

UNIVERSITÀ DEGLI STUDI DI MILANO

SCUOLA DI DOTTORATO
Terra Ambiente e Biodiversità

Dipartimento di Scienze veterinarie e sanità pubblica - DIVET-
Dottorato di Ricerca in Biologia Animale, XXV Ciclo

Hard ticks (ACARINA, IXODIDAE): species identification, phylogeny and symbioses

- VET/06 – BIO/05 -

MATTEO MONTAGNA

Tutore

prof. CLAUDIO BANDI

Coordinatore del Dottorato

prof. MARCO FERRAGUTI

2011-2012

Statement to my Thesis

This work was carried out in the group of Prof. Claudio Bandi at the Università degli Studi di Milano.

My PhD thesis committee consisted of:

Prof. GIORGIO BAVESTRELLO

Prof. CLAUDIO BANDI

Dott.sa MARIA GABRIELLA MARIN

My thesis is written in a cumulative format. It consists of a synopsis covering several aspects concerning to my research works on ticks (Acarina, Ixodidae) with a global approach, from morphological/molecular taxonomy to bacteria harbored and transmitted by ticks, to genomic, expression and population genetic studies. This is followed by result chapters presenting my research consisting of four published research article, a submitted manuscript, an accepted manuscript and one manuscript in preparation. Finally, I have reported other aspect of my research work during my PhD experience (i.e. other published papers, research activities and grants received).

1. Introduction and aim of the thesis	6
1.1 References	10
2. Results	13
2.1 Research article I: " <i>Evaluation of DNA Barcoding as tool to identify hard ticks (genus AMBLYOMMA Koch 1844, Acarina, Ixodidae)</i> " (manuscript in preparation)	13
2.1.1 Summary	13
2.1.2 Manuscript	13
2.1.3 References	21
2.1.4 Appendix	23
2.2 Research article II: " <i>Screening for bacterial DNA in the hard tick Hyalomma marginatum (Ixodidae) from Socotra Island (Yemen): detection of Francisella-like endosymbiont</i> " (published on Journal of Acarological and Entomological Research)	26
2.2.1 Summary	26
2.2.2 Manuscript	26
2.2.3 References	31
2.3 Research article III: " <i>Phylogenomic evidence for the presence of a flagellum and cbb3 oxidase in the free-living mitochondrial ancestor</i> " (published on Molecular Biology and Evolution)	34
2.3.1 Summary	34
2.3.2 Manuscript	34
2.3.3 References	50
2.3.4 Supplementary materials	56
2.4 Research article IV: " <i>Midichloria and like organisms form an ecologically widespread clade of intracellular alpha-proteobacteria</i> " (manuscript submitted to Applied and Environmental Microbiology)	61
2.4.1 Summary	61
2.4.2 Manuscript	61
2.4.3 References	76

2.5 Research article IV: " <i>Tick-Box for 3'-end formation of mitochondrial transcripts in Ixodida, basal chelicerates and DROSOPHILA</i> " (published on PlosOne)	80
2.5.1 Summary	80
2.5.2 Manuscript	80
2.5.3 References	104
2.5.4 Supplementary materials	110
2.6 Research article V: " <i>Localization of the bacterial symbiont Candidatus Midichloria mitochondrii within the hard tick Ixodes ricinus by whole-mount FISH staining</i> " (published on Ticks and Tick-Borne Diseases)	119
2.6.1 Summary	119
2.6.2 Manuscript	119
2.6.3 References	132
2.7 Research article VII: " <i>Multiple independent data reveals an unusual response to Pleistocene climatic changes in the hard tick IXODES RICINUS</i> " (accepted on Molecular Ecology)	135
2.7.1 Summary	135
2.7.2 Manuscript	135
2.7.3 References	156
2.7.4 Appendix	165
3. Other relevant publications produced during my PhD (a color version of the pdf documents are stored in a separate folder on CD-ROM attached to the printed copy)	167
3.1 Research article: " <i>PACHYBRACHIS SASSII, a new species from the Mediterranean Giglio Island (Italy) (Coleoptera, Chrysomelidae, Cryptocephalinae). ZooKeys (2011), 155: 51–60</i> "	167
3.2 Research article: " <i>Molecular Evidence for Multiple Infections as Revealed by Typing of ASAIA Bacterial Symbionts of Four Mosquito Species. Applied and Environmental Microbiology (2010), 76(22): 7444–7450</i> "	167

3.3 Research article: " <i>The Beetle (Coleoptera) and True bug (Heteroptera) species pool of the Alpine "Pian di Gembro" wetland (Villa di Tirano, Italy) and its conservation. Journal of Entomological and Acarological Research (2011) 43: 7-22</i> "	167
3.4 Research article: " <i>Insect community structure and insect biodiversity conservation in an Alpine wetland subjected to an intermediate diversified management regime. Ecological Engineering (2012), 47: 242–246</i> "	167
3.5 Research article: " <i>Molecular typing of bacteria of the genus ASAIA in malaria vector ANOPHELES ARABIENSIS Patton, 1905. Journal of Entomological and Acarological Research (2012) 44: 33-36</i> "	167
3.6 Research article: " <i>A study on the presence of flagella in the order Rickettsiales: the case of 'Candidatus Midichloria mitochondrii'. Microbiology (2012), 158, 1677–1683</i> "	167
4 Grants and projects	168
5 Curriculum vitae	223

1 Introduction and aim of the thesis

Ticks (Acari, Parasitiformes, Ixodida) are obligate blood-sucking ectoparasites able to parasitize a great variety of terrestrial vertebrates (mammals, birds, reptiles and amphibians) (Sonenshine 1991, 1993). Fossil members of the Parasitiformes appear in the late Cretaceous, 100.5 - 66 Mya (Poinar and Brown 2003), but the ancestor of ticks evolved probably in pre-mid Cretaceous period, 145 - 120 Mya (Klompen *et al.* 1996; Nava *et al.* 2009). The primary hosts may have been reptiles or amphibians (Nava *et al.* 2009), but the major adaptive radiation probably occurred in mid-Cenozoic following the radiation of their hosts (Poinar 1995; Klompen *et al.* 1996). Nowadays approximately 870 species of ticks are described (Horak *et al.* 2002; Keirans 2009) and are subdivided into three families: Argasidae, or soft ticks because of the lack of a sclerotized dorsal scutum, and with the capitulum positioned ventrally; Ixodidae (or hard ticks) possess a sclerotized dorsal scutum and an apical positioned capitulum; the monotypic family of Nuttalliellidae, in which *Nuttalliella namaqua* Bedford, 1931 shows features of both hard and soft ticks, in fact it possess a partly sclerotized pseudo-scutum and an apical positioned capitulum (Keirans *et al.* 1976; El Shoura 1990). Ixodidae can be divided in two morphological groups on the basis of the anal groove position, the prostriata (including only the genus *Ixodes*) are considered the basal lineage and they possess the anal groove anteriorly to the anus; the metastriata (including all the other 11 genera) possess the anal groove posteriorly to the anus (Nava *et al.* 2009). Phylogenetic analyses based both on morphological (using up to 125 characters) and molecular characters (using different markers, the nuclear 18S rRNA and 28S rRNA genes, and the mitochondrial 12S rRNA and 16S rRNA genes and the 11 concatenated mitochondrial protein coding genes) agree in considering the tick clade as monophyletic (Klompen *et al.* 2000; Xu *et al.* 2003; Jayaprakash & Hoy 2009; Mans *et al.* 2011; Burger *et al.* 2012). Phylogenetic relationships among the three families remained unresolved until the recent work of Mans and colleagues (2011) that collected alive specimens of *N. namaqua* (Nuttalliellidae). The analyses based on the nuclear gene 18S rRNA for different parasitiformes species highlighted as basal the branch leading to *N. namaqua* indicating that the common ancestor shares features of both soft and hard ticks (Mans *et al.* 2011).

Excluding mosquitoes, able to transmit different species of *Plasmodium* (Apicomplexa), filarial nematodes and a variety of viruses, ticks are nowadays considered the most important group of arthropod vectors, capable of transmitting a variety of pathogens vertebrate animals including humans (Jongejan & Uilenberg 2004). Microorganisms such as bacteria (e.g. *Rickettsia* spp., *Borrelia burgdorferi* sensu lato, *Ehrlichia* spp. and *Francisella* spp.), protozoa (e.g. *Babesia* spp.) and viruses (like Crimean-Congo hemorrhagic fever, Tick Borne Encephalitis) can be transmitted to

the host as a result of a tick bite (Sonenshine 1991; Parola & Raoult 2001; Jongejan & Uilenberg 2004). In the last decades the scenarios become even more serious as result of the climate change and as a consequence, ticks expanded their distributional range especially towards higher altitudes and latitudes with the spreading of the transmitted pathogens even in highly populated environments (Lindgren *et al.* 2000; Cumming 2002; Danielova *et al.* 2006; Hubálek 2009; Pistone *et al.* 2011). Furthermore ticks can play an important role as reservoirs of these bacteria in nature (Parola & Raoult, 2001). In Europe and North America the Lyme disease, caused by the pathogenic *Borrelia burgdorferi* sensu lato and transmitted respectively by *Ixodes ricinus* (Linnaeus, 1758) and *Ixodes scapularis* Say, 1821, is the most prevalent arthropod-borne disease (Rizzoli *et al.* 2011; Wright *et al.* 2012). In Europe every years approximately 65.500 people suffer of Lyme disease symptoms (Hubálek 2009). The medical diagnosis of Lyme diseases (and in general of tick-borne diseases) could be intricate and only the appearance of erythema migrans rash could be decisive for the Lyme diagnosis in absence of laboratory analysis, that are anyway not always conclusive (Wright *et al.* 2012). In fact after the clinical examination and the patient anamnesis, the used laboratory diagnostic protocols involve blood sampling and serological analysis for indirect diagnosis; serological analysis requires the seroconversion time, and can thus be effected after no less than 6 weeks after the tick bite. This last feature associated with the fact that tick-borne pathogens are generally highly specific for a given tick genus (e.g. the etiological agent of the Lyme disease, *Borrelia burgdorferi*, can be transmitted only by the ticks of the genus *Ixodes*) highlight the needs on one hand to perform a PCR screening on the tick sample in order to evaluate the presence of pathogens, and on the other to correctly identify the tick taxon, even starting from a fraction of the whole organism obtained from patients that had not been properly preserved. Identification of tick samples based on morphology requires experience and a high level of knowledge of these arthropods, furthermore most discriminating morphological features are sometimes visible only in adult specimens, making species identification of immature stages not always possible. In this view, based on a DNA barcoding approach, I have developed a database of sequences of the mitochondrial gene *Cytochrome oxidase subunit I* (COI), useful for molecular identification of the species of the genus *Amblyomma* Koch, 1844. The genus *Amblyomma*, for which are nowadays described 131 species (Nava *et al.* 2009), encompasses species able to transmit a variety of pathogens in North and South America, like Q-fever and Brazilian spotted fever. In the field of bacteria transmitted by ticks, I have worked within a health care program of the United Nations (UNDP) on a preliminary study on bacterial pathogens transmitted by tick specimens collected from livestock in the Socotra Island (Yemen).

Concerning the ticks of the European fauna, as mentioned above the most widespread and important vector of diseases is *Ixodes ricinus*. Aside from the etiological agent of Lyme disease

Borrelia burgdorferi s.l., *I. ricinus* can transmit a variety of pathogens such as the virus of the Tick-borne encephalitis (TBE), *Babesia* spp., *Francisella tularensis* and different species of *Rickettsia* (Parola & Raoult 2001; Rizzoli *et al.* 2011). Moreover, the ability of *I. ricinus* to transmit pathogens is improved by the fact that it shows a low host specificity: larvae and nymphs parasitize small rodents, birds and lizards, while adults feed on wild mammals (as deer, roe deer, badger) as well as livestock and occasionally humans (Manilla 1998). Furthermore, the biological interests around this species increased since 2004 when an unknown α -proteobacterium was discovered in the intermembrane space (IMS) of *I. ricinus* ovarian cells mitochondria (Beninati *et al.* 2004), after named '*Candidatus* *Midichloria mitochondrii*' (hereafter *M. mitochondrii*; Sasser *et al.* 2006). Until now, *I. ricinus* is the only known Metazoa that harbors a bacterium in the IMS of its mitochondria. Recently the genome of *M. mitochondrii* was sequenced (Sasser *et al.* 2011). *M. mitochondrii* genome reveals interesting features, such as the capability to synthesize the flagellum and the cytochrome *cbb3* oxidases complex. In this field of research, during my PhD program, I have worked to: a) understand the phylogenetic position of *M. mitochondrii* within α -proteobacteria using a phylogenomic approach; b) perform phylogenetic analysis on the genes coding for the cytochrome oxidase *cbb3* complex, in order to test the hypothesis of descent by common ancestor of this complex versus horizontal gene transfer; c) perform a multi-gene phylogeny on the genes coding for the flagella apparatus in order to exclude the hypothesis of horizontal gene transfer, with the possible implications for the eukaryogenesis theory.

My phylogenetic-phylogenomic work on *M. mitochondrii* confirms that this bacterium belongs to the order of *Rickettsiales* Gieszczykiewicz, 1939 (as obtained by previous works based on 16S rRNA gene; Sasser *et al.* 2006; Vannini *et al.* 2010) and revealed that it belongs to a well separated lineages respect members of *Rickettsiaceae* Pinkerton 1936 and *Anaplasmataceae* Philip 1957. In addition to these results, recently published works aiming to characterize the microbioma of different organisms (based on 16S rRNA gene) found bacteria belonging to the *Midichloria* clade harboured by different hosts scattered in the eukaryotes tree of life, such as ticks, fleas, stink bugs but also cnidarians, amoebae, ciliates, and fishes (Fritsche *et al.* 1999; Beninati *et al.* 2004; Mediannikov *et al.* 2004; Erickson *et al.* 2009; Fraune and Bosch 2007; Lloyd *et al.* 2008; Vannini *et al.* 2010; Matsuura *et al.* 2012). In order to avoid a lack and confusion in the bacterial taxonomy, using a multidisciplinary approach based on almost complete ribosomal 16S rRNA sequences (i.e. phylogenetic inference methods, general mixed Yule-coalescent [GMYC] model, analysis of molecular variance, principal coordinates analysis and analysis of similarities [ANOSIM]), I demonstrated that members of *Midichloria* clade could deserve to be elevated at the family rank, and I submitted a manuscript with the proposal of a novel family of the *Rickettsiales*, i.e. the *Midichloriaceae* fam. nov..

Beside its medical and economic importance, in an evolutionary perspective *I. ricinus* can be considered as a unique model of a three levels relationship: the vertebrate host, the tick ectoparasite and the intra-mitochondrial bacterium *M. mitochondrii*. Up to now the biological role played by this bacterium and the reasons of its presence in the IMS of mitochondria is far from being understood. Only speculations can be formulated about its role as energetic scavenger or as energetic mutualist that could be involved in the reactive oxygen species metabolism during the tick blood meal. Within this framework I was focusing: to understand if the presence of *M. mitochondrii* in the IMS of mitochondria could have shaped the evolution of the tick mitochondrial genome; to understand the spatial distribution and the prevalence (i.e. total amount) of *M. mitochondrii* within the ovary of *I. ricinus* during its development life-cycle; to understand the effect of the removal of *M. mitochondrii* on the fitness of the tick (following to antibiotic treatment). In order to answer to the first question I sequenced the mitochondrial genome of *I. ricinus*, starting from a sample collected in a wild population, and performed comparative analyses respect others available ticks and arthropods mitochondrial genomes. To address the second question I conducted fluorescence in situ hybridization (FISH) essays in order to perform experiments on *I. ricinus* specimens at different life stages. Considering the long life cycle of *I. ricinus* (up to five years in nature and 5/6 months under laboratory conditions), the experiments on the tick fitness are still in progress.

Within populations of the sheep tick *I. ricinus*, interestingly, *M. mitochondrii* is harbored in 100% of females and 44% of males (Sassera *et al.* 2006); recently this bacterium was found to be widespread in many tick genera but with no evidence for host-symbiont co-cladogenesis (Epis *et al.* 2008; Dergousoff *et al.* 2011; Williams-Newkirk *et al.* 2012). During my PhD program I have worked also to test the hypothesis that the presence of *M. mitochondrii* in the IMS of *I. ricinus* mitochondria could have influenced the genetic structure of the latter at populations level (and viceversa) leading to host-symbiont genetic co-variation. At first the attention has been focused to population genetic of *I. ricinus* with the use of nuclear and mitochondrial gene markers. A Multilocus Sequence Typing (MLST) approach on five highly evolving genes of *M. mitochondrii* (data not reported in the present PhD thesis) was adopted for the second purpose.

1.1 References

- Beninati T, Lo N, Sacchi L, Genchi C, Noda H, Bandi C, 2004. A novel alpha-Proteobacterium resides in the mitochondria of ovarian cells of the tick *Ixodes ricinus*. *Appl Environ Microbiol* 70: 2596-2602.
- Burger TD, Shao R, Beati L, Miller H, Barker SC, 2012. Phylogenetic analysis of ticks (Acari: Ixodida) using mitochondrial genomes and nuclear rRNA genes indicates that the genus *Amblyomma* is polyphyletic. *Mol Phylogenet Evol.* 64: 45–55.
- Cumming GS, 2002. Comparing climate and vegetation as limiting factors for species ranges of African ticks. *Ecology* 83: 255–268.
- Danielova V, Rudenko N, Daniel M, Holubova J, Materna J, Golovchenko M, Schwarzova L, 2006. Extension of *Ixodes ricinus* ticks and agents of tick-borne diseases to mountain areas in the Czech Republic. *Int. J. Med. Microbiol.* 296: 48–53.
- Dergousoff SJ, Chilton NB, 2011. Novel genotypes of *Anaplasma bovis*, "Candidatus Midichloria" sp. and *Ignatzschineria* sp. in the Rocky Mountain wood tick, *Dermacentor andersoni*. *Vet Microbiol.* 150: 100–106.
- Epis S, Sasser D, Beninati T, Lo N, Beati L, Piesman J, Rinaldi L, McCoy KD, Torina A, Sacchi L, Clementi E, Genchi M, Magnino S, Bandi C, 2008. Midichloria mitochondrii is widespread in hard ticks (Ixodidae) and resides in the mitochondria of phylogenetically diverse species. *Parasitology* 135: 485–94.
- Erickson DL, Anderson NE, Cromar LM, Jolley A, 2009. Bacterial communities associated with flea vectors of plague. *J Med Entomol* 46: 1532–1536.
- El Shoura SM, 1990. *Nuttalliella namaqua* (Acarina: Ixodoidea: Nuttalliellidae) redescription of the female morphology in relation to the families Argasidae and Ixodidae. *Acarologia* 31: 349–355.
- Fraune S, Bosch TC, 2007. Long-term maintenance of species-specific bacterial microbiota in the basal metazoan *Hydra*. *Proc Natl Acad Sci U S A* 104: 13146–13151.
- Fritsche TR, Horn M, Seyedirashti S, Gautom RK, Schleifer KH, Wagner M, 1999. In situ detection of novel bacterial endosymbionts of *Acanthamoeba* spp. phylogenetically related to members of the order Rickettsiales. *Appl Environ Microbiol* 65: 206–212.
- Horak IG, Camicas JL, Keirans JE, 2002. The Argasidae, Ixodidae and Nuttalliellidae (Acari: Ixodida): a world list of valid tick names. *Exp Appl Acarol* 28: 27–54.
- Hubálek Z, 2009. Epidemiology of Lyme borreliosis. *Curr Probl Dermatol.* 37: 31–50.

- Jeyaprakash A, Hoy MA, 2009. First divergence time estimate of spiders, scorpions, mites and ticks (subphylum: Chelicerata) inferred from mitochondrial phylogeny. *Exp Appl Acarol.* 47: 1–18.
- Jongejan F, Uilnberg G, 2004. The global importance of ticks. *Parasitology* 129: S3–S14.
- Keirans JE, Clifford CM, Hoogstraal H, Easton ER, 1976. Discovery of *Nuttalliella namaqua* Bedford (Acarina: Ixodoidea: Nuttalliellidae) in Tanzania and redescription of the female based on scanning electron microscopy. *Ann Entomol Soc Amer* 69: 926–932.
- Keirans JE, 2009. Order Ixodida. In Krantz G.W., Walter D.E. (Eds.) *A manual of acarology* 3rd ed, Texas Tech Univ. Press, pp. 111–123.
- Klompen, JSH, Black WC, Keirans JE and Norris DE, 2000. Systematic and biogeography of hard ticks, a total evidence approach. *Cladistic* 16: 79–102.
- Lindgren E, Tälleklint L, Polfeldt T, 2000. Impact of climatic change on the northern latitude limit and population density of the disease-transmitting European tick *Ixodes ricinus*. *Environ. Health. Perspect.* 108:119–123.
- Lloyd SJ, LaPatra SE, Snekvik KR, St-Hilaire S, Cain KD, Call DR, 2008. Strawberry disease lesions in rainbow trout from southern Idaho are associated with DNA from a *Rickettsia*-like organism. *Dis Aquat Organ* 82: 111–118.
- Manilla G, 1998. Fauna d'Italia Ixodoidea. Calderini, Bologna, Italia: 280 pp.
- Mans BJ, de Klerk D, Pienaar R, Latif AA, 2011. *Nuttalliella namaqua*: a living fossil and closest relative to the ancestral tick lineage: implications for the evolution of blood-feeding in ticks. *PLoS One* 6: e23675.
- Matsuura Y, Kikuchi Y, Meng XY, Koga R, Fukatsu T, 2012. Novel clade of alphaproteobacterial endosymbionts associated with stinkbugs and other arthropods. *Appl Environ Microbiol.* 78:4149–4156.
- Mediannikov O, Ivanov LI, Nishikawa M, Saito R, Sidel'nikov Iu N, Zdanovskaia NI, Mokretsova EV, Tarasevich IV, Suzuki H, 2004. Microorganism "Montezuma" of the order Rickettsiales: the potential causative agent of tick-borne disease in the Far East of Russia. *Zh Mikrobiol Epidemiol Immunobiol*: 7-13.
- Nava S, Guglielmo AA, Mangold AJ, 2009. An overview of systematics and evolution of ticks. *Front. Biosci.* 14: 2857–2877.
- Parola P, Raoult D, 2001. Tick-borne bacterial diseases emerging in Europe. *Clin. Microbiol. Infect.* 7: 80–83.
- Pistone D, Pajoro M, Fabbi M, Vicari N, Marone P, Genchi C, Novati S, Sassera D, Epis S, Bandi C, 2010. Lyme borreliosis, Po River Valley, Italy. *Emerg Infect Dis.* 16:1289–1291.

- Poinar GO, Brown AE, 2003. A new genus of hard ticks in Cretaceous Burmese amber (Acari: Ixodida: Ixodidae). *Syst Parasitol* 54: 199–205.
- Poinar GO, 1995. First fossil soft ticks, *Ornithodoros antiquus* n. sp. (Acari: Argasidae) in Dominican amber with evidence of their mammalian host. *Experientia* 51: 384–387.
- Rizzoli A, Hauffe H, Carpi G, Vourc HG, Neteler M, Rosa R, 2011. Lyme borreliosis in Europe. *Euro Surveill* 16: 19906.
- Sassera D, Beninati T, Bandi C, Bouman EA, Sacchi L, Bandi C, 2006. *Candidatus* *Midichloria mitochondrii*, an endosymbiont of the tick *Ixodes ricinus* with a unique intramitochondrial lifestyle. *Int J Syst Evol Microbiol* 56: 2535–2540.
- Sassera D, Lo N, Epis S, D'Auria G, Montagna M, Comandatore F, Horner D, Peretó J, Luciano AM, Franciosi F, Ferri E, Crotti E, Bazzocchi C, Daffonchio D, Sacchi L, Moya A, Latorre A, Bandi C, 2011. Phylogenomic evidence for the presence of a flagellum and *cbb(3)* oxidase in the free-living mitochondrial ancestor. *Mol Biol Evol.* 28: 3285–3296.
- Sonenshine DE, 1991. *Biology of ticks*. Vol. 1. New York: Oxford Univ. Press. 447 pp.
- Sonenshine DE, 1993. *Biology of ticks*. Vol. 2 New York: Oxford Univ. Press. 465 pp.
- Vannini C, Ferrantini F, Schleifer KH, 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 Environ Microbiol* 76: 4047–4054.
- Williams-Newkirk AJ, Rowe LA, Mixson-Hayden TR, Dasch GA, 2012. Presence, genetic variability, and potential significance of "*Candidatus* *Midichloria mitochondrii*" in the lone star tick *Amblyomma americanum*. *Exp Appl Acarol.* 58: 291–300.
- Wright WF, Riedel DJ, Talwani R, Gilliam BL, 2012. Diagnosis and management of Lyme disease. *Am Fam Physician.* 85: 1086–1093.
- Xu G, Fang QQ, Keirans JE, Durden L, 2003. Molecular phylogenetic analyses indicate that the *Ixodes ricinus* complex is a paraphyletic group. *J. parasitol.* X: 452-457.

2 RESULTS

2.1 Research article. *Evaluating the accuracy of DNA barcoding for species identification in the tick genus AMBLYOMMA Koch 1844 (Acari: Ixodidae)*

2.1.1 Summary

DNA barcoding, recently introduced as a tool for rapid identification of organisms usually at species level, is sometimes considered to be preferable to traditional phenotypic identification because it does not require extensive taxonomic morphological expertise. In addition, in some cases the DNA barcoding has proven to be useful in revealing taxonomic critical situations or cryptic species in morphologically undistinguishable taxa. Because of the importance of hard ticks vectors in human and veterinary medicine, the implementation of an alternative method for tick identification, particularly for personnel not familiar with tick taxonomy, is appealing. Nevertheless, barcoding methods cannot be applied blindly to any group of taxa, without first being validated. Herein, we present a first evaluation of the DNA barcoding method for ticks belonging to the genus *Amblyomma* Koch 1844.

2.1.2 Manuscript

Introduction

DNA barcoding is becoming a popular molecular method for the identification of species and the demarcation of cryptic species (Hebert *et al.* 2003; Hebert *et al.* 2004; Hajbabaei *et al.* 2006, Hendrich *et al.* 2010). The method is generally based on the comparison of a fragment of 650 base pairs (bp) of the mitochondrial gene *Cytochrome c oxidase subunit I* (COI), considered to be the “universal” specific marker for metazoan taxa (Hebert *et al.* 2003; Hebert *et al.* 2004). Ideally, a genetic marker for DNA barcoding must combine low intraspecific sequence variability with high divergences between closely related taxa (Ward *et al.* 2005; Hajbabaei *et al.* 2006). The morphological identification of ticks (Acari: Ixodida) is often time consuming and requires extensive expertise, only gained after years of practice. A further advantage of the DNA barcoding approach is the possibility to be applied on immature specimens or fragments of ticks.

Ticks, obligate blood feeding ectoparasites, are a group of taxa, which includes major livestock pests and important vectors of pathogens of humans and animals (Jongejan and Uilenberg 2004; Barros-Battesti *et al.* 2006). The metastriate hard-tick genus *Amblyomma* Koch, 1844 encompasses

about 130 species (Nava *et al.* 2009). The genus is more diverse and species-rich in the Southern hemisphere, but has a North-South range extending approximately between 40° of latitude N and S (Robinson 1926; Barros-Battesti *et al.* 2006). Species of this genus are usually three-host ticks and show a wide variety of host belonging to different orders of mammals, reptiles, and even amphibians. Birds are prevalently hosts of immature stages (nymphs and larvae) (Robinson 1926). Several species of *Amblyomma* are important vectors of a number of diseases of medical and veterinary interest (Jongejan and Uilenberg 2004). *A. cajennense* (Fabricius, 1787) and *A. aureolatum* (Pallas, 1772) are recognized vectors of *Rickettsia rickettsii*, the etiological agent of Rocky Mountain (also called Brazilian) spotted fever, one of the most severe spotted fever group rickettsioses (Labruna *et al.* 2008). Other *Amblyomma* species of medical importance are *A. americanum* (Linnaeus, 1758) vector of *Ehrlichia chaffeensis* and *Francisella tularensis*; *A. hebraeum* Koch, 1844 and *A. variegatum*, vectors of *R. africae*, agent of African tick-bite fever (Kelly *et al.* 1996), and *Ehrlichia ruminantium* (Walker and Olwage, 1987). In order to better grasp the importance of each tick species in transmitting diseases, it is crucial to be able to identify specimens found on infected humans or animals correctly. Although the adults of *Amblyomma* ticks have been extensively studied, morphological taxonomic keys for the immatures of the genus are at best fragmentary (Martins *et al.* 2010; Keirans and Durden, 1998).

The main goals of this study are: a) to evaluate DNA barcoding as species-specific identification tool for hard ticks of the genus *Amblyomma*, by analyzing a number of specimens previously identified morphologically and by sequencing of additional molecular genes; 2) test the reliability of DNA barcoding in cryptic species detection, taking into account the “species complex” of *A. cajennense* (Fabricius, 1787) (hereafter “*A. cajennense* complex”) recently discovered to be formed of 5 species (study based on morphological characters; data not published).

Materials and Methods

- Biological samples, DNA extraction, PCR conditions, and sequencing -

Our sample included 74 of the known *Amblyomma* species. Most species (44/74; 59%) were represented by more than one specimen. For the remaining species only one sequence was obtained either because only one specimen was available, or because the specimens used for extraction were old and did not yield any usable DNA material. Although for barcoding purposes it is best to have at least 3-5 sequences/species, we decided to include the species represented by a single sequence because they provided a better representation of the whole genus. In addition, we analyzed DNA from 44 specimens of *Amblyomma cajennense*, from 12 different geographical localities, which have recently been used to redefine an *A. cajennense* taxonomic complex.

DNA was extracted from each specimen by using DNeasy Blood & Tissue Kit (Qiagen, Chatsworth, California) and following the modified protocol in Beati and Keirans (2001). A fragment of 620-640 bp of the mitochondrial gene, which encodes for *Cytochrome c oxidase subunit I* (COI), was amplified with primers Chelicerate-F1 (5' – TACTCTACTAATCATAAAGACATTGG – 3') and Chelicerate R1 (5'- CCTCCTCCTGAAGGGTCAAAAAATGA – 3') (Barrett and Hebert, 2005). For PCR, 2 µl of tick DNA were amplified in a 20 µl reaction mix containing 1.25 µl of each primers (10pmole/µl), 5µl TaqMaster Enhancer (5x), 2.5 µl Taq Buffer with Mg²⁺ (10x, with 1.5 mM Mg²⁺), 1 µl of dNTPs (0.2 mM), 0.2 µl of Taq DNA Polymerase (5U/µl) (MasterTaq Kit, 5PRIME, Gaithersburg, MD - USA), and 6.8µl of molecular grade H₂O. The PCR conditions were as follows: 5 min of denaturation at 94°C, 35 cycles of denaturation at 94°C for 20 sec, annealing at 54.5°C for 45 sec, and elongation at 72°C for 45 sec. The elongation was completed by a further step at 72°C for 5 min and the PCR products were then kept refrigerate at 4 C. The complementary strands of each amplicon were sequenced by using the same primers used for PCR at the Ht-Seq High Throughput Genomics Center (University of Washington, Seattle, WA). The complementary strands were assembled by using Sequencher 4.10.1 and the fragments corresponding to the primers sequences were excised before further analysis. The resulting 218 nucleotide consensus sequences, were translated into corresponding peptide sequences using the tool Transeq in EMBOSS in order to verify the quality of the sequences through the reading frames (Rice *et al.* 2000). The DNA sequences were deposited in the EMBL Data Library according to the EBI Barcoding Procedure. The sequence alignment was obtained automatically with Muscle (Edgar 2004) and refined manually and according to codon organization by using MEGA 5 (Tamura *et al.* 2011). No gaps are present into the final alignment. Nucleotides dataset composition were performed with MEGA 5.1 (Tamura *et al.* 2011).

- *Definition of analysis datasets* -

The 218 DNA sequences of the specimens, previously identified morphologically, were organized in three datasets (hereafter referred as A, B and C) for the analyses to be performed. These three dataset were analyzed with a DNA barcoding analysis (see paragraph below) to test: 1) coherence between morphological and molecular identification (dataset A); 2) the reliability of the approach in detecting the 5 cryptic species of "*A. cajennense* complex" (*dataset B and C*). Dataset A is formed by 44 species (43 species of *Amblyomma* and 1 species of *Bothriocroton* Keirans, King & Sharrad, 1994), for total of 188 sequences of 604 bp long. Specimens of "*A. cajennense* complex" were treated as belonging to the 5 species of the complex (*A. cajennense* Fabricius, 1787 and 4 n. sp). Dataset B is formed by all the 70 species (66 species of *Amblyomma* and 4 species of *Bothriocroton*), with the specimens of "*A. cajennense* complex" treated as belonging to one single

species. Dataset C is formed by 74 species (70 species of *Amblyomma* and 4 species of *Bothriocroton*) for a total of 218 sequence of 604 bp long. Specimens of "*A. cajennense* complex" were treated as belonging to the 5 species of the complex (*A. cajennense* Fabricius, 1787 and 4 n. sp).

- DNA barcoding analysis -

Standard DNA barcoding analysis, based on comparison between intraspecific and interspecific nucleotide divergence (Ferri et al. 2009), were performed on the 604 bp fragment of COI, to compare the "molecular identification method" (DNA barcoding) with the traditional morphological ones and also to evaluate the DNA barcoding as a tool to detect critic situations in hard tick systematics.

The nucleotide sequence divergences were calculated using the Kimura-two-parameter (K2P) model (Kimura 1980), considered as an adequate evolutionary nucleotide model when p-distances between sequences are low (Nei & Kumar 2000). These analyses were implemented in MEGA 5 (Tamura et al. 2011). The distributions of intraspecific and interspecific nucleotide divergences were used to evaluate an intra/inter-species molecular threshold. The comparison of nucleotide divergence from two specimens could generate two types of error: false positive (type I) when conspecific specimens show a genetic divergence greater than the threshold value; false negative (type II) when genetic divergence between specimens from two different species is lower than the threshold value (Ferri et al. 2009). I used the cumulative error plots to evaluate the optimum threshold values of nucleotide divergence (OT), that correspond to the value that minimizes the function Σ [error type I + error type II] (Wiemers and Fiedler 2007). In order to further test the procedure of tick identification, I used a leave-one-out test based on similarity score, obtained as complement to 1 of the nucleotide distance between each specimen. This analysis was performed in order to check if each specimen has the other conspecific specimens as the "best similar". I also examined the relationships among the taxa using a phenetic approach. Phenetic trees were generated with Unweighted Pair Group Method with Arithmetic Mean (UPGMA), Neighbour joining (NJ) and Maximum likelihood (ML) approaches (options UPGMA: K2P; options NJ: K2P, 1000 bootstrap). For ML the evolutionary model best-fitting to the analyzed datasets were selected with jModeltest 0.1.1 (Posada 2008) according to the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The nucleotide substitution model selected for the further ML analysis was the General Time Reversible (GTR; Lanave et al. 1984) with gamma distribution (Γ). Inference with UPGMA and NJ were implemented in MEGA 5 (Tamura et al. 2011), while ML analysis was performed in PhyML version 3.0 (Guindon et al. 2010).

Results

DNA barcoding analysis on dataset A shows a mean nucleotide distance within species of 1.79% (sd= 2.59) and a mean nucleotide distance between species of 18.57% (sd= 3.1). The intra/inter species mean nucleotide distance for dataset B are respectively 9.64% (sd= 7.3) and 18.79% average (sd= 3.2). The analysis for all the 218 sequences of *Amblyomma* and *Bothriocroton* species (dataset C) show a mean intraspecific nucleotide distance 1.78% (sd=2.6) and a mean interspecific genetic distance of 18.9% (sd=3.1).

The leave-one-out test on dataset A shows that all the 187 specimens included in this dataset (belonging to 44 species) have as the genetic “best similar” the conspecific specimens. The cumulative error plot on dataset C (figure 1) show an optimal intra/inter species threshold at nucleotide distance value of 9%.

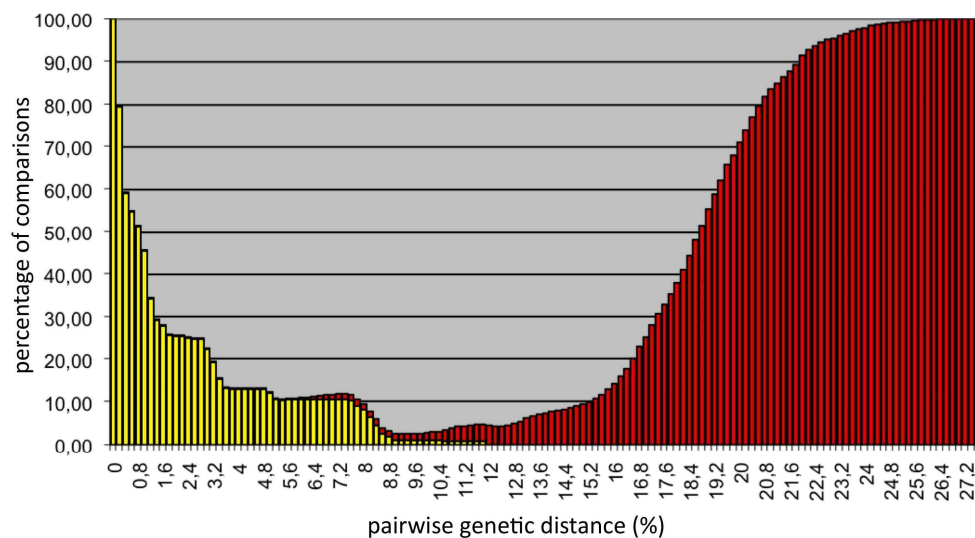


Figure 1. Cumulative error plot on dataset C for the evaluation of the Optimal threshold (OT). X and Y axis represent respectively genetic divergence (%) and cumulative errors. Yellow and red bars show respectively intraspecific and interspecific comparison.

The frequency distribution plot of intraspecific and interspecific genetic divergences in morphologically identified tick specimens of dataset B (i.e. considering the specimens of “*A. cajennense* complex” as members of the same species) reported in figure 2 underlines some critical situations (yellow plots between 13-18.6% pairwise genetic distance). In fact the specimens of “*A. cajennense* complex” show a intra-specific nucleotide distance (17.5%) higher than the dataset

mean value, falling within the range of variability for inter-specific nucleotide distance. Considering specimens of “*A. cajennense* complex” as members of the four new identified species plus *A. cajennense* (dataset C), the intra-specific nucleotide distance resulted in the range of within-species variability (figure 3).

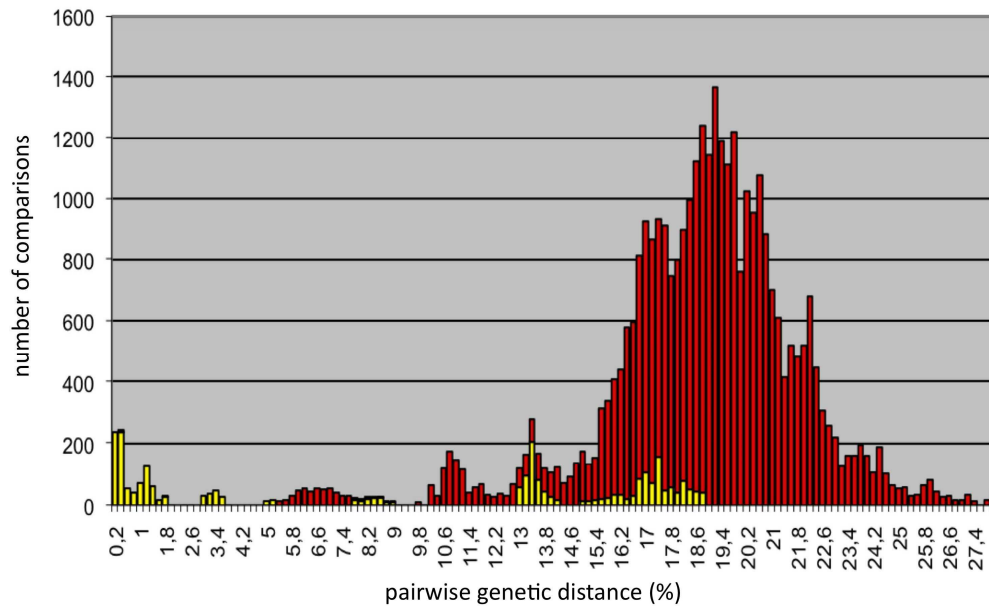


Figure 2. Distance graph of *coxI* *Amblyomma* ticks, dataset B. Frequency distribution of intraspecific (yellow) and interspecific (red) genetic divergences in morphologically identified ticks. The yellow plots in the graph, from value 13% of genetic distance and 18.6% are the specimens of *A. cajennense* considered as single species. X and Y axis represent respectively genetic divergence (%) and number of comparisons.

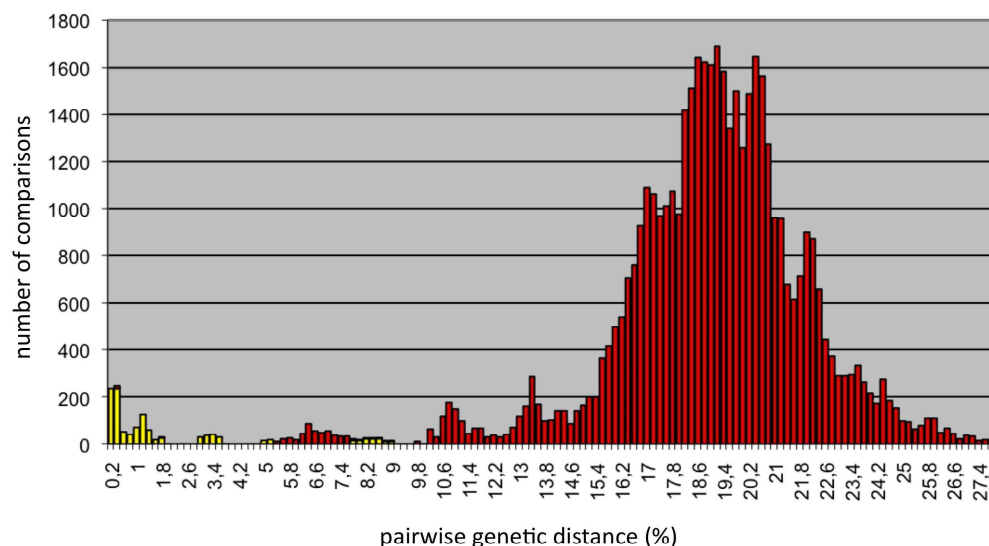


Figure 3. K2P distance graph of *coxI* *Amblyomma* ticks, dataset C. Frequency distribution of intraspecific (yellow) and interspecific (red) genetic divergences in morphologically identified ticks. X and Y axis represent respectively genetic divergence (%) and number of comparisons.

DNA barcoding analysis, based on nucleotide genetic threshold, also highlights that specimens of *A. oblongoguttatum* and *A. parvum* show an intraspecific nucleotide mean distance of 10.2% and 12% respectively, out of the range of intraspecific genetic divergence.

The phylogenetic tree inferred with ML approach on the dataset C is reported in figure 4-6) and in newick format in Appendix 1. Most of the tick species resulted monophyletic.

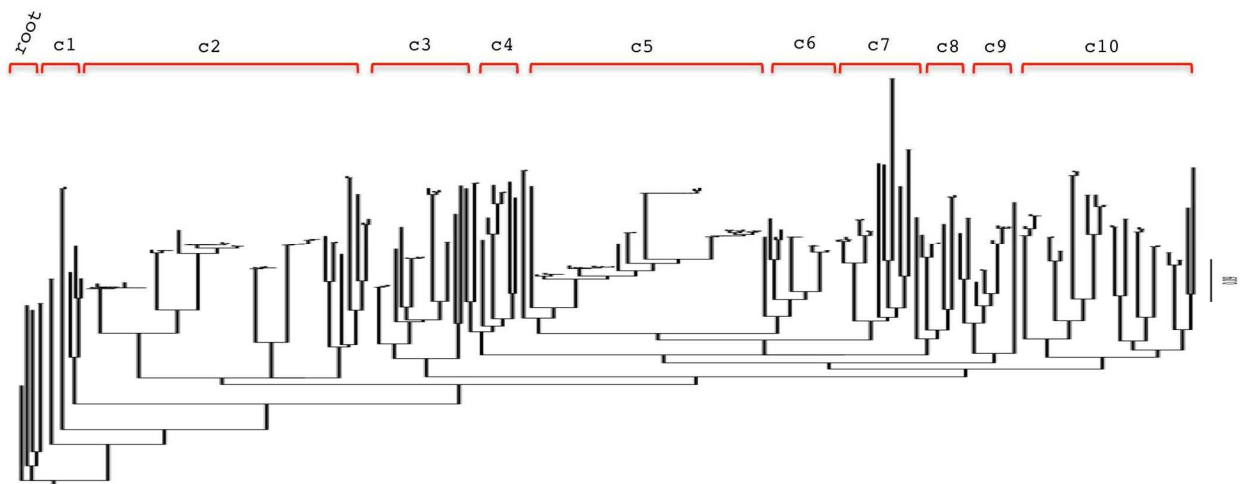


Figure 4. Maximum likelihood tree, obtained from the COI sequences of C dataset, depicting the identified cluster (the species of the genus *Bothriocroton* as root plus cluster from c1 to c10). Clusters are highlighted by red brackets.



Figure 5. Enlargement of maximum likelihood tree obtained from the COI sequences of C dataset. The root plus clusters from one to six are reported. Names on the tips represent the species groups.

In figure 5, within the cluster 5, are reported the two cases in which molecular analysis disagree with morphology: *A. dissimile* and *A. scutatum*. Two separate lineages were identified for *A. dissimile*, one in Argentina/Brazil and one in Southern Mexico. Based on my analysis (on the basis of a 650 bp of COI gene), *A. scutatum* resulted not monophyletic, in fact *A. rotundatum* clusters between two different populations of *A. scutatum*. It is interesting to note that the species group *A. rotundatum*-*A. scutatum* are considered a cryptic clade also by morphological point of view (Beati pers. communication). Excluding these two taxa, all the remaining species groups detected by phenetic analyses were consistent with the current knowledge on the taxonomy of the genus *Amblyomma*.

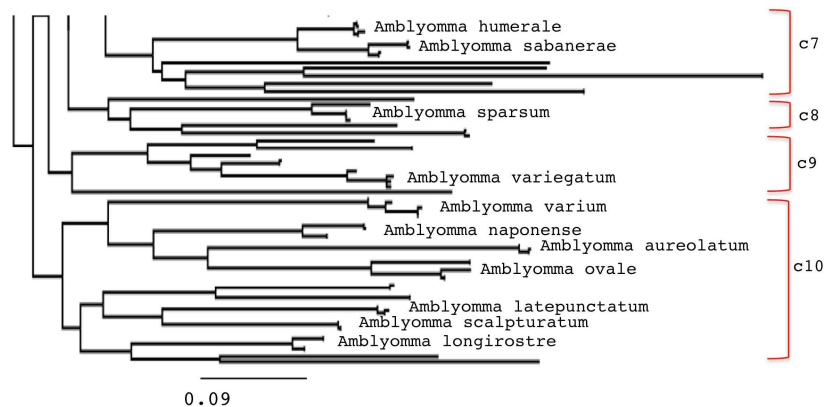


Figure 6. Enlargement of maximum likelihood tree obtained from the COI sequences of C dataset. The clusters from seven to ten are reported. Names on the tips represent the species groups.

Discussion

In this study I tested the DNA barcoding on 70 species of ticks of the genus *Amblyomma* and 4 species of genus *Bothriocroton*. The analysis show that DNA barcoding approach is useful in the identification of ticks at species level, in fact all the species represented in the dataset by more than one specimen were correctly identified by this approach. The same type of control was made also on the full dataset of species. I found that DNA barcoding analysis could also be used as a first tool to uncover cryptic species and could represent a method to reveal critical taxonomic situations. In fact the DNA barcoding approach, based on 650 bp fragment of COI highlight the case of the “*A. cajennense* complex”, recently solved with an integrated approach (i.e. morphology, ecology and genetics). The DNA barcoding analysis on my datasets points out the case of *A. oblongoguttatum*

and *A. parvum*. These two species show an intraspecific nucleotide mean distance of 10.2% and 12%, value that are above the intra-inter species optimal threshold. The phylogenetic approach underline the non monophyletic status of the two species *A. rotundatum* and *A. scutatatum*. The four species are considered by taxonomists problematic species that could hide criptic species (Beati, personal communication). An approach to test this possibility could be the application of the general mixed Yule-coalescent model (GMYC; Pons *et al.* 2006) for delimiting species from single-locus DNA trees; in addition, an approach to better analyze the genetic variability of these taxa could be the analysis of different genetic markers (in particular at the nuclear level) and examine more specimens from different geographic areas.

2.1.3 References

Beati L and Keirans JE, 2001. Analysis of the systematic relationship amongs ticks of the genera *Rhipicephalus* and *Boophilus* (Acari: Ixodidae) based on mitochondrial 12S ribosomal DNA gene sequences and morphological characters. *J Parasitol* 87: 32-48

Barrett RDH and Hebert PDN 2005. Identifying spiders through DNA barcodes. *Can. J. Zool.* 83: 481-491.

Barros-Battesti DM, Arzua M, Bechara GH, 2006. Carrapatos de Importancia Médico-Veterinaria de Regiao Neotropical: Um guia ilustrado para identificacao de especies. Sao Paulo, Vox/ICTTD-3/Butantan, 223 pp.

Ferri E, Barbuto M, Bain O, Galimberti A, Uni S, Guerrero R, Ferté H, Bandi C, Martin C, Casiraghi M, 2009. Integrated taxonomy: traditional approach and DNA barcoding for the identification of filarioid worms and related parasites (Nematoda). *Front Zool* 6: 1.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O, 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59: 307-321.

Hajbabaie M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN, 2006. DNA barcodes distinguish species of tropical Lepidoptera. *Proc Natl Acad Sci Unit States Am* 103: 968–971.

Hall TA 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 41: 95-98.

Hebert PDN, Cywinska A, Ball SL *et al.* 2003 Biological identifications through DNA barcodes. *Proc Roy Soc Lond B Biol Sci* 270: 313-321.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W, 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proc Natl Acad Sci Unit States Am* 101: 14812-14817.

- Hebert PDN, Stoeckle MY, Zemplak TS, Francis CM, 2004b. Identification of birds through DNA barcodes. *PLoS Biology* 2: e312.
- Hendrich L, Pons J, Ribera I, Balke M, 2010. Mitochondrial *cox1* sequence data reliably uncover patterns of insect diversity but suffer from high lineage-idiosyncratic error rates. *PLoS One*. 5(12):e14448.
- Jongejan F, Uilenberg G, 2004. The global importance of ticks. *Parasitology* 129: S3-S14.
- Kimura M 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16: 111-120.
- Labruna MB, Ogrzewalska M, Martins TF, Pinter A, Horta MC, 2008. Comparative Susceptibility of Larval Stages of *Amblyomma aureolatum*, *Amblyomma cajennense*, and *Rhipicephalus sanguineus* to Infection by *Rickettsia rickettsii*. *J Med Entomol* 45: 1156-1159
- Lanave C, Preparata G, Saccone C, Serio G, 1984. A new method for calculating evolutionary substitution rates. *J Mol Evol* 20: 86-93.
- Nava S, Guglielmo AA, Mangold J, 2009. An overview of systematics and evolution of ticks. *Front Biosci* 14: 2857-2877
- Nei M, Kumar S, 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.
- Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, Kamoun S, Sumlin WD, Vogler AP, 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst Biol* 55: 595-609.
- Posada D, 2008. jModelTest: phylogenetic model averaging. *Mol Biol Evol* 25: 1253-1256.
- Rice P, Longden I, Bleasby A, 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 16: 276-277.
- Robinson LE, 1926. The genus *Amblyomma*. Ticks, A Monograph of the Ixodoidea, Part 4 (ed. by G. H. F. Nuttall, C. Warburton and L. E. Robinson), pp. 155–188. Cambridge University Press, Cambridge.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S, 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731-2739.
- Ward RD, Zemplak TS, Innes BH, Last PR, Hebert PD, 2005. DNA barcoding Australia's fish species. *Philos Trans R Soc Lond B Biol Sci* 360: 1847–1857.
- Wiemers M, Fiedler K, 2007. Does the DNA barcoding gap exist? a case study in blue butterflies (Lepidoptera: Lycaenidae). *Front Zool* 4: 8.

2.1.4 Appendix

Maximum Likelihood tree (newick format):

```
((((Amblyomma_triguttatum_123210:0.34441,(Amblyomma_albolineatum_83002:0.20899,Amblyomma_limbatum_122622:0.07757)0.99:0.22716)0.97:0.19112,((Amblyomma_clypeolatum_49429:0.60911,((((Amblyomma_tuberculatum_187:0.0,Amblyomma_tuberculatum_rostal:0.00348)1.00:0.47158,(Amblyomma_usingeri_1:0.06076,(Amblyomma_pilosum_64603:0.07708,(Amblyomma_macfarlandi_f1:0.0,Amblyomma_macfarlandi_f3:0.00364)0.75:0.04266)0.97:0.13026)1.00:0.35583)0.82:0.10255,(((Amblyomma_pacae_7767b:0.32611,((Amblyomma_calcaratum_6789:0.16807,(Amblyomma_calcaratum_50:0.01244,Amblyomma_calcaratum1:0.03944)0.96:0.08204)0.99:0.14809,((Amblyomma_dubitatum_7396D:0.0,(Amblyomma_dubitatum_7396C:0.0,(Amblyomma_dubitatum_coe7396:0.0,Amblyomma_dubitatum_7396B:0.0)0.00:0.0)0.00:0.0)1.00:0.21608,((Amblyomma_coelebs:0.00356,(Amblyomma_coelebs_113123:0.0,Amblyomma_coelebs_123909:0.0)0.00:0.0)0.92:0.02403,(Amblyomma_coelebs_7645B:0.0,Amblyomma_coelebs_7645A:0.0)0.69:0.00622)0.99:0.16146)0.35:0.0358)0.87:0.06762)0.84:0.05182,((Amblyomma_flavomaculatum_122543:0.42061,Amblyomma_exophthalmum_123673:0.47334)0.86:0.12731,(Amblyomma_argentinae_mx25_1:0.50597,((Amblyomma_dissimile_mx7_1:0.0071,((Amblyomma_dissimile_MX14:0.0,Amblyomma_dissimile_MX15:0.0)0.81:0.00354,(Amblyomma_dissimile_Mx6:0.0,(Amblyomma_dissimile_Mex6063:0.0,Amblyomma_dissimile_MX5:0.0)0.00:0.0)0.00:0.0)0.82:0.00734)0.76:0.01222,((Amblyomma_dissimile_hansL_1:0.01178,((Amblyomma_dissimile_119249:0.0036,(Amblyomma_dissimile_MX20:0.00352,Amblyomma_dissimile_Mx17:0.0)0.00:0.0)0.00:0.0,((Amblyomma_dissimile_mx16_1:0.0,Amblyomma_dissimile_Mx13:0.0)0.86:0.00352,(Amblyomma_dissimile_Mx11:0.00352,(Amblyomma_dissimile_120001_ex1f11:0.0,Amblyomma_dissimile_ex3_hon1:0.00351)0.00:0.0)0.00:0.0)0.85:0.00351)0.67:0.0026)0.94:0.02895,(Amblyomma_scutatum_Mx12:0.10904,((Amblyomma_scutatum_006045_2:0.0,(Amblyomma_scutatum_Mex60451:0.0,Amblyomma_scutatum_Mx22:0.00348)0.00:0.0)1.00:0.12551,((Amblyomma_rotundatum_Mx9:0.0,(Amblyomma_rotundatum_Mx8:0.0,(Amblyomma_rotundatum_Mx4:0.0,(Amblyomma_rotundatum_Mx3:0.0,(Amblyomma_rotundatum_Mx1:0.0,(Amblyomma_rotundatum_MX10:0.0,(Amblyomma_rotundatum_1:0.0,((Amblyomma_rotundatum_118473_Ex5br1:0.0,(Amblyomma_rotundatum_120600:0.0,(Amblyomma_rotundatum_mx2_1:0.0,(Amblyomma_rotundatum_123118:0.00727,Amblyomma_rotundatum_731022Darci:0.01062)0.90:0.01035)0.00:0.0)0.00:0.0)0.00:0.0)0.00:0.0)0.00:0.0)0.00:0.0)0.00:0.0)0.00:0.0)1.00:0.26649,(Amblyomma_scutatum_119299:0.0,(Amblyomma_scutatum_119297_87E:0.0,(Amblyomma_scutatum_06055_341854:0.0,(Amblyomma_scutatum_119297_1:0.0,((Amblyomma_scutatum_mex23_1:0.00406,Amblyomma_scutatum_006051:0.00348)0.87:0.00815,(Amblyomma_scutatum_006064_2:0.00354,Amblyomma_scutatum_006064_1:0.00711)0.78:0.00571)0.65:0.00711,(Amblyomma_scutatum_123047139:0.00358,(Amblyomma_scutatum_151cr_1:0.00358,Amblyomma_scutatum_lacr:0.00354)0.75:0.0036)0.85:0.00915)0.80:0.00535)0.00:0.0)0.00:0.0)0.79:0.00354)0.97:0.0928)0.70:0.01574)0.66:0.02872)0.85:0.02126)0.73:0.01068)0.93:0.08241)0.82:0.07351)0.76:0.03364)0.76:0.0293,((((Amblyomma_humerale_Fg1:0.0,Amblyomma_humerale_Fg2:0.00363)0.74:0.00443,(Amblyomma_humerale_118042_F:0.01125,Amblyomma_humerale_7679a:0.0)0.75:0.00668)0.92:0.09299,((Amblyomma_sabanerae_cr1:0.00252,Amblyomma_sabanerae_cr1m:0.00458)0.97:0.06429,(Amblyomma_sabanerae_canada81:0.00359,Amblyomma_sabanerae_Canada2:0.0)0.69:0.01383)0.61:0.11288)0.98:0.23012,(Amblyomma_albopictum_123034_1:0.41583,(Amblyomma_antillorum_1174814:0.0,Amblyomma_antillorum_25:0.0)1.00:0.70623)0.90:0.23506)0.73:0.0625)0.85:0.04425)0.73:0.01993,(Amblyomma_pattoni_123065:0.4579,(Amblyomma_breviscutatum:0.60009,(Amblyomma_babirussae_99038:0.0,Amblyomma_babirussae_99038_F10:0.00356)1.00:0.50449)0.66:0.03608)0.98:0.14884)0.73:0.02409,(Amblyomma_testudinarium_50074:0.41737,(Amblyomma_tollonii_119481:0.49609,((Amblyomma_nuttallii_116888:0.09397,(Amblyomma_sparsum_2zim:0.0,(Amblyomma_sparsum_zim157:0.0,Amblyomma_sparsum_zim139:0.00717)0.00:0.0)0.92:0.05385)1.00:0.28039,(Amblyomma_chabaudi_122317:0.34103,(Amblyomma_latum_121580:0.00403,Amblyomma_latum_17_20:0.00711)1.00:0.4487)0.53:0.0876)0.00:
```

0.0282)0.62:0.0379)0.72:0.02003)0.90:0.0377,((((Amblyomma_gemma_88460:0.18017,(Amblyomma_haebraeum_sa_71:0.0,Amblyomma_haebraeum_zim:0.0)1.00:0.24877)0.95:0.15116,(Amblyomma_cohaerens_38418:0.08936,((Amblyomma_splendidum_Coh164:0.00377,Amblyomma_splendidum_zim162:0.0)0.98:0.09422,(Amblyomma_variegatum_ta19m:0.01464,(Amblyomma_variegatum_ma11f:0.01156,(Amblyomma_variegatum_ma8m:0.00734,Amblyomma_variegatum_usi5f:0.00755)0.00:0.0)0.99:0.06362)1.00:0.19787)0.88:0.05331)0.91:0.08672)0.97:0.13699,((Amblyomma_varium_112852:0.0,(Amblyomma_varium_117764:0.01066,(Amblyomma_varium_8922203_1:0.00709,(Amblyomma_varium_cnc4970f:0.0,Amblyomma_varium_ibsp8922:0.0)0.00:0.0)0.99:0.05105)0.88:0.02758)1.00:0.40751,(((Amblyomma_naponense_101:0.0,Amblyomma_naponense_102:0.00373)0.97:0.09721,(Amblyomma_naponense_124017:0.0,Amblyomma_naponense_Frenchguyi:0.0)0.81:0.03849)0.99:0.23941,((Amblyomma_aureolatum_28marcelo:0.0,(Amblyomma_aureolatum_30marcelo:0.00371,Amblyomma_aureolatum_1397:0.0)0.91:0.01517)1.00:0.48059,((Amblyomma_ovale_7134:0.0,Amblyomma_ovale_isbp71348:0.0)0.94:0.15347,(Amblyomma_ovale_018402cr_1:0.04698,(Amblyomma_ovale_123527:0.0,Amblyomma_ovale_123527II:0.0)0.61:0.00595)0.91:0.11446)0.99:0.26242)0.84:0.08863)0.81:0.06921)0.74:0.04883)0.63:0.04344,((((Amblyomma_brasiliense_7171_1:0.00717,Amblyomma_brasiliense_cnc522_1:0.0)1.00:0.26879,(Amblyomma_incisum_cnc659_1:0.0,Amblyomma_incisum_cnc659_2:0.0)1.00:0.31483)0.97:0.17945,(((Amblyomma_latepunctatum_labru:0.0,(Amblyomma_latepunctatum_labru23:0.00753,Amblyomma_latepunctatum_labru2:0.0)0.49:0.01147)1.00:0.33482,(Amblyomma_sculpturatum_b3:0.0,(Amblyomma_sculpturatum_labr2:0.0,Amblyomma_sculpturatum_labru2:0.0)0.76:0.00376)0.99:0.28901)0.82:0.05775,(Amblyomma_torrei_123609f:0.36865,(Amblyomma_rhinocerotis_b:0.0,Amblyomma_rhinocerotis_a:0.0)1.00:0.51469)0.89:0.13859)0.77:0.03409)0.59:0.04991,(((Amblyomma_longirostre_anle32_1:0.0,Amblyomma_longirostre_122698_1:0.0)0.94:0.04942,(Amblyomma_longirostre_120064:0.0,Amblyomma_longirostre_7322:0.0)0.76:0.01848)0.99:0.25264,(Amblyomma_geayi_117373_1:0.36128,Amblyomma_parkeri_Labruna_1:0.50168)0.60:0.14147)0.88:0.09098)0.71:0.02619)0.76:0.01906,(((Amblyomma_pseudoconcolor_7436_124:0.0,(Amblyomma_pseudoconcolor_7436:0.0,(Amblyomma_auricularium_3581:0.0,Amblyomma_auricularium_6973_22:0.00369)0.85:0.00369)0.00:0.0)1.00:0.2125,((Amblyomma_parvum_12838_2:0.0,(Amblyomma_parvum_11905711_1:0.00327,(Amblyomma_parvum_12838_1:0.00329,Amblyomma_parvum_495cr_1:0.00325)0.79:0.00328)0.00:0.0)1.00:0.20181,((Amblyomma_pseudoparvum_123652:0.32155,Amblyomma_parvum_123529:0.26019)0.74:0.05018,((Amblyomma_maculatum_122675:0.0255,(Amblyomma_triste_f39:0.00365,Amblyomma_triste_32ibsp7297:0.00756)0.63:0.01008)1.00:0.44291,(Amblyomma_inornatum_1jeff:0.0,Amblyomma_inornatum_2jeff:0.0)0.99:0.2316)0.88:0.07816)0.62:0.02319)0.94:0.09486)0.72:0.05135,(Amblyomma_sylvaticum_119155:0.45904,Amblyomma_neumanni_1:0.54228)0.84:0.13627)0.90:0.08329)0.80:0.0255)0.14:0.02775)0.78:0.02691,((((Amblyomma_mixtum_ecu106:0.0,(Amblyomma_mixtum_CRXIBN:0.00334,(Amblyomma_mixtum_Mex1:0.01399,(Amblyomma_mixtum_Mex8:0.0,(Amblyomma_mixtum_Mex6:0.0,(Amblyomma_mixtum_Mex2:0.0,Amblyomma_mixtum_Mex5:0.0)0.00:0.0)0.00:0.0)0.00:0.0)0.81:0.00334)0.00:0.0,(Amblyomma_mixtum_tx2_1:0.02199,(Amblyomma_mixtum_CRXIIBF:0.0,(Amblyomma_mixtum_CRXF:0.0,(Amblyomma_mixtum_CRVIII AF:0.0,Amblyomma_mixtum_ecu107:0.0)0.00:0.0)0.00:0.0)0.00:0.0)0.00:0.0)0.80:0.00334)0.99:0.1744,((Amblyomma_cajennense_rom4:0.0122,(Amblyomma_cajennense_124080_3:0.0,(Amblyomma_cajennense_Fgdrag3:0.0,(Amblyomma_cajennense_fg124080_1:0.0,Amblyomma_cajennense_124017:0.0)0.00:0.0)0.00:0.0)0.75:0.00927)1.00:0.22442,(Amblyomma_sculptum_mgm4_1:0.09704,((Amblyomma_sculptum_Arg12:0.0,((Amblyomma_sculptum_Arg8:0.00334,(Amblyomma_sculptum_Arg9:0.0,(Amblyomma_sculptum_Arg15:0.0,(Amblyomma_sculptum_Arg14:0.0,(Amblyomma_sculptum_Arg13:0.0,Amblyomma_sculptum_Arg10:0.0)0.00:0.0)0.00:0.0)0.00:0.0)0.00:0.0)0.00:0.0,(Amblyomma_sculptum_sam4:0.00682,Amblyomma_sculptum_spf6_1:0.00344)0.87:0.00671)0.00:0.0)0.83:0.01316,(Amblyomma_sculptum_122965_1:0.00337,(Amblyomma_sculptum_122954:0.0,Amblyomma_sculptum_122965:0.0)0.00:0.0)0.76:0.00756)0.69:0.02126)1.00:0.22848)0.49:0.09148)0.99:0.17807,(((Amblyomma_interandinum_Pe59:0.0,(Amblyomma_interandinum_Pe50:0.00342,((Amblyomma_interandinum_Pe11:0.0,Amblyomma_interandinum_Pe61:0.00341)0.86:0.00341,(Amblyomma_interandinum_Pe51:0.0,Amblyomma_interandi

num_Pe58:0.0)0.00:0.0)0.00:0.0)0.72:0.00341)0.99:0.2833,(Amblyomma_tonelliae_Arg5:0.0,(Amblyomma_tonelliae_Arg6:0.0,(Amblyomma_tonelliae_Arg7:0.0,(Amblyomma_tonelliae_Arg11:0.0,(Amblyomma_tonelliae_Arg4:0.0,(Amblyomma_tonelliae_Arg3:0.0,(Amblyomma_tonelliae_Arg123780:0.0,Amblyomma_tonelliae_Arg2:0.0)0.00:0.0)0.84:0.00366)0.98:0.01499)0.80:0.00364)0.00:0.0)0.79:0.00364)1.00:0.39176)0.94:0.14044,((Amblyomma_oblongoguttatum_01691cr1:0.18825,(Amblyomma_oblongoguttatum_124016:0.0,Amblyomma_oblongoguttatum_124016_II:0.0035)0.98:0.15732)0.99:0.2623,(Amblyomma_pecarium_48958:0.36101,((Amblyomma_americanum_123785:0.00666,Amblyomma_americanum:0.0)1.00:0.53304,(Amblyomma_tapirellum_117089:0.35035,(Amblyomma_imitator_123601:0.00757,Amblyomma_imitator_123598_76:0.02249)0.97:0.22442)0.61:0.11551)0.89:0.13954)0.00:0.00801)0.96:0.12684)0.00:0.0)0.81:0.0322)0.88:0.06896)0.92:0.09046,((Amblyomma_sphenodonti_123587:0.0,Amblyomma_sphenodonti_123589:0.0)1.00:0.66706,(Bothriocroton_hydrosauri:0.38116,(Bothriocroton_undatum_HH62276:0.6313,(Bothriocroton_concolor_hh62244_1:0.55223,(Bothriocroton_oudemansi_123614m:0.0,Bothriocroton_oudemansi_12361F:0.0)1.00:0.58801)0.26:0.06332)0.34:0.05877)0.90:0.14154)0.85:0.06888)0.63:0.33759,(Amblyomma_elaphense_123651:0.0,Amblyomma_elaphense_123651_IIF:0.00749)0.97:0.57704,Amblyomma_transversale_121579:1.85401);

2.2 Research article. Screening for bacterial DNA in the hard tick *Hyalomma marginatum* (Ixodidae) from Socotra Island (Yemen): detection of *Francisella*-like endosymbiont

2.2.1 Summary

Thirty-four adult ticks collected from livestock on Socotra Island (Yemen) were identified as *Hyalomma marginatum* using traditional morphological characteristics. Morphological identification was confirmed for all the collected specimens using a molecular approach targeting a fragment of the mitochondrial gene 12S rRNA. All the specimens were examined for the presence of tick-borne pathogens and the tick endosymbiont "*Candidatus* Midichloria mitochondrii" using polymerase chain reaction. Three specimens out of the 34 analyzed tested positive to the presence of *Francisella* spp. leading to the first detection of these bacteria in *H. marginatum* on Socotra Island. The phylogenetic analyses conducted on a 660 bp fragment of the ribosomal gene 16S rRNA of *Francisella* spp. (including *F. philomiragia* as outgroup, the four subspecies of *F. tularensis* and the *Francisella*-like endosymbiont of ticks) confirm that the newly detected *Francisella* strains cluster into the *Francisella*-like endosymbionts of ticks. Interestingly, the detected *Francisella*-like endosymbiont, shows a different genotype to that previously isolated from *H. marginatum* collected in Bulgaria. No specimen was positive for the presence of *Rickettsia* spp., *Coxiella burnetii*, *Borrelia burgdorferi* or *M. mitochondrii*.

2.2.2 Manuscript

Introduction

The ticks are blood-sucking ectoparasites able to parasite a multitude of terrestrial vertebrates as mammals, birds, reptiles and amphibians (Sonenshine, 1991 and 1993). Nowadays ticks are considered the group of arthropods that can transmit the wider variety of pathogenic agents to humans and animals (Jongejan and Uilenberg, 2004). Microorganisms such as bacteria (e.g. *Rickettsia* spp., *Borrelia burgdorferi*, *Erlichia* spp. and *Francisella* spp.), protozoa and viruses (like Crimean-Congo hemorrhagic fever, Tick Borne Encephalitis) can be transmitted to host as a result of a tick bite (Sonenshine, 1991, Jongejan and Uilenberg, 2004). Ticks play also an important role as reservoirs for population of these bacteria in nature (Parola and Raoult, 2001). Recently, the intra mitochondrial bacterium "*Candidatus* Midichloria mitochondrii" (hereafter *M. mitochondrii*), originally discovered in the tick *Ixodes ricinus* (for which was recently sequenced the mitochondrial genome, Montagna *et al.* 2012), was found to be widespread in many tick genera (Lo *et al.*, 2006; Epis *et al.*, 2008).

Approximately 870 species of ticks are described (Nava *et al.*, 2009), subdivided into three families: Argasidae, Ixodidae and Nuttalliellidae (Horak *et al.*, 2003; Nava *et al.*, 2009). The 26 species belonging to the genus *Hyalomma* Koch, 1844 are widespread in Palearctic and Afrotropical biogeographic regions (Horak *et al.*, 2003; Apanaskevich and Horak, 2008; Estrada-Peña *et al.*, 2012). In details *Hyalomma marginatum* Koch, 1844 is widespread in Central and Southern Europe, Northern Africa and in Asia east to Iran (Manilla, 1998; Apanaskevich and Horak, 2008). *H. marginatum* is a two-host species showing a low host specificity, in fact the adults feed on different species of large mammals (ungulates and livestock), while the immature stage feed on birds or small mammals (Manilla, 1998) increasing its ability to spread. This species can transmit a variety of pathogens for human and animal (Hoogstraal, 1956) and it is considered one of the most important tick species involved in the transmission of the virus of the Crimean-Congo haemorrhagic fever (Hoogstraal, 1979; Estrada-Peña *et al.*, 2012); furthermore it is known to transmit bacteria of the genus *Rickettsia* (e.g. *R. conori* the causative agent of the Mediterranean spotted fever) and *Coxiella burnetii*, the causative agent of Q-fever (Hoogstraal, 1956). Recently Ivanov and colleagues (2011) isolates *Francisella*-like endosymbionts (FLEs) in *H. marginatum* collected from Bulgaria.

The paper deals with a first study on ticks (Acarina, Ixodida) collected from Socotra Island (this Indian Ocean archipelago has a peculiar fauna since it has been isolated 35–41 Million Years ago; Girdler and Styles, 1974) reporting the bacterial community associated with these ticks in this area.

Material and Methods

- Sample collection, morphological identification and images acquisition -

A total of 34 adult ticks specimens were collected in Socotra Island (Yemen) directly from livestock (sheep and goats) during field research (December 2010). All the collected specimens were immediately stored in absolute ethanol for further DNA extraction. Genomic DNA was extracted from all specimens individually following a procedure that allow to preserve the morphology for further analyses. Specimens manipulation were completed using the stereo microscope Leica MS5. All ticks were identified using standard taxonomic keys (Starkoff, 1958; Manilla, 1998; Apanaskevich and Horak, 2008). Male and female images were acquired by a machinery made and optimized in order to scan the sample at different focus layers that were mounted with Zerene Stacker 1.0 64 bit (Student Edition).

- DNA extraction and polymerase chain reaction (PCR) -

Total genomic DNA was extracted and purified individually using Qiagen DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany). All the tick preserved in ethanol were washed with distilled water and dried before DNA extraction. Afterwards, ticks were cut with a scalpel along the idiosome and left for 12 hours at 56°C into 180 µl of ATL lysis buffer (Qiagen) with 200 ng/ml proteinase K (Sigma Aldrich, St Louis, USA). The following extraction steps were performed according to the manufacturer's instructions. Extracted DNAs were quantified with Nanodrop 1000 (Thermo Scientific, Wilmington, USA). In order to confirm the morphological identification of the ticks, a fragment of the mitochondrial ribosomal small subunit 12S rRNA gene was amplified (Beati and Keirans, 2001) and sequenced for all the samples. The extracted DNAs were examined, for the presence of *Francisella* spp., *Rickettsia* spp., *C. burnetii*, *B. burgdorferi* and *M. mitochondrii*, using specific PCR protocols. The primers used for the screening of bacterial species are reported in table 1. PCR amplification were performed in 25 µl reaction mix containing 1 µl of each primers (1 µM), 5 µl of GoTaq reaction Buffer (1x) with Mg²⁺ (1.5 mM MgCl₂), 0.5 µl of dNTPs (0.2 mM each dNTP), 0.2 µl of GoTaq DNA Polymerase (1.25 U). Successful amplification was determined by gel electrophoresis. Positive and unambiguous PCR products were directly sequenced in both strand by ABI technology (Applied Biosystems, Foster City, CA, USA). The obtained sequences were manually corrected using Geneious Pro 5.3 and deposited in the EMBL data library (accession numbers HE819515 for *H. marginatum* partial 12S rRNA gene and HE819516 for *Francisella*-like endosymbiont partial 16S rRNA gene).

Table 1. Primers used for bacterial screening in the present study.

Organism	Target gene	Primer sets (5' - 3')	References
<i>Borrelia burgdorferi</i>	16S rRNA	ATGCACACTTGGTGTAACTA GACTTATCACCGGCAGTCTTA	Marconi <i>et al.</i> 1992
<i>Coxiella burnetii</i>	Transposon-like repetitive region	TATTGTATCCACCGTAGCCAGTC CCCAACAACACCTCCTTATTC	Willems <i>et al.</i> 1994; Berri <i>et al.</i> 2000
<i>Francisella</i> spp.	16S rRNA	CAAGGTTAATAGCCTTGGGGGA GCCTTGTCAGCGGCAGTCTTA	Forsman <i>et al.</i> 1994
<i>Midichloria mitochondrii</i>	16S rRNA	GTACATGGGAATCTACCTTGC CAGGTCGCCCTATTGCTTCTTT	Epis <i>et al.</i> 2008
<i>Rickettsia</i> spp.	Citrate synthase <i>gltA</i>	GCAAGTATCGGTGAGGATGTAAT GCTTCCTTAAATTCATAAATCAGGAT	Labruna <i>et al.</i> 2004

- Bioinformatic and phylogenetic analyses -

The tick mitochondrial 12S rRNA and the bacterial 16S rRNA consensus sequences obtained by sequencing were subjected to BLAST analysis (<http://www.ncbi.nlm.nih.gov/blast>) and compared to the sequences available in GeneBank (<http://www.ncbi.nlm.nih.gov/genbank/>). A 16S rRNA sequences of *Francisella* spp. was retrieved from GeneBank in order to perform phylogenetic

analyses. Sequences belonging to the four subspecies of *Francisella tularensis*, *Francisella*-like endosymbiont (FLEs) of tick and other *Francisella* spp. were included in the dataset (accession numbers are reported on the figure 2). The obtained 18 sequences were aligned using MUSCLE (Edgar, 2004) then trimmed with Gblocks (Castresana, 2000) and analyzed with jModelTest 0.1.1 (Posada, 2008) to choose the most suitable model of nucleotide evolution. Phylogenetic reconstructions were performed with Bayesian inferences using MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001). Bayesian analyses were performed using GTR (Lanave *et al.*, 1984) as model of evolution + I+ G; two parallel analyses, each composed of one cold and three incrementally heated chains were run for 2.5 million generations. Trees were sampled every 100 generations and burn-in fraction was calculated as 25% of total sampled trees, according to the lnL stationary analyses. The majority rule consensus tree was rooted with the branch leading to *F. philomiragia* and *F. noatunensis*, node with values of Bayesian Posterior Probability (BPP) less than 0.5 were collapsed. Pairwise p-distance was calculated between the FLEs sequence obtained in this work and the two closely FLEs of *Rhipicephalus sanguineus* (HQ705171) and of *H. m. marginatum* (HQ705170).

Results

All the collected ticks, 11 males and 23 semi-engorged females, were morphologically identified as *Hyalomma marginatum*. Male and female images in dorsal view were reported in Figure 1 (respectively B and A).

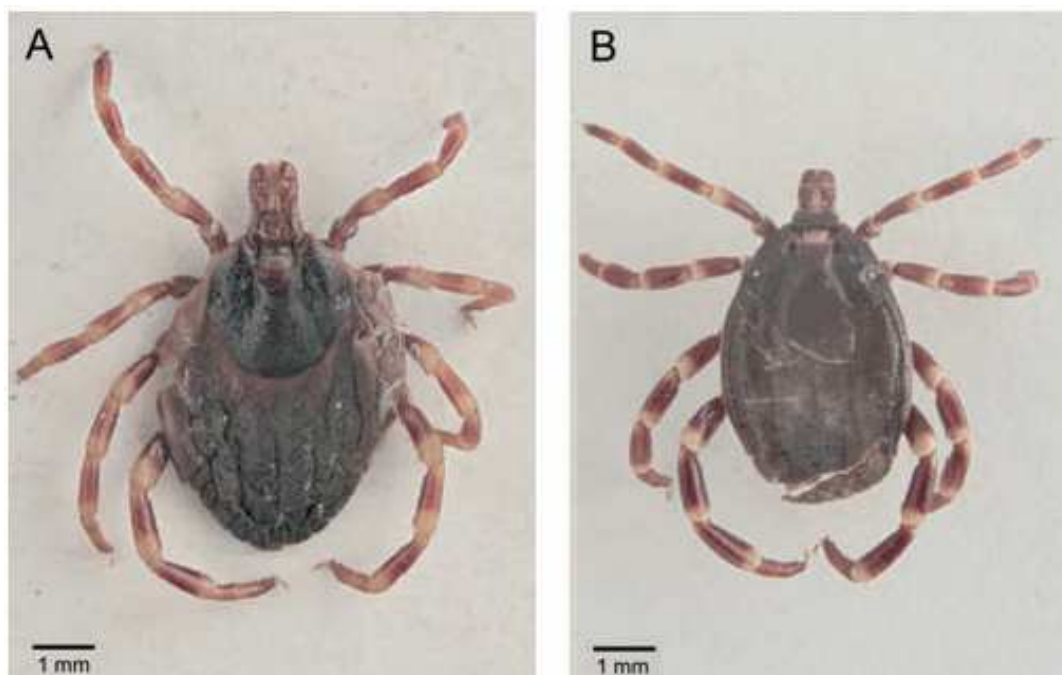


Figure 1. Dorsal view of *Hyalomma marginatum*: (A) female and (B) male.

The DNA extracted from the 34 specimens, quantified by Nanodrop 1000, result in concentration ranging from 40 to 110 ng/μL. All tick samples were positive for 12S rRNA PCR; the PCR products were sequenced and morphological identification were confirmed by BLAST analysis (100% identity with *H. marginatum marginatum*, accession number AF150034).

All the specimens tested negative in PCR for the presence of *Rickettsia* spp., *Coxiella burnetii*, *Borrelia burgdorferi* and *M. mitochondrii*; while three specimens collected from two different hosts (2 females and 1 male, 8.8% of prevalence) were positive for *Francisella* 16S rRNA amplicons. No nucleotide differences were recovered between the three consensus sequences after a pairwise comparison. BLAST analysis performed on the three sequences confirms their identity (99%) with *Francisella*-like endosymbiont. The best hit of the sequences resulted to the FLEs of *Rhipicephalus sanguineus* (HQ705171) and of *H. m. marginatum*, (HQ705170), from which they differs by one nucleotide (pairwise p-distance = 0.15%).

Phylogenetic analyses were performed on a dataset of a 660 bp of the bacterial 16S rRNA composed of eighteen taxa belonging to *Francisella* spp. from different origin (e.g. pure culture, soil samples, seawater, tick endosymbionts) in order to understand the relationships of the newly sequenced bacterial strains. Bayesian analysis (Figure 2) confirm that *Francisella* spp. harboured by *H. marginatum* collected from livestock in Socotra Island cluster within the group of tick FLEs. In detail, the new sequence clusters with a BPP of 1 within a well supported group formed by two FLEs previously detected from *H. marginatum* and *Rhipicephalus sanguineus* collected from Bulgaria.

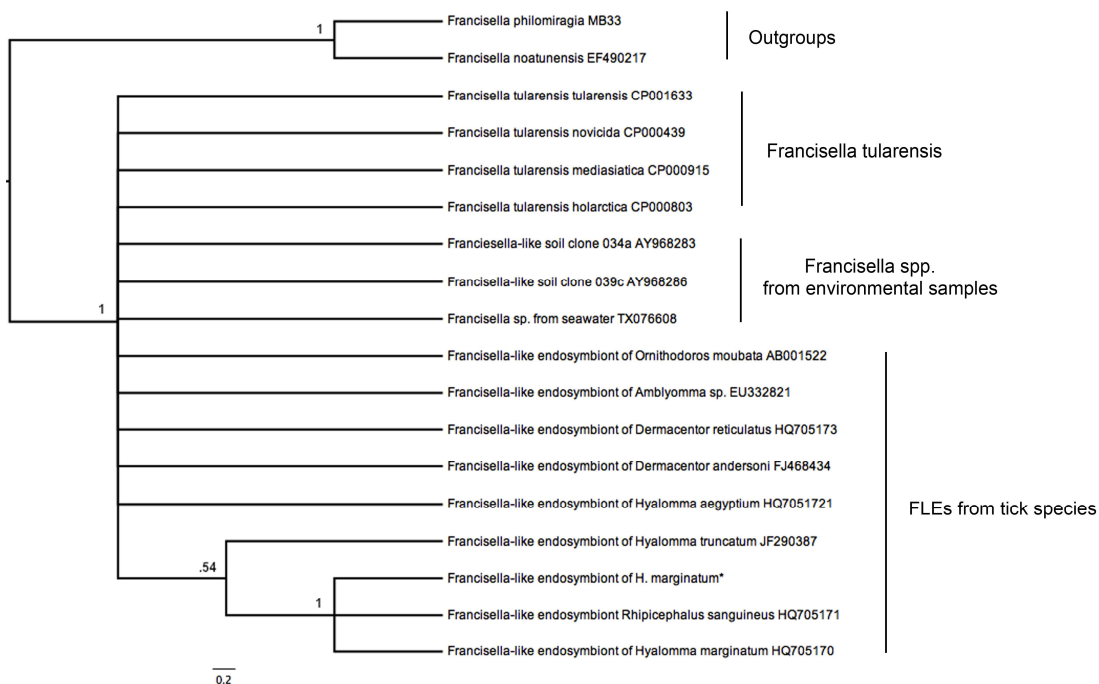


Figure 2. Bayesian consensus cladogram of *Francisella* spp. 16S rRNA gene, Bayesian posterior probability (BPP) values are reported above each node; branches with BPP values less than 0.5 were collapsed. *The sequence obtained in the present study.

Discussion

This study is the first detection of bacteria harboured in tick collected from livestock in Socotra Island. Gram-negative bacteria belonging the genus *Francisella* are known to be distributed mainly in Northern Hemisphere (Foley and Nieto, 2010). Within this group there are bacteria of medical and veterinary importance as the etiologic agent of Tularemia, *F. tularensis* and the FLEs of tick. At present FLEs have been identified in both soft (genus *Ornithodoros*) and hard ticks (*Amblyomma*, *Dermacentor*, *Rhipicephalus*, *Hyalomma*), furthermore their pathogenic rule is unknown even if genes implicated in the pathogenicity of *F. tularensis* have been detected (Machado-Ferreira *et al.*, 2009). The obtained results disagree with the trees obtained in previous work that have suggested the monophyly of the *Francisella*-like tick endosymbionts. This discrepancies is probably due to the fact that we have used a fragment of the 16S rRNA that is not enough informative to resolve these relationships.

Considering the importance to human health of bacteria of the genus *Francisella* our results need to be taken in consideration for the control of new emerging diseases in Socotra Island. In fact, this case provides the first evidence of *Francisella* in Socotra Island and should alert physicians and veterinarians working within the region to the possibility of infection with this organism. Further researches are necessary to elucidate the role that ticks might play as natural reservoirs or vectors of *Francisella* in this country.

2.2.4 References

- Apanaskevich DA, Horak IG, 2008. The genus *Hyalomma* Koch, 1844: V- re-evaluation of the taxonomic rank of taxa comprising the H. (*Euhyalomma*) *marginatum* Koch complex of species (Acari: Ixodidae) with redescriptions of all parasitic stages and notes on biology. *Int. J. Acarol.* 34: 13-42.
- Beati L, Keirans JE, 2001. Analysis of the systematic relationships among ticks of the genera *Rhipicephalus* and *Boophilus* (Acari: Ixodidae) based on mitochondrial 12S ribosomal DNA gene sequences and morphological characters. *J. Parasitol.* 87 (1): 32-48.
- Berri M, Laroucau K, Rodolakis A, 2000. The detection of *Coxiella burnetii* from ovine genital swabs, milk and fecal samples by the use of a single touchdown polymerase chain reaction. *Vet. Microbiol.* 72 (3-4): 285-293.
- Castresana J, 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17 (4): 540-552.

- Edgar RC, 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. 5:113.
- Epis S, Sasser D, Beninati T, Lo N, Beati L, Piesman J, Rinaldi L, McCoy KD, Torina A, Sacchi L, Clementi E, Genchi M, Magnino S, Bandi C, 2008. *Mitochondria* is widespread in hard ticks (Ixodidae) and resides in the mitochondria of phylogenetically diverse species. *Parasitology* 135 (4): 485-94.
- Estrada-Peña A, Jameson L, Medlock J, Vatansever Z, Tishkova F, 2012. Unraveling the Ecological Complexities of Tick-Associated Crimean-Congo Hemorrhagic Fever Virus Transmission: A Gap Analysis for the Western Palearctic. *Vector Borne Zoonotic Dis.* 12: 10.1089/vbz.2011.0767.
- Foley JE, Nieto NC, 2010. Tularemia. *Vet. Microbiol.* 140 (3-4): 332-338.
- Girdler RW, Styles P, 1974. Two stage Red Sea floor spreading. *Nature* 247: 7-11.
- Forsman M, Sandstrom G, Sjostedt A, 1994. Analysis of 16s Ribosomal DNA Sequences of Francisella Strains and Utilization for Determination of the Phylogeny of the Genus and for Identification of Strains by PCR. *Int. J. Syst. Bacteriol.* 44 (1): 38-46.
- Hoogstraal H, 1956. African Ixodoidea. I. Ticks of the Sudan (with special reference to Equatoria Province and with preliminary reviews of the genera *Boophilus*, *Margaropus* and *Hyalomma*). U.S. Navy, Washington D.C., US: 1101 pp.
- Hoogstraal H, 1979 The epidemiology of tick-borne Crimean-Congo hemorrhagic fever in Asia, Europe and Africa. *J. Med. Entomol.* 15 (4): 307-417.
- Horak IG, Camicas JL, Keirans JE, 2003. The Argasidae, Ixodidae and Nuttalliellidae (Acari: Ixodida): a world list of valid tick names. *Exp. Appl. Acarol.* 28 (1-4): 27-54.
- Huelsenbeck JP, Ronquist F, 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*. 17(8):754-755.
- Ivanov IN, Mitkova N, Reye AL, Hübschen JM, Vatcheva-Dobrevska RS, Dobрева EG, Kantardjiev TV, Muller CP, 2011. Detection of new Francisella-like tick endosymbionts in *Hyalomma* spp. and *Rhipicephalus* spp. (Acari: Ixodidae) from Bulgaria. *Appl. Environ. Microbiol.* 77 (15): 5562-5565.
- Jongejan F, Uilnberg G, 2004. The global importance of ticks. *Parasitology* 129: S3–S14.
- Labruna MB, Whitworth T, Horta MC, Bouyer DH, McBride JW, Pinter A, Popov V, Gennari SM, Walker DH, 2004. *Rickettsia* species infecting *Amblyomma cooperi* ticks from an area in the state of São Paulo, Brazil, where Brazilian spotted fever is endemic. *J. Clin. Microbiol.* 42 (1): 90-98.
- Lanave C, Preparata G, Saccone C, Serio G, 1984. A new method for calculating evolutionary substitution rates. *J. Mol. Evol.* 20: 86-93.

- Lo N, Beninati T, Sassera D, Bouman EA, Santagati S, Gern L, Sambri V, Masuzawa T, Gray JS, Jaenson TG, Bouattour A, Kenny MJ, Guner ES, Kharitonov IG, Bitam I, Bandi C, 2006. Widespread distribution and high prevalence of an alpha-proteobacterial symbiont in the tick *Ixodes ricinus*. *Environ. Microbiol.* 8: 1280-1287.
- Machado-Ferreira E, Piesman J, Zeidner NS, Soares CA, 2009. Francisella-like endosymbiont DNA and Francisella tularensis virulence-related genes in Brazilian ticks (Acari: Ixodidae). *J. Med. Entomol.* 46 (2): 369-374.
- Manilla G, 1998. Fauna d'Italia Ixodoidea. Calderini, Bologna, Italia: 280 pp.
- Marconi RT, Garon CF, 1992. Identification of a third genomic group of *Borrelia burgdorferi* through signature nucleotide analysis and 16S rRNA sequence determination. *J. Gen. Microbiol.* 138 (3): 533-536.
- Nava S, Guglielmo AA, Mangold AJ, 2009. An overview of systematics and evolution of ticks. *Front. Biosci.* 14: 2857-2877.
- Parola P, Raoult D, 2001. Tick-borne bacterial diseases emerging in Europe. *Clin. Microbiol. Infect.* 7 (2): 80-83.
- Posada D, 2008. jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* 25 (7): 1253-1256.
- Sonenshine DE, 1991. Biology of ticks, Volume 1. Oxford Univ. Press., New York, US: 447 pp.
- Sonenshine DE, 1993. Biology of ticks, Volume 2. Oxford Univ. Press., New York, US: 465 pp.
- Starkoff O, 1958. Ixodoidea d'Italia, studio monografico. Il pensiero scientifico, Roma, Italia: 384 pp.
- Willems H, Thiele D, Frhlich-Ritter R, Krauss H, 1994. Detection of *Coxiella burnetii* in cow's milk using the polymerase chain reaction (PCR). *J. Vet. Med. Ser. B* 41 (9): 580-587.

2.3 Research article. *Phylogenomic evidence for the presence of a flagellum and cbb3 oxidase in the free-living mitochondrial ancestor*

2.3.1 Summary

The initiation of the intracellular symbiosis that would give rise to mitochondria and eukaryotes was a major event in the history of life on earth. Hypotheses to explain eukaryogenesis fall into two broad and competing categories: those proposing that the host was a phagocytotic proto-eukaryote that preyed upon the free-living mitochondrial ancestor (hereafter FMA) and those proposing that the host was an archaeobacterium that engaged in syntrophy with the FMA. Of key importance to these hypotheses are whether the FMA was motile or non-motile, and the atmospheric conditions under which the FMA thrived. Reconstructions of the FMA based on genome content of Rickettsiales representatives - generally considered to be the closest living relatives of mitochondria - indicate that it was non-motile and aerobic. We have sequenced the genome of *Candidatus Midichloria mitochondrii*, a novel and phylogenetically divergent member of the Rickettsiales. We found that it possesses unique gene sets found in no other Rickettsiales, including 26 genes associated with flagellar assembly, and a *cbb₃*-type cytochrome oxidase. Phylogenomic analyses show that these genes were inherited in a vertical fashion from an ancestral α -proteobacterium, and indicate that the FMA possessed a flagellum, and could undergo oxidative phosphorylation under both aerobic and microoxic conditions. These results indicate that the FMA played a more active and potentially parasitic role in eukaryogenesis than currently appreciated, and provide an explanation for how the symbiosis could have evolved under low levels of oxygen.

2.3.2 Manuscript

Introduction

The symbiosis that would ultimately give rise to mitochondria and their eukaryotic hosts is recognized as one of the major transitions in the history of life on earth (Margulis 1970; Maynard Smith and Szathmary 1997; de Duve 2005; Lane and Martin 2010). All eukaryotes examined thus far have been shown to contain mitochondria, or modified versions of this organelle (hydrogenosomes or mitosomes) (Bui *et al.* 1996; Tovar *et al.* 2003; Embley and Martin 2006). In the wake of the genomic revolution, it is widely accepted that mitochondria arose only once, from a free-living bacterium that took up residence in its host's cytoplasm (Embley and Martin 2006; Kurland *et al.* 2006; de Duve 2007). Recent phylogenomic studies indicate that this bacterium - the

last free-living common ancestor of mitochondria (hereafter FMA) – was closely related to members of the Rickettsiales, within the α -proteobacteria (Fitzpatrick *et al.* 2006; Williams *et al.* 2007).

Despite these advances, consensus has yet to be reached over a number of issues concerning eukaryogenesis (Embley and Martin 2006; Poole and Penny 2006). Chief among these are: 1) the nature of the amitochondriate host; 2) how the FMA was engulfed; and, 3) whether eukaryogenesis occurred under oxic, microoxic, or anoxic conditions. On the nature of the amitochondriate host, hypotheses fall into two general categories. The first, more traditional, view is that the amitochondriate host contained most eukaryotic features, such as a nucleus, cytoskeleton, endomembrane system, and the ability to phagocytose (de Duve 2007; Cavalier-Smith 2009). This “proto-eukaryote” is considered by some authors to have been phylogenetically distinct from the lineages leading to extant archaea and eubacteria (Hartman and Fedorov 2002; Kurland *et al.* 2006; Poole and Penny 2006; Gribaldo *et al.* 2010). The second group of hypotheses posit the amitochondriate eukaryote host as an archaeobacterium (Rivera and Lake 1992; Martin and Muller 1998; Vellai *et al.* 1998; Cox *et al.* 2008; Davidov and Jurkevitch 2009). Under these hypotheses, eukaryotic features evolved after the fusion of the FMA and its archaeobacterial host.

On the issue of engulfment, models of eukaryogenesis typically assume that the FMA was non-motile, playing a relatively passive role in engulfment (Andersson and Kurland 1999; Embley and Martin 2006) (see Figures therein). This is in agreement with the absence of flagella in close relatives of mitochondria (Williams *et al.* 2007), including all examined Rickettsiales (Andersson *et al.* 1998; Wu *et al.* 2004; Dunning Hotopp *et al.* 2006; Cho *et al.* 2007), and *Pelagibacter ubique* (Giovannoni *et al.* 2005). Proto-eukaryote models usually assume that the FMA acted as non-motile prey for its phagocytic host (Cavalier-Smith 2009). This idea is consistent with the common occurrence of phagocytosis of prokaryotes by extant unicellular eukaryotes (Matz and Kjelleberg 2005; Kurland *et al.* 2006; Poole and Penny 2007). In the case of archaeobacterial host models, explaining engulfment of the FMA is challenging, due to the absence of phagocytosis in prokaryotes. However, some prokaryotes are known to exist within other prokaryotes (von Dohlen *et al.* 2001; Davidov and Jurkevitch 2009), indicating that engulfment may occur under some circumstances. An alternative to the general view of passive engulfment of the FMA is the view that the FMA displayed flagellar motility, and played a more active role in eukaryogenesis. This idea has been widely overlooked, with a few rare exceptions (Guerrero *et al.* 1986; Davidov and Jurkevitch 2009).

A further unresolved issue is whether eukaryogenesis occurred under oxic, microoxic, or anoxic conditions. In a number of hypotheses, the FMA is considered to have been aerobic, rescuing or detoxifying its initially anaerobic host from dramatically rising oxygen levels some 2.2 Gyr ago

(Margulis 1970; Andersson and Kurland 1999; Dyall *et al.* 2004; de Duve 2005). This idea is consistent with the presence of similar respiratory chains in mitochondria and members of the Rickettsiales. A challenge, however, for rescue and “ox-tox” hypotheses is explaining how the FMA would have survived in the microoxic or anaerobic conditions under which its host presumably existed. In contrast with oxygen rescue hypotheses, some archaeobacterial host hypotheses propose that the symbiosis originated under anaerobic conditions, for example between a hydrogen-producing FMA, and a hydrogen-utilizing archaeon (Martin and Muller 1998).

Phylogenomic studies of diverse eukaryotes continue to provide important insights into the nature of the eukaryotic ancestor (Dacks and Doolittle 2001; Slamovits and Keeling 2006; Brinkmann and Philippe 2007; Fritz-Laylin *et al.* 2010). Similarly, the genomes of numerous Rickettsiales and other α -proteobacteria have significantly advanced our understanding of the nature of the FMA (Andersson *et al.* 1998; Gabaldón and Huynen 2003; Darby *et al.* 2007). However, unlike the case of the eukaryotes, where representatives of all six main lineages have been sequenced (Fritz-Laylin *et al.* 2010), genome studies in the Rickettsiales have probably not properly sampled the entire diversity of this bacterial order (Darby *et al.* 2007). Recent 16S rDNA-based phylogenetic analyses show the Rickettsiales contains a number of novel lineages (Vannini *et al.* 2010), which are highly divergent from the two well-studied families, the Anaplasmataceae and the Rickettsiaceae. One such lineage contains the recently described *Candidatus* Midichloria mitochondrii (hereafter *M. mitochondrii*), so named because it is the only described bacterium able to enter the mitochondria of any multicellular organism (Lo *et al.* 2006; Sasser *et al.* 2006) (Figure 1 A, B). *M. mitochondrii* is an intracellular symbiont of the tick *Ixodes ricinus*; phylogenetically related endosymbionts have been found in other tick species and other invertebrates (Epis *et al.* 2008; Beninati *et al.* 2009). In *I. ricinus*, *in situ* hybridization and electron microscopic studies indicate that the bacterium is primarily restricted to the cells of the ovary (Zhu *et al.* 1992; Beninati *et al.* 2004; Sacchi *et al.* 2004). The role of *M. mitochondrii* in host biology is unclear, but various lines of evidence suggest it is a facultative mutualist. The prevalence of the symbiont is 100% in female *Ixodes ricinus* ticks collected in the field. This is typical of a mutualist or of a manipulator of the host reproduction. However, the fact that the 100% prevalence is observed throughout the entire geographical distribution of *I. ricinus* and the lack of evidence for alterations of the reproduction in this tick species argue against the second hypothesis. Furthermore, the loss of the symbiont in tick colonies maintained in the laboratory indicates that the symbiosis is not obligate, suggesting a facultative mutualism (for a complete description of the known characteristics of *M. mitochondrii*, see Lo *et al.* 2006; Epis *et al.* 2008; Sasser *et al.* 2008).

Here we report the complete genome sequence of *M. mitochondrii*, and the results of phylogenomic comparisons with related bacteria, as well as mitochondria. We focus on two novel gene

complements present in *M. mitochondrii* which have key implications for our understanding of eukaryogenesis, and for addressing the unresolved issues discussed above.

Materials and Methods

- Transmission electron microscopy -

Twenty-five females of the hard tick *I. ricinus* were collected from different hosts (dogs, roe deer, sheep) in different areas of Italy (counties of Novara, Varese, Parma, Trento, Ascoli Piceno). Ovaries were dissected in saline solution, then fixed, post-fixed, dehydrated and embedded in Epon 812 resin as described (Epis *et al.* 2008). Semi- and ultra-thin sections were stained and examined under optical and transmission electron microscopes.

- DNA purification and sequencing -

A semi-engorged *I. ricinus* adult female was collected from a dog in the county of Varese (Italy) and *M. mitochondrii* DNA was purified and amplified as described previously (Epis *et al.* 2010). Briefly, following tick washing, the ovary was dissected under a stereomicroscope and single oocytes were mechanically detached. Increasing quantities of collagenase A (Sigma-Aldrich, St. Louis, USA) 20 mg/ml diluted in 50% sterile water and 50% PBS were added to cause osmotic breakage of the membrane of the oocytes. Twenty pools of 10-12 cytoplasms were collected using a microcapillary, avoiding the nuclei trapped in the remnants of the cytoplasmic membranes. Cytoplasmic preparations were stained with Hoechst 33342, to discard preparations containing undesired nuclei. The remaining 15 cytoplasm preparations were used as templates for Multiple Displacement Amplification (MDA) using Repli-g Mini Kit (Qiagen, Hilden, Germany). MDA products were purified using a QIAEX II Gel Extraction Kit (Qiagen) and used as templates for a previously described set of real-time PCRs, to assess whether the genome of *M. mitochondrii* was amplified with limited biases, and to exclude the possible residual presence of detectable *I. ricinus* nuclear DNA contamination. For a detailed description of primers and PCR conditions see Epis *et al.* (2010).

Thirteen preparations showed uniform amplification of the seven *M. mitochondrii* loci, and two of these exhibited no sign of *I. ricinus* nuclear DNA contamination. These two MDA products were pooled and used as template for pyrosequencing and for 1-3 kbp plasmid library construction, to be sequenced using Sanger-based technology. One half GS-FLX Titanium run and one quarter 3 kb paired-ends GS-FLX run were performed, netting a total of 544,141 and 156,393 reads respectively. Sanger sequences were obtained from both ends of 856 clones.

- Assembly and annotation -

Data resulting from the two pyrosequencing runs and the Sanger sequences were used for assembly using the Mira assembler v2.9 (Chevreux *et al.* 1999), obtaining 196 large contigs (from 2005 to 124,307 bp). Contigs were manually checked and joined with gap4 (Staden *et al.* 2000) and thus reduced to 154 contigs. A scaffold was then generated using Bambus (Pop *et al.* 2004) and the scaffold was used as a backbone for a second assembly with Mira to obtain 38 large contigs (from 13,459 bp to 424,616 bp). Although the Mira assembler provides an algorithm for chimera recognition and removal, all contigs were visually examined in gap4, with specific attention to all low coverage regions (<5x). Manual correction was performed to close sequence gaps. Eighteen remaining gaps were then closed by PCR and inverse PCR as described by Hartl and Ochman (1996).

Putative coding regions were identified with Glimmer v3.02 for protein-coding genes (Delcher *et al.* 1999), tRNAscan-SE for tRNAs (Schattner *et al.* 2005) and BlastN for ribosomal RNAs. All putative protein-coding genes were analyzed in local with blastP and rpsblast algorithms searching the pfam, eggnog, nr, uniprot and omniome databases. Results from the different databases were compared and manually curated in order to assign gene functions. Forty-seven regions of interest (possible pseudogenes, flagellar genes, *cbb₃* genes, genes coding for surface proteins) were amplified using specific primers on genomic DNA not subjected to MDA, and then sequenced using the Sanger method. Metabolic pathways were reconstructed using the KEGG automated annotation server KAAS (Kaneisha and Goto 2000) and manually checked. Sequences of 112 regions of interest (possible pseudogenes, all flagellar genes, all *cbb₃* genes, genes coding for surface proteins, genes containing homopolymers) were re-obtained by Sanger sequencing after specific PCR amplification from DNA not subjected to MDA. All of the sequences generated after PCR and Sanger-sequencing were identical to the previous consensus, with the exception of six homopolymeric regions that were corrected for the generation of the consensus sequence. Expression of five flagellar genes and three *cbb₃* genes was investigated by reverse transcription PCR on cDNA obtained from *I. ricinus* adult females and egg batches.

- Phylogenetics -

For phylogenomic analysis, an alignment of 88 conserved genes from 72 organisms, previously used in a global α -proteobacteria phylogeny (Williams *et al.* 2007) was kindly provided by K.P. Williams, Sandia National Laboratories, Livermore, California. Orthologous *M. mitochondrii* genes were then added to the aforementioned alignment. For the study of the flagellum phylogeny, two datasets were constructed, one with 14 core flagellar genes and one with 24 conserved bacterial genes (as described in Liu and Ochman 2007), using alignments kindly provided by R. Liu,

University of California, Riverside. These alignments were modified via the addition of extra proteobacterial representatives, including *M. mitochondrii* orthologous genes, for a total of 24 α -, nine β -, 13 γ -, seven δ -, six ϵ -proteobacteria, and three outgroups. A fourth dataset was prepared using 12 highly conserved bacterial proteins that have homologs in mitochondrial genomes (as described in Williams *et al.* 2007), retrieved from 55 taxa, including seven outgroups, eight mitochondria, and 14 members of the Rickettsiales. *cbb*₃ phylogenetic analyses were performed on an alignment of the three genes coding for the fundamental subunits of this complex (*ccoN*, *ccoO*, *ccoP*) for 35 taxa, with the addition of *M. mitochondrii*.

For all phylogenetic analyses, genes were aligned after addition of the *M. mitochondrii* orthologs using Muscle (Edgar 2004). Alignments were trimmed with Gblocks (Castresana 2000) with optimization of the parameters for each single gene alignment, based on the number of taxa for which the amino acid sequence was present: -b1= (N/2)+1; -b2= (N/2)+1; -b3=(N/2); -b4=2; -b5=h, with N = number of taxa. Each alignment was analyzed with ProtTest 2.1 (Abascal *et al.*, 2005) to choose the most suitable model of evolution. PhyML version 3.0 (Guindon and Gascuel 2003) and MrBayes 3.1.2 (Huelsenbeck and Ronquist 2001) were then used for construction of maximum likelihood and Bayesian analysis trees respectively, both on single flagellar genes and on the concatenated gene alignments of the four datasets. PhyML was used to construct maximum likelihood-based phylogenetic trees with non parametric bootstrap analysis (100 bootstrap replicates), with the following options: LG as the amino-acid substitution model; optimized proportions of invariable sites; amino-acid frequency estimation counting the number of different amino-acids observed in the data; optimized rate variation across sites into 8 substitution rate categories; the best of NNI and SPR tree searching operation; computation of the starting tree with BioNJ. Bayesian analyses were performed on the partitioned concatenated alignment to allow parameter optimization for each gene. All the Markov Chain Monte Carlo analyses were implemented into two runs with four chains each (one hot and three cold). Each analysis had samplefreq=10. We used Whelan and Goldman (WAG) substitution model + proportions of invariable sites (I) and gamma (Γ) with 8 categories. The convergence of each run was verified with Tracer 1.4 (Rambaut and Drummond 2007).

In order to investigate the influence of base composition on our phylogenetic reconstructions, we performed a Bayesian phylogenetic analysis of alpha-proteobacterial relationships under a model of heterogeneous composition as implemented in the software P4 (Foster 2004). These analyses recovered topologies that were entirely consistent with the homogeneous Bayesian and ML analyses presented in the figures (results not shown).

- Genome analysis -

Amino acid sequences derived from the genomes of *Wolbachia* endosymbiont of *Drosophila melanogaster*, *Rickettsia prowazekii* str. Madrid, *Rickettsia conorii*, *Rickettsia bellii* OSU 85-389, *Orientia tsutsugamushi* str. Boryong, *Anaplasma phagocytophilum* HZ and *Ehrlichia chaffensis* were downloaded from <ftp://ftp.ncbi.nih.gov/genomes/Bacteria/>. Two sets of four genomes were subjected to a five-way comparison with the genome of *M. mitochondrii* by OrthoMCL v1.4 (Li *et al.* 2003) using default parameters, in order to find orthologous groups conserved between the Rickettsiales. COG functional group assignment was obtained for *M. mitochondrii* and four other Rickettsiales proteins by local rpsblast on the COG database. A graphic representation of the *M. mitochondrii* genome was constructed using the software genomeviz (Ghai *et al.* 2004), using the COG functional group assignment and other features extracted from the genome annotation. Metabolic pathways for the synthesis of vitamins and cofactors were manually inferred from the annotation of the *M. mitochondrii* genome. Repeat content in the *M. mitochondrii* genome was assessed using the ugene repeat finder tool and compared with the repeat content of other selected Rickettsiales. Codon usage bias analysis was performed using the local version of CAIcalc (Puigbo *et al.* 2008) using default parameters, in order to calculate the codon adaptation index (CAI) of all the protein coding genes of the *M. mitochondrii* genome and to evaluate if the CAI values of *cbb₃* and flagellar genes is statistically divergent from the CAI expected.

Results and Discussion

- Characteristics of the *M. mitochondrii* genome, and its phylogenetic relationship to other α -proteobacteria and mitochondria -

The genome of *M. mitochondrii* consists of a single 1,183,732 bp circular chromosome with a G+C content of 36.6% (Figure 1 C). Most of the genome content of *M. mitochondrii* is typical of other members of the Rickettsiales (Andersson *et al.* 1998; Wu *et al.* 2004; Dunning Hotopp *et al.* 2006) (Table 1)(Figure S1). Thus, *M. mitochondrii* possesses a relative scarcity of genes encoding amino acid and nucleotide biosynthesis pathways, compared with free-living α -proteobacterial relatives. Although *M. mitochondrii* has diminished biosynthetic capabilities, like many other Rickettsiales (Dunning-Hotopp *et al.* 2006) it does possess genes for the production of several cofactors, including coenzyme A, biotin, lipoic acid, tetrahydrofolate, panthotenate, heme, and ubiquinone (Figure 2).

Table 1 General genome properties of selected Rickettsiales. MID, *Midichloria mitochondrii*; APH, *Anaplasma phagocytophilum* HZ; ECH, *Ehrlichia chaffensis*; RPR, *Rickettsia prowazekii* str. Madrid; WME, *Wolbachia* endosymbiont of *Drosophila melanogaster*.

Properties	Organisms				
	MID	APH	ECH	RPR	WME
Genome size	1,183,732	1,471,282	1,176,248	1,111,523	1,267,782
GC%	36.6	41.6	30.1	29.1	35.2
ORFs	1,245	1,369	1,115	834	1,271
tRNAs	35	37	37	33	34
rRNAs	3	3	3	3	3
Average gene length	698	775	840	1,005	855
Coding %	73.7	72.2	79.9	75.4	85.7
Assigned function	634	747	604	523	719
Conserved hypothetical	123	82	111	NR	123
Hypotetical	391	458	314	208	337

NOTE.—MID, *Midichloria mitochondrii*; APH, *Anaplasma phagocytophilum* HZ; ECH, *Ehrlichia chaffensis*; RPR, *Rickettsia prowazekii* str. Madrid; and WME, *Wolbachia* endosymbiont of *Drosophila melanogaster*; NR, not reported; ORFs, open reading frames.

M. mitochondrii may supply host cells with these essential cofactors. *M. mitochondrii* is inferred to have a functional Krebs cycle, gluconeogenesis pathway and pyruvate dehydrogenase complex, and almost all enzymes required for glycolysis (Figure S2). *M. mitochondrii* is thus able to synthesize ATP, and the presence of a gene coding for an ATP/ADP translocase indicates that it may also be able to import/export ATP from/to the host.

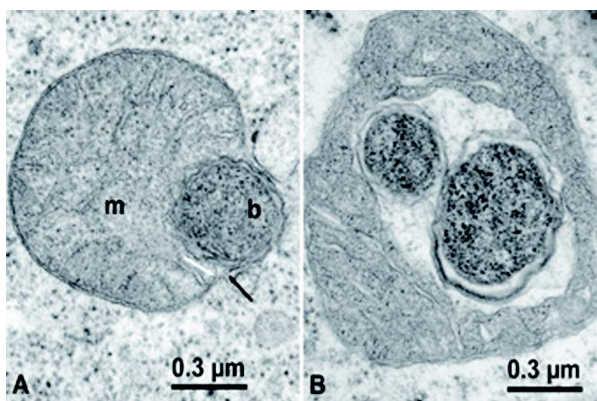
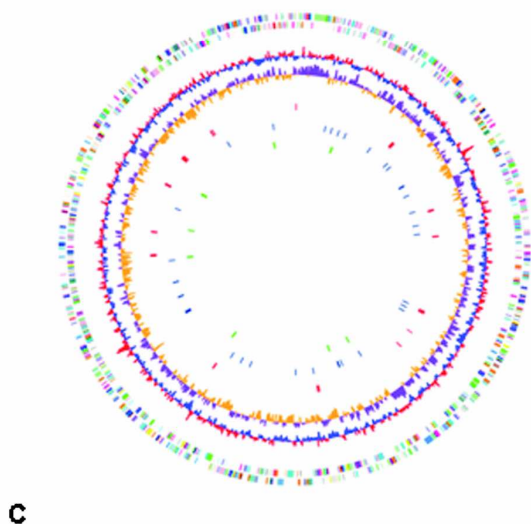


Figure 1. Transmission electron microscopic images of *Midichloria mitochondrii* in the mitochondria (**A**, **B**), and circular genome map (**C**). **A**, initial step of infection showing a bacterium (b) between the outer (black arrow) and inner (white arrow) membranes of the mitochondrion (m). **B**, a mitochondrion harbouring two bacteria; the matrix appears to be partially consumed. **C**, The two outermost circles show all predicted coding regions respectively on the plus and minus strand, divided by color based on COG functional groups; For a complete color-code explanation see Figure S5. The third circle shows the GC content, the fourth circle shows the GC skew. The fifth circle shows the location of the flagellar genes in red, the sixth circle shows the structural RNAs in blue and the innermost circle shows the putative pseudogenes in green.



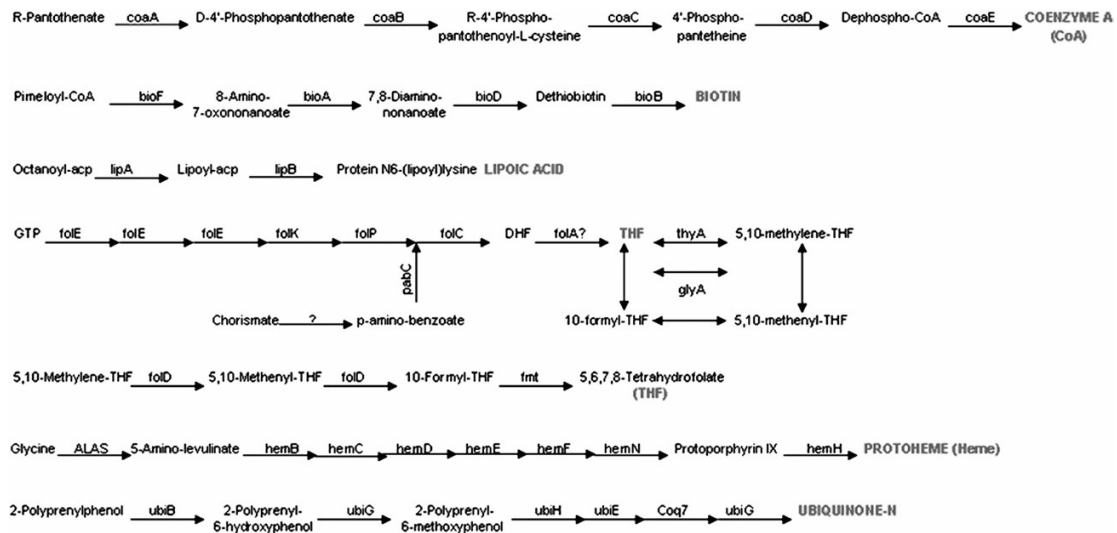


Figure 2. Vitamin and cofactor metabolic pathways present in *M. mitochondrii*, manually inferred from the annotation of the genome.

M. mitochondrii's genome contains a high number of imperfect repeats, possibly due to a past proliferation of mobile elements (Table S1). However, few intact genes associated with mobile and extrachromosomal functions were found. Most identified sequences of this nature (phage capsid proteins, transposases) appear to be truncated pseudogenes. Like other Rickettsiales, *M. mitochondrii* possesses genes encoding type IV and Sec-independent protein secretion systems, ankyrin repeat proteins, and various putative membrane-associated proteins. These genes may be associated with the symbiont's unique ability to invade host mitochondria. Overall, *M. mitochondrii*'s genome does not provide a clear answer to the question of whether it engages in a mutualistic or a parasitic relationship with its tick host. Targeted experimental studies (e.g. examining the fitness of the host with and without the symbiont, and functional genomic studies) are required to address this question.

Phylogenetic analyses of an alignment of 88 conserved genes from 66 α -proteobacterial representatives plus 6 outgroups confirmed the divergent phylogenetic position of *M. mitochondrii* among the Rickettsiales (Figure 3 A).

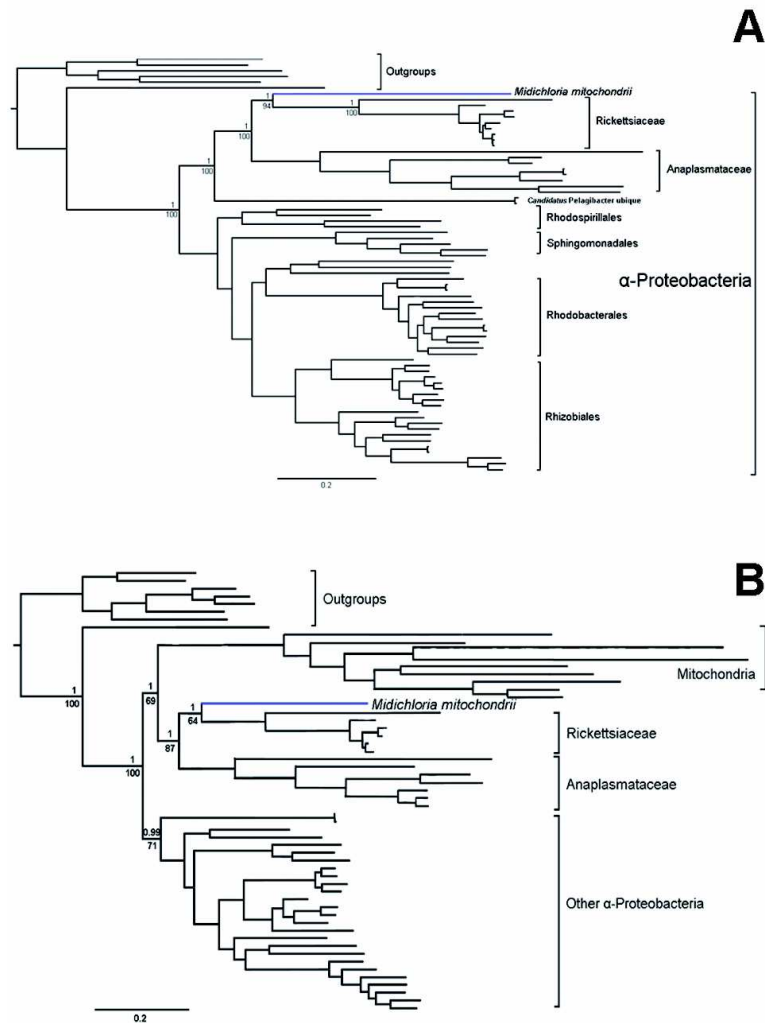


Figure 3. Phylogenetic analysis based on conserved proteins from (A) diverse representatives of the α -proteobacteria (88 single copy proteins), and (B) α -proteobacteria and mitochondria (12 single copy proteins). Concatenated alignments were analyzed using MrBayes (topology shown) and PhyML. Posterior probabilities and PhyML bootstrap supports are respectively shown above and below nodes of interest. In each case, *M. mitochondrii* is placed as the sister group of the Rickettsiaceae (a family within the order Rickettsiales). In B, the Rickettsiales are placed as sister group to mitochondria. For simplicity, trees are shown without taxon names; more detailed trees, with names of the taxa, are provided in the supplementary information – Figure S3.

We found high support for *M. mitochondrii* forming a sister group with the Rickettsiaceae (which includes the genera *Rickettsia* and *Orientia*). The analyses also confirmed the early branching position of the Rickettsiales relative to other orders among the α -proteobacteria (Dunning Hotopp *et al.* 2006; Williams *et al.* 2007). We performed additional phylogenetic analyses to examine the relationship of *M. mitochondrii* to mitochondria. Orthologous genes from *M. mitochondrii* were added to a previous alignment of 12 genes (Williams *et al.* 2007) present both in large mitochondrial genomes and diverse α -proteobacteria, including 14 representatives of Rickettsiales.

M. mitochondrii was found as the sister group to the Rickettsiaceae in the 12-gene analysis (Figure 3 B), equivalent to its position in the 88 gene analysis. The Rickettsiales were placed as the sister group of the mitochondria, which together formed a deep branch within the α -proteobacteria, again in accordance with Figure 3 A.

- *Genes putatively encoding flagellar proteins in the M. mitochondrii genome* -

Unexpectedly, we found 26 genes in the *M. mitochondrii* genome that putatively encode a flagellum, including all key components such as the hook, filament, and basal body. Such genes are found in none of the other ~20 Rickettsiales genomes sequenced thus far (Darby *et al.* 2007), nor are they found in *P. ubique*, which has phylogenetic affinities with the Rickettsiales (Williams *et al.* 2007). Electron microscopic examinations of *M. mitochondrii* in the ovarian cells of 25 wild-collected ticks revealed no evidence for a flagellum (Figure 1 A, B), confirming the results of previous studies (Zhu *et al.* 1992; Sacchi *et al.* 2004). The function of these genes is, therefore, unclear. There is evidence that *M. mitochondrii* is transmitted horizontally among ticks, most likely via their vertebrate hosts (Epis *et al.* 2008). The flagellum might be assembled in as-yet-unexplored stages of the bacterium's life cycle. It may also play a role in the invasion of tick mitochondria by *M. mitochondrii*.

Each of the flagellar genes is of a similar length to orthologs in other eubacteria and contains no stop codons. Genomic and reverse transcriptase PCR assays on a subset of these genes (*fliC*, *fliD*, *flgL*, *flgK*, *flgE*) in multiple *I. ricinus* specimens confirmed their uniform presence in *M. mitochondrii*, and indicated that they are transcribed (data not shown). Thus, there is no indication that they are pseudogenes. GC content of the 26 flagellar genes found in the *M. mitochondrii* genome ($35.6\% \pm 2.3\%$) was not significantly different to the GC content of other *M. mitochondrii* protein coding genes ($36.8\% \pm 4\%$; Mann-Whitney U test, $P=0.137$), indicating that the flagellar genes do not derive from recent lateral gene transfer. Additionally, codon adaptation index (CAI) analysis (Puigbo *et al.* 2008) indicated no statistically significant difference between the expected CAIs for flagellar genes and all other genes ($P<0.05$). We performed phylogenetic analysis to investigate the evolutionary origin of *M. mitochondrii*'s flagellar genes. Based on comparisons of phylogenetic trees inferred from 14 concatenated conserved flagellar genes with those inferred from 24 core eubacterial genes, Liu and Ochman (2007) previously demonstrated that flagellar genes have been inherited in a largely vertical fashion throughout eubacterial diversification. We repeated these analyses on expanded gene datasets, including *M. mitochondrii* and additional proteobacterial representatives. Our analyses revealed the same pattern of vertical inheritance of flagellar genes found by Liu and Ochman (2007). Thus, *M. mitochondrii* is placed as the deepest branch relative to flagellated α -proteobacteria, both in trees inferred from flagellar genes (Figure 4 A), and in trees

inferred from core genes (Figure 4 B). For the 14 flagellar single-gene analyses, all nodes with >75% and 0.8 PP values (PhyML and MrBayes respectively) agreed with those in the concatenated tree, indicating that the examined flagellar genes have followed a common evolutionary pattern since their origin. To test the robustness of the position of *M. mitochondrii* in the concatenated PhyML tree, we performed the SH (Shimodaira and Hasegawa, 1999), two sided and one sided KH (Kishino and Hasegawa 1989; Goldman *et al.* 2000) and ELW (Strimmer, 2002) statistical tests implemented in the program TreePuzzle (Schmidt *et al.* 2002). We forced the *M. mitochondrii* position with either the outgroups, the γ -proteobacteria, or the β -proteobacteria. All the performed statistical tests indicated that the original topology was a significantly better explanation of the data than the alternative solutions ($P < 0.05$).

The position of *M. mitochondrii* in Figure 4 is in accordance with the position of the Rickettsiales as an early branch of α -proteobacteria (Figure 3), and indicates that *M. mitochondrii* has inherited its flagellar genes vertically from the ancestor of the α -proteobacteria. Figure 4 also shows that the ancestor of all α -proteobacteria possessed a flagellum, as demonstrated previously on the basis of genome content comparisons (Boussau *et al.* 2004). This result, taken together with the results shown in Figure 3, indicates that flagellar genes were present in the ancestor of the Rickettsiales and mitochondria (Figure 5). Flagellar genes are known to be lost upon transition from the free-living state to the intracellular state (Toft and Fares 2008). We infer that such loss occurred during the transition of the FMA to the intracellular state. Independent loss of flagellar genes also occurred in the lineages leading to present day Rickettsiaceae and Anaplasmataceae (Figure 5).

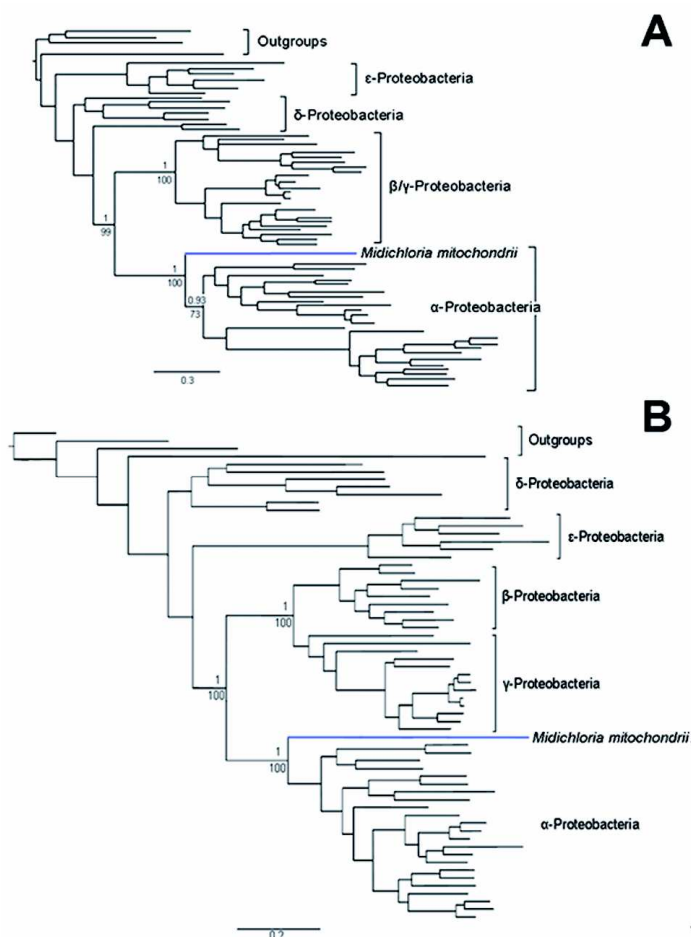


Figure 4. Phylogenetic analysis of 14 conserved flagellar proteins (A) and 24 core proteins (B) from diverse representatives of the proteobacteria phylum. Concatenated alignments were analyzed using MrBayes (topology shown) and PhyML. Posterior probabilities and PhyML bootstrap supports are respectively shown above and below nodes of interest. *M. mitochondrii* is placed as the deepest branch of the α -proteobacteria in each case. For simplicity, trees are shown without taxon names; more detailed trees are provided in the supplementary information – Figure S4.

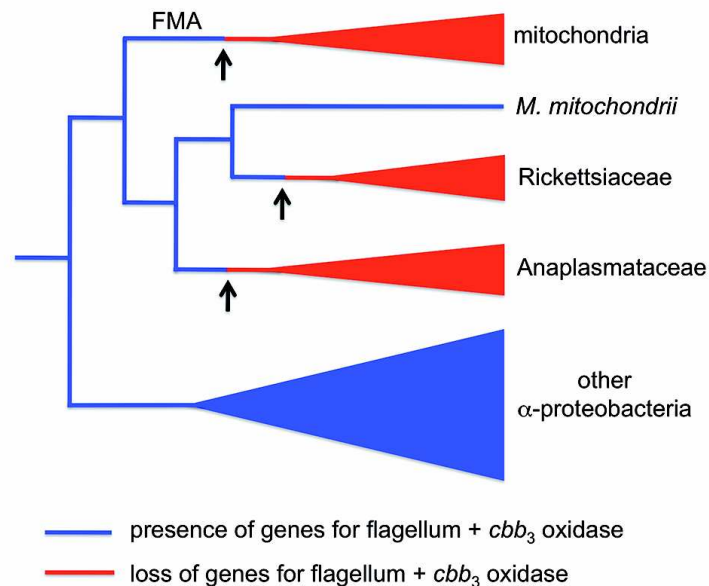


Figure 5. Presence of flagellar and *cbb*₃ oxidase genes in the ancestor of the α -proteobacteria, and inferred losses (indicated by arrows, and red) in lineages leading to mitochondria, Rickettsiaceae, and Anaplasmataceae. The scheme is based on the results of Figures 3, 4 and 6. The FMA is inferred to have possessed flagellar and *cbb*₃ oxidase genes.

- *Genes putatively encoding a *cbb*₃ oxidase in the *M. mitochondrii* genome* -

A second set of genes present in *M. mitochondrii*, but in no other examined members of the Rickettsiales, is predicted to encode a cytochrome *cbb*₃ oxidase, and proteins associated with its assembly. The *cbb*₃ oxidases belong to the C-family of heme-copper oxidases (HCOs). C-family HCOs are phylogenetically divergent from the more common A-family HCOs (*aa*₃-type), which are widespread in prokaryotes and mitochondria, and from the B-family HCOs (*bo*₃-type, or quinol oxidases), which are present mainly in crenarcheota (Ducluzeau *et al.* 2008; Buschmann *et al.* 2010). *cbb*₃ oxidases reduce O₂ with lower efficiency compared to A- and B-family HCOs, but have a higher affinity for it. This characteristic is used by pathogenic proteobacteria that infect microaerobic host tissues, and by symbiotic diazotrophs that are able to simultaneously undertake aerobic respiration and nitrogen fixation (which requires an oxygen-sensitive nitrogenase) (Pitcher and Watmough 2004). In ticks, oogenesis leads to a dramatic increase in oxygen use (Aboul-Nasr and Bassal 1972). The presence of a *cbb*₃ oxidase may enable *M. mitochondrii* to synthesize ATP at oxygen concentrations that are suboptimal for the mitochondrion. We might thus speculate that *M. mitochondrii* could serve as an additional ATP source for the host cell during oogenesis. Transfer of ATP to the host cell might occur through the ATP/ADP translocase of *M. mitochondrii*, as proteins of this family are known to have a reversible function, based on ATP/ADP concentrations and proton gradient.

The genes *ccoN*, *ccoO*, and *ccoP*, which encode the three fundamental subunits of *cbb*₃, are found together in *M. mitochondrii*, in the same order as *Pseudomonas stutzeri* strain Zobell (Buschmann

et al. 2010), though the *M. mitochondrii ccoN* gene is split into two ORFs. Each of the key histidine and methionine residues associated with coordination of heme irons are present (Buschmann *et al.* 2010). Two additional genes, *ccoI* and *ccoG*, reported to be accessory proteins for assembly of the *cbb₃* complex, are also present in the genome. Reverse-transcription PCR assays demonstrated that *ccoN*, *ccoO*, and *ccoP* are transcribed (data not shown). GC content and CAI analysis were performed for *cbb₃* genes as described above for flagellar genes; no evidence for recent horizontal transfer of these genes was found (GC of *cbb₃* genes: 39.9% ± 1.6%, Mann-Whitney U test, P=0.161; CAI test P<0.05).

The *cbb₃* oxidases are distributed primarily in the proteobacteria, and are believed to have evolved in an early progenitor of this phylum, or perhaps earlier (Ducluzeau *et al.* 2008). All known α -proteobacterial *cbb₃* oxidase were previously shown to form a monophyletic clade among other proteobacterial clades, suggesting that they have been inherited vertically from an α -proteobacterial ancestor. Our phylogenetic analyses of *ccoN*, *ccoO* and *ccoP* confirmed these findings, and showed that the genes from *M. mitochondrii* represent the deepest branch among the α -proteobacteria (Figure 6), in accordance with the results from the flagellar and core gene phylogenies (Figure 4). These results indicate that this gene set, like the flagellar gene set, was inherited vertically from the α -proteobacterial ancestor to *M. mitochondrii*. We therefore infer that the ancestor of Rickettsiales, and the ancestor of the FMA, possessed a *cbb₃* oxidase gene set (Figure 5). These genes were later lost during the evolution of mitochondria, and in the lineages leading to the Rickettsiaceae, and the Anaplasmataceae.

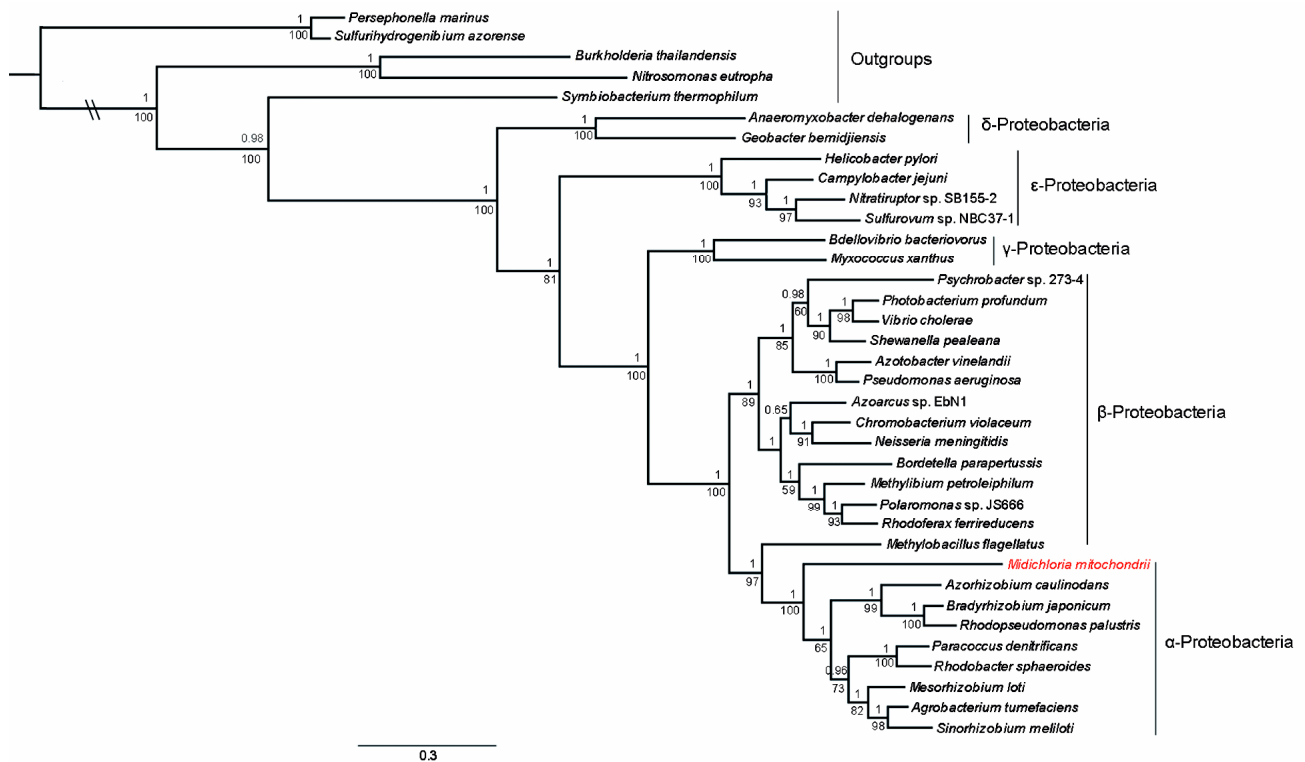


Figure 6. Phylogenetic analysis of *cbb₃* proteins (*ccoN*, *ccoO*, *ccoP*) from diverse representatives of the proteobacteria phylum. Concatenated alignments of the 3 proteins were analysed using MrBayes (topology shown) and PhyML. *M. mitochondrii* is placed as the deepest branch of the α -proteobacteria. Posterior probabilities and PhyML bootstrap supports are respectively shown above and below nodes.

Most α -proteobacteria that possess *cbb₃* oxidases also have the more common *aa₃* oxidase; the presence of both gene sets thus appears to be a primitive trait. Interestingly, although *M. mitochondrii* does not contain a complete *aa₃* oxidase, it does contain one *aa₃* oxidase gene fragment with high similarity to *aa₃* oxidases in other Rickettsiales. Thus the lineage leading to *M. mitochondrii* may have lost the complete set of *aa₃* oxidase genes relatively recently. The *aa₃* oxidases present in the ancestor of the Rickettsiales and the FMA would eventually evolve into the mitochondrial cytochrome oxidase IV complex (Figure 5).

- A flagellum and *cbb₃* oxidase in the FMA: implications for eukaryogenesis -

Hypotheses of eukaryogenesis have generally assumed the absence of a flagellum in the FMA (Andersson and Kurland 1999; Embley and Martin 2006; Cavalier-Smith 2009). In proto-eukaryote host hypotheses, the FMA is considered to have acted as passive prey to a phagocytic, predatory host. This idea is appealing due to the numerous extant examples of single-celled eukaryotes that feed on prokaryotes, and is also in accord with the universally accepted model of chloroplast evolution via the phagocytosis of a member of the cyanobacteria (whose members all lack flagella)

(Douglas 1998). The results presented here represent the first evidence that the FMA possessed a flagellum, and thus challenge this assumption. A FMA with flagellar motility is likely to have played more active role in eukaryogenesis than currently appreciated. For example, feeding by some single-celled phagotrophic eukaryotes is dependent upon motility in their prokaryotic prey (Fenchel 1987). Heliozoan predators extend their axopods in all directions, and wait for motile prey cells to collide with them. Flagella are also known to play a key role in adhesion and engulfment in many kinds of eukaryotic cells, from protist to mammalian (Mahenthiralingam and Speert 1995; Ottemann and Miller 1997; Inglis *et al.* 2003; Ramos *et al.* 2004). In addition to the traditional idea that the FMA acted as prey, the possibility that the FMA acted as a parasite of its amitochondriate host should also be considered. Parasitism of protists by motile prokaryotes is a common phenomenon, and opens up opportunities for facultative mutualism to evolve.

In the case of archaeobacterial host hypotheses, the presence of a flagellum in the FMA may help to explain host-entry (in the absence of any phagocytotic mechanism known in prokaryotes). The re-emerging “predatory hypothesis” (Davidov and Jurkevitch 2009) for eukaryogenesis is based on the ability of some extant flagellated proteobacteria to penetrate the periplasm or cytoplasm of other bacteria. However, long-term persistence of such bacteria in their “prey” (other than as resting bdelloccysts) awaits discovery, as does a case of an archaean representative acting as prey or host.

Phylogenomic analysis of *cbb₃* oxidase genes from *M. mitochondrii* enables us to infer, for the first time, the presence of a *cbb₃* oxidase in the FMA. The presence of this enzyme complex in the FMA has important implications for understanding the ecological context in which eukaryogenesis occurred. Oxygen rescue and ox-tox hypotheses propose that the FMA played a key role in allowing their anaerobic, amitochondriate eukaryote hosts to adapt to the oxygenation of the earth which occurred some 2.2 Gyr ago (Margulis 1970; Andersson and Kurland 1999; Kurland and Andersson 2000). The presence of a *cbb₃* oxidase helps to explain how the FMA could have survived in the microoxic conditions under which the interaction began. Following the establishment of the symbiosis, and the adaptation of eukaryotes to aerobic conditions, *cbb₃* oxidase genes would have become less important, and eventually lost. The presence of a *cbb₃* oxidase in the FMA is also relevant to anaerobic eukaryogenesis models (Martin and Muller 1998), in that it helps to explain how the symbiont transitioned to anaerobic conditions.

The origin of the eukaryotic cell is one of the most mysterious and challenging problems facing modern biology. The results herein illustrate the importance of surveying the genomes of diverse relatives of the FMA, many of which probably remain to be discovered. Such genome surveys will allow testing of the hypotheses presented here, and may shed further light on the nature of the FMA.

2.3.3 References

- Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21:2104-2105.
- Aboul-Nasr AE, Bassal TTM. 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.
- Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, Podowski RM, Naslund AK, Eriksson AS, Winkler HH, Kurland CG. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 396:133-140.
- Andersson SGE, Kurland CG. 1999. Origins of mitochondria and hydrogenosomes. *Curr. Opin. Microbiol.* 2:535-541.
- Beninati T, Lo N, Sacchi L, Genchi C, Noda H, Bandi C. 2004. A novel α -proteobacterium resides in the mitochondria of ovarian cells of the tick *Ixodes ricinus*. *Appl. Environ. Microbiol.* 70:2596-2602.
- Beninati T, Riegler M, Vilcins I, Sacchi L, McFadyen R, Krockenberger M, Bandi C, O'Neill SL, Lo N. 2009. Absence of the symbiont *Candidatus* Midichloria mitochondrii in the mitochondria of the tick *Ixodes holocyclus*. *FEMS Microbiol. Lett.* 299:241-247.
- Boussau B, Karlberg EO, Frank AC, Legault BA, Andersson SGE. 2004. Computational inference of scenarios for α -proteobacterial genome evolution. *Proc. Natl. Acad. Sci. U S A.* 101:9722-9727.
- Brinkmann H, Philippe H. 2007. The diversity of eukaryotes and the root of the eukaryotic tree. In: Back N, Cohen IR, Kritchinsky D, Lajtha A, Paoletti R, editors. *Eukaryotic membranes and cytoskeleton*. Springer New York: p. 20-37.
- Bui ET, Bradley PJ, Johnson PJ. 1996. A common evolutionary origin for mitochondria and hydrogenosomes. *Proc. Natl. Acad. Sci. U S A.* 93:9651-9656.
- Buschmann S, Warkentin E, Xie H, Langer JD, Ermler U, Michel H. 2010. The structure of *cbb₃* cytochrome oxidase provides insights into proton pumping. *Science* 329:327-330.
- Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17:540-552.
- Cavalier-Smith, T. 2009. Predation and eukaryote cell origins: a coevolutionary perspective. *Int. J. Biochem. Cell. Biol* 41:307-322.

- Chevreur, B, Wetter, T, Suhai, S. 1999. Genome sequence assembly using trace signals and additional sequence information. *Computer science and biology: proceedings of the german conference on bioinformatics (GCB)* 99:45-56.
- Cho NH, Kim HR, Lee JH, Kim SY, Kim J, Cha S, Kim SY, Darby AC, Fuxelius HH, et al. (19 coauthors). 2007. The *Orientia tsutsugamushi* genome reveals massive proliferation of conjugative type IV secretion system and host cell interaction genes. *Proc. Natl. Acad. Sci. U S A.* 104:7981-7986.
- Cox CJ, Foster PG, Hirt RP, Harris SR, Embley TM. 2008. The archaeobacterial origin of eukaryotes. *Proc. Natl. Acad. Sci. U S A.* 105:20356-20361.
- Dacks JB, Doolittle WF. 2001. Reconstructing/deconstructing the earliest eukaryotes: how comparative genomics can help. *Cell* 107:419-425.
- Darby AC, Cho NH, Fuxelius HH, Westberg J, Andersson SGE. 2007. Intracellular pathogens go extreme: genome evolution in the Rickettsiales. *Trends Genet.* 23:511-520.
- Davidov Y, Jurkevitch E. 2009. Predation between prokaryotes and the origin of eukaryotes. *BioEssays* 31:748-757.
- de Duve C. 2005. *Singularities: landmarks on the pathways of life.* Cambridge: Cambridge University Press.
- de Duve C. 2007. The origin of eukaryotes: a reappraisal. *Nat. Rev. Genet.* 8, 395-403.
- Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* 27:4636-4641.
- Douglas SE. 1998. Plastid evolution: origins, diversity, trends. *Curr. Opin. Genet. Dev.* 8:655-661.
- Ducluzeau AL, Ouchane S, Nitschke W. 2008. The *cbb₃* oxidases are an ancient innovation of the domain bacteria. *Mol. Biol. Evol.* 25:1158-1166.
- Dunning Hotopp JC, Lin M, Madupu R, Crabtree J, Angiuoli SV, Eisen J, Seshadri R, Ren Q, Wu M, Utterback TR, et al. (40 coauthors). 2006. Comparative genomics of emerging human ehrlichiosis agents. *PLoS Genet* 2:e21.
- Dyall SD, Brown MT, Johnson PJ. 2004. Ancient invasions: from endosymbionts to organelles. *Science* 304:253-257.
- Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.
- Embley TM, Martin W. 2006. Eukaryotic evolution, changes and challenges. *Nature* 440:623-630.
- Epis S, Sassera D, Beninati T, Lo N, Beati L, Piesman J, Rinaldi L, McCoy KD, Torina A, Sacchi L, et al. (14 coauthors). 2008. *Midichloria mitochondrii* is widespread in hard ticks (Ixodidae) and resides in the mitochondria of phylogenetically diverse species. *Parasitology* 135:485-494.

- Epis S, Luciano AM, Franciosi F, Bazzocchi C, Crotti E, Pistone D, Bandi C, Sassera D. 2010. A novel method for the isolation of DNA from intracellular bacteria, suitable for genomic studies. *Ann. Microbiol.* 60:455-460.
- Fenchel T. 1987. *Ecology of Protozoa*. Berlin: Springer Verlag.
- Foster P. 2004. Modeling compositional heterogeneity. *Syst. Biol.* 53:485-495.
- Fitzpatrick DA, Creevey CJ, McInerney JO. 2006. Genome phylogenies indicate a meaningful α -proteobacterial phylogeny and support a grouping of the mitochondria with the Rickettsiales. *Mol. Biol. Evol.* 23:74-85.
- Fritz-Laylin LK, Prochnik SE, Ginger ML, Dacks JB, Carpenter ML, Field MC, Kuo A, Paredez A, Chapman J, Pham J, et al. (24 coauthors). 2010. The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. *Cell* 140:631-642.
- Gabaldón T, Huynen MA. 2003. Reconstruction of the proto-mitochondrial metabolism. *Science* 301:609.
- Ghai R, Hain T, Chakraborty T. 2004. GenomeViz: visualizing microbial genomes. *BMC Bioinformatics* 5, 198.
- Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, Bibbs L, Eads J, Richardson, T.H Noordewier M, et al. (14 co-authors) 2005. Genome streamlining in a cosmopolitan oceanic bacterium. *Science* 309:1242-1245.
- Goldman N, Anderson JP, Rodrigo AG. (2000). Likelihood-based tests of topologies in phylogenetics. *Syst. Biol.* 49:652-670.
- Gribaldo S, Poole AM, Daubin V, Forterre P, Brochier-Armanet CL. 2010. The origin of eukaryotes and their relationship with the Archaea: are we at a phylogenomic impasse? *Nat. Rev. Micro.* 8:743-752.
- Guerrero R, Pedros-Alio C, Esteve I, Mas J, Chase D, Margulis L. 1986. Predatory prokaryotes: predation and primary consumption evolved in bacteria. *Proc. Natl. Acad. Sci. U S A.* 83:2138-2142.
- Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52:696-704.
- Hartman H, Fedorov A. 2002. The origin of the eukaryotic cell: a genomic investigation. *Proc. Natl. Acad. Sci. U S A* 99:1420-1425.
- Hartl DL, Ochman H. 1996. Inverse polymerase chain reaction. *Methods Mol. Biol.* 58:293-301.
- Huelsenbeck JP, Ronquist F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754-755.

- Inglis TJJ, Robertson T, Woods DE, Dutton N, Chang BJ. 2003. Flagellum-mediated adhesion by *Burkholderia pseudomallei* precedes invasion of *Acanthamoeba astronyxis*. *Infect. Immun.* 71:2280-2282.
- Kanehisa M, Goto S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 8:27-30.
- Kishino H, Hasegawa M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *J. Mol. Evol.* 29:170-179.
- Kurland CG, Andersson SGE. 2000. Origin and evolution of the mitochondrial proteome. *Microbiol. Mol. Biol. Rev.* 64:786-820.
- Kurland CG, Collins LJ, Penny D. 2006. Genomics and the irreducible nature of eukaryote cells. *Science.* 312:1011-1014.
- Lane N, Martin W. 2010. The energetics of genome complexity. *Nature* 467:929-934.
- Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13:2178-2189.
- Liu R, Ochman H. (2007). Stepwise formation of the bacterial flagellar system. *Proc. Natl. Acad. Sci. U S A.* 104:7116–7121.
- Lo N, Beninati T, Sacchi L, Bandi C. 2006. An α -proteobacterium invades the mitochondria of the tick *Ixodes ricinus*. In: *Insect Symbiosis Volume 2*, Bourtzis K, Miller TA, editors, Boca Raton, Taylor & Francis, p. 25-37.
- Mahenthalingam E, Speert DP. 1995. Nonopsonic phagocytosis of *Pseudomonas aeruginosa* by macrophages and polymorphonuclear leukocytes requires the presence of the bacterial flagellum. *Infect. Immun.* 63:4519-4523.
- Margulis L. 1970. *Origin of eukaryotic cells*. New Haven: Yale University Press.
- Martin W, Muller M. 1998. The hydrogen hypothesis for the first eukaryote. *Nature* 392:37-41.
- Matz C, Kjelleberg S. 2005. Off the hook - how bacteria survive protozoan grazing. *Trends Microbiol.* 13:302-307.
- Maynard Smith J, Szathmary E. 1997. *The major transitions in evolution*. Oxford: Oxford University Press.
- Ottemann KM, Miller JF. 1997. Roles for motility in bacterial–host interactions. *Mol. Microbiol.* 24:1109-1117.
- Pitcher RS, Watmough NJ. 2004. The bacterial cytochrome *cbb*₃ oxidases. *Biochimic Biophys. Acta (BBA) - Bioenergetics* 1655:388-399.
- Poole AM, Penny D. 2006. Evaluating hypotheses for the origin of eukaryotes. *BioEssays* 29:74-84.
- Poole AM, Penny D. 2007. Engulfed by speculation. *Nature* 447:913.

- Pop M, Kosack DS, Salzberg SL. 2004. Hierarchical scaffolding with Bambus. *Genome Res.* 14:149-159.
- Puigbo P, Bravo IG, Garcia-Vallve S. 2008. CAIcal: a combined set of tools to assess codon usage adaptation. *Biol. Direct* 3:38.
- Ramos HC, Rumbo M, Sirard JC. 2004. Bacterial flagellins: mediators of pathogenicity and host immune responses in mucosa. *Trends Microbiol.* 12:509-517.
- Rambaut A, Drummond AJ. 2007. <http://beast.bio.ed.ac.uk/Tracer>
- Rivera MC, Lake JA. 1992. Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science* 257:74-76.
- 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.
- Sassera D, Beninati T, Bandi C, Bouman EA, Sacchi L, Fabbi M, Lo N. 2006. '*Candidatus* *Midichloria mitochondrii*', an endosymbiont of the tick *Ixodes ricinus* with a unique intramitochondrial lifestyle. *Int. J. Syst. Evol. Microbiol.* 56:2535-2540.
- Sassera D, Lo N, Bouman E, Epis S, Mortarino M, Bandi C. 2008. *Midichloria* endosymbionts bloom after blood meal of the host, the hard tick *Ixodes ricinus*. *Appl. Environ. Microbiol.* 74:6138-6140.
- Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* 33:686-689.
- Schmidt HA, Strimmer K, Vingron M, von Haeseler A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18:502-504.
- Shimodaira H, Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* 16:1114-1116.
- Slamovits C, Keeling P. 2006. A high density of ancient spliceosomal introns in oxymonad excavates. *BMC Evol. Biol.* 6:34.
- Staden R, Beal KF, Bonfield JK. 2000. "The Staden package, 1998". *Methods Mol. Biol.* 132:115-130.
- Strimmer KR, 2002. Inferring confidence sets of possibly misspecified gene trees. *Proc. R. Soc. Lond. B* 269:137-142.
- Toft C, Fares MA. 2008. The evolution of the flagellar assembly pathway in endosymbiotic bacterial genomes. *Mol. Biol. Evol.* 25:2069-2076.

- Tovar J, Leon-Avila G, Sanchez LB, Sutak R, Tachezy J, van der Giezen M, Hernandez M, Muller M, Lucocq JM. 2003. Mitochondrial remnant organelles of *Giardia* function in iron-sulphur protein maturation. *Nature* 426:172-176.
- Vannini C, Ferrantini F, Schleifer KH, 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. Environ. Microbiol.* 76:4047-4054.
- Vellai T, Takács K, Vida G. 1998. A new aspect to the origin and evolution of eukaryotes. *J. Mol. Evol.* 46:499-507.
- von Dohlen CD, Kohler S, Alsop ST, McManus WR. 2001. Mealybug beta-proteobacterial endosymbionts contain gamma-proteobacterial symbionts. *Nature* 412:433-436.
- Williams KP, Sobral BW, Dickerman AW. 2007. A robust species tree for the α -proteobacteria. *J. Bacteriol.* 189:4578-4586.
- Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, McGraw EA, Martin W, Esser C, Ahmadinejad N, et al. (30 co-authors). 2004. Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol* 2:E69.
- Zhu Z, Aeschlimann A, Gern L. 1992. *Rickettsia*-like microorganisms in the ovarian primordia of molting *Ixodes ricinus* (Acari: Ixodidae) larvae and nymphs. *Ann Parasitol. Hum. Comp.* 67:99-110.

3.2.4 Supplementary materials

Table S1 COG categories and repeat content in selected Rickettsiales. MID, *Midichloria mitochondrii*; APH, *Anaplasma phagocytophilum* HZ; ECH, *Ehrlichia chaffensis*; OTS, *Orientia tsutsugamushi* Boryong; RPR, *Rickettsia prowazekii* str. Madrid; WME, *Wolbachia* endosymbiont of *Drosophila melanogaster*.

COG functional group	MID	APH	ECH	OTS	RPR	WME
Translation. ribosomal structure and biogenesis	140 (11.1%)	148 (12.5%)	132 (13.7%)	133 (10.6%)	131 (15.1%)	137 (11.4%)
RNA processing and modification	9 (0.7%)	7 (0.6%)	2 (0.2%)	1 (0.1%)	0 (0%)	5 (0.4%)
Transcription	49 (3.9%)	31 (2.6%)	34 (3.5%)	72 (5.77%)	26 (3%)	46 (3.8%)
Replication. recombination and repair	88 (7%)	80 (6.8%)	62 (6.4%)	118 (9.4%)	68 (7.8%)	122 (10.2%)
Chromatin structure and dynamics	2 (0.2%)	1 (0.1%)	3 (0.3%)	0 (0%)	0 (0%)	3 (0.2%)
Cell cycle control. cell division. chromosome partitioning	22 (1.7%)	22 (1.9%)	14 (1.4%)	21 (1.7%)	19 (2.2%)	29 (2.4%)
Nuclear structure	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Defense mechanisms	17 (1.3%)	11 (0.9%)	5 (0.5%)	9 (0.7%)	10 (1.1%)	13 (1.1%)
Signal transduction mechanisms	30 (2.4%)	32 (2.7%)	21 (2.2%)	83 (6.6%)	15 (1.7%)	20 (1.7%)
Cell wall/membrane/envelope biogenesis	84 (6.6%)	36 (3%)	34 (3.5%)	38 (3%)	78 (9%)	52 (4.3%)
Cell motility	36 (2.8%)	14 (1.2%)	9 (0.9%)	18 (1.4%)	3 (0.3%)	10 (0.8%)
Cytoskeleton	1 (0.1%)	4 (0.3%)	2 (0.2%)	3 (0.2%)	0 (0%)	1 (0.1%)
Extracellular structures	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Intracellular trafficking. secretion. and vesicular transport	50 (3.9%)	64 (5.4%)	43 (4.5%)	91 (7.3%)	40 (4.6%)	51 (4.3%)
Posttranslational modification. protein turnover. chaperones	70 (5.5%)	76 (6.4%)	64 (6.6%)	75 (5.9%)	58 (6.7%)	73 (6.1%)
Energy production and conversion	92 (7.3%)	106 (9%)	79 (8.2%)	97 (7.7%)	80 (9.2%)	97 (8.1%)
Carbohydrate transport and metabolism	59 (4.7%)	43 (3.6%)	37 (3.8%)	51 (4%)	59 (7.8%)	25 (2.9%)
Aminoacid transport and metabolism	68 (5.4%)	48 (4.1%)	53 (5.5%)	51 (4%)	36 (4.2%)	49 (4.1%)
Nucleotide transport and metabolism	31 (2.4%)	51 (4.3%)	42 (4.4%)	24 (1.9%)	19 (2.2%)	39 (3.3%)
Coenzyme transport and metabolism	58 (4.6%)	79 (6.7%)	62 (6.4%)	43 (3.5%)	30 (3.5%)	52 (4.3%)
Lipid transport and metabolism	41 (3.2%)	33 (2.8%)	30 (3.1%)	35 (2.8%)	31 (3.6%)	31 (2.6%)
Inorganic ion transport and metabolism	43 (3.4%)	46 (3.9%)	37 (3.8%)	35 (2.8%)	30 (3.5%)	35 (2.9%)
Secondary metabolites biosynthesis. transport and catabolism	20 (1.6%)	19 (1.6%)	15 (1.6%)	11 (0.9%)	11 (1.3%)	14 (1.2%)
General function prediction only	127 (10%)	119 (10.1%)	101 (10.5%)	120 (9.6%)	81 (9.3%)	148 (12.4%)
Function unknown	128 (10.1%)	110 (9.3%)	83 (8.6%)	110 (8.8%)	75 (8.7%)	115 (9.6%)
Repeat content						
>200bp 100% identity	268	292	18	7622	4	328
>200bp 95% identity	2491	867	73	19659	5	781
Tandem repeats >8bp >2 repeats	559	706	1119	1768	1022	696

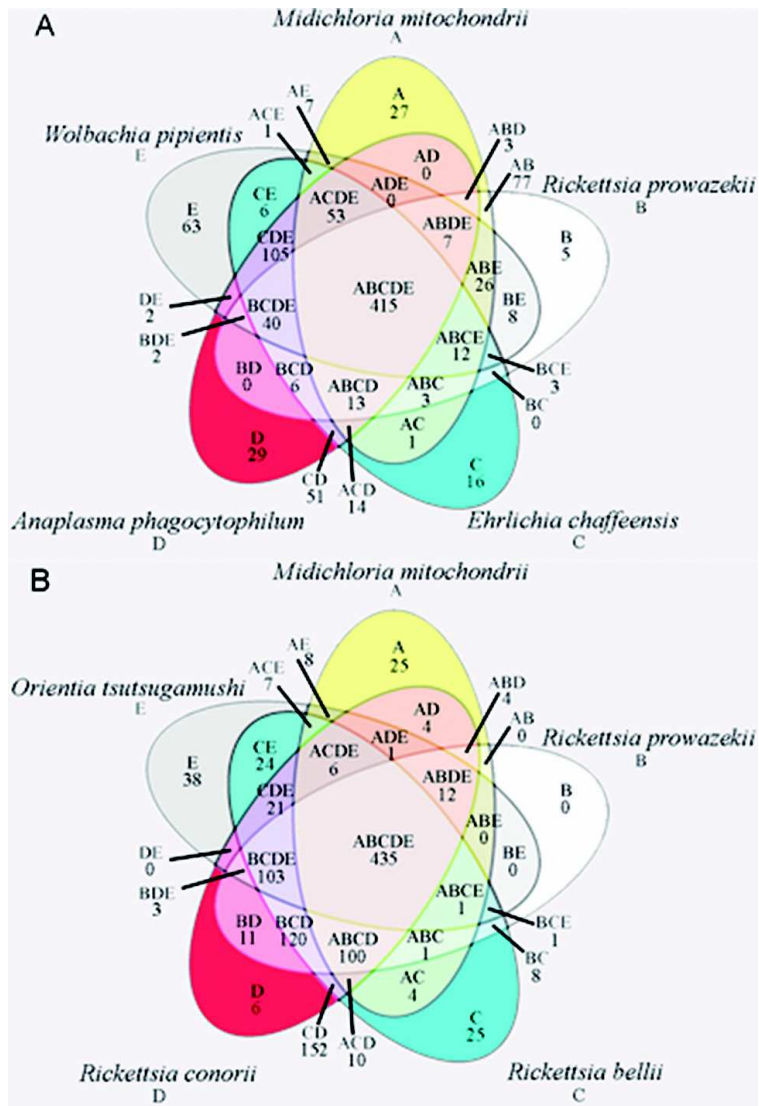


Figure S1. Venn diagrams showing ortholog groups conserved in two set of five selected Rickettsiales genomes.

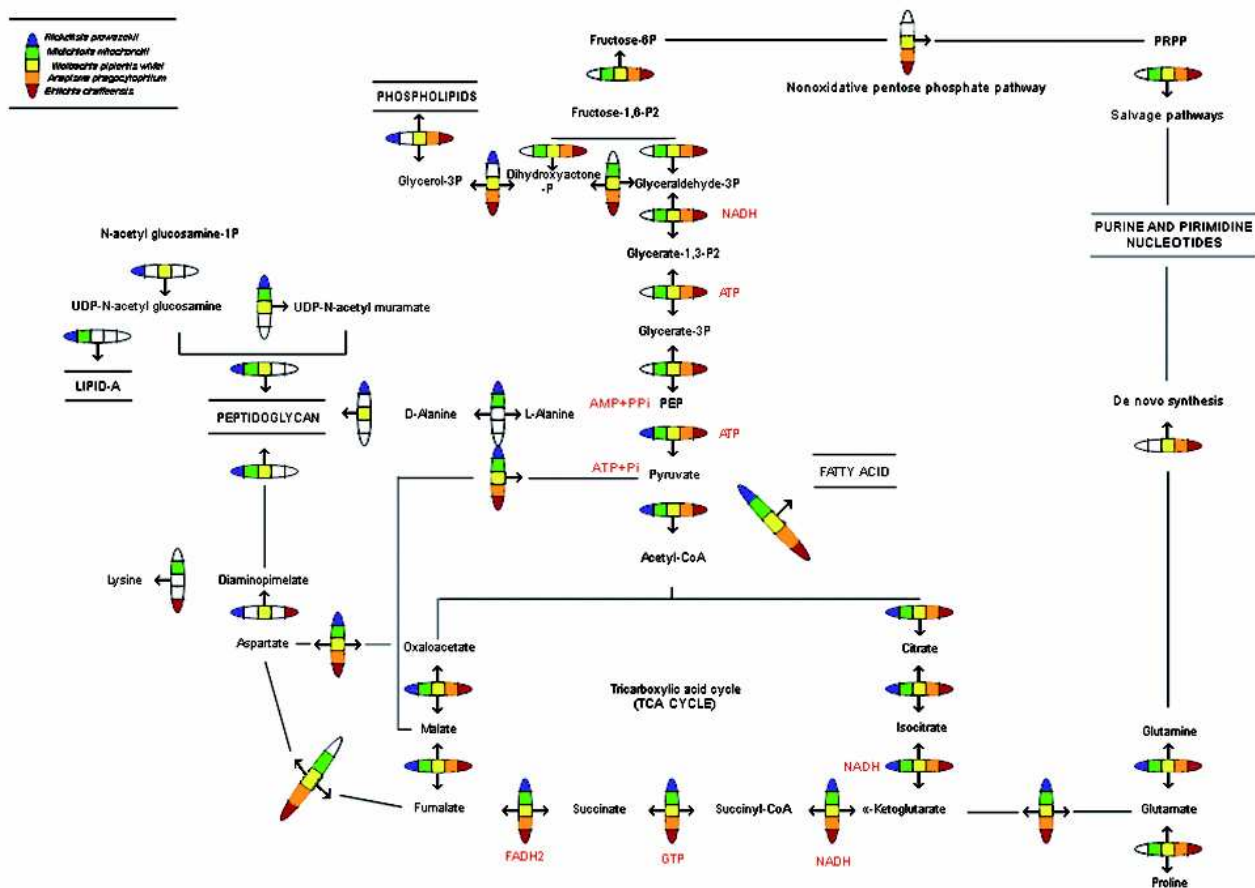
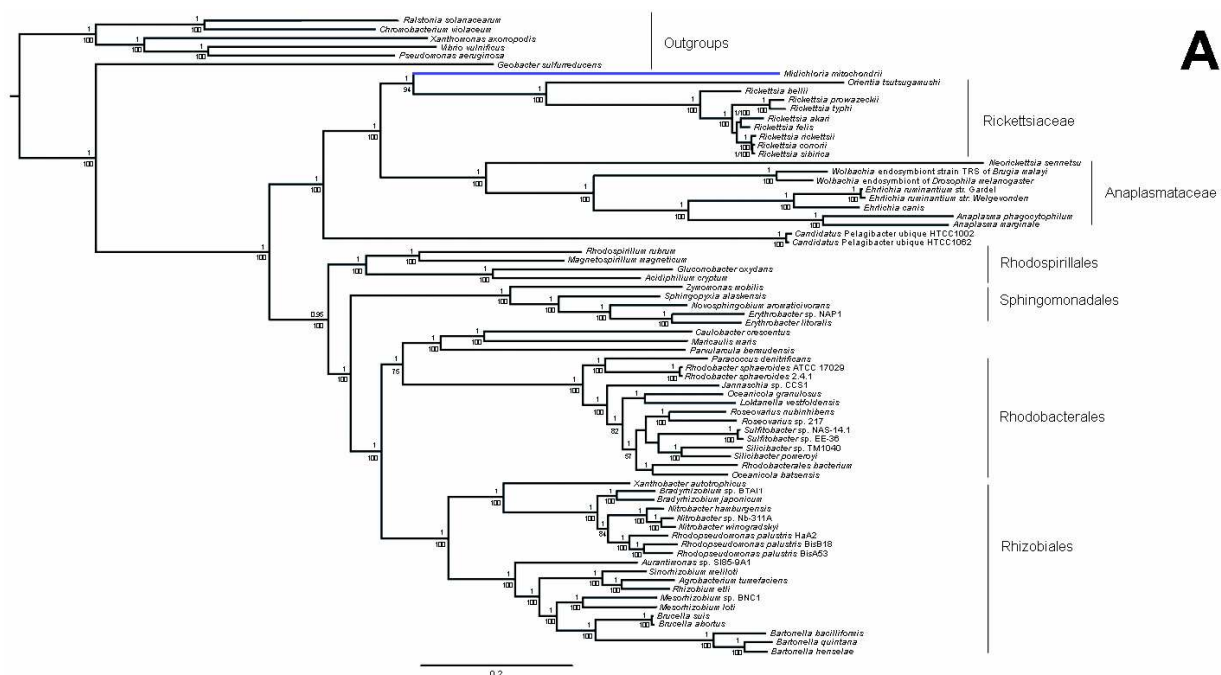
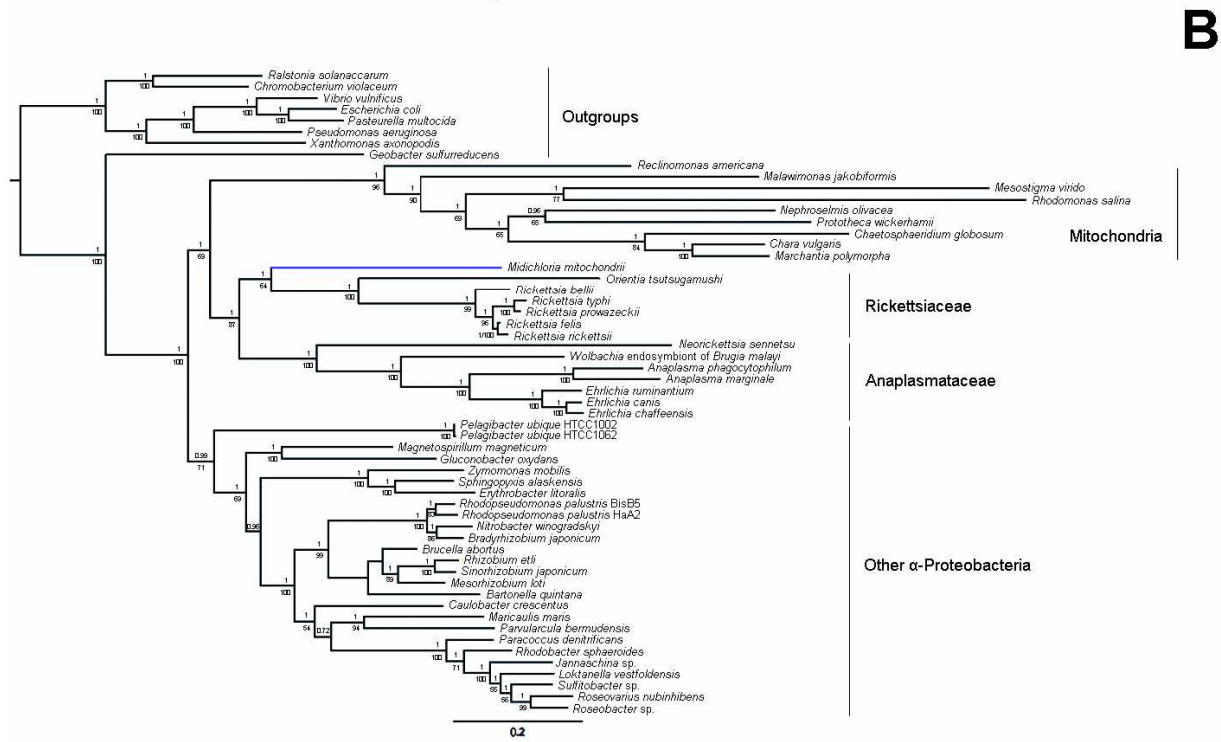


Figure S2. Metabolic pathway maps of five selected Rickettsiales, showing the presence/absence of enzymes for TCA, glycolysis, gluconeogenesis, biosynthesis of purines and pyrimidines, biosynthesis of peptidoglycans and phospholipids.



A



B

Figure S3. Phylogenetic analysis based on conserved proteins from **(A)** diverse representatives of the α -proteobacteria (88 single copy bacterial proteins), and **(B)** α -proteobacteria and mitochondria (12 single copy proteins). Concatenated alignments were analyzed using MrBayes (topology shown) and PhyML. Posterior probabilities and PhyML bootstrap supports are respectively shown above and below nodes.

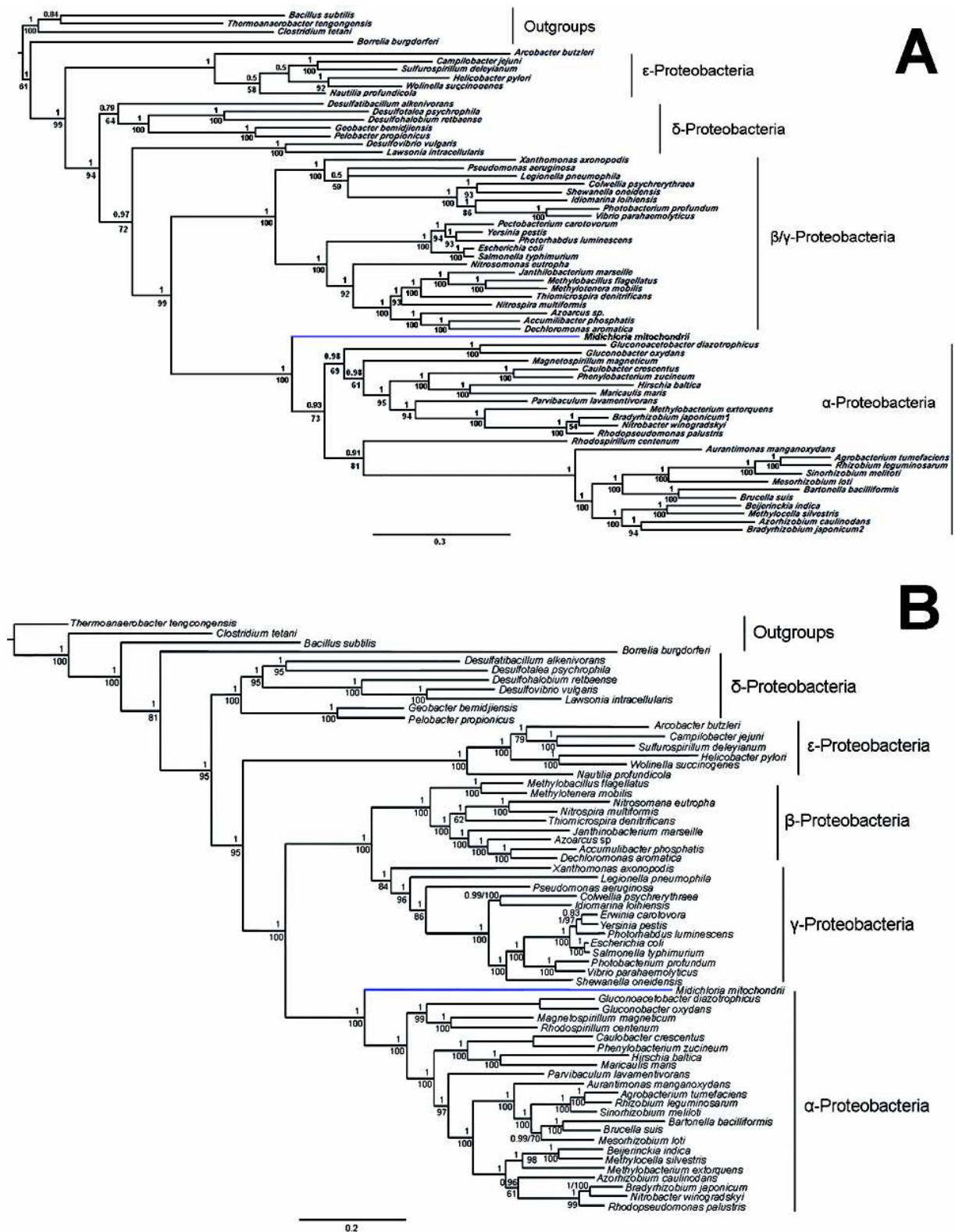


Figure S4. Phylogenetic analysis of 14 conserved flagellar proteins (**A**) and 24 core proteins (**B**) from diverse representatives of the proteobacteria phylum. Concatenated alignments were analyzed using MrBayes (topology shown) and PhyML. Posterior probabilities and PhyML bootstrap supports are respectively shown above and below nodes. In **A**, *Bradyrhizobium japonicum* 1 and 2 represent the positions inferred for the primary and secondary flagellar systems of this bacterium.

2.4 Research article. *Midichloria* and like organisms form an ecologically widespread clade of intracellular alpha-proteobacteria

2.4.1 Summary

Midichloria mitochondrii is an intramitochondrial bacterium of the order *Rickettsiales*, associated with the sheep tick *Ixodes ricinus*. Bacteria phylogenetically related with *M. mitochondrii* (*Midichloria* and like organisms, MALOs) have been shown to be associated with a wide range of hosts, from amoebae to arthropods and invertebrates, including humans. Despite numerous studies focused on specific members of the MALO group, no comprehensive phylogenetic and statistical analyses have so far been performed on the group as a whole. Here we present a multidisciplinary investigation based on 16S rRNA sequences, using both phylogenetic and statistical methods, thereby analysing MALOs in the overall framework of the *Rickettsiales*. This study revealed that 1) MALOs form a monophyletic group, 2) the MALO group is structured into distinct subgroups, verifying current genera as significant evolutionary units and identifying several subclades that could represent novel genera, 3) the MALO group ranks at the level of described *Rickettsiales* families, leading to proposal of the *Midichloriaceae* fam. nov. In addition, based on the phylogenetic trees generated, we present an evolutionary scenario to interpret the distribution and life history transitions of these microorganisms associated with highly divergent eukaryotic hosts.

2.4.2 Manuscript

Introduction

"*Candidatus* *Midichloria mitochondrii*" (hereafter *M. mitochondrii*) is an intracellular bacterium associated with the sheep tick *Ixodes ricinus* (Sassera *et al.*, 2006), the main vector of Lyme borreliosis and other diseases in Europe (Rizzoli *et al.*, 2011). *M. mitochondrii* presents an unusual life style of surviving and multiplying inside the tick mitochondria (Beninati *et al.*, 2004). Analysis of 16S rRNA gene sequences revealed that this bacterium constitutes a deep branch of the order *Rickettsiales* (Sassera *et al.*, 2006), and is not attributable to any of the families/main lineages of this order (Rodriguez-Ezpeleta and Embley, 2012). In the last decade, projects aiming at detecting and cataloguing the bacterial diversity in environmental and biological samples have unveiled that *M. mitochondrii* is likely just "the tip of an iceberg" of an emerging novel group of intracellular bacteria. These bacterial lineages phylogenetically allied to *M. mitochondrii* are associated with a wide range of hosts, scattered throughout the eukaryotic tree of life, from arthropods such as ticks,

fleas, and stink bugs, to deuterostomes including cnidarians, sponges, amoebae, ciliates, fishes, and humans (Beninati *et al.*, 2004; Erickson *et al.*, 2009; Fraune and Bosch, 2007; Fritsche *et al.*, 1999; Lloyd *et al.*, 2008; Mariconti *et al.*, 2012; Matsuura *et al.*, 2012; Mediannikov *et al.*, 2004; Sunagawa *et al.*, 2010; Vannini *et al.*, 2010). Following Mariconti *et al.* (2012), we will refer to the bacteria phylogenetically related to *M. mitochondrii* as MALOs (from: *Midichloria* and like organisms). Besides the ecological and evolutionary interest of MALOs, that derives from their widespread distribution and peculiar intramitochondrial life style (at least in ticks), we emphasize that there is growing evidence for the infectivity of these bacteria to vertebrates, including humans, and for their immunological and potentially pathogenic roles (Lloyd *et al.*, 2008; Mariconti *et al.*, 2012; Mediannikov *et al.*, 2004). As for the intra-mitochondrial niche of *M. mitochondrii*, analysis of both the symbiont and the tick mitochondrial genomes have not yet revealed any clues towards understanding the biology of this peculiar type of symbiosis (Montagna *et al.*, 2012; Sassera *et al.*, 2011).

Despite numerous investigations focused on specific members of the MALO group associated with novel hosts, and the description of novel genera and species allied to this group, no comprehensive phylogenetic, statistical and ecological studies have so far been performed on the MALOs as a whole, in the overall framework of the *Rickettsiales*. Here we present a study aiming at defining whether the MALO group represents a coherent and monophyletic clade, whether it is structured into subgroups that could represent genera, and whether the overall clade could represent a novel family of the order *Rickettsiales*.

In the second volume of Bergey's Manual of Systematic Bacteriology, the order *Rickettsiales* encompasses the three families: *Rickettsiaceae*, *Anaplasmataceae* and *Holosporaceae* (Dumler and Walker, 2005). According to the above classification, all members of this order are obligate intracellular bacteria of eukaryotes. Recently, oceanic environmental bacteria known as the SAR11 group have been proposed to form a fourth family of *Rickettsiales*, the *Pelagibacteraceae* ((Thrash *et al.*, 2011); see also (Georgiades *et al.*, 2011; Weinert *et al.*, 2009). The phylogenetic position of the SAR11 group is still debated (Rodriguez-Ezpeleta and Embley, 2012; Viklund *et al.*, 2012). The controversy around the positioning of the SAR11 group is not relevant to our current study. In order to make our work more comprehensive we included the SAR11 group in the analyses, keeping the *Pelagibacteriaceae* label, i.e. treating them as a fourth family of the *Rickettsiales*.

Our study is based on analysis of all of the 16S rRNA gene sequences available for taxonomically described *Rickettsiales* (at August 2012), plus 16S rRNA sequences available for MALOs, selecting only entire or almost entire gene sequences (see methods). We performed: 1) phylogenetic analyses aimed at verifying the monophyly of the MALO group; 2) analyses aimed at detecting evolutionary significant units within the *Rickettsiales* and the MALO clades (through the GMYC method;

(Fontaneto *et al.*, 2007; Pons *et al.*, 2006)); 3) a principal coordinate analysis (PCoA) on the 16S rRNA sequences; 4) an analysis of molecular variance (AMOVA) between the four *Rickettsiales* families (Dumler and Walker, 2005; Thrash *et al.*, 2011), with MALOs examined as a separate family, or as members of existing families, with tests on the significance of the groups defined.

The obtained results allowed us to: 1) confirm the monophyly of MALOs, and map eukaryotic hosts onto the MALO tree (with a discussion of possible life history transitions); 2) demonstrate that the MALO group is structured into subgroups, identifying current genera as evolutionary significant units (ESUs), and evidencing subclades that could represent novel genera; 3) show that the MALO group ranks at the level of described *Rickettsiales* families, and thus to propose a novel family for this bacterial order.

Materials and Methods

- Taxa selection, alignment strategies and phylogenetic analyses -

A dataset of complete/almost complete 16S rRNA gene sequence for well taxonomically identified taxa belonging to *Rickettsiales* and from MALOs were retrieved from GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>; August 2012). The 16S rRNA starting dataset was composed of 103 species/OTUs (operational taxonomic units) consisting of 7 non *Rickettsiales* Alphaproteobacteria, 8 OTUs from *Pelagibacteraceae*, 3 OTUs from *Holosporaceae*, 32 species/OTUs of *Rickettsiaceae* (genera *Rickettsia*, *Orientia*, *Cryptoprodotis*), 23 species of *Anaplasmataceae* (genera *Neorickettsia*, *Ehrlichia*, *Neoehrlichia*, *Anaplasma*, *Wolbachia*) plus 1 undescribed OTU and 29 species/OTUs from MALOs. All of the species/OTUs used in the analysis, the accession numbers and their taxonomic status are reported in Table S1 of the supplemental material. The retrieved dataset was subsequently subjected to different alignment methods in order to explore variations of the informative sites in the alignments. The dataset was aligned using two algorithms for multiple sequence alignment, Muscle (Edgar, 2004) with default settings and Mafft version 6 (<http://mafft.cbrc.jp/alignment/server/>; (Katoh *et al.*, 2005)) using G-INS-i search strategy (Katoh and Toh, 2008), 200PAM scoring matrix (considering the divergence between the analyzed organisms) and default parameters for remaining settings. The obtained datasets, respectively identified with “A” for the one obtained using Muscle and “B” for the one obtained with Mafft, were trimmed using Gblocks (Castresana, 2000) with different parameters: “conservative” (c) mode that do not allow gap positions within the final block, and “liberal” (l) mode that allow gap positions within the final block and allow less strict flanking positions. These approaches resulted in four alignments on which the phylogenetic analyses were performed. The four alignments were designed as “Ac” (aligned by Muscle plus Gblocks strict mode), “Al” (aligned

by Muscle plus Gblocks weak mode), “Bc” (aligned by Mafft plus Gblocks strict mode) and “BI” (aligned by Mafft plus Gblocks weak mode). Phylogenetic analyses were performed on each of the datasets using the following methods: the distance-matrix-based method Neighbour joining (NJ), and the character-state-based approaches Maximum Likelihood (ML) and Bayesian inference (BI). NJ tree was inferred using MEGA 5 (Tamura *et al.*, 2011) implementing Tamura-Nei (Tamura and Nei, 1993) model of nucleotide substitutions including transitions and transversions, rate variation among sites modelled with a gamma distribution (shape parameter =1) and gaps treated with partial deletion. Nodes support were estimated using 1000 bootstraps replicates. For ML and BI, the evolutionary model best-fitting to the analyzed datasets were selected with jModeltest 0.1.1 (Posada, 2008) according to the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The nucleotide substitution model selected for the further phylogenetic analysis was the General Time Reversible (GTR; (Lanave *et al.*, 1984)) with proportions of invariable sites (I) and gamma distribution (Γ). ML analyses were performed using PhyML version 3.0 (Guindon *et al.*, 2010) with the following options: GTR (Lanave *et al.*, 1984) as nucleotide substitution model, optimized proportions of invariable sites; estimated nucleotide frequency; optimized rate variation across sites into six substitution rate categories; estimated gamma-shape parameter; the best of NNI and SPR tree searching operation and with nonparametric bootstrap analysis (1000 replicates). Bayesian inference was performed with MrBayes 3.2 (Ronquist *et al.*, 2012) on the web-based Bioportal (Kumar *et al.*, 2009). In MrBayes all the Markov Chain Monte Carlo analyses were implemented into two runs of 10 M generations (gens) with four chains each, sample frequency settled every 1000 gens and the GTR substitution model (Lanave *et al.*, 1984), estimated proportions of invariable sites and gamma distribution with 6 categories as parameters of the likelihood model. The convergence of each run was verified with Tracer 1.4 (Drummond and Rambaut, 2007) and 0.25% of the sampled trees were discarded as burn-in.

- *Detection of Evolutionary Significant Units (ESUs)* -

The generalized mixed Yule coalescent (GMYC) method (Fontaneto *et al.*, 2007; Pons *et al.*, 2006) was applied to 16S rRNA gene tree obtained with ML on alignment A1 in order to identify ESUs (Barraclough *et al.*, 2009). The null hypothesis assumes that all samples belong to a single entity, while the alternative to be tested assumes that the samples are divided into n independently evolving units; the log-likelihood ratio test is performed to compare the likelihood of the two models. The method is implemented in the R package "GMYC" in "splits" (SPecies Limits by Threshold Statistics, available at <http://r-forge.r-project.org/projects/splits/>). The ML tree inferred on the alignment A1 was processed in order to drop all branches with 0 length. This operation resulted in a tree with 71 terminal nodes and 69 internal nodes. The tree is converted to ultrametric

using penalized likelihood with a smoothing parameter of 1 (selected after cross-validation of values between 0.1 to 100) as implemented in r8s 1.7 (Sanderson 2003).

- *Principal coordinate analysis (PCoA)* -

The principal coordinate analysis (PCoA; (Gower, 1966)) resulted the most suited ordination method for the input data, consisting in a distance matrix (Sneath and Sokal, 1973), in order to place the analyzed species/OTUs in a new coordinate system. The pairwise nucleotide distance matrix was calculated on alignment A1, that displayed the highest number of likelihood informative sites, using MEGA 5 (Tamura *et al.*, 2011) following the removal of the outgroup taxa (i.e. on the remaining 96 species/OTUs). The implemented model of nucleotide substitutions was the Tamura-Nei (Tamura and Nei, 1993) including transitions and transversions, rate variation among sites modelled with a gamma distribution (shape parameter =1) and gaps treated with partial deletion. The PCoA analysis was performed with the software MVSP 3.1 (Kovach, 1999) using a Euclidean metric and no data transformation.

- *Analysis of Molecular Variance (AMOVA)* -

Analysis of molecular variance (AMOVA) was performed to quantify the genetic variation between and within the families of Rickettsiales plus MALOs. AMOVA (Excoffier *et al.*, 1992; Weir, 1984; Wier, 1996) was performed on the nucleotide distance matrix used as input for PCoA analysis (that used as input alignment A1). AMOVA was performed considering three different possible scenarios, based on phylogenetic results, in order to explore the variability of the molecular variances between/within the families of Rickettsiales+MALOs in the three different situations. The three scenarios are defined as follows: 1) MALOs assumed to be part of *Rickettsiaceae*, and the dataset divided into four groups corresponding to the four described families; 2) MALOs considered part of *Anaplasmataceae*, again with four groups corresponding to the families; 3) MALOs considered separately, with the dataset divided into five groups, corresponding to the four described families plus MALOs.

- *Analysis of similarity between the described families + MALOs* -

In order to assess whether the five groups identified by phylogenetic analysis, GMYC and PCoA analysis were significantly different and to validate our decision to elevate MALOs to the family rank, the genetic pairwise distance matrix used for PCoA and AMOVA analyses was subjected to a non-parametric, one-way analysis of similarity (ANOSIM; (Clarke, 1993)) with permutations (999). The *R value*, obtained by the ANOSIM analysis, is a measure of similarity between groups; an *R value* of 1 indicates that members of the selected group are more similar to each other respect to

members of other groups. ANOSIM analysis was performed using the R package "ANOSIM" in "vegan" (Dixon, 2003).

- Between- and within-group nucleotide distances -

For the five groups examined, *i.e.* MALOs and the four families of the *Rickettsiales*, the intra-group mean nucleotide distances and the between-group mean distances were calculated on the *p*-distance matrix with a complete deletion of gaps, using MEGA 5 (Tamura *et al.*, 2011).

Results and Discussion

The different alignment algorithms and masking strategies resulted in the following four aligned datasets: two alignments without gaps (Ac and Bc of 868 and 867 bp in size, respectively) and two alignments with gaps (Al and Bl of 1396 and 1361 bp in size, respectively). Features of the four datasets obtained with different alignment algorithms and masking procedures (*i.e.* the size, the likelihood informative sites, the parsimony informative sites and sites without polymorphism) are reported in Table 1.

Table 1. Features of the 16S rRNA alignment datasets. Four datasets with different characteristics with respect to alignment algorithms, Gblock strategies, size, likelihood and parsimony informative sites, and sites without polymorphism, were generated.

Dataset	Alignment algorithm	Gblocks options	Size (bp)	Likelihood informative sites	Parsimony informative sites	Sites without polymorphism
Ac	MUSCLE	conservative	868	403 (46.4%)	360 (41.5%)	448 (51.6%)
Al	MUSCLE	liberal	1396	813 (58.2%)	820 (58.7%)	635 (45.5%)
Bc	MAFFT	conservative	867	403 (46.5%)	362 (41.8%)	447 (51.6%)
Bl	MAFFT	liberal	1361	787(57.8%)	787 (57.8%)	574 (42.2%)

Phylogenetic analyses based on the four datasets consistently indicated that MALOs are monophyletic, as are four described families of the *Rickettsiales* (*Rickettsiaceae*, *Anaplasmataceae*, *Holosporaceae* and *Pelagibacteraceae*; Figures 1 and 2). Almost all of the phylogenetic analyses based on the four datasets using different inference methods (BI, ML and NJ) agreed with each other in terms of the relationships between the *Rickettsiales* families (Fig. 1). Figure 1 shows the consensus cladogram of the 11 of 12 phylogenetic relationships inferred from the four datasets analyzed by three methods (one topology was slightly different in that MALOs formed a sister group of *Rickettsiaceae*). Differences in the values of branch support were found among the trees inferred from the four datasets, as reported in figure 1. The alignments obtained by the “conservative” masking strategy (Ac and Bc) contained less informative sites (403 likelihood

informative sites) than the alignments obtained using a “liberal” masking strategy (Al and B1 with 813 and 787 likelihood informative sites, respectively), resulting in lower phylogenetic resolution in the former than in the latter at family level and below. The obtained results (excluding the position of MALOs) were in agreement with previously published topologies inferred from concatenated gene/protein alignments (Sassera *et al.*, 2011; Thrash *et al.*, 2011; Williams *et al.*, 2007).

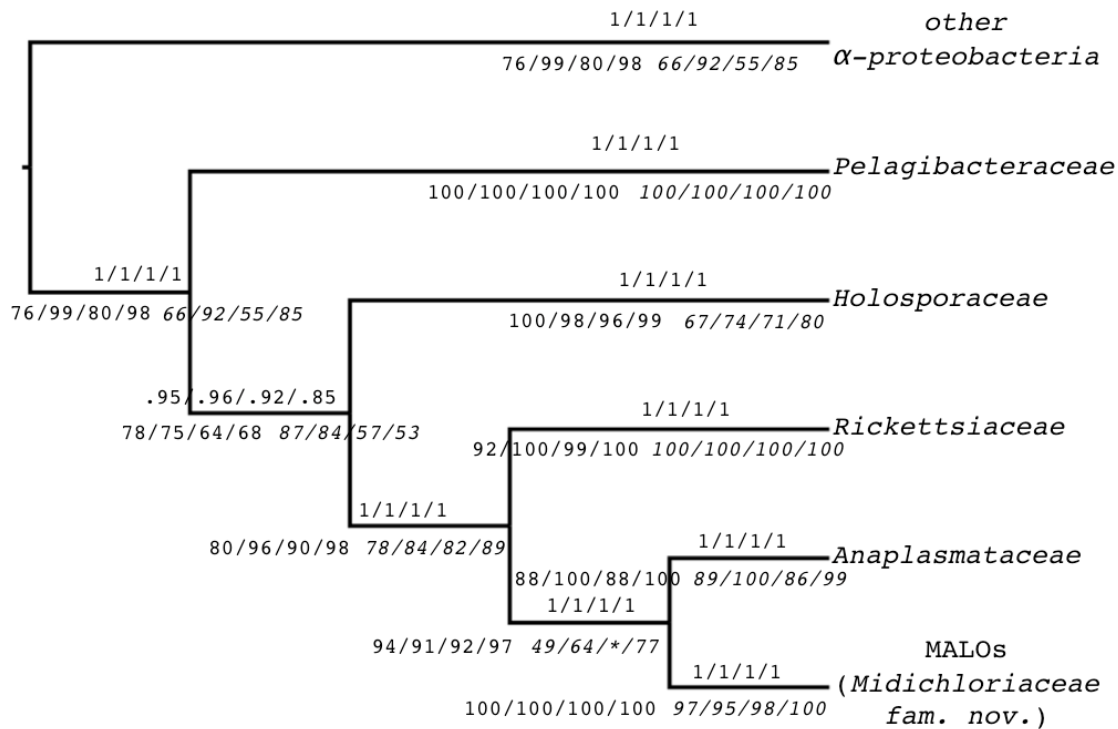


Figure 1. Majority-rule consensus cladogram of the four *Rickettsiales* families plus MALOs based on the four 16S rRNA datasets (Ac, Al, Bc and B1) implementing BI, ML and NJ inferring methods. Support values inferred from the four aligned datasets are reported on the branches: above, the Bayesian posterior probability; below, the maximum likelihood (in regular) and NJ (in italic) bootstrap values. The star indicates the only partition that does not agree with the majority-rule consensus cladogram (i.e. MALOs were found to be the sister clade of *Rickettsiaceae* inferred with NJ on dataset Bc).

Figure 2 shows the unrooted phylogeny inferred from Al (the dataset with the highest number of informative sites, 813) using Bayesian methods. In our analyses based on 16S rRNA gene sequences, MALOs were placed as the sister group of *Anaplasmataceae*, which is concordant with some reports based on 16S rRNA gene sequences (e.g. (Sassera *et al.*, 2006; Vannini *et al.*, 2010)) and multiple protein sequence alignments (Rodriguez-Ezpeleta and Embley, 2012), but differs with other reports that found MALOs as a sister group of *Rickettsiaceae* (Matsuura *et al.*, 2012; Rodriguez-Ezpeleta and Embley, 2012; Sassera *et al.*, 2011). Phylogenetic relationships within the

groups *Holosporaceae*, *Anaplasmataceae* and MALOs were well resolved, while the exact positioning of some lineages in *Rickettsiaceae* was elusive. The results inferred from all the four datasets with both Bayesian and Maximum Likelihood methods exhibited consistent relationships between the taxa/OTUs within the MALO group (Fig. 2 and 3).

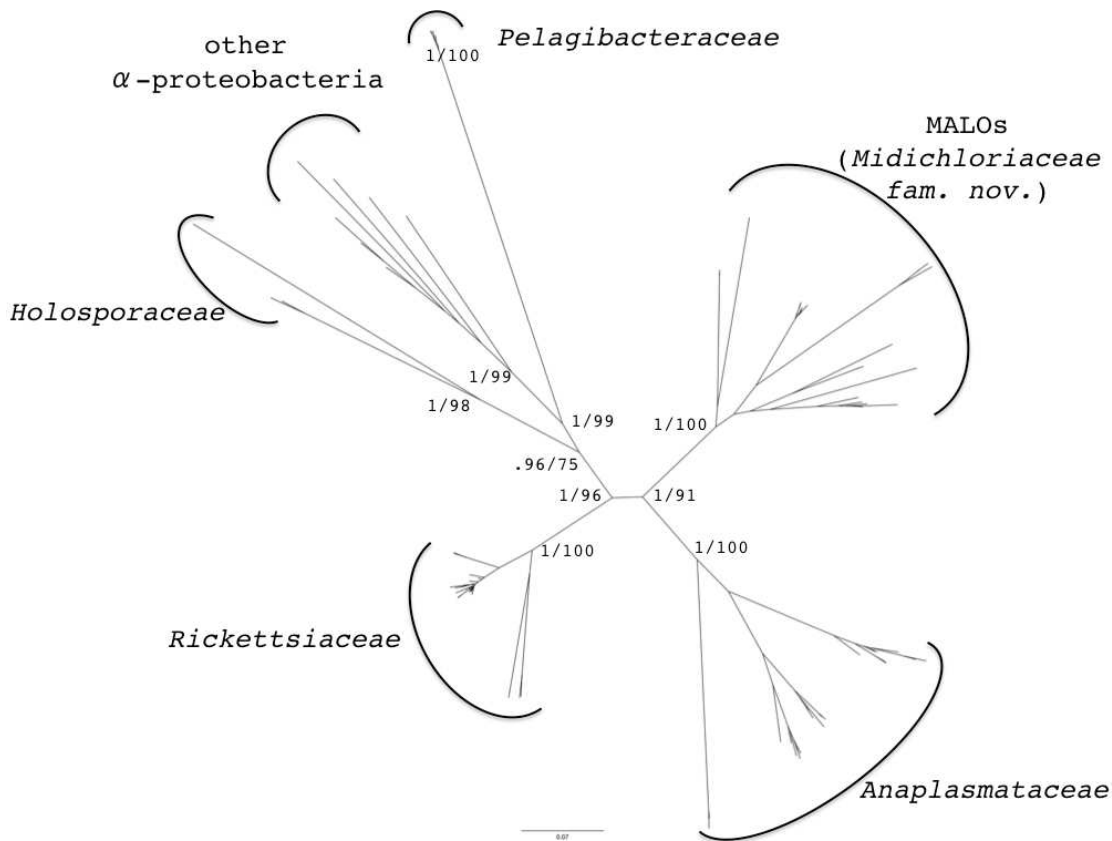


Figure 2. Unrooted *Rickettsiales* Bayesian phylogram calculated on the 16S rRNA A1 alignment. Bayesian posterior probability and ML bootstrap values are reported for the main lineages.

Figure 3 shows the phylogenetic relationship of MALOs only, along with their source hosts/environmental origins. While the phylogenetic relationship among the MALO taxa/OTUs were generally well-resolved and supported (Figure 2 and 3), polytomies were found within the groups "*Candidatus* *Cyrtobacter comes*" (from protists) and "*Candidatus* *Lariskella arthropodarum*" (from arthropods). The basal branch within MALOs is the lineage leading to bacteria harboured by sea water ciliates and bacteria from water samples of Kelike Lake in Tibet (it is possible that these are symbionts of some aquatic eukaryote). It is notable that host taxonomic groups were generally coherent: for example, we identified MALO clades exclusively associated with cnidarians, arthropods, amoebae, and ciliates respectively. However, there are some exceptions, e.g. "*Candidatus* *Anadelfobacter veles*" was not grouped with the other symbionts of ciliates (i.e. *Ca. C. comes*) but clustered with the group of MALOs associated with ticks. The taxonomic distribution of

MALO hosts was quite diverse, ranging from amoebae to vertebrates. We speculate that aquatic eukaryotic microorganisms like ciliates and amoebae, which feed on bacteria and are likely prone to establish stable interactions with phagocytosed bacteria, might represent a sort of evolutionary reservoir of MALOs, from which the lineages infecting animals could have evolved, plausibly more than once. The deepest branch in the MALO tree leads to symbionts associated with aquatic protists (Fig. 3, clade 1). Furthermore, a sister group relationship is also observed between symbionts of other protists and those of metazoans, i.e. vertebrates and ticks (figure 3, clade 2). Finally, there is also a group of MALOs associated with filter-feeding aquatic invertebrates, that is sister group to arthropod-associated MALOs (Fig. 3, clade 3). These relationships indicate the potential for the origin of metazoa-associated MALOs from water-dwelling organisms. Regarding clade 2 in Figure 3, one hypothesis to explain the trajectory of MALOs from aquatic/environmental protista/amoebae to ticks could be the infection (even transient) of a vertebrate host, from which ticks obtain their blood meal. Amoebae demonstrated to harbour MALOs have been isolated from humans (Fritsche et al., 1999), and belong to the genus *Acanthamoeba*, a well-known group of environmental protists capable of infecting vertebrates. The possibility that bacteria from the MALO group are infectious to vertebrates is supported by several independent studies, from the well established case of the rainbow trout (Lloyd et al., 2008), to the results indicating a circulation of *M. mitochondrii* in humans ((Mariconti et al., 2012); see also introduction). Therefore, although it is not certain at this stage that MALOs are infectious and pathogenic to vertebrates, it is reasonable to conclude that vertebrates are transient hosts for these bacteria (in Figure 3 we have thus indicated a vertebrate host, i.e. *Homo sapiens*, as a possible host for MALOs from clades 2 and 3). In summary, the scenario that we propose is that aquatic/environmental protista have acted as evolutionary reservoirs of MALOs, from which one or more lineages have evolved, with the capacity of infecting metazoa.

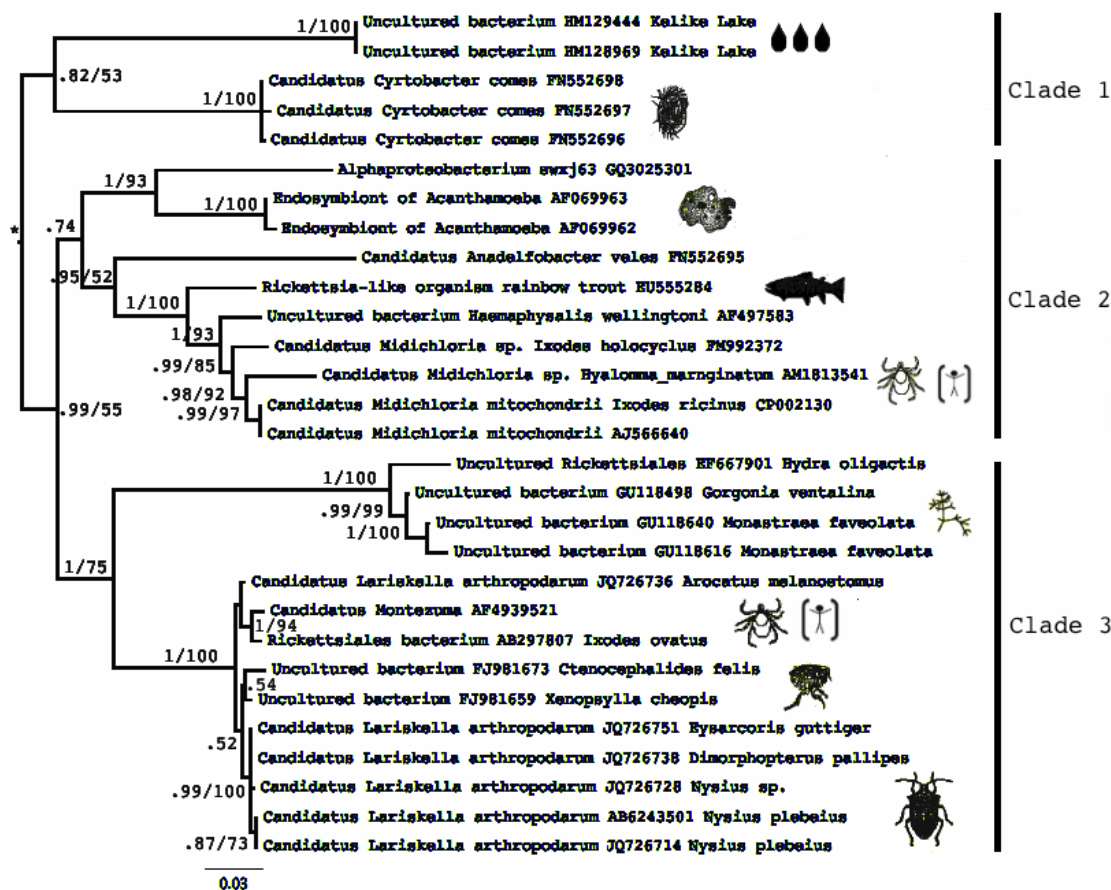


Figure 3. MALOs Bayesian phylogram obtained from the 16S rRNA A1 dataset. Values on the nodes are the Bayesian posterior probabilities and maximum likelihood bootstrap percentages (values below 50% were not reported). Hosts from which bacteria were extracted are mapped beside the terminal tips.

In addition to the above phylogenetic analyses, we also applied the GMYC method (Pons *et al.*, 2006), a tool developed for the detection of evolutionary significant units (ESUs) within a given data set, based on a maximum-likelihood approach, that has recently also been applied to a 16S rRNA data set from prokaryotes (Barracough *et al.*, 2009). The main results of GMYC analysis are shown in Figure 4. The GMYC model exhibited a significantly better likelihood than the null model ($\log L_{\text{GMYC}} = 303$, $\log L_{\text{NULL}} = 295.4$; $2\Delta L = 15.2$, χ^2 test, $P = 0.0016$), and identified a transition (T1) in branching rate at the threshold time -0.094 from the present which detected 12 ML ESUs clusters for a total of 22 ML entities (clusters range 11-14; entities range 18-26). Analyzing the peaks on the likelihood plot, we derived a second threshold line (T2, threshold time -0.7 from the present, $\log L = 297.5$). The two threshold lines, albeit generated independently from any taxonomic/nomenclature information, correspond respectively to the genus level (T1) and to the family rank (Fig. 4). Considering T1 in detail, there is an overall correspondence throughout the order *Rickettsiales* between the ESUs identified by this analysis, and described genera. Minor differences are observed in the genus *Rickettsia*, where GMYC recognizes two ESUs, the first encompassing most of the

species so far described (*Rickettsia* cluster 1, which includes human pathogens), and a second (*Rickettsia* cluster 2), which encompasses *Rickettsia* endosymbiont of *Torix tukubanatorix* and *Rickettsia* endosymbiont of *Hemiclepsis marginata* (leaches), and *R. limoniae* (associated with a dipteran insect). We emphasize that 16S rRNA sequences from these rickettsiae had already been recognized to be highly divergent (Kikuchi *et al.*, 2002). Based on our GMYC analysis, this clade of organisms currently classified as *Rickettsia* could indeed be considered a separate novel genus of the family *Rickettsiaceae*. We hypothesize that other *Rickettsiaceae* present in ciliates could represent a further genus within the family ((Vannini *et al.*, 2005); not included in our current study). With regard to MALOs, GMYC recognizes three groups of sequences that are defined at the T1 threshold as ESUs, i.e. MALOs cluster 2 (from *Ixodes ricinus* and other ticks), MALOs cluster 3a (from cnidarians and sponge), MALOs cluster 3b (from Hemiptera bugs, from fleas and from ticks). In addition, GMYC identified six single-lineage ESUs at T1 level, corresponding to two described genera (i.e. *Cyrtobacter* and *Anadelfobacter* from ciliates) and four undescribed lineages (from two types of environmental samples; from *Achantamoeba* sp.; from the rainbow trout). In summary, threshold level T1, which shows a good correspondence with the genus level for all of the other *Rickettsiales*, indicates the existence of nine lineages that might deserve to be ranked at the genus level within the MALO group, supporting the four published genera within this group (i.e. *Midichloria*, *Lariskella*, *Anadelfobacter* and *Cyrtobacter*), and indicating that five further genera could be proposed. Interestingly, T2, the second identified likelihood peak, identifies five evolutionary lineages, that correspond to the four described families of the *Rickettsiales*, plus MALOs (Fig. 4). Thus, at the family level the GMYC method also generates results that match with the current classification of *Rickettsiales*, and indicates that the MALO group should to be elevated at the family rank. We thus performed a series of analyses to determine whether this group deserves to be elevated to the family rank.

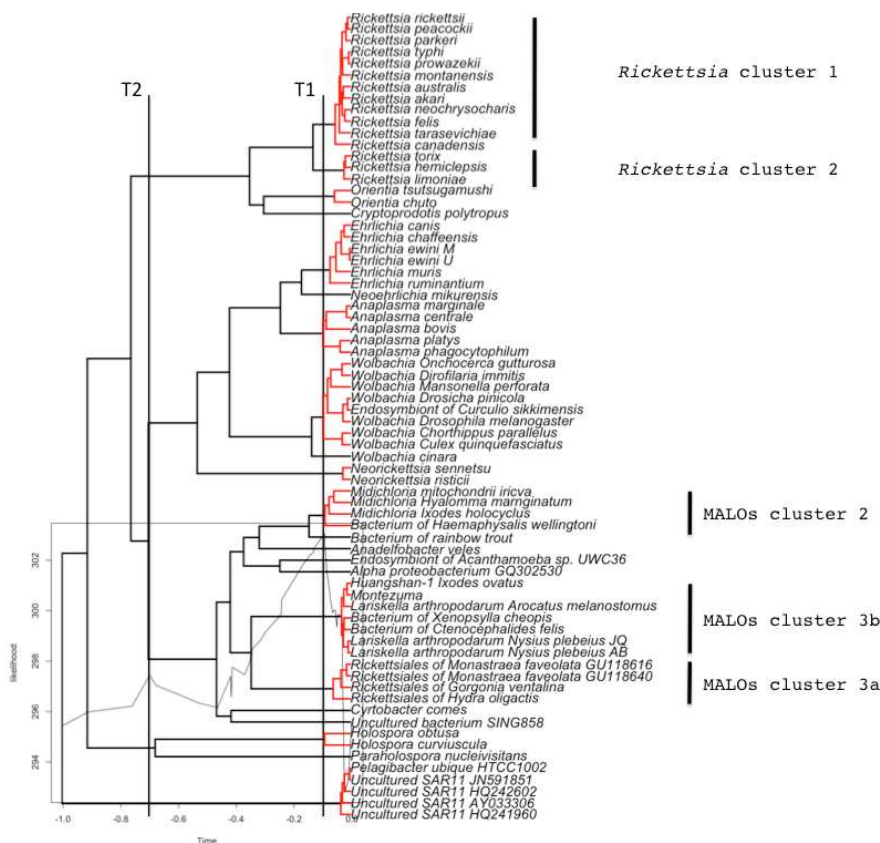


Figure 4. Maximum likelihood ultrametric tree obtained from the 16S rRNA A1 dataset, depicting the identified Evolutionary Significant Units (ESUs) within *Rickettsiales* plus MALOs (see Materials and Methods). Clusters of ESUs are highlighted in red, black terminal tips indicate the identified single entities. The likelihood through time plot is mapped on the tree. The vertical black line T1 shows the maximum likelihood transition point of the switch in branching rates. T2 shows a second likelihood peak corresponding to the family rank on the tree.

The results of the PCoA performed on the distance matrix generated on A1 are presented graphically in Figure 5 (the PCoA case score plot). After the analysis of the eigenvalues and of the scree plot, the first three principal components were considered. These components explain a total of 89.1% of the variation (1st component = 52.6%; 2nd component = 24.2% and the 3rd component = 12.3%). Figure 5 reports the case score plot for the first three principal components, in which five groups of variables are displayed, with different shapes and colors, and correspond to the five clades of *Rickettsiales* (i.e. the four described families plus MALOs). We note that all the five clades of *Rickettsiales* are well isolated. Interestingly all of the members of MALOs are grouped together, while there are members of *Rickettsiaceae* and *Anaplasmataceae* that are located rather far from the centroid of the corresponding family. Within *Rickettsiaceae*, the deviating taxa (indicated by an arrow) are the two species of *Orientia* and *Ca. Cryptoprodotis polytropus*. In agreement with this result, in the phylogenetic tree these species stem as long basal branches and are quite divergent from the members of the genus *Rickettsia* (Fig. 2). A similar situation arises in both

Anaplasmataceae and *Holosporaceae*, where the two species of *Neorickettsia* and *Ca. Paraholospora nucleivisitans*, which constitute basal and a long phylogenetic branches within the families (Fig. 2), appear quite divergent in the PCoA case score plot (respectively identified by asterisk and cross).

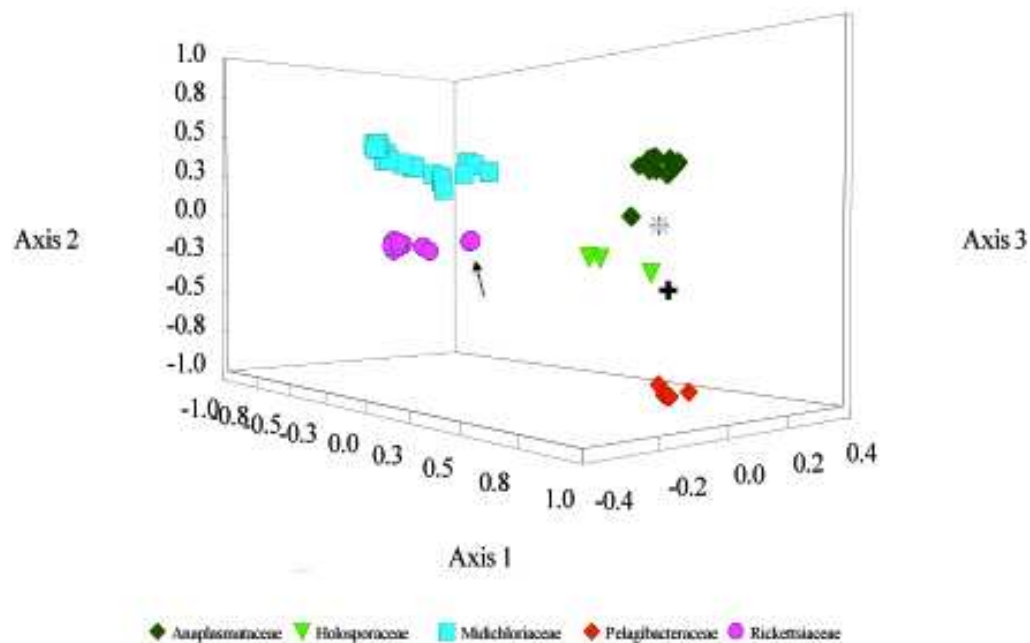


Figure 5. Principal coordinate analysis based on euclidean distances of the pairwise nucleotide distance matrix obtained from the 16S rRNA A1 dataset (explained variances 1st axis: 52.6%; 2nd axis: 24.2%; 3rd axis: 12.3%). The four *Rickettsiales* families plus MALOs are labelled as follows: *Anaplasmataceae* dark green diamonds; *Holosporaceae* light green triangles; MALOs blue squares; *Pelagibacteraceae* red diamonds; *Rickettsiaceae* purple circles. The arrow, the cross and the asterisk indicate the outliers: respectively the two species of *Orientia* and *Ca. Cryptoprodotis polytropus*, and the two species of *Neorickettsia*, *Ca. Paraholospora nucleivisitans*.

AMOVA analyses were performed on the nucleotide pairwise genetic distance matrix (implementing Tamura-Nei (Tamura and Nei, 1993) model of nucleotide substitutions, see Materials and Methods) estimated on A1 (after removal of the outgroups) simulating three possible scenarios on the ranking and positioning of MALOs: 1) assumed to be part of the *Rickettsiaceae*; 2) assumed to be part of the *Anaplasmataceae*; 3) assumed to be a separate family (see Materials and Methods for details). The results of AMOVA analyses (reported in Table 3) indicate that when MALOs are considered as a separate family, the explained percentage of variance among families is 63.1%, i.e. well above that explained when MALOs were grouped into the *Anaplasmataceae* (55.5%) or the *Rickettsiaceae* (49.6%). In parallel, the explained percentage of variation within families is lower when MALOs are considered as a separate group (see Table 2). In other words, the

assumption of MALOs as a separate group at the family rank maximizes the inter-group variance, and minimizes that within the groups.

Table 2. Results from the analysis of molecular variance (AMOVA).

Structure	Groups (number of taxa)	Source of variation	Variance component	% variation explained
MALOs assumed as part of <i>Anaplasmataceae</i>	<i>Anaplasmataceae</i> + MALOs (53) <i>Rickettsiaceae</i> (32) <i>Holosporaceae</i> (3) <i>Pelagibacteraceae</i> (8)	Among groups Within groups	V _a = 0.60 V _b = 0.48	55.5 44.5
MALOs assumed as part of <i>Rickettsiaceae</i>	<i>Rickettsiaceae</i> + MALOs (61) <i>Anaplasmataceae</i> (24) <i>Holosporaceae</i> (3) <i>Pelagibacteraceae</i> (8)	Among groups Within groups	V _a = 0.54 V _b = 0.55	49.6 50.4
MALOs assumed as separate group	1. Members of the new family (29) 2. <i>Anaplasmataceae</i> (24) 3. <i>Rickettsiaceae</i> (32) 4. <i>Holosporaceae</i> (3) 5. <i>Pelagibacteraceae</i> (8)	Among groups Within groups	V _a = 0.63 V _b = 0.37	63.1 36.9

The ANOSIM analysis performed on the estimated genetic pairwise distance matrix on alignment A1 confirms a significant diversity between the five groups examined here, i.e. the four families of *Rickettsiales* plus MALOs (R = 0.976, significance P = 0.001).

-Concluding remarks-

In summary, the results obtained by different approaches (i.e. phylogeny, GMYC, PCoA and AMOVA) indicate that bacteria belonging to the MALO group deserve to be ranked at the family level. We thus propose to create a novel family in the order *Rickettsiales*, i.e. the *Midichloriaceae* fam. nov. The name *Midichloriaceae* (*mi.di.chlo'ria.cea*) derives from the first described member of this family, Ca. *Midichloria mitochondrii*, the intra-mitochondrial bacterium of the sheep tick *Ixodes ricinus*. Table 3 presents a summary of the distances within and between the *Rickettsiales* families.

Table 3. Nucleotide p-distances within and between the four families of *Rickettsiales* plus MALOs. Below the diagonal: values of p-distances between the five clades and standard deviation (%). On the diagonal: within-group p-distances, standard deviations and the min-max values (%). Above the diagonal: values of net p-distances between the five clades and standard deviations (%).

Family	<i>Midichlor.</i>	<i>Anaplasma.</i>	<i>Ricketts.</i>	<i>Pelagibact.</i>	<i>Holospor.</i>
<i>Midichloriaceae</i>	8.7 (sd ± .5) [0,14.7]	5.9 (sd ± .6)	8.1 (sd ± .7)	14.8 (sd ± 1)	7 (sd ± .7)
<i>Anaplasmataceae</i>	14.6 (sd ± .9)	9 (sd ± .5) [0.1,15.6]	9.3 (sd ± .7)	15.6 (sd ± 1)	6.9 (sd ± .6)
<i>Rickettsiaceae</i>	13.6 (sd ± .9)	14.8 (sd ± .9)	2.7 (sd ± .2) [0,10.2]	18.3 (sd ± 1.1)	8.9 (sd ± .7)
<i>Pelagibacteraceae</i>	19.4 (sd ± 1.1)	20.2 (sd ± 1.1)	19.8 (sd ± 1.2)	0.8 (sd ± .2) [0.1,1.3]	14.4 (sd ± 1)
<i>Holosporaceae</i>	17.7 (sd ± .9)	17.6 (sd ± .9)	16.5 (sd ± 1)	21.1 (sd ± 1.2)	12.7 (sd ± .7) [3.2,17.7]

The new family encompasses bacteria associated with a wide range of hosts, from protists to vertebrates including humans. The *Midichloriaceae* are thus to be regarded as organisms of potential medical and veterinary relevance, considering their capacity to induce an immune response in tick-exposed human subjects (Mariconti *et al.*, 2012). The results obtained by phylogenetic analyses for members of *Midichloriaceae* and their hosts suggest a possible scenario of life history transitions, from MALOs infecting water-dwelling protists to those infecting different types of metazoa (as discussed above). The idea that arthropod- and vertebrate-associated intracellular bacteria arose from aquatic/environmental protists is not new (Fritsche *et al.*, 1999). In addition to the scenario here discussed for the *Midichloriaceae*, there is also evidence for the deep branching of a lineage of the *Rickettsiaceae* that infects aquatic protista (Vannini *et al.*, 2005). Considering that current evidence places the ciliate-infecting bacteria *Holosporaceae* as the sister group of the lineage leading to *Rickettsiaceae*, *Anaplasmataceae* and *Midichloriaceae*, there is an overall evidence indicating that intracellular *Rickettsiales* were originally associated with aquatic/environmental protista, that served (and potentially still serves) as an ecological and evolutionary reservoir for the *Rickettsiales* infecting animals.

2.4.3 References

- Barraclough TG, Hughes M, Ashford-Hodges N, Fujisawa T, 2009. Inferring evolutionarily significant units of bacterial diversity from broad environmental surveys of single-locus data. *Biol Lett.* 5: 425-428.
- Beninati T, Lo N, Sacchi L, Genchi C, Noda H, Bandi C, 2004. A novel alpha-Proteobacterium resides in the mitochondria of ovarian cells of the tick *Ixodes ricinus*. *Appl Environ Microbiol.* 70: 2596-2602.
- Castresana J, 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol.* 17: 540-552.
- Clarke K, 1993. Non-parametric multivariate analysis of changes in community structure. *Australian Journal of Ecology* 18: 117-143.
- Dixon P, 2003. VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* 6:927-930.
- Drummond AJ, Rambaut A, 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 7: 214.
- Dumler J, Walker D, 2005. Order II. Rickettsiales Gieszczykiewicz 1939, 25AL emend. Dumler, Barbet, Bekker, Dasch, Palmer, Ray, Rikihisa and Rurangirwa 2001, 2156, In: Garrity, G., Brenner, D., Krieg, N., Staley, J. (Eds.) *Bergey's manual of systematic bacteriology*, 2nd ed. Springer-Verlag, New York, pp. 96-160.
- Edgar RC, 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics.* 5, 113.
- Erickson DL, Anderson NE, Cromar LM, Jolley A, 2009. Bacterial communities associated with flea vectors of plague. *J Med Entomol.* 46: 1532-1536.
- Excoffier L, Smouse PE, Quattro JM, 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics.* 131: 479-491.
- Fontaneto D, Herniou EA, Boschetti C, Caprioli M, Melone G, Ricci C, Barraclough TG, 2007. Independently evolving species in asexual bdelloid rotifers. *PLoS Biol.* 5: e87.
- Fraune S, Bosch TC, 2007. Long-term maintenance of species-specific bacterial microbiota in the basal metazoan *Hydra*. *Proc Natl Acad Sci U S A.* 104: 13146-13151.
- Fritsche TR, Horn M, Seyedirashiti S, Gautom RK, Schleifer KH, Wagner M, 1999. In situ detection of novel bacterial endosymbionts of *Acanthamoeba* spp. phylogenetically related to members of the order Rickettsiales. *Appl Environ Microbiol.* 65: 206-212.

- Georgiades K, Madoui MA, Le P, Robert C, Raoult D, 2011. Phylogenomic analysis of *Odyssella thessalonicensis* fortifies the common origin of Rickettsiales, *Pelagibacter ubique* and *Reclimonas americana* mitochondrion. *PLoS One*. 6: e24857.
- Gower J, 1966. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53: 325-338.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O, 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 59: 307-321.
- Katoh K, Kuma K, Toh H, Miyata T, 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res*. 33: 511-518.
- Katoh, K, Toh, H, 2008. Improved accuracy of multiple ncRNA alignment by incorporating structural information into a MAFFT-based framework. *BMC Bioinformatics*. 9: 212.
- Kikuchi Y, Sameshima S, Kitade O, Kojima J, Fukatsu T, 2002. Novel clade of *Rickettsia* spp. from leeches. *Appl Environ Microbiol*. 68: 999-1004.
- Kovach WL, 1999. M.V.S.P. A Multi-Variate Statistical Package for Windows, Services, K.C., ed. (Wales, Penarth).
- Kumar S, Skjaeveland A, Orr RJ, Enger P, Ruden T, Mevik BH, Burki F, Botnen A, Shalchian-Tabrizi K, 2009. AIR: A batch-oriented web program package for construction of supermatrices ready for phylogenomic analyses. *BMC Bioinformatics*. 10: 357.
- Lanave C, Preparata G, Saccone C, Serio G, 1984. A new method for calculating evolutionary substitution rates. *J Mol Evol*. 20: 86-93.
- Lloyd SJ, LaPatra SE, Snekvik KR, St-Hilaire S, Cain KD, Call DR, 2008. Strawberry disease lesions in rainbow trout from southern Idaho are associated with DNA from a *Rickettsia*-like organism. *Dis Aquat Organ*. 82: 111-118.
- Mariconti M, Epis S, Gaibani P, Dalla Valle C, Sasseria D, Tomao P, Fabbi M, Castelli F, Marone P, Sambri V, Bazzocchi C, Bandi C, 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? *Pathogens and Global Health* 0: 1-6.
- Matsuura Y, Kikuchi Y, Meng XY, Koga R, Fukatsu T, 2012. Novel clade of alphaproteobacterial endosymbionts associated with stinkbugs and other arthropods. *Appl Environ Microbiol*. 78: 4149-4156.
- Mediannikov O, Ivanov LI, Nishikawa M, Saito R, Sidel'nikov Iu N, Zdanovskaia NI, Mokretsova EV, Tarasevich IV, Suzuki H, 2004. Microorganism "Montezuma" of the order Rickettsiales: the potential causative agent of tick-borne disease in the Far East of Russia. *Zh Mikrobiol Epidemiol Immunobiol*. 1: 7-13.

- Montagna M, Sasser D, Griggio F, Epis S, Bandi C, Gissi C, 2012. Tick-box for 3'-end formation of mitochondrial transcripts in ixodida, Basal chelicerates and *Drosophila*. *PLoS One*. 7: e47538.
- Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, Kamoun S, Sumlin WD, Vogler AP, 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst Biol*. 55: 595-609.
- Posada D, 2008. jModelTest: phylogenetic model averaging. *Mol Biol Evol*. 25: 1253-1256.
- Rizzoli A, Hauffe H, Carpi G, Vourc HG, Neteler M, Rosa R, 2011. Lyme borreliosis in Europe. *Euro Surveill*. 16: 19906.
- Rodriguez-Ezpeleta N, Embley TM, 2012. The SAR11 group of alpha-proteobacteria is not related to the origin of mitochondria. *PLoS One*. 7: e30520.
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna 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.
- Sanderson MJ, 2003. r8s: inferring absolute rates of molecular evolution, divergence times in the absence of a molecular clock. *Bioinformatics* 19: 301-302.
- Sasser D, Beninati T, Bandi C, Bouman EA, Sacchi L, Fabbi M, Lo N, 2006. 'Candidatus *Midichloria mitochondrii*', an endosymbiont of the tick *Ixodes ricinus* with a unique intramitochondrial lifestyle. *Int J Syst Evol Microbiol*. 56: 2535-2540.
- Sasser D, Lo N, Epis S, D'Auria G, Montagna M, Comandatore F, Horner D, Pereto J, Luciano AM, Franciosi F, Ferri E, Crotti E, Bazzocchi C, Daffonchio D, Sacchi L, Moya A, Latorre A, Bandi C, 2011. Phylogenomic evidence for the presence of a flagellum and *cbb(3)* oxidase in the free-living mitochondrial ancestor. *Mol Biol Evol*. 28: 3285-3296.
- Sneath PHA, Sokal RR, 1973. *Numerical Taxonomy, the principles and practice of numerical classification*. Freeman W.H. and Company, San Francisco, 573 p.
- Sunagawa S, Woodley CM, Medina M, 2010. Threatened Corals Provide Underexplored Microbial Habitats. *PLoS ONE* 5: e9554.
- Tamura K, Nei M, 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol*. 10: 512-526.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S, 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 28, 2731-2739. Epub 2011 May 2734.
- Thrash JC, Boyd A, Huggett MJ, Grote J, Carini P, Yoder RJ, Robbertse B, Spatafora JW, Rappe MS, Giovannoni SJ, 2011. Phylogenomic evidence for a common ancestor of mitochondria and the SAR11 clade. *Sci Rep*. 1: 13.

- Vannini C, Ferrantini F, Schleifer KH, 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 Environ Microbiol.* 76, 4047-4054. Epub 2010 Apr 4030.
- 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.
- Viklund J, Ettema TJ, Andersson SG, 2012. Independent genome reduction and phylogenetic reclassification of the oceanic SAR11 clade. *Mol Biol Evol.* 29: 599-615.
- Weinert LA, Werren JH, Aebi A, Stone GN, Jiggins FM, 2009. Evolution and diversity of *Rickettsia* bacteria. *BMC Biol.* 7: 6.
- Weir BS, 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358-1370.
- Wier BS, 1996. *Genetic Data Analysis II: Methods for Discrete Population Genetic Data* (Sunderland, Sinauer Assoc).
- Williams KP, Sobral BW, Dickerman AW, 2007. A robust species tree for the alphaproteobacteria. *J Bacteriol.* 189: 4578-4586.

2.5 Research article*. *Tick-Box for 3'-end formation of mitochondrial transcripts in Ixodida, basal chelicerates and Drosophila*

*NOTE: considering the length and the complexity of this research, the references are numbered throughout the manuscript and citations will be numbered in the order in which they appear.

2.5.1 Summary

According to the tRNA punctuation model, the mitochondrial genome (mtDNA) of mammals and arthropods is transcribed as large polycistronic precursors that are matured by endonucleolytic cleavage at tRNA borders and RNA polyadenylation. Starting from the new mtDNA of *Ixodes ricinus* and using a combination of mitogenomics and transcriptional analyses, we found that in all tick lineages (Prostriata, Metastriata and Argasidae) the 3'-end of the polyadenylated *nad1* and *rrnL* transcripts does not follow the tRNA punctuation model and is located upstream of a degenerate 17-bp DNA motif. A slightly different motif is also present downstream the 3'-end of *nad1* transcripts in the primitive chelicerate *Limulus polyphemus* and in *Drosophila* species, indicating the ancient origin and the evolutionary conservation of this motif in arthropods. The transcriptional data suggest that this motif directs the 3'-end formation of the *nad1/rrnL* mature RNAs, likely working as a transcription termination signal or a processing signal of the precursor transcripts. Although this signal is not exclusive of ticks, making a play on word it has been named "Tick-Box", since it is a check mark that has to be verified for the 3'-end formation of some mt transcripts and its consensus sequence has been here best defined in ticks. Finally, as most regulatory elements, Tick-Box is characterized by a taxon-specific evolution and in some taxa it is also located in additional mtDNA positions. Indeed, in the whole tick mtDNA the Tick-Box is always present downstream of *nad1* and *rrnL*, mainly in noncoding regions (NCRs) and occasionally within *trnL(CUN)*. However, some metastrates present a third Tick-Box at an intriguing site - inside the small NCR located at one end of a 3.4 kb translocated region, the other end of which exhibits the *nad1* Tick-Box - hinting that this motif could have been involved in metastriate gene order rearrangements.

2.5.2 Manuscript

Introduction

Chelicerates constitute a major lineage within Arthropoda and encompass both taxa of evolutionary interest, such as the deep-branching lineage Xiphosura including the living fossil *Limulus*

polyphemus, and species of medical relevance, such as the Arachnida (e.g. ticks, mites, scorpions, spiders). Mitogenomic studies of these organisms are thus conducted both to elucidate their evolutionary biology and to derive mitochondrial sequences for use in species identifications. Ticks (Ixodida) are obligate blood-sucking ectoparasites that originated in early/middle Permian (300-260 million years ago, Mya) [1,2,3] and now parasitize a variety of terrestrial vertebrates [4,5]. The approximately 870 described species of ticks are subdivided into three families: Argasidae, Ixodidae and Nuttalliellidae [6]. Ixodidae (hard ticks) can be divided in two morphological groups: Prostriata, including only the genus *Ixodes*, and Metastrata, including the remaining 11 genera [7]. Ticks can transmit a variety of pathogenic agents to humans and animals [4]. In particular, the sheep tick *Ixodes ricinus* (Linnaeus 1758), the most common blood-feeding ectoparasite in Europe, is the vector of Lyme disease and other bacteria and viruses [8]. *I. ricinus* is also of particular interest in that it harbours a symbiont, “*Candidatus* Midichloria mitochondrii” [9], that resides in the intermembrane space of mitochondria. It can thus be considered a model of a three-levels relationship: the vertebrate host, the tick ectoparasite, and the intra-mitochondrial bacterium “*Candidatus* M. mitochondrii”.

Currently the complete mitochondrial genome (mtDNA) has been sequenced in 55 chelicerate species, including the living fossil *L. polyphemus*, whose gene order is considered to be ancestral for all arthropods [10,11]. Chelicerate mitogenomes show several distinctive features compared to other arthropods: bizarre tRNA structures [12,13,14,15]; unusual rRNAs [16,17]; fast nucleotide substitution rate [18]; and extensive gene order rearrangements even between closely related species [17,19,20,21,22,23,24]. Indeed, among the 55 complete mtDNAs of chelicerates, the primitive gene order of *L. polyphemus* is shared only by two whip spiders from the order Amblypygi (*Phrynus* sp. and *Damon diadema*), the mesothel spider *Heptathela hangzhouensis*, the scorpion *Uroctonus mordax*, and several tick species (suborder Ixodida) (see <http://www.caspur.it/mitozoa>).

Given the general interest in the Ixodida, we sequenced the complete mtDNA of the sheep tick *I. ricinus*. The comparison to other tick mtDNAs highlighted several oddities in the *nad1* and *rrnL* genes that prompted us to investigate the transcription of these genes in all major tick lineages (Prostriata, Metastrata and Argasidae). Therefore, we carried out 3' RACE experiments in *I. ricinus*, and mapped the exact 3'-end of these transcripts in several other ticks, using thousands of available tick EST sequences and according to the strategy described in Gissi et al. [25].

In this paper, after a brief summary of the main features of the *I. ricinus* mtDNA, we describe the identification of a degenerate 17 bp sequence motif directing the 3'-end formation of *nad1* and *rrnL* transcripts in all major tick lineages. This motif represents an exception to the tRNA punctuation model, which predicts that arthropod mtDNA is transcribed in large polycistronic RNA

precursors matured through endonucleolytic cleavages and polyadenylation at sites immediately adjacent to tRNA genes [26,27,28]. We also demonstrate the presence of a similar sequence motif, playing a similar function, only downstream *nadI* in the basal chelicerate *L. polyphemus* and in the model hexapod *Drosophila melanogaster*. Finally, we illustrate a possible evolutionary scenario of this motif from chelicerates to hexapods. Making a play on word, we have named this motif “Tick-Box”, since it is a “check mark” that has to be verified for the 3’-end formation of *nadI* and sometimes also *rrnL* transcripts, and its consensus sequence has been best characterized here, for the first time, for the “tick” group.

Materials and Methods

- I. ricinus mtDNA annotation and analyses -

The amplification and sequencing of the complete mtDNA of *I. ricinus* is described in Supporting File S1. The mt sequence was deposited at EMBL database under accession number JN248424.

Protein coding genes (PCG) of *I. ricinus* were annotated by sequence similarity to the orthologous PCGs of other ticks. Partial stop codons were assumed only to avoid overlap with a downstream gene located on the same strand, while the 3’-end of *nadI* was experimentally identified by 3’ RACE and EST analyses (see below). Overlaps between genes located on the same strand were kept as short as possible. tRNA annotation was performed comparing the predictions of tRNAscan-SE [29] and ARWEN [30] to the tRNAs annotated in other ticks (LocARNA multi-alignment [31]). Small (*rrnS*) and large (*rrnL*) ribosomal subunit rRNAs were identified by sequence similarity and their boundaries were settled as adjacent to those of the flanking genes. As an exception, the 3’-end of *rrnL* was experimentally determined by 3’ RACE and EST analyses (see below).

For the *I. ricinus* mtDNA analyses, the gene boundaries of the 10 previously published mtDNAs of Ixodida (Table 1) were revised based on sequence multi-alignment, transcriptional data, and the criterion of “minimum gene overlap”. Using this approach, we optimized the annotation of a total of 93 genes in 10 species, that is 60 tRNAs and 33 PCGs, with up to 14 gene boundaries modified in *I. hexagonus* (data available on request).

Secondary structures of the major non-coding region, the control region (CR), were predicted with Mfold [32].

Exact direct repeats longer than 9 bp were searched in the mtDNA sequences with RepFind [33], setting the P-value cut-off at 0.01 and with no filter for low-complexity sequences.

The Tick-Box motif was defined and searched in complete and partial mt sequences using

PatSearch [34,35]. The Tick-Box consensus sequence (ttgyrtchwwwtwgda; see Figures 3-6) was defined as the sequence with the highest sensitivity in Patsearch analyses within all analysed Ixodida species. Tick-Box searches in the whole mtDNA sequences of the 45 non-Ixodida chelicerates and 14 *Drosophila* species were carried out allowing mismatches and/or indels to the original consensus sequence. Tick-Box sequence logos [36] were generated by WebLogo [37] using all occurrences of the Tick-Box in the analysed species (Table S1). The possible presence of conserved secondary structure around the Tick-Box was verified by LocARNA [31].

Gene order, non-coding regions, and gene sequences of all mtDNAs analysed in this study were retrieved from MitoZoa Rel. 9.1 [38,39] (<http://www.caspur.it/mitozoa>), a database collecting one representative and manually-curated mtDNA entry for each metazoan species. Therefore, the 474 complete *nad1* sequences of arthropods were retrieved from MitoZoa Rel. 9.1

Table 1. Completely sequenced mitochondrial genomes of Ixodida.

Family	Group	Subfamily	Species	mtDNA
Ixodidae	Prostriata	Ixodinae	<i>Ixodes ricinus</i>	This study
"	"	"	<i>Ixodes hexagonus</i>	NC_002010
"	"	"	<i>Ixodes persulcatus</i>	NC_004370
"	Australasian Prostriata	"	<i>Ixodes holocyclus</i>	NC_005293
"	"	"	<i>Ixodes uriae</i>	NC_006078
"	Metastriata	Amblyomminae	<i>Amblyomma triguttatum</i>	NC_005963
"	"	Haemaphysalinae	<i>Haemaphysalis flava</i>	NC_005292
"	"	Rhipicephalinae	<i>Rhipicephalus sanguineus</i>	NC_002074
Argasidae		Ornithodorinae	<i>Carios capensis</i>	NC_005291
"		"	<i>Ornithodoros moubata</i>	NC_004357
"		"	<i>Ornithodoros porcinus</i>	NC_005820

doi:10.1371/journal.pone.0047538.t001

- 3' RACE of *rrnL* and *nad1* genes -

Since both *rrnL* and *nad1* transcripts are polyadenylated in *Drosophila melanogaster* [26,27,40], the 3'-end of these transcripts was identified by 3' RACE (Random Amplification of cDNA End) or by identification of polyA tail start site in mitochondrial ESTs, according to the method used in [25]. The 3' RACE of *nad1* and *rrnL* transcripts of *I. ricinus* was carried out using gene-specific inner (*nad1*-620pr and *rrnL*-1050pr) and outer (*nad1*-173pr and *rrnL*-850pr) primers (see Supporting File S1).

One partially engorged adult female of *I. ricinus* was collected in Monte Bollettone (Como, Italy) and the total RNA was extracted following the total RNA isolation procedure of the mirVana™ miRNA Isolation Kit (Ambion). RNA was retrotranscribed to cDNA using an adaptor-ligated oligo (dT)-primer (FirstChoice RML-RACE Kit, Invitrogen) and the reverse transcriptase of the QuantiTect Reverse Transcription Kit (Qiagen). The first PCR reaction was assembled coupling

the 3' RACE outer adaptor primer (FirstChoice RML-RACE Kit, Invitrogen) with the nad1-173pr or rrnL-850pr primer. The second nested PCR reaction was assembled coupling the 3' RACE inner adaptor primer (FirstChoice RML-RACE Kit, Invitrogen) with the nad1-620pr or rrnL-1050pr primers. All PCR reactions were performed in a total volume of 25 µl with 1.25 units of GoTaq (Promega), according to the manufacturer's protocol. A single band of approximately the expected size was observed as product of the inner and outer PCRs in both the *nad1* and *rrnL* 3' RACE. In order to identify possible alternative polyadenylation sites located few nucleotides apart, nested PCR products were cloned (CloneJET PCR Cloning Kit, Fermentas) and a total of six positive clones were sequenced for each fragment. The partial RNA sequences of *I. ricinus rrnL* and *nad1* were deposited at EMBL database under accession numbers HE798553, HE798554 and HE798555.

- EST analyses of *rrnL* and *nad1* genes -

EST sequences highly similar to *rrnL* and *nad1* of a given tick species were identified by Blast search [41] using as a probe the mt gene sequence of the same or of a congeneric species. Blast searches were carried out against the "Est_other" database that, at February 2012, included 297,856 ESTs of 20 Ixodida species. ESTs with statistically significant matches were assembled together with the corresponding mitogenomic sequence using Geneious [42]. The polyA start site was identified by visual inspection of the assembly. In particular, "A" or "T" stretches > 10 bp located at the end of EST sequences were considered equivalent to the polyA tail of a mature transcript. In some cases, the lack of EST quality data and/or the presence of A stretches on the genomic DNA allowed mapping this site in a range of 2-5 nucleotides, rather than with single-nucleotide resolution. The *rrnL* polyA site of *Boophilus microplus* and *Dermacentor andersoni*, and the *nad1* polyA site of *L. polyphemus* were determined by analysis of the original untrimmed ESTs, kindly provided by the authors.

Phylogenetic analyses

Phylogenetic analyses were performed on the 13 PCGs of the 10 complete mtDNA of Ixodida (Table 1), using Argasidae as outgroup species. PCGs were aligned at the amino acid level with Muscle [43], and the equivalent nucleotide alignments were generated by "back-translation". Ambiguous alignment regions were trimmed with Gblocks [44] using default parameters. The single PCG alignments were then concatenated with SEAVIEW [45].

Bayesian phylogenetic analyses were carried out on both amino acid and nucleotide alignments. The evolutionary models best fitting to the analyzed datasets were selected with ProtTest 1.4 [46] for amino acid, and ModelTest [47] for nucleotide datasets, according to the Akaike Information Criterion (AIC). The selected substitution model was the MtArt [48] with a proportion of invariant sites (I) and a gamma distribution for rate heterogeneity across sites (Γ) for

the amino acid dataset, and the GTR+I+ Γ for the nucleotide dataset [49]. Bayesian trees were calculated using MrBayes 3.1.2 [50]. Due to the absence of MtArt, the more general GTR and MtRev [51] model were applied in amino acid analyses. Two different partitions based on the 13 genes and on the 3 codon positions, were used in the nucleotide dataset analysis. One partition based on the 13 different proteins, was used for the amino acid dataset. Two parallel analyses, each composed of one cold and three incrementally heated chains, were run for 2.5 million generations. Trees were sampled every 100 generations and burn-in fraction was calculated as 25% of total sampled trees, according to the InL stationary analyses.

Results and Discussion

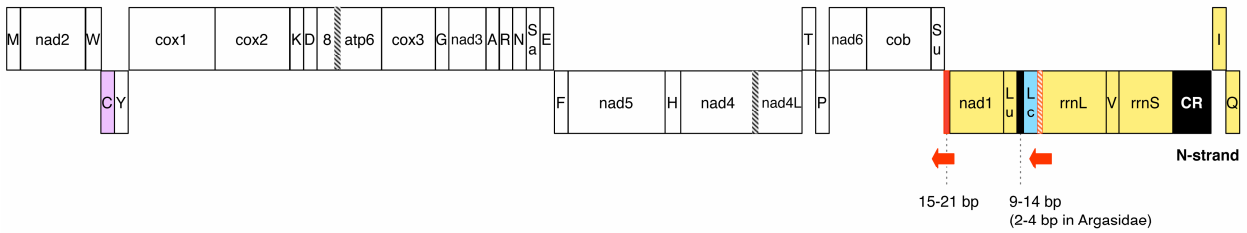
- Ixodes ricinus genome organization and phylogeny -

The mtDNA of *I. ricinus* is 14,566 bp long and encodes the 37 mt genes typical of other metazoans. The general features of this genome, together with peculiarities of the protein-coding genes (PCGs), the tRNA genes, the control region, and the small non-coding regions (NCRs) are illustrated in the Supporting File S1, Supplementary Figure S1 and Supplementary Figure S2.

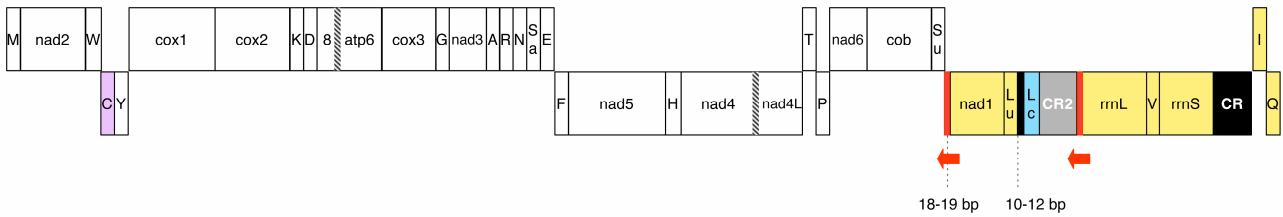
Figure 1 compares the genome organization of all available complete mtDNAs of ticks, taking also into account the location of the control region (CR), which contains the regulatory elements of mt transcription and replication. The genome organization of *I. ricinus* is identical to that found in all other available non-Australasian *Ixodes* species and in Argasidae (Figure 1). Since it is also shared with *L. polyphemus*, this organization is considered to be ancestral to all arthropods [10,11,52]. Australasian *Ixodes* species (*I. uriae* and *I. holocyclus*) have a genome organization very similar to that of other *Ixodes* and Argasidae, but possess a duplicate control region (CR2) between *trnL(CUN)* and *rrnL* (Figure 1), suggesting possible differences in mtDNA replication/transcription mechanisms [23]. With respect to the ancestral genome organization, the Metastriata exhibit: (1) translocation of a large genomic block comprising 7 genes and the CR (yellow block in Figure 1); (2) translocation plus inversion of *trnC*; (3) presence of a duplicate CR2 between *trnL(CUN)* and *trnC* (grey blocks in Figure 1) [24,53]. It is noteworthy that, in both Metastriata and Australasian *Ixodes*, the duplicate CR2s exhibit concerted evolution and probably originated, together with the observed genome rearrangements, through two distinct events of tandem duplication and random gene losses [23,53].

***Ixodes ricinus*, *I. hexagonus*, *I. persulcatus* (non-Australasian Prostriata) - Argasidae**

J-strand



Australasian *Ixodes* (Prostriata)



Metastrata

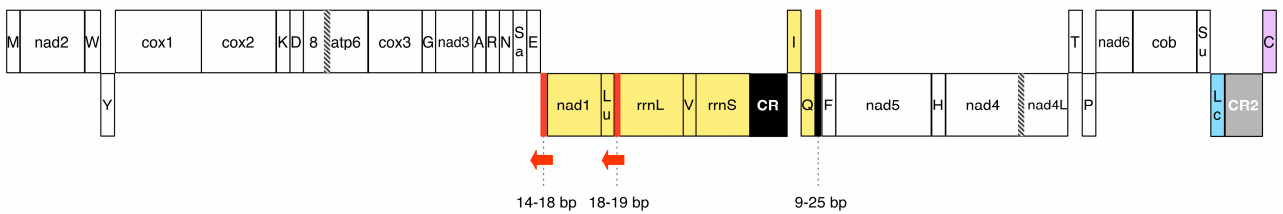


Figure 1. Mitochondrial gene arrangement of *Ixodes ricinus* and 10 other Ixodida species. Translocated genes are reported in the same colour. Black block: non-coding regions ≥ 9 bp in all species of a taxonomic group, with bp range indicated by dashed lines; red block: Tick-Box within a non-coding region; red-hatched block: Tick-Box overlapped to *trnL(CUN)*; red arrow: direction of the Tick-Box; black-hatched block: overlaps between genes; grey block: duplicated control region. Gene abbreviations: 8, atp6: subunits 8 and 6 of the F₀ ATPase; cox1-3: cytochrome c oxidase subunits 1-3; cob: cytochrome b; nad1-6 and nad4L: NADH dehydrogenase subunits 1-6 and 4L; rrnS and rrnL: small and large subunit rRNAs. tRNA genes are indicated by the one-letter code of the transported amino acid, with Lu: *trnL(UUR)*; Lc: *trnL(CUN)*; Sa: *trnS(AGN)*; Su: *trnS(UCN)*. Analysed mtDNAs are listed in Table 1.

Bayesian phylogenetic analyses of Ixodida, carried out on the 13 PCGs at both nucleotide and amino acid level, support the monophyly of the major Ixodida lineages. The phylogenetic tree unambiguously identify *I. ricinus* as sister taxon to *I. persulcatus*, with non-Australasian *Ixodes* positioned in a distinct highly supported clade (Figure 2A). This topology is in agreement with previous phylogenies based on molecular data [2,3,54] or based on both morphological characters and nucleotide sequences (18S and 28S nuclear rRNAs; 16S mt rRNA) [55].

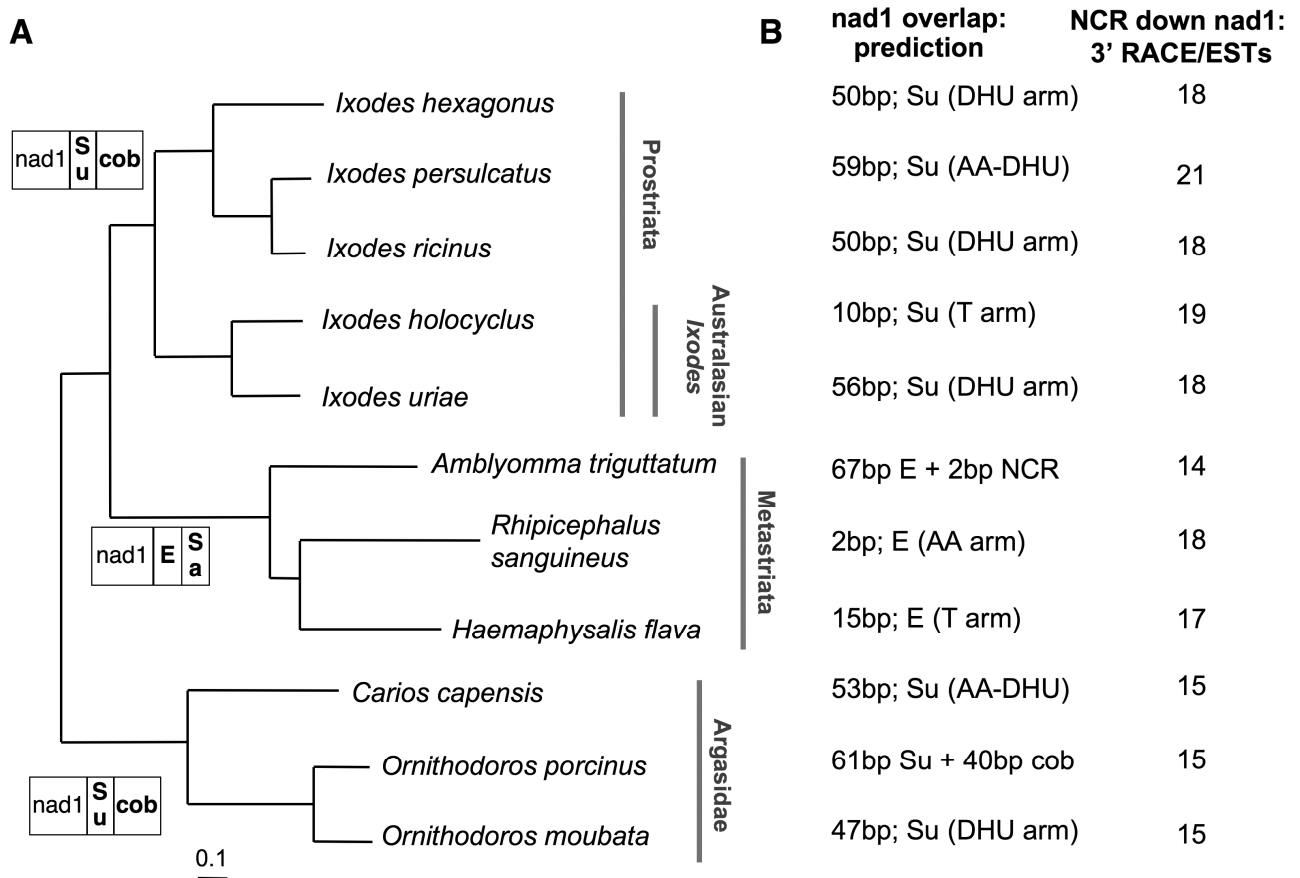


Figure 2. Features of the *nad1* 3'-end and the downstream non-coding region, mapped on the Ixodida phylogeny. (A) Ixodida Bayesian tree calculated on the nucleotide sequence of the 13 mt protein-coding genes, and gene order downstream of *nad1*. Bayesian tree was calculated according to the GTR+I+gamma model, using 13 partitions, and all branches have a posterior probability value equal to 1. (B) Predicted overlap between *nad1* and the downstream gene, and length of the non-coding region experimentally identified downstream of *nad1* by transcriptional data. tRNA regions containing the *nad1* complete stop codon are indicated in brackets. In gene order scheme, the genes encoded by the strand opposite to that of *nad1* are reported in bold. DHU: DHU arm; AA: amino acid acceptor arm; AA-DHU: spacer between the AA and DHU arms. Gene abbreviations as in Figure 1.

- *Partial stop codons and nad1 annotation* -

In the mtDNA, partial stop codons are completed by polyadenylation of mature transcripts that are produced by endonucleolytic cleavages of precursor RNAs at sites immediately adjacent to tRNA genes [26,27,28]. It should be also noted that the usage of a partial stop codon eliminates the overlap between two consecutive genes (a PCG and a tRNA) encoded by the same strand, allowing the production of two full-length transcripts by cleavage of the same polycistronic RNA precursor. Thus, partial stop codons are commonly predicted according to the presence of an abutted tRNA gene and to the rule of “minimum overlap” between genes encoded by the same strand. In *I. ricinus*, the partial stop codons of five PCGs can be predicted according to the above-described rules (“T” in *cox2*, *cox3*, *nad5*, and *cob*; “TA” in *nad2*). On the contrary, the identification of the correct stop codon of *nad1* is quite tricky because the 3’ end of this gene has unique peculiarities that do not fit to the known transcript maturation process and the derived annotation rules. In particular, *nad1* is the only PCG followed by a gene encoded on the opposite strand (Figure 1). Therefore, based on the punctuation model of transcript maturation, the annotation of a partial stop codon is not strictly required in this case, since *nad1* and the downstream gene are transcribed by two different strands. Moreover, the complete stop codon of the *nad1* ORF is surprisingly located well inside the opposite strand-encoded *trnS(UCN)* gene, producing a large gene overlap of 50 bp.

Strikingly, even the 3’-ends of the *nad1* genes/proteins currently annotated in all other ticks present similar unusual features. Firstly, the currently annotated *nad1* end of almost all published ticks [23,24,53] has a complete stop codon located inside the first or even the second downstream, opposite strand-encoded, gene. This annotation gives rise to a gene overlap whose size is highly variable and ranges from 2 to 101 bp (Figure 2B). The most extreme case is in the argasid tick *Ornithodoros porcinus*, where the annotated *nad1* contains the reverse complement of the entire downstream *trnS(UCN)*, and the complete *nad1* stop codon is located inside the following *cob* gene. Similarly, the *nad1* ORF of the metastriate *A. triguttatum* contains the entire *trnE* gene. It is noteworthy that the predicted *nad1* overlap size is not related to species phylogeny or gene order around *nad1* (Figure 2), and that the currently annotated *nad1* complete stop codons fall in different regions of the downstream tRNA gene, depending on the species (Figure 2B).

Secondly, assuming the veracity of these complete stop codons, the *nad1* protein of Ixodida should have an extra C-terminal tail compared to the *nad1* of *D. melanogaster*, ranging from 6 to 38 amino acids (20 amino acids in *I. ricinus*). The analysis of a multi-alignment of 474 *nad1* proteins belonging to different arthropod species (see Materials and Methods) shows that this putative C-terminal tail is Ixodida-specific, being absent in all other available chelicerates (45 species) and in 96% of the whole arthropod dataset. Finally, this putative C-terminal tail has a low amino acid sequence similarity even within Ixodida (data not shown).

All these peculiarities prompted us to experimentally determine the actual *nad1* stop codon of Ixodida by 3' RACE in *I. ricinus*, and by identification of the polyA start sites in *nad1* ESTs in all other tick species for which EST data are available.

In *I. ricinus*, the 3' RACE analysis shows that the *nad1* mRNA ends with a TAA stop codon created by the polyA tail and located exactly at the same position of the complete DNA-encoded stop codon of *D. melanogaster* (Figure 3). *nad1* ESTs confirm this site in *I. ricinus* and in seven additional tick species (*Ixodes scapularis*, three Argasidae and three Metastricata species; see Table 2 and Figure 3). These data unambiguously demonstrate that, in all major Ixodida lineages, the putative C-terminal tail and the gene overlap of *nad1*, predicted *in silico*, result from the misannotation of the actual *nad1* stop codon. Most importantly, the accurate annotation of the *nad1* 3'-end by transcriptional data identifies an unexpected NCR between *nad1* and the downstream tRNA encoded by the opposite J-strand (i.e., *trnS(UCN)* in Argasidae and Prostricata, and *trnE* in Metastricata). This NCR has been identified in all the 11 analysed complete mtDNAs (Figure 2B) and in all the partial mt sequences available for other 17 tick species (Figure 3 and Table S1 in Supporting Information). We can thus conclude that the NCR downstream of *nad1* is a common and ancestral character of the Ixodida mtDNA.

	Partial stop/polyA start site		NCR	Downstream gene	
	nad1				
<i>Ixodes ricinus</i>	AGT ATT T	↓	GTGTCCTTTTTAGAAAAA	trnS(UCN)	
<i>Ixodes hexagonus</i>	ATA TTT T		GTGTCCTTTTTAAGAAAAA		
<i>Ixodes persulcatus</i>	AAT GTT T		GTATCCTTTTTAGAAAAAGA		
<i>Ixodes holocyclus</i>	ATA GTT T		GTATCAATTTTTAGAAAAAA		
<i>Ixodes uriae</i>	CTA ATT T		GTtTCAATTTTTGAATTA		
<i>Carios capensis</i>	TTT AAT T		GTATCAAAATTAGAA		Prostriata, Argasidae
<i>Ornithodoros moubata</i>	TTT GTT T		GTGTCATTTTTAGAA		
<i>Ornithodoros porcinus</i>	TTT GTT T		GTATCATTTTTAGAA		
<i>Amblyomma hebraeum</i>	TTT ATT T		GCATCAATTTTTGTAATT		trnE
<i>Amblyomma vikirri</i>	TTT ATT T		GCATCAATTTTTGTAATT		
<i>Amblyomma triguttatum</i>	TTT TTT T		GCATCAATTTTTGG		
<i>Aponomma concolor</i>	AGT TTT T		GCATCAATTTTTGGACTTCTTAAAGGGCTA		
<i>Aponomma undatum</i>	AGT TTT T		GCATCAATTTTTGGACGTTAT		
<i>Boophilus annulatus</i>	TTT ATT T		GCATCAATTTTTGTAATT		
<i>Boophilus decoloratus</i>	TTT ATT T		GCATCAATTTTTGTAATT		
<i>Boophilus geigy</i>	TTT ATT T		GCATCAATTTTTGTAATT		
<i>Boophilus kohlsi</i>	TTT ATT T		GCAaCATTTTTTTGTAATTAATT		
<i>Boophilus microplus</i> (5 rpt)	TTT ATT T		GCATCAATTTTTGTAATT		
<i>Dermacentor variabilis</i>	TTT ATT T		GCATCATTTTTTTGGAATTATAA		
<i>Rhipicephalus appendiculatus</i>	TTT ATT T		GCATCAATTTTTGTAATTATTA		
<i>Rhipicephalus evertsi</i>	TTT ATT T		GCATCAATTTTTGTAATTAATT		
<i>Rhipicephalus pulchellus</i>	TTT ATT T		GCATCAATTTTTGTAATT		
<i>Rhipicephalus punctatus</i>	TTT ATT T		GCATCAATTTTTGTAATT		
<i>Rhipicephalus sanguineus</i>	TTT ATT T		GCATCAATTTTTGTAATT		
<i>Hyalomma truncatum</i>	TTT ATT T		GCATCAATTTTTGTAATAATT		
<i>Haemaphysalis flava</i>	AGT TTT T		GCATCAATTTTTGGAAT		
<i>Haemaphysalis humerosa</i>	AGT TTT T		GCATCAATTTTTGGAAT		
<i>Haemaphysalis longicornis</i>	AGT TTT T		GCATCAATTTTTGGAAT	Metastricata	
Tick-Box	T T		GyrTChwwwTwwGdA	trnS(UCN)	
<i>Limulus polyphemus</i>	GTT GTT T		GCATaTAATTAAGAATA		
<i>Tachypleus tridentatus</i>	ATT ATa T		GCATaTAATTTAGAAAT		
<i>Drosophila melanogaster</i>	TTA TTA TAG T		GaAttTTTTTTAGTAA		

Figure 3. Non-coding region between *nad1* and *trnS(UCN)/trnE*, and the Tick-Box degenerate consensus sequence. Bold face: species listed in Table 2, for which the 3'-end of the *nad1* transcript was experimentally determined by ESTs or 3' RACE. When the DNA sequence of a given species was unknown, the 3'-end of *nad1* reconstructed by ESTs was mapped on the sequence of a congeneric species, and only the genus name was reported in bold. The *nad1* 3'-end of the argasid *Argas monolakensis* (Table 2) was mapped on the sequence of the argasid *Carios capensis*. Red colour: last DNA-encoded nucleotide preceding the *nad1* polyA tail. Underlined nucleotides: complete stop codons predicted *in silico*; bold lower case nucleotides with grey background: differences to the Tick-Box consensus sequence; rpt: presence of a repeated sequence containing the Tick-Box (see main text). Degenerate nucleotide symbols according to the IUPAC code. Analyses species and sequence accession numbers are listed in Table S1 of Supporting Information. Gene abbreviations as in Figure 1.

Taxon	Species	Available ESTs ^a	<i>nad1</i> ESTs		mtDNA ^c
			tot	polyA ^b	
Prostriata	<i>Ixodes ricinus</i>	1,969	2	2	This study
Prostriata	<i>Ixodes scapularis</i>	193,773	42	3	congeneric
Argasidae	<i>Ornithodoros coriaceus</i>	923	1	1	congeneric
Argasidae	<i>Ornithodoros parkeri</i>	1,563	1	1	congeneric
Argasidae	<i>Argas monolakensis</i>	2,914	7	6	–
Metastricata	<i>Amblyomma rotundatum</i>	1,230	2	1	congeneric
Metastricata	<i>Dermacentor variabilis</i>	2,090	8	7	AY059254s1
Metastricata	<i>Boophilus microplus</i>	52,901	12	4	AF110621
Xiphosura	<i>Limulus polyphemus</i>	8,488	12	7	NC_003057

^aESTs publicly available at 20 Feb, 2012.

^bESTs with a terminal polyA stretch >10 bp, mapping to the end of *nad1*.

Polyadenylated ESTs mapping well inside the *nad1* gene have not been considered, as they are mostly cDNA artefacts originated from the annealing of the oligo-dT primer to an A-rich inner gene region during the cDNA first strand synthesis.

^caccession number of the mt genomic sequence, if available. "Congeneric" means that only the sequence of a congeneric species is available, as reported in Table S1.

doi:10.1371/journal.pone.0047538.t002

Table 2. ESTs matching to the *nad1* gene, and *nad1* ESTs with a polyA stretch corresponding to the polyA tail of the mature transcript.

As shown in Figure 3, this NCR is AT-rich (mean AT%= 76%), ranges from 14 to 30 bp in length, and is characterized by the presence of a degenerate 17 bp motif that includes the two last conserved nucleotides of *nad1*. Moreover, it can be observed that:

- 1) this degenerate motif is associated with the 3'-end of *nad1* even when *nad1* is translocated in Metastricata (Figure 1);
- 2) this motif is located at the boundaries between two large blocks of genes encoded by opposing genomic strands (Figure 1);
- 3) the polyA start site of the *nad1* mRNA does not map at the boundary of the downstream tRNA gene in any analysed tick (Figure 3), thus excluding a *nad1* transcript maturation according to the tRNA punctuation model [27,28];
- 4) this motif is absent in the *nad1* mature transcript, thus its sequence is either un-transcribed or quickly removed from the *nad1* precursor transcript.

All these data suggest that this motif, that we have named the "Tick-Box", directs the 3'-end formation of the polyadenylated *nad1* transcripts in Ixodida, and likely works as a maturation signal for the cleavage of a large precursor mt transcript, or as a transcription termination signal.

We need to stress that this motif has been originally included inside the *nad1* gene, and its identification has been made possible starting from the observations of: i) unusual position of the complete stop codon; ii) unusually large overlap between genes encoded by opposite strand; iii) an extra not-conserved C-terminal tail in a *nad1* protein. Thus, far from being a simple case of *nad1* misannotation, this is an emblematic case that emphasizes how detailed analyses of unusual gene

features can help to identify hidden functional element, and how gene misannotations can hamper the recognition of conserved regulatory elements.

- *The Tick-Box downstream of rrnL* -

Sequences similar to the Tick-Box motif were sought along the entire mt sequences of all 11 ticks (Table 1) using pattern matching software, and were found to be present in only two or three fixed genomic positions (red blocks in Figure 1):

- 1) downstream of *nad1*;
- 2) near the 3'-end of *rrnL*;
- 3) inside a small NCR located between *trnQ* and *trnF* in some Metastriata.

Available partial mt sequences of 41 additional prostriates and metastriates (Table S1 in Supporting Information) contain Tick-Box motifs only in these genomic positions.

The exact location of Tick-Box motif near the 3'-end of *rrnL* depends on the taxa, indeed this Tick-Box falls:

- 1) in the DHU and anticodon arms of *trnL(CUN)* in Argasidae and non-Australasian *Ixodes* lineages (Figures 1 and 6C);
- 2) at the end of CR2 in Australasian *Ixodes* (Figure 1);
- 3) a few bp upstream of the 3'-end of the currently annotated *rrnL* in Metastriata (Figure 1).

In order to study the potential functional role of the *rrnL* associated Tick-Box, we experimentally determined the 3'-end of *rrnL* transcripts through 3' RACE in *I. ricinus*, and by using EST data in 10 other species (Table 3).

In *I. ricinus*, the *rrnL* polyadenylated transcript ends at two alternative sites, separated by 1 bp and located inside *trnL(CUN)*, immediately before the 5'-end of the Tick-Box motif (red sites in Figure 4). Indeed, most *rrnL* 3' RACE clones stop 14-19 bp inside *trnL(CUN)*, while only one clone stops 11-12 bp inside *trnL(CUN)*: the presence of one/multiple "A" nucleotides on the mitogenomic sequence prevents precisely mapping these polyA start sites. Even *rrnL* ESTs of *I. ricinus* confirm these two alternative 3'-ends of *rrnL*. Moreover, these ESTs do not provide support for the existence of *rrnL* transcripts terminating at the 5'-end of *trnL(CUN)*, as predicted by the tRNA punctuation model.

In *I. scapularis*, EST data identify the 3'-end of *rrnL* at two sites corresponding exactly to those found in *I. ricinus* (Table 3 and red sites in Figure 4). In *Ornithodoros* (Argasidae) and in all analysed metastriates, *rrnL* terminates always at the beginning of the Tick-Box. Only in some species an additional *rrnL* 3'-end site can be observed very close to the 5'-end of the nearby *trnL(CUN)* or *trnL(UUR)* gene, as predicted by the tRNA punctuation model (Table 3, and red sites in Figure 4). However, in each analysed species the majority of ESTs support the *rrnL* 3'-end

located at the beginning of Tick-Box motif (Table 3), suggesting that this site could be used more frequently than the other (given the different nature of the original cDNA libraries, definitive quantitative data cannot be inferred. Moreover, in some species the lack of EST quality data and/or of the mitogenomic sequence does not allow mapping of the *rrnL* polyA start site at single-nucleotide resolution). The lack of ESTs for Australasian *Ixodes* precludes validation of the 3'-end of *rrnL* in this lineage. However, based on sequence similarity to other Prostriata and on the lack of a tRNA abutted to *rrnL*, we hypothesize that in Australasian *Ixodes* the 3'-end of *rrnL* occurs immediately before the identified Tick-Box motif (Figure 4).

In conclusion, as for *nadI*, transcriptional data are consistent with an essential functional role for the Tick-Box sequences in the 3'-end formation of polyadenylated *rrnL* transcripts. Indeed, in all analysed species the *rrnL* polyA tail starts immediately before or within the first 5 nt of the Tick-Box motif, independently of the gene/NCR downstream of *rrnL*. All additional *rrnL* polyadenylation sites, observed mainly in Metastricata, conform to the predictions of the tRNA punctuation model (i.e., they fall at the 5'-end of the downstream tRNA gene, considering the ambiguities due to EST quality) and appear infrequently used, as roughly estimated by the number of supporting ESTs (Table 3).

The presence of a Tick-Box near the 3'-end of *rrnL* is intriguing since a transcription termination signal has been functionally identified downstream of *rrnL* in Mammalia: this signal is a tridecamer sequence entirely contained in the *trnL(UUR)* gene [56,57] and functions as a binding site for the mitochondrial transcription termination factor (mTERF) [58,59]. Based only on sequence similarity to this tridecamer sequence, Valverde et al. [60] identified downstream of *rrnL* a "TGGCAGA" heptamer conserved from mammals to insects and protozoans, and hypothesized its function as an "rRNA termination box". However, later functional studies have not validated the Valverde's signal as a binding site to the mTERF homologs of sea urchin and *D. melanogaster* [61,62,63,64]. We need to stress that our Tick-Box does not coincide with the Valverde's rRNA termination box either in sequence or exact genomic position. Moreover, unlike the rRNA termination box, our motif has been defined using both sequence similarity and transcriptional data. Finally, it should be noted that in Argasidae and non-Australasian *Ixodes* the exact location of the *rrnL* Tick-Box generates an overlap between *rrnL* and *trnL(CUN)* (dashed line in Figure 4), recalling the overlap between *rrnL* and *trnL(UUR)* found in mammals because of the presence of the *rrnL* transcription termination signal inside *trnL(UUR)* [65].

As in the case of *nadI*, the determination of the *rrnL* 3'-end by transcriptional data has allowed the discovery of: i) an unexpected NCR downstream of *rrnL* in Metastricata (11-22 bp long); ii) an overlap between *rrnL* and *trnL(CUN)* in Argasidae and non-Australasian *Ixodes* (12-19 bp long; see dashed line in Figure 4); iii) the misannotations of *rrnL* in most ticks (Figure 4).

However, we need to emphasize that the determination of the exact boundaries of *rrnL* only by comparative analyses is complicated by difficulties in the prediction of the rRNA secondary structure and by the low sequence conservation at both ends of this gene.

Taxon	Species	Available ESTs ^a	<i>rrnL</i> ESTs	polyA ESTs ^b	polyA start at:		mtDNA ^c
					5'-end of Tick-Box	Other	
Prostriata	<i>Ixodes ricinus</i>	1,969	23	13	13	0	This study
Prostriata	<i>Ixodes scapularis</i>	193,773	33	8	8	0	AB161439
Argasidae	<i>Ornithodoros coriaceus</i>	923	20	5	4	1	congeneric
Metastriata	<i>Amblyomma americanum</i>	6,480	852	30	30	0	congeneric
Metastriata	<i>Amblyomma rotundatum</i>	1,230	18	14	12	2	congeneric
Metastriata	<i>Amblyomma tuberculatum</i>	387	17	2	2	0	congeneric
Metastriata	<i>Hyalomma anatolicum</i>	736	5	5	2	3	congeneric
Metastriata	<i>Hyalomma marginatum</i>	2,110	27	24	18	6	congeneric
Metastriata	<i>Dermacentor andersoni</i>	1,387	67	25	16	9	congeneric
Metastriata	<i>Boophilus microplus</i>	52,901	296	130	115	15	AF110619
Metastriata	<i>Rhipicephalus sanguineus</i>	2,899	428	57	50	7	NC_002074
Xiphosura	<i>Limulus polyphemus</i>	8,488	2	2	0	2	NC_003057

^aESTs publicly available at 20 Feb, 2012.

^bESTs with a terminal polyA stretch >10 bp, mapping to the end of *rrnL*. Polyadenylated ESTs mapping well inside the *rrnL* gene have not been considered, as they are mostly cDNA artefacts originated from the annealing of the oligo-dT primer to an A-rich inner gene region during the cDNA first strand synthesis.

^caccession number of the mt genomic sequence, if available. "Congeneric" means that only the sequence of a congeneric species is available, as reported in Table S1. doi:10.1371/journal.pone.0047538.t003

Table 3. ESTs matching to the *rrnL* gene, and *rrnL* ESTs with a polyA stretch corresponding to the polyA tail of the mature transcript.

	rrnL	rrnL + trnL(CUN)	trnL(CUN) (non-Australasian Ixodes; Argasidae)	trnL(CUN) (Australasian)	trnL(CUN) (Metastriata)
<i>I_ricinus</i>	AAAGAAATATTTTTTTTAA	ACTAATTTGGCAGAAAAA	TTGTATCAAATTTAGAA	TTTGAAATGGGTTA	CCAATTAGTA
<i>I_acutitarsus</i>	GTTTTATAAAAAAATAA	ACTAATTTGGCAGAAAAA	TTGTATCAAATTTAGAA	TTTGAATATGGGAA	ACCAATTAGTA
<i>I_auritulus</i>	ATTATAAT-AAAAAATAA	ACTAATTTGGCAGAAAAA	TTGTATCAAATTTAGAA	TTTGAATATGGGTT	TCCAATTAGTA
<i>I_hexagonus</i>	TAATATATACATTTTATTA	ACTAATTTGGCAGAAAAA	TTGTGTCAAATTTAGAA	TTTGAGTATGGAT-ACC	CCAATTAGTA
<i>I_loricatus</i>	TTGATAATAGTAAATATTTT	GTTAATTTGGCAGAAAAA	TTGTGTCAAATTTAGAA	TTTGAGTATGGGCA	ACCAATTAGTA
<i>I_persulcatus</i>	TTAATTTTAAAAAGTATTA	ACTAATTTGGCAGAAAAA	TTGTATCAAATTTAGAA	TTTGAATATGGGTTA	CCAATTAGTA
<i>I_pilosus</i>	ATTTTAATTAATTTATGTTT	ACTAATTTGGCAGAAAAA	TTGTATCAAATTTAGAA	TTTGATATGGGTTA	TCCAATTAGTA
<i>I_scapularis</i>	AAAAATATATTTTATATTA	ACTAATTTGGCAGAAAAA	TTGTATCAAATTTAGAA	TTTGAAATGGTTT	ACCAATTAGTG
<i>I_simplex</i>	GTTTAAATTTTATAAATAG	ACTAATTTGGCAGAAAAA	TTGTATCAAATTTAGAA	TTTGAATATGGGATAA	TACCAGTTAGTG
<i>Carios_ca</i>	TTTTGAATATAAGGTTAATTT	TCTAATTTGGCAGATTA	TTGTATCAAATTTAGAA	TTTGAGGATGAATCTA	TTCAAATAGAA
<i>Ornithodoros_mo</i>	TATTAGTTGCAATGTTAGTT	ATTGGCTTGGCAGATTA	TTGTGTCAAATTTAGAA	TTTGAAGATGGTTT	ACCAATTAGTA
<i>Ornithodoros_po</i>	TGGGTATTTGATGTTAATA	GTTGCTTTGGCAGATTA	TTGTATCAAATTTAGAA	TTTGAAGATGGAA	ACCAGGCAATA
<i>I_cordifer</i>	GGGATACATATTTG TT		TTTGTGTCAAATTTAGAA		TTTTTAAACCCCCCAAGGT
<i>I_cornuatus</i>	TCTAAAGAAATTCCTT-TTATT		TTTGTCTTTTTAGAA	AAAAAGATGATAAAAGGGCTTTTTTCCATAAATTAATTT	
<i>I_hirsti</i>	TTAAAATAATATATTA-TTATT		TTTGTGTCTTTAGAA	ATTTTAAAGAGTTTATTTTAAAAAARAGTAAAAAACG	
<i>I_holocyclus</i>	TTTGTAGTAAAGTATATTTTTT		TTTGTCTTAATTTAGAA	AAAAATCAAAGATTTAATTTGAATTTTTTGGAAAAA-CT	
<i>I_myrmecobii</i>	AAAAATATTAATTTATAT		TACTTTGGGTCTATTTAGAA	TTTATT-AAAAAAGTATATTA-TT	
<i>I_trichosuri</i>	TTTTATTTTAAAAAATATTAT		TTTGTGTCTATTTAGAA	AAATTAGTTAA-TATTTTATTTAAAAAGAAATATTA	
<i>I_uriae</i>	AAAAAACAAGTAAAAATAACT		TTTGTCTAAATTTAGAA	repeat	TACCGTTTT
<i>Amblyomma_he</i>	GTTTTATTTATTTATTTTAT		TTGCATCAAATTTTGAATT	ATTAAAGTGGCAGAAA	ATGCAAGGAATTTAA
<i>Amblyomma_tr</i>	TTTTAATTTATTTTACTTTTT		TTGCATCAAATTTTGAATT	ATTAAAGTGACAGAAA	ATGTGAGGAATTTAA
<i>Amblyomma_vi</i>	TAAATATTTGAACACTCTTTT		TTGCATCAAATTTTGAATT	ATTAAAGTGACAGAAA	ATGTGAGGAATTTAA
<i>Aponomma_co</i>	ATTAATTAATCTAACAGTTT		TTGCATCAAATTTTGAATT	ATTAAAGTGCCAGAAA	AATGCGGGAATTTAA
<i>Aponomma_fi</i>	TTTAGTATTATTTATTTTAT		TTGCATCAAATTTTGAATT	ATTAAAGTGCCAGAAA	AAATGCAAGGAATTTAA
<i>Aponomma_un</i>	ATTAATTAATCTAACAGTTT		TTGCATCAAATTTTGAATT	ATTAAAGTGCCAGAAA	AATGCGGGAATTTAA
<i>H_flava</i>	ATACAGAT-ACAACAGTTT		TTGCATCAAATTTTGAATT	ATTAAAGTGACAGAT	ATATGCTGGAATTTAA
<i>H_longicornis</i>	TTTCTGATTT-ACCAGTTT		TTGCATCAAATTTTGAATT	ATTAAAGTGACAGATA	AAATGCGGGAATTTAA
<i>Boophilus_de</i>	ATTAGGATAAACCACTTTTAT		TTGCATCAAATTTTGAATT	ATTAAAGTGACAGATTTAAATGCGGGAATTTAA	
<i>Boophilus_mi</i>	TTTAGTATTTATTTATTTTAT		TTGCATCAAATTTTGAATT	ATTAAAGTGCCAGAAA	AAATGCAAGGAATTTAA
<i>Dermacentor_re</i>	ATTAATAATTAATTTTAT		TTGCATCAAATTTTGAATT	ATTAAAGTGCCAGAAA	AATGCGGGAATTTAA
<i>Dermacentor_va</i>	ATTAATGAAACAACCTTTAT		TTGCATCAAATTTTGAATT	ATTAAAGTGCCAGAA	TTATGCTAGGAATTTAA
<i>Hyalomma_ae</i>	ATTAAGATAATCAACTTTAT		TTGCATCAAATTTTGAATT	ATTAAAGTGCCAGAAA	AAATGCAAGGAATTTAA
<i>Rhipicephalus_nu</i>	TTTTAAATAAATTCGCTTTAT		TTGCATCAAATTTTGAATTAAA	ATTAAAGTGACAGACA	TAATGTAAGGAATTTAA
<i>R_appendiculatus</i>	ATTAGGATAAACCAACTTTAT		TTGCATCAAATTTTGAATT	ATTAAAGTGCCAGACA	T- AAATGCAAGGAATTTAA
<i>R_compositus</i>	ATTAGGATAAACCAACTTTAT		TTGCATCAAATTTTGAATT	ATTAAAGTGCCAGAAATG	CTATGCGGGAATTTAA
<i>R_evertsi</i>	TTTAAATAAACCAACTTTAT		TTGCATCAAATTTTGAATT	ATTAAAGTGCCAGATA	AAATGCGGGAATTTAA
<i>R_maculatus</i>	ATTGGGATAATCGACTTTAT		TTGCATCAAATTTTGAATT	ATTAAAGTGACAGATA	AAATGCGGGAATTTAA
<i>R_pulchellus</i>	ATTAGGATAAATCAACTTTAT		TTGCATCAAATTTTGAATT	ATTAAAGTGACAGATA	AAATGCGGGAATTTAA
<i>R_punctatus</i>	ATTAGGTAACCAACTTTAT		TTGCATCAAATTTTGAATT	ATTAAAGTGACAGAA	TAATGCGGGAATTTAA
<i>R_simus</i>	ATTAGGATAAACCAACTTTAT		TTGCATCAAATTTTGAATT	ATTAAAGTGCCAGACCACAAATGCGGGAATTTAA	
<i>R_sanguineus</i>	ATTAGGATAAACCAACTTTAT		TTGCATCAAATTTTGAATT	ATTAAAGTGCCAGAACCTAAATGCAAGGAATTTAA	
<i>R_turanicus</i>	ATTAGGATAAACCAACTTTAT		TTGCATCAAATTTTGAATT	ATTAAAGTGCCAGAGAA	ATATGCGGGAATTTAA
<i>R_zambeziensis</i>	ATTAGGATAAACCAACTTTAT		TTGCATCAAATTTTGAATT	ATTAAAGTGCCAGAACT	TAATGCGGGAATTTAA
<i>R_zumpti</i>	ATTAGGATAAACCAACTTTAT		TTGCATCAAATTTTGAATT	ATTAAAGTGCCAGACCACAAATGCGGGAATTTAA	
<i>Tick-Box</i>			TTGyrTChwvwTvwGda		
<i>Limulus_p</i>	TTTATTTTATTTTATGTTA	ATTAAATCTGGCAGATAA	GTCTATAGATTTAGAA	TCTAAGGATGGGAATAATAAGTTTCTCGGTAGTG	
<i>Tachypleus_tr</i>	TAAATGTTAATTTAAAAATA	ATTAAATCTGGCAGATAA	GTCTATAGATTTAGAA	TCTAACAATGAAG	TTTATGTTCTCGATTAATA
<i>D_melanogaster</i>	TAAAGAAATATTTAATATAA	ACTATTTTGGCAGATTA	GTGCAATAAATTTAGAA	TTTATATATGGA	TTT-TTATACAAATAGTA

trnL(CUN) (Xiphosura and Drosophila)

Figure 4. The Tick-Box motif located downstream of rrnL.

Bold face: species listed in Table 3, for which the 3'-end of the rrnL transcript was experimentally determined by ESTs or 3' RACE. When the DNA sequence of a given species was unknown, the 3'-end of rrnL reconstructed by ESTs was mapped on the sequence of a congeneric species, and only the genus name was reported in bold. The rrnL 3'-end of Hyalomma anatolicum and marginatum (Table 3) were both mapped on the sequence of Hyalomma aegypticum (Hyalomma_ae). Red colour: last DNA-encoded nucleotide preceding the rrnL polyA tail. Dashed line: overlap between rrnL and trnL(CUN). Genus names were abbreviated only for Ixodes (I), Haemaphysalis (H), Rhipicephalus (R) and Drosophila (D). Underlined nucleotides: tRNA anticodon; bold lower case nucleotide with grey background: differences to the Tick-Box consensus sequence; blue lines and dots: original annotation of the rrnL 3'-end; "repeat": 71 bp-long inverted repeat located in the CR2 and rrnL gene of I. uriae (position 12431-12501 and 12606-12676, respectively, of NC_006078). Degenerate nucleotide symbols according to the IUPAC code. Analyses species and sequence accession numbers are listed in Table S1 of Supporting Information. Gene abbreviations as in Figure 1.

- The Tick-Boxes of Metastriata -

As shown in Figure 5, a third Tick-Box motif is located in the NCR between trnQ and trnF in 9 out of 13 analysed metastriates (complete and partial mtDNAs, see Table S1 in Supporting Information). In the remaining 4 metastriates, the trnQ-trnF NCR is always shorter than 12 bp, and does not contain an even partial Tick-Box sequence.

	trnQ		trnF
Haemaphysalis_fl	CGTGC-CTTACACCAAAGATTA	-aacatc-----ttgaa-----	TATCTTTAAGAA-AATTTCT
Aponomma_co	TGTGC-TTAACACCCAAAGCTA	--taat-----	TGTTTTTANCAGCANAATTT
Amblyomma_am	CGTGCTTTTACACTAAAAGTTA	ttt-----	TGTTTTTAAATAAAAATATTT
Amblyomma_tr	CGTGCTTTTACACTAAAAGTTA	-----atataaaaa-----	TATTTTCAAAAATAATTATTT
Amblyomma_he	TGTG--ATAACACTCAAAGTTA	TTTGCATCATTTTTTTGTAAATTAATA	TGTTTTTAAA-TAATNTTNT
Boophilus_de	TGTG--ATAACACTCAAAGTTA	-TTGCATCATTTTTTTGTAAATTAATA	TGTTTTTAAATA-ATAATTT
Boophilus_mi	TGTG--ATAACACCCAAAGTTA	TTTGCATCATTTTTTTGTAAATTAATA	TGTTTTTAAATA--ATATTT
Rhipicephalus_ap	TGTG--ATAACACCCAAAGTTA	TTTGCATCAATTTTTTGTAAATTAATA	TGTTTTTAACCT--AAATCT
Rhipicephalus_ev	TGTG--ATAACACCCAAAGTTA	TTTGCATCAATTTTTTGTAAATTAATA	TGTTTTTAACCC--TTATTT
Rhipicephalus_pr	TGTG--ATAACACCCAAAGTTA	TTTGCATCAATTTTTTGTAAATTAATA	TGTTTTTAACCT--ATATCT
Rhipicephalus_pu	TGTG--ATAACACCCAAAGTTA	TTTGCATCAATTTTTTGTAAATTAATA	TGTTTTTAACCT--ATATCT
Rhipicephalus_pu	TGTG--ATAACACCCAAAGTTA	TTTGCATCAATTTTTTGTAAATTAATA	TGTTTTTAATTTCT-CTATTT
Rhipicephalus_sa	TGTG--ATAACACCCAAAGTTA	ATTGCATCAATTTTTTGTAAATTAATA	TGTTTTTAACAT--ATATCT
Tick-Box		TTGyrTChwwTwWGDa	

Figure 5. Tick-Box motif located in the NCR between *trnQ* and *trnF* of Metastricata.

Bold lower case with grey background: differences to the Tick-Box consensus sequence. Analysed species and sequence accession numbers are listed in Table S1 of Supporting Information.

This third Tick-Box is characterized by several oddities:

- 1) It is always on the opposite strand compared to the two Tick-Boxes situated downstream of *nad1* and *rrnL* in the same genome (red arrows in Metastricata of Figure 1);
- 2) It is located in a NCR shared only by Metastricata, since the *trnQ-trnF* gene adjacency is specific of the metastricate gene rearrangement. Thus, if present, this third Tick-Box sequence gives rise to an inverted repeat (21 bp-long) that flanks the large translocated mt region of Metastricata ranging from *nad1* to *trnQ* (yellow block in Metastricata of Figure 1). Even more surprisingly, in *B. microplus* [66] this large translocated mt region is preceded by a fivefold tandem repeat (126 bp unit), composed of *trnE*+Tick-Box+3'-end of *nad1*, and is followed by a single inverted copy of the Tick-Box sequence.
- 3) The phylogenetic distribution of this third Tick-Box is quite erratic, since it is absent in Haemaphysalinae, present in Rhipicephalinae, and present/absent even in congeneric species of Amblyomminae (Figure 5, and Table S1 in Supporting Information). Thus, it is difficult to discriminate between ancient or recent origins of this third Tick-Box.

As further peculiarity, the Tick-Box sequences present in individual metastricate genome are almost identical (maximum of 2 nt differences, observed only in one among the 13 analysed species) while the Tick-Boxes present in the same genome of Argasidae and Prostricata differ for 3-6 nucleotides. More interestingly, in the three complete mtDNAs of metastricates, the Tick-Boxes downstream of *nad1* and *rrnL* are located inside a perfect direct repeat of 28-30 bp. On the contrary, perfect direct repeats of the same size are absent in Argasidae and Prostricata. These data suggest that the Metastricata Tick-Box motifs likely undergo to concerted evolution, as the duplicated CR2 of these taxa [24,53]. It should be noted that this observation does not hold for Australasian *Ixodes*,

where the intra-genome Tick-Boxes differ for 4-5 nucleotides and the identified duplicated CR2s also evolve by concerted evolution [23]. Although we have no convincing explanations for this observation, we hypothesize that the strong intra-genome Tick-Box conservation in *Metastricata* is related to the peculiar mt gene arrangement of this taxon.

The functional role of this third Tick-Box is enigmatic, and the absence of EST data for the *trnQ-trnF* region complicates the verification of its function. However, since the sequence of this third Tick-Box is almost identical to that of functional Tick-Boxes identified in the same genome, we suggest that this motif is functional. We could tentatively hypothesize that this third Tick-Box motif plays the role of terminating the transcription of the J-strand, started at the CR, downstream of *trnI*. Indeed, in *metastricates* the movement of *trnI* far away from the cluster of other J-encoded genes makes J-strand transcription after *trnI* pointless (compare the J-strand gene distribution of *Prostricata/Argasidae* to *Metastricata* in Figure 1). Such a role in the rearranged mtDNAs might have represented a selective constraint for the conservation of the third Tick-Box. Finally, the presence of the Tick-Box motif at both ends of the large translocated mt regions of *Metastricata* (yellow blocks in Figure 1) might suggest its involvement in recombination events responsible for genome rearrangements. Indeed, signs of recombination have been found in several chelicerates based on the observation of concerted evolution, gene conversion, and translocation of genes to the opposite strand [17,20,23,67].

- *Origin and evolution of Tick-Box* -

Figure 6A shows the consensus sequence of the Tick-Box motif and few variants, differing only in 1 or 2 positions. Noteworthy, the Tick-Box consensus sequence is quite degenerate, showing nucleotide ambiguity codes in almost half of the 17 sites (Figure 6A).

This relatively high degeneration of the Tick-Box consensus is in accordance with its nature of regulatory element, and can be related to its possible functioning through interactions with one or more nuclear-encoded proteins. Thus, as usual for regulatory elements, the precise sequence of the Tick-Box is quite different from one species to the other, and we expect this element to be subject to a taxon-specific evolution. In this respect, the Tick-Box is very similar to the CR, a mt region also known to evolve in a taxon-specific way [68]. Remarkably, in addition to the control region, the Tick-Box is the only NCR conserved in all *Ixodida* species (Figure 1), while all other NCRs of ticks are unalignable (even those located at the same relative position in different species) and mainly shorter than 9 bp (Figure 1 and Supporting File S1).

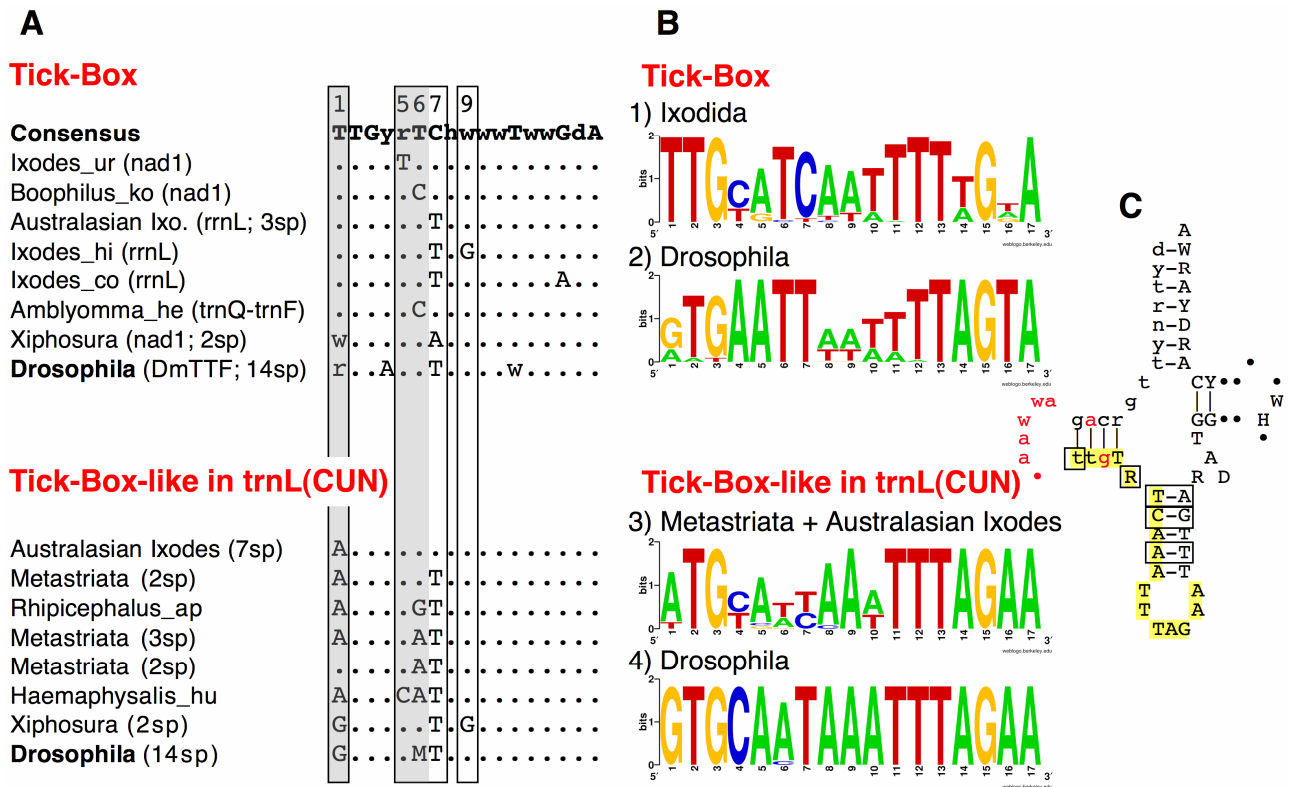


Figure 6. Tick-Box and Tick-Box-like sequences inside *trnL(CUN)* of Chelicerata and *Drosophila* species.

(A) Consensus and variants Tick-Box motifs of Ixodida, Xiphosura, and *Drosophila* species, together with non-functional Tick-Box-like sequences overlapping *trnL(CUN)*. Boxes: positions with nucleotide differences between Tick-Box and Tick-Box-like sequences; grey background: crucial positions discriminating functional Tick-Box from non-functional Tick-Box-like sequences (see main text). In brackets is reported the genomic position of the sequence (i.e., downstream *nad1*, downstream *rrnL*, or between *trnQ-trnF*), and the number of species (sp) showing that sequence. DmTTF: consensus binding sites of DmTTF in *Drosophila*, considering sequences located downstream *nad1* and between *trnE-trnF*. Analysed species and sequences are listed in Table S1 of Supporting Information. (B) Sequence logo for: (1) Tick-Box sequences of Ixodida; (2) DmTTF binding site of 14 *Drosophila* species, including both the sequences downstream of *nad1* and between *trnE-trnF*; (3) Tick-Box-like sequences inside *trnL(CUN)* of Metastriata and Australasian *Ixodes*; (4) Tick-Box-like sequences inside *trnL(CUN)* of 14 *Drosophila* species. Sequence logos were generated as described in Materials and Methods, using sequences listed in Table S1 of Supporting Information. (C) Consensus sequence and secondary structure of the *trnL(CUN)* genes of Argasidae and non-Australasian *Ixodes* containing a functional Tick-Box. Boxes: positions with nucleotide differences between Tick-Box and Tick-Box-like sequences; yellow background: Tick-Box motif; red colour: polyA starts sites determined by 3' RACE or ESTs in Argasidae and non-Australasian *Ixodes*; lower case: overlap region between *rrnL* and *trnL(CUN)*; dot symbol: indels. Degenerate nucleotide symbols according to the IUPAC code. Analysed sequences are listed in Table S1 of Supporting Information.

The analysis of the Tick-Box motifs indicates that the Tick-Box does not form a secondary structure, neither alone nor including surrounding sequences. The only exceptions are the few Tick-Boxes located downstream *rrnL* in non-Australasian *Ixodes* and Argasidae, that are characterized by the overlapping with *trnL(CUN)* (Figure 6C). In these cases, the identified secondary structure has been evolutionary preserved because of the functional constraints of the tRNA gene rather than of the presence of the Tick-Box post-transcriptional regulatory elements (see also below). Therefore, the Tick-Box appears very different from the few hypothesized mt transcript processing sites, where the absence of a tRNA punctuation mark has been supposed to be compensated by stem-loop structures resembling a tRNA portion [28].

In order to define the evolutionary origin of the Tick-Box, we have carefully investigated the presence of the Tick-Box in the basal chelicerate Xiphosura and in *Drosophila*, a highly derived insect genus belonging to the relatively recent Diptera lineage (divergence 228-245 Mya [69,70]). These taxa have been selected due to their peculiar phylogenetic position and also because of the availability of a large amount of ESTs, useful for mt transcripts analyses. Moreover, there are several functional studies on the mt transcription of *D. melanogaster* [26,27,40,64,71], and the complete mtDNA is available for 14 congeneric *Drosophila* species (Table S1 in Supporting Information).

As for Xiphosura, we have considered the horseshoe crabs *L. polyphemus* (for which mtDNA and ESTs are available) and *Tachypleus tridentatus* (for which only the mtDNA is available). Our Xiphosura analyses show that:

- 1) Both species have a Tick-Box sequence (not perfectly matching to the consensus) near the 3'-end of *nad1*. This Tick-Box includes the predicted *nad1* complete stop codon and a short downstream NCR (Figure 3). The ESTs of *L. polyphemus* show that the mRNA of *nad1* terminates immediately upstream of the Tick-Box sequence with a partial stop codon located at the same position of that of Ixodida (Table 2 and Figure 3). Thus, in Xiphosura the existence of a functional Tick-Box motif downstream of *nad1* is supported by both transcriptional and sequence data.
- 2) A divergent Tick-Box sequence (3 mismatches compared to the Ixodida consensus) can be identified near the 3'-end of *rrnL*, exactly inside *trnL(CUN)*, in both horseshoe crabs (Figure 4). However, ESTs of *L. polyphemus* show that the 3'-end of *rrnL* transcript is not located at the beginning of the Tick-Box sequence but just at the 5'-end of *trnL(CUN)*, i.e., at the site predicted by the tRNA punctuation model (Figure 4 and Table 3). In conclusion, in Xiphosura a functional Tick-Box motif is absent downstream of *rrnL*, and the similar sequence identified inside *trnL(CUN)* probably results from functional constraints on *trnL(CUN)*.

Based on these data, we suggest that the Tick-Box downstream of *nad1* is an ancient signal that has been functionally conserved, in spite of the sequence changes, at least over the time

separating Xiphosura from Ixodida (about 400 million years), while the Tick-Box downstream of *rrnL* is a specific invention of Ixodida (Figure 7). We hypothesize that the Tick-Box downstream of *rrnL* has evolved from a portion of *trnL(CUN)*, through acquisition of a new function related to post-transcriptional regulation (Figure 7). After this gain-of-function, the *trnL(CUN)* and the Tick-Box have become overlapped elements and have coevolved in Ixodida for long time, until genome rearrangement events have disrupted the adjacency *rrnL-trnL(CUN)* (two independent events: one in Metastriata and the other in Australasian *Ixodes*). In these rearranged mtDNAs, the sequence including the two overlapped Tick-Box and *trnL(CUN)* elements has been duplicated, and then the two copies have started diverging. In particular, due to the need to regulate the *rrnL* 3'-end formation, the Tick-Box function has been preserved at the position immediately downstream of *rrnL*, where the *trnL(CUN)* function was instead lost. On the contrary, the Tick-Box function has been disrupted in the position actually preserving the *trnL(CUN)* function (Figure 7). Based on the proposed evolutionary scenario, the Tick-Box sequence downstream of *rrnL* in Metastriata and Australasian *Ixodes* should be the only remnant of a duplicated *trnL(CUN)*/Tick-Box sequence that has lost all but the essential *rrnL* post-transcriptional regulatory motif.

As for *Drosophila*, no sequence identical to the Tick-Box consensus motif is present in the whole mtDNA of *D. melanogaster* and congeneric species. However, the *D. melanogaster nadl* gene is followed by a 17 bp-long NCR that is one of the two binding sites of the DmTTF transcription termination factor, the other site being an almost identical sequence located between *trnE* and *trnF* [64,71]. Moreover, the *D. melanogaster nadl* transcript is not 3'-processed at the site predicted by the tRNA punctuation model [27] but it terminates 16 bp upstream of the 5'-end of the downstream *trnS(UCN)* and 1 bp after the stop codon (red colour in Figure 3). It should be also noted that *nadl* ends with a partial stop codon in 6 out of the 13 additional *Drosophila* mtDNAs, and is followed by a NCR ranging from 15 to 25 bp (data for the 14 *Drosophila* species) [72]. Notably, in the 14 *Drosophila* species this NCR has 41-65% identity to the *I. ricinus* NCR downstream of *nadl*, and the consensus of the DmTTF binding site for the 14 *Drosophila* species matches to the Tick-Box degenerate consensus in all but 3-4 positions (Figure 6A; logos n°1 and n°2 in Figure 6B). We conclude that the Tick-Box downstream of *nadl* is present in *Drosophila* but have a sequence quite divergent from the Ixodida consensus (Figure 7). This sequence variability between taxa follows the expected evolutionary pattern for a regulatory element, thus it is likely that the Tick-Box signal downstream of *nadl* is also present in other arthropod lineages with even more divergent sequences. We need also to emphasize that the Tick-Box of *Drosophila* functions as a binding site of DmTTF [64,71].

In *D. melanogaster* the 3'-end of the *rrnL* polyadenylated transcript falls exactly at the site predicted by the tRNA punctuation model [27] (Figure 4), and no DmTTF binding site is present

immediately downstream of *rrnL* [64]. However, we identified a sequence similar to the Tick-Box inside the *trnL(CUN)* gene that is located in all 14 *Drosophila* species just downstream of *rrnL*. This sequence shows 3 mismatches to the Tick-Box consensus sequence (Figure 4, Figure 6A, logo n° 4 of Figure 6B) and 76% identity to the *I. ricinus* Tick-Box inside *trnL(CUN)*. As for horseshoe crabs, we conclude that in *Drosophila* there is no functional Tick-Box downstream of *rrnL* (Figure 7) and that the observed sequence conservation is due to the functional constraints of *trnL(CUN)*. It should be also noted that, for the comparison *D. melanogaster* - *I. ricinus*, the identity percentage is higher in the Tick-Box-like sequences overlapping *trnL(CUN)* than in the functional Tick-Boxes of the NCR downstream *nad1* (76% and 65%, respectively). This indicates that the coevolution of Tick-Box with *trnL(CUN)* and the degenerate nature of this signal can lead to misinterpretation of the Tick-Box presence/absence, especially in taxa phylogenetically distant from Ixodida and especially when only sequence similarity data are taken into account.

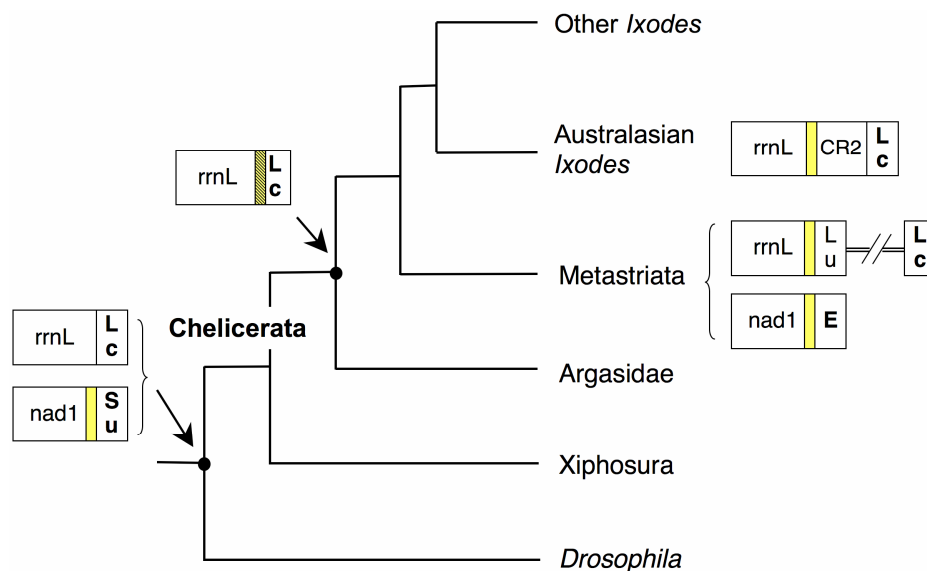


Figure 7. Evolutionary scenario of the Tick-Box motif in Ixodida and other arthropods. Tree topology according to [79]. Yellow block: Tick-Box motif; hatched yellow background: Tick-Box overlapped to *trnL(CUN)*; bold case: genes encoded by the J-strand.

- Coevolution of Tick-Box and *trnL(CUN)* -

To better investigate the coevolution of Tick-Box and *trnL(CUN)*, we have compared the identified Tick-Box motif to the similar sequences found inside the *trnL(CUN)* genes that lack a functional Tick-Box (Figure 6A). This comparison can assist in the identification of nucleotide positions discriminating functional Tick-Boxes from non-functional Tick-Box-like sequences overlapping *trnL(CUN)*. The Tick-Box-like sequences overlapping *trnL(CUN)* were defined as non-functional based on EST data and mismatches to the Tick-Box consensus, and are present in the *trnL(CUN)* of Metastricata, Australasian *Ixodes*, Xiphosura, and all *Drosophila* species (see also

Figures 3-4). The logo of the Tick-Box-like sequences located inside *trnL(CUN)* is shown in Figure 6B, separately for ticks (logo n° 3) and *Drosophila* species (logo n° 4). Moreover, Figure 6C illustrates the consensus of functional Tick-Boxes located inside *trnL(CUN)* in non-Australasian *Ixodes* and Argasidae: as already discussed, these are the only Tick-Box elements showing a conserved secondary structure, since they are superimposed to a functional tRNA gene (see above).

As shown in Figure 6C (yellow background), the Tick-Box sequence inside *trnL(CUN)* is superimposed to half of the DHU and anticodon stems, plus the entire anticodon loop. According to the typical tRNA substitution pattern, the nucleotide substitutions observed in the *trnL(CUN)* with Tick-Box (boxed positions in Figure 6C) are mainly compensatory substitutions falling in the stem regions, while they are carefully avoid the anticodon loop. As reported in Figure 6A, the Tick-Box-like sequences inside *trnL(CUN)* differ in 2-4 positions from the consensus Tick-Box. As an exception, the Tick-Box-like sequence of Australasian *Ixodes* differs from the consensus Tick-Box only for a single substitution (T->A) at position 1. Among positions with differences, we observe that positions 7 and 9 share the same substitution types in both Tick-Box variants and Tick-Box-like sequences (Figure 6A), thus these positions seem not to be crucial for the Tick-Box functionality. On the contrary, positions 5 and 6 (grey background in Figure 6A) have different nucleotides in Tick-Box and Tick-Box-like sequences, indicating that they can be discriminating positions for the Tick-Box functionality. Finally, nucleotide substitutions at the first position of the consensus seem to inactivate the Tick-Box depending on the substitution type, the substitutions co-occurring in other positions, and the taxon (i.e., compare Xiphosura and *Drosophila* Tick-Box and Tick-Box-like sequences in Figure 6A). Thus, from the comparison between functional Tick-Boxes and Tick-Box like sequences, we can conclude that positions 1, 5 and 6 are the most important sites for the functionality of Tick-Box.

Overall, these data further support the hypothesis that Tick-Box is a highly dynamic and degenerate signal whose sequence variability is due to its specific regulatory function (and the possible interaction with regulatory proteins encoded by the nuclear genome) and also to the overlap with coding sequences.

Conclusion

In this study we describe the identification of the Tick-Box, a degenerate 17-bp DNA motif involved in post-transcriptional processes. In particular, Tick-Box directs the 3'-end formation of *nad1* and *rrnL* transcripts in all Ixodida lineages, as well as the 3'-end formation of the single *nad1* transcript in basal chelicerates of the Xiphosura order and in Diptera insects of the *Drosophila* genus. Although this motif is not restricted to tick species, it has been named "Tick-Box" because

its consensus sequence has been here best characterized in Ixodida and because it is a “tick box” necessary for the 3’-end formation of some mt transcripts. In the present study, we have not investigated in details the phylogenetic distribution of this motif in Chelicerata and Arthropoda, however its presence in *Drosophila* and *Limulus* suggest that it could be a quite ubiquitous signal, whose existence has been obscured by its taxon-specific evolutionary pattern and by its nature of post-transcriptional regulatory element. Indeed, as most regulatory elements, Tick-Box is a short and degenerate motif, showing a low sequence similarity between Ixodida and even lower sequence conservation in the more distant species of *Limulus* and *Drosophila*. Therefore, additional studies combining sequence similarity and transcriptional analyses are needed to define the exact consensus sequence and to clarify the phylogenetic distribution of the Tick-Box within the main arthropod groups.

With regard to the exact Tick-Box function, this element is associated to the 3’-end of the *nadI* and *rrnL* genes independently of the downstream gene/NCR. Moreover, it is absent in the mature transcripts. Therefore, we suggest that Tick-Box is either un-transcribed or quickly removed from the primary precursor transcripts of *nadI* and *rrnL*. According to this observation, Tick-Box might be one of the few exceptions to the tRNA punctuation model of mt transcript maturation [28], or a transcription termination signal whose existence was originally hypothesized in *D. melanogaster* by Berthier [26]. Remarkably, the Tick-Box downstream of *nadI* found in *D. melanogaster* has been functionally described some time ago as one of the two binding sites of the DmTTF transcription termination factor [71]. Far from reducing the novelty of this study, the homology between the DmTTF binding site of *D. melanogaster* and the Tick-Box downstream *nadI* of Ixodida supports the functional role of the post-transcriptional signal here identified. Moreover, it testifies the poor link between functional and evolutionary studies on the mtDNA, and the difficulties of mere mt comparative analyses toward the detection of regulatory elements. Indeed, to our knowledge, after its functional characterization, the binding site of DmTTF has not been further investigated at level of taxonomic distribution, consensus sequence or exact genomic location(s) within arthropods.

The discrimination between the two hypothesized Tick-Box functions, precursor transcript maturation or transcription termination, can be experimentally tested in Prostriata and Metastriata by qualitative and quantitative analyses of the whole mt transcriptome and/or experiments aimed at demonstrating the binding of this motif by mt regulatory proteins, such as members of the MTERF protein family [73,74]. The availability of cell lines for both these taxa can also help the analyses [75,76]. We would like to emphasize that the small Tick-Box and the large mt control region are the only non-coding regions conserved in all mtDNAs of ticks. To our knowledge, there is only one other small NCR conserved in all mtDNAs of a large metazoan group, i.e. the vertebrate L-strand

replication origin (oriL) [77,78]. The oriL is a 20-30 bp sequence embedded in a tRNA cluster and forming a stable stem-loop structure partially overlapped to *trnC*. On the contrary the Tick-Box is a degenerate DNA motif that does not show a conserved secondary structure and, like the control region, is characterized by a taxon-specific evolution. Finally, based on the presence of a third Tick-Box in Metastriata and of a second DmTTF binding site in *D. melanogaster*, we could anticipate the presence of Tick-Box in different mitogenomic positions depending on the overall genome organization and on the details of the transcriptional process (i.e., number and type of transcriptional units).

2.5.3 References

1. Dunlop JA, Selden PA, 2009. Calibrating the chelicerate clock: a paleontological reply to Jeyaprakash and Hoy. *Exp Appl Acarol* 48: 183-197.
2. Jeyaprakash A, Hoy MA, 2009. First divergence time estimate of spiders, scorpions, mites and ticks, subphylum: Chelicerata. inferred from mitochondrial phylogeny. *Exp Appl Acarol* 47: 1-18.
3. Mans BJ, de Klerk D, Pienaar R, Latif AA, 2011. *Nuttalliella namaqua*: a living fossil and closest relative to the ancestral tick lineage: implications for the evolution of blood-feeding in ticks. *PLoS One* 6: e23675.
4. Sonenshine DE, 1991. *Biology of ticks*. Vol. 1. New York: Oxford Univ. Press. 447 p.
5. Sonenshine DE, 1993. *Biology of ticks*. Vol. 2 New York: Oxford Univ. Press. 465 p.
6. Horak IG, Camicas JL, Keirans JE, 2002. The Argasidae, Ixodidae and Nuttalliellidae, Acari: Ixodida): a world list of valid tick names. *Exp Appl Acarol* 28: 27-54.
7. Nava S, Guglielmone AA, Mangold AJ, 2009. An overview of systematics and evolution of ticks. *Front Biosci* 14: 2857-2877.
8. Keirans JE, Needham GR, Oliver JH, 1999. The *Ixodes*, *Ixodes. ricinus* complex worldwide: Diagnosis of species in the complex, host and distribution. In: Glen R, Needham, Mitchell R, Horn DJ, Welbourn WC, editors. *Acarology IX*. Columbus, Ohio: The Ohio Biological Survey. pp. 344.
9. Sassera D, Beninati T, Bandi C, Bouman EA, Sacchi L, et al., 2006. Candidatus *Midichloria mitochondrii*, an endosymbiont of the tick *Ixodes ricinus* with a unique intramitochondrial lifestyle. *Int J Syst Evol Microbiol* 56: 2535-2540.
10. Boore JL, Collins TM, Stanton D, Daehler LL, Brown WM, 1995. Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. *Nature* 376: 163-165.

11. Boore JL, Lavrov DV, Brown WM, 1998. Gene translocation links insects and crustaceans. *Nature* 392: 667-668.
12. Klimov PB, Oconnor BM, 2009. Improved tRNA prediction in the American house dust mite reveals widespread occurrence of extremely short minimal tRNAs in acariform mites. *BMC Genomics* 10: 598.
13. Masta SE, 2000. Mitochondrial sequence evolution in spiders: intraspecific variation in tRNAs lacking the TPsiC Arm. *Mol Biol Evol* 17: 1091-1100.
14. Masta SE, Boore JL, 2004. The complete mitochondrial genome sequence of the spider *Habronattus oregonensis* reveals rearranged and extremely truncated tRNAs. *Mol Biol Evol* 21: 893-902.
15. Masta SE, Boore JL, 2008. Parallel evolution of truncated transfer RNA genes in arachnid mitochondrial genomes. *Mol Biol Evol* 25: 949-959.
16. Klimov PB, Knowles LL, 2011. Repeated parallel evolution of minimal rRNAs revealed from detailed comparative analysis. *J Hered* 102: 283-293.
17. Masta SE, 2010. Mitochondrial rRNA secondary structures and genome arrangements distinguish chelicerates: comparisons with a harvestman, Arachnida: Opiliones: *Phalangium opilio*). *Gene* 449: 9-21.
18. Park SJ, Lee YS, Hwang UW, 2007. The complete mitochondrial genome of the sea spider *Achelia bituberculata*, Pycnogonida, Ammotheidae): arthropod ground pattern of gene arrangement. *BMC Genomics* 8: 343.
19. Gissi C, Iannelli F, Pesole G, 2008. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity* 101: 301-320.
20. Shao R, Barker SC, Mitani H, Takahashi M, Fukunaga M, 2006. Molecular mechanisms for the variation of mitochondrial gene content and gene arrangement among chigger mites of the genus *Leptotrombidium*, Acari: Acariformes). *J Mol Evol* 63: 251-261.
21. Jones M, Gantenbein B, Fet V, Blaxter M, 2007. The effect of model choice on phylogenetic inference using mitochondrial sequence data: lessons from the scorpions. *Mol Phylogenet Evol* 43: 583-595.
22. Choi EH, Park SJ, Jang KH, Hwang W, 2007. Complete mitochondrial genome of a Chinese scorpion *Mesobuthus martensii*, Chelicerata, Scorpiones, Buthidae). *DNA Seq* 18: 461-473.
23. Shao R, Barker SC, Mitani H, Aoki Y, Fukunaga M, 2005. Evolution of duplicate control regions in the mitochondrial genomes of metazoa: a case study with Australasian *Ixodes* ticks. *Mol Biol Evol* 22: 620-629.
24. Black WC, Roehrdanz RL, 1998. Mitochondrial gene order is not conserved in arthropods: prostriate and metastriate tick mitochondrial genomes. *Mol Biol Evol* 15: 1772-1785.

25. Gissi C, Pesole G, 2003. Transcript mapping and genome annotation of ascidian mtDNA using EST data. *Genome Res* 13: 2203-2212.
26. Berthier F, Renaud M, Alziari S, Durand R, 1986. RNA mapping on *Drosophila* mitochondrial DNA: precursors and template strands. *Nucleic Acids Res* 14: 4519-4533.
27. Stewart JB, Beckenbach AT, 2009. Characterization of mature mitochondrial transcripts in *Drosophila*, and the implications for the tRNA punctuation model in arthropods. *Gene* 445: 49-57.
28. Ojala D, Montoya J, Attardi G, 1981. tRNA punctuation model of RNA processing in human mitochondria. *Nature* 290: 470-474.
29. Schattner P, Brooks AN, Lowe TM, 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res* 33: W686-689.
30. Laslett D, Canback B, 2008. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* 24: 172-175.
31. Smith C, Heyne S, Richter AS, Will S, Backofen R, 2010. Freiburg RNA Tools: a web server integrating INTARNA, EXPARNA and LOCARNA. *Nucleic Acids Res* 38: W373-377.
32. Zuker M, 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31: 3406-3415.
33. Betley JN, Frith MC, Graber JH, Choo S, Deshler JO, 2002. A ubiquitous and conserved signal for RNA localization in chordates. *Curr Biol* 12: 1756-1761.
34. Grillo G, Licciulli F, Liuni S, Sbisa E, Pesole G, 2003. PatSearch: A program for the detection of patterns and structural motifs in nucleotide sequences. *Nucleic Acids Res* 31: 3608-3612.
35. Pesole G, Liuni S, D'Souza M, 2000. PatSearch: a pattern matcher software that finds functional elements in nucleotide and protein sequences and assesses their statistical significance. *Bioinformatics* 16: 439-450.
36. Schneider T, Stephens R, 1990. Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res* 18: 6097-6100.
37. Crooks G, Hon G, Chandonia J, Brenner S, 2004. WebLogo: a sequence logo generator. *Genome Res* 14: 1188-1190.
38. D'Onorio de Meo P, D'Antonio M, Griggio F, Lupi R, Borsani M, et al., 2012. MitoZoa 2.0: a database resource and search tools for comparative and evolutionary analyses of mitochondrial genomes in Metazoa. *Nucleic Acids Res* 40: D1168-1172.
39. Lupi R, D'Onorio De Meo P, Picardi E, D'Antonio M, Paoletti D, et al., 2010. MitoZoa: a curated mitochondrial genome database of metazoans for comparative genomics studies. *Mitochondrion* 10: 192-199.

40. Benkel BF, Duschesnay P, Boer PH, Genest Y, Hickey DA, 1988. Mitochondrial large ribosomal RNA: an abundant polyadenylated sequence in *Drosophila*. *Nucleic Acids Res* 16: 9880.
41. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, 1990. Basic local alignment search tool. *J Mol Biol* 215: 403-410.
42. Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, et al., 2010. Geneious v5.3 <http://www.geneious.com>.
43. Edgar RC, 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5: 113.
44. Castresana J, 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17: 540-552.
45. Galtier N, Gouy M, Gautier C, 1996. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput Appl Biosci* 12: 543-548.
46. Abascal F, Zardoya R, Posada D, 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21: 2104-2105.
47. Posada D, Crandall KA, 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.
48. Abascal F, Posada D, Zardoya R, 2007. MtArt: a new model of amino acid replacement for Arthropoda. *Mol Biol Evol* 24: 1-5.
49. Lanave C, Preparata G, Saccone C, Serio G, 1984. A new method for calculating evolutionary substitution rates. *J Mol Evol* 20: 86-93.
50. Huelsenbeck JP, Ronquist F, 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754-755.
51. Adachi J, Hasegawa M, 1996. Model of amino acid substitution in proteins encoded by mitochondrial DNA. *J Mol Evol* 42: 459-468.
52. Lavrov DV, Boore JL, Brown WM, 2000. The complete mitochondrial DNA sequence of the horseshoe crab *Limulus polyphemus*. *Mol Biol Evol* 17: 813-824.
53. Shao R, Aoki Y, Mitani H, Tabuchi N, Barker SC, et al., 2004. The mitochondrial genomes of soft ticks have an arrangement of genes that has remained unchanged for over 400 million years. *Insect Mol Biol* 13: 219-224.
54. Xu G, Fang QQ, Keirans JE, Durden LA, 2003. Molecular phylogenetic analyses indicate that the *Ixodes ricinus* complex is a paraphyletic group. *J Parasitol* 89: 452-457.
55. Klompen JSH, Black IV WC, Keirans JE, Norris DE, 2000. Systematics and biogeography of hard ticks: a total evidence approach. *Cladistics* 16: 79-102.

56. Christianson TW, Clayton DA, 1986. In vitro transcription of human mitochondrial DNA: accurate termination requires a region of DNA sequence that can function bidirectionally. Proc Natl Acad Sci USA 83: 6277-6281.
57. Christianson TW, Clayton DA, 1988. A tridecamer DNA sequence supports human mitochondrial RNA 3'-end formation in vitro. Mol Cell Biol 8: 4502-4509.
58. Fernandez-Silva P, Martinez-Azorin F, Micol V, Attardi G, 1997. The human mitochondrial transcription termination factor, mTERF. is a multizipper protein but binds to DNA as a monomer, with evidence pointing to intramolecular leucine zipper interactions. Embo J 16: 1066-1079.
59. Kruse B, Narasimhan N, Attardi G, 1989. Termination of transcription in human mitochondria: identification and purification of a DNA binding protein factor that promotes termination. Cell 58: 391-397.
60. Valverde JR, Marco R, Garesse R, 1994. A conserved heptamer motif for ribosomal RNA transcription termination in animal mitochondria. Proc Natl Acad Sci U S A 91: 5368-5371.
61. Fernandez-Silva P, Loguercio Polosa P, Roberti M, Di Ponzio B, Gadaleta MN, et al., 2001. Sea urchin mtDBP is a two-faced transcription termination factor with a biased polarity depending on the RNA polymerase. Nucleic Acids Res 29: 4736-4743.
62. Loguercio Polosa P, Roberti M, Musicco C, Gadaleta MN, Quagliariello E, et al., 1999. Cloning and characterisation of mtDBP, a DNA-binding protein which binds two distinct regions of sea urchin mitochondrial DNA. Nucleic Acids Res 27: 1890-1899.
63. Roberti M, Mustich A, Gadaleta MN, Cantatore P, 1991. Identification of two homologous mitochondrial DNA sequences, which bind strongly and specifically to a mitochondrial protein of *Paracentrotus lividus*. Nucleic Acids Res 19: 6249-6254.
64. Roberti M, Loguercio Polosa P, Bruni F, Musicco C, Gadaleta MN, et al., 2003. DmTTF, a novel mitochondrial transcription termination factor that recognises two sequences of *Drosophila melanogaster* mitochondrial DNA. Nucleic Acids Res 31: 1597-1604.
65. Van Etten RA, Bird JW, Clayton DA, 1983. Identification of the 3'-ends of the two mouse mitochondrial ribosomal RNAs. The 3'-end of 16S ribosomal RNA contains nucleotides encoded by the gene for transfer RNA^{Leu}UUR. J Biol Chem 258: 10104-10110.
66. Campbell NJH, Barker SC, 1999. The novel mitochondrial gene arrangement of the cattle tick, *Boophilus microplus*: fivefold tandem repetition of a coding region. Mol Biol Evol 16: 732-740.
67. Shao R, Mitani H, Barker SC, Takahashi M, Fukunaga M, 2005. Novel mitochondrial gene content and gene arrangement indicate illegitimate inter-mtDNA recombination in the chigger mite, *Leptotrombidium pallidum*. J Mol Evol 60: 764-773.

68. Pesole G, Gissi C, De Chirico A, Saccone C, 1999. Nucleotide substitution rate of mammalian mitochondrial genomes. *J Mol Evol* 48: 427-434.
69. Friedrich M, Tautz D, 1997. Evolution and phylogeny of the Diptera: a molecular phylogenetic analysis using 28S rDNA sequences. *Systematic Biology* 46: 674-698.
70. Krzeminski W, Krzeminska E, 2003. Triassic Diptera: description, revisions, and phylogenetic relations. *Acta Zoologica Cracoviensia* 46Supp: 153-184.
71. Roberti M, Bruni F, Loguercio Polosa P, Gadaleta MN, Cantatore P, 2006. The *Drosophila* termination factor DmTTF regulates in vivo mitochondrial transcription. *Nucleic Acids Res* 34: 2109-2116.
72. Montooth KL, Abt DN, Hofmann JW, Rand DM, 2009. Comparative genomics of *Drosophila* mtDNA: Novel features of conservation and change across functional domains and lineages. *J Mol Evol* 69: 94-114.
73. Roberti M, Loguercio Polosa P, Bruni F, Manzari C, Deceglie S, et al., 2009. The MTERF family proteins: mitochondrial transcription regulators and beyond. *Biochim Biophys Acta* 1787: 303-311.
74. Linder T, Park CB, Asin-Cayuela J, Pellegrini M, Larsson NG, et al., 2005. A family of putative transcription termination factors shared amongst metazoans and plants. *Curr Genet* 48: 265-269.
75. Munderloh UG, Liu Y, Wang M, Chen C, Kurtti TJ, 1994. Establishment, maintenance and description of cell lines from the tick *Ixodes scapularis*. *J Parasitol* 80: 533-543.
76. Najm N-A, Silaghi C, Bell-Sakyi L, Pfister K, Passos L, 2012. Detection of bacteria related to "Candidatus" *Midichloria mitochondrii* in tick cell lines. *Parasitology Research* 110: 437-442.
77. Hixson JE, Wong TW, Clayton DA, 1986. Both the conserved stem-loop and divergent 5'-flanking sequences are required for initiation at the human mitochondrial origin of light-strand DNA replication. *J Biol Chem* 261: 2384-2390.
78. Macey JR, Larson A, Ananjeva NB, Fang Z, Papenfuss TJ, 1997. Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome. *Mol Biol Evol* 14: 91-104.
79. Meusemann K, von Reumont BM, Simon S, Roeding F, Strauss S, et al., 2010. A phylogenomic approach to resolve the arthropod tree of life. *Mol Biol Evol* 27: 2451-2464.

2.5.4 Supplementary Materials

Table S1. Accession numbers of the 98 mt sequences of the 68 species analysed in this study. trnQ-trnF: region between trnQ and trnF; down. nad1: region downstream nad1; down. rrnL: region downstream rrnL. CR/CR2: control region. Species with complete mtDNA sequences are reported in bold.

Taxon	Species	trnQ-trnF	Notes	Down. nad1	Down. rrnL	CR/CR2
Argasidae; Ornithodorinae	Carios capensis		diff. go	NC_005291	NC_005291	NC_005291
Argasidae; Ornithodorinae	Ornithodoros moubata		diff. go	NC_004357	NC_004357	NC_004357
Argasidae; Ornithodorinae	Ornithodoros porcinus		diff. go	NC_005820	NC_005820	NC_005820
Ixodidae; Amblyomminae	Amblyomma americanum	DQ168138	3bp spacer		DQ168139 *	
Ixodidae; Amblyomminae	Amblyomma hebraeum	AY059170S3		AY059170S1		
Ixodidae; Amblyomminae	Amblyomma triguttatum	NC_005963	9bp spacer	NC_005963	NC_005963	NC_005963
Ixodidae; Amblyomminae	Amblyomma vikirri	AY059177S3		AY059177S1		
Ixodidae; Amblyomminae	Aponomma concolor	AY059185S3	4bp spacer	AY059185S1		AY059177S2
Ixodidae; Amblyomminae	Aponomma fimbriatum	AY059189S2				AY059185S2
Ixodidae; Amblyomminae	Aponomma undatum	*				AY059189S1
Ixodidae; Amblyomminae	Haemaphysalis flava		12bp spacer	AY059193S1	AY059193S2	
Ixodidae; Haemaphysalinae	Haemaphysalis humerosa	NC_005292		NC_005292	NC_005292	NC_005292
Ixodidae; Haemaphysalinae	Haemaphysalis longicornis	AY059256S2		AY059256S1		
Ixodidae; Ixodinae	Ixodes acutitarsus		diff. go			AY059260S2
Ixodidae; Ixodinae	Ixodes auritulus		diff. go			AB105167
Ixodidae; Ixodinae; Australasian	Ixodes cordifer		diff. go			AB161425
Ixodidae; Ixodinae; Australasian	Ixodes cornuatus		diff. go			AB161426
Ixodidae; Ixodinae	Ixodes hexagonus		diff. go	NC_002010	NC_002010	NC_002010
Ixodidae; Ixodinae; Australasian	Ixodes hirsti		diff. go			AB161429
Ixodidae; Ixodinae; Australasian	Ixodes holocyclus		diff. go	NC_005293	NC_005293	NC_005293
Ixodidae; Ixodinae	Ixodes loricatus		diff. go			AB161431
Ixodidae; Ixodinae; Australasian	Ixodes myrmecobii		diff. go			AB161433
Ixodidae; Ixodinae	Ixodes persulcatus		diff. go	NC_004370	NC_004370	NC_004370
Ixodidae; Ixodinae	Ixodes pilosus		diff. go			AB161435
Ixodidae; Ixodinae	Ixodes ricinus		diff. go	This study	This study	This study
Ixodidae; Ixodinae	Ixodes scapularis		diff. go			AB161439
Ixodidae; Ixodinae	Ixodes simplex		diff. go			AB161441
Ixodidae; Ixodinae; Australasian	Ixodes trichosuri		diff. go			AB161443
Ixodidae; Ixodinae; Australasian	Ixodes uriae		diff. go	NC_006078	NC_006078	NC_006078
Ixodidae; Rhipicephalinae	Boophilus annulatus			AY059196		
Ixodidae; Rhipicephalinae	Boophilus decoloratus	AY059197S3		AY059197S1	AY059197S2	
Ixodidae; Rhipicephalinae	Boophilus geigy			AY059200		
Ixodidae; Rhipicephalinae	Boophilus kohlsi			AY059201		
Ixodidae; Rhipicephalinae	Boophilus microplus	AF110621		AF110618	AF110619	
Ixodidae; Rhipicephalinae	Dermacentor marginatus			AY059251S1*		

Ixodidae; Rhipicephalinae	<i>Dermacentor reticulatus</i>			AY059253
	<i>Dermacentor variabilis</i>		AY059254S1	AY059254S2
Ixodidae; Rhipicephalinae	<i>Hyalomma aegyptium</i>		AY059268*	AY059264S6
Ixodidae; Rhipicephalinae	<i>Hyalomma truncatum</i>	AY059270S2*	AY059270S1	
Ixodidae; Rhipicephalinae	<i>Rhipicentor nuttalli</i>			AY059234
	<i>Rhipicephalus appendiculatus</i>	AY059214S3	AY059214S1	AY059214S2
Ixodidae; Rhipicephalinae	<i>Rhipicephalus compositus</i>			AY059218
Ixodidae; Rhipicephalinae	<i>Rhipicephalus evertsi</i>	AY059219S3	AY059219S1	AY059219S2
	<i>Rhipicephalus maculatus</i>			AY059222
Ixodidae; Rhipicephalinae	<i>Rhipicephalus pravus</i>	AY059227		
	<i>Rhipicephalus pulchellus</i>	AY059228S3	AY059228S1	AY059228S2
Ixodidae; Rhipicephalinae	<i>Rhipicephalus punctatus</i>			AY059231S2
	<i>Rhipicephalus sanguineus</i>	NC_002074	NC_002074	NC_002074 NC_002074
Ixodidae; Rhipicephalinae	<i>Rhipicephalus simus</i>			AY059223
	<i>Rhipicephalus turanicus</i>			AY059224
Ixodidae; Rhipicephalinae	<i>Rhipicephalus zambeziensis</i>			AY059225
Ixodidae; Rhipicephalinae	<i>Rhipicephalus zumpti</i>			AY059226
Xiphosura; Limulidae	<i>Limulus polyphemus</i>	diff. go	NC_003057	NC_003057
	<i>Tachypleus tridentatus</i>	diff. go	FJ860267	FJ860267
Xiphosura; Limulidae	<i>Drosophila ananassae</i>	diff. go	BK006336	BK006336
Insecta; Diptera	<i>Drosophila erecta</i>	diff. go	BK006335	BK006335
Insecta; Diptera	<i>Drosophila grimshawi</i>	diff. go	BK006341	BK006341
Insecta; Diptera	<i>Drosophila littoralis</i>	diff. go	FJ447340	FJ447340
	<i>Drosophila mauritiana</i>	diff. go	NC_005779	NC_005779
Insecta; Diptera	<i>Drosophila melanogaster</i>	diff. go	NC_001709	NC_001709
	<i>Drosophila mojavensis</i>	diff. go	BK006339	BK006339
Insecta; Diptera	<i>Drosophila persimilis</i>	diff. go	BK006337	BK006337
	<i>Drosophila pseudoobscura</i>	diff. go	FJ899745	FJ899745
Insecta; Diptera	<i>Drosophila sechellia</i>	diff. go	NC_005780	NC_005780
Insecta; Diptera	<i>Drosophila simulans</i>	diff. go	NC_005781	NC_005781
Insecta; Diptera	<i>Drosophila virilis</i>	diff. go	BK006340	BK006340
Insecta; Diptera	<i>Drosophila willistoni</i>	diff. go	BK006338	BK006338
Insecta; Diptera	<i>Drosophila yakuba</i>	diff. go	NC_001322	NC_001322

Legend: *: presence/absence of Tick-Box cannot be determined because of the lack of all or part of trnF; diff. go: sequence absent because of different gene order; x bp spacer: size of the trnQ-trnF spacer, reported only in case of Tick-Box absence

Materials and Methods

- Tick collection and mtDNA amplification -

An adult female of *Ixodes ricinus* was collected from Monte Cornizzolo (Como, Italy). The sample was washed in distilled water and heated at 95°C for 5 minutes in lysis buffer (Qiagen) to inactivate DNases. DNA extraction was performed with the DNeasy Blood & Tissue Kit (Qiagen).

The whole mtDNA of *I. ricinus* was amplified in 11 non overlapped fragments, ranging from 350 to 5500 bp, using 17 mitochondrial-specific primers (Table S2) designed on the available sequences of other *Ixodes* species. These primers were used in several combinations in both standard and long PCR reactions using the GoTaq (Promega) and the Crimson LongAmp *Taq* (New England Biolabs), respectively. Successful single-band amplifications were sequenced directly or by primer walking, according to the Sanger method and using an Applied Biosystems Sequencer. Single reads were manually checked and assembled with gap4 [1]. Ambiguous mt regions were confirmed by additional PCRs.

Table S2. Primers used to analyse the mitochondrial genome of *Ixodes ricinus*.

Primer ^a	Sequence (5' -> 3')	Usage
IRM_11857	AAAGCAACTCTTACTAAAACAC	PCR/Sequencing
IRM_13741	TTAGATACCCTATTATTTAAGC	PCR/Sequencing
IRM_10310	AGTTGATAATAATACACTCAC	PCR/Sequencing
IRM_11862	CTTTGTGTTTTAGTAAGAGTT	PCR/Sequencing
IRM_11857	AAAGCAACTCTTACTAAAACAC	PCR/Sequencing
IRM_1250	TTCCAATGTCTTTATGGTTAGTAG	PCR/Sequencing
IRM_3303	TTTTCCCTTGCTTCACGC	PCR/Sequencing
IRM_4988	GWHCCAAAAATTCTGTCTCT	PCR/Sequencing
IRM_4833	TAATCTCTTCAGGDATTCA	PCR/Sequencing
IRM_6776	TTATTTTTATGCGCGGGGTTA	PCR/Sequencing
IRM_8690	AAATHCCCCTTGTAACG	PCR/Sequencing
IRM_11862	CTTTGTGTTTTAGTAAGAGTT	PCR/Sequencing
IRM_10312	TCGAGTGTATTATTATCAACTG	PCR/Sequencing
IRM_13364	ATGTTACGACTTATCTCACGG	PCR/Sequencing
IRM_149	TTCTAAGGATATTCATAGGGC	PCR/Sequencing
IRM_1380	GGTTGTCCTAATTCAGTTCG	PCR/Sequencing
IRM_1252	CTACTAACCATAAAGACATTGG	PCR/Sequencing
nad1_173pr	TTTCAACCTTTAAGAGATGCTGT	3'RACE
nad1_620pr	CGTAGTCCATTTGATTTAACTGA	3'RACE
rrnL_850pr	AAATTAGGGACAAGAAGACC	3'RACE
rrnL_1050pr	AATACTCTAGGGATAACAGCGT	3'RACE

a: in the sequencing primers, numbers refer to the position on the mtDNA of *Ixodes persulcatus*; in 3' RACE primers, numbers refer to the position on the corresponding gene of *I. ricinus*

Results and discussion

- General Features of the *I. ricinus* mitochondrial genome -

The mtDNA of *I. ricinus* is 14,566 bp long, thus comparable in length to the mtDNA of other Ixodida (Table S3). The genome is AT-rich, with a 78.7% AT-content similar to that of other Ixodida (average AT% = 75.2 ± 2.7 %) (Table S3). The high AT-content is reflected in codon usage by the preference towards the usage of A and T at the third position of synonymous codons, and by the abundance of AUU (Ile) and UUU (Phe) codons (409 and 380 occurrences, respectively).

- Peculiarities of overlapped protein-coding genes -

Both the *atp8/atp6* and *nad4/nad4L* gene pairs of *I. ricinus* exhibit a 7 bp overlap, thus supporting the synthesis of a mature bicistronic mRNA for each of these gene pairs [2,3]. The reason for conservation of a heptamer as overlap sequence is obscure, although we can hypothesize the existence of functional constraints related to the translation of a bicistronic mRNA. Surprisingly, the size and sequence of this overlap is conserved for both these gene pairs in all 11 analysed ticks (Table 1), with nucleotide (nt) differences observed only at the fourth and at the last position of the overlapped sequence (consensus heptamer: ATGATAr in *atp8/atp6*; ATGyTAr in *nad4/nad4L*; ATGhTAr in the two gene pairs together). The occurrence of nt substitutions at only these two positions can be explained considering that the heptamer is subjected to the functional constraints of two different ORFs. Indeed, the first nt triplet needs to be conserved since it is the start codon of the second ORF of the bicistronic mRNA (*atp6* and *nad4L*); the last nt triplet needs to be conserved as TAr because it is the stop codon of the first ORF of the bicistronic mRNA (*atp8* and *nad4*); finally, the fourth position of the heptamer behaves as a first or a third codon position, depending on the ORF considered, thus it can more easily tolerate nt substitutions.

- Transfer RNAs -

The tRNA genes of *I. ricinus* range in size from 56 to 70 bp, and show the typical cloverleaf secondary structure of animal mt tRNAs [4] (Figure S1). As expected, only *trnS(AGN)* lacks of DHU arm. The *trnC* gene, deficient of the DHU arm in metastriata [5], has the canonical four-arm cloverleaf structure found in all other Prostriata and Argasidae ticks.

As shown in Figure S1, the anticodon (AC) sequence is preceded by a T and followed by an A in all tRNAs except for *trnE*, *trnH* and *trnL(UUR)*. In these three tRNAs, the anticodon is followed by a G. The stems of DHU and T arms are quite variable in length, being their size 3-4 and 2-5 bp, respectively. This length variability is very similar to that observed in the variable loops, ranging from 2 to 5 bp. The AC and the amino acid acceptor (AA) stems have a constant size of 7 and 6 bp, respectively, in accordance with the animal tRNA model proposed by Kumazawa and Nishida [4]. In total, only eight unpaired base pairs have been found in the stem regions of all *I. ricinus* tRNAs. Indeed, single mismatched base pairs are present in: the AC stem of *trnK*, *trnM* and

trnT; the DHU stem of *trnC* and *trnI*; the T stem of *trnD* and *trnL(UUR)*; and the first position of the AA stem of *trnM*. Since the mispairing in the AA stem of *trnM* is present in all *Ixodes* species, a short *trnM* gene with an AA stem of only 6 bp has been previously annotated in the *Ixodes* species. However, it should be noted that the first position of the AA stem is well paired in the *trnM* of all Metastriata and of one Argasidae species, and that one mispairing is tolerated in the stem of functional tRNAs [4,6]. Consequently, we have annotated the *trnM* of *I. ricinus* with a canonical 7 bp AA arm.

The comparison of *I. ricinus* tRNAs with the homologous genes of other 4 *Ixodes* species (Table 1 in the main text) shows that the loops of DHU and T arms are the most variable tRNA regions, both in sequence and length. In general, indels and nt substitutions are very frequently in these loops, while compensatory substitutions are prevalent in the stem regions. Moreover, most nt substitutions in tRNAs clearly distinguish Australasian from non-Australasian *Ixodes*.

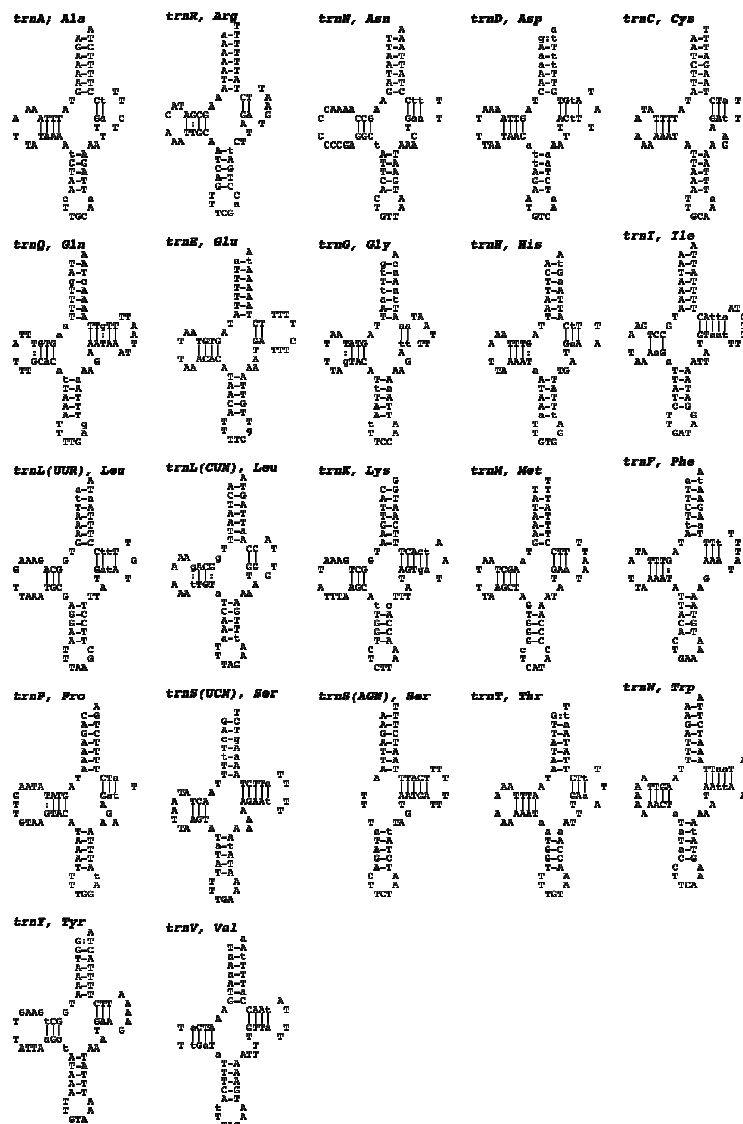


Figure S1. Putative secondary structure of the 22 tRNAs of *I. ricinus*. Lower case indicates positions with nucleotide substitutions in the homologous tRNAs of the other 4 analysed *Ixodes* species (Table 1 of the main paper). Substitutions present in the DHU, T and variable loops are not reported. Canonical and G:T base pairs are differently indicated.

- *The control region* -

The longest non-coding region of arthropods is often indicated as control region (CR), as it usually contains the regulatory elements for mtDNA replication and transcription. The CR of *I. ricinus* is 354 bp long and is located between *rrnS* and *trnI*, the ancestral position of arthropods [7,8,9]. As shown in Table S3, its size is comparable to the CR of other ticks and to the duplicated CR2 of Australasian *Ixodes* and Metastriata. Although the CR is often reported as the most AT-rich region of the arthropod mtDNA, the CR of *I. ricinus* shows almost the same AT% of the whole mt genome (Table S3). This observation holds for almost all Ixodida, except two metastriates whose CR and CR2 have an AT% even lower than that of the entire mtDNA (*Rhipicephalus* and *Haemaphysalis* in Table S3). Therefore, the CRs of ticks do not follow the typical trend of other arthropods towards an increase of AT%, and in Ixodida the term “AT-rich region” should not be used as synonymous of CR.

The CR of *I. ricinus* was aligned and compared to the CR and CR2 fully sequenced in 15 tick species (Table S1), allowing identifying four sequence motifs, including one hairpin structure conserved in some taxa (Figure S2). Motifs I and III (Figure S2) are short sequences perfectly conserved in all analysed *Ixodes* species and were identified by Shao et al. [10] as specific of Australasian *Ixodes*. Motif IV is a 9 bp C-rich sequence present in the CR and CR2 of all analysed tick species, although with a different consensus depending on the taxonomic group (Figure S2). Finally, motif II is a hairpin secondary structure with a 5-8 bp stem and a loop of 4-6 bp. In *I. ricinus*, this hairpin is a stable structure with a $\Delta G = -2.93$ kcal/mol (MFold). This secondary structure is conserved only in *Ixodes* and Argasidae (Mfold analyses), and encompasses two close elements identified as specific of Australasian *Ixodes* by Shao et al. [10].

The conservation of these motifs indicates that they may have specific roles in the tick mtDNA replication and/or transcription. It should be noted that several Authors have hypothesized the association of hairpin structures to the mtDNA replication origin(s) [11,12].

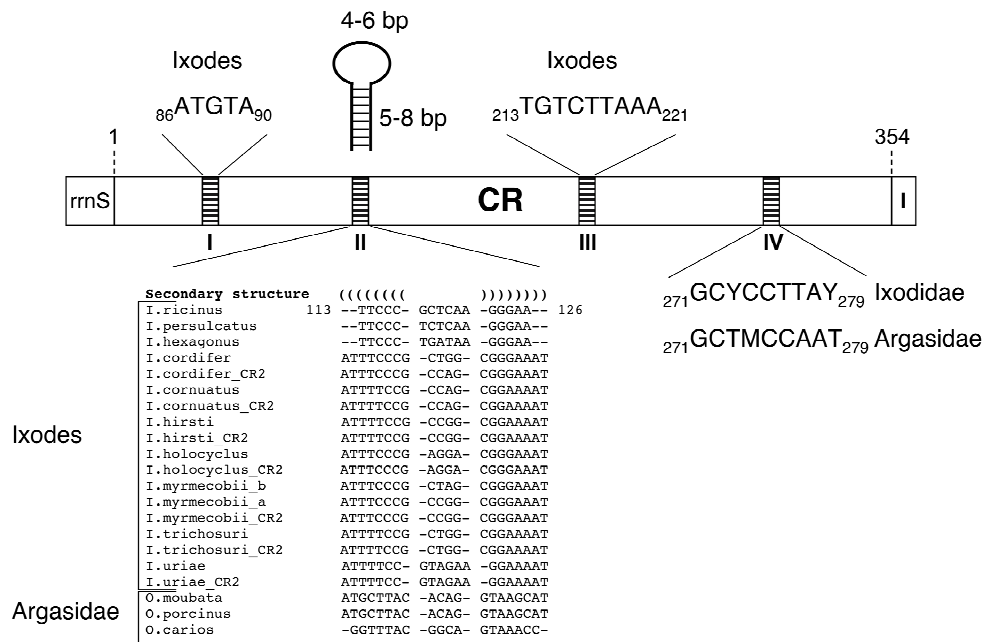


Figure S2. Conserved motifs and secondary structures of the tick control region, mapped on the *I. ricinus* sequence. Dashed blocks indicate conserved motifs/secondary structures, with numbers referring to their position in the *I. ricinus* control region. The two tandem repeats of *Ixodes myrmecobii* are indicated as “_a” and “_b”. Gene abbreviations are reported as in Figure 1 of the main paper. Accession number of the analysed sequences is reported in Table S1.

- Small non-coding regions -

The *I. ricinus* mtDNA has only 11 small NCRs, for a total of 72 bp corresponding to 0.49% of the entire mtDNA (Table S3). These values are similar to those observed in other Ixodida, except for *R. sanguineus* and Argasidae, where the number and length of single NCRs increase and drastically decrease, respectively (Table S3).

A detailed analysis of the non-coding regions (excluding the CR and the Tick-Box) shows that most tick NCRs are 1-6 bp long and have a different length from one species to the other, thus they can be considered as simple gene spacers with no specific functions. As an exception, the NCR between *trnL(UUR)* and *trnL(CUN)* ranges from 9 to 14 bp in *Ixodes* species, is 2-4 bp long in Argasidae, and is absent in Metastricata due to the gene order rearrangement (Figure 1 of the main text). Thus, this NCR is an AT-rich sequence (AT% >83%) showing a tendency to size variation related to the taxonomic group.

In the three complete mtDNAs of Metastricata, the NCR between *trnQ* and *trnF* is 9-25 bp long but its size ranges from 3 to 36 bp when partial sequences are also considered (Table S1). As reported in the main paper, this NCR is located in a gene adjacency restricted only to Metastricata.

Moreover, in some species it includes a third copy of the Tick-Box motif in the reverse-complement orientation compared to the Tick-box motifs downstream of *nad1* and *rrnL*.

On the overall, these analyses show that the mtDNA of *I. ricinus* conforms to that of other *Ixodes* species in size and distribution of small NCRs. Moreover, there are no small conserved NCRs other than those containing the Tick-Box.

Table S3. Size and base composition of mtDNA, control region (CR), and small non-coding regions (NCR) of Ixodida.

	mtDNA		CR		CR2		Small NCR		
	bp	AT%	bp	AT%	bp	AT%	N°	bp	%
<i>Ixodes ricinus</i>	14566	78.6	354	78.5	none		11	72	0.49
<i>Ixodes hexagonus</i>	14539	72.7	358	71.9	none		15	82	0.56
<i>Ixodes persulcatus</i>	14539	77.3	352	77.6	none		12	79	0.54
<i>Ixodes holocyclus</i>	15007	77.4	352	78.4	450	80.0	14	93	0.62
<i>Ixodes uriae</i>	15053	74.8	388	77.1	476	71.0	14	93	0.62
<i>Amblyomma triguttatum</i>	14740	78.4	307	71.6	307	71.7	14	89	0.60
<i>Haemaphysalis flava</i>	14686	76.9	310	66.8	310	66.5	14	76	0.52
<i>Rhipicephalus sanguineus</i>	14710	78.0	304	66.6	303	67.3	18	130	0.88
<i>Carios capensis</i>	14418	73.5	342	71.4	none		9	40	0.28
<i>Ornithodoros moubata</i>	14398	72.3	342	71.6	none		6	37	0.26
<i>Ornithodoros porcinus</i>	14378	71.0	338	69.5	none		6	37	0.26

CR: ancestral control region; CR2: duplicated control region

References

1. Staden R, Beal KF, Bonfield JK, 2000. The Staden package, 1998. *Methods Mol Biol* 132: 115-130.
2. Berthier F, Renaud M, Alziari S, Durand R, 1986. RNA mapping on *Drosophila* mitochondrial DNA: precursors and template strands. *Nucleic Acids Res* 14: 4519-4533.
3. Stewart JB, Beckenbach AT, 2009. Characterization of mature mitochondrial transcripts in *Drosophila*, and the implications for the tRNA punctuation model in arthropods. *Gene* 445: 49-57.
4. Kumazawa Y, Nishida M, 1993. Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. *J Mol Evol* 37: 380-398.
5. Shao R, Aoki Y, Mitani H, Tabuchi N, Barker SC, et al., 2004. The mitochondrial genomes of soft ticks have an arrangement of genes that has remained unchanged for over 400 million years. *Insect Mol Biol* 13: 219-224.

6. Watanabe K, 2010. Unique features of animal mitochondrial translation systems. The non-universal genetic code, unusual features of the translational apparatus and their relevance to human mitochondrial diseases. *Proc Jpn Acad Ser B Phys Biol Sci* 86: 11-39.
7. Lavrov DV, Boore JL, Brown WM, 2000. The complete mitochondrial DNA sequence of the horseshoe crab *Limulus polyphemus*. *Mol Biol Evol* 17: 813-824.
8. Boore JL, Lavrov DV, Brown WM, 1998. Gene translocation links insects and crustaceans. *Nature* 392: 667-668.
9. Boore JL, Collins TM, Stanton D, Daehler LL, Brown WM, 1995. Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. *Nature* 376: 163-165.
10. Shao R, Barker SC, Mitani H, Aoki Y, Fukunaga M, 2005. Evolution of duplicate control regions in the mitochondrial genomes of metazoa: a case study with Australasian *Ixodes* ticks. *Mol Biol Evol* 22: 620-629.
11. Clary DO, Wolstenholme DR, 1987. *Drosophila* mitochondrial DNA: Conserved sequences in the A + T -rich region and supporting evidence for a secondary structure model of the small ribosomal RNA *J Mol Evol* 25: 116-125.
12. Zhang DX, Szymura JM, Hewitt GM, 1995. Evolution and structural conservation of the control region of insect mitochondrial DNA. *J Mol Evol* 40: 382-391.

2.6 Research article. *Localization of the bacterial symbiont Candidatus Midichloria mitochondrii within the hard tick Ixodes ricinus by whole-mount FISH staining*

2.6.1 Summary

Here, I present an investigation on the spatial distribution of the bacterial symbiont '*Candidatus Midichloria mitochondrii*' within *Ixodes ricinus*, by whole mount fluorescence in situ hybridization (FISH). '*Candidatus M. mitochondrii*' is a peculiar, recently discovered bacterium that resides in the mitochondria of female ticks. I applied a rapid and specific FISH protocol with oligonucleotide probes targeted on the 16S rRNA of '*Candidatus M. mitochondrii*', 12S rRNA of tick mitochondria, and a probe revealing active mitochondria. In this report that represents the first application of whole mount FISH on ticks were observed strong, specific fluorescence signals in all the examined life stages, as the optimized protocol allowed to overcome the autofluorescence interference of the cuticle. Cellular localization and quantification of the symbionts were also assessed with electron microscopy and specific real-time PCR, respectively.

2.5.2 Manuscript

Introduction

In temperate zones, ixodid ticks are the most important vectors of animal and human pathogens (viruses, bacteria, protozoa). The tick *Ixodes ricinus* is regarded as the main vector of the causative agents of Lyme borreliosis, tick-borne encephalitis (TBE), and various ehrlichioses and rickettsioses in Europe (Parola and Raoult, 2001; Socolovschi *et al.*, 2009). "*Candidatus Midichloria mitochondrii*" (hereafter called *M. mitochondrii*) is an intracellular alpha-proteobacterial symbiont of *I. ricinus*. This symbiosis is peculiar since the bacterium not only resides in the cytoplasm of the tick cells, but also possesses the capacity of entering the mitochondria of the host (Beninati *et al.*, 2004; Epis *et al.*, 2010). Previous studies found *M. mitochondrii* to be present in 100% females and 44% males of *I. ricinus* throughout the geographical distribution of this species in Europe and North Africa (Lo *et al.*, 2006). A specific

real-time PCR, previously described by Sasser and colleagues (2008), was used to quantify *M. mitochondrii* within its host *I. ricinus*, revealing the presence of a high number of bacteria in larvae, nymphs, and adult females. The distribution of Midichloria-related symbionts was investigated in 22 other hard tick species, and 8 of them with a positive result, albeit with different levels of prevalence. *Rhipicephalus sanguineus*, the most widely distributed tick in the world (Dantas-Torres, 2008), was considered devoid of *M. mitochondrii* due to the absence of PCR amplification with specific primers (Epis *et al.*, 2008). The data previously collected on the presence of *M. mitochondrii* in *I. ricinus* suggest that this symbiont plays an important role in the biology of this tick. But what is the localization, the abundance, and the kind of activity of this microorganism in the different life stages? The aim of the present study was to study the spatial distribution of *M. mitochondrii* bacteria during the life cycle of *I. ricinus*. This was obtained by fluorescent in situ hybridization (FISH) and electron microscopy (TEM), supplemented by specific real-time PCR. FISH is a powerful technique, widely used to detect host–bacteria associations in arthropods (Fritsche *et al.*, 1999; Moter and Göbel, 2000; Gómez-Valero *et al.*, 2004; Gottlieb *et al.*, 2006; Skaljac *et al.*, 2010; Gonella *et al.*, 2011). At present, only 2 reports describe the application of FISH in ticks in order to evaluate the presence of arthropod-associated bacteria investigating *Borrelia burgdorferi* in *Ixodes ricinus* (Hammer *et al.*, 2001) and a *Coxiella*-type symbiont in *Amblyomma americanum* (Klyachko *et al.*, 2007). However, these 2 studies are limited to dissected tissues or whole-body sections, never examining whole mount samples. We used whole-mount FISH at various stages of development, to investigate, for the first time, the specific localization of *M. mitochondrii* and mitochondria in the body of the tick *I. ricinus*.

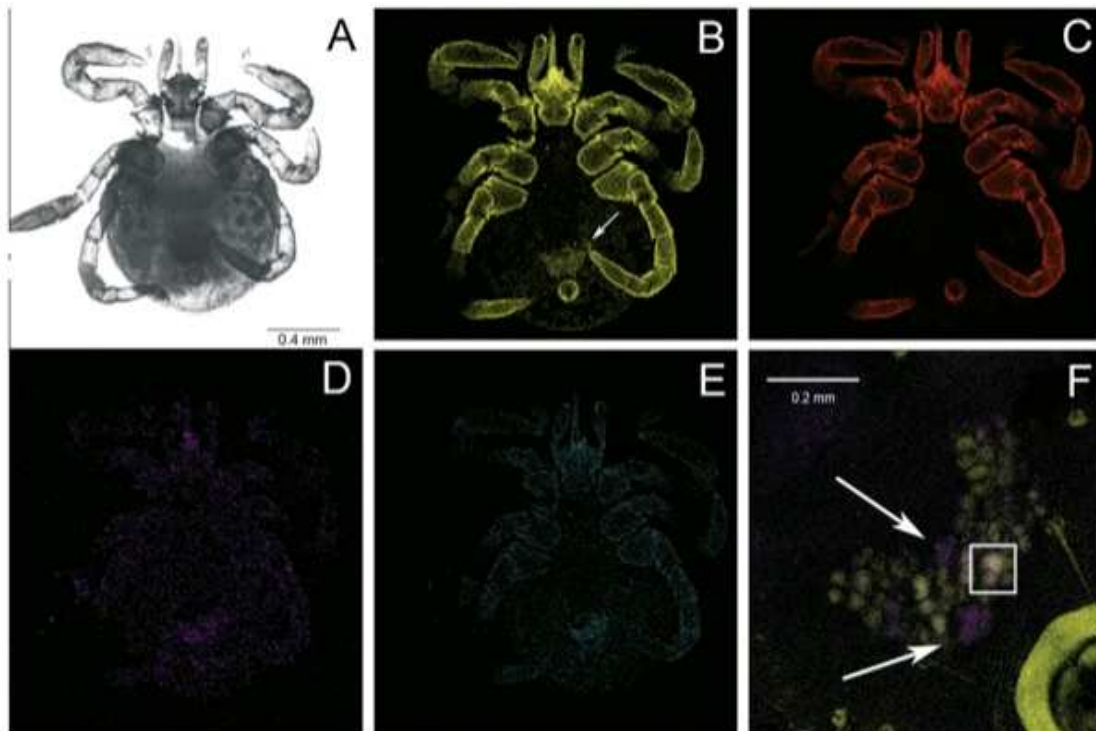


Figure 1. Detection of "*Candidatus Midichloria mitochondrii*" symbionts in larvae of *Ixodes ricinus*. (A) Image obtained from light transmission. (B) 16S rRNA probes for *M. mitochondrii* labeled with Cy5. The arrow indicates fluorescence foci. (C) Tissue auto-fluorescence. (D) 12S rRNA probes for *I. ricinus* mitochondria detection. (E) Detection of vitality of mitochondria by Mitotracker Red. (F) Higher magnification of the primordial ovary. Overlay image with yellow signal for *M. mitochondrii* and pink signal for mitochondria detection. Arrows indicate the mitochondria without symbiont; in the square a merging of *M. mitochondrii* and the mitochondria signals.

Materials and methods

- Ticks -

Ten *I. ricinus* and 2 *R. sanguineus* engorged females were collected from animals (goats and dogs) in the Bergamo province (northern Italy). Thirty non-engorged *I. ricinus* nymphs were also collected from the field by flagging the vegetation in the same province. All ticks were identified using standard taxonomic keys (Manilla, 1998). Five *I. ricinus* and 2 *R. sanguineus* ticks were maintained at 26°C, 85% relative humidity, and 12L:12D photoperiod in order to allow oviposition. Eggs were maintained at the same conditions until hatching. The ovaries extracted from the remaining 5 *I. ricinus* females were divided in 3 parts, and each part was subjected to FISH, TEM, or qPCR analyses. Ten *I. ricinus* nymphs, 10 *I. ricinus* larvae, and 5 *R. sanguineus* larvae were processed for FISH; 20 *I. ricinus* nymphs, 20 *I. ricinus* larvae,

and 5 *R. sanguineus* larvae were processed for qPCR. DNA purification and qPCR Twenty *I. ricinus* and 10 *R. sanguineus* larvae, 20 *I. ricinus* nymphs, and 5 portions of ovary from adult *I. ricinus* females were individually crushed and subjected to DNA extraction using Qiagen tissue kit (Qiagen, Chatsworth, California), as previously described (Epis *et al.*, 2008). Quantitative PCR (qPCR) specific for the *M. mitochondrii* gene coding for gyrase B (*gyrB*) and for the host nuclear gene coding for calreticulin (*cal*) were used to measure symbiont loads in various stages of the host life as previously described (Sassera *et al.*, 2008).

- *FISH* -

We utilized 16S rRNA probes for *M. mitochondrii* (Midi0066 5'-GCTACAGCTCTTGCCCGT-3' and Midi1410 5'-CAAAACCGACTCCCATGGC-3') labeled at the 5' with the fluorochrome Cy5 (indodicarbocyanine) (absorption and emission at 650 nm and 670 nm, respectively). Probe Midi0066 had 11, 12, and 8 mismatches, and Midi1410 had 3, 5, and 10 mismatches with 16S rRNA sequences of other alpha-proteobacteria that commonly parasitize *I. ricinus* ticks, *Rickettsia conorii* (accession no. AF541999), *Anaplasma phagocytophilum* (accession no. AY055469), and *Borrelia burgdorferi* (accession no. AE001147) (Beninati *et al.*, 2004). The oligonucleotide sequences were further checked for their suitability with the software Ribosomal Database Project (Cole *et al.*, 2005). In order to detect tick mitochondria, 12S rRNA mitochondria-specific probes labeled at the 5' end with the fluorochrome Cy3 (indocarbocyanine) (absorption and emission at 550 nm and 570 nm, respectively) were designed on 12S rRNA sequence obtained from *I. ricinus* (1Mito12S 5'-AAATCTGAAACATTGTTCCCTTG-3' and 2Mito12S 5'-GTTTTAGAGATGACCAGCTTTCTTCAC-3'). Additional probe EUB338 (5'-GCTGCCTCCCGTAGGAGT-3'), routinely used as a universal bacterial probe, was employed as a bacterial-positive control (Amann *et al.*, 1990). Probe Gam1019, labeled with Cy5, specific for gamma-proteobacteria was used in this study as a negative control (Nielsen *et al.*, 1999). Before hybridization with the probes, ovaries were dissected from engorged ticks under sterile conditions; larvae and nymphs were cut near the capitulum area utilizing a microclipper for dissection.

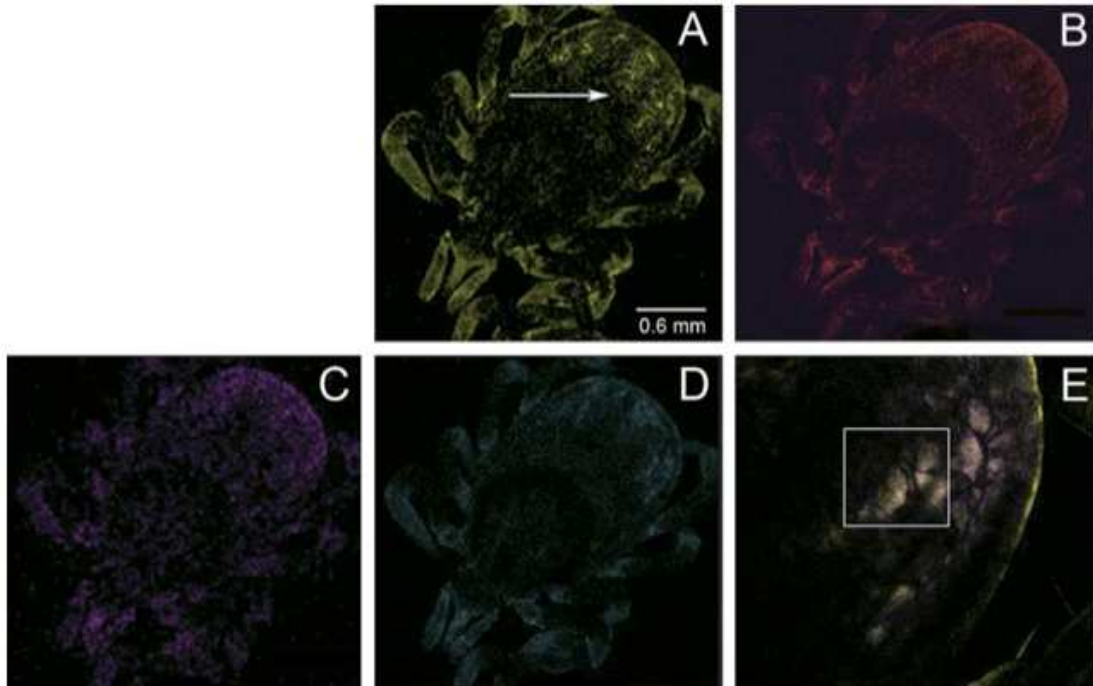


Figure 2. Fluorescent in situ hybridization of an *Ixodes ricinus* nymph. (A) 16S rRNA probes for "*Candidatus* Midichloria mitochondrii" detection. The arrow indicates the concentration of bacteria in the region of the ovary. (B) Tissue autofluorescence. (C) 12S rRNA probes for mitochondria detection. (D) Mitotracker Red detecting living mitochondria. (E) Higher magnification of overlay obtained from *M. mitochondrii* (yellow) and mitochondria (pink) detection; within the square a merge of the 2 probes is shown (white). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article).

After rapid washing in 1x PBS buffer, all samples were maintained in 500 μ l 50 nM Mitotracker Red (Molecular Probe, Eugene, Oregon) in 1x PBS for 30 min at room temperature (RT) in the absence of light.

We determined mitochondrial membrane integrity by analyzing the membrane potential in all samples using Mitotracker Red, a fluorophore that concentrates in mitochondria with high membrane potential and is retained during further processing, thus allowing to assess the vitality of mitochondria in fluorescence microscopy.

After treatment with Mitotracker Red, all tick samples were fixed in 4% paraformaldehyde for 30 min at room temperature and washed in 1x PBS twice. Dissected organs were incubated for 15 min. A prehybridization treatment with proteinase K (1 μ g/ml) was performed for 15 min for larvae and nymphs, for 10 min for ovary samples. Subsequently, samples were washed twice in a solution of 1x PBS containing 1% Tween 20 and once in 1x PBS for 5 min at RT. Hybridization was carried in the dark for 3 h at 37°C, with 20 μ l of hybridization buffer (10x SSC, 50%

formamide, Denhardt's solution 0.5%, 30 ng/ml probes). After hybridization, all tick samples were washed in 500 µl of washing buffer 2x SSC for 10 min, then in 500 µl of 0.1x SSC for 10 min. 75 ng/ml of 4',6-diamidino-2-phenylindole (DAPI) were added for nuclei detection and incubated for 15 min at RT. After a wash in 1x PBS at room temperature, samples were mounted in glass slides with 40 µl of 1,4-diazabicyclo [2.2.2] octane (DABCO) and observed using a laser-scanning confocal microscope SP2-AOBS (Leica, Wetzlar, Germany).

- *Electron microscopy* -

Five ovary portions from *I. ricinus* females were chosen for transmission electron microscopy (TEM) examination. Samples were fixed in 0.1 M cacodylate buffer (pH 7.2) containing 2.5% glutaraldehyde for 3 h at 4°C. The samples were then washed in the same buffer and post-fixed in 1% OsO₄ in the same buffer for 1.5 h at 4°C. Successively, all samples were dehydrated in ethanol and embedded in Epon 812. The semi-thin sections (1 µm) for light microscopy were stained with 0.5% toluidine blue; thin sections (80 nm) were stained with uranyl acetate and lead citrate and examined under an EM900 transmission electron microscope (Zeiss).

Results and discussion

A previously developed qPCR assay (Sassera *et al.*, 2008) was employed to quantify *M. mitochondrii* in all *I. ricinus* samples (5 ovary portions, 20 larvae obtained in the lab, and 20 nymphs collected from the field) prior to FISH and TEM analysis. qPCR results are expressed as absolute quantification of *gyrB* (i.e., the *M. mitochondrii* gene) and *cal* (the *I. ricinus* gene) obtained by interpolating the threshold cycle (ct) values of each sample on the equation obtained with the serial dilutions of samples containing known copy numbers of each gene fragment, cloned in a plasmid vector. The total number of *cal* gene copies in the different life stages of *I. ricinus* was coherent with previous results, indicating a high quality of the DNA extraction and qPCR protocols (data not shown). The quantifications of *gyrB* in single larvae are comparable to previously obtained results (Sassera *et al.*, 2008) with a mean value of 2.48×10^4 copies. Statistical analysis performed in larvae did not show a significant variation of *M. mitochondrii* genome copies between samples (Kruskal–Wallis test, $P = 0.189$). A different result was obtained when evaluating the uniformity of the *M.*

mitochondrii load in nymphs (Kruskal–Wallis test, $P < 0.05$). These data indicate that the *M. mitochondrii* load is variable in nymphs in the analyzed *I. ricinus* population. The loads in nymphs range from 2.98×10^3 to 1.29×10^5 copies, with a mean value of 4.75×10^4 . The load values exhibited a bimodal distribution, with 2 groups above and below the mean value (respective mean values of 1.19×10^4 and 1.13×10^5 copies). We analyzed these 2 groups (Mann–Whitney U test, $P < 0.001$) and obtained a statistically significant difference between the two. This result is congruent with the variability observed among individuals in FISH fluorescence levels (see below for discussion).

We thus performed FISH and TEM analysis on the remaining larvae and nymphs. Ovary samples exhibited a high content of *M. mitochondrii* (mean 6.2×10^7 per genome). These values are comparable to the loads previously observed in qPCR performed on whole engorged *I. ricinus* females. Previous observations (Sacchi et al., 2004) confirm that the ovary is the organ exhibiting the highest concentration of *M. mitochondrii* symbionts within ticks, making it an excellent candidate for microscopy and hybridization analysis.

Five *I. ricinus* ovary portions, 10 *I. ricinus* nymphs, 10 *I. ricinus* larvae, and 5 *R. sanguineus* larvae were processed for FISH; we used Mitotracker Red, 2 probes specific for *M. mitochondrii*, a probe for 12S mitochondrial gene, and DAPI, EUB338, and Gam1019 as controls.

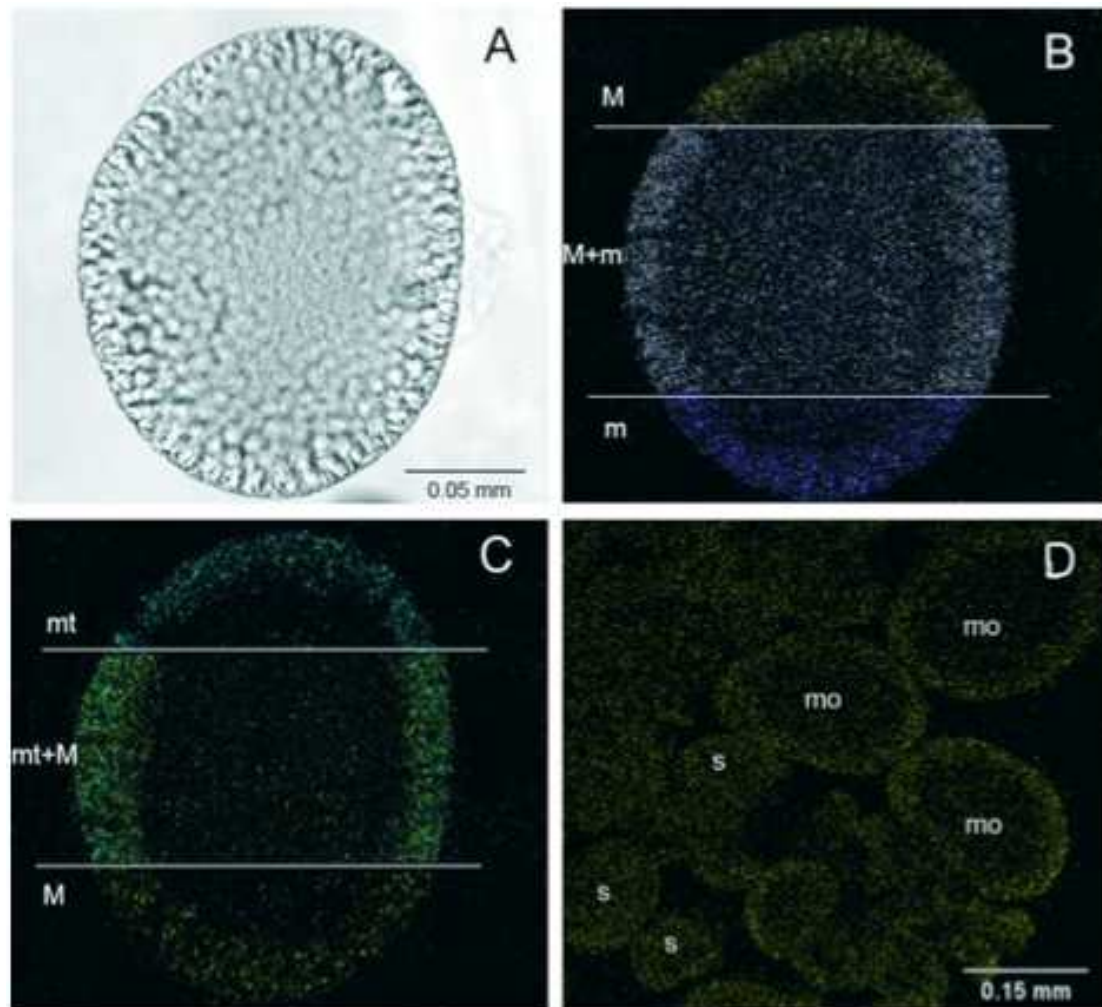


Figure 3. Hybridization on oocytes obtained from an ovary of a semi-engorged *Ixodes ricinus*. (A) Single mature oocyte in light transmission. (B) The same mature oocyte in 3 different fluorescence sections: in the upper part 16S rRNA probes for *Candidatus Midichloria mitochondrii* indicated by M; in the lower part 12S rRNA probes for mitochondria detection (m); in the central part, the merge obtained from the 2 probes (M+m). (C) The same oocyte in 3 other fluorescence sections: in the upper part Mitotracker detection of live mitochondria indicated by mt; in the lower part the 16S rRNA probes for *M. mitochondrii* (M); in the central part, the merge obtained from the 2 probes (mt+M). (D) Different levels of polarization of symbiont presence in oocytes at different stages: symbionts uniformly distributed in small oocytes (s) and polarized at the periphery in mature oocytes (mo).

FISH was performed on larvae and nymphs in whole mount; in both sample types, tissue auto-fluorescence clearly revealed the silhouette of the tick. Hybridization with the EUB338 probe was performed on 2 larvae, 2 nymphs, and 2 ovaries generating a positive fluorescent signal in all cases. The signal obtained from EUB338 is comparable in intensity and localization to the signal obtained with the probes specific for *M. mitochondrii*, indicating that the vast majority of the bacterial biomass

is composed by these symbionts (Fig. 4D–F). Gam1019, specific probe for Gamma-proteobacteria, utilized as a negative control, did not generate detectable fluorescence signals in any of the samples (data not shown). FISH was also performed on 5 larvae of *R. sanguineus* ticks with the 2 probes specific for *M. mitochondrii* (Fig. 4A–C); representing a negative control (i.e., a tick known to be negative for *M. mitochondrii*). None of these samples exhibited any specific fluorescence. A strong hybridization signal (specific for *M. mitochondrii*) was observed near the anal opening in all analyzed larvae (Fig. 1B).

When this region was observed by light transmission microscopy (Fig. 1A), a dense mass of round bodies was detected. In unfed and engorged larvae, this structure represents the primordial ovary composed of large, spherical, or polygonal primordial oocytes with round nuclei (Balashov, 1968; Sonenshine, 1993). The same region was observed with fluorescence microscopy at higher magnifications (Fig. 1F), revealing that the *M. mitochondrii*-specific probes hybridize only within the round bodies of the primordial reproductive organ. Furthermore, some of the round bodies are only stained with the mitochondrial probe and are negative for the *M. mitochondrii* probe, indicating that only a portion of the round bodies is infected with the bacterial symbionts (Fig. 1F). This same result (i.e., only some of the round bodies are *M. mitochondrii*-positive) was observed in nymphs, as highlighted in Fig. 2E. In nymphs, positive hybridization signals were observed at the level of cox IV, where the reproductive organs are located. Assuming, based on literature, that the sex ratio is balanced (Kiszewski et al., 2001) and thus that the nymphs likely represent both sexes, we can conclude that the probe stained *Midichloria* in both males and females (Balashov, 1968) (Fig. 2).

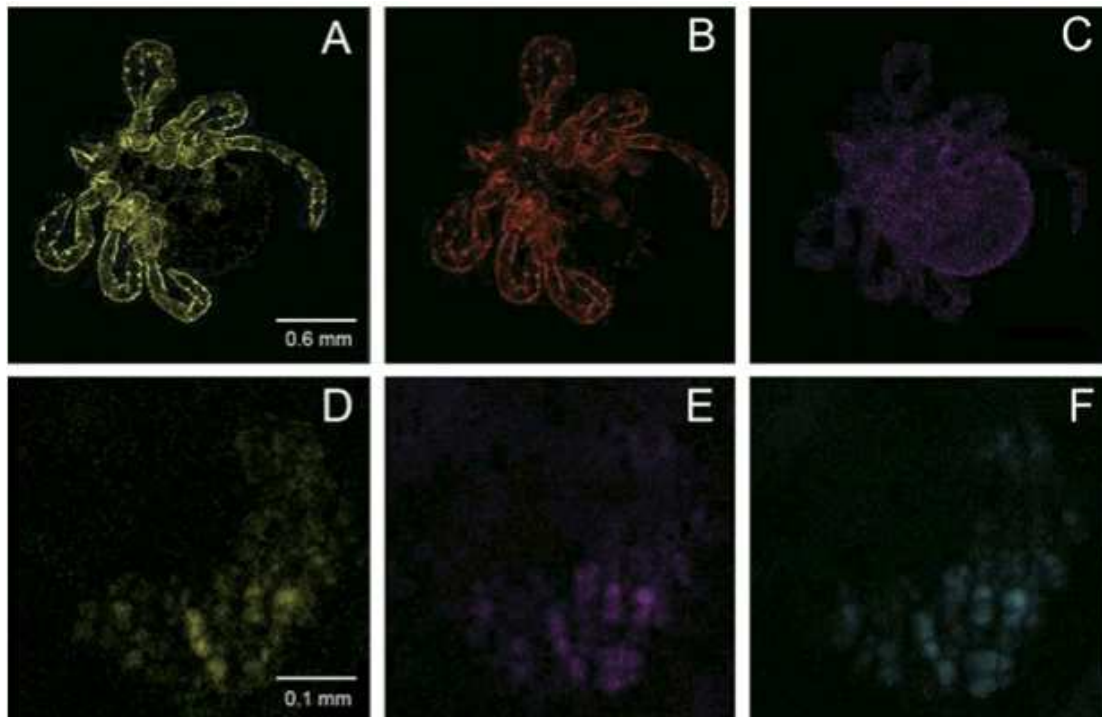


Figure 4. Fluorescence controls in a larva of *Rhipicephalus sanguineus* (A–C) and in an ovary of *Ixodes ricinus* (D–F). (A) 16S rRNA probes for *Candidatus Midichloria mitochondrii* (negative). (B) Tissue autofluorescence. (C) 12S rRNA probes for mitochondria detection. (D) 16S rRNA probes for *M. mitochondrii*. (E) Universal eubacterial probe EUB338. (F) Mitotracker Red.

However, 6 nymphs showed high specific fluorescence in the reproduction system, while the remaining 4 showed a low signal in the same region. Previous results indicate a low mean load in adult males (3.17×10^3 copies) and a high *M. mitochondrii* mean load in non-engorged adult females (2.33×10^6 copies) (Sassera et al., 2008). Both qPCR and FISH results here presented show 2 distinct groups of values in the *M. mitochondrii* load in nymphs. Based on these results, we can hypothesize that nymphs that will become females have a higher *M. mitochondrii* load compared to the ones that will become males.

Since nymphs cannot be sexed based on morphological characteristics and no genetic sex determination test is currently available, the measurement of the *M. mitochondrii* load, either by qPCR or by FISH, could be evaluated as a tool to sex these immature stages in *I. ricinus*.

FISH was also performed on 5 ovary portions from 5 engorged *I. ricinus* ticks (Fig. 3). All the examined ovary portions showed specific fluorescence signals with the *M. mitochondrii*-specific probes. Particularly bright signals were observed in the smallest oocytes, confirming a strong *M. mitochondrii* presence in these cells, as previously

reported by TEM (Sacchi *et al.*, 2004). When analysing higher magnifications, it is possible to note that the signal of M. mitochondrii is stronger in the periphery of the examined oocytes (Fig. 3B and C). The high fluorescence level in the oocytes was confirmed by TEM pictures; in particular, the periphery of the oocytes observed by TEM appeared to be rich of intramitochondrial symbionts (Figs. 5 and 6). We can explain this evidence as a reorganization of the symbionts in developed oocytes similar to that in *Wolbachia* of *Drosophila* during oogenesis. Experiments conducted in *Drosophila* indicated that *Wolbachia* disperse from the anterior of the oocyte suggesting that this symbiont uses the host's microtubule cytoskeleton and transport system to ensure its transmission (Ferree *et al.*, 2005). The result obtained from the merging between Mitotracker and specific M. mitochondrii fluorescence shows a non-perfect overlap between the 2 signals. This indicates that some live mitochondria do not present symbionts (Fig. 3C), in agreement with the results on immatures presented above. On the other hand, the merge obtained from the mitochondria-specific probe, which stains all mitochondria, and the M. mitochondrii probe shows a clear overlap (Fig. 3B). The differences between the stainings obtained with Mitotracker and with the mitochondria-specific probe can be explained considering that Mitotracker only stains live mitochondria.

These results suggest that the symbionts not only reside in live mitochondria, but that they can also be found in dead organelles or at least in organelles whose function was inhibited enough to reduce their reactivity with Mitotracker (Fig. 3B). It remains to be elucidated if the symbionts can invade dead mitochondria or if they only enter live mitochondria and then cause the death of the organelles, or any reduction of their function. This second explanation fits better with previous TEM results (Sacchi *et al.*, 2004), which showed a progressive degradation of the mitochondrial matrix following the multiplication of M. mitochondrii symbionts therein. In conclusion, FISH experiments using Cy3- and Cy5-labeled probes have led to precisely determine the selective localization of endosymbionts and mitochondria throughout the life stages of *I. ricinus* tick showing an excellent fluorescence signal. The protocol here presented is specifically designed for whole-mount FISH in ticks. Here, we used it to study the distribution of M. mitochondrii in different life stages of *I. ricinus*, but it is applicable to other tick-borne bacteria, to other tick species, or for additional studies in this system such as accurate description of symbiont localization following response to antibiotic treatment.

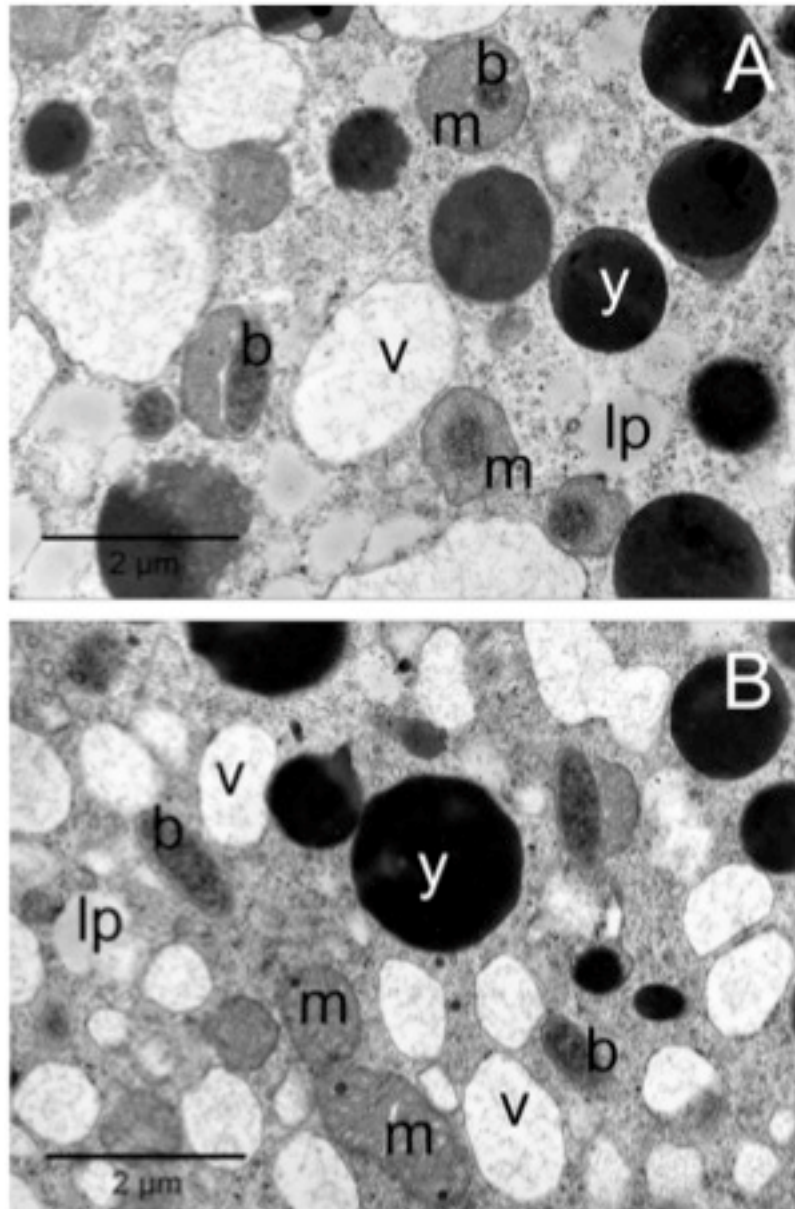


Figure 5. Transmission electron microscopy of the periphery of a vitellogenic oocyte. (A) In the foreground, 2 mitochondria (m) parasitized by *Candidatus Midichloria mitochondrii* (b). (B) Some bacteria are also observed in the cytoplasm and inside the mitochondria near the chorium (c). ld, lipid droplets; tp, tunica propria; v, vesicles; y, yolk sphere.

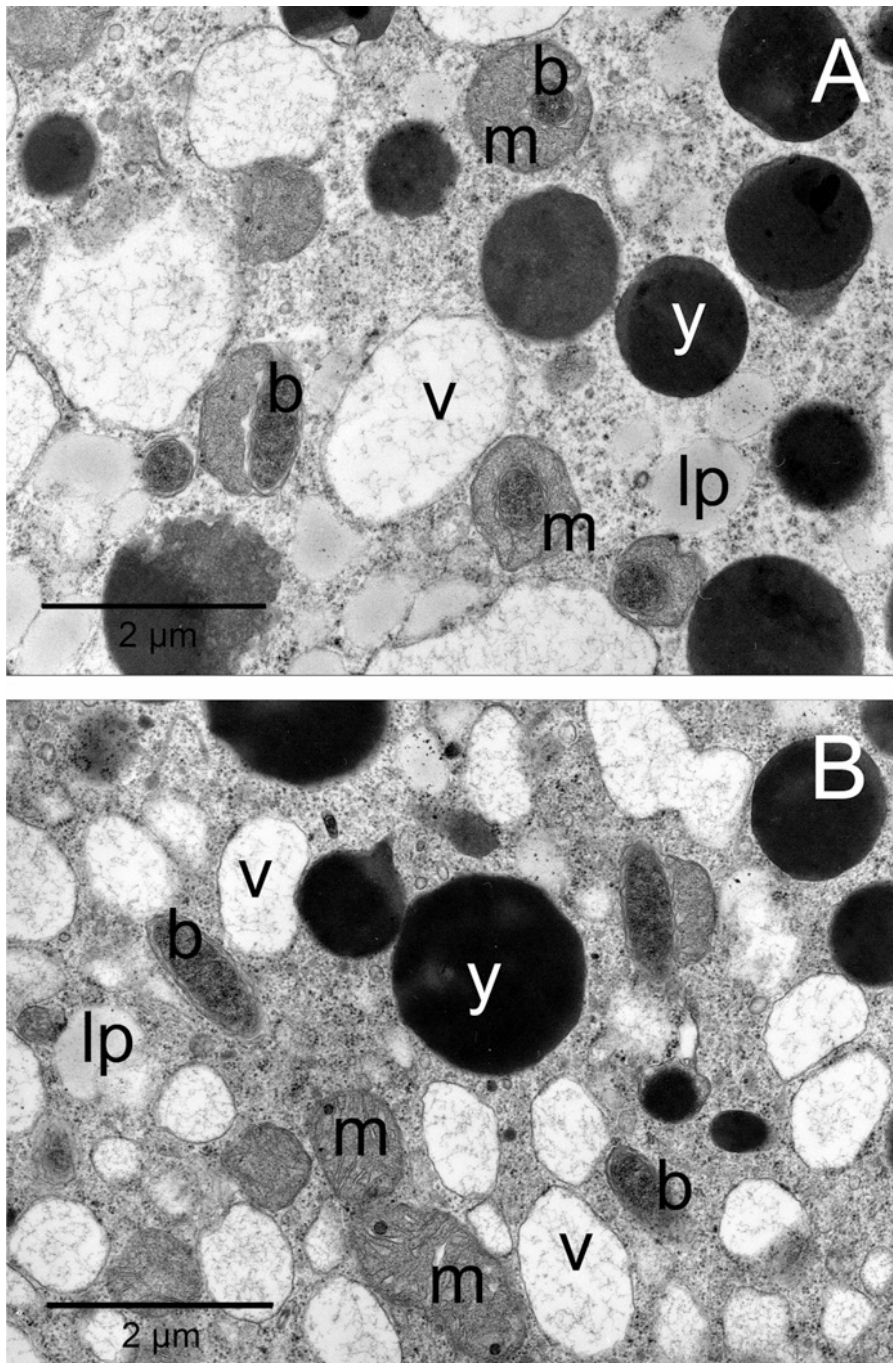


Figure 6. (A, B) Magnification of numerous *Candidatus Midichloria mitochondrii* bacteria (b) inside and outside the mitochondria (m) in the periphery of the developing oocyte obtained by TEM. lp, lipid droplets; tp, tunica propria; v, vesicles; y, yolk sphere.

2.6.3 References

- Amann RI, Krumholz L, Stahl DA, 1990. Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. *J. Bacteriol.* 172: 762–770. 312
- Balashov YuS, 1968. Bloodsucking Ticks (Ixodoidea) – Vectors of Diseases of Man and Animals. Nauka Publishers, Leningrad, pp. 314
- Beninati T, Lo N, Sacchi L, Genchi C, Noda H, Bandi C, 2004. A novel alpha-proteobacterium resides in the mitochondria of ovarian cells of the tick *Ixodes ricinus*. *Appl. Environ. Microbiol.* 70: 2596–2602. 317
- Cole JR, Chai B, Farris, RJ, Wang Q, Kulam SA, McGarrell DM, Garrity GM, Tiedje JM, 2005. The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res.* 1: 33.
- Dantas-Torres F, 2008. The brown dog tick, *Rhipicephalus sanguineus* (Latreille, 1806) (Acari: Ixodidae): from taxonomy to control. *Vet. Parasitol.* 152: 173–185.
- Epis S, Sassera D, Beninati T, Lo N, Beati L, Piesman J, Rinaldi L, McCoy KD, Torina A, Sacchi L, Clementi E, Genchi M, Magnino S, Bandi C, 2008. *Midichloria mitochondrii* is widespread in hard ticks (Ixodidae) and resides in the mitochondria of phylogenetically diverse species. *Parasitology* 135, 485–494:
- Epis S, Luciano AM, Franciosi F, Bazzocchi C, Crotti E, Pistone D, Bandi C, Sassera D, 2010. A novel method for the isolation of DNA from intracellular bacteria, suitable for genomic studies. *Ann. Microbiol.* 60: 455–460.
- Ferree PM, Frydman HM, Li JM, Cao J, Wieschaus E, Sullivan W, 2005. *Wolbachia* utilizes host microtubules and dynein for anterior localization in the *Drosophila* oocyte. *PLoS Pathog.* 1: e14.
- Fritsche TR, Horn M, Seyedirashti S, Gautom RK, Schleifer KH, Wagner M, 1999. In situ detection of novel bacterial endosymbionts of *Acanthamoeba* spp. phylogenetically related to members of the order Rickettsiales. *Appl. Environ. Microbiol.* 65: 206–212. 336
- Gómez-Valero L, Soriano-Navarro M, Pérez-Brocal V, Heddi A, Moya A, García-Verdugo JM, Latorre A, 2004. Coexistence of *Wolbachia* with *Buchnera*

- aphidicola and a secondary 339 symbiont in the aphid *Cinara cedri*. *J. Bacteriol.* 186: 6626–6633.
- Gonella E, Negri I, Marzorati M, Mandrioli M, Sacchi L, Pajoro M, Crotti E, Clementi E, Tedeschi R, Bandi C, Alma A, Daffonchio D, 2011. Bacterial endosymbiont localization in *Hyalesthes obsoletus*, the insect vector of Bois Noir in *Vitis vinifera*. *Appl. Environ. Microbiol.* 77: 1423–435.
- Gottlieb Y, Ghanim M, Chiel E, Gerling D, Portnoy V, Steinberg S, Tzuri G, Horowitz AR, Belausov E, Mozes-Daube N, Kontsedalov S, Gershon M, Gal S, Katzir N, Zchori-Fein E, 2006. Identification and localization of a *Rickettsia* sp. in *Bemisia tabaci* (Homoptera: Aleyrodidae). *Appl. Environ. Microbiol.* 72: 3646–3652.
- Hammer B, Moter A, Kahl O, Alberti G, Göbel UB, 2001. Visualization of *Borrelia burgdorferi* sensu lato by fluorescence in situ hybridization (FISH) on wholebody sections of *Ixodes ricinus* ticks and gerbil skin biopsies. *Microbiology* 147: 1425–1436.
- Kiszewski AE, Matuschka FR, Spielman A, 2001. Mating strategies and spermiogenesis in ixodid ticks. *Annu. Rev. Entomol.* 46: 167–182.
- 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.
- Lo N, Beninati T, Sassera D, Bouman EA, Santagati S, Gern L, Sambri V, Masuzawa T, Gray JS, Jaenson TG, Bouattour A, Kenny MJ, Guner ES, Kharitonov IG, Bitam I, Bandi C, 2006. Widespread distribution and high prevalence of an alpha-proteobacterial symbiont in the tick *Ixodes ricinus*. *Environ. Microbiol.* 8: 1280–1287.
- Manilla G, 1998. *Acari Ixodida*. Ed. Calderini Bologna.
- Moter A, Göbel UB, 2000. Fluorescence in situ hybridization (FISH) for direct visualization of microorganisms. *J. Microbiol. Methods* 41: 85–112.
- Nielsen AT, Liu WT, Filipe C, Grady L, Molin S, Stahl DA, 1999. Identification of a novel group of bacteria in sludge from a deteriorated biological phosphorus removal reactor. *Appl. Environ. Microbiol.* 65: 1251–1258.
- Parola P, Raoult D, 2001. Ticks and tickborne bacterial diseases in humans: an emerging infectious threat. *Clin. Infect. Dis.* 32: 897–928.

- 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.
- Sassera D, Lo N, Bouman EA, 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.
- Skaljac M, Zanic K, Ban SG, Kontsedalov S, Ghanim M, 2010. Co-infection and localization of secondary symbionts in two whitefly species. *BMC Microbiol.* 10: 142. 380
- Socolovschi C, Mediannikov O, Raoult D, Parola P, 2009. The relationship between spotted fever group *Rickettsiae* and ixodid ticks. *Vet. Res.* 40: 34.
- Sonenshine DE, 1993. *Biology of Ticks*. Oxford University Press, USA.

2.7 Research article. *Multiple independent data reveals an unusual response to Pleistocene climatic changes in the hard tick Ixodes ricinus*

2.7.1 Summary

In the last few years, improved analytical tools and the integration of genetic data with multiple sources of information have shown that temperate species exhibited more complex responses to ice ages than previously thought. In this study, we investigated how Pleistocene climatic changes affected the current distribution and genetic diversity of European populations of the tick *Ixodes ricinus*, an ectoparasite with high ecological plasticity. We first used mitochondrial and nuclear genetic markers to investigate the phylogeographic structure of the species and its Pleistocene history using coalescent-based methods; then we used species distribution modelling to infer the climatic niche of the species at Last Glacial Maximum; finally, we reviewed the literature on the *I. ricinus* hosts to identify the locations of their glacial refugia. Our results support the scenario that during the last glacial phase *I. ricinus* never experienced a prolonged allopatric divergence in separate glacial refugia, but persisted with interconnected populations across Southern and Central Europe. The generalist behavior in host choice of *I. ricinus* would have played a major role in maintaining connections between its populations. Although most of the hosts persisted in separate refugia, from the point of view of *I. ricinus* they represented a continuity of “bridges” among populations. Our study highlights the importance of species-specific ecology in affecting responses to Pleistocene glacial–interglacial cycles. Together with other cases in Europe and elsewhere, it contributes to setting new hypotheses on how species with wide ecological plasticity coped with Pleistocene climatic changes.

2.7.2 Manuscript

Introduction

During recent decades, phylogeographic studies have greatly improved our knowledge of where temperate species persisted during glacial phases (Holderegger

and Thiel-Egenter 2009) and how they re-colonized northern regions during post-glacial phases (Beheregaray 2008). In the Western Palaearctic, early phylogeographic studies identified the Mediterranean peninsulas of Iberia, Italy, and the Balkans as major regions of glacial refugia for most taxa, and also shed light on some common patterns of post-glacial re-colonization (Hewitt 1996, 2000; Taberlet *et al.* 1998; Schmitt 2007). In recent years, however, several departures from this 'southerly refugia model' have emerged that have highlighted the occurrence of more complex scenarios and the possibility of a wider range of species responses to ice ages than previously thought (Gómez and Lunt 2007; Valdiosera *et al.* 2007; Beheregaray 2008; Rull 2009; Varga 2009; Stewart *et al.* 2010; Bisconti *et al.* 2011; Hewitt 2011; Porretta *et al.* 2011; Teacher *et al.* 2011).

Factors that have improved our ability to detect biogeographic patterns and infer the processes underlying their origin include the development and wider use of analytical tools to examine different evolutionary scenarios (Knowles 2009; Hickerson *et al.* 2010); the adoption of appropriate sampling strategies (Feliner 2011); and the integration of multiple sources of information that allows the inferences from one data set to be investigated and potentially corroborated by another (Keppel *et al.* 2012). With respect to this latter issue, phylogeographic studies have been integrated with species distribution modelling (SDM) to identify putative locations of Pleistocenic refugia or to provide information about past dispersal corridors (Richards *et al.* 2007; Waltari *et al.* 2007; Svenning *et al.* 2011). In few other cases, information from genetic and/or SDM were integrated with morphological and ecological data (Schlick-Steiner *et al.* 2006; Bardy *et al.* 2010; Barata *et al.* 2012).

Here, we used multiple independent data sets to investigate the Pleistocene evolutionary history of a species, the hard tick *Ixodes ricinus* (Linnaeus 1758), that previous studies based on genetic data alone (McLain *et al.* 2001; Casati *et al.* 2008; Noureddine *et al.* 2010) have suggested to conform to the 'southerly refugia model' (Hewitt 1996, 2000).

Genetic studies on *I. ricinus* populations showed the occurrence of a genetic discontinuity between European and North African populations, while no phylogeographic structure was found across Europe (McLain *et al.* 2001; Casati *et al.* 2008; Noureddine *et al.* 2010). To explain the observed genetic pattern across Europe, a scenario of range contraction into one or more glacial refugia in southern peninsulas

and post glacial re-colonization was hypothesized (McLain *et al.* 2001; Nouredine *et al.* 2010). However, an alternative scenario could be hypothesized on the basis of the ecology of *I. ricinus*. This species is an ectoparasite with wide ecological plasticity. It is distributed in deciduous woodland and mixed forest throughout the Western Palaearctic, where it survives under various environmental conditions and considerable temperature variation (Fournier *et al.* 2000). Although a recent tendency towards a host specialisation has been proposed (Kempf *et al.* 2011), *I. ricinus* parasitizes a large variety of terrestrial vertebrates, including mammals, birds, and reptiles (Milne 1950; Sonenshine 1991, 1993; Manilla 1998). Accordingly, its fate has been not closely linked to that of a single host species, and during glacial periods it could have survived by exploiting alternative host species, even after the extinction of major original hosts (Poulin and Keeney 2008). For *I. ricinus*, therefore, a scenario of persistence across both Southern and Central Europe in interconnected populations can be hypothesized.

In this study, we aimed to distinguish between the two above scenarios by integrating multiple sets of information. First, we used mitochondrial and nuclear genetic markers to investigate the phylogeographic structure of the species and to infer its Pleistocene history. Bayesian phylogeographic approach (Lemey *et al.* 2009) was used to infer ancestral geographic range of the species, and MIGRATE-N (Beerli & Felsenstein 2001) was used to test different migration patterns among population groups. Second, we used species distribution modelling (SDM) to infer the climatic niche of the species under current and Last Glacial Maximum (LGM) conditions (Waltari *et al.* 2007; Kozak *et al.* 2008; Svenning *et al.* 2011). Third, we reviewed the literature on the Pleistocene history of *I. ricinus* hosts to identify the locations of their glacial refugia.

Under this framework, we used SDM to support the evolutionary scenarios hypothesized and to define the population groups used in the genetic analyses. The investigation of phylogeographic pattern and the use of coalescent-based methods as Bayesian phylogeography and MIGRATE-N allowed us to test whether genetic data conform to predictions of the two scenarios hypothesized. According to the 'southerly refugia model' we should observe genetic signatures of range contraction and re-colonization of northernmost regions (e.g. clines of genetic diversity from south to north, genetic lineages that diverged in allopatry, species ancestral range located in

southern regions, etc.). Likewise, SDM would show the occurrence of more suitable climatic conditions in southern European regions than in central regions. According to the second scenario, instead, a lack of the above genetic signatures would be expected, and SDM should show the occurrence of continuous suitable climatic conditions across Southern and Central Europe. Finally, we integrated the SDM and genetic data with information about host glacial refugia because the hosts are a key component of the ecological niche of *I. ricinus* (Sonenshine 1991, 1993). Therefore, their presence needs to be taken into account to assess the plausibility of any inferred evolutionary scenario.

Materials and Methods

- Genetic analyses -

A total of 210 individuals of *Ixodes ricinus* ticks (larvae, nymphs and adults) were collected by flagging on vegetation from 22 localities during 2008-2009 throughout the range of distribution of the species in Europe (Table 1, Fig. 1). They were stored in ethanol 100% and kept at 4 °C until morphological identification using dicotomic key (Manilla 1998), and genetic analyses. Total genomic DNA was extracted following the protocol of Collins *et al.* (1987) and used as template for Polymerase Chain Reaction (PCR) amplifications. Two mitochondrial gene fragments and two nuclear gene fragments already used in previous studies on *I. ricinus* populations (Casati *et al.* 2008; Nouredine *et al.* 2010) were amplified and sequenced: the mitochondrial genes *Cytochrome Oxidase I* (COI) and *Cytochrome Oxidase II* (COII) (in 210 individuals from 22 populations); an intronic region of the nuclear gene *Defensin* (in a subset of 70 individuals from 22 populations); an intronic region of the nuclear gene *TROSPA* (in a subset of 40 individuals from 14 populations) (Table 1 and 2).

Table 1. Geographic location for the 22 sampled populations of *Ixodes ricinus* and haplotypes/alleles observed in each sample at mitochondrial and nuclear genes. In brackets is shown the how many times the haplotype/allele was found in the population.

Locality	Lat.	Long.	mtDNA COI-COII	TROSPA	Defensin
Uppsala, SE	59°51'N	17° 38'E	h1; h2(4); h3; h4(2); h5(2); h6; h7; h8(2); h9; h10	T1; T2; T3(2)	D1(2); D2(3); D3
Kaunas, LTU	54°33'N	23°53'E	h11; h12; h13; h14; h15 h2(6); h5; h21; h30; h31; h32; h33; h34	-	D2; D4 (3); D5(2); D6; D7
Pribice, CZ	48°57'N	16° 33'E	h2(5); h21(2)	T1; T4(2); T5; T6; T7	-
Kosice, SVK	48°43'N	21° 15'E	h2(2); h3; h12; h70; h71; h72	T3(2)	D8; D9
Budapest, HUN	47°30'N	19° 02'E	h2(4); h73; h74(2); h75; h76	T1; T3; T8(2); T9; T10	D4; D7; D10; D11; D12; D13
Cluj-Napoca, ROU	46°46'N	23° 36'E	h2(2); h68; h69	T1; T7; T11(2)	D1; D2; D4(5); D14; D15; D16
Zagabria, HRV	45°49'N	15° 58'E	h2(3); h5; h35; h36	-	D4(2)
Monaco, DEU	48°8'N	11° 34'E	h2(6); h3; h5; h37; h38; h39; h40; h41; h42	-	D3; D4(4); D17
Zurigo, CHE	47°22'N	08°32'E	h2(2); h16; h17(2); h18; h19; h20; h21; h22	T3(2); T12; T13; T14; T15	D1; D4(3); D6; D10
Villecartier, FR	48°44'N	0° 43'W	h2(2); h5(2); h16; h21; h23; h24	T3(2); T13; T16; T17; T18; T19(2)	D2; D4(3); D6; D18; D19; D20
Gardouch, FR	43°23'N	01°41'W	h2; h5(3); h21(5); h25(2); h2(8); h16; h21; h43; h44; h45; h46; h47; h48	T1(2); T3(3); T7; T13; T20; T21; T22	D1; D2(2); D6; D20; D21
Galizia, ESP	42°46'N	07°31'W	h2(8); h3; h13; h21; h25; h40; h49; h50; h51	T1; T2; T7; T11; T19; T23	D1; D4(4); D10; D22; D23 D4(2); D6(2); D21(2); D24; D25
Domodossola, IT	46°07'N	08° 18'E	h2(3); h25; h33; h52; h53; h54	-	D4(3); D26
Varese, IT	48°49'N	08° 50'E	h2(3); h5; h21(2); h24; h55; h56; h57; h58; h59	T24; T25	D2; D4(3); D6; D7
Trento, IT	46°04'N	11°07'E	h2(8); h5; h21; h60; h61; h62	-	D4(4); D6; D18; D21; D24 D4(2); D6; D21; D27; D28; D29; D30
Parma, IT	44°44'N	10° 12'E	h2(5); h21; h63(2); h64; h65	T1; T3(2); T7; T16; T26; T27; T28(2); T29	D2(2); D4(4); D31(2); D36(2); D37(2)
Pistoia, IT	43°56'N	10° 55'E	-	-	D2; D4; D1; D11
Potenza, IT	40°38'N	15° 47'E	h2(4); h66	T11; T16	D21(2); D31(2)
Stilo, IT	38°28'N	16° 27'E	h2; h77; h78; h79; h80; h81; h82	-	D4(2); D34; D35; D36; D37
Barcellona, ESP	38°09'N	15° 13'E			
Corleone, IT	37°49'N	13° 18'E			
Silvri, TUR	39°06'N	29° 33'E		T3(2); T7; T28	D2; D4; D28; D38

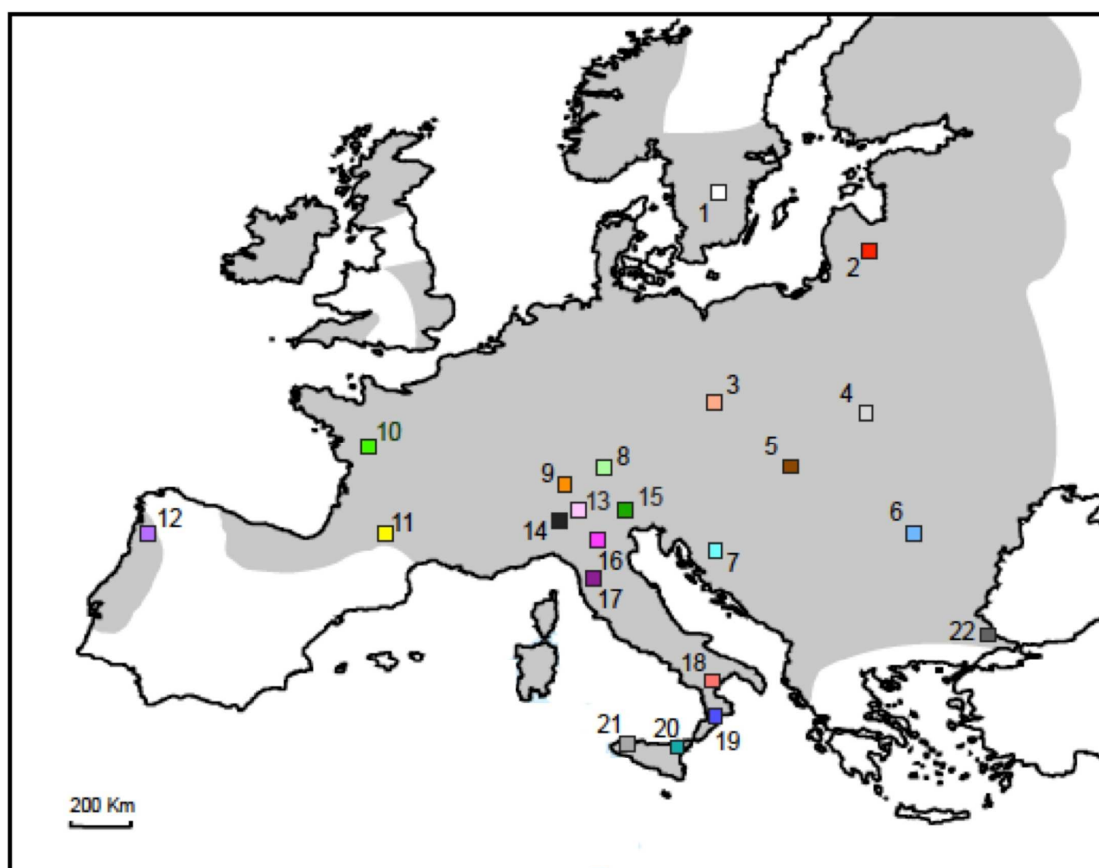


Figure 1. Map showing the sampling sites of *Ixodes ricinus*. The grey areas show the range of distribution of the species in the study area (Fournier *et al.* 2000). Each population is shown in a different colour. The detailed data of collection localities are reported in Table 1.

We decided to analyse a subset of individuals at *Defensin* and *TROSPA* genes on the basis of the concordance among the phylogeographic patterns observed between markers and among our results and those of previous studies on *I. ricinus* populations (Casati *et al.* 2008; Nouredine *et al.* 2010). The selection of the individuals of the subset was any way carried out to be representative of the entire study area.

PCR amplifications of mitochondrial gene fragments were carried out using the primer pairs C1-J-1484/C1-N-2678 which amplify part of COI gene and the pairs C1-J-2797/TK-N-3785 which amplify part of the COI and the COII gene (Casati *et al.* 2008). On the basis of the first sequences obtained, we subsequently designed a new specific primer pair for the COI gene, (z-coI-for 5'-TTTTAATTCTGAACTGAATTA

GGACAA-3' and z_coI_rev 5'-TCATCAATAAATCCTAAAAATCCAA-3'), which worked better than C1-J-1484/C1-N-2678 and was used to amplify all the other individuals analysed. The *Defensin* gene fragment was amplified using the primers Def-12-for 5'-ATGAAGGTCCTTGCCGTCTC-3' and Def-12-rev 5'-CAGCGATGTAGTGCCCATGT-3' (Noureddine *et al.* 2010). The *TROSPA* gene fragment was amplified using the primers Tro-for 5'-GCTACGGACACGGTGGTT-3' and Tro-rev 5'-TGGTTTCCCTTTGAGATG-3' designed by us on the complete sequence of the *I. ricinus* *TROSPA* gene available in GenBank (accession number: EU384705). PCR cycling procedure was: 95° C for 5 min followed by 34 cycles at 93° C for 1 min, 57° C, 52° C, 59° C and 55° C (respectively for the COI, COII, *Defensin* and *TROSPA* regions) for 1 min, 72° C for 1 min 30s, and a single final step at 72° C for 10 min. PCR sequences were obtained using ABI PRISM 3700 DNA sequencer by Macrogen Inc. All individuals were double sequenced using both forward and reverse primers to check for consistency. Sequences were edited and aligned using the software CHROMAS 2.31 and CLUSTAL X 2.0, respectively. All sequences were deposited in GenBank. For the nuclear genes, the heterozygous nucleotide positions were identified by double peaks in the electropherograms (Brumfield *et al.* 2003). We then used the Phase algorithm (Stephens and Donnelly 2003), implemented in DNASP v. 5 with default settings, to resolve haplotypes of heterozygous individuals. Between the haplotypes pairs reconstructed, we considered exclusively those with Bayesian posterior probabilities (BBP) value > 95%. The pairwise homoplasy index (PHI statistic, Bruen *et al.* 2006) implemented in SPLITSTREE v.4.11 was used to assess the possible occurrence of recombination for each nuclear gene.

Polymorphisms of both nucleotide and amino-acidic sequences were assessed by means of the software MEGA 5.0. The COI and COII genes were concatenated for all the subsequent analyses. Haplotype diversity (**h**) and nucleotide diversity (π), as defined by Nei (1987), at mitochondrial and nuclear genes were estimated for the overall dataset and for each population using the software ARLEQUIN 3.1. The genealogical relationships between the haplotypes found at mitochondrial and nuclear genes were investigated by constructing a phylogenetic network (Posada and Crandall 2001). We used the statistical parsimony algorithm described by Templeton *et al.* (1992) and implemented in the software TCS, applying a 95% cutoff for the probability of a parsimonious connection. The loops in the resulting phylogenetic networks were resolved by applying the criteria described by Pfenninger and Posada

(2002). To check for consistency of statistical parsimony networks, we also used the median-joining (MJ) network algorithm as implemented in the software NETWORK 4.6.1.0. The software JMODELTEST 0.1.1 was used to find the optimal model of sequence evolution for our data sets using the Akaike Information Criterion. The best-fit models were TrN+I with the proportion of invariable sites $I = 0.832$ for mitochondrial genes; HKY+I with $I = 0.697$ and $I = 0.768$ for the nuclear genes *Defensin* and *TROSPA*, respectively. To assess the existence of a significant phylogeographic structure, we first compared the differentiation indices G_{ST} (Nei 1987) and N_{ST} (Pons and Petit 1996; Grivet and Petit 2002) as implemented in the software PERMUT (<http://www.pierroton.inra.fr/genetics/labo/Software>). G_{ST} only considers haplotype frequencies, whereas N_{ST} considers both haplotype frequencies and their genetic divergence. The comparison of N_{ST} vs G_{ST} thus allows us to test for correspondence between haplotype phylogenies and their geographic distribution. A higher N_{ST} than G_{ST} indicates that the haplotypes that are co-occurring within populations are on average more similar to each other than random sets of haplotypes at the total species level, supporting the presence of a phylogeographic structure (Pons and Petit 1996). Second, population genetic structure was investigated by spatial analysis of molecular variance (SAMOVA) using the software SAMOVA. It uses a simulated annealing procedure to define, without a priori hypotheses of the expected structure, K groups of populations that are genetically homogenous (among population differentiation index, F_{SC} , minimized) and maximally differentiated from each other (among groups differentiation index, F_{CT} , maximized). We tested K values ranging from 2 to 21, 21 and 13 respectively for mitochondrial COI-COII, *Defensin* and *TROSPA* genes. To check for consistency we conducted each analysis starting from five different initial conditions, running the algorithm for 10,000 steps. To discriminate between the evolutionary scenarios hypothesized we used three different methods. First, we carried out Approximate Bayesian Computation (ABC) to explicitly analyzing several demographic scenarios. However, this method did not allow us to discriminate among the tested models, as the data did not provide strong support for any of them (see ABC analysis). Second, to infer ancestral geographic range of *I. ricinus* we used the recently developed Bayesian phylogeographic approach based on a discrete-states diffusion model (Lemey et al. 2009) as implemented in BEAST v1.6.1. The analysis was conducted for mtDNA, *Defensin* and *TROSPA* genes separately. For each sequence we assigned a character state that

reflects its geographic location in one of the four geographic areas defined as follow: (1) Iberian Peninsula (locality 12, Table 1); (2) France (localities 10, 11); (3) Central-Eastern Europe (localities 3-6, 8,9); (4) Italian Peninsula, Balkans and Southeast-Europe (localities 7, 13-22). This grouping scheme is based on the classic southern refugia (Schmitt 2007), adjusted for *I. ricinus* on the basis of spatial proximity of the sampled localities and the SDM results. The Italian Peninsula and the Balkans (group four) were indeed connected at LGM, and SDM showed the occurrence of a continuous suitable area between them and Southeast-Europe (Fig. 3B-C). The Sweden and Latvian localities (1 and 2, Table 1) were excluded by the analysis because they are outside the range of the inferred species paleodistribution and they were covered by the ice sheet at LGM (see Fig. 3). Following explorative analyses, MCMC (Markov Chains Monte Carlo) were run for 40 million generations, sampling every 1000th generation, with a coalescent-constant size tree prior, random starting tree, uncorrelated lognormal relaxed molecular clock, substitution rate of 0.001 and 0.01 substitutions/site/lineage/Myr for nuclear and mtDNA markers (Powell *et al.* 1986), respectively, and default prior distributions on other model parameters. Models of sequence evolution for each gene were determined using JModelTest as described above. Convergence was assessed using Tracer 1.5 (part of BEAST 1.6.1) and two runs were conducted for each gene and combined using LogCombiner v1.5.3 with a burn-in of 10%. All parameter estimates showed Effective Sample Size (ESS) values higher than 400. Maximum clade credibility consensus (MCC) trees were made using TreeAnnotator v1.6.1 with a burn-in of 25%, and visualized using FigTree v1.3.1. Third, the coalescence-based method implemented in MIGRATE-N 3.2.6 was used to compare alternative migration pattern between southern and northern regions. It allows to infer the scaled mutation rate theta ($\theta = xN_e\mu$) for all populations and the migration rates M (m/μ) (mutational-scaled immigration rate) between them (Beerli and Felsenstein 2001). The populations were grouped as described above for the Bayesian phylogeography analysis. In particular, we compared a source-sink migration matrix and a full migration matrix (Fig. 2B). In the first model, migration was unidirectional, from southern to northern regions. In this migration model, southern populations acted as a source for northern populations, which originated as a consequence of re-colonisation from southern regions. In the second model, migration was bidirectional among all populations groups. According to a scenario of contraction and re-colonisation, we expect that the northern populations originated

from the re-colonisation of the southern regions; therefore, the source-sink matrix model is probably the most likely scenario. Each model was run using random genealogy and values of the parameters θ and M generated by F_{ST} calculation as start condition. Bayesian search strategy was conducted using the following parameters: 5×10^4 recorded steps, increment per step = 100, number of replicates = 5, burn-in = 5×10^5 . The prior distribution for the parameters was uniform with boundaries defined after explorative runs. A static heating scheme with four different temperatures (1.00; 1.50; 3.00; 100,000) was employed, where acceptance–rejection swaps were proposed every step. The model comparison was done by means of the Bayes Factor (BF). Models were ordered by BF, and the model probability was calculated by following the procedure described in the MIGRATE-N Tutorial (https://molevol.mbl.edu/wiki/index.php/Migrate_tutorial).

- *Species Distribution Modelling* -

Species Distribution Models (SDMs) were constructed using current climatic conditions and those of the Last Glacial Maximum (LGM) through the “maximum entropy” approach as implemented in the software package MAXENT 3.3.3 (Phillips *et al.* 2006; Phillips and Dudík 2008). MAXENT uses occurrence-only data in conjunction with environmental data to estimate the probability of species occurrence on the basis of a uniform probability distribution (maximum entropy) and on the presence data provided by the user. *I. ricinus* occurrence points in the study area were taken from records in scientific literature and from the online database “Global Biodiversity Information Facility” (<http://data.gbif.269.org>). We used all georeferenced localities. For occurrence points without geographical coordinates, we chose only those in which the locality has clearly been stated. These are then georeferenced using the software Google Earth. A total of 191 occurrence points was obtained to be used for SDM construction. Nineteen bioclimatic variables with a resolution of 2.5 min (~ 5 km) were downloaded from the WorldClim database (www.worldclim.org). To assess the possible correlation among these variables, we performed pairwise correlation comparisons between them using the Pearson’s correlation coefficient. For pairs that were highly correlated (correlation coefficient ≥ 0.75) we chose the variable which is most biologically meaningful for the species (Sonenshine 1991, 1993). Finally, to construct the climatic niche model, five variables were chosen: BIO4 = Temperature Seasonality (Coefficient of Variation); BIO5 =

Max Temperature of Warmest Period; BIO13 = Precipitation of Wettest Period; BIO15 = Precipitation Seasonality (Coefficient of Variation); BIO17 = Precipitation of Driest Quarter. We used the default values of all parameters with the exception of the regularization parameter beta (β). As suggested by Warren & Seifert (2011), we ran the model using different values of β (1, 3, 5, 7, 9, 11, 13, 15, 17 and 19), then we chose the best one using the corrected Akaike Information Criterion (AICc) scores, as implemented in the software ENMTools. To obtain accurate prediction we run the model making 10 replicates under the cross-validation form of replication. This approach randomly split the data into equal-size groups (“folds”) and creates models leaving out each fold in turn and using them for evaluation (Phillips and Dudík 2008). Finally, 75% of the localities were used to build the model and 25% were randomly selected to test it.

To construct niche model for LGM conditions, we projected the current species’ bioclimatic niche onto past climate layers, downloaded from the WorldClim database at a resolution of 2.5 min as in the analysis above. We used two models: the Community Climate System Model (CCSM 3; Collins *et al.* 2004), and the Model for Interdisciplinary Research on Climate (MIROC Version 3.2; Hasumi & Emori 2004). Both models were run in MAXENT using the settings chosen for current conditions. The area under receiver operating characteristic (ROC) curve (AUC) was then used to evaluate the model performance. An AUC score above 0.75 is considered good model performance (Elith 2002). All SDM predictions were then visualized in QUANTUM GIS 1.0.2 (<http://www.qgis.org>).

- *Glacial refugia of Ixodes ricinus’ hosts* -

We reviewed the literature about the location of glacial refugia of *I. ricinus* hosts. *I. ricinus* exploits a huge number of different vertebrate host species, including mammals, birds and, to a lesser extent, reptiles (Sonenshine 1993; Manilla 1998; Gern and Humair 2002). We based our review on a total of 156 species (65 mammals, 86 birds, 5 reptiles) recognized as *I. ricinus* hosts (Nosek *et al.* 1967; Nosek & Sixl 1972; Gern 2009; Vichová *et al.* 2010; Santos-Silva *et al.* 2011). For each species we searched in the bibliographic database Google Scholar (www.scholar.com) and in the database for peer-reviewed literature, Scopus (www.scopus.com). We used the keywords: “scientific name of the species, ice age, glacial refugia”. When the common name of the species was available, we included it in the keywords. For the

Scopus database we chose the options: search in “all fields”; document type, “all”; data range, “published from all years to present”; subject areas, “life sciences”. The search was updated on the 1st December 2011.

Results

- Genetic data -

For the mitochondrial DNA marker, an alignment of 1476 base pairs (bp) was obtained by concatenating the COI and COII sequences (respectively 822 and 654 base pairs). Eighty-two haplotypes were found in 210 individuals analysed, identified by 92 nucleotide substitutions (78 transitions and 14 transversions; 74 synonymous and 18 non-synonymous), 43 of which were parsimony informative. Mean haplotype diversity (h) and nucleotide diversity (π) estimates were respectively 0.834 (SD ± 0.026) and 0.0032 (SD ± 0.002), while h and π estimates for each locality are shown in Table 2. For the nuclear markers, 140 sequences of 288 bp were obtained for the *Defensin* gene, and 80 sequences of 479 bp were obtained for the *TROSPA* gene. The Phi test did not find evidence for recombination neither in *Defensin* ($P = 0.082$) nor in *TROSPA* ($P = 0.085$). For *Defensin* gene, 38 haplotypes were found, identified by 34 nucleotide substitutions (22 transitions and 12 transversions). For *TROSPA* gene, 29 haplotypes were found, identified by 29 nucleotide substitutions (17 transitions and 12 transversions). Mean haplotype diversity and nucleotide diversity estimates were respectively: 0.838 (0.025) and 0.0032 (0.002) for *Defensin*; 0.912 (0.021) and 0.0081 (0.005) for *TROSPA*. In Table 2 the h and π estimates are shown for each locality.

Table 2 Genetic diversity estimates at mitochondrial and nuclear markers of the *Ixodes ricinus* sampled populations. h haplotype diversity (\pm SD); π nucleotide diversity (\pm SD); N = number of sequences for each population.

Locality	mtDNA			Defensin			TROSPA		
	N	h	π	N	h	π	N	h	π
Uppsala,SE	16	0.925(0.047)	0.0050(0.0028)	6	0.733(0.155)	0.0106(0.007)	4	0.833(0.222)	0.0066(0.005)
Kaunas,LTU	5	1.000(0.127)	0.0064(0.0041)	8	0.857(0.108)	0.0056(0.004)	-	-	-
Pribice,CZ	13	0.808(0.113)	0.0025(0.0015)	4	0.667(0.204)	0.0023(0.003)	6	0.933(0.122)	0.0124(0.008)
Kosice,SVK	7	0.476(0.171)	0.0009(0.0007)	4	0.833(0.222)	0.0080(0.006)	4	0.833(0.222)	0.0084(0.063)
Budapest,HUN	7	0.952(0.095)	0.0012(0.0008)	6	1000(0.096)	0.0118(0.008)	6	0.933(0.121)	0.0153(0.009)
Cluj-Napoca,ROU	9	0.806(0.120)	0.0017(0.0011)	10	0.778(0.137)	0.0089(0.006)	4	0.833(0.222)	0.0042(0.003)
Zagabria,HRV	4	0.833(0.222)	0.0017(0.0014)	4	0.667(0.204)	0.0023(0.003)	4	0.833(0.222)	0.0084(0.063)
Monaco,DEU	6	0.800(0.172)	0.0043(0.0027)	6	0.600(0.215)	0.0035(0.003)	-	-	-
Zurigo,CHE	14	0.835(0.101)	0.0023(0.0014)	6	0.800(0.172)	0.0062(0.005)	6	0.933(0.122)	0.0081(0.005)
Villecartier,FR	10	0.956(0.059)	0.0059(0.0033)	8	0.893(0.111)	0.0067(0.005)	8	0.928(0.084)	0.0085(0.005)
Gardouch,FR	8	0.929(0.084)	0.0058(0.0034)	6	0.933(0.121)	0.0118(0.008)	10	0.911(0.077)	0.0076(0.005)
Galizia,ESP	15	0.867(0.067)	0.0053(0.0029)	8	0.786(0.151)	0.0118(0.008)	6	1.000(0.096)	0.0065(0.004)
Domodossola,IT	16	0.767(0.113)	0.0018(0.0012)	8	0.893(0.086)	0.0071(0.005)	-	-	-
Varese,IT	16	0.767(0.113)	0.0022(0.0013)	4	0.500(0.265)	0.0069(0.006)	4	0.833(0.222)	0.0077(0.005)
Trento,IT	8	0.893(0.111)	0.0034(0.0021)	6	0.800(0.172)	0.0035(0.003)	-	-	-
Parma,IT	12	0.939(0.058)	0.0046(0.0026)	8	0.786(0.151)	0.0052(0.004)	-	-	-
Pistoia,IT	13	0.641(0.150)	0.0027(0.0016)	8	0.964(0.077)	0.0058(0.004)	-	-	-
Potenza,IT	10	0.756(0.130)	0.0021(0.0013)	10	0.800(0.089)	0.0062(0.004)	10	0.956(0.060)	0.0092(0.006)
Stilo,IT	3	-	-	6	0.933(0.122)	0.0183(0.001)	-	-	-
Barcelona,ESP	5	0.400(0.237)	0.0003(0.0003)	4	0.667(0.204)	0.0046(0.004)	4	0.833(0.222)	0.0063(0.005)
Corleone,IT	6	0.600(0.215)	0.0007(0.0006)	6	0.933(0.122)	0.0108(0.007)	-	-	-
Silvri,TUR	7	1000(0.076)	0.0024(0.0016)	4	1.000(0.177)	0.0046(0.004)	4	0.833(0.222)	0.0094(0.007)

The geographical distribution of mitochondrial and nuclear haplotypes is presented in Table 1. Of the 82 mtDNA haplotypes found, 12 were shared by two or more populations and the remaining were private or unique haplotypes. Three haplotypes (h2 and h5 and h17) accounted for over 50% of the individuals examined (39%, 5.8% and 7.7% respectively), and were the most widespread, being found in populations from Western to Eastern and from Southern to Northern Europe. In particular, the haplotype h2 was found in 21 out of the 22 localities studied for mtDNA markers. For the nuclear gene encoding for *Defensin*, of the 38 haplotypes found 14 were shared by two or more populations. The haplotype D4 was the most frequent, being found in 51 out of the 140 sequences obtained (36%) and the most geographically widespread, being found in 19 out of the 22 populations studied at this locus. Also the haplotypes D1, D2, D6 and D21 were widely shared among populations across all the study area (Table 1). For the nuclear gene encoding for *TROSPA*, of the 29 haplotypes found, 9 were shared by two or more populations. The haplotype T3 was the most frequent being found in 20 out of the 80 sequences obtained (25%). It was, as the haplotypes

T1, T7, T11 and T13, geographically spread across all the study area (Fig. S2C, Table 1). The parsimony networks showing the genealogical relationships among haplotypes found at mitochondrial and nuclear genes are presented in Fig. S2. No differences were found between the topologies of parsimony and MJ networks (data not shown). The sequence divergence between mitochondrial haplotypes ranged from 0.07 to 1.29% (overall mean sequence divergence 0.44%). For Defensin, the sequence divergence between haplotypes ranged from 0.35 to 4.17% (overall mean sequence divergence 1.46%). For TROSPA, the sequence divergence between haplotypes ranged from 0.21 to 2.71% (overall mean sequence divergence 1.15%). The estimates of the **NST** and **GST** differentiation indices were respectively 0.040 and 0.031 for mitochondrial COI-COII genes, 0.027 and 0.016 for TROSPA and 0.055 and 0.038 for Defensin genes. No significant differences were found between the **NST** and **GST** for any of the three markers (all **P** values > 0.05). Spatial analysis of molecular variance (SAMOVA), using mitochondrial and nuclear markers, revealed no groups of genetically distinct populations. For mtDNA the **FCT** values showed a very narrow range, and the highest value (**FCT** =0.154) was observed for both **K**=3 and **K**=21, when all populations were separated (Fig. S3). A very narrow range of values was observed also for Defensin and TROSPA genes, and the highest values were observed when all populations were separated (Fig. S3). The majority of variation was always found within populations (between 80 and 90%, data not shown). All repetitions conducted (see Materials and Methods Section) provided very similar values of the fixation indices, so we showed only results of the first replicate. The results of Bayesian phylogeographic analyses to infer ancestral geographic range of *I. ricinus* are shown in Fig. 2A. For all markers no single geographic area was found with significantly higher root posterior probability than the others. Indeed, the root posterior probability values showed a very narrow range and were comprised between 0,217 and 0,280 for mtDNA (Iberian Peninsula and Central-Eastern Europe, respectively), between 0,240 and 0,262 for Defensin (Iberian Peninsula and Italian Peninsula-Balkans-Southeast Europe, respectively), and between 0,247 and 0,252 for TROSPA (Italian Peninsula and Central-Eastern Europe, respectively) (Fig. 2A). Maximum Clade Credibility (MCC) trees are available under request to the authors. In Fig. 2B we showed the population grouping schemes and the migration models tested using MIGRATE-N. the model comparison highly supported (**P** = 1.0) Model B (i.e. bidirectional migration between all groups).

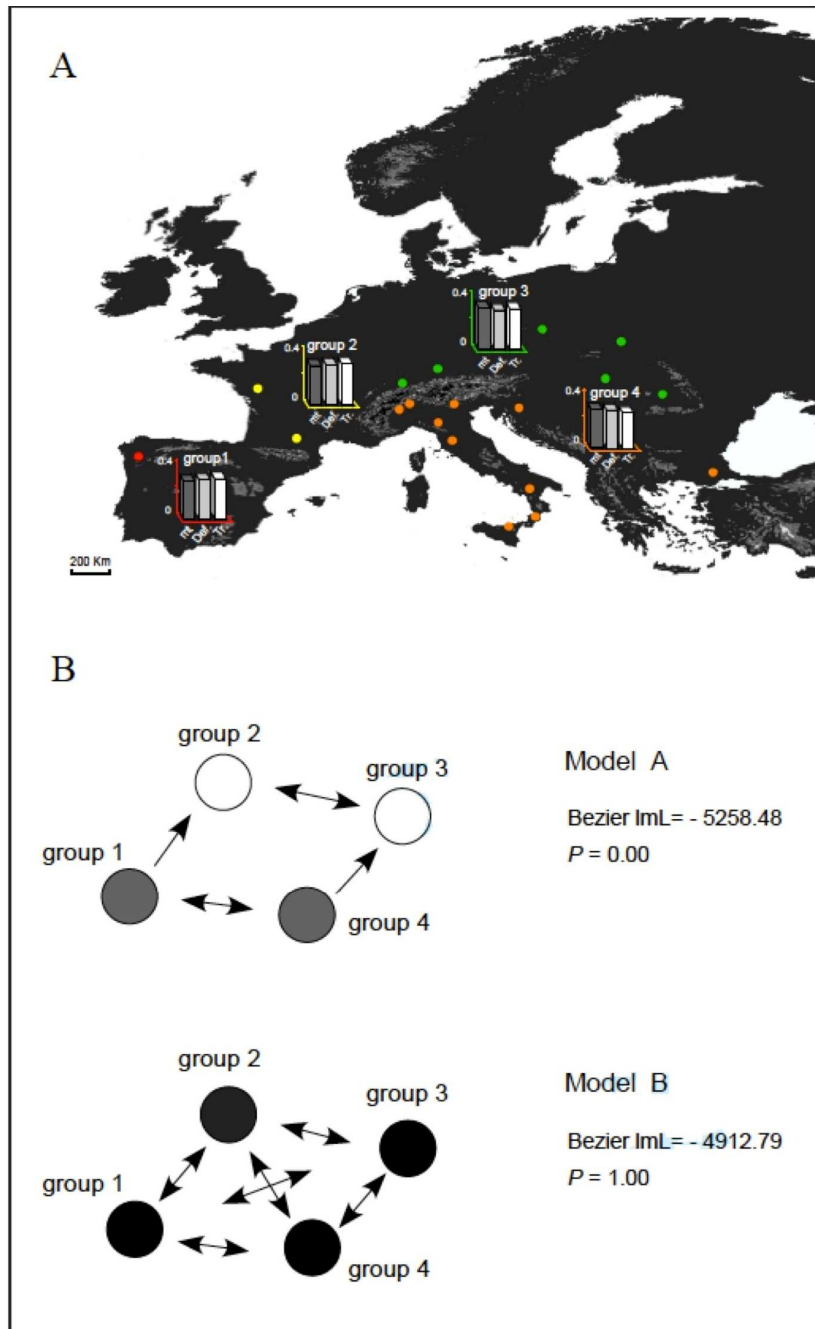


Figure 2. (A) Bayesian phylogeography: root posterior probability values for each geographic area in which sampling localities have been grouped (see Materials and Methods for population grouping). Red dot = group 1; yellow dots = group 2; green dots = group 3; orange dots = group 4. The histograms in the graphs show the root posterior probability values for mtDNA (dark grey), Defensin (light grey) and TROSPA (white) markers. (B) Migration models tested using MIGRATE-N. Model-A: southern populations acted as a source for northern populations. We left the possibility to exchange migrants between the southern groups on the basis of the Species Distribution Modelling Results (see 553 Fig. 3B-C). Grey circles denote the source populations; white circles denote sink populations; Model-B: the migration is bidirectional between all populations groups. Arrows indicate the assumed

directions of gene flow. Bezier ImL = Bezier logmarginal Likelihood; P = modelprobability.

- *Species Distribution Modelling* -

The model built with the regularization parameter $\beta=1$ received the lowest value of AICc, outperforming the remaining nine models. The average value of AUC for the replicate runs is 0.809 (SD ± 0.019), indicating a good performance of the model. The averaged distributions for current and LGM conditions predicted by this model are thus represented in Fig. 3A-C. The potential distribution for the present range of *I. ricinus*, predicted by MAXENT is very similar to the actually known geographic distribution of the species (Fig. 1). Fig. 3B-C showed instead the potential distribution of *I. ricinus* predicted by CCSM and MIROC models during the Last Glacial Maximum. Concordantly both CCSM and MIROC models showed the occurrence of suitable climatic conditions not only in the three Mediterranean Peninsulas (Spain, I and Balkans), but also in Central European regions (42-50°Latitude Nord). Interestingly, as shown in Fig. 3B-C, all these suitable areas were not isolated each other, but connected by land bridges, that emerged following the glacial-induced marine regression (i.e. the Po Plain that connected I and Balkans). Both models, indeed, predicted these land bridges as well as the emerged corridors (between Iberian Peninsula, France and I; or between Iberian Peninsula, France, Ireland and United Kingdom) being climatic suitable for *I. ricinus*. Therefore, on the whole, SDMs reconstructed for the LGM showed the occurrence of a continuous climatically suitable area across the European continent, characterized by a suitability that never fell below 40% (Fig. 3B-C).

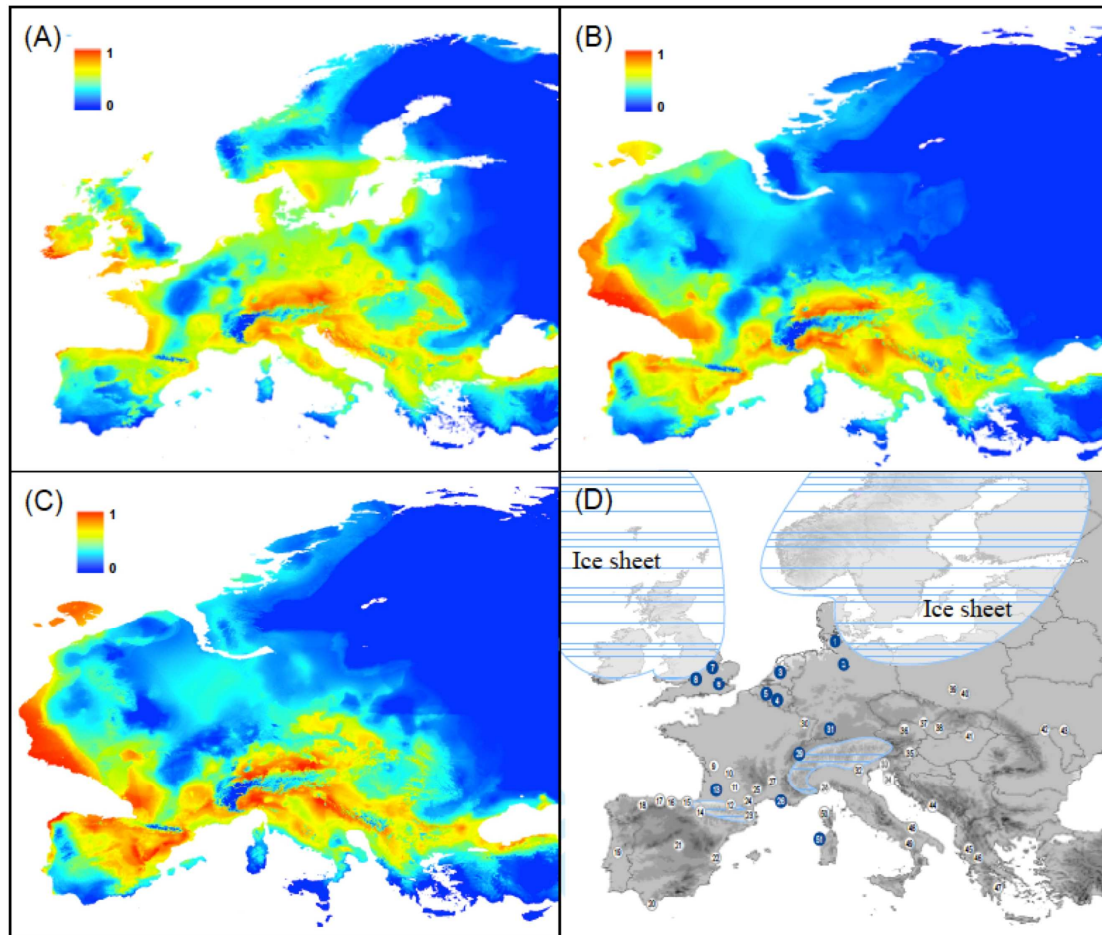


Figure 3. Bioclimatic models for *Ixodes ricinus* in the study area. Models estimated for present-day conditions (A) and for the Last Glacial Maximum, based on both CCSM Model (B) and MIROC model (C). Warmer colours show areas with predicted conditions more suitable for *I. ricinus*. (D) Fossil records of the *I. ricinus* vertebrate hosts. White circles identify fossil localities dated at LGM (21,000-18,000 BP); blue circles identify fossil localities dated between 75,000 and 15,000 years BP. For details see Appendix 1.

- Glacial refugia of *Ixodes ricinus*' hosts -

Among the 156 *I. ricinus* host species that we used in our search in the databases, we found literature records about glacial refugia for 44 species (21 mammals, 20 birds, 3 reptiles) (Appendix). In these studies, the areas of putative glacial refugia were inferred on the basis of fossil records, and/or genetic data, and/or species distribution modelling data (Appendix). In some cases the location of refugia was stated explicitly (i.e. Italo-Balkan area, see for example *Apodemus flavicollis*), while in others the locations of glacial refugia were generic (i.e. Eastern, Western Europe, see *Clethrionomys glareolus*) (Appendix). For this reason we classified the location of

refugia as southern (S) and northern (N). S denotes survival in one or more southern peninsulas (Iberian, Italian and Balkan Peninsulas), N denotes persistence in refugia outside the southern peninsulas. As shown in Appendix, for some species glacial refugia were inferred only in southern peninsulas, for others only northern refugia were inferred; for some other species both southern and northern refugia were inferred. To show graphically a general picture of where *I. ricinus* hosts occurred during the Last Glacial Maximum, we mapped the fossil records in Fig. 3D.

Discussion

The integration of multiple sources of information may greatly improve our ability to detect biogeographic patterns and to infer the processes underlying their origin (Richards *et al.* 2007; Waltari *et al.* 2007; Svenning *et al.* 2011). In this paper, we investigated the Pleistocene evolutionary history of European populations of the hard tick *Ixodes ricinus*, that previous studies based on genetic data alone have suggested to conform to the “southerly refugia model” (McLain *et al.* 2001; Casati *et al.* 2008; Nouredine *et al.* 2010). The integration of genetic data with species distribution modelling and with data about glacial refugia of *I. ricinus* hosts, allowed us to support a different scenario that previously hypothesized and that is unusual for Western Palearctic region.

We observed a lack of phylogeographic structure throughout the study area in all the studied markers. Indeed, no evident association between haplotypes and geography was observed (Table 1), and, as would be expected if a phylogeographic structure was present, no significantly higher NST values than GST values were found. Likewise, the spatial analysis of population structure carried out with SAMOVA identified no grouping of populations that could explain data appreciably better than others. This pattern agrees with the observations of McLain *et al.* (2001), Casati *et al.* (2008), and more recently of Nouredine *et al.* (2010). Among the possible factors underlying the lack of phylogeographic structure, historical factors linked to Pleistocene ice-age cycles were suggested (McLain *et al.* 2001; Casati *et al.* 2008; Nouredine *et al.* 2010). Since a part of the current distribution range of the species was covered by glaciers during the glacial phases, a scenario of contraction into one or more glacial refugia in southern peninsulas and post glacial re-colonization was hypothesized (McLain *et al.* 2001; Nouredine *et al.* 2010).

Our genetic data do not support the above hypotheses to explain the lack of phylogeographic structure observed in *I. ricinus*. Indeed, we observed neither hotspots/coldspots nor clines of genetic diversity from south to north (Table 2), as would be expected if a northward expansion from one or more southern peninsular refugia had occurred (Hewitt 1996, 2000). It might be possible that admixture between populations previously restricted in separate refugia has obscured prior allopatric fragmentation events. However, in this case, we would expect the existence of a genetic signature (i.e. the occurrence of diverging monophyletic lineages or secondary contact zones), which we did not observe (Table 1). On the contrary, the genetic pattern observed best fits with a scenario of persistence in both Southern and Central Europe in interconnected populations that never experienced a prolonged allopatric divergence in separate glacial refugia.

The Approximate Bayesian Computation (ABC) did not facilitate discrimination among the tested models. In contrast, the Bayesian phylogeographic analysis and MIGRATE-N allowed us to gain strong support for the scenario of persistence in interconnected populations. Bayesian phylogeographic analysis showed that southern regions do not have higher root posterior probability of being the ancestral range for the species than the Central European regions (Fig. 2A). Likewise, the comparison of source sink vs full migration matrices by MIGRATE-N supported a full migration matrix model, which is concordant with a scenario of persistence in both Southern and Central European regions.

When genetic data are integrated with species distribution modelling and the location of the host refugia, this scenario of persistence seems highly supported. SDMs concordantly predicted the occurrence of suitable climatic conditions for *I. ricinus* during the LGM in southern peninsulas, central regions and the lands that emerged during the glacial phase (Fig. 3B-C). The marine regression induced by glaciations led to the widening of most coastal plains along the Mediterranean Basin and the Atlantic coast (Thiede 1978; Amorosi *et al.* 2004), which acted as land bridges between the southern peninsulas and between these peninsulas and Western Europe (see Fig. 3B-C). One of the most impressive cases occurred in the northern part of the Adriatic Sea, where the widening of the Po Plain connected the Italian peninsula to the Balkans (Amorosi *et al.* 2004). The advances of coastlines also led to the formation of corridors between the Iberian Peninsula, Southern France, and I, and between the Iberian Peninsula, Northern France, Ireland, and the United Kingdom

(see Fig. 3B-C). The SDM projections for the LGM period showed that these corridors were climatically suitable for *I. ricinus* (Fig. 3B-C); therefore, there would have been a wide, continuous suitable area for *I. ricinus* that stretched from South to Central Europe (Fig. 3B-C).

Our review of the glacial refugia of *I. ricinus* hosts showed their occurrence in both Southern and Central Europe (Fig. 3D, and Appendix). The rates and patterns of gene flow and dispersal of *I. ricinus*, like those of ticks in general, are largely determined by host movements during infestation (Milne 1950; Gray 1985; Korch 1994; Carroll and Schmidtman 1996; Hasle *et al.* 2009). *I. ricinus* is an ectoparasite able to exploit a huge number of vertebrate species, including highly mobile mammals and birds (Milne 1950; Manilla 1988; Sonenshine 1991, 1993; Kempf *et al.* 2011; Pfäffle *et al.* 2011). All the three developmental stages of the tick (larva, nymph, and adult) are parasites and can feed on different hosts at both the individual and species level. In particular, smaller bodied hosts are exploited by larval and nymphal stages, whereas larger bodied animals can carry all life stages (Gern 2009). Not only could the hosts have facilitated the persistence of *I. ricinus* in both southern and central regions, but they could also have played a major role in maintaining interconnection between the populations. Some of the hosts, such as the brown bear and red fox, maintained continuous distributions with interconnected populations across Southern and Central Europe (Valdiosera *et al.* 2007; Teacher *et al.* 2011), while other hosts persisted in separate southern and/or central regions (Appendix). From the point of view of *I. ricinus*, however, the different hosts on the whole would have represented a continuity of “bridges” among populations. *I. ricinus* could have exploited the hosts not only to move within the suitable areas and to cross the land corridors between them, but also to cross the known phylogeographic barriers between Southern and Central Europe (i.e. the Alps and Pyrenees mountain chains). De Meeûs *et al.* (2002), using microsatellite markers, did not detect genetic differentiation between populations on either side of the Alps, and have highlighted the role played by hosts in crossing such geographic barriers. On the whole, the integration of multiple data sets supports the view that during the last glacial phase *I. ricinus*, (i) persisted in both Southern and Central Europe, and (ii) never experienced a prolonged allopatric divergence in separate glacial refugia, but instead has had a prolonged persistence with interconnected populations. The generalist behaviour in host choice of *I. ricinus* and the continuous availability of hosts across Europe would have played a major role

in maintaining the interconnectedness between tick populations.

Conclusions

Similar scenarios to that inferred for *I. ricinus* have recently emerged in Europe and elsewhere for other species that have wide ecological plasticity and high dispersal ability. For example, the brown bear *Ursus arctos* survived not only in the southern peninsulas, but also in mainland Europe, maintaining a continuous gene-flow between its populations (Valdiosera *et al.* 2007; Davison *et al.* 2011). Further, the red fox *Vulpes vulpes* (Teacher *et al.* 2011) and the mosquito *Aedes caspius* (Porretta *et al.* 2011) have recently been shown to not conform to a scenario of contraction and fragmentation during glacial phases, but instead remained interconnected throughout Europe. Furthermore, the wide adaptability and ecological flexibility of these species have been cited as the major features enabling them to persist in a continuous distribution.

Although these study cases must be regarded as exceptions to the general trend observed, they do deserve some attention. Indeed, there is a bias in phylogeographic studies towards species with low to moderate dispersal ability and/or restricted habitats (Hickerson *et al.* 2010). Furthermore, similar patterns have also recently been reported beyond the Western Palaearctic and in disparate taxa that share some life-history traits (i.e. high dispersal and/or wide ecological flexibility), including the frog *Litoria aurea* in Australia (Burns *et al.* 2007), the fruit bat *Cynopterus brachyotis* in South-East Asia (Campbell *et al.* 2006), and the tiger mosquito *Aedes albopictus* in East Asia (Porretta *et al.* 2012).

The increase in studies focused on species with such ecological characteristics will allow us to determine if the above exceptions could have a more general value. The *I. ricinus* case study contributes to this issue and highlights the importance of using multiple sets of information to disentangle different hypotheses. This is particularly necessary when no fossil records are available, as is true for most arthropod species.

2.7.3 References

- Amorosi A, Colalongo ML, Fiornini F et al. 2004. Palaeogeographic and palaeoclimatic evolution of the Po Plain from 150-ky core records. *Global and Planetary Change*, 40, 55–78.
- Banks WE, d'Errico F, Peterson AT. 2008. Reconstructing ecological niches and geographic distributions of caribou. *Rangifer tarandus*. and red deer. *Cervus elaphus*. during the Last Glacial Maximum. *Quaternary Science Reviews*, 27, 2568–2575.
- Barata M, Perera A, Martínez-Freiría F, Harris DJ. 2012. Cryptic diversity within the Moroccan endemic day geckos *Quedenfeldtia*. (Squamata: Gekkonidae): a multidisciplinary approach using genetic, morphological and ecological data. *Biological Journal of Linnean Society*, 106, 828–850.
- Bardy KE, Albach DC, Schneeweiss GM, Fischer MA, Schönswetter P. 2010. Disentangling phylogeography, polyploid evolution and taxonomy of a woodland herb. *Veronica chamaedrys* group, Plantaginaceae s.l. in southeastern Europe. *Molecular Phylogenetics and Evolution*, 57, 771–786.
- Beerli P, Felsenstein J. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences, USA*, 98, 4563–4568.
- Beheregaray LB. 2008. Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology*, 17, 3754–3774.
- Bilton DT, Mirol PM, Mascheretti S, Fredga K, Zima J, Searle JB. 1998. Mediterranean Europe as an area of endemism for small mammals rather than a source for northwards postglacial colonization. *Proceedings of the Royal Society of London B*, 265, 1219–1226.
- Bisconti R, Canestrelli D, Colangelo P, Nascetti G. 2011. Multiple lines of evidence for demographic and range expansion of a temperate species. *Hyla sarda*. during the last glaciation. *Molecular Ecology*, 20, 5313–5327.
- Bruen TC, Hervé P, Bryant D. 2006. A simple and robust statistical test for detecting the presence of recombination. *Genetics*, 172, 2665–2681.
- Brumfield RT, Beerli P, Nickerson DA, Edwards SV. 2003. The utility of single nucleotide polymorphisms in inferences of population history. *Trends in*

- Ecology and Evolution, 18, 249-256.
- Brunhoff C, Albreath KEG, Fedorov VB, Ook JAC, Jaarola A. 2003. Holarctic phylogeography of the root vole (*Microtus oeconomus*): implications for late Quaternary biogeography of high latitudes. *Molecular Ecology*, 12, 957–968.
- Burns EM, Eldridge MDB, Crayn DM, Houlden BA. 2007. Low phylogeographic structure in a wide spread endangered Australian frog (*Litoria aurea*, Anura: Hylidae). *Conservation Genetics*, 8, 17–32.
- Campbell P, Schneider CJ, Adnan AM, Zubaid A, Kunz TH. 2006. Comparative population structure of *Cynopterus* fruit bats in peninsular Malaysia and southern Thailand. *Molecular Ecology*, 15, 29–47.
- Carroll JF, Schmidtman ET. 1996. Dispersal of blacklegged tick. *Acari: Ixodidae*. at the woods–pasture interface. *Journal of Medical Entomology*, 33, 554–558.
- Casati S, Bernasconi MV, Gern L, Piffaretti JC. 2008. Assessment of intraspecific mtDNA variability of European *Ixodes ricinus sensu stricto*. *Acari: Ixodidae*). *Infection, Genetics and Evolution*, 8, 152–158.
- Collins FH, Mendez MA, Rasmussen MO, Mehaffey PC, Besansky NJ, Finnerty 609 V. 1987. A ribosomal RNA gene probe differentiates member species of the *Anopheles gambiae* complex. *American Journal of Tropical Medicine and Hygiene*, 37, 37-41.
- Collins WD, Blackmon M, Bitz C, et al. 2004. The community climate system model: CCSM3. *Journal of Climate*, 19, 2122–2143.
- Davison A, Birks JDS, Brookes RC, Braithwaite TC, Messenger JE., 2001. On the origin of faeces: morphological versus molecular methods for surveying rare carnivores from their scats. *Journal of Zoology*. (London), 257, 141-143.
- De Meeûs T, Beati L, Delaye C, Aeschlimann A, Renaud F. 2002. Sex-biased genetic structure in the vector of Lyme disease, *Ixodes ricinus*. *Evolution*, 56, 1802-1807.
- Deffontaine V, Libois R, Kotlik P et al. 2005. Beyond the Mediterranean peninsulas: evidence of central European glacial refugia for a temperate forest mammal species, the bank vole (*Clethrionomys glareolus*). *Molecular Ecology*, 14, 1727–1739.
- Elith J. 2002. Quantitative methods for modelling species habitat: comparative performance and an application to Australian plants. In: *Quantitative Methods for Conservation Biology*. eds. Ferson S, Burgman M), pp. 39-58. Springer,

New York.

- Feliner GN. 2011. Southern European glacial refugia: A tale of tales. *TAXON*, 60. 2), 365–372.
- Fløjgaard C, Normand S, Skov F, Svenning C. 2009. Ice age distributions of European small mammals: insights from species distribution modelling. *Journal of Biogeography*, 36, 1152-1163.
- Fournier PE, Grunnenberger F, Jaulhac B, Gastinger G, Raoult D. 2000. Evidence of *Rickettsia helvetica* Infection in Humans, Eastern France. *Emerging Infectious Diseases*, 6(4), 389-392.
- Gern L, Humair PF. 2002. Ecology of *Borrelia burgdorferi* sensu lato in Europe. In: *Lyme borreliosis: biology, epidemiology and control*. eds Gray JS, Kahl O, Lane RS, Stanek G), pp. 149–174. CAB International, Wallingford, Oxon, UK.
- Gern L. 2009. Life cycle of *Borrelia burgdorferi* sensu lato and transmission to humans. In: *Lyme borreliosis. Current Problem in Dermatology*,. eds Lipsker D, Jaulhac B), pp. 19-30. Karger, Basel.
- Godinho R, Crespo EG, Ferrand N, Harris DJ. 2005. Phylogeny and evolution of the green lizards, *Lacerta* spp. Squamata: Lacertidae. based on mitochondrial and nuclear DNA sequences. *Amphibia-Reptilia*, 26, 271–430.
- Gómez A, Lunt DH. 2007. Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: *Phylogeography in southern European refugia: evolutionary perspectives on the origins and conservation of European biodiversity*. eds Weiss S, Ferrand N), pp. 155–188. Springer Verlag, Dordrecht, The Netherlands.
- Gray JS. 1985. Studies on the larval activity of the tick *Ixodes ricinus* L. in Co. Wicklow, Ireland. *Experimental and Applied Acarology*, 1, 307–316.
- Grivet D, Petit RJ. 2002. Phylogeography of the common ivy. *Hedera* sp. in Europe: genetic differentiation through space and time. *Molecular Ecology*, 11, 1351–1362.
- Hasle G, Bjune G, Edvardsen E et al. 2009. Transport of ticks by migratory passerine birds to Norway. *Journal of Parasitology*, 95, 1342-1351.
- Hasumi H, Emori S. 2004. K-1 Coupled Model. MIROC. Description. K-1 technical report, Center for Climate System Research, University of Tokyo, 34 pp.
- Haynes S, Jaarola M, Searle JB. 2003. Phylogeography of the common vole. *Microtus*

- arvalis. with particular emphasis on the colonization of the Orkney archipelago. *Molecular Ecology*, 12, 951–956.
- Heckel G, Burri R, Fink S, Desmet J-F, Excoffier L. 2005. Genetic structure and colonization processes in European populations of the common vole *Microtus arvalis*. *Evolution*, 59, 2231–2242.
- Hewitt GM. 1996. Some genetic consequences of ice ages and their role in divergence and speciation. *Biological Journal of the Linnean Society*, 58, 247-276.
- Hewitt GM. 1999. Post-glacial recolonization of European Biota. *Biological Journal of the Linnean Society*, 68, 87–112.
- Hewitt GM. 2000. The genetic legacy of the Quaternary ice ages. *Nature*, 405, 907–913.
- Hewitt GM. 2011. Mediterranean peninsulas – the evolution of hotspots. In: *Biodiversity hotspots*. eds Zachos FE, Habel JC), pp. 123–148. Springer, Amsterdam.
- Hickerson MJ, Carstens BC, Cabender-Bares J et al. 2010. Phylogeography's past, present, and future: 10 years after Avise, 2000. *Molecular Phylogenetics and Evolution*, 54, 291–301.
- Holderegger R, Thiel-Egenter C. 2009. A discussion of different types of glacial refugia used in mountain biogeography and phylogeography. *Journal of Biogeography*, 36, 476–480.
- Hürner H, Krystufek B, Sarà M et al. 2010. Mitochondrial phylogeography of the edible dormouse. *Glis glis*. in the Western Palearctic Region. *Journal of Mammalogy*, 91, 233–242.
- Kempf F, De Meeûs T, Vaumourin E, et al. 2011. Host races in *Ixodes ricinus*, the European vector of Lyme borreliosis. *Infection, Genetics and Evolution*, 11, 2043–2048.
- Keppel G, Van NK, Wardell-Johnson GW et al. 2012. Refugia: identifying and understanding safe havens for biodiversity under climate change. *Global Ecology and Biogeography*, 21(4), 393-404.
- Knowles LL. 2009. Statistical Phylogeography. *Annual Review of Ecology, Evolution and Systematics*, 40, 593-612.
- Korch GW. 1994. *Geographic dissemination of tick-borne zoonoses*. New York, Oxford: Oxford University Press.
- Kotlik P, Deffontaine V, Mascheretti S et al. 2006. A northern glacial refugium for

- bank voles. *Clethrionomys glareolus*). Proceedings of the National Academy of Sciences, USA, 103, 14860–14864.
- Kozak KH, Graham CH, Wiens JJ. 2008. Integrating GIS-based environmental data into evolutionary biology. *Trends in Ecology and Evolution*, 23, 141-148.
- Jaarola M, Searle JB. 2002. Phylogeography of field voles. *Microtus agrestis*. in Eurasia inferred from mitochondrial DNA sequences. *Molecular Ecology*, 11, 2613–2621.
- Jaarola M, Searle JB. 2004. A highly divergent mitochondrial DNA lineage of *Microtus agrestis* in southern Europe. *Heredity*, 92, 228–234.
- Joger U, Fritz U, Guicking D, Kalyabina-Hauf S, Nagye ZT, Wink M. 2007. Phylogeography of western Palaearctic reptiles – Spatial and temporal speciation patterns. *Zoologischer Anzeiger*, 246, 293-313.
- Lebarbenchon C, Poitevin F, Arnal V, Montgelard C. 2010. Phylogeography of the weasel. *Mustela nivalis*. in the western-Palaearctic region: combined effects of glacial events and human movements. *Heredity*, 105, 449-462.
- Lemey P, Rambaut A, Drummond AJ, Suchard MA. 2009. Bayesian phylogeography finds its roots. *PLoS Computational Biology*, 5, e1000520.
- Lorenzini R, Lovari S. 2006. Genetic diversity and phylogeography of the European roe deer: the refuge area theory revisited. *Biological Journal of the Linnean Society*, 88, 85–100.
- Manilla G. 1998. *Acari, Ixodida*, vol. XXXVI. Bologna: Calderini.
- McLain DK, Li J, Oliver JH. 2001. Interspecific and geographical variation 704 in the sequence of rDNA expansion segment D3 of *Ixodes* ticks. *Acari, Ixodidae*). *Heredity*, 86, 234–242.
- Milne A. 1950. The ecology of the sheep tick *Ixodes ricinus* L., microhabitat economy of the adult tick. *Parasitology*, 40, 15-34.
- Michaux JR, Magnanou E, Paradis E, Nieberding C, Libois R. 2003. Mitochondrial phylogeography of the Woodmouse. *Apodemus sylvaticus*. in the Western Palearctic Region. *Molecular Ecology*, 12, 685–697.
- Michaux JR, Libois R, Paradis E, Filippucci MG. 2004. Phylogeographic history of the yellow-necked fieldmouse. *Apodemus flavicollis*. in Europe and in the Near and Middle East. *Molecular Phylogenetics and Evolution*, 32, 788–798.
- Michaux JR, Libois R, Filippucci MG. 2005. So close and so different: comparative

- phylogeography of two small mammal species, the Yellownecked fieldmouse. *Apodemus flavicollis*. and the Woodmouse. *Apodemus sylvaticus*), in the Western Palearctic region. *Heredity*, 94, 52–63.
- Nei M. 1987. *Molecular evolutionary genetics*. Columbia University Press, New York.
- Nosek J, Kozuch O, Ernek E, Lichard M. 1967. Uebertragung des Zecken - encephalitis-Virus. TBE. durch die Weibchen von *Ixodes ricinus* und die Nymphen von *Haemaphysalis inermis* auf die Rehkitzet. *Capreolus capreolus*). *Zentralblatt Fur Bakteriologie*, 203, 162-166.
- Nosek J, Sixl W. 1972. Central-European ticks. *Ixodoidea*). Key for determination. *Mitt. Abt. Zool. Landesmus Joanneum*, 1, 61–92.
- Nouredine R, Chauvin A, Plantard O. 2010. Lack of genetic structure among Eurasian populations of the tick *Ixodes ricinus* contrasts with marked divergence from north-African populations. *International Journal for Parasitology*, 41, 183–192.
- Pfäffle M, Petney T, Skuballa J, Taraschewski H. 2011. Comparative population dynamics of a generalist. *Ixodes ricinus*. and specialist tick. *I. hexagonus*. species from European hedgehogs. *Experimental and Applied Acarology*, 54, 151–164.
- Pfenninger M, Posada D. 2002. Phylogeographic history of the land snail *Candidula unifasciata*. *Poiret 1801*). *Helicellinae, Stylommatophora*): fragmentation, corridor migration and secondary contact. *Evolution*, 56, 1776-1788.
- Phillips SJ, Dudík M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, 31, 161-175.
- Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modelling of species geographic distributions. *Ecological Modelling*, 190, 231–259.
- Pons O, Petit RJ. 1996. Measuring and testing genetic differentiation with ordered vs. unordered alleles. *Genetics*, 144, 1237–1245.
- Porretta D, Canestrelli D, Urbanelli S et al. 2011. Southern crossroads of the Western Palearctic during the Late Pleistocene and their imprints on current patterns of genetic diversity: insights from the mosquito *Aedes caspius*. *Journal of Biogeography*, 38, 20-30.
- Porretta D, Mastrantonio V, Bellini, Somboon P, Urbanelli S. 2012. Glacial History of a Modern Invader: Phylogeography and Species Distribution Modelling of

- the Asian Tiger Mosquito *Aedes albopictus*. *Plos One*, 7(9), e44515. doi:10.1371/journal.pone.0044515.
- Posada D, Crandall KA. 2001. Intraspecific gene genealogies: trees grafting 749 into networks. *Trends in Ecology and Evolution*, 16, 37–45.
- Poulin R, Keeney DB. 2008. Host specificity under molecular and experimental scrutiny. *Trends in Parasitology*, 24, 24–28.
- Powell JR, Caccone A, Amato GD, Yoon C. 1986). Rates of nucleotide substitution in *Drosophila* mitochondrial DNA and nuclear DNA are similar. *Proceeding of the National Academy of Science of USA*, 83, 9090–9093.
- Randi E, Alves PC, Carranza J et al. 2004. Phylogeography of roe deer. *Capreolus capreolus*. populations: the effects of historical genetic subdivisions and recent nonequilibrium dynamics. *Molecular Ecology*, 13, 3071–3083.
- Richards CL, Carstens BC, Knowles LL. 2007. Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. *Journal of Biogeography*, 34, 1833–1845.
- Rull V. 2009. Microrefugia. *Journal of Biogeography*, 36, 481–484.
- Sanchez Marco A. 2004. Avian zoogeographical patterns during the quaternary in the Mediterranean region and paleoclimatic interpretation. *Ardeola*, 51, 91–132.
- Santos-Silva MM, Beati L, Santos AS et al. 2011. The hard-tick fauna of mainland Portugal. *Acari: Ixodidae*): an update on geographical distribution and known associations with hosts and pathogens. *Experimental and Applied Acarology*, 55, 85–121.
- Santucci F, Emerson BC, Hewitt GM. 1998. Mitochondrial DNA phylogeography of European hedgehogs. *Molecular Ecology*, 7, 1163–1172.
- Schlick-Steiner BC, Steiner FM, Moder K et al. 2006. A multidisciplinary approach reveals cryptic diversity in Western Palearctic *Tetramorium* ants. (Hymenoptera: Formicidae). *Molecular Phylogenetics and Evolution*, 40, 259–273.
- Schmitt T. 2007. Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology*, 4, 11.
- Skog A, Zachos FE, Rueness EK. 2009. Phylogeography of red deer. *Cervus elaphus*. in Europe. *Journal of Biogeography*, 36, 66–77.
- Sommer R, Benecke N. 2004. Late- and Post-Glacial history of the Mustelidae in

- Europe. *Mammal Review*, **34**, 249–284.
- Sommer R, Benecke N. 2005. The recolonization of Europe by brown bears *Ursus arctos* Linnaeus, 1758 after the Last Glacial Maximum. *Mammal Review*, **35**, 156–164.
- Sommer RS, Nadachowski A. 2006. Glacial refugia of mammals in Europe: evidence from fossil records. *Mammal Review*, **36**, 251–265.
- Sommer RS, Zachos FE, Street M et al. 2008. Late Quaternary distribution dynamics and phylogeography of the red deer. *Cervus elaphus*. in Europe. *Quaternary Science Reviews*, **27**, 714–733.
- Sonenshine DE. 1991. *Biology of ticks*. Vol. 1. Oxford University Press. New York.
- Sonenshine DE. 1993. *Biology of ticks*. Vol. 2. Oxford University Press. New York.
- Stephens M, Donnelly P. 2003. A comparison of bayesian methods for haplotype reconstruction. *American Journal of Human Genetics*, **73**, 1162–1169.
- Stewart JR, Lister AM, Barnes I, Dalén L. 2010. Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 661–671.
- Surget-Groba Y, Heulin B, Guillaume C-P, et al. 2001. Intraspecific phylogeography of *Lacerta vivipara* and the evolution of viviparity. *Molecular Phylogenetics and Evolution*, **18**, 449–459.
- Svenning JC, Fløjgaard C, Marske KF, Nógues-Bravo D & Normand S. 2011. Applications of species distribution modeling to paleobiology. *Quaternary Science Reviews*, **30**, 2930–2947.
- Swenson JE, Taberlet P, Bellemain E. 2011. Genetics and conservation of European brown bears *Ursus arctos*. *Mammal Review*, **41**, 87–98.
- Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF. 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.
- Teacher AGF, Thomas JA, Barnes I. 2011. Modern and ancient red fox. *Vulpes vulpes*. in Europe show an unusual lack of geographical and temporal structuring, and differing responses within the carnivores to historical climatic change. *BMC Evolutionary Biology*, **11**, 214–223.
- Templeton AR, Crandall KA, Sing CF. 1992. A cladistics analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, **132**, 619–633.

- Thiede J. 1978. A glacial Mediterranean. *Nature*, 276, 680–683.
- Valdiosera CE, Nuria G, Anderung C et al. 2007. Staying out in the cold: glacial refugia and mitochondrial DNA phylogeography in ancient European brown bears. *Molecular Ecology*, 16, 5140–5148.
- Varga Z. 2009. Extra-Mediterranean, post-glacial vegetation history and area dynamics in Eastern Central European refugia. In: *Relict Species*, Vol. 1. eds Habel JC, Assmann T), pp. 57–87. Springer Berlin Heidelberg, Berlin.
- Vega R, Amori G, Aloise G et al. 2010. Genetic and morphological variation in a Mediterranean glacial refugium: evidence from Italian pygmy shrews, *Sorex minutus*. *Mammalia: Soricomorpha*). *Biological Journal of the Linnean Society*, 100, 774–787.
- Vichová B, Majiláthová V, Nováková M et al. 2010. First molecular detection of *Anaplasma phagocytophilum* in European Brown Bear. *Ursus arctos*). *Vector Borne of Zoonotic Disease*, 10, 543–545.
- Waltari E, Hijmans RJ, Peterson AT, Nyári ÁS, Perkins SL, Guralnick RP. 2007. Locating Pleistocene refugia: comparing phylogeographic and ecological niche model predictions. *Plos ONE*, 2, 563-574.
- Warren DL, Seifert S. 2011. Environmental niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications*, 21, 335-342.

2.7.4 Appendix

Species	Glacial Refugia	Fossil records		Genetic data	SDM
		Locality records	References	References	References
Mammals					
<i>Alces alces</i>	S,N	28;32;33;34;37;42	Sommer & Nadachowski (2006)	Michaux et al. (2004);	
<i>Apodemus sylvaticus</i>	S,N	11;23;32	Sommer & Nadachowski (2006)	Michaux et al. (2003); Michaux et al. (2005)	Fløjgaard et al. (2009)
<i>Apodemus flavicollis</i>	S			Michaux et al. (2005)	Fløjgaard et al. (2009)
<i>Capreolus capreolus</i>	S,N	9;11;12;14;15; 16;17;18;19;23; 27;28;32;34;41; 42;44;46;48	Sommer & Nadachowski (2006)	Lorenzini & Lovari (2006); Randi et al. (2004)	
<i>Cervus elaphus</i>	S,N	9;10;11;12;14; 15;16;17;18;19; 22;23;27;28;32; 33;34;36;37;39; 40;42;43;44;45; 46;47;48	Sommer & Nadachowski (2006); Sommer et al. (2008)	Skog et al. (2009) Deffontaine et al. (2005); Kotlik et al. (2006)	Banks et al. (2008)
<i>Clethrionomys glareolus</i>	S,N	39	Sommer & Nadachowski (2006)		Fløjgaard et al. (2009)
<i>Erinaceus europaeus</i>	S	14;17;28;49	Sommer & Nadachowski (2006)	Santucci et al. (1998); Hewitt (1999)	
<i>Felis silvestris</i>	S	19;28;34;46;48	Sommer & Nadachowski (2006)		
<i>Glis glis</i>	S,N	14;23;39	Sommer & Nadachowski (2006)	Hürner et al. (2010)	
<i>Martes martes</i>	S,N	33;39;40	Sommer & Benecke (2004); Sommer & Nadachowski (2006)	Davison et al. (2001)	
<i>Meles meles</i>	S	12;19;20;28;33;45	Sommer & Nadachowski (2006)		
<i>Microtus agrestis</i>	S,N			Jaarola and Searle (2002); Jaarola & Searle (2004)	Fløjgaard et al. (2009)
<i>Microtus arvalis</i>	S,N			Haynes et al. (2003); Heckel et al. (2005)	Fløjgaard et al. (2009)
<i>Microtus oeconomus</i>	N			Brunhoff et al. (2003)	Fløjgaard et al. (2009)
<i>Mustela nivalis</i>	N	1;2;3;4;5;6;7;8;13; 26;29;51	Sommer & Benecke (2004)	Lebarbenchon et al. (2010)	
<i>Mustela putorius</i>	S,N	14;15;28;31;42	Sommer & Benecke (2004); Sommer & Nadachowski (2006)	Davison et al. (2001)	
<i>Sorex araneus</i>	S	14;46	Sommer & Nadachowski (2006)	Hewitt (1999)	
<i>Sorex minutus</i>	S,N			Bilton et al. (1998); Vega et al. (2010)	Vega et al. (2010)
<i>Sus scrofa</i>	S,N	9;11;12;14;15; 16;19;27;28;33; 44;45;46;47;48	Sommer & Nadachowski (2006)		
<i>Ursus arctos</i>	S,N	14;16;19;35; 36;38;42;43;48	Sommer & Benecke (2005)	Swenson et al. (2011) and references therein	

		9:11;12;14;15; 16;17;18;19;28; 33;36;37;38;39; 40;41;42;43;44; 48;49	Sommer & Nadachowski (2006)	Teacher et al. (2011)
<i>Vulpes vulpes</i>	S,N			

Birds

<i>Accipiter nisus</i>	S	34;50	Sanchez Marco (2004)
<i>Asio otus</i>	S	34	Sanchez Marco (2004)
<i>Buteo buteo</i>	S	21;50	Sanchez Marco (2004)
<i>Coccothraustes coccothraustes</i>	S,N	34	Sanchez Marco (2004)
<i>Corvus corone</i>	S,N	5;34;50	Sanchez Marco (2004)
<i>Corvus monedula</i>	S,N	21;24;34	Sanchez Marco (2004)
<i>Coturnix coturnix</i>	S,N	24;34	Sanchez Marco (2004)
<i>Emberiza citrinella</i>	S	21;50	Sanchez Marco (2004)
<i>Erethaceus rubecola</i>	S	34	Sanchez Marco (2004)
<i>Falco tinnunculus</i>	S,N	21;24;25;34;50	Sanchez Marco (2004)
<i>Fringilla coelebs</i>	S	34	Sanchez Marco (2004)
<i>Lanius collurio</i>	S	34	Sanchez Marco (2004)
<i>Perdix perdix</i>	S,N	21;24;25	Sanchez Marco (2004)
<i>Pica pica</i>	N	50	Sanchez Marco (2004)
<i>Pyrrhula pyrrhula</i>	S	34	Sanchez Marco (2004)
<i>Scolopax rusticola</i>	S,N	21;25;34	Sanchez Marco (2004)
<i>Sturnus vulgaris</i>	S	34	Sanchez Marco (2004)
<i>Turdus merula</i>	S	50	Sanchez Marco (2004)
<i>Turdus pilaris</i>	S	50	Sanchez Marco (2004)
<i>Turdus philomelos</i>	S	50	Sanchez Marco (2004)

Reptiles

<i>Lacerta agilis</i>	N		Joger et al. (2007)
<i>Lacerta viridis</i>	S		Godinho et al. (2005)
<i>Lacerta vivipara</i>	S,N		Surget-Groba et al. (2001); Joger et al. (2007)

3. Other relevant publications produced during my PhD (a color version of the pdf documents are stored in a separate folder on CD-ROM attached to the printed copy)

3.1 Research article. "*PACHYBRACHIS SASSII, a new species from the Mediterranean Giglio Island (Italy) (Coleoptera, Chrysomelidae, Cryptocephalinae). ZooKeys (2011), 155: 51–60*"

3.2 Research article. "*Molecular Evidence for Multiple Infections as Revealed by Typing of ASAIA Bacterial Symbionts of Four Mosquito Species. Applied and Environmental Microbiology (2010), 76(22): 7444–7450*"

3.3 Research article. "*The Beetle (Coleoptera) and True bug (Heteroptera) species pool of the Alpine "Pian di Gembro" wetland (Villa di Tirano, Italy) and its conservation. Journal of Entomological and Acarological Research (2011) 43: 7-22*"

3.4 Research article. "*Insect community structure and insect biodiversity conservation in an Alpine wetland subjected to an intermediate diversified management regime. Ecological Engineering (2012), 47: 242–246*"

3.5 Research article. "*Molecular typing of bacteria of the genus ASAIA in malaria vector ANOPHELES ARABIENSIS Patton, 1905. Journal of Entomological and Acarological Research (2012) 44: 33-36*"

3.6 Research article. "*A study on the presence of flagella in the order Rickettsiales: the case of 'Candidatus Midichloria mitochondrii'. Microbiology (2012), 158, 1677–1683*"

4 Grants and research projects

During my PhD experience, in 2011 and 2012 I have received two SYS-RESOURCE grants (DE-TAF-1869 and DE-TAF-1839 respectively) to work in the Department of Entomology at Museum für Naturkunde in Berlin. The project, focused on a revision of the genus *Metallactus* Suffrian, 1866, aims to fill an important gap in the systematic of Neotropical Cryptocephalini (Coleoptera, Chrysomelidae). The first step in the revision is the images' acquisition of general habitus (using a stacking procedure) and morphometric data (total body length; length of pronotum; width of elytra; width of pronotum) of the type specimens. The internal sexual characters (aedeagus, spermatheca and kotpresse) are considered also.

In 2010 I have received a fellowship by the Regional Natural Reserve "Parco delle Orobie Bergamasche" (Bergamo, Italy) in order to study the phylogeography and the genetic structure, using molecular tools, of the endemic species *Cryptocephalus barii* Burlini, 1948. In the same year I have activated, with the partnership of the Biodiversity Institute of Ontario, a project on DNA barcoding of the Leaf Beetles (Coleoptera, Chrysomelidae) that inhabit the Mediterranean Region. This project aims to obtain a DNA repository for a fragment of approximately 650 base pairs of the mitochondrial gene coding for Cytochrome c oxidase subunit I for all the Leaf Beetles species present in the Mediterranean Region. At present, the DNA sequences are available for 759 specimens belonging to 239 species. More data and information on my project are available at the web site www.c-bar.org.

5 Curriculum vitae

Personal Data

Name: Matteo Montagna
Position: PhD student in Animal Biology
Work Address: DIVET University of Milan Via Celoria 2-10, I-20133 Milano (Italy)
Place and date of birth: Como, 8 february 1980
Nationality: Italian
Contact: +393405403214 (mobile); +39 02 50318094 (office)

Education

§ He took the scientific graduation (secondary school) at the end of school year 1998/99 at Liceo statale “Galileo Galilei” of Erba (CO).

§ 01.10.2003 he registered at Corso di Laurea in Valorizzazione e Tutela dell’Ambiente e del Territorio Montano (Class 20), University of Milan, Facoltà di Agraria.

§ 19.10.2006 he registered at Corso di Laurea in Scienze Agroambientali (Agroenvironmental science) (Classe 74/S). University of Milan, Facoltà di Agraria.

§ 27.11.2006 he took the first degree, with honors and distinction (110/110 cum laude), in Valorizzazione e Tutela dell’Ambiente e del Territorio Montano (Class 20- Science and agritechnology, agrifood and forest). Title of dissertation: “Preliminary research of entomofauna of Natural Reserve of Pian di Gembro (Valtellina)”.

§ 20.10.2008 he took the second degree, with honors and distinction (110/110 cum laude), in Agroenvironmental Science (Classe 74/S- Science and management of country and forestal resources). Title of dissertation: ”Analysis and valuation of entomological biodiversity of SIC “Pian Gembro” (Sondrio, Lombardia).

§ November 2009 he received a PhD grant in Animal Biology at the University of Milan.

§ January 2010 he started the PhD program.

Grants

§ 2012 - European Union SYS-RESOURCE Fellowship (FP7 Programme, Access to Research Infrastructures). Department of Entomology, Museum fur Naturkunde, DE. 4 March - 10 March. DE-TAF-1939: A revision of the genus *Metallactus* Suffrian, 1866.

§ Prin 2010-2011 cod. 2010PJJ4W3 -Operative Units of Milano -

Title: Inhibition of multidrug-resistance efflux pumps as a novel strategy for the sustainable control of ticks and tick-borne diseases.

§ 2011 - European Union SYS-RESOURCE Fellowship (FP7 Programme, Access to Research Infrastructures). Department of Entomology, Museum fur Naturkunde, DE. 4 March - 10 March. DE-TAF-1869: A revision of the genus *Metallactus* Suffrian, 1866.

§ 2010 - Research Collaboration with the Regional Natural Reserve “Parco delle Orobie Bergamasche” (Bergamo, Italy) to study the phylogeography and the genetic structure of the endemic species *Cryptocephalus barii* Burlini, 1948.

Publications

Journal article

Montagna M., Chouaia B., Pella F., Mariconti M., Fasola M., Epis S. 2012. Bacterial screening in *Hyalomma marginatum* (Ixodida, Ixodidae) from Socotra Island (Yemen), detection of new isolated Francisella-like tick endosymbiont. *Journal of Entomological and Acarological Research*, 44: 60-63. DOI: 10.4081/jear.2012.e13

Epis S., Mandrioli M., Genchi M., **Montagna M.**, Sacchi L., Pistone D., Sasser D. Whole-mount FISH staining for the localization of the bacterial symbiont *Midichloria mitochondrii* in the hard tick *Ixodes ricinus*. *Ticks and Tick-borne Diseases*, doi: 10.1016/j.ttbdis.2012.06.005. [Epub ahead of print]

Montagna M., Sasser D., Griggio F., Epis S., Bandi C., Gissi C. Tick-Box for 3'-end formation of mitochondrial transcripts in Ixodida, basal chelicerates and *Drosophila*. *PlosOne*: 7: e47538.

Montagna M., Lozzia G.C., Giorgi A., Baumgärtner J. 2012. Insect community structure and insect biodiversity conservation in an Alpine wetland subjected to an intermediate diversified management regime. *Ecological Engineering* 47 (2012) 242–246.

Epis S., **Montagna M.**, Comandatore F., Damiani C., Diabaté A., Daffonchio D., Chouaia B., Favia G. 2012. Molecular typing of bacteria of the genus *Asaia* in malaria vector *Anopheles arabiensis* Patton, 1905. *Journal of Entomological and Acarological Research* 44:e7, doi:10.4081/jear.2012.e7.

Mariconti M, Epis S, Sacchi L, Biggiogera M, Sasser D, Genchi M, Alberti E, **Montagna M**, Bandi C, Bazzocchi C. 2012. A study on the presence of flagella in the order Rickettsiales: the case of 'Candidatus *Midichloria mitochondrii*'. *Microbiology* 158:1677-83.

Pistone D, Marone P, Pajoro M, Fabbi M, Vicari N, Daffara S, Dalla Valle C, Gabba S, Sasser D, Verri A, **Montagna M**, Epis S, Monti C, Strada EG, Grazioli V, Arrigoni N, Giacosa A, Bandi C. 2012. *Mycobacterium avium* paratuberculosis in Italy: commensal or emerging human pathogen? *Dig Liver Dis*. 2012 Jun;44(6):461-

5.

Montagna M, Lozzia GC, Baumgartner J 2011. The Beetle (Coleoptera) and True bug (Heteroptera) species pool of the alpine "Pian di Gembro" wetland (Villa di Tirano, Italy) and its conservation. *Journal of entomological and acarological research* 43: 7-22.

Sassera D, Lo N, Epis S, D'Auria G, **Montagna M**, Comandatore F, Horner D, Peretó J, Luciano AM, Franciosi F, Ferri E, Crotti E, Bazzocchi C, Daffonchio D, Sacchi L, Moya A, Latorre A, Bandi C 2011. Phylogenomic evidence for the presence of a flagellum and cbb(3) oxidase in the free-living mitochondrial ancestor. *Mol Biol Evol.* 28:3285-96.

Montagna M 2011. *Pachybrachis sassii*, a new species from the Mediterranean Giglio Island (Italy) (Coleoptera, Chrysomelidae, Cryptocephalinae). *Zookeys* 155:51-60.

Montagna M, E. Ferri, L. Beati, D. Sassera, M. Casiraghi, S. Epis, L. Sacchi, C. Genchi, C. Bandi (2010). DNA barcoding in ixodid ticks. *PARASSITOLOGIA*, vol. 52, p. 209, ISSN: 0048-2951.

Chouaia B, Rossi P, **Montagna M**, Ricci I, Crotti E, Damiani C, Epis S, Faye I, Sagnon N, Alma A, Favia G, Daffonchio D, Bandi C 2010. Molecular evidence for multiple infections as revealed by typing of *Asaia* bacterial symbionts of four mosquito species. *Appl Environ Microbiol.* 76:7444-50.

Montagna M, Dioli P, Regalin R 2009. Indagini sull'entomofauna del Sito di Importanza Comunitaria Pian Gembro. *Naturalista Valtellinese*, vol. 19, p. 140, ISSN: 1120-6519.

Montagna M, Dioli P, Regalin R 2009. Gli Eterotteri terrestri (Insecta: Heteroptera) della torbiera di Pian Gembro (Sondrio, Lombardia). *Naturalista Valtellinese*, vol. 19, p. 83-111, ISSN: 1120-6519.

Montagna M (2009). Osservazioni sui Crisolmelidi delle riserve naturali "Agoraie di Sopra e Moggetto" (Liguria, Genova) e "Guadine Predaccio" (Emilia Romagna, Parma) (Coleoptera Chrysomelidae). *Bollettino dell'associazione Romana di Entomologia*, vol. LXIV, p. 163-181, ISSN: 0004-6000.

Journal abstract

M. Pajoro, D. Pistone, S. Epis, D. Sassera, **M. Montagna**, C. Luzzago, P. Lanfranchi, R. Viganò, M.C. Cerutti, N. Vicari, M. Fabbi, C. Bandi, C. Genchi

(2010). A survey for bacterial pathogens associated with *Ixodes ricinus* in north-western Italy. *Parassitologia*, vol. 52, p. 178, ISSN: 0048-2951

Book chapter

Montagna M, Sacchi L, Lo N, Clementi E, Daffonchio D, Alma A, Sassera D, Bandi C 2011. Intracellular primary symbionts: the Bacteroidetes *Blattabacterium* and *Sulcia*. In: *Manipulative Tenants: Bacteria Associated with Arthropods*. Zchori-Fein E. & Bourtzis K. (eds). *Frontiers in Microbiology*, CRC Press, 306 pp.

Sassera D., Beninati T., Epis S, Bandi C, Beati L, **Montagna M**, Alba M, Genchi C, Sacchi L, Lo N 2010. '*Candidatus* Midichloria mitochondrii', formerly IricES1, a symbiont of the tick *Ixodes ricinus* that resides in the host mitochondria. In: *Proceedings of International Congress of Acarology*.

F. Comandatore, D. Sassera, N. Lo, S. Epis, G. D'Auria, **M. Montagna**, D.S. Horner, J. Pereto, A.M. Luciano, F. Franciosi, E. Ferri, E. Crotti, C. Bazzocchi, D.G. Daffonchio, L. Sacchi, A. Moya, A. Latorre, C. Bandi (2011). Symbionts, within symbionts, within parasites : a key to thrive in suboxic conditions?. In: *Microbial Diversity 2011 : Environmental Stress and Adaptation*. Milano, 2011, p. 90-92, SIMTREA.

Work activities, short scientific mission

Since 1-3-2009 had been working in molecular biology lab of prof. Claudio Bandi in the Dipartimento di Patologia Animale, Igiene e Sanità Pubblica Veterinaria della Facoltà di Veterinaria, Università degli Studi di Milano.

June 2009 worked with prof. Lorenza Beati on DNA barcoding of ticks genus *Amblyomma* (Ixodidae). U.S. National Tick Collection (USNTC) e l'Institute of Arthropodology & Parasitology (Statesboro GA, USA).

January 2010. European Course on Comparative Genomics, ENS- 2010 University of Lyon (France).

From 26 August to 1 September 2010. Italian workshop on phylogenetic methods and applications. SIBE-University of Milan (Italy)-University of California at Los Angeles (USA).

Since 2010 he is responsible of the Mediterranean Leaf Beetles barcoding project (www.c-bar.org) with the BOLD consortium as partner.

February 10 - 18 February 2012 he spent a short scientific mission in the

laboratory of pro. Jesus Gomez-Zurita (Institut de Biologia Evolutiva CSIC-UPF, Barcelona, Spain), working on a project on the phylogeography of the *Cryptocephalus marginellus* (Coleoptera, Chrysomelidae) species complex.

March 4 - 10 March 2012 he spent a short scientific mission in the Department of Entomology, Museum fur Naturkunde, DE. Project code DE-TAF-1869: A revision of the genus *Metallactus* Suffrian, 1866.

July 22 - August 1 2012 Workshop on Molecular Evolution - Marine Biological Laboratory - Woods Hole, MA - USA

Extra curricular activities

07.2006 Entomological campaign in North Greece to look for find same individuals of new species of *Labidostomis* Germar, 1822.

May 2009 Entomological campaign in Tunisia North

Summer 2010 Entomological campaign in Islands of Archipelago Toscano (IT)

Summer 2011 Entomological campaign in the Var Region (France) and in the Pyrénées-Orientales (France/Spain) to collect specimens of *Cryptocephalus marginellus* complex.

Summer 2012 Entomological campaign in the Var Region (France), Calanques Massif (France) to collect specimens of *Cryptocephalus marginellus* species complex.

Membership of Scientific Societies

SOCIETÀ ITALIANA DI ENTOMOLOGIA

ENTOMOLOGICAL SOCIETY OF AMERICA

WILLY HENNIG SOCIETY

SYSTEMATIC ASSOCIATION

SOCIETY OF SYSTEMATIC BIOLOGISTS

SOCIETÀ ITALIANA DI BIOLOGIA EVOLUZIONISTICA

