The low diverse gastric microbiome of the jellyfish *Cotylorhiza tuberculata* is dominated by four novel taxa

Tomeu Viver[1], Luis H. Orellana[2], Janet K. Hatt[2], Mercedes Urdiain[1], Sara Díaz[1], Michael Richter[4], Josefa Antón[5], Massimo Avian[7], Rudolf Amann[6], Konstantinos T. Konstantinidis[2,3], Ramon Rosselló-Móra[1*]

1- Marine Microbiology Group, Mediterranean Institute for Advanced Studies (IMEDEA; CSIC-UIB), E-07190 Esporles, Spain
2- School of Civil and Environmental Engineering, Georgia Institute of Technology, 311 Ferst Dr. NW, Atlanta, GA 30332, USA
3- School of Biological Sciences, Georgia Institute of Technology, 950 Atlantic Dr. NW, Atlanta, GA 30332, USA
4- Ribocon GmbH, 28359 Bremen, Germany
5- Department of Physiology, Genetics and Microbiology, and Multidisciplinary Institute for Environmental Studies Ramon Margalef, University of Alicante, Alicante, Spain
6- Department of Life Science, University of Trieste, Via L. Giorgieri 10, 34127 Trieste, Italy
7- Department of Molecular Ecology, Max-Planck-Institut für Marine Mikrobiologie, D-28359 Bremen, Germany

***Corresponding Author**:

Ramon Rosselló-Móra
Marine Microbiology Group
Mediterranean Institute for Advanced Studies (IMEDEA; CSIC-UIB)
E-07190 Esporles
Spain
Tel: +34 971 611 826
Email: rossello-mora@uib.es

Running Title: Low diverse microbiome of *C. tuberculata*

Sequence deposits: The new data has been deposited in the European Nucleotide Archive (ENA) under the following accession numbers: 16S rRNA gene sequences LT599034 to LT599043; 23S rRNA gene sequence LT599033. Raw data and binned genomes were submitted under the study accession code PRJEB14783 and respectively under the sample accession codes ERS1246994 to ERS1246997.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13763

Originality-Significance Statement: *Cotylorhiza tuberculata* is one of the most abundant jellyfishes in the Mediterranean Sea. Its massive population blooms have important consequences for economy (e.g., tourism and fisheries) and carbon flow, and may be an indication of the ecosystem decay. Previous studies suggested a very low diversity of the gastric microbial community, which make it a good model for simple digestive microbiomes. However, the diversity, abundance and function of bacteria in the gastric cavity of this scyphozoan are poorly characterized.Here, we employed direct sequencing metagenomics combined with fluorescence in situ hybridization to characterize the jellyfish microbial communities. The microbiome included novel bacterial taxa, some of them present in other marine invertebrates. Metagenomic reconstruction allowed the recovery of the almost complete genomes of key community members. Analysis of these genomes provided testable hypotheses of their role of the microbiome in jellyfish biology.

Summary

Cotylorhiza tuberculata is an important scyphozoan jellyfish producing population blooms in the Mediterranean probably due to pelagic ecosystem's decay. Its gastric cavity can serve as a simple model of microbial – animal digestive associations, yet poorly characterized. Using state-of-the-art metagenomic population binning and catalyzed reporter deposition fluorescence *in situ* hybridization (CARD-FISH), we show that only four novel clonal phylotypes were consistently associated with multiple jellyfish adults. Two affiliated close to *Spiroplasma* and *Mycoplasma* genera, one to chlamydial '*Candidatus* Syngnamydia', and one to bacteroidetal *Tenacibaculum*, and were at least one order of magnitude more abundant than any other bacteria detected. Metabolic modeling predicted an aerobic heterotrophic lifestyle for the chlamydia, which were found intracellularly in *Onychodromopsis*-like ciliates. The *Spiroplasma*-like organism was predicted to be an anaerobic fermenter associated to some jellyfish cells, whereas the *Tenacibaculum*-like as free-living aerobic heterotroph, densely colonizing the mesogleal axis inside the gastric filaments. The association between the jellyfish and its reduced microbiome was close and temporally stable, and possibly related to food digestion and protection from pathogens. Based on the genomic and microscopic data we propose three candidate taxa: '*Candidatus* Syngnamydia medusae', '*Candidatus* Medusoplasma mediterranei' and '*Candidatus* Tenacibaculum medusae'.

Introduction:

Massive blooms of jellyfishes are often associated to ecosystem decay due to overfishing, eutrophication and climate change (Richardson et al., 2009). These promote a trophic chain shift, favoring adapted predators (Utne-Palm et al., 2010) and outcompeting some fish species to ultimately modify the structure of the marine communities (Richardson et al., 2009). One of the most relevant jellyfish increasingly generating massive blooms in the Mediterranean waters is *Cotylorhiza tuberculata* (Macri, 1778), a scyphozoan of the phylum *Cnidaria*, (Prieto *et al*., 2010). Its life cycle encompasses a stage with pelagic medusae occurring in late summer after

abrupt temperature increases. Such jellyfish blooms generate tons of biomass in confined waterbodies that ultimately sink and are consumed by benthic scavengers (Yamamoto et al., 2008). The life span of this species in the pelagic jellyfish stage is one year, and it often ends in mass mortality events (Kikinger, 1992, Prieto *et al*., 2010). The cause of this mass mortality is unknown, but the resulting deposition of decaying jellyfish triggers environmental shifts due to the release of organic and inorganic nutrients (Pitt *et al*., 2009). This includes the activation of bacterial degradation and pronounced changes in the composition of the planktonic microbial community (Dinasquet *et al*., 2012; Tinta *et al*., 2010; Tinta *et al*., 2012). Jellyfish blooms have important consequences for tourism and fisheries (Palmieri *et al*., 2014), and *C. tuberculata* medusae have been reported to be carriers of potential fish pathogens (Cortés-Lara *et al*., 2015; Delannoy *et al*., 2011). Despite its relevance for animal and potentially also human health, aquaculture and tourism, not much is known about the microbiota associated with *C. tuberculata* and its relevance. A few studies have addressed the molecular microbial ecology of ctenophores (Daniels and Breitbart, 2012; Hao *et al*., 2015), which includes the detection of fish pathogens in cnidarians (Delannoy *et al*., 2011; Ferguson *et al*., 2010; Fringuelli *et al*., 2012). Only recently, the microbiomes of two scyphozoan species (*C. tuberculata* and *Aurelia aurita*) have been addressed using 16S rRNA gene amplicon massive sequencing (Cortés-Lara *et al*., 2015; Weiland-Bräuer *et al*., 2015). In both cases, the microbial composition was shown to be specific for the jellyfish and different from the surrounding waters, and in the case of *A. aurita* a compartmentalization (i.e. exumbrella mucus and gastric cavity) with distinct communities within a given exemplar was also demonstrated.

A pilot study on the gastric cavity microbiomes of *C. tuberculata* by means of 16S rRNA gene amplicon sequencing, produced a very intriguing picture of a very low diverse composition, mainly represented by members of the genus *Spiroplasma sensu stricto* that, together with less frequent *Thalassospira* and *Tenacibaculum* phylotypes, accounted for >95% of the retrieved bacterial diversity (Cortés-Lara *et al*., 2015). The gastric cavity of the scyphozoan medusa can be taken as a simple model system for digestive microbiomes. It works as a “blind gut” in where food enters and waste exits through the same numerous mouth arm openings. Inside, the finger-like gastric cirri secrete digestive enzymes, absorb the products of digestion and contain nematocysts that subdue preys still alive. In addition, the mesogleal tissue in the cavity of the gastric filaments transports oxygen and food products to the tissues. The scyphozoa feed on microplanktonic organisms as *Ciliata*, *Crustacea* and *Gastropoda* whose remains have been found among the particulate food items (Kikinger, 1992). Besides, some potential fish pathogens from the genus *Vibrio* had been isolated from the same samples (Cortés-Lara *et al*., 2015), making possible that this pelagic organism could act as dispersal mechanism of such pathogens. This was the very first approach to reveal a putative digestive microbiome in scyphozoan jellyfishes.

In this present study, the same organs in the gastric cavity were submitted to a shotgun metagenomics approach (direct DNA sequencing) to reveal the genomic repertoire of the major members of the associated microbial community and create hypotheses about their role for

jellyfish biology and ecology. Catalyzed reporter deposition-fluorescence *in situ* hybridization (CARD-FISH) was also employed to visualize the microbial cells and identify their preferred location in adult jellyfishes captured in two sampling campaigns between 2013 and 2016. This analysis provided insights into the genetic repertoire, potential metabolic roles and pathogenicity factors, as well as *in situ* localization of the major microbial community members, allowing for their classification as three new *Candidatus* taxa (Konstantinidis and Rosselló-Móra, 2015). We also described the challenges associated with population binning and visualization of the cells within the animal host, which could serve as a guide for future studies of this or similar symbiosis systems.

Results:

Microbial diversity based on rRNA gene fragments

DNA extracts from samples of disintegrated gastric filament – gonad organs (Supplementary Figure S1) of the four specimens (M1-M4) previously studied by 16S rRNA amplicon pyrosequenciation (Cortés-Lara *et al*., 2015), were submitted to metagenome sequencing. In addition, four medusae (C1-C4) captured in September 2015 and two (C5-C6) in September 2016 were studied by CARD-FISH to reveal the community stability between years. In all sampling dates, the jellyfishes were swimming close to the surface waters (within one meter from the surface), and the water temperature was approximately 25ºC. Six specimens were females and four were males (Table 1), yet the total weight of the 2015 medusae was only about half that of the M1-M4 (2013) and C5-C6 (2016) specimens. We sequenced between 2.2 to 2.5 Gb per sample, which translated to between 7.0 $x10^6$ to 8.3 $x10^6$ single reads after trimming (Supplementary Table S1). Assembly yielded between 1,046 and 43,649 contigs larger than 1 kb, and the larger assembled contigs ranged between 165 and 247 kb. The average sequencing coverage of the microbial community provided by each metagenomic dataset (calculated by Nonpareil curves; Supplementary Figure S2) was between 82.3% and 85.4%, meaning these fractions of the total extracted DNA per sample were sequenced.

Small subunit rRNA gene fragments were identified on the raw unassembled metagenomic reads for each jellyfish sample. A total of 12,358 SSU rRNA-encoding reads were obtained, with reads from an individual sample ranging from 1,222 to 5,033 in M2 and M3, respectively (Table 2). The read lengths ranged between 41 to 250 nucleotides (mean 149 $\pm$ 59; and median 145) covering all different areas of the gene. The parsimony insertion into the non-redundant SILVA SSURef_NR99_123 default tree (Quast *et al*., 2013) allowed the affiliation of almost all fragments (282 were non-16S rRNA and 677 corresponded to 18S rRNA gene sequences). Most of the reads (35.4% in M2 to 56.2% in M1) affiliated with the *Chlamydia* phylogenetic branch, with *Simkania negevensis* being the closest type strain sequence. In addition, three major groups of bacterial sequences could be detected, two of which affiliated with the *Mollicutes* phylum. One was close to *Spiroplasma sensu stricto* (i.e. the branch comprising the type strain *S. citri*, in relative abundances ranging between 2.4% in M2 and 23.6% in M1) and

the other to a *Mycoplasma*–like phylogenetic branch only comprised of environmental sequences generally obtained from molluscs (Duperron *et al*., 2013) or corals (Kimes *et al*., 2013). The third sequence cluster affiliated with the phylogenetic branch comprising all the *Tenacibaculum* species, with the exception of the type species of the genus *T. maritimum* that actually affiliated with the *Polaribacter* genus. The few remaining bacterial sequences (between 2.8% and 6.8% of the total) affiliated with various clades (Supplementary Table S2). Between 1.4% and 28.5% of the reads corresponded to eukaryotic 18S rRNA gene fragments (Supplementary Table S3). The male jellyfish M2 had the highest number of eukaryotic fragments, most of which corresponded to either unclassified *Eumetazoa* (35.9% of 18S reads) or to *Cnidaria/Scyphozoa* (23.5%). These sequences most probably originated from the sperm cells. Other reads corresponded to *Copepoda/Calanoida* (7.3% of 18S reads), probably originating from ingested crustaceans, which were also observed in the gastric cavity by microscopy (see below in the CARD-FISH section). Approximately 10% of the sequences affiliated with different groups of *Alveolata* of which one could be clearly affiliated to the dinoflagellate *Symbiodinium*, also identified in the gastric cavity by microscopy. Approximately half of the *Alveolata* sequences affiliated closely to the *Oxytrichidae* protozoa (Supplementary Figure S3) *Onychodromopsis flexilis* (seq. AM412764), *Paraurosomoida indiensis* (JX139117), and especially to environmental clone AB695447 originating from a 18S rRNA gene sequencing survey of the Antarctic moss pillar *Leptobryum* sp. (Nakai *et al*., 2012). The remaining *Alveolata* only affiliated with environmental sequences of uncultured organisms.

The almost full-length sequence of the 16S rRNA and 23S rRNA genes of the four major phylotypes could be recovered from most of the bins, encoded within the larger contigs of each bin, and flanked by genes whose annotation clearly indicated that the contig was not chimeric. The phylogenetic reconstruction of the assembled rRNA gene sequences mirrored the major observations made with single-read fragments. Four almost complete 16S rRNA gene sequences were recovered (Figure 1). The dominant sequence cluster present in all four specimens M1 - M4 affiliated with the lineage comprising *S. negevensis* Z (U68460) as the single species with a validly published name. The most closely related sequences had been obtained from the gills of the marine fish *Symphodus melops* in Norway (KC608868, 99.4% sequence identity) (Nylund *et al*., 2015), followed by '*Candidatus* Syngnamydia salmonis' (KF768762), '*Candidatus* Syngnamydia venezia' (KC182514), and the '*Candidatus* Fritschea' genus (AY140910) with 97.3%, 96.4% and 95.4% identity, respectively. In addition, almost full-length 16S rRNA sequences of the bacteroidetal genus *Tenacibaculum* were assembled from two metagenomes (M1 and M4), although gene fragments of this group were also found in M2 and M3 (see above). The closest relative was *T. aestuarii* (DQ314760) with 96.7% 16S rRNA gene sequence identity (Figure 1). It was also possible to assemble complete mollicutal 16S rRNA sequence variants. One was loosely affiliated to the *Spiroplasma sensu stricto* branch harboring the type species of the genus with identity values of <88%. Although *Spiroplasma*-like sequences were present in all four metagenomes, complete 16S rRNA sequences could only be assembled in M1, M3 and M4. The second mollicutal sequence was only assembled from

M3, and affiliated loosely with a *Mycoplasma*-like branch only consisting of environmental sequences. This sequence shared 82.1% identity with sequence HE663394 obtained from the digestive tract of the mollusc *Leptochiton boucheti* (Duperron *et al*., 2013). In all cases, except for *Tenacibaculum* sp., the identity values were well below 94.5%, which has been suggested as a threshold for the genus category (Yarza *et al*., 2014). These genes were searched and found in several 16S rRNA gene amplicon libraries of corals, gorgonian and sponges in study in our laboratory (Rubio-Portillo et al., unpublished), and identical matches for the *Simkania*-like and *Tenacibaculum*-like sequences were found in the recently published "global sponge microbiome" (Thomas et al., 2016).

Gene insertion within the 23S rRNA gene of the Simkania-like bins.

In the chlamydial bins, a 3,798 nucleotide-long sequence of the 23S rRNA gene was assembled. This gene had an 895 nucleotide insertion at position 1,984, closely related to group I introns present in the chloroplasts and mitochondria of algae and amoeba (Everett *et al*., 1999). Its bioinformatics annotation suggested that it encoded a putative homing endonuclease (Supplementary Figure S4). This intron was most closely related to that of a chlamydial symbiont of the primitive bilaterian *Xenoturbella* (Israelsson, 2007), with 76.3% identity. A phylogenetic reconstruction based on the almost complete 23S rRNA gene reproduced the same topology as the 16S rRNA analysis (data not shown).

Visualization and quantification of phylotypes

Bright field microscopy of disintegrated gastric filament and gonad tissues, and the fluid content of the cavity (Supplementary Figures S1 and S5A) revealed that most of the cells were spherical with a diameter of approximately 5 µm, and larger cells in much lower abundances were observed with an ovoid shape approximately 10 µm in length. In addition, all samples showed the presence of sperm-like cells, which were most abundant (orders of magnitude higher; data not shown) in the male cavities. Finally, the presence of planulae in different developmental stages and parts of calanoid copepods probably degraded by digestion could be observed among the female samples (Supplementary Figure S5D).

In order to microscopically identify, localize and quantify the four major phylotypes, fluorescence *in situ* hybridization (FISH) and catalyzed reporter deposition-FISH (CARD-FISH) were applied using the already established probes for Eukarya (EUK516), Bacteria (EUB338 I-III), Archaea (ARCH915), *Bacteroidetes* (CF319a) and the negative control oligonucleotide NON338. In addition, a set of new probes were designed and optimized targeting the major groups of 16S rRNA gene sequences detected here: *Simkania*-like, Simk174; *Tenacibaculum*-like, Tena1432; *Spiroplasma*-like, Spiro199; *Mycoplasma*-like Myco738. Also, oligonucleotide probes were designed for the 18S rRNA of some scyphozoan species of the *Rhizostomeae,* order, among them *C. tuberculata* (Cotu193 and Cotu1453), and the *Onychodromopsis flexilis* – *Parausomodia indiensis* clade (Onyc1121) (Table 3; Supplementary Figure S6). *In situ* visualization of target bacteria in medusa samples by FISH with fluorescently mono-labeled probes suffered from high background fluorescence and non-specific attachment of probes, as

indicated by NON338 binding (data not shown). Reliable detection required the more sensitive CARD-FISH assay. As none of the target organisms of the new probes could be enriched and/or cultured, sensitivity and stringency optimization of all new probes had to be assayed using the original samples as test material, in relation to the phylotype hybridized contrasted with the detection with EUB338 probe and the absence of signal with NON338 (see experimental procedures section). The probes consistently hybridized with the expected morphotypes (i.e. abundances, common morphology and localization).

All probes except for ARCH915 and Myco738 detected cells in fixed jellyfish samples (Figure 2 and supplementary Figure S5). There were at least four types of eukaryotic cells that could be differentiated based on probe hybridization, autofluorescence and DNA staining with DAPI: *C. tuberculata* cells hybridizing with EUK516, as well as both Cotu193 and Cotu1453 (Figures 2B, 2I and 2J); *Symbiodinium*-like cells exhibiting red autofluorescence due to the presence of photosynthetic pigments (Figures 2C, 2D and 2J); cells of a ciliate, closely related to *Onychodromopsis*, hybridizing with EUK516 and the newly designed probe Onyc1121, but not with Cotu193 or Cotu1453 (Figure 2F, 2G); and a fourth type of cell, with arrow-shaped nuclear organelles (Figures 2B and 2D), not hybridizing with any of the probes used, which most likely represented sperm cells that could be clearly visualized with phase-contrast microscopy (Supplementary Figure S5). *Onychodromopsis*-like cells with 4-5 µm large pleomorphic nuclei occupying most of the cytoplasm could be easily distinguished from cells hybridizing with the *C. tuberculata* probes in which the spherical nuclei had a diameter of 2 µm (Figures 2B, 2C, 2F, 2G and 2J). The shape of the *Onychodromopsis*-like cells was often ovoid with a length of approximately 10 µm.

The Simk174 probe hybridized to small pleomorphic cells with a diameter of approximately 0.6 µm, which were located in the cytoplasm of the *Onychodromopsis*-like eukaryotic cells (Figures 2A, 2B, 2C and 2D, and Supplementary Figure S5). When a ciliate cell was infected, there were generally >20 Simk174-positive cells in the cytoplasm. In addition, these eukaryotic cells exhibited additional intra-cytoplasmic DAPI signals not hybridizing with probe Simk174 (Figures 2B and 2D), suggesting the presence of more intracellular bacteria. The identity of the latter signals could not be revealed as none of the probes used here showed a positive signal. In all jellyfish studied, except C3 (Table 1), *Onychodromopsis*-like ciliates containing Simk174-positive cells were detected. There were no clear differences between male and female medusae.

Probe Tena1432 targeting the *Tenacibaculum*-like organism bound exclusively to extracellular rod-shaped cells with a length of approximately 5 µm. These long rods also hybridized with the CF319a probe, which confirmed their affiliation to *Bacteroidetes* (Supplementary Figure S5). The number of Tena1432-positive cells was low freely swimming in the gastric fluid (see below). However, these organisms were forming dense colonies, scattered within the mesogleal axis of the gastric filaments (Figure 2C and 2E).

We could only observe a positive signal with the Spiro199 probe for sample C2 collected in 2015, although sequences of the *Spiroplasma*-like organism were retrieved from all jellyfishes in 2013 (M1 to M4; Table 2). In C2, the Spiro199-positive cells always appeared pleomorphic in shape, were associated with jellyfish cells (Figure 2H and Supplementary Figure S5) and had a size of approximately 0.6 µm. No *Spiroplasma*-like cells could be detected in cells of *Symbiodinium* or *Onychodromopsis*-like ciliates (Supplementary Figure S5). Finally, despite testing all specimens, no positive signals were obtained with probe Myco738 targeting the *Mycoplasma*-like organism. Thus, the probe application parameters could not be optimized. Therefore, the lack of hybridization with probe Myco738 should not be taken as evidence of the absence of *Mycoplasma*-like cells, since cells could also have remained undetected due to low abundance, consistent with the finding based on sequencing.

Due to the complex tissue structure, we could only calculate abundances in the fluids of the gastric cavity of the two jellyfishes captured in last campaign (C5-C6). In these suspensions the number of *C. tuberculata* cells was 6.04 $x10^6$ ($\pm$1.7 $x10^6$) cell/mL; the *Onychodromopsis*-like ciliates 4.8 $x10^5$ ($\pm$1.35 $x10^5$) cell/mL; *Symbiodinium* cells 4.53 $x10^5$ (1.28 $x10^5$) cell/mL; *Tenacibaculum*-like cells 6.53 $x10^4$ ($\pm$1.07 $x10^5$); and the free – living bacteria 5.92 $x10^5$ ($\pm$1.67$x10^5$) cell/mL. All *Onychodromopsis*-like ciliates exhibited infection, thus the numbers of *Simkania*-like cells could reach 9.6 $x10^6$ (3.4 $x10^6$) cell/mL. However, intact gastric filaments exhibited *Tenacibaculum*-like cells forming dense colony-like colonization of the mesogleal axis in the gastric filaments (Figure 2C and 2E and Supplementary Figure S5). This phylotype represented >99% of the bacterial fraction in these filaments. In the disintegrated gastric filament and gonad tissues the ratio of *Tenacibaculum-like* was nearly 1:1 of the *C. tuberculata* cells, confirming that this organism was mostly inhabiting the organ and not freely swimming in the gastric cavity fluids.

Metagenome population binning and functional annotation

Each of the four metagenomes was binned separately using MaxBin (Wu *et al*., 2014) to avoid making chimeric sequences of related but distinct microbial populations found in each jellyfish (Supplementary Table S4). One bin with high genome completion (>94%) and essentially no heterogeneity or contamination, belonging to the *Simkania*-like phylotype, was obtained in all samples (Figure 1). The four *Simkania*-like bins (one from each sample) were highly related to each other (>99.97% average nucleotide identity) (Supplementary Table S5) and likely represented one homogenous population. Additional bins were recovered, although not from all four metagenomes. These bins usually showed high heterogeneity and were therefore submitted to a second round of binning, where reads mapping on contigs of the bin (98% nucleotide identity level) were reassembled, and the newly formed contigs were manually checked and removed from the bin when they showed unusual coverage and/or phylogenetic origin (see Methods for further details). After removing heterogeneities and contamination, six additional high quality bins were recovered (Table 4), which encoded the 16S rRNA genes shown in Figure 1. Bins representing the *Spiroplasma*-like lineage were extracted from

metagenomes M1, M3 and M4 with a completeness of 55%, 13% and 94%, respectively. M1 and M4 yielded bins of the *Tenacibaculum*-like lineage with a completeness of 93% and 29%, respectively. A bin of 74% completeness corresponding to the *Mycoplasma*-like lineage was obtained from metagenome M3.

Metabolic modeling of Simkania-like bins

The almost complete *Simkania*-like bins had sizes between 2.1 to 2.3 Mb (Table 4 and Table 5). Their G+C% mol content was 39.8% for M1, M2 and M3 and 39.6% for M4. The number of contigs in each bin ranged between 48 and 123. The Tetra Correlation Search (TCS) analysis (Richter *et al*., 2015) showed *S. negevensis* Z as the closest available genome, sharing a total of 1,344 (64%) genes, with an amino acid average identity (AAI) of 59.6%. The number of predicted coding sequences (CDSs) on each *Simkania*-like bin ranged between 2,058 and 2,114, and among these approximately 42% corresponded to hypothetical CDSs. In this regard, 705 of these CDSs were specific to the *Simkania*-like bin when compared to their closest relative, and 490 of these (70% of the bin-specific genes; 23% of the total bin genes) were hypothetical CDSs. The annotated bin-specific genes (Supplementary Table S6) were in general related to transport and metabolism, as well as plasmid functions such as the partitioning protein ParA and virulence plasmid protein pGP6-D. In addition, we checked for the presence and absence of certain virulence factors relevant in *Chlamydiaceae* (Collingro *et al*., 2011; Supplementary Table S7), such as the type III effector proteins, proteins containing eukaryotic domains, DNA-binding protein from starved cells, and M4 thermolysin protein family. Most of these features were present in the bin and thus, shared with the close *Chlamydiaceae* relatives. As in the other *Chlamydiaceae*, except for *Parachlamydia*, no chemotaxis systems could be annotated. In addition, the bin-specific genetic repertoire was mainly comprised of hypothetical CDSs.

According to the metabolic model built from KEGG and RAST annotation, the *Simkania*-like bin sequences appeared to represent aerobic heterotrophic bacteria. They encoded the complete Embden-Meyerhof (glycolysis) pathway, the pyruvate-oxidizing enzymes of the pentose phosphate pathway required for the synthesis of acetyl-CoA, and the Krebs cycle. They also encoded for several electron accepting and donating enzymes related to aerobic respiration, such as cytochrome d and o, ubiquinol oxidases, cytochrome C oxidases, Na(+)-translocating NADH-quinone oxidoreductase and the rnf-like group of electron transport complexes, and respiratory dehydrogenases. The *Simkania*-like bacteria were predicted to be able to synthesize NAD^+, but not tryptophan. As for *S. negevensis*, the *Simkania*-like bins did not encode for the known *Chlamydiaceae* outer membrane complex proteins (OMC; Supplementary Table S8), such as PorB, OprB and the Pmps group of proteins, but encoded for the major outer membrane protein OmpA. As observed for the *Simkania*, *Waddlia* and *Parachlamydia* (Collingro *et al*., 2011), the *Simkania*-like bin also contained a large set of 17 genes encoding for eukaryotic protein domains (Supplementary Table S9), as well as genes encoding the DNA-binding protein from starved cells (Dps) and proteins of the thermolysin family M4. The

metabolic model based on the gapfilling tool (included in DOE Systems Biology Knowledgebase server – KBase web-based interface) predicted that the *Simkania*-like bacteria would actively import (in decreasing relevance; Supplementary Table S10A): 2-oxoglutarate, glycerol-3-phosphate, D-glucosamine, L-proline, and oxygen (among others). The model also predicted that they would actively secrete (in decreasing relevance; Supplementary Table S10B): glycerol, pyrophosphate, L-glutamate, CO_2, L-malate, and D-mannose (among others). The bins encompassed genes involved in biotin, thiamin, menaquinone, riboflavin and folate biosynthesis. Based on the metabolic model, the *Simkania*-like bacteria would take up several amino acids from the environment: methionine, alanine, glycine, cysteine, L-valine, L-isoleucine, L-tyrosine, L-leucine, and L-lysine (Supplementary Table S10A). In addition, the *Simkania*-like bins contained transporters to also take up several ions from the environment (with flux higher than 0): Mg^{2+}, Mn^{2+}, K^{+}, Cl^{-}, Zn^{2+}, Cu^{2+}, Co^{2+} and Fe^{3+}, and they encoded genes related to copper homeostasis and tolerance, as well as zinc resistance.

Metabolic modeling of the Spiroplasma-like bin

The high quality, non-contaminated *Spiroplasma*-like bin retrieved from metagenome M4 was 94.4% complete and had a size of 0.68 Mb in 68 contigs (Table 4 and Table 5). The ANI values with the other two *Spiroplasma*-like bins of lower completion were >99.97% (Supplementary Table S5), suggesting that the same population was present in the gastric cavities of specimens M1, M3 and M4. The G+C% content of these bins was approximately 23.5%, which was in the typical range for *Spiroplasma*. The TCS analysis showed *Spiroplasma kunkelii* CR2-3x, *Mycoplasma iowae* 695, *M. lipofaciens* ATCC 35015 and *Mesoplasma grammopterae* ATCC 49580 as the closest available genomes. The *Spiroplasma*_M4 bin shared 418 genes with *Spiroplasma kunkelii* CR2-3x (47.5% AAI), 410 with *M. iowae* strain DK-PCA (41.3% AAI), 409 with *M. iowae* strain 695 (40.7% AAI), 399 with *M. grammopterae* (45.2% AAI) and 369 with *M. lipofaciens* (41.3% AAI). The six genomes shared a total of 254 genes, of which 25 were hypothetical proteins. From the total of 977 predicted genes in the *Spiroplasma*_M4 bin, 463 were exclusive to this genome. The annotation indicated that the *Spiroplasma*_M4 bin encoded for the complete Embden-Meyerhof (glycolysis) pathway, the non-oxidative phase of the pentose phosphate pathway (synthesis of ribose-5P from fructose-6P), and the enzyme ribose-phosphate pyrophosphokinase for the synthesis of phosphoribosyl pyrophosphate (PRPP) from ribose-5P. Moreover, it encoded the enzymes for galactose, alanine and aspartate metabolism, lysine biosynthesis and the metabolism of seleno-compounds. The bin encoded only nine genes related to respiration, seven of which were involved in the F0F1-type ATP synthase and two were related to respiratory dehydrogenase. No cytochrome oxidases could be annotated. Based on the gapfilling metabolic model (KBase web-based interface) the respective bacteria would not take up O_2 from the environment (Supplementary Table S10A), and thus not perform aerobic respiration, rather they would probably gain energy through fermentation. The metabolic model predicted that as major uptake compounds, the *Spiroplasma*-like organism would actively import (in decreasing relevance; Supplementary Table S10A): 2-D-glucosamine, L-arginine, trehalose and glycerol (among others), as well as take up ions such as H^{+}, Cl^{-}, K^{+}, Mg^{2+}, Fe^{2+}

and Fe^{3+}. In addition, the model predicted that it would overproduce and actively secrete (in decreasing relevance; Supplementary Table S10B): D-glucose, CO2, and N-acetyl-d-glucosamine (among others). The *Spiroplasma*_M4-bin did not encode for genes involved in cell wall and capsule synthesis, nor those required for motility and chemotaxis.

Metabolic modeling of the Tenacibaculum-like bin

Binning of metagenome M1 rendered a 92.8% complete, non-contaminated genome with a size of 3.027 Mb in 695 contigs (Tables 4 and 5). The ANI value with a second smaller bin (29.4% complete) retrieved from specimen M4 was >98.9% (Supplementary Table S5). The G+C% mol content of the larger and almost complete bin was 31.5%. The TCS analysis showed *T. mesophilum* HMG1 and *T. maritimum* NBRC 15946 as the closest available genomes. The *Tenacibaculum*_M1 bin encoded a total of 2,649 predicted genes, 1,086 of which were hypothetical proteins. The bin shared 1,485 genes with *T. maritimum* (66.3% AAI), and 1,501 with *T. mesophilum* (67.8% AAI). The modeling based on KEGG predicted that the organism would be aerobic due to the presence of terminal cytochrome C oxidases and genes involved in the biogenesis of c-type cytochromes, cbb3-type cytochrome C oxidases, succinate dehydrogenases and respiratory dehydrogenases. In addition, this bin encoded for the complete glycolysis pathway (Embden-Meyerhof) and Krebs cycle, the non-oxidative phase of the pentose phosphate pathway, as well as the phosphate acetyltransferase-acetate kinase pathway for acetate assimilation from Acetyl-CoA, pyruvate oxidation, PRPP biosynthesis, propanoyl-CoA metabolism, serine and cysteine biosynthesis, with fatty acid beta-oxidation as the principal metabolic trait. Moreover, the cell would also be able to synthesize the amino acids valine, leucine, isoleucine and ornithine; and the vitamins thiamin (vitamin B_1), biotin (vitamin B_7), menaquinone (vitamin K_2), riboflavin (vitamin B_2) and pyridoxine (vitamin B_6). *Tenacibaculum* did not encode genes involved in motility and chemotaxis. In addition, based on the metabolic modeling, the cell would uptake Cu^{2+}, K^{+}, Fe^{3+}, Cl^{-}, Mn^{2+} and Mg^{2+} ions (Supplementary Table S10A), and encode for genes involved in copper and potassium homeostasis and the hemin transport system. The cell encoded for acetylornithine aminotransferase and acetylornithine deacetylase for ornithine production, one of the excreted metabolites according to the gapfilling model. In addition, ornithine was involved in the urea cycle, for the production of urea, another compound that the cell would overproduce and excrete to the environment (Supplementary Table S10B). The cell also appeared to excrete uridine, deoxyuridine, and ornithine. We could not find (as well as in the closely related *Tenacibaculum* genomes) susC/susD homologues (Larsbrink *et al.,* 2014) relevant for polysaccharide utilization in *Bacteroidetes*. However, a balanced ratio was detected between genes related to peptidases or proteinases (49 CDSs) and genes related to the catalysis of carbohydrates (61 CDSs; Supplementary Table S11).

Finally, and in order to understand whether the distinct populations of the major taxa detected here were clonal, reciprocal recruitment plots between the bins and the counterpart metagenomes were calculated (Supplementary Figure S7). In all cases, almost all recruited

reads (>80% identity with the bin contigs) showed identity values >99% identity indicating that the four jellyfishes carried the same clonal populations.

Discussion:

A recently published pilot study based on 16S rRNA gene amplicon pyrosequencing indicated that *C. tuberculata* organs in the gastric cavity (gastric filament and gonad tissues) are inhabited by low diversity microbiomes (Cortés-Lara *et al*., 2015). By a combination of direct sequencing metagenomics and microscopic phylotype identification, we not only corroborated this finding but generated a large dataset allowing us to predict the genomic potential of four candidate taxa and define their taxonomic affiliations, as well as to hypothesize about their putative ecological niches and functional roles. Interestingly, the dominant recovered bin from all four specimens M1-M4 corresponded to a chlamydial population closely related to *S. negevensis* Z (Everett *et al*., 1999), which presence of very low in the earlier PCR-based 16S rRNA gene analyses (Cortés-Lara *et al*., 2015). In addition, *Tenacibaculum*-like, *Spiroplasma*-like and *Mycoplasma*-like bins that had already been detected in our previous study (Cortés-Lara *et al*., 2015), were retrieved in the metagenomics analysis.. It is not unusual that metagenomics retrieves phylotypes not detected by 16S rRNA gene amplicon sequencing, since this method is not submitted to specific primer amplification biases and the impact of high ribosomal operon copy numbers (Logares *et al*., 2014 Rubio-Portillo et al., 2016).

The combination of comparative metagenomics and phylotype identification by CARD-FISH yields data on genomic potential, taxonomy and ecology sufficient for the definition of *Candidatus* taxa (Konstantindis and Rosselló-Móra, 2015). In this study, this approach was applied to the three recovered, novel phylotypes. The most frequently retrieved sequences originated from a new *Simkania*-like lineage of the phylum *Chlamydiae*, affiliated with the candidate genus '*Candidatus* Syngnamydia' (Fehr *et al*., 2013), in particular with '*Candidatus* Syngnamydia salmonis' (Nylund *et al*., 2015), with 16S rRNA gene identities appropriate for describing a new candidate species. Surprisingly, the *Simkania*-like cells appeared to inhabit non-jellyfish cells only endocellularly. These eukaryotic cells could be identified by CARD-FISH as an uncultured ciliate affiliated with the *O. flexilis – P. indiensis* clade, which was morphologically and phylogenetically similar to *Onychodromopsis flexilis* (Schmidt *et al*., 2007). The respective ciliate 18S rRNA was most closely related to an environmental clone that originated from an 18S rRNA gene sequencing survey of the Antarctic moss *Leptobryum* sp. (Nakai *et al*., 2012), and it is tempting to speculate on the role of the ciliate in the gastric cavity. Its detection in high abundances in the specimens is similar in amounts to the dinoflagellate *Symbiodinium* (Davy *et al*., 2012) in the same samples. This dinoflagellate seems to play an important role in the survival of cnidarians by supplying nutrients and substrates from primary production. However, it was also demonstrated that *Symbiodinium* cells could have a predatory role (Jeong *et al*., 2012). The *Onychodromopsis*-like ciliate detected here in the gastric cavities of *C. tuberculata* could also help in controlling the free-living microbial populations in the gastric

cavity through grazing. *C. tuberculata* feeds on small copepods (Kikinger, 1992), fragments of which could be detected microscopically, and ciliate grazing could reduce bacterial competition for this food source.

It is even more difficult to predict the role of the *Simkania*-like chlamydiae within the ciliates. The life cycle of chlamydiae is known to involve two stages (Collingro *et al*., 2011) of which we only found intracellular reticulate bodies, since no extracellular elementary bodies could be detected. However, and according to the bin dominance in the metagenomes, the calculated cell abundances of this organism in the gastric cavity were nearly two orders of magnitude higher than those of free-living bacteria. Although chlamydiae in *C. tuberculata* cells were not detected, we cannot rule out that they could infect other organisms, as demonstrated for the close relative *S. negevensis* (Kahane *et al*., 2002), or exhibit a potential pathogenic capability as shown for the species of '*Candidatus* Syngnamydia' (Fehr *et al*., 2013; Nylund *et al*., 2015). Our metagenomic data indeed provided a similar repertoire of virulence factor genes as predicted for this close relative (Collingro *et al*., 2011). However, the association could also be symbiotic, as hypothesized for another and even closer relative found in *Xenoturbella* individuals (Israelsson, 2007). The genomes of the *Simkania*-like bins had a size and a genetic repertoire that agreed with a versatile lifestyle, tuned toward unicellular hosts able to cope with changing environmental conditions, adapted to the metabolic host dependency (Collingro *et al*., 2011). Therefore, we speculate that the relationship between the chlamydial cells and their host may be mutually beneficial, and thus this case would represent a nested symbiosis of the *Simkania*-like cells with the *Onychodromopsis*-like ciliates in *C. tuberculata*.

Besides the *Simkania*-like organisms, our metagenomic data revealed the presence of at least three additional major components. Similar to that observed in *A. aurita* (Weiland-Bräuer et al., 2015), we could detect in one metagenome the presence of a member of the *Mollicutes*, loosely affiliated to the *Mycoplasma* genus in sufficient abundance to bin a 74% complete genome. The *A. aurita* mycoplasma affiliated with a different branch than the *C. tuberculata* bin based on initial comparison of short sequences (data not shown). Therefore, since CARD-FISH visualization also failed, we decided to refrain from describing a candidate taxon in this instance.

Three *Spiroplasma*-like bins were obtained from metagenomes M1, M3 and M4, although the respective 16S rRNA gene reads could also be retrieved from the specimen M2 (Cortés-Lara *et al*., 2015). Visualization of the respective phylotype by CARD-FISH was successful only in the cytoplasm of *C. tuberculata* cells obtained from the female specimen C2. These inconsistencies suggested that the *Spiroplasma*-like phylotype might have been present in rather low concentrations. *Spiroplasma* has been reported to occur mainly extracellularly, colonizing insect guts and plant surfaces, generally not causing any pathogenic effect, but controlling the sex ratio in some of them (Anbutsu and Fukatsu, 2011), and generating a range of symptoms in the infected plants (Regassa, 2014). Our bin was monophyletic with *Spiroplasma sensu stricto* and the 16S rRNA gene identity was <94.4%, indicating that it could represent a new candidate

genus (Rosselló-Móra and Amann, 2015). The estimated genome size of 0.7 Mb was smaller than any of the currently known genomes of *Spiroplasma* spp. (Regassa, 2014). It is tempting to speculate that this small genome is the result of an intracellular lifestyle, which was indicated by the microscopic observations of such cells apparently in the cytoplasm of *C. tuberculata* C2. An intracellular life style is also common for most mycoplasmas *sensu lato* (May *et al*., 2014). This hypothesis is also compatible with the predicted anaerobic metabolism encoded by this bin. The healthy appearance of all captured specimens and the visualization in one younger example captured in 2015 leads to the hypothesis that the intracellular *Spiroplasma*-like bacteria are commensals of *C. tuberculata* rather than pathogens, as seems to happen with insects (Regassa, 2014). In any case, the occurrence of the same phylotype in specimens separated by a four-year interval is probably not random and we therefore describe a candidate genus and candidate species in this specific case.

The *Tenacibaculum*-like phylotype was detected in all medusae. However, only the M1 bin was of high quality and the genome was predicted to be 92.8% complete. In this case, the 16S rRNA gene identity with members of the same monophyletic branch (always <97% with *T. aestuarii* as the closest species) indicated that it would represent a new species of this genus. All 20 species of *Tenacibaculum* hitherto classified (http://www.bacterio.net/tenacibaculum.html) have been isolated from marine environments, most of them from water (five species) or sediment samples (four species), but others (*T. adriaticum and T. crassostreae*) were associated with apparently healthy oysters (Lee *et al*., 2009), bryozoan (Heindl *et al*., 2008) or sponges and green algae (Suzuki *et al*., 2001). Only the three species *T. dicentrarchi*, *T. discolor* and *T. soleae* have been isolated from diseased marine fauna (Piñero-Vidal *et al*., 2012, Wang *et al*., 2008). Conspicuously, we could find at least two orders of magnitude higher free-living *Tenacibaculum*-like cells forming colony-like structures within the mesogleal axis within the gastric filaments than free swimming in the gastric cavity content. Despite the frequency of carbohydrate and protein catalysis genes dos not represent a strong enrichment compared to the average free-living bacterium genome with a similar genome size (Konstantinidis and Tiedje, PNAS 2004), the balanced gene repertoire, with >100 genes in total devoted to these two functional categories, it cannot be discarded that part of the degraded polymers could be a source of carbon and energy for the host jellyfish. This fact leads us to speculate that there is a specific commensal or symbiotic association with the host given its conspicuous abundance in the mesogleal that ultimately connect with all jellyfish tissues to provide their carbon and energy sources.

In summary, the gastric cavity of the scyphozoan jellyfishes can be understood as a simple model for microbial – animal digestive association with very structured microniches within the gastric community. The metagenomic approach combined with fluorescence microscopy and probe design allowed us to show that the microbiome associated with the organs of the *C. tuberculata* gastric cavity (gastric filament and gonad tissues, and gastric fluids) was very reduced, and to propose hypotheses concerning the functional role of each abundant member of this microbiome. The described species here appeared to be clonal between exemplars and

highly enriched by at least one order of magnitude compared to the accompanying low-abundance microbial community members, probably acquired through the mouth openings (Kikinger, 1992). Most bacteria seemed to have an intracellular lifestyle and may have established a cooperative relationship with their host. It was remarkable that the most relevant microorganism was a putative endosymbiont of a ciliate that was probably symbiotic with the jellyfish, which could be understood as a nested symbiosis. The content of the gastric cavity was mostly dominated by eukaryotic cells (single and aggregated *C. tuberculata*, *Onychodromopsis*-like and *Symbiodinium* cells) and their infecting bacteria (*Spiroplasma*-like and *Simkania*-like). These findings, together with the high abundance of the *Tenacibaculum*-like bacteria associated to the mesogleal axis of the gastric filaments, could be responsible for the digestion of ingested planktonic copepods and bacteria, which ultimately would be a source of carbon and energy for the jellyfish organs through the connected mesogleal tissues. One striking finding is that the *Simkania*-like and the *Tenacibaculum*-like phylotypes described here were also detected in other marine invertebrates as corals, gorgonian and sponges of the Mediterranean (Rubio-Portillo et al., unpublished results), as well as in a recent survey of "global sponge microbiome" (Thomas et al., 2016) that together, with the presence of mycoplasmas and spiroplasmas in other jellyfishes (Daniels and Breitbart, 2012; Vega-Orellana, 2014; Weiland-Bräuer et al., 2015), molluscs (Duperron *et al*., 2013) corals (Kimes *et al*., 2013), point to a common "metamicrobiome" of some marine invertebrates.

For the newly detected taxa based on the combination of metagenome binning and CARD-FISH visualization we therefore propose the new '*Candidatus* Syngnamydia medusae' sp. nov., '*Candidatus* Medusoplasma mediterranei' gen. nov., sp. nov., and '*Candidatus* Tenacibaculum medusae' sp. nov.

Description of the new candidate taxa:

***'Candidatus* Syngnamydia medusae' sp. nov.**

Syngnamydia medusae (me.du'sae. N.L. gen. n. medusae of a jellyfish).

Cells appear associated to a ciliate, closely related to *Onychodromopsis* sp., found in the digestive cavity of *Cotylorhiza tuberculata*. Intracellular reticular bodies in abundances >20 units per cell show a pleomorphic shape of 0.6 µm located in the cytoplasm of the host. The 16S rRNA gene sequence (LT599036) affiliates with the members of the genus '*Candidatus* Syngnamydia', and the cells can be visualized with the Simk174 probe (5'-CCGGACCTCCTCATTCGG-3') targeting the small ribosomal subunit. The metagenome bin (ERZ325008) showed a genome with a size of 2.1 Mb, and a G+C mol content of 39.8%. This genome bin is the type material for the new taxon. The protologue has been submitted to the Digital Protologue database (http://imedea.uib-csic.es/dprotologue/) under the Taxonumber CA00009.

***'Candidatus* Medusoplasma' gen. nov.**

Medusoplasma (Me.du.so.plas'ma. Gr. fem. n. *Medusa* a Gorgon in Greek mythology; N.L. pl. n. *Medusozoa* jellyfish; Gr. neut. n. *plasma*, anything formed or molded, image, figure; N.L. neut. n. *Medusoplasma*, *Spiroplasma*-like organisms living in jellyfishes).

The bacteria appear associated with cnidarian eukaryotic cells. The 16S rRNA gene (LT599042) loosely affiliates with the phylogenetic branch harboring the *Spiroplasma sensu stricto* genus. The type species of the genus is '*Candidatus* Medusoplasma marinum'.

***'Candidatus* Medusoplasma mediterranei' sp. nov.**

Medusoplasma mediterranei (N.L. gen. n. *me.di.ter.ra.ne'i,* from L. *Mediterraneum mare*, of the Mediterranean Sea)

Bacteria appear associated to *Cotylorhiza tuberculata* cells. Cells appear pleomorphic in shape and number >20 units per cell. The 16S rRNA gene (LT599042) loosely affiliates these organisms with the phylogenetic branch harboring the *Spiroplasma sensu stricto* genus, and the cells can be visualized with the Spiro199 probe (5'-TCTTTAGCGACGCAAACG-3') targeting the small ribosomal subunit. The metagenome bin (ERZ325010) showed a genome with a size of 0.68 Mb, and a G+C mol content of 23.5%. This genome bin is the type material for the new taxon. The protologue has been submitted to the Digital Protologue database (http://imedea.uib-csic.es/dprotologue/) under the Taxonumber CA00018.

***'Candidatus* Tenacibaculum medusae' sp. nov.**

Tenacibaculum medusae (me.du'sae. N.L. gen. n. medusae of a jellyfish).

Members of this species show rod-shaped cells with a length of approximately 5 µm occurring in the mesogleal axis of the gastric filaments of *Cotylorhiza tuberculata*. The 16S rRNA gene (LT599038) affiliates these organisms with the phylogenetic branch harboring most of the *Tenacibaculum* species hitherto classified, and the cells can be visualized with the Tena1432 probe (5'-CCTCACGGTAACCGACTT-3') targeting the small ribosomal subunit. The metagenome bin (ERZ325011) showed a genome with a size of 3 Mb, and a G+C mol content of 31.5%. This genome bin is the type material for the new taxon. The protologue has been submitted to the Digital Protologue database (http://imedea.uib-csic.es/dprotologue/) under the Taxonumber CA00019.

Experimental procedures:

Jellyfish specimens

Four specimens of *C. tuberculata* (M1 to M4) were captured in September 2013 (Cortés-Lara *et al*., 2015), four (C1 to C4) in September 2015, and two (C5-C6) in September 2016 (Table 1) using a landing net. In all cases, sampling was carried out in Alcúdia Bay, in the north of the Island of Mallorca, approximately 2 miles from the shore (39º45'00''N - 3º13'10''E in 2013; 39º45'37''N - 3º16'13''E in 2015; and 39º45'37''N - 3º16'13''E in 2016). All specimens were swimming close together in an area of approximately 10 m^2, and 0.5 m below the surface. Gastric cavity organs (Supplementary Figure S1) containing the gastric filaments tightly bound to the gonads were excised using a sterile scalpel, and the fluid gastric content in the stomach was collected with a syringe. The biomass was collected and either stored at -80ºC for DNA extraction purposes, or fixed with formaldehyde for CARD-FISH.

DNA extraction and genome assembly analysis

DNA extraction from the gastric cavity organs of the M1-M4 jellyfishes was performed as detailed by Cortés-Lara *et al*. (2015), and the same samples were used for the metagenome library construction. As indicated in the original paper, the DNA originated from disaggregated gastric filament and gonad tissues removing by mild centrifugation tissue debris. DNA sequencing libraries were prepared using the Illumina Nextera XT DNA library prep kit according to the manufacturer's instructions, except that the protocol was terminated after isolation of cleaned double-stranded libraries. Library concentrations were determined by fluorescent quantification using a Qubit HS DNA kit and Qubit 2.0 fluorometer (ThermoFisher Scientific formerly Life Technologies) and samples were run on a High Sensitivity DNA Chip using the Bioanalyzer 2100 instrument (Agilent) in order to determine the library insert sizes. An equimolar mixture of the libraries (10 pM) was sequenced using a MiSEQ reagent v2 kit for 500 cycles (2 x 250 bp paired end run) on an in-house Illumina MiSeq instrument (Georgia Institute of Technology) running the MiSEQ control software v2.4.0.4 (MCS). Adapter trimming and demultiplexing of sequenced samples was carried out by the MCS.

Sequence trimming, assembly, binning and annotation.

Metagenomic reads were trimmed using SolexaQA (Cox *et al*., 2010) using a threshold quality of 20 and discarding sequences shorter than 50 bp after trimming. Nonpareil (Rodriguez-R and Konstantinidis, 2014) was used to estimate the coverage of the community sampled by each metagenomic dataset with default parameters. *De novo* assemblies of trimmed reads were performed using the IDBA assembler (Peng *et al*., 2012) with the "–pre_correction" option. MaxBin software (Wu *et al*., 2014) was used to recover draft genomes (bins) from metagenomes. Contigs longer than 1,000 bp were grouped in bins using default parameters. The quality of the recovered bins (completeness and contamination) was analyzed with the CheckM tool (Parks *et al*., 2014), which was also used to identify lineage-specific marker genes in each bin. Bins showing high values of contamination and heterogeneity were submitted to a

second round of binning with the objective of separating the different populations. For this, reads mapping to contigs with identities higher than 98% were identified, re-assembled using IDBA and re-binned using MaxBin. The resulting bins were further checked for improved quality, and manual removal of dubious contigs (e.g. contigs with unusual coverage and/or phylogenetic origin; also see below). Genes were predicted on contigs longer than 1,000 bp using MetaGeneMark.hmm (Zhu *et al*., 2010) and functions were annotated using the RAST Server (Rapid Annotations using Subsystems Technology; Aziz *et al*., 2008). Genes from each bin were annotated using JCoast software (Richter *et al*., 2008) to flag and remove possible contigs from the eukaryotic domain. Abundance of recovered draft genomes was assessed by recruiting metagenomic reads against the corresponding bin sequence using a 98% identity cutoff for a match. The ANI (average nucleotide identity) calculations between bins and the closest relative genomes present in the databases were performed using the JSpecies WS online program (Richter *et al*., 2015) and AAI (average amino acid identity) calculations using the webserver available through http://enve-omics.gatech.edu/ (Rodriguez-R 2016).

Metabolic modeling was based on RAST annotation as well as KAAS-KEGG (Moriya *et al*. 2007). Complete pathways were identified using MinPath (Ye and Doak, 2009). Cell-compound fluxes based on gene prediction were calculated using the APP "Gapfill Metabolic Model" included in the DOE Systems Biology Knowledgebase server (KBase web-based interface). Eukaryotic domains in predicted proteins were performed as described by Collingro *et al*. (2011).

Tree reconstructions based on 16S and 23S rRNA genes.

16S rRNA short fragments from raw metagenome trimmed reads were extracted using Parallel-META v2.4 software. Single reads were aligned using the SINA tool (SILVA Incremental Aligner; Pruesse *et al*., 2012) implemented in ARB software and inserted by parsimony in the non-redundant SILVA SSURef_NR99_123 database (Cortés-Lara *et al*., 2015). In addition, the almost-complete 16S and 23S rRNA genes recovered from the draft population genomes (bins) were extracted using the RNAmmer 1.2 Server (Lagesen *et al*., 2007). The genes were added by parsimony to the SILVA SSURef_NR99_123 and LSURef_123 databases, respectively, in order to recognize the closest relatives. The almost complete gene sequences were used to reconstruct *de novo* trees using the neighbor-joining algorithm with the Jukes-Cantor correction together with the selected close relative sequences from the SILVA database (Quast *et al*., 2013). All trees were reconstructed using the ARB program package (Ludwig *et al*., 2004)

CARD-FISH and probe design.

Disintegrated gastric filaments and gonads, gastric fluids and intact gastric filaments were fixed by suspending 0.5 g in 1 mL of 1x PBS with 4% formaldehyde and storing for 16 h at 4ºC. After fixation, the samples were centrifuged for 5 min at 13,000 rpm (excepting the filaments that

were just collected with tweezers) and washed twice with sterile 1x PBS, and finally stored in 1 mL ethanol : 1x PBS (1:1). For CARD-FISH, 50 µL of the fixed biomass were dispersed in 10 mL 1x PBS and filtered onto 0.2 µm pore size filters (GTTP) of 47 mm diameter. Single small pieces of the filter were used for hybridization. CARD-FISH was performed using standard protocols (Pernthaler *et al*., 2002), and the commonly used probes for Eukarya (EUK516; Amann *et al*., 1990), Bacteria (EUB338-I; Amann *et al*., 1990); EUB338-II and –III (Daims *et al*., 1999), Archaea (ARCH915; Raskin *et al*., 1994), negative control (NON338; Wallner *et al*., 1993) and *Bacteroidetes* CF319a (Manz *et al*., 1996) were used with the recommended stringency conditions.

In addition, oligonucleotide sequences targeting the major groups detected in this study were designed for CARD-FISH purposes, and they are listed in Table 3. In all cases, except probes Onic1121, HOnic1103 and HOnic1136, the design was carried out using the almost full-length 16S rRNA gene sequences obtained after the assembly procedure. The other three probes mentioned were designed against the closest relative almost full-length sequences available in the SILVA SSURef_NR99_123 that affiliated together with the single short reads obtained from the raw metagenomic data. For *C. tuberculata*, the best matching probe also hybridized with three additional *Rhizostomeae* members (*Cassiopea sp. Rhizostoma sp* and *Phyllorhiza sp*.), but no other non-scyphozoan organism. Probe design was carried out using the ARB program with default conditions, not allowing any unspecific match using the updated SILVA SSURef_NR99_123 database. The probe specificity was re-checked against the SSURef_NR99_128 of the SILVA database after the update in September 28nd, 2016, and prior to the submission of the manuscript (Quast *et al*., 2013; Supplementary Figure S6).

In all cases, the probes were synthesized with an HRP label at the 5' end (Biomers.net GmbH, Germany), except for the helper oligonucleotides HSpiro181 and HSpiro217 (helpers of Spiro199), HMyco720 and HMyco756 (helpers of 738), and HOnic1103 and HOnic1136 (helpers of Onic1121). These only facilitated the access of the probe to the corresponding target sites (Fuchs *et al*., 2000). As only the samples themselves were containing the target organisms, positive signals were considered when matching expected morphologies and locations (free-living or cell-associated), and contrasted with the EUB338 and NON338 probes as positive and negative controls respectively. The stringency of the new probes (Table 3) was optimized by hybridizing the samples with different formamide concentrations, choosing the highest stringency exhibiting the best fluorescence intensity.

The cells were counterstained with 1 mg mL^{-1} 4,6-diamidino-2-phenylindol (DAPI) for 10 min at RT in the dark. For microscopy counts, filter pieces were embedded in a 4:1 mixture of low fluorescence glycerol mountant (Citifluor AF1, Citifluor Ltd., London, United Kingdom) and Vecta Shield mounting fluid (Vecta Laboratories, Burlingame, CA, USA). Samples were observed with an Axioskop 2 mot plus epifluorescence microscope equipped with 100x/1.4 and 40x/1.3 plan-apochromatic oil immersion objectives (Carl Zeiss, Germany), and filter sets F36-499, F36-525 and F46-006. Image acquisition was carried out using an AxioCam MRm CCD

camera (Carl Zeiss, Germany). For super-resolution, structured illumination microscopy was used with an ELYRA PS.I microscope and a 63x/1.4 plan-apochromatic oil immersion objective (Carl Zeiss, Germany). Raw image acquisition was carried out with an iXon885 EM-CCD camera (Andor, Belfast) using five phase shifts and three rotations. DAPI staining was visualized with a 405 nm laser (50 mW) and an emission filter with a band width of 420-480 nm, Alexa488 used a 488 nm laser (100 mW) and a band pass filter 502-538 nm, and autofluorescence was recorded with a 561 nm laser (100 mW) and a 573-613 nm band pass filter. The image calculation was performed using the structured illumination algorithm of the ZEN software package (Carl Zeiss, Germany).

Acknowledgements:

The authors want to thank Antonio Rosselló Nadal for lending his boat and being our skipper for the sampling campaigns, to Laura Prieto for her help in interpreting the biology and anatomy of the jellyfishes, to Bernhard Fuchs, Jörg Wulf and Andreas Ellrott for their assistance in the fluorescence microscopy application, and to Carles Saus and Enrique Serra from the Son Espases Hospital for his help in the gastric filament microtome section's preparation. Esther Rubio-Portillo is also acknowledged for sharing her unpublished results, and Ana Belén Marín Cuadrado for her help in bioinformatics. We also thank Aharon Oren and Bernhard Schink for helping in the correctness of the new candidate taxa etymology. This research from RRM's group was partially supported by the Spanish Ministry of Economy projects CGL2012-39627-C03-03 and CLG2015_66686-C3-1-P, which were also supported with European Regional Development Fund (FEDER) funds. KTK's research was supported, in part, by the U.S. National Science Foundation (award No. 1241046). RRM acknowledges the economic support of grant PR2015-00008 included in the program Salvador de Madariaga of the Ministry of Education, Culture and Sports in order to undertake a research stay at the MPI-MM in Bremen. TVP acknowledges the predoctoral fellowship of the Ministerio de Economía y Competitividad of the Spanish Government for the FPI fellowship (Nr BES-2013-064420) supporting his research activities.

The authors declare that there is no conflict of interests to declare in relation with the current manuscript.

References:

Amann, R.I., Binder, B.J., Olson, R.J., Chisholm, S.W., Devereux, R., Stahl, D.A. (1990) Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. Appl Environ Microbiol **56:** 1919-1925.

Anbutsu, H., Fukatsu, T. (2011) Spiroplasma as model insect endosymbiont. Environ. Microbiol. Rep. **3**: 144-153

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., *et al.* (2008) The RAST Server: rapid annotations using subsystems technology. BMC Genomics **9:** 75.

Collingro, A., Tischler, P., Weinmaier, T., Penz, T., Heinz, E., Brunham, R.C., *et al.* (2011) Unity in variety – the pan-genome of the *Chlamydiae*. Mol Biol Evol **28:** 3253-3270.

Cortés-Lara, S., Urdiain, M., Mora-Ruiz, M., Prieto, L., Rosselló-Móra, R. (2015) Prokaryotic microbiota in the digestive cavity of the jellyfish *Cotylorhiza tuberculata*. Syst Appl Microbiol **38:** 494-500.

Cox, M.P., Peterson, D.A., Biggs, P.J. (2010) SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics **11:** 485.

Daims, H., Bruhl, A., Amann, R., Schleifer, K.H., Wagner, M. (1999) The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set. Syst Appl Microbiol **22:** 434-444.

Daniels, C., Breitbart, M. (2012) Bacterial communities associated with the ctenophores *Mnemiopsis leidyi* and *Beroe ovata*. FEMS Microbiol Ecol **82:** 90-101.

Davy, S.K., Allemand, D., Weis, V.M. (2012) Cell biology of the cnidarian-dinoflagellate symbiosis. Microbiol Mol Biol Revs **76:** 229-261.

Delannoy, C.M.J., Houghton, J.D.R., Fleming, N.E.C., Ferguson, H.W. (2011) Mauve stingers (*Pelagia noctiluca*) as carriers of the bacterial fish pathogen *Tenacibaculum maritimum*. Aquaculture **311:** 255-257.

Dinasquet, J., Granhag, L., Riemann, L. (2012) Stimulated bacterioplankton growth and selection for certain bacterial taxa in the vicinity of the ctenophore *Mnemiopsis leidyi*. Frontiers Microbiol **3:** 302.

Duperron, S., Pottier M.A., Léger, N., Gaudron, S.M., Pullandre, S., Le Prieur, S., *et al.* (2013) A tale of two chitons: is habitat specialization linked to distinct associated bacterial communities? FEMS Microbiol Ecol **83:** 552-557.

Everett, K.D.E., Bush, R.M., Andersen, A.A. (1999) Emended description of the order *Chlamydiales*, proposal of *Parachlamydiaceae* fam. nov. and *Simkaniaceae* fam. nov., each containing one monotypic genus, revised taxonomy of the family *Chlamydiaceae*, including a new genus and five new species, and standards for the identification of organisms. Int J Syst Bacteriol **49:** 415-440.

Fehr, A., Walther, E., Schmidt-Posthaus, H., Nufer, L., Wilson, A., Svercel, M., *et al.* (2013) *Candidatus* Syngnamydia venezia, a novel member of the phylum *Chlamydiae* from the broad nosed pipefish *Syngnathus typhle*. PLoS One **8:** e70853.

Ferguson, H.W., Delannoy, C.M.J., Hay, S., Nicholson, J., Sutherland, D., Crumlish, M. (2010) Jellyfish as vectors of bacterial diseases for farmed salmon (*Salmo salar*). J Vet Diagn Invest **22:** 376-382.

Fringuelli, E., Savage, P.D., Gordon, A., Baxter, E.J., Rodger, H.D., Graham, D.A. (2012) Development of a quantitative real-time PCR for the detection of *Tenacibaculum maritimum* and its application to field samples. J Fish Dis 35: 759-790.

Fuchs, B.M., Glöckner, F.O., Wulf, J., Amann, R. (2000) Unlabeled helper oligonucleotides increase the in situ accessibility to 16S rRNA of fluorescently labeled oligonucleotide probes. Appl Environ Microbiol **66:** 3603-3607.

Hao, W., Gerdts, G., Pepplies, J., Wichels, A. (2015) Bacterial communities associated with four ctenophore genera from the German Bight (North Sea). FEMS Microbiol Ecol **91:** 1-11.

Heindl, H., Wiese, J., Imhoff, J.F. (2008) *Tenacibaculum adriaticum* sp. nov, from a bryozoan in the Adriatic sea. Int J Syst Evol Microbiol **58:** 542-547.

Israelsson, O. (2007) Chlamydial symbionts in the enigmatic *Xenoturbella* (*Deuterostomia*). J Inv Path **96:** 213-220.

Jeong, H.J., Yoo, Y.D., Kang, N.S., Lim, A.S., Seong, K.A., Lee, S.Y., *et al.* (2012) Heterotrophic feeding as a newly identified survival strategy of the dinoflagellate *Symbiodinium*. Proc Natl Acad Sc USA **109:** 12604-12609.

Kahane, S., Kimmel, N., Friedman, M.G. (2002) The growth cycle of *Simkania negevensis*. Microbiology **148:** 735-742.

Kikinger, R. (1992). *Cotylorhiza tuberculata* (*Cnidaria: Scyphozoa*) - life history of a stationary population. Mar Ecol **13:** 333-362.

Kimes, N.E., Johnson, W.R., Torralba, M., Nelson, K.E., Weil, E., Morris, P.J. (2013) The *Montastraea faveolata* microbiome: ecological and temporal influences on a Caribbean reef-building coral in decline. Environ Microbiol **15:** 282-2094.

Konstantinidis, K., Rosselló-Móra, R. (2015) Classifying the uncultivated microbial majority: a place for metagenomic data in the *Candidatus* proposal. Syst Appl Microbiol **38:** 223-230.

Lagesen, K., Hallin, P.F., Rødland, E.A., Stæerfeldt, H.H., Rognes, T., Ussery, D.W. (2007) RNAmmer: consistent annotation of rRNA genes in genomic sequences. Nucleic Acids Res **35:** 3100-3108.

Larsbrink, J., Rogers, T.E., Hemsworth, G.R., McKee, L.S., Tauzin, A.S., Spadiut, O., *et al.* (2014) A discrete genetic locus confers xyloglucan metabolism in select human in select human gut Bacteroidetes. Nature **506:** 498-502.

Lee, Y.S., Baik, K.S., Park, S.Y., Kim, E.M., Lee, D.H., Kahng, H.Y., *et al.* (2009) *Tenacibaculum crassostreae* sp. nov, isolated from the Pacific oyster, *Crassostrea gigas*. Int J Syst Evol Microbiol **59:** 1609-1614.

Logares, R., Sunagawa, S., Salazar, G., Cornejo-Castillo, F.M., Ferrera, I., Sarmento, H., *et al.* (2014) Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon

sequencing to explore diversity and structure of microbial communities. Environ Microbiol **16:** 2659-2671.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar., *et al.* (2004) ARB: a software environment for sequence data. Nucleic Acids Res **32:** 1363–1371.

Manz, W., Amann, R., Ludwig, W., Vancanneyt, M., Schleifer, K.H. (1996) Application of a suite of 16S rRNA-specific oligonucleotide probes designed to investigate bacteria of the phylum cytophaga-flavobacter-bacteroides in the natural environment. Microbiol **142:** 1097-1106.

May, M., Balish, M.F., Blanchard, A. (2014) The order *Mycoplasmatales*. The Prokaryotes – Firmicutes and Tenericutes. Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., Thompson, F. (eds). Berlin-Heidelberg: Springer, pp 515-550.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A., Kanehisa, M. (2007) KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res **35:** W182-W185.

Nakai, R., Abe, T., Baba, T., Imura, S., Kagoshima, H., Kanda, H., *et al.* (2012) Eukaryotic phylotypes in aquatic moss pillars inhabiting a freshwater lake in East Antarctica, based on 18S rRNA gene analysis. Polar Biol **35:** 1495-1504.

Nylund, S., Steigen, A., Karlsbakk, E., Plarre, H., Andersen, L., Klarsen, M., *et al.* (2015) Characterization of "*Candidatus* Syngnamydia salmonis" (Chlamydiales, Simkaniaceae), a bacterium associated with epitheoliocystis in Atlantic salmon (*Salmo salar* L.). Arch Microbiol **197:** 17-25.

Palmieri, M.G., Barausse, A., Luisetti, T., Turner, K. (2014) Jellyfish blooms in the northern Adriatic Sea: Fishermen's perceptions and economic impacts on fisheries. Fish Res **155:** 51-58.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W. (2014) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res **25:** 1043-1055.

Peng, Y., Leung, H.C., Yiu, S.M., Chin, F.Y. (2012) IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics **28:** 1420-1428.

Pernthaler, A., Pernthaler, J., Amann, R. (2002) Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria. Appl Environ Microbiol **68:** 3094–3101.

Piñeiro-Vidal, M., Gijón, D., Zarza, C., Santos, Y. (2012) *Tenacibaculum dicentrarchi* sp. nov, a marine bacterium of the family *Flavobacteriaceae* isolated from European sea bass. Int J Syst Evol Microbiol **62:** 425-429.

Pitt, A.K., Welsh, D.T., Condon, R.H. (2009) Influence of jellyfish on carbon, nitrogen and phosphorus cycling and plankton production. Hydrobiologia **616:** 133-149.

Prieto, L., Astorga, D., Navarro, G., Ruiz, J. (2010) Environmental control of phase transition and polyp survival of a massive-outbreaker jellyfish. PLoS One **5:** e13793.

Pruesse, E., Peplies, J., Glockner, F.O. (2012) SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics **28:** 1823–1829.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., *et al.* (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res **41:** D590-596.

Raskin, L., Stromley, J.M., Rittmann, B.E., Stahl, D.A. (1994) Group-specific 16S rRNA hybridization probes to describe natural communities of methanogens. Appl Environ Microbiol **60:** 1232-1240.

Regassa, L.B. (2014) The family Spiroplasmataceae. The Prokaryotes – Firmicutes and Tenericutes. Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., Thompson, F. (eds). Berlin-Heidelberg: Springer, pp. 551-567.

Richardson, A.J., Bakun, A., Hays, G.C., Gibbons, M. J. (2009). The jellyfish joyride: causes, consequences and management responses to a more gelatinous future. Trends Ecol. Evol. **24**: 312-322

Richter, M., Lombardot, T., Kostadinov, I., Kottmann, R., Duhaime, M.B., Peplies, J., Glöckner, F.O. (2008) JCoast – a biologist – centric software tool for data mining and comparison of prokaryotic (meta)genomes. BCM Bioinformatics **9:** 177.

Richter, M., Rosselló-Móra, R., Glöckner, F.O., Pepplies, J. (2015) JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics. **32:** 929-931.

Rodriguez-R, L.M., Konstantinidis, K.T. (2014) Nonpareil: a redundancy-based approach to assess the level of coverage in metagenomic datasets. Bioinformatics **30:** 629-635.

Rodriguez-R, L.-M., and Konstantinidis, K. T. (2016). "The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes." PeerJ Preprints(e1900v1).

Rosselló-Móra, R., Amann, R. (2015) Past and future species definitions for *Bacteria* and *Archaea*. Syst Appl Microbiol **38:** 209-216.

Rubio-Portillo, E., Santos, F., Martínez-García, M., de Los Ríos, A., Ascaso, C., *et al.* (2016) Structure and temporal dynamics of the bacterial communities associated to microhabitats of the coral *Oculina patagonica*. Environ Microbiol **18:** 4564-4578.

Schmidt, S.L., Bernhard, D., Schlegel, M., Foissner, W. (2007) Phylogeny of the *Stichotrichia* (*Ciliophora; Spirotrichea*) reconstructed with nuclear small subunit rRNA gene sequences: discrepancies and accordances with morphological data. J Eukaryot Microbiol **54:** 201-209.

Suzuki, M., Nakagawa, Y., Harayama, S., Yamamoto, S. (2001) Phylogenetic analysis and taxonomic study of marine *Cytophaga*-like bacteria: proposal for *Tenacibaculum* gen. nov.

with *Tenacibaculum maritimum* comb. nov. and *Tenacibaculum ovolyticum* comb. nov, and description of *Tenacibaculum mesophilum* sp. nov, and *Tenacibaculum amylolyticum* sp. nov. Int J Syst Evol Microbiol **51:** 1639-1652.

Thomas, T., Moitinho-Silva, L., Lurgi, M., Björk, J.R., Easson, C., Astudillo-García, C., et al. (2016) Diversity, structure and convergent evolution of the global sponge microbiome. Nat Commun **7**:11870. doi: 10.1038/ncomms11870

Tinta, T., Kogovsek, T., Malej, A., Turk, V. (2012) Jellyfish modulate bacterial dynamic and community structure. PLoS One. **7:** e39274.

Tinta, T., Malej, A., Kos, M., Turk, V. (2010) Degradation of the Adriatic medusa *Aurelia* sp. by ambient bacteria. Hydrobiologia **645:** 179-191.

Utne-Palm, A.C., Salvanes A.G.V., Currie, B., Kaartvedt, S., Nilsson, G.E., Braithwaite, V.A., Stecyk, J.A.W., Hundt, M., van der Bank, M., Flynn, V., Sandvik, G.K., Klevjer, T.A., Sweetman, A.K., Brüchert, V., Pittman, K., Peard, K.R., Lunde, I.G., Strandabø, R.A.U., Gibbons, M.J. (2010). Trophic structure and community stability in an overfished ecosystem. Science. **329**: 333-336

Vega-Orellana, O.M. 2014 Estudio de microorganismos de la clase Mollicutes en organismos marinos, Ph.D. thesis of the University of Las Palmas de Gran Canarias. (http://hdl.handle.net/10553/12198)

Wallner, G., Amann, R., Beisker, W. (1993) Optimizing fluorescent *in situ* hybridization with rRNA-targeted oligonucleotide probes for flow cytometric identification of microorganisms. Cytometry **14:** 136-143.

Wang, J.T., Chou, Y.J., Chou, J.H., Chen, C.A., Chen, W.M. (2008) *Tenacibaculum aiptasiae* sp. nov, isolated from a sea anemone *Aiptasia pulchella*. Int J Syst Evol Microbiol **58:** 761-766.

Weiland-Bräuer, N., Neulinger, S.C., Pinnow, N., Künzel, S., Baines, J.F., Schmitz, R.A. (2015) Composition of bacterial communities associated with *Aurelia aurita* changes with compartment, life stage, and population. Appl Environ Microbiol **81:** 6038-6052.

Wu, Y.W., Tang, Y.H., Tringe, S.G., Simmons, B.A., Singer, S.W. (2014) MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome **2:** 26.

Yamamoto, J., Hirose, M., Ohtani, T., Sugimoto, K., Hirase, K., Shimamoto, N., Shimura, T., Honda, N., Fujimori, Y., Mukai, T. (2008) **153**: 311-317

Yarza, P., Ludwig, W., Euzéby, J., Amann, R., Schleifer, K.H., Glöckner, F.O., Rosselló-Móra, R. (2010) Update of the all-species living tree project based on 16S and 23S rRNA sequence analysis. Syst Appl Microbiol **33:** 291-299.

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F.O., Ludwig, W., Schleifer, K.H., *et al.* (2014) Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nature Revs Microbiol **12:** 635-645.

Ye, Y., Doak, T. (2009) A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. PLoS Comput Biol **5:** e1000465.

Zhu, W., Lomsadze, A., Borodovsky, M. (2010) *Ab initio* gene identification in metagenomic sequences. Nucleic Acids Res **38:** e132.

Tables and Figure Legends

Table 1: *C. tuberculata* captured for the study, their weight and the positive or negative signals obtained on disintegrated gastric filaments using CARD-FISH with specific probes (w indicates very low abundances close to the limit of detection).

Specimen	Gender	Weight (K)	Eub338	Simk174	Tena1432	Spiro199	Capture date
M1	Male	1.75	+	+	+	-	Sept 2013
M2	Male	1.5	+	+	+	-	Sept 2013
M3	Female	2.0	+	+	+	-	Sept 2013
M4	Male	1.0	+	+	+	-	Sept 2013
C1	Female	0.65	+	+	+	-	Sept 2015
C2	Female	0.84	+	+w	+	+	Sept 2015
C3	Male	0.98	+	-	+	-	Sept 2015
C4	Female	0.83	+	+w	+	-	Sept 2015
C5	Female	1.11	+	+	+	-	Sept 2016
C6	Female	1.20	+	+	+	-	Sept 2016

Table 2: Number of short sequence reads extracted from each metagenome corresponding to 16S rRNA fragments, and their affiliation.

	Short individual reads				Percentage (%)			
	M1 ♂	M2 ♂	M3 ♀	M4 ♂	M1 ♂	M2 ♂	M3 ♀	M4 ♂
Simkania-like	2495	432	2373	870	56.2	35.4	47.1	52.3
Spiroplasma-like	1047	29	132	370	23.6	2.4	2.6	22.2
Tenacibaculum-like	471	4	45	67	10.6	0.3	0.9	4.0
Mycoplasma-like	16	222	2264	2	0.4	18.2	45.0	0.1
Other	257	83	140	80	5.8	6.8	2.8	4.8
Eukarya	116	348	68	145	2.6	28.5	1.4	8.7
Non-16S	38	104	11	129	0.9	8.5	0.2	7.8
Total	4440	1222	5033	1663	100	100	100	100

Table 3: Oligonucleotide probes designed for this study

Probe name	*E. coli* position	Probe sequence	Organism	GC%	Formamide optimum (%)
Simk174	174	CCGGACCTCCTCATTCGG	*Simkania*-like	66	35
Spiro199	199	TCTTTAGCGACGCAAACG	*Spiroplasma*-like	50	20
HSpiro181	181	CGTCTTTCAATTTCAAAT	*Spiroplasma*-like	27	20
HSpiro217	217	AATACGCCGCACCCCCAT	*Spiroplasma*-like	61	20
Myco738	738	ATGTCAGGAGTAGACCTG	*Mycoplasma*-like	50	n.d.
HMyco720	720	TTAGTCGCCTTCGCTATT	*Mycoplasma*-like	44	n.d.
HMyco756	756	CCGCGCTCTCATGCCTCA	*Mycoplasma*-like	66	n.d.
Tena1432	1432	CCTCACGGTAACCGACTT	*Tenacibaculum*-like	55	20
Cotu193	193	CGGAGCACACGTATTGGC	*C. tuberculata*	61	30
Cotu1453	1453	TCTCGGACTTCCATCTCC	*C. tuberculata*	50	35
Onyc1121	1121	ACGGGTCGACTAGTTAGC	Unc. *Oxytrichidae*	55	20
HOnyc1103	1103	AGGCTAAGGTCTCGTTCG	Unc. *Oxytrichidae*	55	20
HOnyc1136	1136	AGTCGTGCCCGCTTAGCA	Unc. *Oxytrichidae*	61	20

n.d.: no positive signal was detected with this probe, but since there was no clear positive probe to standardize the hybridization conditions it cannot be ruled out that this probe does not work

Table 4: Major bins retrieved from the metagenomes and their characteristics.

	Bin Id	Comp.	Cont.	Het.	Contigs	bp	%GC	N50	N90	Ns	Longest	rRNA	Average seq. depth	CDSs	Hypothetical CDSs	t-RNAs
SIMKANIA-LIKE																
M1	M1_Simk	96.3	0	0	55	2,155,307	39.8	131,345	22,527	0	247,103	1	550.9	2,074	851 (41%)	35
M2	M2_Simk	96.3	0	0	108	2,253,388	39.8	107,968	12,662	0	165,195	1	67.1	2,100	889 (42%)	35
M3	M3_Simk	96.3	0	0	48	2,144,529	39.8	107,968	21,957	0	247,103	1	652	2,058	869 (42%)	35
M4	M4_Simk	96.3	0	0	123	2,294,964	39.6	131,345	11,904	0	247,103	1	178.5	2,114	886 (42%)	35
SPIROPLASMA-LIKE																
M1	M1_Spiro	55.3	0.8	0	94	275,702	25.1	3,509	1,509	0	14,680	1				
M3	M3_Spiro	13.3	0	0	45	95,860	27.4	2,213	1,350	0	5,642	1				
M4	M4_Spiro	94.4	0.03	100	68	681,895	23.5	24,415	4,430	0	52,169	1	30.1	992	554 (56%)	29
TENACIBACULUM-LIKE																
M1	M1_Tena	92.8	0.7	66.7	695	3,027,267	31.5	5,951	1,983	0	27,708	1	27.6	2,608	1099 (42%)	44
M4	M4_Tena	29.4	0.4	0	599	925,760	31.7	1,528	1,072	0	4,181	1				
MYCOPLASMA-LIKE																
M3	M3_Myco	74.1	2.3	0	11	409,158	32.9	69,761	46,043	0	129,584	1	158.1			

Table 5: Genome features of the retrieved bins.

	M1_Simk	M2_Simk	M3_Simk	M4_Simk	M1_Tena	M4_Spiro
Chromosome size (Mb)	2.2	2.3	2.2	2.3	3.0	0.7
Contigs	55	108	48	123	695	68
Percentage completion (%)	96.3	96.3	96.3	96.3	92.8	74.1
Predicted CDSs	2,074	2,100	2,058	2,114	2,699	1,006
G+C mol%	39.8	39.8	39.8	39.8	31.5	23.5
Coding regions	2,036	2,062	2,019	2,072	2,649	915
Average CDS length (bp)	988.9	981.3	991.6	967.2	912.3	385.2
Hypothetical CDSs	851	858	815	858	944	545
Conserved hypothetical CDS	288	289	290	289	519	159
tRNAs	35	35	35	35	44	29
rRNA genes	3	3	3	3	6	2

Figure 1: Phylogenetic reconstruction based on a neighbor-joining algorithm showing the position of the almost complete 16S rRNA gene sequences. For the reconstruction, a 30% conservational filter was used. Bar indicates 10% sequence divergence. The database used was the LTP_123.

Figure 2: CARD-FISH micrographs of the gastric cavity content hybridized with the candidate microorganisms studied here. The scale bar is indicated in each micrograph:

A) M2 hybridized with Simk174 (green stain, specific for the *Simkania*-like cells) and DAPI staining (blue), ATO488 and Cy3 filter sets at 400x magnification using a fluorescence microscope. *Onychodromopsis*-like cells show green stained intracellular *Simkania*-like cells (a); *Symbiodinium*-like cells appear pink (b) and the *C. tuberculata* nucleus appears only DAPI stained.

B) M2 double hybridized with Simk174 (green stain, specific for the *Simkania*-like cells) and Cotu1453 (for *C. tuberculata* cells, staining green) and DAPI staining (blue) and ATO488 filter sets at 1000x magnification using the super-resolution microscope. The *C. tuberculata* cell (a) shows positive green hybridization in the cytoplasm around the nucleus. The *Onychodromopsis*-like cell (b) shows endocellular *Simkania*-like bodies of approximately 0.2 µm in size (also green labeled), as well as DAPI stained non-hybridized particles. The size of the nucleus of the eukaryotic cell is approximately 3x larger than that of *C. tuberculata* cells. The long arrow-shaped DAPI cells correspond to sperm cells (c).

C) C5 intact gastric filament hybridized with EUB338 and DAPI staining (blue) and ATO488 filter sets at 400x using fluorescence microscope. The image shows the presence of a *Symbiodinium*-like cell colony (a), *Onychodromopsis-like* cells with endocytoplasmatic *Simkania*-like bodies (b), and the free-living *Tenacibaculum*-like cells, all them green stained and located in the mesogleal axis inside the gastric filaments of *C. tuberculata* which nucleus are stained in blue with DAPI.

D) M1 hybridized with Simk174 (green stain, specific for the *Simkania*-like cells) and DAPI staining (blue) and ATO488 filter sets at 1000x magnification using the super-resolution

microscope. The *Onychodromopsis*-like cell shows endocellular *Simkania*-like cells (a; green labeled), as well as DAPI stained non-hybridized bodies (b; only blue). The size of the nucleus of the eukaryotic cell is approximately 3x larger than those of the *Symbiodinium*-like cells. The long arrow-shaped DAPI cells correspond to the sperm cells (c), and the red cell corresponds to a *Symbiodinium*-like cell (d).

E) C5 intact gastric filament hybridized with Tena1432 (green stain, specific for the *Tenacibaculum*-like cells) and DAPI (staining) and ATO488 filter sets at 400x magnification fluorescence microscope. *Tenacibaculum*-like cells form a dense colonization of the mesogleal axis within the filament and appear as green-stained rods of 5 µm length, and the filament cell nucleus appear DAPI stained.

F) M2 hybridized with Onyc1121 (green stain, specific for the *Onychodromopsis*-like cells) together with the unlabeled helper probes HOnic1103 and HOnic1136 observed with DAPI and ATO488 filter sets at 1000x magnification using the fluorescence microscope. The *Onychodromopsis*-like cells appear green stained.

G) The same photomicrograph as F, but only DAPI stained where the cells with larger nuclei show intracytoplasmic DAPI positive bodies (a).

H) C2 hybridized with Spiro199 (green stain, specific for the *Spiroplasma*-like cells) together with the unlabeled helper probes HSpiro1 and HSpiro2 observed with DAPI and ATO488 filter sets at 1000x magnification with a fluorescence microscope. *Spiroplasma*-like cells (a) appear as endocellular bodies in the cytoplasm of cells whose nuclei are more condensed and smaller than those of the *Onychodromopsis*-like cells.

I) M2 hybridized with Cotu193 (green stain, specific for the *C. tuberculata* cells) and observed with the DAPI and ATO488 filter sets at 400x magnification with a fluorescence microscope. *C. tuberculata* positive cells (a) appear with positive green hybridization in the cytoplasm around the nucleus. *Spiroplasma*-like carrying cells (b) appear with a condensed nucleus.

J) M1 hybridized with Cotu193 (green stain, specific for the *C. tuberculata* cells) observed with DAPI and ATO488 filter sets at 1000x magnification with a fluorescence microscope. The *C. tuberculata* positive cell (a) appears with positive green hybridization in the cytoplasm around the nucleus. The *Onychodromopsis*-like cells (b) are not ATO488 stained and have a slight DAPI signal visible in the endocellular cells.

Tables and Figure Legends

Table 1: *C. tuberculata* captured for the study, their weight and the positive or negative signals obtained on disintegrated gastric filaments using CARD-FISH with specific probes (w indicates very low abundances close to the limit of detection).

Specimen	Gender	Weight (K)	Eub338	Simk174	Tena1432	Spiro199	Capture date
M1	Male	1.75	+	+	+	-	Sept 2013
M2	Male	1.5	+	+	+	-	Sept 2013
M3	Female	2.0	+	+	+	-	Sept 2013
M4	Male	1.0	+	+	+	-	Sept 2013
C1	Female	0.65	+	+	+	-	Sept 2015
C2	Female	0.84	+	+w	+	+	Sept 2015
C3	Male	0.98	+	-	+	-	Sept 2015
C4	Female	0.83	+	+w	+	-	Sept 2015
C5	Female	1.11	+	+	+	-	Sept 2016
C6	Female	1.20	+	+	+	-	Sept 2016

Table 2: Number of short sequence reads extracted from each metagenome corresponding to 16S rRNA fragments, and their affiliation.

	Short individual reads				Percentage (%)			
	M1 ♂	M2 ♂	M3 ♀	M4 ♂	M1 ♂	M2 ♂	M3 ♀	M4 ♂
Simkania-like	2495	432	2373	870	56.2	35.4	47.1	52.3
Spiroplasma-like	1047	29	132	370	23.6	2.4	2.6	22.2
Tenacibaculum-like	471	4	45	67	10.6	0.3	0.9	4.0
Mycoplasma-like	16	222	2264	2	0.4	18.2	45.0	0.1
Other	257	83	140	80	5.8	6.8	2.8	4.8
Eukarya	116	348	68	145	2.6	28.5	1.4	8.7
Non-16S	38	104	11	129	0.9	8.5	0.2	7.8
Total	4440	1222	5033	1663	100	100	100	100

Table 3: Oligonucleotide probes designed for this study

Probe name	*E. coli* position	Probe sequence	Organism	GC%	Formamide optimum (%)
Simk174	174	CCGGACCTCCTCATTCGG	*Simkania*-like	66	35
Spiro199	199	TCTTTAGCGACGCAAACG	*Spiroplasma*-like	50	20
HSpiro181	181	CGTCTTTCAATTTCAAAT	*Spiroplasma*-like	27	20
HSpiro217	217	AATACGCCGCACCCCCAT	*Spiroplasma*-like	61	20
Myco738	738	ATGTCAGGAGTAGACCTG	*Mycoplasma*-like	50	n.d.
HMyco720	720	TTAGTCGCCTTCGCTATT	*Mycoplasma*-like	44	n.d.
HMyco756	756	CCGCGCTCTCATGCCTCA	*Mycoplasma*-like	66	n.d.
Tena1432	1432	CCTCACGGTAACCGACTT	*Tenacibaculum*-like	55	20
Cotu193	193	CGGAGCACACGTATTGGC	*C. tuberculata*	61	30
Cotu1453	1453	TCTCGGACTTCCATCTCC	*C. tuberculata*	50	35
Onyc1121	1121	ACGGGTCGACTAGTTAGC	Unc. *Oxytrichidae*	55	20
HOnyc1103	1103	AGGCTAAGGTCTCGTTCG	Unc. *Oxytrichidae*	55	20
HOnyc1136	1136	AGTCGTGCCCGCTTAGCA	Unc. *Oxytrichidae*	61	20

n.d.: no positive signal was detected with this probe, but since there was no clear positive probe to standardize the hybridization conditions it cannot be ruled out that this probe does not work

Table 4: Major bins retrieved from the metagenomes and their characteristics.

	Bin Id	Comp.	Cont.	Het.	Contigs	bp	%GC	N50	N90	Ns	Longest	rRNA	Average seq. depth	CDSs	Hypothetical CDSs	t-RNAs
SIMKANIA-LIKE																
M1	M1_Simk	96.3	0	0	55	2,155,307	39.8	131,345	22,527	0	247,103	1	550.9	2,074	851 (41%)	35
M2	M2_Simk	96.3	0	0	108	2,253,388	39.8	107,968	12,662	0	165,195	1	67.1	2,100	889 (42%)	35
M3	M3_Simk	96.3	0	0	48	2,144,529	39.8	107,968	21,957	0	247,103	1	652	2,058	869 (42%)	35
M4	M4_Simk	96.3	0	0	123	2,294,964	39.6	131,345	11,904	0	247,103	1	178.5	2,114	886 (42%)	35
SPIROPLASMA-LIKE																
M1	M1_Spiro	55.3	0.8	0	94	275,702	25.1	3,509	1,509	0	14,680	1				
M3	M3_Spiro	13.3	0	0	45	95,860	27.4	2,213	1,350	0	5,642	1				
M4	M4_Spiro	94.4	0.03	100	68	681,895	23.5	24,415	4,430	0	52,169	1	30.1	992	554 (56%)	29
TENACIBACULUM-LIKE																
M1	M1_Tena	92.8	0.7	66.7	695	3,027,267	31.5	5,951	1,983	0	27,708	1	27.6	2,608	1099 (42%)	44
M4	M4_Tena	29.4	0.4	0	599	925,760	31.7	1,528	1,072	0	4,181	1				
MYCOPLASMA-LIKE																
M3	M3_Myco	74.1	2.3	0	11	409,158	32.9	69,761	46,043	0	129,584	1	158.1			

Table 5: Genome features of the retrieved bins.

	M1_Simk	M2_Simk	M3_Simk	M4_Simk	M1_Tena	M4_Spiro
Chromosome size (Mb)	2.2	2.3	2.2	2.3	3.0	0.7
Contigs	55	108	48	123	695	68
Percentage completion (%)	96.3	96.3	96.3	96.3	92.8	74.1
Predicted CDSs	2,074	2,100	2,058	2,114	2,699	1,006
G+C mol%	39.8	39.8	39.8	39.8	31.5	23.5
Coding regions	2,036	2,062	2,019	2,072	2,649	915
Average CDS length (bp)	988.9	981.3	991.6	967.2	912.3	385.2
Hypothetical CDSs	851	858	815	858	944	545
Conserved hypothetical CDS	288	289	290	289	519	159
tRNAs	35	35	35	35	44	29
rRNA genes	3	3	3	3	6	2

Figure 1: Phylogenetic reconstruction based on a neighbor-joining algorithm showing the position of the almost complete 16S rRNA gene sequences. For the reconstruction, a 30% conservational filter was used. Bar indicates 10% sequence divergence. The database used was the LTP_123.

Figure 2: CARD-FISH micrographs of the gastric cavity content hybridized with the candidate microorganisms studied here. The scale bar is indicated in each micrograph:

A) M2 hybridized with Simk174 (green stain, specific for the *Simkania*-like cells) and DAPI staining (blue), ATO488 and Cy3 filter sets at 400x magnification using a fluorescence microscope. *Onychodromopsis*-like cells show green stained intracellular *Simkania*-like cells (a); *Symbiodinium*-like cells appear pink (b) and the *C. tuberculata* nucleus appears only DAPI stained.

B) M2 double hybridized with Simk174 (green stain, specific for the *Simkania*-like cells) and Cotu1453 (for *C. tuberculata* cells, staining green) and DAPI staining (blue) and ATO488 filter sets at 1000x magnification using the super-resolution microscope. The *C. tuberculata* cell (a) shows positive green hybridization in the cytoplasm around the nucleus. The *Onychodromopsis*-like cell (b) shows endocellular *Simkania*-like bodies of approximately 0.2 µm in size (also green labeled), as well as DAPI stained non-hybridized particles. The size of the nucleus of the eukaryotic cell is approximately 3x larger than that of *C. tuberculata* cells. The long arrow-shaped DAPI cells correspond to sperm cells (c).

C) C5 intact gastric filament hybridized with EUB338 and DAPI staining (blue) and ATO488 filter sets at 400x using fluorescence microscope. The image shows the presence of a *Symbiodinium*-like cell colony (a), *Onychodromopsis-like* cells with endocytoplasmatic *Simkania*-like bodies (b), and the free-living *Tenacibaculum*-like cells, all them green stained and located in the mesogleal axis inside the gastric filaments of *C. tuberculata* which nucleus are stained in blue with DAPI.

D) M1 hybridized with Simk174 (green stain, specific for the *Simkania*-like cells) and DAPI staining (blue) and ATO488 filter sets at 1000x magnification using the super-resolution

microscope. The *Onychodromopsis*-like cell shows endocellular *Simkania*-like cells (a; green labeled), as well as DAPI stained non-hybridized bodies (b; only blue). The size of the nucleus of the eukaryotic cell is approximately 3x larger than those of the *Symbiodinium*-like cells. The long arrow-shaped DAPI cells correspond to the sperm cells (c), and the red cell corresponds to a *Symbiodinium*-like cell (d).

E) C5 intact gastric filament hybridized with Tena1432 (green stain, specific for the *Tenacibaculum*-like cells) and DAPI (staining) and ATO488 filter sets at 400x magnification fluorescence microscope. *Tenacibaculum*-like cells form a dense colonization of the mesogleal axis within the filament and appear as green-stained rods of 5 µm length, and the filament cell nucleus appear DAPI stained.

F) M2 hybridized with Onyc1121 (green stain, specific for the *Onychodromopsis*-like cells) together with the unlabeled helper probes HOnic1103 and HOnic1136 observed with DAPI and ATO488 filter sets at 1000x magnification using the fluorescence microscope. The *Onychodromopsis*-like cells appear green stained.

G) The same photomicrograph as F, but only DAPI stained where the cells with larger nuclei show intracytoplasmic DAPI positive bodies (a).

H) C2 hybridized with Spiro199 (green stain, specific for the *Spiroplasma*-like cells) together with the unlabeled helper probes HSpiro1 and HSpiro2 observed with DAPI and ATO488 filter sets at 1000x magnification with a fluorescence microscope. *Spiroplasma*-like cells (a) appear as endocellular bodies in the cytoplasm of cells whose nuclei are more condensed and smaller than those of the *Onychodromopsis*-like cells.

I) M2 hybridized with Cotu193 (green stain, specific for the *C. tuberculata* cells) and observed with the DAPI and ATO488 filter sets at 400x magnification with a fluorescence microscope. *C. tuberculata* positive cells (a) appear with positive green hybridization in the cytoplasm around the nucleus. *Spiroplasma*-like carrying cells (b) appear with a condensed nucleus.

J) M1 hybridized with Cotu193 (green stain, specific for the *C. tuberculata* cells) observed with DAPI and ATO488 filter sets at 1000x magnification with a fluorescence microscope. The *C. tuberculata* positive cell (a) appears with positive green hybridization in the cytoplasm around the nucleus. The *Onychodromopsis*-like cells (b) are not ATO488 stained and have a slight DAPI signal visible in the endocellular cells.

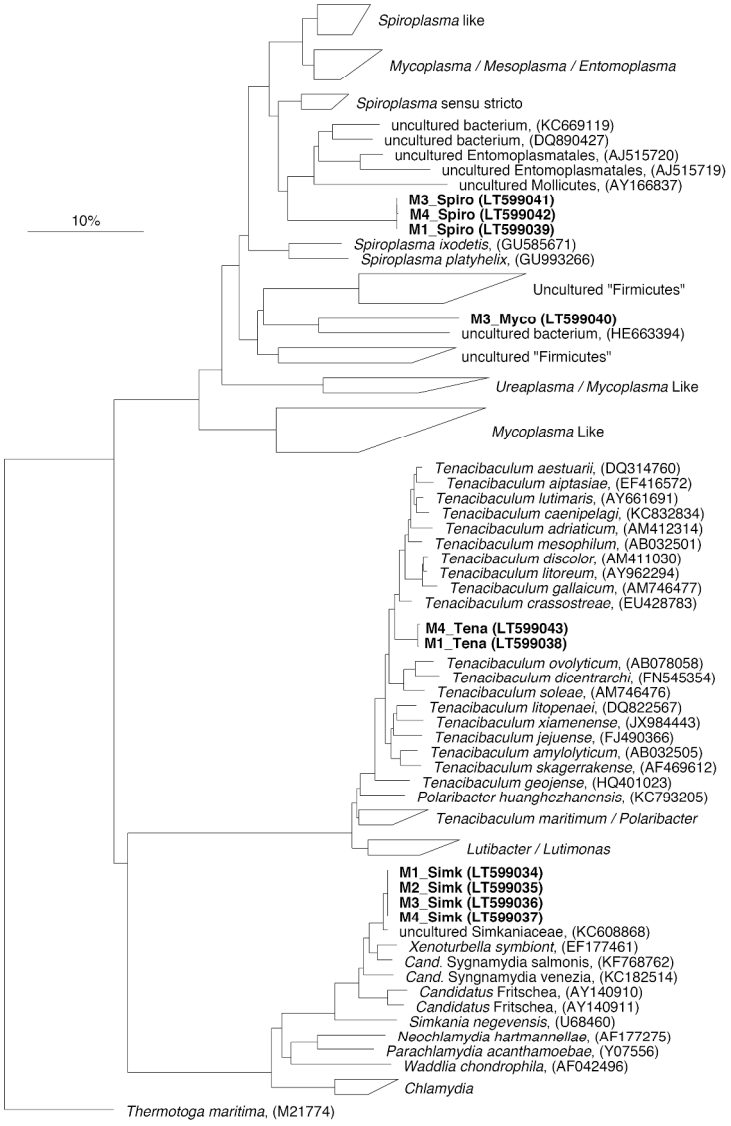

Figure 1: Phylogenetic reconst
236x347mm (300 x 300 DPI)

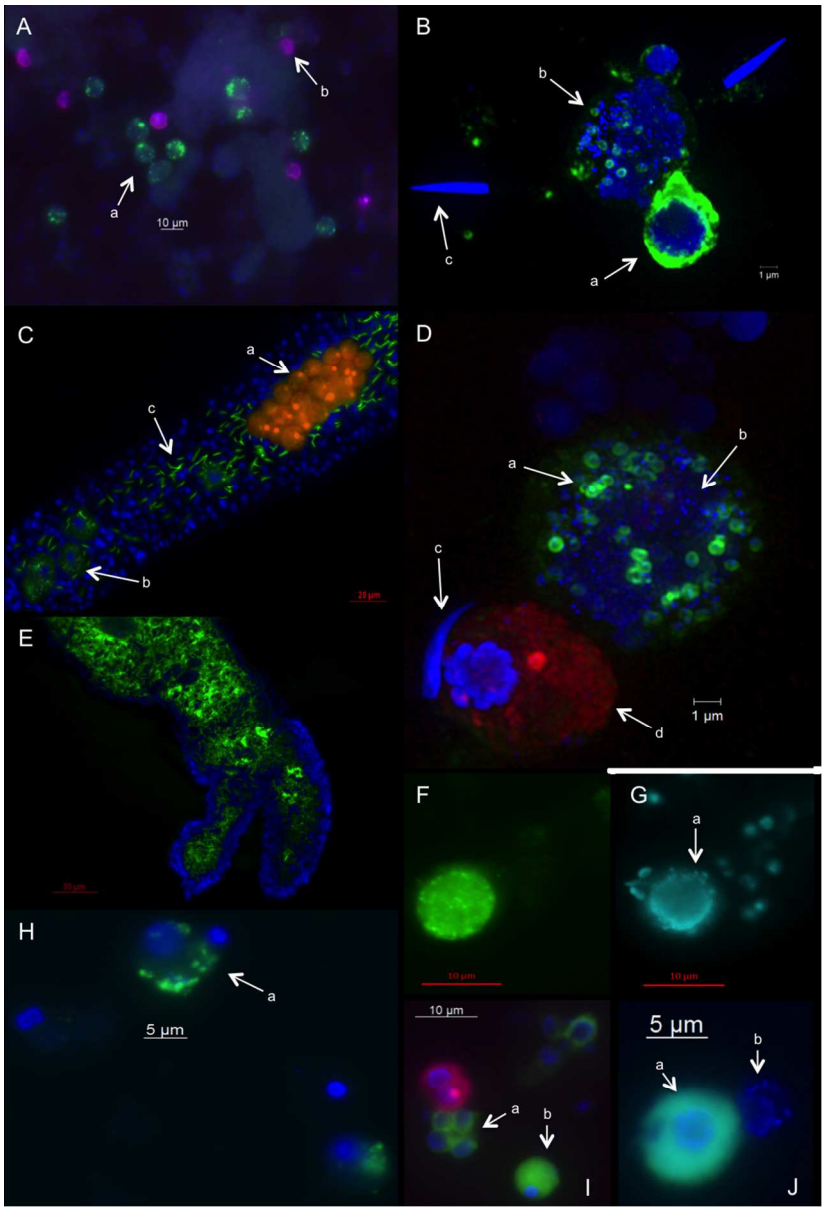

Figure 2: CARD-FISH micrograph