Integrated metatranscriptomics and metaproteomics for the characterization of bacterial microbiota in unfed *Ixodes ricinus*

Angélica Hernández-Jarguín[1&], Sandra Díaz-Sánchez[1&], Margarita Villar[1], José de la Fuente[1,2*]

[1]SaBio. Instituto de Investigación en Recursos Cinegéticos IREC-CSIC-UCLM-JCCM, Ronda de Toledo s/n, 13005 Ciudad Real, Spain.

[2]Department of Veterinary Pathobiology, Center for Veterinary Health Sciences, Oklahoma State University, Stillwater, OK 74078, USA.

[&]The first two authors contributed equally to the work reported in this paper.

*Correspondence: José de La Fuente. SaBio. Instituto de Investigación en Recursos Cinegéticos, IREC-CSIC-UCLM-JCCM, 13005 Ciudad Real, Spain; e-mail: jose_delafuente@yahoo.com

Running title: Metaomics of tick bacterial microbiota

Abstract

An innovative metaomics approach integrating metatranscriptomics and metaproteomics was used to characterize bacterial communities in the microbiota of the Lyme borreliosis spirochete vector, *Ixodes ricinus* (Acari: Ixodidae). Whole internal tissues and salivary glands from unfed larvae and female ticks, respectively were used. Reused *I. ricinus* RNA-sequencing data for metranscriptomics analysis together with metaproteomics provided a better characterization of tick bacterial microbiota by increasing bacteria identification and support for identified bacteria with putative functional implications. The results showed the presence of symbiotic, commensal, soil, environmental, and pathogenic bacteria in the *I. ricinus* microbiota, including previously unrecognized commensal and soil microorganisms. The results of the metaomics approach may have implications in the characterization of putative mechanisms by which pathogen infection manipulates tick microbiota to facilitate infection. Metaomics approaches integrating different omics datasets would provide a better description of tick microbiota compositions, and insights into tick interactions with microbiota, pathogens and hosts.

Keywords: metatranscriptomics; metaproteomics; metaomics; tick; microbiota; biofilm

1. Introduction

The microbiota plays an important role in several processes affecting human and animal health, agriculture, environment, and host-pathogen interactions (Kau et al., 2011; Schwabe and Jobin, 2013; Philippot et al., 2013; Bouchez et al., 2016). Next-generation sequencing or omics technologies can be used for microbiota characterization under different experimental and natural conditions. Metagenomics have been used to characterize the microbiota in different hosts including both model and nonmodel organisms such as humans and tick vectors (Clay et al., 2008; Andreotti et al., 2011; Carpi et al., 2011; Vayssier-Taussat et al., 2015; Qiu et al., 2014; Williams-Newkirk et al., 2014; van Treuren et al., 2015; Narasimhan and Fikrig, 2015; Yoon et al., 2015; Abraham et al., 2017; Heintz-Buschart and Wilmes, 2017; Greay et al., 2017; Varela-Stokes et al., 2017; Xiang et al., 2017). Different metatranscriptomics approaches have been also applied to the study of microbial communities in arthropod vectors and vertebrate hosts (Mäder et al., 2011; Johannson et al., 2013; Vayssier-Taussat et al., 2013; Razzauti et al., 2015; Luo et al., 2017). Recently, metaproteomics and metabolomics have emerged as powerful tools for the characterization of dynamic host-microbiome interactions, particularly in combination with metagenomics and metatranscriptomics approaches (Tanca et al, 2013; 2014; Franzosa et al., 2015; Aguiar-Pulido et al., 2016; Cheng et al., 2017). Furthermore, metaomics or the integration of different omics approaches allows network-based analyses to describe the complexity and function of different biological processes involved in host/tick-pathogen and host/tick-microbiome interactions (Franzosa et al., 2015; Villar et al, 2015; Narasimhan and Fikrig, 2015), and the discovery of new targets for prevention and control of tick-borne diseases (Abraham et al., 2017; Narasimhan et al., 2017; Xiang et al., 2017).

Ixodes ricinus (Linnaeus 1758) (Acari: Ixodidae) are obligate hematophagous ectoparasites and vectors of multiple pathogens such as *Borrelia* spp. (Lyme borreliosis and hard tick-borne relapsing fever), *Anaplasma phagocytophilum* (human granulocytic anaplasmosis), tick-borne encephalitis virus (TBE), and *Babesia* spp. (babesiosis) (de la Fuente et al., 2008; 2017). Additionally, *I. ricinus* have a diverse community of commensal and symbiotic microorganisms which exert multiple effects on tick fitness, nutrition, development, reproduction, defense against environmental stress, immunity and transmission of tick-borne pathogens (Bonnet et al., 2017; de la Fuente et al., 2017). The *I. ricinus* microbiome was first characterized using a metagenomics approach (Carpi et al., 2011; Nakao et al., 2013; Bonnet et al., 2014). Vayssier-Taussat et al. (2013) characterized the bacterial community of *I. ricinus* using a whole transcriptomics approach, resulting in a better identification of previously unknown bacteria and accurate identification of potential pathogens. This method also provides a better understanding of the tick-microbiome interactions when compared to metagenomics. Additionally, reusing RNA sequencing (RNA-seq) data has been also used as an efficient strategy for the screening of pathogens in ticks (Zhuang et al., 2014a).

In this study, we used the integration of metatranscriptomics and metaproteomics for the characterization of the tick bacterial microbiota in unfed *I. ricinus*. Reused *I. ricinus* RNA-seq data for metranscriptomics analysis together with metaproteomics provided a better characterization of tick microbiome by increasing bacterial identification and support for identified bacteria with putative functional implications.

2. Materials and methods

2.1. Tick samples and processing.

Tick samples were obtained and processed as previously described (Genomic Resources Development Consortium et al., 2014). Briefly, *I. ricinus* unfed larvae and adult females were obtained from the reference laboratory colony maintained at the tick rearing facility of the Institute of Parasitology of the Biology Centre of the Academy of Sciences of the Czech Republic. Whole internal tissues and salivary glands from 300 larvae and 30 female ticks, respectively were combined and used for RNA-seq. All ticks were washed with a series of solutions composed of tap water, 3% hydrogen peroxide, two washes of distilled water, 70% ethanol and two more washes with distilled water prior to dissection for DNA, RNA and protein extraction. Total DNA, RNA and proteins were extracted using Tri Reagent (Sigma-Aldrich, St. Louis, MO, USA) according to manufacturer instructions. RNA was further purified with the RNeasy MinElute Cleanup Kit (Qiagen, Valencia, CA, USA) and characterized using the Agilent 2100 Bioanalyzer (Santa Clara, CA, USA) in order to evaluate the quality and integrity of RNA preparations. DNA and RNA concentrations were determined using the Nanodrop ND-1000 (NanoDrop Technologies Wilmington, Delaware USA). Proteins were resuspended in 20 mM Tris-HCl pH 7.5 with 4% SDS and protein concentration was determined using the BCA Protein Assay kit (Thermo Scientific, Rockford, IL, USA) with bovine serum albumin (BSA) as standard.

2.2. Integrated metaomics experimental design.

An integrated metatranscriptomics and metaproteomics approach was developed for the characterization of *I. ricinus* bacterial microbiota (Fig. 1). Reused *I. ricinus* RNA-seq data (Genomic Resources Development Consortium et al., 2014) derived from combined female salivary glands and larvae were the basis for metatranscriptomics analysis that resulted in the database of identified bacterial genera. This database was then used to generate the

Uniprot protein database for application as a variant of the proteomics informed by transcriptomics (PIT) approach (Evans et al., 2012) in the metaproteomics analysis.

2.3. Metatranscriptomics for the identification of bacterial species in the tick microbiome.

A metatranscriptomics pipeline was developed based on the reused *I. ricinus* RNA-seq data (Fig. 1). Tick RNA-seq analysis was conducted as previously described (Genomic Resources Development Consortium et al., 2014). The *I. ricinus* transcriptome, raw reads and assembly results can be accessed at dryad entries doi: 10.5061/dryad.9js92/1 - doi: 10.5061/dryad.9js92/8. The metatranscriptomics database was then generated from the 19,831,942 *I. ricinus* unaligned reads that did not match to the *I. scapularis* reference genome (assembly JCVI_ISG_i3_1.0; http://www.ncbi.nlm.nih.gov/nuccore/NZ_ABJB000000000) (Genomic Resources Development Consortium et al., 2014). The unaligned reads were extracted from the BAM files (the binary version of the SAM file, a tab-delimited text file that contains sequence alignment data) that resulted after the assembly of *I. ricinus* transcriptome using the SAMtools (Li et al., 2009; Li, 2011; http://samtools.sourceforge.net). Then, the unaligned reads were searched against a bacterial sequence database constructed with genome and/or species-specific ribosomal RNA (rRNA) sequences downloaded from the NCBI (https://www.ncbi.nlm.nih.gov) (Supplementary file 1 – Table 1). The bioinformatics approach to identify bacterial sequences was done in two steps. First, the LAST genome-scale sequence comparison tool (http://last.cbrc.jp) was used to search against the bacterial database previously constructed**.** The reads containing poly-A tails were discarded. As cut-off criteria for genome-scale sequence comparison we applied minimum alignment length of 100 nucleotides, e-value 0.001, with a word size of 11, and a minimum of 70% sequence

identity. Then, the putative bacterial reads detected with LAST were further confirmed by BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch) (Frith et al., 2010 a,b; Kiełbasa et al., 2011). BLAST assignments were done by using the 10 best BLAST hits (BBH) for each putative bacteria previously assigned by LAST. The sequences with hits matching to bacteria were confirmed as identified bacterial sequences, discarding the rest. Manual filtering was applied to remove those sequences with similarity to functional domains. The metatranscriptomics database was constructed taking the number of count reads or identifications (IDs) assigned to each identified bacterial sequences, and normalized against the total number of IDs (Supplementary file 2 – Dataset 1).

2.4. Metaproteomics for bacterial protein identification.

The metaproteomics pipeline included *de novo* identification using a protein database constructed based on the bacterial genera identified by metatrascriptomics, a variant of the PIT approach (Evans et al., 2012) (Fig. 1). Protein extracts (150 µg per sample) were on-gel concentrated by SDS-PAGE as previously described (Villar et al., 2015). The unseparated protein band was visualized by staining with GelCode Blue Stain Reagent (Thermo Scientific), excised, cut into 2x2 mm cubes and digested overnight at 37 °C with 60 ng/µl sequencing grade trypsin (Promega, Madison, WI, USA) at 5:1 protein:trypsin (w/w) ratio in 50 mM ammonium bicarbonate, pH 8.8 containing 10% (v/v) acetonitrile (Shevchenko et al., 2006). The resulting tryptic peptides from the gel band were extracted by 30 min-incubation in 12 mM ammonium bicarbonate, pH 8.8. Trifluoroacetic acid was added to a final concentration of 1% and the peptides were finally desalted onto OMIX Pipette tips C18 (Agilent Technologies, Santa Clara, CA, USA), dried-down and stored at -20 °C until mass spectrometry analysis. The desalted protein digests were resuspended in 0.1% formic

acid and analyzed by reverse phase (RP)-liquid chromatography (LC)-mass spectrometry (MS)/MS (RP-LC-MS/MS) using an Easy-nLC II system coupled to an ion trap LCQ Fleet mass spectrometer (Thermo Scientific). The peptides were concentrated (on-line) by reverse phase chromatography using a 0.1x20 mm C18 RP pre-column (Thermo Scientific), and then separated using a 0.075 x 100 mm C18 RP column (Thermo Scientific) operating at 0.3 ml/min. Peptides were eluted using a 180-min gradient from 5 to 35% solvent B in solvent A (solvent A: 0.1% formic acid in water, solvent B: 0.1% formic acid in acetonitrile). Electrospray ionization (ESI) was done using a Fused-silica PicoTip Emitter ID 10 mm (New Objective, Woburn, MA, USA) interface. Peptides were detected in survey scans from 400 to 1600 amu (1 mscan), followed by three data dependent MS/MS scans (Top 3), using an isolation width of 2 mass-to-charge ratio units, normalized collision energy of 35%, and dynamic exclusion applied during 30 sec periods. The MS/MS raw files were searched against a compiled database containing the Uniprot Ixodidae taxonomy (134,957 entries in February 2017) together with a database created from the bacterial genera identified by metatranscriptomics (4,185,346 Uniprot entries in February 2017) using the SEQUEST algorithm (Proteome Discoverer 1.4, Thermo Scientific). The following constraints were used for the searches: tryptic cleavage after Arg and Lys, up to two missed cleavage sites, and tolerances of 1 Da for precursor ions and 0.8 Da for MS/MS fragment ions and the searches were performed allowing optional Met oxidation and Cys carbamidomethylation. A false discovery rate (FDR) < 0.01 was considered as condition for successful peptide assignments and at least two peptides per protein were the necessary condition for protein identification. After discarding *Ixodidae* assignations, peptides corresponding to bacterial genera were grouped and the total number of peptide spectrum matches (PSMs) for each bacterial genera were normalized against the total number of

PSMs (Supplementary file 3 – Dataset 2). The gene ontology (GO) annotations for biological process (BP) were done in proteins identified in tick-borne pathogens (TBPs; *Anaplasma*, *Borrelia*, *Ehrlichia* and *Rickettsia* genera) according to Uniprot (http://www.uniprot.org) (Supplementary file 3 – Dataset 2).

2.5. Phylogenetic and taxonomic abundance analyses.

The metatrasncriptomics and metaproteomics bacteria relative abundance and taxonomic analyses were done using a Heatmap Tool associated with a pruned phylogenetic tree generated with the platform phyloT (http://phylot.biobyte.de) based on NCBI taxonomy, and visualized using the Interactive Tree of Life software v3.4.3 (http://itol.embl.de) (Letunic and Bork, 2011). A correlation analysis was conducted in Microsoft Excel (version 12.0) between bacterial genera identification by metatranscriptomics (IDs) and metaproteomics (PSMs). The Pearson's correlation coefficient was calculated using the Pearson's Correlation Coefficient Calculator (http://www.socscistatistics.com/tests/pearson/Default2.aspx) ($r = 0.5$).

2.6. Validation of tick bacterial microbiota identifications by real-time PCR.

DNA from the same unfed larvae and female salivary glands samples were used for validation of tick bacterial microbiota identification by real-time PCR. Specific oliginucleotide primers for the bacteria *Rickettsia* spp., *Ehrlichia* spp., Spotted Fever Group (SFG) *Rickettsia*, *Anaplasma* spp., *Anaplasma phagocytophilum*, *Borrelia* spp., *Wolbachia* spp., *Candidatus* Midichloria mitochondrii and *Pseudomonas putida* were used (Table 1). The iScript One-Step was used to perform the real-time PCR with SYBR Green and the iQ5 thermal cycler (Bio-Rad, Hercules, CA, USA) following manufacturer's recommendations. A dissociation curve was run at the end of the reaction to ensure that only one amplicon was formed and that the amplicons denatured consistently in the same

temperature range for every sample (Ririe et al., 1997). DNA levels were normalized against tick *16S rRNA* and *Ixodes rps4* genes following the conditions previously reported by Zivkovic et al. (2009) and Koči et al. (2013). Normalization was performed using the genNorm method (ddCT method as implemented by Bio-Rad iQ5 Standard Edition, Version 2.0) (Livak and Schmittgen, 2001).

3. Results and discussion

3.1. Metatranscriptomics bacteria identification in *I. ricinus* microbiota.

The metatranscriptomics analysis identified a total of 450 reads that matched with specific bacterial genomes distributed among 8 phyla and 38 genera, including uncultured bacteria (Table 2, Fig. 2 and Supplementary file 2 – Dataset 1). The most represented phyla identified were Proteobacteria with 21 genera, followed by Actinobacteria and Firmicutes represented by 7 and 4 genera, respectively (Table 2). Other phyla such as Tenericutes, Spirochaetes, Fusobacteria and Bacteroidetes were also identified but with lower diversity (Table 2 and Figure 2). As expected, most of the bacteria identified by metatranscriptomics have been previously described as apart of the microbiota in different tick species (Table 3). However, other genera such as *Actinomyces, Amycolatopsis*, *Bidifidobacterium*, *Giliamella*, *Kurthia*, *Mesorhizobium* and *Variovorax* have not been previously reported in ticks (Table 3). Identified bacteria included tick endosymbionts such as *Candidatus* Midichloria, *Wolbachia*, *Francisella*, *Spiroplasma* and *Rickettsiella*, commensals such as *Escherichia*, *Staphylococcus* and *Streptococcus*, soil and environmental microorganisms such as *Acinetobacter*, *Arthrobacter*, *Bradyrhizobium*, *Sphingomonas* and *Pseudomonas*, human pathogens such as *Brucella* and *Enterococcus*, and TBPs such as *Anaplasma*, *Ehrlichia*, *Rickettsia, Neorickettsia* and *Borrelia* (Table 2, Fig. 2 and Table 3). Among the

newly identified bacteria in tick microbiota, commensal (*Actinomyces*, *Bidifidobacterium*, *Giliamella*) and soil (*Amycolatopsis*, *Kurthia*, *Mesorhizobium* and *Variovorax*) microorganisms were present (Table 2, Fig. 2 and Table 3).

3.2. Integration of metatranscriptomics and metaproteomics approaches.

The metatranscriptomics and metaproteomics results were integrated to provide a metaomics approach to bacteria identifications in the *I. ricinus* microbiota. A total of 10,845 PSMs were assigned to different bacterial genera present in the metatranscriptomics database (Supplementary file 3 – Dataset 2). Metaproteomics not only provided support to metatranscriptomics results by identifying 87% of the identified bacterial genera (Table 2), but also increased bacteria identification for different genera (Fig. 3A, Supplementary file 2 – Dataset 1 and Supplementary file 3 – Dataset 2). However, metaproteomics may results in some peptide assignments that could match to several related species, which requires further analyses with amino acid sequences of peptides used for protein identity for better definition at the species level (Tanca et al., 2013; 2014; Fernández de Mera et al., 2017). Therefore, the integration of metaproteomics with metatranscriptomics provides a better resolution at the species level. For example, although several *Rickettsiella* spp. were identified at metatranscriptomics and metaproteomics levels, only a *Rickettsiella* endosymbiont of *Ixodes* spp. was identified by both analyses (Supplementary file 2 – Dataset 1 and Supplementary file 3 – Dataset 2). These bacteria are closely related to pathogenic *Rickettsiella* spp. (Cordaux et al., 2007), and the metaomics approach provided a better support for the presence of *Rickettsiella* endosymbionts in the *I. ricinus* microbiota. Differences in bacterial identification by metranscriptomics and metaproteomics may be also due to differences in RNA and protein levels (Fig. 2 and Table 2), which could be affected by post-transcriptional and post-translational modifications (Fan et al., 2013;

Ayllón et al., 2015; Villar et al., 2015). This suggestion was supported by the multiple bacterial proteins that were identified by metaproteomics when compared to results of metatranscriptomics analysis (Supplementary file 2 – Dataset 1 and Supplementary file 3 – Dataset 2). Additional support to integrated metatranscriptomics and metaproteomics results was provided at the DNA level by PCR (Fig. 3B).

3.3. Putative functional implications of integrated metaomics results.

This study is a "proof-of-concept" for the metaomics approach to tick microbiome characterization. Integrated metatranscriptomics and metaproteomics results were functionally more relevant than those obtained by metranscriptomics alone, suggesting that identified bacteria might form part of the active microbial community in unfed *I. ricinus*. Differences in bacterial microbiota composition have been attributed to variations between tick species, collection sites, sex and developmental stages, feeding status, and pathogen infection (Williams-Newkirk et al., 2014; van Treuren et al., 2015; Bonnet et al., 2017; Abraham et al., 2017; Xiang et al., 2017). A correlation analysis between metatranscriptomics and metaproteomics results revealed the absence of correlation for the entire bacterial microbiome, and for certain phyla such as Proteobacteria and Actinobacteria (Fig. 3C). However, for Firmicutes and TBPs a positive correlation was obtained between normalized RNA IDs and protein PSMs (Fig. 3C). Most of the metatranscriptomics data corresponded to rRNA, which is the predominant material in the ribosome and essential for protein synthesis (Cole et al., 2003). Therefore, a positive correlation between rRNA and protein levels may reflect that these bacteria were metabolically active in unfed *I. ricinus*. These commensal and environmental bacteria (Firmicutes) and TBPs may interact to affect multiple processes in the tick such as tick fitness, nutrition, development, reproduction, defense against environmental stress,

immunity and vector competence (Narasimhan and Fikrig, 2015; Bonnet et al., 2017; de la Fuente et al., 2017; Abraham et al., 2017).

The GO analysis for BP was conducted on proteins identified in TBPs (*Anaplasma*, *Borrelia*, *Ehrlichia* and *Rickettsia* genera) (Supplementary file 3 – Dataset 2) that showed a positive correlation between metatranscriptomics and metaproteomics results (Fig. 3C). The results showed that excluding unknown proteins, energy metabolism was the most abundant BP in all bacterial genera (Fig. 4). Another BP represented in all bacteria was protein synthesis (Fig. 4). Finally, BPs represented in some but not all bacterial genera included regulation of tick host gene expression, bacteria-tick interactions, DNA replication and transcription, DNA repair, nucleoside metabolism, cell wall biosynthetic process, cell division, and motility (Fig. 4). These results further supported that some of the identified bacteria may be metabolically active, and involved in tick-bacteria interactions.

The *I. ricinus* ticks used in this study were obtained from an uninfected reference laboratory colony. Then, why TBPs such as *A. phagocytophilum* and *Borrelia* spp. were identified by integrated metatranscriptomics and metaproteomics analysis? Two possible responses to this question are (a) that the colony may be infected with transovarially transmitted TBPs or (b) that although these bacteria were identified as TBPs, they may represent non-pathogenic genetic variants of these pathogens. In support to the last suggestion, it has been shown the presence of non-pathogenic species and/or variants in both *A. phagocytophilum* and *Borrelia* spp. (Anderson et al., 1990; Massung et al., 2002; 2003; Portillo et al., 2005; Al-Khedery and Barbet, 2014; Stokes et al., 2016). Nevertheless, their role as commensals or pathogenic potential is unknown.

Recently, Abraham et al. (2017) demonstrated that *A. phagocytophilum* manipulates tick microbiota through induction of *I. scapularis* antifreeze glycoprotein (IAFGP) that results

in alteration of bacterial biofilm formation to facilitate infection. Additionally, they showed that *A. phagocytophilum* alters the composition of the tick microbiota after midgut infection (Abraham et al., 2017). Based on our results and using commensal (*Streptococcus*) and *Anaplasma* bacteria that were identified in the *I. ricinus* microbiota as putatively metabolically active (Firmicutes and TBPs with a positive correlation between normalized RNA IDs and protein PSMs; Fig. 3C), we hypothesized that *Anaplasma* infection may be also facilitated by interfering with biofilm formation through reduction of the levels of biofilm matrix binding proteins (MBPs) and/or the presence of bacteria producing MBPs. Biofilm MBPs play structural roles in the biofilm formation and are produced by several commensal bacteria (Fong and Yildis, 2015), but in this study were identified only in *Streptococcus* (A0A0M4K0V8, A0A0M4JKY5, and E1M3U5; normalized PSMs = 0.06, Supplementary file 3 – Dataset 2). However, this hypothesis needs to be addressed by studies in ticks uninfected and experimentally infected with pathogens such as *A. phagocytophilum*.

4. Conclusions

The innovative metaomics approach used in this study resulted in the characterization of the bacterial microbiota in unfed *I. ricinus*. The approach proposed in this study for the characterization of tick microbiota was based on reused RNA-seq data. Although this is a limitation when compared to metagenomics approaches, it provides the opportunity to reuse already existing data for microboiota characterization. The reuse of RNA-seq data not targeted at bacterial sequences probably affects the identification of some bacterial taxa. However, the integration of metatrancriptomics and metaproteomics approaches provided a better characterization of tick bacterial microbiome when compared to RNA-seq alone by

increasing bacteria identification and support for identified bacteria with putative functional implications. The results corroborated previous reports on the presence of symbiotic, commensal, soil, environmental, and pathogenic bacteria in the *I. ricinus* microbiota. Additionally, previously unrecognized commensal and soil microorganisms were identified in unfed *I. ricinus*. The results of the metaomics approach suggested new mechanisms by which pathogen infection affects biofilm formation to manipulate tick microbiota and facilitate infection. Metaomics approaches integrating different omics datasets from metagenomics, metatranscriptomics and metaproteomics studies would provide a better description of tick microbiota composition, and insights into functional implication of tick interactions with microbiota, pathogens and hosts.

Conflict of interest statement

The authors declare that there are no conflicts of interest.

Acknowledgments

We thank Raquel Tobes and Marina Manrique (Oh no sequences! Research group, Era7 Bioinformatics, Granada, Spain) for technical assistance with the metatranscriptomics analysis. This work was financially supported by the H2020 COllaborative Management Platform for detection and Analyses of (Re-) emerging and foodborne outbreaks in Europe (COMPARE) Grant 643476. MV was supported by the Research Plan of the University of Castilla- La Mancha (UCLM), Spain.

Supplementary information

Supplementary file 1 – Table 1. Bacterial sequence database constructed with genome and/or species-specific rRNA sequences.

Supplementary file 2 - Dataset 1. Metatranscriptomics database of tick bacterial microbiota.

Supplementary file 3 - Dataset 2. Metaproteomics database of tick bacterial microbiota.

References

Abraham, N.M., Liu, L., Jutras, B.L., Yadav, A.K., Narasimhan, S., Gopalakrishnan, V., Ansari, J.M., Jefferson, K.K., Cava, F., Jacobs-Wagner, C., Fikrig, E., 2017. Pathogen-mediated manipulation of arthropod microbiota to promote infection. Proc. Natl. Acad. Sci. U. S. A. 114, E781-E790.

Aguiar-Pulido, V., Huang, W., Suarez-Ulloa, V., Cickovski, T., Mathee, K., Narasimhan, G., 2016. Metagenomics, metatranscriptomics, and metabolomics approaches for microbiome analysis. Evol. Bioinform. Online. 12(Suppl 1), 5-16.

Al-Khedery, B., Barbet, A.F., 2014. Comparative genomics identifies a potential marker of human-virulent *Anaplasma phagocytophilum*. Pathogens. 3, 25-35.

Anderson, J.F., Barthold, S.W., Magnarelli, L.A., 1990. Infectious but nonpathogenic isolate of *Borrelia burgdorferi*. J. Clin. Microbiol. 28, 2693-2699.

Andreotti, R., Perez de Leon, A.A., Dowd, S.E., Guerrero, F.D., Bendele, K.G., Scoles, G.A., 2011. Assessment of bacterial diversity in the cattle tick *Rhipicephalus* (*Boophilus*) *microplus* through tag-encoded pyrosequencing. BMC Microbiol. 11, 6.

Ayllón, N., Villar, V., Galindo, R.C., Kocan, K.M., Šíma, R., López, J.A., Vázquez, J., Alberdi, P., Cabezas-Cruz, A., Kopáček, P., de la Fuente, J., 2015. Systems biology of tissue-specific response to *Anaplasma phagocytophilum* reveals differentiated apoptosis in the tick vector *Ixodes scapularis.* PLoS Genet. 11, e1005120.

Benson, M.J., Gawronski, J.D., Eveleigh, D.E., Benson, D.R., 2004. Intracellular symbionts and other bacteria associated with deer ticks (*Ixodes scapularis*) from Nantucket and Wellfleet, Cape Cod, Massachusetts. Appl. Environ. Microbiol. 70, 616-620.

Bonnet, S., Michelet, L., Moutailler, S., Cheval, J., Hébert, C., Vayssier-Taussat, M., Eloit, M., 2014. Identification of parasitic communities within European ticks using next-generation sequencing. PLoS Negl. Trop. Dis. 8, e2753.

Bonnet, S., Binetruy, F., Hernández-Jarguín, A., Duron, O., 2017. The tick microbiome: Why non-pathogenic microorganisms matter in tick biology and pathogen transmission. Front. Cell. Infect. Microbiol. 7, 236.

Bouchez, T., Blieux, A.L., Dequiedt, S., Domaizon, I., Dufresne, A., Ferreira, S., Godon, J.J., Hellal, J., Joulian, C., Quaiser, A., Martin-Laurent, F., Mauffret, A., Monier, J. M., Peyret, P., Schmitt-Koplin, P., Sibourg, O., D'oiron, E., Bispo, A., Deportes, I., Grand, C., Cuny, P., Maron, P.A., Ranjard, L., 2016. Molecular microbiology methods for environmental diagnosis. Environ. Chem. Lett. 14, 423-441.

Budachetri, K., Browning, R. E., Adamson, S. W., Dowd, S. E., Chao, C.C., Ching, W.M., Karim, S., 2014. An insight into the microbiome of the *Amblyomma maculatum* (Acari: Ixodidae). J. Med. Entomol. 51, 119-129.

Cafiso, A., Bazzocchi, C., De Marco, L., Opara, M.N., Sassera, D., Plantard, O., 2016. Molecular screening for *Midichloria* in hard and soft ticks reveals variable prevalence levels and bacterial loads in different tick species. Ticks Tick Borne Dis. 6, 1186-1192.

Carpi, G., Cagnacci, F., Wittekindt, N. E., Zhao, F., Qi, J., Tomsho, L. P., Drautz, D.I., Rizzoli, A., Schuster, S.C., 2011. Metagenomic profile of the bacterial communities associated with *Ixodes ricinus* ticks. PLoS ONE. 6, e25604.

Cheng, K., Ning, Z., Zhang, X., Li, L., Liao, B., Mayne, J., Stintzi, A., Figeys, D., 2017. MetaLab: an automated pipeline for metaproteomic data analysis. Microbiome. 5, 157.

Clay, K., Klyachko, O., Grindle, N., Civitello, D., Oleske, D., Fuqua, C., 2008. Microbial communities and interactions in the lone star tick, *Amblyomma americanum*. Mol. Ecol. 17, 4371-4381.

Cole, J.R., Chai, B., Marsh, T.L., Farris, R.J., Wang, Q., Kulam, S.A., Chandra, S., McGarrell, D.M., Schmidt, T.M., Garrity, G.M., Tiedje, J.M., Ribosomal Database Project, 2003. The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. Nucleic Acids Res. 31, 442-443.

Cordaux, R., Paces-Fessy, M., Raimond, M., Michel-Salzat, A., Zimmer, M., Bouchon, D., 2007. Molecular characterization and evolution of arthropod-pathogenic *Rickettsiella* bacteria. Appl. Environ. Microbiol. 73, 5045-5047.

Courtney, J.W., Kostelnik, L.M., Zeidner, N.S., Massung, R.F., 2004. Multiplex real-time PCR for detection of *Anaplasma phagocytophilum* and *Borrelia burgdorferi*. J. Clin. Microbiol. 42, 3164-3168.

de la Fuente, J., Estrada-Peña, A., Venzal, J.M., Kocan, K.M., and Sonenshine, D.E., 2008. Overview: Ticks as vectors of pathogens that cause disease in humans and animals. Front. Biosci. 13, 6938-6946.

de la Fuente, J., Antunes, S., Bonnet, S., Cabezas-Cruz, A., Domingos, A.G., Estrada-Peña, A., Johnson, N., Kocan, K.M., Mansfield, K.L., Nijhof, A.M., Papa, A., Rudenko, N., Villar, M., Alberdi, P., Torina, A., Ayllón, N., Vancova, M., Golovchenko, M., Grubhoffer, L., Caracappa, S., Fooks, A.R., Gortazar, C., Rego, R.O.M., 2017. Tick-pathogen interactions and vector competence: identification of molecular drivers for tick-borne diseases. Front. Cell. Infect. Microbiol. 7, 114.

Evans, V.C., Barker, G., Heesom, K.J., Fan, J., Bessant, C., Matthews, D.A., 2012. De novo derivation of proteomes from transcriptomes for transcript and protein identification. Nat. Methods. 9, 1207-1211.

Fan, Y., Thompson, J. W., Dubois, L. G., Moseley, M. A., Wernegreen, J.J., 2013. Proteomic analysis of an unculturable bacterial endosymbiont *(Blochmannia)* reveals high abundance of chaperonins and biosynthetic enzymes. J. Proteome Res. 12, 704-718.

Fernández de Mera, I.G., Ruiz-Fons, F., Mangold, K.J., Gortázar, C., de la Fuente, J., 2013. Spotted Fever Group rickettsiae in questing ticks, Central Spain. Emerg. Infect. Dis. 19, 1163-1165

Fernández de Mera, I.G., Chaligiannis, I., Hernández-Jarguín, A., Villar, M., Mateos-Hernández, L., Papa, A., Sotiraki, S., Ruiz-Fons, F., Cabezas-Cruz, A., Gortázar, C., de la Fuente, J., 2017. Combination of RT-PCR and proteomics for the identification of Crimean-Congo hemorrhagic fever virus in ticks. Heliyon. 3, e00353.

Franzosa, E.A., Hsu, T., Sirota-Madi, A., Shafquat, A., Abu-Ali, G., Morgan, X.C., Huttenhower, C., 2015. Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. Nat. Rev. Microbiol. 13, 360-372.

Frith, M. C., Hamada, M., Horton, P., 2010a. Parameters for accurate genome alignment. BMC Bioinformatics. 11, 80.

Frith, M. C., Wan, R., Horton, P., 2010b. Incorporating sequence quality data into alignment improves DNA read mapping. Nucleic Acids Res. 38, e100.

Gall, C. A., Reif, K. E., Scoles, G. A., Mason, K. L., Mousel, M., Noh, S. M., Brayton, K. A., 2016. The bacterial microbiome of *Dermacentor andersoni* ticks influences pathogen susceptibility. ISME J. 10, 1846-1855.

Genomic Resources Development Consortium, Contreras, M., de la Fuente, J., Estrada-Peña, A., Grubhoffer, L., Tobes, R., 2014. Transcriptome sequence divergence between Lyme disease tick vectors, *Ixodes scapularis* and *Ixodes ricinus.* Genomic Resources Notes accepted 1 April 2014 – 31 May 2014. Mol. Ecol. Resour. 14, 1095.

Greay, T.L., Gofton, A.W., Paparini, A., Ryan, U.M., Oskam, C.L., Irwin, P.J., 2018. Recent insights into the tick microbiome gained through next-generation sequencing. Parasit. Vectors. 11, 12.

Heintz-Buschart, A., Wilmes, P., 2017. Human gut microbiome: function matters. Trends Microbiol. In press.

Johansson, H., Dhaygude, K., Lindström, S., Helanterä, H., Sundström, L., Trontti, K., 2013. A Metatranscriptomic approach to the identification of microbiota associated with the ant (*Formica exsecta)*. PLoS ONE. 8: e79777.

Kau, A.L., Ahern, P.P., Griffin, N.W., Goodman, A.L., Gordon, J.I., 2011. Human nutrition, thegut microbiome and the immune system. Nature. 474, 327-336.

Kiełbasa, S. M., Wan, R., Sato, K., Horton, P., Frith, M. C., 2011. Adaptive seeds tame genomic sequence comparison. Genome Res. 21, 487-493.

Koči, J., Šimo, L., and Park, Y., 2013. Validation of internal reference genes for real-time quantitative polymerase chain reaction studies in the tick, *Ixodes scapularis* (Acari: Ixodidae). J. Med. Entomol. 50, 79-84.

Kondo, N., Ijichi, N., Shimada, M., Fukatsu, T., 2002. Prevailing triple infection with *Wolbachia* in *Callosobruchus chinensis* (Coleoptera: Bruchidae). Mol. Ecol. 11, 167-180.

Kurilshikov, A., Livanova, N. N., Fomenko, N. V., Tupikin, A. E., Rar, V. A., Kabilov, M. R., Tikunova, N. V., 2015. Comparative metagenomic profiling of symbiotic bacterial communities associated with *Ixodes persulcatus*, *Ixodes pavlovskyi* and *Dermacentor reticulatus* ticks. PloS ONE. 10, e0131413.

Letunic, I., Bork, P., 2011. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res. 39, 475-478.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The sequence alignment/map format and SAMtools. Bioinformatics. 25, 2078-2079.

Li, H., 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 27, 2987-2993.

Liu, L., Li, L., Liu, J., Yu, Z., Yang, X., Liu, J., 2016. Population dynamics of multiple symbionts in the hard tick, *Dermacentor silvarum* Olenev (Acari: Ixodidae). Ticks Tick Borne Dis. 7, 188-192.

Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real time quantitative PCR and the $2^{-\Delta\Delta}$ CT method. Methods. 25, 402-408.

Luo, J., Liu, M.X., Ren, Q.Y., Chen, Z., Tian, Z.C., Hao, J.W., Wu, F., Liu, X.C., Luo, J.X., Yin, H., Wang, H., Liu, G.Y., 2017. Micropathogen community analysis

in *Hyalomma rufipes* via high-throughput sequencing of small RNAs. Front. Cell. Infect. Microbiol. 7, 374.

Mäder, U., Nicolas, P., Richard, H., Berriéres, P., Aymerich, S., 2011. Comprehensive identification and quantification of microbial transcriptomes by genome-wide unbiased methods. Curr. Opin. Biotechnol. 22, 32-41.

Martin, A.R., Brown, G.K., Dunstan, R.H., Roberts, T.K., 2005. *Anaplasma platys*: an improved PCR for its detection in dogs. Exp. Parasitol. 109, 176-180.

Massung, R.F., Mauel, M.J., Owens, J.H., Allan, N., Courtney, J.W., Stafford, K.C. 3rd., Mather, T.N., 2002. Genetic variants of *Ehrlichia phagocytophila*, Rhode Island and Connecticut. Emerg. Infect. Dis. 8, 467-472.

Massung, R.F., Priestley, R.A., Miller, N.J., Mather, T.N., Levin, M.L., 2003. Inability of a variant strain of *Anaplasma phagocytophilum* to infect mice. J. Infect. Dis. 188, 1757-1763.

Moreno, C.X., Moy, F., Daniels, T.J., Godfrey, H.P., Cabello, F.C., 2006. Molecular analysis of microbial communities identified in different developmental stages of *Ixodes scapularis* ticks from Westchester and Dutchess Counties, New York. Environ. Microbiol. 8, 761-72.

Nakao, R., Abe, T., Nijhof, A. M., Yamamoto, S., Jongejan, F., Ikemura, T., Sugimoto, C., 2013. A novel approach, based on BLSOMs (Batch Learning Self- Organizing Maps) to the microbiome analysis of ticks. ISME J. 7, 1003-1015.

Narasimhan, S., Fikrig, E., 2015. Tick microbiome: the force within. Trends Parasitol. 31, 315-323.

Narasimhan, S., Rajeevan, N., Liu, L., Zhao, Y.O., Heisig, J., Pan, J., Eppler-Epstein, R., Deponte, K., Fish, D., Fikrig, E., 2014. Gut microbiota of the tick vector *Ixodes scapularis* modulate colonization of the Lyme disease spirochete. Cell Host Microbe. 15, 58-71.

Narasimhan, S., Schuijt, T.J., Abraham, N.M., Rajeevan, N., Coumou, J., Graham, M., Robson, A., Wu, M.J., Daffre, S., Hovius, J.W., Fikrig, E., 2017. Modulation of the tick gut milieu by a secreted tick protein favors *Borrelia burgdorferi* colonization. Nat. Commun. 8, 184.

Noda, A.A., Rodríguez, I., Mondeja, B., Fernándeze, C., 2013. Design, optimization and evaluation of a polymerase chain reaction for detection of *Borrelia* spp. Adv. Clin. Exp. Med. 5, 639-653.

Oteo, J.A., Portillo, A., Santibáñez, S., Blanco, J.R., Pérez-Martínez, L., Ibarra, V., 2006. Cluster of cases of human *Rickettsia felis* infection from Southern Europe (Spain) diagnosed by PCR. J. Clin. Microbiol. 44, 2669-2671.

Philippot, L., Raaijmakers, J.M., Lemanceau, P., van der Putten, W.H., 2013. Going back to the roots: the microbial ecology of the rhizosphere. Nat. Rev. Microbiol. 11, 789-799.

Ponnusamy, L., González, A., Van Treuren, W., Weiss, S., Parobek, C., Juliano, J.J, Knight, R., Roe, R.M., Apperson, C.S., Meshnick, S.R., 2014. Diversity of Rickettsiales in the microbiome of the lone star tick, *Amblyomma americanum*. Appl. Env. Microbiol. 80, 354-359.

Portillo, A., Santos, A.S., Santibáñez, S., Pérez-Martínez, L., Blanco, J.R., Ibarra, V., Oteo, J.A., 2005. Detection of a non-pathogenic variant of *Anaplasma phagocytophilum* in *Ixodes ricinus* from La Rioja, Spain. Ann. N. Y. Acad. Sci. 1063, 333-336.

Qiu, Y., Nakao, R., Ohnuma, A., Kawamori, F., Sugimoto, C., 2014. Microbial population analysis of the salivary glands of ticks: a possible strategy for the surveillance of bacterial pathogens. PLoS ONE. 9, e103961.

Razzauti, M., Galan, M., Bernard, M., Maman, S., Klopp, C., Charbonnel, N., Vayssier-Taussat, M., Eloit, M., Cosson, J.F., 2015. A comparison between transcriptome sequencing and 16S metagenomics for detection of bacterial pathogens in wildlife. PLoS Negl. Trop. Dis. 9, e0003929.

Ririe, K.M., Rasmussen, R.P., Wittwer, C.T., 1997. Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. Anal. Biochem. 245, 154-160.

Ruiz-Fons, F., Fernández de Mera, I., G., Acevedo, P., Gortázar, C., de la Fuente, J., 2012. Factors driving the abundance of *Ixodes ricinus* ticks and the prevalence of zoonotic I. ricinus-borne pathogens in natural foci. Appl. Environ. Microbiol. 8, 2669-2676.

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M.F., Turner, P., Parkhill, J., Loman, N.J., Walker, A.W., 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol. 12, 87.

Schabereiter-Gurtner, C., Lubitz, W., Rolleke, S., 2003. Application of broad-range 16S rRNA PCR amplification and DGGE fingerprinting for detection of tick-infecting bacteria. J. Microbiol. Methods. 52, 251-260.

Schevchenko, A., Tomas, H., Havlis, J., Olsen, J.V., Mann, M., 2006. In-gel digestion for mass-spectrometric characterization of proteins and proteomes. Nat. Protoc. 1, 2856-2860.

Schwabe, R.F., Jobin, C., 2013. The microbiome and cancer. Nat. Rev. Cancer. 13, 800-812.

Selvaratnam, S., Schoedel, B. A., McFarland, B. L., Kulpa, C. F., 1995. Application of reverse transcriptase PCR for monitoring expression of the catabolic *dmpN* gene in a phenol-degrading sequencing batch reactor. Appl. Environ. Microbiol. 61, 3981-3985.

Stokes, J.V., Moraru, G.M., McIntosh, C., Kummari, E., Rausch, K., Varela-Stokes, A.S., 2016. Differentiated THP-1 cells exposed to pathogenic and nonpathogenic *Borrelia* species demonstrate minimal differences in production of four inflammatory cytokines. Vector Borne Zoonotic Dis. 16, 691-695.

Tanca, A., Palomba, A., Deligios, M., Cubeddu, T., Fraumene, C., Biosa, G., Pagnozzi, D., Addis, M.F., Uzzau, S., 2013. Evaluating the impact of different sequence databases on metaproteome analysis: Insights from a lab-assembled microbial mixture. PLoS One. 8, e82981.

Tanca, A., Palomba, A., Pisanu, S., Deligios, M., Fraumene, C., Manghina, V., Pagnozzi, D., Addis, M.F., Uzzau, S., 2014. A straightforward and efficient analytical pipeline for metaproteome characterization. Microbiome. 2, 49.

Trout Fryxell, R., DeBruyn, J., 2016. The microbiome of *Ehrlichia*-infected and uninfected lone star ticks (*Amblyomma americanum*). PLoS ONE. 11, 1-19.

van Overbeek, L., Gassner, F., van der Plas, C. L., Kastelein, P., Nunes-da Rocha, U., Takken, W., 2008. Diversity of *Ixodes ricinus* tick-associated bacterial communities from different forests. FEMS Microbiol. Ecol. 66, 72-84.

Van Treuren, W., Ponnusamy, L., Brinkerhoff, R.J., Gonzalez, A., Parobek, C.M., Juliano, J.J., Andreadis, T.G., Falco, R.C., Ziegler, L.B., Hathaway, N., Keeler, C., Emch, M., Bailey, J.A., Roe, R.M., Apperson, C.S., Knight, R., Meshnick, S.R., 2015. Variation in the

microbiota of *Ixodes* ticks with regard to geography, species, and sex. Appl. Environ. Microbiol. 81, 6200–6209.

Varela-Stokes, A.S., Park, S.H., Kim, S.A., Ricke, S.C., 2017. Microbial communities in North American ixodid ticks of veterinary and medical importance. Front. Vet. Sci. 4, 179.

Vayssier-Taussat, M., Moutailler, S., Michelet, L., Devillers, E., Bonnet, S., Cheval, J., Hébert, C., Eloit, M., 2013. Next deneration sequencing uncovers unexpected bacterial pathogens in ticks in Western Europe. PLoS One. 8, e81439.

Vayssier-Taussat, M., Kazimirova, M., Hubalek, Z., Hornok, S., Farkas, R., Cosson, J.F., Bonnet, S., Vourch, G., Gasqui, P., Mihalca, A.D., Plantard, O., Silaghi, C., Cutler, S., Rizzoli, A., 2015. Emerging horizons for tick-borne pathogens: from the "one pathogen-one disease" vision to the pathobiome paradigm. Future Microbiol. 10, 2033-2043.

Villar, M., Ayllón, N., Alberdi, P., Moreno, A., Moreno, M., Tobes, R., Mateos-Hernández, L., Weisheit, S., Bell-Sakyi, L., de la Fuente, J., 2015. Integrated metabolomics, transcriptomics and proteomics identifies metabolic pathways affected by *Anaplasma phagocytophilum* infection in tick cells. Mol. Cell. Proteomics. 14, 3154-3172.

Williams-Newkirk, A.J., Rowe, L.A., Mixson-Hayden, T.R., Dasch, G.A., 2014. Characterization of the bacterial communities of life stages of free living lone star ticks (*Amblyomma americanum*). PLoS ONE. 9, e102130.

Xiang, L., Pożniak, B., Cheng, T., 2017. Bacteriological analysis of saliva from partially or fully engorged female adult *Rhipicephalus microplus* by next-generation sequencing. Antonie Van Leeuwenhoek. 110, 105-113.

Xu, X. L., Cheng, T. Y., Yang, H., Yan, F., 2015. Identification of intestinal bacterial flora in *Rhipicephalus microplus* ticks by conventional methods and PCR-DGGE analysis. Exp. Appl. Acarol. 66, 257-268.

Yoon, S.S., Kim, E.K., Lee, W.J., 2015. Functional genomic and metagenomic approaches to understanding gut microbiota-animal mutualism. Curr. Opin. Microbiol. 24, 38-46.

Zhuang, L., Zhang, Z., An, X., Fan, H., Ma, M., Anderson, B.D., Jiang, J., Liu, W., Cao, W., Tong, Y., 2014a. An efficient strategy of screening for pathogens in wild-caught ticks and mosquitoes by reusing small RNA deep sequencing data. PLoS One. 9, e90831.

Zhuang, L., Wang, C.Y., Tong, Y.G., Tang, F., Yang, H., Liu, W., Cao, W.C., 2014b. Discovery of *Rickettsia* species in *Dermacentor niveus Neumann* ticks by investigating the diversity of bacterial communities. Ticks Tick Borne Dis. 5, 564-568.

Zivkovic, Z., Blouin, E.F., Manzano-Roman, R., Almazán, C., Naranjo, V., Massung, R.F., Jongejan, F., Kocan, K.M., de la Fuente, J., 2009. *Anaplasma phagocytophilum* and *Anaplasma marginale* elicit different gene expression responses in cultured tick cells. Comp. Funct. Genomics. 2009, 705034.

Figure legends

Figure 1. Metaomics experimental design. An integrated metatranscriptomics and metaproteomics approach was developed for the characterization of *I. ricinus* bacterial microbiota. Reused *I. ricinus* RNA-seq data was the basis for metatranscriptomics analysis that resulted in the database of identified bacterial genera. This database was then use to generate the Uniprot protein database used in metaproteomics analysis.

Figure 2. Phylogenetic and taxonomic abundance analyses. Phylogenetic pruned tree and associated heatmap showing the relative abundance of bacterial genomes (number of count reads or identifications, IDs) and the correspondent peptide assignments (peptide spectrum matches, PSMs) at genus level. The analyses were done using a Heatmap Tool associated with a pruned phylogenetic tree generated with the platform phyloT (http://phylot.biobyte.de) based on NCBI taxonomy, and visualized using the Interactive Tree of Life software v3.4.3 (http://itol.embl.de).

Figure 3. Comparative analysis of metranscriptomics and metaproteomics results. (A) Representation of the bacterial phyla with the corresponding number of genera (species) identified at RNA/protein levels. (B) Validation of integrated metatranscriptomics and metaproteomics results by PCR. Bacterial DNA levels were determined by real-time PCR and normalized against tick *16S rRNA* and *Ixodes rps 4*. Oligonucleotide primers and real-time PCR conditions are described in Table 2. DNA levels are shown in arbitrary units as normalized Ct values. Two experiments were conducted with similar results. (C) Correlation analysis was conducted between normalized number of count reads or identifications (IDs; metatranscriptomics) and peptide spectrum matches (PSMs; metaproteomics) results (Table 2) corresponding to all bacterial genera (N=38),

Proteobacteria (N=21), Actinobacteria (N=7) and Firmicutes (N=4) phyla represented by 4 genera or more, and TBPs (genera *Anaplasma*, *Ehrlichia*, *Rickettsia*, and *Borrelia*). The linear correlation coefficients (R^2) and Pearson correlation coefficients (*r > +0.5) are shown. For reference, *Streptococcus* and *Anaplasma* genera are shown in black and red rhombuses, respectively.

Figure 4. Protein annotation in identified tick-borne pathogens. The GO analysis for BP was conducted on proteins identified in TBPs (*Anaplasma*, *Borrelia*, *Ehrlichia* and *Rickettsia* genera). Quantitative representation of BP abundance (%) was done using the total number of PSMs represented on each BP.

638 **Table 1. Oligonucleotide primers and real-time PCR conditions.**

Organism	Gene	Primer sequences (5-′3′)	Annealing temperature	References
Ixodidae	*16S rRNA*	GACAAGAAGACCCTA ATCCAACATCGAGGT	42 °C	Zivkovic et al., 2009
Ixodes spp.	*rps 4*	GGTGAAGAAGATTGTCAAGCAGAG TGAAGCCAGCAGGGTAGTTTG	54 °C	Koči et al., 2013
Rickettsia spp.	*16S rRNA*	AGAGTTTGATCCTGGCTCAG AACGTCATTATCTTCCTTGC	54 °C	Fernández de Mera et al., 2013
SFG *Rickettsia*	*ompA*	ATGGCGAATATTTCTCCAAAA AGTGCAGCATTCGCTCCCCCT	52 °C	Oteo et al., 2006
Anaplasma spp.	*16S rRNA*	CAGAGTTTGATCCTGGCTCAGAACG GAGTTTGCCGGGACTTCTTCTGTA	42 °C	Ruiz-Fons et al., 2012
Anaplasma phagocytophilum	*msp2*	ATGGAAGGTAGTGTTGGTTATGGTATT TTGGTCTTGAAGCGCTCGTA	60 °C	Courtney et al., 2004
Ehrlichia spp.	*16S rRNA*	GGTACCYACAGAAGAAGTCC TAGCACTCATCGTTTACAGC	54 °C	Martin et al., 2005
Pseudomonas putida	*dmpN*	ATCACCGACTGGGACAAGTGGGAAGACC TGGTATTCCAGCGGTGAAACGGCGG	50 °C	Selvaratnam et al., 1995
Wolbachia spp.	*wsp*	GGGTCCAATAAGTGATGAAGAAAC TTAAAACGCTACTCCAGCTTCTGC	55 °C	Kondo et al., 2002
Candidatus Midichloria mitochondrii	*16S rRNA*	CAAAAGTGAAAGCCTTGGGC TGAGACTTAAAYCCCAACATC	58 °C	Cafiso et al., 2016
Borrelia spp.	*16S rRNA*	TAGATGAGTCTGCGTCTTATTA CTTACACCAGGAATTCTAACTT	58 °C	Noda et al., 2013

Table 2. Relative bacteria IDs and PSMs obtained at genus level.

Phylum	Genus	IDs	PSMs
Proteobacteria	*Acinetobacter*	0.22	0.16
	Anaplasma	10.67	0.73
	Bradyrhizobium	2.44	11.07
	Brevibacterium	1.78	0.77
	Brucella	0.22	0.38
	Candidatus Midichloria	15.55	0.05
	Candidatus Neoehrlichia	8.22	0.00
	Ehrlichia	12.89	0.63
	Escherichia	0.89	4.67
	Francisella	0.22	0.00
	Gilliamella	0.22	1.23
	Klebsiella	0.22	4.24
	Lysobacter	0.22	0.00
	Mesorhizobium	0.22	1.34
	Neorickettsia	0.67	0.06
	Pseudomonas	0.67	14.72
	Rickettsia	0.44	0.18
	Rickettsiella	1.78	0.15
	Sphingomonas	0.22	9.55
	Variovorax	0.22	3.92

	Wolbachia	0.67	0.00
Actinobacteria	*Actinomyces*	1.33	2.90
	Arthrobacter	0.44	4.44
	Amycolatopsis	0.22	0.23
	Bifidobacterium	0.44	0.68
	Corynebacterium	1.55	5.70
	Propionibacterium	0.89	0.31
	Rhodococcus	0.22	7.17
Firmicutes	*Enterococcus*	0.44	5.06
	Kurthia	0.22	0.15
	Staphylococcus	0.44	0.48
	Streptococcus	0.89	12.20
Spirochaetes	*Borrelia*	20.67	1.23
	Treponema	0.22	0.00
Bacteroidetes	*Mucilaginibacter*	0.22	1.72
Fusobacteria	*Leptotrichia*	0.22	0.27
Tenericutes	*Spiroplasma*	10.67	1.03
Uncultured bacteria	Uncultured bacteria	2.22	2.63

642

643 Normalized data is shown as the number of count reads or identifications/ peptide spectrum

644 matches (IDs/PSMs) divided by the total number of IDs/PSMs x 100.

Table 3. Biological information about bacterial genera identified by metatranscriptomics in *I. ricinus*, and previously reported in different tick species.

Bacterial genera	Biological information	Tick spp.	References
Actinomyces	Commensals of the caecum gut flora and oral cavities in human. Responsible of abscesses formation in the mouth, lungs and gastrointestinal tract.	Data not found	
Acinetobacter	Soil organisms. Some species are opportunistic pathogens causing human infections.	*I. scapularis* *I. ricinus* *I. ovatus* *I. persulcatus* *Haemaphysalis flava* *Amblyomma americanum* *Dermacentor niveus*	Benson et al., 2004; Moreno et al., 2006; van Overbeek et al., 2008; Qiu et al., 2014; Clay et al., 2008; Zhuang et al., 2014b; Narashiman et al., 2014; 2017; Abraham et al., 2017
Amycolatopsis	Soil bacteria with antibiotic and anti-inflammatory properties.	Data not found	
Anaplasma	Tick-borne intracellular bacterial pathogens causing diseases in humans and animals.	*I. scapularis* *I. persulcatus* *I. pavlovskyi* *I. ricinus*	Benson et al., 2004; Moreno et al., 2006; Kurilshikov et al., 2014; van Overbeek et al., 2008
Arthrobacter	Soil bacteria associated to bioremediation processes.	*D. niveus*	Zhuang et al., 2014b
Bidifidobacterium	Commensal and symbiotic bacteria in the human body. Produce lactic acid that modulates the intestinal pH.	Data not found	
Borrelia	Vector-borne bacteria responsible for Lyme disease and relapsing fever in mammals.	*Rhipicephalus microplus* *I. ricinus* *I. scapularis* *I. persulcatus* *I. pavlovskyi* *A. americanum*	Andreotti et al., 2011; Carpi et al., 2011;Vayssier-Taussat et al., 2013; van Overbeek et al., 2008; Schabereiter-Gurtner et al., 2003; Moreno et al., 2006; Kurilshikov et al.,

			2014 Clay et al., 2008
Bradyrhizobium	Soil bacteria, present in the roots of plant fixing N_2.	*I. ovatus* *I. persulcatus* *H. flava*	Qiu et al., 2014
Brevibacterium	Soil bacteria, some species can be found in human skin.	*I. ovatus* *I. scapularis*	Qiu et al., 2014; Narashiman et al., 2014; 2017; Abraham et al., 2017
Brucella	Pathogenic bacteria causing disease in human and animals.	*I. ricinus*	Carpi et al., 2011
Candidatus Midichloria	Tick endosymbiotic bacteria.	*A. americanum* *I. ricinus*	Ponnusamy et al., 2014; Trout Fryxell and DeBuyn, 2016; van Overbeek et al., 2008
Candidatus Neoehrlichia	Tick endosymbiotic bacteria.	*I. ricinus* *I. pavlovskyi*	Carpi et al., 2011; Vayssier-Taussat et al., 2013; van Overbeek et al., 2008; Kurilshikov et al., 2014
**Corynebacterium*	Saprophytes, some species are pathogenic for plants and animals.	*R. microplus* *I. ovatus* *I. persulcatus* *H. flava*	Andreotti et al., 2011; Qiu et al., 2014
Ehrlichia	Tick-borne intracellular bacterial pathogens causing diseases in humans and animals.	*I. scapularis* *I. persulcatus* *I. ovatus* *H. flava* *D. reticulatus* *R. microplus* *A. americanum*	Benson et al., 2004; Kurilshikov et al., 2014; Qiu et al., 2014; Xu et al., 2015; Clay et al., 2008
Enterococcus	Commensals of digestive tract, opportunistic pathogens causing septicemia and urinary tract infection in mammals.	*R. microplus* *I. ovatus* *I. persulcatus* *A. americanum*	Andreotti et al., 2011; Qiu et al., 2014; Clay et al., 2008
Escherichia	Commensals of digestive and urinary tracts, opportunistic pathogens causing diarrhea to dysentery in mammals.	*R. microplus* *D. silvarum* *D. niveus* *I. persulcatus*	Andreotti et al., 2011; Liu et al., 2016; Zhuang et al., 2014b; Qiu et al., 2014
Francisella/	Intracellular bacterial	*D. reticulatus*	Kurilshikov et al.,

Francisella-like endosymbiont	pathogens transmitted by vectors (ticks, mosquitoes, flies), and causing tularemia.	*D. andersoni* *I. ricinus* *I. ovatus* *I. persulcatus* *A. maculatum*	2014; Gall et al., 2016; Vayssier-Taussat et al., 2013; Nakao et al., 2013; Budachetri et al., 2014
Giliamella	Bee endosymbiotic bacteria.	Data not found	
Klebsiella	Saprophytes in soil and water, commensals of gastrointestinal tract, opportunistic pathogens responsible for septicemia and pneumonia in mammals	*R. microplus*	Andreotti et al., 2011
Kurthia	Environmental bacteria, present in mammal feces, soil and water. Opportunistic pathogens for humans causing endocarditis.	Data not found	
Leptotrichia	Natural flora in humans, some species cause opportunistic infections.	*I. persulcatus*	Qiu et al., 2014
Mesorhizobium	Soil bacteria, present in the roots of plant fixing N_2.	Data not found	
Mucilaginibacter	Environmental bacteria	*H. longicornis*	Liu et al., 2016
Neorickettsia/ Rickettsia	Intracellular bacteria, transmitted by vectors (ticks, fleas, chiggers, lice), responsible for human diseases such as spotted fever and typhus.	*I. scapularis* *I. ovatus* *I. affinis* *I. persulcatus* *I. ricinus* *I. pavlovskyi* *D. andersoni* *D. reticulatus* *D. niveus* *H. longicornis* *H. formosensis* *H. flava* *A.testudinarium* *A. americanum* *A. maculatum* *R. microplus*	Benson et al., 2004; Moreno et al., 2006; van Treuren et al., 2015; Qiu et al., 2014; Nakao et al., 2013; van Treuren et al., 2015; Kurilshikov et al., 2014; Zhuang et al., 2014b; Williams-Newkirk et al., 2014; Carpi et al., 2011; van Overbeek et al., 2008; Vayssier-Taussat et al., 2013; Schabereiter-Gurtner et al., 2003; Liu et al., 2016; Gall et al., 2016; Ponnusamy et al.,

			2014; Trout Fryxell and DeBuyn, 2016; Clay et al., 2008; Budachetri et al., 2014; Xiang et al., 2017; Xu et al., 2015
Propionibacterium	Commensals of human gut and skin.	*I. ricinus* *I. ovatus* *I. persulcatus* *H. flava*	Carpi et al., 2011; Qiu et al., 2014
Pseudomonas	Saprophytes in soil, opportunistic pathogens for humans and plants, plant growth promoters.	*R. microplus* *I. ricinus* *I. scapularis* *I. persulcatus* *I. pavlovsky* *I. ovatus* *H. longicornis* *D. niveus* *A. americanum*	Andreotti et al., 2011; Xu et al., 2015; Schabereiter-Gurtner et al., 2003; Carpi et al., 2011; Moreno et al., 2006; Qiu et al., 2014; Kurilshikov et al., 2014; Liu et al., 2016; Zhuang et al., 2014b; Clay et al., 2008
**Rhodococcus*	Saprophytes in soil and water, one species is pathogenic for animals causing pneumonia.	*R. microplus* *I. ricinus* *I. persulcatus* *I. ovatus* *I. pavlovsky* *I. scapularis*	Andreotti et al., 2011; Carpi et al., 2011; Schabereiter-Gurtner et al., 2003; Kurilshikov et al., 2014; Qiu et al., 2014; Moreno et al., 2006
Rickettsiella endosymbiont	Tick endosymbiotic bacteria.	*I. pavlovsky* *A. variegatum*	Kurilshikov et al., 2014; Nakao et al., 2013
Sphingomonas	Environmental bacteria and bioremediation agents, some specimens cause clinical infections in humans.	*I. scapularis* *I. ovatus* *I. persulcatus* *H. longicornis* *H. flava*	Benson et al., 2004; Qiu et al., 2014; Liu et al., 2016
Sphingobium	Commonly isolated from soil	*D. niveus* *I. ovatus*	Zhuang et al., 2014b; Qiu et al., 2014
Spiroplasma	Symbionts in the gut hemolymph, few species are pathogenic for mice (cataracts and neurological	*I. ovatus* *I. persulcatus* *H. flava*	Qiu et al., 2014

	damage)		
Staphylococcus	Saprophytes in soil, commensals of skin and mucosal surfaces, opportunistic pathogens (septicemia, food poisoning)	*R. microplus* *D. nievus* *I. ricinus* *I. ovatus* *I. persulcatus* *H. flava*	Andreotti et al., 2011; Xu et al., 2015; Zhuang et al., 2014b;Schabereiter-Gurtner et al., 2003; Qiu et al., 2014
Streptococcus	Saprophytes in soil and water, commensals of skin and mucosal surfaces, opportunistic pathogens (septicemia, meningitis, pneumonia)	*R. microplus* *I. scapularis*	Andreotti et al., 2011; Benson et al., 2004
Variovorax	Soil bacterium associated with bioremediation processes	Data not found	
Wolbachia	Mutualistic bacteria of many insects and nematodes	*R. microplus* *I. scapularis* *I. ricinus*	Andreotti et al., 2011; Benson et al., 2004; Carpi et al., 2011; van Overbeek et al., 2008

649

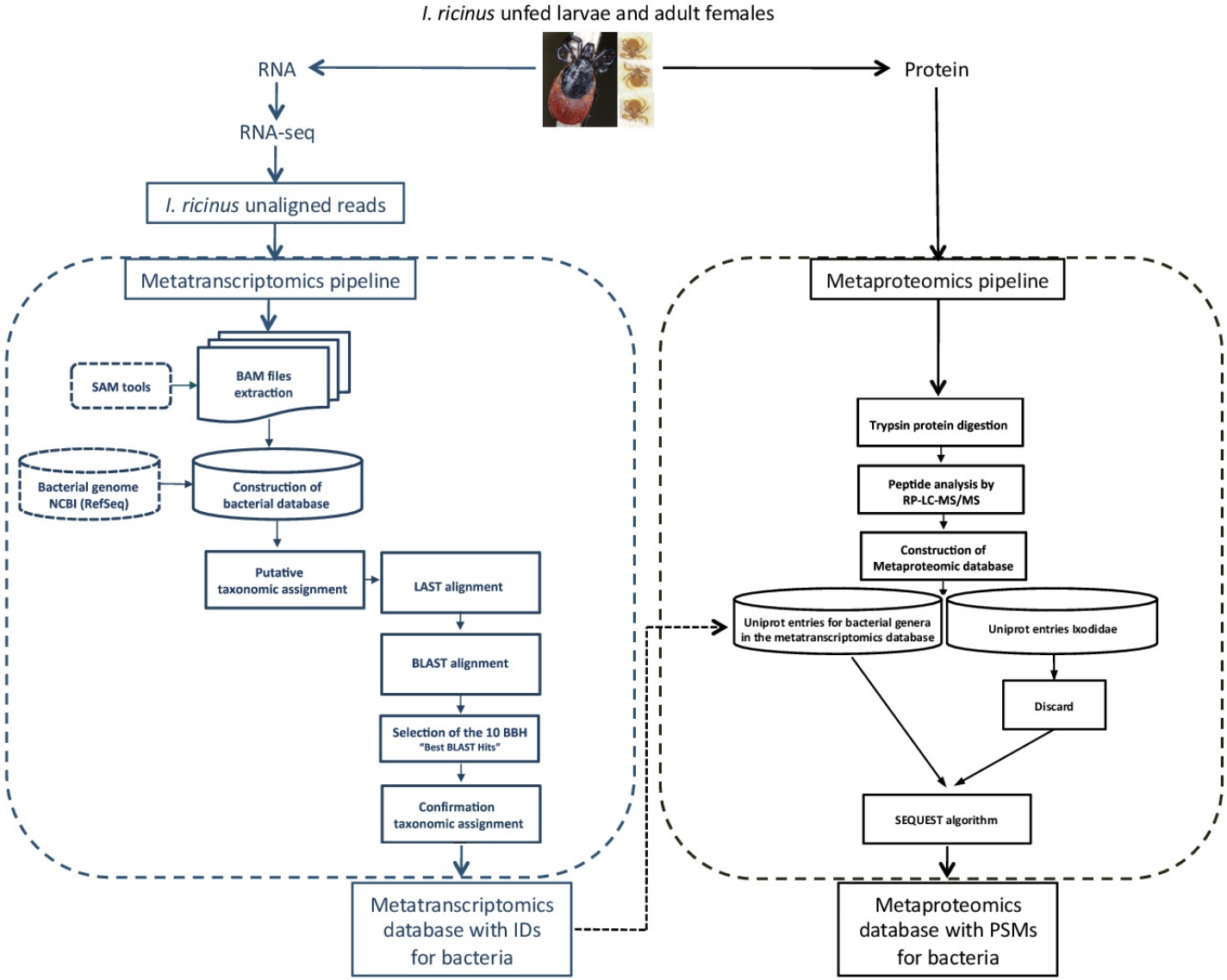
650 Table modified from Razzauti et al. (2015).

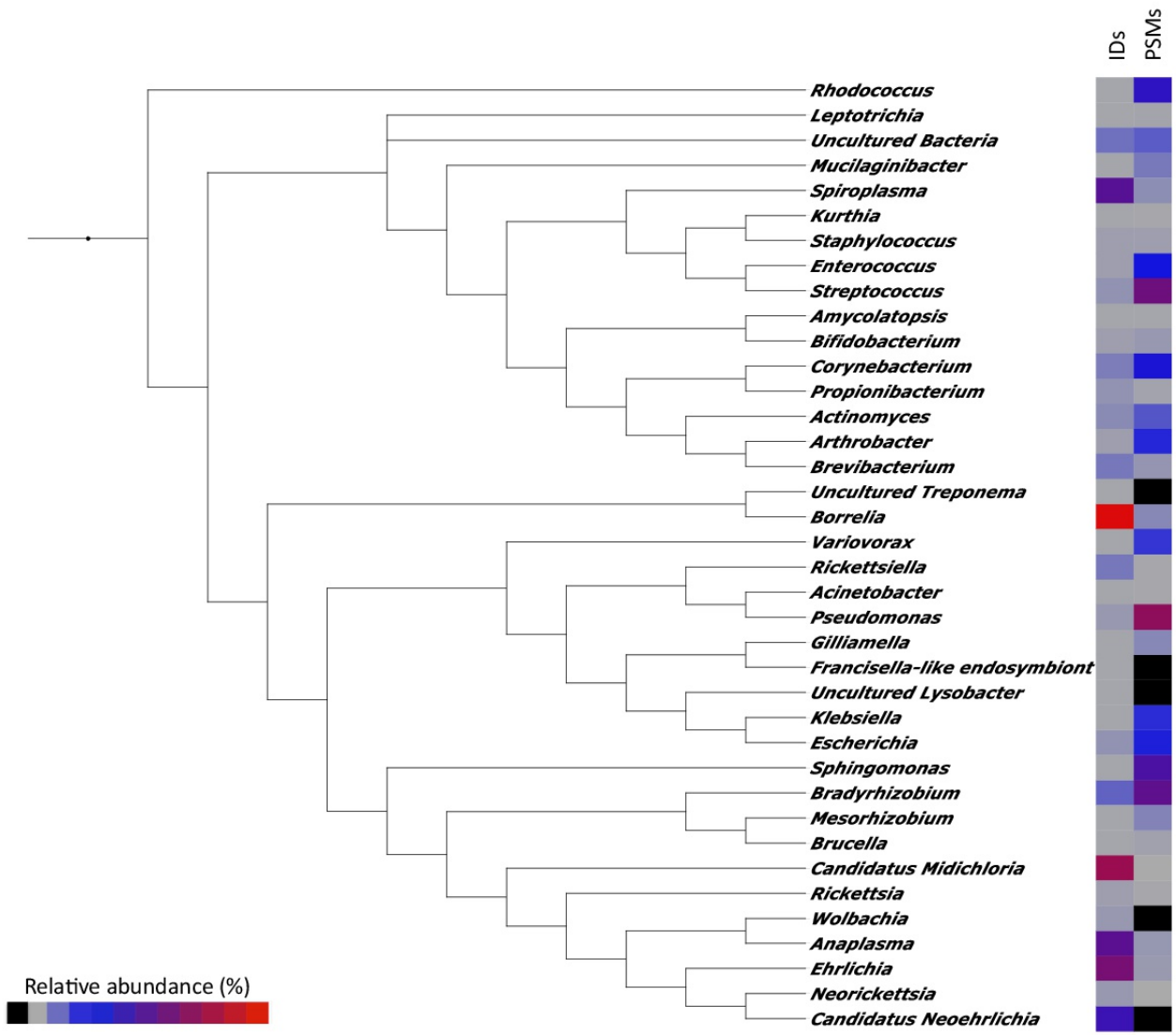
651 *These genera were previously identified as contamination of DNA extraction kits reagents

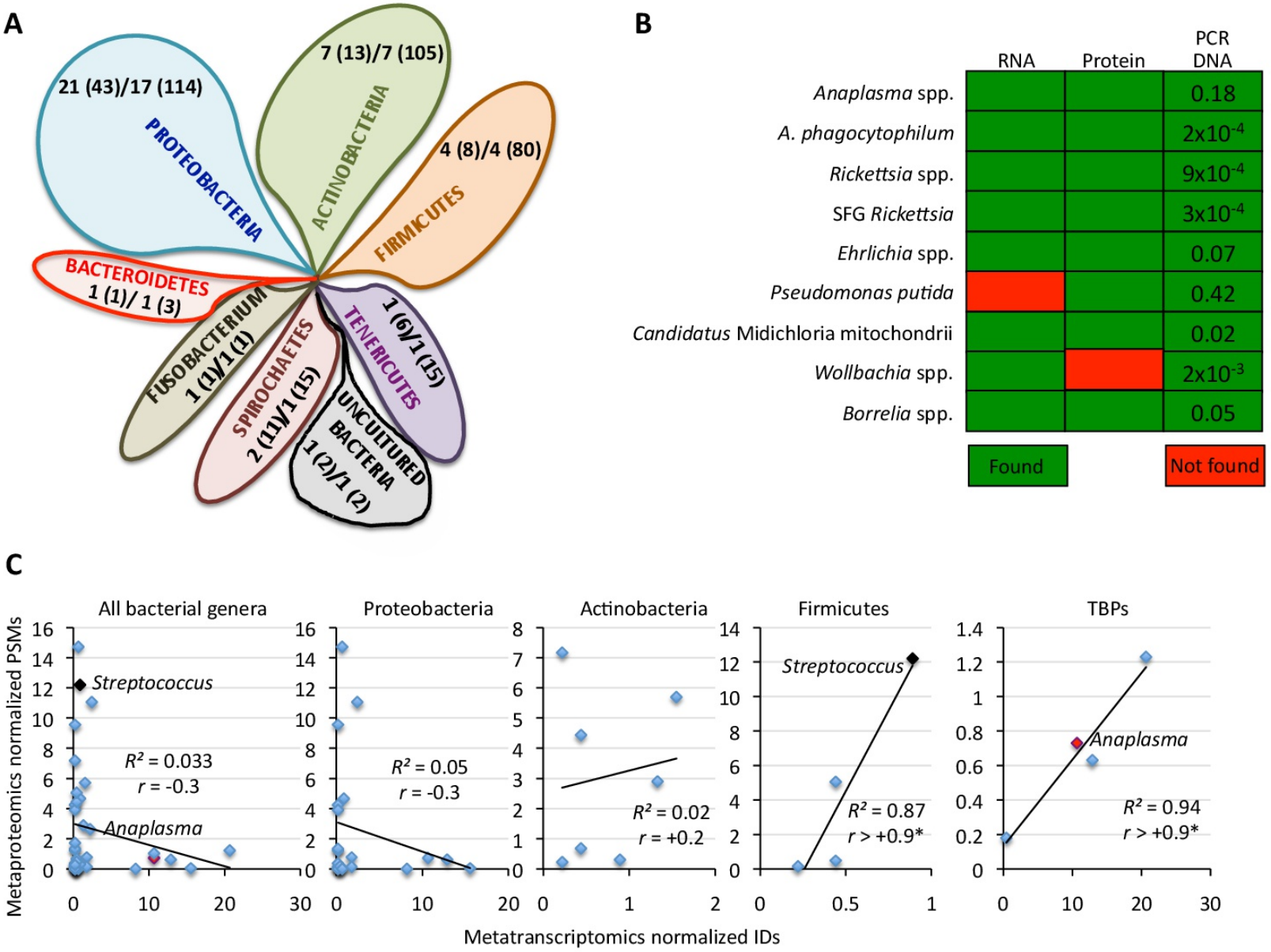
652 and ultrapure water systems, which may lead to erroneous identifications in bacterial

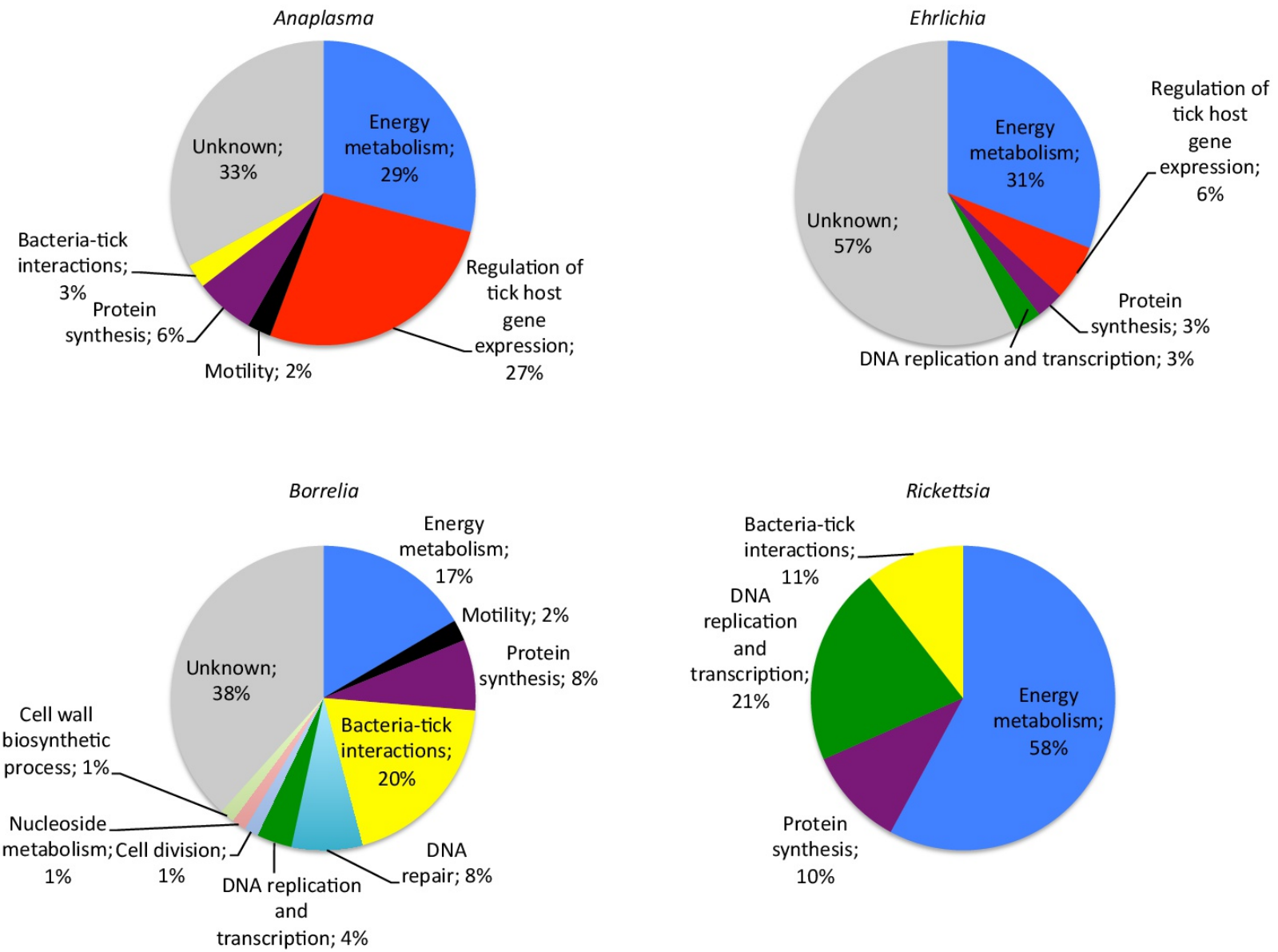
653 assignments (Salter et al., 2014).

654


I. ricinus unfed larvae and adult females
RNA
Protein
RNA-seq
I. ricinus unaligned reads
Metatranscriptomics pipeline
SAM tools
BAM files extraction
Bacterial genome NCBI (RefSeq)
Construction of bacterial database
Putative taxonomic assignment
LAST alignment
BLAST alignment
Selection of the 10 BBH "Best BLAST Hits"
Confirmation taxonomic assignment
Metatranscriptomics database with IDs for bacteria
Metaproteomics pipeline
Trypsin protein digestion
Peptide analysis by RP-LC-MS/MS
Construction of Metaproteomic database
Uniprot entries for bacterial genera in the metatranscriptomics database
Uniprot entries Ixodidae
Discard
SEQUEST algorithm
Metaproteomics database with PSMs for bacteria


IDs
PSMs
Rhodococcus
Leptotrichia
Uncultured Bacteria
Mucilaginibacter
Spiroplasma
Kurthia
Staphylococcus
Enterococcus
Streptococcus
Amycolatopsis
Bifidobacterium
Corynebacterium
Propionibacterium
Actinomyces
Arthrobacter
Brevibacterium
Uncultured Treponema
Borrelia
Variovorax
Rickettsiella
Acinetobacter
Pseudomonas
Gilliamella
Francisella-like endosymbiont
Uncultured Lysobacter
Klebsiella
Escherichia
Sphingomonas
Bradyrhizobium
Mesorhizobium
Brucella
Candidatus Midichloria
Rickettsia
Wolbachia
Anaplasma
Ehrlichia
Neorickettsia
Candidatus Neoehrlichia
Relative abundance (%)


A
21 (43)/17 (114)
PROTEOBACTERIA
7 (13)/7 (105)
ACTINOBACTERIA
4 (8)/4 (80)
FIRMICUTES
BACTEROIDETES
1 (1)/ 1 (3)
FUSOBACTERIUM
1 (1)/1 (1)
SPIROCHAETES
2 (11)/1 (15)
UNCULTURED BACTERIA
1 (2)/1 (2)
TENERICUTES
1 (6)/1 (15)
B
RNA
Protein
PCR DNA
Anaplasma spp.
A. phagocytophilum
Rickettsia spp.
SFG Rickettsia
Ehrlichia spp.
Pseudomonas putida
Candidatus Midichloria mitochondrii
Wollbachia spp.
Borrelia spp.
0.18
2x10-4
9x10-4
3x10-4
0.07
0.42
0.02
2x10-3
0.05
Found
Not found
C
All bacterial genera
Proteobacteria
Actinobacteria
Firmicutes
TBPs
Metaproteomics normalized PSMs
Metatranscriptomics normalized IDs
Streptococcus
Anaplasma
R² = 0.033
r = -0.3
R² = 0.05
r = -0.3
R² = 0.02
r = +0.2
R² = 0.87
r > +0.9*
R² = 0.94
r > +0.9*


Anaplasma
Energy metabolism; 29%
Regulation of tick host gene expression; 27%
Motility; 2%
Protein synthesis; 6%
Bacteria-tick interactions; 3%
Unknown; 33%
Ehrlichia
Energy metabolism; 31%
Regulation of tick host gene expression; 6%
Protein synthesis; 3%
DNA replication and transcription; 3%
Unknown; 57%
Borrelia
Energy metabolism; 17%
Motility; 2%
Protein synthesis; 8%
Bacteria-tick interactions; 20%
DNA repair; 8%
DNA replication and transcription; 4%
Cell division; 1%
Nucleoside metabolism; 1%
Cell wall biosynthetic process; 1%
Unknown; 38%
Rickettsia
Energy metabolism; 58%
Protein synthesis; 10%
DNA replication and transcription; 21%
Bacteria-tick interactions; 11%

Supplementary file 1 – Table 1. Bacterial sequence database constructed with genome and/or species-specific rRNA sequences.

Accession Number (NCBI)	Taxonomic Assignments
NC_019395	*Acidipropionibacterium acidipropionici* ATCC 4875, complete genome
NC_014374	*Acidilobus saccharovorans* 345-15, complete genome
NC_013926	*Aciduliprofundum boonei* T469, complete genome
KU991570.1	*Acinetobacter tjernbergiae* strain T40 16S ribosomal RNA gene, partial sequence
NC_009053	*Actinobacillus pleuropneumoniae* L20 serotype 5b complete genome
CP014232.1	*Actinomyces oris* strain T14V, complete genome
NR_117358.1	*Actinomyces oris* strain ATCC 27044 16S ribosomal RNA gene, partial sequence
LT223615.1	*Actinomyces odontolyticus* partial 16S rRNA gene, strain Marseille-P2379
LT576385.1	*Actinomyces* sp. Marseille-P2825 partial 16S rRNA gene, strain Marseille-P2825
NC_022521	*Aeropyrum camini* SY1 = JCM 12091 DNA, complete genome.
NC_003062	*Agrobacterium fabrum* str. C58 chromosome circular, complete sequence
NC_018011	*Alistipes finegoldii* DSM 17242, complete genome
NC_014318	*Amycolatopsis mediterranei* U32 chromosome, complete genome
CP009110.1	*Amycolatopsis methanolica* 239, complete genome
NC_004842	*Anaplasma marginale* str. St. Maries, complete genome
CP006847.1	*Anaplasma marginale* str. Dawn genome
NC_007797	*Anaplasma phagocytophilum* HZ, complete genome
KX236049.1	*Anaplasma phagocytophilum* isolate LYD52 16S ribosomal RNA gene, partial sequence
CP015376.1	*Anaplasma phagocytophilum* str. Norway variant2, complete genome
KM021418.1	*Anaplasma phagocytophilum* str. BovBat18 23S ribosomal RNA gene, partial sequence
KM021425.1	*Anaplasma platys* str. ChieCal05 23S ribosomal RNA gene, partial sequence
KT364327.1	*Anaplasma* sp. "ivoriense TCI149" 23S ribosomal RNA gene, partial sequence

NC_011567	*Anoxybacillus flavithermus* WK1, complete genome
NC_014218	*Arcanobacterium haemolyticum* DSM 20595, complete genome
NC_008541	*Arthrobacter* sp. FB24, complete genome.
KX168140.1	*Arthrobacter subterraneus* str. SBSK-401 16S ribosomal RNA gene, partial sequence
NC_021171	*Bacillus* sp. 1NLA3E, complete genome
NC_003909	*Bacillus cereus* ATCC 10987, complete genome
NC_005363	*Bdellovibrio bacteriovorus* complete genome, strain HD100
KU593501.1	*Bifidobacterium animalis* str. DR2-1 16S ribosomal RNA gene, partial sequence
CP011965.1	*Bifidobacterium longum* subsp. longum strain CCUG30698, complete genome
NC_008277	*Borrelia afzelii* PKo, complete genome
NR_121981.1	*Borrelia anserina* str. BA2 23S ribosomal RNA gene, complete sequence
KM269460.1	*Borrelia bissettii* str. M7p 23S ribosomal RNA gene
NC_001318	*Borrelia burgdorferi* B31 chromosome, complete genome
NC_017418	Borrelia burgdorferi N40, complete genome
NC_011229	*Borrelia duttonii* Ly, complete genome
NC_010673	*Borrelia hermsii* DAH, complete genome
NC_022079	*Borrelia miyamotoi* LB-2001, complete genome
KU196080.1	*Borrelia miyamotoi* str. Sonom53 16S ribosomal RNA gene, partial sequence
NC_011244	*Borrelia recurrentis* A1, complete genome
NC_008710	*Borrelia turicatae* 91E135, complete genome
NC_006156	*Borreliella bavariensis* PBi chromosome linear, complete sequence
KX034021.1	*Bradyrhizobium* sp. strain TC29 16S ribosomal RNA gene, partial sequence
LC167484.1	*Bradyrhizobium elkanii* gene for 16S ribosomal RNA, partial sequence, str: PHM 1
NC_004463	*Bradyrhizobium japonicum* USDA 110 chromosome, complete genome.
CP014869.1	*Brevibacterium linens* str. BS258, complete genome
KX168131.1	*Brevibacterium oceani* str. SBSK-404 16S ribosomal RNA gene, partial sequence

CP007758.1	*Brucella canis* str. RM6/66 chromosome 1, complete sequence
KX529832.1	*Brucella canis* str. YH-C16 16S ribosomal RNA gene, partial sequence
NC_019751	*Calothrix* sp. PCC 6303, complete genome.
NC_008599	*Campylobacter fetus* subsp. *fetus* 82-40, complete genome
NC_017299	*Clostridium botulinum* H04402 065, complete genome sequence
NC_017096	*Caldisericum* exile AZM16c01 DNA, complete genome.
NC_019791	*Caldisphaera lagunensis* DSM 15908, complete genome.
NC_002163	*Campylobacter jejuni* subsp. *jejuni* NCTC 11168 = ATCC 700819 chr. complete genome
NC_013194	*Candidatus Accumulibacter phos*phatis clade IIA str. UW-1, complete genome
NC_013771	*Candidatus Atelocyanobacterium thalassa* isolate ALOHA, complete genome
NC_010482	*Candidatus Korarchaeum cryptofilum* OPF8, complete genome
NC_015722	*Candidatus Midichloria mitochondrii* IricVA, complete genome
HF568841.1	*Candidatus Midichloria mitochondrii* partial 16S rRNA gene, strain MA7
KU865475.1	*Candidatus Neoehrlichia mikurensis* isolate LN5 16S rRNA gene, partial sequence
JX406180.1	*Candidatus Rickettsiella isopodorum* str. JKI D244/2012 16S ribosomal RNA (rrs) gene
NC_020135	*Candidatus Uzinura diaspidicola* str. ASNER, complete genome
NC_011916	*Caulobacter crescentus* NA1000, complete genome
NC_014151	*Cellulomonas flavigena* DSM 20109, complete genome
NC_009480	*Clavibacter michiganensis* subsp. *michiganensis* NCPPB 382 complete genome
NC_009495	*Clostridium botulinum* A str. ATCC 3502 chromosome, complete genome
NC_022538	complete chromosome *Acholeplasma palmae*
NC_012590	*Corynebacterium aurimucosum* ATCC 700975, complete genome
NC_002935	*Corynebacterium diphtheriae* NCTC 13129, complete genome
NC_003450	*Corynebacterium glutamicum* ATCC 13032 chromosome, complete genome
NR_121991.1	*Corynebacterium glycinophilum* str. AJ 3170 23S rRNA gene
KU319425.1	*Corynebacterium* sp. C2-18 16S rRNA gene, partial sequence

NC_007164	*Corynebacterium jeikeium* K411 complete genome
NC_014329	*Corynebacterium pseudotuberculosis* FRC41, complete genome
NC_015673	*Corynebacterium resistens* DSM 45100, complete genome
NC_021663	*Corynebacterium terpenotabidum* Y-11, complete genome
NC_019753	*Crinalium epipsammum* PCC 9333, complete genome.
NC_010530	*Cupriavidus taiwanensis* str. LMG19424 chromosome 2, complete genome
NC_019778	*Cyanobacterium stanieri* PCC 7202, complete genome
NC_013939	*Deferribacter desulfuricans* SSM1 DNA, complete genome
NC_013216	*Desulfotomaculum acetoxidans* DSM 771, complete genome
NC_015589	*Desulfotomaculum ruminis* DSM 2154, complete genome
NC_007354	*Ehrlichia canis* str. Jake, complete genome
NC_007799	*Ehrlichia chaffeensis* str. Arkansas, complete genome
KT308164.1	*Ehrlichia chaffeensis* isolate 1246 16S rRNA gene, partial sequence
NC_023063	*Ehrlichia muris* AS145, complete genome.
KP702294.1	*Ehrlichia muris* 23S rRNA gene, complete sequence
NC_005295	*Ehrlichia ruminantium* str. Welgevonden, complete genome
KX185055.1	*Enterococcus faecium* str. LUB950217 16S rRNA gene, partial sequence
NC_015601	*Erysipelothrix rhusiopathiae* str. Fujisawa DNA, complete genome
CP015855.1	*Escherichia coli* str. EDL933-1 genome
J01859.1	*Escherichia coli* 16S rRNA, complete sequence
NC_021019	*Eubacterium cylindroides* T2-87 draft genome
NC_017461	*Fervidicoccus fontis* Kam940 chromosome, complete genome
NC_006570	*Francisella tularensis* subsp. *tularensis* SCHU S4 chromosome, complete genome
JQ740890.1	*Francisella*-like endosymbiont of *Ixodes ricinus* 16S rRNA gene, partial sequence
NR_122000.1	*Gilliamella apicola* str. wkB1 23S rRNA, complete sequence
NC_010125	*Gluconacetobacter diazotrophicus* PAl 5 complete genome

NC_018581	*Gordonia* sp. KTR9, complete genome
NC_015153	*Haemophilus influenzae* F3031 complete genome
NC_013422	*Halothiobacillus neapolitanus* c2, complete genome
NC_015564	*Hoyosella subflava* DQS3-9A1, complete genome
NC_015588	*Isoptericola variabilis* 225, complete genome
NC_013174	*Jonesia denitrificans* DSM 20603, complete genome
KX691737.1	*Klebsiella* sp. strain 17LKA 16S rRNA gene, partial sequence
NC_010617	*Kocuria rhizophila* DC2201 DNA, complete genome
NC_012785	*Kosmotoga olearia* TBF 19.5.1, complete genome
CP013217.1	*Kurthia* sp. 11kri321, complete genome
KC904244.1	*Kurthia* sp. LAM0618 16S rRNA gene, partial sequence
NC_006087	*Leifsonia xyli* subsp. *xyli* str. CTCB07, complete genome
NC_013192	*Leptotrichia buccalis* DSM 1135, complete genome
NC_006055	*Mesoplasma florum* L1 chromosome, complete genome
NC_002678	*Mesorhizobium loti* MAFF303099 DNA, complete genome
NC_023044	*Methanobacterium* sp. MB1 complete sequence
NC_021355	*Methanobrevibacter* sp. AbM4, complete genome
NC_000909	*Methanocaldococcus jannaschii* DSM 2661, complete genome
NC_003551	*Methanopyrus kandleri* AV19, complete genome
NC_009712	*Methanoregula boonei* 6A8, complete genome
NC_019943	*Methanoregula formicica* SMSP, complete genome
NC_007796	*Methanospirillum hungatei* JF-1, complete genome
NC_000916	*Methanothermobacter thermautotrophicus* str. Delta H, complete genome
NC_015636	*Methanothermococcus okinawensis* IH1, complete genome
NC_018485	Methylocystis sp. SC2 complete genome
NC_014246	*Mobiluncus curtisii* ATCC 43063, complete genome

NC_014147	*Moraxella catarrhalis* BBH18, complete genome
NC_020418	*Morganella morganii* subsp. *morganii* KT, complete genome
NC_002944	*Mycobacterium avium* subsp. *paratuberculosis* str. k10, complete genome
NC_023036	*Mycobacterium neoaurum* VKM Ac-1815D, complete genome
NC_011025	*Mycoplasma arthritidis* 158L3-1, complete genome
NC_007633	*Mycoplasma capricolum* subsp. *capricolum* ATCC 27343, complete genome
NC_013511	*Mycoplasma hominis* ATCC 23114 chromosome complete genome
NC_006360	*Mycoplasma hyopneumoniae* 232, complete genome
NC_005364	*Mycoplasma mycoides* subsp. *mycoides* SC str. PG1 chromosome, complete genome
NC_004432	*Mycoplasma penetrans* HF-2 DNA, complete genome
NC_015153	*Mycoplasma suis* KI3806 complete genome
KU597223.1	*Mucilaginibacter* sp. str. JWp32 16S rRNA gene, partial sequence
NR_076635.1	*Neorickettsia risticii* str. Illinois 23S rRNA gene, complete sequence
NC_007798	*Neorickettsia sennetsu* str. Miyayama, complete genome
NC_007406	*Nitrobacter winogradskyi* Nb-255, complete genome
NC_015222	*Nitrosomonas* sp. AL212, complete genome
NC_010085	*Nitrosopumilus maritimus* SCM1 chromosome, complete genome
NC_006361	*Nocardia farcinica* IFM 10152 DNA, complete genome
NC_008699	*Nocardioides* sp. JS614, complete genome
NC_018524	*Nocardiopsis alba* ATCC BAA-2165, complete genome
NC_005303	Onion yellows *phytoplasma* OY-M DNA, complete genome
NC_008711	*Paenarthrobacter aurescens* TC1, complete genome
NC_015702	*Parachlamydia acanthamoebae* UV-7, complete genome
NC_020514	*Paraglaciecola psychrophila* 170, complete genome.
NC_014414	*Parvularcula bermudensis* HTCC2503 str. HTCC2503, complete genome
NC_014537	*Parvularcula bermudensis* HTCC2503 str. HTCC2503, complete genome

NC_016605	*Pediococcus claussenii* ATCC BAA-344, complete genome
NC_008609	*Pelobacter propionicus* DSM 2379, complete genome
NC_012440	*Persephonella marina* EX-H1, complete genome
NC_011144	*Phenylobacterium zucineum* HLK1, complete genome
NC_012962	*Photorhabdus asymbiotica* ATCC43949 complete genome
NC_019689	*Pleurocapsa* sp. PCC 7327, complete genome
NC_015501	*Porphyromonas asaccharolytica* DSM 20707, complete genome
NC_004578	*Pseudomonas syringae* pv. tomato str. DC3000 chromosome, complete genome
NC_012660	*Pseudomonas fluorescens* SBW25 complete genome
NC_023064	*Pseudomonas* sp. TKP, complete genome
NC_021085	*Propionibacterium acnes* HL096PA1, complete genome
KX108930.1	*Propionibacterium acnes* str. IR-TUMS/BPG6 16S ribosomal RNA gene
NC_015177	*Pseudopedobacter saltans* DSM 12145, complete genome
NC_018142	*Pseudopropionibacterium propionicum* F0230a, complete genome
NC_008709	*Psychromonas ingrahamii* 37, complete genome
NC_003364	*Pyrobaculum aerophilum* str. IM2 chromosome, complete genome
NC_015931	*Pyrolobus fumarii* 1A, complete genome
CP016819.1	*Rhodococcus* sp. p52, complete genome
AM179867.1	*Rhodococcus* sp. A83 A83 A83 A83 partial 16S rRNA gene, isolate A83
NC_005296	*Rhodopseudomonas palustris* CGA009 complete genome
NC_007940	*Rickettsia bellii* RML369-C, complete genome
NC_007109	*Rickettsia felis* URRWXCal2, complete genome
NC_006142	*Rickettsia typhi* str. Wilmington, complete genome
KP994764.1	*Rickettsiella* endosymbiont of *Ixodes tasmani* isolate Ixo tasmani3 23S ribosomal
KT697685.1	*Rickettsiella* sp. RKTSLLA_T3262 16S rRNA gene, partial sequence
NC_014920	*Rothia dentocariosa* ATCC 17931, complete genome

NC_013715	*Rothia mucilaginosa* DY-18 DNA, complete genome
NC_009092	*Shewanella loihica* PV-4, complete genome
NC_020527	*Sinorhizobium meliloti* 2011 plasmid pSymA, complete sequence
NC_012587	*Sinorhizobium fredii* NGR234 chromosome, complete genome
KX672814.1	*Sphingomonas* sp. K-16 16S rRNA gene, partial sequence
NC_009511	*Sphingomonas wittichii* RW1, complete genome
LN019399.1	*Spirometra erinaceieuropaei* genome assembly S_erinaceieuropaei
NC_008048	*Sphingopyxis alaskensis* RB2256, complete genome
NC_022998	*Spiroplasma apis* B31, complete genome
NC_021833	*Spiroplasma diminutum* CUAS-1, complete genome
NC_021846	*Spiroplasma taiwanense* CT-1, complete genome
NC_021280	*Spiroplasma chrysopicola* DF-1, complete genome
KP967685.1	*Spiroplasma* sp. Bratislava 1 16S rRNA gene, partial sequence
DQ004912.1	*Spiroplasma ixodetis* str. Y-30 16S rRNA gene and 23S rRNA
FJ824553.1	*Spiroplasma* sp. GSU5508 23S rRNA gene, partial sequence
NC_002976	*Staphylococcus epidermidis* RP62A, complete genome
NC_007168	*Staphylococcus haemolyticus* JCSC1435 DNA, complete genome
NC_020164	*Staphylococcus warneri* SG1, complete genome
NC_003098	*Streptococcus pneumoniae* R6 chromosome, complete genome
NC_012004	*Streptococcus uberis* 0140J complete genome
KX679404.1	*Streptococcus* sp. HTS29 16S rRNA gene, partial sequence
NC_002754	*Sulfolobus solfataricus* P2, complete genome
NC_007775	*Synechococcus* sp. JA-3-3Ab, complete genome
NC_012997	*Teredinibacter turnerae* T7901, complete genome
NC_022093	*Thermofilum* sp. 1910b, complete genome
NC_017954	*Thermogladius cellulolyticus* 1633, complete genome

NC_009616	*Thermosipho melanesiensis* BI429, complete genome
NC_018012	*Thiocystis violascens* DSM 198, complete genome
NC_002967	*Treponema denticola* ATCC 35405 chromosome, complete genome
NC_014158	*Tsukamurella paurometabola* DSM 20162, complete genome
NC_002162	*Ureaplasma parvum* serovar 3 str. ATCC 700970, complete genome
LT607803.1	*Variovorax* sp. HW608 genome assembly, chromosome: I
AB300597.1	*Variovorax boronicumulans* gene for 16S rRNA, partial sequence
NC_014643	*Vulcanisaeta distributa* DSM 14429, complete genome
NC_015759	*Weissella koreensis* KACC 15510, complete genome
NC_010981	*Wolbachia* endosymbiont of *Culex quinquefasciatus* Pel str wPip complete genome
NC_002978	*Wolbachia* endosymbiont of *Drosophila melanogaster*, complete genome
NC_018267	*Wolbachia* endosymbiont of *Onchocerca ochengi* complete genome
NC_006833	*Wolbachia* endosymbiont str. TRS of *Brugia malayi*, complete genome
NC_006526	*Zymomonas mobilis* subsp. *mobilis* ZM4, complete genome
KT151398.1	Uncultured *Lysobacter* sp. clone AT204 16S rRNA gene, partial sequence
JQ654170.1	Uncultured *Treponema* sp. clone 5:7P56 16S rRNA gene, partial sequence
KX630121.1	Uncultured bacterium clone 16S(V3-V4)-7606 16S rRNA gene, partial sequence