes in most COG functional categories. This process could have been triggered by the pseudogenisation of the mismatch repair system (*mutSL*), leaving the FEs of *H. marginatum* with only one DNA repair system, nucleotide excision repair (*uvrABCD*). Absence of the *mutSL* system, coupled with strict clonality and a small population size during transovarial transmission, have probably led the FLE- Hmar genomes to accumulate deleterious mutations, that resulted in maladaptation, and ultimately limited their beneficial contributions to ticks.

In the context of FE genome degradation, co-infection with *Midichloria* prevents the breakdown and collapse of nutritional symbiosis in *H. marginatum*; that is, the *Midichloria* genome contains an intact biotin biosynthesis operon, and thus compensates for genome decay of FLE. Inversely, *Midichloria* has only partial pathways for riboflavin and folate and hence cannot meet the nutritional needs of *H. marginatum* for these B vitamins. Consequently, the co-infection of FE with *Midichloria* forms a cooperative system essential for the nutritional symbiosis in *H. marginatum*. In addition, complementary haem synthesis genes are present in the FEs and *Midichloria* genomes, and these symbionts may be an additional source of haem, perhaps conferring a fitness advantage to *H. marginatum* during periods of starvation. In animals, haem is essential for the function of all aerobic cells, playing an important role in controlling protein synthesis and cell differentiation. Complete biosynthetic haem gene pathways exist in most animals, but not in ticks that need an exogenous source of haem (Gulia-Nuss, 2016; Jia, 2020). Vertebrate blood is usually considered to be their unique source for this essential cofactor (Perner, 2016), however ticks spend most of their life off host, and thus the presence of haem genes in FE and *Midichloria* genomes may also indicate a beneficial symbiont provisioning of haem in *H. marginatum*.

The co-localisation of FE and *Midichloria* in tick ovaries suggests that vertical transmission has led to their stable co-inheritance through tick generations, linking their evolutionary fate. This co-transmission fidelity creates the conditions under which cooperative interactions can emerge. The stable coexistence of FE and *Midichloria* genomes makes various genes redundant, as we observed for B vitamin and haem synthesis pathways. FE is more abundant than *Midichloria*, which can be suggestive of some form of competition, and can be the effect of a host control on their numbers, as the tick can provide limited resources (nutrients, and physical space) for the bacterial symbionts. But the covariance of FE and *Midichloria* densities in ticks is positive, indicating at least that the two symbionts are not in a direct and strict competition. The asymmetrical abundance of the symbionts could imply that lower abundance of *Midichloria* are enough to bridge the biotin requirements, as the host could require a lesser quantity than riboflavin and folate, or a different efficiency of *Midichloria* biotin biosynthesis.

While the nutritional association of FEs with *Hyalomma* species has persisted for millions of years (Azagi *et al*, 2017), the nutritional role of *Midichloria* seems more recent in this tick genus. Co-infection with

Midichloria was reported in at least nine *Hyalomma* species, but in most species, such as *H. excavatum* and *H. impeltatum* (Azagi *et al*, 2017; Selmi, 2019), *Midichloria* is present at much lower frequencies than would be expected for an obligate nutritional mutualist, suggesting that it is instead a facultative (i.e., not essential for the tick) association. Interestingly, the *Hyalomma* species apparently devoid of *Midichloria*, *H. asiaticum*, harbours a FE with intact biotin, folate, and riboflavin synthesis pathways (Buysse, 2021). In *H. marginatum*, the decay of the FE genome may have been accelerated by the presence of *Midichloria* and the redundancy of biotin biosynthesis, a process also reported in sap-feeding insects (Husnik & McCutcheon, 2016; Manzano-Marín *et al*, 2020). While it is clear that two events have led to the current status (loss of biotin genes and acquisition of *Midichloria*), the order by which these happened is unclear. *Midichloria* could have rescued a decaying symbiosis, or could have led, with its entrance in the system, to relaxation of selective constraints on FEs.

IV Correlating population structure with symbiont presence in *Ixodes frontalis*

This is an ongoing project, not yet publishe.

4.1 Introduction

Ixodes frontalis is an understudied ornithophilic tick species, present in a vast area, namely widespread in Europe, western Asia and northern Africa, but more frequent in warmer regions (Hoogstraal *et al*, 1963; Keve *et al*, 2022).

I. frontalis is associated with many common birds in Europe: it mainly parasitises blackbirds (*Turdus merula*), but it has been reported also on many other birds species, such long-eared owl (*Asio otus*), common buzzard *(Buteo buteo*), booted eagle (*Hieraaetus pennatus*), blackcap (*Sylvia atricapilla*), redwing (*Turdus iliacus*), reed bunting (*Emberiza schoeniclus*), large tomtit (*Parus major*), greenfinch (*Carduelis chloris*), whitethroat (*Sylvia communis*), robin (*Erithacus rubecula*), and song thrush (*Turdus philomelos*) (Hornok *et al*, 2016; Santos-Silva *et al*, 2006, 2011; Schorn *et al*, 2011).
Some of these birds (*Turdus merula* and *Erithacus rubecula*) are commonly found in urban areas, and can pose a risk of tick transfer to humans and pets in populated areas (Sándor *et al*, 2014).

Different screenings in Europe have reported the presence of *I. frontalis* in various areas, including peri-urban parks (Drehmann et al, 2019). However, *I. frontalis* was often found in small numbers, usually directly on birds (Estrada-Peña *et al*, 2017). Numbers are especially low when compared to other more common ticks, in particular *Ixodes ricinus*. As *I. frontalis* has a different phenology from *I. ricinus,* its numbers are possibly underestimated: with a presence peak in autumn between October-November, and almost no activity in summer, sampling not directed specifically at *I. frontalis* may risk to retrieve very few individuals (Agoulon *et al*, 2019). Sampling on vegetation has recently been reported to be an efficient way to obtain *I. frontalis* individuals, as many of the birds parasitised by this tick species feed from the ground level, in both open and forested habitats (Keve *et al*, 2022) (Gillingham *et al*, 2020; Kahl *et al*, 2019). A specifically good location to sample *I. frontalis* is under bamboo bushes, where the highest numbers of this species have been reported (Plantard *et al*, 2021).

I. frontalis is implicated in the transmission of bird pathogens, such as the avian tick-related syndrome (TRS) (Monks *et al*, 2006), but its direct pathogenicity potential to humans is generally considered very limited, as *I. frontalis* is specialised on feeding on birds, with extremely rare occurrence of reports of bitten humans (Cull *et al*, 2018). *I. frontalis* is host of the Chizé virus, that is capable of infecting mice in laboratory conditions, but is of unclear pathogenicity (Chastel *et al*, 1999), and hosts and transmits to birds various bacterial pathogens, including: "*Candidatus* Neoehrlichia mikurensis" (Movila *et al*, 2013), *Anaplasma phagocytophilum* (Jahfari *et al*, 2014) and *Borrelia burgdorferi sensu lato* (Heylen *et al*, 2017). Considering that *I. frontalis* occupies a very wide area, infesting many different bird species, and is often transported on long distances by migratory birds (Hoogstraal *et al*, 1963), it could be an agent of horizontal transfer of pathogens, for example through co-feeding, as it shares the environment and hosts with other ticks with direct human pathogenic potential, such as *I. ricinus* (Agoulon *et al*, 2019). The risks of pathogen transmission must be monitored also because they could increase, including even in northern Europe, due to climate warming, with an increase of the area exposed to tick dispersal by migratory birds (Estrada-Peña *et al*, 2017).

Hornok and colleagues recently showed that there are two genetic lineages *Ixodes frontalis*, that present no morphological differences, defined as Haplotype A and B, based on the divergence between the sequences of cytochrome oxidase subunit I (COI) (Hornok *et al*, 2016). The mean COI gene sequence difference between the two haplotypes is 9%, a value that exceeds the 6,1% threshold normally used to define species boundaries in ticks (Lv *et al*, 2014).
A subsequent study performed in Hungary observed (Reynolds *et al*, 2022) that the two lineages were

present in one single sampling place, but with different prevalence, as haplotype A was more common. The COI sequence variability inside the haplotype A was limited, with one predominant internal haplotype. The haplotype B had the opposite situation, with less individuals, but contained more different internal haplotypes. These different patterns were hypothesised to indicate that the haplotype A could be the local population in Central Europe, and haplotype B could be composed only of ticks introduced through migratory birds.

Bacterial symbionts have been shown to be fundamental for ticks populations (see previous chapters). The currently available screening of symbionts present in *I. frontalis* (Duron *et al*, 2017) found the presence of *Rickettsia*, *Spiroplasma* and *Midichloria*, but the sample used was very limited (5 individuals) so a more targeted screening could give a more realistic picture of the situation.

4.2 Aim

The aim of the work presented in this chapter is to add more data on the presence of the understudied tick *I. frontalis* in France and Italy. We investigate the prevalence of the two haplotypes, and characterise the population of symbiotic bacteria in the tick species, looking for differences in the presence of symbionts between the two genetic lineages of the tick.

4.3 Methods

4.3.1 Sampling

Tick sampling was conducted between November 2017 and October 2022, in different localities characterised by the presence of bamboo bushes in France and northern Italy. Six sampling locations were near Nantes (Parc de la Chantrerie, Parc Floral de la Beaujoire, La Hérinière, Grand Blottereau, Jardin des Plantes Rezé, Orvault Provotiere), one in Angers, two in Paris (Parc Floral de Vincennes, and Jardin de Plantes), and one sampling location was in Lombardia, Italy (Parco della Sora). All field samples were collected using the flagging technique, sweeping between bamboo stems. 10 other ticks were collected from live birds in France, 7 at Le Loroux Bottereaux (*Passer domesticus*), 1 at Treillières (*Turdus philomelos*), 1 at Sucé sur Erdre (*Turdus merula*), and 1 at Vigneux de Bretagne (*Columba palumbus*).
Collected ticks were morphologically identified following the keys previously reported (Agoulon et al., 2019; Pfäffle et al., 2017). Ticks were then preserved in ethanol (70%) at 4°C.

4.3.2 DNA extraction

DNA was extracted using the NucleoSpin Plant II (Macherey-Nagel GmbH and co. KG, Düren, Germany), following the manufacturer's protocol, optimised for ticks. The extracted DNA was stored at -20°C until use.

4.3.3 PCR and sequencing

To confirm the morphological identification of ticks, a previously published PCR protocol was performed for the mitochondrial COI gene (Hornok *et al*, 2016). Briefly, a fragment of 710 bp was amplified with the universal primers HCO2198 (5′-TAA ACT TCA GGG TGA CCA AAA AAT CA-3′) and LCO1490 (5′-GGT CAA CAA ATC ATA AAG ATA TTG G-3′). The amplification protocol was the following: an initial denaturation step at 95 °C for 5 min was followed by 40 cycles of denaturation at 94 °C for 40 s, annealing at 48 °C for 1 min, and extension at 72 °C for 1 min. Final extension was performed at 72 °C for 10 min.
The DNA was subjected to Sanger sequencing at Eurofins genomics.

4.3.4 Phylogenetics

The sequences of COI genes were aligned with muscle (Edgar, 2004), and the alignment was used to infer

a tree withIQ-TREE 2.2.0.3 (Nguyen *et al*, 2015), with 1000 ultrafast bootstraps (Minh, Nguyen, and von Haeseler 2013) and 1000 SH-aLRT replicates, using the model inferred by IQ-TREE (HKY+F).

4.3.5 16S Amplicon metagenomics

A subset of nine samples were selected for a 16S metagenomic analysis.
For the procedure the V3-V4 region of 16S rDNA was amplified using the universal primers for prokaryotes (including the 5' adapters for sequencing): Pro341F (5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TCC TAC GGG AGG CAG CAG CCT ACG GGN BGC ASC AG-3') and Pro805R (5'-CAA GCA GAA GAC GGC ATA CGA GAT NNN NNN GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TGA CTA CNV GGG TAT CTA ATC C-3').

For the PCR reaction, 3µL of DNA was amplified in a total volume of 40 µL, containing the following reagents: 24,74 µL of distilled water, 4 µL of Ex Taq Buffer 10X (20 mM Tris HCl - pH 8.0, 100mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.5% Tween 20, 0.5% NP-40, 50% Glycerol, 20 mM Mg^{2+}), 4 µL of dNTP (2.5 mM each of dATP, dTTP, dGTP and dCTP), 2 µL of each primer (0,5mM), 0.26 µLof TaKaRa Ex Taq polymerase (Takara Bio, Japan).

The PCR reaction conditions for amplification of DNA were as follows: initial denaturation at 95°C for 3 min, followed by 30 cycles of denaturation at 95° for 30 sec, annealing at 55°C for 30 sec, and extension at 72°C for 30 sec. Final extension was performed at 72 °C for 5 minutes.

The PCR products were sent to MBR Genetics (Italy) for sequencing, and I analysed the sequences through the R package DADA2 (1.26.0), using the Silva 16S sequence database SSU 138.1.

4.4 Results

4.4.1 Tick samples and lineage identification

233 specimens of different instars were collected from vegetation in different locations. Ten other ticks were collected from live birds in France, seven at Le Loroux Bottereaux (*Passer domesticus*), one at Treillières (*Turdus philomelos*), one at Sucé sur Erdre (*Turdus merula*), and one at Vigneux de Bretagne (*Columba palumbus*).
2022 ticks were collected in France, 47 ticks were collected in Italy. 165 were larvae, 53 nymphs and 15 adults.

All the collected specimens were morphologically identified as *I. frontalis*, and thus COI PCRs were performed to assess the ticks' genetic lineage. The results indicate that the majority of ticks belong to the lineage A: overall, 165 ticks (70%) belong to the lineage A (Hu-A), and 68 specimens (30%) are of the lineage B (Hu-B). The full results are in Table 4.

Most of the ticks sampled in France were of the lineage A, with a minority of lineage B, and that situation holds true for most sampling locations in France,with the opposite situation was found for the Italian sampling site and Rezè in France.

Location	Haplotype A	Haplotype B	Total
Angers_(FR)	17	4	21
Beaujoire_(FR)	19	4	23
Chantrerie_(FR)	18	1	19
Grand Blottereau_(FR)	36	4	40
Jardin des Plantes (Paris-FR)	17	3	20
La_Hérinière_(FR)	14	0	14
Orvault Provotiere_(FR)	5	0	5
Rezé_(FR)	5	15	20
Sora_PV	13	33	46
Vincennes (Parc Floral)_(FR)	12	3	15
Sampled from birds (FR)	9	1	10
Total	165	68	233

Table 4: Haplotypes of the I. frontalis *samples in the different sampling locations*

4.4.2 Phylogenetics

We used the COI sequence data from 82 individuals to obtain a ML tree (Figure 23). The branch lengths inside the lineage B are on average longer than the branches on lineage A, confirming the higher sequence variability within this lineage observed previously (Reynolds *et al*, 2022). This result is compatible with the scenario of multiple importation of ticks of lineage B from locations through bird migration.

4.4.3 Symbionts screening

Concerning the screening for the detection of symbionts, the results indicate that the frequencies of *Midichloria* and *Spiroplasma* are quite similar in the analysed *I. frontalis* populations (18% and 19% respectively).
Results of the frequencies in Nantes (Figure 20): La Herinière (36% of ticks carries *S. ixodetis*, while none resulted positive for *Midichloria*), Parc de la Chantrerie (11% of samples carries *Midichloria*, 21% carries *S. ixodetis*, while 5% carries both symbionts), Orvault Provotiere (80% of samples carries *Midichloria*, while the 20% carries *S. ixodetis*), Beaujoire (21% of samples carries *Midichloria*, 25% carries *S. ixodetis*) and Grand Blottereau (22,2% of ticks carries *Midichloria*, 2,5% carries *S. ixodetis*, while 2,5% carries both symbionts), all the ticks from Rezè are devoid of any known symbionts.
As regards to the other sites in France and Italy(Figure 21): Angers (41,6% of samples carries *Midichloria*, 12,5% carries *S. ixodetis*, while 4,3% carries both symbionts), Jardin de Plantes (5% of ticks carries *Midichloria*, 20% carries *S. ixodetis*, while 5% carries both symbionts), Parc Floral (7% of samples carries *Midichloria*, 13% carries *S. ixodetis*, while 7% carries both symbionts); Sora (9% of ticks carries *Midichloria*, and 38% carries *S. ixodetis*).

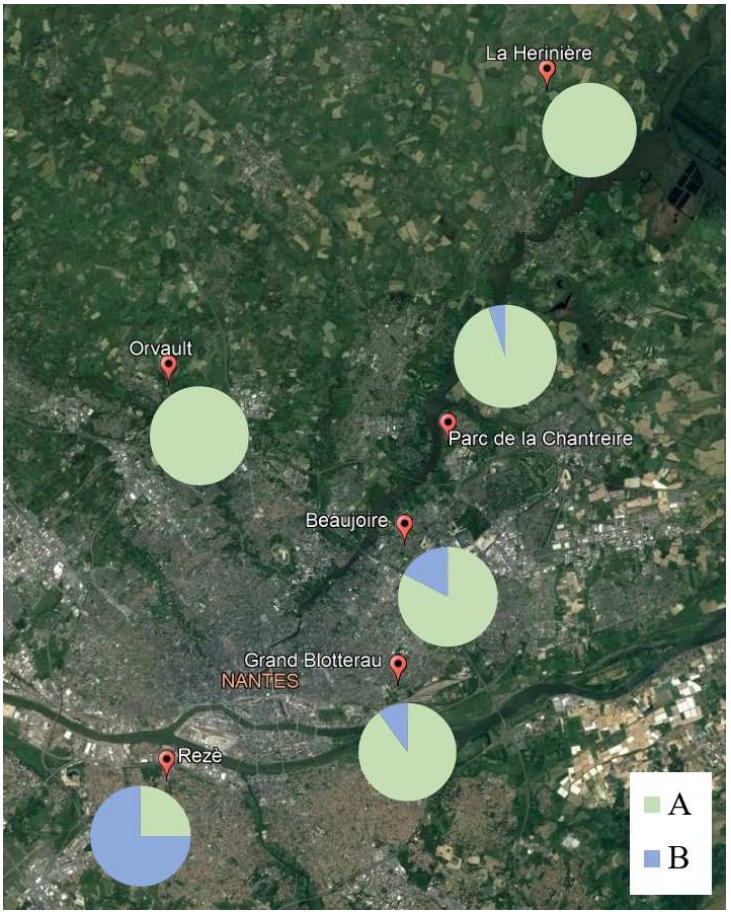


Figure 20. Pie-charts for the lineages' prevalences for the six sampling sites in Nantes. Light green indicates the lineage A, light blue indicates the lineage B.

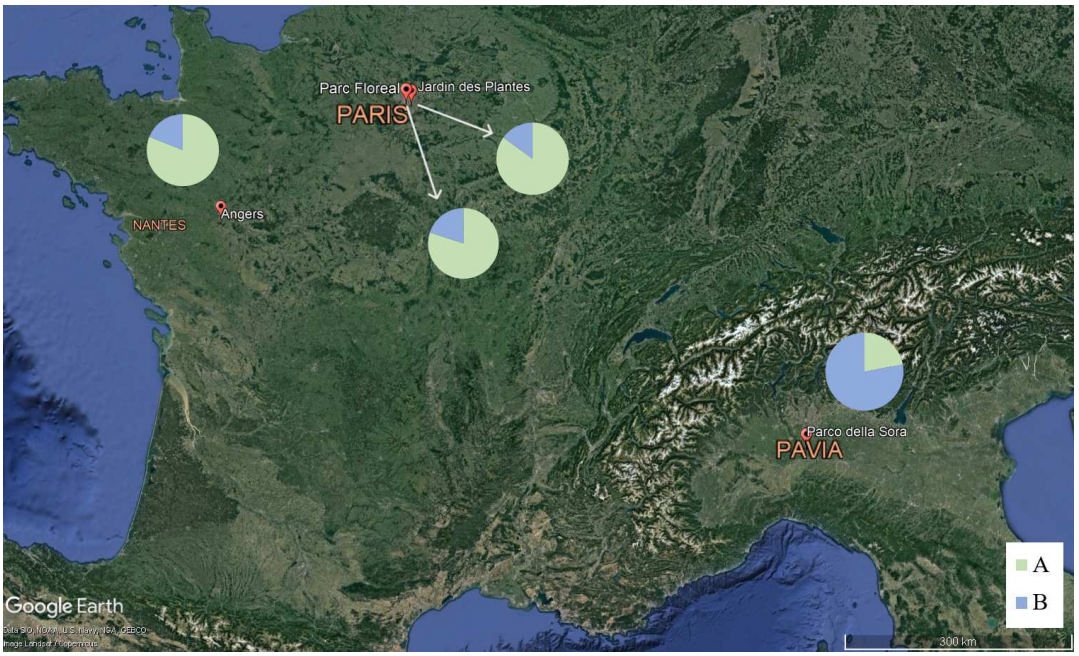


Figure 21. Pie-charts of the two genetic lineages' prevalences for the other sampling sites in Italy and in France. Light green indicates the A lineage, light blue indicates the B lineage.

Concerning the *Rickettsia* screening, all the tested samples resulted negative to the presence of this bacterium.
A high number of screened ticks resulted negative for all tested symbionts, the percentages are listed below, divided per locations: Parc de la Chantrerie (63%), Parc Floral de la Beaujoire (54%), La Hérinière (64%), Grand Blottereau (72,5%), Jardin des Plantes Rezé (100%), Angers (41,6%), Parc Floral de Vincennes (73%) and Jardin de Plantes and Sora (53%).

Lineage	**Null**	**S+M**	***Midichloria***	***S. ixodetis***	***Total***
A	92 (55.7%)	13 (7.8%)	36 (15.4%)	24 (14.5%)	**165**
B	54 (77.9%)	0	1(1,5%)	14(20.6%)	**68**
					233

Table 5: all individuals screened in this work, positives for Midichloria *or* S. ixodetis *include samples that were only positive for that bacteria, and not samples positive for both (S+M)*

4.4.5 16S metagenomics

In order to investigate why a high portion of the samples resulted negative to the PCR screening for the presence of symbionts, a subset of nine samples were selected for a 16S metagenomic analysis. We analysed one sample positive for both symbionts (VIN1), one sample carrying only *Midichloria* (OP3), one carrying only *S. ixodetis* (AN3), six samples that were PCR-negative for the three bacterial symbionts tested, and a negative control (NS) was added (Figure 22).

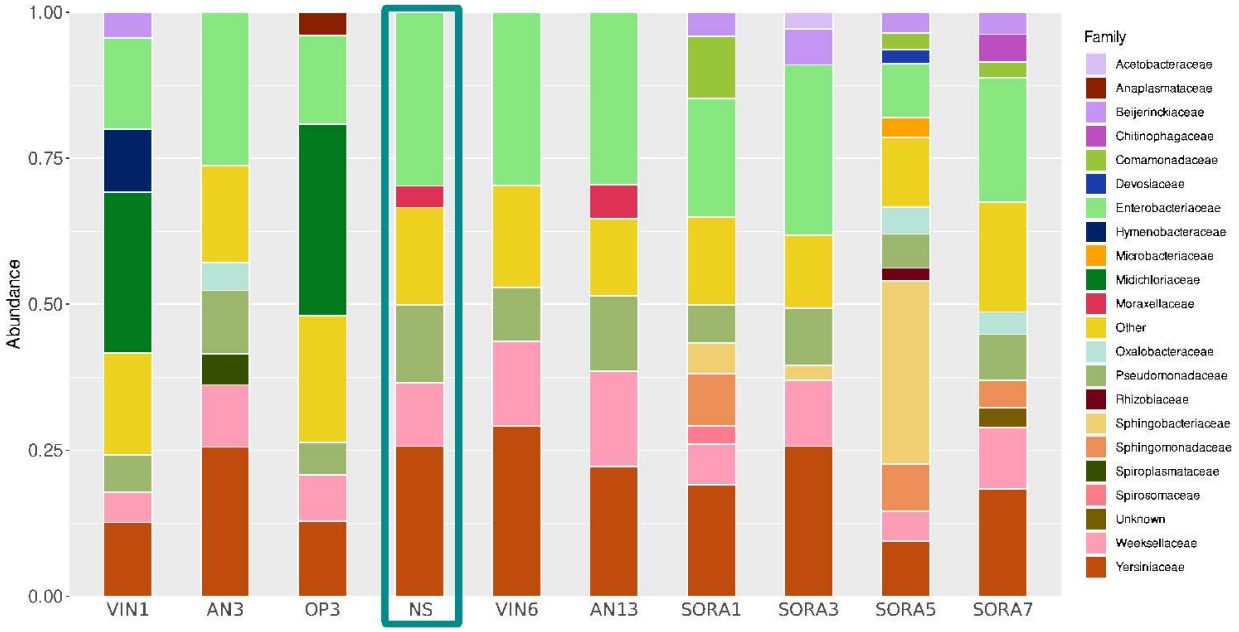


Figure 22: taxa composition of the 16 amplicon metagenomics sample. The negative sample (NS) has a teal coloured square.

The analysis confirmed the presence of *Midichloria* and *Spiroplasma* in the first three samples, while no other known symbiont was detected in the others. Pathogenic bacteria belonging to the genera *Borrelia* and *Ehrlichia* were also detected in two ticks (OP3 and SORA3). The other genera detected are known environmental bacteria that are commonly found in 16S amplicon metagenomics, and probably represent environmental contaminants. The samples that were negative for the presence of symbionts present a

bacterial community similar to that of the negative control, with the differences ascribable to environmental bacteria (i.e. *Massilia* taxon in SORA5). As the samples were composed mostly from nymphs, it is possible that DNA was present at low quantity, not enough to be amplified by the PCR, and that could have happened to all PCR samples negative for the presence of any symbionts.

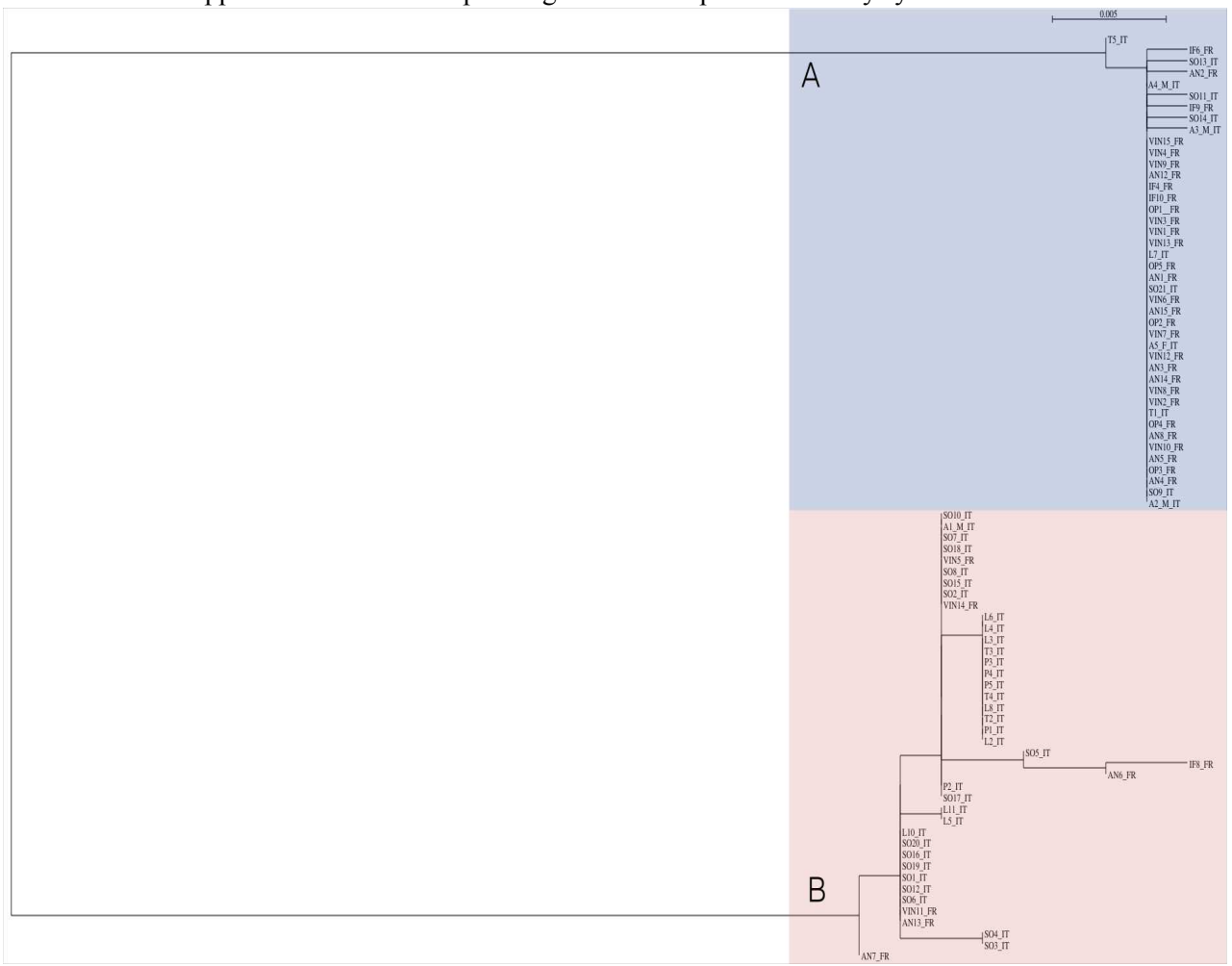


Figure 23: phylogenetic tree of COI sequences. In blue Lineage A, in red Lineage B.

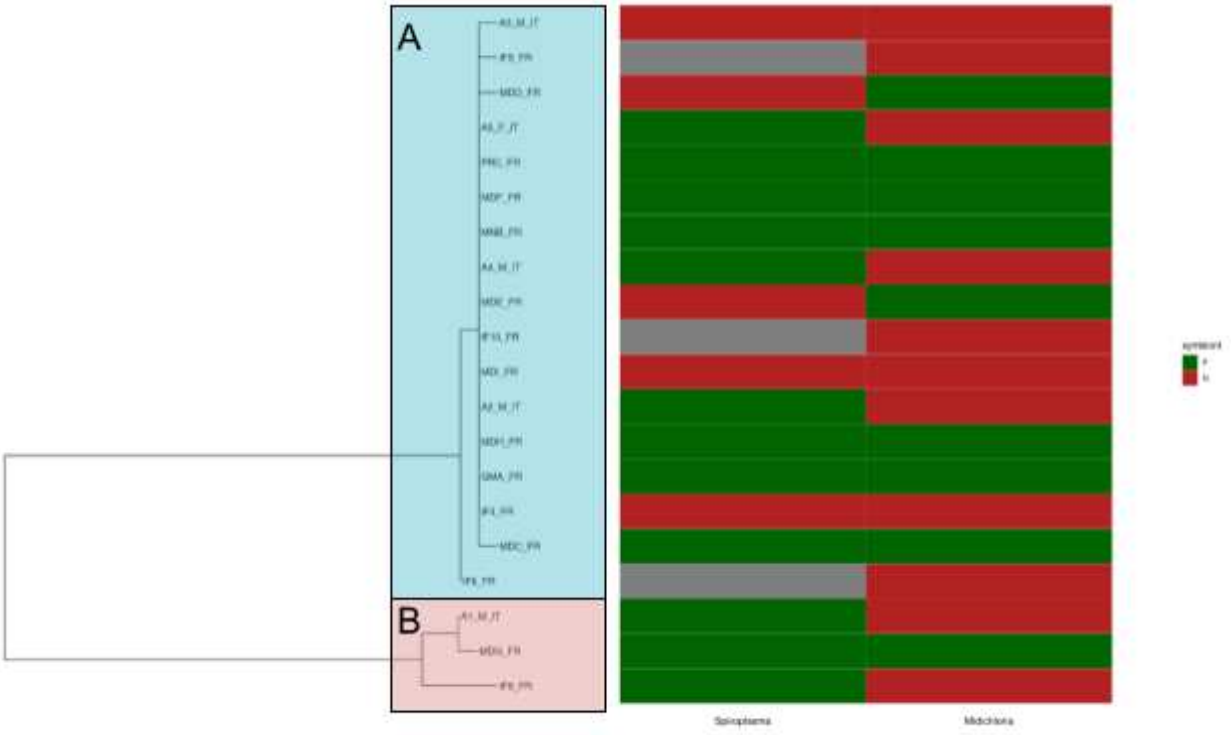


Figure 24: phylogenetic tree of COI sequences including only I. frontalis *adults, with the screened symbionts. In light blue Lineage A, in light red Lineage B. In green are indicated positive samples, in red negative and in grey samples that were not tested.*

4.5 Discussion

In this work we investigated the bacterial symbionts in the ornithophilic hard tick *I. frontalis*, and tried to correlate their presence with the genetic structure of this tick, divided in two lineages.

We observed a different presence of the two lineages of *I. frontalis* in the sampling sites in France and Italy. Lineage A was the majority of sampled ticks in (Reynolds *et al*, 2022) and it was proposed that in Hungary the lineage A was the local population, and lineage B arrived mostly through migratory birds.
Considering our data, we can propose that the two different lineages represent an allopatric separation between the two lineages, with one more common in Central/Western Europe (Haplotype A), and one more common in southern countries (Haplotype B), possibly due to geographical barriers such as the Alps. Phylogenetic analysis confirmed the higher variability within lineage B, and slightly more variability for the lineage A samples from Italy. An explanation for the presence of a majority of lineage B also in a French site could be a low presence of "local" ticks, possibly for some preference of migratory birds for that location.
Analysing these data in conjunction with the available information on tick migratory routes could provide additional clues on the population structure of *I. frontalis*.

We screened all the collected individuals for *S. ixodetis*, *Midichloria* and *Rickettsia*, detecting a low overall prevalence of symbionts. To evaluate whether this was due to PCR issues (e.g. primer specificity) or to the presence of other symbionts, we performed 16S metabarcoding on a subset of samples. This screening confirmed the results of the PCRs, suggesting that the high number of negative samples could be due to insufficient amounts of DNA. Indeed symbiont prevalence in adults was much higher. We know that the *Midichloria* population in *I. ricinus* blooms after blood-meal (Sassera *et al*, 2008), thus it is possible that our *I. frontalis* negative samples had a long time since the last blood-meal before being collected, with a low bacterial load, sufficient for expanding again at the next blood-meal, but not enough to be identified.

Considering the high number of sampled nymphs for which we could not identify the presence of bacterial symbionts, further attempts to screen the *I. frontalis* for symbionts should focus on adults, or on sampling ticks directly from the birds.

Among tick mutualists, we found both *Spiroplasma ixodetis* and *Midichloria* in *I. frontalis*. We did not find any sample positive for *Rickettsia* endosymbiont, differently from what previously reported (Duron *et al*, 2017). It is possible that the presence of *Rickettsia* in *I. frontalis* is dependent on the sample origin, as with other *Rickettsia* symbionts and pathogens in ticks (Niebylski *et al*, 1997). Considering the results from the 16S amplicon metabarcoding, we exclude the presence of symbionts other than *S. ixodetis* and *Midichloria* in *I. frontalis*. In lineage A, taking into consideration samples that were positive for one symbiont, we had a similar prevalence for *S. ixodetis* and *Midichloria*. Instead, for the lineage B only one sample was positive for *Midichloria* out of 68, but we were not able to detect any symbionts for most of them. Taking into account only the data from adults, from which we were able to detect a symbiont in 13 over 15 samples, we have a high prevalence of *S. ixodetis* (12 samples), and a close value for *Midichloria* (9 samples).

For this ongoing project, in the future we aim to first expand the number of adults sampled, to have a more fitting and solid idea of the composition of the symbiont population in *I. frontalis*, seeing if the sampled larvae and nymphs that presented a symbiont were representative of the population, or if was present some bias (e.g. if one the two symbionts is difficult to detect in non adult stages). Also, we aim to detect which pathogens are present in the ticks, if their presence is similar in the different sampling sites, or if geography plays a role into the population of pathogens in *I. frontalis*. Also, we want to see if there is a relation between the presence of one the two mutualists and the pathogenic bacteria, such as exclusion pattern or a positive correlation.

V Deep evolution of *Rickettsiales*: independent development of host association and intracellularity

Currently available on pre-print as: Michele Castelli, **Tiago Nardi**, Leandro Gammuto, Greta Bellinzona, Elena Sabaneyeva, Alexey Potekhin, Valentina Serra, Giulio Petroni, Davide Sassera
Host association and intracellularity evolved multiple times independently in the *Rickettsiales*
bioRxiv 2022.10.13.511287; doi: https://doi.org/10.1101/2022.10.13.511287
Supplementary material available at https://www.biorxiv.org/content/10.1101/2022.10.13.511287v1

The previous three chapters are focused on specific relationships of single lineages of mutualistic bacteria with tick hosts, by identification and comparative evolutionary analyses of the bacteria. In this case, the focus of the project will be from an enlarged evolutionary perspective, namely starting from the whole variability available among *Rickettsiales*. This bacterial order was initially defined as composed of tick - and other vector- borne pathogens, but the views have changed by the progressive discovery of new members of the taxon, associated with a wide range of eukaryotes with variegated interactions. A better understanding of the evolution of the whole *Rickettsiales* will provide insights on the evolution of the tick-associated lifestyle in many representatives, including vertebrate pathogens (e.g. *Rickettsia* spp, Chap 1.1), and mutualists (*Midichloria* spp. Chap. 1.3.2).

5.1 Introduction

Rickettsiales are an early diverging (Muñoz-Gómez *et al*, 2019) and ancient (Wang & Luo, 2021) alphaproteobacterial order. All the experimentally characterised members of this group engage in obligate associations with eukaryotic host cells (Salje, 2021). The most long-term and thoroughly studied *Rickettsiales* include vector-borne pathogens, e.g. *Rickettsia* and *Anaplasma*, causing various diseases in humans and vertebrates (Walker & Ismail, 2008; Rikihisa, 2010; Renvoisé *et al*, 2011), as well as *Wolbachia*, that can establish complex interactions with arthropod and nematode hosts (Werren *et al*, 2008), chiefly reproductive manipulation and mutualism.
In recent years, our knowledge and understanding of the origin, evolution and diversification of *Rickettsiales* have been remarkably improved. We can identify in particular three main advances. The first is the finding of a plethora of novel lineages (over 30 total genera described, grouped into seven families (Castelli *et al*, 2019; Montagna *et al*, 2013; Schön *et al*, 2022; Dumler *et al*, 2015), living in association with a wide variety of hosts (Carrier *et al*, 2021; Davison *et al*, 2022; Gruber-Vodicka *et al*, 2019). Most of those hosts are diverse aquatic unicellular eukaryotes (e.g. ciliates, amoebae, algae) (Yurchenko *et al*, 2018; George *et al*, 2020; Hess, 2017; Castelli *et al*, 2021; Muñoz-Gómez *et al*, 2019; Giannotti *et al*, 2022; Schulz *et al*, 2016), which have been deemed as probable ancestral hosts (Vannini *et al*, 2005; Wang & Luo, 2021; Weinert *et al*, 2009), even though these associations are still poorly investigated.

Second, "*Candidatus* Deianiraea vastatrix" (from now on, *Candidatus* will be omitted from taxonomic names, e.g. *Deianiraea vastatrix*), a fully extracellular *Rickettsiales* bacterium equipped with an unexpectedly large biosynthetic repertoire for amino acids, was discovered (Castelli *et al*, 2019), opening a new perspective on the evolution of *Rickettsiales*. While previous views implied that obligate intracellular association dated back to the last common ancestor of the order ("intracellularity early" hypothesis), this discovery opened a novel alternative scenario. Accordingly, obligate intracellularity could have evolved later and multiple times independently in different sublineages ("intracellularity late" hypothesis), together with a stronger dependence on host cells.

Third, metagenome binning recently allowed the discovery of further *Rickettsiales* sublineages, in particular two early-diverging families (Schön *et al*, 2022). Their genetic repertoire (including nutrient uptake, detoxification, and multiple biosynthetic pathways) led the authors to hypothesise that these

bacteria could be free-living, implying that adaptation to the interaction with host cells would have occurred in more "derived" *Rickettsiales* lineages.

However, many open points still exist on the origin and evolution of the interaction between *Rickettsiales* and their hosts, in particular for what concerns the process(es) of transition between facultative/obligate association and between extracellular/intracellular condition, and whether those were "early" or "late" phenomena. In order to address such salient questions, in this study we collected an unprecedentedly large and representative genomic dataset of *Rickettsiales*, thanks to *de novo* sequencing and selection of metagenomic sequences, thus identifying three novel families and remarkably extending the available diversity within previously known lineages. This allowed us to get a view on the diversity and evolution of host adaptation among *Rickettsiales*, finding multiple and convincing lines of evidence supporting the intracellularity late hypothesis.

5.2 Methods

Sample preparation and sequencing

In this work, the nine novel *Rickettsiales* genome sequences were assembled (for a detailed account on sample preparation, sequencing, and genome assembly, see Supplementary text 3; Supplementary table 1, 4), starting from eight protist host samples. Six of those samples were characterised in previous studies (Vannini *et al*, 2010; Lanzoni *et al*, 2019; Szokoli *et al*, 2016; Boscaro *et al*, 2013; Glöckner *et al*, 2014; Mironov & Sabaneyeva, 2020), while two others, namely the ciliates *Plagiopyla frontata* IBS-3 and *Euplotes woodruffi* NDG2, were newly isolated. Each sample was differentially processed. Briefly, for Illumina sequencing most samples were subjected to whole-genome amplification (WGA) with the REPLI-g Single Cell Kit (Qiagen), either directly from few ciliate host cells (four samples: *Paramecium biaurelia* US_Bl 11III1, *Paramecium nephridiatum* Sr 2-6, *E. woodruffi* NDG2, *P. frontata* IBS-3) or from a previously obtained DNA extract (one sample: *Euplotes harpa* BOD18(Claudia Vannini et al. 2010); Supplementary table 1). Two additional samples (*P. biaurelia* USBL-36I1 and *Paramecium multimicronucleatum* Kr154-4) were processed for bulk DNA extraction (over 200,000 ciliate host cells each), with CTAB and phenol-chloroform protocols, respectively. Each of those seven DNA samples was processed through a Nextera XT library and sequenced by Admera Health (South Plainfield, NJ, USA) on a Illumina HiSeq X machine, producing 2x150 bp paired-end reads. Read quality was assessed with FastQC. NDG2 sample was also subjected to Nanopore sequencing. For this purpose, a bulk (~ 300,000 *Euplotes* cells) extract with the NucleoSpin™ Tissue Kit (Macherey-Nagel™) was processed through a SQK-LSK109 ligation-sequencing library, and sequenced in a FLO-MIN106 flow cell. Basecalling was then performed with guppy 5.0.11. Then, reads were processed with Porechop 0.2.4 (Wick *et al*, 2017b) with default options. Quality of the reads was assessed with NanoPlot 1.23.0 (De Coster *et al*, 2018). The eighth sample consisted in a *Rickettsiales* bacterium associated with the foraminiferan *Reticulomyxa filosa*, already sequenced together with its host in a previous study (Glöckner *et al*, 2014). Sequencing reads were kindly provided by the original authors.

Genome assembly

For each sample, the total Illumina reads were assembled using SPAdes 3.6 (Bankevich *et al*, 2012) with default settings, obtaining a "preliminary assembly". Then, a multi-step procedure was applied, in order to select only those contigs belonging to the symbiont of interest and discard those belonging to the host or to additional organisms present in the sample (e.g. residual food, additional associated bacteria), as described previously (e.g(Castelli *et al*, 2019)). For this purpose, the blobology pipeline was applied (Kumar *et al*, 2013), followed by extensive manual curation. Briefly, preliminary contigs were classified according to their length, GC% content, sequencing coverage, and taxonomy. Reads mapping (Langmead & Salzberg, 2012) on selected contigs were reassembled separately with SPAdes 3.6, or, for NDG2, with Unicycler (Wick *et al*, 2017a) in a hybrid assembly with the respective Nanopore reads. Two samples (*P. multimicronucleatum* 12 and *P. biaurelia* US_Bl 11III1) were subjected to genome finishing, performing

PCR reactions with TaKaRa Ex Taq and reagents (Takara Bio, Japan). Successful results were confirmed by bidirectional Sanger sequencing performed by GATC Biotech (Germany).

Annotation

The newly obtained genomes were all annotated with Prokka 1.10 (Seemann, 2014), using the --rfam option. Afterwards, annotation of the genomes of ciliate symbionts was manually curated by a detailed inspection of blastp hits on NCBI nr and on *Rickettsiales* proteins as described previously (Castelli *et al*, 2019).

Full *Rickettsiales* dataset construction and phylogenomic analyses

Phylogenomic analyses were aimed to collect a representative and comprehensive view on the evolution and diversity of *Rickettsiales*. All sequences were downloaded from NCBI GenBank via ftp (ftp.ncbi.nlm.nih.gov/genomes/all/GCA), and are updated to July 2021. We manually selected a representative set of 36 *Rickettsiales* genomes, including at least one representative per genus. For the phylogeny, other 89 representative non-*Rickettsiales* *Alphaproteobacteria*, as well as 8 *Gammaproteobacteria* and *Betaproteobacteria* as outgroup were employed, taking inspiration from the selection by Muñoz-Gómez and co-authors (Muñoz-Gómez *et al*, 2019). We then aimed to identify all MAGs (metagenome-assembled genomes) which could be assigned to known "core" *Rickettsiales* lineages as of July 2021 (i.e. the four families *Rickettsiaceae*, *Anaplasmataceae*, *Midichloriaceae*, *Deianiraeaceae*), or to any lineage forming a supported monophyletic group with *Rickettsiales* with the exclusion of other (alpha)proteobacterial orders. Identification and representative selection of *Rickettsiales* MAGs was performed by a multi-step procedure (detailed in Supplementary text 4). Briefly, all MAGs assigned to *Rickettsiales* by NCBI taxonomy, those assigned to deep-branching alphaproteobacterial lineages (Martijn *et al*, 2018), plus all additional monophyletic relatives from the gtdb tree (Parks *et al*, 2022), were downloaded (394 total MAGs). MAGs were first filtered by assembly quality, retaining only 314 MAGs having ≥50% single-copy and <5% duplicated of 219 proteobacterial orthologs according to BUSCO 5.0.0 (Simão *et al*, 2015) (Supplementary table 3).

All phylogenomic analyses were performed on concatenated alignments of 179 orthogroups (Supplementary table 5), which were manually selected on purpose (i.e. presence, after manual polishing of paralogs and poorly aligned sequences, in at least 85% of *Rickettsiales* -MAGs excluded-, 85% *Alphaproteobacteria*, 50% outgroup) from the eggNOG orthogroups (Huerta-Cepas *et al*, 2019) predicted with eggnog-mapper 2.0.6 (Cantalapiedra *et al*, 2021). For each organismal dataset (see below), orthogroups were separately aligned (Katoh & Standley, 2013), polished (Criscuolo & Gribaldo, 2010) and concatenated (Borowiec, 2016). All phylogenies were performed with IQ-TREE (Nguyen *et al*, 2015), with 1000 ultrafast bootstraps (Minh *et al*, 2013) and 1000 SH-aLRT replicates, employing the LG+C60+F+R6 model unless specified. A first phylogeny was performed on the full organismal dataset for an initial classification of MAGs, employing ModelFinder (Kalyaanamoorthy *et al*, 2017) for model selection. In order to avoid artefacts due to compositional heterogeneity in the dataset (in particular potential "erroneous" phylogenetic proximity of MAG lineages to "core" *Rickettsiales* due GC/AT biases) (Muñoz-Gómez *et al*, 2019; Viklund *et al*, 2012; Martijn *et al*, 2018), the approach by Muñoz-Gómez and co-authors (Muñoz-Gómez *et al*, 2019) was applied, thus removing 10%, 20%, 30%, 40% or 50% of most heterogeneous sites from the alignment, and performing a separate phylogeny on each trimmed alignment (Supplementary figure 11). Based on resulting monophyly and Average Amino acid Identity (AAI) >0.85, phylogenetically-redundant MAGs were discarded (Supplementary figure 12). Thirteen clades were identified, grouping all those MAGs which could not be directly assigned to "core" *Rickettsiales* or other orders. In order to minimise potential artefacts (e.g. due to long-branch attraction), the phylogenetic position of the MAGs belonging to each clade was tested separately, with respect to core *Rickettsiales* lineages and other *Alphaproteobacteria* (Supplementary figure 13, 14). Phylogenies were performed accounting for compositional heterogeneity as above, and only MAGs in a monophyletic relationship with *Rickettsiales* were retained. Therefore, 68 total *Rickettsiales* MAGs were selected for all the successive analyses, totalling 210 organisms in the final dataset (113 *Rickettsiales*). For the final dataset, we aimed to "balance" the well-ascertained artefacts in alphaproteobacterial phylogeny due to compositional heterogeneity (e.g(Muñoz-Gómez *et al*, 2019; Viklund *et al*, 2012; Martijn *et al*, 2018)) and the loss of a

considerable amount of phylogenetic information (Fan *et al*, 2020), due to the removal of the most-compositionally heterogeneous sites, and likely important for the reconstruction of the inner relationships within *Rickettsiales*. Therefore, we employed the full alignment, using a guide tree (-g option in IQ-TREE) with a minimal set of just two constrained bipartitions (namely those separating *Rickettsiales* from the AT-rich *Holosporales* and *Pelagibacterales*), thus considering the non-compositionally biassed topology among the alphaproteobacterial orders, and leaving at the same time freely unconstrained tree search for what concerns the inner relationships within *Rickettsiales*.

Creation of a set of orthogroups for gene content comparisons

In order to perform gene content comparisons and reconstruct variations along the inferred species tree, a set of orthogroups was obtained for the 210 total organisms in the final phylogeny, starting from the previously obtained eggNOG orthogroups (see Supplementary text 5). The eggNOG database is hierarchically organised by taxonomy, namely each lineage (at domain, phylum, class, order, family levels) has a dedicated set of orthogroups, linked to those of higher level taxa. We aimed to maximise the advantages of such a system (chiefly lineage-specific annotations), and at the same time minimise potential disadvantages, such as, in particular, potential lack of identification of orthologs due to assignment to eggNOG orthogroups belonging to distinct taxa, which may be especially relevant for organisms displaying high sequence evolutionary rates, such as the *Rickettsiales* (Castelli *et al*, 2019). For this purpose, we designed a "telescopic" approach, in order to merge genes into large meaningful orthogroups, while still keeping as much as possible lineage-specific refined annotation. Briefly (see Supplementary text 5 for details), we compared taxonomic paths of all the identified eggNOG orthogroups, and grouped together all the orthogroups sharing at least a partial taxonomic path, assigning each group to the orthogroup at lowest possible shared rank in the taxonomic path leading to *Rickettsiales* (root; *Bacteria*; *Proteobacteria*; *Alphaproteobacteria*; *Rickettsiales*). Applying such a "telescopic" approach, a total of 444,226 genes were assigned to 20,041 orthogroups, 4009 of which present in at least one member of *Rickettsiales*, and 2990 of those present in 4 or more organisms of the total dataset, and thus considered for the following analyses.

Reconstruction of ancestral states of gene copy number

Reconstruction of ancestral states (in terms of gene copy number in each orthogroup) was performed by a tree-reconciliation approach, employing ALE 0.4 (Szöllõsi *et al*, 2013). Briefly, each of the 2990 orthogroups was aligned (Katoh & Standley, 2013), and trimmed(Criscuolo & Gribaldo, 2010). Then, only the 2871 orthogroups ≥30 amino acids after trimming were kept, and, for each of those, a sample of gene trees was obtained running 10000 iterations with PhyloBayes 4.1 (Lartillot & Philippe, 2004). Then, amalgamated likelihood estimation was performed with ALEobserve, run with 10% burn-in, followed by ALEml_undated (Szöllősi *et al*, 2015). For each node of the species tree, predicted events and predicted (or observed at tips) copy numbers were rounded using a 0.3 threshold (Martijn *et al*, 2020), and then summed considering the whole dataset, as well as separately for each eggNOG functional category.

Phylogenetic analyses on biosynthetic pathways for amino acids and nucleotides

Reference biosynthetic pathways for amino acids and nucleotides were obtained from the Biocyc database (Karp *et al*, 2019) (Supplementary table 6). The datasets for the respective phylogenies were extensively manually curated (see Supplementary text 5 for details). Briefly, for each gene the corresponding eggNOG orthogroup was identified by a blastp search, and its composition was refined (e.g. excluding paralogs and poorly aligned sequences) by inspection of the respective alignment and of the respective single-gene tree (Nguyen *et al*, 2015). In order to get more phylogenetically informative datasets for more robust and reliable inferences, we concatenated together the sequences involved in the same pathway, as well as in pathways sharing common reactions. Under similar assumptions, single (or two) gene pathways were not employed in the analyses. We noticed some potential "false positives", namely organisms

(including *Rickettsiales*) displaying only few genes of a given pathway. Alternatively, the apparently missing steps might be "filled" by additional non-specific enzymes, as hypothesised for other pathways in *Rickettsiales* (Driscoll *et al*, 2017). Anyway, due to the "unbalanced" availability of sites, those cases could potentially hamper the accuracy of the phylogenetic inference. Therefore, for each concatenated alignment we opted for two alternative strategies in parallel, namely phylogeny on "full organism dataset", or on a "selected organism dataset", keeping only those organisms displaying at least 50% of the included genes, or, alternatively, at least a significant proportion of selected sub-branches of the pathway (Supplementary table 6). Genes were aligned, trimmed and concatenated as described above, and each concatenated alignment was also processed as described above ("Full *Rickettsiales* dataset construction and phylogenomic analyses"**)** in order to account for compositional heterogeneity (Muñoz-Gómez *et al*, 2019). On each resulting alignment, phylogeny was inferred with IQ-TREE and the LG+C60+F+R6 model, as described above.

Identification of amino acid transporters

All the proteins of the 113 *Rickettsiales* of our final dataset were blasted against the full TCDB database (Saier *et al*, 2021). Then, for each *Rickettsiales* genome the number of proteins having a best significant hit (e-value threshold of 1e-5) on each selected entry were counted (see Supplementary text 5 for details on the selection and refinement of TCDB entries representing putative amino acids transporters and on the rationale for the blast search).

Identification and phylogenetic analysis of the tlc nucleotide translocases

Analyses were focused on the tlc nucleotide translocase transporters (see Supplementary text 5 for details), common in *Rickettsiales* and in other host-associated lineages (Major *et al*, 2017). Briefly, the corresponding eggNOG orthogroup was identified by a blastp search. Then, it was joined together with a selection of the phylogenetic dataset of tlc translocases by Major and co-authors (Major *et al*, 2017), corresponding to the clade of sequences consisting only of the nucleotide transport protein domain. The sequences were then aligned and trimmed as described above, and phylogeny was inferred as described above, employing the LG+C60 model as in (Major *et al*, 2017).

Identification of genes involved in the interaction with host cells

For getting information on the presence and multiplicity of genes involved in multiple features of the interaction of *Rickettsiales* with host cells, the VFDB core reference database was employed (Liu *et al*, 2022)(see Supplementary text 5 for details). The VFDB is quite redundant for orthologs identified in different included pathogens. Thus, in order to make it suitable for analyses on our non-model bacteria, for each VFC (Virulence Factor Class) separately, orthologs were identified within the database with OrthoFinder 2.5.4 (Emms & Kelly, 2019), and manually curated. Then, all proteins of our dataset of *Rickettsiales* were blasted on the full VFDB core database. For each *Rickettsiales* genome, proteins were counted as "assigned" to each curated orthogroup if displaying a significant (evalue 1e-5) best blastp hit on any sequence belonging to the orthogroup.

5.3 Results

5.3.1 Novel genomes

In this work, we obtained nine complete genome sequences of *Rickettsiales* bacteria (belonging to nine species, eight genera, two families). These represent the first sequences for the respective species, with the exception of *Megaira polyxenophila* (Davison *et al*, 2022) . Thus, the evolutionary representativity of available *Rickettsiales* genomes results significantly improved, also considering that all the newly sequenced organisms are hosted by ciliates or other protists. All the assemblies were highly curated, resulting in most cases in a very high quality (or even full closed, with five genomes having L50=1, and one L50=2) (Supplementary table 1). In four cases, the quality of the assembly allowed to clearly determine the presence of plasmids (respectively in two *Rickettsiaceae* and one among *Midichloriaceae*, Supplementary table 1).

5.3.2 Phylogeny

For the successive analyses, we aimed to capture and analyse the widest available diversity of *Rickettsiales* from published sequences, including under-explored (e.g. *Deianiraeaceae*) and possibly yet uncharacterised lineages. To do so, together with a representative set of 36 available *Rickettsiales* genomes, we selected a total of "high-quality" 314 potential *Rickettsiales* metagenome-assembled-genomes (MAGs) from various sources. Their affiliation to *Rickettsiales* was tested by a multi-step phylogeny-based approach, taking advantage of multiple organismal selections tailored on each different MAG lineage to be investigated, as well as of a site-selection approach to counterbalance compositional heterogeneity (Muñoz-Gómez *et al*, 2019). After filtering out phylogenetically redundant MAGs, we identified 68 *Rickettsiales* MAGs, ending up with 113 total *Rickettsiales* for the final analysis, including the nine novel high-quality genomes.

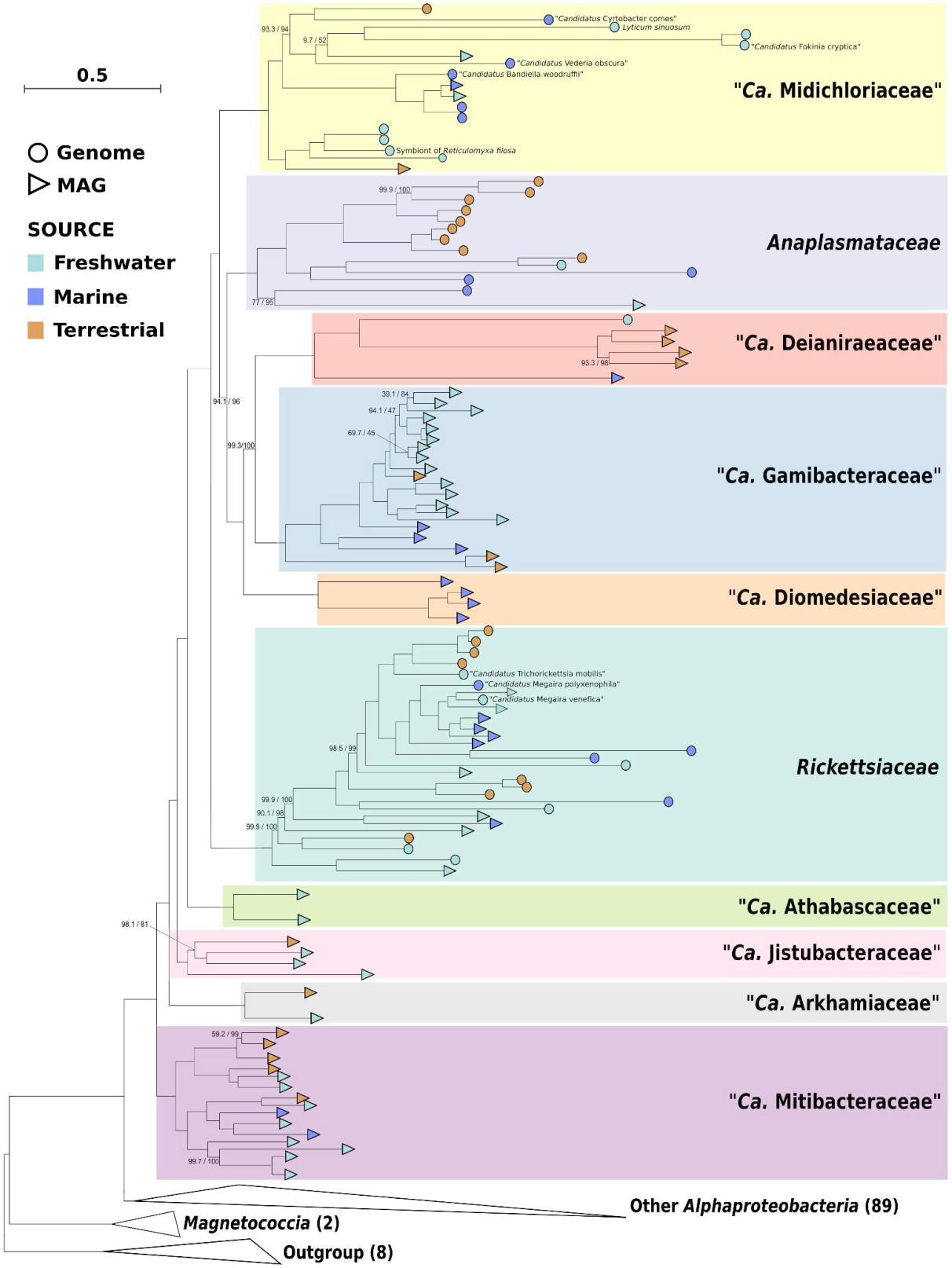


Figure 25: Maximum likelihood phylogenomic tree of 113 Rickettsiales *inferred on 179 concatenated orthologs. Each* Rickettsiales *family is highlighted by a differently coloured box. At tips, round shapes indicate genome assemblies, while triangular shapes indicate metagenome-assembled-genomes (MAGs). Shape fillings show the sample source, namely light blue for freshwater, dark blue for marine, and orange for terrestrial. Due to space constraints, only the names of the nine newly obtained genome assemblies were reported, and non-*Rickettsiales *lineages (including other* Alphaproteobacteria*,* Magnetococcia*, and outgroup) are represented by collapsed triangular shapes, with the respective number of organisms*

reported (full tree is shown in Supplementary figure 1). On each branch, support values by SH-aLRT with 1000 replicates and by 1000 ultrafast bootstraps are reported (full support values were omitted). The tree scale stands for estimated sequence divergence. Ca. is an abbreviation for Candidatus.

The final *Rickettsiales* topology resulted quite robust, with most nodes finding full support, and only eight nodes below the bootstrap threshold of 90% (Figure 25; Supplementary figure 1). The four families that include at least one characterised organism were highly supported, and mostly with a much increased representativity, thanks to novel genomes sequenced and MAGs identified in this study (novel taxa: 14/29 *Rickettsiaceae*; 10/17 *Midichloriaceae*; 1/14 *Anaplasmataceae*; 5/6 *Deianiraeaceae*). The inner relationships within *Rickettsiaceae* and *Anaplasmataceae* are overall consistent with previous phylogenetic and phylogenomic studies (Castelli *et al*, 2019, 2016; Schön *et al*, 2022; Yurchenko *et al*, 2018; George *et al*, 2020; Carrier *et al*, 2021). For *Midichloriaceae*, it was possible to infer a novel inner topology with a much higher support with respect to previous studies (Giannotti *et al*, 2022; Castelli *et al*, 2016; Schulz *et al*, 2016; Gruber-Vodicka *et al*, 2019). Furthermore, it was possible to determine that the *Midichloriaceae* bacterium associated with *Plagiopyla* represents a novel genus and species (Figure 25; from now on, *Vederia obscura*, see Supplementary text 1 for taxonomic description).

The three recently described families composed only by MAGs (Schön *et al*, 2022), namely *Gamibacteraceae*, *Athabascaceae*, and *Mitibacteraceae*, were retrieved with a comparatively higher representativity (Figure 25). Besides, three further MAG-only families were identified for the first time in this study: i) *Diomedesiaceae*, ii) *Jistubacteraceae*, and iii) *Arkhamiaceae*, the first being sister group of *Deianiraeaceae*+*Gamibacteraceae* (we will name these three families together as the "DDG"-clade), while the latter two forming a sequential branching pattern between *Athabascaceae* and *Mitibacteraceae* (Figure 25; Supplementary figure 1, see Supplementary text 1 for taxonomic descriptions).

The phyletic relationships among the families are in general consistent with the most recent and comprehensive studies (Castelli *et al*, 2019; Schön *et al*, 2022), with a single partial exception. The present study indicates a closer relationship between the DDG-clade and *Anaplasmataceae* with respect to *Midichloriaceae*, consistent with 16S rRNA gene phylogenies and some phylogenomic analyses (Castelli *et al*, 2019, 2021), while another study placed *Anaplasmataceae* and *Midichloriaceae* as sister groups (Schön *et al*, 2022). It should be considered that the number of available members of the DDG-clade progressively and largely increased in successive studies (in particular, this is the first study considering *Diomedesiaceae*), with a potential significant influence on the results.

In terms of origin, the vast majority of samples derived from aquatic environments (both freshwater and marine), especially in deeper nodes of the tree (Figure 26). The same holds true within each family, with the significant exceptions of *Anaplasmataceae* and *Deianiraeaceae*, mostly derived from terrestrial environments. Many characterised hosts resulted to be protists (most abundant in all families, except for *Anaplasmataceae*, all hosted by metazoans). For all MAGs, including in particular deeply branching lineages, no conclusive information is available on potential hosts, while in few cases there is loose indication of association, e.g. as originating from rumen (Xie *et al*, 2021; Parks *et al*, 2017, 2).

5.3.3 General genome comparisons

Based on phylogeny, we hereby define as "crown *Rickettsiales*", the smallest monophylum that comprises all characterised organisms (i.e. the one including the six families *Rickettsiaceae*, *Midichloriaceae*, *Anaplasmataceae*, *Deianiraeaceae*, *Gamibacteraceae*, *Diomedesiaceae*), thus corresponding to the classical *Rickettsiales* as defined previously (Schön *et al*, 2022). Conversely, the four other early diverging families will be defined as "basal *Rickettsiales*".

Genome sizes are quite variable between and within *Rickettsiales* families (Supplementary table 2). Crown *Rickettsiales* genomes are mostly in the range 1-1.5 Mb, with some appreciable differences between families, in particular on average *Anaplasmataceae* (1.1 Mb) and *Deianiraeaeaceae* (1.0 Mb), are smaller than others (all averages ≥1.3Mb). Genomes from basal families are much larger (frequently

>2 Mb, and on average ≥1.8 Mb, except *Arkhamiaceae*). It should be taken into account that the likely incompleteness of some MAGs (Supplementary table 3) may influence the average size estimates, especially in DDG-clade and basal families. GC content is in general inversely correlated with genome size (32-36% on average in classical families and 40-52% in basal ones, with a maximum of 61%, in *Mitibacteraceae*).

As expected, gene numbers are consistent with respective genome sizes (Supplementary table 2). Our analyses supported the notions from previous studies that *Rickettsiales* have globally experienced genome reduction trends (Supplementary figure 2), as a putative consequence of adaptation and specialisation to host-associated lifestyles (Driscoll *et al*, 2017; Min *et al*, 2008; Castelli *et al*, 2019; Schön *et al*, 2022; Salje, 2021). In order to investigate on the origin and evolution of such interactions with host cells from a functional and metabolic perspective, we selected a number of relevant traits/functions, in particular biosynthesis and uptake of metabolic precursors (i.e. amino acids and nucleotides), secretion/adhesion/motility apparatuses and putative effector molecules. A detailed overview of gene content variation and evolution in *Rickettsiales*, with a special focus on general "family-level" trends, as well as on the single newly characterised genomes, is presented in (Supplementary text 3).

5.3.4 Secretion, attachment, and motility

Secretion systems are among the components that may exert a central role in regulating interactions with host cells (Green & Mecsas, 2016), potentially enabling specific stages of the bacterial life cycle through the delivery of effectors. In *Rickettsiales*, the hallmark apparatus is type IV secretion system (Supplementary figure 3), representing a probable ancestral horizontal acquisition (Gillespie *et al*, 2010), and a possible prerequisite for establishing interactions with host cells (Schön *et al*, 2022). Only few genomes among *Rickettsiales* are devoid of this apparatus (George *et al*, 2020; Floriano *et al*, 2022), and few others display an incomplete gene set, possibly indicative of ongoing loss, in particular *Bandiella* and the *Midichloriaceae* symbiont of *Reticulomyxa* (Supplementary figure 3). Type VI secretion system is very rare, being found only in two *Rickettsiaceae* (Supplementary figure 3), in both cases encoded on plasmids, a probable indication of horizontal acquisition. In *Sneabacter* this system has possibly functionally replaced the type IV secretion system (George *et al*, 2020), while in *Trichorickettsia* both systems coexist.

Among putative secreted effector proteins (Supplementary figure 4), the "repeat-bearing" ones (ankyrin, tetratricopeptide, leucine-rich or pentapeptide repeats) are overall abundant and enriched in crown families, with many lineage-specific patterns, but are also present in basal families. On the other side, RTX toxins (Linhartová *et al*, 2010; Benz, 2016) are quite abundant in basal families, and uncommon in crown ones. Interestingly, proteins involved in the intracellular invasion of eukaryotes, such as hemolysins, patatin-like and other phospholipases (Borgo *et al*, 2022), are common in some crown families such as *Rickettsiaceae* and *Midichloriacae*, and not uncommon in basal families, while they are rare in DDG clade, especially *Deianiraeaceae* (Supplementary figure 4). Several other putative toxins/effectors, characterised in *Rickettsiales* (Niu et al. 2010) and/or in other bacteria (Lobato-Márquez *et al*, 2016; Davison *et al*, 2022; Alix *et al*, 2012; Nwasike *et al*, 2016; Billington *et al*, 2000; Padmalayam *et al*, 2000; Swart *et al*, 2020; Veyron *et al*, 2018), were found more rarely, showing patterns of presence/absence that appear to be quite lineage-specific, and without sharp differences between basal and crown *Rickettsiales*.

Flagellum might be important especially during horizontal transmission in *Rickettsiales* (Sassera, 2011; Schulz *et al*, 2016). Flagellar genes are common in basal *Rickettsiales* (Supplementary figure 5), likely representing ancestral traits (Sassera, 2011). Conversely, they are absent in the DDG-clade, and very rare in *Anaplasmataceae* (found just in the aquatic *Echinorickettsia* and *Xenolissoclinum*). Within families *Rickettsiaceae* and *Midichloriaceae*, they are extremely rare in terrestrial representatives (the only cases being *Midichloria mitochondrii* and the *Rickettsiaceae* symbiont of *Amblyomma* Ac37b), but quite common in aquatic ones, in particular basal *Rickettsiaceae* and *Midichloriaceae* in general, thereby also confirming the few experimental observations of flagella (Boscaro *et al*, 2013; Lanzoni *et al*, 2019; Mironov & Sabaneyeva, 2020). The poor correlation of the presence of flagellar genes with *Rickettsiales*

phylogeny (including differential cases within the same genus, such as *Megaira* and *Midichloria* (Floriano *et al*, 2022)) would be indicative of multiple independent reduction/loss events (Supplementary figure 5). Similar considerations may hold for chemotaxis, which possibly works in conjunction with flagella for host targeting (Keegstra *et al*, 2022), and is present only in basal *Rickettsiales* and in the basal members of the family *Rickettsiaceae* (Supplementary figure 5).

Type 4 pili may be involved in the adhesion/attachment to host cells in *Rickettsiales* . Their components are present in basal *Rickettsiales*, in the DDG clade, and only rarely in the other crown families, especially in the respective early-divergent and/or aquatic representatives (Supplementary figure 6). Thus, this apparatus was likely ancestral, experiencing multiple independent losses in crown *Rickettsiales*.

Proteins homologous to the FhaBC two-partner secretion system, which were tentatively linked to the attachment and toxicity towards host cells in *Rickettsiales* (Castelli *et al*, 2019) and may also participate in bacterial competition (Guérin *et al*, 2017), were likely ancestral and lost multiple times among *Rickettsiales*, being present in the basal families, *Deianiraeaceae*, *Gamibacteraceae*, and very few representatives of the other crown families (Supplementary figure 6).

For what concerns proteins characterised to be involved in adherence/invasion of host cells (Sears *et al*, 2012; Kahlon *et al*, 2013; Seidman *et al*, 2014) or even immune evasion (Park *et al*, 2003) in *Rickettsiales*, they are present (almost) exclusively in subgroups of the respective families (Supplementary figure 6).

5.3.5 Nucleotide and amino acid metabolism

Nucleotide biosynthesis is stably present (both purines and pyrimidines) in basal *Rickettsiales* (Figure 26; Supplementary figure 7), while, differently from other biosynthetic pathways (Supplementary text 3), its presence in crown *Rickettsiales* is “scattered” along the organismal phylogeny. Indeed, it is ubiquitous in the families *Gamibacteraeae*, *Diomedesiaceae*, *Anaplasmataceae*, present in the earliest diverging *Deianiraeaceae* MAG and in few basal *Rickettsiaceae* (in particular the endosymbiont of *Amblyomma* Ac37b, able to synthesise both purines and pyrimidines), but absent in *Midichloriaceae* and in the remaining and most numerous *Rickettsiaceae* and *Deianiraeaceae*. Phylogenetic analyses of these pathways indicate that they are likely ancestral in *Rickettsiales*, with quite good correspondence with organismal phylogeny at various levels (up to and even above the families) (Supplementary figure 8).

In those crown *Rickettsiales* lacking nucleotide biosynthesis, this absence is counterbalanced by the ability to obtain final products (or intermediates) from their hosts. Indeed, the presence/absence pattern of tlc nucleotide translocases almost perfectly inversely correlates with nucleotide biosynthesis (Figure 26; Supplementary figure 7). These family of transporters include chloroplastic ATP/ADP translocases, as well as a vast array of proteins, able to translocate several different nucleotides (Audia & Winkler, 2006; Daugherty *et al*, 2004), and previously reported to have experienced multiple horizontal gene transfer events between phylogenetically unrelated host-associated bacteria (Schmitz-Esser *et al*, 2004; Major *et al*, 2017).

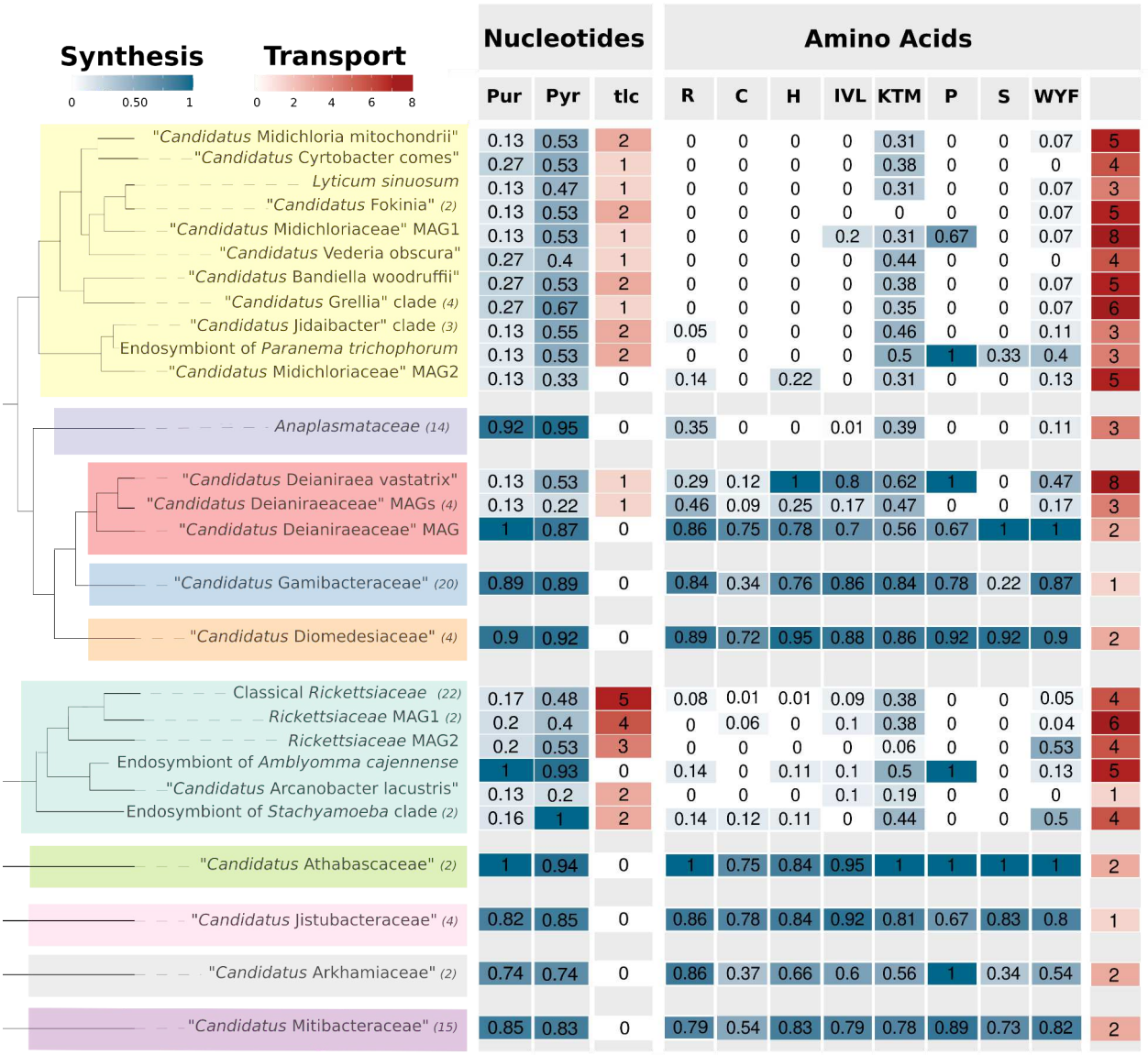


Figure 26: Heat-map showing the presence and abundance of biosynthetic pathways (blue) of nucleotides (purines and pyrimidines) and amino acids (grouped by to their mutually shared enzymatic steps according to BioCyc (Karp et al, 2019)), as well as their respective transporters (red). For biosynthesis, the proportion of the total genes of the pathway is shown (Supplementary table 6), while for transporters, the number of genes is reported, in particular for amino acid transporters the sum of the "characterised hits" (Supplementary figure 10). A cladogram of the organisms is shown on the left, with each Rickettsiales *family highlighted by a differently coloured box. Due to space constraints, selected monophyletic clades with homogeneous gene content were collapsed. For each clade, the number of organisms is shown in brackets (if higher than one), and reported values are averaged, while the complete organism sets are shown in (Supplementary figure 7, 10).*

Comparison of tlc and organismal phylogenies (Supplementary figure 9) indicates that these transporters were acquired multiple independent times by different *Rickettsiales* lineages (up to ten, once among *Deianiraeaceae*, three-five among *Midichloriaceae*, three-four among *Rickettsiaceae*, see Figure 27). We also detected multiple independent events of duplication leading to several paralogs, namely five copies in classical *Rickettsiaceae*, two-four in three distinct basal *Rickettsiaceae* sublineages, and two copies in the *Jidaibacter* lineage among *Midichloriaceae* (Figure 26 and 27; Supplementary figure 7, 8).

The presence/absence pattern of biosynthetic pathways for amino acids shows significant analogies with what we detected for nucleotides (Figure 26; Supplementary figure 10). Indeed, they are quite uniformly present in basal *Rickettsiales*, *Diomedesiaceae*, *Gamibacteraceae*, and partly *Deianiraeceae*. Conversely, they are very rare or fully absent in the non-monophyletic assemblage of *Rickettsiaceae*, *Midichloriaceae* and *Anaplasmataceae*, with very few exceptions, such as arginine in *Ehrlichia* and *Neoehrlichia* (and partly *Anaplasma*), and proline in endosymbionts of *Amblyomma* Ac37b and *Peranema* (Supplementary figure 10). Our phylogeny indicates an overall vertical descendance of these pathways, with exceptions of possible horizontal transfer events with some *Rickettsiales* as recipients for genes belonging to the pathways of cysteine and histidine (Supplementary figure 8).

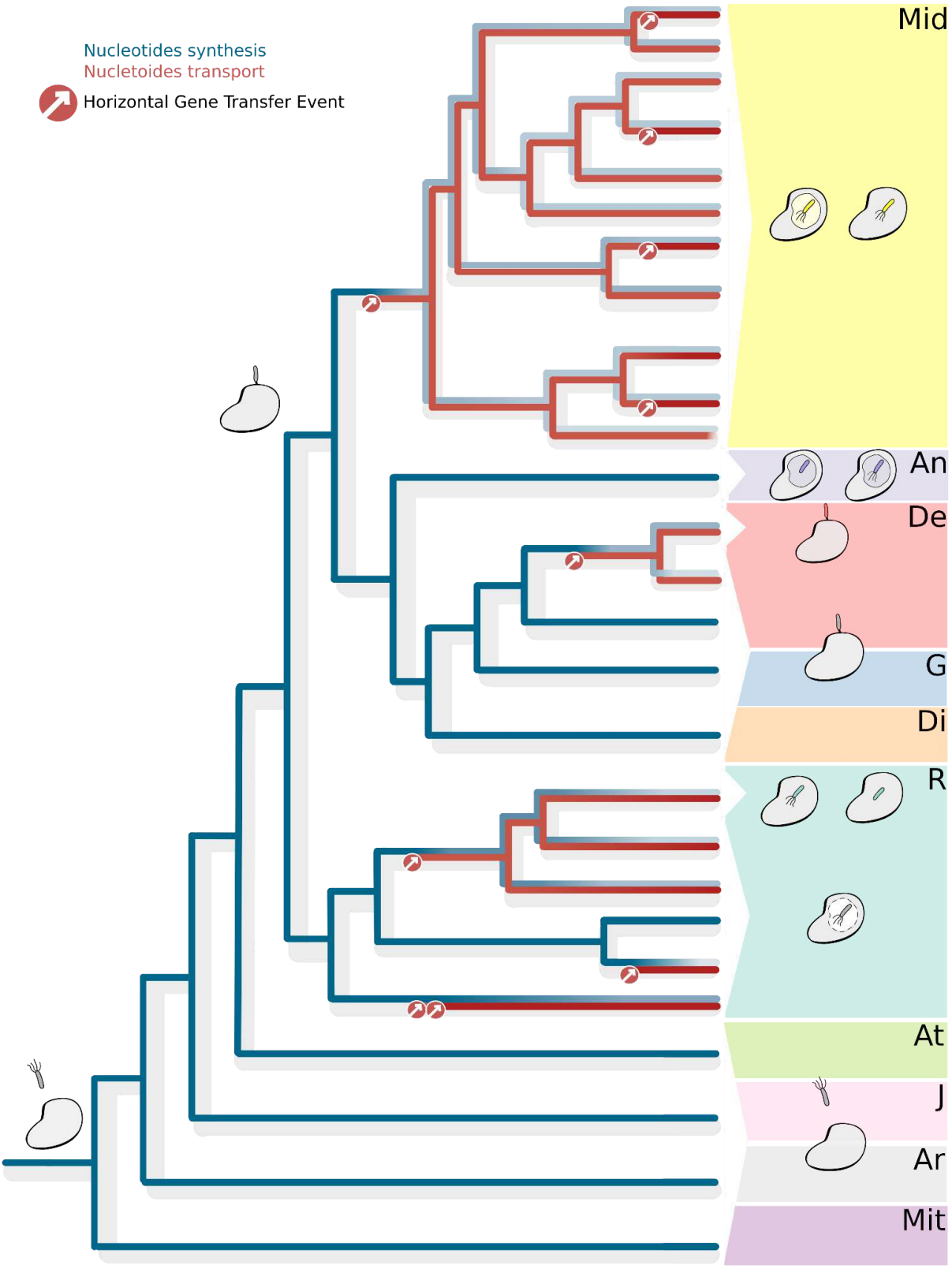


Figure 27: Reconstruction of the main steps of the evolution of Rickettsiales *with a specific focus on the interactions with and dependence on eukaryotic cells. A cladogram of the main lineages (as represented in Figure 26) is shown. The case of the biosynthetic pathways (averaged for purines and pyrimidines, blue) and tlc transporters (red) for nucleotides is represented along the tree by a heat-map-like representation, showing the inferred ancestral conditions and hypothesised steps of variation on each branch. In particular, multiple independent acquisitions of transporters by horizontal gene transfer events*

(red circles with inward arrows) would have led to the progressive reduction/loss of the biosynthesis. Rickettsiales families are highlighted by coloured boxes and by abbreviated names (R: Rickettsiaceae; Mid: Midichloriaceae*; An:* Anaplasmataceae*; De:* Deianiraeaceae*; G:* Gamibacteraceae*; Di:* Diomedesiaceae*; At:* Athabascaceae*; J:* Jistubacteraceae*; Ar:* Arkhamiaceae*; Mit:* Mitibacteraceae*). At tips, groups of tips, and at selected nodes, the drawings represent the known (classical* Rickettsiaceae*,* Anaplasmataceae*,* Midichloriaceae*, and* Deianiraea*) or hypothesised (all other lineages) features of the bacteria and their interaction with eukaryotic hosts, in particular, intracellular or extracellular associations, or lack of association, as well as presence/absence of a vacuole and of flagella. Multiple side-by-side pictures represent alternative conditions/reconstructions for the same organisms.*

The array of putative amino acids transporters is more complex than that for nucleotides, considering the higher numerosity of amino acids and the multiple independent transporter families with variable substrate specificities (Saier *et al*, 2021; Burkovski & Krämer, 2002). This impairs precise homology-based prediction of substrate specificity in *Rickettsiales,* making it impossible to define with certainty which amino acid is imported by which transporter. Nevertheless, the total number of transporters is much higher in the three families that are more deprived in biosynthesis (Supplementary figure 10). This is especially true for those with a higher similarity with ascertained amino acid transporters (Figure 26). Moreover, as seen for nucleotide transporters, the presence of homologs of different amino acid transporters is scattered along the *Rickettsiales* phylogeny (Supplementary figure 10).

We believe that the above presented data clearly indicate a pattern of multiple independent and successive acquisitions of nucleotide and amino acids transporters in different lineages of *Rickettsiales*, events that would have enabled the recipients to efficiently acquire those compounds from their hosts, thus leading to the reduction and eventual loss of the respective biosynthesis, an evolutionary scenario consistent with the intracellularity late hypothesis (Castelli *et al*, 2019).

Discussion

The knowledge and understanding of the evolution of the typically host-associated *Rickettsiales* (Salje, 2021; Castelli *et al*, 2016), in particular its earlier steps, has been hampered by the limited and phylogenetically unbalanced set of genomes available. Here we present a dataset of over one hundred phylogenetically-diverse *Rickettsiales* assemblies, thanks to *de novo* sequencing of nine high-quality genomes from underexplored protist-associated representatives, and to an accurate selection of published genomes and MAGs. We thus obtained an unprecedented representativity of all known families, including those recently described (Schön *et al*, 2022)., and identified three further ones, for a total of ten families in *Rickettsiales*. Leveraging such an extended taxonomic resolution, we investigated whether the obligate association with eukaryotic hosts and specifically intracellular lifestyle were “early” conditions with a single origin, or “late” achievements that evolved multiple times independently in different *Rickettsiales* sublineages.

It is a generally accepted notion that metabolic dependence on the hosts is a key feature in obligate associations such as those involving *Rickettsiales* (Driscoll *et al*, 2017; Min *et al*, 2008; Schön *et al*, 2022; Castelli *et al*, 2019), that evolved as the consequence of the possibility to efficiently acquire metabolites (including precursors and, for energy, ATP). Such ability is likely due to the acquisition of suitable transporters, making the respective biosynthetic and catabolic pathways dispensable, thus leading to their reduction and eventual loss. The pattern of gain of transporters and loss of synthesis pathways can thus be a strong indicator of the state of host dependence through evolution. Based on the herein produced dataset and analyses, the case of nucleotide synthesis and transport is noteworthy among *Rickettsiales*. Indeed, our analysis indicate that the most likely scenario is one of multiple independent horizontal acquisitions (up to ten) of tlc transporters among crown *Rickettsiales*, likely “triggering” independent losses of the ancestral biosynthetic pathways (Figure 26 and 27; Supplementary figure 7, 8, 9). Similar considerations hold for amino acids, even though the impossibility to predict the precise specificity of all transporters impairs a clear reconstruction of single events leading to the multiple independent losses of the ancestral biosynthetic pathways (Figure 26 and 27, Supplementary figure 8, 10).

Besides these more clear-cut cases, detailed analyses of the presence/absence patterns among genes involved in multiple other pathways strongly indicate analogous processes of gradual and independent reduction/losses in different crown *Rickettsiales* sublineages (Supplementary text 3). Sharp differences are present even within single families, such as in the metabolically rich basal *Rickettsiaceae*, as compared to the classical streamlined ones. Thus, in contrast with the more traditional views (Schulz *et al*, 2016; Schön *et al*, 2022; Weinert *et al*, 2009), our analyses provide a clear indication that processes of pathway reduction/loss have not taken place just once in *Rickettsiales*, but instead occurred (and are still possibly occurring) multiple independent times in different crown *Rickettsiales* lineages, also in relation with the host features.

Secretion systems and attachment/invasion molecules are other paramount bacterial components for promoting and actively regulating interactions with host cells(Gillespie et al. 2015, 2010; Green and Mecsas 2016). Interestingly, we found that in crown *Rickettsiales* the repertoire for such systems, as well as for flagellar apparatus, is a substantial subset of the one of basal *Rickettsiales*. This result fortifies previous notions that the ancestral *Rickettsiales* already possessed a quite rich set of proteins to interact with (unicellular) eukaryotes (Gillespie *et al*, 2010; Schön *et al*, 2022). It seems reasonable to hypothesise that such an arsenal could have represented a prerequisite for the establishment of associations with eukaryotic cells, possibly through a partial process of repurposing (Schön *et al*, 2022). (e.g. delivery of effectors molecules active on eukaryotes, motility and chemotaxis involved in horizontal transmission). In this regard, type IV secretion is a good candidate (Gillespie *et al*, 2010; Schön *et al*, 2022), also considering its almost full conservation among *Rickettsiales* (Supplementary figure 3). Other apparatuses, e.g. flagellum and type IV pilus, are more phylogenetically scattered among crown *Rickettsiales* (Supplementary figure 5, 6), likely as a result of multiple independent losses as well. This indicates unique patterns of specialisations along the *Rickettsiales* evolution, with the concurrent lineage-specific losses of traits that were dispensable for the interaction with respective host cells. Interestingly, the correlation with other traits allows the inference of some functional links, such as flagella and chemotaxis in aquatic environments, likely involved in non host-associated stages such as horizontal transmission (Schulz *et al*, 2016), pili for extracellular attachment to host cells in *Deianiraea* (Castelli *et al*, 2019) and possibly other members of the DDG clade. At the same time, alternative/additional functions for these apparatuses due to their homologies with secretion systems (Mattick, 2002; Abby & Rocha, 2012) could be possible, and may account for the exceptions to such correlation patterns (Sassera, 2011; Floriano *et al*, 2022).

Conversely, specialisation in the interaction with host cells has likely implied the expansion of other gene families, in particular those of putative secreted effectors such as the "repeat-containing" ones or acquisition/development of novel ones. In particular, several proteins characterised in pathogenic *Rickettsiales* (e.g. *Rickettsia*, *Anaplasma*) for their direct involvement in the interaction with the host (Sears *et al*, 2012; Seidman *et al*, 2014; Kahlon *et al*, 2013; Park *et al*, 2003; Niu *et al*, 2010), resulted to be lineage-specific (at the family level or even below) rather than conserved in *Rickettsiales* as a whole. Thus, it seems likely that many other still uncharacterised lineage-specific proteins could exist in the other much less investigated *Rickettsiales* (e.g. *Midichloriaceae*, DDG clade). Such a scenario of lineage-specific sets of "interactors" suggests that the mechanisms and conditions of host-association have evolved independently among different (crown) *Rickettsiales* lineages along with their molecular players.

It is more difficult to precisely infer the condition of basal *Rickettsiales* (lacking any experimental data) in terms of potential interaction with eukaryotes. These bacteria are metabolically rich, suggesting independence from possible host cells (Schön *et al*, 2022). At the same time, they bear basically all the putative prerequisite apparatuses for the interaction, and, most significantly, many are also equipped with homologs of effectors typical of crown *Rickettsiales*, such as phospholipases (Borgo *et al*, 2022) and the "repeat-containing" effectors, and they even are selectively enriched of additional potential effectors (Linhartová *et al*, 2010; Benz, 2016)). Conversely, free-living-like traits such as inorganic nutrient transport and detoxification, previously found only among basal *Rickettsiales* (Schön *et al*, 2022)., were retrieved also in some crown representatives (Supplementary text 3). This may be seen as indicative that

those crown *Rickettsiales* retain some "primitive" traits, and, more in general, as another hint at more complex evolutionary trajectories than a simple transition towards obligate association at the root of crown *Rickettsiales*.

Taken together, the above presented data clearly indicate that obligate host-association, as well as intracellularity, were most likely "late" conditions in *Rickettsiales*. Under such a scenario, we propose that at some point in the early *Rickettsiales* evolutionary history their presumably aquatic free-living ancestors were engaged in some kind of facultative interaction with eukaryotes. The starting point could have been defence from protist predators through the release of active effectors, as previously hypothesised (Schön *et al*, 2022). It is possible to envision that such defence mechanisms successively paved the way for the (gradual) development of the (at least occasional) ability to gain further advantages by such interactions, such as the capability to acquire metabolites from the damaged/killed eukaryotes, somehow reversing and taking control of the predator-prey interactions. The lifestyle of *Deianiraea* could be reminiscent of this hypothetical condition (Castelli *et al*, 2019). Most likely, such transition towards facultative associations would have occurred prior to the last common ancestor of crown *Rickettsiales*, which are all obligatorily host-associated.

Conversely, it is not straightforward to precisely place an upper bound for such transition, given the complete lack of direct information on the lifestyles of extant basal *Rickettsiales*. It could have occurred sharply in the common ancestor of crown *Rickettsiales*, or could have been more nuanced, involving also the ancestors of some basal lineages, possibly up to the ancestor of all *Rickettsiales*. Nevertheless, it cannot be excluded that any potential host-associated representative of basal *Rickettsiales* could be the result of a convergent and independent evolution with respect to crown *Rickettsiales*.

In any case, for what concerns crown *Rickettsiales*, we can envision that a single initial facultative association would have differentiated in the descendants, becoming tighter and tighter (and at some point obligate) through parallel successive steps of acquisition/development of metabolite transporters and interactor molecules. Such further transitions would have occurred separately and independently in different *Rickettsiales* lineages, thus supporting the conclusion of a late origin for the obligate association with hosts. The order and kind of such steps, although somehow similar, would have not been unique in each of the different crown *Rickettsiales* phyletic lines, as reflected in the present-day lineages, which exhibit differential features of metabolic dependence (e.g. for most amino acids but not nucleotides in *Anaplasmataceae*, and vice versa in most *Deianiraeaceae*), as well as differential mechanisms and conditions of interaction (chiefly intracellularity vs extracellarity).

All the above would specifically support the intracellularity late hypothesis, as this trait would have evolved multiple independent times and with differential features in some crown *Rickettsiales* (at least the *Rickettsiaceae*, *Midichloriaceae*, and *Anaplasmataceae* ancestors, with the exclusion of *Deianiraea* and possibly the whole DDG clade). An alternative reconstruction which cannot be excluded is that the ancestor of crown *Rickettsiales* had already undergone an early transition towards obligate association through some yet unidentified step(s). Even in such a case, analysed data would still robustly support the reconstruction that most further genome reduction/adaptation steps would have occurred later and independently, so that in particular intracellularity would be anyway designated as a late condition.

Our reconstruction may also provide a novel perspective on the origin and evolution of other host-associated bacterial lineages that, similarly to *Rickettsiales*, present the prerogative to "hold the control" of the interaction and to switch hosts by horizontal transmission, and were thus termed "professional symbionts" (Husnik & Keeling, 2021) (e.g. *Chlamydiae* (Dharamshi *et al*, 2020)), *Legionellales* (Hugoson *et al*, 2022) and *Holosporales* (Muñoz-Gómez *et al*, 2019). These lineages could share similarities in the initial establishment and successive stepwise and "late" evolutionary development.

From a more general perspective, it seems worthily a comparison of *Rickettsiales* (and possibly other "professional symbionts") with the more "canonical" genome evolution models among obligate symbionts, namely nutritional mutualists in insects (with some parallels also in protists (Husnik &

Keeling, 2021)). Such symbionts undergo relatively rapid streamlining as a result of an initial host restriction, followed by a more or less prolonged stasis, and are somehow "doomed" to extinction after replacement. Conversely "professional symbionts" would be the controllers of the interaction from its evolutionary beginning, retaining the ability to horizontally change hosts, and undergoing much more gradual and "flexible" streamlining processes, depending also on the external environmental conditions and not just on the features of a single host. Interestingly, the abundance of lineages preferentially associated with marine invertebrates and showing poor co-cladogenenesis with their hosts (Boscaro *et al*, 2022, 20) indicates that there may be more bacteria sharing traits of professional symbionts than currently recognised.

Large-scale comparative genomic analyses such as those presented here and elsewhere (Hugoson *et al*, 2022; Schön *et al*, 2022; Dharamshi *et al*, 2020) have huge potential to provide major advances in our understanding of functional traits and the underlying evolutionary processes. However, they also face inherent predictive limits, being quite suitable for deriving metabolic dependencies, and less for the inference of more complex and possibly not-yet-documented traits, such as mechanisms of interaction and subcellular (or extracellular) localisation. For example, it could have been burdensome and highly speculative to infer the so far uniquely observed extracellular condition in *Deianiraea* (Castelli *et al*, 2019) (representing the first basis to propose the "intracellularity late" hypothesis) only from its genome. Therefore, considering that we still completely lack any experimental data for six out of ten *Rickettsiales* families (including all basal ones), we strongly invoke the need for further experimental investigations. As in other cases (Imachi *et al*, 2020), these may provide additional and otherwise unpredictable insights on the lifestyle of present-day organisms, and represent the basis for refining existing hypotheses and inferring novel ones on *Rickettsiales* evolution.

VI Conclusions

My work analysed different aspects of the symbiotic interactions between mutualistic bacteria, their hosts, and other bacteria associated with the same host.
I performed four different projects, mostly focused on symbiosis in ticks. The first one is the characterisation of a *Coxiella* endosymbiont in a tick host (*Amblyomma nuttalli*), and a comparative analysis of the CEs, which are typically two-partner symbioses. As a result, I saw the effect of a parallel genome reduction, with indications of multiple independent events of adaptation from a pathogenic lifestyle to a mutualistic one. Also, observed a typical example of the model of genome erosion in nutritional symbionts.
The second project considered a multi-partner relationship between the tick *Hyalomma marginatum* and two different bacteria, *Midichloria* and *Francisella* endosymbionts. In this case, the combination of selective pressures and Muller's Ratchet, ended up locking together the two different bacteria and their host in a mutualistic relationship. The resulting scenario of multiple symbionts locked in an obligatory co-presence is similar to what is observed in various nutritional symbionts in aphids, but the first one described in ticks.
The third project, still ongoing, has the aim to characterise the structure of the symbiont population in *Ixodes frontalis*. Other than *Midichloria*, common in the *Ixodes* genus, is also present *Spiroplasma*, a symbiont that some lineages of ticks acquire horizontally, and we are exploring possible differences in the symbiont presence in relation to the two sub-lineages of *I frontalis*.
To sum up, I investigated instances of different forms of symbiosis in ticks, characterising different evolutionary histories of the development and loss of mutualistic interactions, including loss of pathogenicity, competition and replacement between symbionts or interaction between them.
Finally, the broader and comprehensive evolutionary history of bacteria-eukaryote relationship was the focus of the last project of my PhD: I obtained a comprehensive evolutionary view of the order *Rickettsiales*, to which are affiliated many of the aforementioned tick-associated bacteria investigated in the other projects on my PhD, and reconstructed the deep phylogeny of this lineage, inferring the origin from non-host-dependent symbionts, with numerous adaptations to a wide range of hosts.

VII References

Abby SS & Rocha EPC (2012) The non-flagellar type III secretion system evolved from the bacterial flagellum and diversified into host-cell adapted systems. *PLoS Genet* 8: e1002983

Aboul-Nasr AE & Bassal TT (1972) Biochemical and physiological studies of certain ticks (Ixodoidea). Effect of mating, feeding, and oogenesis on oxygen consumption of Hyalomma (H.) dromedarii Koch (Ixodidae). *J Parasitol* 58: 828–831

Abraham NM, Liu L, Jutras BL, Yadav AK, Narasimhan S, Gopalakrishnan V, Ansari JM, Jefferson KK, Cava F, Jacobs-Wagner C, *et al* (2017) Pathogen-mediated manipulation of arthropod microbiota to promote infection. *Proc Natl Acad Sci* 114: E781–E790

Agoulon A, Hoch T, Heylen D, Chalvet-Monfray K & Plantard O (2019) Unravelling the phenology of Ixodes frontalis, a common but understudied tick species in Europe. *Ticks Tick-Borne Dis* 10: 505–512

Akhter SA Ramy K; Edwards, Robert (2012) PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic Acids Res* 40: e126–e126

Aksoy S (1995) Wigglesworthia gen. nov. and Wigglesworthia glossinidia sp. nov., Taxa Consisting of the Mycetocyte-Associated, Primary Endosymbionts of Tsetse Flies. *Int J Syst Evol Microbiol* 45: 848–851

Alix E, Chesnel L, Bowzard BJ, Tucker AM, Delprato A, Cherfils J, Wood DO, Kahn RA & Roy CR (2012) The capping domain in RalF regulates effector functions. *PLoS Pathog* 8: e1003012

Allen JM, Reed DL, Perotti MA & Braig HR (2007) Evolutionary Relationships of "Candidatus Riesia spp.," Endosymbiotic Enterobacteriaceae Living within Hematophagous Primate Lice. *Appl Environ Microbiol* 73: 1659–1664

Alonge MS Sebastian; Ramakrishnan, Srividya; Wang, Xingang; Goodwin, Sara; Sedlazeck, Fritz J; Lippman, Zachary B; Schatz, Michael C (2019) RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biol* 20: 224–224

Al-Tawfiq JA & Memish ZA (2017) Alkhurma hemorrhagic fever virus. *Microbes Infect* 19: 305–310

Anbutsu H & Fukatsu T (2006) Tissue-specific infection dynamics of male-killing and nonmale-killing spiroplasmas in Drosophila melanogaster. *FEMS Microbiol Ecol* 57: 40–46

Andreotti R, Pérez de León AA, Dowd SE, Guerrero FD, Bendele KG & Scoles GA (2011) Assessment of bacterial diversity in the cattle tick Rhipicephalus (Boophilus) microplusthrough tag-encoded pyrosequencing. *BMC Microbiol* 11: 6

Angelakis E, Mediannikov O, Jos S-L, Berenger J-M, Parola P & Raoult D (2016) Candidatus Coxiella massiliensis Infection. *Emerg Infect Dis* 22: 285–288

Anselme C, Pérez-Brocal V, Vallier A, Vincent-Monegat C, Charif D, Latorre A, Moya A & Heddi A (2008) Identification of the Weevil immune genes and their expression in the bacteriome tissue. *BMC Biol* 6: 43

Anstead CA & Chilton NB (2014) Discovery of Novel Rickettsiella spp. in Ixodid Ticks from Western Canada. *Appl Environ Microbiol* 80: 1403–1410

Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y & Wishart DS (2016) PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 44: W16-21

Arthur DR (1965) Ticks in Egypt in 1500 B.C.? *Nature* 206: 1060–1061

Audia JP & Winkler HH (2006) Study of the five Rickettsia prowazekii proteins annotated as ATP/ADP translocases (Tlc): Only Tlc1 transports ATP/ADP, while Tlc4 and Tlc5 transport other ribonucleotides. *J Bacteriol* 188: 6261–6268

Azagi T, Klement E, Perlman G, Lustig Y, Mumcuoglu KY, Apanaskevich DA & Gottlieb Y (2017) Francisella-Like Endosymbionts and Rickettsia Species in Local and Imported Hyalomma Ticks. *Appl Environ Microbiol* 83: e01302-17

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, *et al* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol J Comput Mol Cell Biol* 19: 455–477

Bary A de (1879) Die Erscheinung der Symbiose: Vortrag gehalten auf der Versammlung Deutscher Naturforscher und Aerzte zu Cassel Trübner

Bates DM; M Martin; Bolker, Benjamin M; Walker, Steven C (2015) Fitting Linear Mixed-Effects

Models Using lme4. *J Stat Softw* 67: 1–48
Bennett GM & Moran NA (2015) Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. *Proc Natl Acad Sci* 112: 10169–10176
Ben-Yosef M, Rot A, Mahagna M, Kapri E, Behar A & Gottlieb Y (2020) Coxiella-Like Endosymbiont of Rhipicephalus sanguineus Is Required for Physiological Processes During Ontogeny. *Front Microbiol* 11
Benz R (2016) Channel formation by RTX-toxins of pathogenic bacteria: Basis of their biological activity. *Biochim Biophys Acta* 1858: 526–537
Billington SJ, Jost BH & Songer JG (2000) Thiol-activated cytolysins: structure, function and role in pathogenesis. *FEMS Microbiol Lett* 182: 195–205
Binetruy F, Buysse M, Lejarre Q, Barosi R, Villa M, Rahola N, Paupy C, Ayala D & Duron O (2020) Microbial community structure reveals instability of nutritional symbiosis during the evolutionary radiation of Amblyomma ticks. *Mol Ecol* 29: 1016–1029
Björn LO (2020) Peptidoglycan in eukaryotes: Unanswered questions. *Phytochemistry* 175: 112370
Bodnar JL, Fitch S, Rosati A & Zhong J (2018) The folA Gene from the Rickettsia endosymbiont of Ixodes pacificus Encodes a Functional Dihydrofolate Reductase Enzyme. *Ticks Tick-Borne Dis* 9: 443–449
Bogovic P & Strle F (2015) Tick-borne encephalitis: A review of epidemiology, clinical characteristics, and management. *World J Clin Cases WJCC* 3: 430–441
Bolger AL Marc; Usadel, Bjoern (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinforma Oxf Engl* 30: 2114–2120
Borgo GM, Burke TP, Tran CJ, Lo NTN, Engström P & Welch MD (2022) A patatin-like phospholipase mediates Rickettsia parkeri escape from host membranes. *Nat Commun* 13: 3656
Borowiec ML (2016) AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4: e1660
Boscaro V, Holt CC, Van Steenkiste NWL, Herranz M, Irwin NAT, Àlvarez-Campos P, Grzelak K, Holovachov O, Kerbl A, Mathur V, *et al* (2022) Microbiomes of microscopic marine invertebrates do not reveal signatures of phylosymbiosis. *Nat Microbiol* 7: 810–819
Boscaro V, Schrallhammer M, Benken KA, Krenek S, Szokoli F, Berendonk TU, Schweikert M, Verni F, Sabaneyeva EV & Petroni G (2013) Rediscovering the genus Lyticum, multiflagellated symbionts of the order Rickettsiales. *Sci Rep* 3: 3305
Brenner AE, Muñoz-Leal S, Sachan M, Labruna MB & Raghavan R (2021) Coxiella burnetii and Related Tick Endosymbionts Evolved from Pathogenic Ancestors. *Genome Biol Evol* 13: evab108
Bright M & Bulgheresi S (2010) A complex journey: transmission of microbial symbionts. *Nat Rev Microbiol* 8: 218–230
Buchner P, 1886-, Buchner P & 1886- (1965) Endosymbiosis of animals with plant microorganisms Interscience
Burgdorfer W, 1925-, Anacker RL & 1926- (1981) Rickettsiae and rickettsial diseases. In Academic Press
Burkovski A & Krämer R (2002) Bacterial amino acid transport proteins: occurrence, functions, and significance for biotechnological applications. *Appl Microbiol Biotechnol* 58: 265–274
Busby AT, Ayllón N, Kocan KM, Blouin EF, de la Fuente G, Galindo RC, Villar M & de la Fuente J (2012) Expression of heat shock proteins and subolesin affects stress responses, Anaplasma phagocytophilum infection and questing behaviour in the tick, Ixodes scapularis. *Med Vet Entomol* 26: 92–102
Buysse M, Binetruy F, Leibson R, Gottlieb Y & Duron O (2022) Ecological Contacts and Host Specificity Promote Replacement of Nutritional Endosymbionts in Ticks. *Microb Ecol* 83: 776–788
Buysse MD Olivier (2021) Evidence that microbes identified as tick-borne pathogens are nutritional endosymbionts. *Cell* 184: 2259–2260
Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P & Huerta-Cepas J (2021) eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol Biol Evol* 38: 5825–5829
Capella-Gutierrez SS-M José M; Gabaldón, Toni (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinforma Oxf Engl* 25: 1972–1973
Carrier TJ, Leigh BA, Deaker DJ, Devens HR, Wray GA, Bordenstein SR, Byrne M & Reitzel AM

(2021) Microbiome reduction and endosymbiont gain from a switch in sea urchin life history. *Proc Natl Acad Sci U A* 118
Castelli M, Lanzoni O, Nardi T, Lometto S, Modeo L, Potekhin A, Sassera D & Petroni G (2021) ' *Candidatus* Sarmatiella mevalonica' endosymbiont of the ciliate *Paramecium* provides insights on evolutionary plasticity among *Rickettsiales*. *Environ Microbiol* 23: 1684–1701
Castelli M, Sabaneyeva E, Lanzoni O, Lebedeva N, Floriano AM, Gaiarsa S, Benken K, Modeo L, Bandi C, Potekhin A, *et al* (2019) Deianiraea, an extracellular bacterium associated with the ciliate Paramecium, suggests an alternative scenario for the evolution of Rickettsiales. *ISME J* 13: 2280–2294
Castelli M, Sassera D & Petroni G (2016) Biodiversity of "Non-model" Rickettsiales and Their Association with Aquatic Organisms. *Rickettsiales*: 59–91
Chastel C, Chandler L, Le Goff F, Chastel O, Tesh R & Shope R (1999) Chizé virus, a new phlebovirus isolated in France from Ixodes (Trichotoixodes) frontalis. *Acta Virol* 43: 279–283
Chen HB Paul C (2011) VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* 12: 35–35
Chong RA & Moran NA (2018) Evolutionary loss and replacement of Buchnera, the obligate endosymbiont of aphids. *ISME J* 12: 898–908
Cibichakravarthy B, Oses-Prieto JA, Ben-Yosef M, Burlingame AL, Karr TL & Gottlieb Y (2022) Comparative Proteomics of Coxiella like Endosymbionts (CLEs) in the Symbiotic Organs of Rhipicephalus sanguineus Ticks. *Microbiol Spectr* 10: e01673-21
Citelli M, Lara FA, Vaz I da S & Oliveira PL (2007) Oxidative stress impairs heme detoxification in the midgut of the cattle tick, Rhipicephalus (Boophilus) microplus. *Mol Biochem Parasitol* 151: 81–88
Clark HF (1974) The suckling mouse cataract agent (SMCA). A slow mycoplasma-like agent? *Prog Med Virol Fortschritte Med Virusforsch Progres En Virol Medicale* 18: 307–322
Conway JR, Lex A & Gehlenborg N (2017) UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33: 2938–2940
Cordaux R, Paces-Fessy M, Raimond M, Michel-Salzat A, Zimmer M & Bouchon D (2007) Molecular Characterization and Evolution of Arthropod-Pathogenic Rickettsiella Bacteria. *Appl Environ Microbiol* 73: 5045–5047
Costa IR, Thompson JD, Ortega JM & Prosdocimi F (2014) Metazoan remaining genes for essential amino acid biosynthesis: sequence conservation and evolutionary analyses. *Nutrients* 7: 1–16
Criscuolo A & Gribaldo S (2010) BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* 10: 210
Cull B, Pietzsch ME, Hansford KM, Gillingham EL & Medlock JM (2018) Surveillance of British ticks: An overview of species records, host associations, and new records of Ixodes ricinus distribution. *Ticks Tick-Borne Dis* 9: 605–614
Dantas-Torres F, Chomel BB & Otranto D (2012) Ticks and tick-borne diseases: a One Health perspective. *Trends Parasitol* 28: 437–446
Darling ACE, Mau B, Blattner FR & Perna NT (2004) Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Res* 14: 1394–1403
Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B & Flouri T (2019) ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Mol Biol Evol*
Daugherty RM, Linka N, Audia JP, Urbany C, Neuhaus HE & Winkler HH (2004) The nucleotide transporter of Caedibacter caryophilus exhibits an extended substrate spectrum compared to the analogous ATP/ADP translocase of Rickettsia prowazekii. *J Bacteriol* 186: 3262–3265
Daveu R, Laurence C, Bouju-Albert A, Sassera D & Plantard O (2021) Symbiont dynamics during the blood meal of Ixodes ricinus nymphs differ according to their sex. *Ticks Tick-Borne Dis* 12: 101707
Davison HR, Pilgrim J, Wybouw N, Parker J, Pirro S, Hunter-Barnett S, Campbell PM, Blow F, Darby AC, Hurst GDD, *et al* (2022) Genomic diversity across the Rickettsia and 'Candidatus Megaira' genera and proposal of genus status for the Torix group. *Nat Commun* 13: 2630
De Coster W, D'Hert S, Schultz DT, Cruts M & Van Broeckhoven C (2018) NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34: 2666–2669

Delsing CE, Kullberg BJ & Bleeker-Rovers CP (2010) Q fever in the Netherlands from 2007 to 2010. *Neth J Med* 68: 382–387
Dharamshi JE, Tamarit D, Eme L, Stairs CW, Martijn J, Homa F, Jørgensen SL, Spang A & Ettema TJG (2020) Marine Sediments Illuminate Chlamydiae Diversity and Evolution. *Curr Biol* 30: 1032-1048.e7
Di Lecce I, Bazzocchi C, Cecere JG, Epis S, Sassera D, Villani BM, Bazzi G, Negri A, Saino N, Spina F, *et al* (2018) Patterns of Midichloria infection in avian-borne African ticks and their trans-Saharan migratory hosts. *Parasit Vectors* 11: 106
Di Venere M, Fumagalli M, Cafiso A, De Marco L, Epis S, Plantard O, Bardoni A, Salvini R, Viglio S, Bazzocchi C, *et al* (2015) Ixodes ricinus and Its Endosymbiont Midichloria mitochondrii: A Comparative Proteomic Analysis of Salivary Glands and Ovaries. *PloS One* 10: e0138842
Douglas AE (1998) Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria Buchnera. *Annu Rev Entomol* 43: 17–37
Douglas AE (2017) The B vitamin nutrition of insects: the contributions of diet, microbiome and horizontally acquired genes. *Curr Opin Insect Sci* 23: 65–69
Drehmann M, Chitimia-Dobler L, Lindau A, Frank A, Mai S, Fachet K, Hauck D, Knoll S, Strube C, Lühken R, *et al* (2019) Ixodes frontalis: a neglected but ubiquitous tick species in Germany. *Exp Appl Acarol* 78: 79–91
Driscoll TP, Verhoeve VI, Guillotte ML, Lehman SS, Rennoll SA, Beier-Sexton M, Rahman MS, Azad AF & Gillespie JJ (2017) Wholly ! Reconstructed Metabolic Profile of the Quintessential Bacterial Parasite of Eukaryotic Cells. *MBio* 8
Dumler JS, Stephen Dumler J & Walker DH (2015) Rickettsiales. *Bergeys Man Syst Archaea Bact*: 1–1
Duron O, Binetruy F, Noël V, Cremaschi J, McCoy KD, Arnathau C, Plantard O, Goolsby J, León AAP de, Heylen DJA, *et al* (2017) Evolutionary changes in symbiont community structure in ticks. *Mol Ecol* 26: 2905–2921
Duron O, Cremaschi J & McCoy KD (2016) The High Diversity and Global Distribution of the Intracellular Bacterium Rickettsiella in the Polar Seabird Tick Ixodes uriae. *Microb Ecol* 71: 761–770
Duron O, Doublet P, Vavre F & Bouchon D (2018a) The Importance of Revisiting Legionellales Diversity. *Trends Parasitol* 34: 1027–1037
Duron O, Morel O, Noël V, Buysse M, Binetruy F, Lancelot R, Loire E, Ménard C, Bouchez O, Vavre F, *et al* (2018b) Tick-Bacteria Mutualism Depends on B Vitamin Synthesis Pathways. *Curr Biol CB* 28: 1896-1902.e5
Duron O, Noël V, McCoy KD, Bonazzi M, Sidi-Boumedine K, Morel O, Vavre F, Zenner L, Jourdain E, Durand P, *et al* (2015) The Recent Evolution of a Maternally-Inherited Endosymbiont of Ticks Led to the Emergence of the Q Fever Pathogen, Coxiella burnetii. *PLoS Pathog* 11
Duron OG Yuval (2020) Convergence of Nutritional Symbioses in Obligate Blood Feeders. *Trends Parasitol* 36: 816–825
Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5: 113
European Centre for Disease Prevention and Control (2022). Tick-borne encephalitis. ECDC: Annual epidemiological report for 2020. Stockholm: ECDC; 2022.
EFSA Panel on Animal Health and Welfare (AHAW) (2010) Scientific Opinion on Geographic Distribution of Tick-borne Infections and their Vectors in Europe and the other Regions of the Mediterranean Basin. *EFSA J* 8: 1723
Eldin C, Mélenotte C, Mediannikov O, Ghigo E, Million M, Edouard S, Mege J-L, Maurin M & Raoult D (2017) From Q Fever to Coxiella burnetii Infection: a Paradigm Change. *Clin Microbiol Rev* 30: 115–190
Elliman JR & Owens L (2020) Confirmation that candidatus Coxiella cheraxi from redclaw crayfish (Cherax quadricarinatus) is a close relative of Coxiella burnetii, the agent of Q-fever. *Lett Appl Microbiol* 71: 320–326
Emms DM & Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 20: 238
Estrada-Peña A, Mihalca AD & Petney TN eds. (2017) Ticks of Europe and North Africa Cham: Springer International Publishing

Fan L, Wu D, Goremykin V, Xiao J, Xu Y, Garg S, Zhang C, Martin WF & Zhu R (2020) Phylogenetic analyses with systematic taxon sampling show that mitochondria branch within Alphaproteobacteria. *Nat Ecol Evol* 4: 1213–1219

Felsheim RF, Kurtti TJ & Munderloh UG (2009) Genome sequence of the endosymbiont Rickettsia peacockii and comparison with virulent Rickettsia rickettsii: identification of virulence factors. *PloS One* 4: e8361

Feng Y, Napier BA, Manandhar M, Henke SK, Weiss DS & Cronan JE (2014) A Francisella virulence factor catalyses an essential reaction of biotin synthesis. *Mol Microbiol* 91: 300–314

Ferrarini MG, Dell'Aglio E, Vallier A, Balmand S, Vincent-Monégat C, Hughes S, Gillet B, Parisot N, Zaidman-Rémy A, Vieira C, *et al* (2022) Efficient compartmentalization in insect bacteriomes protects symbiotic bacteria from host immune system. *Microbiome* 10: 156

Floriano AM, Biffignandi GB, Castelli M, Olivieri E, Clementi E, Comandatore F, Rinaldi L, Opara M, Plantard O, Palomar AM, *et al* (2022) The origin and evolution of mitochondrial tropism in Midichloria bacteria. 2022.05.16.490919 doi:10.1101/2022.05.16.490919 [PREPRINT]

Foray V, Grigorescu AS, Sabri A, Haubruge E, Lognay G, Francis F, Fauconnier M-L, Hance T & Thonart P (2014) Whole-Genome Sequence of Serratia symbiotica Strain CWBI-2.3T, a Free-Living Symbiont of the Black Bean Aphid Aphis fabae. *Genome Announc* 2: e00767-14

Galeotti M, Volpatti D, Byadgi O, Beraldo P, Orioles M, Sarti M, Ciulli S & Magi GE (2021) Red mark syndrome (RMS) in farmed rainbow trout: First report of outbreaks in Bosnia and Herzegovina. *J Fish Dis* 44: 627–631

Gall CA, Reif KE, Scoles GA, Mason KL, Mousel M, Noh SM & Brayton KA (2016) The bacterial microbiome of Dermacentor andersoni ticks influences pathogen susceptibility. *ISME J* 10: 1846–1855

Galperin MY, Makarova KS, Wolf YI & Koonin EV (2015) Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res* 43: D261–D269

Garcia-Vozmediano A, Tomassone L, Fonville M, Bertolotti L, Heylen D, Fabri ND, Medlock JM, Nijhof AM, Hansford KM, Sprong H, *et al* (2022) The Genetic Diversity of Rickettsiella Symbionts in Ixodes ricinus Throughout Europe. *Microb Ecol* 84: 613–626

Gasparich GE (2002) Spiroplasmas: evolution, adaptation and diversity. *Front Biosci J Virtual Libr* 7: d619-640

George EE, Husnik F, Tashyreva D, Prokopchuk G, Horák A, Kwong WK, Lukeš J & Keeling PJ (2020) Highly Reduced Genomes of Protist Endosymbionts Show Evolutionary Convergence. *Curr Biol* 30: 925-933.e3

Gerardo NM, Altincicek B, Anselme C, Atamian H, Barribeau SM, de Vos M, Duncan EJ, Evans JD, Gabaldón T, Ghanim M, *et al* (2010) Immunity and other defenses in pea aphids, Acyrthosiphon pisum. *Genome Biol* 11: R21

Gerhart JG, Auguste Dutcher H, Brenner AE, Moses AS, Grubhoffer L & Raghavan R (2018) Multiple Acquisitions of Pathogen-Derived Francisella Endosymbionts in Soft Ticks. *Genome Biol Evol* 10: 607–615

Gerhart JG, Moses AS & Raghavan R (2016) A Francisella-like endosymbiont in the Gulf Coast tick evolved from a mammalian pathogen. *Sci Rep* 6

Giannotti D, Boscaro V, Husnik F, Vannini C & Keeling PJ (2022) The 'Other' : an Overview of the Family ' Midichloriaceae'. *Appl Env Microbiol* 88: e0243221

Giladi MA-P Neta; Levin, Itay; Levy, Liat; Mevarech, Moshe (2003) FolM, A New Chromosomally Encoded Dihydrofolate Reductase in Escherichia coli. *J Bacteriol* 185: 7015–7018

Gillespie JJ, Brayton KA, Williams KP, Diaz MAQ, Brown WC, Azad AF & Sobral BW (2010) Phylogenomics reveals a diverse Rickettsiales type IV secretion system. *Infect Immun* 78: 1809–1823

Gillingham EL, Hansford KM, Meadows S, Henney J, Wieckowski F, Hernández-Triana LM, Muscat I, Muscat J, Beckert C, Nikolova NI, *et al* (2020) Ticks on the Channel Islands and implications for public health. *Ticks Tick-Borne Dis* 11: 101405

Glöckner G, Hülsmann N, Schleicher M, Noegel AA, Eichinger L, Gallinger C, Pawlowski J, Sierra R, Euteneuer U, Pillet L, *et al* (2014) The genome of the foraminiferan Reticulomyxa filosa. *Curr Biol* 24: 11–18

Gottlieb Y, Lalzar I & Klasson L (2015) Distinctive Genome Reduction Rates Revealed by Genomic

Analyses of Two Coxiella- Like Endosymbionts in Ticks. *Genome Biol Evol* 7: 1779–1796
Gray J, Zintl A, Hildebrandt A, Hunfeld K-P & Weiss L (2010) Zoonotic babesiosis: Overview of the disease and novel aspects of pathogen identity. *Ticks Tick-Borne Dis* 1: 3–10
Greay TL; G Alexander W; Paparini, Andrea; Ryan, Una; Oskam, Charlotte L; Irwin, Peter J (2018) Recent insights into the tick microbiome gained through next-generation sequencing. *Parasit Vectors* 11: 12–12
Green ER & Mecsas J (2016) Bacterial Secretion Systems: An Overview. *Microbiol Spectr* 4
Gruber-Vodicka HR, Leisch N, Kleiner M, Hinzke T, Liebeke M, McFall-Ngai M, Hadfield MG & Dubilier N (2019) Two intracellular and cell type-specific bacterial symbionts in the placozoan Trichoplax H2. *Nat Microbiol* 4: 1465–1474
Guérin J, Bigot S, Schneider R, Buchanan SK & Jacob-Dubuisson F (2017) Two-Partner Secretion: Combining Efficiency and Simplicity in the Secretion of Large Proteins for Bacteria-Host and Bacteria-Bacteria Interactions. *Front Cell Infect Microbiol* 7: 148
Guizzo MG, Neupane S, Kucera M, Perner J, Frantová H, da Silva Vaz I, de Oliveira PL, Kopacek P & Zurek L (2020) Poor Unstable Midgut Microbiome of Hard Ticks Contrasts With Abundant and Stable Monospecific Microbiome in Ovaries. *Front Cell Infect Microbiol* 10: 211
Guizzo MG, Parizi LF, Nunes RD, Schama R, Albano RM, Tirloni L, Oldiges DP, Vieira RP, Oliveira WHC, Leite M de S, *et al* (2017) A Coxiella mutualist symbiont is essential to the development of Rhipicephalus microplus. *Sci Rep* 7
Gulia-Nuss MN Andrew B; Meyer, Jason M; Sonenshine, Daniel E; Roe, R Michael; Waterhouse, Robert M; Sattelle, David B; de la Fuente, José; Ribeiro, José MC; Megy, Karyn; Thimmapuram, Jyothi; Miller, Jason R; Walenz, Brian P; Koren, Sergey; Hostetler, Jessica B; Thiagarajan, Mathangi; Joardar, Vinita; Hannick, Linda; Bidwell, Shelby L; Hammond, Martin; Young, Sarah; Zeng, Qiandong; Abrudan, Jenica L; Almeida, Francisca C; Ayllón, Nieves; Bhide, Ketaki; Bissinger, Brooke W; Bonzón-Kulichenko, Elena; Buckingham, Steven D; Caffrey, Daniel R; Caimano, Melissa J; Croset, Vincent; Driscoll, Timothy P; Gilbert, Don; Gillespie, Joseph J; Giraldo-Calderón, Gloria I; Grabowski, Jeffrey M; Jiang, David; Khalil, Sayed MS; Kim, Donghun; Kocan, Katherine M; Koči, Juraj; Kuhn, Richard J; Kurtti, Timothy J; Lees, Kristin; Lang, Emma G; Kennedy, Ryan C; Kwon, Hyeogsun; Perera, Rushika; Qi, Yumin; Radolf, Justin D; Sakamoto, Joyce M; Sánchez-Gracia, Alejandro; Severo, Maiara S; Silverman, Neal S; Šimo, Ladislav; Tojo, Marta; Tornador, Cristian; Van Zee, Janice P; Vázquez, Jesús; Vieira, Filipe G; Villar, Margarita; Wespiser, Adam R; Yang, Yunlong; Zhu, Jiwei; Arensburger, Peter; Pietrantonio, Patricia V; Barker, Stephen C; Shao, Renfu; Zdobnov, Evgeny M; Hauser, Frank; Grimmelikhuijzen, Cornelis JP; Park, Yoonseong; Rozas, Julio; Benton, Richard; Pedra, Joao HF; Nelson, David R; Unger, Maria F; Tubio, Jose MC; Tu, Zhijian Jake; Robertson, Hugh M; Shumway, Martin; Sutton, Granger G; Wortman, Jennifer R; Lawson, Daniel; Wikel, Stephen K; Nene, Vishvanath; Fraser, Claire M; Collins, Frank H; Birren, Bruce W; Nelson, Karen E; Caler, Elisabet; Hill, Catherine A (2016) Genomic insights into the Ixodes scapularis tick vector of Lyme disease. *Nat Commun* 7: 10507–10507
Gurevich AS Vladislav; Vyahhi, Nikolay; Tesler, Glenn (2013) QUAST: quality assessment tool for genome assemblies. *Bioinforma Oxf Engl* 29: 1072–1075
Guy LK Jens Roat; Andersson, Siv GE (2010) genoPlotR: comparative gene and genome visualization in R. *Bioinforma Oxf Engl* 26: 2334–2335
Hamilton PT, Peng F, Boulanger MJ & Perlman SJ (2016) A ribosome-inactivating protein in a Drosophila defensive symbiont. *Proc Natl Acad Sci U S A* 113: 350–355
He Z, Wang P, Shi H, Si F, Hao Y & Chen B (2013) Fas-associated factor 1 plays a negative regulatory role in the antibacterial immunity of Locusta migratoria. *Insect Mol Biol* 22: 389–398
Heddi A, Grenier A-M, Khatchadourian C, Charles H & Nardon P (1999) Four intracellular genomes direct weevil biology: Nuclear, mitochondrial, principal endosymbiont, and Wolbachia. *Proc Natl Acad Sci* 96: 6814–6819
Herrmann C & Gern L (2010) Survival of Ixodes ricinus (Acari: Ixodidae) under challenging conditions of temperature and humidity is influenced by Borrelia burgdorferi sensu lato infection. *J Med Entomol* 47: 1196–1204
Herrmann C & Gern L (2012) Do the level of energy reserves, hydration status and Borrelia infection influence walking by Ixodes ricinus (Acari: Ixodidae) ticks? *Parasitology* 139: 330–337

Herrmann C & Gern L (2013) Survival of Ixodes ricinus (Acari: Ixodidae) nymphs under cold conditions is negatively influenced by frequent temperature variations. *Ticks Tick-Borne Dis* 4: 445–451
Herrmann C & Gern L (2015) Search for blood or water is influenced by Borrelia burgdorferi in Ixodes ricinus. *Parasit Vectors* 8: 6
Hess S (2017) Description of Hyalodiscus flabellus sp. nov. (Vampyrellida, Rhizaria) and Identification of its Bacterial Endosymbiont, 'Candidatus Megaira polyxenophila' (Rickettsiales, Alphaproteobacteria). *Protist* 168: 109–133
Heylen D, Krawczyk A, Lopes de Carvalho I, Núncio MS, Sprong H & Norte AC (2017) Bridging of cryptic Borrelia cycles in European songbirds. *Environ Microbiol* 19: 1857–1867
Hoogstraal H, Kaiser MN, Traylor MA, Guindy E & Gaber S (1963) Ticks (Ixodidae) on birds migrating from Europe and Asia to Africa, 1959-61. *Bull World Health Organ* 28: 235–262
Hornok S, Flaisz B, Takács N, Kontschán J, Csörgő T, Csipak Á, Jaksa BR & Kováts D (2016) Bird ticks in Hungary reflect western, southern, eastern flyway connections and two genetic lineages of Ixodes frontalis and Haemaphysalis concinna. *Parasit Vectors* 9: 101
Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, *et al* (2019) eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 47: D309–D314
Hugoson E, Guliaev A, Ammunét T & Guy L (2022) Host Adaptation in Legionellales Is 1.9 Ga, Coincident with Eukaryogenesis. *Mol Biol Evol* 39: msac037
Hugoson EL Wai Tin; Guy, Lionel (2020) miComplete: weighted quality evaluation of assembled microbial genomes. *Bioinforma Oxf Engl* 36: 936–937
Husnik F & Keeling PJ (2021) The fate of obligate endosymbionts: reduction, integration, or extinction. *Curr Opin Genet Dev* 58–59: 1–8
Husnik F & McCutcheon JP (2016) Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis. *Proc Natl Acad Sci* 113: E5416–E5424
Imachi H, Nobu MK, Nakahara N, Morono Y, Ogawara M, Takaki Y, Takano Y, Uematsu K, Ikuta T, Ito M, *et al* (2020) Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature* 577: 519–525
Jahfari S, Coipan EC, Fonville M, van Leeuwen AD, Hengeveld P, Heylen D, Heyman P, van Maanen C, Butler CM, Földvári G, *et al* (2014) Circulation of four Anaplasma phagocytophilum ecotypes in Europe. *Parasit Vectors* 7: 365
Jia NW Jinfeng; Shi, Wenqiang; Du, Lifeng; Sun, Yi; Zhan, Wei; Jiang, Jia-Fu; Wang, Qian; Zhang, Bing; Ji, Peifeng; Bell-Sakyi, Lesley; Cui, Xiao-Ming; Yuan, Ting-Ting; Jiang, Bao-Gui; Yang, Wei-Fei; Lam, Tommy Tsan-Yuk; Chang, Qiao-Cheng; Ding, Shu-Jun; Wang, Xian-Jun; Zhu, Jin-Guo; Ruan, Xiang-Dong; Zhao, Lin; Wei, Jia-Te; Ye, Run-Ze; Que, Teng Cheng; Du, Chun-Hong; Zhou, Yu-Hao; Cheng, Jing Xia; Dai, Pei-Fang; Guo, Wen-Bin; Han, Xiao-Hu; Huang, En-Jiong; Li, Lian-Feng; Wei, Wei; Gao, Yu-Chi; Liu, Jing-Ze; Shao, Hong-Ze; Wang, Xin; Wang, Chong-Cai; Yang, Tian-Ci; Huo, Qiu-Bo; Li, Wei; Chen, Hai-Ying; Chen, Shen-En; Zhou, Ling-Guo; Ni, Xue-Bing; Tian, Jun-Hua; Sheng, Yue; Liu, Tao; Pan, Yu-Sheng; Xia, Luo-Yuan; Li, Jie; Genome, Tick; Zhao, Fangqing; Cao, Wu-Chun (2020) Large-Scale Comparative Analyses of Tick Genomes Elucidate Their Genetic Diversity and Vector Capacities. *Cell* 182: 1328-1340.e13
Kahl O, Bulling I & Chitimia-Dobler L (2019) Questing Ixodes frontalis larvae in a forest close to Berlin (Germany) in November 2018. *Ticks Tick-Borne Dis* 10: 513–514
Kahlon A, Ojogun N, Ragland SA, Seidman D, Troese MJ, Ottens AK, Mastronunzio JE, Truchan HK, Walker NJ, Borjesson DL, *et al* (2013) Anaplasma phagocytophilum Asp14 is an invasin that interacts with mammalian host cells via its C terminus to facilitate infection. *Infect Immun* 81: 65–79
Kalmár Z, Cozma V, Sprong H, Jahfari S, D'Amico G, Mărcuțan DI, Ionică AM, Magdaș C, Modrý D & Mihalca AD (2015) Transstadial transmission of Borrelia turcica in Hyalomma aegyptium ticks. *PloS One* 10: e0115520
Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A & Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14: 587–589
Kanehisa M & Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28:

27–30
Kanehisa MS Yoko; Morishima, Kanae (2015) BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol* 428: 726–731
Karp PD, Billington R, Caspi R, Fulcher CA, Latendresse M, Kothari A, Keseler IM, Krummenacker M, Midford PE, Ong Q, *et al* (2019) The BioCyc collection of microbial genomes and metabolic pathways. *Brief Bioinform* 20: 1085–1093
Katoh K & Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30: 772–780
Katoh KM Kazuharu; Kuma, Kei-ichi; Miyata, Takashi (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30: 3059–3066
Keegstra JM, Carrara F & Stocker R (2022) Publisher Correction: The ecological roles of bacterial chemotaxis. *Nat Rev Microbiol* 20: 505
Kernif T, Leulmi H, Raoult D & Parola P (2016) Emerging Tick-Borne Bacterial Pathogens. *Microbiol Spectr* 4
Keve G, Sándor AD & Hornok S (2022) Hard ticks (Acari: Ixodidae) associated with birds in Europe: Review of literature data. *Front Vet Sci* 9: 928756
Klompen H & Grimaldi D (2001) First Mesozoic Record of a Parasitiform Mite: a Larval Argasid Tick in Cretaceous Amber (Acari: Ixodida: Argasidae). *Ann Entomol Soc Am* 94: 10–15
Klubal R, Kopecky J, Nesvorna M, Sparagano OAE, Thomayerova J & Hubert J (2016) Prevalence of pathogenic bacteria in Ixodes ricinus ticks in Central Bohemia. *Exp Appl Acarol* 68: 127–137
Klyachko O, Stein BD, Grindle N, Clay K & Fuqua C (2007) Localization and visualization of a coxiella-type symbiont within the lone star tick, Amblyomma americanum. *Appl Environ Microbiol* 73: 6584–6594
Krause PJ, McKay K, Gadbaw J, Christianson D, Closter L, Lepore T, Telford SR, Sikand V, Ryan R, Persing D, *et al* (2003) Increasing health burden of human babesiosis in endemic sites. *Am J Trop Med Hyg* 68: 431–436
Krawczyk AI, Röttjers S, Coimbra-Dores MJ, Heylen D, Fonville M, Takken W, Faust K & Sprong H (2022) Tick microbial associations at the crossroad of horizontal and vertical transmission pathways. *Parasit Vectors* 15: 380
Krow-Lucal ER, Lindsey NP, Fischer M & Hills SL (2018) Powassan Virus Disease in the United States, 2006–2016. *Vector-Borne Zoonotic Dis* 18: 286–290
Kumar S, Jones M, Koutsovoulos G, Clarke M & Blaxter M (2013) Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. *Front Genet* 4
Kupper M, Gupta SK, Feldhaar H & Gross R (2014) Versatile roles of the chaperonin GroEL in microorganism-insect interactions. *FEMS Microbiol Lett* 353: 1–10
Kurtti TJ, Felsheim RF, Burkhardt NY, Oliver JD, Heu CC & Munderloh UGY 2015 (2015) Rickettsia buchneri sp. nov., a rickettsial endosymbiont of the blacklegged tick Ixodes scapularis. *Int J Syst Evol Microbiol* 65: 965–970
Kuzminov A (1999) Recombinational Repair of DNA Damage inEscherichia coli and Bacteriophage λ. *Microbiol Mol Biol Rev* 63: 751–813
Lalzar I, Friedmann Y & Gottlieb Y (2014) Tissue tropism and vertical transmission of Coxiella in Rhipicephalus sanguineus and Rhipicephalus turanicus ticks. *Environ Microbiol* 16: 3657–3668
Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9: 357–359
Lanzoni O, Sabaneyeva E, Modeo L, Castelli M, Lebedeva N, Verni F, Schrallhammer M, Potekhin A & Petroni G (2019) Diversity and environmental distribution of the cosmopolitan endosymbiont 'Candidatus Megaira'. *Sci Rep* 9: 1179
Lara FA, Lins U, Paiva-Silva G, Almeida IC, Braga CM, Miguens FC, Oliveira PL & Dansa-Petretski M (2003) A new intracellular pathway of haem detoxification in the midgut of the cattle tick Boophilus microplus: aggregation inside a specialized organelle, the hemosome. *J Exp Biol* 206: 1707–1715
Lartillot N & Philippe H (2004) A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol* 21: 1095–1109
Lefèvre T & Thomas F (2008) Behind the scene, something else is pulling the strings: emphasizing parasitic manipulation in vector-borne diseases. *Infect Genet Evol J Mol Epidemiol Evol Genet*

Infect Dis 8: 504–519
Li HD Richard (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinforma Oxf Engl* 25: 1754–1760
Li L-H, Zhang Y, Zhu D & Zhou X-N (2018) Endosymbionts Alter Larva-to-Nymph Transstadial Transmission of Babesia microti in Rhipicephalus haemaphysaloides Ticks. *Front Microbiol* 9
Linhartová I, Bumba L, Mašín J, Basler M, Osička R, Kamanová J, Procházková K, Adkins I, Hejnová-Holubová J, Sadílková L, *et al* (2010) RTX proteins: a highly diverse family secreted by a common mechanism. *FEMS Microbiol Rev* 34: 1076–1112
Liu B, Zheng D, Zhou S, Chen L & Yang J (2022) VFDB 2022: a general classification scheme for bacterial virulence factors. *Nucleic Acids Res* 50: D912–D917
Lo N, Beninati T, Sassera D, Bouman E a. P, Santagati S, Gern L, Sambri V, Masuzawa T, Gray JS, Jaenson TGT, *et al* (2006) Widespread distribution and high prevalence of an alpha-proteobacterial symbiont in the tick Ixodes ricinus. *Environ Microbiol* 8: 1280–1287
Lobato-Márquez D, Díaz-Orejas R & García-Del Portillo F (2016) Toxin-antitoxins and bacterial virulence. *FEMS Microbiol Rev* 40: 592–609
Luedtke BE, Mahapatra S, Lutter EI & Shaw EI (2017) The Coxiella Burnetii type IVB secretion system (T4BSS) component DotA is released/secreted during infection of host cells and during in vitro growth in a T4BSS-dependent manner. *Pathog Dis* 75
Lv J, Wu S, Zhang Y, Chen Y, Feng C, Yuan X, Jia G, Deng J, Wang C, Wang Q, *et al* (2014) Assessment of four DNA fragments (COI, 16S rDNA, ITS2, 12S rDNA) for species identification of the Ixodida (Acari: Ixodida). *Parasit Vectors* 7: 93
Machado-Ferreira EV Vinicius F; Balsemão-Pires, Emilia; Moerbeck, Leonardo; Gazeta, Gilberto Salles; Piesman, Joseph; Voloch, Carolina M; Soares, Carlos AG (2016) Coxiella symbionts are widespread into hard ticks. *Parasitol Res* 115: 4691–4699
Madison-Antenucci S, Kramer LD, Gebhardt LL & Kauffman E (2020) Emerging Tick-Borne Diseases. *Clin Microbiol Rev* 33: e00083-18
Major P, Embley TM & Williams TA (2017) Phylogenetic Diversity of NTT Nucleotide Transport Proteins in Free-Living and Parasitic Bacteria and Eukaryotes. *Genome Biol Evol* 9: 480–487
Manzano-Marín A, D'acier AC, Clamens A-L, Cruaud C, Barbe V & Jousselin E (2022) Co-obligate symbioses have repeatedly evolved across aphids, but partner identity and nutritional contributions vary across lineages. 2022.08.28.505559 doi:10.1101/2022.08.28.505559 [PREPRINT]
Manzano-Marín A, Simon J-C & Latorre A (2016) Reinventing the Wheel and Making It Round Again: Evolutionary Convergence in Buchnera-Serratia Symbiotic Consortia between the Distantly Related Lachninae Aphids Tuberolachnus salignus and Cinara cedri. *Genome Biol Evol* 8: 1440–1458
Manzano-Marín A, Coeur d'acier A, Clamens A-L, Orvain C, Cruaud C, Barbe V & Jousselin E (2020) Serial horizontal transfer of vitamin-biosynthetic genes enables the establishment of new nutritional symbionts in aphids' di-symbiotic systems. *ISME J* 14: 259–273
Mariconti M, Epis S, Gaibani P, Dalla Valle C, Sassera D, Tomao P, Fabbi M, Castelli F, Marone P, Sambri V, *et al* (2012) Humans parasitized by the hard tick Ixodes ricinus are seropositive to Midichloria mitochondrii: is Midichloria a novel pathogen, or just a marker of tick bite? *Pathog Glob Health* 106: 391–396
Martijn J, Schön ME, Lind AE, Vosseberg J, Williams TA, Spang A & Ettema TJG (2020) Hikarchaeia demonstrate an intermediate stage in the methanogen-to-halophile transition. *Nat Commun* 11
Martijn J, Vosseberg J, Guy L, Offre P & Ettema TJG (2018) Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature* 557: 101–105
Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17: 10–12
Massengo-Tiassé RP & Cronan JE (2008) Vibrio cholerae FabV Defines a New Class of Enoyl-Acyl Carrier Protein Reductase. *J Biol Chem* 283: 1308–1316
Mattick JS (2002) Type IV pili and twitching motility. *Annu Rev Microbiol* 56: 289–314
McCutcheon JP; M Nancy A (2012) Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* 10: 13–26
Mertens K, Lantsheer L, Ennis DG & Samuel JE (2008) Constitutive SOS expression and

damage-inducible AddAB-mediated recombinational repair systems for Coxiella burnetii as potential adaptations for survival within macrophages. *Mol Microbiol* 69: 1411–1426
Meseguer AS, Manzano-Marín A, Coeur d'Acier A, Clamens A-L, Godefroid M & Jousselin E (2017) Buchnera has changed flatmate but the repeated replacement of co-obligate symbionts is not associated with the ecological expansions of their aphid hosts. *Mol Ecol* 26: 2363–2378
Min C-K, Yang J-S, Kim S, Choi M-S, Kim I-S & Cho N-H (2008) Genome-based construction of the metabolic pathways of Orientia tsutsugamushi and comparative analysis within the Rickettsiales order. *Comp Funct Genomics*: 623145
Minh BQ, Nguyen MAT & von Haeseler A (2013) Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol* 30: 1188–1195
Mironov T & Sabaneyeva E (2020) A Robust Symbiotic Relationship Between the Ciliate and the Bacterium . Trichorickettsia Mobilis. *Front Microbiol* 11: 603335
Monks D, Fisher M & Forbes NA (2006) Ixodes frontalis and avian tick-related syndrome in the United Kingdom. *J Small Anim Pract* 47: 451–455
Montagna M, Sassera D, Epis S, Bazzocchi C, Vannini C, Lo N, Sacchi L, Fukatsu T, Petroni G & Bandi C (2013) 'Candidatus Midichloriaceae' fam. nov. (Rickettsiales), an ecologically widespread clade of intracellular alphaproteobacteria. *Appl Environ Microbiol* 79: 3241–3248
Moran NA, McCutcheon JP & Nakabachi A (2008) Genomics and Evolution of Heritable Bacterial Symbionts. *Annu Rev Genet* 42: 165–190
Movila A, Alekseev AN, Dubinina HV & Toderas I (2013) Detection of tick-borne pathogens in ticks from migratory birds in the Baltic region of Russia. *Med Vet Entomol* 27: 113–117
Muñoz-Gómez SA, Hess S, Burger G, Lang BF, Susko E, Slamovits CH & Roger AJ (2019) An updated phylogeny of the Alphaproteobacteria reveals that the parasitic Rickettsiales and Holosporales have independent origins. *eLife* 8: e42535
Nakao R, Shinjo K, Sakiyama T, Ogata S, Kusakisako K, Kinoshita G, Naguib D, Chatanga E, Mohamed WMA, Moustafa MAM, *et al* (2021) Amblyomma testudinarium infestation on a brown bear (Ursus arctos yesoensis) captured in Hokkaido, a northern island of Japan. *Parasitol Int* 80: 102209
Narasimhan S, Rajeevan N, Liu L, Zhao YO, Heisig J, Pan J, Eppler-Epstein R, Deponte K, Fish D & Fikrig E (2014) Gut microbiota of the tick vector Ixodes scapularis modulate colonization of the Lyme disease spirochete. *Cell Host Microbe* 15: 58–71
Needle DB, Agnew DW, Bradway DS, Nordhausen RW & Garner MM (2020) Avian coxiellosis in nine psittacine birds, one black-browed barbet, and one paradise tanager. *Avian Pathol* 49: 268–274
Neelakanta G, Sultana H, Fish D, Anderson JF & Fikrig E (2010) Anaplasma phagocytophilum induces Ixodes scapularis ticks to express an antifreeze glycoprotein gene that enhances their survival in the cold. *J Clin Invest* 120: 3179–3190
Nguyen L-T, Schmidt HA, von Haeseler A & Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32: 268–274
Niebylski ML, Schrumpf ME, Burgdorfer W, Fischer ER, Gage KL & Schwan TG (1997) Rickettsia peacockii sp. nov., a new species infecting wood ticks, Dermacentor andersoni, in western Montana. *Int J Syst Bacteriol* 47: 446–452
Niu H, Kozjak-Pavlovic V, Rudel T & Rikihisa Y (2010) Anaplasma phagocytophilum Ats-1 is imported into host cell mitochondria and interferes with apoptosis induction. *PLoS Pathog* 6: e1000774
Nogge G (1976) Sterility in tsetse flies (Glossina morsitans Westwood) caused by loss of symbionts. *Experientia* 32: 995–996
Nwasike C, Ewert S, Jovanovic S, Haider S & Mujtaba S (2016) SET domain-mediated lysine methylation in lower organisms regulates growth and transcription in hosts. *Ann N Acad Sci* 1376: 18–28
Nyholm SV & McFall-Ngai MJ (2004) The winnowing: establishing the squid-vibrio symbiosis. *Nat Rev Microbiol* 2: 632–642
Ogata S, Mohamed WMA, Kusakisako K, Thu MJ, Qiu Y, Moustafa MAM, Matsuno K, Katakura K, Nonaka N & Nakao R (2021) Spiroplasma Infection among Ixodid Ticks Exhibits Species Dependence and Suggests a Vertical Pattern of Transmission. *Microorganisms* 9: 333
Oliver KM, Moran NA & Hunter MS (2006) Costs and benefits of a superinfection of facultative symbionts in aphids. *Proc R Soc B Biol Sci* 273: 1273–1280

Oliver KM, Russell JA, Moran NA & Hunter MS (2003) Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proc Natl Acad Sci* 100: 1803–1807
Olivieri E, Epis S, Castelli M, Varotto Boccazzi I, Romeo C, Desirò A, Bazzocchi C, Bandi C & Sassera D (2019) Tissue tropism and metabolic pathways of Midichloria mitochondrii suggest tissue-specific functions in the symbiosis with Ixodes ricinus. *Ticks Tick-Borne Dis* 10: 1070–1077
Otten C, Brilli M, Vollmer W, Viollier PH & Salje J (2018) Peptidoglycan in obligate intracellular bacteria. *Mol Microbiol* 107: 142–163
Padmalayam I, Karem K, Baumstark B & Massung R (2000) The gene encoding the 17-kDa antigen of Bartonella henselae is located within a cluster of genes homologous to the virB virulence operon. *DNA Cell Biol* 19: 377–382
Pais IS, Valente RS, Sporniak M & Teixeira L (2018) Drosophila melanogaster establishes a species-specific mutualistic interaction with stable gut-colonizing bacteria. *PLoS Biol* 16: e2005710
Pais R, Lohs C, Wu Y, Wang J & Aksoy S (2008) The Obligate Mutualist Wigglesworthia glossinidia Influences Reproduction, Digestion, and Immunity Processes of Its Host, the Tsetse Fly. *Appl Environ Microbiol* 74: 5965–5974
Panek H & O'Brian MR (2002) A whole genome view of prokaryotic haem biosynthesis. *Microbiology* 148: 2273–2282
Park J, Kim KJ, Grab DJ & Dumler JS (2003) Anaplasma phagocytophilum major surface protein-2 (Msp2) forms multimeric complexes in the bacterial membrane. *FEMS Microbiol Lett* 227: 243–247
Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A & Hugenholtz P (2022) GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res* 50: D785–D794
Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, Hugenholtz P & Tyson GW (2017) Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* 2: 1533–1542
Paules CI, Marston HD, Bloom ME & Fauci AS (2018) Tickborne Diseases — Confronting a Growing Threat. *N Engl J Med* 379: 701–703
Pennington JE, Goldstrohm DA & Wells MA (2003) The role of hemolymph proline as a nitrogen sink during blood meal digestion by the Mosquito Aedes aegypti. *J Insect Physiol* 49: 115–121
Perner JS Roman; Sima, Radek; Konvičková, Jitka; Sojka, Daniel; de Oliveira, Pedro Lagerblad; Hajdusek, Ondrej; Kopáček, Petr (2016) Acquisition of exogenous haem is essential for tick reproduction. *eLife* 5: NA-NA
Philip CB (1953) Tick Talk. *Sci Mon* 76: 77–84
Pierantoni U (1927) Nuove osservazioni sull'organo simbiotico di Calandra oryzae. *Rendiconto DellAccademia Delle Sci Fis E Mat*
Plantard O, Hoch T, Daveu R, Rispe C, Stachurski F, Boué F, Poux V, Cebe N, Verheyden H, René-Martellet M, *et al* (2021) Where to find questing Ixodes frontalis ticks? Under bamboo bushes! *Ticks Tick-Borne Dis* 12: 101625
Price DRG, Bartley K, Blake DP, Karp-Tatham E, Nunn F, Burgess STG & Nisbet AJ (2021) A Rickettsiella Endosymbiont Is a Potential Source of Essential B-Vitamins for the Poultry Red Mite, Dermanyssus gallinae. *Front Microbiol* 12: 695346
Randolph SE, Gern L & Nuttall PA (1996) Co-feeding ticks: Epidemiological significance for tick-borne pathogen transmission. *Parasitol Today* 12: 472–479
Reinhardt C, Aeschlimann A & Hecker H (1972) Distribution of rickettsia-like microorganisms in various organs of an Ornithodorus moubata laboratory strain (Ixodoidea, Argasidae) as revealed by electron microscopy. *Z Für Parasitenkd* 39: 201–209
Renvoisé A, Merhej V, Georgiades K & Raoult D (2011) Intracellular Rickettsiales: Insights into manipulators of eukaryotic cells. *Trends Mol Med* 17: 573–583
Reynolds C, Kontschán J, Takács N, Solymosi N, Sándor AD, Keve G & Hornok S (2022) Shift in the seasonality of ixodid ticks after a warm winter in an urban habitat with notes on morphotypes of Ixodes ricinus and data in support of cryptic species within Ixodes frontalis. *Exp Appl Acarol* 88: 127–138

Říhová J (2017) Legionella Becoming a Mutualist: Adaptive Processes Shaping the Genome of Symbiont in the Louse Polyplax serrata. *Genome Biol Evol* 9: 2946–2957
Rikihisa Y (2010) Anaplasma phagocytophilum and Ehrlichia chaffeensis: subversive manipulators of host cells. *Nat Rev Microbiol* 8: 328–339
Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA & Huelsenbeck JP (2012) MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst Biol* 61: 539–542
Rosenberg R, Lindsey NP, Fischer M, Gregory CJ, Hinckley AF, Mead PS, Paz-Bailey G, Waterman SH, Drexler NA, Kersh GJ, *et al* (2018) Vital Signs: Trends in Reported Vectorborne Disease Cases — United States and Territories, 2004–2016. *Morb Mortal Wkly Rep* 67: 496–501
Rosenwald LC, Sitvarin MI & White JA (2020) Endosymbiotic Rickettsiella causes cytoplasmic incompatibility in a spider host. *Proc R Soc B Biol Sci* 287: 20201107
Ross BD, Hayes B, Radey MC, Lee X, Josek T, Bjork J, Neitzel D, Paskewitz S, Chou S & Mougous JD (2018) Ixodes scapularis does not harbor a stable midgut microbiome. *ISME J* 12: 2596–2607
Roux V, Bergoin M, Lamaze N & Raoult DY 1997 (1997) Reassessment of the Taxonomic Position of Rickettsiella grylli. *Int J Syst Evol Microbiol* 47: 1255–1257
Sacchi L, Bigliardi E, Corona S, Beninati T, Lo N & Franceschi A (2004) A symbiont of the tick Ixodes ricinus invades and consumes mitochondria in a mode similar to that of the parasitic bacterium Bdellovibrio bacteriovorus. *Tissue Cell* 36: 43–53
Saier MH, Reddy VS, Moreno-Hagelsieb G, Hendargo KJ, Zhang Y, Iddamsetty V, Lam KJK, Tian N, Russum S, Wang J, *et al* (2021) The Transporter Classification Database (TCDB): 2021 update. *Nucleic Acids Res* 49: D461–D467
Salje J (2021) Cells within cells: Rickettsiales and the obligate intracellular bacterial lifestyle. *Nat Rev Microbiol* 19: 375–390
Sándor AD, Mărcuţan DI, D'Amico G, Gherman CM, Dumitrache MO & Mihalca AD (2014) Do the Ticks of Birds at an Important Migratory Hotspot Reflect the Seasonal Dynamics of Ixodes ricinus at the Migration Initiation Site? A Case Study in the Danube Delta. *PLoS ONE* 9: e89378
Sandström J & Moran N (1999) How nutritionally imbalanced is phloem sap for aphids? In *Proceedings of the 10th International Symposium on Insect-Plant Relationships*, Simpson SJ Mordue AJ & Hardie J (eds) pp 203–210. Dordrecht: Springer Netherlands
Santos-Silva M, Sousa R, Santos AS, Lopes D, Queijo E, Doreta A, Vitorino L & Bacellar F (2006) Ticks and tick-borne Rickettsiae surveillance in Montesinho Natural Park, Portugal. *Ann N Y Acad Sci* 1078: 137–142
Santos-Silva MM, Beati L, Santos AS, De Sousa R, Núncio MS, Melo P, Santos-Reis M, Fonseca C, Formosinho P, Vilela C, *et al* (2011) The hard-tick fauna of mainland Portugal (Acari: Ixodidae): an update on geographical distribution and known associations with hosts and pathogens. *Exp Appl Acarol* 55: 85–121
Sassera D, Lo N, Bouman EAP, Epis S, Mortarino M & Bandi C (2008) "Candidatus Midichloria" Endosymbionts Bloom after the Blood Meal of the Host, the Hard Tick Ixodes ricinus. *Appl Environ Microbiol* 74: 6138–6140
Sassera DL Nathan; Epis, Sara; D'Auria, Giuseppe; Montagna, Matteo; Comandatore, Francesco; Horner, David S; Peretó, Juli; Luciano, Alberto M; Franciosi, Federica; Ferri, Emanuele; Crotti, Elena; Bazzocchi, Chiara; Daffonchio, Daniele; Sacchi, Luciano; Moya, Andrés; Latorre, Amparo; Bandi, Claudio (2011) Phylogenomic evidence for the presence of a flagellum and cbb3 oxidase in the free-living mitochondrial ancestor. *Mol Biol Evol* 28: 3285–3296
Schmitz-Esser S, Linka N, Collingro A, Beier CL, Neuhaus HE, Wagner M & Horn M (2004) ATP/ADP translocases: a common feature of obligate intracellular amoebal symbionts related to Chlamydiae and Rickettsiae. *J Bacteriol* 186: 683–691
Schön ME, Martijn J, Vosseberg J, Köstlbacher S & Ettema TJG (2022) The evolutionary origin of host association in the Rickettsiales. *Nat Microbiol*
Schorn S, Schöl H, Pfister K & Silaghi C (2011) First record of Ixodes frontalis collected by flagging in Germany. *Ticks Tick-Borne Dis* 2: 228–230
Schrallhammer M, Schweikert M, Vallesi A, Verni F & Petroni G (2011) Detection of a Novel Subspecies of Francisella noatunensis as Endosymbiont of the Ciliate Euplotes raikovi. *Microb Ecol* 61: 455–464

Schulz F, Martijn J, Wascher F, Lagkouvardos I, Kostanjšek R, Ettema TJG & Horn M (2016) A Rickettsiales symbiont of amoebae with ancient features. *Env Microbiol* 18: 2326–2342
Sears KT, Ceraul SM, Gillespie JJ, Allen ED Jr, Popov VL, Ammerman NC, Rahman MS & Azad AF (2012) Surface proteome analysis and characterization of surface cell antigen (Sca) or autotransporter family of Rickettsia typhi. *PLoS Pathog* 8: e1002856
Seemann T (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30: 2068–2069
Seidman D, Ojogun N, Walker NJ, Mastronunzio J, Kahlon A, Hebert KS, Karandashova S, Miller DP, Tegels BK, Marconi RT, *et al* (2014) Anaplasma phagocytophilum surface protein AipA mediates invasion of mammalian host cells. *Cell Microbiol* 16: 1133–1145
Selmi RS Mourad Ben; Mamlouk, Aymen; Yahia, Houcine Ben; Messadi, Lilia (2019) Molecular detection and genetic characterization of the potentially pathogenic Coxiella burnetii and the endosymbiotic Candidatus Midichloria mitochondrii in ticks infesting camels (Camelus dromedarius) from Tunisia. *Microb Pathog* 136: 103655–103655
Seppey M, Manni M & Zdobnov EM (2019) BUSCO: Assessing Genome Assembly and Annotation Completeness. *Methods Mol Biol Clifton NJ* 1962: 227–245
Siguier P, Perochon J, Lestrade L, Mahillon J & Chandler M (2006) ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* 34: D32-36
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV & Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31: 3210–3212
Sonenshine DE & Roe RM (2013) Biology of Ticks Volume 1 OUP USA
Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313
Stanek G & Strle F (2018) Lyme borreliosis–from tick bite to diagnosis and treatment. *FEMS Microbiol Rev* 42: 233–258
Sterkel M, Oliveira JHM, Bottino-Rojas V, Paiva-Silva GO & Oliveira PL (2017) The Dose Makes the Poison: Nutritional Overload Determines the Life Traits of Blood-Feeding Arthropods. *Trends Parasitol* 33: 633–644
Stothard PW David S (2004) Circular genome visualization and exploration using CGView. *Bioinforma Oxf Engl* 21: 537–539
Süss J (2011) Tick-borne encephalitis 2010: Epidemiology, risk areas, and virus strains in Europe and Asia—An overview. *Ticks Tick-Borne Dis* 2: 2–15
Swart AL, Gomez-Valero L, Buchrieser C & Hilbi H (2020) Evolution and function of bacterial RCC1 repeat effectors. *Cell Microbiol* 22: e13246
Syberg-Olsen M, Garber A, Keeling P, McCutcheon J & Husnik F (2020) Pseudofinder.
Szokoli F, Castelli M, Sabaneyeva E, Schrallhammer M, Krenek S, Doak TG, Berendonk TU & Petroni G (2016) Disentangling the Taxonomy of Rickettsiales and Description of Two Novel Symbionts ('Candidatus Bealeia paramacronuclearis' and 'Candidatus Fokinia cryptica') Sharing the Cytoplasm of the Ciliate Protist Paramecium biaurelia. *Appl Env Microbiol* 82: 7236–7247
Szöllősi GJ, Davín AA, Tannier E, Daubin V & Boussau B (2015) Genome-scale phylogenetic analysis finds extensive gene transfer among fungi. *Philos Trans R Soc Lond B Biol Sci* 370: 20140335
Szöllõsi GJ, Rosikiewicz W, Boussau B, Tannier E & Daubin V (2013) Efficient exploration of the space of reconciled gene trees. *Syst Biol* 62: 901–912
Talavera G & Castresana J (2007) Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. *Syst Biol* 56: 564–577
Tamas IW Jennifer J; Nystedt, Björn; Kauppinen, Seth N; Darby, Alistair C; Gomez-Valero, Laura; Lundin, Daniel; Poole, Anthony M; Andersson, Siv GE (2008) Endosymbiont gene functions impaired and rescued by polymerase infidelity at poly(A) tracts. *Proc Natl Acad Sci U S A* 105: 14934–14939
Tanner JJ (2008) Structural Biology of Proline Catabolism. *Amino Acids* 35: 719–730
Theiler G & Salisbury L (1959) Ticks in the South African Zoological Survey Collection. Part IX. 'the Amblyomma marmoreum group'. *Onderstepoort J Vet Res* 28: 77
Tsementzi D, Gordillo JC, Mahagna M, Gottlieb Y & Konstantinidis KT (2018) Comparison of closely related, uncultivated Coxiella tick endosymbiont population genomes reveals clues about the mechanisms of symbiosis. *Environ Microbiol* 20: 1751–1764

Vannini C, Ferrantini F, Schleifer K-H, Ludwig W, Verni F & Petroni G (2010) 'Candidatus anadelfobacter veles' and 'Candidatus cyrtobacter comes,' two new rickettsiales species hosted by the protist ciliate Euplotes harpa (Ciliophora, Spirotrichea). *Appl Env Microbiol* 76: 4047–4054

Vannini C, Petroni G, Verni F & Rosati G (2005) A Bacterium Belonging to the Rickettsiaceae Family Inhabits the Cytoplasm of the Marine Ciliate Diophrys appendiculata (Ciliophora, Hypotrichia). *Microb Ecol* 49: 434–442

Vautrin E & Vavre F (2009) Interactions between vertically transmitted symbionts: cooperation or conflict? *Trends Microbiol* 17: 95–99

Veyron S, Peyroche G & Cherfils J (2018) FIC proteins: from bacteria to humans and back again. *Pathog Dis* 76

Vial LS Frédéric; Leblond, Agnès; Huber, Karine; Vourc'h, Gwenaël; René-Martellet, Magalie; Desjardins, Isabelle; Balança, Gilles; Grosbois, Vladimir; Pradier, Sophie; Gély, Marie; Appelgren, Anaïs; Estrada-Peña, Agustín (2016) Strong evidence for the presence of the tick Hyalomma marginatum Koch, 1844 in southern continental France. *Ticks Tick-Borne Dis* 7: 1162–1167

Vigneron A, Masson F, Vallier A, Balmand S, Rey M, Vincent-Monégat C, Aksoy E, Aubailly-Giraud E, Zaidman-Rémy A & Heddi A (2014) Insects Recycle Endosymbionts when the Benefit Is Over. *Curr Biol* 24: 2267–2273

Viklund J, Ettema TJG & Andersson SGE (2012) Independent Genome Reduction and Phylogenetic Reclassification of the Oceanic SAR11 Clade. *Mol Biol Evol* 29: 599–615

Vogel KJ & Moran NA (2013) Functional and evolutionary analysis of the genome of an obligate fungal symbiont. *Genome Biol Evol* 5: 891–904

Walker DH & Ismail N (2008) Emerging and re-emerging rickettsioses: endothelial cell infection and early disease events. *Nat Rev Microbiol* 6: 375–386

Walter DE & Proctor HC (1998) Feeding behaviour and phylogeny: observations on early derivative Acari. *Exp Appl Acarol* 22: 39–50

Wang D, Huang Z, Billen J, Zhang G, He H & Wei C (2021a) Structural diversity of symbionts and related cellular mechanisms underlying vertical symbiont transmission in cicadas. *Environ Microbiol* 23: 6603–6621

Wang D, Liu Y, Su Y & Wei C (2021b) Bacterial Communities in Bacteriomes, Ovaries and Testes of three Geographical Populations of a Sap-Feeding Insect, Platypleura kaempferi (Hemiptera: Cicadidae). *Curr Microbiol* 78: 1778–1791

Wang S & Luo H (2021) Dating Alphaproteobacteria evolution with eukaryotic fossils. *Nat Commun* 12: 3324

Weinert LA, Werren JH, Aebi A, Stone GN & Jiggins FM (2009) Evolution and diversity of Rickettsia bacteria. *BMC Biol* 7: 6

Werren JH, Baldo L & Clark ME (2008) Wolbachia: master manipulators of invertebrate biology. *Nat Rev Microbiol* 6: 741–751

Wick RR, Judd LM, Gorrie CL & Holt KE (2017a) Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13: e1005595

Wick RR, Judd LM, Gorrie CL & Holt KE (2017b) Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb Genom* 3: e000132

Wick RR, Schultz MB, Zobel J & Holt KE (2015) Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* 31: 3350–3352

Xie F, Jin W, Si H, Yuan Y, Tao Y, Liu J, Wang X, Yang C, Li Q, Yan X, *et al* (2021) An integrated gene catalog and over 10,000 metagenome-assembled genomes from the gastrointestinal microbiome of ruminants. *Microbiome* 9: 137

Xie Z & Tang H (2017) ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics* 33: 3340–3347

Yabsley MJ & Shock BC (2013) Natural history of Zoonotic Babesia: Role of wildlife reservoirs. *Int J Parasitol Parasites Wildl* 2: 18–31

Yurchenko T, Ševčíková T, Přibyl P, El Karkouri K, Klimeš V, Amaral R, Zbránková V, Kim E, Raoult D, Santos LMA, *et al* (2018) A gene transfer event suggests a long-term partnership between eustigmatophyte algae and a novel lineage of endosymbiotic bacteria. *ISME J* 12: 2163–2175

Zhong J, Jasinskas A & Barbour AG (2007) Antibiotic Treatment of the Tick Vector Amblyomma americanum Reduced Reproductive Fitness. *PLoS ONE* 2

Zhong Z, Zhong T, Peng Y, Zhou X, Wang Z, Tang H & Wang J (2021) Symbiont-regulated serotonin biosynthesis modulates tick feeding activity. *Cell Host Microbe* 29: 1545-1557.e4

Zug R & Hammerstein P (2015) Wolbachia and the insect immune system: what reactive oxygen species can tell us about the mechanisms of Wolbachia–host interactions. *Front Microbiol* 6

Acknowledgements

Science is never the results of one person alone, so, without surprise, pursuing a doctorate relies upon the help of many people to whom I owe many thanks.

First, I am grateful to my supervisor, Prof. Davide Sassera and to Dr. Michele Castelli for supervising me during my PhD, supporting me to grow as a scientist.
I have to thank Dr. Simonetta Gribaldo and Dr. Najwa Taib for their supervision during my experience at the Pasteur Institut in Paris.

I am grateful to my lab colleagues for the time spent together, in particular Greta, Gherard and Anna. I I am also grateful to all people in Dr. Gribaldo's lab at Pasteur, in particular Basile, and all the Stapa people.

From a more personal point of view, I have to thank my family for supporting me.
I owe a great deal to my usual ND squad: Marco, Tullio and Fra, and to Chero, Fofo and Horr.
I could not have undertaken this PhD, and many other things, without the support, time and tea from my dearest Silvia and Isa.
I have to extend thanks to my Parisian household: Clio and Giulia with the (mostly) helpful presence of M. v. Richthofen, A. Garibaldi and M. Raskova.