

# SCIENTIFIC REPORTS



OPEN

## Diversity and environmental distribution of the cosmopolitan endosymbiont “*Candidatus Megaira*”

Olivia Lanzoni<sup>1</sup>, Elena Sabaneyeva<sup>2</sup>, Letizia Modeo<sup>1</sup>, Michele Castelli<sup>3</sup>, Natalia Lebedeva<sup>4</sup>, Franco Verni<sup>1</sup>, Martina Schrällhammer<sup>5</sup>, Alexey Potekhin<sup>2</sup> & Giulio Petroni<sup>1</sup> 

Members of the order *Rickettsiales* are often found in association with ciliated protists. An interesting case is the bacterial endosymbiont “*Candidatus Megaira*”, which is phylogenetically closely related to the pathogen *Rickettsia*. “*Candidatus Megaira*” was first described as an intracellular bacterium in several ciliate species. Since then it has been found in association with diverse evolutionary distantly-related hosts, among them other unicellular eukaryotes, and also algae, and metazoa, such as cnidarians. We provide the characterization of several new strains of the type species “*Candidatus Megaira polyxenophila*”, and the multidisciplinary description of a novel species, “*Candidatus Megaira venefica*”, presenting peculiar features, which highlight the diversity and variability of these widespread bacterial endosymbionts. Screening of the 16S rRNA gene short amplicon database and phylogenetic analysis of 16S rRNA gene hypervariable regions revealed the presence of further hidden lineages, and provided hints on the possibility that these bacteria may be horizontally transmitted among aquatic protists and metazoa. The phylogenetic reconstruction supports the existence of at least five different separate species-level clades of “*Candidatus Megaira*”, and we designed a set of specific probes allowing easy recognition of the four major clades of the genus.

The family *Rickettsiaceae* (*Rickettsiales*, *Alphaproteobacteria*) comprises obligate intracellular bacteria hosted by eukaryotic organisms. *Rickettsia* and *Orientia* are the most studied genera and they are etiological agents of many human diseases, such as spotted fever or typhus<sup>1</sup>. Within the *Rickettsiaceae* family, several other bacteria are not of direct health concern and have been also referred to as “neglected *Rickettsiaceae*”<sup>2</sup>. These bacteria are important for understanding key evolutionary and ecological phenomena, such as host-symbiont interactions<sup>3,4</sup>, the emergence of pathogenicity<sup>5,6</sup>, and even the evolution of mitochondria, which are considered by several authors related to *Rickettsiales*<sup>7</sup>.

Traditionally, members of the *Rickettsiaceae* family were known to colonize only arthropods and vertebrates<sup>8</sup>, but with the discovery and characterization of new species and genera phylogenetically affiliated to this family, this conviction vanished. Indeed, recent studies<sup>2,9–14</sup> unraveled the existence of many “neglected *Rickettsiaceae*”, often hosted by protists. Considering the interwoven phylogenetic relationships among bacteria harbored by diverse hosts<sup>15</sup>, these data provided indirect evidence that some *Rickettsiaceae* can be frequently transmitted among hosts from evolutionary far related lineages<sup>9</sup>.

An interesting bacterial genus within “neglected *Rickettsiaceae*” is the endosymbiont “*Candidatus Megaira*”<sup>2</sup>. The first molecular report of this endosymbiont was in 2005 from the ciliate *Diophrys oligothrix*, where it was mentioned as a member of *Rickettsiaceae*<sup>16</sup>. Subsequently, this bacterium was found in many different hosts ranging from multicellular organisms, like cnidarians<sup>17,18</sup> to green algae including chlorophytes and streptophytes<sup>19–21</sup>, and protists - amoebae<sup>11</sup>, and numerous ciliates<sup>2,12,22</sup>. In some cases, bacteria were associated to the gut content of animals, such as ascidians<sup>23</sup>, fish<sup>24</sup>, or worms<sup>25</sup>. As these studies dealt mainly with molecular approaches, the true

<sup>1</sup>Department of Biology, University of Pisa, Pisa, Italy. <sup>2</sup>Faculty of Biology, Saint Petersburg State University, Saint Petersburg, Russia. <sup>3</sup>Centro Romeo ed Enrica Invernizzi Ricerca Pediatrica, Dipartimento di Bioscienze, Università degli studi di Milano, Milan, Italy. <sup>4</sup>Core Facilities Centre “Culture Collections of Microorganisms”, Saint Petersburg State University, Saint Petersburg, Russia. <sup>5</sup>Microbiology, University of Freiburg, Freiburg, Germany. Correspondence and requests for materials should be addressed to G.P. (email: [giulio.petroni@unipi.it](mailto:giulio.petroni@unipi.it))

nature of these associations was not clarified. Some full-length 16S rRNA gene sequences belonging to the genus “*Ca. Megaira*” have been retrieved from environmental samples, such as freshwater lakes<sup>26</sup>, aquaria<sup>27</sup>, wastewater treatment plants<sup>28</sup>, and soil<sup>29</sup>, but these studies do not allow conclusions regarding their association with possible hosts.

Notably, the type species “*Ca. Megaira polyxenophila*” was retrieved in diverse species of ciliates belonging to different classes<sup>2</sup>, including the fish parasite *Ichthyophthirius multifiliis*<sup>12</sup>, and in a wide range of additional, evolutionary distant, organisms<sup>11,17–21</sup>. In ciliates, this bacterial endosymbiont is able to colonize different host cell compartments (e.g. macronucleus or cytoplasm) according to the host species. Colonization of ciliate cells by bacteria can be tentatively explained accounting that these protists, as phagotrophic predators, can engulf bacteria, which may escape from digestion and colonize one of many cell compartments available<sup>30,31</sup>. Considering that ciliates and other protists are at the basis of many trophic chains<sup>32</sup>, it is tempting to hypothesize that they may transmit their symbionts to other hosts at higher trophic levels<sup>2</sup>.

Herein we report 14 new strains of “*Ca. Megaira polyxenophila*” and describe the novel species “*Ca. Megaira venefica*”. We also provide a critical revision of the genus “*Ca. Megaira*”, the already described species, and new clades it comprises. We designed and validated a new genus-specific probe for fluorescence *in situ* hybridization (FISH) experiments in order to allow an easy recognition of members of the genus. Moreover, we designed *in silico* four clade/species specific probes targeting the same hypervariable 16S rRNA region, and validated two of them for which strains of “*Ca. Megaira polyxenophila*” and “*Ca. Megaira venefica*” were available. Screening of 16S rRNA gene amplicons applying the Integrated Microbial Next Generation Sequencing (IMNGS)<sup>33</sup> platform demonstrates the high diffusion of this bacterial genus in many ecosystems.

## Results

**Host identification.** A preliminary screening for the presence of bacterial symbionts was performed by Differential Interference Contrast (DIC) microscopy on paramecia residing in the Core Facilities Centre “Culture Collections of Microorganisms”. When a bacterial symbiont was detected, usually all *Paramecium* hosts were infected and then molecularly characterized. A list of ciliate cultures harbouring bacterial endosymbionts affiliated to “*Ca. Megaira*” is provided (Table 1). As for the host species not present in the culture collection, infection of “*Ca. Megaira*” was recorded only in some of the isolated ciliates (e.g. *Colpidium striatum* ASP\_B; *Paramecium caudatum* RFL1, RanNy1602-AP18 and Mue14b).

The traits used for species recognition within ciliate genera should be considered reliable only if they diverge sufficiently from one species to another one, and are considered as a whole and not individually<sup>34</sup>. With this perspective, in this study, ciliates were identified up to morphospecies level, combining morphological investigation on features, such as the cell shape, size, and micronuclear appearance and number (data not shown) with molecular characterization through 18S rRNA gene sequencing (for further details see Supplementary Table S1). In case of hosts belonging to the *Paramecium aurelia* complex, sequencing of cytochrome oxidase *c* subunit I (COI) gene and internal transcribed spacer (ITS1-5.8S-ITS2) was also performed to identify the species (Supplementary Table S1).

18S rRNA gene sequence of algal endosymbionts of *Paramecium bursaria* strains was 99.6% identical to *Micractinium conductrix* (Accession Number KF887344).

**Molecular characterization of bacterial endosymbionts.** 16S rRNA gene sequencing and sequence comparison. The majority of the nearly full-length 16S rRNA gene sequences characterized in this study displayed identity higher than 99.7% with “*Ca. Megaira polyxenophila*” (Supplementary Table S2), and were consequently assigned to this species. In case of *P. caudatum* Sp9-41, a second symbiont morphologically similar to *Holospira undulata* was present in the micronucleus (Supplementary Fig. S1). Its 16S rRNA gene (accession number MG563930, 1404 bp) was obtained after cloning, and confirmed the assignment to *H. undulata*.

The sequences of the novel “*Ca. Megaira*” species inhabiting the cytoplasm of *P. bursaria*, *P. nephridiatum* and *P. putrinum* strains were 1413 bp long and shared the highest identity with members of the “*Ca. Megaira*” Clade C, namely uncultured *Rickettsiales* bacteria obtained from *Hydra oligactis*, several uncultured bacteria associated to *Ichthyophthirius multifiliis*<sup>12</sup>, and bacterial sequences retrieved from environmental samples of lakes and soil (range 98.6–99.8%, Supplementary Table S3). The newly characterized sequences presented few nucleotides of difference between symbionts from different host species (Supplementary Table S3).

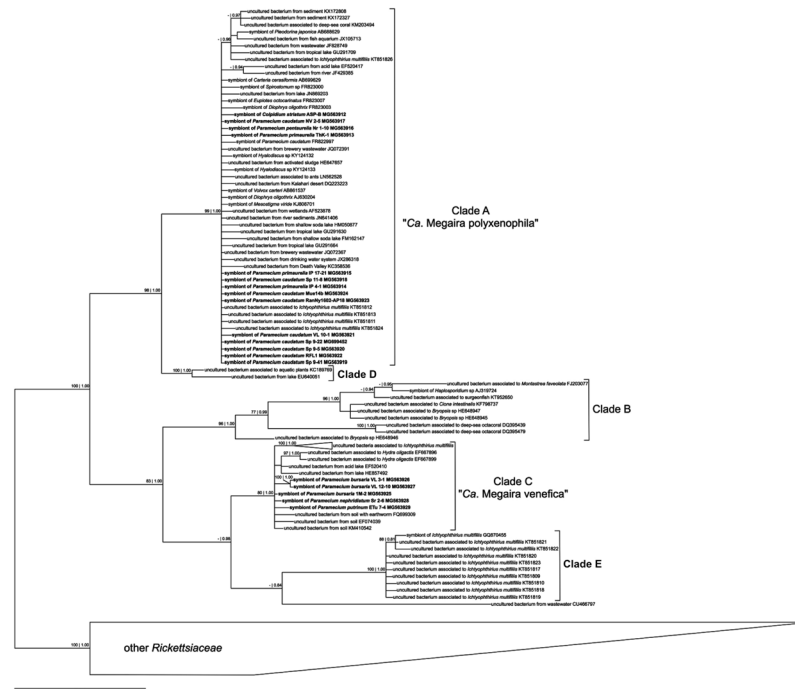
16S rRNA gene sequence identities were calculated for all members of the genus “*Ca. Megaira*” (Supplementary Table S4), and confirmed the existence of at least five clades within the genus.

A critical manual inspection of available sequences from NCBI nucleotide database showed the presence of 9 previously unrecognized chimeras, and 20 misclassified sequences. Sequences were firstly screened with nucleotide BLAST; if identity was more pronounced only on one side of the sequence, this was split in two parts in the supposed breaking region. The two sequence parts were independently blasted, and if only one of them provided BLAST results similar to the whole sequence, the other one being significantly different from the best hit, the original sequence was considered a chimera and consequently reannotated. Four identified chimeras were particularly hard to detect, as they originated from two diverse “*Ca. Megaira*” clades (AM159487; HQ691997; KT851814; KT851825). This observation suggests that representatives of two “*Ca. Megaira*” clades were co-occurring in the same host/sample. Some of these chimeras have been included in previous phylogenies of the genus (e.g. AM159487; HQ691997; KT851814; KT851825); these sequences were excluded from the following analyses (Supplementary Table S5) and their putative chimeric origin has been communicated to GenBank to implement public annotation. In our strains, we detected the presence of a single “*Ca. Megaira*” species at once. On the contrary, a critical screening of “*Ca. Megaira*” sequences retrieved from different *I. multifiliis* strains evidenced sometimes the presence of representative from two “*Ca. Megaira*” clades, namely Clade A and Clade E in the same host. Indeed, in our chimera screening, in these hosts we also detected chimeric sequences between these two clades.

Symbiont	Host	Culture index	Intracellular localization	Origin	Sampling geographical coordinates
"Candidatus Megaira polyxenophila"	<i>Colpidium striatum</i>	ASP_B (polyclonal)	Cytoplasm	Italy, Perugia (fish-farm)	42.763018, 12.861376
	<i>Paramecium primaurelia</i>	ThK-1 (monoclonal)	Cytoplasm	Thailand, Phi-Phi Don islands, (stream)	7.737428, 98.773924
	<i>Paramecium primaurelia</i>	IP 4-1 (monoclonal)	Cytoplasm	Italy, Pisa (drain)	43.432147, 10.224717
	<i>Paramecium primaurelia</i>	IP 17-21 (monoclonal)	Cytoplasm	Italy, Pisa (drain)	43.441793, 10.225463
	<i>Paramecium pentaurelia</i>	Nr1-10 (monoclonal)	Cytoplasm	Russia, Novorossiysk (wastewater)	44.423800, 37.470800
	<i>Paramecium caudatum</i>	NV 2-5 (monoclonal)	Macronucleus	Russia, Velikiy Novgorod, (pond)	58.311554, 31.161758
	<i>Paramecium caudatum</i>	Sp 11-8 (monoclonal)	Macronucleus	Spain, Madrid, (drain)	40.244900, 3.410472
	<i>Paramecium caudatum</i>	Sp 9-41 (monoclonal)	Macronucleus (presence of <i>Holospora undulata</i> in micronucleus)	Spain, Madrid, (pond)	40.250417, 3.410606
	<i>Paramecium caudatum</i>	Sp 9-5 (monoclonal)	Macronucleus (presence of betaproteobacterium in cytoplasm)	Spain, Madrid, (pond)	40.250417, 3.410606
	<i>Paramecium caudatum</i>	Sp 9-22 (monoclonal)	Macronucleus (presence of betaproteobacterium in cytoplasm)	Spain, Madrid, (pond)	40.250417, 3.410606
	<i>Paramecium caudatum</i>	VL 10-1 (monoclonal)	Macronucleus	Russia, Vladimir, (wastewater pond)	56.101171, 40.434921
	<i>Paramecium caudatum</i>	RFL1 (polyclonal)	Macronucleus	Russia, Ropsha (fish-farm)	59.725603, 29.858563
	<i>Paramecium caudatum</i>	RanNy1602-AP18	Macronucleus	Germany, Rangsdorf (lake)	52.289597, 13.407075
	<i>Paramecium caudatum</i>	Mue14b	Macronucleus	Germany, Muenster (canal)	51.968526, 7.632249
"Candidatus Megaira venefica"	<i>Paramecium bursaria</i>	1M-2 <sup>(T)</sup> (monoclonal)	Cytoplasm	Russia, Peterhof (wastewater drain)	59.879020, 29.851318
	<i>Paramecium bursaria</i>	VL 3-1 (monoclonal)	Cytoplasm	Russia, Vladimir (pond)	56.159550, 40.360211
	<i>Paramecium bursaria</i>	VL 12-10 (monoclonal)	Cytoplasm	Russia, Vladimir (pond)	56.157310, 40.357658
	<i>Paramecium nephridiatum</i>	Sr 2-6 (monoclonal)	Cytoplasm	Russia, White sea (brackish pool)	66.287654, 33.666561
	<i>Paramecium putrinum</i>	ETu 7-4 (monoclonal)	Cytoplasm	Estonia, Tartu (canal)	58.369089, 26.751568

**Table 1.** List of ciliate cultures harbouring bacterial endosymbionts.

In accordance with previous studies<sup>2</sup>, members of Clade A "Ca. Megaira polyxenophila" showed identity values above 98.6%. Members of Clade D had 99.2% of identity among themselves, while they shared identity values between 97.3% and 98.5% with Clade A (when only full-length sequences were considered), and these values are below the threshold established (98.65-98.7%) to discriminate different bacterial species according to 16S rRNA gene sequence<sup>35</sup>. The situation of Clade B was more heterogeneous; indeed, all members shared a minimum identity of 96.5%, which may reveal further inner species level subdivision. On the contrary, Clade C and Clade E were more homogeneous and had identity range higher than 98.6% and 99.5%, respectively. Identities of members of Clade C with other "Ca. Megaira" clades did not exceed 98.3%. Therefore, the newly characterized symbionts together with previously available sequences allowed to define Clade C as a new species, which we named "Ca. Megaira venefica" (a formal description is provided as Supplementary Text S1).



**Figure 1.** Bayesian inference phylogenetic tree of the genus “*Candidatus Megaira*” based on 16S rRNA gene sequences inferred with the GTR + I + G model. Numbers associated to each node represent bootstrap values inferred after 1000 maximum likelihood pseudo-replicates and Bayesian posterior probabilities (values below 70|0.70 are not shown). Sequences in bold were characterized in this study, and “Ca.” stands for “*Candidatus*”. The scale bar represents an estimated genetic distance of 0.3.

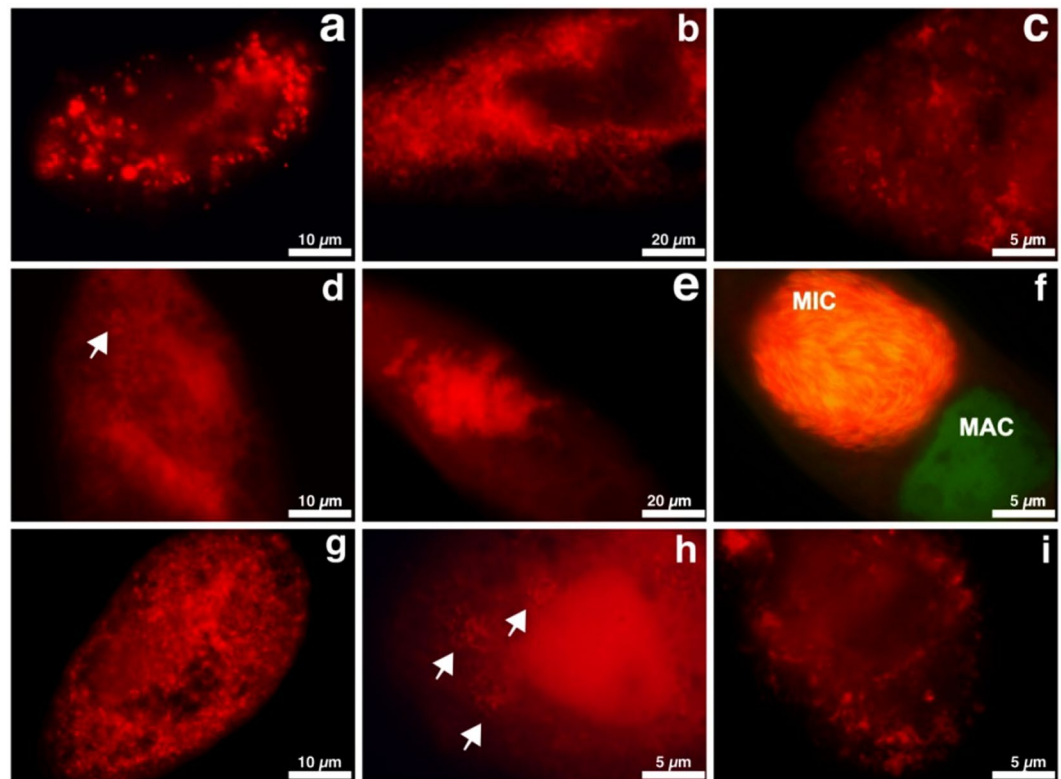
**Phylogenetic analyses.** Our phylogenetic reconstruction of “*Ca. Megaira*” shows an updated scenario compared to previous studies<sup>2,11,12</sup>, where only up to four clades were reported. Indeed, we confirmed Clade A and B, which were retrieved by all previous authors<sup>2,11,12</sup>, we followed the suggestion by Zaila *et al.*<sup>12</sup> to split Clade C of Schrallhammer *et al.*<sup>2</sup> into two and named them respectively Clade C and E. The presence of an additional clade suggested by Hess<sup>11</sup>, named Clade D, was confirmed. Additionally, we identified one sequence from wastewater (CU466797), which is clearly not affiliated to any of the described clades, but instead forms a sister group of Clade E, thus suggesting the possible existence of an additional clade (Fig. 1). This hypothesis is also sustained by phylogenetic analysis on hypervariable regions presented below, where region V4–V6 of this sequence is embedded in a large clade. At variance with previous studies, we could recognize several published sequences as possible chimeras, and removed them from the phylogenetic analyses.

The clade comprising the type species “*Ca. Megaira polyxenophila*”, referred to as Clade A by previous authors<sup>2,11,12</sup>, is highly supported (99% bootstrap value for Maximum Likelihood (ML) and 1.00 of posterior probability), and includes all our characterized symbionts assigned to this species. Clade D, sister group of Clade A, is fully supported (100% bootstrap value for ML and 1.00 of posterior probability for BI), and consists of two sequences isolated from aquatic plants and a freshwater lake, respectively. Clade B is well supported (96% bootstrap value for ML and 1.00 of posterior probability for BI), and contains almost only bacteria associated to marine hosts, such as *Bryopsis*<sup>19</sup>, corals<sup>18</sup>, and a *Haplosporidium* species infecting a marine mollusc<sup>36</sup>. Clade C is moderately supported (80% bootstrap value for ML and 1.00 of posterior probability for BI), and includes the new species “*Ca. Megaira venefica*”, represented by symbionts found in *Paramecium* (current study), bacteria associated to *Hydra oligactis*, *I. multifiliis*, and by uncultured environmental bacteria. Finally, Clade E comprises exclusively bacteria detected in the fish parasite *I. multifiliis* and is highly supported (100% bootstrap value for ML and 1.00 of posterior probability for BI).

**Fluorescence *in situ* hybridization.** One new genus-specific probe was designed for FISH experiments, which targeted the symbionts of interest with high specificity with one exception (Table 2): i.e. the probe designed for the genus (Megenus\_487) matched *in silico* all available members of “*Ca. Megaira*” present in our phylogenetic reconstruction (Fig. 1) but also had a single unpecific hit with an uncultured gammaproteobacterium sequence (Table 2). Other organisms positive to probe Megenus\_487 in Ribosomal Database Project (RDP)<sup>37</sup> turned out to be misclassified “*Ca. Megaira*” sequences (often recorded as *Orientia*, or uncultured *Alphaproteobacteria*), or chimeric sequences (Supplementary Table S5). Additionally, four specific probes were additionally designed to recognize the most abundant clades A + D, B, C and E; all these probes showed a high *in silico* specificity for the targeted sequences with no unpecific hits. The hypervariable region chosen as target site was basically the same, and it was comprised between position 66 and position 95 according to *E. coli* 16S rRNA gene reference. Only two clade-specific probes (A + D, C) could be experimentally tested with FISH experiments using available strains (i.e. “*Ca. Megaira*”).

Probe name	Target	Sequence	RDP	SILVA
Megenus_487	All genus members	5'-GCCGGGGCTTTTCTGTGGT-3' ( $T_m = 61.8^\circ\text{C}$ )	1	0
MegPoly_66	" <i>Candidatus Megaira polyxenophila</i> " Clade A and " <i>Candidatus Megaira</i> " Clade D	5'-GCAAGCCCAATTTGTTCGT-3' ( $T_m = 57.9^\circ\text{C}$ )	0	0
MegairaB_76	" <i>Candidatus Megaira</i> " Clade B	5'-YCTGAAGCAAGCTCCAGC-3' ( $T_m = 57.1^\circ\text{C}$ )	0	0
MegVene_95	" <i>Candidatus Megaira venefica</i> " Clade C	5'-CCGTTGCCACTAACGAC-3' ( $T_m = 56.0^\circ\text{C}$ )	0	0
MegairaE_69	" <i>Candidatus Megaira</i> " Clade E	5'-GGTGCTTCGTCCAAAGGCATC-3' ( $T_m = 61.8^\circ\text{C}$ )	0	0

**Table 2.** *In silico* matching of "*Candidatus Megaira*" probes against bacterial 16S rRNA gene sequences available from RDP (release 11, update 4) and SILVA (release 123) databases. The reported numbers indicate the number of non-specific target sequences detected by the probe against the total number of sequences matching the probe with 0 mismatches.

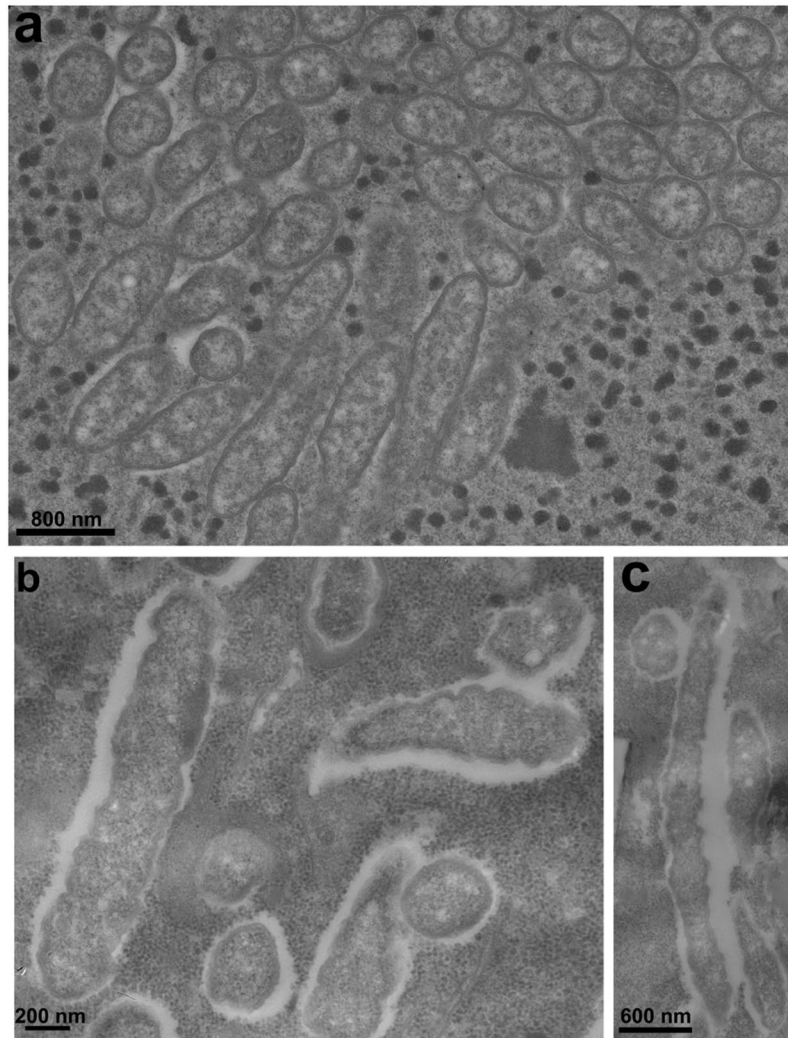


**Figure 2.** Fluorescence *in situ* hybridization of "*Candidatus Megaira polyxenophila*" and "*Candidatus Megaira venefica*". "*Ca. Megaira polyxenophila*" targeted by probe MegPoly\_66 (Cy3-labelled, red signal) from the cytoplasm of *Colpidium striatum* ASP-B (a), *P. primaurelia* ThK-1 (b), *P. primaurelia* IP 17-21 (c), *P. pentaurella* Nr1-10 (d), and the macronucleus of *P. caudatum* Sp 11-8 (e). Double nuclear infection in *P. caudatum* Sp 9-41: *Holospora undulata* (Cy3-labelled specific probe for *Holospora*, red signal) in the micronucleus (MIC), and "*Ca. Megaira polyxenophila*" (MegPoly\_66 probe labelled with fluorescein, green signal) in the macronucleus (MAC) (f). "*Ca. Megaira venefica*" in the cytoplasm of its respective hosts *P. bursaria* (g), *P. nephridiatum* (h), and *P. putrinum* (i), targeted by probe MegVene\_95 (Cy3-labelled, red signal). White arrows indicate the presence of symbionts within the host cells.

*polyxenophila*", and "*Ca. Megaira venefica*") at different formamide concentrations (0%, 15%, and 30%). For the other untested probes, we assumed that they should work properly as they were designed on the same 16S rRNA gene region, which should be accessible as in the tested closely related "*Ca. Megaira*" species. Nevertheless, future controls are recommended once strains of these clades will be available. The signal for the two tested probes resulted visible at all formamide levels, and signal intensity was optimal at 0% formamide (Fig. 2).

"*Ca. Megaira polyxenophila*" strains were located either in the cytoplasm in case of *Colpidium striatum* and members of the *P. aurelia* complex, or in the macronucleus in case of *P. caudatum* (Fig. 2a-e). Infection level

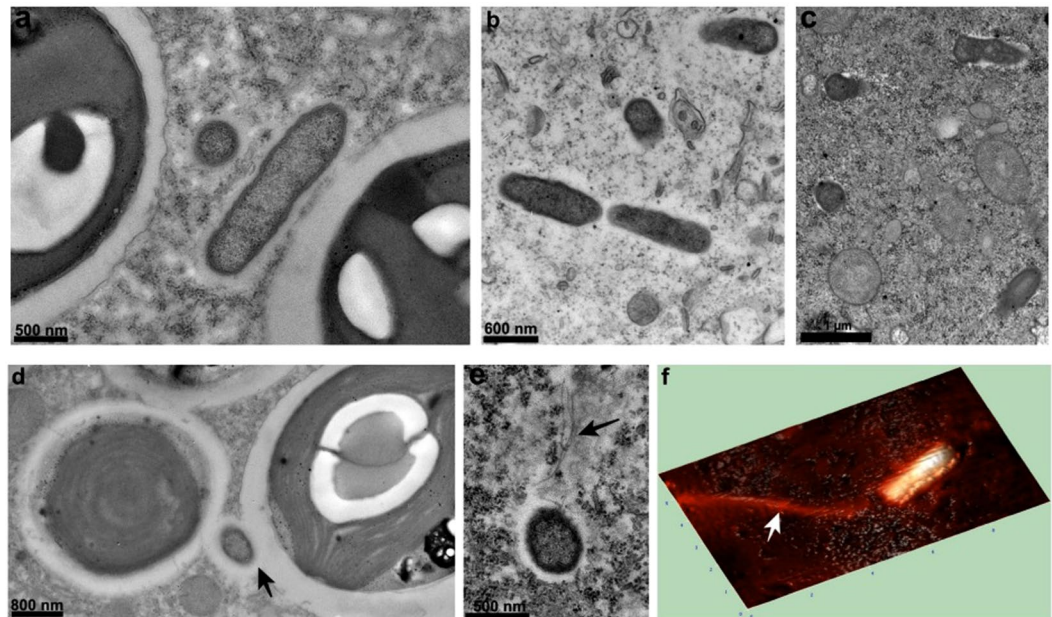




**Figure 3.** Transmission electron microscopy of “*Candidatus Megaira polyxenophila*”. View of macronuclear infection of “*Ca. Megaira polyxenophila*” in *P. caudatum* Sp 9-5 (a), longitudinal and transverse sections of “packed” “*Ca. Megaira polyxenophila*” in the cytoplasm of strain *P. primaurelia* ThK-1 (b), longitudinal and transverse sections of “giant” endosymbionts found in the cytoplasm of strain *P. primaurelia* ThK-1 (c).

and prevalence (number of infected individuals) was usually very high in each culture screened. In *P. pentastrella* Nr1–10, we observed fluctuations in the number of symbionts per cell, ranging from only few bacteria per cell to densely filled ciliates. In three strains of *P. caudatum*, other intracellular bacteria were detected: strain Sp 9–41 harbored *H. undulata* in its micronucleus (Fig. 2f), while strains Sp 9-5 and Sp 9-22 hosted in their cytoplasm a betaproteobacterium, which is presently under investigation (manuscript in preparation, data not shown). *H. undulata* and the betaproteobacterium were not labelled by the employed “*Ca. Megaira*” probes at any formamide concentration. All *P. bursaria* strains and *P. putrinum* harboring “*Ca. Megaira venefica*” hosted symbionts homogeneously distributed in the cytoplasm (Fig. 2g,i), while cytoplasmic symbionts of *P. nephridiatum* clustered together in small groups (Fig. 2h). A further FISH control was performed using “*Ca. Megaira polyxenophila*” and “*Ca. Megaira venefica*” in the same experiment, and probes did not cross-react, thus showing good specificity (data not shown).

**Morphological and ultrastructural description of endosymbionts “*Candidatus Megaira polyxenophila*”.** All “*Ca. Megaira polyxenophila*” strains studied presented the classical “*Ca. Megaira polyxenophila*” ultrastructure<sup>2,21</sup>, consisting of double membrane (typical of Gram-negative bacteria), homogeneous cytoplasm, a clear halo surrounding the cells, and size dimensions in the range ( $\sim 1.5 \mu\text{m} \times 0.6 \mu\text{m}$ ). Cytoplasm of *P. primaurelia* strain ThK-1 and macronucleus of *P. caudatum* strain Sp 9-5 were densely packed with bacteria (Fig. 3a,b), but also some bacteria-free space was always left. Some bacterial cells from the cytoplasm of strain ThK-1 appeared “giant” when compared to previous descriptions of “*Ca. Megaira polyxenophila*”, reaching length of  $\sim 4.8 \mu\text{m}$  (Fig. 3c).

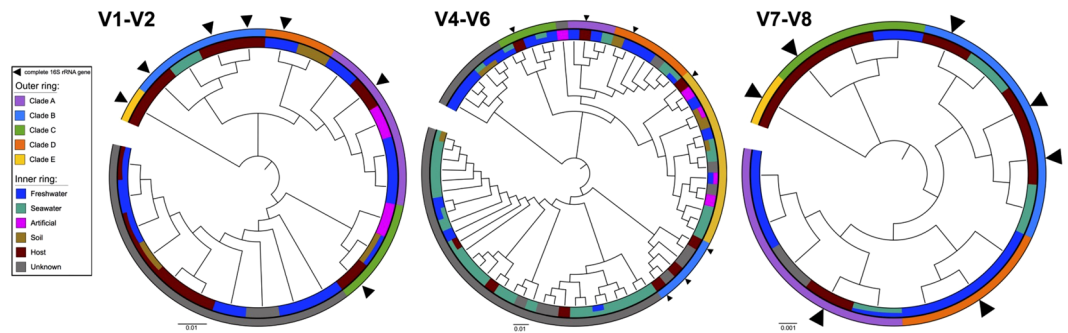


**Figure 4.** Ultrastructure of “*Candidatus Megaira venefica*”. TEM images of two “*Ca. Megaira venefica*” cells in the cytoplasm of *P. bursaria* strain 1M-2<sup>(T)</sup>, shown in transverse and longitudinal sections (a). Endosymbionts in the cytoplasm of *P. putrinum* (b), and in *P. nephridiatum* (c). “*Ca. Megaira venefica*” associated to the perialgal vacuole (black arrow) in *P. bursaria* strain 1M-2<sup>(T)</sup> (d), cell of “*Ca. Megaira venefica*” reporting flagella (black arrow) detected in the cytoplasm of *P. bursaria* strain VL 12-10 (e). Atomic force microscopy image of “*Ca. Megaira venefica*” in strain VL 12-10, showing the putative flagellum (white arrow; f).

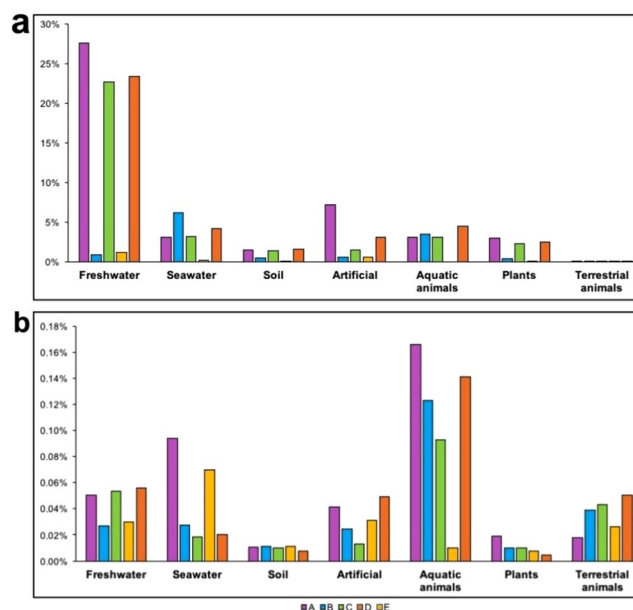
**Novel “*Candidatus Megaira*” species: “*Candidatus Megaira venefica*”.** Symbionts identified as the novel “*Ca. Megaira*” species appeared rod-shaped and presented a typical Gram-negative ultrastructure consisting of two membranes with a homogeneous cytoplasm. The length of bacteria ranged from ~1.9 μm in *P. bursaria* to 1.4 μm in *P. putrinum*, while width was constantly 0.4 μm. Bacteria were mostly present in the ciliate cytoplasm, but their abundance varied according to the *Paramecium* species: symbionts in *P. bursaria* and in *P. putrinum* were scattered, whereas in *P. nephridiatum* they were grouped together in rare clusters (Fig. 4a–c).

In *P. bursaria*, half of bacteria were also found closely associated to the membrane of perialgal vacuole containing symbiotic algae symbionts of *P. bursaria* (Fig. 4d). The possible localization of bacteria inside the perialgal membrane was detected only once, so we cannot rule out the possibility of fixation artefacts as well (Supplementary Fig. S2). However, bacteria were never noticed to invade the symbiotic algal cells. Generally, no pili and flagella were observed, except for the strain VL 12-10, where structures resembling flagella were occasionally detected (Fig. 4e,f).

**Diversity and environmental distribution of “*Candidatus Megaira*”.** “*Ca. Megaira*” diversity and environmental distribution in previous metabarcoding studies was examined using IMNGS platform<sup>33</sup>. The output of IMNGS similarity search with a threshold of 95% identity with “*Ca. Megaira*” sequences produced a total number of 11373 centroid sequences (identity higher than 99%, as by IMNGS default settings) originating from 4673 samples. After removal of short and low quality sequences, all remaining 11306 sequences were pooled and re-clustered in 194 OTUs using a 99% identity threshold. OTUs were divided in three different groups according to which hypervariable region of 16S rRNA gene they belonged to and phylogenetic trees were constructed for each hypervariable region. In all trees, the five major clades of “*Ca. Megaira*” are distinguishable and their relationships are in agreement with full-length 16S rRNA gene phylogeny: Clade A and D are sister groups, as well as Clade C and Clade E, whereas Clade B is always separate (Fig. 5). However, many not yet described lineages are present in addition to the previously identified clades, for example in the tree based on hypervariable region V4–V6, at least six additional clades can be observed (Fig. 5). It was impossible to assign each clade to a specific environment according to OTU environmental distribution (Fig. 5), thus a more detailed analysis was performed to investigate the ecology of different “*Ca. Megaira*” clades. Frequency of occurrence (calculated as the number of microbiome samples positive for “*Ca. Megaira*” divided the total number of samples of each category) and relative abundance (determined as the number of “*Ca. Megaira*” reads divided total number of reads for each sample) of the five “*Ca. Megaira*” clades were estimated for environments and host categories (Fig. 6, for further details see Methods). Clades A, C, D had maximal frequencies of occurrence in freshwater samples reaching more than 20% (Fig. 6a), whereas their frequencies of occurrence were below 5% for the rest of environmental categories and less than 1% in terrestrial animals (Fig. 6a). On the other side, Clade B showed an overall lower occurrence, in particular, it was much less frequently retrieved in freshwater (less than 1% samples), while it had the highest frequency of occurrence in seawater (about 7%) (Fig. 6a). Clade E seemed to be very rare in the environment, thus lacking any significant tendency (Fig. 6a). All “*Ca. Megaira*” clades, all except Clade E, had a higher relative



**Figure 5.** Diversity of “*Candidatus Megaira*” based on 16S rRNA gene amplicon search in IMNGS. Phylogenetic trees of different 16S rRNA gene hypervariable regions. IMNGS hits longer than 300 bp were clustered in OTUs with 99% identity and separated for each hypervariable region taken into analysis. A total number of 194 OTUs were obtained: 42 for V1–V2, 135 for V4–V6, 17 for V7–V8. Complete 16S rRNA gene sequences were employed in the analysis to enlighten the diversity of each “*Ca. Megaira*” clade. OTUs were also assigned to environmental compartments according to the sequence origin (outer ring).



**Figure 6.** Environmental distribution of “*Candidatus Megaira*” based on 16S rRNA gene amplicon search in IMNGS. Frequency of occurrence estimated as the number of times with which “*Ca. Megaira*” occurred in all samples. On the x axis the categories of environment and host samples are represented, whereas in the y axis the percentage of frequency of occurrence is shown (a). Relative abundance of “*Ca. Megaira*” expressed as the ratio between “*Ca. Megaira*” positive samples and the total number of samples for a certain environment or host. On the x axis the categories of environment and host samples are represented, whereas in the y axis the percentage of relative abundance is shown (b).

abundance in aquatic animals (more than 0.09%). In most cases, relative abundances were comprised between 0.01% and 0.06% including also environments where bacteria had a high frequency of occurrence (e.g. Clades A, C, D, in freshwater) or in in which they had a minimal frequency of occurrence (e.g. all clades in terrestrial organisms). Clade E showed very low relative abundances in all categories considered (Fig. 6b) except seawater, where it reached 0.06%; similarly, Clade A that has a low frequency of occurrence in seawater showed in the same environment a relatively high abundance of 0.09%.

## Discussion

The study presents the description of a novel “*Ca. Megaira*” species and characterization of new “*Ca. Megaira polyxenophila*” strains, together with screening of “*Ca. Megaira*”-related sequences in environmental 16S rRNA gene amplicons databases, providing further understanding of “*Ca. Megaira*” wide host range adaptation.

The new species “*Ca. Megaira venefica*” has been found in several strains of *Paramecium* isolated from different Russian and Estonian locations. Bacterial cells were usually located in the cytoplasm of their hosts, and,



in case of *P. bursaria* and *P. putrinum*, symbionts were extremely numerous (Fig. 2). An interesting feature of bacteria found in *P. bursaria* strains was the predilection for being closely associated to the endosymbiotic algae inhabiting the ciliate cytoplasm (Fig. 4d; Supplementary Fig. S2). A similar “behavior” has been already reported for “*Ca. Sonnebornia yantaiensis*”, another cytoplasmic bacterial symbiont of *P. bursaria*<sup>38</sup>. Probably, bacterial endosymbionts benefit from this intracellular localization, as they find protection from host auto-digestion<sup>31</sup>, and at the same time, they may exploit nutrients and metabolites synthesized and eventually secreted by the endosymbiotic algae<sup>39</sup>.

The novel “*Ca. Megaira*” species displays morphological diversity, as some bacteria showed the presence of flagella (Fig. 4c,d). Some members of the order *Rickettsiales* possess either flagella or a set of flagellar genes<sup>3,10,40–42</sup>, and the presence of flagella is known to be an ancient trait of *Alphaproteobacteria*<sup>43</sup>, lost in many *Rickettsiales* probably due to adaptation to intracellular lifestyle. The maintenance of flagella/flagellar genes has been suggested to be involved in the infection process, symbiosis preservation, and transmission of the symbiont to other hosts, either as a motility organelle or as a type III secretion system<sup>40,41,44</sup>. Since flagella in “*Ca. Megaira venefica*” are not always present, it is tempting to hypothesize a connection with some unknown life-cycle stage. For example, flagella might be useful to colonize new hosts, possibly in conjunction with other players involved in molecular interactions, such as a type IV secretion system and a rOmbB-like surface protein, which have been detected in “*Ca. Megaira polyxenophila*” from green algae, and are putatively involved in pathogenesis in other *Rickettsiaceae*<sup>21</sup>.

Our ultrastructural investigations revealed that members of the “*Ca. Megaira*” genus possess variability even in cell dimensions. While the usual recorded cell size of “*Ca. Megaira polyxenophila*” was 1.0–1.6 µm in length and 0.3–0.4 µm in width<sup>2,16,21,45</sup>, in this work, the giant symbionts detected in the cytoplasm of *P. primaurelia* strain ThK-1 were almost three times longer (Fig. 3c). Cell size of “*Ca. Megaira venefica*” varied among strains from different *Paramecium* host species, as *P. nephridiatum* and *P. putrinum* symbionts were smaller than those of *P. bursaria*. Variation in cell size is known for bacteria from the order *Rickettsiaceae*, for example different size has been described within the same *Rickettsia* species<sup>46</sup>.

In ciliates screening studies, the most frequently found species of the genus is “*Ca. Megaira polyxenophila*”, which is able to colonize diverse hosts, and shows only limited intraspecific variation in the 16S rRNA gene. We characterized several strains of this bacterial symbiont both from novel hosts like *Colpidium striatum* and species of the *P. aurelia* complex, and from already known hosts such as *P. caudatum*. Our investigation supported previous studies<sup>2</sup> showing that this symbiont displays cellular compartment specificity according to the host. In particular, “*Ca. Megaira polyxenophila*” inhabits the cytoplasm of the *P. aurelia* species, while in *P. caudatum* this symbiont was always found in the macronucleus (Table 1). Thus, even in the closely related hosts this bacterium demonstrates variability of interaction, probably due to peculiarities of the host biology: in *P. aurelia* the macronucleus is not a stable compartment being frequently resorbed in autogamy, while in *P. caudatum* autogamy is absent<sup>47</sup>.

Our sequence comparison analysis and phylogenetic reconstruction allows drawing a more defined scenario within the genus “*Ca. Megaira*” with respect to the previous studies<sup>2,11,12</sup>. Clade D is phylogenetically positioned as a different species from Clade A with high support (Fig. 1), despite data on this clade are still scarce, as only two complete 16S rRNA gene sequences are available so far, thus only hinting at the real variability within this clade. However, phylogenetic analysis of hypervariable regions discloses a much wider diversity of Clade D (Fig. 5). Clade B is highly supported in 16S rRNA gene phylogeny (Fig. 1). Nevertheless, its members displayed relatively high genetic diversity (lowest identity 96.5%, Supplementary Table S4), thus allowing to divide this clade into at least three separate sub-groups. 16S rRNA gene hypervariable region phylogenies confirm even greater diversity within the Clade B lineage, as many different OTUs are present within the single compact clade (Fig. 5). All these data taken together suggest that diversity of Clade B resides mainly in hypervariable regions. In any case, we have to take into account that part of the internal variability present in each “*Ca. Megaira*” clade could be partly attributed to PCR errors derived from cloning procedures that could artificially determine lower identity values<sup>48</sup>. A critical analysis of 16S rRNA gene hypervariable region phylogenies points out the presence of at least ten “*Ca. Megaira*” species-level clades (Fig. 5). Moreover, the single “stand-alone” sequence (CU466797) evidenced formed a separate clade in hypervariable region V4–V6 tree (Fig. 5). This hypervariable region highlighted the existence of at least other five clades, possibly representing new species, for which full-length 16S rRNA sequences are still missing. Thus, we can consequently predict as a minimum ten species within “*Ca. Megaira*” genus, half of which currently lack considerable molecular data.

We failed to assign directly each clade of the 16S rRNA gene hypervariable region phylogenies to a certain environment or host by checking OTUs ecological origins. A more detailed environmental investigation of a 16S rRNA gene short amplicon database was applied considering frequencies of occurrence and relative abundances for each “*Ca. Megaira*” clade (Fig. 6). Frequency of occurrence analysis clearly showed a preference of Clades A, C, and D for freshwater environments, whereas Clade B displayed a preference for marine ones. On the contrary, Clade E was rarely detected in all environments, and it is possible to speculate that this could be related to a higher host specificity of Clade E bacteria. Indeed, up to now, full-length sequences have been found exclusively associated to the fish parasitic ciliate *Ichthyophthirius multifiliis*<sup>12</sup>. Interestingly, when present, Clade E bacteria showed a relative abundance similar to those of other clades or, in seawater, even higher (up to 0.07%). In general, relative abundance, which is higher in aquatic animals, suggests that, in these samples, if present, “*Ca. Megaira*” is also relatively abundant. Intriguingly, this is true also for terrestrial animals in which “*Ca. Megaira*” is seldom found (frequency of occurrence < 1%), but, when present, has relative abundance higher than in soil (i.e. the environment generally explored by terrestrial animals). The high relative abundance of clades A and E in seawater if compared to low frequency of occurrence could be explained as a specific abundance in unidentified microeukaryote present in those specific samples. Unfortunately, no reliable estimate is available on the potential presence of microeukaryotes existing in

water samples used for these metabarcoding studies. In general, the comparison of data on frequency occurrence with those on relative abundance are coherent with the idea that “*Ca. Megaira*” is associated to eukaryotic hosts.

Nearly all studies reporting “*Ca. Megaira*” presence dealt with aquatic environments, and these bacteria were reported as endosymbionts of diverse hosts, both unicellular, namely ciliates<sup>2,16,22</sup> and amoebae<sup>11</sup>, and multicellular ones, such as cnidarians<sup>17,18</sup>, and green algae<sup>19</sup>. Protozoa and unicellular green algae are known to be frequent hosts of “*Ca. Megaira*” symbionts<sup>2</sup>. However, these hosts are so far underrepresented in 16S rRNA gene metabarcoding studies and, interestingly, our results revealed the absence of “*Ca. Megaira*” in the few samples associated to unicellular organisms. At the same time, we strongly suppose that many positive environmental hits from aquatic samples or gut samples of aquatic organisms might be explained by presence of unnoticed protists hosting “*Ca. Megaira*” endosymbionts, especially considering that “*Ca. Megaira*” – like all *Rickettsiales* – has never been shown to have a free-living stage. On the contrary, terrestrial organisms’ samples are very numerous, but “*Ca. Megaira*” sequences were extremely rare in these datasets. Relative abundance of “*Ca. Megaira*” sequence is generally low, as it is an endosymbiotic bacterium and sampling procedures were not aimed to isolate only host organisms. The relative abundance of endosymbionts in ciliates can reach high values only in studies specifically designed to analyse the microbiome of single cells<sup>49</sup>. Although aquatic ecosystems appear to be preferential environments for these bacteria (Fig. 6a), our results show that also terrestrial plants (e.g. crops, *Arabidopsis thaliana*, tropical trees), or their associated microbiomes, are other ecological niches where “*Ca. Megaira*” can occur (Fig. 6b). This is also true for terrestrial animals where “*Ca. Megaira*” is rarely found (Fig. 6a), albeit significantly abundant in samples (Fig. 6b). Differently from ciliates’ endosymbionts, intracellular bacteria of metazoans are usually located only in some specific organ or tissue<sup>50</sup>. Up to now, the diversity of such ecosystems has probably not been sufficiently investigated, in particular concerning host species, but recent<sup>23,51,52</sup> and ongoing studies of “terrestrial” hosts may contribute in widening our knowledge about transmission of this poorly known symbiont between different hosts and environments. Most of the known *Rickettsiales* genera are usually associated to aquatic organisms including protists<sup>15</sup>, which strongly suggests that the last *Rickettsiales* common ancestor was hosted by an aquatic eukaryote. The finding of some *Rickettsiales* species from different families hosted by arthropods<sup>40,53</sup> indicates that this “terrestrial” shift of *Rickettsiales* has evolved independently several times in some lineages<sup>4,15</sup>. Consistently with this consideration, also “*Ca. Megaira*” could have been able to shift to “terrestrial” environment. Notably, protists are often associated to “*Ca. Megaira*”, and might serve as environmental vectors of these bacteria. Indeed, each of four major clades of “*Ca. Megaira*” has at least one representative found in symbiotic association with unicellular eukaryotes.

## Methods

**Hosts isolation, cultivation and identification.** Several strains of ciliates were used in this study (Table 1). Ciliate cultures were maintained in lettuce medium inoculated with *Klebsiella aerogenes* at +18 °C (Sanyo climatic chamber). For *P. bursaria* illumination regime of 10 hours darkness and 14 hours light by 2000 lx lamps was used. Cultures are available and maintained at the RR CCM Core Facilities Centre “Culture Collections of Microorganisms” of St. Petersburg State University, Russia. Identification of ciliate species was performed morphologically<sup>54</sup>, and then confirmed by molecular analysis.

**DNA extraction and molecular characterization.** Approximately 100 starved ciliate cells were repeatedly washed in sterile distilled water to minimize bacterial contamination, and then fixed in 70% ethanol. Total genomic DNA was extracted employing the NucleoSpin™ Plant II kit (Macherey-Nagel GmbH & Co., Dueren NRW, Germany) using the protocol for fungal DNA extraction.

Polymerase chain reactions (PCRs) were performed in a C1000™ Thermal Cycler (BioRad, Hercules, USA) with the high-fidelity TaKaRa Ex Taq (TaKaRa Bio, Inc., Otsu, Japan). All PCRs consisted of 35 cycles with a preliminary denaturation step at 94 °C for 3 min, then for each cycle denaturation at 94 °C for 30 seconds, annealing at 55 °C for 45 seconds and elongation at 72 °C for 90 seconds, and a final elongation step at 72 °C for 5 minutes<sup>55</sup>. PCR products were estimated through electrophoresis on 1% agarose, purified with EuroGold Cycle Pure Kit (EuroClone®, Milano, Italy), and then sequenced with appropriate internal primers at GATC Biotech (Cologne, Germany) (for further details see Supplementary Text S2).

**Phylogenetic analyses.** The obtained 16S rRNA gene sequences were aligned automatically with the automatic aligner of the ARB software package<sup>56</sup> using the SILVA database release 128, and then manually edited in order to optimize base-pairing in the predicted rRNA stem regions. A total number of 130 sequences were selected from the family *Rickettsiaceae* (118 belonging to “*Ca. Megaira*”, used as ingroup, and 12 others as outgroup). Firstly, 119 sequences longer than 1397 bp were used to build the character matrix, then 11 short sequences were added, and phylogeny was inferred.

The optimal substitution model was determined with jModelTest<sup>57</sup> according to the Akaike Information Criterion. Phylogenetic analyses were performed using both ML and BI methods. ML tree was inferred with 1000 bootstrap pseudoreplicates using PhyML software 3.0<sup>58</sup>, while Bayesian Inference was carried out with MrBayes 3.2<sup>59</sup> employing three runs, with one cold and three heated Monte Carlo Markov chains each, for 1000000 generations with a burn-in of 25%. Runs were stopped after checking that the average standard deviation of split frequencies reached a value below 0.01.

**Probe design and fluorescence *in situ* hybridization.** Based on the obtained almost complete 16S rRNA gene sequences of bacterial symbionts, several probes were designed, targeting the whole “*Ca. Megaira*” genus, and for each of the most common clades A + D, B, C, and E, respectively (Table 2). Specificity was tested *in silico* both on RDP<sup>35</sup> and on TestProbe tool 3.0 (SILVA rRNA database<sup>60</sup>) allowing 0 and 1 mismatches (Table 2). Probes were synthesized and labelled with Cy3 or fluorescein by Eurofins GMBH (Ebersberg, Germany).

The protocol used for all fluorescence *in situ* hybridization (FISH) experiments was the one described by Szokoli *et al.*<sup>61</sup>. Experiments always included negative controls, namely experiments without the use of any probe, and hybridizations were performed using several formamide concentrations in the hybridization buffer (0, 15 and 30% v/v) to test the different stringency levels for the newly designed FISH probes. The almost universal bacterial probe EUB338 (5'-GCTGCCTCCCGTAGGAGT-3'<sup>62</sup>) was used to verify the presence of other intracellular bacteria in the host cells.

**Transmission electron microscopy.** *P. primaurelia* ThK-1, *P. caudatum* Sp 9-5, *P. bursaria* strains 1M-2<sup>(T)</sup>, VL3-1, VL12-10, *P. nephridiatum* Sr 2-6, and *P. putrinum* ETu 7-4 were prepared for electron microscopy as described by Nitla *et al.*<sup>63</sup>. Briefly, cells were fixed in 2.5% glutaraldehyde and 1.6% paraformaldehyde in phosphate buffer (0.1 M, pH 7.4), with a post-fixation in 1.5% OsO<sub>4</sub>. Afterwards cells were dehydrated at increasing percentages of ethanol solutions, and finally embedded in Epoxy embedding medium (Fluka, BioChemika). Ultrathin sections were stained with uranyl acetate followed by lead citrate. Samples were observed using JEOL JEM-1400 (JEOL, Ltd., Tokyo, Japan) electron microscope.

**Atomic force microscopy.** *P. bursaria* VL12-10 cells were briefly washed in water, collected in a small drop, squashed on a cover slip and air-dried. The images were obtained with NTEGRA AURA microscope in a semi-contact mode.

**Screening of 16S rRNA gene datasets.** Raw sequencing 16S rRNA gene reads derived from more than 111000 microbiome samples were screened using the platform IMNGS<sup>33</sup>, in order to use the results for investigating both diversity and environmental distribution of “*Ca. Megaira*” in the environments and in potential hosts. The almost complete 16S rRNA genes of Clade A - “*Ca. Megaira polyxenophila*” (Sp 11-8, AB688629), Clade B (HE648945, HE648946, FJ203077, DQ395479), Clade C - “*Ca. Megaira venefica*” (1M-2<sup>(T)</sup>, KT851791), Clade D (EU640051, KC189769), Clade E (KT851820, GQ870455), and the stand-alone sequence from wastewater (CU466797) were selected as representatives of “*Ca. Megaira*” genus and used as queries.

Sequences longer than 300 bp, and having at least 95% of identity with the queries were selected for analysis. Sequences retrieved from all queries were pooled together, and were separated in three datasets according to hypervariable regions V1-2, V4-6, V7-8, and then clustered in OTUs with a threshold at 99% using UCLUST<sup>64</sup>. Afterwards, OTUs were aligned with MUSCLE<sup>65</sup>, and phylogenetic analyses were performed using FastTree<sup>66</sup>. To attribute an environmental provenance to the OTUs, each sequence clustering in the same OTU was screened and assigned to its environment, according to the provenance of the original sample. Thus, a single environment was assigned to the respective OTU when more than 50% of sequences had the same origin, while two environments were considered when two ecosystems were equally dominant. The category “Unclassified” was used when more than two environments were dominant.

In order to further investigate environmental distribution of each “*Ca. Megaira*” clade (Clade A - “*Ca. Megaira polyxenophila*”, Clade B, Clade C - “*Ca. Megaira venefica*”, Clade D, and Clade E), two diverse indices were calculated: frequency of occurrence and relative abundance. Frequency of occurrence was estimated as the number of microbiome samples positive for “*Ca. Megaira*” in each category divided the total number of samples of the category (i.e. Freshwater; Seawater; Soil; Artificial/Anthropogenic including sequences associated to wastewater, food, activated carbon, bioreactor and activated sludge; Aquatic animals; Plants; Terrestrial animals). Relative abundance of “*Ca. Megaira*” for each specific environment was calculated as the average of relative abundances in “*Ca. Megaira*” positive samples (i.e. the number of “*Ca. Megaira*” reads divided total number of reads for each sample; for further details see Supplementary Text S3).

## References

- McQuiston, J. H. & Paddock, C. D. Public health: Rickettsial infections and epidemiology. In *Intracellular Pathogens II: Rickettsiales* (pp. 40–83). American Society of Microbiology (2012).
- Schrallhammer, M. *et al.* *Candidatus Megaira polyxenophila* gen. nov., sp. nov.: considerations on evolutionary history, host range and shift of early divergent *Rickettsiales*. *PLoS One* **8**, e72581 (2013).
- Mieog, J. C. *et al.* The roles and interactions of symbiont, host and environment in defining coral fitness. *PLoS One* **4**, e6364 (2009).
- Weinert, L. A., Werren, J. H., Aebi, A., Stone, G. N. & Jiggins, F. M. Evolution and diversity of *Rickettsia* bacteria. *Bmc Biol* **7**, 6 (2009).
- Martijn, J. *et al.* Single-cell genomics of a rare environmental alphaproteobacterium provides unique insights into *Rickettsiaceae* evolution. *ISME J* **9**, 2373–2385 (2015).
- Balraj, P., Renesto, P. & Raoult, D. Advances in *Rickettsia* pathogenicity. *Ann NY Acad Sci* **1166**, 94–105 (2009).
- Wang, Z. & Wu, M. An integrated phylogenomic approach toward pinpointing the origin of mitochondria. *Sci Rep-UK* **5**, 7949 (2015).
- Perlman, S. J., Hunter, M. S. & Zchori-Fein, E. The emerging diversity of *Rickettsia*. *P R Soc Lond B Bio*, **273**, 2097–2106 (2006).
- Ferrantini, F. *et al.* “*Candidatus Cryptoprodotis polytropus*,” A Novel *Rickettsia*-Like Organism in the Ciliated Protist *Pseudomicrothorax dubius* (Ciliophora, Nassophorea). *J Eukaryot Microbiol* **56**, 119–129 (2009).
- Vannini, C. *et al.* Flagellar movement in two bacteria of the family *Rickettsiaceae*: a re-evaluation of motility in an evolutionary perspective. *PLoS One* **9**, e87718 (2014).
- Hess, S. Description of *Hyalodiscus flabellus* sp. nov. (Vampyrellida, Rhizaria) and Identification of its Bacterial Endosymbiont, “*Candidatus Megaira polyxenophila*” (*Rickettsiales*, *Alphaproteobacteria*). *Protist* **168**, 109–133 (2017).
- Zaila, K. E. Diversity and Universality of Endosymbiotic *Rickettsia* in the Fish Parasite *Ichthyophthirius multifiliis*. *Front Microbiol* **8**, 189 (2017).
- Castelli, M. *et al.* The Hidden World of *Rickettsiales* Symbionts: “*Candidatus Spectririckettsia obscura*,” a Novel Bacterium Found in Brazilian and Indian *Paramecium caudatum*. *Microb ecol*, <https://doi.org/10.1007/s00248-018-1243-8> (2018).
- Yurchenko, T. *et al.* A gene transfer event suggests a long-term partnership between eustigmatophyte algae and a novel lineage of endosymbiotic bacteria. *ISME J* **12**, 2163–2175 (2018).
- Castelli M., Sasser D. & Petroni G. Biodiversity of “Non-model” *Rickettsiales* and their association with aquatic organisms. In *Rickettsiales biology, molecular biology, epidemiology, and vaccine development*, (pp. 59–94). Springer (2016).

16. Vannini, C., Petroni, G., Verni, F. & Rosati, G. A bacterium belonging to the *Rickettsiaceae* family inhabits the cytoplasm of the marine ciliate *Diophrys appendiculata* (Ciliophora, Hypotrichia). *Microb ecol* **49**, 434–442 (2005).
17. Fraune, S. & Bosch, T. C. Long-term maintenance of species-specific bacterial microbiota in the basal metazoan. *Hydra*. *P Natl Acad Sci USA* **104**, 13146–13151 (2007).
18. Sunagawa, S. Bacterial diversity and White Plague Disease-associated community changes in the Caribbean coral *Montastraea faveolata*. *ISME J* **3**, 512–521 (2009).
19. Hollants, J., Leliaert, F., Verbruggen, H., Willems, A. & De Clerck, O. Permanent residents or temporary lodgers: characterizing intracellular bacterial communities in the siphonous green alga *Bryopsis*. In *Proc. R. Soc. B*, **280**, 20122659 (2013).
20. Kawafune, K. *et al.* Two different rickettsial bacteria invading *Volvox carteri*. *PLoS One* **10**, e0116192 (2015).
21. Yang, A., Narechania, A. & Kim, E. Rickettsial endosymbiont in the “early-diverging” streptophyte green alga *Mesostigma viride*. *J Phycol* **52**, 219–229 (2016).
22. Sun, H. Y. Endosymbiotic bacteria in the parasitic ciliate *Ichthyophthirius multifiliis*. *Appl Environ Microb* **75**, 7445–7452 (2009).
23. Dishaw, L. J. *et al.* The gut of geographically disparate *Ciona intestinalis* harbors a core microbiota. *PLoS One* **9**, e93386 (2014).
24. Miyake, S., Ngugi, D. K. & Stingl, U. Phylogenetic Diversity, Distribution, and Cophylogeny of Giant Bacteria (*Epulopiscium*) with their Surgeonfish Hosts in the Red Sea. *Front Microbiol* **7**, 285 (2016).
25. Murakami, T. *et al.* Bacterial Microbiota Associated with the Glacier Ice Worm Is Dominated by Both Worm-Specific and Glacier-Derived Facultative Lineages. *Microbes Environ* **32**, 32–39 (2017).
26. Frischer, M. E. *et al.* Bacterial community structure of acid-impacted lakes: what controls diversity? *Appl Environ Microb* **74**, 1856–1868 (2008).
27. Vlahos, N. *et al.* Changes of bacterioplankton apparent species richness in two ornamental fish aquaria. *SpringerPlus* **2**, 66 (2013).
28. Chouari, R. *et al.* Molecular analyses of the microbial community composition of an anoxic basin of a municipal wastewater treatment plant reveal a novel lineage of *Proteobacteria*. *Microb ecol* **60**, 272–281 (2010).
29. Jangid, K. *et al.* Relative impacts of land-use, management intensity and fertilization upon soil microbial community structure in agricultural systems. *Soil Biol Biochem* **40**, 2843–2853 (2008).
30. Görtz, H. D. Intracellular bacteria in ciliates. *Int Microbiol* **4**, 143–150 (2010).
31. Fokin, S. I. Bacterial endocytobionts of ciliophora and their interactions with the host cell. *Int Rev Cytol* **236**, 181–249 (2004).
32. Anderson, O. R. *Comparative protozoology: ecology, physiology, life history*. Springer Science & Business Media (2013).
33. Lagkouravdos, I. *et al.* IMGs: A comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. *Sci Rep-UK*, **6** (2016).
34. Modeo, L., Petroni, G., Lobban, C. S., Verni, F. & Vannini, C. Morphological, ultrastructural, and molecular characterization of *Euplotidium rosati* n. sp. (Ciliophora, Euplotida) from Guam. *J Eukaryot Microbiol* **60**(25–36), 2013 (2013).
35. Kim, M., Oh, H. S., Park, S. C. & Chun, J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Micr* **64**, 346–351 (2014).
36. Hine, P. M., Wakefield, S., Diggles, B. K., Webb, V. L. & Maas, E. W. Ultrastructure of a haplosporidian containing *Rickettsia*, associated with mortalities among cultured paua *Haliotis iris*. *Dis Aquat Organ* **49**, 207–219 (2002).
37. Cole, J. R. *et al.* Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res*, gkt1244 (2013).
38. Gong, J., Qing, Y., Guo, X. & Warren, A. “*Candidatus* Sonnebornia yantaiensis”, a member of candidate division OD1, as intracellular bacteria of the ciliated protist *Paramecium bursaria* (Ciliophora, Oligohymenophorea). *Syst Appl Microbiol* **37**, 35–41 (2014).
39. Sommaruga, R. & Sonntag, B. Photobiological aspects of the mutualistic association between *Paramecium bursaria* and *Chlorella*. In *Endosymbionts in Paramecium* (pp. 111–130). Springer Berlin Heidelberg (2009).
40. Schulz, F. *et al.* A *Rickettsiales* symbiont of amoebae with ancient features. *Environ Microbiol*. <https://doi.org/10.1111/1462-2920.12881> (2015).
41. Sasser, D. *et al.* Phylogenomic evidence for the presence of a flagellum and *cbb 3* oxidase in the free-living mitochondrial ancestor. *Mol Biol Evol* **28**, 3285–3296 (2011).
42. Sabaneyeva, E. *et al.* Host and symbiont intraspecific variability: The case of *Paramecium calkinsi* and “*Candidatus* Trichorickettsia mobilis”. *Eur j protistol* **62**, 79–94 (2018).
43. Boussau, B., Karlberg, E. O., Frank, A. C., Legault, B. A. & Andersson, S. G. Computational inference of scenarios for  $\alpha$ -proteobacterial genome evolution. *P Natl Acad Sci USA* **101**, 9722–9727 (2004).
44. Boscaro, V. *et al.* Rediscovering the genus *Lyticum*, multiflagellated symbionts of the order *Rickettsiales*. *Sci Rep-UK* **3**, 3305 (2013).
45. Kawafune, K., Sato, M., Toyooka, K. & Nozaki, H. Ultrastructure of the rickettsial endosymbiont “MIDORIKO” in the green alga *Carteria cerasiformis* as revealed by high-pressure freezing and freeze-substitution fixation. *Protoplasma* **250**, 949–953 (2013).
46. Anderson, D. R., Hopps, H. E., Barile, M. F. & Bernheim, B. C. Comparison of the ultrastructure of several *Rickettsiae*, ornithosis virus, and *Mycoplasma* in tissue culture. *J Bacteriol* **90**, 1387–1404 (1965).
47. Potekhin, A. *et al.* Complex life cycle, broad host range and adaptation strategy of the intranuclear *Paramecium* symbiont *Preeria caryophila* comb. nov. *FEMS microbiol ecol*, <https://doi.org/10.1093/femsec/fiy076> (2018).
48. Haas, B. J. *et al.* Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome research* **21**, 494–504 (2011).
49. Park, T. & Yu, Z. Do ruminal ciliates select their preys and prokaryotic symbionts? *Front Microbiol*, **9** (2018).
50. Wang, M. *et al.* Tissue localization and variation of major symbionts in *Haemaphysalis longicornis*, *Rhipicephalus haemaphysaloides* and *Dermacentor silvarum* in China. *Appl Environ Microb*, AEM-00029 (2018).
51. Bodenhausen, N., Horton, M. W. & Bergelson, J. Bacterial communities associated with the leaves and the roots of *Arabidopsis thaliana*. *PLoS one* **8**, e56329 (2013).
52. Koneru, S. L., Salinas, H., Flores, G. E. & Hong, R. L. The bacterial community of entomophilic nematodes and host beetles. *Mol Ecol* **25**, 2312–2324 (2016).
53. Mediannikov, O. *et al.* High quality draft genome sequence and description of *Occidentia massiliensis* gen. nov., sp. nov., a new member of the family *Rickettsiaceae*. *Stand Genomic Sci* **9**, 9 (2014).
54. Fokin, S. I. *Paramecium* genus: biodiversity, some morphological features and the key to the main morphospecies discrimination. *Protistology*, **6** (2010).
55. Modeo, L. *et al.* Morphology, ultrastructure, and molecular phylogeny of the ciliate *Sonderia vorax* with insights into the systematics of order Plagiopylida. *BMC microbiol* **13**, 40 (2013).
56. Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Res* **32**, 1363–1371 (2004).
57. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* **9**, 772–772 (2012).
58. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307–321 (2010).
59. Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* **61**, 539–542 (2012).
60. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*, gks1219 (2012).
61. Szokoli, F. *et al.* “*Candidatus* Fokinia solitaria”, a novel “stand-alone” symbiotic lineage of *Midichloriaceae* (*Rickettsiales*). *PLoS One* **11**, e0145743 (2016).



62. Amann, R. I. *et al.* Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microb* **56**, 1919–1925 (1990).
63. Nitla, V. *et al.* Critical revision of the family Plagiopylidae (Ciliophora: Plagiopylea), including the description of two novel species, *Plagiopyla ramani* and *Plagiopyla narasimhamurtii*, and redescription of *Plagiopyla nasuta* Stein, 1860 from India. *Zool J Linn Soc*, <https://doi.org/10.1093/zoolinnea/zly041> (2018).
64. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
65. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797 (2004).
66. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* **26**, 1641–1650 (2009).

## Acknowledgements

We would like to thank Simone Gabrielli for his technical assistance in TEM sections preparation, Konstantin Benken for his help with atomic force microscopy, and Prof. Fabio Bulleri for his helpful suggestion with environmental distribution part. Experiments and sequencing were partly performed using the equipment of the Core Facility Centers “Cultivation of Microorganisms”, “Molecular and Cell Technologies”, “Microscopy and Microanalysis” of St. Petersburg State University. Work of O.L. in St. Petersburg State University was supported by RFBR 16-34-50043. Molecular analyses and TEM experiments were partly performed with support from RFBR grants 16-04-01195 to A.P., from 15-04-06410 to E.S., and from University of Pisa project PRA\_2016\_58 to G.P. Fluorescent microscopy observations were made possible with support of RSF grant 16-14-10157 to A.P. European Commission FP7-PEOPLE-2009-IRSES project CINAR PATHOBACTER (247658) supported mobility to O.L., E.S., N.L., M.S., A.P., G.P.

## Author Contributions

Conceived and designed the experiments: O.L., G.P. Performed the molecular investigations: O.L., M.S. Transmission electron microscopy analysis: O.L., E.S., L.M. Phylogenetic analysis: O.L., M.C. Diversity and environmental investigations: O.L., Field work: O.L., N.L., M.S., A.P. Analyzed data: O.L., G.P. Contributed reagents/materials/analysis tools: F.V., M.S., A.P., G.P. Wrote the paper: O.L. All authors revised the final version of the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-37629-w>.

**Competing Interests:** The authors declare no competing interests.

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019