UNIVERSITÀ DI PISA

RESEARCH DOCTORATE SCHOOL

IN

BIOLOGICAL AND MOLECULAR SCIENCES

DOCTORAL DISSERTATION

GENES, GENOMES, AND TRANSCRIPTOMES OF CILIATES AND THEIR PROKARYOTIC ENDOSYMBIONTS

PH.D. CANDIDATE:

Vittorio Boscaro

SUPERVISOR:

Prof. Franco Verni

Chapter 1.

Studying symbiosis: models, methods, and angles

Symbiosis is a biological phenomenon of huge importance but also difficult to circumscribe. Different authors employ the term with different meanings, all arbitrary to a certain extent; choosing one depends more on practical purposes than true profound theoretical differences. The most complete definition I know of is that of the International Society of Symbiosis, reciting: "symbiosis is an intimate, permanent relationship between partners of different species, that results in new structures or new metabolisms, and that can involve genic exchange". It is a good definition that provides objective parameters and stresses the most important outcome of the process: the promotion of emerging features. However, as is often the case with complex definitions in biology, it would be disingenuous to pretend that this is the one most often applied in practical cases. Scientific accuracy would require years of study before proving that a certain relationship actually falls within these boundaries. What we usually consider symbiosis (and the definition employed in this thesis) coincides with the older, broader concept of Anton De Bary: "the living together of unlike organisms". Or, more technically put, "a temporally protracted and spatially close relationship between organisms belonging to different species".

Definitions used by symbiontologists, thus, do not include at all the *type of effect* that the relationship produces on the partners involved. Ecological terms like commensalism, parasitism, or mutualism may all be included in symbioses, when the interaction between the organisms is sufficiently strict (in contrast with some of the more "medical" interpretations, that equalize symbiosis and mutualism). This is in itself an indication of the huge diversity that symbiotic systems may show. A second dimension of variability is that of the actual organisms participating. Symbioses truly are ubiquitous, and all major lineages include countless examples. All vertebrates, to name one, require extremely complex communities of essential prokaryotes to survive – and harbor just as numerous other less well-meaning species. Symbiosis

is literally everywhere; it is one of the most important drives of evolution; and can be extremely varied.

For all these reasons, it is difficult to define a general field of "symbiontology" – although a few common patterns and keys to understand the phenomenon do exist. Researchers are more often defined according to the group they study.

Ciliates (phylum Ciliophora) are relatively well-known exponents of a generally neglected universe, that of protozoa – or, using a more current term, protists. "Protists" themselves are tricky to define (I personally still prefer the now disused definition of "all eukaryotes that are not plants, fungi, or metazoans". It is unambiguous, if lacking in descriptive power), but ciliates are a rare example of a microbial group that won the trials of time, and were always more or less considered a solid evolutionary unit. Ciliates are unicellular (sometimes colonial) eukaryotes, relatively large, extremely common, and useful as models for a variety of studies, including those on symbiosis. Probably because of their phagotroph heterotrophic feeding, they are often colonized by "symbiotic" endocellular prokaryotes (there are also striking examples of ectosymbiosis, with bacteria adhering to the external surface of the host, but they will not be treated in this thesis but for cursory citations). The study of these interactions, pioneered among others by Tracy Sonneborn, Anthony Soldo, John and Louise Preer and later led by Klaus Heckmann, Hans-Dieter Görtz, Masahiro Fujishima and Sergei Fokin, is about a century old, and is currently expanding. Symbionts of ciliates are intriguing for several reasons, expanded later in this thesis, and are extremely varied in their taxonomic affiliation and type of effect on their host ("symbiont" and "host" may be ambiguous terms, sometimes – but conventionally, the prokaryote is always called "symbiont" and the ciliate "host").

The study of ciliates, having a considerably long history, evolved with the rest of biology (maybe lagging only a bit behind). Traditionally investigated by zoologists, ciliates were first examined at the light microscope, then dissected with electron microscopy and subjected to genetic inquiry, and eventually targeted by molecular methods. Molecular markers are now essential tools in most studies dealing with these microbial eukaryotes, and complete genomes and transcriptomes are being sequenced.

The methodologies targeting their symbionts, of course, followed a similar trajectory. When looking at the literature, there is a clear demarcation between the early discovered "classic symbionts" and those only recently described. Until the last decade of the XX century, there were no molecular sequences of ciliate-inhabiting prokaryotes, and they were thus often difficult to identify. It was already known that most ciliates harbor "particles" resembling bacteria, but many of them were almost featureless. On the contrary, some exhibited eye-catching morphological features, and were much more intensely studied. In the last two decades, molecular biology allowed to explore the elusive diversity, and a large number of novel species were detected and described.

In the projects presented in this thesis, the approach is mainly molecular. At the beginning, I employed the standard molecular techniques applied to ciliates and bacterial symbionts in the protistology lab of the University of Pisa, including DNA extraction, PCR, cloning, and Fluorescence In Situ Hybridization. Then I became particularly interested in refining phylogenetic methods, studying Bayesian inference and nucleotide substitution models. Eventually, during my PhD, my work shifted to approaches never applied before on these systems, like performing genomics on symbionts and single-cell transcriptomics on hosts.

Another characteristic of most of my projects was multidisciplinarity. While personally focusing on molecular techniques, I learned, used and worked with students on more traditional methods, like morphological and ultrastructural observation to provide more complete and robust results. I will argue in later chapters about the huge importance of this approach in characterizations.

Symbiosis is thus an intriguing topic, ciliates good models to investigate it, and molecular biology an appropriate way to do it. Theoretically, many angles could be interesting and profitable. There is, however, a serious problem: not much is known about these systems to begin with. There is a specialized literature, but it is much poorer than what would be desirable. Studying the complex relationship between two or more organisms is in itself a complex issue, with countless variables. It would be totally fruitless without a sound framework on which assumptions could be made.

That is what characterizations are for. In order to study symbiosis and interactions, we must first know something – as much as we can – about the organisms involved. We should, to some extent, *describe* them. This is the first step of any scientific enterprise, a necessary one when data is clearly lacking, and the angle I chose to follow during my PhD work. The next chapters will all deal with characterizations of ciliates and their bacterial endosymbionts, at different levels and with different approaches, starting with single-marker phylogenies and proceeding with analyses of entire genomes and transcriptomes. During the travel, many new things were discovered, hypotheses advanced, and the foundations for future projects laid.

Structure of the thesis

The main part of this thesis is organized in 8 Chapters distributed in 4 Parts. Each Chapter derives from a different paper that I have coauthored during the years of my PhD research (2012-2014). Some Chapters are followed by Supplementary Materials, with additional elements that were separated from the main text for reasons of fluency, and Afterwords reporting significant developments occurred since the publication of the original manuscript. Three papers were omitted because I played only a secondary role in their execution, but are briefly cited and summarized when closely related to subjects at hand. A complete list of my publications and contributions is reported in the Appendix.

Each Part is introduced by a short foreword introducing the topic and summarizing the following Chapters. **Part I** collects my results on "traditional" characterizations of bacterial symbionts harbored by ciliates, employing the "Full Cycle rRNA Approach" and ancillary techniques. **Part II** is dedicated to the major initial goal of my PhD project, the first study of a symbiont at the genomic level. As the first two Parts are focused on bacteria, the other two deal with ciliates. **Part III** reports the characterization and survey of symbiont-bearing ciliate genera, while **Part IV** summarizes my most important result in the field of multi-marker characterizations: the development of a single-cell transcriptomic method for protists.

PART I

Characterization of Bacterial Symbionts

Bacterial symbionts of ciliates do not, by any means, form a coherent group. They are strongly heterogeneous in terms of their taxonomic affiliation, morphology, effect on their hosts and subcellular localization. The protocols employed to characterize them, however, usually consist of similar steps. Thus, the field of study dealing with these prokaryotes is mostly defined methodologically, by a set of well-established techniques.

The basic characterization of a bacterial symbiont requires information on the host and the harboring compartment; a general description of the symbiont's morphology, preferably including ultrastructural features; and a high-quality, confirmed sequence of a reliable molecular marker – conventionally, that of the 16S rRNA gene. The last condition is met through the Full-Cycle rRNA Approach.

The Full-Cycle rRNA Approach is widely used to identify uncultivable symbionts, and has been employed on ciliates since 1990. The cycle requires a sufficient amount of ciliate cells fixed in two ways: in tubes (usually with ethanol) to perform DNA extractions, and on slides (with paraformaldehyde or osmium tetroxide, for example) for fluorescence in situ hybridization (FISH) experiments. Assuming that no previous information is available, oligonucleotidic fluorescent probes with broad specificity are employed to assess the presence of bacteria inside the eukaryotic cells, and provide preliminary cues on their identity (at the phylum or class level). The collected insights are then used to obtain the 16S rRNA gene sequence through PCR on the extracted DNA employing low-specificity primers, followed by direct sequencing (in the easier cases) or cloning of the amplicons and sequencing of the clones. It is then possible to design a new, species-specific (in practical terms, sequence-specific) probe that can be tested on the slides. The fluorescent signal should confirm the correctness of the sequence and its belonging to the presumed symbiont.

A successful Full-Cycle rRNA Approach confirms the presence of symbionts (and their approximate localization and number), assesses their identity, and assures repeatability: the complete 16S rRNA gene sequence and the validated sequence-specific probe will serve as diagnostic characters to allow the identification of the taxon in future investigations. Nevertheless, additional information is always welcomed when describing a new species – in particular, morphological features. Morphology is generally not very reliable on prokaryotes because the number of observable characters is limited and plesiomorphy common; as a whole, bacteria exhibit a lamentably shallow range of forms. There are impressive exceptions, though,

and plenty of them. Conspicuity of the look is certainly a biased, human-centric measure of an organism's interest, but it is a useful one nonetheless. Not only it provides an easy way to recognize taxa, but it also reflects adaptations and hints at possible functions. Several symbionts, and among them some symbionts of ciliates, exhibit striking ultrastructural features. The instruments to capture them are considerably older than the Full-Cycle rRNA Approach.

Light microscopy is still the first tool to detect bacteria inside ciliates. Bacterial symbionts are visible at dissecting microscope magnifications, although most just look like plain rod-like particles. Some, however, are clearly identifiable even at this stage, like the huge, compartmentalized infectious forms of *Holospora* or the refractive R-body bearing individuals of *Caedibacter*. And some catch the eye for a different reason: their ability to move, swimming (sometimes spectacularly fast) within the host in an unsettling parasite dance, and yet rarely killing it, or damaging it in any obvious way. More features are revealed, of course, through transmission electron microscopy. Not only conspicuous structures can be scrutinized in more detail, but other characters can be checked, like the type of cell wall and the granularity of the cytoplasm. Consequently, more surprises can arise: the presence of flagella, for example, or viral capsid-like particles that forms a putative tripartite symbiosis.

Strictly speaking, the Full-Cycle rRNA Approach ends with the validation of a molecular sequence, but it is phylogenetic inference that leads to the correct identification of the organism. Single-marker phylogenies are now within the reach of any computer, but performing them properly is a different (and lengthy) matter. While artifacts and systematic errors are relatively well known issues, phylogenetic trees (and the reliability of their nodes) should also be tested for subtler factors. Classic methods like bootstrapping are universally used, but emphasis is rarely placed on the influence of the character matrix, for example. Phylogenetic workflows have intrinsic flaws, and only the most robust of results, obtained with several independent methods, should be presented as conclusive. That said, correctly interpreted phylogenetic trees are powerful tools, and they tend to slowly converge toward a consensus over time. Several working hypotheses can come to mind just looking at a tree, thus they are almost always associated with characterization studies today.

Phylogenetic inferences are also the most important factor in systematics, and they guide taxonomy. Nearly any other topic may seem more intriguing to the average biologist, but correctly naming organisms is indeed an important task. As is recognizing monophyletic groups,

and from time to time (but not too often) rearranging the existing classification to match the improved understanding of the relationships among lineages. Nomenclatural rules, although essential to proceed with good order, are not the core of it. The most important thing to keep in mind is that classification, like it or not, guides biological thinking. Whenever a conclusion is made from even the least taxonomically inclined experimenter, it is classification that creates the categories allowing him or her to make generalizations. A mouse is not a rat, rodents and humans are not Mammalia, and *Escherichia coli* is not *Bacteria*.

Speaking of nomenclatural rules, a recently novel one, widely used although a bit confusing, is that uncultivable bacteria must be described only as "*Candidatus*" taxa. The "*Candidatus*" status follows a set of separate rules and has a different list of requirements for being used, but it is at least an improvement after a period when new obligate symbionts were not named, because they could not be kept in pure cultures.

Full-Cycle rRNA Approach, ultrastructural observations and phylogenetic inference on a single marker form the core of the works presented in the next four chapters. Most of these studies started in a similar fashion: sampled ciliates were preliminary screened at the dissecting microscope and, when promising, with a set of fluorescent probes. If an interesting signal was present, the complete approach was performed. Currently, no systematic work has been performed to quantitatively assess the frequency and type of ciliate/prokaryote relationships in a certain area, or in a broad taxonomic group (although reviews are periodically released, summarizing the available data). This may seem odd, but there are actually good explanations for it. Some are practical: the methodologies are well established, but also time-consuming, requiring a lot of handwork and well-behaving ciliate populations. Additionally, there are few researchers working on this topic. But in addition to that, the field is still relatively in its infancy; the impression is that new characterizations result in novel taxa essentially at the same rate as before, suggesting that we are far from a good coverage of the total biodiversity we are dealing with. Granted, some patterns start to emerge and "common" symbionts are found more or less regularly; but the number of "rare" ones is ever-growing. Thus, there is simply no complete set of primers and probes to readily screen ciliates in search of known symbionts. Nor the database of already characterized taxa would be enough for culture-independent approaches.

Admittedly, similar problems entail that the organisms actually found are not, by any

stretch, unbiasedly distributed. Not all potential hosts are investigated with the same frequency, to begin with. This is not related as much to the actual abundance of ciliate species in the environment, as it is to their predisposition for culturing. *Paramecium* and *Euplotes* are easy to grow in lab compared with most other ciliates, and for this reason alone we have much more data on their symbiotic relationships than on all other species combined. Secondly, it is faster to detect bacteria that are similar to other found previously, because of the availability of diagnostic instruments. It is thus inevitable to search preferentially certain groups during pilot studies, and consequently to find more bacteria belonging to these groups. The result is that the taxonomic distribution of known symbionts is also strongly biased, and at this point very few conclusions can be drawn on this subject.

Summarizing, there is currently no alternative approach to fill the gaps in our knowledge about the biodiversity of prokaryotes "inhabiting" ciliates. The other face of the coin, however, is that there remains a lot to discover, and any project about the identification of a bacterial symbiont rarely ends up in disappointment.

I have participated in such projects during the last six years in the unit of Protistology of the University of Pisa, where the general protocols (with the possible exception of phylogenetic inference) were already refined when I started. Some of the work presented here was completely performed in Pisa, while others took advantage from the collaboration of researchers in Europe and Russia. This line of inquiry is still active, and was a constant during the entire period of my PhD work, intended as the first step toward larger and more ambitious projects.

Chapter 2 presents the first description of "*Candidatus* Nebulobacter yamunensis", one of the relatively few gammaproteobacterial symbionts found in ciliates so far. The other two – a *Francisella noatunensis* subspecies and *Caedibacter taeniospiralis* – are part of the same family-like clade, although they all clearly became symbionts of ciliates independently. "*Candidatus* Nebulobacter yamunensis" is an example of a taxon without a (known) defining feature that set it immediately apart from others, but its characterization was an occasion to tackle a still unorganized group and better define it from a taxonomic perspective.

Different is the case of "*Candidatus* Gortzia infectiva", presented in **Chapter 3**. This organism belongs to the so called "*Holospora*-like bacteria" (HLB), renowned ciliate symbionts

with a striking morphology and a peculiar, infective life cycle. The description took this into account and surveyed some aspects of the bacterium's biology, like its transmission ability; but the work itself was also intended as an updated review on HLB, their evolution and taxonomy.

Like HLB, the bacteria characterized in **Chapter 4** – "*Candidatus* Defluviella procrastinata" and "*Candidatus* Cyrtobacter zanobii" – belong to *Alphaproteobacteria*, but in a different lineage: the candidate family "*Midichloriaceae*". "*Candidatus* Midichloriaceae" seems a rich source of symbionts of ciliates, but also sponges, cnidarians, arthropods and other eukaryotes. Their phylogenetic history, as presented in the chapter, is only resolved in one aspect: there is no congruence whatsoever between the branching of hosts and symbionts. This in itself is a strong evidence of the infective ability of this clade's members, even in the (sad) absence of direct experimental evidence.

To "*Candidatus* Midichloriaceae" belongs the genus *Lyticum*, too, discussed in **Chapter 5**. There is however almost no other similarity between the two works. First, the genus *Lyticum* is not newly established (thus does not require the label "*Candidatus*"), but is indeed a very old one, named more than 50 years ago (but known well before then) and now described with modern methods for the first time. More importantly, this genus is another classic example of morphologically unique symbionts, displaying a conspicuous flagellar system whose function is unknown. Its assignment to the order *Rickettsiales*, lacking any detectable trace of flagella until a few years ago, is an important cornerstone: in 2011, non-functional flagellar genes were found in the genome of a related bacterium; then *Lyticum* provided a full, but non-motile, flagellar apparatus; and shortly after another symbiont of ciliates belonging to the same order was formally described by our unit (as discussed in the Afterword of Chapter 5), possessing flagella *and* flagellar movement. These evidences piled up to suggest a novel inference on the features of the most recent common ancestor of *Rickettsiales*, which is, in fact, also the ancestor of mitochondria. This provides an interesting link between the study of something so specialized as bacterial symbionts of ciliates and the evolution of eukaryotic organelles.

Chapter 2.

"*Candidatus* Nebulobacter yamunensis"

Original publication:

Characterization of "*Candidatus* Nebulobacter yamunensis" from the cytoplasm of *Euplotes aediculatus* (Ciliophora, Spirotrichea) and emended description of the family *Francisellaceae*

Vittorio Boscaro, Claudia Vannini, Sergei I Fokin, Franco Verni, Giulio Petroni

Systematic and Applied Microbiology 35(7): 432-440. October 2012.

Abstract

Our knowledge of ciliate endosymbionts occurrence and diversity greatly expanded in the last decades, due to the development of characterization methods for uncultivable bacteria. Symbionts related to human pathogens such as rickettsiae and francisellae have been detected inside the cytoplasm of different ciliate species. In the present work, we have characterized a novel *Francisella*-related bacterium inside the rich prokaryotic community harbored by a population of *Euplotes aediculatus* (Ciliophora, Spirotrichea). Following the "Full-Cycle rRNA Approach" we obtained the almost full-length 16S rRNA gene sequence of this bacterium, and developed probes for diagnostic fluorescence *in situ* hybridizations. Attempts to culture the endosymbiont outside of its host failed. We classified this novel organism in a new taxon for which we propose the name "*Candidatus* Nebulobacter yamunensis". In order to investigate its evolutionary relationships, we have also performed phylogenetic analyses on the class *Gammaproteobacteria* and the order *Thiotrichales*, which include the monogeneric family *Francisellaceae*. We found highly supported evidences for the establishment of a new monophyletic taxon including *Francisella* species, other organisms currently *incertae sedis*, and "*Candidatus* Nebulobacter yamunensis". These organisms form a clade sharing a signature

sequence not present in other *Thiotrichales* bacteria. Moreover, most of them have developed an intracellular life cycle inside eukaryotic organisms. We emended the original description of family *Francisellaceae* in order to encompass all members of the described clade.

Introduction

Bacteria inhabiting the cytoplasm or a different compartment of an eukaryotic cell are usually labeled as "endosymbionts". The biodiversity and frequency of this category of bacteria in ciliates (Ciliophora, Alveolata) have been recently proven to be very high[1-17]. One of the most striking features of these ciliate-bacteria relationships is their variability. The endosymbiont can act as a parasite[18] as well as a mutualist providing advantages for its host in different systems[19,20] or under different conditions[21-23]. It can be highly infectious or rely on vertical inheritance[18,24]. In most cases, these intracellular bacteria show some kind of dependence upon their host, and don't grow on standard culture media outside of their natural habitat in the eukaryotic cell[16]. As can be expected from such different behaviors, they belong to different taxonomic groups[4,6-8,10,11,13], and their adaptations to the endosymbiotic life-style have evolved many times independently.

Some of the bacterial endosymbionts of ciliates have received more attentions than others[16]. Bacteria belonging to the genera *Holospora* and *Caedibacter*, for example, have been studied for decades because of their peculiar morphological characteristics and the relevant effects that they produce on their hosts[25,26]. They are also harbored by one of the most studied ciliate taxon, the genus *Paramecium*. The genus *Euplotes*, another very well known ciliate, is usually cited by symbiontologists regarding the presence of the obligate endosymbiont *Polynucleobacter necessarius* in one clade of fresh- and brackish-water species[12,24,27]. Strains of these *Euplotes* species cannot divide properly when the bacteria are removed with ampicillin, and eventually die[20,27].

Since the introduction of reliable molecular methods to study uncultivable bacteria[28,29], a vast number of less-known and much more diverse organisms have been described. In most cases their ecological role, if any, is unknown. Most of them belong to the phylum *Proteobacteria*, and specifically to the classes *Alphaproteobacteria*, *Betaproteobacteria* and *Gamma-proteobacteria*[4,6,8,9,24,29]. Even those endosymbionts belonging to the same order can be only distantly related to each other[2,10]. Because the number of papers on this subject is growing, it is probable that the taxonomic diversity of bacterial endosymbionts of ciliates will increase.

One of the intriguing aspects of this field of research is that some of the recently described ciliate endosymbionts are phylogenetically related to known pathogenic bacteria. Some are

assigned to the group of "*Rickettsia*-like organisms"[2,8], bacteria closely related to the family *Rickettsiaceae* (*Alphaproteobacteria*), which includes the etiologic agents of diseases like epidemic typhus (*Rickettsia prowazekii*), Rocky Mountain spotted fever (*Rickettsia rickettsii*) and scrub typhus (*Orientia tsutsugamushi*)[30]. Until now, only one example is known of a ciliate endosymbiont belonging to the genus *Francisella* (*Francisellaceae*, *Gammaproteobacteria*), which also includes *Francisella tularensis*, the etiologic agent of tularemia[31]. "*Candidatus* Francisella noatunensis subsp. endociliophora" was discovered inside the species *Euplotes raikovi*[6]. *Caedibacter taeniospiralis*, an endosymbiont of *Paramecium tetraurelia*, is also related to family *Francisellaceae*, although less closely[32]. To our knowledge there are currently no dedicated studies, but it would be very interesting to test the infective capabilities of those ciliate endosymbionts related to human pathogens. This could even allow to uncover new environmental reservoirs of old diseases.

In this work, we have characterized another *Euplotes* endosymbiont related to francisellae, for which we propose the name "*Candidatus* Nebulobacter yamunensis", in accordance to the current rules of nomenclature for uncultivated bacteria[33,34]. We employed the "Full Cycle rRNA Approach"[28] to obtain its 16S rRNA gene sequence, and we developed a specific fluorescent probe in order to diagnose the presence of this novel endosymbiont in future studies. Moreover, we performed detailed phylogenetic analyses aiming not only to elucidate the phylogenetic relationships of the endosymbiont, but also to revise the taxonomy of family *Francisellaceae*. We have identified a more comprehensive taxon whose monophyly is highly supported by all phylogenetic methods employed. This clade includes the genus *Francisella*, *Caedibacter taeniospiralis*, the free-living bacterium *Fangia hongkongensis* and the newly characterized "*Candidatus* Nebulobacter yamunensis". We thus propose to extend the boundaries of the currently monogeneric family *Francisellaceae* Sjöstedt 2005 to encompass all these related organisms.

Materials and Methods

Ciliate sampling, culture and identification. *Euplotes aediculatus* population In was collected on February 2007 in an eutrophic freshwater pond near the Yamuna River (New Delhi; India). Several cells were isolated from the sample and cultured in the original medium

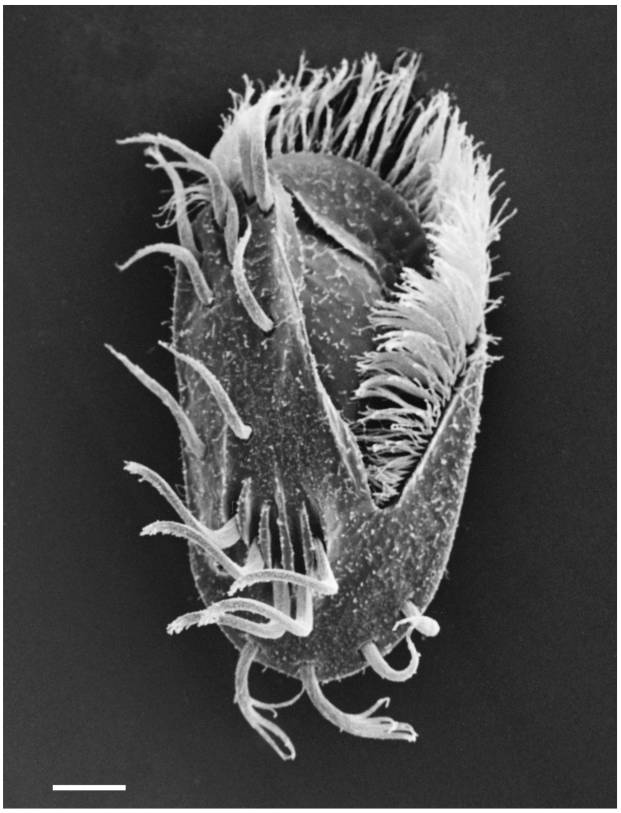


Figure 2.1 Ventral view of a *Euplotes aediculatus* cell of the population In at the scanning electron microscope. The bar corresponds to 10 µm.

periodically enriched with boiled rice grains and diluted with SMB (Synthetic Medium for *Blepharisma*[35]). The polyclonal culture was maintained at 19-20 °C on a 12:12 h irradiance of 200 µmol photons m^{-2} s^{-1}. All attempts to isolate monoclonal strains growing under controlled feeding conditions repeatedly failed (single cells didn't divide or divided only a few times when fed with *Dunaliella tertiolecta*, *Phaeodactylum tricornutum* or *Enterobacter aerogenes*). The species identification was performed through living and fixed observations with optical and scanning electron microscopes (Figure 2.1), as described by Modeo and coworkers[36]. The pattern of diagnostic characters, especially the number and position of frontoventral cirri, was in accordance with that of *E. aediculatus*[37]. The characterization of the 18S rRNA gene sequence (accession number: FR873713) of this ciliate is reported by Vannini and coworkers[9] and confirmed the identification.

DNA extraction and 16S rRNA gene sequencing. About 80 cells of the population In were individually collected, washed three times in sterile distilled water and stored in ethanol 70% at -22 °C. Total genomic DNA (tgDNA) was extracted using the NucleoSpin™ Plant II DNA extraction kit (Macherey-Nagel GmbH & Co., Düren NRW, Germany), following the protocol for fungal DNA extraction.

Polymerase chain reactions (PCRs) were performed in a C1000™ Thermal Cycler (BioRad, Hercules, CA) with the TaKaRa Ex *Taq* (TaKaRa Bio Inc., Otsu, Japan). All PCRs started with a preliminary denaturation step at 94 °C for 3 min and a final elongation step at 72 °C for 6 min; denaturation (94 °C), annealing and elongation (72 °C) steps were respectively 30 s, 30 s

and 90 s long in each cycle. A negative control was included in every reaction. 5 µL of PCR products were evaluated through electrophoresis on 1% agarose gel (*Gelly*Phor LE, EuroClone, Milano, Italy) subsequently stained with ethidium bromide. The remaining products were purified for subsequent uses with the NucleoSpin™ Extract II kit (Macherey-Nagel).

A first attempt to obtain the 16S rRNA gene sequence of the bacterium with primers 16S alfa F19a (5'-CCTGGCTCAGAACGAACG-3'[13]) and 1492R (5'-GGNWACCTT GTTACGACTT-3'; modified from Lane[38]; annealing temperature: 50 °C, 30 cycles) resulted in a low-quality product. After cloning (kit TOPO TA Cloning®; Invitrogen, Carlsbad, CA) we recovered a putative sequence of the gammaproteobacterial symbiont. We designed two primers of narrower specificity (as described by Petroni and coworkers[39]) on this sequence. The forward primer Neb_F203 (5'-CTTTAGGGCAGTCGCTATAC-3') and the reverse primer FNF_R759 (5'-CCACGCTTTCGTCCCTC-3') were then employed in two new PCR amplifications of the stored tgDNA, the first one with primers Neb_F203 and 1492R (annealing temperature: 50 °C, 30 cycles); the second one with primers 27F (5'-AGAGTTTGATYMTGGCTCAG-3'; modified from Lane[38]) and FNF_R759 (annealing temperature: 55 °C, 35 cycles). Direct sequencing of the two PCR products was performed according to Vannini and coworkers[13] with three internal primers: 16S R515 ND (5'-ACCGCGGCTGCTGGCAC-3'), 16S F343 ND (5'-TACGGGAGGCAGCAG-3'), 16S F785 ND (5'-GGATTAGATACCCTGGTA-3') and the resulting sequences were assembled in order to obtain an almost complete 16S rRNA gene sequence of the gammaproteobacterial symbiont.

The NCBI BLASTN software[40] was employed for a preliminary sequence comparison.

Sequence availability and phylogenetic analysis. The characterized sequence is available from DDBJ/EMBL/GenBank databases under the accession number HE794998.

The 16S rRNA gene sequence was first aligned against more than 450,000 prokaryotic sequences from the SILVA 104 database[41] with the automatic aligner of the ARB software package[42], and then manually edited to optimize base-paring in the predicted rRNA stem regions. Two sets of phylogenetic analyses were performed, at two different taxonomic levels.

The first one included the characterized sequence and 62 homologous sequences from representatives of class *Gammaproteobacteria*. At least one species of the type genus for each of the 42 families (belonging to 15 orders) were chosen; families with more than 5 genera were

represented by more than one species. 12 sequences from organisms belonging to the class *Alphaproteobacteria* were included as outgroup. Sequence lengths were reduced to that of the shortest one. Columns of the alignment including only one non-gap character were excluded. The final character matrix consisted of 1,319 columns.

The second phylogenetic analysis compared the sequence of the In endosymbiont with those of 51 species and subspecies belonging to the order *Thiotrichales* (*Gammaproteobacteria*). 49 of them are formally classified inside one of the three families of the order (*Francisellaceae*, *Piscirickettsiaceae* and *Thiotrichaceae*). *Caedibacter taeniospiralis* (accession number: AY102612) and *Fangia hongkongensis* (accession number: AB176554) are members of the order but don't belong to any of the aforementioned families. The sequence of *Wolbachia persica* (accession number: M21292; formally belonging to the class *Alphaproteobacteria*) was included because of its high similarity with those of genus *Francisella* (*Francisellaceae*). 4 sequences of uncultured organisms (accession numbers: FJ425613, FJ202084, DQ889939, GU118319) were also included because of their similarity with that of the characterized endosymbiont. 5 sequences of organisms belonging to other orders of the class *Gammaproteobacteria* were included as outgroup. Sequence lengths were reduced to that of the shortest one. Columns of the alignment including only one or two non-gap characters were excluded. The final character matrix consisted of 1,310 columns.

The classification of Euzéby[43] (updated on March 4th 2012) has been followed.

jModelTest[44,45] was employed to select the substitution model that fits best the data. The TREE-PUZZLE[46] likelihood mapping method was employed with 100,000 randomly chosen quartets to test the amount of evolutionary information contained in the character matrix. Tree reconstruction was performed through different inferring methods: the Phylip DNAPARS[47] and PHYML[44] software for maximum parsimony (MP) and maximum likelihood (ML) analyses respectively were provided by the ARB package[42]. Statistical reliability of nodes was evaluated through bootstrap analysis with 1,000 pseudoreplicates for the ML method. Similarity matrices and Neighbor Joining trees (NJ[48]) were built with the software ARB NJ from the same package, using the "similarity" and "felsenstein" correction respectively[42]. Similarity values were calculated on the same character matrices employed for phylogenetic analyses. Bayesian Inference analysis (BI) was performed with MrBayes[49], using three different Markov Chain Monte Carlo

runs, with one cold chain and three heated chains each, running for 2,500,000 (*Gammaproteobacteria* phylogeny) and 1,000,000 (*Thiotrichales* phylogeny) generations.

Fluorescence *in situ* hybridization (FISH). For FISH experiments, cells starved for a few days were individually collected and washed several times in sterile distilled water; they were then fixed on slides with formaldehyde (4% in PBS) and dehydrated in ethanol. Experiments were performed as described by Manz and coworkers[50], with varying degree of formamide in the hybridization buffers. All experiments included a negative control. Double-probe hybridizations were also performed in order to better check the specificity of the used probes. The slides were observed with a Zeiss Axioplan light microscope (Carl Zeiss, Oberkochen, Germany) equipped for epifluorescence.

Preliminary experiments were performed to assess the biodiversity of the bacterial community of endosymbionts living in the In cells. The oligonucleotide probes ALF1b, BET42a and GAM42a (5'-CGTTCGYTCTGAGCCAG-3', 5'-GCCTTCCCACTTCGTTT-3' and 5'-GCCTTCCCACATCGTTT-3' respectively[50]), which match the most common eubacterial classes found so far inside ciliates[2,6,9-11] were employed. The probe PnecC-16S-445 (5'-GAGCCGGTGTTTCTTCCC-3'[51]) was also employed to detect the presence of *Polynucleobacter necessarius* (*Burckholderiales*, *Betaproteobacteria*). In order to prove that the 16S rRNA gene sequence obtained belongs to one of the endosymbionts, slightly modified versions of the primers Neb_F203 and FNF_R759 were labeled with fluorescent dies and used as probes. The probes NebProb203 (5'-ATAGCGACTGCCCTAAAG-3') and FNFProb759 (5'-CCCACGCTT TCGTCCCT-3') were designed and their specificities were checked on both the SILVA 104[41] and RDP[52] databases. In order to optimize the protocol for their use, different formamide concentrations in the hybridization buffer were tested (0%, 15%, 30%, 40%). Probe NebProb203 produced brighter signals and kept its specificity in formamide-free experiments; probe FNFProb759 was employed with formamide concentrations in the range between 15% and 30%. Both the newly designed probes were deposited at ProbeBase[53].

Cultivation experiments. Growth experiments were performed in order to isolate and cultivate the endosymbiont of population In outside of its ciliate host. Ciliate cells were harvested and washed twice in sterile SMB. They were then treated overnight with chloramphenicol 0.2 mg/mL in SMB in order to eliminate possible contaminants and washed again. Ciliate cell lysis was performed mechanically using a closed glass micro-pipette.

Homogenate was then used as inoculum in three different concentrations: no dilution, dilution 1:10, dilution 1:100 (10 µL were used for each run). Culture attempts were performed on a) solid and liquid Yeast Peptone-Sea Water (YP-SW) medium[54], with SMB substituting sea water: 0.4% yeast extract (w/v), 0.2% peptone (w/v), 75% SMB (v/v), 25% distilled water (1.5% agar w/v for solid medium); b) SMB added with 0.1% of yeast extract. Both media were tested at 19 °C and at 30 °C. Growth by the endosymbiont was checked by FISH with the species-specific probe NebProb203 at days 1°, 3°, 7°, 14° and 28°.

Results

Molecular characterization of the symbiont. Preliminary FISH experiments performed with probes ALF1b, BET42a, PnecC-16S-445 and GAM42a gave positive results in all In cells inspected, pointing out the coexistence of at least three different types of bacteria inside the ciliate's cytoplasm. BET42a and PnecC-16S-445 matched the 16S rRNA of the obligate and mutualistic endosymbiont *Polynucleobacter necessarius* (*Betaproteobacteria*)[9]. ALF1b targeted bacteria belonging to the class *Alphaproteobacteria.* Cytoplasmic symbionts matching the probe GAM42a, targeting bacteria of the *Gammaproteobacteria* class, are the object of the present study. Similar experiments were repeated periodically in a three-year period and always gave the same results, proving that the bacterial community composition was stable.

We obtained an almost complete 16S rRNA gene sequence of 1,456 bp length through direct sequencing. BLASTN showed that the most similar sequence available is that of the 16S rRNA gene of *Fangia hongkongensis* (similarity: 92.5%; accession number: AB176554[55]), belonging to the order *Thiotrichales* (*Gammaproteobacteria*). FISH experiments with the sequence-specific probe NebProb203 were performed in order to fulfill the "Full-Cycle rRNA Approach"[28] and gave positive signals in all In cells observed (Figure 2.2). Signals from the probes NebProb203 and GAM42a coincided, thus excluding the presence of additional gammaproteobacterial endosymbionts. Probe FNFProb759 also gave positive results.

According to all employed 16S rRNA databases, the sequence matching the probe NebProb203 is unique to the newly characterized endosymbiont. The sequence of probe FNFProb759 is not specific. It matches more than 17,000 bacterial sequences available in the RDP database, most of which (~13,500) belong to the phylum *Bacteroidetes*. Of the 434 hits

reported for the class *Gammaproteobacteria*, 387 (89.2%) belong to the order *Thiotrichales* and show the highest similarity either to the genus *Francisella* or to *Fangia hongkongensis* and *Caedibacter taeniospiralis*. Other hits in the class (10.8%) belong to uncultured organisms, with

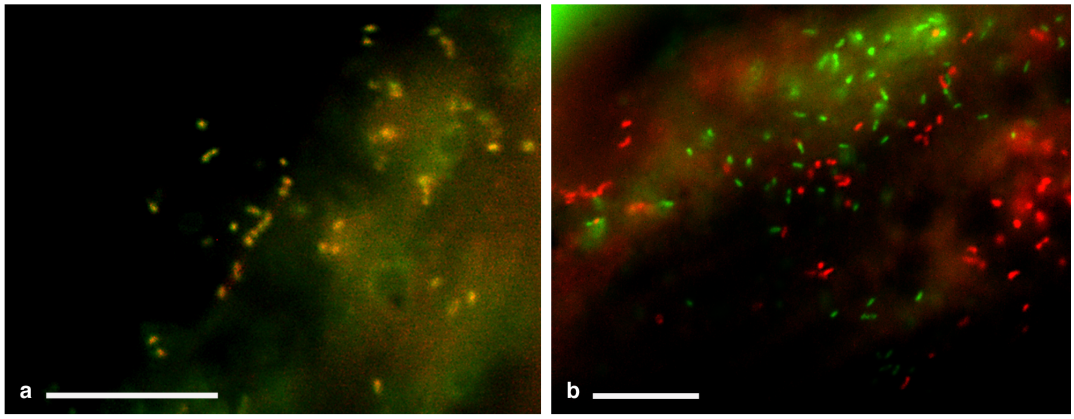


Figure 2.2 FISH results on fixed In cells. Two combination of probes are shown: GAM42a and NebProb203 (a), respectively labeled with fluorescein (green) and Cy3 (red); ALF1b and NebProb203 (b), respectively labeled with AlexaFluor® 488 (green) and Cy3. Completely overlapping signals are visible in (a). The bars correspond to 10 µm.

the exception of two genera of the order *Methylococcales* (*Methylohalobius* and *Methylothermus*). The hits found in the SILVA 104 database are 2,850, mostly scattered in the prokaryotes phylogenetic tree.

Phylogenetic analyses. In the analysis of *Gammaproteobacteria*, the Akaike Information Criterion (AIC) calculated by jModelTest selected the GTR+I+G model of substitution. TREE-PUZZLE likelihood mapping estimated that about 90.4% of the quartets had well-defined topologies, less than 2.12% of the characters in the matrix were either gaps (98.5%) or ambiguous (1.5%) and that only 2 out of 75 sequences had a statistically significant difference in base composition (*Acidithiobacillus thiooxidans* and *Thiocapsa roseopersicina*).

Most of the nodes in the obtained ML tree (Figure 2.3) are supported by low values of bootstrap or Posterior Probability (PP), thus the phylogenetic relationships that they imply are not reliable. The monophyletic status has been recovered for only 3 out of 13 orders represented

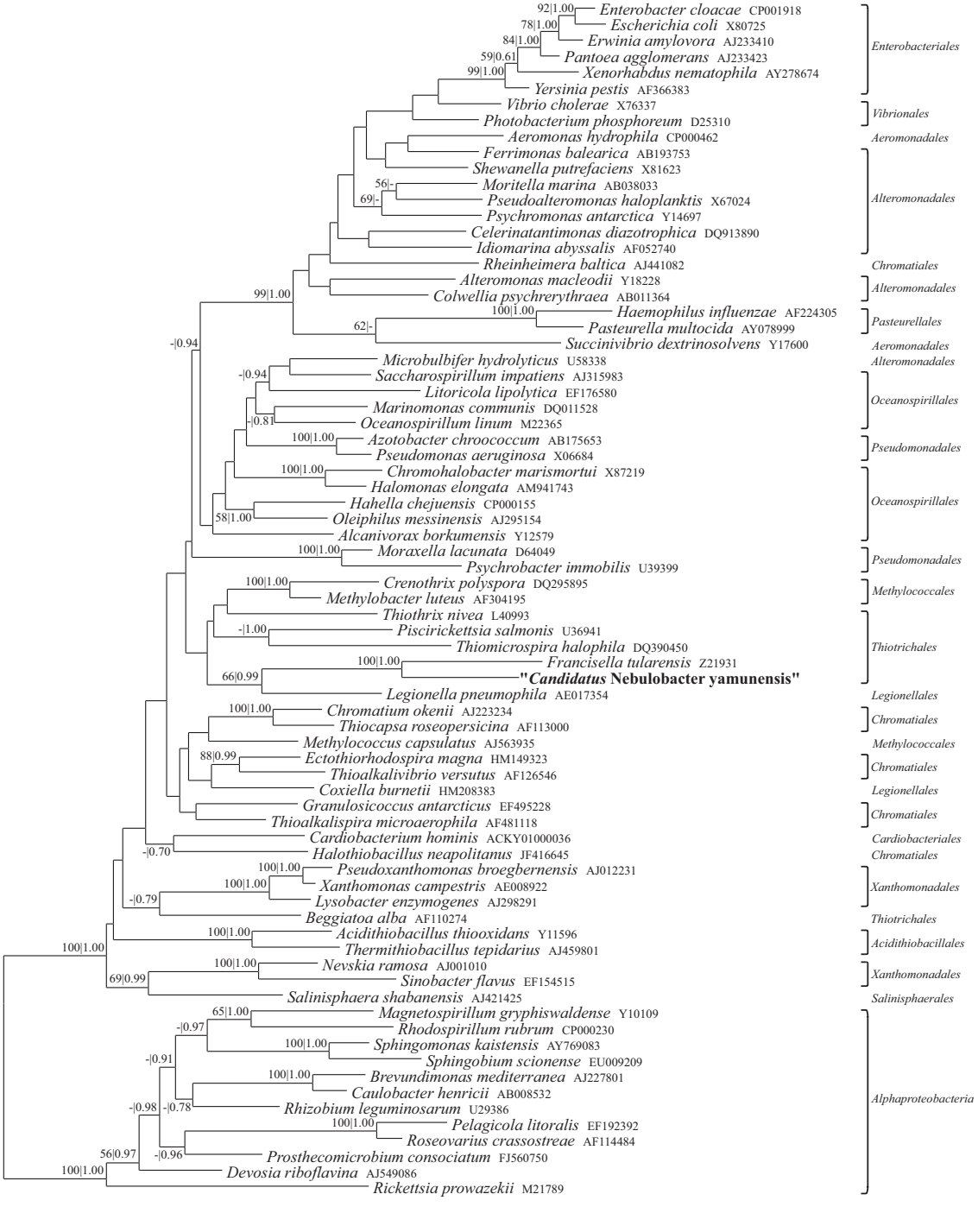


Figure 2.3 Maximum likelihood phylogenetic tree of the class *Gammaproteobacteria*. Species names, classification and accession numbers of the sequences employed are shown. The parameters associated with nodes represent bootstrap and Posterior Probability (values below 50|0.60 are not shown). The 16S rRNA gene sequence of "*Candidatus* Nebulobacter yamunensis" characterized in this study is in bold characters. The bar corresponds to an estimated sequence divergence of 10%.

by more than one genus: *Acidithiobacillales*, *Enterobacteriales* and *Pasteurellales*. A supra-ordinal clade that received high support is the "crown group" including *Enterobacteriales*, *Vibrionales*, *Pasteurellales*, *Aeromonadales*, most of *Alteromonadales* and *Rheinheimeria baltica* (*Chromatiales*). The In gammaproteobacterial endosymbiont associates with *Francisella tularensis* (*Thiotrichales*) with maximal statistical support. The clade formed by these two organisms is the sister group of *Legionella pneumophila* (this conclusion is supported by statistical values of 66%/0.99). Other members of *Thiotrichales* are more distantly related.

MP and NJ trees show a topology largely similar to that obtained through ML. In particular, all nodes supported by statistical values of 50%/0.60 or higher in the ML tree were recovered by both, the only exception being that MP analysis failed to cluster *Succinivibrio dextrinosolvens* (*Aeromonadales*) with the other bacteria of the "crown group".

In the analysis of *Thiotrichales*, the AIC calculated by jModelTest selected the GTR+I+G model of substitution. TREE-PUZZLE likelihood mapping estimated that about 93.7% of the quartets had well-defined topologies, about 2.00% of the characters in the matrix were either gaps (96.0%) or ambiguous (4.0%) and that only 2 out of 62 sequences had a statistically significant difference in base composition (*Sulfurivirga caldicuralii* and *Thiothrix eikelboomii*).

The obtained ML tree (Figure 2.4) has the majority of its nodes well supported by statistical values. The only genus of family *Francisellaceae*, *Francisella*, is monophyletic and closely related to *Wolbachia persica*. Both families *Piscirickettsiaceae* and *Thiotrichaceae* appear non-monophyletic, although only for the former this conclusion is statistically supported. The In endosymbiont is most closely related to *Caedibacter taeniospiralis*, an endosymbiont of *Paramecium tetraurelia*. These two organisms, along with *Fangia hongkongensis* and 4 uncultured bacteria characterized in studies on diseased corals (two of which are unpublished)[56,57], form the "*Caedibacter taeniospiralis* clade". There is maximal statistical support for this clade and for its association to genus *Francisella* and *Wolbachia persica*. All of these organisms have in common the signature sequence 5'-CCCACGCTTTCGTCCCT-3' matched by the probe FNFProb759, which is not shared by any other bacterium in the analysis. The most closely related organism is *Piscirickettsia salmonis* (with statistical values of 85%/1.00).

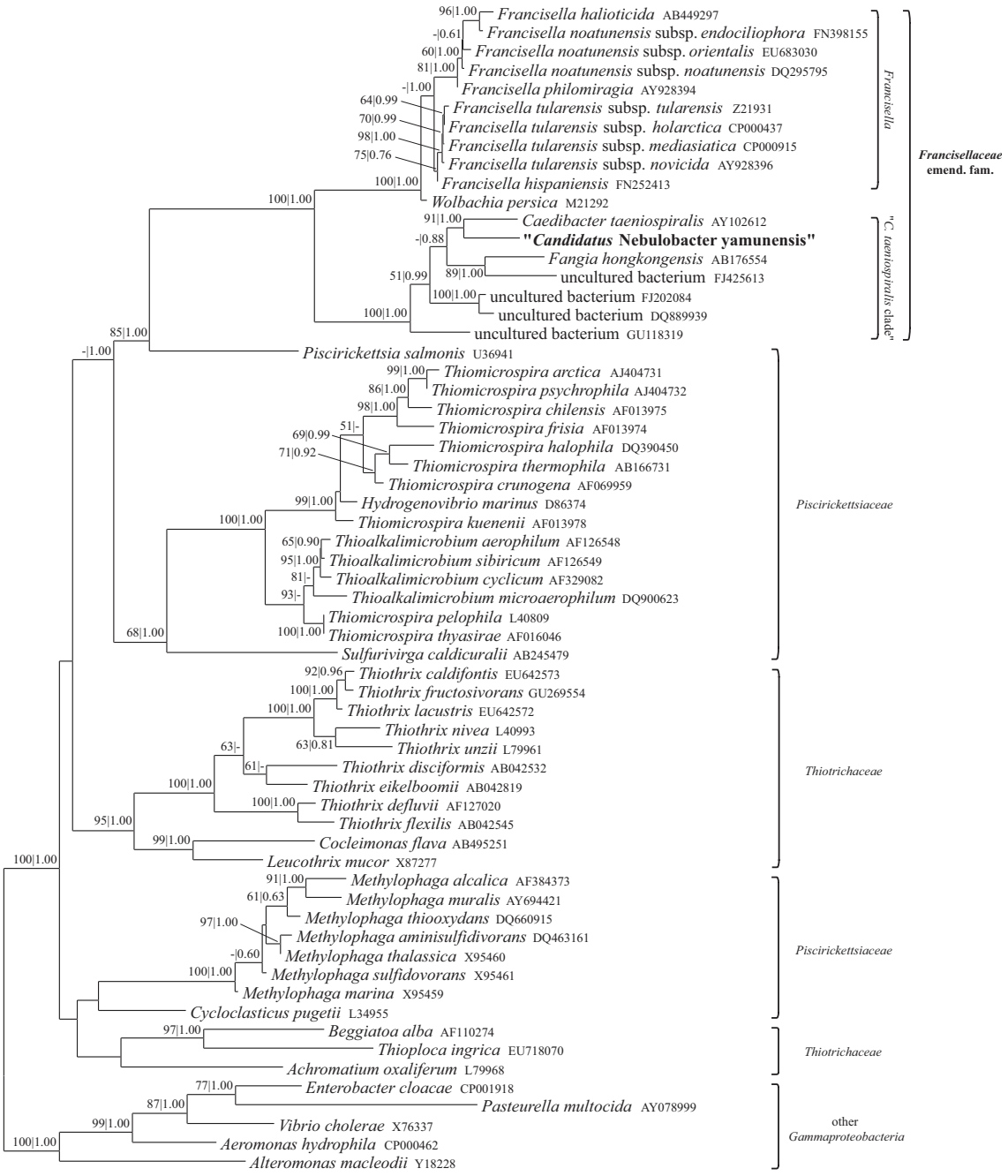


Figure 2.4 Maximum likelihood phylogenetic tree of the order *Thiotrichales*. Species names, classification and accession numbers of the sequences employed are shown. The parameters associated with nodes represent bootstrap and Posterior Probability (values below 50|0.60 are not shown). The 16S rRNA gene sequence of "*Candidatus* Nebulobacter yamunensis" characterized in this study is in bold characters. The bar corresponds to an estimated sequence divergence of 10%.

All nodes of the ML tree supported by statistical values of 75%/0.90 or higher are recovered also with the NJ and MP methods (with only one minor exception: the NJ analysis doesn't associate *Methylophaga alcalica* and *Methylophaga muralis*).

The clade of *Francisella* and *Wolbachia persica* includes sequences with similarity values above the threshold of 95%, higher than those shared between members of other genera like *Thiothrix* (88%), *Thiomicrospira* (90%) and *Methylophaga* (94%). The similarity threshold separating representatives of the "*Caedibacter taeniospiralis* clade" from other bacteria is 90%. The lowest similarity value within the clade including *Francisella*, *Wolbachia persica* and the "*Caedibacter taeniospiralis* clade" is above 86%. None of the sequences share higher similarity values with more distantly related bacteria. Sequences within each of the two families *Piscirickettsiaceae* and *Thiotrichaceae* share comparatively lower similarity values of 83% and 82% respectively.

Cultivation experiments. No detectable growth of the gammaproteobacterial endosymbiont was observed in any of the performed cultivation attempts, at any of the analyzed dates.

Discussion

Characterization of the endosymbiont. The gammaproteobacterial endosymbiont of the *Euplotes aediculatus* population In belongs to the class *Gammaproteobacteria*, according to its 16S rRNA gene sequence and the positivity to probe GAM42a. There are only two currently known ciliate endosymbionts belonging to this class, namely *Caedibacter taeniospiralis* (a cytoplasmic endosymbiont of *Paramecium tetraurelia*)[32] and the recently described "*Candidatus* Francisella noatunensis subsp. endociliophora" found inside the cytoplasm of *Euplotes raikovi*[6]. These three endosymbionts belong to the order *Thiotrichales*, and *C. taeniospiralis* is indeed the closest relative of the In endosymbiont. However, sections of In cells observed at the transmission electron microscope (TEM) showed no bacterial endosymbiont containing an R-body (data not shown), a typical character of genus *Caedibacter*[26].

Our research demonstrates that the characterized bacterium was only one component of a stable and conspicuous bacterial community living inside the cytoplasm of the host. The

betaproteobacterial species *Polynucleobacter necessarius* is an obligate endosymbiont for *E. aediculatus*[9,24,27], and usually other coexisting bacteria are called "secondary symbionts"[10]. In the In population the gammaproteobacterium here described is not the only secondary symbiont, because representatives of the class *Alphaproteobacteria* were also detected. The ecological role played by these bacteria, if any, is not known.

We didn't succeed in cultivating the gammaproteobacterium outside the host on media which resembled those suitable for *Fangia hongkongensis*, its closest cultivable relative. This suggests some degree of dependence of the endosymbiont on the ciliate. The infection rate was very high, considering that no In cell lacking of it was detected in three years of periodical FISH experiments.

The low 16S rRNA gene sequence similarity values shared by the endosymbiont and all the other currently characterized bacteria suggest the establishment of a new genus and species[58]. Following the rules valid for uncultivable bacteria[33,34], we propose the name "*Candidatus* Nebulobacter yamunensis". A diagnostic description of the new taxon is available at the end of this section.

Systematic and taxonomy of *Francisellaceae*. The order *Thiotrichales* Garrity et al. 2005 formally includes three families: *Francisellaceae* (with the single genus *Francisella*), *Piscirickettsiaceae* and *Thiotrichaceae*[59]. The order as well as the last two families have no unifying phenotypical characters, and were established only on the basis of 16S rRNA gene sequences. According to the List of Prokaryotic names with Standing in Nomenclature[43], the names *Thiotrichales* and *Thiotrichaceae* are illegitimate because they contain the type genera of previously established families and orders (namely *Achromatium*, *Beggiatoa* and *Leucothrix*). Furthermore, in our analysis based on the 16S rRNA gene the order is not monophyletic, although the statistical support underlying this conclusion is not high. The non-monophyly of both family *Piscirickettsiaceae* and *Thiotrichaceae* is also suggested. Our conclusions are very similar to those obtained by phylogenomic studies[60,61].

The monophyly of genus *Francisella* (and thus of family *Francisellaceae*) was recovered and supported by high statistical values. The species *Wolbachia persica* Suitor and Weiss 1961, that falls very close to the *Francisella* clade, has been repeatedly proven to be a misidentified taxon, that should be reclassified inside the genus *Francisella*[62,63].

Francisellaceae as currently defined is thus a coherent and monophyletic taxon, although this is largely due to its being a monogeneric family including only very similar and well-known bacteria. In our analysis we have defined a more comprehensive taxon including the genus *Francisella* (with *Wolbachia persica*) and the "*Caedibacter taeniospiralis* clade". The monophyly of this taxon is highly supported by statistical values; the major splits within it are also well-defined; the similarity values of the less related sequences are above those calculated for other families. Its members can also be recognized, although not univocally, using the combination of probes GAM42a and FNFProb759.

We propose to extend the definition of family *Francisellaceae* in order to include this entire clade. The monophyly of the family so defined is strongly proven, while that of other formally established genera, families and orders in the class *Gammaproteobacteria* are not. There are molecular diagnostic markers shared by all members of the new family. Most of its characterized bacteria (the exception being *Fangia hongkongensis*[55]) have facultative or obligate intracellular life cycles.

An emended description of family *Francisellaceae* is available at the end of this section. We have classified it tentatively in the order *Thiotricales*, although our analyses suggest a further revision of this taxon as well.

Description of "*Candidatus* Nebulobacter yamunensis". *Nebulobacter yamunensis* (Ne.bu.lo.bac'ter ya.mu.nen'sis; N.L. adj. *nebulosus*, misty; N.L. masc. n. *bacter*, rod; N.L. masc. n. *Nebulobacter*, misty rod; N.L. adj. *yamunensis*, coming from the Yamuna River region, India).

Rod-shaped bacterium. Cytoplasmic endosymbiont of the ciliate *Euplotes aediculatus* (Ciliophora, Spirotrichea). Phylogenetically related to *Caedibacter taeniospiralis* and *Fangia hongkongensis*. Basis of assignment: 16S rRNA gene sequence (accession number: HE794998) and positive match with the specific FISH oligonucleotide probe NebProb203 (5'-ATAGCGACTGCCCTAAAG-3') and the gammaproteobacterial probe GAM42a (5'-GCCTTCCCACATCGTTT-3'). Identified in the *E. aediculatus* population In, isolated from a freshwater sample collected near the Yamuna River (New Delhi, India). Uncultured thus far.

Emended description of family *Francisellaceae* Sjöstedt 2005. The species *Caedibacter taeniospiralis*, *Fangia hongkongensis* and "*Candidatus* Nebulobacter yamunensis"

are placed inside the family *Francisellaceae* together with the genus *Francisella*. Defining characters listed in the original description apply integrally only to the genus *Francisella*. The family forms a highly supported monophyletic group, according to the 16S rRNA gene sequence phylogeny. The nucleotide sequence 5'-CCCACGCTTTCGTCCCT-3' inside the 16S rRNA gene distinguishes members of the family from other bacteria of the order *Thiotrichales*. With the exception of *Fangia hongkongensis*, all members of the family that received a binomial name are obligate or facultative endosymbionts of eukaryotic cells.

References

1. Eschbach E, Pfannkuchen M, Schweikert M, Drutschmann D, Brümmer F, *et al.* (2009) "*Candidatus* Paraholospora nucleivisitans", an intracellular bacterium in *Paramecium sexaurelia* shuttles between the cytoplasm and the nucleus of its host. Syst Appl Microbiol. 32: 490-500
2. Ferrantini F, Fokin SI, Modeo L, Andreoli I, Dini F, *et al.* (2009) "*Candidatus* Cryptoprodotis polytropus", a novel *Rickettsia*-like organism in the ciliated protist *Pseudomicrothorax dubius* (Ciliophora, Nassophorea). J Eukaryot Microbiol. 56: 119-129
3. Fokin SI. (2012) Frequency and biodiversity of symbionts in representatives of the main classes of Ciliophora. Eur J Protistol. 48: 138-148
4. Irbis C, Ushida K. (2004) Detection of methanogens and proteobacteria from a single cell of rumen ciliate protozoa. J Gen Appl Microbiol. 50: 203-212
5. Schrallhammer M, Fokin SI, Schleifer KH, Petroni G. (2006) Molecular characterization of the obligate endosymbiont "*Caedibacter macronucleorum*" Fokin and Görtz, 1993 and of its host *Paramecium duboscqui* strain Ku4-8. J Eukaryot Microbiol. 53: 499-506
6. Schrallhammer M, Schweikert M, Vallesi A, Verni F, Petroni G. (2011) Detection of a novel subspecies of *Francisella noatunensis* as endosymbiont of the ciliate *Euplotes raikovi*. Microbial Ecol. 61: 455-464
7. Shinzato N, Watanabe I, Meng XY, Sekiguchi Y, Tamaki H, *et al.* (2007) Phylogenetic analysis and fluorescence *in situ* hybridization detection of archaeal and bacterial endosymbionts in the anaerobic ciliate *Trimyema compressum*. Microbial Ecol. 54: 627-636
8. Sun HY, Noe J, Barber J, Coyne RS, Cassidy-Hanley D, *et al.* (2009) Endosymbiotic bacteria in the parasitic ciliate *Ichthyophthirius multifiliis*. Appl Environ Microbiol. 75: 7445-7452
9. Vannini C, Ferrantini F, Ristori A, Verni F, Petroni G. (2012) Betaproteobacterial symbionts of the ciliate *Euplotes*: origin and tangled evolutionary path of an obligate microbial association. Environ Microbiol. 14: 2553-2563
10. Vannini C, Ferrantini F, Schleifer KH, Ludwig W, Verni F, Petroni G. (2010) "*Candidatus* Anadelfobacter veles" and "*Candidatus* Cyrtobacter comes", two new rickettsiales species hosted by the protist ciliate *Euplotes harpa* (Ciliophora, Spirotrichea). Appl Environ Microbiol. 76: 4047-4054

11. Vannini C, Petroni G, Verni F, Rosati G. (2005) A bacterium belonging to the *Rickettsiaceae* family inhabits the cytoplasm of the marine ciliate *Dioprhys appendiculata* (Ciliophora, Hypotrichia). Microbial Ecol. 49: 434-442
12. Vannini C, Petroni G, Verni F, Rosati G. (2005) *Polynucleobacter* bacteria in the brackish-water species *Euplotes harpa* (Ciliata Hypotrichia). J Eukaryot Microbiol. 52: 116-122
13. Vannini C, Rosati G, Verni F, Petroni G. (2004) Identification of the bacterial endosymbionts of the marine ciliate *Euplotes magnicirratus* (Ciliophora, Hypotrichia) and proposal of "*Candidatus* Devosia euplotis". Int J Syst Evolution Microbiol. 54: 1151-1156
14. Fokin SI. (2004) Bacterial endocytobionts of Ciliophora and their interactions with the host cell. Int Rev Cytol. 236: 181-249
15. Görtz HD. (1996) Symbiosis in ciliates. In: Hausmann K, Bradbury PS (eds), Ciliates. Cells as organisms, Fisher, Stuttgart, New York, pp 441-462
16. Görtz HD. (2006) Symbiotic associations between ciliates and prokaryotes. In: Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E (eds), The Prokaryotes. Third Edition, vol. 1, Springer-Verlag, New York, NY, pp 364-402
17. Vannini C, Petroni G, Schena A, Verni F, Rosati G. (2003) Well-established mutualistic associations between ciliates and prokaryotes might be more widespread and diversified than so far supposed. Eur J Protistol. 39: 481-485
18. Fels D, Kaltz O. (2006) Temperature-dependent transmission and latency of *Holospora undulata*, a micronucleus-specific parasite of the ciliate *Paramecium caudatum*. P Roy Soc Lond B Bio. 273: 1031-1038
19. Kusch J, Czubatinsky L, Wegmann S, Hübner M, Alter M, Albrecht P. (2002) Competitive advantages of *Caedibacter*-infected paramecia. Protist. 153: 47-58
20. Vannini C, Lucchesi S, Rosati G. (2007) *Polynucleobacter*: symbiotic bacteria in ciliates compensate for a genetic disorder in glycogenolysis. Symbiosis. 44: 85-91
21. Hori M., Fujii K, Fujishima M. (2008) Micronucleus-specific bacterium *Holospora elegans* irreversibly enhances stress gene expression of the host *Paramecium caudatum*. J Eukaryot Microbiol. 55: 515-521
22. Hori M, Fujishima M. (2003) The endosymbiotic bacterium *Holospora obtusa* enhances heat-shock gene expression of the host *Paramecium caudatum*. J Eukaryot Microbiol. 50: 293-298
23. Smurov AO, Fokin SI. (1999) Resistance of *Paramecium* species (Ciliophora, Peniculia) to salinity of environment. Protistology. 1: 43-53
24. Vannini C, Pöckl M, Petroni G, Wu QL, Lang E, *et al.* (2007) Endosymbiosis *in statu nascendi*: close phylogenetic relationship between obligately endosymbiotic and obligately free-living *Polynucleobacter* strains (*Betaproteobacteria*). Environ Microbiol. 9: 347-359
25. Fokin SI, Görtz HD. (2009) Diversity of *Holospora* bacteria in *Paramecium* and their characterization. In: Fujishima M (ed), Endosymbionts in *Paramecium*. Microbiology Monograph, Münster, pp 161-199
26. Schrallhammer M, Schweikert M. (2009) The killer effect of *Paramecium* and its causative agents. In: Fujishima M. (ed), Endosymbionts in *Paramecium*. Microbiology Monograph, Münster, pp 227-246
27. Heckmann K, Schmidt HJ. (1987) *Polynucleobacter necessarius* gen. nov., sp. nov., an obligately endosymbiotic bacterium living in the cytoplasm of *Euplotes aediculatus*. Int J Syst Bacteriol. 37: 456-457
28. Amann RI, Ludwig WG, Schleifer KH. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev. 59: 143-169

29. Amann RI, Springer N, Ludwig W, Görtz HD, Schleifer KH. (1991) Identification *in situ* and phylogeny of uncultured bacterial endosymbionts. Nature. 351: 161-164

30. Yu X, Walker DH. (2005) *Rickettsiaceae* Pinkerton 1936, 186^{AL} emend. Dumler, Barbet, Bekker, Dasch, Palmer, Ray, Rikihisa and Rurangirwa 2001. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds), Bergey's Manual of Systematic Bacteriology, Second Edition, Vol. II, Part C. Springer-Verlag, New York, NY, pp 96-116

31. Sjöstedt AB. (2005) *Francisellaceae* fam. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds), Bergey's Manual of Systematic Bacteriology, Second Edition, Vol. II, Part B. Springer-Verlag, New York, NY, pp 199-210

32. Beier CL, Horn M., Michel R., Schweikert M, Görtz HD, Wagner M. (2002) The genus *Caedibacter* comprises endosymbionts of *Paramecium* spp. related to the *Rickettsiales* (*Alphaproteobacteria*) and to *Francisella tularensis* (*Gammaproteobacteria*). Appl Environmen Microbiol. 68: 6043-6050

33. Murray RGE, Schleifer KH. (1994) A proposal for recording the properties of putative taxa of procaryotes. Int J Syst Bacteriol. 44: 174-176

34. Murray RGE, Stackebrandt E. (1995) Implementation of the provisional status *Candidatus* for incompletely described prokaryotes. Int J Syst Bacteriol. 45: 186-187

35. Miyake A. (1981) Cell-cell interaction by gamones in *Blepharisma*. In: O'Day DH, Horgen PA (eds), Sexual interactions in eukaryotic microbes. Academic Press, New York, pp 95-129

36. Modeo L, Rosati G, Andreoli I, Gabrielli S, Verni F, Petroni G. (2006) Molecular systematics and ultrastructural characterization of a forgotten species: *Chattonidium setense* (Ciliophora, Heterotrichea). P Jpn Acad B-Phys. 82: 359-374

37. Curds CR. (1975) A guide to the species of the genus *Euplotes* (Hypotrichida, Ciliatea). Bull Br Mus Nat Hist (Zool) 28: 3-61

38. Lane DJ. (1991) 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M (eds), Nucleic Acid Techniques in Bacterial Systematics, Wiley, New York, pp 115-147, 1991

39. Petroni G, Rosati G, Vannini C, Modeo L, Dini F, Verni F. (2003) *In situ* identification by fluorescently labeled oligonucleotide probes of morphologically similar, closely related ciliate species. Microbial Ecol. 45: 156-162

40. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, *et al.* (1997) Gapped BLAST and PSI BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389-3402

41. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, *et al.* (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 35: 7188-7196

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

43. Euzéby JP. (1997) List of Bacterial Names with Standing in Nomenclature: a folder available on the Internet. Int J Syst Bacteriol 47: 590-592 (List of Prokaryotic Names with Standing in Nomenclature. Available at: http://www.bacterio.net)

44. Guindon S, Gascuel O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 52: 696-704

45. Posada D. (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol. 25: 1253-1256

46. Schmidt HA, Strimmer K, Vingron M, von Haeseler A. (2002) TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics. 18: 502-504.

47. Felsenstein J. (1989) PHYLIP-Phylogeny Inference Package (version 3.2). Cladistics. 5: 164-166

48. Saitou N, Nei M. (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol. 4: 406-425

49. Huelsenbeck JP, Ronquist F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754-755

50. Manz W, Amann RI, Ludwig W, Wagner M, Schleifer KH. (1992) Phylogenetic oligodeoxynucleotide probes for the major subclasses of Proteobacteria: problems and solutions. Syst Appl Microbiol. 15: 593-600

51. Hahn MW, Pöckl M, Wu QL. (2005) Low intraspecific diversity in a *Polynucleobacter* subcluster population numerically dominating bacterioplankton of a freshwater pond. Appl Environ Microbiol. 71: 45394547

52. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, *et al.* (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res. 37: D141-D145

53. Loy A, Maixner F, Wagner M, Horn M. (2007) probeBase – an online resource for rRNA-targeted oligonucleotide probes: new features 2007. Nucleic Acids Res. 35: D800-D804

54. Lau KWK, Ng CY, Ren J, Lau SC, Qian PY, Wong PK, *et al.* (2005). *Owenweeksia hongkongensis* gen. nov., sp. nov., a novel marine bacterium of the phylum "*Bacteroidetes*". Int J Syst Evol Microbiol. 55: 1051-1057

55. Lau KWK, Ren J, Fung MC, Woo PC, Yuen KY, *et al.* (2007) *Fangia hongkongensis* gen. nov., sp. nov., a novel gammaproteobacterium of the order *Thiotrichales* isolated from coastal seawater of Hong Kong. Int J Syst Evolution Microbiol. 57: 2665-2669

56. Sunagawa S, DeSantis TZ, Piceno YM, Brodie EL, DeSalvo MK, *et al.* (2009) Bacterial diversity and White Plague Disease-associated community changes in the Caribbean coral *Montastrea faveolata*. ISME J. 3: 512-521

57. Sunagawa S, Woodley CM, Medina M. (2010) Threatened corals provide underexplored microbial habitats. PLoS ONE 5: e9554

58. Ludwig W, Strunk O, Klugbauer S, Klugbauer N, Weizenegger M, *et al.* (1998) Bacterial phylogeny based on comparative sequence analysis. Electrophoresis. 19: 554-568

59. Garrity GM, Bell JA, Lilburn T. (2005) *Thiotrichales* ord. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds), Bergey's Manual of Systematic Bacteriology, Second Edition, Vol. II, Part B. Springer-Verlag, New York, NY, pp 131-210

60. Cutiño-Jiménez AM, Martins-Pinheiro M, Lima WC, Martín-Tornet A, Morales OG, Menck CFM. (2010) Evolutionary placement of *Xanthomonadales* based on conserved protein signature sequences. Mol Phylogenet Evol. 54: 524-534

61. Gao B, Mohan R, Gupta RS. (2009) Phylogenomics and protein signatures elucidating the evolutionary relationships among the *Gammaproteobacteria*. Int J Syst Evol Microbiol. 59: 234-247

62. Dumler JS, Barbet AF, Bekker CPJ, Dasch, GA, Palmer GH, *et al.* (2001) Reorganization of genera in the families *Rickettsiaceae* and *Anaplasmataceae* in the order *Rickettsiales*: unification of some species of *Ehrlichia* with *Anaplasma*, *Cowdria* with *Ehrlichia* and *Ehrlichia* with *Neorickettsia*, descriptions of six new species combinations and designation of *Ehrlichia equi* and 'HGE agent' as subjective synonyms of *Ehrlichia phagocytophila*. Int J Syst Evolution Microbiol. 51: 2145-2165

63. Forsman M, Sandström G, Sjöstedt A. (1994) Analysis of 16S ribosomal DNA sequences of *Francisella* strains and utilization for determination of the phylogeny of the genus and for identification of strains by PCR. Int J Syst Bacteriol. 44: 38-46

Chapter 3.

"*Candidatus* Gortzia infectiva"

Original publication:

Revised systematics of *Holospora*-like bacteria and characterization of "*Candidatus* Gortzia infectiva", a novel macronuclear symbiont of *Paramecium jenningsi*

Vittorio Boscaro, Sergei I Fokin, Martina Schrallhammer, Michael Schweikert, Giulio Petroni

Microbial Ecology 65(1): 255-267. January 2013.

Abstract

The genus *Holospora* (*Rickettsiales*) includes highly infectious nuclear symbionts of the ciliate *Paramecium* with unique morphology and life cycle. To date, nine species have been described, but a molecular characterization is lacking for most of them. In this study we have characterized a novel *Holospora*-like bacterium (HLB) living in the macronuclei of a *Paramecium jenningsi* population. This bacterium was morphologically and ultrastructurally investigated in detail, and its life cycle and infection capabilities were described. We also obtained its 16S rRNA gene sequence and developed a specific probe for fluorescence *in situ* hybridization experiments. A new taxon, "*Candidatus* Gortzia infectiva", was established for this HLB according to its unique characteristics and the relatively low DNA sequence similarities shared with other bacteria. The phylogeny of the order *Rickettsiales* based on 16S rRNA gene sequences has been inferred, adding to the available data the sequence of the novel bacterium and those of two *Holospora* species (*Holospora obtusa* and *Holospora undulata*) characterized for the purpose. Our phylogenetic analysis provided molecular support for the monophyly of HLBs and showed a possible pattern of evolution for some of their features. We suggested to classify inside the family *Holosporaceae* only HLBs, excluding other more distantly related and phenotypically different *Paramecium* endosymbionts.

Introduction

An increasing amount of studies[1-20] accounts for the diversity and frequency of bacterial symbionts of ciliates (Alveolata; Ciliophora). Those bacteria that possess an intracellular life-style are termed "endosymbionts" and have been found hosted by species belonging to various classes, and in virtually all subcellular compartments[3,16]. They can play opposite roles, from mutualists to parasites or anything in the continuum between the extremes[21-25]. The association with their hosts can be accidental or obligate (meaning that they cannot complete their life cycle without the ciliate); the bacteria can be vertically transmitted or highly infectious[21,23,26]. Often they are inconspicuous and difficult to detect without careful ultrastructural observations or the employment of molecular techniques like fluorescence *in situ* hybridization (FISH)[27,28].

On the other hand, there is a striking example of eye-catching bacterial symbionts known since the end of the nineteenth century. The genus *Holospora* comprises very peculiar and easily recognizable infectious bacteria, mostly found inside the macronucleus (MA) or (more rarely) the micronuclei (MI) of *Paramecium* species[29-33]. *Holospora* bacteria are specialized for horizontal transmission, and in most cases show a high level of host and nuclear type specificity[34]. They can be regarded as parasites 21-23; but see also[37-39]. The small (1-3 µm) reproductive forms (RFs), that are inherited by the host's daughter cells during division, have a typical rod-shaped, gram-negative morphology. The infectious forms (IFs) are very large (5-20 µm), elongated and with a hypertrophied, osmiophilic periplasmic space[38-41]. The cytoplasm and inner membrane occupy one end of the IF; opposite, there is a structure called recognition tip, which is implicated in the infection process[40-43]. The life cycle can be summarized as follows[17,18,30,44]: the RFs divide by binary fission inside the MA or MI, and some of them differentiate into specialized IFs. During nuclear division, IFs concentrate in a vesicle called the "connecting piece", which detaches from the dividing nucleus and fuses with the plasmalemma of the ciliate, thus releasing the IFs in the external environment[45]. An IF phagocytosed by another *Paramecium* cell can leave the digestive vacuole when acidosomes fuse with it and lower its pH[46]. The IF is not motile, hence it utilizes the host cell cytoskeleton[47,48] for reaching the targeted nuclear envelope, penetrates it with the recognition tip ahead and then divides by multiple fission, de-differentiating again into RFs. The most common exception observed to the pattern described is that some *Holospora* species do not induce the formation of the connecting piece,

and leave the host nucleus through a different mechanism that appears like a reversal of the infection sequence[16,49,50].

The biological characters described above can be easily observed. In the following, the expression *Holospora*-like bacteria (HLBs) will be employed with reference to all those bacteria that present a similar set of features, regardless of their taxonomic position.

In the literature, there are currently nine species classified inside the genus *Holospora*[29,30], but only four of them are validly described according to the rules of bacterial nomenclature. The different species can be discriminated by a set of diagnostic features including host species, morphology of the IFs, nuclear localization and capability of inducing the connecting piece[30]. The latter criterion has been used by some authors, together with other evidences, to divide the genus into two groups. Species presenting this feature were considered more derived, and arguably more specialized for infection, or completely unrelated to other HLBs[49].

A molecular phylogeny of the HLBs is still lacking. Currently there are only four 16S rRNA gene sequences available: two for *Holospora obtusa* (accession numbers: X58198[28] and JF713682[51]), one for "*Holospora curviuscula*" (accession number: JF713683[51]) and a relatively short one for *Holospora elegans* (accession number: AB297813[35]). These three species belong to the connecting piece-inducing group, the so-called "classical holosporas". It has been shown that *H. obtusa* belongs to the class *Alphaproteobacteria*[28], and the family *Holosporaceae* was established within the order *Rickettsiales*[30]. While the original description of the family listed the characters typically defining HLBs, also other, very different endosymbionts of *Paramecium* were included in *Holosporaceae* as *incertae sedis*. One example is the genus *Caedibacter*, responsible for the so-called killer trait of infected *Paramecium* strains[52]. Also placed *incertae sedis* within family *Holosporaceae* are "*Candidatus* Paracaedibacter" and "*Candidatus* Odyssella", parasites of various *Acanthamoeba* species[53,54], and other bacterial genera for which no DNA sequence is available, like *Lyticum*[*] and *Tectibacter* (endosymbionts of species of the *Paramecium aurelia* complex[19,33]). More recently, the bacterium "*Candidatus* Paraholospora nucleivisitans" was described[1]. It is phylogenetically related to *Holospora obtusa* and *Caedibacter caryophilus*, but its morphology and life cycle differ from both. Its classification inside the family *Holosporaceae* was implied, but not formalized.

[*] *Lyticum* sequences were obtained later (see Chapter 5)

This paper deals with the multidisciplinary characterization of a novel HLB found inside the MA of *Paramecium jenningsi*, a prevalently tropical species for which no symbiont was described before[29,55], and *Paramecium quadecaurelia*. We performed morphological and ultrastructural observations as well as a phylogenetic analysis based on 16S rRNA gene sequences. Cross-species infection experiments were carried out and the life cycle of the bacterium was observed. Here we also provide the first 16S rRNA gene sequence of the otherwise well-known *Holospora undulata* (a "classical holospora" infecting the MI of *Paramecium caudatum* and type species of the genus). Our aim is to set the standards for further characterizations of similar organisms, in order to accomplish all the requirements of current rules in bacterial taxonomy. With this paper the accumulation of a significant set of molecular data, which we hope will grow soon, starts to form. Some up-to-date systematic and phylogenetic considerations on this remarkable group of symbionts are therefore presented.

Materials and Methods

***Paramecium* sampling, culture and identification.** A ciliate community with organisms belonging to the morphospecies *Paramecium jenningsi* and *Paramecium aurelia* was isolated from a water sample taken in Thailand (Chaweng Lake, Samui Island, September 2010). On the basis of mating reactions and molecular markers, *P. aurelia* cells were classified as *Paramecium quadecaurelia* (E. Przyboś, personal communication). About 70% of the cells in the established laboratory culture manifested evidences of macronuclear prokaryotic infection one month after sampling. From this culture, one population for each morphospecies was established and designated as TS-j (*P. jenningsi*) and TS-a (*P. quadecaurelia*). These were used as starting material for the generation of monoclonal lines of *P. jenningsi* (TSj1-8) and *P. quadecaurelia* (TSa1-7). The monoclonal lines were used for morphological and experimental investigations of the newly found macronuclear symbionts.

The *Paramecium* cells were grown at room temperature (18-24 °C) on lettuce medium[56] inoculated with *Enterobacter aerogenes*. To immobilize the living cells, drops of culture medium were placed in a special device[57]. This allowed a detailed inspection of cell morphology and MA infection of the living specimen. *Paramecium* spp. identification was primarily done according to a set of morphological features, first of all the type of MI[58,59].

Monoclonal cultures of *Paramecium caudatum* infected with either *Holospora undulata* (*P. caudatum* strain Stb) or *Holospora obtusa* (*P. caudatum* strain 27aG3) were obtained from O. Kaltz (Institut des Sciences de l'Evolution, University of Montpellier 2, France). Cells were cultivated at 20 °C in Cerophyll medium inoculated with *Raoultella planticola* as food bacteria. Wheat grass pellets (GSE Vertrieb GmbH) were used instead of Rye grass Cerophyll. The medium contained 0.25% Wheat grass, 2.8 mM Na_2HPO_4, 0.85 mM $NaH_2PO4 \times 2\ H_2O$, 1.8 mM NaCl, 1.6 mM $MgSO_4 \times 7\ H_2O$, 4.2 mM $MgCl_2$ x 6 H_2O, 0.12 mM $CaCl_2$, 0.31 mM KCl, and 5 ng/mL Stigmasterol.

Light microscopy observation. Infected and uninfected paramecia were immobilized for observation with the help of the above mentioned device. Micrographs of the material were taken using an Orthoplan Leitz microscope equipped with differential interference contrast (DIC) microscopy, as well as a Leica DMR microscope at 300× - 1250× magnifications.

Transmission electron microscopy. The cells were fixed with a mixture of 1.6% paraformaldehyde and 2.5% glutaraldehyde in phosphate buffer (pH 7.2) for 1 h at room temperature, washed in phosphate buffer containing 125 mg/mL sucrose and postfixed with 1.5% OsO_4 for 1h at 4 °C. The cells were then embedded in 3% agar, agar blocks were dehydrated by increasing ethanol concentrations and acetone and finally embedded in Epon. The blocks were sectioned with a LKB or a Leica UCT Ultracut ultramicrotome. Ultrathin sections were stained with uranyl acetate and lead citrate.

Cross-infections experiments. For the experiments, clonal cultures of aposymbiotic (that have lost their symbionts) *P. jenningsi* (TS-j4) and *P. quadecaurelia* (TS-a3) and symbionts-free *Paramecium schewiakoffi* (Sh1-38), *Paramecium sonneborni* (Ps-a) and *Paramecium caudatum* (IP-5) were used. Experimental infection was carried out using a homogenate prepared from infected cells according to Preer[60]. *Paramecium* spp. cells were infected by mixing equal volumes of a dense cell culture and the homogenate in a 3 mL depression slide, and maintained at 18–24 °C. The number of infectious bacterial cells was 3×10^4 to 1×10^5 bacteria mL^{-1} in various infection experiments.

To check the infection status, a set of living cells (n=10) was observed by DIC 1, 2, 24 and 48 h after mixing with homogenate. Experimentally infected cultures were checked every 7 days. Series of experiments were carried out using each of the five *Paramecium* spp. stocks. For each combination the infection experiment was repeated three times.

Table 3.1 List of primers and probes employed

	Name	Sequence	Reference
Forward primer	27F	5'-AGAGTTTGATYMTGGCTCAG-3'	90[a]
Forward primer	16S alfa F19a	5'-CCTGGCTCAGAACGAACG-3'	15
Forward primer	16S F343 ND	5'-TACGGGAGGCAGCAG-3'	15
Forward primer	16S F785 ND	5'-GGATTAGATACCCTGGTA-3'	15
Forward primer	16S F114HoloCaedi[b]	5'-TGAGTAACGCGTGGGAATC-3'	This work
Forward primer	16S F1142Holo[b]	5'-GAGAACTTTAAGAAGACTGCC-3'	This work
Reverse primer	16S R515 ND	5'-ACCGCGGCTGCTGGCAC-3'	15
Reverse primer	16S R1328HoloCaedi[b]	5'-TAGCGATTCCAACTTCATG-3'	This work
Reverse primer	16S R1488 Holo[b]	5'-TACCTTGTTACGACTTAACC-3'	This work
Reverse primer	16S R1522	5'-GGAGGTGATCCAACCGCA-3'	67
Probe	EUB338	5'-GCTGCCTCCCGTAGGAGT-3'	91
Probe	ALF1b	5'-CGTTCGYTCTGAGCCAG-3'	72
Probe	H16-23a	5'-TTCCACTTTCCTCTACCG-3'	28
Probe	GortProb659	5'-TTCCGTTTTCCTCTACCA-3'	This work

[a] Slightly modified from the original version

[b] These primers were developed on sequences of HLBs and other *Rickettsiales* in order to improve amplification and sequencing, but are not specific and their sequences cannot be employed as diagnostic markers

DNA extraction and 16S rRNA gene sequencing. About 50 infected cells from TS-j, TS-a and the *P. caudatum* cultures were individually collected, washed several times in sterile water and stored in ethanol 70% at -22 °C. Total DNA was extracted using the NucleoSpin™ Plant II DNA extraction kit (Macherey-Nagel GmbH & Co., Düren NRW, Germany), following the protocol for fungal DNA extraction.

Polymerase chain reactions (PCRs) were performed in a C1000™ Thermal Cycler (BioRad, Hercules, CA) with the high-fidelity (error rate of 8.7×10^{-6}) TaKaRa Ex *Taq* (TaKaRa Bio Inc., Otsu, Japan). A negative control without DNA was included in each experiment. 5 µL of PCR products were evaluated through electrophoresis on 1% agarose gel (*Gelly*Phor LE, EuroClone, Milano, Italy) and subsequent ethidium bromide staining. The remaining products were purified for subsequent uses with the NucleoSpin™ Extract II kit (Macherey-Nagel). A list of the primers employed in this work is shown in Table 3.1.

Almost full-length 16S rRNA genes were amplified with the forward primer 16S alfa F19a and the reverse primer 16S R1488 Holo; 3 min denaturation of the DNA at 94 °C were followed by 35 cycles at 94 °C (30 s), 57 °C (30 s) and 72 °C (120 s) and a final elongation step at 72 °C for 6 min. The PCR product from TS-j was directly sequenced in both directions according to Vannini and coworkers[15] with three internal primers: 16S R515 ND, 16S F343 ND and 16S F785 ND. The PCR product from TS-a was inserted in a pCR®2.1-TOPO® plasmid vector (TOPO TA

Cloning®; Invitrogen, Carlsbad, CA). Competent *E. coli* cells Mach1®-T1^{R} (Invitrogen) were transformed with the recombinant plasmids. PCR-amplified inserts obtained from 19 correctly transformed clones were screened through restriction fragment length polymorphism (RFLP) analysis with Bsu*RI* (Hae*III*; Fermentas International Inc., Canada) as the restriction enzyme. RFLP results revealed a main restriction pattern, labeled as "A" (6 clones), two less represented patterns, labeled "B" and "C" respectively (2 clones each), and nine unique patterns. Plasmid DNA was extracted (PureLink™ Quick Plasmid Miniprep Kit, Invitrogen) from three representative clones for pattern "A", one for pattern "B" and one for pattern "C". Inserts were sequenced full-length with vector-specific primers.

For the "classical holosporas" *H. obtusa* and *H. undulata* harbored by *P. caudatum*, the 16S rRNA genes were amplified with the primers 27F and 16S R1522b. Two semi-nested PCRs were performed using a 1/100 dilution of the first PCR product to obtain sufficient material for direct sequencing, applying primers 16S F114HoloCaedi and 16S R1488 Holo. A touchdown PCR program[61] with an annealing temperature of 60 °C for the first 5 cycles, then 58 °C (10 cycles) and finally 55 °C (15 cycles) was carried out. As sequencing primers 16S F114HoloCaedi, 16S F1142Holo, and 16S R1328HoloCaedi were used.

The NCBI BLASTN software[62] was employed for preliminary sequence comparison.

Sequence availability and phylogenetic analysis. The characterized sequences are available from DDBJ/EMBL/GenBank databases under the following accession numbers: HE797905-HE797912.

The sequences of *Holospora obtusa*, *Holospora undulata*, and those obtained from TS-j and TS-a were first aligned against more than 450,000 prokaryotic sequences from the SILVA 104 database[63] with the automatic aligner of the ARB software package[64]. This alignment was then manually edited to optimize base-paring in the predicted rRNA stem regions. The sequences obtained from TS-j and TS-a were identical, hence the sequence from TS-a was discarded in order to avoid redundancy. A detailed phylogenetic analysis was then performed with 32 additional sequences clustering in the order *Rickettsiales* according to the ARB general tree, and 7 other sequences belonging to the class *Alphaproteobacteria* as outgroup. Only sequences without ambiguous bases were chosen. The sequence belonging to *H. elegans* 16S rRNA gene (accession number: AB297813[35]) was not included in the main analysis due to its short length.

In order to obtain a provisional hypothesis on its phylogenetic position, we added it to the final tree with the ARB Quick-Add function.

Sequence lengths were reduced to that of the shortest one. Columns of the alignment including only one non-gap character were excluded. The final character matrix consisted of 1,190 nucleotide columns. jModelTest[65,66] was employed to select the substitution model that fits best the data; the TREE-PUZZLE[67] likelihood mapping method was employed to test the amount of evolutionary information contained in the character matrix. Tree reconstructions were performed through different inferring methods: the Phylip DNAPARS[68] and PHYML[65] software for maximum parsimony (MP) and maximum likelihood (ML) analyses respectively were provided by the ARB package. Statistical reliability of nodes was evaluated through bootstrap analysis with 500 pseudoreplicates for the ML method. Similarity matrix and Neighbor Joining tree (NJ[69]) were built with the software ARB NJ from the same package, using the "similarity" and "felsenstein" correction respectively. Similarity values were calculated on the same character matrices employed for phylogenetic analyses. Bayesian Inference analysis (BI) was performed with MrBayes[70], using three different Markov Chain Monte Carlo runs, with one cold chain and three heated chains each, running for 500,000 generations.

Fluorescence *in situ* hybridization (FISH). The macronuclear endosymbionts of TS-a were lost soon after the population was established (see Results). Hence, FISH experiments were performed only on TS-j. Cells were transferred in SMB (Synthetic Medium for *Blepharisma*[71]) and starved for a few days; they were then fixed on slides using formaldehyde (4% in PBS) or osmium tetraoxide vapors (4% in H_20) and dehydrated in ethanol. Experiments were performed as described by Manz and coworkers[72] with 0% and 30% formamide in the hybridization buffer. All experiments included a negative control without probes to test for autofluorescence. The slides were observed with a Zeiss Axioplan (Carl Zeiss, Oberkochen, Germany) and a Leica DMR light microscopes equipped for epifluorescence. A list of the probes employed in this work is shown in Table 3.1.

Preliminary experiments were performed with the oligonucleotide probe H16-23a. The probe GortProb659 was appositely designed to match only the 16S rRNA sequences obtained from TS-j and TS-a. Its specificity was tested *in silico* using the RDP (Ribosomal Database Project) database[73]. GortProb659 was used together with the universal eubacterial probe

EUB338 or the alphaproteobacterial-specific probe ALF1b in order to verify that no other bacterial symbiont inhabits the ciliate cells.

Results

Bacterial morphology and life cycle. The MA of both *P. jenningsi* and *P. quadecaurelia* from the native community were found to be infected with a new kind of HLB (Figures 3.1a-c, e, f). All the established *P. quadecaurelia* clonal lines (n=7) and the experimental population TS-a lost the infection in 1-2 weeks. Further investigations were performed with infected *P. jenningsi* lines (n=8), which manifested a stable infection over a period of at least one year. Two types of straight non-motile bacteria with different size and structure were observed in the infected MA (Figures 3.1 and 3.2). Small and short (1.0-3.0 x 0.7-0.8 μm) rod-like forms manifested the typical homogenous prokaryotic cytoplasm (Figures 3.1e-g; 3.2a, c). The second type consisted of longer and a bit wider straight bacteria (4.0-8.0 x 0.9-1.0 μm) with slightly tapered ends, differentiated cytoplasmic and periplasmic parts and a recognition tip-like structure (Figures 3.1e-g; 3.2b, c). The recognition tip contained less osmiophilic material than those of *H. obtusa* and *H. elegans* and was further subdivided (Figures 3.2b, c). Some intermediate forms (large, but not compartmentalized) could also be recorded in the same infected nucleus (Figure 3.1g).

The microorganism has a typical HLB life cycle with IF (large) and RF (small) forms, which could be completed (at least during experimental infection) in 5-7 days. Sometimes, infected MA may be overpopulated by bacteria (several hundred cells) and become distinctively larger in size than uninfected nuclei (Figures 3.1a, b).

During division of infected MA the HLB never produced the connecting piece (Figure 3.1d), an equatorial part of the dividing host nucleus where the majority of IFs collects, the feature manifested by "classical holosporas". The daughter cells always inherited both bacterial forms in their MA. However, some IFs were observed in the cytoplasm at different stages of the host cell cycle, possibly after releasing from the infected nucleus by an unknown mechanism (Figure 3.1f).

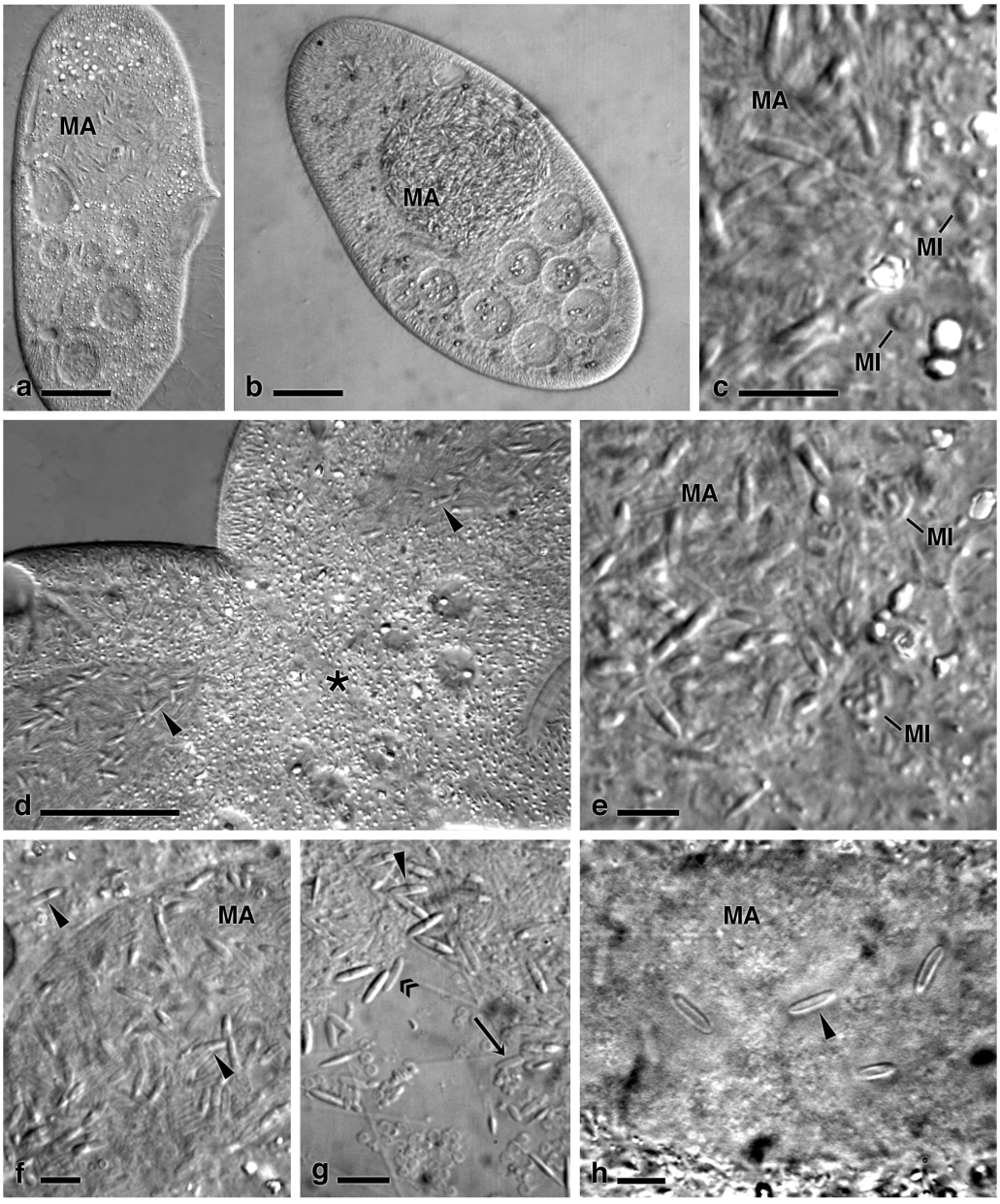


Figure 3.1 Light microscopy observations of *Paramecium jenningsi* (a-b, d-h) and *Paramecium quadecaurelia* (c). *P. jenningsi* cell of the original population with infected (a) or hyperinfected (b) macronucleus (MA); detail of micronuclei (MI) and infected MA of *P. quadecaurelia* (c); a cell undergoing division shows the absence of the equatorial connecting piece (d); bacterial reproductive and infectious forms in the MA (e); infectious forms in the ciliate's cytoplasm (f); crashed infected MA (g); newly infected MA with dedifferentiating infectious forms (h). Arrowheads indicate infectious forms, arrows reproductive forms and double arrowheads transitional forms. The asterisk indicates the cleavage furrow. Bars stand for 20 µm (a, b, d) and 5 µm (c, e-h).

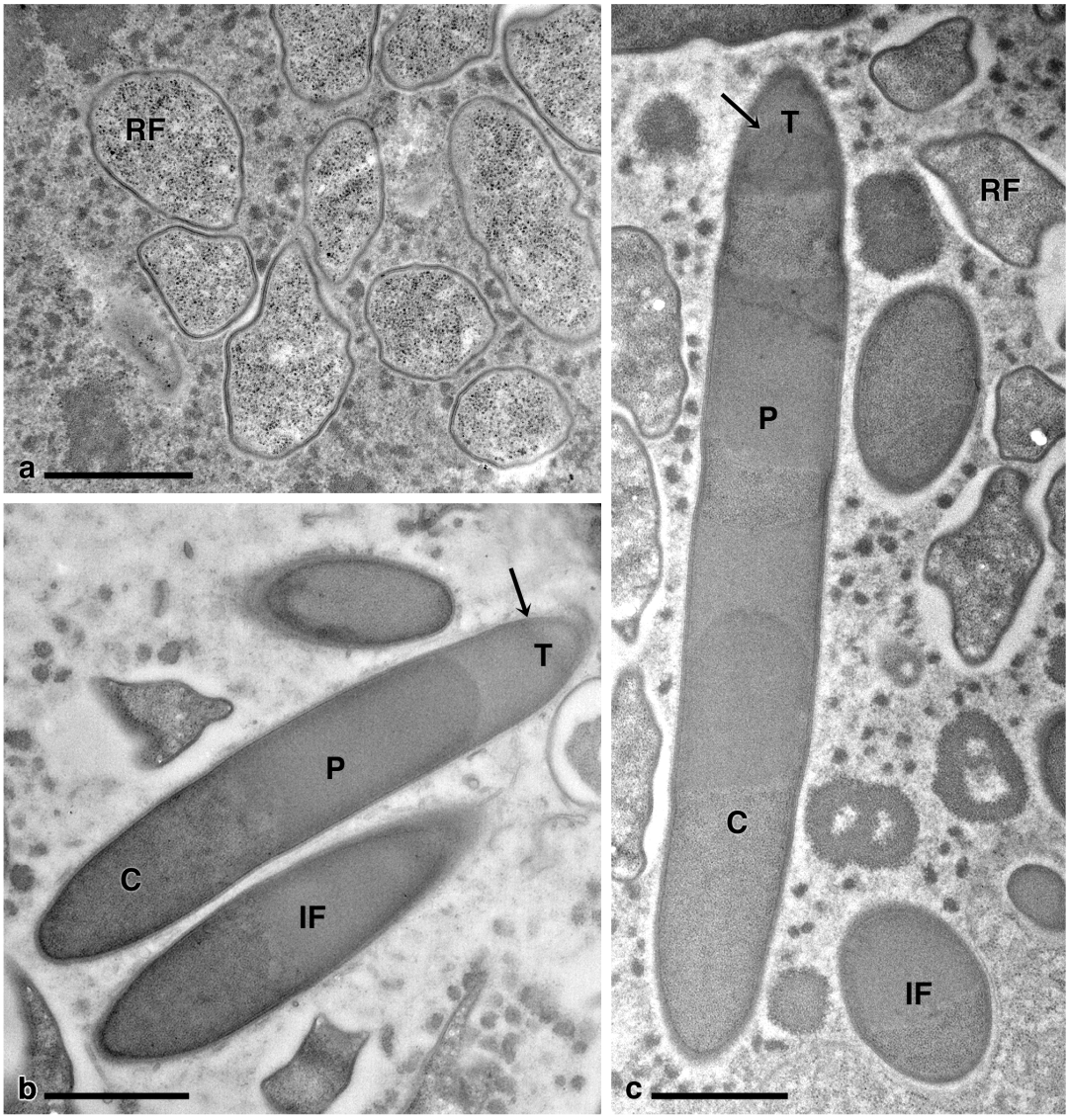


Figure 3.2 Ultrastructural morphology of the novel *Holospor*a-like bacteria harbored in the macronucleus of *Paramecium jenningsi*. Reproductive forms (RF) (a, c); infectious forms (IF) showing compartmentalization (b, c). C, bacterial cytoplasm, P, periplasmic space, T, recognition tip. The arrowhead indicates the subdivision of T. Bars stand for 1 μm.

Infection capabilities. Aposymbiotic cells of *P. jenningsi* could be experimentally infected by the HLB in 1-2 h (Figure 3.1h). In the experiments, entrance in the target nucleus was also recorded for *P. schewiakoffi, P. quadecaurelia*, *P. sonneborni* and even the less closely related *P. caudatum*, but only in *P. jenningsi* the HLB could complete its life cycle: 48 h after the entrance in the MA only RFs could be detected (Figure 3.2a); then, a fraction of them differentiated into IF during the next 72-96 h.

P. quadecaurelia and *P. sonneborni* revealed higher susceptibility for the experimental infection. In all three repetitions IFs were recorded in the MA of all paramecium cells after 1-2 h. *P. schewiakoffi* cells were infected in two out of three repetitions (70% and 56% of cells infected), and in the case of *P. caudatum* successful experimental infection was recorded only once (27% of cells infected). In most cases, the mean number of IFs 2 hours after the mixing with homogenate was more than 5 per nucleus.

No mortality of *Paramecium* cells has been observed during the experiments.

Molecular characterization of the novel HLB. A 16S rRNA gene sequence of 1,398 bp length was obtained for the population TS-j through direct sequencing. Sequences from the three clones representatives of the dominant RFLP pattern "A" were compared to obtain a consensus sequence for the 16S rRNA gene of the TS-a endosymbiont. The comparison of the bacterial sequences deriving from the two different hosts revealed 100% identity. These sequences shared only a modest similarity (90.5% according to BLASTN) with one of the available sequence of *H. obtusa* 16S rRNA gene (accession number: X58198[28]). Sequences from clones belonging to RFLP patterns "B" (accession number: HE797911) and "C" (accession number: HE797912) showed the highest similarity with the *Acetobacteraceae* bacterium strain SHB-3 (99.9%; accession number: HQ687487) and *Magnetospirillum bellicus* strain VDY (95.8%; accession number: EF405824) respectively. These sequences probably derive from contaminating *Rhodospirillales* bacteria in the medium.

No positive signals were observed in FISH experiments with probe H16-23a. This was expected, as the corresponding region of the 16S rRNA contains 3 mismatches. The sequence-specific probe GortProb659 bound to IF- and RF-like bodies inside the TS-j macronucleus in FISH observations, thus demonstrating that the characterized 16S rRNA gene sequence actually derives from the HLB. The number of bacterial cells in macronuclei varied greatly (from less than 5 to hundreds), but all inspected TS-j cells (n=50) hosted at least two HLB cells. Double hybridization experiments with probes GortProb659 and EUB338 or ALF1b (Figure 3.3) suggested that the HLBs are the only bacteria harbored by TS-j cells. As negative control, GortProb659 was also tested on the *Paramecium biaurelia* strain FGC3 infected with macronuclear *Holospora caryophila* (Vitali & Schrallhammer, personal communication). In this experiment, no positive signal was observed (data not shown).

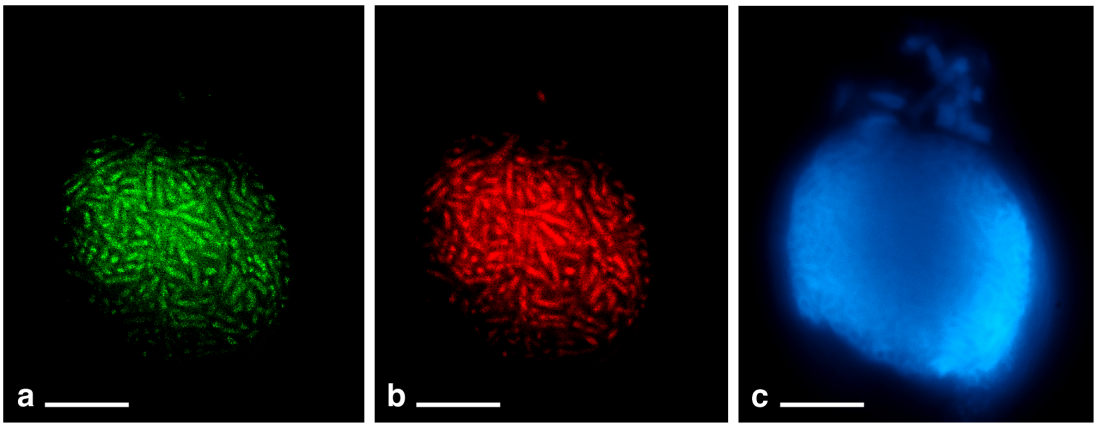


Figure 3.3 FISH results on fixed *Paramecium jenningsi* TS-j cells. Positive signals of the probes ALF1b conjugated with AlexaFluor488® (a) and GortProb659 conjugated with fluorescein (b) are shown. The ciliate macronucleus stained by 4',6-diamidino-2-phenylindole (DAPI) is shown in (c). Bars correspond to 10 µm.

The sequence of probe GortProb659 matches only 10 sequences already present in the RDP database, and all belong to uncultured bacteria. The probe contains two central mismatches with the 16S rRNA gene sequence of "*Holospora curviuscula*" and three with all the available sequences of *H. obtusa*, *H. elegans* and *H. undulata*. The sequence of the newly designed probe was deposited at probeBase[74].

Molecular characterization of *H. obtusa* and *H. undulata*. Sequences of 1,312 bp and 1,293 bp were obtained for *H. obtusa* and *H. undulata* respectively. The similarity values calculated by BLASTN with the sequence of *H. obtusa* X58198 were 99.5% and 98.5%.

Phylogenetic analysis. The Akaike Information Criterion (AIC) calculated by jModelTest selected the GTR+I+G model of substitution. TREE-PUZZLE likelihood mapping estimated that about 92.2% of the quartets had well-defined topologies, less than 1.9% of the characters in the matrix consisted of gaps and that only 2 out of 42 sequences had a statistically significant difference in base composition ("*Candidatus* Hepatincola porcellionum", accession number: AY189806; *Wolbachia pipientis*, accession number: X61768).

Figure 3.4 Maximum likelihood phylogenetic tree of the order *Rickettsiales*. Accession numbers of the sequences employed are shown. The parameters associated with nodes represent bootstrap and Posterior Probability (values below 70|0.85 are not shown). The 16S rRNA gene sequences characterized in this study are in bold characters. The bar corresponds to an estimated sequence divergence of 10%.

Most of the nodes in the ML tree (Figure 3.4) are supported by good statistical values of bootstrap/Posterior Probabilities. The monophyly of families *Anaplasmataceae*, *Rickettsiaceae*and of the "*Candidatus* Midichloria"* clade[13] is recovered with maximal support. They form a major clade in the order *Rickettsiales*; the relationships within this clade are also well-defined, with *Anaplasmataceae* being most closely related to the "*Candidatus* Midichloria"

* Currently named "*Candidatus* Midichloriaceae", see next Chapters

clade. The other major division of *Rickettsiales* includes many protists' endosymbionts as well as "*Candidatus* Hepatincola porcellionum", an endosymbiont of the common woodlouse *Porcellio scaber*[75]. The two sequences of "*Candidatus* Paracaedibacter acanthamoebae" and "*Candidatus* Paracaedibacter symbiosus" don't cluster together. "*Candidatus* P. acanthamoebae" is indeed more closely related to another *Acanthamoeba* endosymbiont, "*Candidatus* Odyssella thessalonicensis".

The HLBs form a monophyletic and highly supported group. Sequences of *H. obtusa* collected by different authors cluster together, and form the sister group of *H. undulata*. The divergence of "*H. curviuscula*" is more ancient. The endosymbiont of TS-j associates to the cluster of *Holospora* species. Sequences of uncultured bacteria collected from non-aquatic environments are only distantly related, but even more distant are 16S rRNA gene sequences of other *Paramecium* endosymbionts like "*Candidatus* Paraholospora nucleivisitans", *Caedibacter caryophilus* and "*Caedibacter macronucleorum*". MP and NJ trees have slightly different topologies, but all the highly supported (statistical values of 75%/0.90 or higher) nodes of the ML tree are recovered in both (data not shown).

The sequence of *H. elegans* added to the final ML by Quick-Add clusters with that of *H. undulata* (Figure 3.5).

The similarity values calculated by ARB NJ are all above 99% for different sequences of *H. obtusa*, above 96% for the *Holospora* genus and above 91% for the HLBs clade.

Discussion

Characteristics and taxonomy of the novel *Holospora*-like bacterium. The HLBs found in the naturally coexisting *P. jenningsi* and *P. quadecaurelia* Thai populations (Figures 3.1a-f) have the same general morphology and identical 16S rRNA gene sequences. While the infection was maintained for months in the isolated *P. jenningsi* clones, the nuclear bacteria quickly disappeared from cultures of *P. quadecaurelia* isolated from the original sample. Moreover, it was demonstrated that living HLBs obtained from *P. jenningsi* infected cells were capable of entering the aposymbiotic *P. quadecaurelia* MA (as well as those of other tested *Paramecium* spp.), but not of maintaining the infection for more than a few days. All these

observations show that we are dealing with a single infectious bacterial strain, well adapted for the life cycle inside the Thai *P. jenningsi* MA. Its presence in the natural population of *P. quadecaurelia* was probably due to the continual exposure to the source of infection, the *P. jenningsi* population.

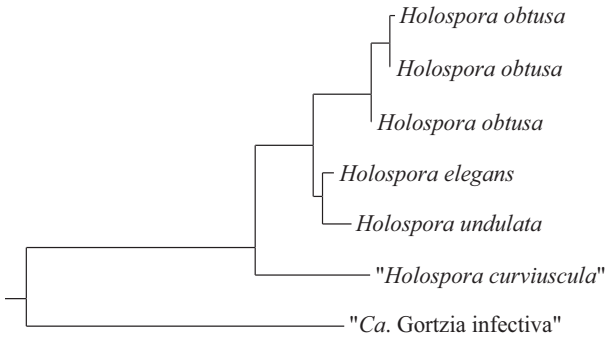


Figure 3.5 The HLB clade as inferred in the ML tree, with the sequence of *Holospora elegans* (accession number: AB297813) added using the Quick-Add function of the ARB software package[64].

The characterized bacterium shares many morphological and life cycle similarities with *Holospora*. The pattern of features usually employed to identify HLBs is unique and sets it apart from currently described species. The compartmentalized ultrastructure of the IF recognition tip is especially noteworthy and not previously described. The novel HLB belongs to the group of those unable to induce the formation of the connecting piece[49]. It is also the first HLB found inside the morphospecies *P. jenningsi*[19,29]. Our experiments gave evidences of some degree of adaptation of the symbiont to its host, because this novel HLB cannot maintain the infection inside different, although closely related, *Paramecium* species. Due to the survival of all cells during the infection experiments, we can conclude that the HLB do not produce any killer effect on the potential ciliate hosts.

The molecular characterization and phylogenetic analysis concur with morphological observations in separating the *P. jenningsi* endosymbiont from "classical holosporas". It is noticeable that all available 16S rRNA gene sequences of bacteria classified in the *Holospora* genus belong to the connecting piece-inducing group, and share similarity values higher than 95%. This threshold has been informally proposed as a good indicator for uniting different bacterial species in a single genus[76]. On the contrary, the sequence from the novel HLB is significantly different, with similarity values around 91%-92% with *Holospora* species. On the basis of these considerations, we have classified the here described macronuclear endosymbiont of *P. jenningsi* in a new species and genus. Because HLBs are not cultivable outside their hosts, a complete culture-dependent characterization is not possible; hence, we propose the provisional

name (according to Murray and Schleifer[77], Murray and Stackebrandt[78]) "*Candidatus* Gortzia infectiva" in honor of Professor emeritus Hans-Dieter Görtz, the prominent specialist in the field of ciliate's prokaryotic symbionts investigations and our appreciated colleague.

A diagnostic description of the new taxon follows at the end of this section.

Systematics and taxonomy of *Holospora*-like bacteria. The present study is a significant step forward on enlightening the phylogeny and diversity of *Holospora* and *Holospora*-like bacteria. We provided a third sequence of *H. obtusa*, adding further evidence that morphological and molecular identification of this *Holospora* morphospecies are in accordance. We also obtained the first sequence of *H. undulata*, the type species of the genus, and demonstrated that it falls inside the clade of "classical holosporas", those able to induce the connecting piece formation. Finally, this represents the first molecular characterization of an HLB unable to induce the connecting piece, and its classification in a different genus, closely related to *Holospora*.

Our phylogenetic analysis is the first molecular indication that HLBs of both groups previously identified (those that are able to induce the connecting piece and those that are not) form a clade. According to our data, the connecting piece induction capability is an apomorphy of the "classical holosporas". Future researches on *Holospora* species lacking this character, like *H. caryophila*, will help in clarifying this interpretation. *H. obtusa*, *H. undulata* and *H. elegans*, the three species exclusively found in *P. caudatum*[30], are most closely related to each other, although the position of *H. elegans* is provisional and should be confirmed obtaining the full sequence of its 16S rRNA gene. It appears that the specific relationships with *P. caudatum* was developed once in the common ancestor of these organisms, and never lost. "*H. curviuscula*", that is able to infect both the MA and the MI of its host *Paramecium bursaria*[16,79], falls basally to the clade of *Holospora* species that typically show an exclusive nuclear localization. This suggests that the nuclear specialization is more recent than the host specialization. The more universal HLB features like the peculiar morphology of the infectious form, the overall life cycle and the ability to reproduce only inside the host's nuclear apparatus, clearly share a single origin, and strongly separate these bacteria from other *Paramecium* endosymbionts like "*Candidatus* Paraholospora nucleivisitans" or *Caedibacter*.

Caedibacter[30,33,52] supplies its hosts with the so called "killer trait" against sensitive strains, providing it a competitive advantage[24,80,81]. Bacteria of this genus can be identified by a

cytoplasmic inclusion known as the "R-body", a proteinaceous ribbon tightly coiled inside the cell[82,83]. *Caedibacter* species are not infectious under natural conditions[24,80,81,84], and differ in their cellular localization[52,85]. The genus itself is polyphyletic, because *C. caryophilus* and "*C. macronucleorum*" belong to *Rickettsiales* (*Alphaproteobacteria*), while *Caedibacter taeniospiralis* clusters inside the class *Gammaproteobacteria*[8,86,87]. No molecular data on other species is currently available. "*Candidatus* Paraholospora nucleivisitans" was only recently investigated in *Paramecium sexaurelia*[1]. It is unusual in its capability of residing both inside the cytoplasm and the macronucleus, probably shuttling between them. It is not infectious, and it has a unique sigmoid-shaped morphology.

Caedibacter caryophilus, "*Caedibacter macronucleorum*", "*Candidatus* Paraholospora nucleivisitans" and "*Candidatus* Gortzia infectiva" are all members of the order *Rickettsiales*, whose phylogeny has been here reconstructed. The topology of our tree is generally in good accordance with those of previous papers[2,13]. "*Candidatus* Paraholospora nucleivisitans" would indeed branch basally with respect to the HLB clade, if no uncultured bacteria were included in the analysis. This could lead to the intriguing hypothesis that its ability to live in different compartments of the eukaryotic cell represents an intermediate stage between exclusively cytoplasmic and exclusively nuclear life-style. A more thorough analysis shows, however, that many sequences from uncultured bacteria fall nearer to the HLB clade. In our analysis, some representatives were included, coming from prairie dog's fleas[88] and plant-associated soil[89], both environments very unlikely for *Paramecium* species. *C. caryophilus* and "*C. macronucleorum*" are even more distantly related to HLBs. These observations suggest the conclusion that the superficial characters uniting these very different symbionts (like their presence inside the same host genus or the sporadic ability to infect its nuclear apparatus) probably arose several times independently.

Additional support for the argumentation that the *Paramecium* symbionts should not be lumped together comes from the abundance of less-known ciliate endosymbionts recently discovered, that fall inside the order *Rickettsiales*[2,12,20]. It is likely that even more will be found in the future, when research on protists' bacterial endosymbionts will be more widespread. All current evidences support the view that these bacteria established their life style independently, or perhaps as a consequence of the preadaptation for endocellularity present in the common ancestor of the entire order.

For these reasons, we think that it would be better to opt for a conservative view, and formally describe family *Holosporaceae* as only including those bacteria possessing the HLB traits. Although it is generally practical to establish more comprehensive taxa, the unusually well-defined biology of this group of bacteria suggests otherwise. Their characters can be easily diagnosed by morphological observation alone, and until now have proven to be in very good accordance with molecular phylogeny. They seem to be true synapomorphies. On the contrary, a family so extended as to include "*Candidatus* Paraholospora nucleivisitans" and the *Caedibacter* species belonging to *Rickettsiales* would present no reliable shared characters. Its monophyly would also be less supported by our analysis. We strongly suggest to adhere to the original description of the family *Holosporaceae*, and to consider the other cited *Paramecium* symbionts as *incertae sedis* in order *Rickettsiales*. Their taxonomic position should be revised only when more data will be presented. This conclusion is even more cogent for distantly related bacterial endosymbionts like "*Candidatus* Odyssella" and "*Candidatus* Paracaedibacter" that were provisionally included *incertae sedis* in the family *Holosporaceae*.

Description of "*Candidatus* Gortzia infectiva". *Gortzia infectiva* (Gor'tzi.a in.fec.ti'va; N.L. fem. n. *Gortzia*, in honor of Professor emeritus Hans-Dieter Görtz; N.L. adj. *infectivus*, corrupting, infectious).

Rod-shaped gram-negative bacteria, with differentiated reproductive (RF) and infectious (IF) forms. RF 1.0-3.0 µm x 0.7-0.8 µm with homogeneous cytoplasm without visible inclusions. IF 4.0-8.0 x 0.9-1.0 µm straight rods, with slightly tapered ends and extensive periplasmic space including a recognition tip divided in two parts with different osmiophilic density. Macronuclear endosymbiont of the free-living protist *Paramecium jenningsi*, identified in a sample taken from Chaweng Lake, Samui Island (Thailand). Capable of horizontal and vertical transmission in the host species. Can temporarily infect the macronuclei of *Paramecium quadecaurelia*, *Paramecium schewiakoffi*, *Paramecium sonneborni* and *Paramecium caudatum*. Does not induce the formation of a distinctive "connecting piece" during host cell division. It has no killing activities. Basis of assignment: 16S rRNA gene sequence (DDBJ/ EMBL/ GenBank accession number: HE797907) and positive matching with the 16S rRNA-targeting oligonucleotide probe GortProb659 (5'-TTCCGTTTTCCTCTACCA-3'); morphological characters pattern as above. Uncultured thus far.

References

1. Eschbach E, Pfannkuchen M, Schweikert M, Drutschmann D, Brümmer F, *et al.* (2009) "*Candidatus* Paraholospora nucleivisitans", an intracellular bacterium in *Paramecium sexaurelia* shuttles between the cytoplasm and the nucleus of its host. Syst Appl Microbiol. 32: 490-500
2. Ferrantini F, Fokin SI, Modeo L, Andreoli I, Dini F, *et al.* (2009) "*Candidatus* Cryptoprodotis polytropus", a novel *Rickettsia*-like organism in the ciliated protist *Pseudomicrothorax dubius* (Ciliophora, Nassophorea). J Eukaryot Microbiol. 56: 119-129
3. Fokin SI. (2012) Frequency and biodiversity of symbionts in representatives of the main classes of Ciliophora. Eur J Protistol. 48:138-148
4. Irbis C, Ushida K. (2004) Detection of methanogens and proteobacteria from a single cell of rumen ciliate protozoa. J Gen Appl Microbiol. 50: 203-212
5. Petroni G, Spring S, Schleifer KH, Verni F, Rosati G. (2000) Defensive extrusive ectosymbionts of *Euplotidium* (Ciliophora) that contain microtubule-like structures are bacteria related to *Verrucomicrobia*. Proc Natl Acad Sci USA 97: 1813-1817
6. Rinke C, Schmitz-Esser S, Loy A, Horn M, Wagner M, Bright M. (2009) High genetic similarity between two geographically distinct strains of the sulfur-oxidising symbiont "*Candidatus* Thiobios zoothamnicoli". FEMS Microbiol Ecol. 67: 229-241
7. Rinke C, Schmitz-Esser S, Stoecker K, Nussbaumer AD, Molnár DA, *et al.* (2006) "*Candidatus* Thiobios zoothamnicoli", an ectosymbiotic bacterium covering the giant marine ciliate *Zoothamnium niveum*. Appl Environ Microbiol. 72: 2014-2021
8. Schrallhammer M, Fokin SI, Schleifer KH, Petroni G. (2006) Molecular characterization of the obligate endosymbiont "*Caedibacter macronucleorum*" Fokin and Görtz, 1993 and of its host *Paramecium duboscqui* strain Ku4-8. J Eukaryot Microbiol. 53: 499-506
9. Schrallhammer M, Schweikert M, Vallesi A, Verni F, Petroni G. (2011) Detection of a novel subspecies of *Francisella noatunensis* as endosymbiont of the ciliate *Euplotes raikovi*. Microb Ecol. 61: 455-464
10. Shinzato N, Watanabe I, Meng XY, Sekiguchi Y, Tamaki H, *et al.* (2007) Phylogenetic analysis and fluorescence *in situ* hybridization detection of archaeal and bacterial endosymbionts in the anaerobic ciliate *Trimyema compressum*. Microbial Ecol. 54: 627-636
11. Sun HY, Noe J, Barber J, Coyne RS, Cassidy-Hanley D, *et al.* (2009) Endosymbiotic bacteria in the parasitic ciliate *Ichthyophthirius multifiliis*. Appl Environ Microbiol. 75: 7445-7452
12. Vannini C, Ferrantini F, Ristori A, Verni F, Petroni G. (2012) Betaproteobacterial symbionts of the ciliate *Euplotes*: origin and tangled evolutionary path of an obligate microbial association. Environ Microbiol. 14: 2553-2563
13. Vannini C, Ferrantini F, Schleifer KH, Ludwig W, Verni F, Petroni G. (2010) "*Candidatus* Anadelfobacter veles" and "*Candidatus* Cyrtobacter comes", two new rickettsiales species hosted by the protist ciliate *Euplotes harpa* (Ciliophora, Spirotrichea). Appl Environ Microbiol. 76: 4047-4054
14. Vannini C, Petroni G, Verni F, Rosati G. (2005) A bacterium belonging to the *Rickettsiaceae* family inhabits the cytoplasm of the marine ciliate *Dioprhys appendiculata* (Ciliophora, Hypotrichia). Microbial Ecol. 49: 434-442

15. Vannini C, Rosati G, Verni F, Petroni G. (2004) Identification of the bacterial endosymbionts of the marine ciliate *Euplotes magnicirratus* (Ciliophora, Hypotrichia) and proposal of "*Candidatus* Devosia euplotis". Int J Syst Evolution Microbiol. 54: 1151-1156
16. Fokin SI. (2004) Bacterial endocytobionts of Ciliophora and their interactions with the host cell. Int Rev Cytol. 236: 181-249
17. Görtz HD. (1996) Symbiosis in ciliates. In: Hausmann K, Bradbury PS (eds), Ciliates. Cells as organisms, Fisher, Stuttgart, New York, pp 441-462
18. Görtz HD. (2006) Symbiotic associations between ciliates and prokaryotes. In: Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E (eds), The Prokaryotes. Third Edition, vol. 1, Springer-Verlag, New York, NY, pp 364-402
19. Görtz HD, Fokin SI. (2009) Diversity of endosymbiotic bacteria in *Paramecium*. In: Fujishima M (ed.), Endosymbionts in *Paramecium*. Microbiology Monograph, Münster, pp 131-160
20. Vannini C, Petroni G, Schena A, Verni F, Rosati G. (2003) Well-established mutualistic associations between ciliates and prokaryotes might be more widespread and diversified than so far supposed. Eur J Protistol. 39: 481-485
21. Duncan AB, Fellous S, Kaltz O. (2011) Temporal variation in temperature determines disease spread and maintenance in *Paramecium* microcosm populations. Proc R Soc B. 278: 3412-3420
22. Fellous S, Quillery E, Duncan AB, Kaltz O. (2011) Parasitic infection reduces dispersal of ciliate host. Biol Lett. 7: 327-329
23. Fels D, Kaltz O. (2006) Temperature-dependent transmission and latency of *Holospora undulata*, a micronucleus-specific parasite of the ciliate *Paramecium caudatum*. P Roy Soc Lond B Bio. 273: 1031-1038
24. Kusch J, Czubatinsky L, Wegmann S, Hübner M, Alter M, Albrecht P. (2002) Competitive advantages of *Caedibacter*-infected paramecia. Protist. 153: 47-58
25. Vannini C, Lucchesi S, Rosati G. (2007) *Polynucleobacter*: symbiotic bacteria in ciliates compensate for a genetic disorder in glycogenolysis. Symbiosis. 44: 85-91
26. Vannini C, Pöckl M, Petroni G, Wu QL, Lang E, *et al.* (2007) Endosymbiosis *in statu nascendi*: close phylogenetic relationship between obligately endosymbiotic and obligately free-living *Polynucleobacter* strains (*Betaproteobacteria*). Environ Microbiol. 9: 347-359
27. Amann RI, Ludwig WG, Schleifer KH. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev. 59: 143-169
28. Amann RI, Springer N, Ludwig W, Görtz HD, Schleifer KH. (1991) Identification *in situ* and phylogeny of uncultured bacterial endosymbionts. Nature. 351: 161-164
29. Fokin SI, Görtz HD. (2009) Diversity of *Holospora* bacteria in *Paramecium* and their characterization. In: Fujishima M (ed), Endosymbionts in *Paramecium*. Microbiology Monograph, Münster, pp 161-199
30. Görtz HD, Schmidt HJ. (2005) *Holosporaceae* fam. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds), Bergey's Manual of Systematic Bacteriology, 2nd edn, vol II, part C. Springer-Verlag, New York, pp 146-160
31. Gromov BV, Ossipov DV. (1981) *Holospora* (*ex* Hafkine 1890) nom. rev., a genus of bacteria inhabiting the nuclei of paramecia. Int J Syst Bacteriol. 31: 348-352
32. Hafkine MW. (1890) Maladies infectieuses des paramécies. Ann Inst Pasteur Paris 4: 363-379

33. Preer JR, Preer LB. (1982) Revival of names of protozoan endosymbionts and proposal of *Holospora caryophila* nom. nov. Int J Syst Bacteriol. 32: 140-141

34. Fokin SI. (2000) Host specificity of *Holospora* and its relationships with *Paramecium* phylogeny. Jpn J Protozool. 33: 94

35. Hori M., Fujii K, Fujishima M. (2008) Micronucleus-specific bacterium *Holospora elegans* irreversibly enhances stress gene expression of the host *Paramecium caudatum*. J Eukaryot Microbiol. 55: 515-521

36. Hori M, Fujishima M. (2003) The endosymbiotic bacterium *Holospora obtusa* enhances heat-shock gene expression of the host *Paramecium caudatum*. J Eukaryot Microbiol. 50: 293-298

37. Smurov AO, Fokin SI. (1999) Resistance of *Paramecium* species (Ciliophora, Peniculia) to salinity of environment. Protistology. 1: 43-53

38. Fujishima M, Hoshide K. (1988) Light and electron microscopic observations of *Holospora obtusa*: a macronucleus-specific bacterium of the ciliate *Paramecium caudatum*. Zool Sci. 5: 791-799

39. Fujishima M, Sawabe H, Iwatsuki K. (1990) Scanning electron microscopic observations of differentiation from the reproductive short form to the infectious long form of *Holospora obtusa*. J Protozool. 37: 123-128

40. Görtz HD, Ahlers N, Robenek H. (1989) Ultrastructure of the infectious and reproductive forms of *Holospora obtusa*, a bacterium infecting the macronucleus of *Paramecium caudatum*. J Gen Microbiol. 135: 3079-3085

41. Görtz HD, Dieckmann J. (1980) Life cycle and infectivity of *Holospora elegans* Haffkine, a micronucleus-specific symbiont of *Paramecium caudatum* (Ehrenberg). Protistologica. 16: 591-603

42. Dohra H, Fujishima M. (1999) Cell structure of the infectious form of *Holospora*, an endonuclear symbiotic bacterium of the ciliate *Paramecium*. Zool Sci. 16: 93-98

43. Iwatani K, Dohra H, Lang BF, Burger G, Hori M, Fujishima M. (2005) Translocation of an 89-kDa periplasmic protein is associated with *Holospora* infection. Biochem Biophys Res Commun. 337: 1198-1205

44. Fujishima M. (2009) Infection and maintenance of *Holospora* species in *Paramecium caudatum*. In: Fujishima M (ed), Endosymbionts in *Paramecium*. Microbiology Monograph, Münster, pp 201-225

45. Wiemann M, Görtz HD. (1989) Release of the endonucleobiotic bacterium *Holospora elegans* from its host cell *Paramecium caudatum*. Eur J Protistol. 25: 100-108

46. Fujishima M, Kawai M. (1997) Acidification in digestive vacuoles is an early event required for *Holospora* infection of *Paramecium* nucleus. In: Achenk HEA, Herrmann RG, Jeon KW, Müller NE, Schwemmler W (eds), Eukaryotism and symbiosis. Springer, Berlin, pp 367-370

47. Sabaneyeva EV, Derkacheva ME, Benken KA, Fokin SI, Vainio S, Skovorodkin IN. (2009) Actin-based mechanism of *Holospora obtusa* trafficking in *Paramecium caudatum*. Protist. 160: 205-219

48. Sabaneyeva EV, Fokin SI, Gavrilova EV, Kornilova ES. (2005) Nocodazole inhibits macronuclear infection with *Holospora obtusa* in *Paramecium caudatum*. Protoplasma. 226: 147-153

49. Fokin SI, Brigge T, Brenner J, Görtz HD. (1996) *Holospora* species infecting the nuclei of *Paramecium* appear to belong into two groups of bacteria. Eur J Protistol. 32: 19-24

50. Fokin SI, Sabaneyeva EV. (1997) Release of endonucleobiotic bacteria *Holospora bacillata* and *Holospora curvata* from the macronucleus of their host cells *Paramecium woodruffi* and *Paramecium calkinsi*. Endocyt Cell Res. 12: 49-55

51. Vakkerov-Kouzova ND, Rautian MS. (2011) Obtaining and characterization of "*Holospora curviuscula*" and *Holospora obtusa*, bacterial symbionts of the macronuclei of *Paramecium bursaria* and *Paramecium caudatum*. Microbiology. 80: 728-732

52. Schrallhammer M, Schweikert M. (2009) The killer effect of *Paramecium* and its causative agents. In: Fujishima M. (ed), Endosymbionts in *Paramecium*. Microbiology Monograph, Münster, pp 227-246
53. Birtles RJ, Rowbotham TJ, Michel R, Pitcher DG, Lascola B, *et al.* (2000) "*Candidatus* Odyssella thessalonicensis" gen. nov., sp. nov., an obligate intracellular parasite of *Acanthamoeba* species. Int J Syst Evol Microbiol. 50: 63-72
54. Horn M, Fritsche TR, Gautom RK, Schleifer KH, Wagner M. (1999) Novel bacterial endosymbionts of *Acanthamoeba* spp. related to the *Paramecium caudatum* symbiont *Caedibacter caryophilus*. Environ Microbiol. 1: 357-367
55. Przyboś E, Fokin S, Stoeck T, Schmidt HJ. (1999) Occurrence and ecology of *Paramecium jenningsi* strains. Folia Biol – Krakow. 47: 53-59
56. Wichterman R. (1953) The biology of Paramecium. Blakiston, McGraw-Hill, New York, p 527
57. Skovorodkin IN. (1990) A device for immobilizing biological objects in the light microscope studies. Tsitologiya. 32: 301-302 (In Russian with English summary)
58. Fokin SI. (1997) Morphological diversity of the micronuclei in *Paramecium*. Arch Protistenkd. 148: 375-387
59. Fokin SI. (2011) *Paramecium* genus: biodiversity, some morphological features and the key to the main morphospecies discrimination. Protistology. 6: 227-235
60. Preer LB. (1969) Alpha, an infectious macronuclear symbiont of *Paramecium aurelia*. J Protozool. 16: 570-578
61. Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS. (1991) "Touchdown" PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res. 19: 4008
62. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, *et al.* (1997) Gapped BLAST and PSI BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389-3402
63. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, *et al.* (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 35: 7188-7196
64. Ludwig W, Strunk O, Westram R, Richter L, Meier H, *et al.* (2004) ARB: a software environment for sequence data. Nucleic Acids Res. 32: 1363-1371
65. Guindon S, Gascuel O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 52: 696-704
66. Posada D. (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol. 25: 1253-1256
67. Schmidt HA, Strimmer K, Vingron M, von Haeseler A. (2002) TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics. 18: 502-504
68. Felsenstein J. (1989) PHYLIP-Phylogeny Inference Package (version 3.2). Cladistics. 5: 164-166
69. Saitou N, Nei M. (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol. 4: 406-425
70. Huelsenbeck JP, Ronquist F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754-755
71. Miyake A. (1981) Cell-cell interaction by gamones in *Blepharisma*. In: O'Day DH, Horgen PA (eds.), Sexual interactions in eukaryotic microbes. Academic Press, New York, pp 95-129
72. Manz W, Amann RI, Ludwig W, Wagner M, Schleifer KH. (1992) Phylogenetic oligodeoxynucleotide probes for the major subclasses of Proteobacteria: problems and solutions. Syst Appl Microbiol. 15: 593-600

73. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, *et al.* (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res. 37: D141-D145

74. Loy A, Maixner F, Wagner M, Horn M. (2007) probeBase – an online resource for rRNA-targeted oligonucleotide probes: new features 2007. Nucleic Acids Res. 35: D800-D804

75. Wang Y, Stingl U, Anton-Erxleben F, Zimmer M, Brune A. (2004) "*Candidatus* Hepatincola porcellionum" gen. nov., sp. nov., a new, stalk-forming lineage of *Rickettsiales* colonizing the midgut glands of a terrestrial isopod. Arch Microbiol. 181: 299-304

76. Ludwig W, Strunk O, Klugbauer S, Klugbauer N, Weizenegger M, *et al.* (1998) Bacterial phylogeny based on comparative sequence analysis. Electrophoresis. 19: 554-568

77. Murray RGE, Schleifer KH. (1994) A proposal for recording the properties of putative taxa of procaryotes. Int J Syst Bacteriol. 44: 174-176

78. Murray RGE, Stackebrandt E. (1995) Implementation of the provisional status *Candidatus* for incompletely described prokaryotes. Int J Syst Bacteriol. 45: 186-187

79. Borchsenius ON, Skoblo II, Ossipov DV. (1983) *Holospora curviuscula* – a new species of macronuclear symbiotic bacteria of *Paramecium bursaria*. Tsitologiya. 25: 91-97 (In Russian with English summary)

80. Landis WG. (1981) The ecology, role of the killer trait, and interactions of five species of *Paramecium aurelia* complex inhabiting the littoral zone. Can J Zool. 59: 1734-1743

81. Landis WG. (1987) Factors determining the frequency of the killer trait within populations of the *Paramecium aurelia* complex. Genetics. 115: 197-205

82. Preer JR, Stark PS. (1953) Cytological observations on the cytoplasmic factor "kappa" in *Paramecium aurelia*. Exp Cell Res. 5: 478-491

83. Schrallhammer M, Galati S, Altenbuchner J, Schweikert M, Görtz HD, Petroni G. (2012) Tracing the role of R-bodies in the killer trait: Absence of toxicity of R-body producing recombinant *E. coli* on paramecia. Eur J Protistol. 48: 290-296

84. Preer JR, Preer LB, Jurand A. (1974) Kappa and other endosymbionts in *Paramecium aurelia*. Bacteriol Rev. 38: 113-163

85. Kusch J, Stremmel M, Breiner HW, Adams V, Schweikert M, Schmidt HJ. (2000) The toxic symbiont *Caedibacter caryophila* in the cytoplasm of *Paramecium novaurelia*. Microb Ecol. 40: 330-335

86. Beier CL, Horn M., Michel R., Schweikert M, Görtz HD, Wagner M. (2002) The genus *Caedibacter* comprises endosymbionts of *Paramecium* spp. related to the *Rickettsiales* (*Alphaproteobacteria*) and to *Francisella tularensis* (*Gammaproteobacteria*). Appl Environmen Microbiol. 68: 6043-6050

87. Springer N, Ludwig W, Amann R, Schmidt HJ, Görtz HD, Schleifer KH. (1993) Occurrence of fragmented 16S ribosomal-RNA in an obligate bacterial endosymbiont of *Paramecium caudatum*. Proc Natl Acad Sci USA. 90: 9892-9895

88. Jones RT, McCormick KF, Martin AP (2008) Bacterial communities of *Bartonella*-positive fleas: diversity and community assembly patterns. Appl Environ Microbiol. 74: 1667-1670

89. Lesaulnier C, Papamichail D, McCorkle S, Ollivier B, Skiena S, *et al.* (2008) Elevated atmospheric CO_2 affects soil microbial diversity associated with trembling aspen. Environ Microbiol. 10: 926-941

90. Lane DJ. (1991) 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M (eds), Nucleic Acid Techniques in Bacterial Systematics, Wiley, New York, pp 115-147, 1991

91. Amann R, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA. (1990) Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. Appl Environ Microbiol. 56: 1919-1925

Chapter 4.

Symbionts of the "*Candidatus* Midichloriaceae" family

Original publication:

"*Candidatus* Defluviella procrastinata" and "*Candidatus* Cyrtobacter zanobii", two novel ciliate endosymbionts belonging to the "*Midichloria* clade"

Vittorio Boscaro, Giulio Petroni, Alessandro Ristori, Franco Verni, Claudia Vannini

Microbial Ecology 65(2): 302-310. February 2013.

Abstract

The "*Midichloria* clade" is a recently discovered but well established evolutionary lineage clustering inside the order *Rickettsiales* (*Alphaproteobacteria*). Not much is known about the biology of these organisms. The best characterized ones are endocellular symbionts of very different eukaryotic hosts, ranging from arthropods to protists. "*Candidatus* Midichloria mitochondrii", the most studied organism of the group, is an interesting object of study because of its unique capability to infect metazoans' mitochondria and the presence of flagellar genes in its genome. With this work we aim at increasing the knowledge on the biodiversity and phylogeny of the "*Midichloria* group". We characterized according to the "full cycle rRNA approach" two novel endosymbionts of ciliated protozoa, i.e. *Paramecium nephridiatum* and *Euplotes aediculatus*. According to the nomenclatural rules for uncultivated prokaryotes, we established the novel taxa "*Candidatus* Defluviella procrastinata" and "*Candidatus* Cyrtobacter zanobii" for the two bacterial symbionts. Our phylogenetic analysis based on 16S rRNA gene sequences confirms that the evolutionary histories of "*Midichloria* clade" representatives and of their hosts are very different. This suggests that the symbiotic processes arose many times independently, perhaps through ways of transmission still not described in *Rickettsiales*.

Introduction

The bacterial order *Rickettsiales* (*Alphaproteobacteria*) contains obligate intracellular symbionts of eukaryotic organisms, often highly specialized for parasitism[1]. It is generally acknowledged that they are the living eubacteria most closely related to the ancestor of mitochondria[2,3]. The order includes three formally recognized families: *Rickettsiaceae*, *Anaplasmataceae* and *Holosporaceae*[1]. Bacteria belonging to a fourth major lineage were repeatedly detected in recent years[4-6]. The most studied and first described organism of this novel clade is "*Candidatus* Midichloria mitochondrii"[7], an endosymbiont reported in many species of "hard ticks" of the family Ixodidae[4,7-10].

Organisms related to "*Candidatus* M. mitochondrii" were found in association with acanthamoebas[11], sponges[12,13], cnidarians[14,15], insects[5,16-18] and other arthropods[19]. Two were detected in different strains of the ciliated protist *Euplotes harpa*: "*Candidatus* Anadelfobacter veles" and "*Candidatus* Cyrtobacter comes"[6]. These two organisms, along with "*Candidatus* M. mitochondrii" and the recently described "*Candidatus* Lariskella arthropodarum"[5], are the only members of this family-like taxon that received a binomial name, although a provisional one.

The so-called "*Midichloria* clade"[6] is still much less known than the three families of the order *Rickettsiales*, but there are several reasons to think that it will become a deeply interesting field of research in the future. Hints on the involvement of some of these bacteria in human or animal diseases have already been found[19,20]. The genome of "*Candidatus* M. mitochondrii" contains 26 flagellar genes[21] that are at least partially expressed[22]; phylogenomic analyses suggest that the ancestor of mitochondria could have been a motile bacterium even though all known *Rickettsiales* are not capable of flagellar movement[21]. Moreover, "*Candidatus* M. mitochondrii" possesses the unique feature of infecting mitochondria in some of its metazoan hosts[7-9,23]. It also has a strongly sex-biased prevalence in ticks, being more common in females than in males, and preferentially infecting ovarian tissues where it secures its vertical transmission[4,8]. These last characteristics are reminiscent of those of the well known bacterium *Wolbachia pipientis* (*Anaplasmataceae*), which is a powerful manipulator of the reproductive biology of its hosts, and probably constitutes one of the driving forces in their evolution[24].

The "*Midichloria* clade" is a monophyletic, well-supported group. It is considered the sister group of *Anaplasmataceae* in most papers[4,6,7,9], although in a single analysis based on 16S

rRNA gene sequences[5] and one multigenic phylogeny[21] it clustered with *Rickettsiaceae*. The phylogenetic relationships within the clade are less clear. The evolution of symbionts and hosts are not congruent even in the two most investigated groups, ticks and stinkbugs[5,9]. The presence of environmental 16S rRNA gene sequences clustering in this clade, as well as those of symbionts of very different protists (like acanthamoebas and ciliates) obscures the picture even more. As data accumulate, the diversity of its members and their hosts will likely increase; the widespread distribution of at least "*Candidatus* M. mitochondrii" has already been proven[4,8,9].

In this work, we provide the characterization according to the "full cycle rRNA approach"[25] of two new bacteria belonging to the "*Midichloria* clade" and infecting different ciliates: *Paramecium nephridiatum* (Ciliophora, Oligohymenophorea) and *Euplotes aediculatus* (Ciliophora, Spirotrichea). We also performed a detailed phylogenetic analysis of the whole bacterial clade, based on 16S rRNA sequences. Following the current rules of prokaryotic nomenclature for uncultivated bacteria[26,27], we propose the name "*Candidatus* Defluviella procrastinata" for the endosymbiont of *P. nephridiatum* and "*Candidatus* Cyrtobacter zanobii" for the endosymbiont of *E. aediculatus*.

Materials and Methods

Sampling, identification and culturing of the ciliates. *P. nephridiatum* was sampled in January 2010 from the sludge aerobic digestion tank of the wastewater treatment plant of Sabaudia (Latina, Italy) and identified according to morphological features[28]. About 30 ciliate cells were isolated from the original medium and singly washed several times in distilled water in order to minimize the presence of contaminating organisms. The obtained polyclonal population (PAR) was then maintained and grown in artificial brackish water (5‰ salinity), at a fixed temperature of 19-20 °C, on a 12:12 h irradiance of 200 µmol photons m^{-2} s^{-1} and fed with the green alga *Dunaliella tertiolecta*.

Data on the population In of *E. aediculatus* can be found in the papers of Vannini and coworkers[29] and Boscaro and coworkers[30].

DNA extraction and 18S/16S rRNA gene sequencing. For both PAR and In, about 50 ciliate cells were individually harvested from the culture medium, put to starvation for

a few days, then washed several times in distilled water and fixed in 70% ethanol. Total genomic DNA was extracted from cell pellets using the NucleoSpin™ Plant II DNA extraction kit (Macherey-Nagel GmbH & Co., Düren NRW, Germany). All Polymerase Chain Reactions (PCRs) were performed in a C1000™ Thermal Cycler (BioRad, Hercules, CA) with the high-fidelity TaKaRa Ex *Taq* (TaKaRa Bio Inc., Otsu, Japan). Each PCR ran for 25-35 cycles, with 30'' denaturation at 94 °C, 30'' annealing at variable temperature (between 50 °C and 63 °C) and 60-120'' elongation at 72 °C.

The 18S rRNA gene sequence of PAR was obtained through PCR with the eukaryotic primers forward 18S F9 Euk 5'-CTGGTTGATCCTGCCAG-3'[31] and reverse 18S R1513 Hypo 5'-TGATCCTTCYGCAGGTTC-3'[32]. The PCR products were directly sequenced with three internal primers according to Rosati and coworkers[33]. The 18S rRNA gene sequence of In was already available (accession number: FR873713[29]).

In order to obtain the sequence of the PAR endosymbiont a PCR was performed using the eubacterial primers forward 27f 5'-AGAGTTTGATYMTGGCTCAG-3' and reverse 1492r 5'-GGNWACCTTGTTACGACTT-3', both modified from Lane[34]. The PCR products were inserted in pCR®2.1-TOPO® plasmidic vectors (TOPO TA Cloning®; Invitrogen, Carlsbad, CA), and the plasmids were transformed in competent *Escherichia coli* cells Mach1®-T1^{R} (Invitrogen). 72 colonies containing the plasmid and the insert were screened through restriction fragment length polymorphism (RFLP) analysis. 12.5% of the screened clones showed the same RFLP pattern. Three of these clones (labeled 5, 19 and 26) were sequenced in both directions, and the resulting sequences were compared to build a consensus that was employed in subsequent analyses.

In order to obtain the sequence of the In endosymbiont a first PCR product (forward primer 27f and reverse primer for *Rickettsia*-like organisms 16S_Rick_R697 5'-GTGTTCCTCCTAATATCTAAG-3') was sequenced with the internal primer 16S_Midich_R550 5'-CCCAATAATTCCGAGTAAC-3'. To obtain a longer portion of the gene, a forward primer matching the obtained partial sequence was developed and employed in a new PCR amplification (forward primer 16S_CyrIn_F140 5'-TAGTACGAAATAACTATTG GAAAC-3' and reverse primer 16S alfa R1517 5'-TGATCCAGCCGCAGGTTC-3'[35]). Two semi-nested PCRs were then performed (first one: forward 16S_CyrIn_F140 and reverse R1390_Uni 5'-GACGGGCGGTGTGTRCAA-3' modified from Amann and coworkers[25]; second one: forward

16S_CLMid_F428 5'-GTAAAGCTCTTTCAGTGGG-3' and reverse 16S alfa R1517) on the amplicon. These products were sequenced with primers 16S_CLMid_F428 and 16S F785 ND 5'-GGATTAGATACCCTGGTA-3'[35] respectively. The three partially overlapping electropherograms were compared and assembled.

Preliminary comparisons were performed with the NCBI BLASTN software[36]. All sequences were submitted to the EMBL database, and are available under the accession numbers HE978247-9 (endosymbiont of PAR; 16S rRNA gene, three clones), HE978250 (endosymbiont of In; 16S rRNA gene) and HE978251 (PAR; 18S rRNA gene).

Phylogenetic analyses. 23 representative sequences of the "*Midichloria* clade", 30 of other *Rickettsiales* taxa and 16 of non-*Rickettsiales* alphaproteobacteria (as outgroup) were aligned with the "Fast Aligner" provided by the ARB software package[37]. The alignment was then manually edited in order to optimize base-paring in the predicted stem regions of the 16S rRNA gene.

Phylogenetic inferences were performed with four different methods: Neighbor Joining (NJ), Maximum Parsimony (MP), Maximum Likelihood (ML) and Bayesian Inference (BI). The software Phylip NEIGHBOR[38], Phylip DNAPARS[38], PHYML[39] and MrBayes[40] were employed respectively. Bootstrap analyses (1,000 pseudoreplicates) were performed for MP and ML methods. Three different runs, with three cold and one heated chains each, ran for 1,000,000 generations in BI analysis. Sequence lengths were reduced to that of the shortest one. Modified character matrices were produced and analyzed with each method in order to test the robustness of the results: gaps were coded as a fifth character and all columns were kept (matrix a); only those columns where the most conserved base was present in at least 2% (matrix b), 20% (matrix c) and 40% (matrix d) of sequences were kept; only columns without gaps were kept (matrix e). The evolutionary models that fit best the data were chosen according to the Akaike Information Criterion (AIC) calculated by jModelTest[39,41]. The Likelihood Mapping function of TREE-PUZZLE[42] was used for calculating base frequencies, checking the suitability of data in the character matrices and generating the distance matrices for NEIGHBOR. TREE-PUZZLE was also employed for performing the SH test[43] on obtained topologies.

All similarity values were calculated on the character matrix (a).

Fluorescence *In Situ* Hybridization (FISH). In order to verify the presence of bacterial symbionts inside the ciliate cells preliminary FISH experiments were performed using probes EUB338 5'-GCTGCCTCCCGTAGGAGT-3'[44] and ALF1b 5'-CGTTCGYT CTGAGCCAG-3'[45], targeting, respectively, most of the organisms belonging to *Eubacteria* and to the class *Alphaproteobacteria*. On the basis of the obtained 16S rRNA gene sequences, the specific probes Deflu_197 5'-TCTTATAGCGACTTTCGTCT-3' and CyrIn_142 5'-CGTTTCCAATAGTTATTTCGTAC-3' were designed, synthesized and labeled with Cy3 by Eurofins MWG Operon (Ebersberg, Germany). The specificity of these new probes was tested *in silico* both on the SILVA 104[46] and on the RDP[47] databases. The newly designed probes were then tested in FISH experiments on the ciliate cells at different stringency levels, i.e. using different formamide concentration in the hybridization buffer (0%, 20%, 30%, 40%, v/v). The optimum stringency condition for probe Deflu_197 was assessed at 40% (v/v) of formamide concentration, while the probe CyrIn_142 was employed without formamide in the hybridization buffer. Concerning previously designed probes, experimental conditions recommended by the authors or information available at probeBase[48] were used. FISH were then performed following Manz and coworkers[45]. Negative controls (known bacteria not targeted by the employed probes) were always included; when available, positive controls (known bacteria targeted by the employed probes) were also included. Sequences of the newly designed probes were deposited at probeBase.

Results

***P. nephridiatum* 18S rRNA gene sequence.** The 1695 bp long 18S rRNA gene sequence of the PAR population is identical to those of strains WS97-1 and BB3-9 (accession numbers: AF100317 and AF100316 respectively), the only two currently available for *P. nephridiatum*, and thus confirms the morphological identification.

Preliminary FISH experiments. The results of preliminary FISH experiments clearly showed the presence of alphaproteobacterial symbionts inside the ciliate cells of the PAR population of *P. nephridiatum*. The same result was obtained for cells of the In population of *E. aediculatus*, where the presence of already characterized betaproteobacterial[29] and gammaproteobacterial[30] symbionts was also shown.

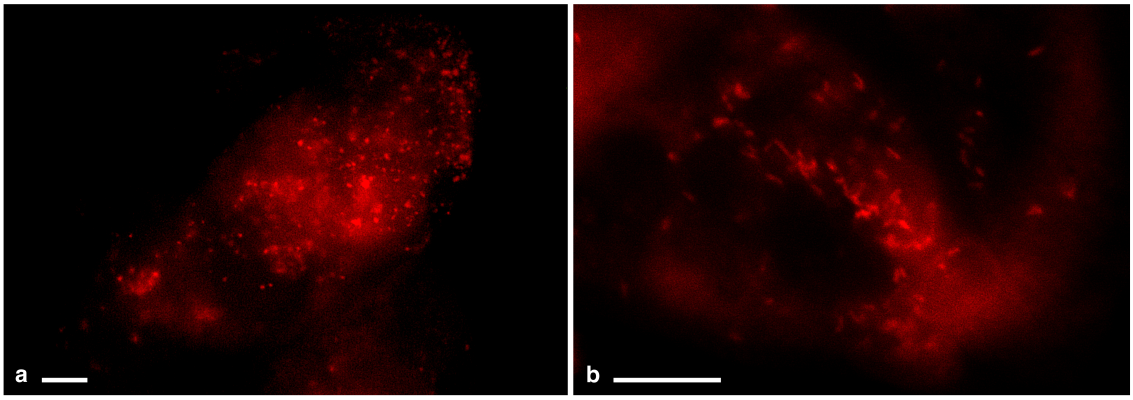


Figure 4.1 FISH results using sequence-specific oligonucleotide probes labeled with Cy3. Signals from one representative cell of the *P. nephridiatum* PAR population targeted by Deflu_197 (a) and one representative cell of the *E. aediculatus* In population targeted by CyrIn_142 (b). The bars correspond to 10 µm.

Characterization of the *P. nephridiatum* endosymbiont. An almost-complete 16S rRNA gene sequence was obtained from each of the sequenced clones (1446 bp each). The consensus sequence shows a similarity value of 83.0% with the 16S rRNA gene sequence of "*Candidatus* Midichloria mitochondrii". The values are 81.2% and 82.2% with the sequences of "*Candidatus* Anadelfobacter veles" and "*Candidatus* Cyrtobacter comes", respectively. The newly designed probe Deflu_197 showed a high *in silico* specificity, matching no other sequences neither on the SILVA 104 nor the RDP databases, even if a single base difference is allowed. FISH experiments performed with this probe always gave clear and positive signals in the cytoplasm of ciliate cells belonging to the PAR population (Figure 4.1a), thus confirming the validity of the obtained sequence.

Characterization of the *E. aediculatus* endosymbiont. The almost-complete 16S rRNA gene sequence obtained through PCR and direct sequencing is 1447 bp long, and it shows the highest similarity value (98.3%) with the sequence of "*Candidatus* Cyrtobacter comes". Similarity values with other bacteria of the "*Midichloria* clade" are: 89.1% with "*Candidatus* Midichloria mitochondrii", 87.3% with "*Candidatus* Anadelfobacter veles", 82.3% with the novel endosymbiont of *P. nephridiatum*. The probe CyrIn_142 doesn't match any other sequence available in the databases, and matches only three uncultured proteobacteria if a single base difference is allowed on RDP. In particular, it shows two central mismatches with the sequence of "*Candidatus* Cyrtobacter comes". On the contrary, the already developed probe Cyrt_1438[6] matches both "*Candidatus* Cyrtobacter comes" and the *E. aediculatus* In symbiont. FISH

experiments on In cells performed with probe CyrIn_142 gave positive signals, highlighting the bacteria in the cytoplasm (Figure 4.1b).

Phylogenetic analyses. Model choice and data on model parameters, base frequencies and composition of the character matrices are shown in Supplementary Table S4.1. According to the SH test, NJ topologies were always considered worse explanation of the data ($p<0.05$) than MP, ML and BI topologies, and were thus discarded.

All trees recovered with strong statistical support the monophyly of families *Rickettsiaceae*, *Anaplasmataceae* and *Holosporaceae* (*sensu* Boscaro and coworkers[49]), as well as that of the "*Midichloria* clade". The "*Midichloria* clade" and *Anaplasmataceae* are sister groups, forming a taxon related to *Rickettsiaceae*.

The relationships within the "*Midichloria* clade" are much less reliable. The topology shown in Figure 4.2 is obtained with only minor differences when the character matrices (a), (b) and (c) are employed, but changes remarkably (and in different ways according to the inference method) when the character matrices (d) and (e) are employed instead. Only the most recent nodes are recovered by all trees. In particular, the sequence of the In endosymbiont is associated to that of "*Candidatus* Cyrtobacter comes"; the sequence of the PAR endosymbiont is associated to that of an uncultured bacterium from an American lake (accession number: FJ437943).

Discussion

The establishment of stable and intimate relationships between prokaryotes and ciliates, commonly labeled as symbiotic phenomena, provides an unexpected source of biodiversity. Bacteria belonging to at least three classes of the phylum *Proteobacteria* (*Alpha-*[6,35,49-55], *Beta-*[29,56] and *Gammaproteobacteria*[30,57-59]) and to the phyla *Firmicutes*[60] and *Verrucomicrobia*[61] have now been characterized with the modern molecular approach. Bacterial-ciliates symbiosis clearly originated on many independent occasions.

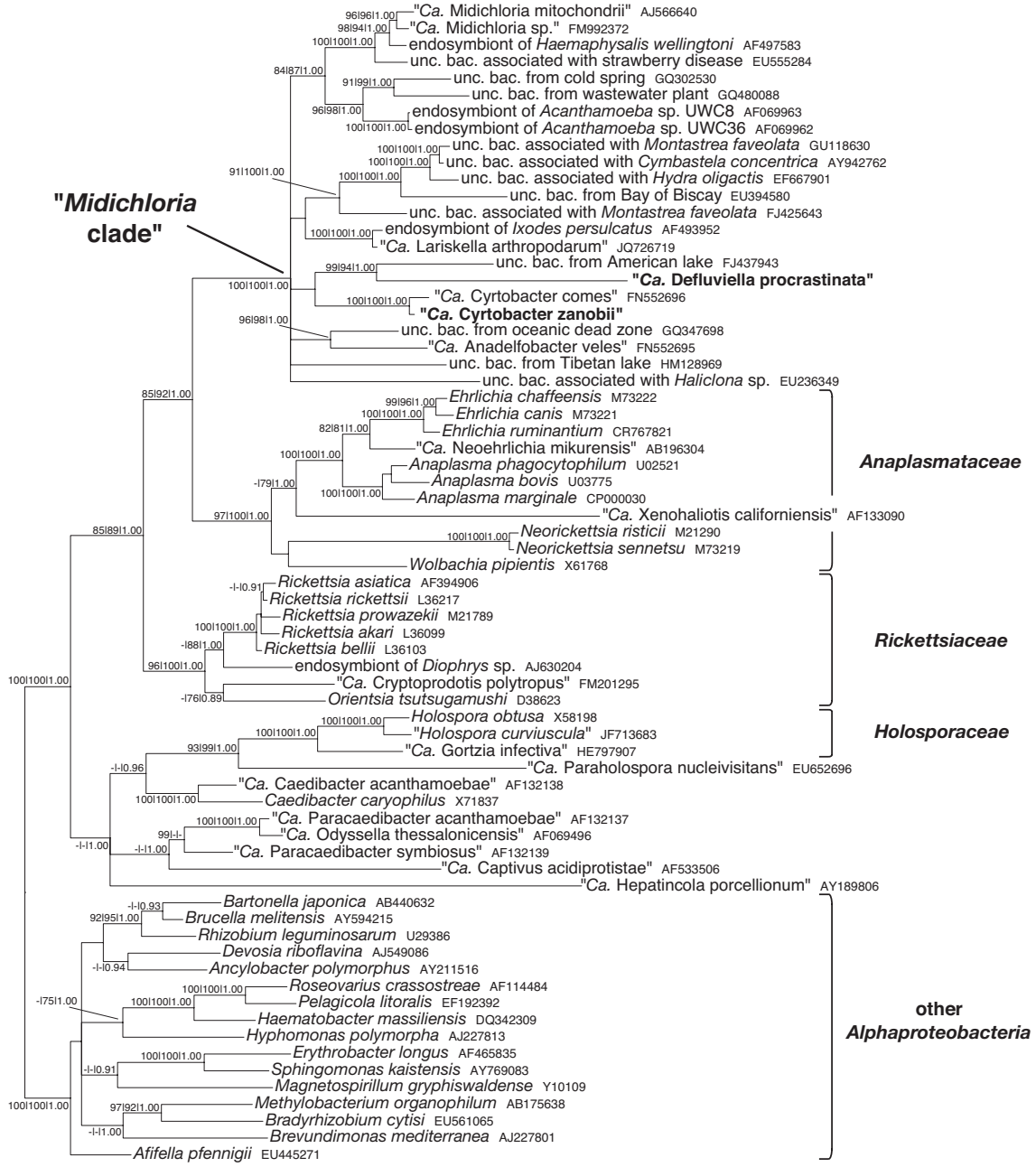


Figure 4.2 Bayesian phylogenetic tree of the order *Rickettsiales* inferred from 16S rRNA gene sequences. The character matrix a (see text) included 69 sequences and 1360 sites; 7.05% of the characters were either gaps (99.86%) or wildcards (0.16%). The GTR+I+G (4 discrete categories) model was employed. 94.5% of 100,000 randomly chosen quartets had well-defined topologies. Numbers associated to each node represent bootstrap values for MP and ML and posterior probability, respectively. The sequences characterized in this study are in bold characters; for the others, the accession number is shown. The bar corresponds to an estimated sequence divergence of 10%. "*Ca.*" stands for "*Candidatus*"; "unc. bac." stands for "uncultured bacterium".

The data here reported on two novel ciliate symbionts belonging to the "*Midichloria* clade" further strengthen this view. The symbiont of the PAR population is the first organism of this clade which has been found in the genus *Paramecium*. The considerably low 16S rRNA gene sequence similarity it shares with all other known bacteria allows the establishment of a new genus and species, for which we propose the name "*Candidatus* Defluviella procrastinata". The symbiont of the In population is closely related to that of the *Euplotes harpa* strain HS11/7 ("*Candidatus* Cyrtobacter comes"[6]), but the similarity value between their 16S rRNA gene sequences is below the suggested species threshold of 98.7-99.0%[62]. Moreover, the probe CyrIn_142 here developed is specific for the novel endosymbiont. Thus, we propose the establishment of a second species in the same genus, namely "*Candidatus* Cyrtobacter zanobii". The diagnostic descriptions of both novel symbionts are provided at the end of this section.

Different symbionts of ciliates are scattered in all the families of the order *Rickettsiales*, with the curious exception of *Anaplasmataceae*. The case of the "*Midichloria* clade", which is a family-like taxon still lacking formal standing in nomenclature, is explicative: of the four symbionts of ciliates belonging to this clade, none but two are closely related. Furthermore, there is no reason to assume that even the two species of the "*Candidatus* Cyrtobacter" genus descend from a common ancestor that was itself a symbiont of *Euplotes*. In fact, there are *E. aediculatus* and *E. harpa* strains that do not contain any alphaproteobacterium (data not shown); moreover, another bacterium of the "*Midichloria* clade" has been indeed found in a different *E. harpa* strain ("*Candidatus* Anadelfobacter veles" from strain BOD18[6]), but it is not related to the genus "*Candidatus* Cyrtobacter". Although more data are needed to draw definitive conclusions, the picture is compatible with a series of independent gaining of symbionts, as in the more thoroughly studied case of the betaproteobacterial symbiont of *Euplotes*[29,63].

All *Rickettsiales* are, to our knowledge, obligate intracellular bacteria. The findings of so many *Rickettsiales* organisms in phylogenetically distant eukaryotes such as arthropods and ciliates, with the evolutionary histories of symbionts and hosts so utterly different, strongly suggest multiple occurrences of horizontal transmission. Moreover, the environments inhabited by ciliates, as well as those of some cnidarians and sponges harboring bacteria of the "*Midichloria* clade", suggest that the transmission of *Rickettsiales* bacteria, perhaps even some pathogenic ones, may be more reliant on water as a vector than we currently suspect.

Our work highlights the necessity to continue the characterization of new symbiotic systems, where a large part of the still unknown prokaryotic biodiversity may dwell. More efforts on understanding the ecological role of the symbionts is also needed; nevertheless, it is likely that these roles are very diverse, as proved by more investigated examples[64]. The possibility that some of these *Rickettsiales* representatives are capable of horizontal transmission (as it is largely proven for *Holospora* in *Paramecium*[65]) is especially worth testing. As a further note, our phylogenetic analyses highlighted a peculiar difficulty in untangling the relationships inside the "*Midichloria* clade". In order to better address this issue, there is a need for more data, coming both from novel organisms and from different markers.

Description of "*Candidatus* Defluviella procrastinata". *Defluviella procrastinata* (Def.lu.vi.el'la pro.cras.ti.na'ta; L. v. *defluo*, to flow, because it was sampled in wastewater flow, L. v. *procrastino*, to delay, because the characterization of this symbiont was completed several months after its first finding; N.L. fem. n. *Defluviella*; N.L. adj. *procrastinata*).

Short rod-shaped bacterium. Cytoplasmic endosymbiont of the ciliate *Paramecium nephridiatum* (Ciliophora, Oligohymenophorea). Belonging to the "*Midichloria* clade" in the order *Rickettsiales* (*Alphaproteobacteria*). Basis of assignment: 16S rRNA gene sequence (consensus from three clones, accession numbers: HE978247-9) and positive match with the specific FISH oligonucleotide probe Deflu_197 (5'-TCTTATAGCGACTTTCGTCT-3'). Identified in the *P. nephridiatum* population PAR, isolated from a wastewater treatment plant in Sabaudia (Latina, Italy). Uncultured thus far.

Description of "*Candidatus* Cyrtobacter zanobii". *Cyrtobacter zanobii* (Cyr.to.bac'ter za.no'bi.i; Gr. adj. *kyrtòs*, humped, N.L. masc. n. *bacter*, rod, from the shape of the first described species of the genus; N.L. masc. n. *Cyrtobacter*; N.L. adj. *zanobii*, of Zanobi, because the first sequence was obtained on the celebration day of Saint Zanobi, bishop of Florence).

Rod-shaped bacterium. Cytoplasmic endosymbiont of the ciliate *Euplotes aediculatus* (Ciliophora, Spirotrichea). Closely related to "*Candidatus* Cyrtobacter comes" (endosymbiont of *Euplotes harpa*) and belonging to the "*Midichloria* clade". Basis of assignment: 16S rRNA gene sequence (accession number: HE978250) and positive match with the specific FISH oligonucleotide probe CyrIn_142 (5'-CGTTTCCAATAGTTATTTCGTAC-3'). Identified in the

E. aediculatus population In, isolated from a freshwater sample collected near the Yamuna River (New Delhi, India). Uncultured thus far.

References

1. Dumler JS, Walker DH. (2005) *Rickettsiales* Gieszczykiewicz 1939, 25^{AL} emend. Dumler, Barbet, Bekker, Dasch, Palmer, Ray, Rikihisa and Rurangirwa 2001. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds), Bergey's Manual of Systematic Bacteriology, 2nd edn, vol II, part C. Springer-Verlag, New York, pp 96-160
2. Andersson SGE, Zomorodipour A, Andersson JO, Sicheritz-Pontén T, Alsmark UCM, *et al.* (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature. 396:133-143
3. Fitzpatrick DA, Creevey CJ, McInerney JO. (2006) Genome phylogenies indicate a meaningful alpha-proteobacterial phylogeny and support a grouping of the mitochondria with the *Rickettsiales*. Mol Biol Evol. 23: 74-85
4. Lo N, Beninati T, Sassera D, Bouman EAP, Santagati S, *et al.* (2006) Widespread distribution and high prevalence of an alpha-proteobacterial symbiont in the tick *Ixodes ricinus*. Environ Microbiol. 8: 1280-1287
5. Matsuura Y, Kikuchi Y, Meng XY, Koga R, Fukatsu T. (2012) Novel clade of alphaproteobacterial endosymbionts associated with stinkbugs and other arthropods. Appl Environ Microbiol. 78: 4149-4156
6. Vannini C, Ferrantini F, Schleifer KH, Ludwig W, Verni F, Petroni G. (2010) "*Candidatus* Anadelfobacter veles" and "*Candidatus* Cyrtobacter comes", two new rickettsiales species hosted by the protist ciliate *Euplotes harpa* (Ciliophora, Spirotrichea). Appl Environ Microbiol. 76: 4047-4054
7. Sassera D, Beninati T, Bandi C, Bouman EAP, Sacchi L, *et al.* (2006) "*Candidatus* Midichloria mitochondrii", an endosymbiont of the tick *Ixodes ricinus* with a unique intramitochondrial lifestyle. Int J Syst Evol Microbiol. 56: 2535-2540
8. Beninati T, Lo N, Sacchi L, Genchi C, Noda H, Bandi C. (2004) A novel alpha-proteobacterium resides in the mitochondria of ovarian cells of the tick *Ixodes ricinus*. Appl Environ Microbiol. 70: 2596-2602
9. Epis S, Sassera D, Beninati T, Lo N, Beati L, *et al.* (2008) *Midichloria mitochondrii* is widespread in hard ticks (Ixodidae) and resides in the mitochondria of phylogenetically diverse species. Parasitology. 135: 485-494
10. Parola P, Cornet JP, Sanogo YO, Miller RS, Van Thien H, *et al.* (2003) Detection of *Ehrlichia* spp., *Anaplasma* spp., *Rickettsia* spp., and other eubacteria in ticks from the Thai-Myanmar border and Vietnam. J Clin Microbiol. 41: 1600-1608
11. Fritsche TR, Horn M, Seyedirashti S, Gautom RK, Schleifer KH, Wagner M. (1999) In situ detection of novel bacterial endosymbionts of *Acanthamoeba* spp. phylogenetically related to members of the order *Rickettsiales*. Appl Environ Microbiol. 65: 206-212
12. Longford SR, Tujula NA, Crocetti GR, Holmes AJ, Holmström C, *et al.* (2007) Comparisons of diversity of bacterial communities associated with three sessile marine eukaryotes. Aquat Microb Ecol. 48: 217-229
13. Sipkema D, Holmes B, Nichols SA, Blanch HW. (2009) Biological characterization of *Haliclona* (?*gellius*) sp.: sponge and associated microorganisms. Microb Ecol. 58: 903-920
14. Fraune S, Bosch TCG. (2007) Long-term maintenance of species-specific bacterial microbiota in the basal metazoan *Hydra*. Proc Natl Acad Sci USA. 104: 13146-13151

15. Sunagawa S, Woodley CM, Medina M. (2010) Threatened corals provide underexplored microbial habitats. PLoS One 5: e9554

16. Erickson DL, Anderson NE, Cromar LM, Jolley A. (2009) Bacterial communities associated with flea vectors of plague. J Med Entomol. 46: 1532-1536

17. Hornok S, Földvári G, Elek V, Naranjo V, Farkas R, de la Fuente J. (2008) Molecular identification of *Anaplasma marginale* and rickettsial endosymbionts in blood-sucking flies (Diptera: Tabanidae, Muscidae) and hard ticks (Acari: Ixodidae). Vet Parasitol. 154: 354-359

18. Richard S, Seng P, Parola P, Raoult D, Davoust B, Brouqui P. (2009) Detection of a new bacterium related to "*Candidatus* Midichloria mitochondrii" in bed bugs. Clin Microbiol Infect. 15: 84-85

19. Mediannikov Olu, Ivanov LI, Nishikawa M, Saito R, Sidel'nikov IuN, *et al.* (2004) [Microorganism "Montezuma" of the order *Rickettsiales*: the potential causative agent of tick-borne disease in the Far East of Russia]. Zh Mikrobiol Epidemiol Immunobiol. 1: 7-13 (In Russian with English summary)

20. Lloyd SJ, LaPatra SE, Snekvik KR, St-Hillaire S, Cain KD, Call DR. (2008) Strawberry disease lesions in rainbow trout from southern Idaho are associated with DNA from a *Rickettsia*-like organism. Dis Aquat Organ. 20: 111-118

21. Sassera D, Lo N, Epis S, D'Auria G, Montagna M, *et al.* (2011) Phylogenomic evidence for the presence of a flagellum and cbb_3 oxidase in the free-living mitochondrial ancestor. Mol Biol Evol. 28: 3285-3296

22. Mariconti M, Epis S, Sacchi L, Biggiogera M, Sassera D, *et al.* (2012) A study on the presence of flagella in the order *Rickettsiales*: the case of "*Candidatus* Midichloria mitochondrii". Microbiology. 158: 1677-1683

23. Sacchi L, Bigliardi E, Corona S, Beninati T, Lo N, Franceschi A. (2004) A symbiont of the tick *Ixodes ricinus* invades and consumes mitochondria in a mode similar to that of the parasitic bacterium *Bdellovibrio bacteriovorus*. Tissue Cell 36: 43-53

24. Werren JH, Baldo L, Clark ME. (2008) *Wolbachia*: master manipulators of invertebrate biology. Nature Rev Microbiol. 6: 741-751

25. Amann RI, Ludwig WG, Schleifer KH. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev. 59: 143-169

26. Murray RGE, Schleifer KH. (1994) A proposal for recording the properties of putative taxa of procaryotes. Int J Syst Bacteriol. 44: 174-176

27. Murray RGE, Stackebrandt E. (1995) Implementation of the provisional status *Candidatus* for incompletely described prokaryotes. Int J Syst Bacteriol. 45: 186-187

28. Fokin SI. (2010/11) *Paramecium* genus: biodiversity, some morphological features and the key to the main morphospecies discrimination. Protistology. 6: 227-235

29. Vannini C, Ferrantini F, Ristori A, Verni F, Petroni G. (2012) Betaproteobacterial symbionts of the ciliate *Euplotes*: origin and tangled evolutionary path of an obligate microbial association. Environ Microbiol. 14: 2553-2563

30. Boscaro V, Vannini C, Fokin SI, Verni F, Petroni G (2012) Characterization of "*Candidatus* Nebulobacter yamunensis" from the cytoplasm of *Euplotes aediculatus* (Ciliophora, Spirotrichea) and emended description of the family *Francisellaceae*. Syst Appl Microbiol. 35: 432-440

31. Medlin L, Elwood HJ, Stickel S, Sogin ML. (1988) The characterization of enzymatically amplified 16S-like rRNA-coding regions. Gene. 71: 491-499

32. Petroni G, Dini F, Verni F, Rosati G. (2002) A molecular approach to the tangled intrageneric relationships underlying phylogeny in *Euplotes* (Ciliophora, Spirotrichea). Mol Phylogenet Evol. 22: 118-130
33. Rosati G, Modeo L, Melai M, Petroni G, Verni F. (2004) A multidisciplinary approach to describe protists: a morphological, ultrastructural, and molecular study on *Peritromus kahli* Villeneuve-Brachon, 1940 (Ciliophora, Heterotrichea). J Eukaryot Microbiol. 51: 49-59
34. Lane DJ. (1991) 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M (eds), Nucleic Acid Techniques in Bacterial Systematics, Wiley, New York, pp 115-147, 1991
35. Vannini C, Rosati G, Verni F, Petroni G. (2004) Identification of the bacterial endosymbionts of the marine ciliate *Euplotes magnicirratus* (Ciliophora, Hypotrichia) and proposal of "*Candidatus* Devosia euplotis". Int J Syst Evolution Microbiol. 54: 1151-1156
36. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, *et al.* (1997) Gapped BLAST and PSI BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389-3402
37. Ludwig W, Strunk O, Westram R, Richter L, Meier H, *et al.* (2004) ARB: a software environment for sequence data. Nucleic Acids Res. 32: 1363-1371
38. Felsenstein J. (1989) PHYLIP-Phylogeny Inference Package (version 3.2). Cladistics. 5: 164-166
39. Guindon S, Gascuel O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 52: 696-704
40. Huelsenbeck JP, Ronquist F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754-755
41. Posada D. (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol. 25: 1253-1256
42. Schmidt HA, Strimmer K, Vingron M, von Haeseler A. (2002) TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics. 18: 502-504
43. Shimodaira H, Hasegawa M. (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol. 16: 1114-1116
44. Amann R, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA. (1990) Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. Appl Environ Microbiol. 56: 1919-1925
45. Manz W, Amann RI, Ludwig W, Wagner M, Schleifer KH. (1992) Phylogenetic oligodeoxynucleotide probes for the major subclasses of Proteobacteria: problems and solutions. Syst Appl Microbiol. 15: 593-600
46. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, *et al.* (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 35: 7188-7196
47. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, *et al.* (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res. 37: D141-D145
48. Loy A, Maixner F, Wagner M, Horn M. (2007) probeBase – an online resource for rRNA-targeted oligonucleotide probes: new features 2007. Nucleic Acids Res. 35: D800-D804
49. Boscaro V, Fokin SI, Schrallhammer M, Schweikert M, Petroni G. (2013) Revised systematics of *Holospora*-like bacteria and characterization of "*Candidatus* Gortzia infectiva", a novel macronuclear symbiont of *Paramecium jenningsi*. Microb Ecol. 65: 255-267
50. Amann RI, Springer N, Ludwig W, Görtz HD, Schleifer KH. (1991) Identification *in situ* and phylogeny of uncultured bacterial endosymbionts. Nature. 351: 161-164

51. Eschbach E, Pfannkuchen M, Schweikert M, Drutschmann D, Brümmer F, *et al.* (2009) "*Candidatus* Paraholospora nucleivisitans", an intracellular bacterium in *Paramecium sexaurelia* shuttles between the cytoplasm and the nucleus of its host. Syst Appl Microbiol. 32: 490-500

52. Ferrantini F, Fokin SI, Modeo L, Andreoli I, Dini F, *et al.* (2009) "*Candidatus* Cryptoprodotis polytropus", a novel *Rickettsia*-like organism in the ciliated protist *Pseudomicrothorax dubius* (Ciliophora, Nassophorea). J Eukaryot Microbiol. 56: 119-129

53. Springer N, Ludwig W, Amann R, Schmidt HJ, Görtz HD, Schleifer KH. (1993) Occurrence of fragmented 16S ribosomal-RNA in an obligate bacterial endosymbiont of *Paramecium caudatum*. Proc Natl Acad Sci USA. 90: 9892-9895

54. Sun HY, Noe J, Barber J, Coyne RS, Cassidy-Hanley D, *et al.* (2009) Endosymbiotic bacteria in the parasitic ciliate *Ichthyophthirius multifiliis*. Appl Environ Microbiol. 75: 7445-7452

55. Vannini C, Petroni G, Verni F, Rosati G. (2005) A bacterium belonging to the *Rickettsiaceae* family inhabits the cytoplasm of the marine ciliate *Dioprhys appendiculata* (Ciliophora, Hypotrichia). Microbial Ecol. 49: 434-442

56. Springer N, Amann R, Ludwig W, Schleifer KH, Schmidt H. (1996) *Polynucleobacter necessarius*, an obligate bacterial endosymbiont of the hypotrichous ciliate *Euplotes aediculatus*, is a member of the ß-subclass of *Proteobacteria*. FEMS Microbiol Lett. 135: 333-336

57. Beier CL, Horn M., Michel R., Schweikert M, Görtz HD, Wagner M. (2002) The genus *Caedibacter* comprises endosymbionts of *Paramecium* spp. related to the *Rickettsiales* (*Alphaproteobacteria*) and to *Francisella tularensis* (*Gammaproteobacteria*). Appl Environmen Microbiol. 68: 6043-6050

58. Rinke C, Schmitz-Esser S, Loy A, Horn M, Wagner M, Bright M. (2009) High genetic similarity between two geographically distinct strains of the sulfur-oxidising symbiont "*Candidatus* Thiobios zoothamnicoli". FEMS Microbiol Ecol. 67: 229-241

59. Schrallhammer M, Schweikert M, Vallesi A, Verni F, Petroni G. (2011) Detection of a novel subspecies of *Francisella noatunensis* as endosymbiont of the ciliate *Euplotes raikovi*. Microb Ecol. 61: 455-464

60. Shinzato N, Watanabe I, Meng XY, Sekiguchi Y, Tamaki H, *et al.* (2007) Phylogenetic analysis and fluorescence *in situ* hybridization detection of archaeal and bacterial endosymbionts in the anaerobic ciliate *Trimyema compressum*. Microbial Ecol. 54: 627-636

61. Petroni G, Spring S, Schleifer KH, Verni F, Rosati G. (2000) Defensive extrusive ectosymbionts of *Euplotidium* (Ciliophora) that contain microtubule-like structures are bacteria related to *Verrucomicrobia*. Proc Natl Acad Sci USA 97: 1813-1817

62. Stackebrandt E, Ebers E. (2006) Taxonomic parameters revisited: tarnished gold standards. Microbiol Today. nov06, 153-154

63. Vannini C, Pöckl M, Petroni G, Wu QL, Lang E, *et al.* (2007) Endosymbiosis *in statu nascendi*: close phylogenetic relationship between obligately endosymbiotic and obligately free-living *Polynucleobacter* strains (*Betaproteobacteria*). Environ Microbiol. 9: 347-359

64. Görtz HD. (2006) Symbiotic associations between ciliates and prokaryotes. In: Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E (eds), The Prokaryotes. Third Edition, vol. 1, Springer-Verlag, New York, NY, pp 364-402

65. Fujishima M. (2009) Infection and maintenance of *Holospora* species in *Paramecium caudatum*. In: Fujishima M (ed), Endosymbionts in *Paramecium*. Microbiology Monograph, Münster, pp 201-225

Chapter 4 Supplementary Information.

Supplementary Table S4.1 Parameters of character matrices and molecular substitution models calculated for phylogenetic analyses

	Matrix [a]*	Matrix [b]	Matrix [c]	Matrix [d]	Matrix [e]
Sequences	69	69	69	69	69
Columns	1360	1296	1274	1236	1178
Model selected	GTR+I+G	GTR+I+G	GTR+I+G	GTR+I+G	GTR+I+G
Well-defined quartets (%)	94.5	94.9	94.7	95.3	95.4
Wildcards	9	9	9	9	9
Gaps (number)	6609	2263	856	326	0
Gaps (%)	7.04	2.58	0.97	0.38	0.00
A content (ModelTest)	0.2776	0.2744	0.2771	0.2724	0.2830
A content (TreePuzzle)	0.2680	0.2680	0.2680	0.2680	0.2680
A content (MrBayes)	0.2782	0.2784	0.2780	0.2774	0.2824
C content (ModelTest)	0.1837	0.1826	0.1833	0.1824	0.1821
C content (TreePuzzle)	0.2090	0.2090	0.2090	0.2100	0.2120
C content (MrBayes)	0.1823	0.1815	0.1821	0.1814	0.1834
G content (ModelTest)	0.2540	0.2569	0.2580	0.2689	0.2537
G content (TreePuzzle)	0.2970	0.2970	0.2980	0.2990	0.2950
G content (MrBayes)	0.2474	0.2500	0.2508	0.2570	0.2500
T content (ModelTest)	0.2847	0.2831	0.2816	0.2763	0.2811
T content (TreePuzzle)	0.2250	0.2250	0.2250	0.2230	0.2250
T content (MrBayes)	0.2921	0.2900	0.2891	0.2842	0.2842
a [AC] (ModelTest)	1.0515	1.0639	1.0541	1.1628	1.1092
a [AC] (MrBayes)	1.0900	1.0899	1.0838	1.1691	1.1270
b [AG] (ModelTest)	5.1107	5.1036	5.0481	5.1592	5.5725
b [AG] (MrBayes)	5.2468	5.2269	5.1892	5.3804	5.9325
c [AT] (ModelTest)	1.7744	1.7903	1.7702	1.8566	1.6486
c [AT] (MrBayes)	1.6673	1.6789	1.6552	1.7362	1.6717
d [CG] (ModelTest)	1.0820	1.0787	1.0675	1.0895	1.0405
d [CG] (MrBayes)	1.0612	1.0538	1.0455	1.1172	1.0701
e [CT] (ModelTest)	6.1596	6.2420	6.2152	6.9639	6.9385
e [CT] (MrBayes)	6.1679	6.2296	6.1800	7.0919	7.1385
f [GT] (ModelTest)	1.0000	1.0000	1.0000	1.0000	1.0000
f [GT] (MrBayes)	1.0000	1.0000	1.0000	1.0000	1.0000
Invariant sites (ModelTest)	0.3470	0.3490	0.3530	0.3570	0.3620
Invariant sites (MrBayes)	0.3351	0.3376	0.3414	0.3523	0.3522
Alpha (ModelTest)	0.6590	0.6600	0.6610	0.6680	0.6510
Alpha (MrBayes)	0.6809	0.6827	0.6831	0.7192	0.6866

* The tree shown in the chapter was calculated on this character matrix

Chapter 5.

The genus *Lyticum*

Original publication:

Rediscovering the genus *Lyticum*, multiflagellated symbionts of the order *Rickettsiales*

Vittorio Boscaro[*], Martina Schrallhammer[*], Konstantin A Benken, Sascha Krenek, Franziska Szokoli, Thomas U Berendonk, Michael Schweikert, Franco Verni, Elena V Sabaneyeva, Giulio Petroni

Scientific Reports 3: 3305. November 2013.

Abstract

Among the bacterial symbionts harbored by the model organism *Paramecium*, many still lack a recent investigation that includes a molecular characterization. The genus *Lyticum* consists of two species of large-sized bacteria displaying numerous flagella, despite their inability to move inside their hosts' cytoplasm. We present a multidisciplinary redescription of both species, using the deposited type strains as well as newly collected material. On the basis of 16S rRNA gene sequences, we assigned *Lyticum* to the order *Rickettsiales*, that is intensely studied because of its pathogenic representatives and its position as the extant group most closely related to the mitochondrial ancestor. We provide conclusive proofs that at least some *Rickettsiales* possess actual flagella, a feature that has been recently predicted from genomic data but never confirmed. We give support to the hypothesis that the mitochondrial ancestor could have been flagellated, and provide the basis for further studies on these ciliate endosymbionts.

[*] These authors contributed equally to the work

Introduction

In the mid-twentieth century, T. Sonneborn revealed two features of *Paramecium* that had a lasting impact beyond the field of protozoology. First, he identified many morphologically identical strains of "*Paramecium aurelia*" that were not sexually compatible[1]. This observation provided one of the first and most extreme cases of a "sibling species" complex – fifteen different species have been described within the *P. aurelia* species complex until now[2,3]. Another important discovery was that of "cytoplasmic particles" of various kinds, found many times in several *Paramecium* species and often able to confer non-genetically inherited traits[4]. Years later, all these particles were identified as bacterial endosymbionts[5].

Many of these symbionts have peculiar biological properties, and sometimes remarkably distinctive morphologies. Examples include the infectious *Holospora* with its specialized nucleus-invading form[6], and the "killer" symbionts, that confer to infected paramecia the ability to kill uninfected "sensitive" strains present in the same culture medium[5,7]. Much interest was directed to the unusual bacteria belonging to genus *Caedibacter* (formerly "kappa particles") and their complex cytoplasmic inclusions, the "R-bodies"[8-13]. Other equally intriguing killer symbionts were characterized, and among them were those belonging to the genus *Lyticum*[4,5].

Lyticum bacteria appear as large rods (2.0-10.0 µm long) harbored in the hundreds in the cytoplasm of three different species of the *P. aurelia* complex[4,5,14,15]. They are non-motile, despite being covered by numerous flagella[16]. The two species were formally described as *Lyticum flagellatum* (formerly "lambda particle", type species of the genus) and *Lyticum sinuosum* ("sigma particle")[17]. They differ in shape (respectively, straight *vs.* curved rods) and host specificity (respectively, *Paramecium tetraurelia* or *Paramecium octaurelia vs. Paramecium biaurelia*)[5,18].

The original descriptions of *Lyticum* and many other symbionts detected in the last century left many questions unanswered. One of the most important issues from an evolutionary point of view concerns the phylogenetic relationships of these bacteria.

The study of prokaryotic symbionts of protozoa is currently attracting a renewed interest, and is performed with the aid of molecular tools complementing ultrastructural methods like electron microscopy[19-25]. In recent years, the focus has shifted to the remarkable biodiversity of

these organisms and their close relationships with human pathogens, like *Rickettsia*[26-29] and *Francisella*[30,31].

In this work, we have characterized the symbionts of *P. octaurelia* strain 299 and *P. biaurelia* strain 114 following a multidisciplinary approach. They represent the type strains of *L. flagellatum* and *L. sinuosum*, respectively. We also reported a recently sampled environmental isolate (*P. biaurelia* USBL-36I1) infected by *L. sinuosum*, for which only one host strain was known so far. Morphology, ultrastructure and killer capabilities of the two bacterial species were investigated and molecular tools for their identification developed and tested. Moreover, we established their phylogenetic relationships, placing them inside the order *Rickettsiales* (*Alphaproteobacteria*) together with other obligate intracellular symbionts. This discovery not only clarifies the *Lyticum* affiliation, but also provides evidence supporting the hypothesis that *Rickettsiales*, the extant bacteria most closely related to the mitochondrial ancestor[32-35], were ancestrally flagellated[36]. This finding provides a relevant contribution in inferring the features of the free-living ancestor of both *Rickettsiales* and mitochondria, supporting the view that it was motile.

Materials and Methods

Hosts identification and culture. The *P. octaurelia* strain 299 and the *P. biaurelia* strain 114 were kindly provided by T. G. Doak and M. Lynch (Indiana University). The *P. biaurelia* strain USBL-36I1 was collected in 2011 from a small pond near Spencer (IN, USA, 39°17'45"N, 86°48'1"W). In order to confirm the identity of the host strains, morphological diagnostic features were checked[37] and the mitochondrial cytochrome *c* oxidase subunit 1 gene (*cox1*) was sequenced according to Barth and coworkers[38]; sequences are available at EMBL database with the accession numbers HF969031-3. The cultures were maintained at 19 °C on a 12:12 h light/dark cycle and fed with *Raoultella planticola* inoculated in modified Cerophyl medium[25] or, alternatively, with *Enterobacter aerogenes* inoculated in lettuce medium at room temperature. Strain 114 was obtained several times, but the symbionts were always lost shortly after the paramecia started to propagate. Thus, Transmission Electron Microscopy (TEM), fluorescence *in situ* hybridizations (FISH) and killer tests could not be performed on this strain.

Transmission electron microscopy. Ciliate cells were harvested by gentle centrifugation and fixed with 2.5% glutardialdehyde in 0.1M Cacodylate buffer (pH 7.4) for 1 hour at room temperature. After washing in buffer, cells were postfixed in 1% OsO_4 in 0.1M Cacodylate buffer (1 hour at room temperature). Three washing steps in this buffer were performed prior to dehydration in an acetone series and consecutive infiltration into Spurrs resin[39]. After ultrathin sectioning, sections were post-stained with 1% aqueous uranyl acetate and lead citrate[40]. Images were taken with a Zeiss EM 10 electron microscope at 60kV. Alternatively, the cells were fixed in a mixture containing 2.5% glutaraldehyde and 1.6% paraformaldehyde in 0.1M phosphate buffer (pH 7.4) for 2 hours at room temperature followed by a wash in the same buffer containing 12.5% sucrose and post-fixation in 1.6% OsO_4 (1 hour at 4 °C). The cells were dehydrated through a graded series of alcohol and acetone and embedded in Epoxy embedding medium (Fluka, BioChemika). Polymerization was carried out according to the manufacturer's protocol. Ultrathin sections were cut using a Reichert-Jung Ultracut E or Leica UC6, and stained with aqueous 1% uranyl acetate and 1% lead citrate. The samples were visualized using a Jeol JEM-1400 at 89 kV.

For negative staining of bacteria, several *Paramecium* cells were briefly washed in distilled water, squashed with a thin glass capillary in a drop of water, and a drop of the resulting suspension was placed on a Pioloform coated grid. Bacteria were allowed to precipitate for 2-3 min, then a drop of 1% uranyl acetate in distilled water was added for no longer than 1 min. The liquid was then absorbed with filter paper and the grid was air-dried.

16S rRNA gene sequencing. The almost complete 16S rRNA gene sequences were obtained through several PCR amplifications of overlapping regions and direct sequencing of the products (299 symbiont), or through cloning of PCR products, RFLP analyses and sequencing of 5 clones showing the most represented pattern to produce a consensus (for details of primers and PCR reactions, see Supplementary Text S5.1). The sequences are available at EMBL with the accession numbers HF969034-44.

FISH. Hybridizations were performed according to the protocol of Manz and coworkers[41] on individually collected *Paramecium* cells fixed with 2% paraformaldehyde (w/v). Preliminary FISH experiments were performed with the eubacterial probe EUB338[42] and the alphaproteobacterial probe ALF1b[41]. Oligonucleotide probes specific for the obtained 16S rRNA gene sequences were developed [LytiProb_433 5'-TATCTTCCCCACCAAAAGAAC-3', genus

Lyticum specific; Lflag_268 5'-GCTAAAGATCGAAGCCTTGGTAA-3', *L. flagellatum* specific; Lsinu_268 5'-GCTAAAGATCGTAGCCTTGGTAA-3', *L. sinuosum* specific]. These novel probes were tested with a wide range (0-50%) of formamide concentrations in the hybridization buffer. *Paramecium* strains containing different alphaproteobacterial symbionts were employed as negative controls. Probe specificities were checked also *in silico* with the ProbeMatch tool of the Ribosomal Database Project (RDP) website[43] and probe data were deposited at probeBase[44].

Phylogenetic analyses. Non-identical 16S rRNA gene sequences obtained were aligned with 42 homologous sequences of *Rickettsiales* bacteria and 16 of non-*Rickettsiales* alphaproteobacteria (as outgroup) using the ARB software package[45]. Sequence lengths were reduced to that of the shortest one, then multiple character matrices were produced according to Boscaro and colleagues[46]; unless otherwise stated, similarity values were calculated on the unmodified dataset. Maximum likelihood (ML) phylogenetic analyses were performed with PhyML[47], employing bootstrap analysis (1,000 pseudoreplicates) to evaluate the reliability of nodes. Bayesian inference (BI) analyses were performed with MrBayes[48], using three different runs with three heated and one cold chain each, iterating for 1,000,000 generations. The evolutionary model was selected according to the Akaike information criterion calculated by jModelTest[49].

Killer tests. 5 cells of the putative killer strains (299 or USBL-36I1) and 5 cells of putative sensitive *Paramecium* strains (Table 5.1) were put together in a depression slide containing 50 µL of sterile Cerophyl or lettuce medium. Numbers of motile cells were checked after 30 and 60 minutes. 10 cells of putative sensitives were employed as controls in each experiment, which was independently repeated three times. Attempts with sterile water instead of medium and/or extended observation periods were

Table 5.1 Strains employed during killer tests

Putative sensitive strain	Species
562	*Paramecium biaurelia*
325	*Paramecium triaurelia*
Alpha4-2	*Paramecium tetraurelia*
Hp1-9	*Paramecium pentaurelia*
159	*Paramecium sexaurelia*
GFG-1	*Paramecium octaurelia*
223	*Paramecium decaurelia*
UV	*Paramecium dodecaurelia*
Rs12	*Paramecium dodecauarelia*
209	*Paramecium tredecaurelia*

also performed.

Results

Morphology and ultrastructure. The cytoplasmic symbionts of *P. octaurelia* 299 (*L. flagellatum*) are straight rod-shaped bacteria 0.6-0.9 x 2.0-4.0 µm in size (Figures 5.1a-c), while those harbored by *P. biaurelia* USBL-36I1 are bigger – up to 1.1 x 7.8 µm – and curved (Figures 5.1d-f), perfectly fitting the description of *L. sinuosum*. Both are covered by numerous thick, peritrichous flagella about 4 µm long, clearly visible in TEM sections and negative staining. Nevertheless, *in vivo* observations did not show any sign of motility. The cytoplasm of both kinds of bacteria is homogeneous, with no visible inclusion. They both feature a Gram-negative type cell, with two membranes, and the symbionts are enclosed in a membrane-bound vesicle, often with several bacteria inside the same vesicle (Figures 5.1b,f). These results are in good accordance with previous descriptions[5].

Molecular characterization. The 16S rRNA gene sequences of the symbionts harbored by strains 114 and USBL-36I1 are identical. They differ by 6 out of 1331 (0.5%) sites from the homologous sequence of the 299 symbiont. The most similar sequences available according to NCBI blastn are those of the *Acanthamoeba* spp. UWC8 and UWC36 symbionts (87.1-88.0% similarity), which belong to the "*Candidatus* Midichloriaceae" family within *Rickettsiales*[50].

Hybridizations with the genus-specific oligonucleotide probe LytiProb_433 (that provides no match on RDP) gave clear signals deriving from bacteria localized in both 299 and USBL-36I1 cells at formamide concentrations in the range of 0-50% (with an optimum at 30%). *Lyticum* bacteria were always present in all examined paramecia, usually numbering in the hundreds, but sometimes far fewer – especially in the case of 299. *L. flagellatum* bacteria in 299 were often found concatenated in groups of 2 or more cells. Double hybridizations with the eubacterial probe EUB338 demonstrated that there are no other intracellular bacteria hosted by the *Paramecium* strains studied (Figures 5.2a,b). The species-specific probes Lflag_268 (providing only 2 matches on RDP, both corresponding to uncultured *Proteobacteria*) and Lsinu_268 (providing 546 matches on RDP, but only 5 inside the order *Rickettsiales*) used in competition were able to discriminate between 299 and USBL-36I1 symbionts at formamide

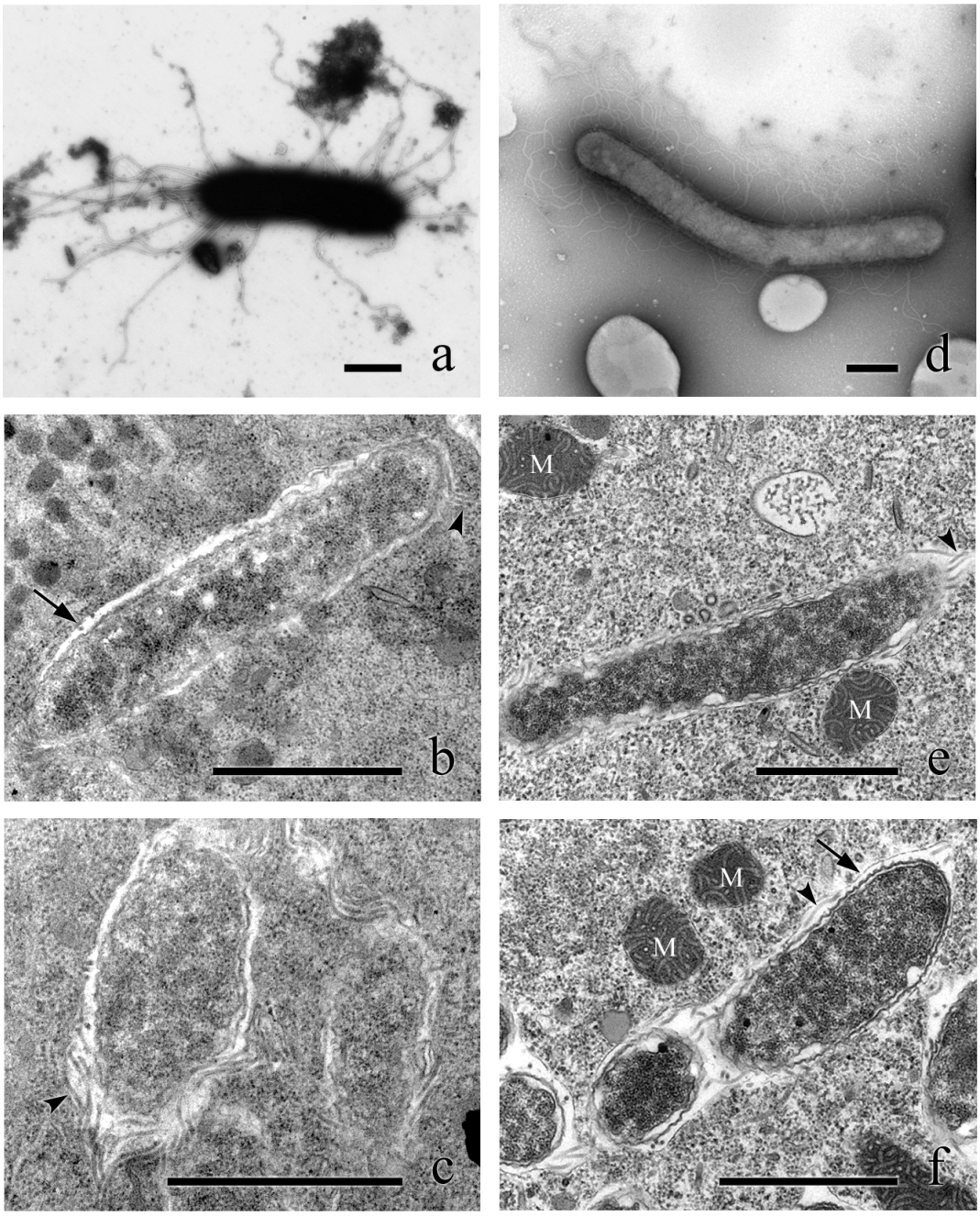


Figure 5.1 Morphology and ultrastructure of *Lyticum* species. Negative staining (a) and ultrathin sections (b, c) of *L. flagellatum* harbored by *P. octaurelia* strain 299. Negative staining (d) and ultrathin sections (e, f) of *L. sinuosum* harbored by *P. biaurelia* strain USBL-36I1. Bars stand for 1 µm. Arrowheads highlight some of the flagella, arrows point at symbiosomal membranes. M, mitochondria.

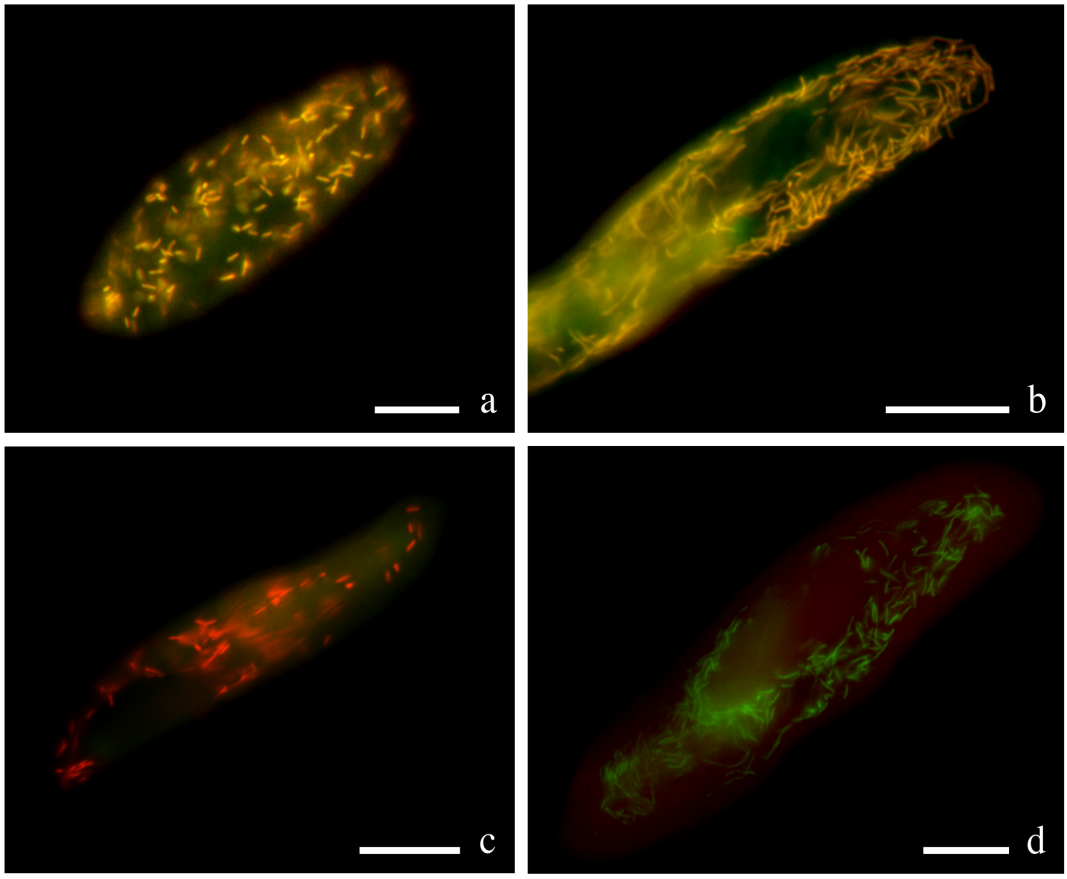


Figure 5.2 Genus- and species-specific *in situ* detection of *Lyticum flagellatum* and *Lyticum sinuosum*. Merge of the signals from probes EUB338 marked with fluorescein (green) and LytiProb_433 marked with Cy3 (red) on *P. octaurelia* strain 299 (a) and *P. biaurelia* strain USBL-36I1 (b). The signals coincide, and *Lyticum* bacteria appear yellowish. Merge of the signals from probes Lflag_268 marked with Cy3 (red) and Lsinu_268 marked with fluorescein (green) on *P. octaurelia* strain 299 (c) and *P. biaurelia* strain USBL-36I1 (d). At 20% formamide concentration, the probes used in competition are able to discriminate between the species. Bars stand for 20 µm.

concentrations in the range of 10-20% (with an optimum at 20%; Figures 5.2c,d).

Molecular phylogeny. Details of tree topology differ, according to the method and the character matrix employed, especially within the families "*Ca.* Midichloriaceae" and *Holosporaceae sensu lato*[18]. Nevertheless, all trees recover the monophyly of the *Rickettsiales* families, including *Holosporaceae sensu stricto*[18,25], and their relative positions, confirming other recent 16S rRNA analyses[46,50]. All trees confirm the monophyly of the *Lyticum* genus, as suggested by high similarity values between the strains, and its association to the candidate family *Midichloriaceae* within *Rickettsiales* (Figures 5.3). The exact relationships within this

family are not clearly resolved; however, the four genera of ciliate symbionts affiliated to this clade ("*Ca.* Anadelfobacter", "*Ca.* Cyrtobacter", "*Ca.* Defluviella" and *Lyticum*) do not form a monophyletic group.

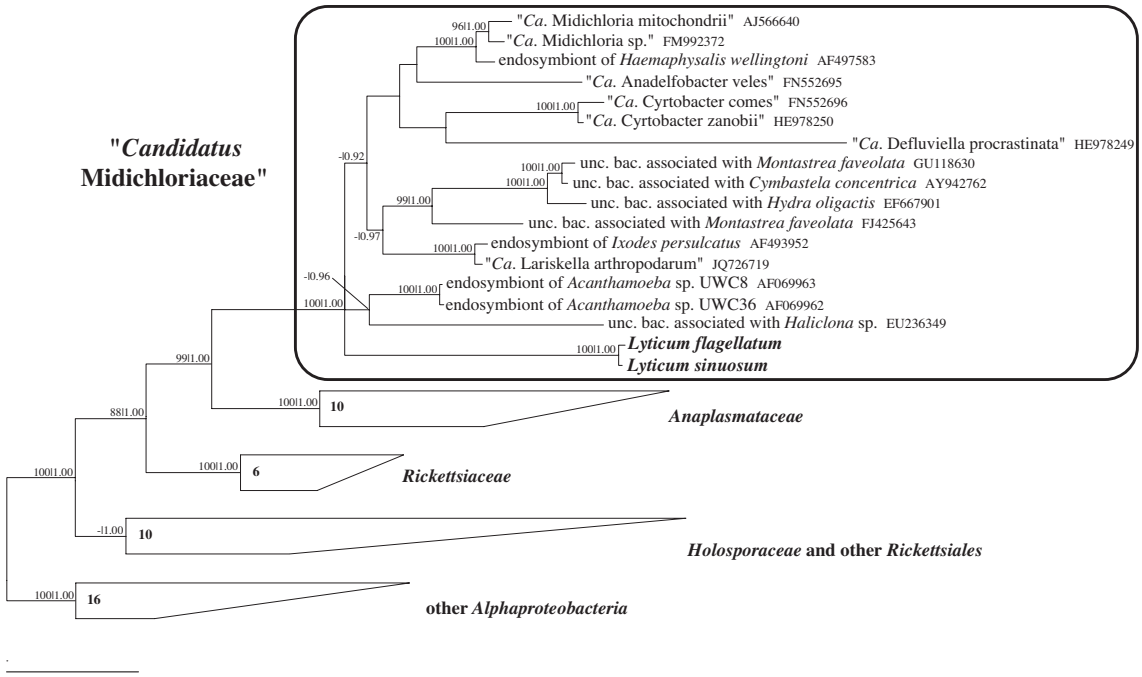


Figure 5.3 Phylogenetic position of *Lyticum* species. Bayesian tree built on the unmodified character matrix (60 sequences, 1331 characters) employing the GTR+I+G model (with the continuous gamma distribution approximated by 4 discrete categories). Numbers associated to each node correspond to ML bootstrap values and posterior probability values (values below 70|0.85 are omitted); numbers inside trapezoids show the number of sequences used to represent that clade. The bar stands for an estimated sequence divergence of 10%. *Ca.*, *Candidatus*; unc. bac., uncultured bacterium.

Killer effect. No killer effect was detected in any of the performed experiments. The number of living cells did not decrease in the treatments nor in the controls, and the pre-lethal symptoms described by Jurand and coworkers[51] were never observed.

Discussion

The infected *Paramecium* strains 299 and 114 were sampled almost a century ago[4]. Nevertheless, cultures of these ciliates still retain their original symbionts, although those of strain 114 are almost instantly lost after adaptation to standard cultivation conditions. On the

other hand, the stability of the *L. flagellatum-P. octaurelia* 299 relationship supports the hypothesis that the symbiosis is obligate for the host, which possibly depends on metabolites provided by the bacterium[52].

L. sinuosum has been reported so far only in *P. biaurelia* 114. We obtained a new environmental isolate of *P. biaurelia* which is infected by the same bacterial species, as can be inferred by morphology and the identity of 16S rRNA gene sequences. Interestingly, the monoclonal strain *P. biaurelia* USBL-36I1 was established from a water sample collected in the surroundings of the Indiana University, where T. Sonneborn was working at the time of his *Lyticum* description.

The morphological difference between the two *Lyticum* species corresponds to a difference in their 16S rRNA gene sequences, albeit small. Due to the diagnostic characters separating the two bacteria and the species-specific probes herein developed we recommend maintaining their status of separate but closely related species.

Although the identification of the described symbionts is sound, we could not repeat previous results concerning the killer trait. This was not entirely unexpected: the original literature describes the death of non-infected paramecia induced by *Lyticum* as extremely rapid (10-40 minutes), but triggered only in some *Paramecium* strains belonging to *P. triaurelia*, *P. pentaurelia* and *P. novaurelia*. Those strains were not available for the killer tests performed in this study. Therefore, our results suggest that those sensitive strains were the exception, and not the rule. The common adaptive explanation of the killer trait as a competitive advantage for the hosts[11,53] would not apply to *Lyticum*, which apparently has no effect on most strains of the *P. aurelia* complex, including those belonging to the same species as their hosts (*P. biaurelia*, *P. tetraurelia* and *P. octaurelia*). It is also possible, of course, that the *Lyticum* killer effect requires specific physiological conditions in the sensitive, the killer and/or its symbiotic bacteria, and that those requirements were not met in our experiments. However, also the recently sampled strain USBL-36I1 did not act as a killer. This result makes it highly unlikely that an "ageing" effect of the cultures is responsible for the loss of killer activity.

Lyticum clearly belongs to the recently established candidate family "*Ca.* Midichloriaceae"[50] within *Rickettsiales*, like several other symbionts of ciliates[24,46], amoebas[54] and metazoa[55,56]; a member of this group was also associated to fish disease[57]. The present study enables, for the first time, the assignment of a valid genus to this clade. Like other cytoplasmic

bacteria belonging to "*Ca.* Midichloriaceae"[24,56] and *Anaplasmataceae*[58], but in contrast to members of *Rickettsiaceae*[26,27,58], *Lyticum* symbionts are enclosed with an additional membrane, likely of host origin.

On the basis of genome annotations and phylogenomic analyses recently performed on "*Candidatus* Midichloria mitochondrii", a hypothesis concerning the presence of flagella and motility in the *Rickettsiales*-mitochondria ancestor was proposed[36], even though none of the so far characterized *Rickettsiales* bacteria actually possesses flagellar structures. Additionally to genome-derived evidences, further support is provided by the expression of flagellar genes on RNA and in one case also on protein level by "*Ca.* Midichloria mitochondrii"[59]. This hypothesis would confer an important role to motility in the establishment of the ancient symbiotic relationship that turned free-living bacteria into organelles. Our results support this view, revealing for the first time that heavily flagellated bacteria can be found among members of the order, and suggesting that the last common ancestor of *Rickettsiales*, or at least of "*Ca.* Midichloriaceae", possessed flagella. The next step required for corroborating this scenario would be obtaining the sequence of *Lyticum* flagellar genes, and comparing them with those found in the "*Ca.* Midichloria mitochondrii" genome to test the alternative hypothesis that they were acquired independently.

"*Ca.* Midichloria mitochondrii" displays no flagella and is non-motile. Curiously enough, the *Lyticum* species do not use their flagella for movement. The question arises whether flagella or single flagellar proteins can also serve other than locomotion related functions. In a syntrophic symbiosis between a fermentative bacterium and a methanogenic archaeon, the significant role of the flagellar cap protein FliD to synchronize their metabolism was described[60]. One might speculate about an involvement of the numerous *Lyticum* flagella in establishment or maintenance of the symbiosis with *Paramecium*, hence this question awaits future analyses.

References

1. Sonneborn TM. (1938) Mating types in *P. aurelia*: diverse conditions for mating in different stocks; occurrence, number and interrelations of the types. Proc Amer Phil Soc. 79: 411-434
2. Sonneborn TM. (1975) The *Paramecium aurelia* complex of fourteen sibling species. Trans Amer Micros Soc. 94: 155-178

3. Aufderheide KJ, Daggett PM, Nerad TA. (1983) *Paramecium sonneborni* n. sp., a new member of the *Paramecium aurelia* species-complex. J Protozool. 30, 128-131
4. Sonneborn TM. (1959) Kappa and related particles in *Paramecium*. Advan Virus Res. 6: 229-356
5. Preer JR, Preer LB, Jurand A. (1974) Kappa and other endosymbionts in *Paramecium aurelia*. Bacteriol Rev. 38: 113-163
6. Fokin SI, Görtz HD. (2009) Diversity of *Holospora* bacteria in *Paramecium* and their characterization. In: Fujishima M (ed), Endosymbionts in *Paramecium*. Microbiology Monograph, Münster, pp 161-199
7. Schrallhammer M, Schweikert M. (2009) The killer effect of *Paramecium* and its causative agents. In: Fujishima M. (ed), Endosymbionts in *Paramecium*. Microbiology Monograph, Münster, pp 227-246
8. Pond FR, Gibson I, Lalucat J, Quackenbush RL. (1989) R-body producing bacteria. Microbiol Rev. 53: 25-67
9. Springer N, Ludwig W, Amann R, Schmidt HJ, Görtz HD, Schleifer KH. (1993) Occurrence of fragmented 16S ribosomal-RNA in an obligate bacterial endosymbiont of *Paramecium caudatum*. Proc Natl Acad Sci USA. 90: 9892-9895
10. Beier CL, Horn M., Michel R., Schweikert M, Görtz HD, Wagner M. (2002) The genus *Caedibacter* comprises endosymbionts of *Paramecium* spp. related to the *Rickettsiales* (*Alphaproteobacteria*) and to *Francisella tularensis* (*Gammaproteobacteria*). Appl Environmen Microbiol. 68: 6043-6050
11. Kusch J, Czubatinsky L, Wegmann S, Hübner M, Alter M, Albrecht P. (2002) Competitive advantages of *Caedibacter*-infected paramecia. Protist. 153: 47-58
12. Schrallhammer M, Fokin SI, Schleifer KH, Petroni G. (2006) Molecular characterization of the obligate endosymbiont "*Caedibacter macronucleorum*" Fokin and Görtz, 1993 and of its host *Paramecium duboscqui* strain Ku4-8. J Eukaryot Microbiol. 53: 499-506
13. Schrallhammer M, Galati S, Altenbuchner J, Schweikert M, Görtz HD, Petroni G. (2012) Tracing the role of R-bodies in the killer trait: Absence of toxicity of R-body producing recombinant *E. coli* on paramecia. Eur J Protistol. 48: 290-296
14. van Wagtendonk WJ, Clark JAD, Godoy GA. (1963) The biological status of lambda and related particles in *Paramecium aurelia*. Proc Natl Acad Sci USA. 50: 835-838
15. Soldo AT, van Wagtendonk WJ, Godoy GA. (1970) Nucleic acid and protein content of purified endosymbiont particles of *Paramecium aurelia*. Biochim Biophys Acta. 204: 325
16. Jurand A, Preer LB. (1969) Ultrastructure of flagellated lambda symbionts in *Paramecium aurelia*. J Gen Microbiol. 54: 359-364
17. Preer JR, Preer LB. (1982) Revival of names of protozoan endosymbionts and proposal of *Holospora caryophila* nom. nov. Int J Syst Bacteriol. 32: 140-141
18. Görtz HD, Schmidt HJ. (2005) *Holosporaceae* fam. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds), Bergey's Manual of Systematic Bacteriology, 2nd edn, vol II, part C. Springer-Verlag, New York, pp 146-160
19. Petroni G, Spring S, Schleifer KH, Verni F, Rosati G. (2000) Defensive extrusive ectosymbionts of *Euplotidium* (Ciliophora) that contain microtubule-like structures are bacteria related to *Verrucomicrobia*. Proc Natl Acad Sci USA 97: 1813-1817
20. Vannini C, Rosati G, Verni F, Petroni G. (2004) Identification of the bacterial endosymbionts of the marine ciliate *Euplotes magnicirratus* (Ciliophora, Hypotrichia) and proposal of "*Candidatus* Devosia euplotis". Int J Syst Evolution Microbiol. 54: 1151-1156

21. Rinke C, Schmitz-Esser S, Stoecker K, Nussbaumer AD, Molnár DA, *et al.* (2006) "*Candidatus* Thiobios zoothamnicoli", an ectosymbiotic bacterium covering the giant marine ciliate *Zoothamnium niveum*. Appl Environ Microbiol. 72: 2014-2021

22. Shinzato N, Watanabe I, Meng XY, Sekiguchi Y, Tamaki H, *et al.* (2007) Phylogenetic analysis and fluorescence *in situ* hybridization detection of archaeal and bacterial endosymbionts in the anaerobic ciliate *Trimyema compressum*. Microbial Ecol. 54: 627-636

23. Eschbach E, Pfannkuchen M, Schweikert M, Drutschmann D, Brümmer F, *et al.* (2009) "*Candidatus* Paraholospora nucleivisitans", an intracellular bacterium in *Paramecium sexaurelia* shuttles between the cytoplasm and the nucleus of its host. Syst Appl Microbiol. 32: 490-500

24. Vannini C, Ferrantini F, Schleifer KH, Ludwig W, Verni F, Petroni G. (2010) "*Candidatus* Anadelfobacter veles" and "*Candidatus* Cyrtobacter comes", two new rickettsiales species hosted by the protist ciliate *Euplotes harpa* (Ciliophora, Spirotrichea). Appl Environ Microbiol. 76: 4047-4054

25. Boscaro V, Fokin SI, Schrallhammer M, Schweikert M, Petroni G. (2013) Revised systematics of *Holospora*-like bacteria and characterization of "*Candidatus* Gortzia infectiva", a novel macronuclear symbiont of *Paramecium jenningsi*. Microb Ecol. 65: 255-267

26. Vannini C, Petroni G, Verni F, Rosati G. (2005) A bacterium belonging to the *Rickettsiaceae* family inhabits the cytoplasm of the marine ciliate *Dioprhys appendiculata* (Ciliophora, Hypotrichia). Microbial Ecol. 49: 434-442

27. Ferrantini F, Fokin SI, Modeo L, Andreoli I, Dini F, *et al.* (2009) "*Candidatus* Cryptoprodotis polytropus", a novel *Rickettsia*-like organism in the ciliated protist *Pseudomicrothorax dubius* (Ciliophora, Nassophorea). J Eukaryot Microbiol. 56: 119-129

28. Sun HY, Noe J, Barber J, Coyne RS, Cassidy-Hanley D, *et al.* (2009) Endosymbiotic bacteria in the parasitic ciliate *Ichthyophthirius multifiliis*. Appl Environ Microbiol. 75: 7445-7452

29. Schrallhammer M, Ferrantini F, Vannini C, Galati S, Schweikert M, *et al.* (2013) "*Candidatus* Megaira polyxenophila" gen. nov., sp. nov.: considerations on evolutionary history, host range and shift of early divergent rickettsiae. PLoS ONE. 8: e72581

30. Schrallhammer M, Schweikert M, Vallesi A, Verni F, Petroni G. (2011) Detection of a novel subspecies of *Francisella noatunensis* as endosymbiont of the ciliate *Euplotes raikovi*. Microb Ecol. 61: 455-464

31. Boscaro V, Vannini C, Fokin SI, Verni F, Petroni, G. (2012) Characterization of "*Candidatus* Nebulobacter yamunensis" from the cytoplasm of *Euplotes aediculatus* (Ciliophora, Spirotrichea) and emended description of the family *Francisellaceae*. Syst Appl Microbiol. 35: 432-440

32. Andersson SGE, Zomorodipour A, Andersson JO, Sicheritz-Pontén T, Alsmark UCM, *et al.* (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature. 396:133-143

33. Gray MW, Burger G, Lang BF. (1999) Mitochondrial evolution. Science. 283: 1476-1481

34. Fitzpatrick DA, Creevey CJ, McInerney JO. (2006) Genome phylogenies indicate a meaningful alpha-proteobacterial phylogeny and support a grouping of the mitochondria with the *Rickettsiales*. Mol Biol Evol. 23: 74-85

35. Derelle R, Lang BF. (2012) Rooting the eukaryotic tree with mitochondrial and bacterial proteins. Mol Biol Evol. 29: 1277-1289

36. Sassera D, Lo N, Epis S, D'Auria G, Montagna M, *et al.* (2011) Phylogenomic evidence for the presence of a flagellum and cbb_3 oxidase in the free-living mitochondrial ancestor. Mol Biol Evol. 28: 3285-3296

37. Fokin SI. (2010/2011) *Paramecium* genus: biodiversity, some morphological features and the key to the main morphospecies discrimination. Protistology. 6: 227-235

38. Barth D, Krenek S, Fokin SI, Berendonk TU. (2006) Intraspecific genetic variation in *Paramecium* revealed by mitochondrial cytochrome *c* oxidase I sequences. J Eukaryot Microbiol. 53: 20-25

39. Spurr AR. (1969) A low-viscosity epoxy resin embedding medium for electron microscopy. J Ultrastruct Res. 26: 31-43

40. Venable JH, Coggeshall RA. (1965) simplified lead citrate stain for use in electron microscopy. J Cell Biol. 25: 407-408

41. Manz W, Amann RI, Ludwig W, Wagner M, Schleifer KH. (1992) Phylogenetic oligodeoxynucleotide probes for the major subclasses of Proteobacteria: problems and solutions. Syst Appl Microbiol. 15: 593-600

42. Amann R, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA. (1990) Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. Appl Environ Microbiol. 56: 1919-1925

43. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, *et al.* (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res. 37: D141-D145

44. Loy A, Maixner F, Wagner M, Horn M. (2007) probeBase – an online resource for rRNA-targeted oligonucleotide probes: new features 2007. Nucleic Acids Res. 35: D800-D804

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

46. Boscaro V, Petroni G, Ristori A, Verni F, Vannini C. (2013) "*Candidatus* Defluviella procrastinata" and "*Candidatus* Cyrtobacter zanobii", two novel ciliate endosymbionts belonging to the "*Midichloria* clade". Microb Ecol. 65: 302-310

47. Guindon S, Gascuel O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 52: 696-704

48. Huelsenbeck JP, Ronquist F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754-755

49. Posada D. (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol. 25: 1253-1256

50. Montagna M, Sassera D, Epis S, Vannini C, Lo N, *et al.* (2013) "*Candidatus* Midichloriaceae" fam. nov. (*Rickettsiales*), an ecologically widespread clade of intracellular alphapreoteobacteria. Appl Environ Microbiol. 79: 3241-3248

51. Jurand A, Rudman BM, Preer JR. (1982) Prelethal symptoms of the killing action of some non-kappa killers in the *Paramecium aurelia* complex. J Exp Zool. 220: 135-145

52. Soldo AT, Godoy GA. (1973) Observations on the production of folic acid by symbiont lambda particles of *Paramecium aurelia* stock 299. J Protozool. 20: 502

53. Landis WG. (1987) Factors determining the frequency of the killer trait within populations of the *Paramecium aurelia* complex. Genetics. 115: 197-205

54. Fritsche TR, Horn M, Seyedirashti S, Gautom RK, Schleifer KH, Wagner M. (1999) In situ detection of novel bacterial endosymbionts of *Acanthamoeba* spp. phylogenetically related to members of the order *Rickettsiales*. Appl Environ Microbiol. 65: 206-212

55. Sassera D, Beninati T, Bandi C, Bouman EAP, Sacchi L, *et al.* (2006) "*Candidatus* Midichloria mitochondrii", an endosymbiont of the tick *Ixodes ricinus* with a unique intramitochondrial lifestyle. Int J Syst Evol Microbiol. 56: 2535-2540

56. Matsuura Y, Kikuchi Y, Meng XY, Koga R, Fukatsu T. (2012) Novel clade of alphaproteobacterial endosymbionts associated with stinkbugs and other arthropods. Appl Environ Microbiol. 78: 4149-4156

57. Lloyd SJ, LaPatra SE, Snekvik KR, St-Hillaire S, Cain KD, Call DR. (2008) Strawberry disease lesions in rainbow trout from southern Idaho are associated with DNA from a *Rickettsia*-like organism. Dis Aquat Organ. 20: 111-118

58. Dumler JS, Walker DH. (2005) *Rickettsiales* Gieszczykiewicz 1939, 25^{AL} emend. Dumler, Barbet, Bekker, Dasch, Palmer, Ray, Rikihisa and Rurangirwa 2001. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds), Bergey's Manual of Systematic Bacteriology, 2nd edn, vol II, part C. Springer-Verlag, New York, pp 96-160

59. Mariconti M, Epis S, Sacchi L, Biggiogera M, Sassera D, *et al.* (2012) A study on the presence of flagella in the order *Rickettsiales*: the case of "*Candidatus* Midichloria mitochondrii". Microbiology. 158: 1677-1683

60. Shimoyama T, Kato S, Ishii S, Watanabe K. (2009) Flagellum mediates symbiosis. Science. 323: 1574

Chapter 5 Supplementary Information.

Supplementary Text S5.1 – 16S rRNA gene sequencing

16S rRNA gene sequencing of *Lyticum flagellatum*. The almost complete 16S rRNA gene sequence of the type strain of *L. flagellatum* (endosymbiont of *P. octaurelia* 299) was obtained through two touchdown PCRs (annealing temperature: 60 °C, 5 cycles; 58 °C, 10 cycles; 55 °C, 20 cycles; employing the high-fidelity TaKaRa Ex *Taq* and its buffer mix (TaKaRa Bio Inc., Otsu)) and direct sequencing of the amplified products. The amplification primers were: 16S F35 OdyHolo (5'-GCTGGCGGCATGCTTAAC-3') and 1492R (5'-GGNWACCTTGTTACGACTT-3', modified from Lane[61]) in the first PCR; 16S Lyti F832 (5'-AGGTGTTAAAATTTTTTTAGTGCC-3') and 16S R1522a (5'-GGAGGTGATCCAGCCGCA-3', modified from Schrallhammer and coworkers[12]) in the second PCR. The amplified products were sequenced with the internal primers 16S R785 ND (5'-TACCAGGGTATCTAATCC-3'[20]), 16S F515 ND (5'-GTGCCAGCAGCCGCGGT-3'[20]), 16S F49 AlphaSym (5'-TAACACATGCAAGTCGAAC-3') (first amplicon) and 16S Alfa R1517 (5'-TGA TCCAGCCGCAGGTTC-3'[27]) (second amplicon). The resulting sequences were compared and assembled.

16S rRNA gene sequencing of *Lyticum sinuosum*. The almost complete 16S rRNA genes of the type strain (endosymbiont of *P. biaurelia* 114) and the recently sampled strain (endosymbiont of *P. biaurelia* USBL-36I1) of *L. sinuosum* were PCR amplified using the universal primers Bac16SFor (5'-AAGAGTTTGATCCTGGCTC-3') and Bac16SRev (5'-TACGGCTACCTTGTTACGAC-3'[62]). Each reaction mix contained 1x PCR buffer (Colorless GoTaq Reaction Buffer, Promega) with 200 µM dNTPs, 2.5 mM $MgCl_2$, 1 U *Taq*-polymerase (GoTaq, Promega) and 1 U DNase I (RQ1 DNase, Promega) in a total volume of 20.5 µL. This mixture was incubated for 30 min at 37 °C to degrade any containing bacterial DNA. Afterwards, 1 µL of RQ1 DNase Stop Solution (20 mM EGTA, Promega) was added and the mixture was incubated for another 15 min at 65 °C to inactivate the DNase. In a second step, 10 pmol of each primer and 2.5 µL of Chelex® extracted DNA was added to reach a final volume of 25 µL. The PCRs were performed under touchdown conditions in which a 5 min initial

denaturation at 95 °C was followed by 35 cycles of denaturation at 95 °C for 45 s, annealing at 63-55 °C for 45 s and extension at 72 °C for 90 s; the initial annealing temperature of 63 °C was decreased by 0.7 K at each of the first 10 cycles to reach 55 °C for the final 25 cycles. A final extension step at 72 °C for 5 min was added to facilitate ligation of PCR products into the pGEM®-T Vector (Promega).

PCR products were analyzed by electrophoresis on 1% agarose gel (SeaKem® LE Agarose, Lonza) stained with GelRed™ (Biotium). Fragments of the right size (approx. 1.5 kb) were excised, purified with the NucleoSpin™ Extract II kit (Macherey-Nagel) and subsequently cloned and transformed into competent *E. coli* JM109 cells using the pGEM®-T Vector System (Promega) and following the manufacturer's instructions.

Picked colonies from agar plates were resuspended into 20 µL ddH_2O and incubated for 15 min at 95 °C for cell lysis. 10 µL of this suspension was used for colony PCR with the following components: 1x PCR buffer (Colorless GoTaq Reaction Buffer, Promega) with 50 µM dNTPs, 2 mM $MgCl_2$, 1 U *Taq*-polymerase (GoTaq, Promega) and 5 pmol of both M13 forward and reverse primer in a total volume of 20 µL. PCR conditions were: 3 min initial denaturation at 94 °C, followed by 35 cycles of 30 s at 94 °C, 30 s at 60 °C, 90 s at 72 °C and a final extension step of 5 min at 72 °C.

To identify colonies containing the 16S rRNA gene fragments of interest (derived from endosymbionts and not bacterial contaminants), PCR products were digested with the restriction enzyme MboI (Promega). Five clones derived from DNA of either strain 114 or USBL-36I1 showing the same and most represented restriction pattern were sequenced from both directions using standard M13 primers. The resulting five sequences of each strain were used to produce a consensus sequence for each of the two endosymbionts.

Additional References

61. Lane DJ. (1991) 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M (eds), Nucleic Acid Techniques in Bacterial Systematics, Wiley, New York, pp 115-147, 1991
62. Neilan BA, Jacobs D, Del Dot T, Blackall LL, Hawkins PR, *et al.* (1997) rRNA sequences and evolutionary relationships among toxic and nontoxic cyanobacteria of the genus *Microcystis*. Int J Syst Bacteriol. 47: 693-697

Chapter 5 Afterword.

As already mentioned, features of the bacterial order *Rickettsiales* are of particular interest for the pathogenic nature of its representatives and the accepted theory that the ancestor of mitochondria belonged to this clade. Identifying the apomorphies of the order would thus allow, according to the parsimony criterion, to infer some of the characteristics of the ancestor of eukaryotes' respiratory organelles.

No confirmed *Rickettsiales* with flagella was ever found until a few years ago. In 2011, the detection of non-functional flagellar genes in the genome of "*Candidatus* Midichloria mitochondrii", a member of the recently described *Rickettsiales* clade "*Candidatus* Midichloriaceae", provided a first exception to the rule. But it was a single case, unable to support generalized claims. The discovery that the flagellated genus *Lyticum* belongs to "*Candidatus* Midichloriaceae", reported in Chapter 5, added more evidences. Now a case could be made for the presence of flagella at least in the most recent common ancestor of this candidate family, that may indeed be the most closely related to mitochondria according to some phylogenomic analyses.

Another step forward was made shortly afterward, with the publication of a paper describing novel flagellated bacteria (Vannini et al., 2014 – see Appendix). Dr. Claudia Vannini coordinated the characterization of 6 strains, in Pisa ad elsewhere, including another symbiont of the *Euplotes* population In performed by myself. Our paper proved that there are flagellated bacteria in a second *Rickettsiales* family (*Rickettsiaceae*) and that they are capable of movement in the hosts' compartments (including the nucleus!). Additionally, these bacteria also contained, at least sometimes, capsid-like particles arranged in geometrical order. All details can be found online, including some staggering videos of bacteria swimming in mysterious channels in the macronuclear chromatin, without any apparent damage to the host.

The paper on the "flagellated rickettsiae" served multiple purposes. The newly described symbionts were previously unknown, and yet they seem to be relatively widespread, and are conspicuous enough to compete with "classic symbionts" like *Holospora* and *Caedibacter* for their allure. They opened a lot of lines of inquiry – why do they move inside the host? How can

the chromatin rearrangement not effect the ciliate? Are the flagellar genes indeed homologs of those found in "*C.* Midichloria"? What is the function, if any, of the capsid-like particles? Furthermore, the paper built a much stronger case for the presence of flagella in the ancestor of all *Rickettsiales* (mitochondria included). To definitely confirm the hypothesis, the genomes of flagellated symbionts should be obtained, and hopefully will be in the near future.

Another consequence of the paper that is less remarkable for science, but rewarding at a personal level, is the official description of a fourth symbiont in the cytoplasm of the *Euplotes* population In, alongside *Polynucleobacter necessarius*, "*Candidatus* Nebulobacter yamunensis", and "*Candidatus* Cyrtobacter zanobii" (see Chapters 2 and 4). This particular strain of "*Candidatus* Trichorickettsia mobilis" is not flagellated nor motile, but was able to survive for at least a couple of years in a stable and complex prokaryotic community of symbionts. The *Euplotes* population In currently holds the record for highest number of endocellular bacteria characterized and, although now extinct, was a great source of research material that never disappointed (despite being frustratingly difficult to maintain). And yet, I suspect that it was not extremely peculiar in this respect, but only more thoroughly investigated than others. If ciliates can indeed harbor, on a common basis, stable communities of 3 or more (probably interacting) prokaryote species, this field has indeed a very long future ahead.

PART II

Symbionts Genomics

As discussed previously, the number of still unknown bacterial symbionts of ciliates is not likely to become low in the near future. This ensures that the field of traditional characterizations (employing ultrastructural analyses and single molecular markers) will be fertile for a while still. However, it is important to stress that characterizations are just the first step in understanding organisms and their biology. The major goal at the beginning of my PhD work was to proceed for the first time to the next "molecular" level, and study the symbionts' genomes.

Genomics of bacterial symbionts is a field with a relatively rich literature. At closer inspection, however, it becomes apparent that the range of studied host organisms is strongly biased, dominated by selected examples of that huge (and economic important) class of metazoans which is Insecta. Insects are good models in several ways, but they hardly represent the diversity of eukaryotes. Other groups (especially among protists) are less studied more because of a lack of dedicated resources (human as well as economic) than an intrinsic lower appeal. This is the main reason why genomic studies of even the most well known bacterial symbionts of ciliates did not start earlier.

A second one is the limited number of researchers sharing interest in both topics: ciliate symbionts and genomics. Additional difficulties originate from methodological issues in separating symbionts and hosts in order to obtain reasonably pure DNA extracts for genome sequencing.

All three obstacles were finally overcome during my PhD research years. The protistology group of the University of Pisa, with its know-how on studying and handling bacterial symbionts of ciliates, first cooperated with Prof. Martin Hahn (University of Innsbruck), a microbial ecologist, in order to obtain the resources to sequence one symbiont's genome, and then with Prof. Michael Lynch (Indiana University), a renowned expert of genome evolution. I went to Bloomington (Indiana) in Winter 2012, with the genomic sequence obtained in Pisa from a symbiont of the ciliate *Euplotes*: *Polynucleobacter necessarius*.

Polynucleobacter necessarius is the most recent of the "classic symbionts" of ciliates, namely those discovered and named before the molecular era, and usually displaying conspicuous biological and morphological features – in this case, being essential for its host and possessing multiple dot-like nucleoids in the cytoplasm (hence the name). *P. necessarius* was not only the first case of an obligate symbiont in ciliates, but also the only prominent prokaryote described from a genus other than *Paramecium*, and it is still among the few

Betaproteobacteria found in ciliates.

A longer discussion about the studies on this organism is contained in the introduction of **Chapter 6**, that deals with the results of the genomic analysis. The data and conclusions are organized as a comparison between the newly obtained genome and that of a free-living strain of the same species, published previously. We started comparing the metabolism (as inferred from gene content) and trying to tie it with the physiology of the symbiotic relationship. Then we addressed the topic of genome reduction, which is so pervasive in the literature on endosymbionts. We strongly emphasized the possible effects of each separate DNA-repair systems (and their loss) and were able to generally confirm – but tweak, at the same time – the current model for genome reduction while presenting entirely new hypothetical mechanisms. My way to look at evolution in general, and not only genomic evolution, was changed and inspired by Prof. Lynch's theories, and the data interpretations and conclusions present in the paper equally belong to me and my colleague Michele Felletti, that became my own symbiont in all the phases of the work following annotation. Our speculations on the role of trans-lesion polymerase loss, that eagerly awaits experimental corroboration, is in my opinion the most general and original contribution of this PhD thesis.

Chapter 6.

The genome of *Polynucleobacter necessarius*

Original publication:

Polynucleobacter necessarius, a model for genome reduction in both free-living and symbiotic bacteria

Vittorio Boscaro[*], Michele Felletti[*], Claudia Vannini, Matthew S Ackerman, Patrick SG Chain, Stephanie Malfatti, Lisa M Vergez, Maria Shin, Thomas G Doak, Michael Lynch, Giulio Petroni

Proceedings of the National Academy of Sciences USA 110(46): 18590-18595. November 2013.

Abstract

We present the complete genomic sequence of the essential symbiont *Polynucleobacter necessarius* (*Betaproteobacteria*), which is a valuable case study for several reasons. First, it is hosted by a ciliated protist, *Euplotes*; bacterial symbionts of ciliates are still poorly known because of a lack of extensive molecular data. Second, the single species *P. necessarius* contains both symbiotic and free-living strains, allowing for a comparison between closely related organisms with different ecologies. Third, free-living *P. necessarius* strains are exceptional by themselves because of their small genome size, reduced metabolic flexibility, and high worldwide abundance in freshwater systems. We provide a comparative analysis of *P. necessarius* metabolism, and explore the peculiar features of a genome reduction that occurred on an already streamlined genome. We compare this unusual system with current hypotheses for genome

[*] These authors contributed equally to the work

erosion in symbionts and free-living bacteria, propose modifications to the presently accepted model and discuss the potential consequences of translesion DNA polymerase loss.

Introduction

Symbiosis, defined as a close relationship between organisms belonging to different species[1], is a ubiquitous, diverse and important mechanism in ecology and evolution[2-4]. In extreme cases, through the establishment of symbiotic relationships, quite unrelated lineages can functionally combine their genomes and generate advantageous emergent features, or initiate parasite/host arms races. Ciliates, common unicellular protists of the phylum Ciliophora, are extraordinary receptacles for prokaryotic ecto- and endosymbionts[5,6] that provide varied examples of biodiversity and ecological roles[6]. Nevertheless most of these symbionts are understudied, partially due to the scarcity of available molecular data and the absence of sequenced genomes. And yet, thanks to their various biologies and the ease of sampling and cultivating their protist hosts, they are excellent potential models for symbioses between bacteria and heterotrophic eukaryotes. Until recently, this field was dominated by studies on endosymbionts of invertebrates, especially insects[7], although unicellular systems like amoebas[8,9] have been shown to be suitable models.

Polynucleobacter necessarius was first described as a cytoplasmic endosymbiont of the ciliate *Euplotes aediculatus*[10,11]. Further surveys detected its presence in a monophyletic group of fresh and brackish water *Euplotes* species[12,13]. All of the investigated strains of these species die soon after being cured of the endosymbiont[10,12,13]. In the few cases where *P. necessarius* is not present, a different and rarer bacterium apparently supplies the same function[12,14]. No attempt to grow symbiotic *P. necessarius* outside their hosts has yet been successful[15], strongly suggesting that the relationship is obligate for both partners, in contrast to most other known prokaryote/ciliate symbioses[6].

Thus, the findings of many environmental 16S rRNA gene sequences similar to that of the symbiotic *P. necessarius*[16] but belonging to free-living freshwater bacteria came as a surprise. These free-living strains, which have been isolated and cultivated[17], are ubiquitous and abundant in the plankton of lentic environments[17,18]. They are smaller in size and do not show the most prominent morphological feature of the symbiotic form, the presence of multiple nucleoids each containing one copy of the genome[10,11]. It is clear that free-living and endosymbiotic *P. necessarius* are not different life stages of the same organism[15]. Nevertheless, these strikingly different bacteria, occupying separate ecological niches, exhibit >99% 16S rRNA gene sequence identity, and phylogenetic analyses fails to separate them into two distinct

groups[15]. Rather, several lines of evidence point to multiple, recent origins of symbiotic strains from the free-living bacterial pool[14,15].

Thus, the *Euplotes-Polynucleobacter* symbiosis provides a promising system for the study of changes promoting or caused by the shift to an intracellular life-style. The remarkably small (2.16 Mb) genome of the free-living strain QLW-P1DMWA-1 has been sequenced and studied, especially for features that would explain the success of this lineage in freshwater systems worldwide[19,20]. Phylogenies based on the 16S rRNA gene[13,14] and multiple-gene analyses[19,21,22] consistently cluster *Polynucleobacter* with bacteria of the family *Burkholderiaceae* (*Betaproteobacteria*), either in a basal position or as the sister group of *Ralstonia* and *Cupriavidus*.

Here, we provide the complete genomic sequence of a symbiotic *P. necessarius* harbored in the cytoplasm of *E. aediculatus* and present a comparative analysis of the two sequenced *Polynucleobacter* genomes, addressing the possible biological bases of the *Euplotes-Polynucleobacter* symbiosis. We also provide insights into the evolution of the unique two-step genome reduction in this bacterial species, the first step involving streamlining in a free-living ancestor, and the second a more recent period of genome erosion confined to the symbiotic lineage.

Materials and Methods

Purification of the endosymbiont DNA. Cultures of the ciliate host *Euplotes aediculatus* strain STIR1 were starved, filtered and treated with chloramphenicol in order to minimize the amount of contaminating bacteria in the culture medium. The cells were mechanically ruptured to release the symbionts, which were then separated from the eukaryote's cellular fragments through centrifugations at increasing accelerations. Total genomic DNA was extracted from the supernatant after the final centrifugation (protocols and quality controls are detailed in Supplementary Text S6.1).

Sequencing, assembling and annotation. The genome of this symbiotic strain of *P. necessarius* was selected for sequencing on the basis of the DOE Joint Genome Institute Community Sequencing Program 2006, and is publicly available (accession number:

NC_010531.1). Sequencing, assembling and annotation was performed as described elsewhere for the conspecific free-living strain QLW-P1DMWA-1T genome[20]. Pseudogenes were identified on the base of reading frame interruptions and/or substantially shorter sequence length with respect to orthologs, using one more conservative and one more permissive threshold (Table 6.1; for details, see Supplementary Text S6.2).

Table 6.1 Genomic features

	Symbiotic strain	**Free-living strain**	**Ratio**
Number of chromosomes	1	1	
Genome size	1,560,469 bp	2,159,490 bp	72.3%
GC content	45.6%	44.8%	
Coding size	1,142,322 bp	1,977,881 bp	57.8%
Protein-coding genes	1,279	2,075	61.6%
Average protein genes length	886 bp	949 bp	93.4%
RNA-coding genes	44	44	
Pseudogenes[a]	231	11	
Pseudogenes[b]	460	13	
Rest of non-coding DNA size	129,615 bp	176,194 bp	73.6%

[a] Conservative estimate (see Supplementary Text S6.2)
[b] Permissive estimate (see Supplementary Text S6.2)

Functional analysis. Predicted ORFs from symbiotic and free-living *P. necessarius* were compared against the non-redundant protein sequences database using BlastP[23]. A list of best results was produced for each ORF, and the putative protein product inferred. The KAAS – KEGG Automatic Annotation Server[24] was used as an aid in the interpretation of metabolic pathways in *Polynucleobacter*, and to comparatively screen a set of complete genomes from other *Burkholderiaceae* bacteria (see Table 6.2). Data on available genomes were obtained from the NCBI microbial genomes webpage (www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial_taxtree.html).

Substitution rates. Homologous sequences were defined by closed groups of reciprocal best hits determined from BlastP[23] among the pooled genes belonging to the genomes listed in Table 6.2. Amino acid sequences were aligned with MUSCLE[25] and nonsynonymous substitution rates were inferred from using PAML[26].

Table 6.2 Non-*Polynucleobacter* reference genomes*

	Genome accession number	Chromosome and plasmid number	Genome size (Mb)	GC content (%)	Protein coding genes
Burkholderia ambifaria AMMD	NC_010551-3, NC_010557	4	7.53	66.8	6,610
Burkholderia cenocepacia HI2424	NC_08542-5	4	7.70	66.8	6,919
Burkholderia gladioli BSR3	NC_015376-8, NC_015381-3	6	9.05	67.4	7,411
Burkholderia glumae BGR1	NC_012718, NC_012720-1, NC_012723-5	6	7.28	67.9	5,773
Burkholderia mallei ATCC 23344	NC_006348-9	2	5.84	68.5	5,023
Burkholderia multivorans ATCC17616	NC_010801-2, NC_010804-5	4	7.01	66.7	6,258
Burkholderia phenoliruptrix BR3459a	NC_018672, NC_018695-6	3	7.65	63.1	6,496
Burkholderia phymatum STM815	NC_010622-3 NC_010625 NC_010627	4	8.68	62.3	7,496
Burkholderia phytofirmans PsJN	NC_010676, NC_010679, NC_010681	3	8.21	62.3	7,241
Burkholderia pseudomallei K96243	NC_006350-1	2	7.25	68.1	5,728
Burkholderia rhizoxinica HKI 454	NC_014718, NC_014722-3	3	3.75	60.7	3,870
Burkholderia thailandensis E264	NC_007650-1	2	6.72	67.6	5,632
Burkholderia vietnamiensis G4	NC_009254-6, NC_009226-30	8	8.39	65.7	7,617
Burkholderia xenovorans LB400	NC_007951-3	3	9.73	62.6	8,702
Ralstonia eutropha JMP134	NC0073367, NC007347-8	4	7.26	64.5	6,446
Ralstonia solanacearum GMI1000	NC_003295-6	2	5.81	67.0	5,113
Ralstonia pickettii 12D	NC_012849, NC_012851, NC_012855-7	5	5.69	63.3	5,361
Cupriavidus metallidurans CH34	NC007971-4	4	6.91	63.5	6,477
Cupriavidus necator H16	NC_015723-4, NC_015726-7	3	7.42	66.4	6,626
Cupriavidus taiwanensis LMG19424	NC_010528-30	3	6.48	67.0	5,896

* All data come from the NCBI Microbial Genome website
http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial_taxtree.html

Results and Discussion

General features of the genome. The circular chromosome (Figure 6.1) of the symbiotic *P. necessarius* strain is 1.56 Mb long and contains about 1,279 protein-coding genes (Text S1, Table S1). The reduction in genome size that has occurred since the establishment of symbiosis can be estimated by comparing this genome with that of the *P. necessarius* free-living strain QLW-P1DMWA-1, an acceptable procedure given that the symbiont's gene set is largely a subset of that of the free-living strain (Figure 6.2; Supplementary Text S6.2). This reduction is more apparent as a decrease in the amount of coding DNA (42.3%) than in genome size (27.7%) because of the massive number of pseudogenes in the symbiont (Figure 6.3). There are virtually no genes related to mobile elements and extremely few recently duplicated genes in the symbiotic isolate. Horizontally-transmitted genes, mostly originated before the split between the two strains, are present (Supplementary Text S6.2), but there is no sign of DNA exchange between the symbiont and its host *Euplotes* nor any other eukaryote. A total of 277 genes in the symbiont genome (105 of which are not shared with the free-living relative) have unknown functions (Supplementary Text S6.2).

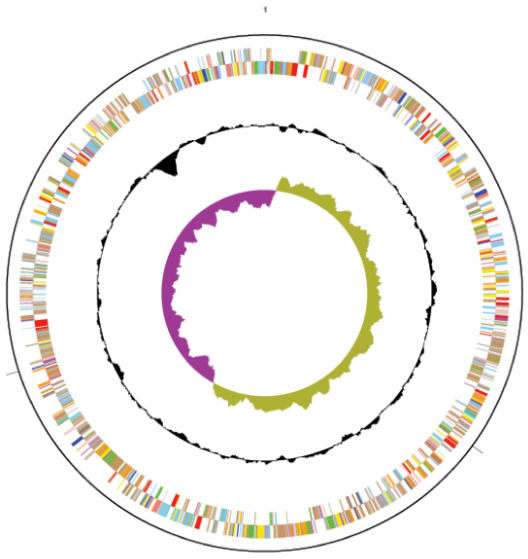


Figure 6.1 Circular map of the chromosome. From outside to the center: genes on forward strand, genes on reverse strand, GC content, GC skew. Gene colors reflect COG categories.

Metabolism. *Central metabolism and carbon sources.* Both the symbiont and the free-living *P. necessarius* lack a glycolytic pathway. They do not possess the central regulatory enzyme of the Embden-Meyerhof pathway (6-phosphofructokinase) nor enzymes specific to the Entner-Doudoroff variant, which is employed by most *Burkholderiaceae* bacteria. They also lack any enzyme that could phosphorylate glucose to glucose-6-phosphate or be involved in the assimilation of other

monosaccharides. The inability to exploit sugars as carbon or energy sources reflects a general poverty in catabolic pathways.

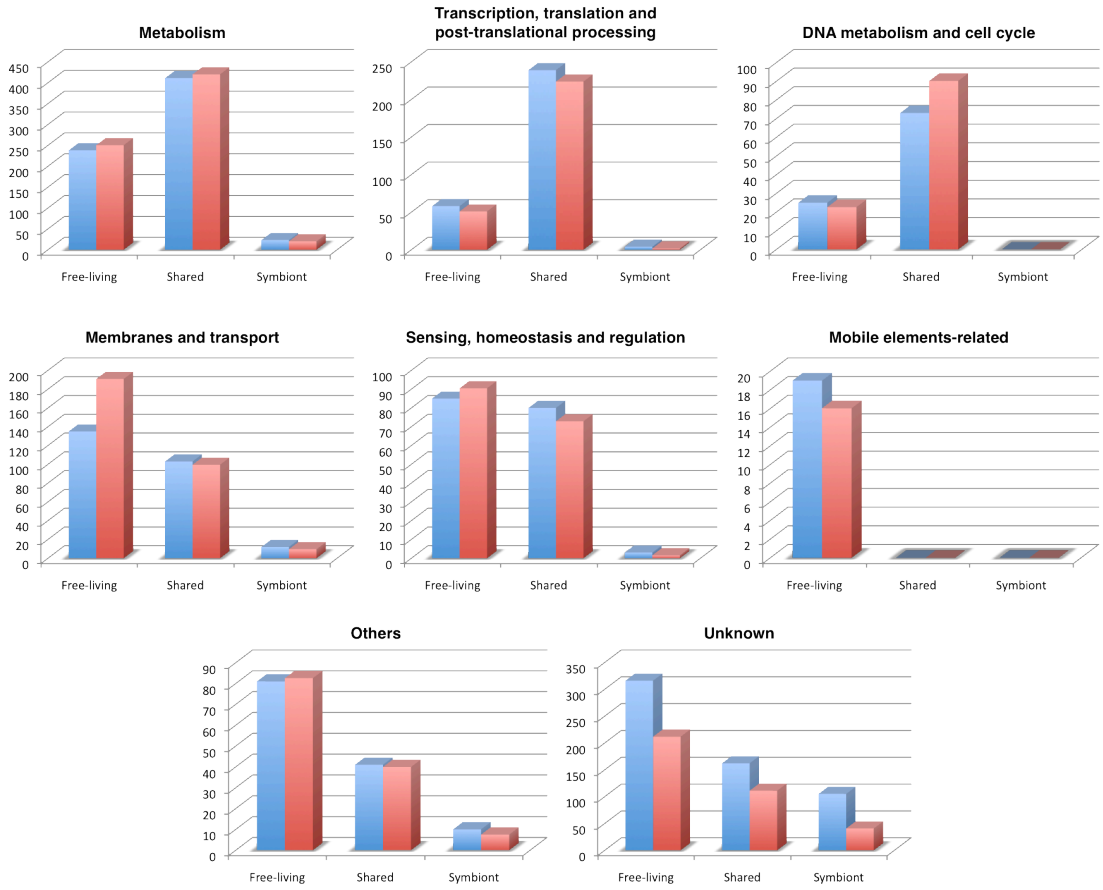


Figure 6.2 Bar graphs showing the putative functional genes exclusive to the free-living strain, the symbiont, and shared (with a close ortholog in each genome). Blue bars indicate gene number, red bars indicate the total length of the genes (in kbp, averaged for the shared genes).

Genomic analysis (Figure 6.4, Supplementary Text S6.3; see also Hahn and coworkers[19]) suggests that the principal carbon sources for the free-living strain are pyruvate, acetate, carboxylic acids and probably compounds convertible to them. These can be metabolized to acetyl-CoA, which is the key intermediate of both energy production and anabolism thanks to a complete glyoxylate cycle, tricarboxylic acids cycle (TCA), and gluconeogenesis pathways. Acetyl-CoA can additionally be directed to the synthesis of polyhydroxybutyrate (PHB), a storage polymer whose production has been investigated in the bacterium *Ralstonia eutropha*, also a

member of *Burkholderiaceae*[27]. A total of 8 amino acids and all TCA intermediates can also be converted to glucose.

The symbiont possesses all the aforementioned enzymatic paths with the exception of the glyoxylate cycle, the metabolic link between acetyl-CoA and biosynthetic pathways. Notably, genes involved in the polymerization, depolymerization and metabolic regulation of PHB are still present. Thus, non-TCA carboxylic acids are not exploitable as sole carbon sources, and most enzymes acting on related compounds are missing. Only 3 amino acids can potentially be converted to glucose in the symbiont (Supplementary Text S6.3). Thus, it either relies on a very small range of compounds as sole carbon sources, or directly imports various metabolic precursors from its host.

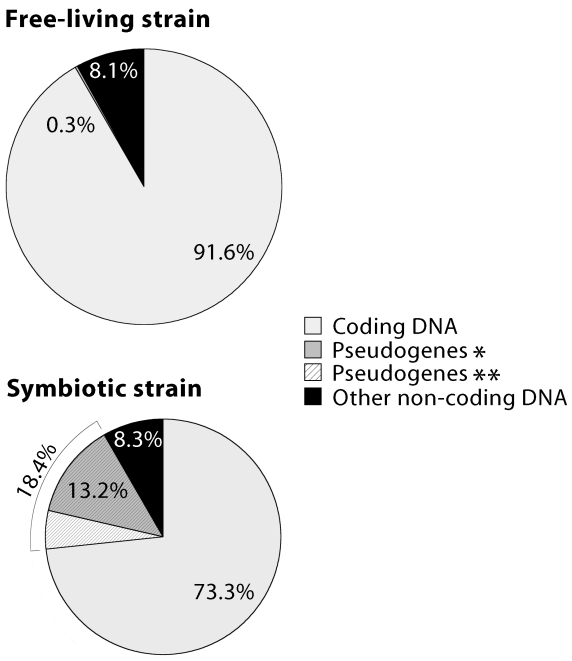


Figure 6.3 Percentage allocation of coding and non-coding DNA in the free-living and symbiotic *Polynucleobacter* genomes. The graph areas are proportional to genome sizes (2.16 and 1.56 Mb, respectively). Pseudogenes were identified with a more conservative (*) and a more permissive (**) approach as detailed in Supplementary Text S6.2.

Nitrogen and sulfur metabolism. The free-living *P. necessarius* can perform the assimilatory reduction of nitrate imported from the environment, but the symbiont has lost this ability. An operon including all necessary enzymes for import and catabolism of urea is present only in the free-living strain. Both the free-living and symbiotic *Polynucleobacter* do not possess enzymes involved in nitrification, denitrification, or nitrogen fixation.

The free-living strain can assimilate elemental sulfur and sulfate, while the pathway is absent in the symbiont. Moreover, as discussed by Hahn and coworkers[19], the free-living strain possesses an entire set of *sox* genes, and hence can probably obtain electrons from hydrogen

sulfide (chemolithoheterotrophy). Many of these genes remain in the symbiont, but it is unclear if the pathway is still functional.

Electron transport chain. Both genomes encode the entire electron transport chain complex and an F-type ATPase. Electrons must come mostly from carboxylic acids and, in the free-living strain, also from hydrogen sulfide. In addition to the most widespread cytochrome *c* oxidase complex, there is a variant in both – the *cbb-3* complex – and additionally a *bd* complex in the free-living strain. The *cbb-3* complex is present in the genera *Ralstonia* and *Cupriavidus*, and the *bd* complex is present also in *Burkholderia*; both are used in microaerophilic respiration. The symbiont apparently does not posses enzymes able to exploit terminal electron acceptors other than oxygen, while the free-living strain possesses a set of alcohol dehydrogenases,

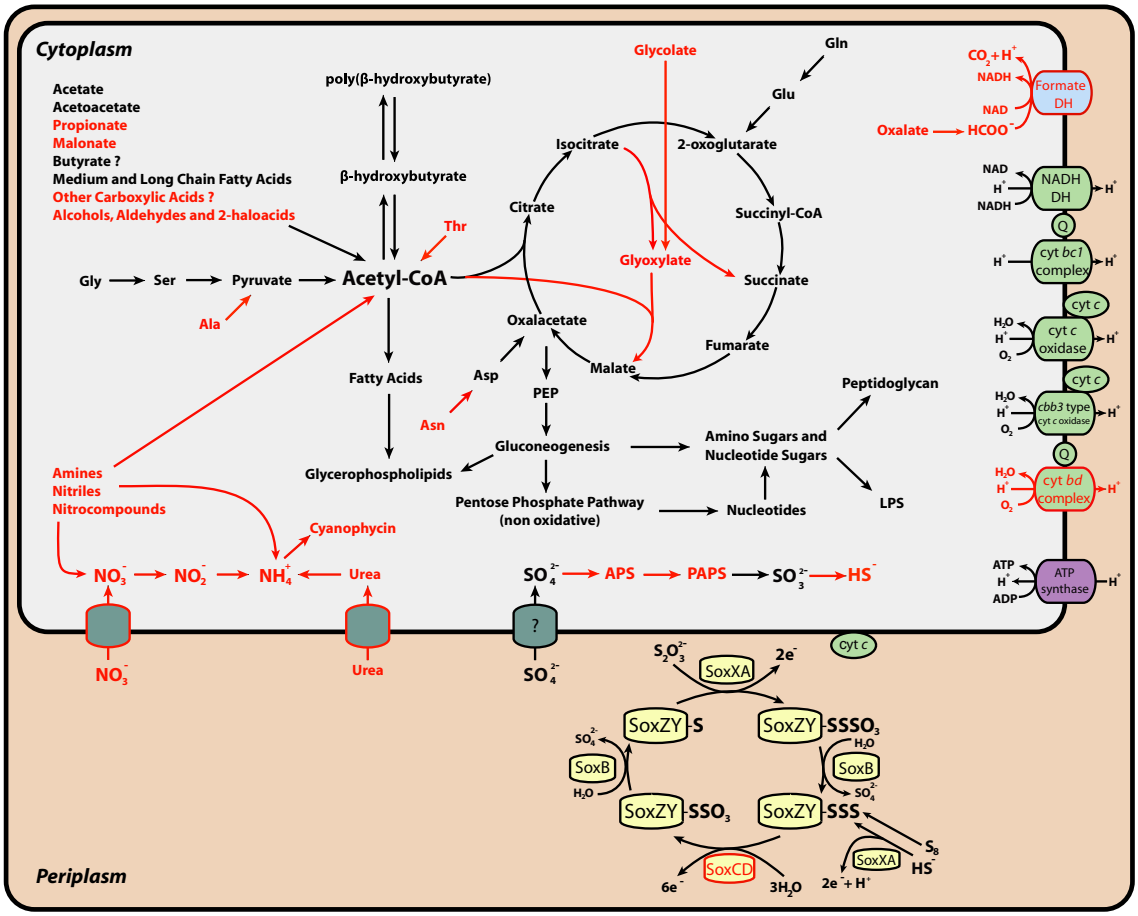


Figure 6.4 Schematic drawing of selected pathways of *Polynucleobacter necessarius* metabolism as inferred by genomic analysis. Elements in red are exclusive to the free-living strain's genome. *DH*, dehydrogenase; *cyt*, cytochrome.

suggesting the possibility of energy production through fermentation under anaerobic conditions.

Amino acid and cofactor metabolism. Experimental evidence finds that the free-living isolate can grow on single carbon sources (like acetate) with a few cofactors[19], so we must assume that all amino acid biosynthetic pathways are somehow present in the genome, although it appears that in certain cases the bacterium uses unconventional synthetic strategies (Supplementary Text S6.3). The biosynthesis of four amino acids (alanine, aspartate, serine and cysteine) has been lost in the symbiont. In a similar way, the symbiont cannot produce many cofactors, although the free-living strain lacks the ability to synthesize some of them as well (Supplementary Text S6.3).

Other biosynthetic pathways. The major biosynthetic pathways stemming from TCA and gluconeogenesis (fatty acids biosynthesis, *de novo* and salvage nucleotide biosynthesis, and all those providing the necessary intermediates, *e.g.* the non-oxidative part of the pentose phosphate pathway) are present in both bacteria. As suggested by encoded genes, *P. necessarius* can synthesize phosphatidylserine, phosphatidylethanolamine, phosphatidylglycerol and cardiolipin, but not phosphatidylinositol nor lecithin, as membrane phospholipids. Typical bacterial structures of the cell wall, peptidoglycan and lipopolysaccharide, can be synthesized as well, although the symbiont apparently has a reduced set of genes involved in the recycling and modification of both.

Other genomic features. *Gene classification.* We classified the genes of both genomes in broad functional categories (Supplementary Text S6.2), highlighting the differential amounts of gene loss in the symbiont compared with the free-living strain. Genes involved in core metabolism, protein metabolism and growth are relatively more retained, while those related to membranes and transport, sensing and regulation and other/unknown functions are much more reduced (Figure 6.2).

DNA repair systems. DNA repair systems are considerably underrepresented in both *Polynucleobacter* strains. While base and nucleotide excision repair pathways are intact, the enzymes for mismatch repair (MMR) are missing. Of the DNA polymerases able to perform translesion synthesis and rescue replication from arrest at points of chromosomal lesions (Pol II, Pol IV and Pol V), only the error-prone Pol V is present in the free-living strain, and none in the symbiont.

No intact homologous recombination pathway is present, either: the *recBCD* system is completely missing, while the *recFOR* system lacks the essential gene *recF*. The Holliday junction resolution system (*ruvABC* and *recG*), however, is present. This suggests that in *Polynucleobacter* illegitimate recombination mechanisms are present. Indeed, the genes *recJ*, *recO*, *recR* and *recA*, which are upregulators of illegitimate recombination[28], are maintained despite the absence of *recF*.

Membranes and transport. There are less than 100 putatively functional genes for transport-related proteins in the symbiotic *P. necessarius* (mostly belonging to large families like ABC, TRAP-T and major facilitators). In both strains, the phosphotransferase system is reduced to non-specific components and a few putative cytoplasmic enzymes. The specificity of most transporters is difficult to assess (see also Hahn and coworkers[19]).

Fewer differences between the strains were found in genes related to protein export. Both a Sec and Sec-independent pathways were found, as well as enzymes involved in the recognition of signal peptides, but no complete bacterial secretion system complex is encoded in either strain. Other membrane-related proteins are limited to lipoproteins and very conserved complexes involved in outer membrane integrity (Tol/Par operon), or outer membrane proteins' correct folding and targeting (YaeT complex). A huge putative membrane protein (10,429 aa) of unknown function is encoded by the free-living strain's genome[19], but is absent in the symbiont.

Cell-cycle, sensing, and stress resistance. While genes involved in DNA metabolism (including replication) and cell cycle are relatively conserved between the *P. necessarius* strains, those involved in sensing and resistance were largely lost in the symbiont, and are already few in number in the free-living isolate. Many genes of the two-component and TolB systems are likely nonfunctional in the symbiont. This suggests a strongly reduced ability to react to changes in the environment – the ciliate cytoplasm – which is probably more stable than the water column. Among environmental defenses, oxidative stress response systems are particularly reduced – the symbiont lacks, for example, glutathione synthetase, notwithstanding the presence of enzymes requiring glutathione (*e.g.* glutathione S-transferases).

Physiological bases of the *Polynucleobacter-Euplotes* symbiosis. The metabolic profile provides a clear explanation for the inability to grow symbiotic *P. necessarius* strains outside their hosts[15]. The symbiont relies on the ciliate at least for carbon sources, organic nitrogen and sulfur, and other essential molecules, including many cofactors. In contrast with

ancient or extremely specialized symbionts[25], though, it can still perform its own basic anabolic processes and energy production. This condition is similar to that of the other symbiotic member of the family *Burkholderiaceae* with a severely reduced genome: "*Candidatus* Glomeribacter gigasporarum"[22], a beneficial endosymbiont of arbuscular mychorrhizal fungi. *P. necessarius* and "*Ca.* G. gigasporarum" underwent independent events of genome reduction (see also the phylogenetic tree of Ghignone and coworkers[22]) that produced the loss of different metabolic pathways. For example, "*Ca.* G. gigasporarum" has conserved more amino acid degradation pathways than *P. necessarius* (Supplementary Text S6.3), but lacks instead the beta-oxidation pathway[22]. Both share the loss of glycolytic pathways.

The symbiotic *P. necessarius* is probably specialized for the intracellular environment in other aspects too, as suggested by its reduced set of genes for sensing and stress resistance. Defensive and regulation mechanisms, as well as membrane and cell wall plasticity appear to be weakened, providing other plausible reasons for the inability of the bacterium to grow outside its host.

It is more difficult to understand why the symbiont is essential for *Euplotes* survival[12,15]. Many obligate symbionts of eukaryotes described as mutualists serve as a source of essential molecules[7,29]. The possibility that *P. necessarius* provides at least some metabolites to its host cannot be completely ruled out, but we consider it unlikely. *Euplotes* are heterotrophic algal and bacterial feeders, and can probably obtain all required amino acids and cofactors from their diet, unlike specialized feeders like sap-feeding insects. More likely, the ciliate host requires *Polynucleobacter* in order to fix a catabolic deficiency (*e.g.* in compound degradation) in a pathway usually conserved in both bacteria and eukaryotes but lost in the clade of *Euplotes* species harboring *Polynucleobacter*. Vannini and coworkers[30] worked on a similar premise and provided evidence for a possible role of these bacteria in glycogen depolymerization; nevertheless, we found nothing in the genome supporting this hypothesis, so the real catabolic pathway involved remains uncertain. The genome of a *Polynucleobacter*-harboring *Euplotes* and further experimental investigation will be able to better address the matter, now that the genomic bases of *Polynucleobacter* biology are established.

It has been pointed out[31] that obligate bacterial symbionts generally do not have "symbiotic genes" coding for exotic functions. Their genomes are fundamentally a subset of those of free-living relatives, and the functional role of symbiosis is better explained by a

metabolic cooperation between partners[32,33]. An interesting exception are bacterial secretion systems (BSS), that are probably involved in the ancestral invasion process and are often found in an active or degraded form in symbiont genomes[29]. The symbiotic *P. necessarius* strain does not encode a complete BSS, contrary to some more parasitic-like bacteria that infect amoebas[8]. Nevertheless, we found a region similar to a pathogenicity island [positions 625,503-639,686] including 16 ORFs, 10 of which show similarities with type II and type IV secretion system assembly protein genes (the other 6 ORFs do not share significant similarities with any available sequence). This region is of horizontal origin (Supplementary Text S6.2) and is absent in the free-living strain genome. Most of the genes are considerably shorter (range: 13-78%) than their closest homologs, and are possibly nonfunctional. An ancestral free-living strain may have acquired these "invasion genes", which allowed it to survive ingestion and digestion in a predatory *Euplotes*, and was then trapped in the cytoplasm. The ubiquity of free-living *P. necessarius*, the high frequency of BSS genes horizontal transmission, together with no apparent requirement for the evolution of novel functions (at least at the genomic level), may have facilitated the multiple origins and complex pattern of replacement inferred for these essential *Euplotes* symbionts[14].

Genomic reduction in the symbiont. The symbiotic *P. necessarius* isolate possesses one of the smallest genomes observed so far in *Betaproteobacteria*, being surpassed only by two exceptional cases that blur the distinction between organism and organelle – "*Candidatus* Tremblaya princeps" (0.14 Mb, 148 genes)[34] and "*Candidatus* Zinderia insecticola" (0.21 Mb, 232 genes)[33].

Progressive genome reduction is the rule for obligate symbionts[34,35]. This process has been explained by non-selective mechanisms in insects' symbionts: either relaxed selection and enhanced genetic drift[36,37], or an increase in mutation rates[38]. The first hypothesis stems from the decreased number of essential functions in symbionts and their small population sizes, reduced gene exchange and frequent occurrence of bottlenecks. The prediction is for a higher nonsynonymous/synonymous (dN/dS) ratio in symbiotic lineages, which has often been reported[36,37], although a potential bias resulting from synonymous-site saturation has been suggested[38,39]. In *P. necessarius*, dN is slightly higher in the symbiotic than in the free-living strain (Figure 6.5), in accordance with expectations. Nevertheless, it was impossible to obtain confident estimates of dS in the two lineages despite their unparalleled level of sequence similarity, because synonymous sites were still saturated on the branch leading to the closest

outgroup. When more *P. necessarius* genomes of both free-living and symbiotic strains become available, this problem will almost certainly be circumvented with the help of an intraspecific outgroup.

Moran and coworkers[35] and McCutcheon and Moran[21] described the steps of genome

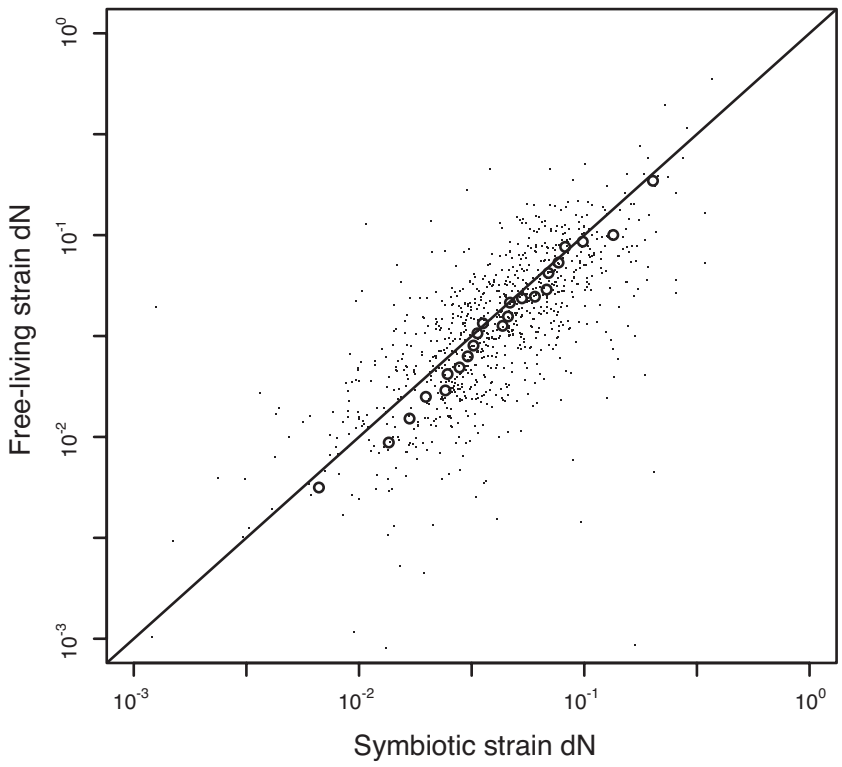


Figure 6.5 Nonsynonymous divergence between 904 homolog functional genes in the free-living and symbiotic *Polynucleobacter necessarius*. The rate of divergence from the outgroup is higher in the symbiont (averages: 0.0556 SE:0.0049 *vs* 0.0479 SE:0.0035; p=0.006, two-tail t-test). Axes are in logarithmic scale and each circle represents the average position of 40 gene-windows projected onto the diagonal.

erosion: the crucial turning point is the loss of DNA repair mechanisms, which brings increased mutation rates, an A+T bias, and massive gene inactivation driven by the spread of mobile elements. Under these conditions, elimination of DNA in non-coding regions is expected to be due to an intrinsic deletion bias[34,41]. Most of our findings are compatible with this scenario, notwithstanding the differences between the systems studied: *P. necessarius* doesn't experience a bottleneck during the host asexual division, and ciliate effective population sizes are arguably

larger than those of insects. Where our system stands in striking contrast to the prior model is the virtual absence of mobile elements (MEs), paired with the abundance of pseudogenes. MEs and pseudogenes should be a signature of the first stages of genome erosion, and until now have always been found together and in large numbers in recently evolved symbionts[34,39,40], but in very low quantity, or entirely missing, in the extremely reduced genomes of more ancient symbionts[40]. In this regard, the *P. necessarius* case is an exception in recent symbionts, and may be explained by the pre-existing small size of the ancestral genome, or by a relatively large effective population size. It demonstrates, however, that the multiplication of MEs is not a required step of genome erosion.

An aspect that has been less investigated is the role of translesion DNA polymerases (TLPs), and we propose here a scenario that may deserve attention in future studies. The symbiotic *P. necessarius* lacks all TLPs. While the loss of MMR and homologous recombination may increase the rate of mildly deleterious mutations, the loss of TLPs introduces the risk that a single damaged nucleotide, even in non-coding regions, can entirely block replication. This expectation is supported by the severe reduction in survival of bacteria with artificially impaired translesion synthesis under DNA-damaging conditions[42,43] and reduced fitness under non-stress conditions[44]. A small genome, however, provides less target sites for damage. If the last TLP was lost passively during the early stages of gene inactivation, this would have exerted an additional pressure towards deletions in sequences whose function is not selected for maintenance. The *umuC* and *umuD* genes (encoding the two subunits of Pol V, the only TLP present in the free-living strain) are not even recognizable as pseudogenes in the genome of the symbiotic *Polynucleobacter*, suggesting that the loss of translesion synthesis happened relatively early during the genome erosion process.

Not all proximate causes of genome reduction are known, but among them there is illegitimate recombination, apparently more so when coupled with MMR loss[28]. Although the molecular pathways involved are not entirely characterized, TLPs actually participate in deletion accumulation during illegitimate recombination in *Salmonella*[45]. However, other mechanisms have been proposed in different systems[46]. Illegitimate recombination and absence of MMR likely constitute the main source of deletions in *Polynucleobacter* through TLP-independent mechanisms.

The probability that an aborted chromosome replication leads to cell death could be alleviated by multiple copies of the genome. It is thus intriguing to link the loss of TLPs with the presence of multiple nucleoids in *Polynucleobacter*. Polyploidy has also been observed in other intracellular symbionts with reduced genomes, such as *Buchnera*[31], and this explanation might apply to them, too (and is not in contradiction with the possibility that a multiple-copy genome is advantageous because of increased gene expression). An alternative hypothesis is that multiple nucleoids arose in *P. necessarius* before the loss of the last remaining TLP, maybe as byproduct of an unbalanced cell cycle, and paved the way for reduced selection on translesion synthesis maintenance. However, the proximal causes of the polyploidy in the symbiont are not immediately apparent from genomic sequences. The symbiont and the free-living *P. necessarius* share a very similar set of genes involved in chromosome segregation and cell division (including the ParAB-*parS*, SMC and Fts systems[47]).

Genomic reduction in the free-living progenitor. As stressed above, the genome of the symbiotic *P. necessarius* strain is largely a subset of that of its conspecific relative. But the free-living *P. necessarius* strain QLW-P1DMWA-1T already possesses a remarkably small genome[19], comparable in size to that of betaproteobacterial obligate pathogen like *Neisseria* (2.09-2.28 Mb) and "*Ca.* G. gigasporarum" (about 1.72 Mb) and much smaller than those of other free-living or facultative symbiotic *Burkholderiaceae* (range: 3.75-9.73, see also Table S2). Thus, a first event of gene loss in the *Polynucleobacter* lineage occurred before the establishment of the symbiosis, presumably in a free-living ancestor.

Genome streamlining in free-living bacteria is less understood than genome erosion in symbionts, but has drawn attention in *Prochlorococcus*[48] and "*Candidatus* Pelagibacter ubiquae"[49], two marine taxa with huge global populations. Most authors have argued for adaptive explanations of genome streamlining in free-living bacteria[48,50], and selection-driven gene loss has been reported for experimental populations[51,52]. In particular, mechanisms requiring relaxed selection and increased drift are generally considered unrealistic because of the huge population sizes of these bacteria.

Nevertheless, the reduced genomes of symbionts and the streamlined genomes of free-living bacteria share many analogies, like a higher AT content and reduced DNA repair mechanisms, in particular MMR loss (reported here for *P. necessarius*, and also observed in strains of *Prochlorococcus*[48] and "*Candidatus* Pelagibacter"[53]). This suggests that non-adaptive

mechanisms may have shaped the genomes of these free-living lineages too, maybe acting in the past when the population sizes were smaller. Although likely detrimental during the first stages, a genome erosion driven by the same mechanisms invoked for symbionts may produce a compact, specialized and less costly metabolism that could be very successful in the right environment.

References

1. de Bary A. (1879) Die Erscheinung der Symbiose, Trübner, Strassburg
2. Dyall SD, Brown MT, Johnson PJ. (2004) Ancient invasions: from endosymbionts to organelles. Science. 304: 253-257
3. Gilbert SF, Sapp J, Tauber AI. (2012) A symbiotic view of life: we have never been individuals. Q Rev Biol. 87: 325-341
4. Moran NA. (2007) Symbiosis as an adaptive process and source of phenotypic complexity. Proc Natl Acad Sci USA. 104: 8627-8633
5. Fokin SI. (2012) Frequency and biodiversity of symbionts in representatives of the main classes of Ciliophora. Eur J Protistol. 48: 138-148
6. Görtz HD. (2006) Symbiotic associations between ciliates and prokaryotes. In: Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E (eds), The Prokaryotes. Third Edition, vol. 1, Springer-Verlag, New York, NY, pp 364-402
7. Moran NA, McCutcheon JP, Nakabachi A. (2008) Genomics and evolution of heritable bacterial symbionts. Annu Rev Genet. 42: 165-190
8. Horn M, Collingro A, Schmitz-Esser S, Beier CL, Purkhold U, *et al.* (2004) Illuminating the evolutionary history of chlamydiae. Science. 304: 728-730
9. Schmitz-Esser S, Tischler P, Arnold R, Montanaro J, Wagner M, *et al.* (2010) The genome of the amoeba symbiont "*Candidatus* Amoebophilus asiaticus" reveals common mechanisms for host cell interaction among amoeba-associated bacteria. J Bacteriol. 192: 1045-1057
10. Heckmann K. (1975) *Omikron*, ein essentieller Endosymbiont von *Euplotes aediculatus*. J Protozool. 22: 97-104
11. Heckmann K, Schmidt HJ. (1987) *Polynucleobacter necessarius* gen. nov., sp.. nov., an obligately endosymbiotic bacterium living in the cytoplasm of *Euplotes aediculatus*. Int J Syst Bacteriol. 37: 456-457
12. Heckmann K, Ten Hagen R, Görtz HD. (1983) Freshwater *Euplotes* species with a 9 type 1 cirrus pattern depend upon endosymbionts. J Protozool. 30: 284-289
13. Vannini C, Petroni G, Verni F, Rosati G. (2005) *Polynucleobacter* bacteria in the brackish-water species *Euplotes harpa* (Ciliata Hypotrichia). J Eukaryot Microbiol. 52: 116-122
14. Vannini C, Ferrantini F, Ristori A, Verni F, Petroni G. (2012) Betaproteobacterial symbionts of the ciliate *Euplotes*: origin and tangled evolutionary path of an obligate microbial association. Environ Microbiol. 14: 2553-2563

15. Vannini C, Pöckl M, Petroni G, Wu QL, Lang E, *et al.* (2007) Endosymbiosis *in statu nascendi*: close phylogenetic relationship between obligately endosymbiotic and obligately free-living *Polynucleobacter* strains (*Betaproteobacteria*). Environ Microbiol. 9: 347-359
16. Zwart G, Crump BC, Kamst-van Agterveld MP, Hagen F, Han SK. (2002) Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. Aquat Microb Ecol. 28:141-155
17. Hahn MW. (2003) Isolation of strains belonging to the cosmopolitan *Polynucleobacter necessarius* cluster from freshwater habitats located in three climatic zones. Appl Environ Microbiol. 69: 5248-5254
18. Jezberová J, Jezbera J, Brandt U, Lindström ES, Langenheder S, Hahn MW. (2010) Ubiquity of *Polynucleobacter necessarius* ssp. *asymbioticus* in lentic freshwater habitats of a heterogeneous 2000 km^2 area. Environ Microbiol. 12: 658-669
19. Hahn MW, Scheuerl T, Jezberová J, Koll U, Jezbera J, *et al.* (2012) The passive yet successful way of planktonic life: genomic and experimental analysis of the ecology of a free-living *Polynucleobacter* population. PLoS ONE. 7: e32772
20. Meincke L, Copeland A, Lapidus A, Lucas S, Berry KW, *et al.* (2012) Complete genome sequence of *Polynucleobacter necessarius* subsp. *asymbioticus* type strain (QLW-P1DMWA-1^T). Stand Genomic Sci. 6:74-83
21. McCutcheon JP, Moran NA. (2010) Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. Genome Biol Evol. 2: 708-718
22. Ghignone S, Salvioli A, Anca I, Lumini E, Ortu G, *et al.* (2012) The genome of the obligate endobacterium of an AM fungus reveals an interphylum network of nutritional interactions. ISME J. 6: 136-145
23. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990) Basic Local Alignment Search Tool. J Mol Biol. 215: 403-410
24. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. (2007) KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35: W182-W185
25. Edgar RC. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32: 1792-1797
26. Yang Z. (2007) PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24: 1586-1591
27. Pohlmann, Fricke WF, Reinecke F, Kusian B, Liesegang H, *et al.* (2006) Genome sequence of the bioplastic-producing "Knallgas" bacterium *Ralstonia eutropha* H16. Nat Biotechnol. 24: 1257-1262
28. Nilsson AI, Koskiniemi S, Eriksson S, Kugelberg E, Hinton JCD, Andersson DI. (2005) Bacterial genome size reduction by experimental evolution. Proc Natl Acad Sci USA. 102: 12112-12116
29. Dale C, Moran NA. (2006) Molecular interactions between bacterial symbionts and their hosts. Cell. 126: 453-465
30. Vannini C, Lucchesi S, Rosati G. (2007) *Polynucleobacter*: symbiotic bacteria in ciliates compensate for a genetic disorder in glycogenolysis. Symbiosis. 44: 85-91
31. Moran NA, Degnan PH. (2006) Functional genomics of *Buchnera* and the ecology of aphid hosts. Mol Ecol. 15: 1251-1261
32. Hansen AK, Moran NA. (2011) Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. Proc Natl Acad Sci USA. 108: 2849-2854

33. McCutcheon JP, von Dohlen CD. (2011) An interdependent metabolic patchwork in the nested symbiosis of mealybugs. Curr Biol. 21: 1366-1372
34. Moran NA. (2003) Tracing the evolution of gene loss in obligate bacterial symbionts. Curr Opin Microbiol. 6: 512-518
35. Moran NA, McLaughlin HJ, Sorek R. (2009) The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. Science. 323: 379-382
36. Moran NA. (1996) Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. Proc Natl Acad Sci USA. 93: 2873-2878
37. Wernegreen JJ, Moran NA. (1999) Evidence for genetic drift in endosymbionts (*Buchnera*): analyses of protein-coding genes. Mol Biol Evol. 16: 83-97
38. Itoh T, Martin W, Nei M. (2002) Acceleration of genomic evolution caused by enhanced mutation rate in endocellular symbionts. Proc Natl Acad Sci USA. 99: 12944-12948
39. Burke GR, Moran NA. (2011) Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. Genome Biol Evol. 3: 195-208
40. McCutcheon JP, Moran NA. (2012) Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol. 10: 13-26
41. Kuo CH, Ochman H. (2009) Deletional bias across the three domains of life. Genome Biol Evol. 1: 145-152
42. Jarosz DF, Cohen SE, Delaney JC, Essigmann JM, Walker GC. (2009) A DinB variant reveals diverse physiological consequences of incomplete TLS extension by a Y-family DNA polymerase. Proc Natl Acad Sci USA. 106: 21137-21142
43. Kuban W, Vaisman A, McDonald JP, Karata K, Yang W, *et al.* (2012) *Escherichia coli* UmuC active site mutants: effects on translesion DNA synthesis, mutagenesis and cell survival. DNA Repair. 11: 726-732
44. Yeiser B, Pepper ED, Goodman MF, Finkel SE. (2002) SOS-induced DNA polymerases enhance long-term survival and evolutionary fitness. Proc Natl Acad Sci USA. 99: 8737-8741
45. Koskiniemi S, Andersson DI. (2009) Translesion DNA polymerases are required for spontaneous deletion formation in *Salmonella typhimurium*. Proc Natl Acad Sci USA. 106: 10248-10253
46. Ikeda H, Shiraishi K, Ogata Y. (2004) Illegitimate recombination mediated by double-strand break and end-joining in *Escherichia coli*. Adv Biophys. 38: 3-20
47. Reyes-Lamothe R, Nicolas E, Sherrat DJ. (2012) Chromosome replication and segregation in bacteria. Annu Rev Genet. 46: 121-143
48. Dufresne A, Salanoubat M, Partensky F, Artiguenave F, Axmann IM, *et al.* (2003) Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. Proc Natl Acad Sci USA. 100: 10020-10025
49. Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, *et al.* (2005) Genome streamlining in a cosmopolitan oceanic bacterium. Science. 309: 1242-1245
50. Morris JJ, Lenski RE, Zinser ER. (2012) The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. MBio. 3: e00036-12
51. Lee MC, Marx CJ. (2012) Repeated, selection-driven genome reduction of accessory genes in experimental populations. PLoS Genet. 8: e1002651
52. Koskiniemi S, Sun S, Berg OG, Andersson DI. (2012) Selection-driven gene loss in bacteria. PLoS Genet. 8: e1002787

53. Viklund J, Ettema TJ, Andersson SG. (2012) Independent genome reduction and phylogenetic reclassification of the oceanic SAR11 clade. Mol Biol Evol. 29: 599-615

Chapter 6 Supplementary Information.

Supplementary Text S6.1 – Details of methods

Purification of the endosymbiont DNA. The ciliate host *Euplotes aediculatus* strain STIR1 was collected in the freshwater Stirone river (Italy)[54]. It was cultured in artificial brackish water (5‰ salinity), in an incubator with constant temperature (19-20 °C) and 12:12 h irradiance of 200 µmol photons m^{-2} s^{-1}, and fed with the green alga *Dunaliella salina*.

The ciliates were starved for at least one week before performing DNA extraction. The day before extraction, about 5 L of culture were washed and treated in order to eliminate possible contaminants. Ciliates were placed in a cylindric filter with pores of 10 µm and rinsed with ~2.5 L of flowing sterile culture medium, thus concentrating the *Euplotes* cells and diluting algal and bacterial contaminants. The remaining volume (about 200 mL) was added with chloramphenicol to a final concentration of 0.2 mg/mL and maintained at 15 °C overnight. Pilot experiments, followed by *in situ* hybridizations (performed according to Vannini and coworkers[15]), showed that such a treatment was efficient in removing bacteria in the medium without affecting *Polynucleobacter* symbionts. After this step, ciliate cells were washed again through passage in the cylindric filter and then through three centrifuges (250 g, 10 min) and the pelleted cells transferred to sterile culture medium. After the third wash, the pellet was suspended in a volume of 2 mL and transferred to a glass grinder for mechanical lysis on ice. Successful rupture of the cellular membranes and release of the symbionts were checked by microscope.

The bacteria were separated from most of the eukaryote's cellular fragments through centrifugations at increasing acceleration. The homogenate was loaded at the top of a PBS solution and centrifuged at 50, 100, 200 and 300 g (10 min), discarding the pellet after each step. A centrifugation at 11,300 g (10 min) was performed to collect the purified bacteria, followed by a lysis step in 1% SDS. A final centrifugation, again at 11,300 g, separated the *Polynucleobacter* DNA, in the supernatant, from DNA-containing-residue of eukaryotic organelles in the pellet. Total genomic DNA was extracted from this last supernatant with a modified CTAB method (see below). In order to check and optimize this method, DNA extraction from different pellets and supernatants were evaluated through electrophoresis on a

1% agarose gel stained with ethidium bromide. The host DNA appears as a smear of high-to-low size fragments. In the final supernatant, the smear is not visible on the gel. DNA extracted from a total volume of ~30 L of ciliate cultures was used for genome sequencing.

CTAB method protocol

- Add lysozyme (100 mg/ml) and mix
- Incubate for 5 min at RT
- Add 10% SDS and mix
- Add Proteinase K (10 mg/ml) and mix
- Incubate for 1 hr at 37 °C
- Add 5 M NaCl and mix
- Add CTAB/NaCl (4.1 g NaCl and 10 g CTAB dissolved in 100 ml water heated to 65 °C) and mix
- Incubate 65 °C for 10 min
- Add chloroform:isoamyl alcohol (24:1) and mix
- Spin at max speed for 10 min at RT
- Transfer aqueous phase to clean microcentrifuge tubes
- Add phenol:chloroform:isoamyl alcohol (25:24:1) and mix
- Spin at max speed for 10 min at room temperature
- Transfer aqueous phase and add 0.6 vol isopropanol (-20 °C)
- Incubate at room temperature for 30 min
- Spin at max speed for 15 min
- Wash pellet with 70% ethanol, spin at max speed for 5 min
- Discard the supernatant and let pellet dry for 5-10 min at RT
- Resuspend in TE plus RNAse A (conc. 10 mg/ml) 99:1
- Transfer to sterile microcentrifuge tubes
- Incubate at 37 °C for 20 min

RT, room temperature

Supplementary Text S6.2 – Genome composition analysis

Both *Polynucleobacter necessarius* genomes are encoded by a single circular chromosome, 2.16 Mb long in the free-living strain QLW-P1DMWA-1 and 1.56 Mb long in the symbiotic strain harbored by *Euplotes aediculatus* STIR1.

Pseudogenes were identified with the software GenePRIMP[55] according to two criteria: interruptions of the reading frame (stop codons, frameshift mutations) and/or severely reduced length (less than 50% of the 5 best hit homologs, defined as non-fused genes with at least 40% sequence identity with the query gene). In addition to this, we considered as likely non-functional those genes whose length in the symbiont genome is less then 75% that of the ortholog in the free-living strain genome (and vice-versa). All the genes in this category encode single enzymes in otherwise missing multiple-gene pathways or transporter operons (that are instead completely encoded in the free-living *Polynucleobacter* genome), and thus likely represent non-functional remnants of the process of genome erosion.

The remaining, putatively functional protein-coding genes were divided in 8 categories:

1. **Metabolism**, coding for known enzymes performing reactions involved in core metabolism, energy production, nitrogen and sulfur metabolism and biosynthetic/catabolic processes.
2. **Transcription, translation and post-translational protein processing** (including protein degradation).
3. **DNA metabolism and cell cycle**.
4. **Membranes and transport**.
5. **Sensing, homeostasis and regulation**.
6. **Mobile elements-related**, coding for transposases, integrases and/or of phage- or plasmid-origin.
7. **Others**, coding for enzymes belonging to known families but whose precise role in cell physiology is not known.
8. **Unknown**, putative protein genes that could not be unambiguously assigned.

Table 6.1 reports comparative data, while Figure 6.2 visually depicts the degree of gene loss in the symbiont's genome for each protein-coding gene category.

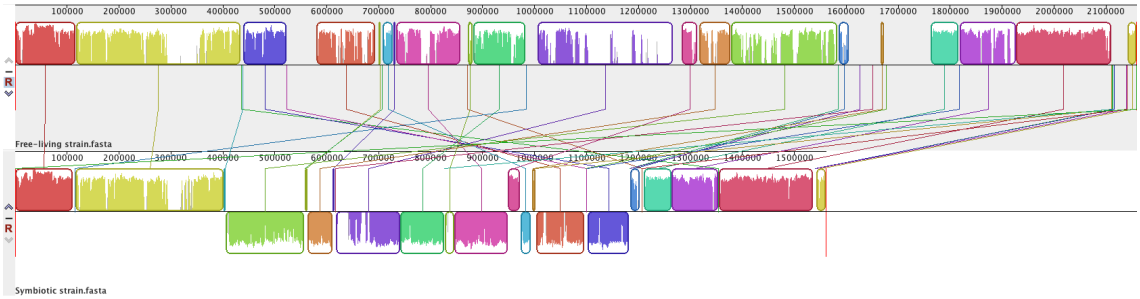
The most apparent feature of the symbiont's genome is its huge number of pseudogenes compared to that of the free-living strains. This strongly reduces the percentage of coding DNA, which is high in the free-living *Polynucleobacter* genome (91.6%). The reduction in the rest of non-coding DNA is instead similar to the genome size reduction (26.4%).

Concerning the categories, all those related to metabolic processes (1-3) are relatively more conserved, with a high percentage of shared genes and a comparatively lower number of genes exclusive to the free-living strain. All other categories are strongly underrepresented in the symbiont. Genes included in category 8 ("Unknown") are generally shorter than those in other categories, often less than 300 bases long in both genomes (average in the symbiont: 542.0 bp). 105 predicted genes in the symbiont are not shared with the free-living strain, but these are even shorter (average: 384 bp) and 75 do not share significant sequence similarity with any other available gene. It is plausible that many of them are not actually active, thus reducing the number of proteins expressed by the bacteria, and consequently the probability that important features went unnoticed in our analysis.

In all categories, the percentage of genes exclusive to the symbiotic strain is negligible, thus allowing to use the related free-living strain's genome as an approximation of that of the common ancestor. Orthologous gene sequences tend to be shorter in the symbiont (average: 93.4%), too.

Horizontally-transmitted genes mostly originated before the split between the two strains: the best BlastP hit still is the ortholog in the conspecific genome, but all the others belong to non-betaproteobacterial prokaryotes. There are about 80 of these genes in the symbiont's genome; approximately half have no known function. All genes in the putative pathogenicity island discussed in the main text are of horizontal origin (BlastP best hits include unrelated bacteria like *Pseudomonas*, "*Candidatus* Pelagibacter" and *Legionella*), and do not share an ortholog in the free-living strain's genome.

Whole-genome alignment performed with Mauve software[56] revealed one major inversion event and several minor ones since the divergence of the two strains (Supplementary Figure S6.1).



Supplementary Figure S6.1 Free-living / symbiotic *Polynucleobacter* whole-genome alignment performed with the Mauve software, showing a major inversion event and several minor ones.

Supplementary Text S6.3 – Details on metabolic analysis

Carbon sources. According to the analysis of coded enzymes, the principal carbon sources for the free-living strain are pyruvate, tricarboxylic acid cycle (TCA) intermediates, acetate, carboxylic acids and carboxylic acid derivatives. Some of those compounds constitute the low molecular weight photo-oxidation products of humic substances (HS) on which Hahn and coworkers[19] observed growth of bacterial cultures. In such a metabolism acetyl-CoA represents the key intermediate of both energy production and anabolism, thanks to complete glyoxylate cycle, TCA and gluconeogenesis pathways. No enzymes involved in mono- and polysaccharides catabolism were found, with the notably exception of a cellulase which in the free living could be involved in the regulation of biofilm formation in antagonism of cellulose synthase (see below).

Pyruvate and gluconeogenesis. Pyruvate can be directly converted to phosphoenolpyruvate (PEP) with PEP synthetase only in the free-living *Polynucleobacter*. Thus, due to the lack of a complete glyoxylate cycle, pyruvate cannot be used as a sole carbon source in the symbiont. The same is true for all the carbon sources that are converted to acetyl-CoA as intermediate. Oxalacetate can be used as a substrate for gluconeogenesis through phosphoenolpyruvate carboxykinase in both strains, thus TCA intermediates are glucogenic.

Acetate and acetoacetate derivatives. In both symbiotic and free-living *Polynucleobacter*, acetate can be directly assimilated through acetyl-CoA synthetase. In many bacteria this enzyme was shown to act on propionate and butyrate as well[57]. In both strains, acetoacetate can be

transferred on CoA and metabolized thanks to a 3-oxoacid-CoA transferase, whereas 3-hydroxybutyrate, coming from the degradation of poly-β-hydroxybutyrate, can be used as a further carbon source after oxidation to acetoacetate. Chloroacetate and other 2-halocids could be metabolized by the 2-haloacid dehalogenase present in both bacteria.

Propionate. In the free-living *Polynucleobacter*, propionate can be directly assimilated through a propionyl-CoA synthetase. Apparently the methylmalonyl-CoA pathway for the propionate assimilation is not complete due to the lack of methylmalonyl-CoA epimerase, however propionate could be metabolized, at least in the free-living strain, through a modified β-oxidation pathway involving malonic semialdehyde or 3-oxopropionyl-CoA/malonyl-CoA as intermediates. The ability to catabolize propionate is uncertain in the symbiont because it lacks propionyl-CoA synthetase (although the reaction could be performed by acetyl-CoA synthetase). The β-oxidation pathway variant involving malonic semialdehyde is not functional, while the variant involving 3-oxopropionyl-CoA/malonyl-CoA as intermediates might be present, but is not well characterized from the enzymatic and genetic point of view.

Fatty acids and lipids. A complete ß-oxidation pathway is present in both the free-living and symbiotic strain, even though they don't seem to be able to metabolize short chain fatty acids (C6, C8, C10) because they lack the short chain acyl-CoA synthetase. This is in accordance with the assimilation tests performed by Hahn and coworkers[19,58] on the free living bacterium. On the other hand medium and long chain fatty acids can by transferred on the CoA by a synthetase. More complex lipids such as triacylglicerids and some phospholipids can be metabolized as well due to the presence of three secretory lipases and three putative patatin-like phospholipases.

Other carboxylic acids. In the free-living strain's genome the presence of a gene encoding for a CoA transferase, which in *Ralstonia* catalyzes reversible transfer reactions of a CoA group from CoA-thioesters to free acids using acetyl-CoA as a CoA donor[59], could be responsible for the assimilation of other carboxylic acids. Malonate can be assimilated thanks to the presence of a specific malonyl-CoA transferase. Only in the free-living strain, three organic acids (pyruvate, malate and L-lactate) can be oxidized by specific oxidoreductases which direct the reducing equivalents directly into the electron transport chain: pyruvate and malate oxidoreductases reduce the quinone, L-lactate dehydrogenase reduces the cytochrome.

Glycolate and tartrate. As previously shown by Hahn and coworkers[19], the free-living

strain has the genetic potential to assimilate glycolate and phospho-glycolate (products of algal photorespiration), which are important components of algal exudates. Furthermore, the free-living bacterium seems to have the enzymatic activities (tartrate dehydrogenase/decarboxylase) for the conversion of tartrate, another important component of vegetal exudates, to D-glycerate. It has also the enzyme for the conversion of hydroxypyruvate to D-glycerate. Surprisingly it was not possible to find any gene encoding for a glycerate kinase, which catalyzes the synthesis of the glycolysis intermediate 3-phosphoglycerate. No functional gene for glycolate and tartrate degradation were found in the symbiont.

Oxalate and formate. Only in the free-living bacterium, oxalate can be used as an additional source of electrons (but not as a carbon source). Oxalate can be converted to formate through the concurrent action of the formyl-CoA transferase and oxalyl-CoA decarboxylase. Formate can be further oxidized to CO_2 with the production of NADH, due to the presence of at least one complete set of genes, present only in the free-living bacterium, encoding for a functional formate dehydrogenase.

Alcohols and aldehydes. Apparently only the free-living *Polynucleobacter* strain has the genetic potential to assimilate organic compounds such as alcohols and aldehydes, which can be converted in carboxylic acids. Indeed we have found different alcohol dehydrogenases (Fe-dependent, Zn-dependent, short chain dehydrogenase) and an aldehyde dehydrogenase.

Nitrogen containing compounds. Some nitrogen containing compounds can be used as additional carbon and nitrogen sources. The free-living bacterium has the possibility to degrade amines and nitroalkanes through an amine oxidase and a nitronate monoxygenase respectively. These two enzymes remove the nitrogen containing group producing the relative aldehyde. The hydrolysis of acetamide and formamide is also possible only in the free-living *Polynucleobacter* thanks to an acetamidase/formamidase.

Aromatic compounds. Finally a series of enzymes involved in the aromatic compounds degradation pathways were found only in the free-living strain's genome. In particular we identified several enzymes of the 3-oxoadipate pathway, especially of the ortho-cleavage of catechole, protocatechuate and chlorobenzoate, three putative nitroreductase genes (responsible of the reduction and the degradation of nitroaromatic compounds), an arylesterase and a benzoylformate decarboxylase (involved in the benzoate degradation pathway). The 3-oxoacid-

CoA transferase discussed above could be involved in the late steps of this pathway. Nevertheless it was not possible to identify a complete degradation pathway.

Amino Acids. *Biosynthesis.* Ammonia assimilation is achieved through the concerted activity of glutamate synthase (GOGAT) and glutamine synthase (GS) in both the symbiotic and free-living strain. A monomeric ferredoxin-dependent GOGAT is present in both strains while a dimeric NADPH-dependent glutamate synthase is present only in the free-living strain.

Aspartate transaminase is apparently missing in *Polynucleobacter necessarius*, a feature shared by other bacteria in the *Burkholderiaceae* family (in our screening of the available genomes, we found this enzyme only in *Burkholderia phenoliruptrix* BR3459a, *Burkholderia phytofirmans* PsJN and *Ralstonia solanacearum* GMI1000); one possibility is that aspartate ammonia-lyase, usually a degradation enzyme, is used anabolically as in *Rhizobium lupini*[60]. The symbiont lacks even this enzyme. Alternatively, aspartate could be produced through aromatic-amino-acid aminotrasferase, which has a significant role in aspartate formation in many bacteria[61]. Asparagine synthetase is missing, but the asparagine required for protein synthesis is available through modification of aspartyl-tRNA^{Asn} due to the presence of the genes *gatA* and *gatB*. This indirect pathway is found in most bacteria, but it usually coexists with the standard one[62]. Instead, in *Polynucleobacter* it seems to be the only possibility for the synthesis of the asparaginyl-tRNA^{Asn}, a feature also observed in *Deinococcus radiodurans* and *Thermus thermophilus*[63]. Methionine biosynthesis through the transsulfuration pathway is not possible because of the lack of one required enzyme in the free-living strain (cystathionine gamma-synthase) and two in the symbiont (missing cystathionine beta-lyase, too). The presence of all the genes for the biosynthetic pathway of homoserine, plus a homolog of O-succinylhomoserine sulfhydrylase (*metZ*), suggests the possibility that homocysteine is produced through direct sulfhydrilation of O-acetyl-L-homoserine, as seen in *Pseudomonas putida*[64].

The differences between the symbiont and the free-living strains' genomes are that the symbiont has lost one enzyme necessary for cysteine biosynthesis (serine O-acetyltransferase) and one for serine biosynthesis (D-3-phosphoglycerate dehydrogenase). It also lacks glutathione synthase, although apparently maintains other enzymes necessary for glutathione biosynthesis and those related to glutathione-dependent detoxification.

Degradation. When compared with the other analyzed members of the *Burkholderiaceae* family, both bacteria have extremely reduced amino acid degradation pathways. A common

feature shared with the bacteria belonging to *Burkholderiaceae* is the loss of lysine, tryptophan and methionine degradation pathways (with the only exception of *B. phenoliruptrix* BR3459a, which has a cystathionine gamma-lyase for methionine degradation). Due to the lack of the branched-chain α-ketoacid dehydrogenase complex, the degradation pathways of valine, leucine and isoleucine seem to be absent in many bacteria of the *Burkholderiaceae* family as well as in *P. necessarius* (complete degradation pathways were found only in *B. pseudomallei* K96243, *B. thailandensis* E264, *B. mallei* ATCC 23344, *B. cenocepacia* HI2424, *B. ambifaria* AMMD, *B. glumae* BGR1 and *B. gladioli* BSR3). A few enzymes catalyzing the last reactions of valine, leucine, isoleucine and lysine degradation pathways are retained in the free-living strain probably because they are also involved in carboxylic acid degradation and assimilation processes.

In addition, the degradation pathways for cysteine, proline, arginine, histidine, phenylalanine and tyrosine are missing in both *P. necessarius* strains. Only in a few cases those degradation pathways are lost in other members of the *Burkholderiaceae* family (*Burkholderia rhizoxinica* HKI 454 can not degrade arginine, histidine, phenylalanine and tyrosine and representatives of the genera *Cupriavidus* and *Ralstonia* can not degrade arginine).

Alanine can be catabolized to pyruvate in the free-living strain through alanine-glyoxylate transaminase, alanine transaminase and alanine dehydrogenase. The symbiotic bacterium lacks all three enzymes.

Aspartate can be oxidized to oxalacetate in both the free-living strain and the symbiont thanks to aspartate oxidase (NadB), an enzyme which is involved mainly in NAD^+ biosynthesis and which is present in all the *Burkholderiaceae* genomes analyzed. Moreover, only in the free-living strain, the presence of the aspartate ammonia-lyase enables the degradation of aspartate to fumarate. The symbiont has lost asparaginase, thus it can not degrade asparagine.

As both bacteria lack glutaminase, glutamine can be converted to glutamate only trough the reaction of glutamate synthase. Furthermore, as they are both missing glutamate dehydrogenase, glutamate can be converted to 2-oxoglutarate only through reactions involving glutamate as an amino-group donor (*e.g.* aromatic amino acid aminotransferase, acetyl-ornithine aminotransferase, phosphoserine aminotransferase, branched-chain amino acid aminotransferase).

Both bacteria lack a complete glycine cleavage system, but glycine can be converted to serine thanks to a glycine hydroxymethyltransferase. Threonine dehydratase, which is present in both strains, can degrade serine and threonine to pyruvate and 2-oxobutyrate respectively. The following oxidation of 2-oxobutyrate to propionyl-CoA may be performed in the free-living strain but probably not in the symbiont because the latter lacks a 2-oxoacid dehydrogenase activity. Moreover, as explained before, the ability of catabolize propionyl-CoA is uncertain in the symbiont. Due to the lack of a complete glyoxylate cycle in the symbiont, the carbon skeletons of glycine and serine cannot be used for gluconeogenesis and, therefore, as sole carbon source. Nevertheless these two amino acids might be used as additional energy and carbon sources.

In summary, although only aspartate, glutamate and glutamine could be considered glucogenic in the symbiont, it is unlikely that they are effectively used by the bacterium as carbon and energy sources because their conversions to oxalacetate and 2-oxoglutarate respectively seem to be side reactions in other biosynthetic pathways. On this point, the situation appears very different from that of "*Candidatus* Glomeribacter gigasporarum" (*Burkholderiaceae*), the obligate endobacterium of an arbuscolar mycorrhizal fungus, where serine, glutamate, aspartate and arginine can be used by the symbiont as carbon and energy sources[22]. On the other hand, the amino acids that have the potential to be used as sole carbon sources in the free-living *P. necessarius* are alanine, glycine, serine, threonine, aspartate, asparagine, glutamate and glutamine.

Cofactors. Both the free-living and symbiotic strain have all the genes necessary for S-adenosylmethionine (SAM), polyamines, haem, NAD^+ and $NADP^+$ biosynthesis. Putrescine is probably synthesized from ornithine after decarboxylation through a putative ornithine/lysine/arginine decarboxylase. Differently from the symbiont, the free-living strain can additionally synthesize thiamine pyrophostate (TPP), FAD, CoA, molibdopterin and folate. The symbiont lacks only the last two enzymes for FAD biosynthesis (riboflavine kinase and FAD synthetase), it needs pantotenate or pantoic acid in order to make CoA (this feature is shared with "*Ca.* G. gigasporarum"[22]) and it retains only the enzymes for the folate-mediated one-carbon metabolism.

Apparently, both bacteria are unable to synthesize vitamin B12 and siroheme even if they still possess a few genes belonging to this biosynthetic pathway. The free-living strain's genome

lacks two key genes (*epd* and *pdxB*) and the symbiont four (*epd*, *pdxB*, *pdxA* and *pdxJ*) involved in the biosynthesis of B6 vitamers. The biotin biosynthetic pathway is not complete in the free-living strain and is completely lost in the symbiont.

Our genomic analysis seems to be partially in contrast with the data of Hahn and coworkers[19] which shows that the free-living strain QLW-P1DMWA-1 can grow on acetate as sole carbon and energy source in an inorganic medium exclusively supplemented with vitamin B12.

Additional notes. As expected, there are no genes involved in flagellar assembly or motility in either genome. On the other hand, the free-living *P. necessarius* has genes usually related to aggregation, challenging the notion that this organism is strictly planktonic. These include a putative cellulose synthase – whose role is not entirely understood in bacteria, but whose presence correlates with the formation of biofilms[65]; and diguanylate cyclase, producing the messenger molecule c-di-GMP, which is often involved in triggering and maintaining aggregation[66]. These genes were lost in the symbiont.

Another exclusive, and intriguing, feature of the free-living strain is the presence of cyanophycin synthetase. Cyanophycin is a polymer of aspartate and arginine employed for nitrogen storage, typical of cyanobacteria[67].

Additional References

54. Petroni G, Dini F, Verni F, Rosati G. (2002) A molecular approach to the tangled intrageneric relationships underlying phylogeny in *Euplotes* (Ciliophora, Spirotrichea). Mol Phylogenet Evol. 22: 118-130
55. Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, *et al.* (2010) GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. Nat Methods. 7: 455-457
56. Darling ACE, Mau B, Blattner FR, Perna NT. (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14: 1394-1403
57. Arias-Barrau E, Olivera ER, Sandoval A, Naharro G, Luengo JM. (2006) Acetyl-CoA synthetase from *Pseudomonas putida* U is the only acyl-CoA activating enzyme induced by acetate in this bacterium. FEMS Microbiol Lett. 260: 36-46
58. Hahn MW, Lang E, Brandt U, Wu QL, Scheuerl T. (2009) Emended description of the genus *Polynucleobacter* and the species *Polynucleobacter necessarius* and proposal of two subspecies, *P. necessarius* subsp. *necessarius* subsp. nov. and *P. necessarius* subsp. *asymbioticus* subsp. nov. Int J Syst Evol Microbiol. 59: 2002-2009
59. Lindenkamp N, Schürmann M, Steinbüchel A. (2013) A propionate CoA-transferase of *Ralstonia eutropha* H16 with broad substrate specificity catalyzing the CoA thioester formation of various carboxylic acids. Appl Microbiol Biotechnol. 97: 7699-7709

60. Kretovich WL, Kariakina TI, Weinova MK, Sidelnikova LI, Kazakova OW. (1981) The synthesis of aspartic acid in *Rhizobium lupini* bacteroids. Plants & Soil. 61: 145-156
61. Umbarger HE. (1978) Amino acid biosynthesis and its regulation. Annu Rev Biochem. 47: 532-606
62. Sheppard K, Yuan J, Hohn MJ, Jester B, Devine KM, Söll D. (2008) From one amino acid to another: tRNA-dependent amino acid biosynthesis. Nucleic Acids Res. 36: 1813-1825
63. Min B, Pelaschier JT, Graham DE, Tumbula-Hansen D, Söll D. (2002) Transfer RNA-dependent amino acid biosynthesis: an essential route to asparagine formation. Proc Natl Acad Sci USA. 99: 2678-2683
64. Alaminos M, Ramos JL. (2001) The methionine biosynthetic pathway from homoserine in *Pseudomonas putida* involves the *metW*, *metX*, *metZ*, *metH* and *metE* gene products. Arch Microbiol. 176: 151-154
65. Römling U. (2002) Molecular biology of cellulose production in bacteria. Res Microbiol. 153: 205-212
66. Tamayo R, Pratt JT, Camilli A. (2007) Roles of cyclic diguanylate in the regulation of bacterial pathogenesis. Annu Rev Microbiol. 61: 131-148
67. Oppermann-Sanio FB, Steinbüchel A. (2002) Occurrence, functions and biosynthesis of polyamides in microorganisms and biotechnological production. Naturwissenschaften. 89: 11-22

PART III

Characterization of Ciliates

Ciliates are relatively well studied, compared to other protist groups. They have been observed since the invention of the microscope and are relatively large, often rich in morphological features guiding identification, common and sometimes able to grow under laboratory conditions. Nevertheless, there is still a tremendous amount of things we do not know about them. And in order to explore the gaps, like in all other biological areas, we have to start with classification.

The literature on ciliate taxonomy offers a lot of examples of subtle and complex problematics. First of all, there is no consensus on their expected biodiversity. This is mostly due to the absence of a good species concept that can be unambiguously applied (ciliates possess a form of sexual interaction called conjugation, but mating type systems are often complicated, and performing reliable experiments is tricky and time-consuming). In practice, what is commonly used is a vague morphological criterion. It is vague because there is no agreement on the difference threshold separating species, let alone higher taxa. The result is that the classification of different groups is mostly settled by leading experts of each particular lineage, that are forced to choose independent, "promising" sets of diagnostic features.

A second series of problems concerns the quality of descriptions. It is true that in the last decade or so some standards have emerged, setting the bar for characterizations and the establishment of new taxa. Nevertheless, the sheer number of newly described species entails that not all publications are equally detailed, nor completely comparable. However, the real confusion comes from older descriptions. Ciliates are notorious for a (perceived?) paradox: more than 3,000 species are "known", but only a handful are commonly found, identified and reported. The vast majority of existing binomial names were assigned either very recently or a long time ago, and all kind of contradictions arise. Nomenclatural rules are strict (for good reasons) when it comes to change, use or dismiss old names, and the outcome is that researchers oftentimes tend to ignore them. It is indeed easier to establish a new species than to revise an old one.

Molecular data are now an intrinsic part of the characterization process. The consequent advantages are huge, especially in terms of our understanding of phylogeny and a stark improvement in the comparability of studies. But new issues also became apparent: what should we do when morphology and molecular data do not match? In ciliates, this is largely a group-specific problem, with well-behaving classes and families as well as incredibly chaotic situations.

Concluding this short review on the hindrances that vex ciliate taxonomists, I must cite sibling species complexes. Even if we can find a way to separate all morphotypes, what if some of them are not good biological objects at all? What if their variability is hidden, and can only be detected by fine-tuned characters (such as fast-evolving molecular markers)? This is clearly no idle nor theoretical questions in ciliates, because we know that some of the most studied morphospecies are actually composed of largely undistinguishable – but genetically isolated – lineages. What can we say about the thousands of species that are *not* that well studied?

And yet, identification (guided by a good classification) is utterly essential in order to ask any biological question – be it about physiology, ecology, biogeography, anthropocentric applications… or symbiosis. Aside from the laboratories using clonal long-living cultures of model strains, all biologists working with ciliates must first address systematic issues.

My works on ciliates focused on this topic. While no magic solution exists to overcome the obstacles listed above, it is consistently reported that large amounts of data tend to slowly converge to a (more) stable resolution. It is however true that not all studies can include a long and onerous screening phase; thus, I consider it the role of taxonomists to perform detailed surveys, draw conclusions, organize the revised framework in a concise fashion and highlight the remaining issues, so that researchers less interested in untangling taxonomic knots can use previously tested tools. The projects described in the next chapters revolve around surveying existing taxa, instead of naming new ones; accumulating as much data as possible on narrow systematic groups; and providing a thorough review of existing literature, trying to combine it with the newly obtained results in new paradigms.

As an obvious consequence of my intended goals, the surveys are directed at relatively neglected taxa with a confusing literature. For similar reasons, the approaches are multidisciplinary, addressing the morphological as well as the molecular level.

Chapter 7 deals with a relatively obscure species in the most celebrated ciliate genus, *Paramecium*. *Paramecium duboscqui* is a good example of the risk involved in generalizations, since it displays large differences compared to its famous cousin *P. aurelia*. This work provided

the first molecular survey of this species and general information on its variability and patterns – including the possible presence of sibling species.

Chapter 8 includes a multidisciplinary redescription of an entire genus, *Spirostomum*. Although including large and common species, *Spirostomum* presents all the issues listed at the beginning of this section. On the other side, it proved to be reasonably pliant to careful analyses, and the survey was able to answer many questions and settle most ambiguities, producing what I think is a good identification tool for any non-specialist.

In addition to the projects reported here in detail, I played a minor part in a third ciliate characterization work. Though sharing the same goal of the other papers, the redescription of *Sonderia vorax* (Modeo et al., 2013 – see Appendix) had a more specific target, a single population of this elusive, anaerobic and endosymbiont-bearing species. And yet, the amount of data collected – mostly by Dr. Letizia Modeo and Profs. Giovanna Rosati and Sergei Fokin – was unusually large, especially in the detailed description of ultrastructural features, an aspect of ciliate studies that is dwindling as fast as short characterizations of new species are accruing.

Chapter 7.

Paramecium duboscqui

Original publication:

Survey of *Paramecium duboscqui* using three markers and assessment of the molecular variability in the genus *Paramecium*

Vittorio Boscaro, Sergei I Fokin, Franco Verni, Giulio Petroni

Molecular Phylogenetics and Evolution 65(3): 1004-1013. December 2012.

Abstract

The genus *Paramecium* (phylum Ciliophora) is one of the best-known among protozoa. Nevertheless, the knowledge on the diversity and distribution of species within this genus was remarkably scarce until recent times. In the last years a constantly growing amount of data has formed, especially on the distribution of species and the characterization of molecular markers. Much effort has been made on detecting clades inside each morphospecies, which could suggest the presence of sibling species complexes as in the famous case of *Paramecium aurelia*. In this work we present new data on *Paramecium duboscqui*, one of the morphospecies that have not yet been surveyed employing DNA sequences as markers. We obtained data from nine strains sampled around the world, using the three most commonly employed markers (18S rRNA gene, ITS1-5.8S-ITS2 and COI gene sequences). Moreover, we compared our results with those already available for other *Paramecium* species, and performed phylogenetic analyses for the entire genus. We also expanded the knowledge on the ITS2 secondary structure and its usefulness in studies on *Paramecium*. Our approach, that considers the data of all the species together, highlighted some characteristic patterns as well as some ambiguities that should be further investigated.

Introduction

The genus *Paramecium* (Oligohymenophorea, Peniculida) includes some of the most studied protozoan species; some of them are among the few model organisms belonging to the phylum Ciliophora. More than 40 morphospecies can be found in literature, but only 17 are properly described[1,2,3]. Of even less (12) there is at least one DNA sequence available. Sorting them in accordance with the classification proposed by Fokin and coworkers[2], these are: *Paramecium bursaria* (subgenus *Chloroparamecium*), *Paramecium duboscqui* and *Paramecium putrinum* (subgenus *Helianter*), *Paramecium calkinsi*, *Paramecium nephridiatum*, *Paramecium polycaryum* and *Paramecium woodruffi* (subgenus *Cypriostomum*), *Paramecium aurelia*, *Paramecium caudatum*, *Paramecium multimicronucleatum*, *Paramecium jenningsi* and *Paramecium schewiakoffi* (subgenus *Paramecium*)*. *P. aurelia* is actually a complex of fifteen biological species, where each syngen (group of compatible mating types) received a binomial name[4,5]. For other morphospecies the existence of sibling species complexes has been sometimes suggested (in *P. multimicronucleatum*[6] and *P. jenningsi*[7]) but never formalized, even in those cases where the existence of multiple syngens was demonstrated long ago, like in *P. bursaria*[3]. Although we will use here the term "sibling species" as synonymous of "biological species undistinguishable according to morphological characters", some authors employed different operative definitions[8,9].

Not all the *Paramecium* species have been equally studied. For example, a research on the Web of Knowledge[SM] for "Paramecium tetraurelia" gave 1,165 results, while the sum of all results provided for the species in the *Helianter* and *Cypriostomum* subgenera barely reaches 100**.

Some aspects of *Paramecium* biology in the natural environment are still poorly known; this is rapidly changing thanks to a recent increase of available data, especially molecular ones, on many of the aforementioned morphospecies. While the method of RAPD fingerprinting has been successfully employed starting from the late '90s (on *P. bursaria*[10], *P. duboscqui*[11], *P.*

---

* The 18S rRNA gene sequence of another long forgotten morphospecies, *Paramecium chlorelligerum*, was obtained shortly after this survey (see the Chapter 7 Afterword)

** The trend continues. Using the same parameters, I could find only 7 new papers about these subgenera published since this work appeared, against 190 dealing with *P. tetraurelia* alone

nephridiatum[12], *P. caudatum*[13], *P. jenningsi*[7], *P. aurelia* complex[14,15]), most of the recent studies opted for the sequencing of multiple DNA fragments as molecular markers (on *P. bursaria*[16], *P. calkinsi*[17], *P. multimicronucleatum*[18,19], *P. caudatum*[18], *P. aurelia* complex[20-22]). The latter approach provides results that are more informative and easier to compare, at the expense of lengthier work times. Two common DNA markers for intraspecific molecular surveys are the nuclear region including the internal transcribed sequences 1 and 2 and the gene for 5.8S rRNA (here labeled as ITS1-5.8S-ITS2), and part of the mitochondrial gene encoding the cytochrome *c* oxydase subunit I (COI). The gene for 18S rRNA, that is widely used for high-ranks taxonomy, has been generally neglected for this purpose; one exception is the work of Hosina and coworkers[16] on *P. bursaria*, that showed an unexpectedly high variability of this marker between different strains.

Most of the recent papers employed two markers and investigated strains from different locations. They addressed similar questions: the amount of molecular variability, the possible existence of sibling species complexes and the detection of biogeographic patterns. Especially the last matter is the subject of a great debate concerning the distribution of ciliates and protists in general[23,24].

In this work, we have performed the first molecular survey based on DNA sequences of the morphospecies *P. duboscqui* Chatton and Brachon 1933. After its original description, for a long time this taxon was considered very rare, or even non-valid by some authors[3,25]; it was properly described again by Shi and coworkers[26]. *P. duboscqui* has the truncated body shape of the "*bursaria*" type and a body length of about 80-150 µm[27]. It has an ellipsoidal macronucleus and one-eight (usually two) vesicular micronuclei. Cells swim in a right-handed spiral. Mating experiments performed on Chinese strains by Watanabe and coworkers[28] and Shi and coworkers[26] revealed only one syngen and two mating types; the same result was obtained on strains from Kunashir Island by Fokin and coworkers[11]. *P. duboscqui* sometimes contains bacterial endosymbionts[29-31], although less often than other species[32]. A first multidisciplinary survey on its distribution and variability was carried on by Fokin and coworkers[11]. This morphospecies can be usually found in brackish water sites, and shows a preference for low temperatures, from less than 10 °C to about 20 °C[11,26].

We obtained from nine strains of *P. duboscqui* sampled around the world the sequences of three molecular markers: nuclear 18S rRNA gene and ITS1-5.8S-ITS2 sequence, and

mitochondrial COI gene. We assessed on the base of these markers the amount of molecular variability within the morphospecies, the presence of distinct clades and their distribution. Our results were compared with previously published data on other species, and phylogenetic inferences based on each marker were performed on the entire genus. A comparative analysis of the secondary structure of ITS2 was also performed, expanding the results of Coleman[20]. We could thus produce a critical evaluation of the available data and derive more general conclusions on the molecular variability inside the genus, highlighting some ambiguities and providing suggestions for future researches.

Material and Methods

***P. duboscqui* sampling, identification and culture.** The nine *Paramecium* strains investigated in this work were sampled in different geographic locations (see Table 7.1). They were all classified as *P. duboscqui* on the basis of morphological diagnostic characters according to the first description[27] and the key of Fokin[1]. We couldn't find any morphological trait enabling us to distinguish them from one another.

The monoclonal strains were obtained isolating single *Paramecium* cells from the original populations. These cells were briefly washed several times in sterile distilled water, and then grown at 20 °C in artificial brackish water (salinity 5‰). The green alga *Dunaliella tertiolecta* was employed as food.

DNA extraction. After the establishment of each monoclonal strain, 50-70 cells were harvested with a glass micro-pipette, washed several times in sterile distilled water and stored in ethanol 70% at -22 °C. Total DNA was extracted from the stored cells using the NucleoSpin™ Plant II DNA extraction kit (Macherey-Nagel). All polymerase chain reactions (PCRs) were performed in a C1000™ Thermal Cycler (Bio-Rad) with the TaKaRa Ex *Taq* (TaKaRa Bio Inc.) polymerase on 2 µL of thawed DNA solution.

Sequences obtainment. The nuclear DNA region including the nearly complete sequence of the 18S rRNA gene and the complete ITS1-5.8S-ITS2 sequence was characterized through PCR amplification and direct sequencing. Two PCRs were performed, the first with primers matching sequences on the 18S rRNA gene (forward 18S F9 Euk 5'-CTGGTTGATCCTGCCAG-

3'[33] and reverse 18S R1513 Hypo 5'-TGATCCTTCYGCAGGTTC-3'[34]) and the second one with the forward primer matching a sequence on the 18S rRNA gene and the reverse primer matching a sequence on the 28S rRNA gene (18S F783 5'-GACGATCAGATACCGTC-3'[35] and RGD2 5'-GGTCCGTGTTTCAAGACGGG-3' respectively). Sequencing was performed in both directions with the following primers: 18S R536 5'-CTGGAATTACCGCGGCTG-3'[35], 18S F783, 18S R1052 5'-AACTAAGAACGGCCATGCA-3'[35], FG1400 5'-TTGYACACACCGCCCGTC-3' and RGD2.

The COI gene sequences were obtained through a two-steps approach. A first PCR amplification was performed with the primers forward COI_F543SKmP and reverse COI_R1321SKmP (5'-**TAATACGACTCACTATAGGG**GGTTTTGCTAATTTTTTAATTCCTTATCATGTTGGWKCBAAAGATGTWGC-3' and 5'-**CAGGAAACAGCTATGAC**GTTGGTATAATTATAATATADACYTCAGGGTGACCRAAAAATCA-3' respectively; modified from Strüder-Kypke and Lynn[36]; sequences in bold correspond to those of the universal primers T7 and M13r and were added to optimize the sequencing procedure). Semi-nested PCRs were then performed on the products following the approach of Andreoli and coworkers[37]. The first one employed the primers forward T7 and reverse COI_PdubR1088 (5'-GTTCTTCTAGTAATTAATAAATTMGTAAAAG-3'), and its product was sequenced with the primer T7. The second semi-nested employed primers forward COI_PdubF922 (5'-TCAAGAAGAMGAAAAATWTATTTTAGTAAATGTTC-3') and reverse M13r, and its product was sequenced with the primer M13r.

Electropherograms were visualized using Chromas Lite (Technelysium), compared and assembled in order to obtain the complete sequence.

Sequences availability, comparison and phylogenetic analysis. The characterized sequences are available from the GenBank database under the GenBank IDs listed in Table 7.1.

A preliminary screening was performed using BLASTN[38]. A partial 18S rRNA gene sequence of the *P. duboscqui* strain IG2-1[39], and the COI sequences of strains IG2-1[36] and TuAI-2[40] were already available. We repeated the sequencing in order to confirm previous results. The 18S rRNA gene sequence of the strain Ku4-8 was also available[31]. This strain is now extinct, so we could not obtain its ITS1-5.8S-ITS2 and COI gene sequences.

The 18S rRNA gene sequences were first aligned against more than 3,500 ciliates' sequences from the SILVA 108 database[41] with the Fast Aligner function of the ARB software package[42]. The alignment of all sequences from organisms belonging to the order Peniculida was then manually edited in order to optimize base-pairing in the predicted stem regions.

ITS1-5.8S-ITS2 sequences were aligned with ClustalW[43] together with those already available for other peniculids. The secondary structures of ITS2 were predicted according to Coleman[20,44,45] using the outputs of the mfold website (http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form[46]). Differences between two ITS2 sequences occurring in predicted stem regions were defined as Compensatory Base Changes (CBCs) if occurring on both strands or hemi-CBCs if one-sided[44].

Alignment of COI gene sequences was performed with Fast Aligner on their predicted aminoacidic translation.

All phylogenetic analyses were carried out with both Maximum Likelihood (ML) and Bayesian Inference (BI) methods. The software PHYML[47] as provided by ARB, and MrBayes 3.1.2[48] were employed. Bootstrap analyses with 1,000 pseudoreplicates were performed for the ML method. Three runs, with one cold and three heated chains each, ran for 1,000,000 generations in BI analyses. The ARB NJ algorithm[42] with the "similarity" correction was employed in order to calculate the similarity matrices. Sequence lengths were reduced to that of the shortest one available in each dataset. No pair of identical sequences was included. Gaps were coded as a fifth character. The model of nucleotide evolution that fits best the data was chosen according to jModelTest[47,49]; the Akaike Information Criterion selected the GTR+I+G model (using a discrete, four-categories gamma function) in all analyses. The likelihood mapping function of TREE-PUZZLE[50] was employed in order to check the informational content of each character matrix.

All the comparisons, as well as all site positions given, refer to the same character matrices employed in phylogenetic analyses.

Table 7.1 List of *Paramecium duboscqui* strains employed in this work

Strain	Sampling location	18S haplotype name and ID	ITS1-5.8S-ITS2 haplotype name and ID	COI haplotype name and ID	Additional references
702	Harbin, China	B; HE819871	B; HE819871	bI; HE819880	[11]
BB8	Baltic Sea, Beryozovye Islands, Russia	B; HE819872	B; HE819872	bII; HE819881	[11]
AWH9-4	Atlantic Ocean, Woods Hole, USA	A; HE819873	A1; HE819873	aII; HE819882	[11]
IG2-1	Mediterranean Sea, Giglio Island, Italy	A; HE819874[b]	A1; HE819874	aI; FJ905146[b]	[11,36,39]
IN05	Mediterranean Sea, Naples, Italy	A; HE819875	A2; HE819875	aI; HE819883	-
Pd-2	Mediterranean Sea, Pisa, Italy	A; HE819876	A1; HE819876	aI; HE819884	-
Ppd-3	Mediterranean Sea, Procida Island, Italy	A; HE819877	A2; HE819877	aI; HE819885	-
Tub2	Mediterranean Sea, Orbetello Lagoon, Italy	A; HE819878	A1; HE819878	aI; HE819886	-
TuAI-2	Tozeur, Tunisia	A; HE819879	A2; HE819879	aI; FN421332[b]	[40]
Ku4-8	Pacific Ocean, Kunashir Island, Russia	C; AM236094[a]	-	-	[11,31]

[a] Data already available

[b] Data already available and checked in this study

Results

18S rRNA gene sequences. The sequence of strain IG2-1 obtained in this study differs by one base (site 929) from the one previously published (GenBank ID: AF100312). The electropherogram we obtained was of good quality, and showed no sign of a double peak in this marker. Moreover, this site contains an adenine in all Peniculida as well as in our sequence, and a thymine only in the sequence AF100312. On this basis, we considered likely that sequence AF100312 contained an error and used ours (GenBank ID: HE819874) in further analyses.

The 18S rRNA gene sequences of *P. duboscqui* strains AWH9-4, IG2-1, IN05, Pd-2, Ppd-3, TuAI-2 and Tub2 (group A) are identical, as are those of strains 702 and BB8 (group B). Sequences of the two groups differ in two sites (552 and 962). The homologous sequence of strain Ku4-8 is more divergent and represents group C; it shows an average of 17.5 different sites with groups A and B. The different bases are mostly concentrated in the regions 543-641 and 956-973, that are the least conserved in the genus.

The ML phylogenetic tree based on 18S rRNA gene sequences is shown in Figure 7.1. Groups A and B are more closely related to each other inside the *P. duboscqui* clade. The subgenera *Paramecium* and *Cypriostomum* are monophyletic and sister groups. The phylogenetic relationships between *P. duboscqui*, *P. putrinum* (subgenus *Helianter*) and *P. bursaria* (subgenus *Chloroparamecium*) are less supported by statistical values. The analysis suggests, however, that *Helianter* is not monophyletic because *P. putrinum* is more closely related to *P. bursaria*.

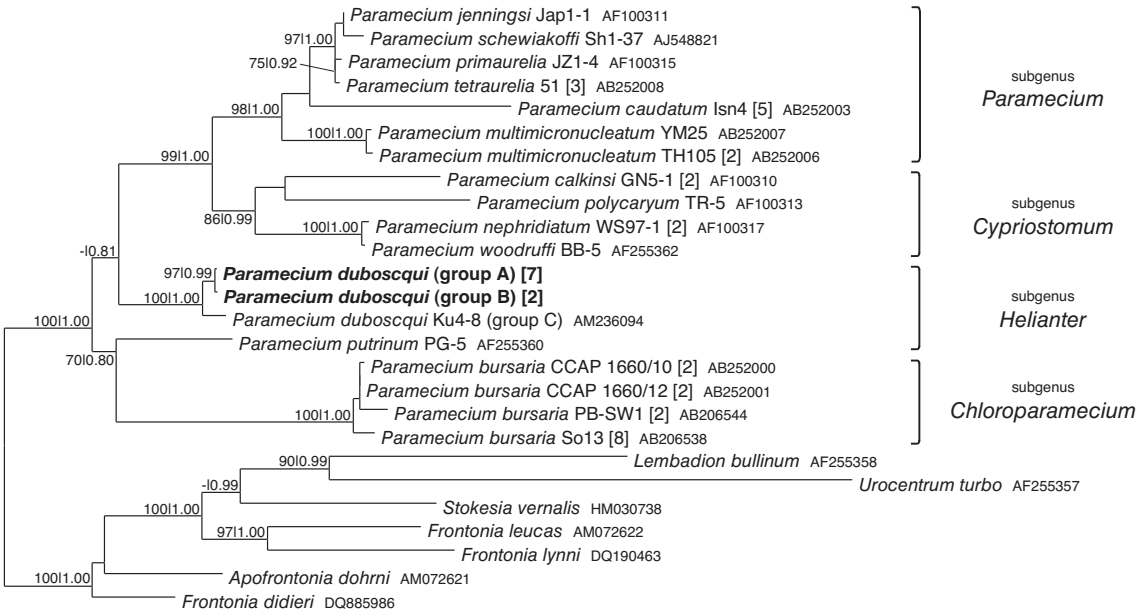


Figure 7.1 ML phylogenetic tree of genus *Paramecium* based on 18S rRNA gene sequences. The character matrix included 26 sequences and 1672 sites; only 2.90% of characters were gaps. Nucleotide frequency estimates were: 28.1% (A), 19.0% (C), 25.6% (G) and 27.4% (T), conserved among all sequences (p>>0.05). 94.1% of 100,000 randomly chosen quartets had well-defined topologies. The root was placed between *Paramecium* and other Peniculida. The numbers associated to each node represent bootstraps and Posterior Probability (values below 70|0.80 are not shown). Sequences in bold were characterized in this study. The number of sequences in GenBank that share the same haplotype is shown in brackets (if higher than one). The bar stands for an estimated genetic distance of 0.1.

The percentage of identical sites between haplotypes A and B is 99.88% (2 different sites), that between A and C is 98.98% (17 different sites) and that between B and C is 98.92% (18 different sites). For comparison, the similarity between *P. primaurelia* and *P. tetraurelia* of the *aurelia* complex is 99.70% (5 different sites). Intraspecific similarity values for the only two

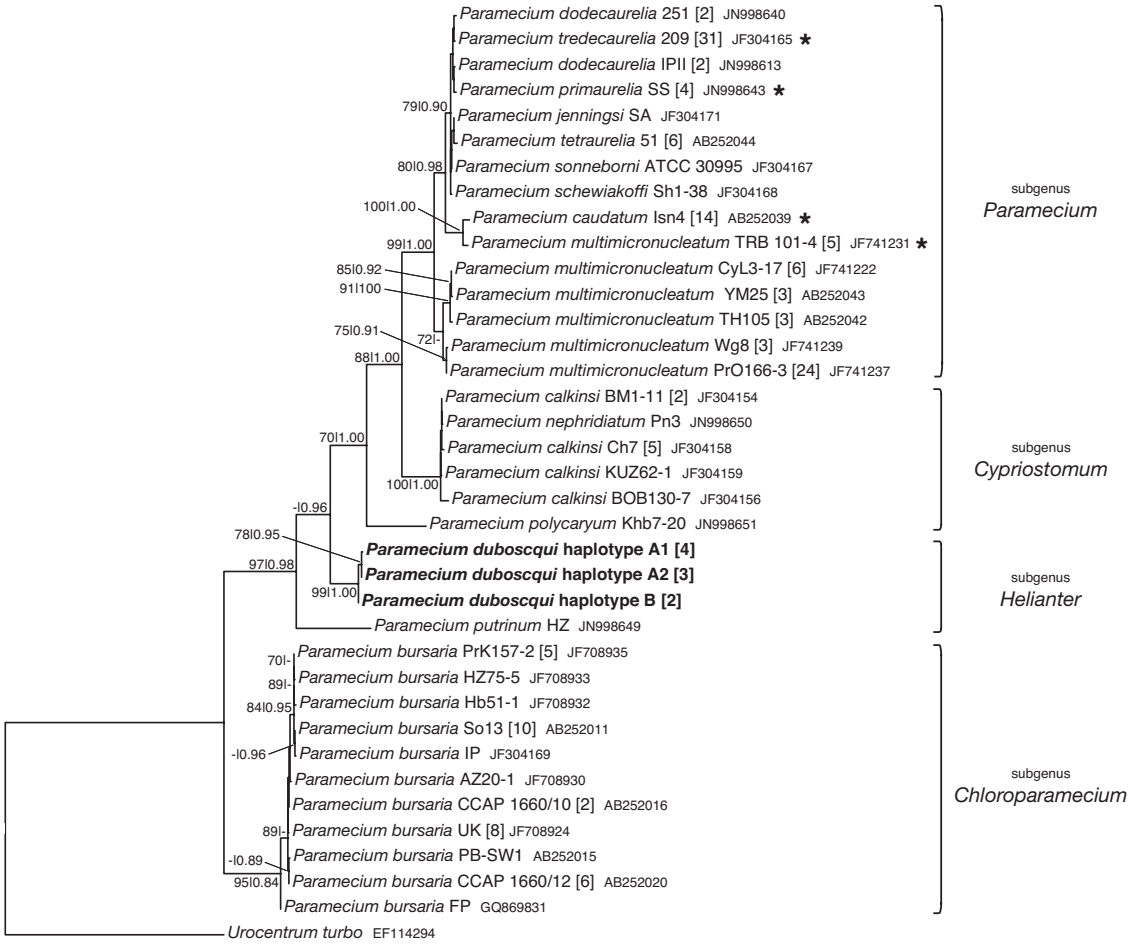


Figure 7.2 ML phylogenetic tree of genus *Paramecium* based on ITS1-5.8S-ITS2 sequences. The character matrix included 37 sequences and 493 sites; 10.20% of characters were gaps. Nucleotide frequency estimates were: 31.7% (A), 17.3% (C), 17.3% (G) and 33.7% (T), but *Paramecium polycaryum* and *Paramecium duboscqui* haplotype A2 sequences significantly deviated from these means ($p<0.05$). 93.3% of 100,000 randomly chosen quartets had well-defined topologies. The root was placed between *Paramecium* and *Urocentrum turbo*. The numbers associated to each node represent bootstraps and Posterior Probability (values below 70|0.80 are not shown). The number of sequences in GenBank that share the same haplotype is shown in brackets (if higher than one). Sequences in bold were characterized in this study. Asterisks mark discordant haplotypes, that are discussed in the text. The bar stands for an estimated genetic distance of 0.5.

other morphospecies for which more than one haplotype is available are 99.70% (*P. multimicronucleatum*, 5 different sites) and 98.39% (*P. bursaria*, 27 variable sites).

ITS1-5.8S-ITS2 sequences. Strains of *P. duboscqui* groups A and B differ in three sites (50 and 98 in ITS1 and 443 in ITS2). One additional polymorphic site (469) hosts a cytosine in

all sequences of group B and some sequences of group A (strains IN05, Ppd-3 and TuAI-2), and a thymine in the other sequences of group A (strains AWH9-4, IG2-1, Pd-2 and Tub2); thus, strains belonging to group A have two slightly different ITS1-5.8S-ITS2 haplotypes (A1 and A2).

The ML phylogenetic tree based on ITS1-5.8S-ITS2 is shown in Figure 7.2. The haplotypes of *P. duboscqui* group A strains cluster together, closely related to the single haplotype of group B. The subgenus *Paramecium* is monophyletic, but the sequence of *P. multimicronucleatum* strain TRB 101-4 doesn't cluster with conspecific sequences, and is more closely related to the single available haplotype of *P. caudatum*; moreover, the sequence of *P. multimicronucleatum* strain Kr 113-3 (GenBank ID: JF741226) is identical to that of *P. caudatum*, and thus was not included in this analysis. Some haplotypes are also shared by *Paramecium dodecaurelia* and other sibling species of the *aurelia* complex, namely *P. primaurelia* and *Paramecium tredecaurelia*. The single *P. polycaryum* sequence available doesn't cluster with others of the subgenus *Cypriostomum*, and *P. nephridiatum* strain Pn3 falls inside the cluster of *P. calkinsi* sequences. *P. duboscqui* and *P. putrinum* don't form a monophyletic group, and *P. bursaria* is the first morphospecies to branch off at the base of the genus.

Haplotypes A1 and A2 of *P. duboscqui* share 99.80% identity, and groups A and B an average of 99.29%. The identity thresholds among strains of other morphospecies vary from 96.75% (*P. bursaria*) to 97.97% (*P. calkinsi*) and 98.17% (*aurelia* complex). *P. multimicronucleatum* sequences share an identity of 97.77% or higher if those of strains TRB 101-4 and Kr 113-3 are excluded, and of 92.29% if they are included.

All the investigated species of the genus possess two conserved helices in their ITS2 secondary structure, as showed by Coleman[20] on a smaller set of species of the *Paramecium* subgenus. These are homologous to the helices II and III of the general eukaryotic ITS2 structure[20,44,45]. In addition, all species except *P. caudatum*, *P. multimicronucleatum* and those belonging to the *P. aurelia* complex show a smaller helix closer to the 5' end of the sequence.

Both *P. calkinsi* and *P. multimicronucleatum* strains show intraspecific hemi-CBCs and CBCs in helix II. *P. duboscqui*, *P. multimicronucleatum* and *P. aurelia* strains show intraspecific hemi-CBCs in helix III, but only that of *P. multimicronucleatum* is in the most conserved region of the helix; moreover, it only separates the anomalous strains TRB 101-4 and Kr 113-3 from strains of the main cluster. These anomalous strains have no CBCs nor hemi-

CBCs with respect to each other or *P. caudatum*. The single-base polymorphism differentiating groups A and B inside the morphospecies *P. duboscqui* is a hemi-CBC in helix III, while the polymorphism differentiating haplotypes A1 and A2 falls in a loop region (see Figure 7.3). There are no intraspecific CBCs in helix III in any *Paramecium* morphospecies. Interspecific CBCs are present in both helices, but only one falls in the conserved helix III region.

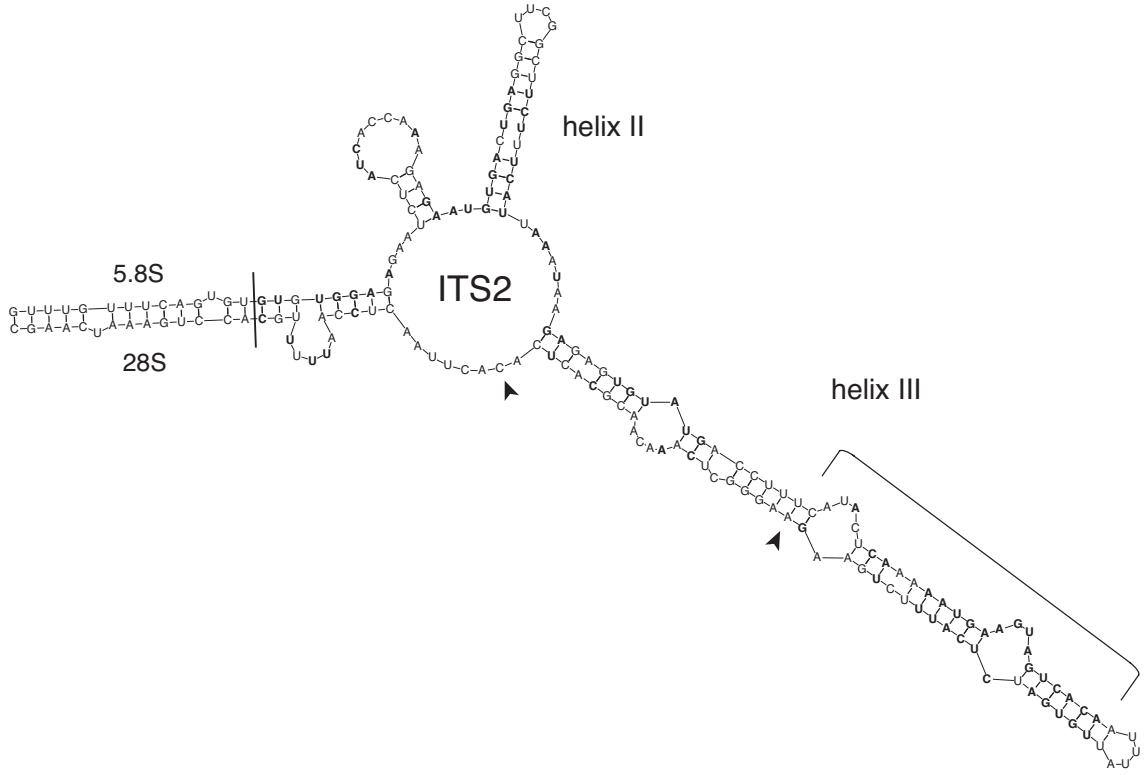


Figure 7.3 Predicted secondary structure of *Paramecium duboscqui* ITS2 (haplotype A2). The bracket highlights the most conserved region of helix III. Site conserved in all *Paramecium* sequences investigated are in bold. Arrowheads mark the polymorphic sites within *P. duboscqui*.

COI gene sequences. The obtained sequences of strains IG2-1 and TuAI-2 are identical to the homologous regions of those previously published (GenBank IDs: FJ905146 and FN421332 respectively) and were not resubmitted to public DNA databases.

P. duboscqui groups A and B differ by 40 bp on average. Inside each group, two haplotypes can be found. Strains BB8 and 702 of group B differ in two sites (120 and 255); the strain AWH9-4 presents one single-base difference with all the other strains of group A (site 12).

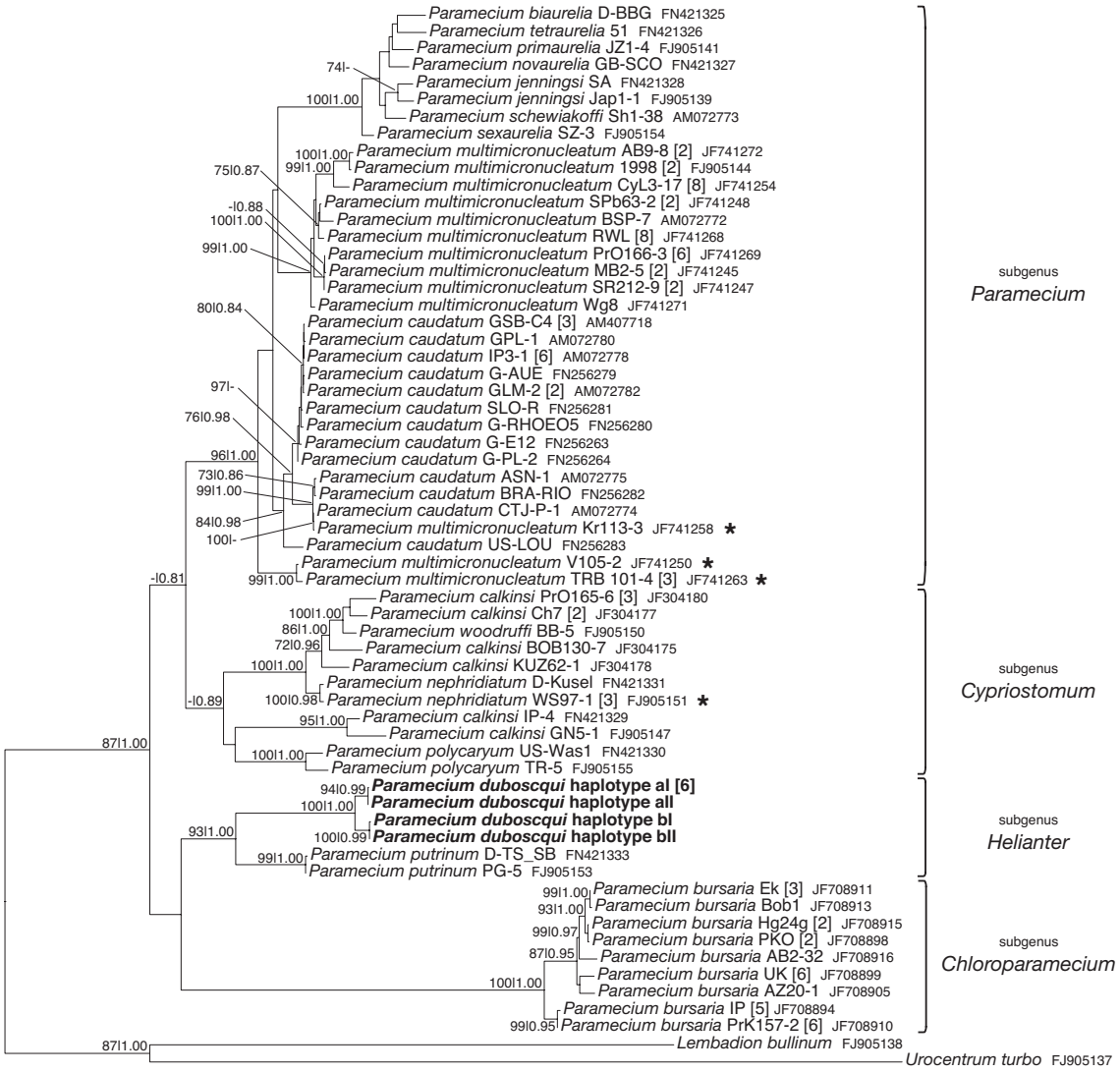


Figure 7.4 ML phylogenetic tree of genus *Paramecium* based on COI gene sequences. The character matrix included 62 sequences and 633 sites; only 1.22% of characters were gaps. Nucleotide frequency estimates were: 33.2% (A), 13.4% (C), 12.9% (G) and 40.5% (T), but 13 sequences significantly deviated from these means ($p<0.05$); *Paramecium aurelia*, *Paramecium jenningsi* and *Paramecium schewiakoffi* sequences are more balanced in their G/C contents, while *Paramecium putrinum* and some *Paramecium bursaria* ones are even more A/T rich. 85.6% of 100,000 randomly chosen quartets had well-defined topologies. The root was placed between *Paramecium* and other Peniculida. The numbers associated to each node represent bootstraps and Posterior Probability (values below 70|0.80 are not shown). Sequences in bold were characterized in this study. The number of sequences in GenBank that share the same haplotype is shown in brackets (if higher than one). Asterisks mark discordant haplotypes, that are discussed in the text. The bar stands for an estimated genetic distance of 0.5.

There are no intra-groups differences in aminoacid sequences, while 6 out of 211 aminoacids differ between the groups A and B.

The ML phylogenetic tree based on COI nucleotide sequences is shown in Figure 7.4. Thefour haplotypes found in *P. duboscqui* cluster together, and form the sister group of *P. putrinum* sequences. All four subgenera appear monophyletic, with subgenera *Paramecium* and *Cypriostomum* as sister groups on one major branch, and *Chloroparamecium* and *Helianter* sister groups on the other (both branches are poorly supported). The sequence of *P. multimicronucleatum* strain Kr 113-3 actually falls in the cluster of *P. caudatum* sequences, and those of *P. multimicronucleatum* strains TRB 101-4 and V 105-2 fall in a basal position inside subgenus *Paramecium*. All the other strains of *P. multimicronucleatum* clusters together. *P. calkinsi* also appears non monophyletic, its phylogeny being entangled with those of the other *Cypriostomum* morphospecies; the COI sequences of *P. calkinsi* strains BM1-11 and BM14-2 (GenBank IDs: JF304173 and JF304174 respectively) are even identical to that of *P. nephridiatum* strain WS97-1, and thus were not included in the analysis.

The *P. duboscqui* intra-group similarity values are 99.84% and 99.68% for groups A and B respectively. The inter-groups averaged similarity value is 93.68%. Inside other monophyletic morphospecies, the values vary from 80.41% (*aurelia* complex) to 88.78% (*P. bursaria*) and 90.84% (*P. caudatum*). *P. jenningsi*, *P. nephridiatum*, *P. polycaryum* and *P. putrinum* are represented by only two haplotypes each, hence their intraspecific similarity values (88.15%, 97.63%, 92.10% and 99.53% respectively) should be considered preliminary estimations. Each of the two clusters of *P. multimicronucleatum* sequences considered separately give values of 90.21% and 98.10%, but the value drops to 84.68% if they are considered together. A similar problem is also encountered with *P. calkinsi*, where the most distant sequences share a similarity value of only 72.35%.

P. bursaria is the only examined species with a longer sequence, containing an insert of 9 nucleotides (sites 304-312) in the middle of the characterized COI gene region. This is a synapomorphy of the *P. bursaria* strains.

Discussion

General remarks on the molecular markers employed. Of the three markers used in this study, the sequence of the nuclear 18S rRNA is the slowest evolving and the mitochondrial COI gene sequence the fastest. This is reflected in the capability of each marker to discriminate between strains, and in its suitability for phylogenetic inferences.

Studies on intraspecific diversity often don't employ the more conserved 18S sequence. However, this study and that of Hoshina and coworkers[16] prove that this marker can show some degree of variability inside the same morphospecies. One of the advantages of 18S is that it is a good target for sequence-specific oligonucleotide probes for fluorescence *in situ* hybridization (FISH), that could represent a fast method for discriminating closely related taxa[51-53].

The ITS1-5.8S-ITS2 sequence is actually composed of three functionally different regions, and some previous studies employed only ITS1[54] or ITS2[16]. In our opinion, the composed sequence contains a more reliable phylogenetic information. However, the ITS2 sequence has some unique characteristics. Coleman[20,45] derived two rules after surveying more than 100 eukaryotic species: 1) the occurrence of one or more CBCs in the conserved region of helix III implies the inability of two organisms to produce a zygote; 2) the total identity of ITS2 correlates with fertility. Thus, the common sequence of evolutionary events happening in two speciating taxa is: (I) appearance of mutations in the less conserved regions of ITS2; (II) development of reproductive isolation; (III) appearance of CBCs in the conserved region of helix III. If further confirmed, this marker could show some promise in the study of cryptic species complexes in all those occasions in which performing mating experiments is not feasible.

The mitochondrial COI gene sequence has received huge attention as a "DNA barcode"[55], especially in animals. In ciliates, this marker was employed in recent years for discriminating and inferring the relationships between closely related species, especially in the genus *Tetrahymena*[56-59]; and in intraspecific surveys trying to identify complexes of sibling species within known morphospecies[17-19,60,61]. Its use in the phylum was critically evaluated by Strüder-Kypke and Lynn[36]. An analysis focused on the whole mitochondrial genome in *Paramecium* was recently published by Barth and Berendonk[40].

***P. duboscqui* molecular variability.** We have found three main groups of *P. duboscqui* (A, B and C) according to molecular data. The 18S rRNA gene sequence is sufficient

to discriminate them and, unfortunately, it is the only marker available for the most divergent group C, represented by the extinct strain Ku4-8. Our results are in accordance with those obtained by Fokin and coworkers[11] using RAPD on a set of strains that overlap our own. They classified the three groups as different species and called them *Paramecium duboscqui I*, *II* and *III* respectively. The mitochondrial marker COI can further identify two haplotypes inside groups A and B, but intra-group similarity values are very high compared to inter-group values. The similarity values shared by groups A and B are generally higher than those found in other *Paramecium* morphospecies. However, where group C data are available the similarity threshold is lower than those existing between different biological species of the *aurelia* complex or different *P. multimicronucleatum* strains.

Sequence similarity alone cannot prove or disprove the existence of a complex of sibling species. No CBCs was found between different strains in the conserved region of helix III of ITS2, although a hemi-CBC was found between groups A and B near this region. Mating experiments should be performed in the future between strains of the three groups, but it is plausible that at least strains of group C will not result compatible with strains of other groups, according to their high genetic divergence.

***P. duboscqui* biogeography.** Although interspecific differences in the patterns of distribution are known[2], other recent molecular surveys failed to detect correlations between molecular clades and sampling locations within the same morphospecies. For *P. duboscqui*, a geographic structure of groups emerges instead. Group C is so far limited to Kunashir Island, in the Pacific Ocean. Group B was found in China and the Baltic Sea while group A was found on both sides of the Mediterranean Sea and on the Atlantic coast of North America. Different COI haplotypes characterize the Chinese and Baltic strains in group B and the American and Mediterranean strains in group A.

The study of Fokin and coworkers[11] suggests that strains coming from the same sampling location are genetically very similar or identical. These authors also investigated strains from the White Sea and the North Sea, that clustered inside the *P. duboscqui II* clade (that corresponds to our group B). As can be seen in Figure 7.5, there is a latitudinal gradient between the groups A and B, with A being located southerly. As far as our data can show, at least for this morphospecies of *Paramecium* there is a correlation between geographic separation and

molecular evolution. This hypothesis can be easily tested sampling in the predicted areals of group A, B or C and recovering the sequence of one of the markers employed here.

Figure 7.5 Map of sampling sites. The five Italian locations are represented by only two symbols. Dark circles – strains of group A; dark triangles – strains of group B; empty triangles – strains that belong to group B according to Fokin et al. (1999a); dark square – strain of group C.

Molecular variability in other morphospecies of the genus *Paramecium*. Our analyses compared the most taxonomically diverse dataset until now. This allowed us to detect some patterns, although we refer to the original papers listed in the introduction for details on each morphospecies.

Not all morphospecies show the same relative increase in evolutionary rate when different markers are considered. *P. bursaria* shows more molecular variability than *P. calkinsi* according to ITS1-5.8S-ITS2, but far less according to COI gene. We also noticed that, although the 18S marker is always the less able to discriminate between different genotypes, COI is not always more able than ITS1-5.8S-ITS2 in performing this duty. In *P. duboscqui*, for example, there is a polymorphic site in ITS2 that can discriminate among Italian strains that have the same COI haplotype. In *P. bursaria* and *P. multimicronucleatum* there are many similar cases, while none has been found in *P. calkinsi* nor in *P. caudatum*.

P. caudatum is consistently shown to be a less variable morphospecies, as highlighted also by other authors using different approaches[9,13,18]. Only one haplotype has been found so far in this taxon for both nuclear markers.

We have here investigated the secondary structure of ITS2 for the first time after the preliminary study of Coleman[20]. We have found that all the morphospecies not examined before have an additional small helix near the 5' end of the sequence, perhaps a vestige of the eukaryotic helix I[44,45], that was lost in some of the most derived lineages. No investigated morphospecies present CBC polymorphisms in the conserved region of helix III, not even those in which non-interbreeding syngens are known (Wichterman, 1986). CBCs in this region are rare even between morphospecies. This means that different lineages of *Paramecium* are mostly in the phase between milestones (I) and (III) of their evolutionary histories, and thus the properties of ITS2 derived by Coleman[45], even if further confirmed, cannot help in clarifying the sibling species problem in this genus.

Phylogeny of the genus *Paramecium*. While the genus itself is clearly monophyletic, some of the interspecific relationships are still unclear[2,16,39]. The subgenus *Paramecium* is well-characterized and monophyletic. The *aurelia* complex is closely related to *P. schewiakoffi* and *P. jenningsi*, to the point that the employed markers fail to clearly discriminate between these taxa (see also Catania and coworkers[21]). The ancestor of this clade underwent some major genomic change[40,62] that set it apart from other lineages. According to our analyses, it cannot be said if *P. multimicronucleatum* or *P. caudatum* is its closest relative.

Subgenera *Paramecium* and *Cypriostomum* are sister groups, although this conclusion is somewhat hindered by the divergent ITS1-5.8S-ITS2 sequence of *P. polycaryum*. This morphospecies, as well as *P. woodruffi*, *P. nephridiatum* and *P. putrinum* has not been thoroughly investigated with DNA sequence markers yet, and thus sound conclusions cannot be reached. We have noticed, however, that ITS1-5.8S-ITS2 sequences of *P. polycaryum* and *P. duboscqui* have a significantly different base composition compared to that of other morphospecies, and this could have caused some kind of artifact in the tree.

The relative phylogenetic positions of *P. putrinum*, *P. duboscqui* and *P. bursaria* are difficult to assess, because each marker give a different topology. According to the molecular characterization, they are quite different from each other and from the other species.

Discordant haplotypes. In our analyses we found two remarkable anomalies. One concerns *P. calkinsi*: the COI sequences obtained from strains of this morphospecies form a polyphyletic assemblage in the subgenus *Cypriostomum* that was also evident in the tree of Przyboś and coworkers[17], and at least one haplotype is shared with *P. nephridiatum*. The non-monophyly is also obtained in the ITS1-5.8S-ITS2 tree. Intraspecific similarity values calculated on COI sequences for *P. calkinsi* are extraordinarily low.

A similar case happens for some *P. multimicronucleatum* strains investigated by Tarcz and coworkers[19]. Our phylogenetic analysis is largely congruent with theirs, but they didn't include other closely related morphospecies, using only *P. bursaria* as outgroup. We have here showed that the divergent group they labeled as "C" is actually unrelated to the other strains, and contains sequences undistinguishable from those of *P. caudatum* (strain Kr 113-3) or of uncertain affiliation (strains TRB 101-4 and V 105-2).

About *P. calkinsi*, the problem should be tackled again when other morphospecies of subgenus *Cypriostomum* will be systematically investigated. One possibility is that the "*calkinsi*" morphotype is the ancestral one for the entire subgenus, conserved in some evolutionary lineages but lost in other, more derived ones. It is likely that what we call *P. calkinsi* is actually a complex of different biological entities. *P. multimicronucleatum* represents perhaps a similar case, although at least the strain Kr 113-3 could have been simply misidentified or contaminated and, in any case, the majority of sequences fall in a single clade.

In the light of these observations, we recommend that future surveys on single morphospecies include a broader range of available data in their comparisons and phylogenetic analyses. Some of the discordant strains, if still available, should also be inspected again.

References

1. Fokin SI. (2010/11) *Paramecium* genus: biodiversity, some morphological features and the key to the main morphospecies discrimination. Protistology. 6: 227-235
2. Fokin SI, Przyboś E, Chivilev SM, Beier CL, Horn M, *et al.* (2004) Morphological and molecular investigations of *Paramecium schewiakoffi* sp. nov. (Ciliophora, Oligohymenophorea) and current status of distribution and taxonomy of *Paramecium* spp. Eur J Protistol. 40: 225-243
3. Wichterman R. (1986) In: The biology of Paramecium, second edition, Plenum, New York and London

4. Aufderheide KJ, Daggett PM, Nerad TA. (1983) *Paramecium sonneborni* n. sp., a new member of the *Paramecium aurelia* species-complex. J Protozool. 30: 128-131
5. Sonneborn TM. (1975) The *Paramecium aurelia* complex of fourteen sibling species. Trans Am Micros Soc. 94: 155-178
6. Allen SL, Nerad TA, Rushford CL. (1983) Comparison of the esterases and acid phosphatases in *Paramecium multimicronucleatum*, syngens 1-5, *P. jenningsi*, *P. caudatum* and the *P. aurelia* complex. J Protozool. 30: 148-154
7. Skotarczak B, Przyboś E, Wodecka B, Maciejewska A. (2004). Sibling species within *Paramecium jenningsi* revealed by RAPD. Acta Protozool. 43: 29-35
8. Hall MS, Katz LA. (2011) On the nature of species: insights from *Paramecium* and other ciliates. Genetica. 139: 677-684
9. Hori M, Tomikawa I, Przyboś E, Fujishima M. (2006) Comparison of the evolutionary distances among syngens and sibling species of *Paramecium*. Mol Phylogenet Evol. 38: 697-704
10. Greczek-Stachura M, Tarcz S, Przyboś E. (2010). Intra-specific differentiation of *Paramecium bursaria* strains by molecular methods – preliminary studies. Folia Biol. (Kraków). 58: 35-45
11. Fokin SI, Stoeck T, Schmidt H. (1999). *Paramecium duboscqui* Chatton, Brachon, 1933. Distribution, ecology and taxonomy. Eur J Protistol. 35: 161-167
12. Fokin SI, Stoeck T, Schmidt HJ. (1999) Rediscovery of *Paramecium nephridiatum* Gelei, 1925 and its characteristics. J Eukaryot Microbiol. 46: 416-425
13. Stoeck T, Welter H, Seitz-Bender D, Kusch J, Schmidt HJ. (2000) ARDRA and RAPD-fingerprinting reject the sibling species concept for the ciliate *Paramecium caudatum* (Ciliophora, Protoctista). Zool Scr. 29: 75-82
14. Przyboś E, Prajer M, Greczek-Stachura M, Skotarczak B, Maciejewska A, Tarcz S. (2007) Genetic analysis of the *Paramecium aurelia* species complex (Protozoa: Ciliophora) by classical and molecular methods. Syst Biodivers. 5: 417-434
15. Stoeck T, Schmidt HJ. (1998) Fast and accurate identification of European species of the *Paramecium aurelia* complex by RAPD-fingerprints. Microbial Ecol. 35: 311-317
16. Hoshina R, Hayashi S, Imamura N. (2006). Intraspecific genetic divergence of *Paramecium bursaria* and re-construction of the paramecian phylogenetic tree. Acta Protozool. 42: 171-181
17. Przyboś E, Tarcz S, Potekhin A, Rautian M, Prajer M. (2012) A two-locus molecular characterization of *Paramecium calkinsi*. Protist. 163: 263-273
18. Barth D, Krenek S, Fokin SI, Berendonk TU. (2006) Intraspecific genetic variation in *Paramecium* revealed by mitochondrial cytochrome *c* oxidase I sequences. J Eukaryot Microbiol. 53: 20-25
19. Tarcz S, Potekhin A, Rautian M, Przyboś E. (2012) Variation in ribosomal and mitochondrial DNA sequences demonstrates the existence of intraspecific groups in *Paramecium multimicronucleatum* (Ciliophora, Oligohymenophorea). Mol Phylogenet Evol. 63: 500-509
20. Coleman AW. (2005) *Paramecium aurelia* revisited. J Eukaryot Microbiol. 52: 68-77
21. Catania F, Wurmser F, Potekhin AA, Przyboś E, Lynch M. (2009) Genetic diversity in the *Paramecium aurelia* species complex. Mol Biol Evol 26: 421-431
22. Tarcz S, Przyboś E, Prajer M, Greczek-Stachura M. (2006) Intraspecific variation of diagnostic rDNA genes in *Paramecium dodecaurelia, P. tredecaurelia* and *P. quadecaurelia* (Ciliophora, Oligohymenophorea). Acta Protozool. 45: 255-263

23. Finlay BJ, Esteban GF, Fenchel T. (2004) Protist diversity is different? Protist. 155: 15-22
24. Foissner W. (2006) Biogeography and dispersal of micro-organisms: a review emphasizing protists. Acta Protozool. 45: 111-136
25. Vivier E. (1974) Morphology, taxonomy, and general biology of the genus *Paramecium*. In: Van Wagtendonk WS (ed), Paramecium: a current survey, Elsevier, New York, pp 1-89
26. Shi X, Jin M, Liu G. (1997) Rediscovery of *Paramecium duboscqui* Chatton and Brachon, 1933, and a description of its characteristics. J Eukaryot Microbiol. 44: 134-141
27. Chatton E, Brachon S. (1933) Sur une paramécie à deux races: *Paramecium duboscqui*, n. sp. C R Soc Biol. 114: 988-991
28. Watanabe T, Shi X, Liu G, Jin M. (1996) Cytological studies of conjugation and nuclear processes in *Paramecium duboscqui* Chatton and Brachon 1933. Eur J Protistol. 32: 175-182.
29. Fokin SI. (1988) A bacterial symbiont of the macronuclear perinuclear space in the ciliate *Paramecium duboscqui*. Cytology (Sankt-Petersburg). 30: 632-635 [in Russian with English summary]
30. Fokin SI, Görtz HD. (1993). *Caedibacter macronucleorum* sp. nov., a bacterium inhabiting the macronucleus of *Paramecium duboscqui*. Arch Protistenkd. 143: 319-324
31. Schrallhammer M, Fokin SI, Schleifer K, Petroni G. (2006) Molecular characterization of the obligate endosymbiont "*Caedibacter macronucleorum*" Fokin and Görtz, 1993 and of its host *Paramecium duboscqui* strain Ku4-8. J Eukaryot Microbiol. 53: 499-506
32. Fokin SI. (2012) Frequency and biodiversity of symbionts in representatives of the main classes of Ciliophora. Eur J Protistol. 48: 138-148
33. Medlin L, Elwood HJ, Stickel S, Sogin ML. (1988) The characterization of enzymatically amplified 16S-like rRNA-coding regions. Gene. 71: 491-499
34. Petroni G, Dini F, Verni F, Rosati G. (2002) A molecular approach to the tangled intrageneric relationships underlying phylogeny in *Euplotes* (Ciliophora, Spirotrichea). Mol Phylogenet Evol. 22: 118-130
35. Rosati G, Modeo L, Melai M, Petroni G, Verni F. (2004) A multidisciplinary approach to describe protists: a morphological, ultrastructural, and molecular study on *Peritromus kahli* Villeneuve-Brachon, 1940 (Ciliophora, Heterotrichea). J Eukaryot Microbiol. 51: 49-59
36. Strüder-Kypke MC, Lynn DH. (2010) Comparative analysis of the mitochondrial cytochrome *c* oxidase subunit I (COI) gene in ciliates (Alveolata, Ciliophora) and evaluation of its suitability as a biodiversity marker. Syst Biodiv. 8: 131-148
37. Andreoli I, Mangini L, Ferrantini F, Santangelo G, Verni F, Petroni G. (2009) Molecular phylogeny of unculturable Karyorelictea (Alveolata, Ciliophora). Zool Scr. 38: 651-662
38. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, *et al.* (1997) Gapped BLAST and PSI BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389-3402
39. Strüder-Kypke MC, Wright AG, Fokin SI, Lynn DH. (2000) Phylogenetic relationships of the genus *Paramecium* inferred from small subunit rRNA gene sequences. Mol Phylogenet Evol. 14: 122-130
40. Barth D, Berendonk TU. (2011). The mitochondrial genome sequence of the ciliate *Paramecium caudatum* reveals a shift in nucleotide composition and codon usage within the genus *Paramecium*. BMC Genomics. 12: 272-283

41. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, *et al.* (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 35: 7188-7196
42. Ludwig W, Strunk O, Westram R, Richter L, Meier H, *et al.* (2004) ARB: a software environment for sequence data. Nucleic Acids Res. 32: 1363-1371
43. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, *et al.* (2007) Clustal W and Clustal X version 2.0. Bioinformatics. 23: 2947-2948
44. Coleman AW. (2003) ITS2 is a double-edged tool for eukaryote evolutionary comparisons. Trends Genet. 19: 370-375
45. Coleman AW. (2009) Is there a molecular key to the level of "biological species" in eukaryotes? A DNA guide. Mol Phylogenet Evol. 50: 197-203
46. Zuker M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31: 3406-3415
47. Guindon S, Gascuel O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 52: 696-704
48. Huelsenbeck JP, Ronquist F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 17: 754-755
49. Posada D. (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol. 25: 1253-1256
50. Schmidt HA, Strimmer K, Vingron M, von Haeseler A. (2002) TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics. 18: 502-504
51. Fried J, Foissner W. (2007) Differentiation of two similar glaucomid ciliate morphospecies (Ciliophora, Tetrahymenida) by fluorescence in situ hybridization with 18S rRNA targeted oligonucleotide probes. J Eukaryot Microbiol. 54: 381-387
52. Petroni G, Rosati G, Vannini C, Modeo L, Dini F, Verni F. (2003) *In situ* identification by fluorescently labeled oligonucleotide probes of morphologically similar, closely related ciliate species. Microbial Ecol. 45: 156-162
53. Schmidt SL, Bernhard D, Schlegel M, Fried J. (2006) Fluorescence in situ hybridization with specific oligonucleotide rRNA probes distinguishes the sibling species *Stylonychia lemnae* and *Stylonychia mytilus* (Ciliophora, Spirotrichea). Protist. 157: 21-30
54. Hoshina R. (2010) Secondary structural analyses of ITS1 in *Paramecium*. Microbes Environ. 25: 313-316
55. Hebert PDN, Cywinska A, Ball SL, deWaard JR. (2003) Biological identifications through DNA barcodes. Proc R Soc Lond B Biol Sci. 270: 313-321
56. Chantangsi C, Lynn DH. (2008) Phylogenetic relationships within the genus *Tetrahymena* inferred from the cytochrome *c* oxidase subunit 1 and the small subunit ribosomal RNA genes. Mol Phylogenet Evol. 49: 979-987
57. Catania F, Wurmser F, Potekhin AA, Przyboś E, Lynch M. (2009) Genetic diversity in the *Paramecium aurelia* species complex. Mol Biol Evol 26: 421-431
58. Kher CP, Doerder FP, Cooper J, Ikonomi P, Achilles-Day U, *et al.* (2011) Barcoding *Tetrahymena*: discriminating species and identifying unknowns using the cytochrome *c* oxidase subunit I (cox-1) barcode. Protist. 162: 2-13
59. Lynn DH, Strüder-Kypke MC. (2006) Species of *Tetrahymena* identical by small subunit rRNA gene sequences are discriminated by mitochondrial cytochrome *c* oxidase I gene sequences. J Eukaryot Microbiol. 53: 385-387

60. Gentekaki E, Lynn DH. (2009) High-level genetic diversity but no population structure inferred from nuclear and mitochondrial markers of the peritrichous ciliate *Carchesium polypinum* in the Grand River Basin (North America). Appl Environ Microbiol. 75: 3187-3195

61. Gentekaki E, Lynn DH. (2010) Evidence for cryptic speciation in *Carchesium polypinum* Linnaeus, 1758 (Ciliophora: Peritrichia) inferred from mitochondrial, nuclear, and morphological markers. J Eukaryot Microbiol. 57: 508-519

62. Aury JM, Jaillon O, Duret L, Noel B, Jubin C, *et al.* (2006) Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. Nature. 444: 171-178

Chapter 7 Afterword.

Similar studies were performed after this, mostly dealing with only one of the markers investigated here, and our results were generally cited and confirmed. The most relevant novelty in our understanding of *Paramecium* diversity since the publication of this work was the redescription of the morphospecies *Paramecium chlorelligerum*.

P. chlorelligerum is an old and neglected species, virtually never appearing in the literature after its original establishment. It harbors symbiotic algae, like *P. bursaria*, but Kreutz, Stoeck, and Foissner clearly demonstrate in their paper, published at the same time as ours on *P. duboscqui* ("Morphological and molecular characterization of *Paramecium* (*Viridoparamecium* nov. subgen.) *chlorelligerum* Kahl 1935 (Ciliophora)", 2012), that it is a separate and valid species. The authors also provided the 18S rRNA gene sequence, showing that *P. chlorelligerum* belongs to a divergent lineage, deserving in their opinion the status of a new subgenus, in unresolved relationship with the others. However, because only few sequences are currently available for this species, it is still impossible to perform a molecular survey and evaluate the intra-specific variability.

The careful (and visually impeccable) work of Kreutz and coworkers is unusual, in the recent literature, in making every effort to associate a new characterization to an old, disused name, instead of creating a new one and increasing the number of taxa. This issue is central in the work presented in the next chapter.

Chapter 8.

Redescription of the genus *Spirostomum*

ORIGINAL PUBLICATION:

Focusing on genera to improve species identification: revised systematics of the ciliate *Spirostomum*

Vittorio Boscaro, Daniela Carducci, Giovanna Barbieri, Marcus VX Senra, Ilaria Andreoli, Fabrizio Erra, Giulio Petroni, Franco Verni, Sergei I Fokin

Protist 165(4): 527-541. August 2014.

Abstract

Although many papers dealing with the description of new ciliate taxa are published each year, species taxonomy and identification in most groups of the phylum Ciliophora remain confused. This is largely due to a scarcity of surveys on the systematics of immediately higher levels (genera and families) providing data for old and new species together. *Spirostomum* is a common and distinctive inhabitant of fresh- and brackish water environments, including artificial and eutrophic ones, and is a good model for applied ecology and symbiosis research. Despite this, only 3 of the numerous species are commonly cited, and no studies have yet confirmed their monophyly, with the consequence that reproducibility of the results may be flawed. In this paper we present morphological and molecular data for 30 *Spirostomum* populations representing 6 different morphospecies, some of which were collected in previously unreported countries. We performed a detailed revision of *Spirostomum* systematics combining literature surveys, new data on hundreds of organisms and statistical and phylogenetic analyses; our results provide insights on the evolution, ecology and distribution of known morphospecies and a novel one: *Spirostomum subtilis* sp. n. We also offer tools for quick species identification.

Introduction

Ciliate taxa have been described for almost two centuries, while the theoretical background underlying their understanding changed so much that organisms once described as "perfect" (multicellular) animalcules[1,2] are now known to be unicellular and more closely related to some algae than to animals[3]. Classes and orders were melted, recombined and renamed numerous times, following different criteria[4]. The molecular revolution set new standards for characterizations, and a constantly increasing number of papers describing new species are published each year (about 20 new species and 4 new monospecific genera were established only in 2013[5-11]. Despite this, a huge confusion still reigns in ciliate taxonomy of low-level taxa. Fewer works deal with the redescription of old species[12], and even less with the integrated systematics of genera – except for a handful of flagship, intensely studied ones, like *Paramecium*[13,14], *Tetrahymena*[15,16] or *Euplotes*[17]. Many old descriptions are so vague that it is virtually impossible to compare new data with the original ones, but because of nomenclatural rules names are hard to dismiss, and pursuing the goal of invalidating them is often not worth the effort. Thus, new species names appear, old ones remain, and confusion increases.

And yet, all applied studies relying on correct species identification (ecology, molecular biology, etc.) would greatly benefit from unambiguous species naming, description and guide to identification. Although the discovery of new species is certainly appealing and informative, the immediately higher hierarchical levels, like genera and families, should be more often reconsidered in order to maintain a holistic view. Integrating new morphological and molecular data from several closely related morphospecies together with a thorough survey of the literature is in fact the best way to detect discrepancies and to better assess biodiversity.

The genus *Spirostomum* Ehrenberg, 1834[1] is a good example of a weak systematic framework's potential issues. *Spirostomum* representatives are large ciliates of the class Heterotrichea with an eye-catching, distinctive shape; they are common in freshwater and low salinity (brackish) environments, sometimes in high abundances[18,19]. They are valuable targets of ecological studies, and some species are often claimed to be good bioindicators for a number of threats to the environment's quality[20,21]. Also, more recently they were investigated as hosts of eukaryotic[22] and prokaryotic[23-25] and symbionts. Nevertheless, only three species are commonly reported in recent works, against the dozen present in literature. Several papers presented

molecular data without morphological descriptions[24-27], or vice-versa[28]. To our knowledge, only the paper of Fernandes and da Silva Neto[29] provided linked morphological and molecular data, but only for two species (represented by one population each). Morphological characters employed to define morphospecies are not numerous, and their variability range somewhat differs according to different sources. Finally, *Spirostomum* poses specific obstacles to detailed morphological studies: it is highly contractile and does not always adapt to growth in laboratory conditions, making the culturing of monoclonal strains unpractical.

In this work, we present and discuss morphological and molecular data on 30 *Spirostomum* populations belonging to 5 known morphospecies and a novel species. The populations were collected in several continents, including previously uninvestigated tropical and northern countries. We provide a survey of the genus based on our newly collected data on hundreds of organisms as well as a critical interpretation of literature reports, providing insights on the variability, phylogeny, evolution and distribution of these ciliates. We took particular care to avoid the proliferation of new species names and to revive instead those already present in older literature and reported here again. We also identified a few easily observable key characters for each morphospecies, in order to facilitate identification for non-taxonomists.

Materials and Methods

Samples and cultures. Table 8.1 list all investigated *Spirostomum* populations and sampling metadata. Aliquots of sediment and water from each sample were observed under the dissecting microscope in order to detect organisms belonging to the genus *Spirostomum* and divide them according to their main morphotype. The resulting populations were then maintained at 18-20 °C in their original medium, periodically enriched with rice grains, lettuce medium and/or modified Cerophyl medium[30] inoculated with *Raoultella planticola* (*Gammaproteobacteria*).

Morphological data. Single *Spirostomum* cells were harvested from the culture medium and observed with a Leitz (Weitzlar, Germany) differential interference contrast microscope equipped with a digital camera (Canon PowerShot S45). The device developed by Skovorodkin[31]

Table 8.1. Selected features of the 30 investigated *Spirostomum* populations.

Population name	Sampling location	Cell color	Cell length (µm)	Cell length/width	Peri./length (%)	CV/length (%)	Kineties number[ab]	CG rows[b]	CG pattern	MAC shape	MAC nodules	MAC length[c] (µm)	MAC length/width[c]	MIC number	MIC size (µm)	
Mdg2-1	MADAGASCAR	Brownish	**278** ± 32	**8.9** ± 1.3	**39** ± 4	**16** ± 4	**7.0** ± 1.2	**3.6** ± 0.6	Homogeneous	Ellipsoid	**1.0** ± 0.0	**29.2** ± 3.4	**4.1** ± 1.2	**1.3** ± 0.5	**1.9** ± 0.2	***Spirostomum teres***
Seef1	AUSTRIA	Brownish	**252** ± 30	**8.0** ± 1.6	**43** ± 5	**17** ± 5	**6.1** ± 0.5	**3.2** ± 0.4	Homogeneous	Ellipsoid	**1.0** ± 0.0	**25.9** ± 3.3	**3.4** ± 0.9	**1.4** ± 0.7	**1.9** ± 0.3	***Spirostomum teres***
StMgN	MADAGASCAR	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	*Spirostomum* sp.
Nor_BG	NORWAY	Brownish	**315** ± 54	**7.9** ± 1.8	**44** ± 5	**15** ± 6	**7.9** ± 1.3	**4.3** ± 0.6	Homogeneous	Ellipsoid	**1.0** ± 0.0	**28.8** ± 2.9	**2.2** ± 0.5	**1.8** ± 0.8	**2.0** ± 0.2	***Spirostomum teres***
Nor_KD	NORWAY	Brownish	**346** ± 55	**11.4** ± 2.3	**41** ± 5	**17** ± 3	**8.2** ± 1.5	**2.9** ± 0.7	Homogeneous	Ellipsoid	**1.0** ± 0.0	**41.1** ± 10.3	**4.6** ± 2.3	**1.1** ± 0.3	**2.0** ± 0.4	***Spirostomum teres***
PFEU3_Sm2	ITALY	Brownish	**336** ± 64	**10.5** ± 2.6	**44** ± 7	**15** ± 4	**7.4** ± 1.3	**3.7** ± 0.5	Homogeneous	Ellipsoid	**1.0** ± 0.0	**38.5** ± 6.2	**3.8** ± 1.0	**2.1** ± 0.8	**2.1** ± 0.3	***Spirostomum teres***
IP1	ITALY	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	*Spirostomum* sp.
StJFp2	BRAZIL	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	*Spirostomum* sp.
PFEU1_Spte	ITALY	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	*Spirostomum* sp.
PF1a	ITALY	Brownish	**366** ± 39	**9.8** ± 1.0	**42** ± 4	**27** ± 7	**9.1** ± 0.8	**2.0** ± 0.0	Heterogeneous	Ellipsoid	**1.0** ± 0.0	**28.8** ± 4.4	**2.8** ± 0.6	**1.2** ± 0.4	**2.2** ± 0.3	***Spirostomum teres***
LarnCyp	REPUBLIC OF CYPRUS	Brownish	**297** ± 36	**7.9** ± 1.5	**45** ± 5	**16** ± 5	**7.4** ± 1.0	**4.1** ± 0.7	Homogeneous	Elongated	**1.0** ± 0.0	**44.7** ± 8.5	**6.1** ± 1.1	**1.3** ± 0.6	**1.9** ± 0.1	***Spirostomum yagiui***
SpirWS	RUSSIA	Brownish	**389** ± 73	**10.2** ± 2.3	**48** ± 4	**12** ± 5	**9.8** ± 1.1	**2.5** ± 0.7	Homogeneous	Elongated	**1.0** ± 0.0	**51.0** ± 9.9	**6.7** ± 1.5	**2.3** ± 1.2	**1.8** ± 0.3	***Spirostomum yagiui***
GNS4	GERMANY	Brownish	**370** ± 41	**9.9** ± 1.9	**53** ± 4	**16** ± 6	**9.4** ± 0.5	**3.6** ± 0.5	Homogeneous	Elongated	**1.0** ± 0.0	**69.8** ± 18.2	**9.0** ± 3.3	**1.9** ± 1.2	**2.0** ± 0.4	***Spirostomum yagiui***
Pozz	ITALY	Brownish	**360** ± 64	**12.1** ± 1.9	**40** ± 4	**17** ± 8	**7.4** ± 0.9	**3.3** ± 0.5	Homogeneous	Elongated	**1.0** ± 0.0	**42.6** ± 3.7	**5.0** ± 1.2	**2.9** ± 1.0	**1.7** ± 0.3	***Spirostomum yagiui***
SAd	SOUTH AFRICA	Dark	**406** ± 58	**11.0** ± 1.8	**55** ± 8	**16** ± 6	**9.7** ± 1.6	**3.8** ± 0.4	Homogeneous	Filiform	**1.0** ± 0.0	**137.9** ± 20.4	**21.8** ± 5.7	**2.9** ± 1.5	**1.5** ± 0.2	***Spirostomum dharwarensis***
Mdg3	MADAGASCAR	Brownish	**449** ± 92	**12.1** ± 3.2	**47** ± 5	**17** ± 6	**9.2** ± 1.9	**4.1** ± 0.4	Homogeneous	Ellipsoid	**1.0** ± 0.0	**26.8** ± 4.7	**2.5** ± 0.7	**1.4** ± 0.7	**1.8** ± 0.2	***Spirostomum teres***
Mdg2-2	MADAGASCAR	Brownish	**342** ± 496	**9.1** ± 1.3	**49** ± 4	**13** ± 3	**8.7** ± 1.1	**3.5** ± 0.6	Homogeneous	Moniliform	**15.7** ± 2.9	**19.7** ± 4.9	**3.4** ± 0.7	**5.5** ± 1.6	**1.8** ± 0.2	***Spirostomum minus***
Mdg4	MADAGASCAR	Brownish	**515** ± 99	**13.1** ± 3.1	**51** ± 3	**14** ± 3	**8.1** ± 1.3	**2.9** ± 0.4	Homogeneous	Moniliform	**13.8** ± 3.7	**29.9** ± 10.8	**4.5** ± 1.4	**2.1** ± 1.4	**1.8** ± 0.2	***Spirostomum minus***
SmJFp1	BRAZIL	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	*Spirostomum* sp.
Thd2	THAILAND	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	*Spirostomum* sp.
Gmn	GERMANY	Brownish	**518** ± 121	**11.0** ± 2.1	**44** ± 4	**17** ± 8	**8.6** ± 1.5	**3.0** ± 0.5	Homogeneous	Moniliform	**12.8** ± 4.4	**28.0** ± 3.1	**3.0** ± 0.8	**2.2** ± 1.2	**2.7** ± 0.6	***Spirostomum minus***
Ind	INDIA	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	*Spirostomum* sp.
SAm	SOUTH AFRICA	Dark	**561** ± 67	**12.2** ± 2.5	**52** ± 6	**12** ± 2	**9.7** ± 1.1	**3.7** ± 0.6	Homogeneous	Moniliform	**16.9** ± 3.2	**17.3** ± 3.8	**4.2** ± 1.1	**2.9** ± 1.5	**2.7** ± 0.5	***Spirostomum minus***
SmPS	ITALY	Dark	**421** ± 47	**9.8** ± 2.0	**51** ± 4	**16** ± 2	**9.4** ± 1.4	**3.9** ± 0.3	Heterogeneous	Moniliform	**9.4** ± 3.4	**29.3** ± 6.4	**2.7** ± 0.7	**3.2** ± 1.1	**1.9** ± 0.2	***Spirostomum minus***
PBG1	ITALY	Dark	**558** ± 74	**9.5** ± 2.1	**53** ± 6	**11** ± 3	**10.0** ± 1.6	**3.7** ± 0.6	Heterogeneous	Moniliform	**13.8** ± 2.5	**24.0** ± 3.9	**3.2** ± 1.0	**1.8** ± 1.2	**2.2** ± 0.5	***Spirostomum minus***
SS5	SWEDEN	Brownish	**816** ± 86	**13.4** ± 2.1	**51** ± 4	**11** ± 2	**9.9** ± 0.9	**3.9** ± 0.3	Heterogeneous	Moniliform	**14.2** ± 2.7	**25.8** ± 5.0	**2.5** ± 0.7	**11.5** ± 3.2	**2.1** ± 0.3	***Spirostomum minus***
SS4-1	SWEDEN	Dark	**955** ± 132	**20.9** ± 3.2	**42** ± 4	**25** ± 6	**9.5** ± 0.6	**1.0** ± 0.0	Homogeneous	N/A	N/A	N/A	N/A	N/A	N/A	***Spirostomum subtilis* sp. nov.**
Zur3	SWITZERLAND	Dark	**837** ± 98	**17.3** ± 3.3	**53** ± 5	**17** ± 7	**11.2** ± 0.7	**1.0** ± 0.1	Homogeneous	Moniliform	**20.4** ± 3.0	**17.3** ± 4.8	**3.1** ± 0.8	**3.1** ± 1.5	**2.9** ± 0.4	***Spirostomum subtilis* sp. nov.**
SS4-2	SWEDEN	Dark	**1,230** ± 180	**12.7** ± 2.5	**61** ± 5	**9** ± 2	**21.5** ± 2.5	**4.2** ± 0.4	Heterogeneous	Moniliform	**21.0** ± 3.1	**35.0** ± 7.1	**4.0** ± 0.8	**9.4** ± 4.7	**1.6** ± 0.2	***Spirostomum ambiguum***
Seef2	AUSTRIA	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	*Spirostomum* sp.

[a] Inferred from the number of CG stripes

[b] Counted in the middle section of the cells

[c] Taken for the biggest nodule for moniliform MACs

Peri., Peristome; **CV**, Contractile Vacuole; **CG**, Cortical Granules; **MAC**, macronucleus; **MIC,** micronucleus

was employed to stop the organisms' movement without altering their shape. Feulgen staining was performed to observe the features of the nuclear apparatus. Length measures, on both living and fixed cells, were taken on collected pictures with the software Macnification v2.0.1 (Orbicule bvba). Pictures are available upon request; fixed slides were deposited in the collection of the Museo della Scienza e del Territorio della Certosa di Calci (Calci, Pisa, Italy).

Discriminant function analyses were performed with the software Statistica v6 on 10 characters from in vivo observations, and 5 characters from Feulgen observations (see additional details online: http://www.sciencedirect.com/science/article/pii/S1434461014000510). Only cells with at least 70% of the parameters measured were employed, resulting in two character matrices of 1,960 and 1,395 data (196 and 279 cells) for in vivo and Feulgen analyses, respectively. Missing data (less than 5% in the in vivo matrix and less than 2% in the Feulgen matrix) were substituted with mean values.

Molecular data. About 30-50 cells from each well-growing population were individually picked with a glass micropipet, washed several times in sterile water and fixed in EtOH 70%. 18S rRNA gene and a sequence containing the complete ITS1+5.8S+ITS2 region and part of the 28S rRNA gene were amplified and sequenced as described previously[13]. When members of a population were not numerous enough, only a few organisms were collected, and PCR mixtures were directly applied on the exsiccated cells as described in Andreoli and coworkers[32]; these amplicons did not require heminested PCR reactions to be sequenced. Standard ambiguity letters were associated to sites that showed double peaks in electropherograms despite repeated sequencing attempts.

All sequences were deposited at the European Nucleotide Archive (ENA) database (accession numbers: HG939524-53).

Phylogenetic analyses. Phylogenetic analyses on the 18S rRNA gene sequence were performed on 127 homologous sequences of Heterotrichea and Karyorelictea, trimmed at the ends in order to obtain a rectangular matrix (matrix *a*, 1,680 characters) and additionally excluding columns containing only one non-gap character (matrix *b*, 1,660 characters) and columns in which the most represented base was present in less than 20% of taxa (matrix *c*, 1,629 characters). Unless where differently specified, similarity values were calculated on unmodified matrices, ignoring ambiguous sites. A 131-sequence matrix was built in order to include some available *Spirostomum* sequences that were too short, or contained too many ambiguous

characters, to be employed in the main analysis; missing data were then treated as gaps. Phylogenetic analyses were also performed on the subset of 22 populations studied in more detail, employing the entire 18S+ITS1+5.8S+ITS2+28S region (2,379 characters), morphological data (18 characters), and a mixed character matrix. Morphological characters employed in the phylogenetic analysis are shown in Supplementary Table S1; qualitative characters were coded in unordered categories, and each quantitative, continuous characters were coded in 6 ordered discrete categories.

Maximum Likelihood analyses were performed with PHYML[33] as implemented in the ARB software package[34]; 100 pseudoreplicates were used for bootstrapping. Bayesian Analyses were performed with MrBayes v3.2.2[35], employing 3 independent runs with 1 cold and 3 heated chains each; runs were iterated for 1,000,000 generations. The best-fitting model for phylogenetic inferences on molecular data was chosen among 88 candidate models according to jModelTest v2.1.4[36]. The AIC criterion always selected the GTR+I+G model; the continuous gamma function was approximated with 4 discrete categories.

Results

Morphological observations. 22 out of the 30 sampled populations analyzed in this work were numerous enough to be subjected to quantitative morphological analysis. Parameters were measured on about 450 living and 420 stained cells (resulting in more than 3,000 and 1,700 raw data, respectively). The main morphological data are summarized in Table 8.1; a more comprehensive dataset can be found online: http://www.sciencedirect.com/science/article/pii/S1434461014000510.

In vivo observations. The worm-like appearance typical of *Spirostomum*, with a long peristomial field parallel to the main body axis and a posteriorly located contractile vacuole (CV), is a feature shared by all populations (Figures 8.1A-F). Despite the importance sometimes attributed to this character's variability, we observed uniformity in the shape of extremities, the anterior one usually being "rounded" and the posterior one "truncated". Quantitative parameters generally vary along a smooth gradient between extreme values, that can be very far

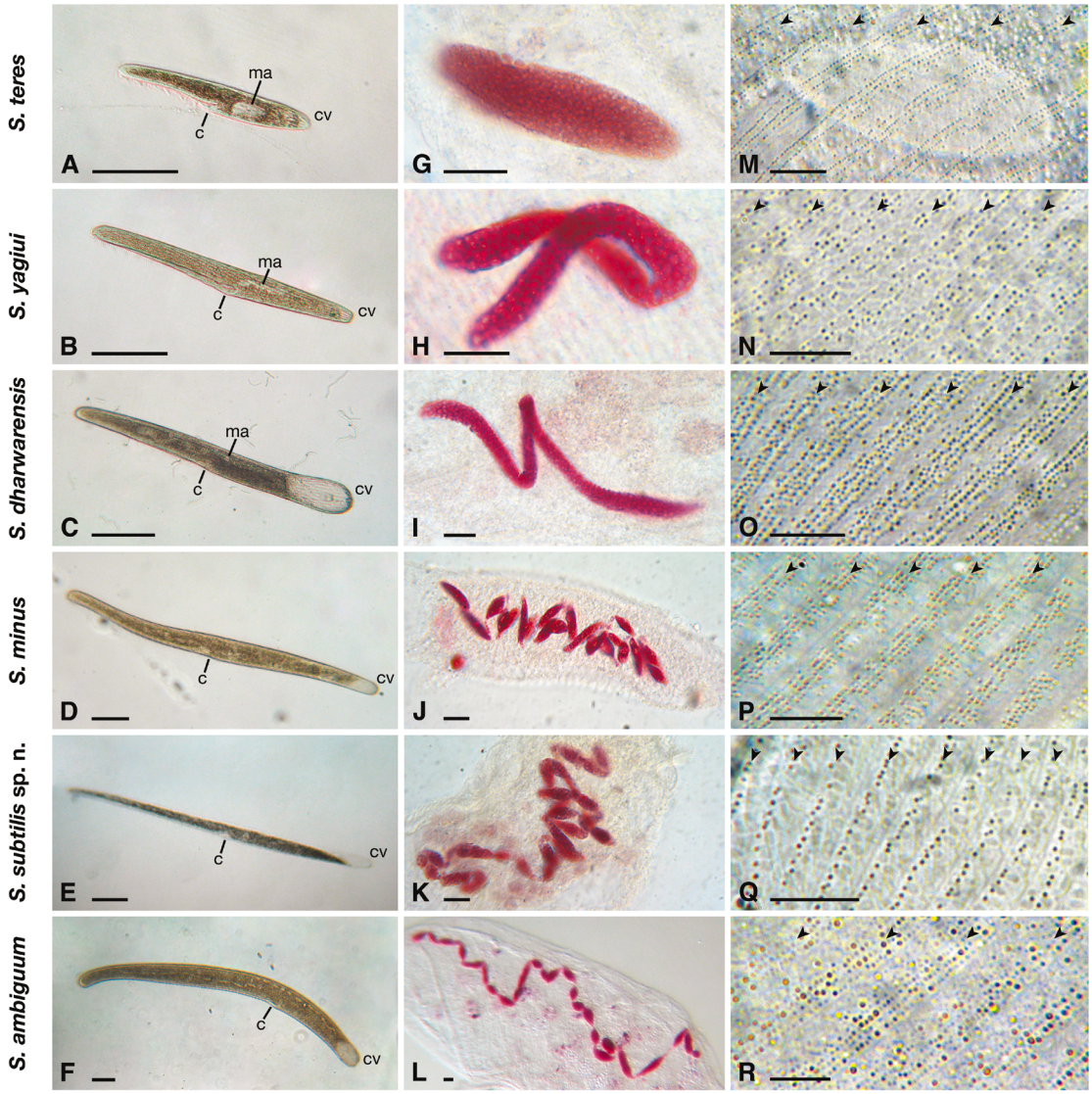


Figure 8.1 Morphological features of *Spirostomum* morphospecies. For each species, pictures taken with a differential interference contrast microscope show representative living cells (A-F), Feulgen-stained macronuclei (G-L) and cortical granule (CG) patterns in vivo (M-R). The novel morphospecies *Spirostomum subtilis* is particularly thin, often with a large contractile vacuole (E); these characters, and the cytostome position, differentiate it from *S. ambiguum* (F). *S. subtilis* may be distinguished from all other species, and in particular by the longest *S. minus* populations, because of its unique CG pattern (Q). Arrowheads mark the interkinetal CG stripes consisting of several CG rows in all species except *S. subtilis*, where a single row per stripe is visible. *S. teres* is represented by population PFEU3_Sm2, *S. yagiui* by population GNS4, *S. dharwarensis* by population SAd, *S. minus* by population SAm, *S. subtilis* sp. n. by population Zur3 and *S. ambiguum* by population SS4-2. Bars stand for 100 µm (A-F) and 10 µm (G-R). c, cytostome, ma; macronucleus (only "single"-type macronuclei are labeled in living cells); cv, contractile vacuole.

apart (especially for cell length). Exceptions are the number of kineties, which is significantly higher in population SS4-2 than in the others, and the CV length : cell length ratio, which strongly departs from the average value only in a few populations (PF1a, SS4-1).

Cortical granules (CG) are arranged in stripes that run parallel to the main body axis. There is always a single stripe between each kinety pair, but it may include one or (more often) several CG rows (Figures 8.1M-R); variation in number of rows for each stripe is almost as large within populations as among different populations. This character may vary slightly in different sectors (anterior, median and posterior) as well as different stripes in the same cell. It is also possible that the CG pattern depends in some degree to the organism's physiological status, because on rare occasions CG rows were almost invisible, despite being clearly detectable in the same population during other observations. Nevertheless, most populations showed 2 or more CG rows in each stripe; the exceptions were populations Zur3 and SS4-1, which virtually never departed from their usual pattern consisting of a single CG row per stripe (Figure 8.1Q). The "homogeneous" or "heterogeneous" CG pattern types refer to the occasional presence of granules with different size and/or color in the heterogeneous patterns.

Discriminant function analysis performed on living observations strongly differentiated only some of the populations (Figure 8.2A). In particular, SS4-2 on one hand and Zur3 and SS4-1 on another are clearly separated from the bulk of other populations. The remaining 19 are arranged in a continuous multidimensional gradient, with several populations bridging those with more extreme features.

Feulgen observations. Macronuclear (MAC) morphotypes can be divided into two main categories: moniliform (formed by a chain of beads, the MAC nodules; Figures 8.1J-L) and single (Figures 8.1G-I). The number of nodules in moniliform MACs vary among populations. On the contrary, variation in nodule shape seems to be more related to the organism's condition: ovoid, elongated and spindle-shaped nodules can usually be found in the same population or even cell. The degree of connection between adjacent nodules (from none to thick channels) is also fairly unstable within each population. The "single" MAC morphotype can be further divided in ellipsoid (length : width ratio < 5, Figure 8.1G), elongated (length : width ratio > 5, Fig. 1H) and filiform (length : width ratio > 20, Figure 8.1I). While the ovoid and filiform MACs show relatively uniform shapes, the elongated MACs are more diverse. This MAC type is only found in the littoral or brackish populations LarnCyp, SpirWS, GNS4 and Pozz,

and in all of them it is present in three variants: rod-like, convoluted and dividing (Figure 8.3). It is likely that these shapes are associated to different stages of the cell-cycle.

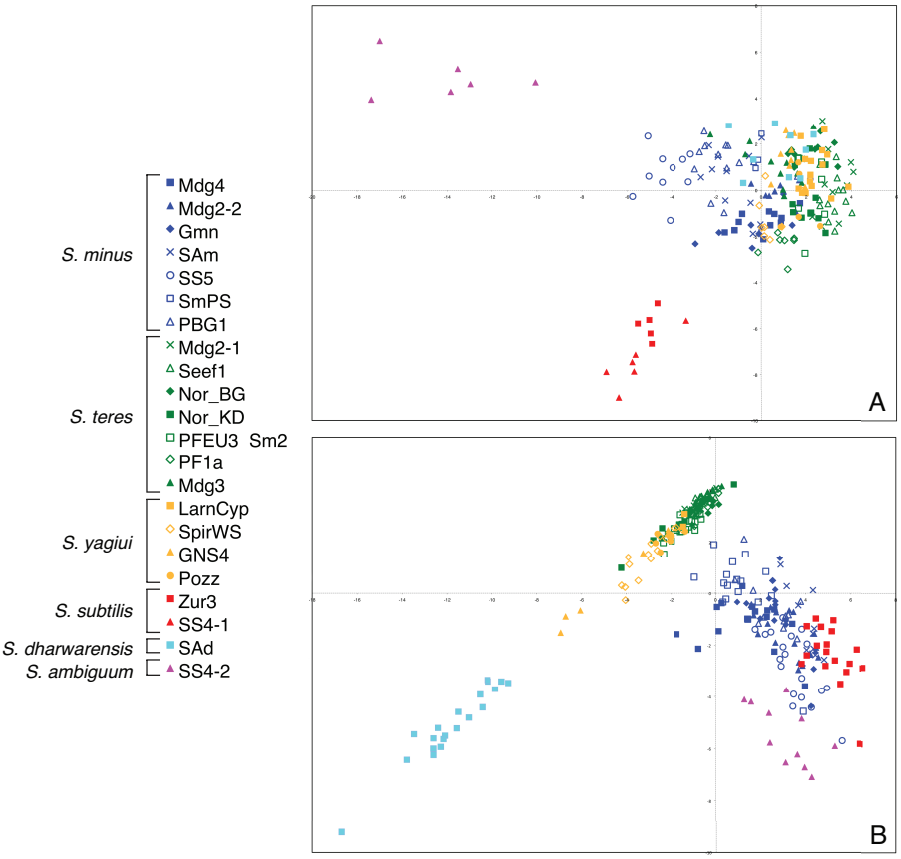


Figure 8.2 Discriminant function analysis of quantitative in vivo (A) and Feulgen staining (B) data. Axes represent the first two discriminant functions.

Micronuclei (MICs) vary in number and, to a lesser extent, size. MIC shape is ovoid or ellipsoid, and MICs are always close to MACs; they tend to be located in more or less deep matching depressions in the "single" type MACs. It is important to note that MIC number might be underestimated: it is possible that some get hidden below MACs in Feulgen pictures, because of their smaller size.

Discriminant function analysis on Feulgen observations (Figure 8.2B) differentiated more groups than the analysis of in vivo data. Populations with a moniliform MAC are clustered, with SS4-2 at one extreme of the gradient and SmPS at the other; another group includes populations with ellipsoid or elongated MACs, the latter being more scattered; population SAd

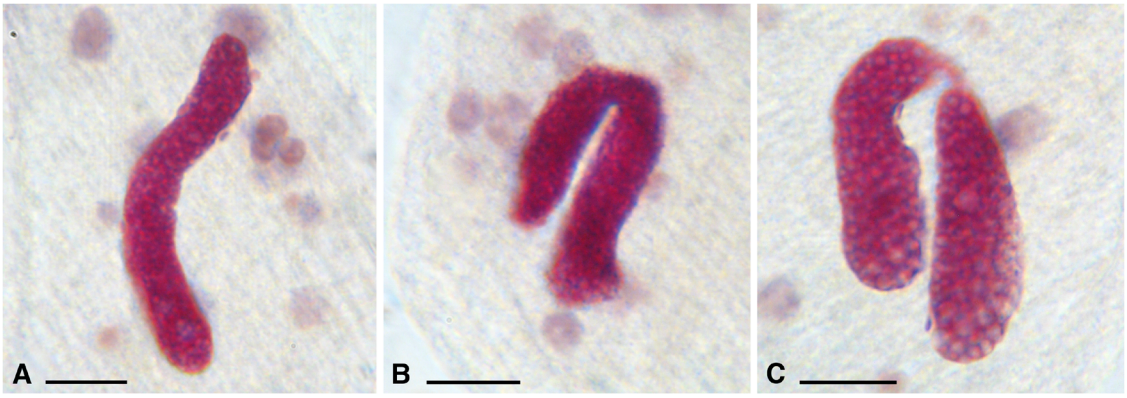


Figure 8.3 The three variants of the "elongated" type macronuclei. A, rod-like; B, convoluted (see also Fig. 1H); C, dividing. Pictures were taken on *S. yagiui* population SpirWS. Bars stand for 10 µm.

is completely separated in virtue of its unique filiform MAC. The two main discriminant functions are both heavily influenced by the number and length of MAC nodules.

Molecular sequences. The almost complete 18S rRNA gene sequences obtained during this work were 1702 bp long, and did not require the introduction of gaps during alignment; some previously released *Spirostomum* sequences slightly differ in this respect, exhibiting a few indels in very short homopolymeric regions. The lowest similarity value shared by all *Spirostomum* sequences is 97.5% (comparing it with those of other heterotrich genera represented by several sequences, the value is similar to the 96.1% of *Stentor* and the 97.3% of *Blepharisma*, but it is higher than the 90.7% of *Condylostoma*).

The ITS1+5.8S+ITS2+28S region amplified is 662-674 bp long, contains short regions with several indels and it is more variable than the 18S gene region: the lowest identity value for all *Spirostomum* sequences is 89.5%; there is no other extensive set of data for this marker in any other heterotrich genera to compare this value with.

Phylogeny. In all 18S rRNA trees, *Spirostomum* strongly clusters within Heterotrichea (Figure 8.4). As in other analyses[37,38], only a few clades above the genus level are resolved, like the "crown-group" formed by *Stentor*, *Blepharisma*, *Maristentor*, *Fabrea* and folliculinids. The relationships among other heterotrich lineages are only weakly supported. Nevertheless, *Spirostomum* sequences never associate with the only sequence of the confamiliar genus *Gruberia* (accession number: L31517[26]). The most closely related sequence, instead, belongs to *Anigsteinia*

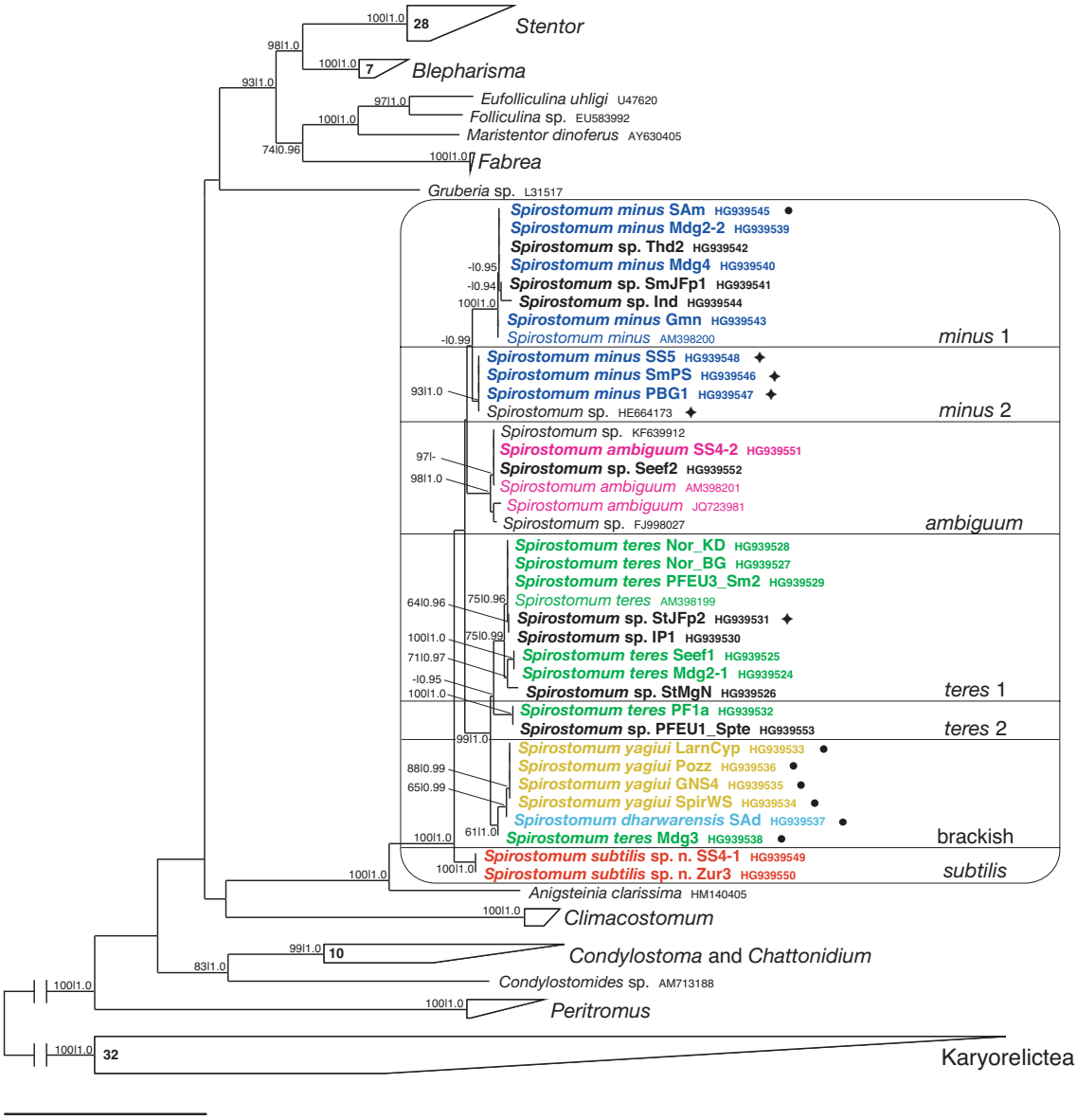


Figure 8.4 18S rRNA gene Maximum Likelihood tree of Heterotrichea, with focus on the genus *Spirostomum*. The analysis was performed on character matrix *a* (see text). Sequences obtained in this work are in bold. Numbers associated with nodes represent bootstrap|posterior probability values; bootstrap values below 60 and posterior probability values below 0.90 were omitted. Numbers in trapezoids show the number of sequences (if more than 2) representing clades collapsed for clarity. Dark circles indicate populations collected in brackish, littoral or estuarine environments; diamonds indicate populations harboring bacterial symbionts (see also Figure 8.5). The bar stands for an estimated divergence of 10%.

(accession number: HM140405; unpublished), which according to traditional taxonomy should be a close relative of *Blepharisma*[4]. However, no morphological observation was provided together with the *Gruberia* and *Anigsteinia* sequences.

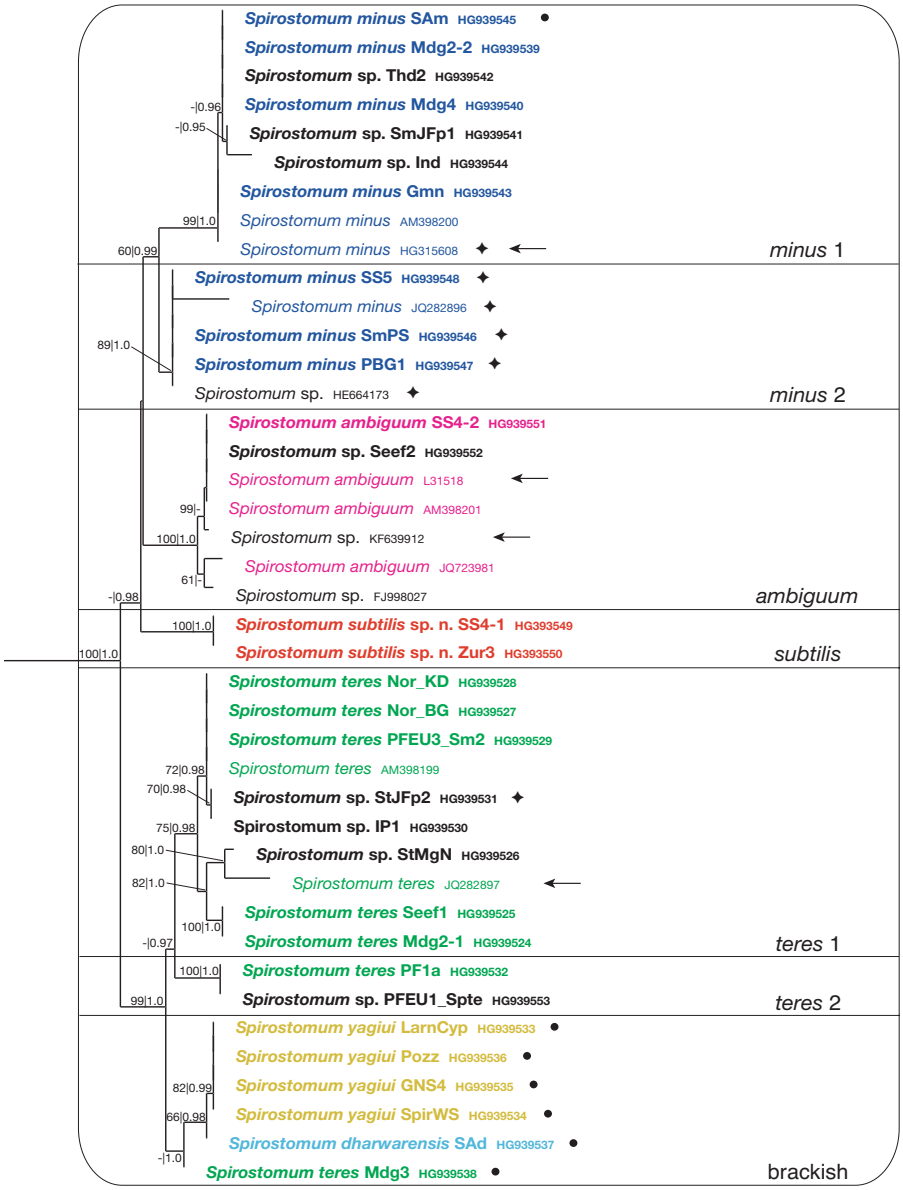


Figure 8.5 18S rRNA gene Maximum Likelihood tree of Heterotrichea, with focus on the genus *Spirostomum*. This tree differs from the one depicted in Figure 8.4 by the inclusion of 4 additional *Spirostomum* sequences of shorter length or lesser quality, indicated by arrows. The position of *S. subtilis* is also different, but still weakly supported. The analysis was performed on a character matrix including 131 sequences and 1,767 characters. Sequences obtained in this work are in bold. Numbers associated with nodes represent bootstrap|posterior probability values; values below 60|0.90 were omitted. Dark circles indicate populations collected in brackish, littoral or estuarine environments; diamonds indicate populations harboring bacterial symbionts.

All inferred 18S rRNA trees displayed 7 highly homogeneous and strongly supported subgeneric clades (Figure 8.4). The general topology linking these clades is also constant in all trees with a single exception: the "*subtilis*" clade clusters with the "*ambiguum*" + "*minus* 1" +

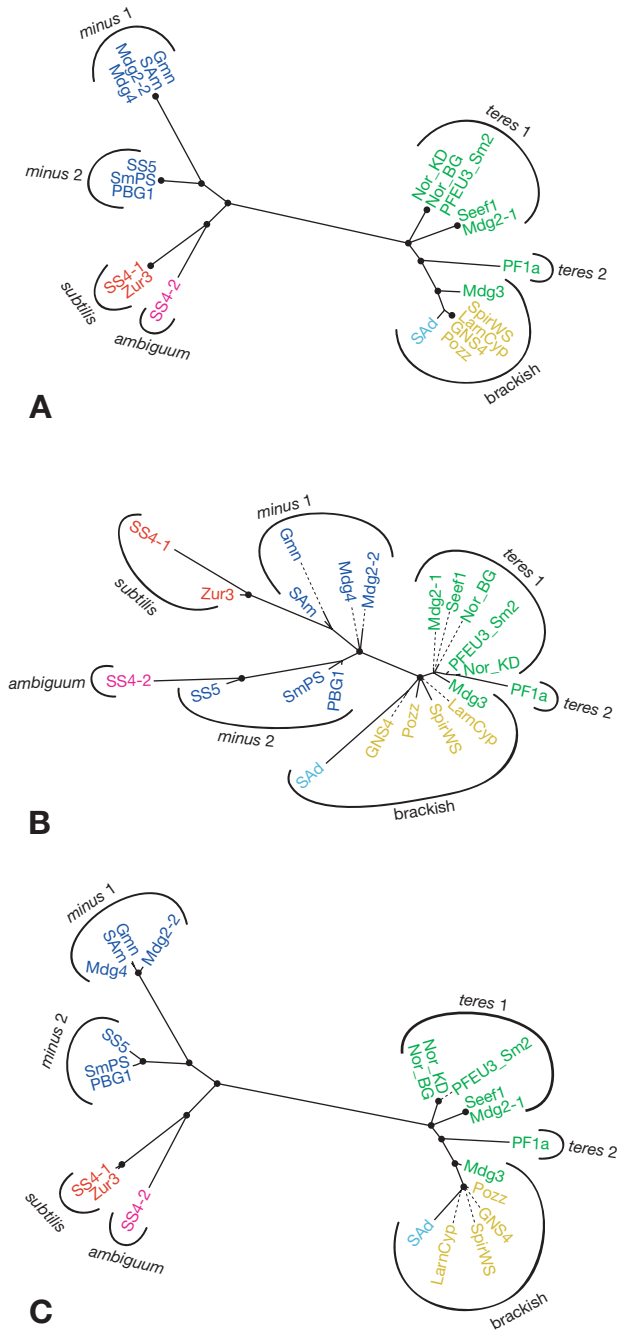


Figure 8.6 Unrooted Bayesian trees of the 22 populations described both morphologically and molecularly in this work. A, based on the 18S+ITS1+5.8S+ITS2+28S region sequence; B, based on morphological characters; C, based on the mixed character matrix. Circles indicate nodes supported by posterior probability values higher than 0.90. Labels around the trees represent subgeneric clades identified by the 18S rRNA analysis for comparison.

"*minus* 2" group in Bayesian trees, but it is the sister group of all other *Spirostomum* lineages in Maximum Likelihood trees. Nodes hierarchically above the 7 clades are supported by low statistical values even when they are present in all trees, and are thus relatively unreliable; the noteworthy exception is the association of clades "*teres* 1", "*teres* 2" and "brackish". Sequence identity values within clades are 99.4-100% in the 18S rRNA gene region (*Stentor* morphospecies exhibit similar values: 99.5-100%) and 95.0-100% in the ITS1-5.8S-ITS2-28S region.

The topology of the Bayesian tree based on the entire rRNA gene cluster sequenced (Figure 8.6A) shows similar associations among the populations, but the clade "*teres* 1" is further split in separated lineages. The morphology-based Bayesian tree (Figure 8.6B) has a slightly different topology. All populations of the molecular "*teres* 1" and "*teres* 2" clades, plus population Mdg3 of the "brackish" clade, are grouped. Population SS5 shows a substantial divergence from other "*minus* 2" populations, and the "*minus* 1" clade is split; also, the "*minus*" clades do no longer cluster together. Fewer nodes are strongly supported by statistical values in the morphological phylogenesis. The

combined tree (Figure 8.6C) is generally in accordance with molecular ones, but the position of population SAd within the "brackish" clade is different.

Discussion

Morphospecies identification. Populations for which we obtained only molecular data were conservatively labeled as "*Spirostomum* sp.". We classified the other 22 populations in morphospecies according to the most recent reviews[39,40] and the original descriptions.

Repak and Isquith[40] recognized 9 valid *Spirostomum* morphospecies. *Spirostomum inflatum*, *Spirostomum loxodes* and *Spirostomum caudatum* are characterized by conspicuous features related to cell shape (respectively, a dilated posterior half, a laterally-oriented anterior "beak" and a long and thin "tail") that were never observed in our populations. Moreover, *S. inflatum* and *S. loxodes* did not appear in published papers after their first description, and are thus either extremely rare, or were originally misidentified (Kahl himself suggested that *S. inflatum*, which he described as a novel taxon in 1932[41], could be an aberrant form of another species). *Spirostomum intermedium* was synonymized by Foissner and coworkers[39] with *S. minus*. We agree with this choice, because no distinctive character was ever described that allowed discrimination between the two morphospecies; the name is virtually absent in recent literature. On similar bases, we would argue that *Spirostomum ephrussi* should be synonymized with *Spirostomum teres*. *S. ephrussi* was never detected after its original description, and a few cells from the *S. teres* populations here characterized actually approaches the peristome : length ratio of 3/5 indicated as the only distinguishing character for *S. ephrussi*.

Spirostomum ambiguum, *S. minus* and *S. teres* are by far the most commonly found and studied morphospecies in the genus. According to keys and descriptions consulted, a single population here characterized could be unambiguously assigned to *S. ambiguum*, 7 populations to *S. minus* and 7 to *S. teres*.

Spirostomum yagiui and *Spirostomum dharwarensis* have complex taxonomical histories. In reporting their nomenclatural vicissitudes, Repak and Isquith[40] lumped together these morphospecies, at that time represented by a single population each and mainly distinguished by the length of their "single"-type MACs. We have found 4 new populations matching the

description of *S. yagiui*[42] and 1 matching the description of *S. dharwarensis*[43,44], and we can now provide arguments against their synonymization. Although these morphospecies can be unambiguously discriminated only by MAC features, we found that these are quite stable. *S. yagiui* possesses an elongated-type MAC that apparently circles between three different states, all of them unique to this morphospecies. On the other hand, *S. dharwarensis* exhibits a distinctive filiform MAC that is significantly longer and slenderer than all other "single"-type MACs. Since several smaller *S. minus* populations can be reliably distinguished from *S. teres* populations only by MAC type, our opinion is that this character is also sufficient to separate *S. yagiui* and *S. dharwarensis* morphospecies.

Spirostomum semivirescens rarely appears in literature, and is not mentioned in the otherwise comprehensive review of Repak and Isquith[40]. Nevertheless, we agree that it is a valid, probably rare, morphospecies, identifiable mainly by its most distinctive characteristic: the presence of zoochlorellae in the cytoplasm[22]. We did not find algae-harboring *Spirostomum* populations in this survey, though.

***Spirostomum subtilis* sp. n.** Populations Zur3 and SS4-1, despite being *minus*-like in appearance (slender, with a moniliform MAC and a peristome reaching approximately half the length of the body), share features that separate them from all others. They possess the highest length : width ratios (17-21 on average), and their mean length is intermediate between those of *S. minus* and *S. ambiguum*. Statistical analysis of characters collected in vivo shows that the overall morphology of these populations is significantly different from all others' (Figure 8.2).

Zur3 and SS4-1 additionally present one strong synapomorphy in their unique CG pattern: they are the only populations with CG stripes consisting of a single CG row (Figure 8.1). Since the number of CG rows in each stripe was never considered before in *Spirostomum* species identification, it is unclear if the description of this morphotype is entirely novel, or if it was lumped by other authors with that of *S. minus*. Our observations, though, provide evidence that the "single row" CG pattern is a distinctive and reliable feature present only in these 2 populations, among those investigated.

The morphological separation is substantiated by molecular data: Zur3 and SS4-1 have identical 18S+ITS1+5.8S+ITS2+28S sequences, and these differ from those of *S. minus* (2.3-3.4% different sites) and *S. ambiguum* (about 1.9%), the other morphospecies with a moniliform-type MAC. Finally, phylogenetic analyses show that these populations belong to a

quite different lineage of uncertain affinity (Figures 8.4, 8.6). For all these reasons, we propose to establish a new specific taxon, *Spirostomum subtilis* sp. n., to include populations Zur3 and SS4-1. A detailed description of its features follows at the end of the Discussion, and the tools for its unambiguous discrimination are presented in Figure 8.1 and the dichotomous key in Supplementary Text S8.1.

Species phylogeny and systematics. Our morphological data and statistical analyses clearly show that representatives of the genus *Spirostomum* are distributed along a generally continuous gradient in the morphospace. No single character is able to discriminate among all species. Despite this, a combination of living observations and nuclear features allows unambiguous identification of several groups that correspond to morphospecies (see also below). Molecular phylogeny is largely congruent with this morphological systematics, but there are some telling points in which they disagree.

S. ambiguum is a well defined, easily recognizable morphospecies whose monophyly is also strongly supported by molecular sequences. *S. minus*, instead, is morphologically variable and is separated in two clades in molecular trees, here called "*minus* 1" and "*minus* 2". These clades appear to be sister groups in molecular and combined trees (although this fact is undermined by low supporting statistical values), possibly meaning that *S. minus* is indeed a monophyletic group that may include two or more cryptic species. We could not find key characters able to reliably discriminate between subspecific groups in the absence of molecular information; morphological data analyses, both statistical and phyogenetic, also do not recognize them, showing instead a tendency for population SS5 to separate from all others *S. minus*. Hence, we prefer not to create new formal names. It is interesting to notice, however, that all known populations belonging to the "*minus* 2" clade harbor macronuclear bacterial symbionts (Figure 8.4), while only one reported population in the "*minus* 1" clade possessed a symbiont[25] (Figure 8.5), and it was a cytoplasmic one. Until data on the identity of the bacteria are available, though, this character should not be considered of diagnostic importance. From a molecular point of view, sequences of clade "*minus* 2" present several ambiguous sites, mostly concentrated in a variable region (bases 608-620) of the 18S rRNA gene.

The inclusion of the two populations Zur3 and SS4-1 in the novel species *S. subtilis* is strongly supported by both molecular and morphological trees. The position of *S. subtilis* in phylogenetic trees is not stable: in some analyses it appears to be more closely related to the

other morphospecies with moniliform MACs, while in others it is the first-branching lineage in the genus.

S. yagiui populations share total sequence identity, and they strongly cluster with *S. dharwarensis*. This gives credit to the choice of Repak and Isquith[40] to unify the two morphospecies; they are distinguishable but closely related, and in our phylogenetic analyses of morphological and mixed data the *S. dharwarensis* population SAd is actually nested within the *yagiui* group (although with very low statistical support).

S. teres sequences are scattered in at least three groups; the interpretation of clades "*teres* 1" and "*teres* 2" mirrors that of the two "*minus*" clades (possibly representing cryptic species in a monophyletic group), but the *S. teres* population Mdg3 is more closely related to *S. yagiui* and *S. dharwarensis* than to other *S. teres*. Also, the structure of "*teres*" clades becomes more complex when fast-evolving sequence regions are introduced in the analyses. Although minor variations in morphological characters can be associated to at least some of the molecularly-identified groups, more populations representing each lineage should be described, before drawing definitive conclusions. On the whole, *S. teres* appears as a relatively well defined but likely non-monophyletic morphospecies that includes several evolutionary lineages.

Character evolution. The low statistical confidence associated with deeper nodes in *Spirostomum* phylogeny – especially the uncertain positioning of *S. subtilis* – makes most inferences on character evolution unreliable. Nevertheless, something can be said about the evolution of MAC shape, which is arguably the most important diagnostic character separating morphospecies. In fact, the clade including all taxa with a "single"-type MAC (*S. teres*, *S. yagiui* and *S. dharwarensis*) is strongly supported. Our results do not allow assessment of which main MAC morphotype was the ancestral one for the genus, but suggest that the elongated and filiform MAC of *S. yagiui* and *S. dharwarensis*, respectively, are derived from the ellipsoid MAC of *S. teres*. In fact, it seems likely that the morphospecies *S. teres* does not represent a separated evolutionary branch, but defines instead a relatively ancestral set of characters, some of which became differentiated in the derived lineages of *S. dharwarensis* and *S. yagiui*.

Distribution and environment. An interactive map of geographic locations for many *Spirostomum* spp. samplings presented here and elsewhere can be found online at http://goo.gl/5C10jN. Molecular data are available only for a small fraction of the represented populations, but there is little correlations between sampling location and molecular distance.

Even populations from remote areas, like some of those reported here for tropical countries and northern Europe, may share a very high sequence similarity. Nevertheless, one clade – the "brackish" one – is intriguingly defined by an environmental character: it includes all the populations sampled in brackish waters, or at least in littoral zones that experience occasional saltwater inputs. All *S. yagiui* populations and one of the *S. teres* ones (Mdg3) are united in this clade (Figure 8.4), thus they probably originate from a common *S. teres*-like ancestor that invaded brackish environments from freshwater ones. *S. dharwarensis* is a more uncertain case: the population described here was found in freshwater, but in an estuarine area where marine and fresh waters likely mix, thus fitting the characteristics of the other "brackish" clade populations. On the other hand, the original Indian populations of *S. dharwarensis*[43,44] were sampled in freshwater environments, distant to the sea. More data are required to establish if this morphospecies is strictly a freshwater inhabitant or if it can be also found in brackish environments.

To support the relevance of water salinity as a diagnostic character, we observed that these 6 populations can survive in freshwater, if gradually acclimated to it, while all those originally sampled in freshwater die when the salinity of their medium increases (data not shown). As a side note, it is to be highlighted that in many old publications, and in some more recent reviews, *Spirostomum* species were considered either "freshwater" or "marine". In our experience, *Spirostomum* is never found in truly marine (above 33‰ salinity) environments, and we suspect that the original samples labeled as "marine" were indeed littoral ones with relatively low salinity. The genus *Spirostomum* likely inhabits only fresh- and brackish-waters.

Concluding remarks. We have presented the first systematic work on the ubiquitous genus *Spirostomum* based on both morphological and molecular characters. According to literature and our results, there are 8 valid morphospecies, one of which is novel, although some may be complexes of sibling species. Morphological and molecular data do agree to a large extent, and a small set of unambiguous characters is sufficient to discriminate among taxa. New data on *S. caudatum* and *S. semivirescens*, the morphospecies still lacking a molecular characterization, should be welcomed in the future to complete this survey.

In addition to our conclusions on the genus *Spirostomum*, we provided a case for the relevance of genera as valuable targets for systematics studies. A lot of ciliate genera are easy to discriminate, and many of them are quite common, but species identification is often a bigger

issue. Whenever possible, we encourage taxonomists to present updated, multidisciplinary and quantitative data on several species at the same time and perform analyses aiming to organize the extant knowledge, with the goal of investigating biodiversity while providing non-taxonomists with the most practical tools for recognizing species.

Revised systematics of *Spirostomum*

In this paragraph we list schematic descriptions of distinguishable *Spirostomum* morphospecies based on the integration of literature observations and our novel morphological and molecular data. The set of characters sufficient to identify each morphospecies is in bold. This section is also presented as a dichotomous key in Supplementary Text S8.1. We strictly limited our discussion of nuclear features to the results of Feulgen reactions, but it is worth to stress that the fundamental MAC type character is usually quite recognizable also in living cells (Figures 8.1A-C).

Spirostomum Ehrenberg, 1834[1]. Medium to very large ciliates (150 µm – several mm) with a worm-like shape, cylindrical or slightly flattened cell body, length : width ratio ranging from 5 to more than 30. Color varies from faint brown or yellowish to very dark-greenish. A single row of well-developed oral membranelles defines the left side of the long, thin peristomial field, which runs parallel to the main body axis from the anterior end to the cytostome, located at 1/4-2/3 of the body length. Somatic kineties (10-50 in number) homogeneously distributed, parallel to the main body axis, but strongly spiraled when the organism contracts due to myonemes action. One stripe of packed cortical granule rows (1-6) between each kinety pair; cortical granules may be of the same or different size. Contractile vacuole always single and posteriorly located, with a collecting canal reaching the anterior end. Macronucleus single (ellipsoid, elongated or filiform) or moniliform. Micronuclei variable in number but generally small (1-3 µm) and associated with the macronucleus. The monophyly of the genus is strongly supported. Common in fresh water, can be found also in brackish environments. Cosmopolitan.

Spirostomum ambiguum Ehrenberg, 1834[1]. [syn: *Trichoda ambiguum* Müller, 1786; *S. ambiguum* var. *major* Roux, 1901[45]] **900 µm – several mm long**. Length : width ratio about 9-17. 15-25 kineties on each side; heterogeneous, numerous (4-5) CG rows per stripe. **Peristome always longer than 1/2 of the body length, often reaching 2/3**. CV much shorter than

body length, rarely exceeding 1/10. The color depends on cytoplasmic granules. Moniliform MAC with 12-50 (avg. 15-25) nodules not exceeding 35-45 µm in length when stained by Feulgen reaction. Numerous (up to 100) MICs 1-2 µm long. Monophyletic. Only found in freshwater. Reported in central and northern Europe, England, Russia, central Africa, USA, Jamaica, India and Japan. It sometimes harbors prokaryotic symbionts in the MAC.

Spirostomum caudatum (Müller, 1786) Delphy, 1939[46]. [syn. *Enchelis caudata* Müller, 1786; *Uroleptus filum* Ehrenberg, 1833; *Spirostomum filum* Dujardin, 1841; *S. teres* var. *caudatum* Zacharias, 1903] **Tapering, thin posterior tail**. 200-700 (avg. 200-400) µm long. 14-16 kineties on each side. **Peristome about 1/4 of the body length**. Ellipsoid MAC. No molecular sequence available. Found in fresh- and saltwater (?). Reported in central Europe, central Africa and Korea.

Spirostomum dharwarensis Desai, 1966[43]. 300-550 (avg. 400) µm long. Length : width ratio about 8-14 (avg. 11). 7-13 (avg. 10) kineties on each side; usually homogeneous 3-4 CG rows per stripe. Peristome variable, from less than 1/2 of the body length to about 2/3. CV usually less than 1/5 of the body length. Dark cytoplasm. **Filiform, convoluted MAC (length : width ratio always > 10, usually > 20; uniform diameter) about 100-200 x 5-10 µm when stained by Feulgen reaction**. Several (1-7) MICs 1-3 µm long. Only one molecular sequence available. Found in freshwater, once in estuarine environment. Reported in southern Africa and India.

Spirostomum minus Roux, 1901[45]. [syn: *S. ambiguum* var. *minor* Roux, 1901; *S. intermedium* Kahl, 1932] **350-900 (avg. 450-600) µm long. Length : width ratio about 7-15 (avg. 9-13)**. 6-12 (avg. 8-10) kineties on each side; homogeneous or heterogeneous CG rows, variable in number per stripe (2-4). Peristome about 1/2 of the body length. CV usually less than 1/5 of the body length. **Moniliform MAC** with 5-25 (avg. 10-15) nodules not exceeding 30-40 µm in length when stained by Feulgen reaction. Variable number (up to 20) of 1.5-3.5 µm long MICs. Molecular analyses identify two clades, probably representing cryptic species with no reliable morphological autapomorphy; the morphospecies might be monophyletic. Commonly found in freshwater. Reported in Europe, England, southern and central Africa, Madagascar, USA, Brazil, China, India, possibly Thailand and Australia. It relatively often harbors prokaryotic symbionts in cytoplasm or MAC.

Spirostomum semivirescens Perty, 1852[47]. **Numerous zoochlorellae in the cytoplasm**. 600-2,000 (avg. 1,200-1,300) µm long; **the posterior half is usually encased in a mucilaginous coating**. Length : width ratio about 17-40 (avg. 30-40). 14-15 kineties on each side. Peristome about 1/2 of the body length. Moniliform MAC with about 12 nodules. Only found in freshwater. No molecular sequence available. Reported in central Europe, England, Russia and Japan.

Spirostomum subtilis sp. n. 700-1,000 µm long. **Length : width ratio about 14-24**. 9-12 kineties on each side; a **single, homogeneous CG row per stripe**. Peristome about 1/2 of the body length. CV often conspicuous, up to 1/3 of the body length, contrasting with the dark cytoplasm of the anterior part. Moniliform MAC with 15-24 nodules not exceeding 20-25 µm in length when stained by Feulgen reaction. Several (1-6) MICs 2-3 µm long. Monophyletic. Only found in freshwater. Reported in central and northern Europe. Type location and features of the type population Zur3 detailed in Table S1. Type materials: one slide of permanent Feulgen stained specimens collected in Zurich (population Zur3) deposited in the collection of the Museo della Scienza e del Territorio della Certosa di Calci (Calci, Pisa, Italy); extracted genomic DNA available upon request. 18S+ITS1+5.8S+ITS2+28S region sequence deposited at the ENA database (accession number: HG939550).

Spirostomum teres Cláparéde and Lachmann, 1858-1859[48]. [syn: *S. ephrussi* Delphy, 1939[46]] 150-650 (avg. 250-450) µm long. Length : width ratio about 5-16 (avg. 8-12). 5-15 (avg. 7-10) kineties on each side; usually homogeneous CG rows, variable in number per stripe (2-4). Peristome from 1/3 to slightly more than 1/2 of the body length. CV usually less than 1/5 of the body length. Often brownish. **Ellipsoid MAC (length : width ratio $<$ 5) in the middle sector of the body, about 20-50 x 5-20 µm when stained by Feulgen reaction**. A few (1-3) MICs 1-2 µm long. Molecular analyses suggest that this morphospecies include phylogenetically diverse lineages, some of which are more closely related to *S. yagiui* and *S. dharwarensis*; no reliable morphological autapomorphy has yet been detected for these lineages. Found in both fresh- and brackish-water environments. Reported in Europe, central Africa, Madagascar, USA, Brazil, Caspian Sea, India, China and Korea. It sometimes harbors cytoplasmic prokaryotic symbionts.

Spirostomum yagiui Shigenaka, 1959[42]. 240-500 (avg. 300-400) µm long. Length : width ratio about 6-17 (avg. 8-14). 6-12 (avg. 7-10) kineties on each side; usually homogeneous CG

rows, variable in number per stripe (2-4). Peristome variable, from about 1/3 of the body length to more than 1/2. CV less than 1/5 of the body length. Usually brownish. **Elongated MAC (length : width ratio > 5) exhibiting different shapes during cell cycle, about 35-90 x 5-15 µm when stained by Feulgen reaction**. A few (up to 6) MICs about 2 µm long. Likely monophyletic. **Only found in brackish water, or in close proximity to saltwater basins**. Found in Mediterranean Sea islands (Sicily and Cyprus), northern Europe, Russia and Japan.

References

1. Ehrenberg CG. (1838) Die Infusionsthierchen als Vollkommene Organismen. Leopold Voss, Leipzig
2. Fokin SI. (2004) A brief history of ciliate studies (late XVII – the first third of the XX century). Protistology. 3: 283-296
3. Adl SM, Simpson AGB, Lane CE, Lukeš J, Bass D, *et al.* (2012) The revised classification of eukaryotes. J Eukaryotic Microbiol. 59: 429-493
4. Lynn DH. (2008) The ciliated protozoa. Characterization, classification, and guide to the literature, Third Edition, Springer, Germany
5. Chantangsi C, Lynn DH, Rueckert S, Prokopowicz AJ, Panha S, Leander BS. (2013) *Fusiforma themisticola* n. gen., n. sp., a new genus and species of apostome ciliate infecting the hyperiid amphipod *Themisto libellula* in the Canadian Beaufort Sea (Arctic Ocean), and establishment of the Pseudocolliniidae (Ciliophora, Apostomatia). Protist. 164: 793-810
6. Foissner W. (2013) Description of *Glaucomides bromelicola* n. gen., n. sp. (Ciliophora, Tetrahymenida), a macrostome forming inhabitant of bromeliads (Bromeliaceae), including redescriptions of *Glaucoma scintillans* and *G. reniformis*. J Eukaryotic Microbiol. 60: 137-157
7. Modeo L, Petroni G, Lobban CS, Verni F, Vannini C. (2013) Morphological, ultrastructural, and molecular characterization of *Euplotidium rosati* n. sp. (Ciliophora, Euplotida) from Guam. J Eukaryotic Microbiol. 60: 25-36
8. Pan H, Gao F, Lin X, Warren A, Song W. (2013a) Three new *Loxophyllum* species (Ciliophora: Pleurostomatida) from China with a brief review of the marine and brackish *Loxophyllum* species. J Eukaryotic Microbiol. 60: 44-56
9. Pan H, Li L, Al-Rasheid KAS, Song W. (2013b) Morphological and molecular description of three new species of the cyrtophorid genus *Chlamydodon* (Ciliophora, Cyrtophoria). J Eukaryotic Microbiol. 60: 2-12
10. Pan Y, Li J, Li L, Hu X, Al-Rasheid KAS, Warren A. (2013c) Ontogeny and molecular phylogeny of a new marine ciliate genus, *Heterokeronopsis* g. n. (Protozoa, Ciliophora, Hypotricha), with description of a new species. Eur J Protistol. 49: 298-311
11. Park KM, Jung JH, Min GS. (2013) Morphology, morphogenesis, and molecular phylogeny of *Anteholosticha multicirrata* n. sp. (Ciliophora, Spirotrichea) with a note on morphogenesis of *A. pulchra* (Kahl, 1932) Berger, 2003. J Eukaryotic Microbiol. 60: 564-577

12. Modeo L, Fokin SI, Boscaro V, Andreoli I, Ferrantini F, *et al.* (2013a) Morphology, ultrastructure, and molecular phylogeny of the ciliate *Sonderia vorax* with insights into the systematics of order Plagiopylida. BMC Microbiol. 13: 40
13. Boscaro V, Fokin SI, Verni F, Petroni G. (2012) Survey of *Paramecium duboscqui* using three markers and assessment of the molecular variability in the genus *Paramecium*. Mol Phylogenet Evol. 65: 1004-1013
14. Strüder-Kypke MC, Wright ADG, Fokin SI, Lynn DH. (2000) Phylogenetic relationships of the genus *Paramecium* inferred from small subunit rRNA gene sequences. Mol Phylogenet Evol. 14: 122-130
15. Chantangsi C, Lynn DH. (2008) Phylogenetic relationships within the genus *Tetrahymena* inferred from the cytochrome *c* oxidase subunit 1 and the small subunit ribosomal RNA genes. Mol Phylogenet Evol. 49: 979-987
16. Kher CP, Doerder FP, Cooper J, Ikonomi P, Achilles-Day U, *et al.* (2011) Barcoding *Tetrahymena*: discriminating species and identifying unknowns using the cytochrome *c* oxidase subunit I (cox-1) barcode. Protist. 162: 2-13
17. Petroni G, Dini F, Verni F, Rosati G. (2002) A molecular approach to the tangled intrageneric relationships underlying phylogeny in *Euplotes* (Ciliophora, Spirotrichea). Mol Phylogenet Evol. 22: 118-130
18. Bradley MW, Esteban GF, Finlay BJ. (2010) Ciliates in chalk-stream habitats congregate in biodiversity hot spots. Res Microbiol. 161: 619-625
19. Finlay BJ, Esteban GF. (1998) Freshwater protozoa: biodiversity and ecological function. Biodiversity Conserv. 7: 1163-1186
20. Nałecz-Jawecki G, Sawicki J. (1999) Spirotox – a new tool for testing the toxicity of volatile compounds. Chemosphere. 38: 3211-3218
21. Twagilimana L, Bohatier J, Groliere CA, Bonnemoy F, Sargos D. (1998) A new low-cost microbiotest with the protozoan *Spirostomum teres*: culture conditions and assessment of sensitivity of the ciliate to 14 pure chemicals. Ecotoxicol Environ Saf. 41: 231-244
22. Esteban GF, Bradley MW, Finlay BJ. (2009) A case-building *Spirostomum* (Ciliophora, Heterotrichida) with zoochlorellae. Eur J Protistol. 45: 156-158
23. Fokin SI, Schweikert M, Brümmer F, Görtz HD. (2005) *Spirostomum* spp. (Ciliophora, Protista), a suitable system for endocytobiosis research. Protoplasma. 225: 93-102
24. Schrallhammer M, Ferrantini F, Vannini C, Galati S, Schweikert M, *et al.* (2013) "*Candidatus* Megaira polyxenophila" gen. nov., sp. nov.: considerations on evolutionary history, host range and shift of early divergent rickettsiae. PLoS One. 8: e72581
25. Vannini C, Boscaro V, Ferrantini F, Benken KA, Mironov TI, *et al.* (2014) Flagellar movement in two bacteria of the family *Rickettsiaceae*: a re-evaluation of motility in an evolutionary perspective. PLoS One. 9: e87718
26. Hirt RP, Dyal PL, Wilkinson M, Finlay BJ, Roberts DM, Embley TM. (1995) Phylogenetic relationships among karyorelictids and heterotrichs inferred from small subunit rRNA sequences: resolution at the base of the ciliate tree. Mol Phylogenet Evol. 4: 77-87
27. Schmidt SL, Treuner T, Schlegel M, Bernhard D. (2007b) Multiplex PCR approach for species detection and differentiation within the genus *Spirostomum* (Ciliophora, Heterotrichea). Protist. 158: 139-145
28. Jang SW, Kwon CB, Shin MK. (2012) First records of two *Spirostomum* ciliates (Heterotrichea: Heterotrichida: Spirostomidae) from Korea. Anim Syst Evol Divers. 28: 29-35
29. Fernandes NM, da Silva Neto ID. (2013) Morphology and 18S rDNA gene sequence of *Spirostomum minus* and

Spirostomum teres (Ciliophora: Heterotrichea) from Rio de Janeiro, Brazil. Zoologia. 30: 72-79

30. Boscaro V, Fokin SI, Schrallhammer M, Schweikert M, Petroni G. (2013) Revised systematics of *Holospora*-like bacteria and characterization of "*Candidatus* Gortzia infectiva", a novel macronuclear symbiont of *Paramecium jenningsi*. Microb Ecol. 65: 255-267
31. Skovorodkin IN. (1990) A device for immobilizing biological objects in the light microscope studies. Tsitologiya. 32: 301-302
32. Andreoli I, Mangini L, Ferrantini F, Santangelo G, Verni F, Petroni G. (2009) Molecular phylogeny of unculturable Karyorelictea (Alveolata, Ciliophora). Zool Scr. 38: 651-662
33. Guindon S, Gascuel O. (2003): A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 52: 696-704
34. Ludwig W, Strunk O, Westram R, Richter L, Meier H, *et al.* (2004) ARB: a software environment for sequence data. Nucleic Acids Res. 32: 1363–1371
35. Ronquist F, Huelsenbeck JP. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19: 1572-1574
36. Darriba D, Taboada GL, Doallo R, Posada D. (2012) jModelTest 2: more models, new heuristics and parallel computing. Nat Methods. 9: 772
37. Miao M, Song W, Clamp JC, Al-Rasheid KAS, Al-Khedhairy AA, Al-Arift S. (2009) Further consideration of the phylogeny of some "traditional" heterotrichs (Protista, Ciliophora) of uncertain affinities, based on new sequences of the small subunit rRNA gene. J Eukaryotic Microbiol. 56: 244-250
38. Schmidt SL, Foissner W, Schlegel M, Bernhard D. (2007a) Molecular phylogeny of the Heterotrichea (Ciliophora, Postciliodesmatophora) based on small subunit rRNA gene sequences. J Eukaryotic Microbiol. 54: 358-363
39. Foissner W, Berger H, Kohmann F. (1992) Taxonomische und ökologische Revision der Ciliaten des Saprobiensystems – Band II: Peritrichida, Heterotrichida, Odontostomatida. Informationsberichte des Bayer. Landesamtes für Wasserwirtschaft, Deggendorf
40. Repak AJ, Isquith IR. (1974) The systematics of the genus *Spirostomum* Ehrenberg, 1838. Acta Protozool. 12: 325-333
41. Kahl A. (1932) Urtiere oder Protozoa. I: Wimpertiere oder Ciliata (Infusoria). 2. Spirotricha. In: Dahl F (ed), Die Tierwelt Deutschland, Gustav Fischer, Jena, pp 437-440
42. Shigenaka Y. (1959) A new marine ciliate, *Spirostomum yagiui* n. sp. Zool Mag Tokyo. 68: 368-372
43. Desai RN. (1966) Nuclear apparatus and binary fission in *Spirostomum dharwarensis* n. sp. Experientia. 22: 35-36
44. Seshachar BR, Padmavathi PB. (1956) The cytology of a new species of *Spirostomum*. J Protozool. 3: 145-150
45. Roux. (1901) Faune infusiorenne de eaux stagnants des environs de Genve. Geneva
46. Delphy J. (1939) Sur les Spirostomes. Archs néerl Zool. 3: 141-145
47. Perty M. (1852) Zur Kenntniss kleinster Lebensformen nach Bau, Funktion, Systematik, mit Specialverzeichniss der in der Schweiz beobachteten. Jent & Reinert, Bern
48. Claparède E, Lachmann J. (1858-1859) Études sur les infusoires et les rhizopodes. Mém Inst Nat Genévois. 5: 1-260

Supplementary Text S8.1 - Dichotomous key for *Spirostomum* morphospecies identification

(1) Macronucleus type
- Single……**(2)**
- Moniliform……**(5)**

(2) Posterior end shape
- Elongated in a thin "tail"……*Spirostomum caudatum*
  200-700 (avg. 200-400) µm long. 14-16 kineties on each side. Peristome about 1/4 of the body length. Ellipsoid MAC
- Rounded or truncated……**(3)**

(3) Feulgen-stained macronucleus length : width ratio
- Less than 5 on average……*Spirostomum teres*
  150-650 (avg. 250-450) µm long. Length : width ratio about 5-16 (avg. 8-12). 5-15 (avg. 7-10) kineties on each side; usually homogeneous CG rows variable in number per stripe (2-4). Peristome from 1/3 to slightly more than 1/2 of the body length. CV usually less than 1/5 of the body length. Often brownish. Ellipsoid MAC in the middle sector of the body, about 20-50 x 5-20 µm when stained by Feulgen reaction. A few (1-3) MICs 1-2 µm long
- More than 5 on average……**(4)**

(4) Feulgen-stained macronucleus length
- about 35-90 µm……*Spirostomum yagiui*
  240-500 (avg. 300-400) µm long. Length : width ratio about 6-17 (avg. 8-14). 6-12 (avg. 7-10) kineties on each side; usually homogeneous CG rows, variable in number per stripe (2-4). Peristome variable, from about 1/3 of the body length to more than 1/2. CV less than 1/5 of the body length. Usually brownish. Elongated MAC exhibiting different shapes during cell cycle. A few (up to 6) MICs about 2 µm long
- about 100-200 µm ……*Spirostomum dharwarensis*
  300-550 (avg. 400) µm long. Length : width ratio about 8-14 (avg. 11). 7-13 (avg. 10) kineties on each side; usually homogeneous 3-4 CG rows per stripe. Peristome variable, from less than 1/2 of the body length to about 2/3. CV usually less than 1/5 of the body length. Dark cytoplasm. Filiform, convoluted MAC. Several (1-7) MICs 1-3 µm long

(5) Zoochlorelle in cytoplasm:
- Present……*Spirostomum semivirescens*
  600-2,000 (avg. 1,200-1,300) µm long; the posterior half is usually encased in a mucilaginous coating. Length : width ratio about 17-40 (avg. 30-40). 14-15 kineties on each side. Peristome about 1/2 of the body length. MAC with about 12 nodules
- Absent……**(6)**

(6) Cortical granule rows per stripe:
- Almost always 1……*Spirostomum subtilis*
  700-1,000 µm long. Length : width ratio about 14-24. 9-12 kineties on each side; a single homogeneous CG row per stripe. Peristome about 1/2 of the body length. CV often conspicuous, up to 1/3 of the body length, contrasting with the dark cytoplasm of the anterior part. MAC with 15-24 nodules not exceeding 20-25 µm in length when stained by Feulgen reaction. Several (1-6) MICs 2-3 µm long
- 2 or more……**(7)**

(7) Length:
- Below 600 µm on average……*Spirostomum minus*
  350-900 (avg. 450-600) µm long. Length : width ratio about 7-15 (avg. 9-13). 6-12 (avg. 8-10) kineties on each side; homogeneous or heterogeneous CG rows, variable in number per stripe (2-4). Peristome about 1/2 of the body length. CV usually less than 1/5 of the body length. MAC with 5-25 (avg. 10-15) nodules not exceeding 30-40 µm in length when stained by Feulgen reaction. Variable number (up to 20) of 1.5-3.5 µm long MICs
- Over 900 µm on average……*Spirostomum ambiguum*
  900 µm – several mm long. Length : width ratio about 9-17. 15-25 kineties on each side; heterogeneous, numerous (4-5) CG rows per stripe. Peristome always longer than 1/2 of the body length, often reaching 2/3. CV much shorter than body length, rarely exceeding 1/10. The color depends on cytoplasmic granules. MAC with 12-50 (avg. 15-25) nodules not exceeding 35-45 µm in length when stained by Feulgen reaction. Numerous (up to 100) MICs 1-2 µm long

PART IV

Ciliate Transcriptomics

Just like those on symbionts, studies on ciliates should only *start* with characterizations, before proceeding toward larger goals. And just as for symbionts, genomes would be a very useful source of information for their eukaryotic hosts, too.

However, the obstacles are even greater. Eukaryotic genomes are large, complex, difficult to assembly and annotate, and the results are not always worth the (costly) effort. Ciliate genomics will certainly be a very important area in the future, and several groups around the world are enrolling in that enterprise, but available genomes are few and far between at present. A similar situation applies to all protist groups, and some years ago an alternative approach was proposed to avoid many of the hindrances of genomics while keeping a similar level of depth: transcriptomics. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) led by Prof. Patrick Keeling and other experts on protist biodiversity started as a collaboration of numerous laboratories around the world; the results, published in 2014, were exceptionally good, with almost a thousand transcriptomes obtained from lineages whose molecular profiles were previously represented by only a handful of single-gene sequences.

The MMETSP data are publicly available, and the groups involved in extracting total RNAs from mass cultures have already started to analyze the transcripts. At the very beginning of my PhD research I performed 5 total RNA extractions on the ciliate species agreed upon by the Pisa unit with a representative scholar of MMETSP, Prof. Denis Lynn, also author of the leading guide on ciliate classification. The transcripts obtained in Pisa were pooled with others and presented with a phylogenetic purpose in a paper (Gentekaki et al., 2014 – see Appendix) providing the first multi-gene tree of ciliates.

Although productive, my involvement in MMETSP stressed one of the limits of a traditional transcriptomic approach: the necessity for a very high amount of input material. The number of required organisms poses serious inconveniences even for easily cultivable species, and completely prevents the study of that vast majority of ciliates that cannot be grown in laboratory condition. Taking advantage of the transcriptomics' benefits, but with less demanding requirements for performing the experiments, would have been a huge boost for ciliate studies.

With that goal in mind, I departed for the University of British Columbia (Vancouver) for a collaboration with Profs. Keeling and Lynn. During two travels there, in the summers of 2013 and 2014, I and the post-doctoral fellows of Keeling lab were able to optimize and test a protocol for single-cell transcriptomics on ciliates and other unicellular protists. The possible applications are countless, and **Chapter 9** briefly summarizes the success of our pilot study.

Chapter 9.

Single-cell RNA-seq

Original publication:

Single-cell transcriptomics for microbial eukaryotes

Martin Kolisko, Vittorio Boscaro, Fabien Burki, Denis H Lynn, Patrick J Keeling

Current Biology 24(22): R1081-R1082. November 2014.

One of the greatest hindrances to a comprehensive understanding of microbial genomics, cell biology, ecology, and evolution is that most microbial life is not in culture. Solutions to this problem have mainly focused on whole-community surveys like metagenomics, but these analyses inevitably loose information and present particular challenges for eukaryotes, which are relatively rare and possess large, gene-sparse genomes[1,2]. Single-cell analyses present an alternative solution that allows for specific species to be targeted, while retaining information on cellular identity, morphology, and partitioning of activities within microbial communities[2]. Single-cell transcriptomics, pioneered in medical research[3], offers particular potential advantages for uncultivated eukaryotes, but the efficiency and biases have not been tested. Here we describe a simple and reproducible method for single- cell transcriptomics using manually isolated cells from five model ciliate species; we examine impacts of amplification bias and contamination, and compare the efficacy of gene discovery to traditional culture-based transcriptomics. Gene discovery using single-cell transcriptomes was found to be comparable to mass-culture methods, suggesting single-cell transcriptomics is an efficient entry point into genomic data from the vast majority of eukaryotic biodiversity.

Ciliates are good models to evaluate single-cell transcriptomics for uncultivated microbial eukaryotes, because they allow many of the potential problems to be examined in a relatively well-controlled fashion. Several species with well-curated whole-genomes or deep-coverage transcriptomes are available to provide points of reference[4-6]. Moreover, cell size varies by orders of magnitude between species[7]. Also, because ciliates are obligate heterotrophs[7], cultures are typically 'contaminated' with food bacteria or other eukaryotes, and even isolated cells can retain potentially misleading remnants of partially digested cells of foreign origin. We have selected five species representing a wide variety of cell sizes (from *Condylostoma* at 500 μm to *Tetrahymena* at 50 μm) and derived from different environments. To evaluate contamination, we include species in axenic culture (*Tetrahymena*), with *endosymbiotic* bacteria (*Euplotes harbours Protistobacter heckmanni*), feeding on mixed prey from a natural environment (*Blepharisma*), or feeding on defined eukaryotes (*Condylostoma* feeding on the diatom *Phaeodactylum tricornutum*, and *Euplotes* and *Paramecium* feeding on the green alga *Dunaliella tertiolecta*). From each species, individual cells were manually isolated and washed, and single-cell cDNA libraries were constructed and sequenced (detailed methods are available in Supplementary Text S9.1).

We assessed three important characteristics of high-throughput sequence datasets: bias introduced by library construction and sequencing; contamination levels; and the effectiveness of gene discovery. Assembly and annotation resulted in between 12,030 and 39,221 contigs (Figure 9.1A). Larger cells yielded more contigs, potentially due to differential bias for reads mapping to individual contigs (Figure 9.1B). For example, in *Condylostoma* (comparatively large cells) the most abundantly represented contig (a cystein protease) accounted for 8% of reads, whereas in *Tetrahymena* (comparatively small cells) 90% of the reads mapped to the LSU rRNA. In contrast, contamination levels were relatively even and uniformly low (2.1%–4.27%; Figure 9.1A). Putative contaminants were most often related to bacteria, but generally not to a single type. In *Euplotes*, contamination from its *endosymbiotic* bacteria was also low — only 0.4% of sequences were identified as being from *Burkholderiacaea*. Interestingly, no sequences from known eukaryotic prey were found, despite the fact that they should be unaffected by polyA selection. It is possible that prey RNA (perhaps unlike DNA) is quickly cleared from feeding cells, so single-cell transcriptomes may offer a manageable solution to contamination in complex natural communities.

To evaluate the success of gene discovery, we first examined the recovered proportion of two defined collections of housekeeping genes: 20 aa-tRNA synthetases, and 248 core eukaryotic proteins[8]. Recovery rates varied between 90 and 100% for aa-tRNA synthetases and 66 and 94% for the 248 core-gene set (Figure 9.1A). Because some of the 248-gene set may not be present in ciliates, this may be a slight underestimate. Second, we compared the single-cell transcriptomes to equivalent data from mass-culture, which was done in three different ways depending on the best available comparators. For *Condylostoma*, a direct comparison with a transcriptome from the same strain[6,9] revealed the single-cell transcriptome actually recovered more unique contigs than the mass- culture transcriptome. Both datasets included more than 3,000 unique contigs, but shared ~19,000 contigs in common, suggesting they comparably reflect the expression status of the cell. No transcriptome is available from the same strain/ species of *Blepharisma*, *Euplotes*, and *Paramecium*, so their transcriptomes were compared with those of closely related congeners against the most similar available genome or complete transcriptome. Here, the single-cell transcriptome yielded about 90% of the genes recovered by culture-based methods and, except for *Paramecium*, both datasets contained similar sets of orthologues when compared to the reference. For *Tetrahymena*, a genome of the same species is available[4], so single-cell and culture-

A

	Contamination levels		Estimation of transcriptome completness		Comparison with culture-based transcriptome
Species (cell lenght)	# Total contigs	# Contaminant contigs	tRNA synthetase	248 gene set	Single cell vs. Culture
Condylostoma magnum (500μm)	39221	1646 (4.2%)	20/20	232/248	102% (19033/3817/3257)
Blepharisma sp. (150μm)	35221	743 (2.12%)	20/20	232/248	96% (3467/774/921)
Euplotes woodruffi (120μm)	23491	1013 (4.31%)	18/20	197/248	80% (2510/986/1873)
Paramecium duboscqui (100μm)	18005	538 (2.99%)	20/20	216/248	90% (3835/3539/4366)
Tetrahymena thermophile (50μm)	12030	351 (2.92%)	19/20	164/248	Single cell: 33% Culture based: 77%

B

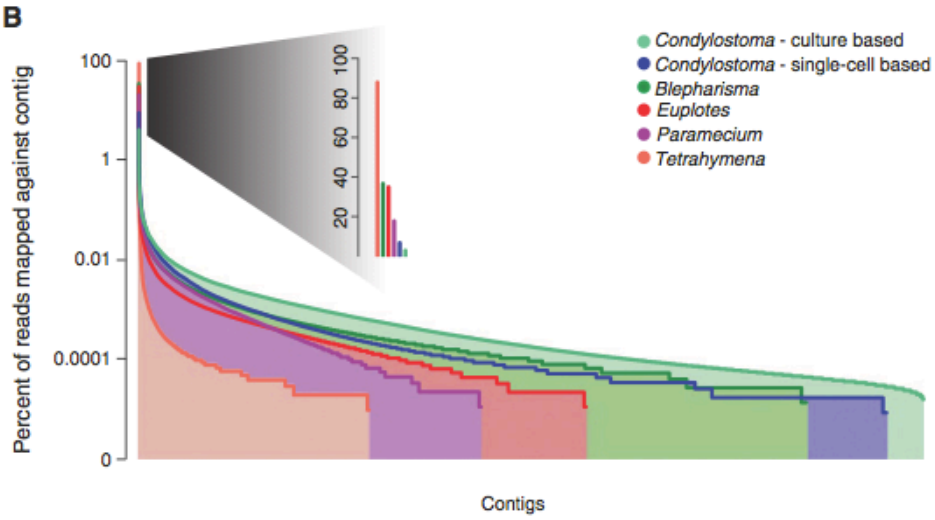


Figure 9.1 Comparison of single-cell and mass-culture transcriptomes from five species. (A) Summary of the general characteristics of the single-cell transcriptome data sets, including levels of identifiable contamination (left columns, see also Figure 9.2), estimation of completeness by comparison with two sets of generally universally present housekeeping genes (middle columns), and a direct comparison of the efficiency of gene discovery with culture-based transcriptome data (right column; the three numbers in brackets represent from left to right: the transcripts shared, those unique to the single-cell data, and those unique to the culture data). (B) Summary of bias (over-representation) of five single-cell transcriptomes (colour coded to the right). The graph shows a log-scale bar chart with the percentage of reads mapping to each contig from each species. Along the X-axis are bars that each represent a contig (colour coded depending on the species). Because there are >10,000 contigs per species, they are packed closely together and are not each visible as discrete bars (except in the blow-up of the top end). The height of each bar (the Y-axis) is a log-scale percentage of reads that map against that particular contig. Contigs are sorted so that moving from left to right corresponds to the largest number of the reads mapped to lowest number of the reads mapped. The blow-up expands the upper portion showing the most over-represented contigs from each species, which vary from as low as 8% in *Condylostoma* to as high as 90% in *Tetrahymena*.

based transcriptomes were mapped directly to its full gene set. The single-cell transcriptome was less efficient due to the rRNA bias described above, but most reads and contigs nevertheless mapped to the genome (Figure 9.2), recovering ~11,000 genes, or 33% of the genome (compared with 77% from a comparable number of reads from a culture-based transcriptome).

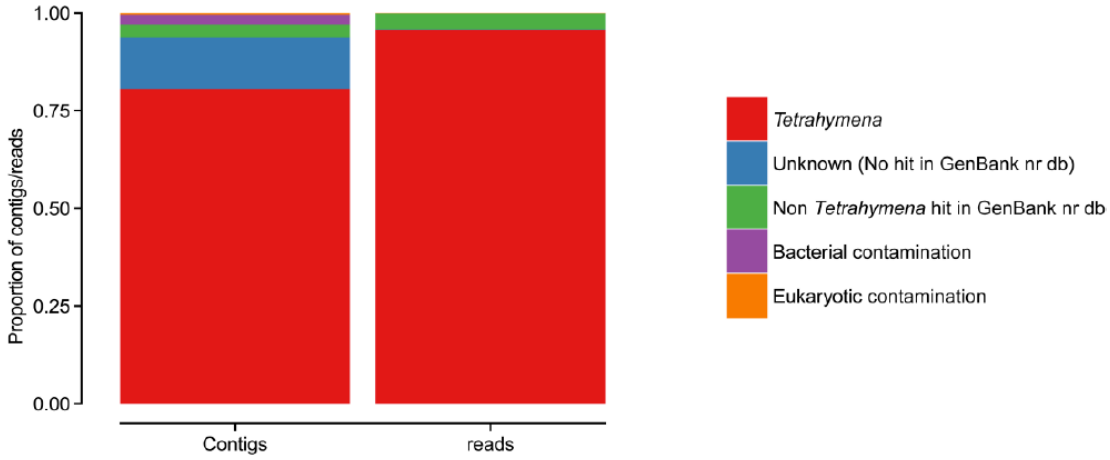


Figure 9.2 Proportion of the sequence data that map to *Tetrahymena*, unknown and putative contaminations in contigs and raw reads.

Single-cell RNAseq is a powerful method to generate large-scale datasets from uncultivated microbial eukaryotes. Comparing data from ciliates revealed some biases, but even in the excessively biased case of *Tetrahymena* (where 90% of the sequence was uninformative), about one-third of the known genes in the genome were still identified. The majority of the single-cell transcriptomes were comparable to those from mass-culture. Because the total number of genes expressed in one cell at one time must be lower than those expressed collectively in cells in mass-culture, these results suggest the method is a very efficient way to recover transcripts in isolated cells. The parsimonious nature of this approach is also noteworthy: each transcriptome required resources comparable to cloning and sequencing 4–5 protein- coding genes, but instead generated tens of thousands of genes. Single-cell transcriptomes are readily applicable to a wide range of questions, the most obvious being the acquisition of data from species that are uncultivated or in complex culture (e.g., obligate predators), or that have uncultured life cycle stages (e.g., parasites). Enabling expression profiling and analysis of genome-wide data from these abundant but poorly studied systems will be key to advancing our understanding of

microbial eukaryotes, their interactions with other microbial life, and the roles they play in natural environments.

References

1. Keeling PJ. (2013) Elephants in the room: protists and the importance of morphology and behaviour. Environ Microbiol Rep. 5: 5-6
2. del Campo J, Sieracki ME, Molestina R, Keeling PJ, Massana R, Ruiz-Trillo I. (2014) The others: our biased perspective of eukaryotic genomes. Trends Ecol Evol 29: 252-259
3. Saliba AE, Westermann AJ, Gorski SA, Vogel J. (2014) Single-cell RNA-seq: advances and future challenges. Nucleic Acids Res. 42: 8845-8860
4. Eisen JA, Coyne RS, Wu M, Wu D, Thiagarajan M, *et al.* (2006). Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. PLoS Biol. 4: e286
5. Aury JM, Jaillon O, Duret L, Noel B, Jubin C, *et al.* (2006) Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. Nature. 444: 171-178
6. Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, *et al.* (2014). The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol. 12: e1001889
7. Lynn DH. (2008) The ciliated protozoa. Characterization, classification, and guide to the literature, Third Edition, Springer, Germany
8. Parra G, Bradnam K, Korf I. (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinfo. 23: 1061-1067
9. Gentekaki E, Kolisko M, Boscaro V, Bright KJ, Dini F, *et al.* (2014). Large-scale phylogenomic analysis reveals the phylogenetic position of the problematic taxon *Protocruzia* and unravels the deep phylogenetic affinities of the ciliate lineages. Mol Phylogenet Evol. 78: 36-42

Chapter 9 Supplementary Materials.

Supplementary Text S9.1 – Supplemental Procedures

cDNA amplification and library costruction. Organisms were selected to represent variability found among ciliates (phylum CIliophora). Cell size ranged from about 500 µm (*Condylostoma*) to about 50 µm (*Tetrahymena*); populations were originally sampled in marine (*Condylostoma*), brackish (*Euplotes*, *Paramecium*), or freshwater (*Blepharisma*, *Tetrahymena*) environments; processed organisms were either isolated from mass cultures (*Condylostoma*, fed with the diatom *Phaeodactylum tricornutum*; *Euplotes* and *Paramecium*, fed with the green alga *Dunaliella tertiolecta*; *Tetrahymena*, axenic) or directly collected from natural samples (*Blepharisma*); both subphyla of the phylum *Ciliophora* were covered (*Condylostoma* and *Blepharisma*: Postciliodesmatophora; *Euplotes*, *Paramecium* and *Tetrahymena*: Intramacronucleata). The *Euplotes* cells contained thousands of essential prokaryotic endosymbionts[10], allowing to test the suitability of the approach in presence of obligate intracellular bacteria.

The protocol was performed in two separate lab areas. A biological safety cabinet with equipment for pre-PCR steps was used as a clean work station, and treated with ethanol 70%, UV light and DNAZap™ (Life Technologies) before each experiment. PCR and post-PCR steps were performed in a general lab environment. No materials were moved from the general lab to the clean workstation, with the exception of the tube containing the single ciliate cell in the Reaction Buffer (see below). All the necessary equipment for collecting the cell (dissecting microscope, an independent set of pipettes and pipette tips, etc.) was prepared on a bench, sterilized before and after each experiment, and employed exclusively for this purpose.

Ciliate cells from monoclonal cultures were isolated the day before the experiment, briefly washed and kept overnight in sterile water of appropriate salinity. *Blepharisma* cells from a sample with a complex community were isolated in the same way just before library preparation. One cell was picked in 1 µL of surrounding liquid and washed three times in distilled water through successive transfers in three-depression slides, diluting the original medium and removing contaminant organisms. Marine and brackish-water organisms were washed in

progressively diluted sterile artificial seawater instead, using distilled water only in the last step in order to avoid osmotic lysis. After the last wash, the cell was transferred to a 0.2 mL tube containing 4 µL of distilled water and 5 µL of Reaction Buffer prepared as described in the SMARTer® Ultra Low Input RNA for Illumina® Sequencing Kit - HV user manual (Clontech Laboratories, Inc.). The cDNA synthesis and amplification were performed following the manufacturer's instruction. The Advantage® 2 PCR Kit (Clontech Laboratories, Inc.) was used for PCR amplification, and the Agencourt AMPure XP PCR Purification Kit (Beckman Coulter) was used for DNA purification steps.

The cDNA libraries obtained were quantified with the Qubit® dsDNA HS Assay Kit (Life Technologies). A diluted aliquot of each cDNA library was converted to an Illumina library with the Nextera® XT DNA Sample Preparation kit (Illumina). The resulting libraries were paired-ends sequenced on a MiSeq platform by Génome Québec. The raw sequencing dataset are deposited in NCBI SRA archive under accession numbers SRX727264, SRX727265, and SRX735142-SRX735144.

Data analyses. Each sequencing dataset was quality trimmed using sickle[11] and leftover transposon sequences (from the tagmentation process during NexteraXT library preparation) were removed using an in-house script. The clean datasets were then assembled using "inchworm" from the Trinity package[12]. All reference datasets were downloaded from SRA archive and assembled by inchworm.

To evaluate contamination levels, all five single-cell assemblies were blasted against NCBI-nr database using blastx. A sequence was considered a putative contamination when it was at least 80% identical over at least 50% the length of the query to a sequence from a non-ciliate organism.

Bowtie 2[13] was used for read mapping back to the assembly and package SAMtools[14] was used to extract the number of reads mapped per contig. The amplification bias was evaluated by plotting the number of reads mapped against each contig (Figure 9.1B).

The comparison between single-cell and culture-based datasets was performed in 3 different manners. For *Condylostoma*, where a culture-based transcriptome is available from the same strain, a simple blastn of all-against-all was performed and sequences were considered identical between the two datasets if they were at least 97% identical and overlap by at least

80% length (over one of the paired sequences).

For *Blepharisma*, *Euplotes*, and *Paramecium*, where only congeneric culture-based RNAseq datasets are available (*Blepharisma japonicum*, *Euplotes harpa* and *Paramecium bursaria*), we used a reference genome/transcriptome dataset: *Condylostoma magnum* RNAseq for *Blepharisma*; *Oxytricha trifallax* genome (http://oxy.ciliate.org) for *Euplotes*; and *Paramecium tetraurelia* (http://paramecium.cgm.cnrs-gif.fr/) for *Paramecium*. All transcriptomes were blasted using blastx against NCBI nr (e-value cut-off e-05) and translated into protein sequence according to the best hit and clustered using CD_HIT (98% identity level). Best-hit-reciprocal blastp was used to identify genes from the reference genome/transcriptome that are present in each transcriptome.

In case of *Tetrahymena* a fully sequenced genome is available, representing a near-complete gene set from the same species. Reads from the single-cell transcriptome and from a culture-based transcriptome (by randomly sampling an identical number of reads as were generated by the single-cell transcriptome) were mapped onto predicted gene CDS sequences. For each transcriptome, the number of genes with at least one mapped read was counted.

To evaluate the transcriptome completeness we searched for the 20 tRNA synthetases and a set of 248 eukaryotic core proteins (http://korflab.ucdavis.edu/Datasets/genome_completeness/). Reciprocal blast was used to identify these genes in each of the five transcriptomes.

Additional References

10. Vannini C, Ferrantini F, Ristori A, Verni F, Petroni G. (2012). Betaproteobacterial symbionts of the ciliate *Euplotes*: origin and tangled evolutionary path of an obligate microbial association. Environ Microbiol. 14: 2553-2563
11. Joshi NA, Fass JN. (2011) Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at https://github.com/najoshi/sickle
12. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, *et al.* (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat Biotechnol. 29: 644-652
13. Langmead B, Trapnell C, Pop M, Salzberg SL. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10: R25
14. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, *et al.* (2009). The Sequence Alignment/Map format and

SAMtools. Bioinfo. 25: 2078-2079

Acknowledgments

I rarely enjoy re-reading my previous texts, but I recently looked again at the acknowledgments chapter of my Master Thesis and found it quite good and heart-felt. It's fair to say, it was written at the end of a happy and hopeful period. Nevertheless, I'm uncomfortable now trying to emulate that feeling. I'll do my best.

Starting as tradition requires with working relationship, the first of the list is undoubtedly Claudia Vannini, superior, coworker, supporter, and friend during the last 6 (!!!) years in the protistology unit. From the In *Euplotes* population to the struggle with (against?) qiime, from the gossipy dinners to the Filter snacks, and through a series of *variegated* co-supervised students, she was one of the few stable anchors in my work life. I have proudly said in other occasions that I was able to productively interact and work with many of the people in the unit "old guard", like Letizia Modeo, Sergei Fokin, Simone Gabrielli and Fabrizio Erra, from each of whom I learned something. Carolina Chiellini deserves a special mention for being the first person to show me the lab, and her contagious love for the job. Nowadays the local fauna is significantly more rich and different in composition, and though I had several occasions to appreciate some of the more recent colleagues, the greatest satisfactions from the work performed in the unit in the last three years came from my small crew of students, chiefly Alessia Rossi with her never-ending doctoral-like thesis and Daniela Carducci with the *very* satisfying results on those "long, thin things".

The greatest difference between my doctoral years and my previous career has certainly come from the experiences abroad. I will be forever thankful to Denis Lynn for giving me one of the greatest opportunity of my life, being such a wonderful and kind person, and renovating my hopes in the academic population. He is the single most important reason why I'm still clinging to science as a profession. Thanks to him, I have known my love-at-first-sight, Vancouver, a precious gem of staggering beauty, where I would gladly spend more time in the future. Meeting and working at different times with Patrick Keeling and Michael Lynch truly broadened my mind and guided my scientific reasoning and inquiry in the last years, and taught me what a true scientific brainstorming is. I am grateful to so many people in the Vancouver group that I will only list them without any further comments: Jan Janouscovec, Nick Irwin, Erick James, Martin Kolisko, Javier del Campo, Fabien Burki and several others. As for Bloomington, I have

to thank Tom Doak for his kindness and hospitality, but also for the clever ways he used to boost the results I obtained there.

On the border between life and work, Antonio Casini and Michele Felletti remained my best friends and daily (not entirely for their choice) support despite the hundreds of kilometers that separate us (except when they don't). Having my most important paper coauthored by Michele was the greatest pleasure of my PhD years, and that is saying something, since the same paper also cost me dearly. I am thankful for all the Bio-Trips (in Italian sounds better), and for each and every weekend Ant subtracted from his hectic work routine to visit me. Thanks to Filippo Lipparini for the culinary breaks and Lorenzo Romeo that hoped to make "a better man" out of me just by becoming my roommate. To Beatrice Pozzetti and our thousand farewells and shared mood swings. A last salute also to the now *completely* scattered remnants of my university years within and outside the Pack: Fabio Miazzi, Nicola Pierazzo (!), Luca Battistella, Giulio Belletti, Umberto Binetti, Beatrice Moretti, Franco Egidi (the most unlikely gym pal). A special thank to the various generations of the Board Game Group, and especially to Alessandra Caraceni, who started it, and Giovanni Paolini, the only member that remained through all the seasons (and all the games). And, as always, thanks to those that have been there for 15 – or 25 – years: Totta, Tia, Teo (I'm coming closer!) and Luca.

A final thank to mom, dad and Valeria for all the support and the travels to and from Tuscany. I'm going to a more interesting, but slightly less easy to reach, place.

I'm beginning to grow tired of Pisa and almost happy to leave it, now… after everybody else. But I loved it for more than 8 years, and I am still able to appreciate its merits (anyone can list the demerits, come on! That's easy). But the best part of it was gone even before starting the PhD. I miss college life, I miss Faedo and Carducci, I miss study halls and laundry room, I miss the courtyard and the community… even the cafeteria! I can't say that I didn't realize what I had at the time – I did – but I now have a long list of new reasons to miss everything.

Many years ago, closing the acknowledgment section of a very different volume, I thanked both friends and foes considering them essential parts of life and growth. I feel my heart was in the right place, at that time. But what's true at 18 is not necessarily so at 27. In the research

world one can be unlucky enough to meet certain types of people: the oppressor, the narrow-minded, the parochial, the arrogant, the incompetent, up to the worst of all: the weak-willed. Not necessarily separated in different bodies. When any of these unsavory characters occupy the wrong position, they poison the joy that a younger person should feel when doing the best work in the world: the one he chose to do. They seldom contribute to anybody else's growth, in any sense of the word, and create an environment that selects only stubbornness and defiance (or lickspittles) – NOT the best proxies for talent. Thus, I today praise only those that, due to bravery and sense of justice, oppose them with good deeds and true hearts. And I can only hope – without any certainty – that I won't be added to the long list of their fallen victims.

APPENDIX

Complete list of publications and contributions (2012-2014)

Articles on peer-reviewed journals:

1. Boscaro V, Vannini C, Fokin SI, Verni F, Petroni G. (2012) Characterization of "*Candidatus* Nebulobacter yamunensis" from the cytoplasm of *Euplotes aediculatus* (Ciliophora, Spirotrichea) and emended description of the family *Francisellaceae*. Syst Appl Microbiol. 35: 432-440

2. Boscaro V, Fokin SI, Verni F, Petroni G. (2012) Survey of *Paramecium duboscqui* using three markers and assessment of the molecular variability in the genus *Paramecium*. Mol Phylogenet Evol. 65: 1004-1013

3. Boscaro V, Fokin SI, Schrallhammer M, Schweikert M, Petroni G. (2013) Revised systematics of *Holospora*-like bacteria and characterization of "*Candidatus* Gortzia infectiva", a novel macronuclear symbiont of *Paramecium jenningsi*. Microb Ecol. 65: 255-267

4. Boscaro V, Petroni G, Ristori A, Verni F, Vannini C (2013) "*Candidatus* Defluviella procrastinata" and "*Candidatus* Cyrtobacter zanobii", two novel ciliate endosymbionts belonging to the "*Midichloria* clade". Microb Ecol. 65: 302-310

5. Modeo L, Fokin SI, Boscaro V, Andreoli I, Ferrantini F, Rosati G, Verni F, Petroni G. (2013) Morphology, ultrastructure, and molecular phylogeny of the ciliate *Sonderia vorax* with insights into the systematics of order Plagiopylida. BMC Microbiol. 13: 40

6. Boscaro V, Felletti M, Vannini C, Ackerman MS, Chain PSG, Malfatti S, Vergez LM, Shin M, Doak TG, Lynch M, Petroni G. (2013) *Polynucleobacter necessarius*, a model for genome reduction in both free-living and symbiotic bacteria. Proc Natl Acad Sci U S A. 110: 18590-18595

7. Boscaro V, Schrallhammer M, Benken KA, Krenek S, Szokoli F, Berendonk TU, Schweikert M, Verni F, Sabaneyeva EV, Petroni G. (2013) Rediscovering the genus *Lyticum*, multiflagellated symbionts of the order *Rickettsiales*. Sci Rep. 3: 3305

8. Vannini C, Boscaro V, Ferrantini F, Benken KA, Mironov TI, Schweikert M, Görtz H-D, Fokin SI, Sabaneyeva EV, Petroni G. (2014) Flagellar movement in two bacteria of the family *Rickettsiaceae*: a re-evaluation of motility in an evolutionary perspective. PLoS One. 9: e87718

9. Gentekaki E, Kolisko M, Boscaro V, Bright KJ, Dini F, Di Giuseppe G, Gong Y, Miceli C, Modeo L, Molestina RE, Petroni G, Pucciarelli S, Roger AJ, Strom SL, Lynn DH. (2014) Large-scale phylogenomic analysis reveals the phylogenetic position of the problematic taxon *Protocruzia* and unravels the deep phylogenetic affinities of the ciliate lineages. Mol Phylogenet Evol. 78: 36-42

10. Boscaro V, Carducci D, Barbieri G, Senra MVX, Andreoli I, Erra F, Petroni G, Verni F, Fokin SI. (2014) Focusing on genera to improve species identification: revised systematics of the ciliate *Spirostomum*. Protist. 165: 527-541

11. Kolisko M, Boscaro V, Lynn DH, Keeling PJ. (2014) Single-cell transcriptomics for microbial eukaryotes. Curr Biol. 24: 1081-1082

Oral presentations performed at congresses and meetings:

1. Boscaro V, Moretti B, Fokin SI, Verni F, Petroni G, Vannini C. Diversity and taxonomy of the prokaryotic community harbored by the population In of *Euplotes aediculatus* (Alveolata, Ciliophora). 7th International Symbiosis Society Congress "The earth's vast symbiosphere". July 22-28 2012, Krakow, Poland

2. Boscaro V, Moretti B, Fokin SI, Verni F, Petroni G, Vannini C. The population In of *Euplotes aediculatus* harbors the most taxonomically diverse endosymbiotic community characterized so far in ciliates. IRSES CINAR-Pathobacter Meeting. August 5-8 2012, Stuttgart, Germany

3. Boscaro V. Multi-marker molecular characterization of ciliates and their endosymbionts. XXIX Italian Society of Protistology National Congress. October 1-4 2012, Roma, Italy

4. Boscaro V, Felletti M, Vannini C, Ackerman MS, Hahn MW, Doak TG, Lynch M, Petroni G. *Polynucleobacter necessarius*, a new model for genome reduction in both free-living and symbiotic bacteria. COST Action BM1102, CNRS GDRE & IRSES CINAR Pathobacter Joint Meeting. May 12-16 2013, Tallinn, Estonia

5. Boscaro V, Felletti M, Vannini C, Ackerman MS, Hahn MW, Doak TG, Lynch M, Petroni G. Functional analysis and evolution of the genome of *Polynucleobacter necessarius*, obligate symbiont of the ciliate *Euplotes aediculatus*. XIV International Congress of Protistology. July 28 - August 2 2013, Vancouver, Canada

6. Boscaro V, Kolisko M, Keeling PJ, Lynn DH. Single-cell RNAseq on ciliates: from sampling to complete transcriptomes in a few weeks. XXX Italian Society of Protistology National Congress "Basic and applied research on protists". October 2-4 2014, Padova, Italy

7. Boscaro V, Kolisko M, Keeling PJ, Lynn DH. Suitability of the single-cell RNAseq approach for ciliates. GDRE, BMBS COST Action BM1102 and IRSES CINAR PATHOBACTER Joint Meeting "Frontiers in ciliate genome evolution, adaptation, and symbiosis". October 6-8 2014, Pisa, Italy

Co-authored congress and meeting abstracts:

1. Fokin SI, Boscaro V. *Spirostomum* spp. and its bacterial symbionts. Biodiversity and perspectives. IRSES CINAR-Pathobacter Meeting. August 5-8 2012, Stuttgart, Germany

2. Sabaneyeva E, Mironov T, Benken K, Schweikert M, Boscaro V, Fokin SI. The intranuclear motile bacterium in *Paramecium multimicronucleatum*: an old acquaintance requiring reinvestigation. IRSES CINAR-Pathobacter Meeting. August 5-8 2012, Stuttgart, Germany

3. Benken KA, Boscaro V, Schrallhammer M, Schweikert M, Sabaneyeva E, Petroni G. Molecular characterization of the symbiont of *Paramecium octaurelia* 299 lambda. IRSES CINAR-Pathobacter Meeting. August 5-8 2012, Stuttgart, Germany

4. Modeo L, Dias RJP, Fernandes N, Tomei G, Boscaro V, Verni F, da Silva-Neto ID, Petroni G. Phylogenetic systematics of the genus *Condylostoma* (Ciliophora, Heterotrichea) by means of a multidisciplinary analytical study of some marine morphospecies. XXIX Italian Society of Protistology National Congress. October 1-4 2012, Roma, Italy

5. Senra MVX, Dias RJP, Boscaro V, da Silva-Neto ID, Petroni G, Soares CAG. Advances in the *Teredinibacter turnerae*/shipworms/*Boveria* interaction: an inter-kingdom tripartite interaction. XXIX Italian Society of Protistology National Congress. October 1-4 2012, Roma, Italy

6. Schrallhammer M, Boscaro V, Benken KA, Krenek S, Szokoli F, Berendonk TU, Schweikert M, Verni F, Sabaneyeva E, Petroni G. Rediscovering the genus *Lyticum*, multiflagellated symbionts of the order *Rickettsiales*. COST Action BM1102, CNRS GDRE & IRSES CINAR Pathobacter Joint Meeting. May 12-16 2013, Tallinn, Estonia

7. Barbieri G, Rossi A, Carducci D, Boscaro V, Senra MVX, Fokin SI. Biodiversity of *Spirostomum* species and its relationships with bacterial endocytobionts. COST Action BM1102, CNRS GDRE & IRSES CINAR Pathobacter Joint Meeting. May 12-16 2013, Tallinn, Estonia

8. Mironov T, Yashenko V, Benken KA, Boscaro V, Fokin SI, Schweikert M, Sabaneyeva E. A triple symbiotic system in *Paramecium multimicronucleatum*. COST Action BM1102, CNRS GDRE & IRSES CINAR Pathobacter Joint Meeting. May 12-16 2013, Tallinn, Estonia [POSTER]

9. Rossi A, Boscaro V, Carducci D, Serra V, Modeo L, Verni F, Fokin SI, Petroni G. Biodiversità delle comunità di ciliati di ambienti dulciacquicoli naturali e antropizzati della provincia di Pistoia: uno studio multidisciplinare. 74° Italian Zoological Union National Congress. September 30 - October 3 2013, Modena, Italy [in Italian] [POSTER]

10. Carducci D, Barbieri G, Boscaro V, Lebedeva NA, Rossi A, Senra M, Verni F, Petroni G, Fokin SI. Caratterizzazione morfologica e molecolare di 19 popolazioni di protisti ciliati appartenenti al genere *Spirostomum* Ehrenberg, 1838. 74° Italian Zoological Union National Congress. September 30 - October 3 2013, Modena, Italy [in Italian] [POSTER]

11. Kolisko M, Boscaro V, Lynn D, Keeling PJ. Single-cell RNAseq from microbial eukaryotes: Test and application of the protocol. CIFAR IMB Meeting. July 2014, Prague, Czech Republic [POSTER]

12. Gentekaki E, Kolisko M, Boscaro V, Bright KJ, Dini F, Di Giuseppe G, Gong Y, Miceli C, Modeo L, Molestina RE, Petroni G, Pucciarelli S, Roger AJ, Strom SL, Lynn DH. Large-scale phylogenomic analysis reveals the phylogenetic position of the problematic taxon *Protocruzia* and unravels the deep phylogenetic affinities of the ciliate lineages. Protist 2014. August 3-8 2014, Banff, Canada [POSTER]

13. Kolisko M, Boscaro V, Lynn DH, Keeling PJ. Single cell transcriptomics as a way to study uncultivable protists – case study on oxymonads. Protist 2014. August 3-8 2014, Banff, Canada [POSTER]

14. Boscaro V, Sigona C, Petroni G, Vannini C. Symbiosis evolution: retracing the tangled path from establishment to obligate association in the *Betaproteobacteria-Euplotes* relationship. XIII Italian Federation on Life Sciences Congress "FISV 2014". September 24-27 2014, Pisa, Italy