A novel approach, based on BLSOMs (Batch Learning Self-Organizing Maps), to the microbiome analysis of ticks

Ryo Nakao[1,a], Takashi Abe[2,3,a], Ard M. Nijhof[4], Seigo Yamamoto[5], Frans Jongejan[6,7], Toshimichi Ikemura[2], Chihiro Sugimoto[1]

[1]*Division of Collaboration and Education, Research Center for Zoonosis Control, Hokkaido University, Kita-20, Nishi-10, Kita-ku, Sapporo, Hokkaido 001-0020, Japan*

[2]*Nagahama Institute of Bio-Science and Technology, Nagahama, Shiga 526-0829, Japan*

[3]*Graduate School of Science & Technology, Niigata University, 8050, Igarashi 2-no-cho, Nishi-ku, Niigata 950-2181, Japan*

[4]*Institute for Parasitology and Tropical Veterinary Medicine, Freie Universität Berlin, Königsweg 67, 14163 Berlin, Germany*

[5]*Miyazaki Prefectural Institute for Public Health and Environment, 2-3-2 Gakuen Kibanadai Nishi, Miyazaki 889-2155, Japan*

[6]*Utrecht Centre for Tick-borne Diseases (UCTD), Department of Infectious Diseases and Immunology, Faculty of Veterinary Medicine, Utrecht University, Yalelaan 1, 3584 CL Utrecht, The Netherlands*

[7]*Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, Private Bag X04, 0110 Onderstepoort, South Africa*

[a]These authors contributed equally to this work.

Keywords: BLSOMs/emerging diseases/metagenomics/microbiomes/symbionts/ticks

Running title: Tick microbiomes revealed by BLSOMs

Subject category: Microbe-microbe and microbe-host interactions

Abstract

Ticks transmit a variety of viral, bacterial and protozoal pathogens, which are often zoonotic. The aim of this study was to identify diverse tick microbiomes, which may contain as-yet unidentified pathogens, using a metagenomic approach. DNA prepared from bacteria/archaea-enriched fractions obtained from seven tick species, namely *Amblyomma testudinarium*, *Amblyomma variegatum*, *Haemaphysalis formosensis*, *Haemaphysalis longicornis*, *Ixodes ovatus*, *Ixodes persulcatus* and *Ixodes ricinus*, was subjected to pyrosequencing after whole-genome amplification. The resulting sequence reads were phylotyped using a Batch Learning Self-Organizing Map (BLSOM) program, which allowed phylogenetic estimation based on similarity of oligonucleotide frequencies, and functional annotation by BLASTX similarity searches. In addition to bacteria previously associated with human/animal diseases, such as *Anaplasma*, *Bartonella*, *Borrelia*, *Ehrlichia*, *Francisella* and *Rickettsia*, BLSOM analysis detected microorganisms belonging to the phylum Chlamydiae in some tick species. This was confirmed by pan-Chlamydia PCR and sequencing analysis. Gene sequences associated with bacterial pathogenesis were also identified, some of which were suspected to originate from horizontal gene transfer. These efforts to construct a database of tick microbes may lead to the ability to predict emerging tick-borne diseases. Furthermore, a comprehensive understanding of

tick microbiomes will be useful for understanding tick biology, including vector competency and interactions with pathogens and symbionts.

Introduction

Ticks (Acari: Ixodida) are amongst the most well-known blood-sucking arthropods. They can transmit a variety of pathogens, including viruses, bacteria and protozoa, to humans and animals. Opportunities for human/animal contact in natural foci of tick-borne infection are increasing, along with expanded distributions of vector ticks due to environmental changes (Beugnet and Marié 2009; Parola 2004). Meanwhile, reports of novel tick-borne pathogens are also increasing (Spolidorio *et al*., 2010; Yu *et al*., 2011). Above all, many of the emerging rickettsial pathogens identified in recent decades were first discovered in ticks and subsequently recognized as the causative agents of animal or human diseases (Parola *et al*., 2005). This indicates that active surveillance for potential pathogens in ticks may be a feasible approach to predict and combat emerging tick-borne diseases. Given the diversity of tick species and their wide geographic distribution (Bowman and Nuttall, 2008), it is likely that as-yet unidentified pathogens are harboured by ticks.

Morphological detection by electron microscopy has shown that *Rickettsia*- and *Wolbachia*-like bacteria infect a variety of tick species (Lewis, 1979; Roshdy, 1968), although these bacteria are not necessarily pathogenic in either mammalian hosts or ticks. The availability of molecular genetic tools such as PCR and real-time PCR has increased the number of reports on a variety of microorganisms in ticks (Benson *et al*., 2004; Noda *et al*., 1997; van Overbeek *et al*., 2008); however, their biological roles and potential to cause disease in mammals remain poorly understood. Interactions between arthropods and their symbionts have been relatively well studied in aphids. A variety of symbionts are reported to have detrimental, neutral or beneficial effects on the aphid hosts; for example, through pathogenicity (Grenier *et al*., 2006), providing nutrition (Nakabachi and Ishikawa, 1999), altering gene expression associated with body colour (Tsuchida *et al*., 2010) and, most intriguingly, displaying protective functions against pathogens and predators (Scarborough *et al*., 2005). It has been suggested that certain symbiotic microbes might also affect the vector competency of their arthropod host (Weiss and Aksoy, 2011). A full characterization of microorganisms harboured by ticks will provide valuable information towards a better understanding of tick biology in terms of the relationship between ticks and their microbiomes, including microbes that are potential pathogens in mammals.

Recent progress in second generation sequencing technologies has led to the discovery of previously unknown organisms in a variety of samples, including arthropods (Cox-Foster *et al*.,

2007; Suen *et al*., 2010; Warnecke *et al*., 2007), since uncultured microorganisms represent the vast majority of symbiotic, commensal and pathogenic microorganisms. Attempts to assess the microbial diversity in two tick species, *Rhipicephalus microplus* and *Ixodes ricinu*s, using novel sequencing technologies, resulted in the identification of microbes (including bacterial species) previously unreported in ticks (Andreotti *et al*., 2011; Carpi *et al*., 2011). However, since the data processing methods used in these two studies were based on conventional homology-based sequence searches, the discoveries were limited to microorganisms for which genomic information was already available in the database used. An alternative approach to characterizing DNA fragments might offer improvements for investigating microbial diversity in ticks that are expected to harbour poorly characterized microorganisms.

Batch Learning Self-Organizing Map (BLSOM) is a bioinformatics method that follows a learning process and generates a map independently of the order of data input (Abe *et al*., 2003; Kanaya *et al*., 2001). This method was designed to separate and cluster genomic sequence fragments based on the similarity of oligonucleotide frequencies without any other taxonomical information. Since BLSOM does not require orthologous sequence sets and sequence alignments, it could provide a new systematic strategy for revealing the microbial diversity and relative abundance of different phylotype members of uncultured microorganisms in a wide variety of environmental metagenomic samples (Abe *et al*., 2005; Uehara *et al*., 2011). Of note, similar

SOM-based methods were recently used to obtain clear phylotype-specific classification of metagenomic sequences (Chan *et al*., 2008; Dick *et al*., 2009; Martin *et al*., 2008; Weber *et al*., 2011).

In the present study, we used BLSOM to characterize diverse microbiomes in seven tick species. The ticks harboured a variety of bacteria, including those previously reported to be associated with human and animal diseases, and others not previously reported from ticks. This study was intended as a first step towards the discovery of emerging tick-borne pathogens and a systematic understanding of the relationship between ticks and their microbiomes.

Materials and methods

Tick species.

Ticks were collected from the field by dragging flannel sheets over the vegetation. Three tick species, namely *Amblyomma testudinarium* (AT), *Haemaphysalis formosensis* (HF) and *Haemaphysalis longicornis* (HL) were collected in March 2009 in Miyazaki, Japan. This is an area where *Rickettsia japonica*, the causative agent of Japanese spotted fever and *Rickettsia tamurae*, the spotted fever group rickettsia recently associated with human disease (Imaoka *et al*.,

2011), are known to be endemic (Morita *et al*., 1990; S Yamamoto, personal communication). Two *Ixodes* species, *Ixodes ovatus* (IO) and *Ixodes persulcatus* (IP), were collected in June 2010 in Hokkaido, Japan, where *Borrelia burgdorferi sensu lato* (*s.l.*) and *Anaplasma phagocytophilum*, the causative agents of Lyme disease and human granulocytic anaplasmosis, respectively, are endemic (Miyamoto *et al*., 1992; Murase *et al*., 2011). *Ixodes ricinus* (IR) was collected in August 2010 in Soesterberg, The Netherlands, where Lyme disease is endemic (Hofhuis *et al*., 2006). *Amblyomma variegatum* (AV) ticks were collected in The Gambia in 2005 and subsequently maintained under laboratory conditions at the Utrecht Centre for Tick-borne Diseases, The Netherlands. Adult AV ticks from the second laboratory generation were used in this study. The collected ticks were pooled according to species, life stage and sex (except for HF and HL) and specific pool IDs (AVf, AVm, IOf, IOm, IPf, IPm, IRf, ATn, HFfmn and HLfmn) were provided as indicated in Table 1 ("f", "m" and "n" represent female, male and nymph, respectively).

DNA preparation.

The live ticks were washed twice with 70% ethanol supplemented with 1% povidone-iodine (Meiji Seika, Tokyo, Japan) solution and rinsed three times with distilled water to decontaminate the tick body surface. Ticks were then ground with 4.8 mm stainless steel beads

(TOMY, Tokyo, Japan) using the Micro Smash MS-100R (TOMY). Tick homogenates were treated with a membrane lysis buffer [10 mM Tris-HCl (pH 8.0), 150 mM NaCl, 10 mM $MgCl_2$, 0.1% IGEPAL CA-630 (Sigma Chemical Co., St. Louis, MO)] for 10 min on ice and centrifuged at 400 × *g* for 25 min. The supernatant was centrifuged at 20,000 × g for 30 min and the pellet containing the bacterial/archaeal cells was resuspended in a DNase buffer [25 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 0.8 U/μL DNase I (Takara, Shiga, Japan)] to remove any contaminating host-cell DNA. After 60 min at 37°C, the reaction was stopped by adding 25 mM EDTA (pH 8.0). The solution was filtered through a 5.0 μm pore-size membrane filter (Millipore, Bedford, MA) by centrifugation at 1,000 × g. The filtrate was centrifuged at 20,000 × g for 30 min and the resulting pellet was washed twice with PBS. All centrifugation steps were performed at 4°C. To lyse a wide range of bacterial/archaeal species, this study utilized an achromopeptidase, which is reported to have broad-spectrum bacteriolytic activity (Ezaki and Suzuki, 1982). The bacterial/archaeal pellet was treated with an achromopeptidase buffer [10 mM Tris-HCl (pH 8.0), 10 mM NaCl, 1 U/μL achromopeptidase (WAKO, Osaka, Japan)] for 60 min at 37°C. Bacterial/archaeal DNA was then extracted using the NucleoSpin Tissue XS Kit (Macherey-Nagel, Düren, Germany) according to the manufacturer's instructions. The genomic DNA was then subjected to whole-genome amplification using the GenomiPhi™ V2 DNA Amplification Kit (GE Healthcare, Chalfont St Giles, UK) according to the manufacturer's instructions. The

DNA concentration was measured using the Quant-iT dsDNA BR assay with a Qubit Fluorometer (Invitrogen, Carlsbad, CA). All procedures were conducted under sterile conditions in a flow cabinet. A schematic flow diagram showing the extraction process of bacterial/archaeal DNA from ticks is presented in Figure 1.

Pyrosequencing.

Genomic DNA from each tick pool was subjected to pyrosequencing on a Roche/454 Genome Sequencer FLX Titanium (Roche Applied Science/454 Life Science, Brandford, CT) at Hokkaido System Science (Sapporo, Japan). Two GS FLX shotgun libraries were prepared using a standard protocol from the manufacturer and analysed in four independent sequencing runs: the first library contained three pools (ATn, HFfmn and HLfmn) with different multiplex identifiers and was analysed on 1/8 of the PicoTiterPlate (Roche Applied Science/454 Life Science), while the second library contained seven pools (AVf, AVm, IOf, IOm, IPf, IPm and IRf) and was analysed on 1/4 of the plate. The metagenomic sequences were deposited in the DNA Data Bank of Japan (DDBJ) (http://www.ddbj.nig.ac.jp) Sequence Read Archive under the accession no. DRA000590.

BLSOM analysis.

Self-Organizing Map (SOM) is an unsupervised neural network algorithm that implements a characteristic nonlinear projection from the high-dimensional space of input data onto a two-dimensional array of weight vectors (Kohonen, 1990). We used the 'Batch Learning SOM' (BLSOM), which is a modified version of the conventional SOM for genome informatics that makes the learning process and creation of the resulting map independent of the order of data input (Abe *et al*., 2003; Kanaya *et al*., 2001). The initial weight vectors were defined by principal component analysis instead of random values. BLSOM learning was conducted as described previously (Abe *et al*., 2003), and the BLSOM program was obtained from UNTROD Inc., Japan (y_wada@nagahama-i-bio.ac.jp).

To estimate phylotypes of the metagenomic sequences, three types of large-scale BLSOMs, namely Kingdom-, Bacteria/Archaea- and Genus group-BLSOM, were constructed in advance, using sequences deposited in DDBJ/EMBL/GenBank as previously described (Abe *et al*., 2003). The clustering conditions were adopted from the previously optimised parameters to minimise the computation time without loss of clustering power (Abe *et al*., 2005). Kingdom-BLSOM was constructed with tetranucleotide frequencies for 5-kb sequences from the whole-genome sequences of 111 eukaryotes, 2,813 bacteria/archaea, 1,728 mitochondria, 110 chloroplasts and 31,486 viruses. To obtain more detailed phylotype information for bacterial/archaeal sequences, Bacteria/Archaea- and Genus group-BLSOM were constructed with

a total of 3,500,000 5-kb sequences from 3,157 species, for which at least 10 kb of sequence was available from DDBJ/EMBL/GenBank.

Mapping of metagenomic sequences longer than 300 bp on Kingdom-BLSOMs, after normalization of the sequence length, was conducted by finding the lattice point with the minimum Euclidean distance in the multidimensional space. To identify further detailed phylogenies of the metagenomic sequences that had been mapped to the bacterial/archaeal territories on Kingdom-BLSOM, these sequences were successively mapped on Bacteria/Archaea-BLSOM. Similar stepwise mappings of metagenomic sequences on the Genus group-BLSOM were subsequently conducted to obtain further detailed phylogenetic information. In order to evaluate the accuracy of BLSOMs for taxonomic classification of metagenomic sequences, the BLSOM analysis was applied to three previously published simulated datasets of varying complexities (Mavromatis *et al*., 2007). When sequences longer than 300 bp were mapped, approximately 90%, 70% and 40%-50% were correctly classified to the kingdom, phylum and genus levels, respectively (Supplementary Figure S1).

Pan-Chlamydia PCR, sequencing and data analysis.

A pan-Chlamydia PCR was conducted on the tick species in which bacteria belonging to the phylum Chlamydiae had been detected by BLSOM analysis. PCR amplifications were

performed using three different sets of primers targeting the 16S rRNA gene (Supplementary Table 1). The successfully amplified products were cloned into a pGEM-T vector (Promega, Madison, WI). Each plasmid clone was sequenced using the BigDye Terminator version 3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) and an ABI Prism 3130x genetic analyzer (Applied Biosystems) according to the manufacturer's instructions. The DNA sequences obtained were submitted to the DDBJ under accession nos. AB725685 to AB725705.

The obtained sequences were aligned using ClustalW2 (Larkin *et al*., 2007) and the resulting alignments were used to generate an uncorrected pairwise distance matrix using the MOTHUR program version 1.25.0 (Schloss *et al*., 2009). An average neighbour clustering algorithm in the MOTHUR program was used to assign sequences into operational taxonomic units (OTUs), with a 97% sequence similarity cut-off. A representative sequence from each OTU was taxonomically identified using the Classifier tool available on the Ribosomal Database Project (RDP) database (Cole *et al*., 2009) and was compared against the NCBI GenBank database using a BLASTn search.

Functional annotation and characterization.

The metagenomic sequences were annotated using BLASTX (Altschul *et al*., 1990) against the Clusters of Orthologous Groups (COG) database (Tatusov *et al*., 1997) with an E-

value threshold of 1×10^{-5} and were classified according to COG functional categories. We also identified putative virulence-associated factors using the MvirDB database (Zhou *et al.*, 2007), which consists of sequence information from multiple microbial databases of protein toxins, virulence factors and antibiotic resistance genes. Sequences associated with "antibiotic resistance", "pathogenicity islands" and "virulence protein" were further phylotyped using BLSOM as described above. All sequence reads were used to identify putative virulence-associated factors in BLASTX-based analyses, while only sequences longer than 300 bp were phylotyped.

Results

Kingdom classification.

Bacterial/archaeal cells in tick homogenates were purified using centrifugation, DNase treatment and filtration. DNA was extracted from bacteria/archaea-enriched fractions and was subjected to pyrosequencing after whole-genome amplification. The total sequence reads for each species ranged from 21,213 (HFfmn) to 42,258 (AVf), with average lengths between 143.5 bp (IPm) and 391.4 bp (AVm) (Table 2). The sequence reads longer than 300 bp were classified as

bacterial/archaeal, eukaryotic, virus, mitochondrial or chloroplast using Kingdom-BLSOM (Figure 2). The percentage of the reads categorized as bacteria/archaea ranged from 37.8% (AVm) to 72.3% (HLfmn), indicating that the present protocol could efficiently enrich bacterial/archaeal cells from tick homogenates. An average of 35% of the eukaryotic reads were clustered together with those of other arthropods, which were included in the Kingdom-BLSOM construction, while the remaining reads were with a variety of eukaryotes including protists and mammals (data not shown).

Phylum classification.

The bacterial/archaeal sequence reads were further classified at the phylum level using Bacteria/Archaea-BLSOM. Between 97.2% (AVf) and 99.4% (HLfmn) of the reads were successfully assigned to specific phyla (Table 2). The number of different phyla obtained from a single library ranged from 20 (HLfmn) to 28 (AVm, IOf and IPf). The composition of bacterial/archaeal phyla in each tick pool is shown in Figure 3. Each tick species had a different microbial composition. Differences were also found between males and females within the same tick species. Sequences classified into the phyla Firmicutes and Gammaproteobacteria constituted nearly half of the total sequences in most tick species. Sequences classified into the phylum

Alphaproteobacteria were also observed in all tick species, while those classified into the phylum Chlamydiae were observed at high abundance only in IPf, IPm and HFfmn.

Genus classification.

Between 85.4% (AVm) and 92.3% (ATn) of the sequences characterized into certain bacterial/archaeal phyla were successfully characterized to the genus level using Genus group-BLSOM (Table 2). The number of different genera obtained from a single library ranged from 133 (IRf) to 223 (AVm). The twenty most prevalent bacterial/archaeal genera within each tick species are listed in Table 3, while a complete list of bacterial/archaeal genera recovered from each tick pool is shown in Supplementary Table 2. A total of 64 different genera were detected and some of the dominant bacterial/archaeal genera were commonly observed between tick species. Some of the sequences could not be assigned to known taxa. For example, 4.7% of the sequences in AVm were characterized to the phylum Euryarchaeota, but could not be characterized at the genus level (Table 3). In accordance with the results of phylum classification, many of the dominant bacteria belonged to genera within the phyla Firmicutes and Gammaproteobacteria. The sequences of known tick-borne pathogens such as *Anaplasma*, *Bartonella*, *Borrelia*, *Ehrlichia*, *Francisella* and *Rickettsia*, as well as those of known tick-symbionts such as *Coxiella*, *Rickettsiella* and *Wolbachia*, were also listed as top 20 genera (Table

3). Only three pools (IPf, IPm and HFfmn) contained bacterial sequences classified into the family *Chlamydiaceae* (genera *Chlamydia* and *Chlamydophila*) as dominant taxa/populations.

Pan-Chlamydia PCR and sequencing analysis.

DNA extracted from five tick pools, IPf, IPm, HFf, HFm and HFn, was tested by pan-Chlamydia PCR using three different primer sets. Only one primer set (16S FOR2 and 16S REV2) produced PCR products of the expected size (approximately 260 bp) in all tested samples (data not shown). A total of 14 different clones of partial 16S rRNA gene sequences were recovered, four of which were identified in both tick species (Table 4). The MOTHUR program assigned these 14 sequences into five OTUs, the representative sequences of which were predicted to belong to the genera *Neochlamydia*, *Parachlamydia* or *Simkania* based on RDP database classification. BLASTn analysis of a representative sequence for each OTU showed highest similarity (88%–98%) to Chlamydial 16S rRNA gene sequences recovered from various animal sources including cockroach, sea bass, koala, cat and leafy seadragon (Table 4).

Functional annotation.

Functional annotation of the bacterial/archaeal sequences using BLASTX against the COG database indicated similar functional gene compositions between pools, except for AVm (Supplementary Figure 2). More than 80% of the AVm sequences were annotated as "Replication, recombination and repair"-related genes. This result may indicate that the sequences in this

library were not randomly amplified during the whole-genome amplification process. Sequences associated with putative virulence-associated factors were identified using BLASTX against the MvirDB database. The number of identified sequences differed greatly between libraries: IOm showed the highest number (6,715) and ATn the lowest (365) (Figure 4). Most of these were categorized as either "antibiotic resistance", "pathogenicity island" or "virulence protein"-related genes. Both IOm and IPf contained a large number of sequences associated with "virulence protein" (*n* = 6,515 and 2,189, respectively), in comparison to other pools. Sequences associated with "pathogenicity island" were observed to some extent in all pools, with the highest number obtained from IPf (*n* = 2,072), followed by IOf, AVm and HLfmn. The origins of these virulence-associated sequences were predicted using Bacteria/Archaea-BLSOM, and five of the most dominant origins of the sequences associated with "antibiotic resistance", "pathogenicity island" and "virulence protein" in each tick pool are shown in Table 5. The most common phyla were Firmicutes, Alphaproteobacteria, Gammaproteobacteria and Chlamydiae, while many of the sequences could not be phylotyped by BLSOM and remained unclassified.

Discussion

The present study investigated the diversity of microbial communities in ticks using pyrosequencing technology coupled with a composition-based data processing approach called BLSOM. In order to provide an overview of microbiomes associated with different tick species, ticks were pooled according to species and subjected to the analyses. We showed that ticks harboured a variety of bacteria, including those previously associated with human/animal diseases such as *Anaplasma*, *Bartonella*, *Borrelia*, *Ehrlichia*, *Francisella* and *Rickettsia*, as well as potential pathogens such as *Chlamydia*. This is a first attempt to apply BLSOM to the detection of potential pathogens. Since this approach can be directly applied to other vector arthropods of medical and veterinary importance, it has great potential for the detection of diverse microbes, including as-yet unidentified microorganisms, and to pre-empt emerging infectious diseases.

Unlike the conventional sequence homology searches used in most metagenomic studies, BLSOM does not require orthologous sequences for phylogenetic classification of metagenomic sequences. This is a great advantage, especially when applied to a group composed of poorly characterized microorganisms (Abe *et al*., 2005). In addition, ticks have highly variable genome sizes, ranging from nearly one third to over two times the size of the human genome, but only one whole-genome sequencing project is underway (Nene, 2009). This indicates that alignment-

based techniques may encounter difficulties in sorting out tick-derived sequences in the metagenomic libraries. On the other hand, BLSOM can separate bacterial/archaeal sequences from those of eukaryotes with high accuracy (Supplementary Figure S1) and thus have great potential for detecting symbiotic and commensal microorganisms in eukaryotic hosts with scarce or no genomic information. Despite the fact that Kingdom-BLSOM did not include tick genome sequences as these were not available at the time of construction, an average of 35% of the eukaryotic sequences were clustered together with those of other arthropods and thus likely to be derived from tick cells. Once it becomes available, the inclusion of a tick genome in BLSOM construction would facilitate the identification of eukaryotic reads as being tick-derived.

In a previous pyrosequencing study, in which the whole bodies of *I. ricinus* were directly subjected to DNA and RNA extraction, only 0.095% of the genomic DNA and 0.30% of the cDNA-derived sequences were assigned to known bacterial taxa (Carpi *et al*., 2011). Although this shotgun strategy seems ideal for capturing a snapshot of a microbial community with minimum bias introduced during sample preparation, there is a concern that genomic data for bacteria/archaea present in very low numbers may be masked by the huge amount of host-genome sequences. To address this, our strategy included a process of microbial enrichment using centrifugation, DNase treatment and filtration, to enrich bacterial/archaeal cells from tick homogenates prior to pyrosequencing. The microbial profiles obtained in this study, therefore,

may not completely reflect the whole picture regarding microbial composition in the tested ticks due to the biases introduced during the microbial enrichment and whole-genome amplification processes, especially in the case of AVm, where the sequences in the library were not likely to be randomly amplified (Supplementary Figure 2). Nonetheless, the fact that nearly half of the sequence reads were identified as bacterial/archaeal, and were associated with a diverse range of bacterial/archaeal phyla, indicated sufficient coverage by our method to obtain an overview of tick microbiomes.

Sequences of genera containing several tick-borne pathogens, such as *Anaplasma*, *Borrelia* and *Rickettsia*, were present in the tested samples (Table 3 and Supplementary Table 2). Further identification to the species level can be achieved by the use of traditional methods such as conventional species-specific PCR. This can be done to confirm the presence of tick-borne bacterial species such as *A. phagocytophilum*, *B. burgdorferi s.l.*, *R. japonica* and *R. tamurae*, which have previously been detected in the tested tick species (Hofhuis *et al*., 2006; Miyamoto *et al*., 1992; Morita *et al*., 1990; Murase *et al*., 2011). The data obtained in the present study, however, are useful for broadening our understanding of phylogenetic diversity among bacterial genera, and the extent of their habitats. For example, the genera *Chlamydia* and *Chlamydophila* were predicted to be harboured by IPf, IPm and HFfmn in high abundance (Table 3). Further pan-Chlamydia PCR and sequencing analysis supported the existence of microorganisms possibly

related to the genera *Neochlamydia*, *Parachlamydia* and *Simkania* (Table 4). Some representatives of these genera are implicated in diseases affecting their host animals (Corsaro *et al*., 2007; Meijer *et al*., 2006; von Bomhard *et al*., 2003). Since only the sequences of the genera *Chlamydia* and *Chlamydophila* were available from the phylum Chlamydiae at the time of BLSOM construction, those of the genera *Neochlamydia*, *Parachlamydia* and *Simkania* were not included in Genus group-BLSOM, so sequences from the phylum Chlamydiae were allocated to either genus *Chlamydia* or *Chlamydophila*. This technical limitation will be overcome as sequence data from a wider range of microorganisms become available. Although possible transmission of Chlamydial bacteria to cattle and humans through ticks was reported in several previous studies (Caldwell and Belden, 1973; Facco *et al*., 1992; McKercher *et al*., 1980), there has been no molecular genetic evidence to support these observational studies. Thus, this is the first molecular genetic evidence supporting the presence of Chlamydial organisms in ticks. Given the wide genetic diversity and host range, including several arthropods, of members of this phylum of bacteria (Horn, 2008), further research is warranted to improve the understanding of their biology, especially with respect to their pathogenic potential for humans and animals.

Genus group-BLSOM identified 133 to 223 different genera from each single tick pool (Table 2). The presence of a variety of bacterial/archaeal genera in ticks was demonstrated in previous metagenomic studies; in *R. microplus* and *I. ricinus*, 120 and 108 different genera were

identified, respectively, using a rDNA-based amplicon pyrosequencing approach (Andreotti *et al*., 2011; Carpi *et al*., 2011). This discrepancy may be explained by the fact that these two approaches rely on different analytic principles. One possible explanation is that BLSOM might have misclassified genome segments introduced through horizontal gene transfer (HGT) (Abe *et al*., 2005). Tamames and Moya (2008) reported evidence of HGT in 0.8%–1.5% and 2%–8% of the sequences in environmental metagenomic libraries using phylogenetic and composition-based methods, respectively, indicating that the present approach may require further refinement; for example, by combining BLSOM with rDNA-based phylogenetic classification as proposed elsewhere (Abe *et al*., 2005). It is also possible that the Genus group-BLSOM has overestimated the microbial diversity in the tested samples as a result of a decreased classification specificity on the genus level (Supplementary Figure S1). This drawback can be remedied by using only longer sequences to improve the specificity of the analysis or by using more genome sequences associated with the habitat under study during BLSOM construction (Abe *et al*., 2005; Weber *et al*., 2011).

The present study included the ticks maintained under laboratory conditions. The microbial diversity of those ticks was comparable to that of those collected in the field in terms of the number of different genus identified (Table 2). Environmental factors and developmental stages were reported to affect microbial communities in ticks (Andreotti *et al*., 2011; Carpi *et al*.,

2011; Clay *et al*., 2008; van Overbeek *et al*., 2008). Therefore, a further comparison of microbial diversity between laboratory colony and field ticks of the same species may be useful to elucidate environmental factors playing key roles on microbial compositions. IO and IP were sampled at the same site and time and have a common host preference for sika deer (*Cervus nippon*) in the sampling area (Isogai *et al*., 1996), allowing a comparison of microbial compositions between these two tick species. IO harboured *Rickettsia* and *Ehrlichia*, but IP harboured *Bartonella*, at high abundance, despite the presence of *Borrelia* and *Francisella* in both tick species at similar abundance (Table 3). These findings may be useful for improving the understanding of the vector potential of different tick species. In addition to these species differences, microbial communities are expected to vary between individual ticks as reported elsewhere (Andreotti *et al*., 2011; Carpi *et al*., 2011). One aspect to be addressed, which may help to explain these differences, is the interaction between microbial lineages. Indeed, antagonistic interactions between microbial lineages in ticks, including pathogenic *Rickettsia*, were demonstrated previously (de la Fuente *et al*., 2003; Macaluso *et al*., 2002). Thus, one of the most interesting challenges for future studies on tick microbiomes might be the association between microbial composition in ticks and the potential for pathogen transmission. The pictures of whole microbiomes harboured by each tick species generated in the present study provide the foundation for a statistical correlation analysis between pathogens and certain microbial lineages in individual ticks. Such studies could lead to

the identification of microbiological factors affecting the prevalence and persistence of pathogenic lineages in ticks.

The sequences associated with putative virulence-associated factors were identified from metagenomic libraries using BLASTX-based methods (Figure 4). It should be noted, however, that some of the factors, such as type III secretion systems, are responsible for both bacterial pathogenesis and symbiosis, including mutualism and commensalism (Coombes, 2009), and thus not all factors are directly or necessarily related to human/animal pathogens. The present study predominantly identified sequences related to antibiotic resistance (Figure 4). It has been recognized for some time that antibiotic resistance genes are not only associated with pathogens but also with a wide variety of environmental microbes (Allen *et al*., 2010). Several lines of evidence indicate that some of the clinically relevant resistance mechanisms in pathogenic organisms originated from such environmental bacteria through HGT (Wright, 2010). Direct contact between pathogens and resistance genes in commensals harboured by ticks could therefore represent a mechanism for the emergence of resistant pathogens.

Several studies described mutualistic relationships between microbes and host arthropods (Currie *et al*., 1999; Kaltenpoth *et al*., 2005; Scott *et al*., 2008). It is plausible that the tick microbiome also affects aspects of tick physiology. This was for example demonstrated by a reduction in reproductive fitness of *Amblyomma americanum* females following antibiotic

treatment, suggesting that the microbe plays a role in tick growth and development (Zhong *et al*., 2007). Further studies should consider the potential contributions of tick microbiomes to the survival of ticks in the environment; this could provide novel targets for the control of ticks and tick-borne pathogens.

BLSOM was used to estimate the phylogenetic origins of the sequences associated with putative virulence-associated factors; however, many of these remained unclassified (Table 5). This may indicate that those sequences had retained their functions, but lost sequence characteristics such as oligonucleotide frequency. One possible explanation is that the unclassified sequences were introduced through ancient HGT, and lost their oligonucleotide frequency in the recipient genome (either tick or other microbes). Indeed, the oligonucleotide composition of sequences introduced through ancient HGT drift over time (Brown, 2003), makes it difficult to correctly phylotype those sequences using composition-based methods such as BLSOM. Furthermore, numerous reports have associated bacterial virulence factors with HGT (Ho Sui *et al*., 2009; Juhas *et al*., 2009; Pallen and Wren, 2007). Another explanation is that these sequences simply originated from as-yet unclassified microorganisms. Regardless, the detection of virulence-associated factors in ticks may suggest either the existence or possible emergence of tick-borne pathogens and, thus, warrants further investigation to assess potential risks to human and animal health.

In conclusion, these results provide a foundation for the construction of a database of tick microbes, which may aid in the prediction of emerging tick-borne diseases. The present approach can be extended to detect or predict emerging pathogens in other vector arthropods such as mosquitoes and flies. A comprehensive understanding of tick microbiomes will also be useful for understanding tick biology, including vector competence and interactions with pathogens.

Acknowledgements

We thank everyone who helped with tick collection, especially Nariaki Nonaka, Jacqueline Schipper, Joseph W. Magona, Joseph M. Kamau, Noboru Sasaki and Noriko Nakao. The first author was supported by a research grant fellowship from the Japanese Society for the Promotion of Science (JSPS) for young scientists. This work was supported by a Grant-in-Aid for JSPS fellows and for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT), the program of Funding Research Center for Emerging and Re-emerging Infectious Disease, MEXT. The BLSOM calculation was done, in part, by the Earth Simulator of the Japan Agency for Marine-Earth Science and Technology.

References

Abe T, Kanaya S, Kinouchi M, Ichiba Y, Kozuki T, Ikemura T. (2003). Informatics for unveiling hidden genome signatures. *Genome Res* **13**: 693-702.

Abe T, Sugawara H, Kinouchi M, Kanaya S, Ikemura T. (2005). Novel phylogenetic studies of genomic sequence fragments derived from uncultured microbe mixtures in environmental and clinical samples. *DNA Res* **12**: 281-290.

Allen HK, Donato J, Wang HH, Cloud-Hansen KA, Davies J, Handelsman J. (2010). Call of the wild: antibiotic resistance genes in natural environments. *Nat Rev Microbiol* **8**: 251-259.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic local alignment search tool. *J Mol Biol* **215**: 403-410.

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

Bowman AS, Nuttall, PA. (2008). *Ticks: Biology, Disease and Control*. Cambridge University Press: Cambridge.

Benson MJ, Gawronski JD, Eveleigh DE, Benson DR. (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.

Beugnet F, Marié JL. (2009). Emerging arthropod-borne diseases of companion animals in Europe. *Vet Parasitol* **163**: 298-305.

Brown JR. (2003). Ancient horizontal gene transfer. *Nat Rev Genet* **4**: 121-132.

Caldwell HD, Belden EL. (1973). Studies of the role of *Dermacentor occidentalis* in the transmission of bovine chlamydial abortion. *Infect Immun* **7**: 147-151.

Carpi G, Cagnacci F, Wittekindt NE, Zhao F, Qi J, Tomsho LP *et al*. (2011). Metagenomic profile of the bacterial communities associated with *Ixodes ricinus* ticks. *PLoS One* **6**: e25604.

Chan C-KK, Hsu AL, Tang S-L, Halgamuge SK. (2008). Using growing self-organising maps to improve the binning process in environmental whole-genome shotgun sequencing. J Biomed Biotechnol 2008, doi:10.1155/2008/513701

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 JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ *et al*. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* **37**: D141-145.

Coombes BK. (2009). Type III secretion systems in symbiotic adaptation of pathogenic and non-pathogenic bacteria. *Trends Microbiol* **17**: 89-94.

Corsaro D, Thomas V, Goy G, Venditti D, Radek R, Greub G. (2007). *'Candidatus* Rhabdochlamydia crassificans', an intracellular bacterial pathogen of the cockroach *Blatta orientalis* (Insecta: Blattodea). *Syst Appl Microbiol* **30**: 221-228.

Cox-Foster DL, Conlan S, Holmes EC, Palacios G, Evans JD, Moran NA *et al*. (2007). A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* **318**: 283-287.

Currie CR, Mueller UG, Malloch D. (1999). The agricultural pathology of ant fungus gardens. *Proc Natl Acad Sci U S A* **96**: 7998-8002.

de la Fuente J, Blouin EF, Kocan KM. (2003). Infection exclusion of the rickettsial pathogen *Anaplasma marginale* in the tick vector *Dermacentor variabilis*. *Clin Diagn Lab Immunol* **10**: 182-184.

Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP *et al*. (2009). Community-wide analysis of microbial genome sequence signatures. *Genome Biol* **10**: R85.

Ezaki T, Suzuki S. (1982). Achromopeptidase for lysis of anaerobic gram-positive cocci. *J Clin Microbiol* **16**: 844-846.

Facco F, Grazi G, Bonassi S, Magnani M, Di Pietro P. (1992). Chlamydial and rickettsial transmission through tick bite in children. *Lancet* **339**: 992-993.

Grenier AM, Duport G, Pagès S, Condemine G, Rahbé Y. (2006). The phytopathogen *Dickeya dadantii* (*Erwinia chrysanthemi* 3937) is a pathogen of the pea aphid. *Appl Environ Microbiol* **72**: 1956-1965.

Ho Sui SJ, Fedynak A, Hsiao WW, Langille MG, Brinkman FS. (2009). The association of virulence factors with genomic islands. *PLoS One* **4**: e8094.

Hofhuis A, van der Giessen JWB, Borgsteede FHM, Wielinga PR, Notermans DW, van Pelt W. (2006). Lyme borreliosis in the Netherlands: strong increase in GP consultations and hospital admissions in past 10 years. *Euro Surveillance: European Communicable Disease Bulletin* **11**.

Horn M. (2008). Chlamydiae as symbionts in eukaryotes. *Annu Rev Microbiol* **62**: 113-131.

Imaoka K, Kaneko S, Tabara K, Kusatake K, Morita E. (2011). The First Human Case of *Rickettsia tamurae* Infection in Japan. *Case Rep Dermatol* **3**: 68-73.

Isogai E, Isogai H, Masuzawa T, Postic D, Baranton G, Kamewaka Y *et al*. (1996). *Borrelia burgdorferi sensu lato* in an endemic environment: wild sika deer (*Cervus nippon yesoensis*) with infected ticks and antibodies. *Microbiol Immunol* **40**: 13-19.

Juhas M, van der Meer JR, Gaillard M, Harding RM, Hood DW, Crook DW. (2009). Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol Rev* **33**: 376-393.

Kaltenpoth M, Göttler W, Herzner G, Strohm E. (2005). Symbiotic bacteria protect wasp larvae from fungal infestation. *Curr Biol* **15**: 475-479.

Kanaya S, Kinouchi M, Abe T, Kudo Y, Yamada Y, Nishi T *et al*. (2001). Analysis of codon usage diversity of bacterial genes with a self-organizing map (SOM): characterization of horizontally transferred genes with emphasis on the *E. coli* O157 genome. *Gene* **276**: 89-99.

Kohonen T. (1990). The self-organizing map. *Proceedings of the IEEE*. pp 1464-1480.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H *et al*. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947-2948.

Lewis D. (1979). The detection of rickettsia-like microorganisms within the ovaries of female *Ixodes ricinus* ticks. *Z Parasitenkd* **59**: 295-298.

Macaluso KR, Sonenshine DE, Ceraul SM, Azad AF. (2002). Rickettsial infection in *Dermacentor variabilis* (Acari: Ixodidae) inhibits transovarial transmission of a second *Rickettsia*. *J Med Entomol* **39**: 809-813.

Martin C, Diaz NN, Ontrup J, Nattkemper TW. (2008). Hyperbolic SOM-based clustering of DNA fragment features for taxonomic visualization and classification. Bioinformatics **24**: 1568–1574.

Mavromatis K, Ivanova N, Barry K, Shapio H, Goltsman E, McHardy AC, *et al*. (2007). Use of simulated data sets to evaluate the fidelity of metagenomic processing methods. *Nat Methods* **4**: 495-500.

McKercher DG, Wada EM, Ault SK, Theis JH. (1980). Preliminary studies on transmission of *Chlamydia* to cattle by ticks (*Ornithodoros coriaceus*). *Am J Vet Res* **41**: 922-924.

Meijer A, Roholl PJ, Ossewaarde JM, Jones B, Nowak BF. (2006). Molecular evidence for association of *chlamydiales* bacteria with epitheliocystis in leafy seadragon (*Phycodurus*

eques), silver perch (*Bidyanus bidyanus*), and barramundi (*Lates calcarifer*). *Appl Environ Microbiol* **72**: 284-290.

Miyamoto K, Nakao M, Uchikawa K, Fujita H. (1992). Prevalence of Lyme borreliosis spirochetes in ixodid ticks of Japan, with special reference to a new potential vector, *Ixodes ovatus* (Acari: Ixodidae). *J Med Entomol* **29**: 216-220.

Morita C, Yamamoto S, Tsuchiya K, Yoshida Y, Yabe T, Kawabata N *et al*. (1990). Prevalence of spotted fever group rickettsia antibody in *Apodemus speciosus* captured in an endemic focus in Miyazaki Prefecture, Japan. *Jpn J Med Sci Biol* **43**: 15-18.

Murase Y, Konnai S, Hidano A, Githaka NW, Ito T, Takano A *et al*. (2011). Molecular detection of *Anaplasma phagocytophilum* in cattle and *Ixodes persulcatus* ticks. *Vet Microbiol* **149**: 504-507.

Nakabachi A, Ishikawa H. (1999). Provision of riboflavin to the host aphid, *Acyrthosiphon pisum*, by endosymbiotic bacteria, *Buchnera*. *J Insect Physiol* **45**: 1-6.

Nene V. (2009). Tick genomics--coming of age. *Front Biosci* **14**: 2666-2673.

Noda H, Munderloh UG, Kurtti TJ. (1997). Endosymbionts of ticks and their relationship to *Wolbachia* spp. and tick-borne pathogens of humans and animals. *Appl Environ Microbiol* **63**: 3926-3932.

Pallen MJ, Wren BW. (2007). Bacterial pathogenomics. *Nature* **449**: 835-842.

Parola P. (2004). Tick-borne rickettsial diseases: emerging risks in Europe. *Comp Immunol Microbiol Infect Dis* **27**: 297-304.

Parola P, Davoust B, Raoult D. (2005). Tick- and flea-borne rickettsial emerging zoonoses. *Vet Res* **36**: 469-492.

Roshdy MA. (1968). A rickettsialike microorganism in the tick *Ornithodoros savignyi*: Observations on its structure and distribution in the tissues of the tick. *J Invertebr Pathol* **11**: 155-169.

Scarborough CL, Ferrari J, Godfray HC. (2005). Aphid protected from pathogen by endosymbiont. *Science* **310**: 1781.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al*. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537-7541.

Scott JJ, Oh DC, Yuceer MC, Klepzig KD, Clardy J, Currie CR. (2008). Bacterial protection of beetle-fungus mutualism. *Science* **322**: 63.

Spolidorio MG, Labruna MB, Mantovani E, Brandao PE, Richtzenhain LJ, Yoshinari NH. (2010). Novel spotted fever group rickettsiosis, Brazil. *Emerg Infect Dis* **16**: 521-523.

Suen G, Scott JJ, Aylward FO, Adams SM, Tringe SG, Pinto-Tomás AA *et al*. (2010). An insect herbivore microbiome with high plant biomass-degrading capacity. *PLoS Genet* **6**: e1001129.

Tamames J, Moya A. (2008). Estimating the extent of horizontal gene transfer in metagenomic sequences. *BMC Genomics* **9**: 136.

Tatusov RL, Koonin EV, Lipman DJ. (1997). A genomic perspective on protein families. *Science* **278**: 631-637.

Tsuchida T, Koga R, Horikawa M, Tsunoda T, Maoka T, Matsumoto S *et al*. (2010). Symbiotic bacterium modifies aphid body color. *Science* **330**: 1102-1104.

Uehara H, Iwasaki Y, Wada C, Ikemura T, Abe T. (2011). A novel bioinformatics strategy for searching industrially useful genome resources from metagenomic sequence libraries. *Genes Genet Syst* **86**: 53-66.

van Overbeek L, Gassner F, van der Plas CL, 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.

von Bomhard W, Polkinghorne A, Lu ZH, Vaughan L, Vögtlin A, Zimmermann DR *et al*. (2003). Detection of novel chlamydiae in cats with ocular disease. *Am J Vet Res* **64**: 1421-1428.

Warnecke F, Luginbühl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT *et al*. (2007). Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**: 560-565.

Weber M, Teeling H, Huang S, Waldmann J, Kassabgy M, Fuchs BM *et al*. (2011). Practical application of self-organizing maps to interrelate biodiversity and functional data in NGS-based metagenomics. *ISME J* **5**: 918-928.

Weiss B, Aksoy S. (2011). Microbiome influences on insect host vector competence. *Trends Parasitol* **27**: 514-522.

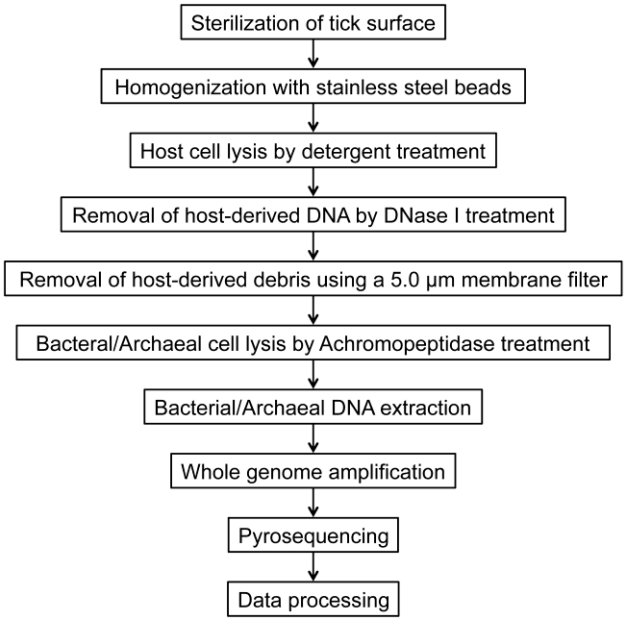
Wright GD. (2010). Antibiotic resistance in the environment: a link to the clinic? *Curr Opin Microbiol* **13**: 589-594.

Yu XJ, Liang MF, Zhang SY, Liu Y, Li JD, Sun YL *et al*. (2011). Fever with thrombocytopenia associated with a novel bunyavirus in China. *N Engl J Med* **364**: 1523-1532.

Zhong J, Jasinskas A, Barbour AG. (2007). Antibiotic treatment of the tick vector *Amblyomma americanum* reduced reproductive fitness. *PLoS One* **2**: e405.

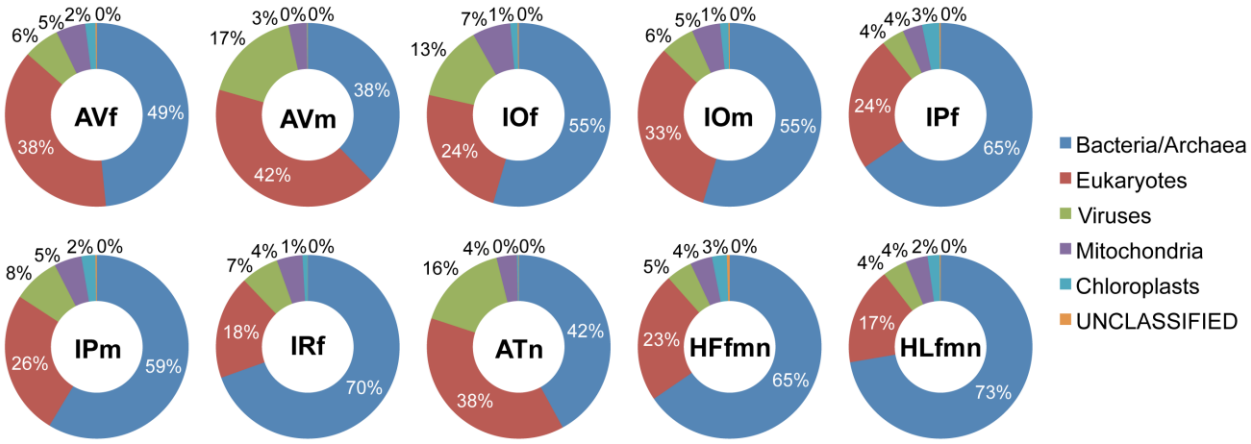
Zhou CE, Smith J, Lam M, Zemla A, Dyer MD, Slezak T. (2007). MvirDB--a microbial database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence applications. *Nucleic Acids Res* **35**: D391-394.

Figure 1. Workflow for bacterial/archaeal purification and metagenomic analysis.



The present strategy employed a process for purifying of bacteria/archaea, which comprised centrifugation, DNase treatment and filtration, to enrich bacterial/archaeal cells from tick homogenates prior to whole-genome amplification and pyrosequencing.

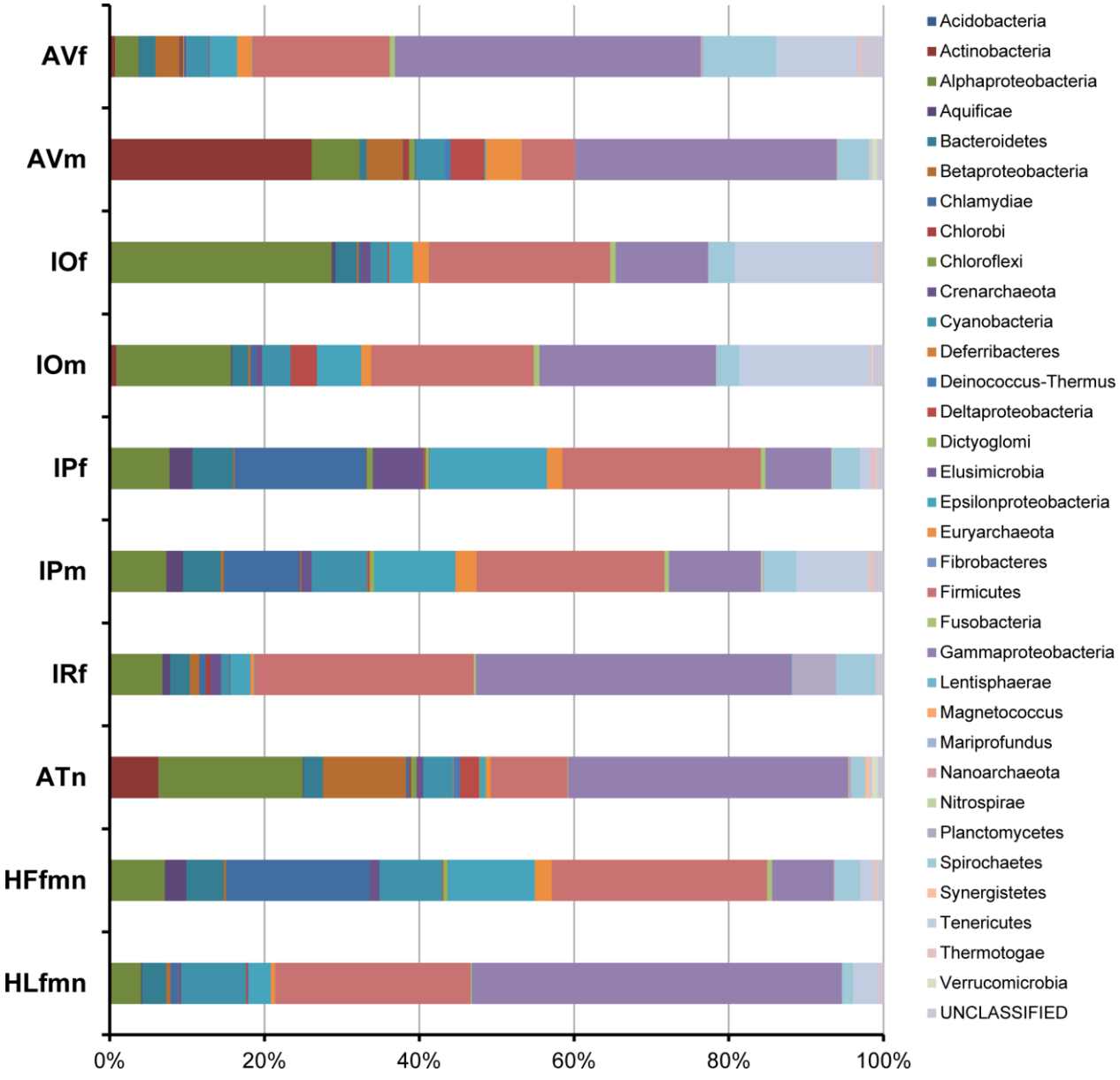
Figure 2. Kingdom classification of the metagenomic sequences from each tick pool.



620 Sequence reads longer than 300 bp were classified into bacteria/archaea, eukaryotes, viruses, mitochondria or chloroplasts using Kingdom-BLSOM. The percentage of sequences in each category is provided. The pool ID is shown in the centre of each pie chart.

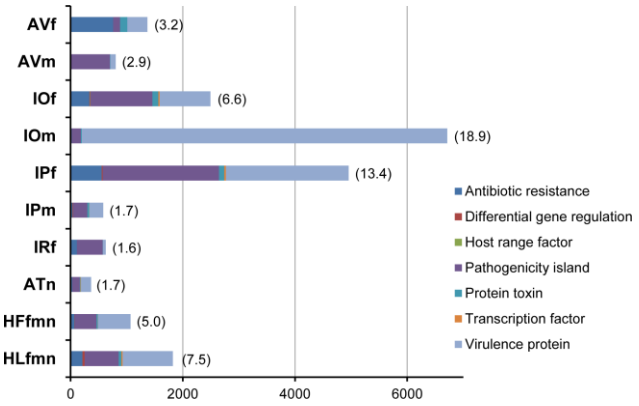
625

Figure 3. Phylum classification of the metagenomic sequences from each tick pool.



Sequence reads classified as bacterial/archaeal using Kingdom-BLSOM were further classified at the phylum level using Bacteria/Archaea-BLSOM. Thirty-three phyla are indicated by different colours. The pool ID is shown on the left of the graph.

Figure 4. Sequences associated with virulence-associated factors.



640 Sequences associated with putative virulence-associated factors were identified by BLASTX against the MvirDB database (Zhou *et al.*, 2007), applying a cut-off value of 1e-5. The resulting data are shown in seven categories denoted by different colours. The numbers at the bottom of the graph indicate the number of sequences. The numbers in parentheses indicate the percentage of total sequence reads. The pool ID is shown on the left of the graph.

645

Table 1. Collection details for the ticks used in the present study.

Pool ID	Tick species	Collection site	Adult		Nymph	Total
			Female	Male		
AVf	*Amblyomma variegatum*	Laboratory colony[a]	20			20
AVm	*Amblyomma variegatum*	Laboratory colony[a]		20		20
IOf	*Ixodes ovatus*	Hokkaido, Japan	59			59
IOm	*Ixodes ovatus*	Hokkaido, Japan		110		110
IPf	*Ixodes persulcatus*	Hokkaido, Japan	56			56
IPm	*Ixodes persulcatus*	Hokkaido, Japan		75		75
IRf	*Ixodes ricinus*	Soesterberg, The Netherlands	55			55
ATn	*Amblyomma testudinarium*	Miyazaki, Japan			57	57
HFfmn	*Haemaphysalis formosensis*	Miyazaki, Japan	32	24	513	569
HLfmn	*Haemaphysalis longicornis*	Miyazaki, Japan	16	20	84	120

[a]This laboratory colony was originally obtained from The Gambia and has been maintained at Utrecht Centre for Tick-borne Diseases, Utrecht University, The Netherlands.

Table 2. Summary of pyrosequencing and BLSOM classifications.

	AVf	AVm	IOf	IOm	IPf	IPm	IRf	ATn	HFfmn	HLfmn
Total no. of reads	42,258	27,582	37,667	35,544	37,136	33,789	39,810	21,639	21,213	24,249
Average read length (nt)	163.6	391.4	379.8	154.6	371.7	143.5	149.9	149.0	185.0	187.5
No. of long reads (≥300 bp)	8,160	22,201	29,804	2,962	28,618	4,210	5,438	2,906	4,502	5,186
No. of reads assigned to bacteria/archaea (%)[a]	3,951 (48.4)	8,389 (37.8)	16,239 (54.5)	1,620 (54.7)	18,711 (65.4)	2,468 (58.6)	3,782 (69.5)	1,222 (42.1)	2,946 (65.4)	3,750 (72.3)
No. of reads assigned to a bacterial/archaeal phyla (%)[b]	3,841 (97.2)	8,272 (98.6)	16,088 (99.1)	1,598 (98.6)	18,556 (99.2)	2,439 (98.8)	3,748 (99.1)	1,213 (99.3)	2,925 (99.3)	3,728 (99.4)
No. of phyla obtained	27	28	28	23	28	25	22	27	24	20
No. of reads assigned to known bacterial/archaeal genera (%)[c]	3,419 (89.0)	7,068 (85.4)	14,410 (89.6)	1,410 (88.2)	16,566 (89.3)	2,131 (87.4)	3,398 (90.7)	1,119 (92.3)	2,605 (89.1)	3,319 (89.0)
No. of reads assigned to candidate bacterial/archaeal genera (%)	71 (1.8)	62 (0.7)	700 (4.4)	66 (4.1)	669 (3.6)	98 (4.0)	76 (2.0)	13 (1.1)	89 (3.0)	75 (2.0)
No. of bacterial/archaeal genera obtained	158	223	215	143	185	162	133	201	138	156

[a]Calculated using the formula: (No. of reads assigned to bacteria/archaea)/{No. of long reads (≥300 bp)} × 100. [b]Calculated using the formula: (No. of reads assigned to any phyla)/(No. of reads assigned to bacteria/archaea) × 100. [c]Calculated using the formula: (No. of reads assigned to any genera)/(No. of reads assigned to any phyla) × 100.

Table 3. Top 20 genera associated with the metagenomic sequences from each tick pool.

	AVf		AVm		IOf		IOm		IPf		IPm		IRf		ATn		HFfmn		HLfmn	
1	*Borrelia*	(8.6)	*Clavibacter*	(13.4)	*Rickettsia*	(19.7)	*Ureaplasma*	(11.0)	*Chlamydophila*	(8.8)	*Ureaplasma*	(7.6)	*Borrelia*	(5.3)	*Pseudoalteromonas*	(17.2)	*Chlamydophila*	(10.5)	*Lactobacillus*	(7.1)
2	*Mycoplasma*	(6.2)	*Pseudoalteromonas*	(5.1)	*Ureaplasma*	(15.4)	*Rickettsia*	(7.3)	*Streptococcus*	(8.5)	*Chlamydophila*	(5.4)	*Lactobacillus*	(5.3)	*Rickettsia*	(10.0)	*Streptococcus*	(8.5)	U (Gammaproteobacteria)	(5.1)
3	*Staphylococcus*	(4.0)	U (Euryarchaeota)	(4.7)	*Mycoplasma*	(8.2)	*Mycoplasma*	(7.1)	*Helicobacter*	(6.9)	*Streptococcus*	(5.0)	*Streptococcus*	(4.7)	*Synechococcus*	(2.1)	*Chlamydia*	(6.7)	*Salmonella*	(4.6)
4	*Actinobacillus*	(4.0)	*Treponema*	(3.6)	*Clostridium*	(5.8)	*Clostridium*	(3.1)	*Chlamydia*	(5.9)	*Campylobacter*	(4.7)	*Ureaplasma*	(4.3)	*Wigglesworthia*	(2.1)	*Helicobacter*	(5.5)	*Grimontia*	(4.3)
5	*Coxiella*	(4.0)	*Synechococcus*	(3.2)	*Ehrlichia*	(4.0)	*Helicobacter*	(3.1)	*Campylobacter*	(5.5)	*Helicobacter*	(4.4)	U (Gammaproteobacteria)	(4.1)	*Clostridium*	(1.8)	*Prochlorococcus*	(4.2)	*Providencia*	(3.7)
6	*Grimontia*	(3.7)	U (Gammaproteobacteria)	(3.1)	*Francisella*	(3.4)	*Desulfovibrio*	(3.1)	*Prochlorococcus*	(3.5)	*Prochlorococcus*	(3.9)	*Grimontia*	(3.9)	U (Alphaproteobacteria)	(1.6)	*Campylobacter*	(3.5)	*Coxiella*	(3.4)
7	*Thiocystis*	(3.7)	*Arcanobacterium*	(2.9)	*Borrelia*	(2.9)	U (Gammaproteobacteria)	(3.1)	*Borrelia*	(3.0)	*Mycoplasma*	(3.4)	*Bacillus*	(3.8)	*Orientia*	(1.6)	*Bacillus*	(3.3)	*Cyanothece*	(3.1)
8	*Luteimonas*	(3.4)	*Actinobacillus*	(2.7)	U (Euryarchaeota)	(2.1)	*Streptococcus*	(2.7)	*Clostridium*	(2.7)	*Borrelia*	(3.1)	*Luteimonas*	(3.0)	*Bordetella*	(1.5)	*Clostridium*	(3.0)	*Streptococcus*	(3.0)
9	*Dichelobacter*	(2.7)	*Micromonospora*	(2.6)	*Campylobacter*	(2.0)	*Phytoplasma*	(2.6)	*Bacillus*	(2.7)	*Chlamydia*	(3.0)	*Vibrio*	(2.9)	*Bacillus*	(1.4)	*Thermoanaerobacter*	(2.8)	*Staphylococcus*	(2.7)
10	*Phytoplasma*	(2.5)	*Salmonella*	(2.6)	*Staphylococcus*	(1.9)	*Francisella*	(2.4)	*Thermoanaerobacter*	(2.6)	U (Euryarchaeota)	(2.8)	*Rickettsia*	(2.7)	U (Gammaproteobacteria)	(1.4)	*Borrelia*	(2.5)	*Bacillus*	(2.7)
11	*Campylobacter*	(2.4)	U (Actinobacteria)	(2.6)	*Orientia*	(1.9)	*Borrelia*	(2.3)	*Sulfurihydrogenibium*	(2.4)	*Bacillus*	(2.4)	*Thiothrix*	(2.5)	*Amycolatopsis*	(1.2)	*Lactobacillus*	(2.3)	*Acinetobacter*	(2.5)
12	U (Tenericutes)	(2.4)	*Hydrogenovibrio*	(2.5)	C (Gammaproteobacteria)	(1.5)	*Thiothrix*	(2.2)	*Francisella*	(2.3)	*Clostridium*	(2.3)	*Actinobacillus*	(2.4)	*Micromonospora*	(1.2)	U (Euryarchaeota)	(2.2)	*Mycoplasma*	(2.5)
13	*Lactobacillus*	(2.4)	*Aeromonas*	(2.3)	*Streptococcus*	(1.4)	*Ehrlichia*	(1.9)	U (Euryarchaeota)	(2.0)	*Thermoanaerobacter*	(2.0)	*Coxiella*	(2.4)	*Francisella*	(1.2)	*Sulfurihydrogenibium*	(2.1)	*Pseudoxanthomonas*	(2.4)
14	*Thiothrix*	(2.1)	U (Alphaproteobacteria)	(1.9)	*Wolbachia*	(1.3)	*Staphylococcus*	(1.6)	C (Chlamydiae)	(2.0)	*Bartonella*	(1.6)	*Enterococcus*	(2.3)	*Streptococcus*	(1.2)	*Bartonella*	(1.5)	*Xenorhabdus*	(2.2)
15	*Salmonella*	(2.1)	*Streptococcus*	(1.7)	*Bacillus*	(1.2)	*Campylobacter*	(1.4)	*Wolbachia*	(1.9)	*Lactobacillus*	(1.6)	*Mycoplasma*	(2.3)	*Xanthomonas*	(1.2)	C (Chlamydiae)	(1.4)	*Actinobacillus*	(1.9)
16	U (Euryarchaeota)	(2.0)	*Hafnia*	(1.6)	*Buchnera*	(1.2)	*Lactobacillus*	(1.4)	*Bartonella*	(1.8)	U (Gammaproteobacteria)	(1.6)	*Colwellia*	(2.3)	*Aeromonas*	(1.0)	*Wolbachia*	(1.3)	*Rickettsia*	(1.9)
17	*Neisseria*	(2.0)	*Geobacter*	(1.4)	U (Gammaproteobacteria)	(1.1)	*Listeria*	(1.4)	*Sulfurimonas*	(1.2)	*Francisella*	(1.5)	*Francisella*	(1.9)	*Borrelia*	(1.0)	*U (Firmicutes)*	(1.2)	*Thiothrix*	(1.8)
18	U (Gammaproteobacteria)	(1.8)	*Rickettsiella*	(1.4)	*Phytoplasma*	(1.1)	C (Alphaproteobacteria)	(1.3)	*Lactobacillus*	(1.2)	*Sulfurihydrogenibium*	(1.4)	*Salmonella*	(1.7)	*Methylobacillus*	(1.0)	*C (Bacteroide-tes)*	(1.2)	*Campylobacter*	(1.7)
19	*Providencia*	(1.8)	*Methylobacterium*	(1.3)	C (Bacteroide-tes)	(1.1)	U (Euryarchaeota)	(1.3)	*Alkaliphilus*	(1.1)	C (Chlamydiae)	(1.3)	*Acinetobacter*	(1.7)	*Shigella*	(1.0)	*Cyanothece*	(1.2)	*Listeria*	(1.5)
20	*Vibrio*	(1.8)	*Actinoalloteichus*	(1.2)	C (Tenericutes)	(1.0)	*Bacillus*	(1.3)	*Lamprocystis*	(1.1)	*Phytoplasma*	(1.3)	*Clostridium*	(1.7)	*Staphylococcus*	(1.0)	*Sulfurimonas*	(1.2)	*Nostoc*	(1.5)

Numbers in parentheses indicate the percentage. U, Unclassified; C, Candidatus.

Table 4. Phylogenetic analysis of 16S rRNA gene sequences amplified by pan-Chlamydia PCR.

Operational taxonomic unit (OTU)	Sequence ID	Clones assigned to each OTU					RDP result[a]		BLASTn search result[a]		
		IPf	IPm	HFf	HFm	HFn	Family (%)[b]	Genus (%)[b]	Identity (%)[c]	Source	Accession no.
1	1-1*	Yes	Yes			Yes	*Simkaniaceae* (80%)	*Simkania* (80%)	*Rhabdochlamydia crassificans* strain CRIB01 (98%)	Cockroach (*Blatta orientalis*)	AY928092
	1-2	Yes									
	1-3	Yes									
	1-4		Yes								
	1-5	Yes									
2	2-1*		Yes	Yes	Yes	Yes	*Parachlamydiaceae* (53%)	*Neochlamydia* (22%)	Uncultured Chlamydiae bacterium clone UFC3 (90%)	Sea bass (*Dicentrarchus labrax*)	FJ376381
	2-2		Yes	Yes							
	2-3			Yes							
	2-4				Yes						
3	3-1*				Yes		*Parachlamydiaceae* (89%)	*Neochlamydia* (73%)	Uncultured Chlamydiales bacterium isolate UKC7 (90%)	Koala (*Phascolarctos cinereus*)	AY167120
	3-2		Yes		Yes						
4	4-1*			Yes			*Parachlamydiaceae* (95%)	*Neochlamydia* (67%)	Uncultured clinical *Neochlamydia* sp. clone WB13 (92%)	Cat (*Felis silvestris catus*)	AY225593
	4-2			Yes							
5	5-1*				Yes		*Parachlamydiaceae* (76%)	*Parachlamydia* (74%)	Uncultured Chlamydiales CRG20 (88%)	Leafy seadragon (*Phycodurus eques*)	AY013396

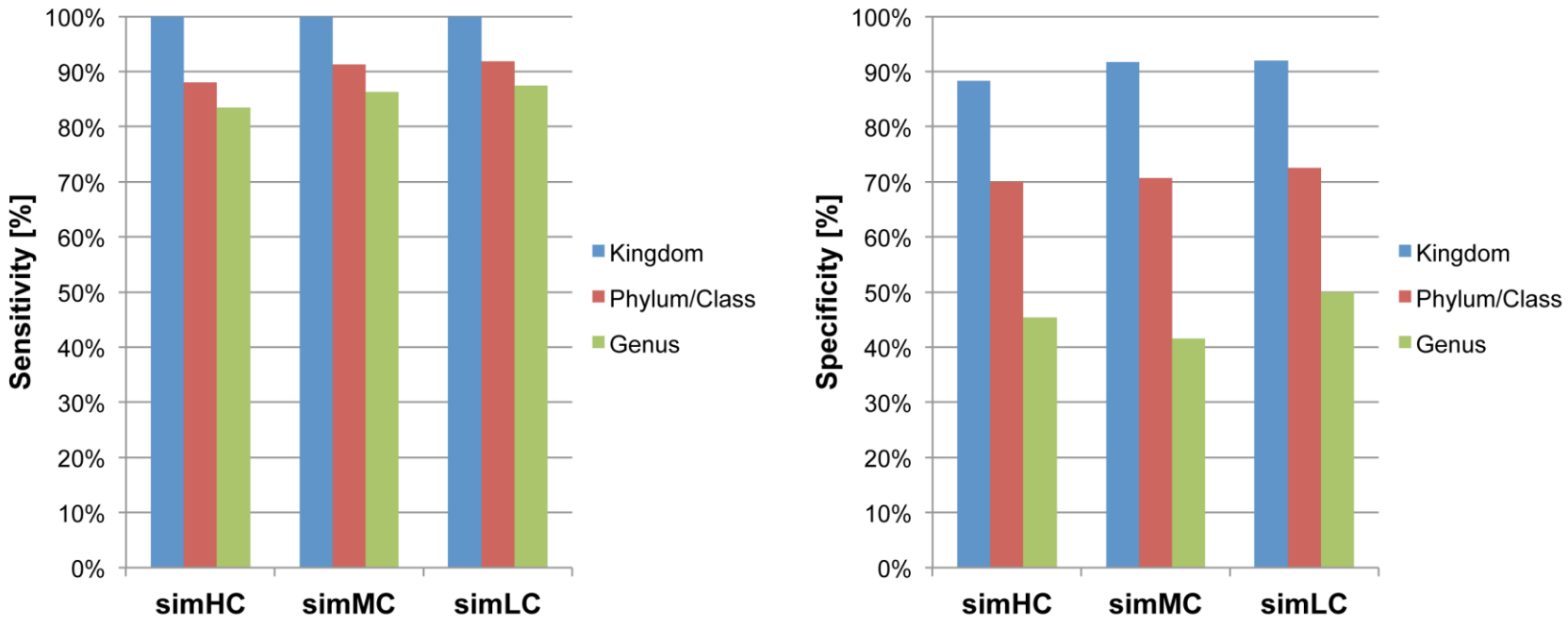
*Representative sequence. [a]Analysis was conducted using a representative sequence for each OTU. [b]Percentages indicate bootstrap values. [c]Percentages indicate the similarity between two sequences.

Table 5. Origins of the sequences of virulence-associated factors.

	AVf		AVm		IOf		IOm		IPf		IPm		IRf		ATn		HFfmn		HLfmn	
Antibiotic resistance																				
1	U	(424)	Actinobacteria	(1)	Alphaproteobacteria	(90)	U	(6)	U	(130)	Firmicutes	(8)	Gammaproteobacteria	(68)	Gammaproteobacteria	(8)	U	(15)	Gammaproteobacteria	(50)
2	Gammaproteobacteria	(85)	Gammaproteobacteria	(1)	U	(82)	Firmicutes	(5)	Firmicutes	(110)	Chlamydiae	(6)	U	(20)	U	(7)	Firmicutes	(9)	U	(39)
3	Firmicutes	(73)			Firmicutes	(49)	Gammaproteobacteria	(4)	Gammaproteobacteria	(50)	U	(5)	Alphaproteobacteria	(6)	Alphaproteobacteria	(5)	Epsilonproteobacteria	(4)	Firmicutes	(16)
4	Spirochaetes	(38)			Gammaproteobacteria	(44)	Alphaproteobacteria	(2)	Chlamydiae	(49)	Gammaproteobacteria	(4)	Firmicutes	(5)	Actinobacteria	(1)	Gammaproteobacteria	(4)	Cyanobacteria	(5)
5	Euryarchaeota	(26)			Crenarchaeota	(17)	Cyanobacteria	(1)	Cyanobacteria	(44)	Cyanobacteria	(3)	Cyanobacteria	(4)	Betaproteobacteria	(1)	Chlamydiae	(3)	Tenericutes	(2)
Pathogenicity island																				
1	Gammaproteobacteria	(46)	U	(505)	U	(329)	U	(86)	U	(546)	U	(61)	Gammaproteobacteria	(164)	Gammaproteobacteria	(30)	U	(69)	Gammaproteobacteria	(183)
2	Firmicutes	(28)	Gammaproteobacteria	(98)	Alphaproteobacteria	(293)	Firmicutes	(29)	Firmicutes	(346)	Firmicutes	(37)	Firmicutes	(149)	U	(24)	Firmicutes	(38)	Firmicutes	(88)
3	U	(16)	Actinobacteria	(29)	Firmicutes	(152)	Gammaproteobacteria	(12)	Chlamydiae	(286)	Chlamydiae	(36)	U	(59)	Actinobacteria	(15)	Chlamydiae	(36)	U	(58)
4	Betaproteobacteria	(9)	Alphaproteobacteria	(18)	Gammaproteobacteria	(80)	Tenericutes	(9)	Gammaproteobacteria	(165)	Bacteroidetes	(19)	Alphaproteobacteria	(12)	Alphaproteobacteria	(13)	Alphaproteobacteria	(23)	Alphaproteobacteria	(13)
5	Alphaproteobacteria	(3)	Cyanobacteria	(12)	Bacteroidetes	(32)	Alphaproteobacteria	(8)	Epsilonproteobacteria	(140)	Epsilonproteobacteria	(17)	Cyanobacteria	(11)	Betaproteobacteria	(7)	Epsilonproteobacteria	(19)	Cyanobacteria	(13)
Virulence protein																				
1	Gammaproteobacteria	(192)	Gammaproteobacteria	(51)	Alphaproteobacteria	(291)	Gammaproteobacteria	(81)	U	(625)	U	(60)	Gammaproteobacteria	(14)	U	(35)	U	(96)	Gammaproteobacteria	(197)
2	U	(55)	U	(14)	U	(238)	U	(34)	Firmicutes	(411)	Firmicutes	(44)	U	(14)	Gammaproteobacteria	(28)	Firmicutes	(68)	U	(116)
3	Firmicutes	(34)	Firmicutes	(12)	Firmicutes	(92)	Alphaproteobacteria	(23)	Chlamydiae	(285)	Chlamydiae	(27)	Firmicutes	(9)	Alphaproteobacteria	(8)	Chlamydiae	(46)	Firmicutes	(86)
4	Betaproteobacteria	(22)	Actinobacteria	(3)	Gammaproteobacteria	(90)	Deltaproteobacteria	(21)	Epsilonproteobacteria	(159)	Gammaproteobacteria	(24)	Alphaproteobacteria	(4)	Firmicutes	(8)	Cyanobacteria	(26)	Cyanobacteria	(35)
5	Bacteroidetes	(7)	Betaproteobacteria	(2)	Tenericutes	(25)	Firmicutes	(15)	Gammaproteobacteria	(117)	Epsilonproteobacteria	(19)	Cyanobacteria	(1)	Betaproteobacteria	(6)	Gammaproteobacteria	(18)	Alphaproteobacteria	(15)

675 Numbers in parentheses indicate the total number of sequences. U, Unclassified.

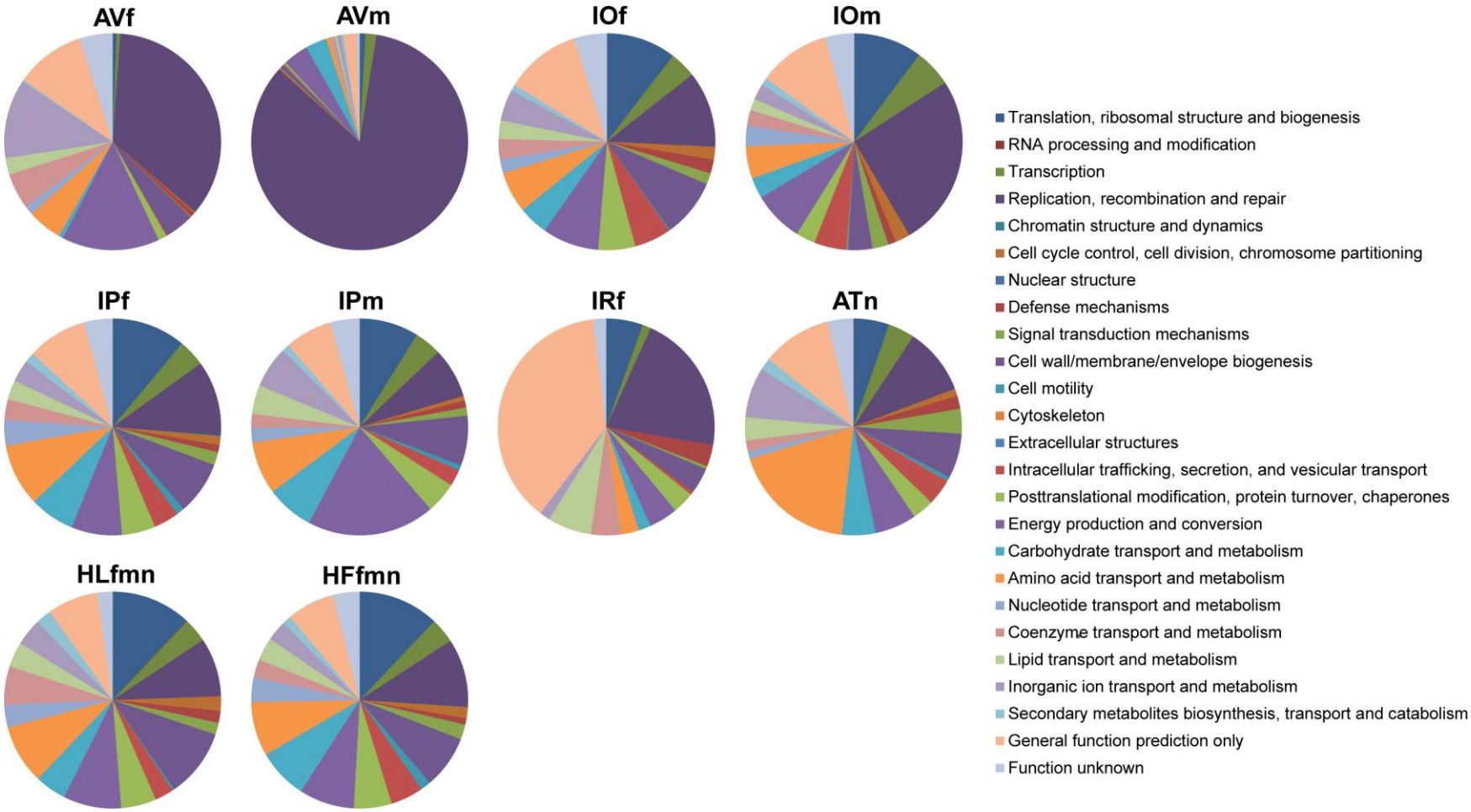
Supplementary Figure S1. Evaluation of BLSOM-based classification using simulated datasets.



Evaluation of Kingdom-, Bacteria/Archaea- and Genus group BLSOM-based classification using simulated datasets of high (simHC), medium (simMC) and low (simLC) complexities (Mavromatis *et al*., 2007) on contigs longer than 300 bp after normalization of sequence length. The left graph shows the sensitivity {true positives/(true positives + false negatives)} (%) and the right graph shows the specificity {true positives/(true positives + false positives)} (%) of the BLSOM classification. Different taxonomic levels are shown in different colours.

Reference: Mavromatis K, Ivanova N, Barry K, Shapio H, Goltsman E, McHardy AC, et al. (2007). Use of simulated data sets to evaluate the fidelity of metagenomic processing methods. *Nat Methods* **4**: 495-500.

Supplementary Figure S2. Functional annotation using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database.



690 Sequence reads were annotated by BLASTX against the Clusters of Orthologous Groups (COG) database with an E-value threshold of 1×10^{-5} and were classified according to COG functional categories. Twenty-four COG categories are indicated by different colours. The pool ID is shown on the top of each pie chart.

Supplementary Table S1. Primers amplifying the 16S rRNA gene of pan-Chlamydial bacteria.

Primer name	Sequence (5' to 3')	Amplicon size (bp)	Reference
CF1	CGTGGATGAGGCATGC(A/G)AGTCG	about 1500	(Corsaro *et al.*, 2002)
CR6	GTCATC(A/G)GCC(T/C)(T/C)ACCTT(A/C/G)(C/G)(A/G)C(A/G)(T/C)(T/C)TCT		
16SIGF	CGGCGTGGATGAGGCAT	about 1500	(Thomas *et al.*, 2006)
rP2Chlam	CTACCTTGTTACGACTTCAT		
16S FOR2	CGTGGATGAGGCATGCAAGTCGA	about 260	(Ossewaarde *et al.*, 1999)
16S REV2	CAATCTCTCAATCCGCCTAGACGTCTTAG		

695

References:

Corsaro D, Venditti D, Valassina M. (2002). New chlamydial lineages from freshwater samples. *Microbiology* **148**: 343-344.

Ossewaarde JM, Meijer A. (1999). Molecular evidence for the existence of additional members of the order *Chlamydiales*. *Microbiology* **145**: 411-417.

700 Thomas V, Casson N, Greub G. (2006). *Criblamydia sequanensis*, a new intracellular *Chlamydiales* isolated from Seine river water using amoebal co-culture. *Environ Microbiol* **8**: 2125-2135.

Supplementary Table S2. Bacterial/archaeal genera recovered from each tick pool.

	AVf		AVm		IDf		IDm		IPf		IPm		IRf		ATm		HFfem		HLfem	
1	*Borrelia*	(329)	*Clavibacter*	(1110)	*Rickettsia*	(3165)	*Ureaplasma*	(176)	*Chlamydophila*	(1642)	*Ureaplasma*	(186)	*Borrelia*	(199)	*Pseudoalteromonas*	(200)	*Chlamydophila*	(388)	*Lactobacillus*	(266)
2	*Mycoplasma*	(238)	*Pseudoalteromonas*	(421)	*Ureaplasma*	(2481)	*Rickettsia*	(116)	*Streptococcus*	(1569)	*Chlamydophila*	(132)	*Lactobacillus*	(199)	*Rickettsia*	(121)	*Streptococcus*	(250)	U (Gammaprotec	(191)
3	*Staphylococcus*	(155)	U (Euryarchaeota)	(389)	*Mycoplasma*	(1317)	*Mycoplasma*	(113)	*Helicobacter*	(1279)	*Streptococcus*	(123)	*Streptococcus*	(176)	*Synechococcus*	(25)	*Chlamydia*	(196)	*Salmonella*	(172)
4	*Actinobacillus*	(154)	*Treponema*	(301)	*Clostridium*	(940)	*Clostridium*	(50)	*Chlamydia*	(1104)	*Campylobacter*	(115)	*Ureaplasma*	(161)	*Wigglesworthia*	(25)	*Helicobacter*	(161)	*Grimontia*	(161)
5	*Cosiella*	(153)	*Synechococcus*	(265)	*Ehrlichia*	(639)	*Helicobacter*	(50)	*Campylobacter*	(1025)	*Helicobacter*	(107)	U (Gammaprotec	(152)	*Clostridium*	(22)	*Prochlorococcus*	(124)	*Providencia*	(137)
6	*Grimontia*	(143)	U (Gammaprotec	(260)	*Francisella*	(545)	*Desulfovibrio*	(49)	*Prochlorococcus*	(642)	*Prochlorococcus*	(94)	*Grimontia*	(145)	U (Alphaprotebo	(10)	*Campylobacter*	(101)	*Cosiella*	(127)
7	*Thiocystis*	(141)	*Arcanobacterium*	(236)	*Borrelia*	(466)	U (Gammaprotec	(49)	*Borrelia*	(548)	*Mycoplasma*	(82)	*Bacillus*	(141)	*Orientia*	(10)	*Bacillus*	(97)	*Cyanothece*	(115)
8	*Luteimonas*	(130)	*Actinobacillus*	(221)	U (Euryarchaeota)	(338)	*Streptococcus*	(43)	*Clostridium*	(497)	*Borrelia*	(75)	*Luteimonas*	(114)	*Bordetella*	(10)	*Clostridium*	(87)	*Streptococcus*	(112)
9	*Dichelobacter*	(105)	*Micromonospora*	(218)	*Campylobacter*	(324)	*Phytoplasma*	(42)	*Bacillus*	(493)	*Chlamydia*	(73)	*Vibrio*	(110)	*Bacillus*	(17)	*Thermosaurorbac*	(82)	*Staphylococcus*	(100)
10	*Phytoplasma*	(96)	*Salmonella*	(217)	*Staphylococcus*	(312)	*Francisella*	(38)	*Thermosaurorbac*	(490)	U (Euryarchaeota)	(68)	*Rickettsia*	(100)	U (Gammaprotec	(17)	*Borrelia*	(74)	*Bacillus*	(99)
11	*Campylobacter*	(94)	U (Actinobacteria)	(214)	*Orientia*	(301)	*Borrelia*	(37)	*Sulfurihydrogenib*	(441)	*Bacillus*	(58)	*Thiothrix*	(92)	*Amycolatopsis*	(15)	*Lactobacillus*	(66)	*Acinetobacter*	(92)
12	U (Tenericutes)	(94)	*Hydrogenovibrio*	(203)	U (Gammaprotec	(234)	*Thiothrix*	(35)	*Francisella*	(422)	*Clostridium*	(55)	*Actinobacillus*	(90)	*Micromonospora*	(15)	U (Euryarchaeota)	(65)	*Mycoplasma*	(92)
13	*Lactobacillus*	(91)	*Aeromonas*	(193)	*Streptococcus*	(232)	*Ehrlichia*	(30)	U (Euryarchaeota)	(372)	*Thermosaurorbac*	(48)	*Cosiella*	(89)	*Francisella*	(14)	*Sulfurihydrogenib*	(61)	*Pseudoxanthomon*	(89)
14	*Thiothrix*	(81)	U (Alphaprotebac	(158)	*Wolbachia*	(212)	*Staphylococcus*	(26)	C (Chlamydiae)	(370)	*Bartonella*	(40)	*Enterococcus*	(88)	*Streptococcus*	(14)	*Bartonella*	(44)	*Xenorhabdus*	(82)
15	*Salmonella*	(79)	*Streptococcus*	(137)	*Bacillus*	(201)	*Campylobacter*	(22)	*Wolbachia*	(353)	*Lactobacillus*	(40)	*Mycoplasma*	(88)	*Xanthomonas*	(14)	C (Chlamydiae)	(41)	*Actinobacillus*	(72)
16	U (Euryarchaeota)	(78)	*Hafnia*	(130)	*Buchnera*	(195)	*Lactobacillus*	(22)	*Bartonella*	(339)	U (Gammaprotec	(39)	*Colwellia*	(85)	*Aeromonas*	(12)	*Wolbachia*	(37)	*Rickettsia*	(70)
17	*Neisseria*	(77)	*Geobacter*	(116)	U (Gammaprotec	(181)	*Listeria*	(22)	*Sulfurimonas*	(217)	*Francisella*	(37)	*Francisella*	(70)	*Borrelia*	(12)	U (Firmicutes)	(36)	*Thiothrix*	(66)
18	U (Gammaprotec	(71)	*Rickettsiella*	(115)	*Phytoplasma*	(178)	C (Alphaprotebac	(21)	*Lactobacillus*	(216)	*Sulfurihydrogenib*	(35)	*Salmonella*	(64)	*Methylobacillus*	(12)	C (Bacteroidetes)	(35)	*Campylobacter*	(64)
19	*Providencia*	(70)	*Methylobacterium*	(107)	C (Bacteroidetes)	(168)	U (Euryarchaeota)	(21)	*Alkaliphilus*	(212)	C (Chlamydiae)	(33)	*Acinetobacter*	(63)	*Shigella*	(12)	*Cyanothece*	(35)	*Listeria*	(57)
20	*Vibrio*	(68)	*Actinosalinibacilus*	(101)	C (Tenericutes)	(167)	*Bacillus*	(20)	*Lamprocystis*	(198)	*Phytoplasma*	(31)	*Clostridium*	(63)	*Staphylococcus*	(12)	*Sulfurimonas*	(34)	*Nostoc*	(55)
21	*Bacillus*	(66)	*Burkholderia*	(96)	U (Tenericutes)	(156)	*Prochlorococcus*	(19)	U (Gammaprotec	(188)	*Rickettsia*	(27)	*Campylobacter*	(50)	*Arcanobacterium*	(11)	*Francisella*	(33)	*Lamprocystis*	(52)
22	*Escherichia*	(56)	*Citricoccus*	(96)	*Staphylothermus*	(140)	C (Gammaprotec	(18)	*Rickettsia*	(185)	C (Bacteroidetes)	(25)	*Staphylococcus*	(50)	*Clavibacter*	(11)	*Nostoc*	(33)	*Picrictkettsia*	(47)
23	*Cyanothece*	(52)	*Bacillus*	(87)	*Lactobacillus*	(144)	*Grimontia*	(18)	C (Bacteroidetes)	(181)	*Caldicellulosiruptu*	(25)	*Listeria*	(57)	*Polaromonas*	(11)	U (Gammaprotec	(33)	*Escherichia*	(44)
24	*Streptococcus*	(44)	*Saccharophagus*	(87)	C (Alphaprotebac	(122)	*Pseudovibrio*	(18)	*Mycoplasma*	(172)	*Nostoc*	(24)	*Staphylothermus*	(51)	*Burkholderia*	(10)	*Mycoplasma*	(32)	U (Firmicutes)	(43)
25	*Flavobacterium*	(43)	*Rhodofirax*	(84)	*Prochlorococcus*	(121)	*Pseudoalteromon*	(17)	U (Firmicutes)	(171)	C (Alphaprotebac	(22)	*Dichelobacter*	(46)	*Dechloromonas*	(10)	*Rickettsia*	(32)	*Helicobacter*	(42)
26	*Leptospira*	(39)	*Wigglesworthia*	(81)	*Fusobacterium*	(116)	*Buchnera*	(16)	*Caldicellulosiruptu*	(170)	*Lamprocystis*	(22)	*Oligobrachia*	(46)	*Delftia*	(10)	*Alkaliphilus*	(27)	*Pecinbacterium*	(42)
27	*Listeria*	(38)	*Ishikawaella*	(80)	U (Firmicutes)	(116)	*Saccharophagus*	(16)	*Flavobacterium*	(164)	*Wolbachia*	(22)	*Marinobacter*	(43)	*Streptomalleicha*	(10)	*Anaplasma*	(27)	*Borrelia*	(39)
28	*Ureaplasma*	(37)	*Neisseria*	(76)	*Vibrio*	(114)	U (Tenericutes)	(16)	*Nitratiruptor*	(131)	*Ehrlichia*	(21)	*Bartonella*	(41)	*Acidovorax*	(9)	U (Alphaprotebac	(27)	*Phytoplasma*	(39)
29	*Clostridium*	(35)	*Melittangium*	(75)	*Trichodesmium*	(91)	*Vibrio*	(16)	*Nostoc*	(145)	*Staphylococcus*	(21)	U (Alphaprotebac	(41)	*Escherichia*	(9)	*Flavobacterium*	(25)	*Rheinheimera*	(39)
30	*Grobacillus*	(34)	*Ralstonia*	(73)	*Bacteroides*	(78)	*Acinetobacter*	(14)	*Cyanothece*	(139)	*Alkaliphilus*	(20)	*Escherichia*	(38)	*Grimontia*	(9)	*Lamprocystis*	(25)	*Prochlorococcus*	(38)
31	C (Alphaprotebac	(32)	*Streptomalleicha*	(73)	*Anaplasma*	(72)	*Alcanivorax*	(13)	*Acinetobacter*	(137)	*Listeria*	(20)	*Thiocystis*	(37)	*Lactobacillus*	(9)	*Listeria*	(25)	*Trichodesmium*	(37)
32	U (Firmicutes)	(32)	*Lactobacillus*	(71)	*Escherichia*	(70)	*Salmonella*	(13)	*Listeria*	(132)	U (Bacteroidetes)	(20)	*Lactococcus*	(36)	*Salmonella*	(9)	*Staphylothermus*	(25)	*Colwellia*	(33)
33	*Enterococcus*	(31)	*Grimontia*	(69)	*Spiroplasma*	(67)	*Fusobacterium*	(12)	U (Alphaprotebac	(132)	*Leptospira*	(17)	*Neisseria*	(34)	U (Betaproteobac	(9)	*Acinetobacter*	(23)	*Flavobacterium*	(32)
34	*Buchnera*	(27)	*Xanthomonas*	(68)	*Salmonella*	(59)	*Trichodesmium*	(12)	*Enterococcus*	(129)	*Salmonella*	(16)	*Orientia*	(34)	*Deinococcus*	(8)	*Leptospira*	(23)	*Plesiomonas*	(32)
35	*Fusobacterium*	(27)	*Xenorhabdus*	(68)	*Arcobacter*	(54)	*Xenorhabdus*	(12)	*Fusobacterium*	(116)	*Staphylothermus*	(16)	*Flavobacterium*	(32)	*Listeria*	(8)	*Staphylococcus*	(21)	C (Chlamydiae)	(31)
36	*Xenorhabdus*	(27)	*Zymobacter*	(56)	*Grimontia*	(53)	C (Chlamydiae)	(11)	U (Epsilonproteob	(116)	*Acinetobacter*	(15)	*Helicobacter*	(32)	*Ralstonia*	(8)	*Caldicellulosiruptu*	(20)	*Enterococcus*	(29)
37	*Acinetobacter*	(24)	*Escherichia*	(55)	*Thiothrix*	(51)	*Spiroplasma*	(11)	*Anaplasma*	(108)	*Dictyoglomus*	(15)	C (Chlamydiae)	(31)	U (Euryarchaeota)	(8)	*Enterococcus*	(20)	*Ureaplasma*	(28)
38	*Oligobrachia*	(24)	*Serratia*	(54)	*Nostoc*	(50)	*Staphylothermus*	(11)	*Salmonella*	(107)	*Sulfurimonas*	(15)	*Phytoplasma*	(31)	C (Bacteroidetes)	(7)	*Salmonella*	(20)	*Clostridium*	(27)
39	*Xanthomonas*	(22)	*Nicoletella*	(53)	*Flavobacterium*	(49)	U (Firmicutes)	(11)	C (Alphaprotebac	(103)	C (Gammaprotec	(14)	*Buchnera*	(28)	*Comamonas*	(7)	*Fusobacterium*	(19)	*Alteromonas*	(24)
40	U (Epsilonproteob	(23)	C (Firmicutes)	(51)	*Sulfurihydrogenib*	(48)	C (Bacteroidetes)	(10)	*Hydrogenobacula*	(100)	*Cyanothece*	(14)	U (Firmicutes)	(27)	*Geobacter*	(7)	*Hydrogenobacula*	(19)	*Anabaena*	(24)
41	*Thermosinus*	(20)	*Deinococcus*	(49)	*Thermosaurorbac*	(47)	*Escherichia*	(10)	*Anaerococcus*	(99)	*Grimontia*	(14)	*Xenorhabdus*	(21)	*Hafnia*	(7)	U (Epsilonproteob	(18)	*Francisella*	(24)
42	C (Gammaprotec	(18)	*Desulfovibrio*	(48)	U (Alphaprotebac	(47)	*Nostoc*	(10)	*Staphylothermus*	(97)	U (Alphaprotebac	(14)	C (Gammaprotec	(20)	*Providencia*	(7)	*Desulfotomaculum*	(17)	*Luteimonas*	(23)
43	*Helicobacter*	(18)	*Anaplasma*	(43)	*Pseudoalteromon*	(45)	*Providencia*	(10)	U (Bacteroidetes)	(97)	U (Firmicutes)	(14)	*Rheinheimera*	(19)	*Staphylothermus*	(7)	*Lysinibacillus*	(16)	*Pediococcus*	(23)
44	*Pseudomonas*	(18)	*Acidithiobacillus*	(41)	*Treponema*	(45)	*Bartonella*	(9)	*Lysinibacillus*	(95)	*Cosiella*	(13)	*Sulfurihydrogenib*	(19)	*Thiothrix*	(7)	*Grimontia*	(13)	*Saccharophagus*	(23)
45	*Trichodesmium*	(18)	*Alteromonas*	(40)	*Listeria*	(43)	*Cyanothece*	(9)	*Polaribacter*	(95)	*Enterococcus*	(13)	*Grobacillus*	(18)	U (Synergistetes)	(7)	U (Bacteroidetes)	(13)	*Rickettsiella*	(22)
46	C (Tenericutes)	(17)	*Congregibacter*	(40)	*Cyanothece*	(41)	*Enterococcus*	(9)	*Staphylococcus*	(95)	*Flavobacterium*	(13)	*Hydrogenobacula*	(18)	*Xenorhabdus*	(7)	U (Tenericutes)	(13)	U (Euryarchaeota)	(22)
47	*Pseudoalteromon*	(17)	U (Firmicutes)	(40)	*Acinetobacter*	(40)	*Orientia*	(9)	*Ferroplasma*	(93)	*Providencia*	(13)	*Lamprocystis*	(18)	*Acinetobacter*	(6)	C (Alphaprotebac	(12)	*Pseudomonas*	(21)
48	*Thermosipho*	(17)	*Alcanivorax*	(36)	*Xenorhabdus*	(40)	U (Alphaprotebac	(9)	*Dictyoglomus*	(84)	U (Epsilonproteob	(13)	U (Euryarchaeota)	(18)	*Brucella*	(6)	*Marinobacter*	(12)	U (Tenericutes)	(21)
49	U (Betaproteobac	(17)	*Bacteroides*	(36)	*Helicobacter*	(37)	*Ishikawaella*	(8)	*Orientia*	(83)	*Vibrio*	(13)	*Ehrlichia*	(17)	*Campylobacter*	(6)	*Polaribacter*	(12)	U (Alphaprotebac	(20)
50	*Ornocaccus*	(15)	*Borrelia*	(32)	*Pseudomonas*	(37)	*Lamprocystis*	(8)	*Dokdonia*	(80)	*Caldivirga*	(12)	*Ornocaccus*	(17)	*Ehrlichia*	(6)	*Anaerococcus*	(11)	*Marinobacter*	(19)
51	*Rhodobacteriales*	(15)	*Thermobifida*	(31)	*Anaerofustis*	(34)	*Nitratiruptor*	(8)	*Phytoplasma*	(79)	*Dokdonia*	(12)	*Thermosaurorbac*	(17)	*Prochlorococcus*	(6)	*Nitratiruptor*	(11)	*Moraxella*	(19)
52	*Rickettsia*	(14)	*Olavius*	(30)	U (Bacteroidetes)	(34)	*Anaplasma*	(6)	*Thiothrix*	(75)	*Fusobacterium*	(12)	*Microscilla*	(16)	*Saccharophagus*	(6)	*Anabaena*	(10)	*Alkilibvio*	(17)
53	*Ehrlichia*	(13)	*Brucella*	(27)	*Alkaliphilus*	(31)	*Bacteroides*	(6)	*Leptospira*	(74)	*Hydrogenobacula*	(12)	*Pseudoxanthomon*	(16)	*Treponema*	(6)	*Bacteroides*	(10)	*Ornocaccus*	(17)
54	*Prochlorococcus*	(13)	Glucanacetobacter	(27)	*Lamprocystis*	(31)	*Clavibacter*	(6)	*Trichodesmium*	(67)	*Trichodesmium*	(12)	*Alteromonas*	(15)	*Desulfovibrio*	(5)	*Cosiella*	(10)	C (Gammaprotec	(16)
55	*Rickettsiella*	(13)	*Rosenbergiella*	(26)	*Chlamydophila*	(30)	*Leptospira*	(6)	U (Chlamydiae)	(64)	U (Tenericutes)	(11)	*Carboxydothermus*	(15)	*Mycoplasma*	(5)	*Crocinbacter*	(10)	*Oligobrachia*	(15)
56	*Nostoc*	(12)	*Vibrio*	(26)	*Bartonella*	(28)	*Pseudomonas*	(6)	U (Tenericutes)	(62)	*Anaerococcus*	(10)	*Chloroherpeton*	(15)	*Rhodothermus*	(5)	*Microcystis*	(10)	*Bartonella*	(14)
57	*Rheinheimera*	(12)	*Paracoccus*	(25)	*Pyrodictium*	(28)	U (Epsilonproteob	(6)	*Bacteroides*	(61)	*Crocinbacter*	(10)	*Rickettsiella*	(14)	U (Actinobacteria	(5)	*Orientia*	(10)	*Lysinibacillus*	(14)
58	U (Bacteroidetes)	(12)	*Chlorobium*	(24)	*Lysinibacillus*	(25)	*Wolbachia*	(6)	*Grimontia*	(60)	*Ferroplasma*	(10)	*Aggregatibacter*	(13)	U (Verrucomicro	(5)	*Pseudomonas*	(10)	*Microcystis*	(14)
59	*Lamprocystis*	(11)	*Chloroflexus*	(24)	U (Epsilonproteob	(25)	*Xanthomonas*	(6)	*Psychroflexus*	(60)	*Polaribacter*	(10)	*Alkaliphilus*	(13)	*Verminephrobacter*	(5)	*Phytoplasma*	(9)	*Pelagibacterr*	(14)
60	*Polynucleobacter*	(11)	*Cosiella*	(24)	*Actinobacillus*	(24)	*Alteromonas*	(5)	*Anabaena*	(55)	*Lysinibacillus*	(9)	*Providencia*	(12)	*Bacteroides*	(4)	*Escherichia*	(7)	*Thiocystis*	(14)
61	*Pseudoxanthomon*	(11)	*Mesorhizobium*	(24)	*Caldicellulosiruptu*	(23)	*Pecinbacterium*	(5)	*Escherichia*	(50)	*Pseudoxanthomon*	(9)	*Gramella*	(11)	*Erythrobacter*	(4)	*Providencia*	(7)	U (Bacteroidetes)	(14)
62	*Synechococcus*	(11)	*Syntrophobacter*	(23)	*Clavibacter*	(22)	C (Tenericutes)	(4)	*Algoriphagus*	(48)	*Buchnera*	(8)	*Methylomicrobium*	(11)	*Grobacillus*	(4)	*Thiothrix*	(7)	C (Alphaprotebac	(13)
63	*Saccharophagus*	(10)	*Variibaculum*	(23)	*Polaribacter*	(22)	*Filobacterium*	(4)	*Ehrlichia*	(47)	*Orientia*	(8)	*Bacteroides*	(10)	*Halomonas*	(4)	U (Thermotogae)	(7)	C (Bacteroidetes)	(13)
64	U (Alphaprotebac	(10)	*Cellvibrio*	(22)	*Hydrogenobacula*	(21)	*Lactococcus*	(4)	*Lactococcus*	(44)	*Xenorhabdus*	(8)	*Fusobacterium*	(10)	*Ishikawaella*	(4)	*Anaerocellum*	(6)	*Carboxydothermus*	(13)
65	*Clavibacter*	(9)	*Polaromonas*	(22)	*Mesoplasma*	(21)	U (Deltaproteoba	(4)	*Marinobacter*	(44)	*Anaplasma*	(7)	*Prochlorococcus*	(10)	*Lamprocystis*	(4)	*Gramella*	(6)	*Leptospira*	(12)
66	*Epulopiscium*	(9)	*Amycolatopsis*	(21)	*Baumannia*	(20)	*Wolinella*	(4)	*Nevskiella*	(43)	*Bacteroides*	(7)	C (Tenericutes)	(9)	*Leptospira*	(4)	*Nevskiella*	(6)	*Leuconostoc*	(12)
67	*Bacteroides*	(8)	*Luteimonas*	(20)	*Saccharophagus*	(20)	*Actinobacillus*	(3)	*Oscillatoria*	(43)	*Crocosphaera*	(7)	*Capnocytophaga*	(9)	*Pecinbacterium*	(4)	*Nitrosopumilus*	(6)	*Bacteroides*	(11)
68	*Bartonella*	(8)	*Erysipelothrix*	(19)	U (Spirochaetes)	(20)	*Anabaena*	(3)	*Eubacterium*	(40)	*Lactococcus*	(7)	*Pseudoalteromon*	(9)	*Promicromonospo*	(4)	*Pseudoxanthomon*	(6)	*Nicoletella*	(11)
69	*Chlorobium*	(8)	*Plesiocystis*	(19)	*Caminibacter*	(19)	*Arcanobacterium*	(3)	*Capnocytophaga*	(38)	*Mesoplasma*	(7)	*Pseudomonas*	(9)	*Rhizobium*	(4)	*Psychroflexus*	(6)	*Grobacillus*	(10)
70	*Exiguobacterium*	(8)	*Grobacillus*	(18)	*Capnocytophaga*	(19)	*Baumannia*	(3)	*Gramella*	(38)	*Microcystis*	(7)	U (Tenericutes)	(9)	*Rhodofirax*	(4)	*Rickettsiella*	(6)	U (Cyanobacteria	(10)
71	*Pedinococcus*	(8)	Opitutaceae	(18)	*Lawsonia*	(19)	*Carboxydothermus*	(3)	*Cosiella*	(34)	*Pseudomonas*	(7)	C (Alphaprotebac	(8)	*Roseiflexus*	(4)	*Trichodesmium*	(6)	*Psychroflexus*	(9)
72	*Wolbachia*	(8)	Opitutus	(18)	*Synechococcus*	(18)	*Cosiella*	(3)	*Crocosphaera*	(33)	*Psychroflexus*	(7)	C (Bacteroidetes)	(8)	*Sorangium*	(4)	U (Cyanobacteria	(6)	*Polaribacter*	(8)
73	*Beggiatoa*	(7)	U (Chlorobi)	(18)	*Grobacillus*	(17)	*Dokdonia*	(3)	*Microscilla*	(33)	*Anabaena*	(6)	*Plesiomonas*	(8)	U (Firmicutes)	(4)	*Capnocytophaga*	(5)	*Wolbachia*	(8)
74	*Moraxella*	(7)	*Enterobacter*	(17)	*Leptospira*	(17)	*Grobacillus*	(3)	*Sulfurovum*	(33)	*Gramella*	(6)	*Nostoc*	(7)	*Verrucomicrobiae*	(4)	*Crocosphaera*	(5)	*Anaplasma*	(7)
75	*Plesiomonas*	(7)	*Pelodictyon*	(17)	*Pedinococcus*	(16)	*Mesoplasma*	(3)	*Crocinbacter*	(34)	*Nitratiruptor*	(6)	*Pedinococcus*	(7)	*Wolbachia*	(4)	*Dokdonia*	(5)	*Caldivirga*	(7)
76	*Rhodopirellula*	(7)	*Desulfobacterium*	(16)	*Thermotoga*	(16)	*Microcystis*	(3)	*Pseudomonas*	(33)	*Saccharophagus*	(6)	*Beggiatoa*	(6)	*Acidithiobacillus*	(3)	*Eubacterium*	(5)	*Desulfotomaculum*	(7)
77	*Arcanobacterium*	(6)	*Lyngbya*	(16)	*Piscirickettsia*	(15)	*Oceanobacillus*	(3)	*Pseudoxanthomon*	(33)	*Thiothrix*	(6)	*Chlorobium*	(6)	*Actinobacillus*	(3)	*Natronarrobius*	(5)	*Dichelobacter*	(7)
78	*Chloroherpeton*	(6)	*Parabacteroides*	(16)	*Parabacteroides*	(14)	*Rickettsiella*	(3)	*Porphyromonas*	(31)	C (Tenericutes)	(5)	*Leptospira*	(6)	*Anaplasma*	(3)	*Ferroplasma*	(5)	*Xanthomonas*	(7)
79	*Francisella*	(6)	*Providencia*	(16)	*Pecinbacterium*	(14)	*Synechococcus*	(3)	U (Cyanobacteria)	(30)	*Eubacterium*	(5)	*Pecinbacterium*	(6)	*Baumannia*	(3)	*Synechococcus*	(5)	*Alkaliphilus*	(6)
80	*Orientia*	(6)	*Clostridium*	(15)	*Colwellia*	(13)	*Syntrophomonas*	(3)	*Treponema*	(28)	*Grobacillus*	(5)	*Spiroplasma*	(6)	*Cellvibrio*	(3)	*Zymomonas*	(5)	*Beggiatoa*	(6)
81	*Pelagibacterr*	(6)	*Dehalococcoides*	(15)	*Natronarrobius*	(13)	*Thermosaurorbac*	(3)	*Microcystis*	(25)	*Rheinheimera*	(5)	U (Spirochaetes)	(6)	*Chloroflexus*	(3)	*Algoriphagus*	(4)	*Carnobacterium*	(6)
82	*Aggregatibacter*	(5)	*Ehrlichia*	(15)	*Porphyromonas*	(13)	*Thermotoga*	(3)	*Providencia*	(25)	*Treponema*	(5)	*Xanthomonas*	(6)	*Cosiella*	(3)	*Carnobacterium*	(4)	*Gramella*	(6)
83	*Burkholderia*	(5)	*Polaromaculum*	(15)	*Crocinbacter*	(12)	*Treponema*	(3)	U (Thermotogae)	(25)	U (Betaproteobac	(5)	*Lyngbya*	(5)	*Cytophaga*	(3)	*Clavibacter*	(4)	*Microscilla*	(6)
84	*Lysinibacillus*	(5)	*Thiocystis*	(15)	*Dokdonia*	(12)	*Alkaliphilus*	(2)	*Thermodesulfovib*	(24)	U (Chlamydiae)	(5)	U (Cyanobacteria	(5)	*Granulibacter*	(3)	*Grobacillus*	(4)	*Orientia*	(6)
85	*Microcystis*	(5)	*Listeria*	(14)	*Epulopiscium*	(12)	C (Acidobacteria	(2)	*Synechococcus*	(22)	U (Spirochaetes)	(5)	U (Epsilonproteob	(5)	*Helicobacter*	(3)	*Microscilla*	(4)	*Polynucleobacter*	(6)
86	*Thermosaurorbac*	(5)	*Nostoc*	(14)	*Xanthomonas*	(12)	*Capnocytophaga*	(2)	*Anaerocellum*	(21)	*Algoriphagus*	(4)	*Wolbachia*	(5)	*Herminiimonas*	(3)	U (Chlamydiae)	(4)	*Thermosaurorbac*	(6)
87	*Agrobacterium*	(4)	*Pseudonocardia*	(13)	*Burkholderia*	(11)	*Colwellia*	(2)	*Carboxydothermus*	(21)	*Carboxydothermus*	(4)	*Alkilibvio*	(4)	*Janthinobacterium*	(3)	*Ureaplasma*	(4)	U (Betaproteobac	(6)
88	*Alkilibvio*	(4)	*Sorangium*	(13)	*Dictyoglomus*	(11)	*Cytophaga*	(2)	*Nitrosopumilus*	(20)	*Desulfotomaculum*	(4)	*Caldicellulosiruptu*	(4)	*Limnobacter*	(3)	*Delftia*	(3)	*Vibrio*	(6)
89	*Alteromonas*	(4)	U (Deltaproteoba	(13)	*Sulfurimonas*	(11)	*Dichelobacter*	(2)	*Pyrodictium*	(20)	*Escherichia*	(4)	*Cytophaga*	(4)	*Mycobacterium*	(3)	*Desulfotomaculum*	(3)	*Fusobacterium*	(5)
90	*Anaerococcus*	(4)	*Paenibacillus*	(12)	*Enterococcus*	(10)	*Grobacter*	(2)	*Oceanobacillus*	(19)	*Marinobacter*	(4)	*Lentisphaera*	(4)	*Nitrobacter*	(3)	*Epulopiscium*	(3)	*Ishikawaella*	(5)
91	*Anoxybacillus*	(4)	*Thiothrix*	(12)	U (Cyanobacteria	(10)	*Hafnia*	(2)	*Carnobacterium*	(18)	*Oligobrachia*	(4)	*Leuconostoc*	(4)	*Nostoc*	(3)	*Lactococcus*	(3)	*Staphylothermus*	(5)
92	*Chloroflexus*	(4)	*Mycoplasma*	(11)	*Anaerococcus*	(9)	*Luteimonas*	(2)	*Alkilibvio*	(17)	*Plesiomonas*	(4)	*Microcystis*	(4)	*Porphyromonas*	(3)	*Nicoletella*	(3)	*Zymomonas*	(5)
93	*Hydrogenovibrio*	(4)	*Sinorhizobium*	(11)	*Carboxydothermus*	(9)	*Marinobacter*	(2)	*Natronarrobius*	(17)	*Pseudoalteromon*	(4)	*Nitrococcus*	(4)	*Pseudoxanthomon*	(3)	*Oceanobacillus*	(3)	*Acholeplasma*	(4)
94	*Ishikawaella*	(4)	U (Verrucomicro	(11)	*Eubacterium*	(9)	*Mesorhizobium*	(2)	*Clavibacter*	(16)	*Rhodopirellula*	(4)	*Piscirickettsia*	(4)	*Roseobacter*	(3)	*Pseudoalteromon*	(3)	*Capnocytophaga*	(4)
95	*Nitrosospira*	(4)	*Caulobacter*	(10)	*Lactococcus*	(9)	*Methylocidiphilum*	(2)	*Epulopiscium*	(15)	*Spiroplasma*	(4)	*Thermotoga*	(4)	*Agrobacterium*	(2)	U (Nanoarchaeota	(3)	*Cellvibrio*	(4)
96	*Paenibacillus*	(4)	*Mannheimia*	(10)	*Polaromonas*	(9)	*Methylobacillus*	(2)	*Luteimonas*	(15)	*Thermosipho*	(4)	U (Fibrobacteres)	(4)	*Alcanivorax*	(2)	*Anaerofustis*	(2)	*Chlamydophila*	(4)
97	*Pecinbacterium*	(4)	*Nitrosospira*	(10)	U (Deltaproteoba	(9)	*Natronarrobius*	(2)	*Ruminococcus*	(15)	U (Cyanobacteria	(4)	*Aeromonas*	(3)	*Alkilibvio*	(2)	*Aquifex*	(2)	*Clavibacter*	(4)
98	*Pyrodictium*	(4)	*Bdellovibrio*	(9)	*Wigglesworthia*	(9)	*Octadecabacter*	(2)	*Thermosipho*	(15)	*Zymomonas*	(4)	*Anoxybacillus*	(3)	*Alkalilimnicola*	(2)	*Arcobacter*	(2)	*Cytophaga*	(4)
99	U (Actinobacteria	(4)	*Porphyromonas*	(9)	C (Chlamydiae)	(8)	*Ornocaccus*	(2)	*Polynucleobacter*	(14)	*Acaryochloris*	(3)	*Arcobacter*	(3)	*Alkaliphilus*	(2)	*Carboxydothermus*	(2)	*Lactococcus*	(4)

100	*Zymomonas*	(4)	*Promicromonospo*	(9)	*Oceanobacillus*	(8)	*Pirias monas*	(2)	*Rheinheimera*	(14)	*Alitvibrio*	(3)	*Burkholderia*	(3)	*Anabaena*	(2)	*Cytophaga*	(2)	*Lyngbya*	(4)	
101	*Alistipes*	(3)	*Staphylococcus*	(9)	*Providencia*	(8)	*Pseudoxanthomon*	(2)	*Actinobacillus*	(13)	*Aquifex*	(3)	*Cellvibrio*	(3)	*Anaerofustis*	(2)	*Ehrlichia*	(2)	*Microcolrus*	(4)	
102	*Alkaliphilus*	(3)	*Bifidobacterium*	(8)	*Sulfuriphaera*	(8)	*Psychroflexus*	(2)	*C.*(Gammaproteo	(13)	*Beijerinckia*	(3)	*Kordia*	(3)	*Anaeromyxobacte*	(2)	*Lawsonia*	(2)	*Neisseria*	(4)	
103	*Caldicellulosirupto*	(3)	*Actinobacter*	(7)	*Sulfurovum*	(8)	*Rhizobium*	(2)	*Streptobolbirichus*	(13)	*Hydrogenivirga*	(3)	*Lysinibacillus*	(3)	*Beijerinckia*	(2)	*Lyngbya*	(2)	*Sulfurihydrogenib*	(4)	
104	*Caminibacter*	(3)	*Anaerotruncus*	(7)	*Thermobifida*	(8)	*Ruminococcus*	(2)	*Ureaplasma*	(13)	*Leptospirillum*	(3)	*Pelagibacter*	(3)	*Bradyrhizobium*	(2)	*Methylacidiphilum*	(2)	*Synechococcus*	(4)	
105	*Citricoccus*	(3)	*Desulfococcus*	(7)	U (Verrucomicro	(8)	*Serratia*	(2)	*Acaryochloris*	(12)	*Oceanobacillus*	(3)	*Treponema*	(3)	C (Chlamydiae)	(2)	*Ornocaccus*	(2)	U (Elusimicrobia)	(4)	
106	*Eubacterium*	(3)	*Helicobacter*	(7)	*Acholeplasma*	(7)	*Sulfurihydrogenib*	(2)	*Desulfotomaculum*	(12)	*Piscirickettsia*	(3)	U (Bacteroidetes	(3)	C (Firmicutes)	(2)	*Oligobrachia*	(2)	*Anoxybacillus*	(3)	
107	*Lactococcus*	(3)	*Pasteurella*	(7)	*Anabaena*	(7)	*Sulfurimonas*	(2)	*Fervidobacterium*	(12)	*Pyrolitium*	(3)	*Zymomonas*	(3)	*Caldivirga*	(2)	*Pectobacterium*	(2)	*Anthrospira*	(3)	
108	*Mesoplasma*	(3)	*Roseobacter*	(7)	*Chlamydia*	(7)	*Symbiobacterium*	(2)	*Geobacillus*	(12)	*Rickettsiella*	(3)	*Acholeplasma*	(2)	*Carboxydothermus*	(2)	*Pedobacter*	(2)	*Buchnera*	(3)	
109	*Nitrosopumilus*	(3)	*Thermosinus*	(7)	*Cosiella*	(7)	U (Acidobacteria	(2)	*Methylacidiphilum*	(12)	U (Actinobacteria	(5)	*Agrobacterium*	(2)	*Chlamydia*	(2)	*Pelotomaculum*	(2)	*Chlamydia*	(3)	
110	*Petrotoga*	(3)	U (Deinococcus-T	(7)	*Desulfitobacterium*	(7)	U (Bacteroidetes)	(2)	U (Beijerinckiace	(12)	U (Thermotogae)	(3)	*Carnobacterium*	(2)	*Caprothermobacte*	(2)	*Riemerella*	(2)	*Crocosphaera*	(3)	
111	*Polaribacter*	(3)	U (Synergistetes)	(7)	*Pseudoxanthomon*	(7)	*Acholeplasma*	(1)	*Acholeplasma*	(11)	*Xanthomonas*	(3)	*Clavibacter*	(2)	*Crocebacter*	(2)	*Sulfurovum*	(2)	*Epulopiscium*	(3)	
112	*Rahnnia*	(3)	U (Thermotogae)	(7)	*Psychroflexus*	(7)	*Acholeplasma*	(1)	*Alteromonas*	(11)	*Anaerocellum*	(2)	*Exiguobacterium*	(2)	*Cyanothece*	(2)	*Thermobifida*	(2)	*Geobacter*	(3)	
113	*Treponema*	(3)	Ureaplasma	(7)	U (Actinobacteria	(7)	*Aeromonas*	(1)	*Colwellia*	(11)	*Chlorobium*	(2)	*Hydrogenovibrio*	(2)	*Desulfuromonas*	(2)	*Thermosipho*	(2)	*Nitrococcus*	(3)	
114	U (Fibrobacteres)	(3)	*Agrobacterium*	(6)	*Acaryochloris*	(6)	*Agrobacterium*	(1)	*Desulfotalea*	(11)	*Desulfitobacterium*	(2)	*Nitrosospira*	(2)	*Dinoroseobacter*	(2)	*Thiocystis*	(2)	*Pedobacter*	(3)	
115	*Anaplasma*	(2)	*Buchnera*	(6)	*Alcanivorax*	(6)	*Anaerofustis*	(1)	*Finegoldia*	(11)	*Epulopiscium*	(2)	*Polymorphobacter*	(2)	*Herpetosiphon*	(2)	*Treponema*	(2)	*Aggregatibacter*	(2)	
116	*Arcobacter*	(2)	*Limnobacter*	(6)	*Arcanobacterium*	(6)	*Arcobacter*	(1)	*Ornocaccus*	(11)	*Fervidobacterium*	(2)	*Saccharophagus*	(2)	*Hydrogenovibrio*	(2)	U (Betaproteobac	(2)	*Anaerococcus*	(2)	
117	C.(Bacteroidetes)	(2)	*Methylacidiphilum*	(6)	*Cytophaga*	(6)	*Ausspirillum*	(1)	*Anaerofustis*	(10)	*Geobacter*	(2)	*Synechocystis*	(2)	*Luteimonas*	(2)	*Xenorhabdus*	(2)	*Baumannia*	(2)	
118	*Cellvibrio*	(2)	*Moritella*	(6)	*Geobacter*	(6)	*Brucella*	(1)	*Pediococcus*	(10)	*Lyngbya*	(2)	*Trichodesmium*	(2)	*Lyngbya*	(2)	*Acaryochloris*	(1)	C (Tenericutes)	(2)	
119	*Colwellia*	(2)	*Nodularia*	(6)	*Ishikawaella*	(6)	*Burkholderia*	(1)	*Saccharophagus*	(10)	*Neorickettsia*	(2)	U (Elusimicrobia)	(2)	*Lysinibacillus*	(2)	*Acholeplasma*	(1)	*Clover*	(2)	
120	*Crocebacter*	(2)	*Roseovarius*	(6)	*Nitratiruptor*	(6)	*Caldicellulosirupto*	(1)	U (Actinobacteria	(10)	*Pediococcus*	(2)	*Wigglesworthia*	(2)	*Magnetospirillum*	(2)	*Aeromonas*	(1)	*Crocebacter*	(2)	
121	*Desulfitobacterium*	(2)	*Solibacter*	(6)	*Parvimonas*	(6)	*Carboxydibrachiu*	(1)	U (Elusimicrobia)	(10)	*Ruminococcus*	(2)	*Algoriphagus*	(1)	*Mannheimia*	(2)	*Beggiatoa*	(1)	*Desulfitobacterium*	(2)	
122	*Geobacter*	(2)	*Zymomonas*	(6)	*Pelagibacter*	(6)	*Carnobacterium*	(1)	*Thermotoga*	(9)	*Stigmatella*	(2)	*Anabaena*	(1)	*Maricaulis*	(2)	*Burkholderia*	(1)	*Desulfotalea*	(2)	
123	*Herpetosiphon*	(2)	*Alicycliobacillus*	(5)	*Petrotoga*	(6)	*Cellvibrio*	(1)	U (Nanoarchaeot	(9)	*Sulfurovum*	(2)	*Anaerotruncus*	(1)	*Melittangium*	(2)	C (Gammaproteo	(1)	*Desulfuromonas*	(2)	
124	*Janthinobacterium*	(2)	*Azoarcus*	(5)	*Thermosipho*	(6)	*Chlamydophila*	(1)	*Veillonella*	(9)	*Synechococcus*	(2)	*Anaplasma*	(1)	*Methylibium*	(2)	*Caminibacter*	(1)	*Diokdonia*	(2)	
125	*Kordia*	(2)	*Cyanothece*	(5)	U (Thermotogae)	(6)	*Caprothermobacte*	(1)	*Xenorhabdus*	(10)	U (Fusobacteria)	(2)	*Arcanobacterium*	(1)	*Methylobacter*	(2)	*Cellulosimicrobiu*	(1)	*Finegoldia*	(2)	
126	*Martinobacter*	(2)	*Exiguobacterium*	(5)	*Citrobacter*	(5)	*Deinococcus*	(1)	*Buchnera*	(8)	U (Nanoarchaeot	(2)	*Anthrospira*	(1)	*Methylobacterium*	(2)	*Chlorobium*	(1)	*Hydrogenovibrio*	(2)	
127	*Micromonospora*	(2)	*Leuconostoc*	(5)	*Coprococcus*	(5)	*Desulfotomaculum*	(1)	U (Deltaproteoba	(8)	*Acetobacter*	(1)	*Bdellovibrio*	(1)	*Nicoletella*	(2)	*Coprococcus*	(1)	*Methylophaga*	(2)	
128	*Natranaerobius*	(2)	*Oligobrachia*	(5)	*Desulfotomaculum*	(5)	*Faecalibacterium*	(1)	U (Spirochaetes)	(8)	*Acholeplasma*	(1)	*Brucella*	(1)	*Ochrobactrum*	(2)	*Desulfitobacterium*	(1)	*Nitrosomonas*	(2)	
129	*Octadecabacter*	(2)	*Stigmatella*	(5)	*Gramella*	(5)	*Finegoldia*	(1)	*Wolinella*	(8)	*Actinoallotichus*	(1)	*Chlorobaculum*	(1)	*Octadecabacter*	(2)	*Desulfotalea*	(1)	*Nodularia*	(2)	
130	*Parabacteroides*	(2)	*Virgisporangium*	(5)	*Marinobacter*	(5)	*Glaconostrebacte*	(1)	U (Elusimicrobia)	(10)	*Aeromonas*	(1)	*Cyanothece*	(1)	*Olavius*	(2)	*Faecalibacterium*	(1)	*Oceanobacillus*	(2)	
131	*Pasteurella*	(2)	*Azorhizobium*	(4)	*Nitrosopumilus*	(5)	*Glaconobacter*	(1)	*Aquifex*	(7)	*Alcanivorax*	(1)	*Desulfitobacterium*	(1)	*Paenibacillus*	(2)	*Fervidobacterium*	(1)	*Petrotoga*	(2)	
132	*Pedobacter*	(2)	*Cellulosimicrobiu*	(4)	*Alitvibrio*	(4)	*Granulibacter*	(1)	*Caminibacter*	(7)	*Alteromonas*	(1)	*Desulfotomaculum*	(1)	*Pasteurella*	(2)	*Fischerella*	(1)	*Thermosipho*	(2)	
133	*Prbdictyon*	(2)	*Dichelobacter*	(4)	*Anaerocellum*	(4)	*Heliobacterium*	(1)	*Chlorobium*	(7)	*Anaerofustis*	(1)	*Diokdonia*	(1)	*Pedobacter*	(2)	*Hafnia*	(1)	*Acaryochloris*	(1)	
134	*Piscirickettsia*	(2)	*Enterococcus*	(4)	*Anaerotruncus*	(4)	*Hydrogenovibrio*	(1)	*Coprococcus*	(7)	*Arcanobacterium*	(1)	*Epulopiscium*	(1)	*Pseudomonas*	(2)	*Halomonas*	(1)	*Acidiphilium*	(1)	
135	*Porphyromonas*	(2)	*Fusobacterium*	(4)	*Carnobacterium*	(4)	*Hyphomonas*	(1)	*Cylindrospermops*	(7)	*Anthrospira*	(1)	*Geobacter*	(1)	*Rhodobacter*	(2)	*Kordia*	(1)	*Aeromonas*	(1)	
136	*Psychroflexus*	(2)	*Glaconobacter*	(4)	*Fervidobacterium*	(4)	*Kingella*	(1)	*Ishikawaella*	(7)	*Bordetella*	(1)	*Mesoplasma*	(1)	*Rhodospirillula*	(2)	*Lentisphaera*	(1)	*Alcanivorax*	(1)	
137	*Thermobifida*	(2)	*Microcystis*	(4)	*Finegoldia*	(4)	*Leuconostoc*	(1)	*Kordia*	(7)	*Brachyspira*	(1)	*Metallosphaera*	(1)	*Rickettsiella*	(2)	*Mannheimia*	(1)	*Algoriphagus*	(1)	
138	U (Cyanobacteria	(2)	*Octadecabacter*	(4)	*Halothermothrix*	(4)	*Liberobacter*	(1)	*Lyngbya*	(7)	*Burkholderia*	(1)	*Microcolrus*	(1)	*Salinivibrio*	(2)	*Mesoplasma*	(1)	*Alistipes*	(1)	
139	U (Elusimicrobia)	(2)	*Parvibaculum*	(4)	*Metallosphaera*	(4)	*Lysinibacillus*	(1)	*Nicoletella*	(7)	C (Firmicutes)	(1)	*Nitrosomonas*	(1)	*Sphingopyxis*	(2)	*Metallosphaera*	(1)	*Anaerocellum*	(1)	
140	*Wigglesworthia*	(2)	*Phytoplasma*	(4)	*Microscilla*	(4)	*Mariprofundus*	(1)	*Piscirickettsia*	(7)	*Caminibacter*	(1)	*Oceanobacillus*	(1)	*Thermobifida*	(2)	*Microcolrus*	(1)	*Anaerotruncus*	(1)	
141	*Acidithiobacillus*	(1)	*Pigeon*	(4)	*Monozwella*	(4)	*Mitsuokella*	(1)	*Coprothermobacte*	(6)	*Capnocytophaga*	(1)	*Parabacteroides*	(1)	*Trichodesmium*	(2)	*Nitrococcus*	(1)	*Aquifex*	(1)	
142	*Actinoallotichus*	(1)	*Piscirickettsia*	(4)	*Ornocaccus*	(4)	*Nicoletella*	(1)	*Cytophaga*	(6)	*Carnobacterium*	(1)	*Parvimonas*	(1)	*Acaryochloris*	(1)	*Nitrosomonas*	(1)	*Brachyspira*	(1)	
143	*Aeromonas*	(1)	*Terrabacter*	(4)	*Oligobrachia*	(4)	*Nitrococcus*	(1)	*Desulfitobacterium*	(6)	*Caprothermobacte*	(1)	*Psychroflexus*	(1)	*Acidiphilium*	(1)	*Pediococcus*	(1)	*Caldicellulosirupto*	(1)	
144	*Algoriphagus*	(1)	U (Betaproteobac	(4)	*Piriasmonas*	(4)	*Parvimonas*	(1)	*Donra*	(6)	*Cytophaga*	(1)	*Rhodoferax*	(1)	*Alteromonas*	(1)	*Piscirickettsia*	(1)	*Chloroflexus*	(1)	
145	*Anabaena*	(1)	U (Spirochaetes)	(4)	*Rhizobium*	(4)	*Pedlococcus*	(1)	*Parabacteroides*	(6)	*Dehalococcoides*	(1)	*Roseobacter*	(1)	*Anaerotruncus*	(1)	*Polaromonas*	(1)	*Ehrlichia*	(1)	
146	*Blastopirellula*	(1)	*Veillonella*	(4)	U (Betaproteobac	(4)	*Piscirickettsia*	(1)	*Pectobacterium*	(6)	*Desulfovibrio*	(1)	*Serratia*	(1)	*Anoxybacillus*	(1)	*Polymorphobacter*	(1)	*Eubacterium*	(1)	
147	*Brucella*	(1)	*Aggregatibacter*	(3)	U (Magnetococci	(4)	*Planctomyces*	(1)	*Pelagibacter*	(6)	*Dichelobacter*	(1)	*Synechococcus*	(1)	*Azoarcus*	(1)	*Porphyromonas*	(1)	*Halothermothrix*	(1)	
148	C.(Chlamydiae)	(1)	*Bradyrhizobium*	(3)	U (Nanoarchaeot	(4)	*Polaribacter*	(1)	*Arcobacter*	(5)	*Enterobacter*	(1)	*Syntrophobacter*	(1)	*Azotobacter*	(1)	*Saccharophagus*	(1)	*Heliobacterium*	(1)	
149	C.(Firmicutes)	(1)	C.(Alphaproteoba	(3)	*Acidithiobacillus*	(3)	*Porphyromonas*	(1)	*Nitrosomonas*	(5)	*Herpetosiphon*	(1)	*Thermobifida*	(1)	*Blastopirellula*	(1)	*Spiroplasma*	(1)	*Herminimonas*	(1)	
150	*Carboxydothermus*	(1)	*Campylobacter*	(3)	*Aeromonas*	(3)	*Pyrolitium*	(1)	*Oligobrachia*	(5)	*Hydrogenovibrio*	(1)	U (Chlorobi)	(1)	C (Alphaproteoba	(1)	*Sulfuriphaera*	(1)	*Kordia*	(1)	
151	*Chlamydia*	(1)	*Carboxydothermus*	(3)	*Aggregatibacter*	(3)	*Roseobacter*	(1)	*Pseudoalteromons*	(5)	*Ishikawaella*	(1)	U (Crenarchaeota	(1)	C (Betaproteobac	(1)	*Syntrophomonas*	(1)	*Methylobacillus*	(1)	
152	*Comamonas*	(1)	*Chlamydia*	(3)	*Alistipes*	(3)	*Salinibacter*	(1)	*Rickettsiella*	(5)	*Luteimonas*	(1)					*Caldicellulosirupto*	(1)	*Thermotoga*	(1)	*Methylomicrobium*	(1)
153	*Caprothermobacte*	(1)	*Desulfitobacillum*	(3)	*Alteromonas*	(3)	*Salinivibrio*	(1)	*Ruinerella*	(5)	*Lentisphaera*	(1)					*Capnocytophaga*	(1)	U (Actinobacteria	(1)	*Mitsuokella*	(1)
154	*Crocosphaera*	(1)	*Methylobacillus*	(3)	*Aquifex*	(3)	*Thermobifida*	(1)	*Bacteriovorax*	(4)	*Liberobacter*	(1)					*Caulobacter*	(1)	U (Deltaproteoba	(1)	*Natranaerobius*	(1)
155	*Desulfotalea*	(1)	*Mitsuokella*	(3)	*Hydrogenivirga*	(3)	*Thermodesulfovib*	(1)	*Halothermothrix*	(4)	*Luteimonas*	(1)					*Chlamydophila*	(1)	U (Elusimicrobia)	(1)	*Nitratiruptor*	(1)
156	*Dictyoglomus*	(1)	*Pelagibacter*	(3)	*Hydrogenovibrio*	(3)	*Thiobacillus*	(1)	*Hydrogenivirga*	(4)	*Mannheimia*	(1)					*Chloroherpeton*	(1)	U (Synergistetes)	(1)	*Nitrosopumilus*	(1)
157	*Fervidobacterium*	(1)	*Prochlorococcus*	(3)	*Leuconostoc*	(3)	U (Thermotogae)	(1)	*Lentisphaera*	(4)	*Methylobacillus*	(1)					*Chromobacterium*	(1)	*Vibrio*	(1)	*Ochrobactrum*	(1)
158	*Halothermothrix*	(1)	*Pseudovibrio*	(3)	*Neorickettsia*	(3)	U (Verrucomicro	(1)	*Paenibacillus*	(4)	*Methylobacterium*	(1)					*Chromohalobacter*	(1)	*Wolinella*	(1)	*Parabacteroides*	(1)
159	*Herminimonas*	(1)	*Pseudoxanthomon*	(3)	*Nicoletella*	(3)	*Wigglesworthia*	(1)	*Piriasmonas*	(4)	*Micromonospora*	(1)					*Citricoccus*	(1)			*Parvimonas*	(1)
160	*Hydrogenobaculu*	(1)	*Rhodococcus*	(3)	*Paenibacillus*	(3)	*Zymomonas*	(1)	*Selenomonas*	(4)	*Microscilla*	(1)					*Congregibacter*	(1)			*Pasteurella*	(1)
161	*Leuconostoc*	(1)	*Salinibacter*	(3)	*Pelotomaculum*	(3)			*Synechocystis*	(4)	*Monozwella*	(1)					*Desulfitobacillum*	(1)			*Pelotomaculum*	(1)
162	*Mesorhizobium*	(1)	*Shigella*	(3)	*Pseudovibrio*	(3)			*Syntrophomonas*	(4)	*Moritella*	(1)					*Desulfitobacterium*	(1)			*Pseudoalteromons*	(1)
163	*Metallosphaera*	(1)	*Sulfuriphaera*	(3)	*Rickettsiella*	(3)			*Thermobifida*	(4)	*Nicoletella*	(1)					*Desulfocaccus*	(1)			*Rhodovulum*	(1)
164	*Methylomicrobium*	(1)	U (Magnetococci	(3)	*Roseobacter*	(3)			*Vibrio*	(4)	*Nitrosopumilus*	(1)					*Dictyoglomus*	(1)			*Ruminococcus*	(1)
165	*Microscilla*	(1)	*Acaryochloris*	(2)	*Sulfitobacter*	(3)			*Arcanobacterium*	(3)	*Nodularia*	(1)					*Edwardsiella*	(1)			*Spiroplasma*	(1)
166	*Nicoletella*	(1)	*Akkermansia*	(2)	U (Fusobacteria)	(3)			*Enterobacter*	(3)	*Ornocaccus*	(1)					*Enterobacter*	(1)			*Spiroplasma*	(1)
167	*Nitratiruptor*	(1)	C.(Betaproteobac	(2)	*Acidovorax*	(2)			*Erysipelothrix*	(3)	*Olavius*	(1)					*Eubacterium*	(1)			*Sulfurimonas*	(1)
168	*Nitrococcus*	(1)	C.(Gammaproteo	(2)	*Aeropyrum*	(2)			*Hafnia*	(3)	*Paenibacillus*	(1)					*Exiguobacterium*	(1)			*Syntrophomonas*	(1)
169	*Planctomyces*	(1)	C.(Tenericutes)	(2)	*Anthrospira*	(2)			*Metallosphaera*	(3)	*Parabacteroides*	(1)					*Flavobacterium*	(1)			*Thermobifida*	(1)
170	*Rahnella*	(1)	*Chlorobaculum*	(2)	*Beggiatoa*	(2)			*Prevotella*	(3)	*Pasteurella*	(1)					*Fusobacterium*	(1)			*Treponema*	(1)
171	*Rhizobium*	(1)	*Chromohalobacter*	(2)	*Bordetella*	(2)			*Rahnnia*	(3)	*Pectobacterium*	(1)					*Gloeobacter*	(1)			U (Deltaproteoba	(1)
172	*Staphylothermus*	(1)	*Comamonas*	(2)	*Bradyrhizobium*	(2)			*Beggiatoa*	(2)	*Polymorphobacter*	(1)					*Heliobacterium*	(1)			U (Synergistetes)	(1)
173	*Sulfurihydrogenib*	(1)	*Dinoroseobacter*	(2)	*Brucella*	(2)			*Caldivirga*	(2)	*Porphyromonas*	(1)					*Hydrogenobaculu*	(1)			U (Thermotogae)	(1)
174	*Thermodesulfovib*	(1)	*Faecalibacterium*	(2)	*Chlorobium*	(2)			*Hydrogenovibrio*	(2)	*Promicromonospo*	(1)					*Hyphomonas*	(1)				
175	*Thermotoga*	(1)	*Flavobacterium*	(2)	*Crocosphaera*	(2)			*Methylophaga*	(2)	*Pyrobaculum*	(1)					*Lactococcus*	(1)				
176	*Thermus*	(1)	*Heliobacterium*	(2)	*Desulfitobacillum*	(2)			*Nodularia*	(2)	*Rahnnia*	(1)					*Mesorhizobium*	(1)				
177	U (Chloroflexi)	(1)	*Hyphomonas*	(2)	*Desulfovibrio*	(2)			*Parvimonas*	(2)	*Rahiginitalea*	(1)					*Methylomicrobium*	(1)				
178	U (Magnetococci	(1)	*Janthinobacterium*	(2)	*Hafnia*	(2)			*Spiroplasma*	(2)	*Roseburia*	(1)					*Methylophaga*	(1)				
179	U (Nanoarchaeot	(1)	*Lamprocystis*	(2)	*Lentisphaera*	(2)			*Thiocystis*	(2)	*Roseiflexus*	(1)					*Microcolrus*	(1)				
180	U (Spirochaetes)	(1)	*Leptospirillum*	(2)	*Luteimonas*	(2)			U (Aquificae)	(2)	*Roseobacter*	(1)					*Microcilla*	(1)				
181	U (Thermotogae)	(1)	*Metallosphaera*	(2)	*Lyngbya*	(2)			*Acidianus*	(1)	*Thermodesulfovib*	(1)					*Moritella*	(1)				
182			*Ornocaccus*	(2)	*Mannheimia*	(2)			*Aeropyrum*	(1)	*Thermotoga*	(1)					*Neisseria*	(1)				
183			*Pectobacterium*	(2)	*Microcystis*	(2)			*Alvinella*	(1)	U (Chlorobi)	(1)					*Nitrococcus*	(1)				
184			*Prdobacter*	(2)	*Micromonospora*	(2)			*Anaeromyxobacter*	(1)	U (Deferribacteres)	(1)					*Nitrosomonas*	(1)				
185			*Pyrobaculum*	(2)	*Mitsuokella*	(2)			*Anoxybacillus*	(1)	*Wigglesworthia*	(1)					*Nitrosopumilus*	(1)				
186			*Rhodobacter*	(2)	*Nitrosomonas*	(2)			*Baumannia*	(1)							*Nitrosospira*	(1)				
187			*Rhodopseudomon*	(2)	*Pasteurella*	(2)			*Beijerinckia*	(1)							*Oceanobacillus*	(1)				
188			*Salinivibrio*	(2)	*Prdobacter*	(2)			C.(Firmicutes)	(1)							*Ornocaccus*	(1)				
189			*Sulfurovum*	(2)	*Planctomyces*	(2)			C.(Tenericutes)	(1)							*Parabacteroides*	(1)				
190			*Thauera*	(2)	*Polymorphobacter*	(2)			*Chloroherpeton*	(1)							*Paracoccus*	(1)				
191			U (Acidobacteria	(2)	*Rheinheimera*	(2)			*Legionella*	(1)							*Prdococcus*	(1)				
192			U (Fibrobacteres)	(2)	*Roseiflexus*	(2)			*Liberobacter*	(1)							*Prlobacter*	(1)				
193			*Acidovorax*	(1)	*Ruminococcus*	(2)			*Luteimonas*	(1)							*Prlodictyon*	(1)				
194			*Aeropyrum*	(1)	*Salinivibrio*	(2)			*Microcolrus*	(1)							*Petrotoga*	(1)				
195			*Alistipes*	(1)	*Serratia*	(2)			*Micromonospora*	(1)							*Piscirickettsia*	(1)				
196			*Alkalilimnicola*	(1)	U (Crenarchaeota	(2)			*Monozwella*	(1)							*Plesiocystis*	(1)				
197			*Anaeromyxobacter*	(1)	*Zymobacter*	(2)			*Neisseria*	(1)							*Polymorphobacter*	(1)				
198			*Bartonella*	(1)	*Agrobacterium*	(1)			*Prdobacter*	(1)							*Pseudovibrio*	(1)				
199			*Beggiatoa*	(1)	*Algoriphagus*	(1)			*Prlodictyon*	(1)							*Rheinheimera*	(1)				

200	*Bordetella*	(1)	*Anoxybacillus*	(1)	*planctomycete*	(1)	*Roseovarius*	(1)
201	C (Bacteroidetes)	(1)	*Blastopirellula*	(1)	*Planococcus*	(1)	*Ruminococcus*	(1)
202	C (Epsilonproteob	(1)	*Brachyspira*	(1)	*Pseudanabaena*	(1)	*Silicibacter*	(1)
203	*Caminibacter*	(1)	*Caldivirga*	(1)	*Pyrobaculum*	(1)	*Stigmatella*	(1)
204	*Chloroherpeton*	(1)	*Cellvibrio*	(1)	*Roseburia*	(1)	*Sulfitobacter*	(1)
205	*Chromobacterium*	(1)	*Chloroflexus*	(1)	*Sulfurisphaera*	(1)	*Synechocystis*	(1)
206	*Chthoniobacter*	(1)	*Chloroherpeton*	(1)	*Thermus*	(1)	*Syntrophobacter*	(1)
207	*Clostridiaceae*	(1)	*Chromobacterium*	(1)	U (Chlorobi)	(1)	*Thermoanaerobac*	(1)
208	*Colwellia*	(1)	*Coprothermobacte*	(1)	U (Crenarchaeota	(1)	*Thermodesulforha*	(1)
209	*Coprothermobacte*	(1)	*Dechloromonas*	(1)	U (Deferribactere	(1)	*Thermosynechocc*	(1)
210	*Dechloromonas*	(1)	*Desulfobacula*	(1)	U (Dictyoglomi)	(1)	*Termus*	(1)
211	*Desulfotalea*	(1)	*Dorea*	(1)	U (Synergistetes)	(1)	U (Bacteroidetes)	(1)
212	*Desulfotomaculum*	(1)	*Enterobacter*	(1)	*Weissella*	(1)	U (Chlorobi)	(1)
213	*Gemmata*	(1)	*Erythrobacter*	(1)	*Xanthomonas*	(1)	U (Deferribactere	(1)
214	*Gloeobacter*	(1)	*Gloeobacter*	(1)			U (Magnetococcu	(1)
215	*Granulibacter*	(1)	*Gluconacetobacte*	(1)			U (Thermotogae)	(1)
216	*Heliobacillus*	(1)	*Herminiimonas*	(1)			U (Deinococcus-T	(1)
217	*Heliophilum*	(1)	*Mesorhizobium*	(1)			*Ureaplasma*	(1)
218	*Herminiimonas*	(1)	*Methylobacillus*	(1)			*Vibrio*	(1)
219	*Hoeflea*	(1)	*Methylobacterium*	(1)			*Victivallis*	(1)
220	*Hydrogenivirga*	(1)	*Methylophaga*	(1)			*Zymobacter*	(1)
221	*Jannaschia*	(1)	*Moorella*	(1)				
222	*Leptospira*	(1)	*Moritella*	(1)				
223	*Maricaulis*	(1)	*Nodularia*	(1)				
224	*Marinobacter*	(1)	*Ochrobactrum*	(1)				
225	*Microscilla*	(1)	*Planktothrix*	(1)				
226	*Nitratiruptor*	(1)	*Rahnella*	(1)				
227	*Nitrobacter*	(1)	*Ralstonia*	(1)				
228	*Nitrosomonas*	(1)	*Rhodobacter*	(1)				
229	*Novosphingobium*	(1)	*Rhodoferax*	(1)				
230	*Oligotropha*	(1)	*Rhodopirellula*	(1)				
231	*Parvimonas*	(1)	*Ruegeria*	(1)				
232	*Planctomyces*	(1)	*Shewanella*	(1)				
233	*Prosthecochloris*	(1)	*Sinorhizobium*	(1)				
234	*Pseudomonas*	(1)	*Syntrophobacter*	(1)				
235	*Rheinheimera*	(1)	*Syntrophomonas*	(1)				
236	*Rhizobium*	(1)	*Terrabacter*	(1)				
237	*Rhodopirellula*	(1)	*Thermosinus*	(1)				
238	*Rhodospirillum*	(1)	U (Chlorobi)	(1)				
239	*Silicibacter*	(1)	U (Elusimicrobia)	(1)				
240	*Symbiobacterium*	(1)	U (Fibrobacteres)	(1)				
241	*Synechocystis*	(1)	U (Nitrospirae)	(1)				
242	*Thermoanaerobac*	(1)	*Verminephrobacte*	(1)				
243	*Thermosynechocc*	(1)						
244	U (Bacteroidetes)	(1)						
245	U (Crenarchaeota	(1)						
246	U (Cyanobacteria	(1)						
247	*Victivallis*	(1)						
248	*Wolbachia*	(1)						
249	*Wolinella*	(1)						

Numbers in parentheses indicate the percentage. U, Unclassified; C, Candidatus.